Private vs. Pooled Transportation: Customer Preference and Congestion Management

Kashish Arora, Fanyin Zheng, and Karan Girotra

1 Introduction

Traffic congestion and its economic and social consequences plague most large urban areas. On-demand taxi services have done little to alleviate these concerns, and, on the contrary, have exacerbated congestion in some of the biggest cities. In recent years, *pooled transportation* has emerged as a cheaper and more environmentally and traffic friendly alternative to on-demand transportation services. Examples of such pooled transportation options range from Uber's UberPool service to the *shuttle services* offered by Via, Chariot, and others in major cities across countries. At the same time, most of these services face challenges as, many customers remain reluctant to switch to pooled services from private on-demand transportation options. The goal of this paper is, therefore, to understand customer preferences in choosing between private and pooled transportation services and to investigate the way that policies can be designed to incentivize customers to use pooled transportation and effectively manage congestion.

From the government's perspective, the most often used policies are congestion surcharges. Local governments in large cities such as New York, Singapore, and London have imposed congestion prices to incentivize increased usage of pooled transportation and less usage of the private rides. For instance, the city of London uses an *all-day* congestion surcharge policy, whereas Singapore uses a *peak hour* surcharge policy. A key challenge for many cities is to assess the potential impact and the relative effectiveness of these policies before they are implemented. Experi-

K. Arora (✉) · K. Girotra
Cornell University, Ithaca, NY, USA
e-mail: ka522@cornell.edu; girotra@cornell.edu

F. Zheng
Columbia University, New York, NY, USA
e-mail: fanyin.zheng@columbia.edu

H. Yang et al. (eds.), *AI and Analytics for Public Health*, Springer Proceedings in Business and Economics, https://doi.org/10.1007/978-3-030-75166-1_2

menting with different policies in practice is extremely expensive in this setting, and, hence, their implementation has been largely on an ad hoc basis. Our paper seeks to provide guidance to evaluate the performance of different policies in incentivizing customers to switch from private to pooled transportation. Specifically, our guidance to policy makers is in terms of two aspects. First, we evaluate the efficacy of three particular types of congestion strategies: (i) *price strategies* based on congestion surcharges applied in the city; (ii) *price strategies* based on providing discounts to pooled transportation services; and (iii) *operational strategies* based on improving the service features of pooled transportation services. Second, our findings highlight the importance of incorporating customer heterogeneity of preferences in designing effective policies to promote pooled transportation usage.

Our analysis proceeds in two steps. First, we estimate customer demand for pooled and private transportation with usage data from *Ola Cabs* in India. The form of pooled transportation we study is the shuttle service, and the private transportation we focus on is cabs. We adopt a structural modeling approach to estimate demand and recover the customer's preferences for price and other operational service features. Then, we evaluate various congestion management strategies using the estimated model via counterfactual analyses. We provide a brief description of each of the two steps and summarize our main findings next.The results from this analysis suggest a substantial degree of substitution between the two services. We then build a structural model to estimate customer preferences over prices and different service features for the cab and shuttle services. Our control function approach with Lasso selected instruments corrects for the bias in the customer price elasticity and wait time sensitivity estimates. The estimated price coefficient is lower than the estimate without the correction, suggesting the presence of route based discounting by the platform to grow the customer base. We estimate an average cost of ₹98.5 ($1.3) for walking an additional km to the shuttle stop ₹3.6 for traveling ten extra minutes on the shuttle. Using the estimates from the model, we provide prescriptive recommendations on reducing congestion through counterfactuals. In the first counterfactual, we apply differential percentage *congestion surcharges* to cab and shuttle services in a congestion zone of the city. We find that a 20% congestion surcharge on cabs achieves a 15.0% overall vehicle reduction on the road. The corresponding vehicle reduction due to customers substituting from cab to shuttle service is around 4.04%. We also apply congestion surcharges to the services in peak hours following similar policies implemented in Singapore. We find that, interestingly, surcharges applied to the evening rush hours achieve around 3 times the %age vehicle reduction as compared to the morning rush hours. Moreover, the evening rush hour surcharge policy achieves a higher %age vehicle reduction than an all-day surcharge policy. In the second counterfactual, we evaluate the impact of providing discounts to shuttle rides on customer choices. We find that a 20% discount on shuttle rides leads to around 1% reduction in vehicles due to customers substituting from cabs to the shuttle service. Moreover, the reduction disproportionately comes from new users with relatively low past shuttle usage and more room for usage growth in the future. Finally, we evaluate *operations based strategies* and estimate the change in congestion levels when the firm improves some of the key shuttle service features. We find that a city could reap a significant portion of the benefits obtained by applying congestion surcharge policies by utilizing the operational levers itself. More importantly, adopting these strategies avoids the deadweight losses associated with the price surcharges due to its tax nature. Specifically, we find that a 20% decrease in customers' walking distance to shuttle pick up stops can achieve 35% of the total effect achieved by the congestion surcharge in terms of the number of customers substituting from the cab to shuttle services. Similarly, a 20% decrease in the shuttle travel time can achieve 6.9% of the total substitution achieved by the congestion surcharge. This result shows that improving the service features of pooled ride services is an important alternative to price-based policies in managing congestion in big cities.

Our paper is closely related to the structural demand estimation of mobility services in operations management and economics. He et al. (2019) and Kabra et al. (2019) study customer demand in bike-share systems. Buchholz (2020) and Ata et al. (2019) study spatial demand for the taxi service. To the best of our knowledge, our work is the first empirical study to investigate customer choices between the on-demand and the pooled ride services, which allows us to quantify the impact of congestion policies on customer choices. Our paper also closely relates to the empirical literature on ride-sharing services in operations management. Cohen et al. (2020b) run field experiments to nudge commuters to carpool using in-app notifications. Also, Cohen et al. (2020a) document the frustrations caused by inconveniences such as longer travel time in pooled services. The main difference between these studies and our paper is that, instead of an experimental approach, we recover customer preferences for choosing the services while directly incorporating the inconveniences associated with the shuttle service in the model.

Our work also contributes to the literature on congestion management in transportation. Han et al. (2019) build a stochastic model to develop a road pricing scheme to curb congestion. Recent work by Ostrovsky and Schwarz (2018) studies the relationship of carpooling, road pricing, and autonomous transportation. The authors highlight the role of road pricing in the adoption of pooled transport. Almost all of these studies are either analytical or predictive, even though the topic is of high practical relevance to both policy makers and ride-sharing platforms. Our work complements the literature using data and empirical methods to estimate customer preferences and provide prescriptive recommendations for the design of congestion policies.

Our work also fits into the growing literature on structural estimation in operations management. We use a discrete choice model with a control function to estimate the customer preference parameters for different ride service features. Similar methods have been applied in Petrin and Train (2010) and Guajardo et al. (2012). To correct for endogeneity, we build on the network type of instrumental variable method used in prior work on demand estimation in operations management (He et al., 2019). To strengthen the relevance of our instruments, we employ selection methods commonly used in machine learning and the causal inference literature (Belloni et al. 2011, 2012).

2 Data

Our study uses data provided to us by the Indian ride-hailing company Ola Cabs.The firm competes directly with (i) other, similar platforms such as Uber; (ii) public transportation; and (iii) city taxis. Its market share in the Indian ride-hailing market was around 65% at the time we collected our data. financing. We use four different sources of data in our analysis: (i) cab rides data; (ii) shuttle rides and trips data; (iii) Google Places API data; and (iv) census data

The cab rides dataset contains over 25 million cab rides from Jan-Aug 2016 in Delhi. Each ride record contains information about (i) the customer's pickup and drop-off locations (latitudes and longitudes); (ii) anonymized ID; (ii) timestamps for the initial ride request, pickup, and drop-off; (iii) prices; and (iv) distance traveled. There are around 3.5 million unique users on the platform. The shuttle data contain information about customer rides and shuttle trips. About 76K unique users took 1.28 million rides in Delhi from Jan-Aug 2016. For each ride record, we observe (i) timestamps for the customer's initial booking request; (ii) anonymized ID; (iii) latitudes and longitudes for pickup and drop-off; (iv) prices; (v) distance traveled; and (vi) latitude and longitude of the customer's mobile device when she makes the initial booking request. We also obtained latitudes and longitudes of around 176K points of interest in the city, collected from Google Places API. The places are classified into 93 classes, including restaurants, museums, libraries, hospitals, and theaters. In addition, we obtain demographic information from the Indian census of 2011.

2.1 Descriptive Evidence

To motivate our main model, we document two sets of descriptive evidence to motivate our main analysis. First, we show that a large number of customers use both the on-demand cab service and the shuttle service, and, hence, the market is not segmented into cab-only and shuttle-only users. Moreover, there is a large heterogeneity in customers' preferences for choosing shuttles and cabs. This motivates the structure of our model in Sect. 3. We control for this heterogeneity by including variables that measure the customer's past usage metrics on the platform.

Second, using a difference-in-differences analysis, we estimate the degree of substitution between the two services, thereby establishing that the services are substitutes. We leverage the fact that the shuttle platform was adding routes as it was expanding over time. The addition of routes over time serves as a quasi-experiment for our difference-in-differences analysis. We run a two-way-fixed-effects (TWFE) model to identify the causal impact of opening up the shuttle route on cab ridership. There is a net reduction of 88 rides per route when the shuttle services are operating. The reduction in cab ridership shows directly that the two services act as substitutes. In summary, we show that customers choose between shuttle and cab services, and we provide evidence of substitution between the two services. In the next section, we outline a richer customer-level choice model that helps us understand customer preferences when they are making choices between the two services.

3 Choice Model

Customer's Utility Customer i, who wants to travel from origin location $j \in \{1, \ldots, L\}$ to destination location $k \in \{1, \ldots, L\}$, chooses to take one of the alternatives, $a \in \{\text{Shuttle}, \text{Cab}\}$ or an outside option. The customers who travel from j to k belong to one market. The utility that customer i traveling in market jk gets from choosing an alternative a at time t is:

$$U_{ijkta} = \alpha_a + \sum_r \beta_r p_{jka} d_{ir} + \gamma w_{jc} + \sum_r X^{1'}_{jka} \Theta_r d_{ir} + X^{2'}_{jk} \Delta$$
$$+ Q^{'}_{iw(t)} \Omega + \mathrm{T}_{as(t)} + \xi_{jk} + \xi_i + \epsilon_{ijkta}. \quad (1)$$

In Eq. (1), α_a represents the baseline preference for the shuttle and the cab service. p_{jka} denotes the average price for service a in market jk. w_{jc} denotes the average wait time for the cab service. We do not include the wait time for the shuttle in our model since the majority of rides arrive at the scheduled time. We allow for the price coefficient to vary across two groups of customers segmented by their total shuttle usage in the past. d_{ir} is an indicator that identifies whether customer i belongs to the low-usage or new users group ($r = 1$) or the high-usage or experienced users group ($r = 2$). We include four sets of service features and control variables in our model as follows:

Market-Alternative-Level Service Features X^1_{jka} is a vector of key service features other than prices and wait time that affect the customer's choice. First, it includes the time and distance traveled on the ride across the two services. For any origin-destination pair jk, the time and distance traveled in shuttles is larger than those in cabs. The extra time and distance traveled and walking are the inconveniences associated with the shuttle service. Like price, we also allow for the vector of sensitivities to the service features to vary across the two customer groups.

Market-Level Controls X^2_{jk} is the vector of market-level control variables. We include a rich set of market variables in the model. Specifically, we include the number of Google Places category counts at both j and k (20 in total), and demographic information such as population densities and working population at both the locations (eight in total).

Customer-Time-Level Controls $Q_{iw(t)}$ is the vector of the time-varying usage history of a customer on the two platforms. $w(t) \in \{1, \ldots, T\}$ is an operator that denotes the number of weeks starting from January 1, 2016. $Q_{iw(t)}$ includes two variables: (i) the cumulative number of shuttle rides taken by the customer up to last week, and (ii) the number of recent rides taken by the customer across the two services. The first usage variable controls for customers' familiarity with the experience of the shuttle service. Since the shuttle service was newly launched during the period of our data, new shuttle customers may not have been fully aware of the experience of using the shuttle service and, hence, may have been less likely to choose that service. Thus, it is important to control for this variable. The second variable captures the effect of the customer's recent activity on the platform. A customer may be more likely to choose the service if she was recently active on the platform. Both variables control for customer heterogeneity in terms of their awareness of the shuttle service.

Time-Level Controls $T_{as(t)}$ are the time fixed effects, and $s(t)$ is an operator that denotes the time slot of the day (morning, evening, night, etc).

Unobservables ξ_{jk} are the market-level unobservables that affect demand. Examples of some of these factors are the unobserved popularity of the route, the level of congestion on the route. ξ_i are the customer-level unobservables that affect demand. Examples of these unobservables include income, age, and any other factors that affect a customer's preference for choosing between the two services. ϵ_{ijkta} are independent and identically distributed idiosyncratic errors that follow extreme value type 1 distribution. Apart from the shuttles and cabs, the customer can also choose to take an outside option. The utility of the outside option o is defined as:

$$U_{ijkto} = u_o + \epsilon_{ijkto}. \tag{2}$$

where u_o is normalized to be zero. The customer chooses the alternative that maximizes her utility. The choice probability of customer i is given by :

$$P_{ijkta} = \frac{\begin{bmatrix} exp(\alpha_a + \sum_r \beta_r p_{jka} d_{ir} + \gamma w_{jc} + \sum_r X^{1'}_{jka}\Theta_r d_{ir} + X^{2'}_{jk}\Delta \\ +Q^{'}_{iw(t)}\Omega + \mathrm{T}_{as(t)} + \xi_{jk} + \xi_i) \end{bmatrix}}{\begin{bmatrix} 1 + \Sigma_{a\in\{c,s\}} exp(\alpha_a + \sum_r \beta_r p_{jka} d_{ir} + \gamma w_{jc} + \sum_r X^{1'}_{jka}\Theta_r d_{ir} \\ +X^{2'}_{jk}\Delta + Q^{'}_{iw(t)}\Omega + \mathrm{T}_{as(t)} + \xi_{jk} + \xi_i). \end{bmatrix}} \tag{3}$$

Our goal is to estimate the unknown scalars (β_r, γ) and vectors $(\Theta_r, \Delta, \Omega)$ in the model.

4 Estimation

4.1 Endogeneity

As in many discrete choice demand estimation settings, some of the observed product attributes, such as price, are often correlated with unobserved product characteristics such as quality and, hence, are endogenously determined. Specifically, a firm that tries to optimize profits or growth adjusts prices for products and services based on features that are observable to itself but not to the researcher. In our setting, routes may be priced by the platform managers based on their popularity. If the firm raises prices on the popular routes to optimize profits, without taking this into account in the estimation, the price coefficient obtained from the model will be underestimated. However, if the firm cuts prices on popular routes to grow the customer base, the price coefficient obtained from the model would be overestimated. In either case, using the biased estimate would lead to unreasonable prescriptive policy recommendations and managerial insights in the counterfactual analyses. Interestingly, in our setting, the shuttle service was in a phase of growth and expansion, whereas, by comparison, the cab service was in a more mature phase. This makes for an interesting setting in which to study the price endogeneity problem. Prices of the two services are not the only endogenous variables in our setting. Wait times for cabs are correlated with the unobserved popularity or congestion level of the route and are determined endogenously. When wait times increase with unobserved popularity or congestion, the coefficient for wait time would be underestimated if we did not take into account the endogeneity problem in the estimation. Moreover, past shuttle usage is correlated with unobserved customer-level characteristics. These unobservables lead to an endogenous selection of users into the low- and high-usage groups. Without considering this endogeneity issue, the coefficient for past usage will be overestimated in magnitude.

4.2 Instruments

Our approach to correct for the biases discussed in the previous section is to find valid instruments for the endogenous variables. We construct instrumental variables for the prices and cab wait time by utilizing the network structure of our data. Then, we select from a large set of valid instruments, the best set, following recent developments in the intersection of the causal inference and machine learning literatures. Second, we construct the instrumental variable for past shuttle usage by utilizing the timing of the introduction of shuttle routes.

Network Instruments The first step here is to define a set of network-based instruments following He et al. (2019). To explain the variation in prices for route jk, we look for exogenous characteristics of h that affect the popularity of j and k. We construct instrumental variables separately for origin j and destination k. Our proposed valid instruments are averages of the exogenous characteristics of all such feasible h that satisfy the relevance and the exclusion restriction criteria.

Lasso Instrument Selection Since the set of valid network instruments is large (1236), selecting the right ones is not trivial. Instead of handpicking some instruments or naively selecting the compete set, we use machine selection methods. We follow the method in Belloni et al. (2012) to select the best set of instruments in our setting. The method involves using a penalty function that penalizes the addition of more instruments to avoid a weak first stage in the IV estimation. The penalization is achieved through the square root Lasso estimator.

Shuttle Route Introduction Based Instrument for Shuttle Usage The instrument should provide an exogenous variation in customer i's past shuttle usage to recover the causal parameter of interest. We construct the instrument by utilizing the timing of the introduction of different shuttle routes in our data. Specifically, to explain the variation in customers' shuttle usage, we utilize the time of introduction of a route od_i attached to the customer i. Let T_{od_i} be the timing of the introduction of route od_i in the sample period and D_{od_i} the number of days from Jan 1 till T_{od_i}. Then, $\mathcal{Z}_i = 244 - D_{od_i}$ is a valid instrument for shuttle usage (Fig. 1).

The rationale for using the above instrument is as follows. The timing of opening up of shuttle operations on customer i's home route gives an exogenous shock to her shuttle usage. Thus, the length of time that the route od_i is active in the sample period affects customer i's shuttle usage. This is the relevance condition for the instrument. Also, the timing of the introduction of shuttle operations on a route is unlikely to be correlated with the unobserved customer characteristics ξ_i conditional on the market characteristics. This gives us the exclusion restriction condition.

4.3 *Control Function Approach to Estimation*

We use a control function approach with the instruments described in Sect. 4.2 to correct for the endogeneity in prices, wait time and past shuttle usage variables, following the method in Petrin and Train (2010). The control functions for the prices

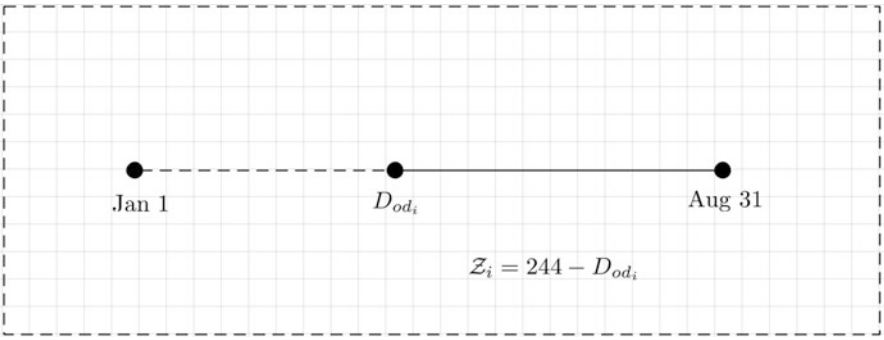

Fig. 1 Instrument for past shuttle usage of a customer. D_{od_i} represents the number of days from the start of the sample period to the introduction of home route od_i. The length of the solid line in days is the magnitude of the instrument

and wait time are specified below in the system of equations:

$$
\begin{aligned}
p_{jks} &= \alpha_{sp} + \kappa_{sp} Z_{sp}^{LASSO} + X_{jks}^{1'} \Lambda_{sp} + X_{jk}^{2'} \Phi_{sp} + \nu_{jks}^{p} \\
p_{jkc} &= \alpha_{sc} + \kappa_{cp} Z_{cp}^{LASSO} + X_{jkc}^{1'} \Lambda_{sp} + X_{jk}^{2'} \Phi_{cp} + \nu_{jkc}^{p} \\
w_{jc} &= \alpha_{wc} + \kappa_{cw} Z_{cw}^{LASSO} + X_{jkc}^{1'} \Lambda_{sp} + X_{jk}^{2'} \Phi_{cw} + \mu_{jc}^{w}.
\end{aligned}
\tag{4}
$$

where, p_{jks}, p_{jkc} are average route-level prices for shuttle and cabs, and w_{jc} is the average cab wait time. $(Z_{sp}^{LASSO}, Z_{cp}^{LASSO}, Z_{cw}^{LASSO})$ are the Lasso selected sets of instrumental variables for the endogenous variables. X_{jk}^{2} is the same vector of exogenous market characteristics used in Eq. (1). Similarly, X_{jks}^{1} and X_{jkc}^{1} are the vectors of exogenous market-alternative level controls used in Eq. (1). $(\Lambda_{sp}, \Lambda_{cp}, \Lambda_{cw})$ and $(\Phi_{sp}, \Phi_{cp}, \Phi_{cw})$ are the associated coefficients. Using the control functions ν_{jkc}^{p}, ν_{jkc}^{p} , μ_{jc}^{p} and μ_{i}^{us} obtained from Eq. (4), the customer utility can be written as:

$$
\begin{aligned}
U_{ijkta} = {} & \alpha_a + \sum_r \beta_r p_{jka} d_{ir} + \gamma w_{jc} + \sum_r X_{jka}^{1'} \Theta_r d_{ir} + X_{jk}^{2'} \Delta + Q_{iw(t)}^{'} \Omega \\
& + \mathrm{T}_{as(t)} + \lambda_1 \nu_{jkc}^{p} + \lambda_2 \nu_{jkc}^{p} + \lambda_3 \mu_{jc}^{p} + \lambda_4 \mu_{i}^{us} + \epsilon_{ijkta}.
\end{aligned}
\tag{5}
$$

where λ_1, λ_2, λ_3 and λ_4 are the coefficients for the control functions. Then the customer's choice probability can be calculated as :

$$
\begin{aligned}
& P_{ijkta}(\beta_r, \gamma, \Theta_r, \Delta, \Omega, \lambda_1, \lambda_2, \lambda_3, \lambda_4) \\
& \quad = \frac{exp(V_{ijkta} + \lambda_1 \nu_{jkc}^{p} + \lambda_2 \nu_{jkc}^{p} + \lambda_3 \mu_{jc}^{p} + \lambda_4 \mu_{i}^{us})}{1 + \Sigma_{a \in \{c,s\}} exp(V_{ijkta} + \lambda_1 \nu_{jkc}^{p} + \lambda_2 \nu_{jkc}^{p} + \lambda_3 \mu_{jc}^{p} + \lambda_4 \mu_{i}^{us}).}
\end{aligned}
\tag{6}
$$

We recover the parameters $(\beta_r, \gamma, \Theta_r, \Delta, \Omega, \lambda_1, \lambda_2, \lambda_3, \lambda_4)$ by maximum likelihood estimation. The log likelihood is computed over all the choices observed in the data.

$$
(\hat{\beta}_r, \hat{\gamma}, \hat{\Theta}_r, \hat{\Delta}, \hat{\Omega}, \hat{\lambda}_1, \hat{\lambda}_2, \hat{\lambda}_3, \hat{\lambda}_4) = \arg\max \mathcal{L}(((\beta_r, \gamma, \Theta_r, \Delta, \Omega, \lambda_1, \lambda_2, \lambda_3, \lambda_4)).
\tag{7}
$$

4.4 Results from the Choice Model

The estimated parameters from the choice model (second stage) are presented in Table 1. First, we note that the price coefficients β_r in the specification without using the IVs are −0.017 and −0.013 for groups 1 and 2, respectively. After using the control function approach, we recover β_r = −0.014 and −0.009. Hence, the

Table 1 Parameter estimates from the choice model. .s and .c are shuttle- and cab-specific coefficients. .1 and .2 are coefficients for the two customer usage groups (*,**,*** indicates statistical significance at 10%,5%,1% level)

Explanatory variable	Without IV	With IV
Shuttle intercept	−13.922*** (0.082)	−13.942*** (0.082)
Cab intercept	−12.887*** (0.080)	−13.032*** (0.080)
Price Paid.1	−0.017*** (0.0001)	−0.014*** (0.0001)
Price Paid.2	−0.013*** (0.0001)	−0.009*** (0.0001)
Wait	−0.129*** (0.001)	−0.135*** (0.001)
Time	−0.002*** (0.0001)	−0.004*** (0.0001)
Commute.1	−1.516*** (0.006)	−1.536*** (0.006)
Commute.2	−0.830*** (0.006)	−0.807*** (0.006)
Distance.1	−0.192*** (0.002)	−0.161*** (0.002)
Distance.2	−0.106*** (0.003)	−0.066*** (0.003)
Controlfunction shuttle price		−0.026*** (0.0002)
Controlfunction cab price		0.002*** (0.0002)
Controlfunction cab wait		0.069*** (0.005)
Morning.s	3.857*** (0.009)	3.879*** (0.009)
Morning.c	−13.754*** (0.082)	−13.766*** (0.081)
Cumulative Shuttle Usage.s	0.091*** (0.0002)	0.087*** (0.0002)
Controlfunction Usage.s		0.004*** (0.0001)
Cumulative Shuttle Usage.c	−3.923 (163.312)	−4.051 (165.234)
Controlfunction Usage.c		0.602 (1.321)
Recent Week Rides.s	0.948*** (0.002)	0.955*** (0.002)
Recent Week Rides.c	0.691*** (0.003)	0.680*** (0.003)
Market controls	Yes	Yes
Observations	1,323,413	1,323,413
McFadden R^2	0.580	0.585

price coefficients are adjusted down in magnitude—i.e., the price coefficient is overestimated without using the IVs. The control function for the shuttle price is significant and negative. The strong negative control function for the shuttle price indicates that the average price of the shuttle in a market is lower than what the observed market characteristics can explain. The control function for cab price is positive and much weaker in magnitude than that for the shuttle price. The positive sign indicates that the average price of cabs on a route is higher than as explained by the observed attributes. This is consistent with our discussion on the endogeneity of price in Sect. 4.1. Moreover, we find that there is substantial heterogeneity in the price coefficients across the two groups. The new users (group 1) are about 1.55 times more price-sensitive than experienced users (group 2). The coefficient for wait time γ in the model without IV is -0.129. The control function for cab wait time corrects for this bias expectedly. The corrected coefficient is -0.135. The coefficients for walking distance to the shuttle are negative and significant for both the user groups. The coefficient on walking distance captures the disutility incurred by the customer in walking to the shuttle pickup stop. Group 1 users are more sensitive to walking than group 2 users. The results allow us to estimate the monetary cost associated with the disutility of walking to the pickup stop. We estimate a disutility of ₹109 ($1.45) and ₹89 ($1.18), respectively, for walking 1 km to the shuttle stop to catch the shuttle for the two groups.

In addition to the operational levers, the customer's usage variables enter our model specification. First, the number of cumulative shuttle rides taken by a customer has a positive and significant effect on the probability of choosing the shuttle. The shuttle-specific coefficient of cumulative shuttle rides by a customer is 0.091 in the specification without IV. This shows that it is important to control for the usage variables when explaining the customer's choice. In the model with correction, we recover an estimate of 0.087. In addition to cumulative usage, the total number of rides on the platform in the recent week positively influences the probability of choosing both the shuttles and the cabs. The fixed effects for morning time slots are also significant. Our base category for the time slots is afternoon hours. As compared to the afternoon slot, people are more likely to choose shuttles and less likely to choose cabs in the morning. Finally, the pseudo R^2 for both specifications is around 0.58, which suggests that our rich model fits the data quite well.

5 Counterfactuals

In this section, we use our estimated model to conduct counterfactual analyses and provide prescriptive guidance to policy makers on congestion management in big cities. The question that we seek to answer is how to effectively increase the usage of pooled ride services and reduce the level of congestion on the roads. Our counterfactuals evaluate the impact of different policy interventions on customers' choices between private and pooled ride services and, therefore, on the number of vehicles on the road. Specifically, we examine three sets of strategies: (i) imposing congestion surcharges on private cabs and shuttles; (ii) providing discounts on shuttle service; and (iii) improving the service features of shuttle service. We call the first two sets of strategies the *price-based strategies* and the latter *operations-based strategies*. From a policy maker's point of view, it is challenging to evaluate the effectiveness of a strategy before implementing it. For example, it is very difficult to know a priori the right level of congestion surcharge to levy on the vehicles. This is where the strength of our model lies. Our estimated model allows us to measure customers' service choices when prices or service features are changed. Hence, using the estimated model, we can evaluate the relative efficacy of the different policies before implementation and, thus, provide the policy maker with a host of prescriptive solutions.

5.1 Applying Percentage Surcharges to a Congestion Zone

Local governments in large cities around the world, such as London, Singapore, and New York, have introduced congestion pricing policies to reduce congestion. In New York, for example, a surcharge is applied to all ride-hailing trips in a pre-determined congestion zone in Manhattan. In this counterfactual, we quantify the impact of imposing percentage congestion surcharges on cabs and shuttles on the number of vehicles on the road. Specifically, for all rides that cross the congestion zone—i.e., $j \in$ Zone 1 or $k \in$ Zone 1, irrespective of the time of the day, we calculate the customers' choices, the number of vehicles on the road, and the platform's revenue under policy $\mathscr{P}^{'}$. We vary the level of the surcharge by applying different price multipliers, $(1+\theta_c)$ and $(1+\theta_s)$, to p_c and p_s for cab and shuttle services, respectively. Here, θ_c and θ_s are the percentage price surcharges. The corresponding θ_s for the shuttle ride is determined by the relationship:

$$\mathcal{M} p_c \theta_c = n_s p_s \theta_s. \tag{8}$$

where $\mathcal{M}$ is a multiplier and n_s is the number of seats on the shuttle. Specifically, in our simulations, we set $\mathcal{M}$ to 2, based on proposed recommendations in New York.[1] The number of seats in the shuttle, n_s, is set to 20, equal to the median number.

Figure 2 shows the effect of applying percentage congestion surcharges to shuttles and cabs. The x-axis is $100\times(\theta_c)$. The status quo policy $\mathscr{P}_{Zone1}(p_c, p_s, X_s^1)$ is on the extreme left ($\theta_c = 0$). The *red line* in Fig. 2 shows the net percentage reduction in the number of total vehicles on the road calculated relative to the status quo level. The number of vehicles at any point is calculated as: # cab rides + $\frac{\text{\# shuttle rides}}{20}$. For this calculation, we assume that the outside option is public transportation which does not affect the total number of vehicles on the road. Although we do not observe the outside option of the customers, this calculation provides an upper bound to the magnitude of the impact of the surcharge policy

[1] https://www.nytimes.com/2019/04/24/nyregion/what-is-congestion-pricing.html.

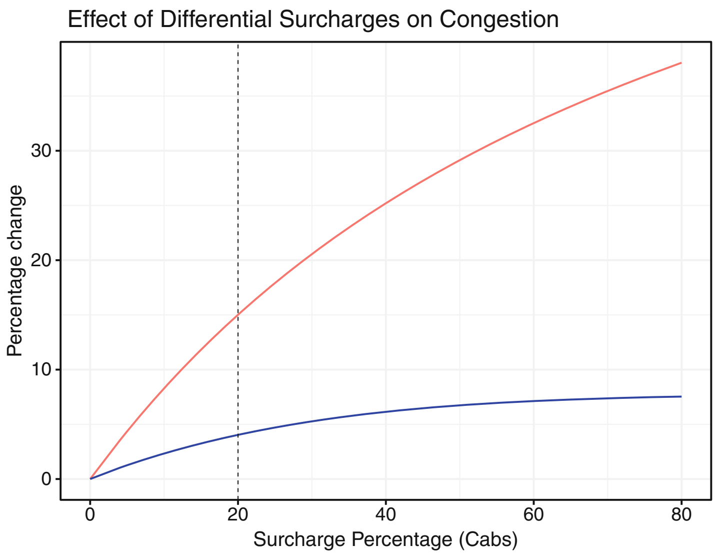

Fig. 2 Percentage reduction in total number of vehicles (red line) and percentage reduction due to pure substitution (blue line) after application of percentage surcharges. Vertical dashed line corresponds to $\theta_c = 0.2$

on congestion reduction. We also quantify the reduction in vehicles due to cab customers substituting to the shuttle service only. The *blue* line shows the net percentage reduction due to the pure substitution between the cabs and the shuttles. For instance, a 20% surcharge on the cabs ($\theta_c = 0.2$) leads to a 15% net vehicle reduction on the road. The percentage vehicle reduction due to the pure substitution effect is around 4.04%. In this calculation, since we focus only on customers substituting from the cab service to the shuttle service without taking into account potential substitutions to the outside option which includes public transportation and other ride services, and that substituting to the outside option is unlikely to increase the number of vehicles on the road given the surcharge raises the price of riding in (or driving) smaller vehicles more than that of the big ones, the blue line provides a lower bound to the impact of the surcharge policy on congestion reduction.

We also disentangle the total reduction (blue line) into two components based on the customer segments that it arises from: (i) *new users* and (ii) *experienced users*. For $\theta_c = 0.2$, the effect from new users (solid black line) is close to 80% of the total congestion reduction. This suggests that customer heterogeneity is crucial in designing an effective congestion policy.

5.2 Offering Discounts to Users

In this counterfactual, we study the service choice of customers and quantify the decrease in the number of vehicles due to the substitution between the services—i.e., the lower bound of the impact on congestion reduction when discounts are offered to shuttle customers. Specifically, we change the shuttle price by applying a discount multiplier $\mathcal{B}$ while keeping the cab price the same as observed in the data. We vary $\mathcal{B}$ in increments of 0.05 over the support of [0.60,1].

Figure 3 shows the substitution effect after applying percentage discounts on the shuttle service. The *red line* in Fig. 3 shows the percentage decrease in the number of vehicles due to pure substitution, calculated over the status quo level ($\mathcal{B} = 0$). At a discount level of 20% ($\mathcal{B} = 0.7$), the corresponding percentage decrease in the number of vehicles is 1.04%. This is about 26% of the corresponding substitution effect in the percentage surcharge counterfactual in Sect. 5.1. We decompose this reduction into two components: the reduction arising from new users (solid black line) and from experienced users (dashed black line). The corresponding reduction from the two groups is 0.7% and 0.3% of the total number of vehicles on the road, respectively. Since the number of new and experienced users is the same, our finding suggests that targeting the new users when providing discounts is more than twice as effective as targeting the experienced users.

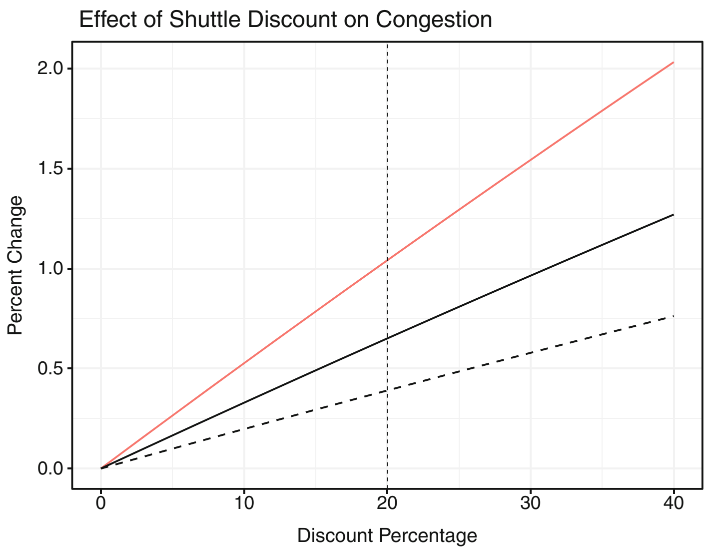

Fig. 3 Percentage reduction in the number of vehicles due to the pure substitution between the cabs and the shuttles (red line). The solid (dashed) black line is the contribution from the new (experienced) user group. Vertical dashed line corresponds to the discount percentage level of 20%

5.3 Improving Service Features

Using price-based strategies is an effective way to curb congestion on the road, but it comes with many drawbacks, such as deadweight loss to overall welfare due to its tax nature. An interesting alternative to achieve the same outcome is to use operations-based strategies, which have two advantages. First, they do not hurt customer welfare. Second, the strategies come "free" for the firm. The firm does not lose any surcharge revenue by employing these strategies. In this counterfactual, we apply multipliers $\mathcal{Q}$ to the service features and estimate the corresponding percentage reduction in the number of vehicles. We vary the multiplier in increments of 0.05 over the support of [0.65,0.95]. We do this exercise separately for the walking distance and the shuttle travel time features. As in the surcharge counterfactuals, we calculate both the total change in vehicles and the change in vehicles due to the substitution between the two services.

We report the corresponding effects in Table 2. We find that a 20% reduction in walking distance and travel time for shuttle rides leads to around a 1.46% and a 0.28% decrease in the number of vehicles on road due to the substitution between the two services. We can compare this decrease with the corresponding substitution in the price-based strategies. Considering the 20% surcharge scenario as the base case (see, Fig. 2), a 20% decrease in walking inconvenience can achieve around 35% of the total substitution achieved by the congestion surcharge. Similarly, a 20% decrease in the shuttle travel time can achieve 6.9% of the total substitution achieved by the congestion surcharge. In other words, a city could reap a significant portion of the benefits of congestion surcharges by utilizing the operational levers of the pooled ride service. The benefits could quickly *stack up* when the improvements in the various inconveniences are combined. Specifically, a 20% reduction in both shuttle ride time and walking inconvenience leads to a cumulative 1.51% vehicle reduction due to substitution (around 37.3% of the reduction achieved by the congestion surcharges in Sect. 5.1). From the policy maker's perspective, we find that improving the pooled ride service features proved to be an effective strategy to reduce congestion, while avoiding the drawbacks of the surcharge policies.

Table 2 Comparison of percentage reductions (total number of vehicles and substitution effect) when shuttle : travel time (left) and walking distances (right) are reduced

	Shuttle travel time inconvenience		Shuttle walking inconvenience	
Multiplier	Vehicle reduction	Substitution	Vehicle reduction	Substitution
0.95	0.05%	0.07%	0.07%	0.37%
0.90	0.10%	0.14%	0.15%	0.74%
0.85	0.15%	0.22%	0.25%	1.11%
0.80	0.21%	0.28%	0.36%	1.46%
0.75	0.26%	0.35%	0.48%	1.80%
0.70	0.31%	0.42%	0.62%	2.15%
0.65	0.36%	0.49%	0.79%	2.48%

6 Conclusion

In this paper, we study customer preferences of private and pooled transportation services and investigate how effective policies can be designed to incentivize customers to use pooled transportation to reduce congestion. Using detailed customer usage data from Ola's on-demand cab and fixed-route shuttle services in India, we estimate customer preferences of key service features using a discrete choice model. We account for the endogeneity of the service features, such as price and wait time, and of customers' past shuttle usage on the platform using the control function approach. We then conduct counterfactual analyses to evaluate the impact of congestion surcharge policies, discount policies, and improved pooled service features on the customers' choices and, therefore, the number of vehicles on the road. We find that, by changing operations levers such as pooled service features, instead of imposing a surcharge policy, cities can reduce a substantial amount of congestion without sacrificing consumer welfare. We also highlight the role of customer heterogeneity in improving the effectiveness of policy design. Our findings provide prescriptive recommendations to cities for designing effective policies for congestion management.

References

Ata, B., Barjesteh, N., & Kumar, S. (2019). Spatial pricing: An empirical analysis of taxi rides in neorking york city. Tech. rep., Working Paper.

Belloni, A., Chernozhukov, V., & Hansen, C. (2011). Lasso methods for gaussian instrumental variables models. arXiv:1012.1297v2

Belloni, A., Chen, D., Chernozhukov, V., & Hansen, C. (2012). Sparse models and methods for optimal instruments with an application to eminent domain. *Econometrica, 80*(6), 2369–2429.

Buchholz, N. (2020). Spatial equilibrium, search frictions and dynamic efficiency in the taxi industry. Working paper, Princeton.

Cohen, M., Fiszer, M.-D., & Kim, B.J. (2020a). Frustration-based promotions: Field experiments in ride-sharing. Working paper, SSRN.

Cohen, M., Fiszer, M.-D., Ratzon, A., & Sasson, R. (2020b). Incentivizing commuters to carpool: A large field experiment with waze. Working paper, SSRN.

Guajardo, J. A., Cohen, M. A., Kim, S.-H., & Netessine, S. (2012). Impact of performance-based contracting on product reliability: An empirical analysis. *Management Science, 58*(5), 961–979.

Han, L., Zhu, C., Wang, D.Z.W., Sun, H., Tan, Z., & Meng, M. (2019). Discrete-time dynamic road congestion pricing under stochastic user optimal principle. *Transportation Research Part E: Logistics and Transportation Review, 131*, 24–36.

He, P., Zheng, F., Belavina, E., Girotra, K. (2019). Customer preference and station network in the London bike share system. Working paper, Columbia Business School.

Kabra, A., Belavina, E., & Girotra, K. (2019). Bike-share systems: Accessibility and availability. *Management Science, 66*(9), 3799–4358.

Ostrovsky, M., & Schwarz, M. (2018). Carpooling and the economics of self-driving cars. Working Paper 24349, National Bureau of Economic Research.

Petrin, A., & Train, K. (2010). A control function approach to endogeneity in consumer choice models. *Journal of Marketing Research, 47*(1), 3–13.