Customer Voice in Two-Sided Platforms: An Examination of Surge Pricing and Customer Complaints

Linli Xu, Yanhao “Max” Wei, Yi Zhu

October 30, 2021

Abstract

Many online platforms help match consumers with service providers (e.g., lodging, ridesharing, labor). Customers are provided with review/rating systems, which the platforms use to monitor and screen service providers. However, customers may mis-direct complaints about platform-responsible features towards individual service providers. We examine the extent and consequence of this issue in the context of a large ridesharing platform, and focus on surge pricing. Our findings are as follows. First, for an average trip, surge pricing makes it 1.13 times more likely for the driver to receive a complaint (via either a low rating or a formal complaint). Second, complaints negatively impact on driver’s future income; surge-induced complaints offset about 25% of a driver’s immediate gain from the surged fare. Third, a policy implemented by the platform to cap the magnitude of surge prices seems to have decreased the extent of surge-induced customer complaints. The policy also led to significantly more customer usage of the platform. The results suggest important interactions between the pricing and review systems on ridesharing platforms, and more broadly, point to the importance for online two-sided platforms to separate and address customer complaints about platform-level features in their review/rating systems.

Keywords: two-sided matching platforms, ridesharing, customer complaints, surge pricing, rating system, review system.

*Linli Xu is Assistant Professor at Carlson School of Management, University of Minnesota. Yanhao “Max” Wei is Assistant Professor at Marshall School of Business, University of Southern California; and Yi Zhu is Associate Professor at Carlson School of Management, University of Minnesota. The authors thank Yiting Deng, Rex Du, Anthony Dukes, Avi Goldfarb, Yongdong Liu, Mitch Lovett, Shijie Lu, and Ken Wilbur for their helpful comments, as well as seminar participants at the Marketing Science Conference, Fudan University, and UC Riverside. The third author gratefully acknowledges financial support from 3M Non-tenured Faculty Grant. E-mail for corresponding author: linlixu@umn.edu.
1 Introduction

In his seminal book, *Exit, Voice, and Loyalty*, Albert Hirschman emphasizes the importance of voice, where customers let a firm know of their dissatisfaction with the quality of the product or service they have purchased, in relation to exit where customers stop purchasing from the firm. The role of voice is particularly salient in online two-sided platforms that match consumers and service providers (e.g., Uber, Lyft, Airbnb, and TaskRabbit), as almost all of these platforms offer review systems through which customers can rate, review, and sometimes submit formal complaints about individual service providers. Such customer feedback can make a substantial impact on the service provider’s income, because low ratings, negative reviews, or formal complaints decrease the likelihood that future consumers choose that service provider, reduce the priority by which platforms match the provider with consumers, and can even get the provider suspended from the platform altogether.\(^1\) For example, ridesharing drivers often post stories on social media about how rider complaints affect their earnings and, consequently, how upsetting it is to receive complaints.

The rating and review systems are designed to help platforms monitor service quality and screen out low-quality service providers. However, the design of these systems also means that sometimes customers’ voice of dissatisfaction about platform-level design (not at the fault of the service providers) may be channeled into the ratings or reviews of individual service providers. For example, on ridesharing platforms, a driver gets matched to a distant rider, increasing the passenger’s wait time. The rider might give a low rating to the driver due to the long wait experience. This low rating may adversely affect the driver’s future income. Further, if the platform does not address the problem but continues to attribute such complaints to individual drivers (albeit unknowingly), the negative experience will persist and start affecting riders’ usage of the platform. As a result, mis-attributed complaints can negatively impact both sides of the matching market and consequently the platform.

The current study examines the extent and consequences of the above issue in the context of ridesharing platforms. We focus on surge pricing, a mechanism implemented by the ridesharing platforms to balance supply and demand in real time. It is a factor that might cause customers to voice their unhappiness of paying a higher-than-normal price, but drivers should not be blamed for it by getting low ratings or complaints.\(^2\) Specifically, we investigate the following questions: (i) To what extent – if any – does surge pricing affect customer complaint rate? (ii) If surge does increase complaint rate, how does the surge-induced complaint influence a driver’s future income, relative to the immediate gain from the surged price? (iii) Is the dissatisfaction with surge substantial enough that it leads to decreases in the customer usage of the platform?

We assemble comprehensive datasets from the largest ride-hailing platform in China. The platform covers 60% of worldwide ride-hailing service in terms of user volume. The data contain

---

1 For example, Uber deactivated drivers from serving on the platform if their ratings were below 4.6 out of 5 (Business Insider, 2015). According to its own community policies, Airbnb could suspend a provider’s listings if the provider fails to meet the basic requirement on overall rating.

2 Drivers may “pursue” surge by appearing at times and locations where surge is more likely to occur. Nevertheless, it seems to us that pursuing surge should not be seen as the drivers’ fault.
trip-level information including trip GPS locations, time stamps, the estimated and actual travel
distance and duration, an indicator for whether surge pricing was applied, customer ratings, as well
as whether the customer submitted a complaint. The data cover a random sample of over 4 million
trips in a major city (hereafter City A) over a four-month period in 2016-2017 and a full record of
over 1 million trips in a smaller city (hereafter City B) during a two-month period in 2016.

We first estimate the effect of surge pricing on the probability of customer complaints. We
define customer complaints as either low ratings or formally submitted complaints. One challenge
in analyzing the complaint rate at the trip level is rider selection. Certain riders may be more
likely to experience surge pricing due to their locations and daily schedules (e.g., commute at rush
hours, visit downtown on Friday night). To account for this, we include extensive controls including
customer fixed effects, trip location, hour of the day, and day of the week. Another similar concern
arises from driver selection. In particular, some drivers work (purposely or not) at hours and
locations in which surge is more frequent. As a result, a systematically higher complaint rate on
surge trips may reflect the inherent performance of this group of drivers, rather than the effect
from surge itself. We address this issue by adding extensive driver fixed effects and trip-level driver
performance measures (e.g., delay in picking up the passenger).

We find that surge pricing indeed increases complaint rates. In City A, driving a surge trip
is 1.13 times more likely to get a complaint compared to driving a non-surge trip. The effect
is even larger in City B. To provide more confidence to these results, we explore an alternative
identification strategy. Specifically, we exploit a discontinuity in the platform’s design of surge
triggering. Ridesharing platforms typically use a hexagonal grid to discretize surge pricing decisions
across different locations. If two trips originate at two sides of a border within the grid, one trip
may experience surge pricing whereas the other does not. Exploiting this possibility, we match
each surge trip with a non-surge trip in close proximity in terms of both the time and location of
origination. We adopt a matching estimator to compare the average complaint probability of the
surge trips (treatment group) with that of the matched non-surge trips (control group). The results
align well with the results from the fixed-effect regression.

Next, we answer the question of how the surge-induced complaints affect a driver economically.
We first investigate the effect of complaints on driver daily income. Quantifying this effect is not
straightforward due to driver selection. For example, some drivers treat ridesharing as a fulltime
job, while others use it more casually. In addition, personal issues (e.g., family or health) may
simultaneously affect a driver’s ability to earn income and her quality of service. To this end, we
leverage a discontinuity in the platform’s design in calculating drivers’ scores. Specifically, a driver’s
score is based on the customer feedback received from the driver’s last 300 trips only. Complaints
from earlier trips are “forgiven” and not included in the score calculation, and thus would not affect
the driver’s income. This discontinuity suggests separating the complaints before and after the
last-300-trips cutoff to estimate causal impact of complaints on income.

We use data from City B for the estimation of the effect of complaints on driver income,

3For details on the hexagonal grid system, see https://eng.uber.com/h3/.

3
because it has the complete history of complaints received by each driver during the data period (In contrast, data from City A is a random sample). We estimate that a percentage-point increase in the complaint rate among the last 300 trips reduces a driver’s daily income by 15.3%, whereas the complaints from earlier trips do not have a significant effect. A back-of-the-envelope calculation shows that if a driver increases her frequency of driving surge trips, about 25% of the direct income lift (from the surge prices) will be offset by the future income loss (due to an increased complaint rate). This offset could discourage some drivers from responding to surge pricing, reducing the supply stimulation effect that surge pricing is originally designed to achieve.

Lastly, we examine the impact of surge pricing on the customer usage of the platform. Ex ante, the impact is not clear. Surge pricing is designed to stimulate supply and in this way increase the platform sales. However, we know that surge causes customer complaints, which may result in a negative effect on the demand side. In addition, the complaints hamper the supply stimulation effect intended by surge pricing. Thus, it is possible that surge pricing has an overall negative impact on platform sales. For empirical investigation, we exploit a policy change implemented by the platform. On March 8 of 2017, the platform set a cap on how much surged fares may exceed regular fares. Our data from City A covers this policy change. We find that the policy seemed to have reduced the effect of surge on complaint rate. In addition, the volume of trips increased at both the platform and individual customer level. These results imply that when the level of surge prices is too high, it can hurt the ridesharing platform – the negative effect of surge on customer satisfaction and demand exceeds the supply-stimulating effect. The policy change benefited the platform as well as its users.

Our findings are relevant for matching platforms in two ways. First, dynamic pricing is an important tool for many two-sided matching platforms.\textsuperscript{4} The pricing and review systems are usually thought as two separate mechanisms that serve different purposes for the platforms. However, we show that this is not the case. Platforms need to take into account the repercussion in customer complaints when designing pricing policies. Second, with the current design of review systems at matching platforms, customer voice about platform-level features (such as surge pricing) can be channeled into complaints on individual service providers. These complaints are mis-attributed and thus unfair to service providers. Moreover, they bring errors into the review system in terms of measuring and monitoring the quality of service providers. However, they are not all bad; they identify platform-level issues. As a result, if a platform can separate these complaints and use them to detect and address the underlying issue, it will likely benefit both sides of the market and ultimately the platform.

The rest of the paper is organized as follows. Section 2 discusses related literature. Section 3 presents the data as well as some descriptive analyses. Section 4 describes our empirical strategy. Section 5 presents the results. The last section discusses managerial implications and concludes.

\textsuperscript{4}Airbnb uses what it calls “smart pricing.” Other examples include OpenTable and Resy; see “Testing, Testing: Get More from your Hottest Seats,” OpenTable, 2015.
This paper connects with several strands of literature. The first includes a stream of research examining consumer complaining behavior. Hirschman (1970) first highlighted the important role that consumer voice plays in markets and illustrated how market structure affects the likelihood that consumers use voice as opposed to exit as a response to declines in product quality. Building on Hirschman’s work, Gans et al. (2021) theoretically predicted that more complaints occur in more concentrated markets and found empirical evidence supporting that prediction using tweets to U.S. airlines. A separate body of research in this literature has investigated how consumer demographics affect complaint behavior (Oster 1980; Singh 1989; Garrett and Toumanoff 2010; Ayres et al. 2014; Raval 2020). The current study contributes to this literature by examining customer complaints on two-sided online platforms where customer voice is proactively collected by the platforms and used to monitor service quality. We provide clear evidence that customer voice of dissatisfaction can be mis-attributed to individual service providers when the underlying cause of dissatisfaction is not service provider’s fault but relates to platform design.

Our paper is also related to an extensive literature on reputation and feedback systems on online platforms (see Tadelis (2016) for a review). A large number of studies have consistently shown that positive reviews lead to higher consumer demand in various empirical settings such as bids and sales on eBay (McDonald and Slawson 2002; Ba and Pavlou 2002; Bajari and Hortacsu 2003; House and Wooders 2006; Lucking-Reiley et al. 2007; Cabral and Hortacsu 2010), book sales on Amazon.com and Barnesandnoble.com (Chevalier and Mayzlin 2006), restaurant reservations on Yelp.com (Anderson and Magruder 2012; Luca 2016); physicians’ appointments on ZocDoc.com (Luca and Vats 2013); and product purchases on an online retailer website (Vana and Lambrecht 2021).

With the fast penetration of peer-to-peer marketplaces, a growing body of research has focused on the rating systems of online sharing platforms. Zervas, Proserpio, and Byers (2015) find that for properties that are posted on both Airbnb and TripAdvisor, Airbnb exhibits much higher ratings. This difference could be explained by the differences in reviewers’ preferences on the two platforms, or potential reviewing bias on Airbnb as a result of its bilateral reviewing system. Proserpio, Xu, and Zervas (2018) show that, under a bilateral reputation system, Airbnb hosts who reciprocate their guests’ efforts more have higher ratings. Kapoor and Tucker (2017) use data from a ridesharing platform in India to investigate how drivers respond to unfair ratings. They show that drivers are more likely to respond negatively to a bad rating, and thus receive subsequent low ratings. Fradkin, Grewal, and Holtz (2021) use experimental and observational data from Airbnb to explore different mechanisms leading to reviewing bias, including sorting into review, strategic reciprocity, and socially induced reciprocity. They find the largest source of reviewing bias is negative experiences being under-reported. Zhang et al. (2019) find that

---

5In addition to online reviews, a separate but related stream of research examines the impact of microblogging, a different type of user-generated content, on product demand (for example, Gong et al. 2017; Seiler, Yao, and Wang 2017).
professionally made property photos on Airbnb have a negative impact on reviews because they create unrealistically high expectations for the guests. As a result, posting professional photos leads to a trade-off for the host between today’s and tomorrow’s demand (which is adversely affected by bad reviews). All of these studies imply the critical role of rating systems for online platforms and the importance of designing their rating systems to reduce potential biases and correct unfair ratings. The current study contributes to this stream of literature by highlighting a particular bias caused by mis-attribution of platform-level problems to individual service providers.

In addition, this paper contributes to the emerging stream of research trying to understand the impact of ridesharing economy in general in various domains, including consumer surplus (Cohen et al. 2016; Lam, Liu, and Hui 2020; Liu, Wan, and Yang 2019), flexible work schedule for drivers (Chen et al. 2019; Hall and Krueger 2018; Hall, Horton, and Knoepfle 2021), driver moral hazard (Liu, Brynjolfsson, and Dowlatabadi 2021), capacity utilization (Creamer and Krueger 2016), spatial distribution of supply (Ghili and Kumar 2020), taxi markets (Frechette, Lizzeti, and Salz 2019; Wang, Wu, and Zhu 2019; Buchholz 2021) and general public welfare such as local consumer mobility and consumption patterns (Zhang and Li 2017), reduced drunk driving (Greenwood and Wattal 2017), and traffic congestion (Li, Hong, and Zhang 2016). The current study contributes to this literature by showing the magnitude of the economic impact of customer complaints on drivers’ earnings.

Lastly, the importance of dynamic pricing for ridesharing platforms has been documented both empirically and theoretically (Hall, Kendrick, and Nosko 2015; Castillo, Knoepfle, and Weyl 2018). However, prior studies examining this phenomenon remain mostly theoretical. They mainly focus on optimal pricing-mechanism designs such as dynamic threshold pricing (Riquelme, Banerjee, and Johari 2015), spatial pricing (Bimpikis et al. 2019), or a dynamic menu of prices and allocations (Abhishek, Dogan, and Jacquillat 2021). This paper is the first to empirically highlight the repercussion of the dynamic pricing on the review system.

3 Data and Descriptive Analysis

3.1 Data

The data in our empirical context come from the largest ridesharing platform in China, whose customer base accounts for over 60% of the worldwide e-hailing service user base. Similar to Uber’s user interface, a potential passenger opens the app, sets the locations for pick-up and drop-off, and requests a ride. Before the passenger sends the request, he/she is able to see an estimated fare for the trip as well as an estimated wait time for the driver’s arrival. The platform searches for available drivers near passenger’s location and then assigns the request to a driver. The matched driver cannot observe the destination of the trip until he/she picks up the passenger. At the end of each trip, the passenger observes the actual fare to be paid for the trip. The passenger can submit a rating on a scale of 1-5 stars and provide feedback about the riding experience. The passenger
can also choose to submit a formal complaint about the driver. The ratings and complaints that a driver received in recent trips are key determinants of the driver’s service score at the platform. A driver with a higher service score enjoys higher priority in trip assignment and is less likely to get cancellations from passengers, who can observe the driver’s service score after being matched with the driver.

The company provided us with data from two cities, both located in southern China. City A is a large city, among the top five in China in terms of the number of ridesharing app users. Due to a large number of trips, we were only given a random sample of 1.6 million trips (including both completed and cancelled trips) per month from January 2017 to April 2017. The company also provided the total number of trips each day over the same time period. We resample the data by day to keep the sampling rate the same across the days. After excluding the canceled trips (around 16.7%) and trips with extreme characteristics (about 1.1%)\(^7\), we end up with about 4.03 million observations for city A, consisting of about 122,000 drivers and 1.05 million riders. Over the four-month period, the average number of trips completed by a driver is 33 and the average number of completed trips for a rider is 4.

For city B, which is a much smaller city, we were given the full records of trips (instead of a random sample) from October 1 to November 30 in 2016. There are over 1.04 million observations consisting of 7,113 drivers and about 215,000 riders, after excluding the canceled trips and trips with extreme characteristics. For the subsequent analyses, we primarily use the data from city A because: (i) it has a larger sample size and (ii) the sample period covers a policy change that we will exploit in our empirical strategies. The only exception is the examination of the effect of complaints on driver income, where we use the data from city B. This is because the full data allows us to calculate the complaint rate from recent trips for each driver, about which we will provide more details in Section 4.

For each completed trip, we observe the GPS coordinates of the trip origin and destination and a set of timestamps (including: when the trip was requested, when the driver accepted the request, when the driver picked up the passenger, when the trip started, and when the trip was completed). In addition, we observe an identifier for the driver, an indicator for whether a surge price was charged, the estimated pickup distance (travel distance between the driver’s location and the passenger’s location), the estimated travel distance and duration of the trip, the actual travel distance and duration, trip fare, a 1-5 star rating if the customer rated the trip, and an indicator for whether the passenger submitted a complaint.\(^8\) Here, note that the estimated travel distance and duration were calculated before the start of the trip by a navigation software (similar to Google Map), which took into account the traffic condition at the time.

Similar to Uber, passengers on this platform predominantly give five stars if they submit a rating after a ride; a rating lower than or equal to three stars is considered a problem and often followed up

\(^6\)To protect passengers, drivers do not observe the identity of the passenger associated with a rating or complaint.

\(^7\)Extreme characteristics include: (i) passengers waited overly long, (ii) travel distance is unusually long, and (iii) fare is abnormally high.

\(^8\)Unfortunately, the data do not contain information on the nature of the complaints submitted by the passengers.
by the platform. For city A, low ratings (lower or equal to three stars) happen on about 0.73% of the trips. Formal complaints are submitted much less frequently than low ratings, on about 0.10% of the trips. A low rating and a formal complaint on the same trip is rare, taking up only 0.02% of the trips. We think these statistics indicate that, from the passenger’s point of view, giving a low rating and submitting a formal complaint are substitute ways to voice dissatisfaction. For this reason, we define our first dependent variable, “customer complaint,” as a dummy equal to 1 if the trip receives either (i) a rating lower than or equal to three stars or (ii) a formal complaint submitted by a passenger. In one robustness check, we use (i) and (ii) separately to define alternative dependent variables and the results are reported in Appendix C.

One main variable of interest is the surge indicator, which equals 1 if surge pricing was triggered for a trip, and 0 otherwise. Surge happens more often in city A (5.2%) than in city B (0.5%). This is likely because city A is a much larger, densely populated city, where the demand for ridesharing can be particularly high for some area-hour combinations.

Finally, we supplement the data with weather information. Rain has a significant influence on traffic and driving conditions. Rainfall is also known to carry effects on general emotion (Coviello et al., 2014), which may translate into higher complaint rates. We collect data on the rain condition of each day during the sample period. Related to this, we also collect data on the air quality index (AQI) of each day. Due to the increasing awareness of pollution issues in recent years, the general public in China has been paying increasingly greater attention to air quality. Air quality influences the public’s decisions on outdoor activities, how long to stay outside, as well as which transportation mode to choose.

### 3.2 Variable construction

We construct several variables at the trip level, including “traffic estimate,” “pickup delay,” “travel detour,” and “arrival delay.” These correspond to factors that could contribute to customer complaints.

Traffic estimate is computed as the ratio between the estimated trip duration (which takes into account the traffic condition at the time) and the duration normally expected for the travel distance of the trip. The expected duration is computed as the predicted value from a regression of the estimated trip duration on the estimated trip distance. Pickup delay is calculated as the log of 1 plus the passenger’s actual wait time (in minutes) for pickup minus the log of 1 plus the wait time normally expected for the pickup distance. The expected wait time is computed as the predicted value from a regression of the actual wait time on the pickup distance. Travel detour is defined as the log of the actual travel distance minus the log of the estimated travel distance of the trip. Finally, arrival delay is defined as the log of the actual trip duration minus the log of the estimated trip duration.

Among the four constructed variables, pickup delay, travel detour, and arrival delay could happen for reasons that are not a driver’s fault, such as unexpected road closures or heavy traffic.

---

9Log transformation is used upon a variable whenever the distribution is highly skewed to the right.
But they could also occur because the driver drives unnecessarily slow or intentionally lengthens the trip distance to increase the fare (which is a known practice at least among taxi drivers). By contrast, a higher level of traffic estimate should not indicate any driver’s fault.

Table 1 provides summary statistics for the main variables to be used in our analyses.

### 3.3 Some descriptive analyses

We provide some model-free evidence of the impact of surge on customer complaints directly from data. Figure 1 shows the daily difference between two complaint rates. The first complaint rate equals the fraction of trips with customer complaints among all surge trips in any given day. The second complaint rate equals the fraction of trips with customer complaints in all non-surge trips in any given day. We see that the difference stays positive for most days, suggesting that surge trips are associated with a higher chance of getting complaints compared to non-surge trips.

The difference, however, dropped significantly after March 8, 2017, when the company began to implement a new policy to cap the magnitude of surge prices. Our conversation with the company indicated that the intention to apply this policy was never revealed to the public before its implementation. The policy was likely introduced in response to the repeated media coverage of events where consumers felt “ripped off” by high surge prices.

Overall, we see Figure 1 as a descriptive analysis providing suggestive evidence that surge pricing has an effect on customer complaints. It, however, does not control for other factors that may contribute to customer complaints, such as driver characteristics, rider characteristics, and poor traffic or weather conditions. We control for these factors in the subsequent empirical analyses.

### 4 Empirical Strategy

We focus on three questions: (i) How does surge (and other factors) affect customer complaint rates? (ii) How do customer complaints affect driver income? (iii) How does capping surge prices impact the platform in terms of customer complaints and customer usage of the platform?

To answer these questions, our empirical strategy makes use of several different approaches exploiting different features of our data. The richness of our data allow us to apply extensive controls including trip location, time of the day, day of the week, and importantly, both driver fixed effects and passenger fixed effects. In addition, we exploit discontinuities in the platform’s designs of its surge pricing mechanism and rating system.

The rest of this section provides more details on the aforementioned empirical approaches.
4.1 How surge affects complaint rates

We first build our estimation on a probability function of consumer complaints at the individual trip level, where we focus on estimating the effect of surge pricing on complaint probabilities. To provide robustness, we apply a matching estimator that exploits a discontinuity in the platform’s design of the surge pricing system.

We choose to apply the linear probability model instead of logit or probit, because: (i) the linear model gives more interpretable coefficient estimates, and (ii) the large number of observations as well as the fixed effects make estimating non-linear models difficult, if feasible at all.\textsuperscript{10} The same reasons have prompted some other studies to adopt the linear probability model (e.g., Goldfarb and Tucker 2011). Specifically, let \( j \) index trips. Use \( i(j) \) to indicate the driver of the trip, and \( k(j) \) to indicate the customer of the trip. Let \( y_j \in \{0,1\} \) represent whether there is a consumer complaint on trip \( j \) (which includes both low ratings and formal complaints; see Section 3 for more details). Let \( \text{Surge}_j \in \{0,1\} \) indicate whether surge pricing was triggered on trip \( j \). We specify

\[
y_j = f_{i(j)} + c_{k(j)} + \delta' F_j + \gamma' X_j + \beta \times \text{Surge}_j + \varepsilon_j. \tag{1}
\]

The main parameter of interest here is \( \beta \), capturing the impact of surge on customer complaints. Term \( f_{i(j)} \) is the driver fixed effect, and \( c_{k(j)} \) is the passenger fixed effect. Aside from surge, other trip-level factors may also affect the complaint rate, and they are included in \( F_j \). Some of these factors are potentially the driver’s fault (e.g., delay in picking up the passenger, the extent of detour in the trip, delay in arriving at the destination), other factors are likely not the driver’s fault (e.g., raining, pickup distance, trip distance, estimated traffic condition). Parameter \( \delta \) captures the effects of these factors. Vector \( X_j \) includes a set of variables controlling for the time and location of the trip. Specifically, these controls include the hour-of-day dummies, day-of-week dummies, holiday dummies, daily trends (linear and quadratic), and the dummies indicating the zone where trip \( j \) was originated. Each zone is a hexagon with a diameter of 3 km.

Several sources of endogeneity could bias the estimate of \( \beta \). First, certain driver characteristics (e.g., gender, age) may be prone to customer complaints and at the same time have a systematically different frequency of encountering surge compared with other drivers (e.g., because they actively seek surge areas and hours). We control for driver characteristics using the driver fixed effects \( f_{i(j)} \). In a robustness test reported in Appendix C, we include driver-time fixed effects to control for driver-specific factors that may vary throughout the day.\textsuperscript{11} Second, certain passengers (e.g., commuters) may be more likely to experience surge than others and at the same time have a systematically different tendency to complain than other passengers. These unobserved passenger characteristics are controlled for by the passenger fixed effects \( c_{k(j)} \). Third, certain areas of a

\textsuperscript{10}One potential concern of utilizing the linear probability model is that the predicted probability could be outside the bounds of \([0,1]\). In our empirical context, only 2.3% of the predicted probabilities (out of 4.03 million observations) is below 0 and none of the predicted probabilities is above 1. Therefore, we consider the advantages of adopting the linear probability model outweighs this potential concern.

\textsuperscript{11}To mitigate the proliferation of fixed effects, we use 6 four-hour windows (instead of 24 one-hour windows) to construct the driver-time fixed effects.
city (e.g., business districts, airports) and certain hours of a day (e.g., 5-7pm) tend to have busy traffic, which means surge is more likely to be triggered at these locations or during these hours. Meanwhile, passengers at these locations or hours may have a systematically different tendency to complain. The variables included in $X_j$ are used to control for such unobserved time and location characteristics. In a robustness test reported in Appendix C, we include zone-time fixed effects to control for unobserved time-varying location characteristics.

In addition to regression analysis, our setting provides a suitable case for matching estimators. One advantage of matching estimators is that they are not subject to functional form misspecification. In this regard, prior literature has suggested using matching estimators for causal inferences (Imbens, 2004; Abadie and Imbens, 2006; Stuart, 2010; Wager and Athey, 2018). Moreover, the large sample size of our data allows us to closely match each surge trip with a non-surge trip with similar starting location and time.

The underlying identification strategy that motivates our matching estimator lies in the way in which surge is triggered. Given our conversation with the platform operators, ridesharing platforms determine surge pricing based on a hexagonal grid (also known as the H3 Grid; see Figure 2 for an illustration). If two trips originated at the two sides of a border in the grid, one trip may experience surge while the other may not. Intuitively, one can estimate the effect of surge by comparing the complaint probabilities on two such trips. The same logic applies to trips immediately before and after a time (instead of spatial) cutoff for surge triggering. Following this logic, we compute the matching estimator as follows. Specifically, let

$$S \equiv \{ j : \text{Surge}_j = 1 \}$$

denote the set of trips with surge pricing. Let $m(j)$ denote a non-surge trip that best satisfies two conditions: (i) the trip’s originating time (day, hour, and minute) was in close proximity to the originating time of trip $j$, and (ii) the trip’s originating GPS location is in close proximity to the originating location of trip $j$. Here, we use the time when the rider requests a trip as the trip’s originating time, and the location for pickup as the trip’s originating location. We estimate the effect of surge on the complaint probability as

$$\frac{1}{|S|} \sum_{j \in S} \left[ y_j - y_{m(j)} \right],$$

where again $y_j \in \{0, 1\}$ indicates whether there was a customer complaint for trip $j$, and $|S|$ is the number of surge trips. In the language of matching estimators, this estimate is known as the average treatment effect of the treated. The treatment in our context is the surge, and the control group consists of trips without surge that are closely matched with surge trips.

[Insert Figure 2 about here]
4.2 How complaints affect driver income

Ridesharing platforms rely on rider feedback to manage the service quality of drivers. Specifically, the platform that provides us with the data keeps a dynamically updated service score for each driver. Ratings and formal complaints by riders are both key determinants of a driver’s service score. The service score can affect driver income mainly through two channels. First, a passenger can observe a driver’s service score when she is matched with the driver. The passenger may cancel the trip if she sees a low score. Second, the platform gives priorities to drivers with higher scores when assigning trips, so it is easier for drivers with higher scores to complete more trips. Given these two channels, our goal here is to quantify the extent to which customer complaints affect driver income. The answer will help us quantify the indirect effect of surge pricing on driver income due to customer complaints.

Our estimation builds on a function of driver’s daily income and exploits a discontinuity in the platform’s design of the review system. Specifically, let $w_{it}$ denote the log of 1 plus driver $i$’s income (in Yuan) on day $t$. We specify

$$w_{it} = \gamma' X_t + f_i + \alpha \times \text{ComplaintRate}_{it} + \varepsilon_{it}. \quad (3)$$

In the above, $X_t$ includes the day-of-week dummies, holiday dummies, time trends (linear and quadratic), raining condition, and air quality. We again include driver fixed effects $f_i$ to control for unobserved driver characteristics. The main coefficient of interest in equation (3) is $\alpha$, measuring the effect of complaints on driver income. The variable associated with this coefficient, ComplaintRate$_{it}$, is the fraction of driver $i$’s recent trips that received customer complaints.

We further exploit a discontinuity in the platform’s design. During the data period, the platform computed a driver’s score based on the complaints received in the driver’s last 300 trips; any complaints received prior to the last 300th trip were not factored into the score. So, conceptually, we can estimate the causal impact of complaints on driver income by comparing the effects of complaints before and after the 300th-trip cutoff. To operationalize, we split the driver’s last 400 trips into four 100-trip intervals and extend equation (3) to include four complaint rates:

$$w_{it} = \gamma' X_t + f_i + \alpha_1 \times \text{ComplaintRate}_{1it} + \alpha_2 \times \text{ComplaintRate}_{2it} + \alpha_3 \times \text{ComplaintRate}_{3it} + \alpha_4 \times \text{ComplaintRate}_{4it} + \varepsilon_{it}. \quad (4)$$

The variable ComplaintRate$_{1it}$ equals the number of complaints from the driver’s 1-100th most recent trips before $t$ divided by 300, ComplaintRate$_{2it}$ equals the number of complaints from the driver’s 101-200th most recent trips before $t$ divided by 300, and so on. Note the common divisor 300 is used for all four complaint rates. If complaints indeed have a significant causal impact on a driver’s income, then we should expect: (i) $\alpha_1$, $\alpha_2$, and $\alpha_3$ are significantly negative and have
similar magnitudes, and (ii) $\alpha_4$ is insignificant. For estimation, we use the data from city B, for which we have the records of all trips. In contrast, the random sample from city A does not allow us to calculate accurate complaint rates for each driver.

### 4.3 How capping surge prices affects the platform

The platform made a policy change on March 8, 2017 that capped the magnitude of surge pricing (both the absolute surge amount and surge multiplier). We examine how this policy change affected: (i) the effect of surge pricing on complaint rates, and (ii) the usage of the platform. Note that without exploiting an exogenous shock such as the policy change, it is generally difficult to answer how surge pricing influences rider usage because the decision to take a surge trip by a passenger is endogenous. The policy change thus provides us a unique opportunity to shed lights on this important question.

To examine (i), we allow the effect of surge to be different before and after policy in the regression equation (1):

\[
y_j = \ldots + \beta \times \text{SurgePricing}_j
\]

\[
\ldots + \rho_1 \times \text{SurgePricing}_j \times \text{Policy}_j
\]

\[
\ldots + \rho_2 \times \text{Policy}_j
\]

The dummy variable “Policy\(_j\)” takes the value of 1 if trip \(j\) happened after March 8, 2017. The key coefficient of interest is \(\rho_1\). If the estimate of \(\rho_1\) is significantly negative, it indicates that the policy mitigated the effect of surge on complaint rate. We estimate equation (1) using the data from city A. The data from city B unfortunately does not cover the date of the policy change.

We also examine the policy change using the matching estimator (see equation 2). Specifically, we apply the matching estimator twice, first on the data before the policy change and then on the data after the policy change. If the policy indeed mitigated the effect of surge on complaint rate, then we would expect the estimated effect based on the data before policy to be larger than the estimated effect based on the data after the policy.\(^{12}\)

To examine (ii), we focus on the number of completed trips at the individual-day level. Trip completion is a key measure of the platform usage and directly relates to the platform’s revenue. Specifically, let \(v_{kt}\) be the log 1 plus the number of trips by customer \(k\) on day \(t\). We specify:

\[
v_{kt} = c_k + \gamma' X_t + \eta \times \text{Policy}_t + \varepsilon_t. \tag{6}
\]

In the above, \(c_k\) is the customer fixed effect, vector \(X_t\) collects controls including the day of week dummies, whether it was raining on day \(t\), air quality index, holiday dummies, and daily trends

\(^{12}\)For a robustness check, we apply an alternative matching criterion. Specifically, we match any surge trip before the policy change to a surge trip after the policy change. Detail and results are reported in Appendix B. We thank a referee for suggesting this approach.
(linear and quadratic). The variable Policy\(_t\) equals to one if day \(t\) is after March 8, 2017. The main coefficient of interest is \(\eta\), capturing the effect of the policy change on the ridership.

5 Results

5.1 Effects of surge on customer complaint rates

The first set of columns in Table 2 reports the parameter estimates of the main regression model (Equation 1), applied on data from city A. All reported standard errors of the estimates are clustered by drivers.\(^{13}\) The key parameter of interest here is \(\beta\), measuring the effect of surge pricing on the complaint rate.

[Insert Table 2 about here]

All coefficients are shown in percentage points. The coefficient on surge pricing is positive and significant (0.101, s.e.=0.03). It implies that driving a surge trip gives an absolute increase of 0.101 percentage point to the probability of getting a customer complaint, as compared to a non-surge trip. Note that this effect is after controlling for the location and time of the trips as well as the driver and passenger fixed effects. The complaint rate among non-surge trips in our data is 0.80 percentage point (see Table 1). The estimate suggests that surge increases the probability of getting a complaint by a factor of 1.13 (\(\simeq 1+0.101/0.80\)).

Table 2 also shows the estimates of other factors that could affect customer complaint rates. Particularly interesting are the factors that normally are not a driver’s fault. We note that the prospect of bad traffic conditions significantly increases the complaint rate. However, a longer pickup distance does not affect the complaint rate. We suspect this insignificance is due to the fact that typical pickup distance is short, with the majority below 1 km and all below 3 km. In fact, the platform matches a driver and a passenger only if they are within 3 km apart.

The factors that are potentially a driver’s fault all have significant and positive effects on complaint rates. These factors include being late picking up a passenger, making a detour on the trip, and arriving at the destination later than expected. We do not make further discussion on these effects – it suffices to say they all have expected magnitudes.

The second set of columns in Table 2 focuses on the effects of the policy change that imposed caps on the magnitude of surge prices. We will discuss it later in Section 5.3.

Matching estimator\(^{14}\) Following Section 4.1, we apply a matching estimator to estimate the effect of surge on complaint, using data from city A.\(^{14}\) In the language of matching estimators, surge trips constitute our “treatment” group, whereas the matched non-surge trips constitute the

\(^{13}\) We have tried clustering the standard errors by driver-locations; the main results hold.

\(^{14}\) Matching estimators trade efficiency for robustness to functional-form assumptions, and as a result, they generally require larger sample size to have the same statistical power as regressions. The matching estimator applied to our data from city B, which is of much smaller size, shows insignificant results (large standard errors). Thus, we do not report them here.
“control” group. We match trips based on when and where the trip originated. Thanks to the large size of our data with city A, the match is close for most of the surge trips. About 57% of the matched pairs are: (i) within 10 minutes in terms of when the passenger requests the trip, and (ii) within 500 meters in terms of the passenger’s location for pickup. Note that condition (i) essentially requires the matched trips to be on the same day (or close to the midnight between two days). The matching percentage increases to 95% as we further relax condition (i) to 20 minutes and condition (ii) to 1 km.

The first column in Table 3 shows the result from the matching estimator. We find that the complaint rate among surge trips is 0.994% compared to 0.883% for the matched non-surge trips. The difference, 0.111%, is statistically significant, implying that surge pricing increases the probability of getting a complaint by a factor of 1.13 (=0.994/0.883). This is the same as the regression results discussed above. The second and third columns of Table 3 report the matching estimator when we split the sample by the policy change. We will discuss these columns later in Section 5.3.

The results so far may give the perception that the bump in the complaint rate due to surge pricing is small (around 0.1 percentage point) and thus inconsequential. However, it is important to note that the baseline complaint rate is low (on the order of 1%). While complaints are rare, a single complaint can have a substantial impact on a driver. We notice that many drivers tell stories on social media about how complaints affect their earnings, and consequently, how upsetting it is to receive complaints. Next, we try to quantify this economic impact of complaints on driver income.

5.2 Effects of surge on driver income through complaints

We now quantify the effect of surge on driver income thanks to the increased possibility of getting a complaint. We first look at how complaints affect individual driver’s daily income. As discussed, customer complaints can negatively affect a driver’s income via two channels. First, the platform gives priority to higher-score drivers when assigning trips. Second, customers sometimes cancel trips after seeing they are matched with a low-score driver. Our goal here is to measure the joint effect of these two channels. We follow Equations (3) and (4) in Section 4.2 and exploit the discontinuity in the platform’s design whereby only the complaint history of the last 300 trips is included in a driver’s score calculation. We use data from city B because the full sample allows us to compute the complaint rates accurately for each driver, which the random sample from City A does not.

Table 4 presents the results. We note three key observations. First, the complaint rate in a driver’s last 300 trips has a significantly negative effect on the driver’s daily income. Second, if we divide the last 300 trips into three 100-trip intervals, the complaints from each of these three intervals have a negative effect on the driver’s daily income.

\[15\] The regression is run at the driver-day level. The complaint rates for driver \(i\) on day \(t\) are based on trips before \(t\) and do not include any trip within day \(t\). As a result, the regression does not account for the possible effect of the same-day complaints on a driver’s daily income.
have about the same effect on driver income. Third, controlling for the complaint rate in the last 300 trips, the complaints from the trips prior to the last 300 trips have an insignificant effect. These observations support a causal interpretation of the estimated negative effect of complaints on driver income.

To further demonstrate how the 300th-trip discontinuity helps us identify the causal impact of complaints, we divide a driver’s past 450 trips into 10-trip intervals and run a regression similar to Equation 4. In Figure 3, we plot the coefficient estimates, one for each 10-trip interval. We see a sharp difference between the estimates prior to the 300th-trip cutoff and the estimates after the cutoff.\(^{16}\) This result is consistent with that in Table 4. Overall, the discontinuity helps us preclude important confounding explanations that cannot be controlled by driver fixed effects (e.g., temporary personal issues that simultaneously affect the driver’s complaint rate and daily earnings).

We are now in a position to calculate the adverse effect of surge on driver income due to the increased complaint rate. Specifically, we are interested in quantifying the extent to which this negative effect offsets the immediate income bonus that drivers typically get from surge pricing. To start, note that the estimates in Table 4 indicate that a one-percentage-point increase in the complaint rate received from recent trips reduces a driver’s daily income by about 15%. Because this result is based on data from city B, we need the effect of surge on complaint rates in city B as well. In the appendix (Table A.1), we find that in city B, driving a surge trip leads to a 0.344 percentage-point increase in the likelihood of receiving a complaint.\(^{17}\) Therefore, if a driver increases the fraction of surge trips in all her trips by \(q\), her last-300-trip complaint rate (in percentage point) will see an increase of \(0.344q\). This increase in the complaint rate means that the driver’s daily income will suffer by \(0.344q \times 15\% \approx 5q\%\). Our conversations with the ridesharing platform indicated that, on average trip surge pricing increases the payment to the driver by about 20%. This bonus translates into a 20\(q\)% increase in the daily income for our driver. Putting these results together, we see that about 25\% of the direct income bonus from surge pricing is offset by the future income loss due to the increased complaint rate.\(^{18}\)

### 5.3 Effects of surge caps on the platform

Our final set of analyses examines the consequences of capping surge prices. As discussed, the platform made a policy change on March 8, 2017 that capped the magnitude of surge pricing.

\(^{16}\)Some may notice the coefficient estimate for the 291-300 trip interval is close to null. The reason is that the complaint rates are computed at the beginning of the day. As a driver starts to complete rides on a given day, his last 291-300 trips as seen at the beginning of the day gradually move past the cutoff to become his last 301st trip, 302nd trip, and so on.

\(^{17}\)We note the effect of surge on complaint rates is larger in city B than city A. We think this is because surge pricing was less common in city B, and as a result, the riders there were less used to it.

\(^{18}\)We acknowledge that the calculation presented here focuses on a single driver at a time and thus ignores equilibrium effects. When a significant fraction of drivers try to drive more surge trips, other drivers will find a lower probability of getting surge trips. The extent and duration of surge pricing may be affected too. A welfare analysis accounting for these effects requires an equilibrium model, which is admittedly beyond the scope of this paper.
We first examine whether the policy change mitigated the effect of surge pricing on complaint rates. This result is given in the second set of columns in Table 2 (based on equation 5 in Section 4.3). Before the policy change, the estimated effect of surge on customer complaint rate is 0.139 percentage point, implying that surge increases the probability of getting a complaint by a factor of 1.17 ($\approx 1 + 0.139/0.8$). The policy decreased the effect of surge on complaint rate by 0.089 percentage point, but this decrease is only marginally significant ($p$-value =0.09). There are two possible explanations. First, there is a lack of statistical power. Though we have a large data set, the number of surge trips as well as the number of complaints are substantially smaller, which makes it difficult to distinguish the pre- and post-policy effects of surge on complaint rates. Second, the effect of surge on complaint rate is partly driven by riders’ dissatisfaction about paying higher than regular fares, regardless of how much higher the surge price is above the regular fare. It would be managerially useful for future research to investigate whether this second explanation is true (e.g., by analyzing the texts of formal complaints submitted by riders).

The above result on the policy change is echoed by the matching estimator. As shown in Table 3, the complaint rate among surge trips is 1.035% before the policy change, compared to 0.881% for the matched non-surge trips. The difference, 0.155%, is statistically significant. This suggests that surge pricing increases the probability of getting a complaint by a factor of 1.18 ($\approx 1 + 0.155/0.881$), which is very close to our regression result from Table 2 (1.17). After the policy change, the difference in complaint rate between surge vs. non-surge trips becomes 0.062%, a decrease of 0.093 percentage point compared to the pre-policy difference (0.093=0.155-0.062). This decrease is very close to our regression result in Table 2 (0.089). We note this decrease is not statistically significant, which should not be surprising given the marginally significant result in Table 2 and the fact that matching estimators are typically more conservative than regressions (thanks to the lack of functional-form assumptions).

Next, we examine the effect of the policy on the platform sales (or usage). As discussed, ex ante, the effect can go either way. On the demand side, riders should become more willing to use the platform knowing that they would not encounter extreme surge prices. On the supply side, lower surge prices probably reduce drivers’ incentives to provide their service (though the incentives are already hampered by the effect of surge-induced complaints on driver income). Our empirical analysis aims to find out which force dominates.\footnote{In the data, we do not directly observe the number of available drivers at a given time and location. As a result, we cannot separate the supply side effect from the demand side. It would be an interesting direction for further research.}

We assemble a panel of the about 1.05 million riders in City A between January 1, 2016 and April 30, 2017. We conduct the analysis at the passenger-day level, with the dependent variable measuring the number of trips completed by a passenger in a day (see equation 6 in Section 4.3). We rely on data from city A because it covers the date of the policy change. One caveat here is that the data is a random sample, and therefore the dependent variable has a measurement error. However, this should not lead to biases in the coefficient estimates but only increases their standard errors.
Table 5 presents the results. The first set of columns shows that an average rider completes more trips per day after the policy change. The increase is about 0.22%. The result implies that the demand side force overcomes the loss in supply stimulation, leading to an increase in trip completions.

We also tested the robustness of the results by examining the impact of the policy across locations and hours. In the second set of columns in Table 5, we use a grid of hexagon zones with 3km diameter and calculate the percentage of trips that experienced surge in the data in each zone. We then split the zones into the high-surge group and low-surge group, where the high-surge group consists of the top quartile zones. Similarly, we split the sample along the time dimension and the the results are reported in the last set of columns in Table 5. We see that the policy leads an increase in trip completions across locations and hours.

The results above have important implications on the implementation of surge pricing. If surge prices are too high, it can actually lead to a decrease in the platform sales. By implementing appropriate surge caps, the platform not only decreased customer complaints, but also gained trip volume and thus revenues.

6 Concluding Remarks

In Hirschman’s framework, customer dissatisfaction leads to Exit, Voice, or Loyalty. The role of “voice” is particularly salient in modern online two-sided platforms, almost all of which provide review systems for their customers to give feedback on individual service providers. Such review systems are mostly designed to help online platforms monitor service quality. However, in this paper, we raise the important question of whether, given a review system, customer voice about platform-responsible features can be mis-channeled towards individual service providers. We examine the consequences of such misattribution on service providers’ income as well as platform’s revenue.

We leverage comprehensive trip-level data from the largest ridesharing platform in China and focus on surge pricing as a platform-level feature that may give rise to passenger complaint but is not at the fault of drivers. We rely on data from two cities and exploit multiple empirical strategies. The findings offer three key takeaways: (1) Surge pricing significantly increases the probability for a driver to receive a customer complaint on a trip (by a factor of 1.13 in our context). (2) Receiving customer complaints substantially decreases a driver’s daily income (by about 15% for each percentage-point increase in the driver’s recent complaint rate). Taking (1) and (2) together, they imply that a driver’s immediate income boost from surge pricing is offset by about 25% due to the expected loss of future income. This economic impact is large, and can weaken the original purpose of surge-pricing, i.e., stimulating supply. (3) Capping surge prices can reduce surge-induced complaints as well as increase the overall customer usage of the platform. In other words, high surge prices not only cause customer complaints but also lower platform sales.

The current paper is the first study to provide evidence that, on a ridesharing platform, the
platform’s design of surge pricing can lead to mis-attributed customer complaints towards individual drivers. This result has important consequences for both the drivers and the platform. Though the current study focuses on one type of online platforms (ridesharing) and one platform-level feature (surge pricing), the implications can be applied to two-sided online platforms in general.

First, although customer reviews (e.g., ratings, reviews, complaints) constitute probably the most important metric to evaluate the quality of service providers, they are not perfect. Customers may complain about a service provider for reasons that are not the service provider’s fault. For example, Airbnb had prices displayed in a way that failed to reflect the fees later passed on to the consumer (e.g., cleaning costs). The guests might blame the hosts for the additional charges by giving low ratings or writing a negative review. Another example is Amazon, which provides optional shipping services on behalf of its sellers. Customers may give sellers bad ratings for delayed deliveries, even when the shipping service is provided through Amazon. Currently, Amazon excludes such ratings from a seller’s profile but the exclusion is on a case-by-case basis upon the seller’s request. The two-sided nature of online matching platforms implies that they need to maintain a healthy relationship with their service providers to be successful. Therefore, it would be in the platforms’ interests to shelter service providers from the complaints caused by factors that are not the providers’ fault.

Second, understanding the effects of different platform-level factors on customer complaints helps platforms redesign their review systems. Although our research does not prescribe specific designs, it does provide directions for the platforms to consider. For example, the platforms can discount complaints or low ratings due to factors that are not the providers’ fault. One way to do so is to statistically adjust factors that contribute to the complaints or low ratings when computing the overall score for a particular provider. Related to this idea, Dai et al. (2018) developed an algorithm to aggregate restaurant reviews on Yelp.com while accounting for the informativeness of each review. Anecdotally, our conversations with a senior data scientist at Uber suggest that the platform uses the residuals from a regression of ratings on various factors, rather than the ratings themselves, to internally evaluate drivers. However, driver-responsible factors and platform-responsible factors do not seem to be considered and addressed separately. The senior scientist was actually surprised to learn that surge pricing can significantly increase the complaint rate. Aside from statistical adjustments, another approach is to use machine learning techniques to analyze the texts of formally submitted complaints and classify the underlying causes. A related approach is to ask customers to select from a list of reasons behind a complaint (e.g., driving erratically, car not clean, being rude, being late). Even though such a list is currently offered on some platforms, most listed factors are designed to pinpoint the fault of service providers. Platforms may want to consider purposely adding factors that are not the providers’ fault (e.g., fee too high, app frozen, traffic was bad) to the list.

Third, while mis-attributed complaints are counterproductive for evaluating service providers, they are useful for the platform to detect and subsequently address platform-level problems. In our

---

study, this problem is high surge prices. Surge pricing helps balance real-time supply and demand. However, service providers can get complaints thanks to price increases, and these complaints negatively impact service providers’ future income. If the complaints are left unaddressed, customers will reduce their platform usage, which could in turn negatively affect the platform’s revenue. In our study, the platform implemented a policy to cap surge prices. However, the policy seems to have been motivated by the negative coverage of high surge prices in social media, rather than the customer complaints from the platform’s own review system. Hence, platforms want to separate the complaints about platform-level features from other complaints, and actively make use of these complaints to detect and correct platform problems (instead of letting all customer complaints fall on service providers).

Finally, our results also have implications for service providers on matching platforms. In the ridesharing case, some drivers actively seek surge trips, for example, through a platform-provided heatmap that indicates where surge pricing is more likely to occur (Lu et al. 2018). Third-party developers also offer surge-tracking apps to drivers to help them pursue surge pricing. On one hand, driving surge trips provides immediate income bonus for drivers. On the other hand, drivers need to be aware that the effect of surge on complaints/ratings can substantially offset that bonus. This knowledge is particularly important for new drivers, who are not yet familiar with the consequences of complaints, and low-score drivers, who have the pressure to improve their reputation on the platform. More broadly, dynamic pricing is deployed by many online platforms to balance supply and demand in real time (e.g., Airbnb, Homeaway, OpenTable, and Resy). As service providers may actively pursue the price increases, they also want to be aware of the potential offsets thanks to the lower ratings and complaints from unhappy customers.

There are several limitations of our study, which opens possibilities for future research. First, we focus on one specific platform-level factor that is not the service providers’ fault, namely surge pricing. A direction for future research is to examine other empirical contexts to explore other factors that are not the service providers’ fault but nevertheless cause customer complaints towards service providers. A list of such common factors will be of tremendous use to guide platforms in the initial design of their review systems. In this respect, the text data of formally submitted complaints may be greatly useful to find potential factors. Second, we do not observe the actual price increase under surge or the exact value of the surge multiplier (e.g., 1.5X of regular price). An interesting direction for future research is to investigate to what extent the complaints are triggered by the fact that surge pricing is applied (extensive margin) vs. the exact difference between the surged price and regular price (intensive margin). Third, another useful direction of research is to explore the other side of the platform, namely, customers. At the time of writing this paper, Uber is introducing a new policy to suspend riders with low ratings (which are rated by drivers). It

---

22See “Uber To Start Banning Passengers With Low Ratings.” NPR, 2019. After Uber announced the new policy implementation, certain riders had already expressed concerns about unfair ratings that are due to factors that are not in their control and the likelihood of switching to use a competitor’s platform. See “No Ride for You? Uber’s Risky Ratings.” The Wall Street Journal, 2019.
would be interesting to explore how different factors, especially the ones that are out of control of the passengers, affect passenger ratings.

Service providers are vital to the prosperity of matching platforms. To this end, platforms create review/rating systems to help control the quality of service providers. However, not all customer complaints are created equal. Some complaints do reflect the fault of service providers but others reflect the problems of platforms. Simply letting all complaints fall upon individual service providers is unfair and keeps the platform blindfolded from its own problems. Different customer “voices” should be used and addressed differently. As we have demonstrated in this paper, doing so can benefit both sides of the market and consequently the platform.
References


Table 1: Summary Statistics

<table>
<thead>
<tr>
<th></th>
<th>City A</th>
<th>City B</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>S.D.</td>
<td>Mean</td>
</tr>
<tr>
<td>Complaint</td>
<td>0.0081</td>
<td>0.011</td>
<td></td>
</tr>
<tr>
<td>Surge indicator</td>
<td>0.052</td>
<td>0.005</td>
<td></td>
</tr>
<tr>
<td>Pickup distance</td>
<td>0.669</td>
<td>0.342</td>
<td>0.493</td>
</tr>
<tr>
<td>Trip-distance estimate</td>
<td>1.781</td>
<td>0.702</td>
<td>1.185</td>
</tr>
<tr>
<td>Traffic estimate</td>
<td>0.962</td>
<td>0.296</td>
<td>0.964</td>
</tr>
<tr>
<td>Pickup delay</td>
<td>-0.149</td>
<td>0.440</td>
<td>-0.042</td>
</tr>
<tr>
<td>Travel detour</td>
<td>0.026</td>
<td>0.204</td>
<td>0.040</td>
</tr>
<tr>
<td>Arrival delay</td>
<td>0.082</td>
<td>0.289</td>
<td>0.258</td>
</tr>
<tr>
<td>Rain</td>
<td>0.401</td>
<td>0.518</td>
<td>0.564</td>
</tr>
<tr>
<td>Air quality index</td>
<td>0.497</td>
<td>0.134</td>
<td>0.420</td>
</tr>
<tr>
<td>Number of drivers</td>
<td>121,873</td>
<td></td>
<td>7,113</td>
</tr>
<tr>
<td>Total observations</td>
<td>$4.03 \times 10^6$</td>
<td>$1.04 \times 10^6$</td>
<td></td>
</tr>
</tbody>
</table>

Notes: Variables are logged if their original distributions are highly right skewed.
1. “Pickup distance” is the log of the travel distance between the driver and the customer when the driver just accepted the trip request.
2. “Trip distance estimate” is the log of the travel distance from the pickup location to the destination as estimated by the navigation software.
3. “Traffic estimate” is the ratio between the trip-duration estimate (calculated by the navigation software accounting for traffic condition at the time) and the duration normally expected for such a trip.
4. “Pickup delay” is the difference between the log of the passenger’s actual wait time for pickup and the log of the wait time normally expected given the pickup distance.
5. “Travel detour” is the difference between the log of the actual travel distance of the trip and the log of the trip-distance estimate.
6. “Arrival delay” is the difference between the log of the actual trip duration and the log of the trip-duration estimate.
7. “Rain” takes a value of 0 (no rain), 1 (small rain), or 2 (big rain).
8. “Air quality index” is the usual AQI scaled down by a factor of 100.
<table>
<thead>
<tr>
<th></th>
<th>Coef.</th>
<th>s.e.</th>
<th>Coef.</th>
<th>s.e.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surge pricing</td>
<td>0.101***</td>
<td>(.027)</td>
<td>0.139***</td>
<td>(.035)</td>
</tr>
<tr>
<td>Surge pricing × Policy</td>
<td></td>
<td></td>
<td>-0.089*</td>
<td>(.052)</td>
</tr>
<tr>
<td>Policy</td>
<td></td>
<td></td>
<td>-0.024</td>
<td>(.022)</td>
</tr>
<tr>
<td>Pickup-distance estimate</td>
<td>0.002</td>
<td>(.017)</td>
<td>0.001</td>
<td>(.017)</td>
</tr>
<tr>
<td>Trip-distance estimate</td>
<td>0.156***</td>
<td>(.010)</td>
<td>0.156***</td>
<td>(.010)</td>
</tr>
<tr>
<td>Traffic estimate</td>
<td>0.382***</td>
<td>(.026)</td>
<td>0.382***</td>
<td>(.026)</td>
</tr>
<tr>
<td>Pickup delay</td>
<td>0.205***</td>
<td>(.014)</td>
<td>0.204***</td>
<td>(.014)</td>
</tr>
<tr>
<td>Travel detour</td>
<td>0.625***</td>
<td>(.043)</td>
<td>0.626***</td>
<td>(.043)</td>
</tr>
<tr>
<td>Arrival delay</td>
<td>0.907***</td>
<td>(.028)</td>
<td>0.907***</td>
<td>(.028)</td>
</tr>
<tr>
<td>Rain</td>
<td>-0.006</td>
<td>(.011)</td>
<td>-0.004</td>
<td>(.011)</td>
</tr>
<tr>
<td>Air quality index</td>
<td>-0.011</td>
<td>(.044)</td>
<td>-0.009</td>
<td>(.044)</td>
</tr>
<tr>
<td>Holiday</td>
<td>-0.027</td>
<td>(.019)</td>
<td>-0.029</td>
<td>(.019)</td>
</tr>
<tr>
<td>Driver fixed effects</td>
<td>Yes.</td>
<td></td>
<td>Yes.</td>
<td></td>
</tr>
<tr>
<td>Passenger fixed effects</td>
<td>Yes.</td>
<td></td>
<td>Yes.</td>
<td></td>
</tr>
<tr>
<td>Location fixed effects</td>
<td>Yes.</td>
<td></td>
<td>Yes.</td>
<td></td>
</tr>
<tr>
<td>Hour fixed effects</td>
<td>Yes.</td>
<td></td>
<td>Yes.</td>
<td></td>
</tr>
<tr>
<td>Day-of-week fixed effects</td>
<td>Yes.</td>
<td></td>
<td>Yes.</td>
<td></td>
</tr>
<tr>
<td>Time trend (quadratic)</td>
<td>Yes.</td>
<td></td>
<td>Yes.</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>$4.03 \times 10^6$</td>
<td></td>
<td>$4.03 \times 10^6$</td>
<td></td>
</tr>
</tbody>
</table>

Notes: The dependent variable is the dummy indicating whether the trip received a customer complaint. Coefficient values are shown in percentage points. “Surge pricing” is a dummy variable indicating whether surge pricing was triggered for the trip. The regression runs from January 1, 2017 to April 30, 2017. All standard errors are clustered by drivers.
Table 3: Matching Estimator of Effect of Surge Pricing on Complaints, City A

<table>
<thead>
<tr>
<th></th>
<th>Whole sample period</th>
<th>Before Mar. 9, 2017</th>
<th>After Mar. 9, 2017</th>
</tr>
</thead>
<tbody>
<tr>
<td>With surge pricing</td>
<td>0.994 %</td>
<td>1.035 %</td>
<td>0.938 %</td>
</tr>
<tr>
<td>Without surge</td>
<td>0.883 %</td>
<td>0.881 %</td>
<td>0.876 %</td>
</tr>
<tr>
<td>Difference</td>
<td>0.111 (.037) %</td>
<td>0.155 (.049) %</td>
<td>0.062 (.056) %</td>
</tr>
<tr>
<td>p-val:</td>
<td>Diff</td>
<td>&gt; 0</td>
<td>0.003</td>
</tr>
<tr>
<td>N</td>
<td>217,106</td>
<td>120,937</td>
<td>95,498</td>
</tr>
</tbody>
</table>

Notes: This table presents the results of the matching estimator. The sample runs from January 1, 2017 to April 30, 2017. We match each surge-trip with a non-surge trip. The first column matches trips among the entire sample period, the second column matches among the pre-policy trips and the last column matches among the post-policy trips. Matching is based on: (i) trip originating time (date, hour, and minute), and (ii) trip originating location. The originating time is taken as the time when the trip was requested (by rider). The originating location is taken as the GPS location to pick up the rider. Condition (i) effectively means that the matched trips happened on the same day (or near the midnight between two days).
Table 4: Effects of Complaints on Log Daily Income, City B

<table>
<thead>
<tr>
<th>Complaint rate, last 300 trips together</th>
<th>Coef.</th>
<th>s.e.</th>
<th>Coef.</th>
<th>s.e.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complaint rate, last 1-100 trips</td>
<td>-15.26***</td>
<td>(2.2)</td>
<td>-15.03***</td>
<td>(3.05)</td>
</tr>
<tr>
<td>Complaint rate, last 101-200 trips</td>
<td>-15.03***</td>
<td>(3.05)</td>
<td>-15.97***</td>
<td>(3.19)</td>
</tr>
<tr>
<td>Complaint rate, last 201-300 trips</td>
<td>-15.97***</td>
<td>(3.19)</td>
<td>-17.96***</td>
<td>(3.46)</td>
</tr>
<tr>
<td>Complaint rate, last 300-400 trips</td>
<td>-3.538</td>
<td>(3.34)</td>
<td>-0.732***</td>
<td>(0.075)</td>
</tr>
<tr>
<td>Rain</td>
<td>0.141***</td>
<td>(0.020)</td>
<td>0.141***</td>
<td>(0.020)</td>
</tr>
<tr>
<td>Air quality index</td>
<td>-0.732***</td>
<td>(0.075)</td>
<td>-0.732***</td>
<td>(0.075)</td>
</tr>
<tr>
<td>Tuesday</td>
<td>-0.021</td>
<td>(0.020)</td>
<td>-0.021</td>
<td>(0.020)</td>
</tr>
<tr>
<td>Wednesday</td>
<td>-0.041**</td>
<td>(0.018)</td>
<td>-0.041**</td>
<td>(0.018)</td>
</tr>
<tr>
<td>Thursday</td>
<td>-0.103***</td>
<td>(0.019)</td>
<td>-0.103***</td>
<td>(0.019)</td>
</tr>
<tr>
<td>Friday</td>
<td>0.125***</td>
<td>(0.024)</td>
<td>0.125***</td>
<td>(0.024)</td>
</tr>
<tr>
<td>Saturday</td>
<td>0.299***</td>
<td>(0.033)</td>
<td>0.298***</td>
<td>(0.033)</td>
</tr>
<tr>
<td>Sunday</td>
<td>0.261***</td>
<td>(0.019)</td>
<td>0.261***</td>
<td>(0.019)</td>
</tr>
<tr>
<td>Driver fixed effects</td>
<td>Yes.</td>
<td></td>
<td>Yes.</td>
<td></td>
</tr>
<tr>
<td>Time trend (quadratic)</td>
<td>Yes.</td>
<td></td>
<td>Yes.</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>25,385</td>
<td></td>
<td>25,385</td>
<td></td>
</tr>
</tbody>
</table>

Notes: The dependent variable is the log income of a driver at a given day. The sample is city B from September 1, 2016 to November 31, 2016. The regressions run on the last two weeks of the sample. The part of the sample prior to these two weeks is used to compute the recent complaint rates. Each complaint rate is calculated as the number of complaints within the associated trip range divided by 300. All standard errors are clustered by drivers.
Table 5: Effects of Policy Change on Ridership, City A

<table>
<thead>
<tr>
<th></th>
<th>Overall</th>
<th>Location</th>
<th>Coef.</th>
<th>s.e.</th>
<th>Coef.</th>
<th>s.e.</th>
<th>Coef.</th>
<th>s.e.</th>
<th>Coef.</th>
<th>s.e.</th>
<th>Coef.</th>
<th>s.e.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Policy</td>
<td>0.0022***</td>
<td>(4.7e-5)</td>
<td>0.0017***</td>
<td>(4.8e-5)</td>
<td>0.0017***</td>
<td>(5.4e-5)</td>
<td>0.0012***</td>
<td>(5.0e-5)</td>
<td>0.0019***</td>
<td>(4.6e-5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rain</td>
<td>-0.0007***</td>
<td>(2.1e-5)</td>
<td>-0.0003***</td>
<td>(2.1e-5)</td>
<td>-0.0007***</td>
<td>(2.4e-5)</td>
<td>-0.0003***</td>
<td>(2.2e-5)</td>
<td>-0.0006***</td>
<td>(2.1e-5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Air quality index</td>
<td>0.0019***</td>
<td>(8.7e-5)</td>
<td>0.0017***</td>
<td>(8.9e-5)</td>
<td>0.0012***</td>
<td>(1.0e-4)</td>
<td>0.0017***</td>
<td>(9.2e-5)</td>
<td>0.0012***</td>
<td>(8.6e-5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Holiday</td>
<td>-0.0104***</td>
<td>(3.0e-5)</td>
<td>-0.0076***</td>
<td>(3.0e-5)</td>
<td>-0.0084***</td>
<td>(3.4e-5)</td>
<td>-0.0067***</td>
<td>(3.0e-5)</td>
<td>-0.0082***</td>
<td>(2.9e-5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Passenger fixed effects</td>
<td>Yes.</td>
<td></td>
<td>Yes.</td>
<td></td>
<td>Yes.</td>
<td></td>
<td>Yes.</td>
<td></td>
<td>Yes.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Day-of-week fixed effects</td>
<td>Yes.</td>
<td></td>
<td>Yes.</td>
<td></td>
<td>Yes.</td>
<td></td>
<td>Yes.</td>
<td></td>
<td>Yes.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time trend (quadratic)</td>
<td>Yes.</td>
<td></td>
<td>Yes.</td>
<td></td>
<td>Yes.</td>
<td></td>
<td>Yes.</td>
<td></td>
<td>Yes.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>1.26 × 10^8</td>
<td></td>
<td>0.87 × 10^8</td>
<td></td>
<td>0.80 × 10^8</td>
<td></td>
<td>0.70 × 10^8</td>
<td></td>
<td>1.07 × 10^8</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: The dependent variable is log 1 plus the number of completed trips at the rider-day level. The “Policy” dummy equals 1 if the date is after the policy change that capped surge prices. The regressions run from January 1, 2017 to April 30, 2017. High/low surge locations are defined based on the hexagon map and the percentage of surge trips in each hexagon in the data. High/low surge hours are defined similarly. All standard errors are clustered by passengers.
Figure 1: Complaint Rate Difference between Surge vs. Non-surge Trips over Time

Notes: This figure plots, for each day, the complaint rate among surge trips minus the complaint rate among non-surge trips. The plots are based on raw data (no controls are used) of City A. The two fitting curves, separated by the policy change date, are both quadratic polynomials.
Figure 2: Surge Intensity on a Hexagon Grid, City A

Notes: Each hexagon zone has a diameter of 3 km. Colors in the graphs represent the intensity of surge pricing. The intensity is computed as the percentage of surge trips in the data at the given hexagon and hour. White color indicates an intensity close to the overall percentage of surge trips in the data. A red color indicates a higher intensity, whereas a blue/grey color indicates a lower intensity.
Figure 3: Effect of Complaints on Driver Income across Past Trips

Notes: We divide a driver’s last 450 trips into 10-trip intervals and plot the income effect of the complaints from each interval. The magnitudes of the effects are estimated from a driver-day income regression same as that in Table 4, except that there are 45 complaint rates instead of 4 complaint rates in the regression.
A Effects of surge on customer complaint rates using data from City B

Table A.1 presents the results of the main model equation (1) using data from city B. Data from city B do not cover the time point when the policy change on surge pricing was implemented. As a result, we cannot conduct the same policy-change analysis (equation 5) as in city A data. The coefficient estimates again suggest that driving surge trips increases the probability of getting a complaint, but at an even larger magnitude compared to that in city A (0.344 vs. 0.140 in percentage points). One possible explanation is that surge pricing is much less common in a small city like city B than a major city like city A (see Table 1), hence passengers are less accustomed to encountering surge pricing.

B Matching estimator: before/after the policy change

We apply an alternative matching estimator to estimate how the policy (capping surge prices) influences the complaint rates. Instead of matching surge trips with non-surge trips, we match a pre-policy trip (i.e., treatment group) with a post-policy trip (i.e., control group). The matching estimator is carried out twice, first among surge trips then among non-surge trips. The nature of this matching estimator means that matched trips can no longer be on the same day. To mitigate this issue, we require the matched trip to be to on the same day of the week, then we find the best match based on the clock time when the trip was requested as well as the pickup location. Among surge trips, 69% of matched pairs are: (i) within 10 minutes and (ii) within 500 meters. This percentage increases to 89% when we relax condition (i) to 20 minutes and condition (ii) to 1 km. Among non-surge trips, the percentages are 96% and 100%, respectively. The results are shown in Table B.1.

Among matched non-surge trip pairs, the complaint rate remained virtually unchanged (0.003 percentage point) before and after the policy change, which is expected. Among matched surge trip pairs, the difference is -0.064 percentage point. Together, they imply that the effect of surge on complaint rate reduced by 0.064+0.003=0.067 percentage point, which is not far from our result in Table 3 (i.e., 0.155-0.062=0.093). However, the reduction is not statistically significant, which is also the case with Table 3.

C Robustness checks

We provide robustness checks for our main regression that shows the effects of surge on customer complaints (Table 2). Table C.1 presents four model specifications. Column (a) and (b) change the definition of the dependent variable. Column (a) defines the dependent variable as whether a trip
receives a rating equal to 3 out of 5 or lower (while ignoring the formally submitted complaints). Column (b) defines the dependent variable as whether a formal complaint is submitted by the passenger (while ignoring the ratings). We see that in both columns the effect of surge is positive and significant. The magnitude of the effect is smaller in Column (b), most likely because formal complaints are more rare in the data (accounting for only 0.1% of the trips). The remaining coefficients are consistent with what are reported in Table 2.

Column (c) and (d) add interactions between fixed effects. Column (c) includes driver-time fixed effects to control for unobserved driver-specific factors that may vary throughout a day. To mitigate the proliferation of fixed effects, we use six 4-hour windows (instead of 24 one-hour windows) to construct these driver-time fixed effects. Column (d) includes zone-time fixed effects to control for unobserved location factors that may vary throughout a day. We see that the coefficients in these two columns are quite similar in size and significance to the ones presented in Table 2.
Table A.1: Effects of Surge Pricing on Customer Complaint Rates, City B

<table>
<thead>
<tr>
<th></th>
<th>Coef.</th>
<th>s.e.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surge pricing</td>
<td>0.344*</td>
<td>(0.198)</td>
</tr>
<tr>
<td>Pickup-distance estimate</td>
<td>0.131***</td>
<td>(0.043)</td>
</tr>
<tr>
<td>Trip-distance estimate</td>
<td>0.660***</td>
<td>(0.030)</td>
</tr>
<tr>
<td>Traffic estimate</td>
<td>0.908***</td>
<td>(0.074)</td>
</tr>
<tr>
<td>Pickup delay</td>
<td>0.196***</td>
<td>(0.033)</td>
</tr>
<tr>
<td>Travel detour</td>
<td>0.579***</td>
<td>(0.081)</td>
</tr>
<tr>
<td>Arrival delay</td>
<td>1.057***</td>
<td>(0.063)</td>
</tr>
<tr>
<td>Rain</td>
<td>-0.038</td>
<td>(0.024)</td>
</tr>
<tr>
<td>Air quality index</td>
<td>0.096</td>
<td>(0.064)</td>
</tr>
<tr>
<td>Holiday</td>
<td>-0.212</td>
<td>(0.083)</td>
</tr>
<tr>
<td>Driver fixed effects</td>
<td>Yes.</td>
<td></td>
</tr>
<tr>
<td>Passenger fixed effects</td>
<td>Yes.</td>
<td></td>
</tr>
<tr>
<td>Location fixed effects</td>
<td>Yes.</td>
<td></td>
</tr>
<tr>
<td>Hour fixed effects</td>
<td>Yes.</td>
<td></td>
</tr>
<tr>
<td>Day-of-week fixed effects</td>
<td>Yes.</td>
<td></td>
</tr>
<tr>
<td>Time trend (quadratic)</td>
<td>Yes.</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>$1.04 \times 10^6$</td>
<td></td>
</tr>
</tbody>
</table>

Notes of Table 2 apply except that the data is from City B and runs from October 1, 2016 to November 31, 2016.

Table B.1: Alternative Matching Estimator, City A

<table>
<thead>
<tr>
<th></th>
<th>Surge trips</th>
<th>Non-surge trips</th>
</tr>
</thead>
<tbody>
<tr>
<td>Before Mar. 9, 2017</td>
<td>0.993 %</td>
<td>0.795 %</td>
</tr>
<tr>
<td>After Mar. 9, 2017</td>
<td>0.929 %</td>
<td>0.798 %</td>
</tr>
<tr>
<td>Difference</td>
<td>-0.064 (0.066) %</td>
<td>0.003 (0.012) %</td>
</tr>
<tr>
<td>p-val:</td>
<td>Diff</td>
<td>&gt; 0</td>
</tr>
<tr>
<td>N</td>
<td>91,622</td>
<td>1,814,646</td>
</tr>
</tbody>
</table>

Notes: This table presents the results of the alternative matching estimator. The sample runs from January 1, 2017 to April 30, 2017. We match each pre-policy trip with a post-policy trip. The first column matches among surge trips and the second column matches among non-surge trips. Matching is based on: (i) trip originating time (including day of week, hour, and minute), and (ii) trip originating location. The originating time is taken as the time when the trip was requested (by rider). The originating location is taken as the GPS location to pick up the rider.
<table>
<thead>
<tr>
<th></th>
<th>Model a</th>
<th>Model b</th>
<th>Model c</th>
<th>Model d</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coef.</td>
<td>s.e.</td>
<td>Coef.</td>
<td>s.e.</td>
</tr>
<tr>
<td>Surge pricing</td>
<td>0.081***</td>
<td>(0.025)</td>
<td>0.026***</td>
<td>(0.010)</td>
</tr>
<tr>
<td>Pickup-distance estimate</td>
<td>-0.005</td>
<td>(0.016)</td>
<td>0.004</td>
<td>(0.006)</td>
</tr>
<tr>
<td>Trip-distance estimate</td>
<td>0.109***</td>
<td>(0.009)</td>
<td>0.053***</td>
<td>(0.004)</td>
</tr>
<tr>
<td>Traffic estimate</td>
<td>0.323***</td>
<td>(0.024)</td>
<td>0.067***</td>
<td>(0.009)</td>
</tr>
<tr>
<td>Pickup delay</td>
<td>0.180***</td>
<td>(0.013)</td>
<td>0.035***</td>
<td>(0.005)</td>
</tr>
<tr>
<td>Travel detour</td>
<td>0.545***</td>
<td>(0.039)</td>
<td>0.129***</td>
<td>(0.019)</td>
</tr>
<tr>
<td>Arrival delay</td>
<td>0.709***</td>
<td>(0.026)</td>
<td>0.237***</td>
<td>(0.012)</td>
</tr>
<tr>
<td>Rain</td>
<td>-0.006</td>
<td>(0.011)</td>
<td>0.001</td>
<td>(0.004)</td>
</tr>
<tr>
<td>Air quality index</td>
<td>-0.019</td>
<td>(0.041)</td>
<td>0.019</td>
<td>(0.016)</td>
</tr>
<tr>
<td>Holiday</td>
<td>-0.025</td>
<td>(0.018)</td>
<td>-0.002</td>
<td>(0.007)</td>
</tr>
<tr>
<td>Driver fixed effects</td>
<td>Yes.</td>
<td></td>
<td>Yes.</td>
<td></td>
</tr>
<tr>
<td>Passenger fixed effects</td>
<td>Yes.</td>
<td></td>
<td>Yes.</td>
<td></td>
</tr>
<tr>
<td>Location fixed effects</td>
<td>Yes.</td>
<td></td>
<td>Yes.</td>
<td></td>
</tr>
<tr>
<td>Hour fixed effects</td>
<td>Yes.</td>
<td></td>
<td>Yes.</td>
<td></td>
</tr>
<tr>
<td>Day-of-week fixed effects</td>
<td>Yes.</td>
<td></td>
<td>Yes.</td>
<td></td>
</tr>
<tr>
<td>Time trend (quadratic)</td>
<td>Yes.</td>
<td></td>
<td>Yes.</td>
<td></td>
</tr>
<tr>
<td>Driver-hour fixed effects</td>
<td>No.</td>
<td></td>
<td>No.</td>
<td></td>
</tr>
<tr>
<td>Location-hour fixed effects</td>
<td>No.</td>
<td></td>
<td>No.</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>4.03 × 10^6</td>
<td></td>
<td>4.03 × 10^6</td>
<td></td>
</tr>
</tbody>
</table>

Notes of Table 2 apply.