Contents lists available at ScienceDirect

Transportation Research Part C

journal homepage: www.elsevier.com/locate/trc

## Highlights

**The impact of ride-hail surge factors on taxi bookings**

*Transportation Research Part C xxx (xxxx) xxx*

Sumit Agarwal, Ben Charoenwong, Shih-Fen Cheng*, Jussi Keppo

- We estimate the cross-price elasticity of 0.26 between ride-hailing and taxi services.
- The results suggest up to 18% of taxi fares are due to cross-platform substitution.
- Using surge factors when guiding taxi drivers leads to 9.4% reduction in vacant time.
- We estimate that such saving could lead to 2.6% increase in the hourly taxi trips.
- Other markets with less comparable taxi service will have smaller effects.

Contents lists available at ScienceDirect

Transportation Research Part C

journal homepage: www.elsevier.com/locate/trc

# The impact of ride-hail surge factors on taxi bookings

Sumit Agarwal[a], Ben Charoenwong[a], Shih-Fen Cheng[b,*], Jussi Keppo[a]

[a] NUS Business School, National University of Singapore, Singapore
[b] School of Computing and Information Systems, Singapore Management University, Singapore

ARTICLE INFO

*Keywords:*
Ride-hailing surge pricing
Taxi demand
Cross-price elasticity of taxi bookings
Taxidemand prediction

ABSTRACT

We study the role of ride-hailing surge factors on the allocative efficiency of taxis by combining a reduced-form estimation with structural analyses using machine-learning-based demand predictions. Where other research study the effect of entry on incumbent taxis, we use higher frequency granular data to study how location-time-specific surge factors affect taxi bookings to bound the effect of customer decisions while accounting for various confounding variables. We find that even in a unique market like Singapore, where incumbent taxi companies have app-based booking systems similar to those from ride-hailing companies like Uber, the estimated upper bound on the cross-platform substitution between ride-hailing services and taxi bookings is only 0.26. On the other hand, we show that incorporating surge price factor improves the precision of demand prediction by 12% to 15%. Our structural analyses based on a driver guidance system finds this improved accuracy in demand prediction reduces drivers' vacant roaming times by 9.4% and increases the average number of trips per taxi by 2.6%, suggesting the price information is valuable across platforms, even if elasticities are low.

## 1. Introduction

The ride-hailing industry has grown from a niche start-up market to the canon of technological disruption, with around US$60 billion in revenue as of 2017 and is projected to grow to $285 billion in revenue by 2030.[1] Moreover, since the entry of Uber, various competitors have emerged, such as Lyft in the United States, Didi in China, and Grab in Southeast Asia. Within three to five taps on a smartphone, a customer can book a ride through another app. So, when the price of one option rises, customers can presumably easily switch to alternative options. However, extant research on inter-platform competition has primarily studied the impact of ride-hailing service entry on incumbents, without precisely studying how dynamic relative prices between different platforms affect consumer decisions through time and across locations (Wang and Yang, 2019).

The aim of this paper is to study whether and to what extent relative prices across ride-booking platforms affect customer choices and transportation network efficiency. More specifically, we study the relation between multiplicative ride-hailing surge factors[2]

* Corresponding author.
*E-mail addresses:* ushakri@yahoo.com (S. Agarwal), bizbgc@nus.edu.sg (B. Charoenwong), sfcheng@smu.edu.sg (S.-F. Cheng), keppo@nus.edu.sg (J. Keppo).

[1] According to a study by Goldman Sachs as reported in a May 27, 2017, MarketWatch article by Caitlin Huston, titled "Ride-hailing industry expected to grow eightfold to $285 billion by 2030". URL: https://web.archive.org/web/20170528220543/http://www.marketwatch.com:80/story/ride-hailing-industry-expected-to-grow-eightfold-to-285-billion-by-2030-2017-05-24 (archived on May 28, 2017).

[2] A multiplicative surge factor is a common dynamic pricing mechanism defined as a price multiplier relative to regular taxi fares, which has been used by Uber and whose effectiveness has and garnered substantial research attention. For example, Zha et al. (2018), Chakraborty et al. (2020), Zuniga-Garcia et al. (2020), and Garg and Nazerzadeh (2021) theoretically and empirically study how travelers react to surge pricing and the implication of the pricing mechanism on the transportation efficiency.

https://doi.org/10.1016/j.trc.2021.103508
Received 30 April 2021; Received in revised form 2 December 2021; Accepted 2 December 2021

and customer's choices, and quantify how having access to this information could improve the prediction quality of demands for competing taxi services. An empirical challenge for studying cross-platform consumer decisions has been acquiring data for multiple competing services. To this end, we use a unique empirical setting and data from Singapore to study the impact of ride-hailing surge prices on conventional taxi booking demand and the allocative efficiency of taxis using granular data.[3] Our goal can be broken down into two sets of analyses. First, to establish an upper bound on the cross-price substitution effect from ride-hailing services to taxi bookings, we use an empirical specification with high-dimensional fixed effects utilizing variation in surge prices and rides in a narrowly-defined geographical region within a half-hour interval within a given day of the week.[4] Second, to study the informational role of surge prices on taxi bookings and allocative efficiency, we use our detailed taxi data with demand forecasts by a random forest model to simulate a counterfactual guidance system that directs taxis to areas with higher taxi demand and study the decrease in vacant roaming times and the increase in the number of rides.

There are three main benefits from the Singapore setting. First, taxis in Singapore already provide similar features as those offered by ride-hailing services, such as cashless payments, location tracking, and booking through a smartphone app. Therefore, to the extent that taxis and ride-hailing services are more easily substitutable in Singapore compared to other countries, our upper bound estimates on the cross-price elasticity of taxi ride demand also provide an upper bound for other countries. Second, the leading ride-hailing service provider in Singapore – Grab – has a ride-hailing feature similar to UberX, which uses a surge factor mechanism to balance supply and demand of rides.[5] Third, the Land and Transport Authority of Singapore maintains a dataset containing the mobility traces of all taxis in Singapore updated every 30 s, showing their location, whether a taxi is occupied or available, and if occupied, whether a trip was initiated by app booking or street pick-up. These three unique characteristics address the data availability challenge to permit a study documenting the relation between surge factors and customer decisions in a precise location at specific times.

Our first empirical study documents the reduced-form relationships between taxi bookings and surge factors, controlling explicitly for taxi supply within a half-hour interval and route. Explicitly controlling for the supply measures addresses typical concerns of simultaneity bias when seeking to estimate demand elasticities. The half-hour interval, weekend, and route fixed effects control for some (but not all) unobserved variables that affect overall ride demand such as daily seasonalities in rides. Therefore, to the extent that confounding correlated demand shocks affecting both Grab and taxi booking remain unaccounted for, the estimated coefficient represents an upper bound on the cross-price elasticity of taxi booking demand relative to ride-hailing surge factors as any remaining omitted variables would affect the demand of both ride-hailing and taxis, introducing positive bias. Yet, we find that a 10% increase in the surge factor predicts an increase in taxi booking of 2.6% within the same region, half-hour interval, and day-of-week. We show that these estimates are not due to constrained taxi supply biasing the estimate downwards. A higher surge factor is not correlated to higher taxi supply in an area but corresponds to a decrease in excess supply and increase in used supply. Therefore, our main estimates likely reflect customer behavior.[6]

Our second empirical study tests whether surge factors affect the demand prediction of taxi bookings. Although the estimated economic effect of surge factors on taxi bookings may be small, it may capture meaningful statistical variation for prediction purposes. Using a random forest model, we find that the surge factor information is even more important in demand forecasting than the distance of the ride, the month, and the amount of rainfall. Including the surge factor in addition to other variables improves the accuracy of taxi booking predictions between 12% and 15% in terms of out-of-sample root-mean-squared error.

Our third empirical study evaluates how the improvement in the statistical predictability of demands maps to transportation network efficiency like taxi driver revenues and the number of filled trips. Taking the perspective of a taxi operator, we simulate a driver guidance system (DGS) for taxi drivers – a realistic structural model that has been implemented into the everyday industry practice – to show via simulations that a 15% improvement in the accuracy of demand predictions leads to 9.4% reduction in the average vacant roaming time. The DGS is evaluated in a microscopic agent-based taxi fleet simulation in which taxis follow roaming guides to maximize the taxi fleet's utilization, and is based on Cheng et al. (2018).[7] The roaming guides in our simulation are generated with demand predictions as inputs. In addition to demand-side information, the DGS also incorporates supply-side information (the current and future distribution of vacant taxis), with the objective of maximizing the expected income of guided drivers. The counterfactual analysis holds all other settings equal and then studies the effects of changes in the demand prediction accuracy.

[3] Although, in theory, other modes of transportation like walking, taking a bus or trains/subways, street pickups could potentially be substitutes as well (for example, see Grahn et al. (2021)), those other modes of transportation are not as close substitutes as taxi booking through an app, mostly due to the difference in expected travel time; for example, for the 90 origin/destination pairs that we identify for surge price data collection, their travel times via public transport are on average almost 3 times that of booked taxis or private cars. Therefore, we focus our study on the relation between Grab surge prices with taxi bookings. We discuss the similarities between the ride-hailing and taxi apps in Section 3 and show some evidence corroborating this premise in Appendix B.

[4] Geographical regions range from around 2 $km^2$ (1 $mi^2$) to 20 $km^2$ (15 $mi^2$), where most regions in Singapore are closer to the smaller end. The larger regions include areas in Singapore consisting of two water-catchment reservoir areas.

[5] Additional details are available at https://web.archive.org/web/20190330163258/https://www.grab.com/sg/justgrab/ (archived on March 30, 2019).

[6] In the Appendix, we also study the effect of trip length on the customers' behavior. We do not find any statistically significant differences in the price elasticity of customers in short-distance versus long-distance trips. A short-distance trip is defined as a ride that is less than 5 km, or about 3.1 miles. This is about 10% of Singapore's length on the east–west dimension (the longer end). Moreover, we find that the relation between surge factors and taxi bookings is concave in Appendix B.2, consistent with customers reacting to the presence of a surge factor.

[7] In practice, the guidance for drivers is delivered via a smartphone app, which is illustrated in Fig. 1. The app highlights the recommended region in red, and when the guided driver enters the recommended region, additional details such as which streets or taxi stands to visit are displayed. This exact DGS model has since been adopted and deployed by the actual taxi fleet that we study.

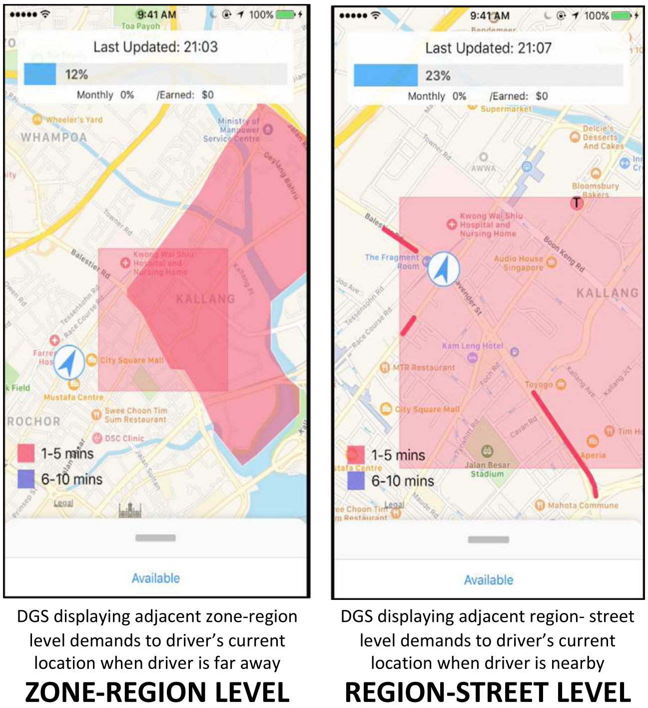



**Fig. 1.** Screenshot of the driver guidance system (DGS) app that implements part of our empirical analysis.

Overall, we find that incorporating the surge factor into demand forecasts increases the taxi fleet's efficiency and consumer welfare. Extrapolating the results observed from the real-world DGS field-trial, we estimate that a 9.4% reduction in the average vacant roaming time could lead to an increase of 2.6% in the number of trips per hour. In addition, the higher counterfactual taxi utilization rate corresponds to a weakly decreasing waiting time for customers as the matching between the riders and taxis improves. For the transportation network as a whole, given that the surge factor is based only on the imbalance between ride-hailing demand and supply, the constrained ride-hailing service is partly mitigated with the improved taxi allocation. Therefore, although the upper bound cross-price elasticity between ride-hailing and taxi bookings is low, incorporating the relative price can improve the efficiency of the whole transportation network for ride-booking.

The rest of the paper is organized as follows. We identify our contributions to the related literature in Section 2. In Section 3, we introduce the taxi and ride-hailing market in Singapore. Section 4 explains the used data and empirical methodology, and then Section 5 gives the main results. Section 6 extends the empirical analysis to taxi bookings forecasting and its policy implications. Finally, Section 7 concludes.

## 2. Related literature

Our research studies the interplay between ride-hailing surge pricing and the taxi booking demand. We illustrate how taxi demand predictions can be improved by incorporating surge pricing information from the ride-hailing platform and quantify how this prediction accuracy improvement can lead to the increase in taxi driver's productivity using simulations. Our findings contribute to the transport economics literature studying the effect of ride-hailing services like Uber, Lyft, and Grab on the taxi industry and customer demand prediction.

*Impact of Ride-hailing on Taxis and Other Modes of Transportation.* The empirical effect of ride-hailing services on the incumbent transportation industry has led to a range of diverse legal views regarding their regulation (e.g., see Posen (2015) and Ross (2015)). Research such as Young and Farber (2019) document that ride-hailing service leads to a decrease in taxi ridership, particularly among young and affluent riders. This finding applies not just to the developed nations but also the developing nations (e.g., Acheampong et al. (2020) documents similar results in Ghana). On the other hand, there are also researchers who point out that although ride-hailing service decreases taxi ridership in general, ride-hailing services and incumbent taxi operators have reached a new equilibrium where two modes complement and co-exist.[8] For example, by using the staggered introduction of Uber in different states in the

[8] This could be true if taxis and ride-hailing services differ in dimensions other than pricing, effectively segmenting the market into different types of riders. For example, Brown and LaValle (2021) compare taxi and ride-hail service quality in Los Angeles and find meaningful differences, Rayle et al. (2016) find that people do not seem to want to use taxis if they did not use ride-hailing, and Schaller Consulting (2017) find that taxis and ride-hailing cars serve different areas. We discuss the Singapore setting relative to this existing literature later in Section 3.

USA, Berger et al. (2017) show that Uber does not seem to cause adverse employment outcomes for taxi drivers but reduces the earnings potential of the incumbent drivers. Similarly, Nie (2017) study the impact of ride-hailing entry on the taxi industry in Shenzhen, China, and find that taxi ridership stabilized at a lower level after initially suffering a sharp decline. Relative to this work, our research is unique in that rather than studying the entry of ride-hailing services, we focus on higher frequency and granular customer decisions using intensive margin variation in surge pricing at specific times and locations.

*Customer Demand Prediction.* The second research area our paper contributes to is the demand prediction for taxi and ride-hailing services. Lu et al. (2018) show that providing a surge factor heat-map raises driver revenues by up to 70%. Together with the advancement in the computational techniques, researchers have moved from offline statistical approach such as time series modeling (e.g., Moreira-Matias et al. (2013)) to a state-of-the-art machine learning approach such as deep learning (e.g., Yao et al. (2018) and Geng et al. (2019)). Given the prevalence of ride-hailing services, there are also recent interests in predicting surge prices (e.g., Battifarano and Qian (2019)). However, to the best of our knowledge, the cross-platform substitution effect between ride-hailing and taxi services has not been studied in the taxi/ride-hailing demand prediction literature.

Finally, our findings also provide implications for the allocation of available taxis based on demand predictions, relating to existing research on the optimal dynamic allocation of scarce resources with uncertain demand in providing transportation and logistics services. For example, Liu et al. (2020) study the impact of demand predictions on the optimization of last-mile delivery, Glaeser et al. (2019) use customer location data to optimize the location of retail pick-ups, and Bimpikis et al. (2019) study optimal pricing in a ride-hailing network. Bian and Liu (2019a,b) analyze optimal ride-sharing allocation with customer-specific requirements, and Ramezani and Nourinejad (2018) study how to utilize model predictive control approach in enabling city-wide taxi dispatch system. Consistent with the above studies, Cheng et al. (2018) conduct a series of field trials with actual taxi drivers and demonstrate that providing guidance to taxi drivers can indeed reduce vacant cruising time significantly. Our work directly builds off of the latter research using the same simulation procedure as a means to quantify the estimated relation between surge factors and taxi bookings.

## 3. Taxi booking and ride-hailing in Singapore

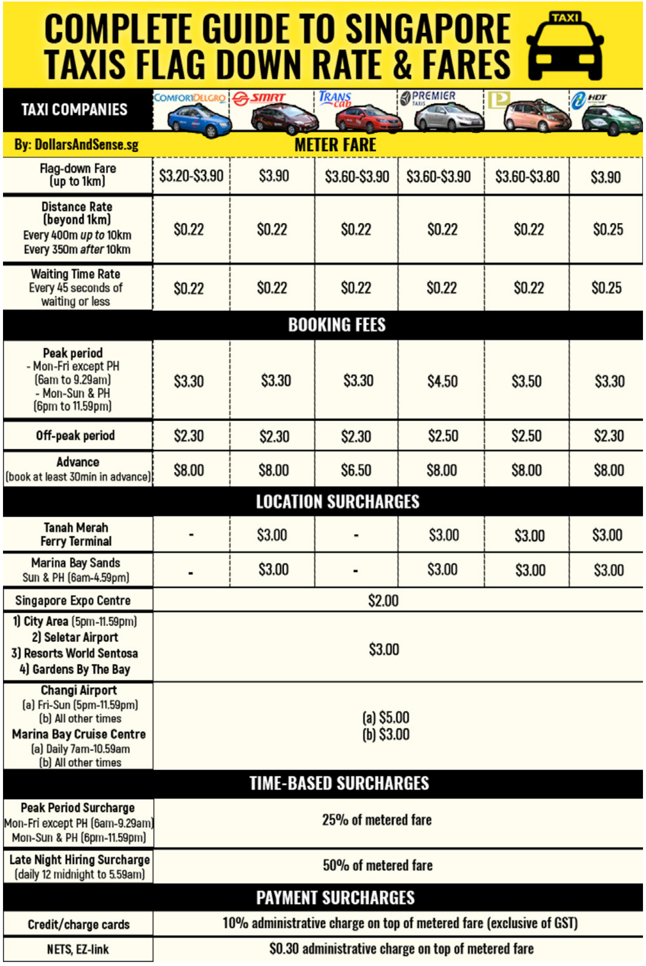
The taxi market in Singapore is regulated and more than 99% of taxis are owned by seven taxi companies, with the largest company having almost 60% market share. Drivers must pass a vocational license test, where only Singapore citizens 30-year old and above can apply, and then drivers must rent a taxi from one of the taxi companies. The rental cost is set by individual operators and covers all vehicle-related expenses, just like non-vocational car rentals. Drivers pay for the variable costs such as fuel, parking, and road tolls. The taxi industry grew steadily in terms of aggregated fleet size until the end of 2014, after which the fleet decreased due to the emergence of ride-hailing firms such as Uber and Grab. Despite the impact of ride-hailing platforms, the taxi industry in Singapore is competitive in terms of fleet sizes (around 27,000 in total at its peak), booking technology (which as shown later, is very similar to Grab/Uber), and costs (ride-hailing firms in fact uses taxi fare as the baseline to price their services). The only difference we can see is that taxi companies do not have access to private cars, and they do not adopt dynamic (surge) pricing during our study period. This is drastically different from other major cities commonly studied in the literature of taxi/ride-hailing industry, where taxis often suffer from the issues of availability and affordability.

Taxi fares in Singapore are set by individual companies, and follows schemes described in Fig. 2, which include the meter fares, booking fees, location surcharges, time surcharges, and payment surges.[9] The fare component requiring some elaboration is "surcharge", which was introduced in 1994 to encourage taxi supply at strategic times and locations. These surcharges can be viewed as "static" price surges to spatiotemporally balance supply and demand. That is, the taxi surcharges do not respond dynamically and geographically to unexpected demand–supply imbalances. Taxi stands at every mall provide the pricing schedule for taxis, and every taxi is required by the Land and Transport Authority (LTA) to have a standardized sticker of the fare schedule on the window. Taxi "peak hour" pricing is decided by individual operators and approved by the Public Transport Council (PTC), while the LTA decides the "area-based" surge pricing (e.g., airports, Marina Bay Sands, and Central Business District). Although it is not clear how much control the PTC exercises over the pricing decisions, the only differences in prices across taxi operators are the initial fixed-cost of the taxi. All remaining rates and peak hour pricing are the same. Meanwhile, the metered fare never has dynamic surge charges and are always charged based on waiting time (on the road) and driving distance mechanically.

Customers can book taxis through a phone call, and from 2014, also book taxis through the company's smartphone app (although not all operators have their own app).[10] All taxi stands have a unique identification number that can be supplied when ordering a taxi through the phone. However, an important point to note is that for this taxi app (just like with Grab, described below), users see the price of the entire trip, shown as either an estimate with a lower and upper range or a flat fare. Although the taxi company does not reveal how the flat fare is calculated, from our observations, it is most likely the average or the median of historical prices from the same origin and destination.

[9] Figure from a personal finance blog Dollars and Sense from August 2018. See https://web.archive.org/web/20180904165850/https://dollarsandsense.sg/complete-guide-singapore-taxis-flag-rate-fares/ (archived on September 4, 2018). In addition, the LTA also provides similar information on taxi fares: https://web.archive.org/web/20210123133326/https://www.lta.gov.sg/content/ltagov/en/getting_around/taxis_private_hire_cars/taxi_fares_payment_methods.html (archived on January 23, 2021).

[10] In recent years, most smaller taxi companies ended their support for their own booking apps and instead started to collaborate with Grab to provide booking service to customers. As of April 2017 (the beginning of our sample), five out of the seven taxi fleet operators had formally reached an agreement with Grab to allow their taxis to be part of the Grab fleet via the JustGrab service. A taxi trip booked via the JustGrab service is priced dynamically within the Grab service platform using Grab's surge pricing scheme, which overrides traditional taxi fares and surcharges.



# COMPLETE GUIDE TO SINGAPORE TAXIS FLAG DOWN RATE & FARES

By: DollarsAndSense.sg

| TAXI COMPANIES | ComfortDelGro | SMRT | Trans cab | Premier Taxis | | HDT |
|---|---|---|---|---|---|---|
| **METER FARE** | | | | | | |
| Flag-down Fare (up to 1km) | $3.20-$3.90 | $3.90 | $3.60-$3.90 | $3.60-$3.90 | $3.60-$3.80 | $3.90 |
| Distance Rate (beyond 1km) Every 400m *up to* 10km Every 350m *after* 10km | $0.22 | $0.22 | $0.22 | $0.22 | $0.22 | $0.25 |
| Waiting Time Rate Every 45 seconds of waiting or less | $0.22 | $0.22 | $0.22 | $0.22 | $0.22 | $0.25 |
| **BOOKING FEES** | | | | | | |
| Peak period - Mon-Fri except PH (6am to 9.29am) - Mon-Sun & PH (6pm to 11.59pm) | $3.30 | $3.30 | $3.30 | $4.50 | $3.50 | $3.30 |
| Off-peak period | $2.30 | $2.30 | $2.30 | $2.50 | $2.50 | $2.30 |
| Advance (book at least 30min in advance) | $8.00 | $8.00 | $6.50 | $8.00 | $8.00 | $8.00 |
| **LOCATION SURCHARGES** | | | | | | |
| Tanah Merah Ferry Terminal | - | $3.00 | - | $3.00 | $3.00 | $3.00 |
| Marina Bay Sands Sun & PH (6am-4.59pm) | - | $3.00 | - | $3.00 | $3.00 | $3.00 |
| Singapore Expo Centre | $2.00 | | | | | |
| 1) City Area (5pm-11.59pm) 2) Seletar Airport 3) Resorts World Sentosa 4) Gardens By The Bay | $3.00 | | | | | |
| Changi Airport (a) Fri-Sun (5pm-11.59pm) (b) All other times Marina Bay Cruise Centre (a) Daily 7am-10.59am (b) All other times | (a) $5.00 (b) $3.00 | | | | | |
| **TIME-BASED SURCHARGES** | | | | | | |
| Peak Period Surcharge Mon-Fri except PH (6am-9.29am) Mon-Sun & PH (6pm-11.59pm) | 25% of metered fare | | | | | |
| Late Night Hiring Surcharge (daily 12 midnight to 5.59am) | 50% of metered fare | | | | | |
| **PAYMENT SURCHARGES** | | | | | | |
| Credit/charge cards | 10% administrative charge on top of metered fare (exclusive of GST) | | | | | |
| NETS, EZ-link | $0.30 administrative charge on top of metered fare | | | | | |

**Fig. 2.** Taxi costs for major taxi companies in Singapore as of August 2018.

Compared to the taxi prices, the fares of ride-hailing service Grab (whose flagship service is called JustGrab) are set dynamically by the ride-hailing app's algorithm to equilibrate the supply and demand of ride-hailing rides. The price is based on a multiplicative surge factor — the ratio of the ride-hailing platform price to a standard taxi price.[11] When Grab demand is high relative to supply, the surge factor is greater than one, meaning that ride-hailing prices are higher than the taxi prices. On the other hand, when Grab supply exceeds demand, the surge factor is equal to or less than one, meaning that ride-hailing prices can be lower than standard taxis. Apart from the surge factor, Grab does have no other surcharges in our sample period. Like taxis, private car drivers are required by the LTA to obtain a license, with the ride-hailing license costing S$400. If they do not comply with this rule, they face a S$10,000 fine. In addition, ride-hailing companies may have further required training. In the remaining of our paper, we use the term "Grab" and "ride-hailing" interchangeably as it is the dominant ride-hailing service in Singapore at this time. In addition, we use the term "taxis" to denote the taxis driving for the operator of interest that does not allow their drivers to participate on Grab (as opposed to "taxis participating in Grab").

The Grab service includes all but two taxi operators[12] as well as private-hire cars.[13] Grab taxis are treated the same as Grab private cars in that (1) when a customer books a trip, a driver is allocated based on the nearest distance, (2) they get paid according to the fare shown in the Grab app which is inclusive of a surge price, (3) customers rate them at the end of the trip, (4) they both

[11] See https://web.archive.org/web/20200924014047/https://www.grab.com/sg/blog/news/askgrab-pricing/ (archived on September 24, 2020).

[12] Both these two operators are owned by ComfortDelGro, which is the largest taxi company in Singapore with close to 60% combined taxi market share. As ComfortDelGro operated an advanced taxi booking service and considered Grab its direct competitor, it forbade its drivers to take jobs from Grab during our period of study.

[13] In April 2017, there were 41,297 chauffeur-driven private-hire cars, which was 56% more than the total taxi population of 26,476. See https://web.archive.org/web/20170525192809/https://www.straitstimes.com/singapore/transport/private-hire-cars-outnumber-taxis-by-a-mile (archived on May 25, 2017).

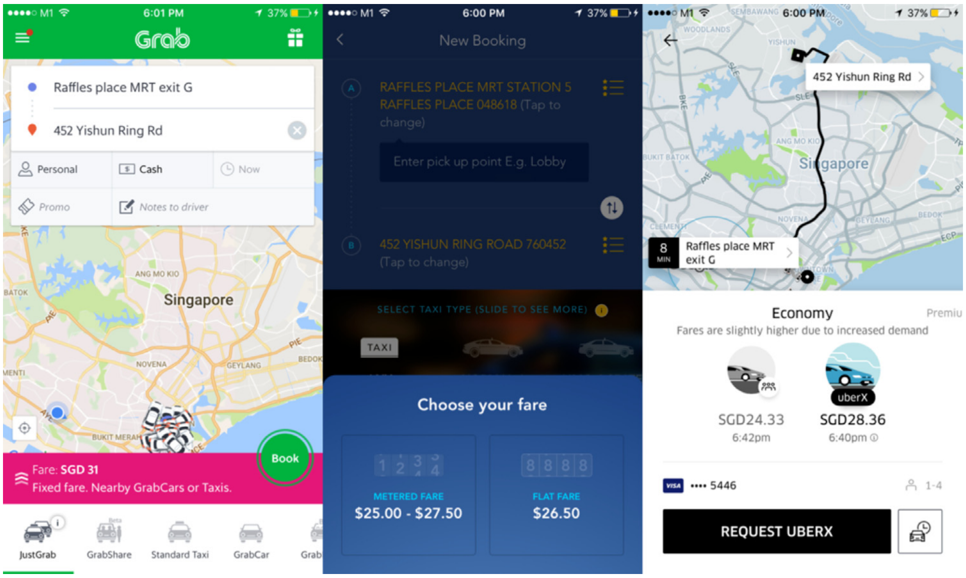



**Fig. 3.** The App interfaces from Grab, ComfortDelGro (the largest taxi fleet operator in Singapore), and Uber, at the time when the JustGrab service was launched.
*Source:* https://web.archive.org/web/20170411203615/https://www.todayonline.com/singapore/comfortdelgro-hopes-woo-riders-flat-fare-rates (archived on April 11, 2017).

are eligible for the same incentive schemes, which contain minimum driver rating as an eligibility criteria. Therefore, private cars and taxis participating in the Grab service are incentivized to have very similar quality.

The service quality for JustGrab service and taxi operators should be similar for the following four reasons. First, both are incentivized to maintain good driver acceptance rates. Taxi operators also have a complaint system which may result in banning drivers with low service quality. Second, Singapore is a small urban country, comparable to New York City or Los Angeles, so the distribution of cars is roughly even. Validating this premise, the 2018 Public Transport Council (a Singapore government agency) survey of point-to-point services, based on 1500 respondents, documents overall customer satisfaction of both taxis and private-hire cars at exactly 7.9 out of 10. In 2017, the averages were 7.9 for taxis and 8.2 for private-hire cars, respectively.[14] Third, the pricing of ride-hailing services in Singapore closely follows that of taxi services. In fact, Grab was founded in 2012 as GrabTaxi, a service exclusively for customers to book taxis on its platform. Because of this, the pricing of its services (in the absence of surge prices) tracks taxi prices closely. This is unlike many other cities where taxis are more expensive and less accessible than ride-hailing services. Finally, the App designs of the taxi operator of interest and ride-hailing platforms are very similar, making it easy for customers to switch if they want to (see Fig. 3 for snapshots of App UIs from Grab, ComfortDelGro, and Uber, right after the launch of JustGrab).

Overall, the Singapore setting provides a unique laboratory to study the impact of relative prices on customer behavior, absent quality differences between taxis and ride-hailing services that have been documented elsewhere in the literature (e.g., Brown and LaValle (2021), Rayle et al. (2016) and Schaller Consulting (2017)). In other markets where such differences exist, we expect lower estimated substitution, which could be confounded by quality and service differences rather than simply due to price. Further, both Grab's and the taxi company of interest's apps use Google maps, making their map accuracies highly comparable.

## 4. Data and methodology

### 4.1. Sample construction

The data used in this study can be broken down into those pertaining to prices and those pertaining to quantities. For prices, we derive location-specific surge factors from the price of the Grab service from selected location pairs based on information on the app, which is available to any potential customer using the Grab app. For quantities, we use two datasets on taxi rides that were either derived from or obtained directly from the LTA.

[14] See https://web.archive.org/web/20211129114754/https://www.ptc.gov.sg/docs/default-source/default-document-library/pcss-2018---point-to-point-services-continue-to-be-rated-highly-by-commuters.pdf (archived on November 29, 2021) and http://web.archive.org/web/20211129114904/https://www.ptc.gov.sg/docs/default-source/news/pcss-press-release-2017-(web).pdf (archived on November 29, 2021).

For the surge factor data collection, we choose representative origin–destination pairs that historically produced the most number of taxi booking requests with some manual adjustments to avoid similar locations being queried too frequently (e.g., the airport).[15] In total we have 90 origin–destination pairs. The query is run at each origin–destination pair at fixed intervals, between 30 to 60 min per pair. The locations are based on the postal sectors used in Singapore, which are the first two digits of the six-digit postal code administered by the Singapore Postal service.[16] Even though customers are only shown the price of the whole ride and not a multiplicative surge factor (unlike Uber, which shows both as of September 2019), we can calculate the baseline price for a typical taxi ride using the published pricing formula and extract the surge factors. This procedure yields a dataset of route-level surge factors at every thirty-minute interval throughout the whole sample period. Section 4.2 defines this construction more explicitly.

For taxis, the first dataset contains the number of taxi trips between an origin and destination region across all operators and trip types at every half-hour interval. We also observe whether the ride was initiated through a booking (either through the call center or app, 18% of rides), street pick-up (59%), third-party limo services or other means (23%).[17] For our analyses, we only use data on the two taxi operators that forbid their drivers from participating in the Grab service as the outcome variable (in the following sections, we refer to this outcome variable as simply the "taxi bookings"), to ensure that no Grab-specific features contaminate the analyses beyond surge factors and unobserved demand shocks. The dataset ranges from April to August 2017, with the exception of July, where the LTA experienced a database malfunction of several taxi operators in certain areas, resulting in severe loss of data. For data consistency, we drop July from our sample.[18]

The second dataset contains the mobility traces of all taxis in Singapore at every half-hour interval. In this dataset, we observe the anonymized taxi identifier, the operator of each taxi, whether the taxi is available or hired, and its location. These micro data allow us to split taxi supply into excess supply (taxis with the "Available" status) and hired supply (taxis with the "Hired" status) by each operator in each region within each half-hour interval. The supply of taxis is based on the operators that forbid their taxis from participating in the JustGrab service.

We merge surge factors to taxi rides and mobility traces by postal sectors and half-hour intervals. To merge the Grab surge factor dataset to the taxi trips dataset, for each half-hour interval, we aggregate surge factors by averaging the surge factor across all available postal codes into a postal sector. Lastly, to study whether taxi supply responds to surge factors and public signals, we also use rainfall data from the National Environmental Agency of Singapore as an exogenous shock in our analysis. The weather station locations in terms of latitude and longitude are merged into the nearest postal code, which is then merged to the corresponding postal sector. Fig. A.2 in the Appendix shows the locations of the weather stations.

## 4.2. Variable construction

The surge factor captures the relative price of ride-hailing services relative to standard taxi bookings and is defined as

$$\text{surge factor} = \frac{P_{Grab}}{P_{Taxi}},$$

exactly the relative price of Grab versus taxis. The taxi price incorporates all fixed cost surcharges as well as multiplicative surcharges.

Because the data spans multiple horizons across many days and narrowly defined regions in Singapore, we have a large heterogeneity across any half-hour intervals. In our sample, the average surge factor is 0.95, meaning that Grab is 5% cheaper than taxis on average. The standard deviation of the surge factor is 0.29, with a minimum observed surge factor of 0.67 and a maximum of 2.40, and around 24% of Grab rides have a surge factor above 1.00.[19]

Our raw taxi supply data contains the geographical region that a specific taxi was in and the number of minutes in each half-hour interval spent in either "Available", "Hired", "Busy", "Changing Shift", or "Other" statuses. Similar to Cramer and Krueger (2016), we construct the taxi supply in a region as the sum of all taxi-minutes in each status in half-hour intervals for each day, scaled by the 30-min interval as

$$\text{Taxi Supply}^{K}_{o,t,i} = \frac{1}{30}\sum_{j} h^{K}_{o,t,i,j},$$

where $K$ is the taxi status and $h^{K}_{o,t,i,j}$ is the number of minutes taxi $j$ was in status $K$ in region $o$ in a half-hour interval $i$ on day $t$. Taxis in the "Available" status are considered excess supply, those in "Hired" are considered hired supply, and those in "Busy",

[15] These locations are plotted in Fig. A.1 in Appendix. The choice of origin and destination regions are representative of the level of transportation and economic activity in Singapore, which is more concentrated towards the south of Singapore, in the Central Business District.

[16] The postal sectors in Singapore were introduced in 1995 to split the island into regions with roughly equal populations in each area for administrative and mailing purposes. Most of the regions are closer to the smaller range of 1-square mile rather than those up to 15-square miles. We adopt this region-based sampling approach because there are over 100,000 postal codes (one for each building), so using all permutations would be computationally prohibitive. In addition, we query the Grab system for surge factor data, so we do not want to submit too many queries, which can impose substantial costs on the app.

[17] In our analyses, we do not include any taxi rides from the latter group, which include third-party bookings, unknown and special pickups, because we do not know the exact reasons behind these classifications. The taxi operators do not have specific rules of how rides get classified into these groups.

[18] Including the few days of July that we have data for does not affect our main results.

[19] Note that the raw surge factors calculated are relative to basic taxi fares only and do not account for additional charges such as peak-hour or area-specific pricing. As such, the raw surge factors tend to overstate the actual surge price, which might introduce a negative bias on our estimate of the elasticity of substitution, and we use surge factors adjusted by the actual metered taxi fares. For more details on adjusting surge factors, please refer to Appendix A.1.

**Table 1**
Summary statistics.

| | Mean | SD | P25 | Median | P75 |
|---|---|---|---|---|---|
| *Taxi & Ride-hailing variables:* | | | | | |
| Surge factor | 0.947 | 0.286 | 0.800 | 0.800 | 0.968 |
| Taxi bookings | 2.60 | 5.44 | 1.00 | 3.00 | 5.00 |
| log(Taxi bookings) | 1.00 | 0.69 | 0.00 | 1.10 | 1.61 |
| Street pick ups | 9.62 | 15.3 | 3.00 | 7.00 | 13.00 |
| Taxi supply | 96.2 | 56.4 | 52.3 | 88.4 | 133.3 |
| Excess taxi supply | 41.1 | 29.9 | 18.6 | 34.0 | 56.4 |
| Scaled response time | 0.424 | 0.637 | 0.194 | 0.293 | 0.476 |
| *Environmental variables:* | | | | | |
| Weekend | 0.395 | 0.489 | 0.000 | 0.000 | 1.00 |
| Rainfall | 0.027 | 0.161 | 0.000 | 0.000 | 0.000 |

“Changing Shift”, or “Other” statuses are not considered in taxi supply. For the rest of our paper, unless explicitly qualified as either excess or hired supply, all references to taxi supply are the sum of excess and hired supply in a region in a half-hour interval. This means that just because a taxi is in a region but is hired still contributes to supply in that area. Specifically, it contributes to the used supply, but not excess supply. We think of each half-hour by geography as a market. For example, if the time interval from 9:00 am to 9:30 am in region A had two taxis that spent 12 min and 28 min respectively in the “Available” status, we calculate an excess taxi supply of 1.33 (= 40/30).

In Table 1 we provide the summary statistics of the variables used in our paper. All rows have 57,486 observations. Taxi bookings and street pick-ups are counts, taxi supply and excess taxi supply are defined in terms of car-minutes per half-hour interval, scaled response time is the fraction of the half-hour block for which the average customer waits (defined as the average number of minutes each ride has in response time divided by 30 min in the half-hour block), and weekend and rainfall are indicators taking the value of one if the date is a weekend and if there is rain in an area, respectively. All values are rounded to three significant digits or three decimal points, whichever is shorter.

### *4.3. Reduced-form empirical specification*

Before stating the empirical specification, we discuss the empirical strategy for our first analyses based on microeconomic fundamentals. We use a unique feature of most taxi markets: the relative price of taxi bookings to ride-hailing services is set in the ride-hailing services market, because taxi prices are set based on a fixed pricing schedule. However, the relative price of booking a taxi versus street pick-up does not change dynamically. We believe the main margin of substitution for Grab demand is taxi booking demand, given that those booking rides are typically more time constrained and delay-sensitive Taylor (2018).

Fig. 4 shows two figures summarizing the ride-booking market, showing the link between the Grab and taxi booking market, where the residual demand defined above is plotted against the surge factor. The left panel shows how the surge factor is determined to clear the Grab market, where supply equals demand. Fig. A.3 shows the cyclicality in the surge factor throughout the day that arises from the cyclicality from supply and/or demand. The right panel shows that the taxi market is not necessarily in equilibrium since nominal prices of taxis do not respond to market conditions. Although the surge factor does not impact the prices that the taxi drivers receive, and hence should not affect supply, the surge factor changes the relative price between Grab and taxi bookings. A higher surge factor will lead to more substitution between Grab and taxi bookings. Therefore, at any given time, the taxi booking market may have excess supply or excess demand. The supply of taxis are from the operators that do not participate on Grab and come from the same pool of taxis available for booking and street pick-ups.[20]

A simple regression of taxi bookings on the surge factor captures (1) the cross-price elasticity between Grab and taxi bookings, (2) confounding demand shocks that simultaneously affect the demand of Grab and taxi bookings, and/or (3) the response of taxi supply to surge factors. The first effect is presumably positive, as consumers should substitute away from Grab towards taxi bookings. The second effect is positive because common shocks affect Grab and taxi bookings are likely positively correlated. The third effect can be ambiguous. If taxi supply increases when surge factors rise then the effect is negative. If taxi supply falls when surge factors rise then the effect is positive. Therefore, in order to establish a proper upper bound on the cross-price elasticity between Grab and taxi bookings, we must know how taxi supply correlates with surge factors. Fortunately, our data allows us to study this latter question directly.

We use this approach because our goal is to statistically describe the relation between surge factors and taxi bookings. So long as we are able to establish that our estimation procedure produces an upper bound both due to the microeconomic setting as well as the Singapore-specific features of the Grab and taxi market, it can still serve a useful purpose for policymakers and future research. More importantly, to quantify the usefulness of the surge factor on the transportation network efficiency, we are agnostic as to whether the surge factor predicts taxi booking demand through each of the factors above.

[20] In Table A.3 in the Appendix, we study the impact of surge prices on street pick-ups and find some evidence that street pick-ups decrease in response to a higher surge factor, as more taxis choose to respond to bookings due the booking fee. Also, taxi supply does not differ based on the destination because in Singapore, street-hailed taxis are not permitted to accept or refuse a ride based on the destination when they are available. Such practice is a finable offense.

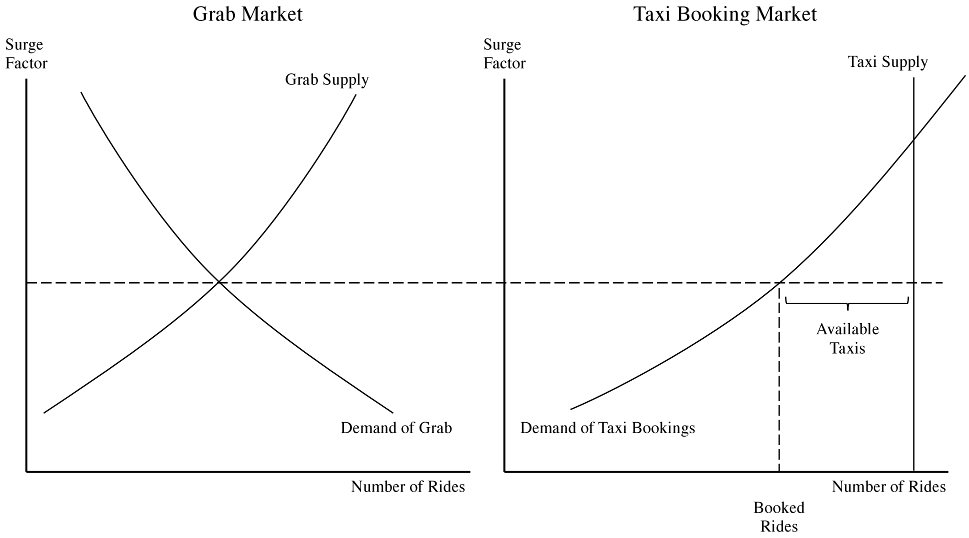



**Fig. 4.** These figures schematically show the interplay between Grab and taxi booking markets. The vertical supply for taxi bookings in the right graph is an assumption of non-binding supply on responses to consumer demand which we verify in Table 4. A similar schema has been used in Button (2020). We note that we do not require the Surge Factor be set to actually equilibrate the Grab market, but draw it so for illustrative purposes. For our empirical analyses, we simply need that the surge factors respond dynamically to the relative Grab supply to demand more so than the fixed-fare taxi prices do.

However, to the extent that we are able to control for demand factors, our estimate gets close to reflecting the cross-price elasticity. We believe our unique setting yields an informative upper-bound estimate of the cross-price elasticity for two reasons. First, the taxi operators we consider do not permit their taxis to participate with Grab and therefore should not respond to the surge factor (we test this assumption explicitly in Section 5.1). To further alleviate the concern that perhaps taxi supply reacts to surge factors through other unobservables, we explicitly control for taxi supply in a region at each point in time. Second, our high frequency taxi data permits a specification that includes high-dimensional fixed effects, controlling for origin region by time interval by day-of-week fixed effects to account factors affecting both supply and demand. Given that the postal sectors used in our study are around 1 to 15 square-mile blocks that were designed to split Singapore into regions with roughly equal portions of people, we believe the fixed effects do not distort our analysis. Any remaining shocks to demand would introduce an upward bias in the estimated cross-price elasticity, as the demand of taxi bookings and ride-hailing services would be positively correlated.

Therefore, our high frequency data on taxi rides, ride-hailing surge factors, and full market-wide taxi supply permits an upper bound estimation of cross-price elasticity of demand controlling for a myriad of potential shocks to demand and explicitly controlling for changes in taxi supply. The large cross section and high frequency of the data allow us to isolate local market factors both in terms of space and time using the following specification:

$$\log \text{Taxi Bookings}_{o,t,i} = \alpha_{o,day(t),i} + \gamma_t + \beta \text{Surge Factor}_{o,t,i} + \phi \log \text{Taxi Supply}_{o,t,i} + \varepsilon_{o,t,i} \quad , \tag{1}$$

where $o$ indexes an origin region, $t$ indexes a date, $i$ denotes a half-hour interval, and $day(t)$ is the day-of-the week of a given date. Routes, day of week, and half-hour interval fixed effects allow us to compare rides in the same route across days at exactly the same half-hour interval. The outcome variable is the log number of taxi bookings through a phone booking or the taxi company's mobile application. We do not include street pick-ups, which may represent a different market for rides, because we assume that the main margin of substitution is between Grab and booking taxis.[21] Any remaining omitted demand shifters that affect both Grab ride-hailing demand and taxi-booking demand would induce a positive bias in our estimates, meaning that our estimates would be an upper bound on the true cross-price elasticity. Fig. A.4 in the Appendix shows surge factors are autocorrelated, so we cluster standard errors by starting region as well as time to allow for cross-sectional correlations of surge factors, demand, and supply as well.[22] In addition to the total number of rides, we also study the relation between response times and the surge factor, which provides more insight into whether the responses of taxi bookings may be constrained by taxi supply.

However, if taxi supply is unable to accommodate for any substitution between ride-hailing and taxis, then the estimated relation between taxi bookings and surge factors would be biased downwards. In our follow-up analysis, we study the relation between surge factors and taxi supply using the following specification:

$$\log \text{Taxi Supply}^{K}_{o,t,i} = \alpha_{o,day(t),timeofday(t),i} + \beta \text{Surge Factor}_{o,t,i} + \varepsilon_{o,t,i} \quad , \tag{2}$$

[21] Although the surge factor is determined only based on the ride-hailing supply and demand at the origin and not taxi supply, our empirical specification also accounts for any potential reverse causality. In fact, comparing the point estimate using a specification with and one without the taxi supply controls is indicative of these potential endogeneities (Table 2 columns (1) and (2)).

[22] Related to this, Cachon et al. (2017) study how surge pricing affects the self-scheduling for rides.

**Table 2**
Surge factors and taxi bookings. This table shows the responses of the log number of taxi bookings and the log scaled response time in response to the level of the surge factor explicitly controlling for both taxi supply. Observations are at the half-hour by start-region level. All regressions include region by time interval by day-of-week fixed effects. Standard errors (in parentheses) are clustered by region and half-hour interval.

| $y_{o,t,i}$ = | log Taxi Bookings | | log Scaled Response Time | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Surge Factor | 0.257*** | 0.256*** | −0.160*** | −0.147*** |
| | (0.027) | (0.027) | (0.025) | (0.022) |
| Taxi Supply | 0.008 | | −0.293** | |
| | (0.053) | | (0.113) | |
| Observations | 57,486 | 57,486 | 57,429 | 57,429 |
| $R^2$ | 0.525 | 0.525 | 0.449 | 0.433 |

*p < 0.1; **p < 0.05; ***p < 0.01.

where $K$ is the taxi status (available or hired) in region $o$ in a half-hour interval $i$ on day $t$. We use taxi supply, which comprises both taxis that are available and taxis that are hired (taxi supply is the sum of excess supply and hired supply). We later also break down supply into excess supply and hired supply, and study both separately to see which responds to the surge factor.

## 5. Ride-hailing surge factors and taxi bookings

Table 2 shows our empirical estimates of the impact of surge factor on taxi bookings when explicitly controlling for taxi supply within the same half-hour interval and the same geographical region. Column (1) suggests that a 10% increase in surge factors, holding constant the local taxi supply, leads to an increase of around 2.6% in taxi bookings within the same half-hour interval. In terms of economic significance, a one-standard-deviation increase of 0.29 in the surge factor corresponds to a 0.25-standard-deviation increase in log taxi bookings (relative to a standard deviation of 0.69). The point estimate on surge factors does not change from Column (1) to (2), which is consistent with the surge factor not being affected by taxi supply.

Columns (3) and (4) show the relation between surge factors and the scaled response times, which is defined as the average fraction of a 30-min interval that a booked taxi spends to pick up the customer who made a booking (i.e., the average number of minutes each ride has in response time divided by 30 min in the half-hour block). In both Columns (3) and (4), we find that a 10% higher surge factor predicts a slightly lower response time of around 1 min (=1.6% × 30 min). In terms of economic significance, a one-standard deviation increase of 0.29 in the surge factor corresponds to a 0.07-standard-deviation decrease in waiting times (relative to a standard deviation of 0.64). Although seemingly counterintuitive, the findings are consistent with an unconstrained taxi supply in an interval-region that picks up slack from the Grab market, since surge factors correspond to high Grab demand periods. Such unconstrained taxi supply can respond to bookings more quickly as more riders substitute to taxi bookings. In Column (3), we also find that a higher taxi supply corresponds to a decrease in response time, as there are more available cars to respond to bookings in a more time-efficient manner.

Additional analyses in the Appendix show our results are robust to various specifications, alternative variable constructions, and different subsamples. Table A.2 shows that the results are robust to excluding the weekend, and falsification tests drawing from the surge factor or same route do not mechanically generate our results. Moreover, Fig. A.5 shows that the point estimates of the elasticity are stable across months from April through August. We also find a negative relation between taxi street pickups and surge factors, likely because taxi drivers prefer booked rides rather than street pick-ups due to the additional booking fee that they receive. Although taxi drivers are legally not permitted to decline street-hailing customers based on their destination, they may ignore street pick-ups. Taxi drivers can easily switch from street pick-ups to bookings. The latter only offer a fixed pricing schedule and taxi drivers earn more when accepting bookings compared to street pick-ups due to the additional booking fee on average.[23]

In the following subsections, we first show that the estimated upper bound on the cross-price elasticity does not seem to be driven by constrained taxi supply. Then, we show that taxi supply does not respond to changes in the surge factor. Jointly, these follow-on analyses rule out alternatives due to simultaneous or reverse causality. Appendix B shows additional robustness and cross-sectional heterogeneity in responses. Appendix B.1 studies the impact of demand shocks through time and the impact of journey distance on the cross-price elasticity. Appendix B.2 studies whether the elasticity of substitution differs based on trip distance and Appendix B.3 studies whether the elasticity is affected by local area income.

[23] Unfortunately, the street pick-up specification is not a good placebo for three reasons. First, the estimated relation between surge factor and street pick-ups may be affected by taxi drivers switching from supplying street pick-ups to taxi bookings. Second, a non-trivial fraction of pick-ups also include rides by tourists and others that may not have a smartphone or the Grab app, who are likely to be unaware of switching between street pick-ups and Grab. Third, unlike street pick-ups, the taxi booking app and Grab app both show the trip price at the time of booking whereas customers only know the price of a street pick-up after completion, making the two services less substitutable due to the uncertainty component.

**Table 3**
Elasticity depending on excess taxi supply. This table shows the response of the log number of available taxis in response depending on the previous half-hour's taxi excess supply. The high surge sample is a subset of data with above median surge factor value, and the projected sample comes from instrumenting for surge factors with the excess supply in each origin destination in each half-hour interval. High excess taxi supply is defined relative to the median and based on the previous half-hour interval in the same region. The observations are at the starting region by half-hour interval level. All regressions include region by time interval by day-of-week fixed effects. Standard errors (in parentheses) are clustered by region and half-hour interval.

| | $y_{o,d,t,i} = \log\left(\text{Taxi Bookings}_{o,d,t,i}\right)$ | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Surge Factor$_t$ | 0.240*** | 0.097** | 0.260*** | 0.213*** |
| | (0.029) | (0.038) | (0.036) | (0.034) |
| High Excess Taxi Supply$_{t-1}$ | −0.051 | −0.041 | −0.056 | −0.043 |
| | (0.032) | (0.080) | (0.044) | (0.047) |
| Surge factor$_t$ × High Excess Taxi Supply$_{t-1}$ | −0.011 | −0.063 | −0.012 | −0.014 |
| | (0.031) | (0.059) | (0.041) | (0.047) |
| Taxi Supply$_t$ | 0.017 | −0.100 | 0.091 | −0.083* |
| | (0.045) | (0.071) | (0.062) | (0.049) |
| Sample | All | Surge > 1 | Weekday | Weekend |
| Observations | 57,474 | 12,447 | 34,766 | 22,708 |
| $R^2$ | 0.527 | 0.668 | 0.535 | 0.516 |

*p < 0.1; **p < 0.05; ***p < 0.01.

### 5.1. Is taxi supply constraining consumer substitution?

In this subsection, we study whether the availability of taxis restricts the consumers' ability to substitute between ride-hailing and taxi booking. We use a specification based on Eq. (1) that also interacts the surge factor with excess taxi supply, defined based on the number of taxis in the "Available" status in the previous half-hour interval. If the substitution is constrained by taxi supply, we would expect the interaction term to impact the substitution effect that we estimated before. Specifically, we would expect a positive and statistically significant interaction term between surge factor and high excess taxi supply.

Table 3 shows that the interaction of the excess taxi supply in the previous half-hour interval with the surge factor is not statistically significant, even when subsetting based on the surge factors above one, weekdays, and weekends. The excess taxi supply is as an indicator of whether the taxi supply is greater than the median number of taxi supply in that same region. Therefore, our results suggest that the estimated upper bound on the cross-price elasticity is not downward biased due to a lack of available taxis to substitute to when surge factors are high.

### 5.2. Does taxi supply respond to surge factors?

In this subsection, we directly test whether taxi supply is correlated with surge factors. There are two main possibilities on the relation of surge factors and taxi supply. First, if non-participating taxi drivers observe ride-hailing surge factors, they could gauge areas with high demand and increase taxi supply in the regions and times of rising surge factors. Second, the surge factor algorithm may increase prices when taxi supply is low. Since these two possibilities counteract, the overall result can be in either direction. However, we believe that the latter case is not possible, as to our knowledge (based on conversations with LTA), Grab does not have a live data feed of the taxi drivers from the two operators that are not participating on Grab. Therefore, we study if the taxi supply responds to surge factors without considering the latter reverse causality.

In this specification, we study how taxi supply responds to surge factors as well as rainfall, an exogenous variable that can affect both ride booking demand and supply. Taxi drivers are unable to observe surge factors but drivers near a rainy area can observe the rain and know whether an area has high demand. In fact, because rain is market-relevant and common information among all the drivers in the area, it may affect the allocation of taxi supply. On one hand, taxi supply may increase in anticipation of a predicted increase in demand in a rainy area. On the other hand, rain may also raise concerns that road conditions are more dangerous or congested.

Column (1) of Table 4 shows that taxi supply (which is the sum of the number of "Available" and "Hired" taxis) increases by 13% during rainy periods, but it does not appear to respond to surge factors within a region in a specific half-hour interval. Columns (2) and (3) show a higher surge factor is associated with more hired taxis and a direct reduction of excess supply of taxis, which is the number of taxis in the "Available" status. As consumers switch from Grab to taxis, the excess supply of taxis in a region appears to completely absorb the substitution.

Therefore, in a half-hour interval, higher demand matches with the available taxis, but the taxi supply (the sum of available and hired taxis) in the area does not change. Together with results from Tables 2 and 3, we conclude there appears to be a pool of available taxis that absorbs the entire increase in demand in a given half-hour interval.

Finally, we note that the supply of taxis available come from the same pool of taxis as those available for street pick-ups. In Table A.3, we show that street pick-ups decrease as surge factors rise. That surge factors have no relation with taxi supply but is associated with an increase in taxi bookings and a small decrease in street pick-ups suggest the substitution of taxi supply from

**Table 4**
Taxi supply and surge factors. This table shows the response of the log taxi supply in response to the level of the surge factor. 1{Rain} is an indicator taking the value of one when there is any rain detected by the weather stations in the region. Rainfall is measured in centimeters. All regressions include region by time interval by day-of-week fixed effects. Standard errors (in parentheses) are clustered by region and half-hour interval.

| $y_{o,t,i}$ = | log Taxi Supply | log Excess Taxi Supply | log Hired Taxi Supply |
|---|---|---|---|
| | (1) | (2) | (3) |
| Surge Factor | −0.024 | $-0.448^{***}$ | $0.212^{***}$ |
| | (0.043) | (0.039) | (0.041) |
| 1{Rain} | $0.133^{***}$ | $0.044^{***}$ | $0.142^{***}$ |
| | (0.035) | (0.030) | (0.034) |
| log(1+Rainfall) | −0.003 | −0.034 | 0.016 |
| | (0.025) | (0.020) | (0.027) |
| Observations | 57,427 | 57,427 | 57,427 |
| $R^2$ | 0.339 | 0.459 | 0.375 |

$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

street pick-ups to bookings. This is consistent with the fact that all else equal, a taxi would prefer a booking due to the additional booking fee of S$3 to S$5 that they receive.[24]

## 6. Demand prediction and structural analyses

The previous section estimated an upper bound on the impact of ride-hailing surge prices on taxi bookings. A back-of-the-envelope calculation using the estimated bound, multiplied by the changes in the surge price and the taxi fare, suggests that up to 18% of the current revenues from taxi bookings are due to this substitution.[25] However, the implication of this effect on taxi booking prediction in a statistical sense is not obvious, since the cross-price demand elasticity is fairly small. To tackle this problem head-on, this section illustrates how much the surge factor improves the predictive accuracy of taxi bookings. This prediction problem is of interest to policy makers as well as taxi companies that seek to minimize congestion and to improve allocation of taxi supply.

### *6.1. Predicting taxi bookings with machine learning*

To allow flexibility in the interactions of variables, we use a random forest model based on the data that is available to policy makers (the LTA in this case). The goal of the random forest prediction exercise is to evaluate the marginal improvement when including the surge factor predicting taxi bookings relative to a model only relying on only 10 environmental variables. The five categorical variables include the month, day of the month, day of the week, starting region, ending region, and half-hour interval. We include these as taxi operators typically know the average rush hour, and which areas have high demand at different times of the day, and revealed preference indeed shows that static taxi surcharges incorporate such information as discussed in Section 3. The numeric variables include the surge factor, distance in kilometers, calendar month over the sample, which represents the summer period, and rainfall, which we discussed above. We use the randomForest package in R to estimate a forest with 500 trees, a maximum depth of 5, a maximum of 3 variables to try at each node, and minimum sample split of 5.[26] Since our samples are representative across different categories of start regions, end regions, and time intervals, we use a simple random sample method for each out-of-bag sample used to evaluate each tree.

We consider two sampling methods of our data in this prediction exercise. In the first sample, we leave out August 2017 as the test data and use data from April through June 2017 as the training sample. The second sample is built by randomly sampling 75% of the observations in each of the four months for the training data and using the remaining 25% of the data as the testing data. Although the first sample closely approximates the real world datasets used in predicting taxi bookings by policy makers and taxi and ride-hailing companies, comparing the results of these two samples provides a robustness check for the stability of the model over our full data period.

[24] Although this does not pose a problem for the Grab to taxi booking margin of adjustment which we study, it would be problematic for other research studying the substitution between ride-hailing services and taxi street-hails. Because taxis may ignore street-hails in favor of bookings, the supply of taxis willing to service street hails may fall, therefore introducing a downward bias in the estimate of the substitution between Grab and street-hailing. This imposes a negative non-pecuniary externality on customers who use street pick-ups. Therefore, studying the economy-wide impact of ride-hailing and overall customer welfare is complex as some consumers benefit from Grab substitution while others are harmed. As a result, for the remainder of the paper, we focus on the perspective of the taxi operator which unambiguously benefits from taking more bookings even if it crowds out street hails due to the additional booking fee.

[25] The calculation is as follows. First, we calculate counterfactual total taxi fares based on route-times if there were zero substitution. To do so, we sum up all the fares from a particular route in that time. Then we assume the estimated 0.26 upper bound on the cross-price elasticity applies equally to the number of trips as well as dollars and multiply any deviation from surge factors above one by the elasticity. We then re-scale the fares in a particular route-trip down by that amount. For example, if a route-trip in a half-hour period had total fares of S$10,000 and the surge factor was 1.5, we assume up to $0.5 \times 0.26 = 13\%$ of the fares came from the substitution. Finally, we sum the counterfactual fares in dollars and divide it by the sum of the actual realized fares.

[26] Fig. A.9 in the Appendix shows the sensitivity of in-sample root-mean squared errors based on the number of trees used.

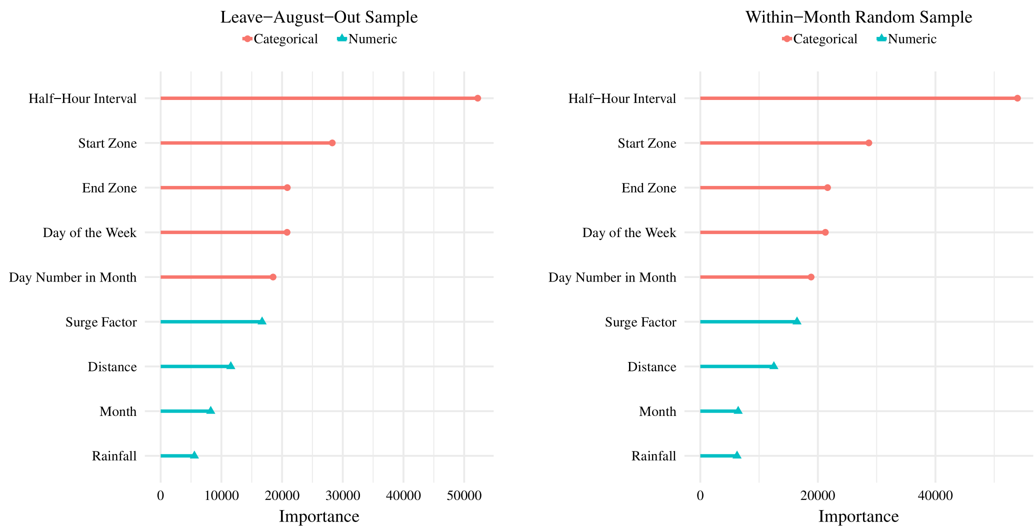



**Fig. 5.** Variable Importance of Taxi Bookings Predictive Model. These figures show the variable importance ranking based on node impurity for each of our two samples. The left panel is for the sample excluding August and the right panel is for the within-month 75% random sample.

**Table 5**
Accuracy of taxi bookings predictive model. This table shows the accuracy of the random forest model in predicting the number of taxi bookings with different samples (both in-sample and out-of-sample). RMSE stands for root-mean-squared error.

| Model test sample | In-sample RMSE | | |
|---|---|---|---|
| | No surge factor | With surge factor | Improvement (%) |
| August | 1.675 | 1.664 | 0.661 |
| 20% of each month | 1.732 | 1.726 | 0.348 |
| **Model test sample** | **Out-of-sample RMSE** | | |
| | No surge factor | With surge factor | Improvement (%) |
| August | 1.286 | 1.089 | 15.319 |
| 20% of each month | 1.488 | 1.306 | 12.231 |

Fig. 5 shows the variable importance plot for both the samples, splitting variables into categorical and numerical variables by color. In both, the surge factor is the 6th most important variable, and the most important among the numerical variables. The most important variables, expectedly, are the time, day of week, and destination variables. In the training sample using all the data before August, the proportion of variance captured by the random forest model is 49.62% for the model with the surge factor and 48.98% without. Likewise, the in-sample improvement using the random subsample is also small from 47.51% without the surge factor to 47.89% with the surge factor. Thus, in this $R^2$ sense, the in-sample improvement due to the surge factor is small.

In terms of root-mean-squared error, incorporating surge factor information increases out-of-sample forecast accuracy by 12% to 15%. Table 5 shows the in-sample and out-of-sample prediction accuracies of the random forest model for both the leave-August-out and within-month random samples. For the leave-August-out sample, we see that although including the surge factor improves the in-sample accuracy by less than 1%, it raises the out-of-sample accuracy by over 15%. Likewise with the within-month random sample, the model with surge factor data improves the in-sample accuracy by less than 0.5%, but the out-of-sample accuracy rises by over 12%. While the performance is similar with or without the surge factor, the factor substantially raises the out-of-sample prediction accuracy by eliminating systematic errors in the forecasts. Our out-of-sample results suggest that the surge factor helps policy makers as well as taxi and ride-hailing companies in forecasting the future taxi bookings.

Although in practice one could query the Grab surge factor in real time, for robustness, we also consider including the half-hour lagged surge factor. Table A.6 shows this specification, documenting similar out-of-sample performances in the root mean-squared error of predicted taxi bookings of 11.5% and 15% for the out-of-sample periods corresponding to August or a random sample of 20% of each month respectively. This similar out-of-sample performance is likely due to the autocorrelation in the surge factor (see Fig. A.4), where a persistent and slowly-decaying mismatch between Grab supply and demand makes the autocorrelated component of the surge factor remains highly relevant for predicting taxi bookings.

However, having better taxi demand prediction itself does not directly contribute to the increase in social welfare. In the section below, we operationalize the demand prediction information to study impact on social welfare by providing guidance to taxi drivers. Such guidance should take into account both the predicted demand occurrence in different parts of the cities and the current and predicted future locations of vacant taxis, and generate a suggested region for the driver to go to in the upcoming time period.

### 6.2. Structural analyses and simulation

To shed light on how the surge price information impacts social welfare, we focus on the reduction of congestion and taxi vacancy times. A more accurate taxi booking demand prediction could help policy makers and ride-hailing and taxi companies to anticipate demand better and reduce congestion. Given a taxi fleet size, decreasing congestion and taxi vacancy times increases both consumer welfare and driver welfare. Therefore, by focusing only on the latter, we will underestimate the impact of the surge factor information on social welfare.

For implementation, the policy makers or companies have to collect both the predicted demands and the whereabouts of vacant taxis, and compute policies on how to position vacant taxis to optimize spatiotemporal demand–supply matching. An academic prototype of this idea, the Driver Guidance System (DGS) (Jha et al., 2018), has been developed and its field trial results with around 500 taxi drivers show that by following centrally generated recommendations, taxi drivers experience, on average, 28.9% reduction in their vacant roaming time from January to May 2018 (Cheng et al., 2018). Further, if the drivers hypothetically follow the DGS guidance all the time, the number of fetched trips could rise by around 10.5%.

#### 6.2.1. Stylized model for taxi fleet coordination

To illustrate the information needed for the implementation, we present a stylized model assuming all taxi trips have the same distance and fare, and taxi supply and demand are balanced. The locations of the demand quantities are random, but could be inferred by observing the surge prices at different locations. This stylized coordination problem for the taxi company can be written as follows:

$$\max_{m_1,m_2,\ldots,m_n} \sum_{i=1}^{n} E\left[\min\{m_i, D_i\}\right] \quad \text{such that} \quad \sum_{i=1}^{n} m_i = m, \tag{3}$$

where $n$ is the number of locations, $m$ is the number of taxis that equals the realized total residual taxi demand quantity, and $m_i$ and $D_i$ are the number of taxis and demand quantity respectively. Let us denote surge price in location $i$ by $s_i$.

Without the surge price information, $D_i$ is a random variable (but the aggregate demand quantity is fixed and known). Therefore, in this case, the probability for all the customers to find matching taxis is less than one, i.e., $Pr(\text{number of rides} = m) < 1$. However, if the surge price information is incorporated, $D_i$ is inferred and, thus, is no longer a random variable. In other words, instead of having $E[\min\{m_i, D_i\}|s_i]$ in the objective function above, we have $\min\{m_i, D_i\}$. Therefore, with the surge price information, the optimal allocation of taxis is simply $m_i = D_i$ for all $i \in \{1, 2, \ldots, n\}$, which gives that the number of rides equals $m$. In this case, all the rides are matched, which maximizes the welfare of consumers within the ride-hailing and taxi-booking market.[27]

Assuming that a taxi driver's objective is revenue maximizing, where the related customer welfare is measured in terms of the numbers of ride-hailing and booked taxi rides, we also see that the impact of taxi companies using surge factors on Grab cars is minimal, as surge factors were already determined due to the matching of Grab supply and demand. A high surge factor implies high Grab demand relative to supply, suggesting that Grab cars will not be able to service all the demand. Of course, this analysis is single-period and does not consider, for instance, the travel time experienced by taxis. To extend this stylized analysis to a more realistic setting, we conduct detailed agent-level simulations, where we evaluate how the improvement of taxi booking demand prediction would positively impact close-to-reality taxi fleet operations.

#### 6.2.2. Simulating the impact of demand prediction accuracy on taxi fleet coordination

To incorporate taxi booking demand prediction and observe how it affects taxi driver's decision making process, we adopt a DGS following Jha et al. (2018), and use a realistic microscopic taxi operation simulation (Cheng and Nguyen, 2011) to evaluate the resulting gains in the social welfare (computed as all taxis' income) if demand predictions were to be improved. Our simulation is microscopic and agent-based, where each taxi is modeled as an agent; we also model the service modes of street hails, taxi queues, and booking. As the simulation runs, it also generates a data stream in real-time to feed to the demand–supply matching engine that computes taxi-specific movement recommendations (based on the demand predictions using the simulated data stream). In Fig. 6, we provide a screenshot of the taxi operation simulation, which illustrates two main features: a) the whole city is modeled as a network of road links (our map topology is obtained from the OpenStreetMap, which is open-source and provides link-level details), and b) our simulation is agent-based, where each agent (taxi) has its own state (represented by different color, capturing the state of "vacant" and "hired"); as noted later, we assume that taxi agents are following region-level guidance provided by the DGS engine, however, their link-level movement patterns are stochastic, following distribution derived from the historical data. When a taxi picks up a passenger, we assume that it will travel to the destination following the shortest path.

By introducing a parameter that deliberately controls the demand prediction accuracy, we then observe the impact of improving demand prediction on the guidance effectiveness. As the computation of driver guidance depends on the demand prediction as well, we conduct a simulation study to test how much the improvement in the demand prediction raises the effectiveness of the spatiotemporal demand–supply matching. In other words, this counterfactual simulation uses the existing simulation from Jha et al. (2018) and exogenously improves the demand prediction accuracy as a way to gauge the economic impact.

[27] However, as noted before in Section 5.2, the overall consumer welfare in all related markets comprises those from riders of both ride-hailing services, taxi bookings, and taxi street-pick-ups. Since we document some negative spillover effects of the relation between ride-hailing and taxi bookings on street pick-ups, the overall welfare is difficult to quantify. Therefore, we take the objective to maximize the expected revenues of the drivers, and the welfare is measured in terms of the numbers of rides of ride-hailing services and booked taxis.

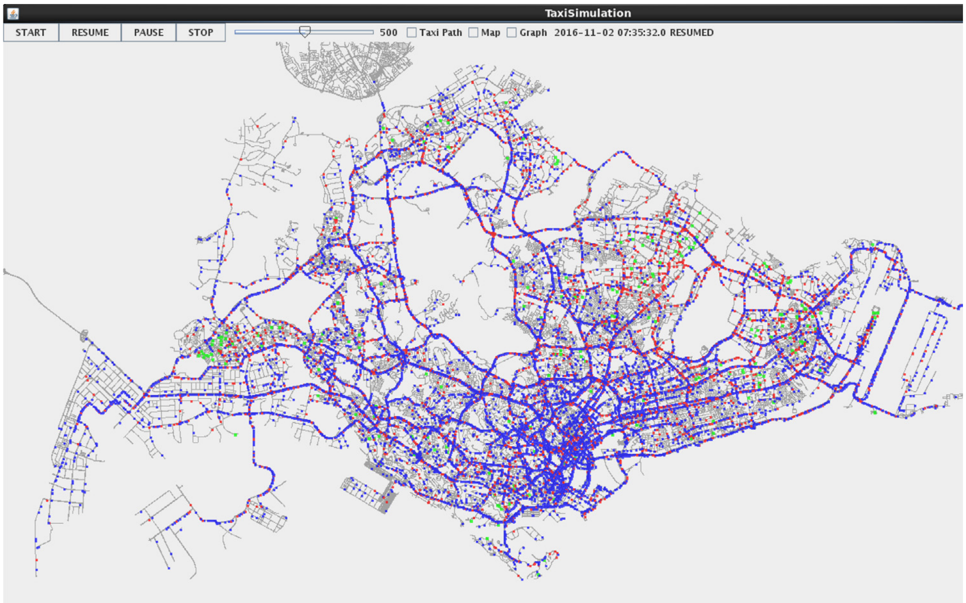



**Fig. 6.** A screenshot of the taxi operation simulation. Taxi agents are denoted as blue (vacant) or red (hired); passenger demands are denoted as green dots. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)


**Fig. 7.** Grid map for Singapore used in the simulation.

The reason why the performance of DGS depends on the accuracy of demand prediction is that there are three major components in the optimization formulation: the immediate movement cost, the expected future revenue, and the expected future movement cost. Better demand prediction should contribute to the estimation accuracy of the expected future revenue. We design the simulation to demonstrate the potential benefits of having better taxi demand prediction. For the simulation, we make the following assumptions:

1. We assume that all taxis in the simulation follow the generated guidance; according to the simulation results in Jha et al. (2018), having all taxis following the guidance generates greatest social welfare. This assumption allows us to estimate the upper bound on the gains that could result from demand prediction.
2. We use the Singapore map grid in Fig. 7 to setup the simulation. To provide appropriate granularity, we define the grid of 1 km-by-1 km to be the minimal geographic unit for demand and supply prediction, and the target of the recommendation.

**Table 6**
Summary of simulation results for $\alpha = 0$ and 0.15, in terms of vacant roaming time (in minutes). Besides mean and standard deviation, we also provide the quantiles of (min, 25%, 50%, 75%, max). The last row is the percentage improvement of $\alpha = 0.15$ over $\alpha = 0$, which demonstrates that better demand prediction is most effective in reducing extra long vacant time.

| $\alpha$ | Mean | Std | Min | Quantiles | | | Max |
|---|---|---|---|---|---|---|---|
| | | | | 25% | 50% | 75% | |
| 0 | 26.3 | 20.2 | 0.9 | 10.6 | 19.9 | 36.8 | 88.3 |
| 0.15 | 23.8 | 16.9 | 0.9 | 10.3 | 19.0 | 33.8 | 71.1 |
| | 9.4% | 16.2% | −0.3% | 2.8% | 4.8% | 8.0% | 19.4% |

Based on this grid definition, we train the demand prediction model then calculate realistic geographical features such as traveling distances and cost between grid regions.[28]

3. The demand pattern in the simulation is derived from historical weekdays. For each grid $i$, each time period $t$ (30 min), we calculate the Poisson arrival rate ($\lambda_{it}$) as the average number of trips occurring in grid $i$, at time $t$. During the actual simulation, Poisson arrivals are simulated in each grid region with the assumption that the arrival rate is $\lambda_{it}$. The demand predictions are generated every 5 min, for the horizon of 6 time periods (30 min). We pre-generate all demand arrivals before the simulation, and store them as the ground truth. During the simulation, the simulated demand arrivals are read from this pre-generated source. The demand predictions are computed on the fly during the simulation, using only realized and observable information. To improve the prediction accuracy, we compare the predicted demand against the pre-generated ground truth and adjust the predicted value so that its difference from the ground truth is improved by $100\alpha\%$. In other words, if $d_{it}$ is the actual demand (unknown to the driver), $\hat{d}_{it}$ is the predicted demand using a deep learning neural network model using only point-in-time observable variables, and the adjusted demand prediction is given by:

$$\hat{d}'_{it} = d_{it} + (\hat{d}_{it} - d_{it}) \times (1 - \alpha),$$

where the original Forecast Error $(\hat{d}_{it} - d_{it})$ is from the empirical distribution of actual forecast errors relative to the original demand prediction. This $(1 - \alpha)$ term corresponds to a decrease in root-mean-squared error of $100\alpha\%$ relative to the baseline model.[29]

4. The guidance from the DGS indicates on the grid region in which a taxi should stay in. The actual movements along the streets are decided by historical frequency: when a simulated taxi reaches a road intersection, the simulator queries the historical frequency on which road segment to turn, and stochastically decides which road segment the taxi should drive to. The constraint is that the choice at the street-level should ensure that the grid-level decision is maintained.

We compare the predicted demands against the corresponding demands. More specifically, we alter the predictions so that they are $100\alpha$ percent closer to the corresponding simulated demands ($\alpha$ is a parameter for the experiment, representing the desired improvement in demand prediction accuracy). As the simulated demand scenarios are all pre-generated, we conduct paired performance comparisons of different demand prediction accuracies. To reflect the improvement in demand prediction accuracies, we have compared the cases with $\alpha = 0$ (the baseline without improvement) and $\alpha = 0.15$ (the case where demand prediction accuracy is 15% better).

We generate 20 demand scenarios and summarize our findings in Table 6. For each demand scenario, we obtain the summary statistics on the vacant roaming time by averaging over all taxi agents (we have around 19,000 taxi agents in all 20 simulations). The summary statistics on the vacant roaming time we obtain from these 20 simulations all look very similar, probably due to the large number of agents per simulation instance. The detail breakdown of the aggregated simulation results are in Table 6; the percentage improvement of $\alpha = 0.15$ over $\alpha = 0$ is summarized in the last row. As illustrated, we can observe that while the reduction in mean vacant roaming time is 9.4%, the improvement is not uniform across the quantiles. The higher the quantile, the higher the reduction percentage (the maximum roaming time is reduced the most, by 19.4%).

To find out the probable real-world impact for a 9.4% reduction in the vacant roaming time, we extrapolate the real-world field trials reported in Cheng et al. (2018). The basic idea of the extrapolation is to extract the actual time components of a trip, which contains roaming, response (when a booked taxi is on the way to pick up a passenger), and service (from pickup to drop-off) times. We then compute and summarize the probable reduced roaming time for $\alpha = 0.15$ in Table 7. The number of trips per hour is computed by

$$\frac{1}{\text{Avg. Roaming Time} + \text{Avg. Response Time} + \text{Avg. Service Time (in hour)}}.$$

[28] We note that this grid definition is different from the definition used in the reduced-form analyses. The choice of a uniform grid is to ensure no bias is mechanically introduced due to varying region sizes. We note that the choice of grid size does impact the effectiveness of the DGS which we are simulating, as it captures a trade-off between demand aggregation (which favors larger regions) and the effectiveness of the recommendation (which favors smaller, more granular regions). However, this is beyond the scope of this paper.

[29] Relative to an original forecast error $var(\hat{d}_{it} - d_{it}) = \sigma^2$, the variance of the adjusted error term is $var((\hat{d}_{it} - d_{it}) \times (1 - \alpha)) = \sigma^2 \times (1 - \alpha)^2$. The relative root-mean-squared error is $\sqrt{(1-\alpha)^2\sigma^2}/\sqrt{\sigma^2} - 1 = -\alpha$, a decrease of $100\alpha\%$.

**Table 7**
The actual performance of taxi drivers who used the DGS against those who did not (data is collected from January 2018 to May 2018, when the DGS App is used most frequently; extracted from Cheng et al. (2018)). The column DGS ($\alpha = 0.15$) is our attempt to extrapolate DGS performance for $\alpha = 0.15$.

| | Non-DGS | DGS | DGS ($\alpha = 0.15$) |
|---|---|---|---|
| Avg. roaming time (min) | 9.44 | 6.71 | 6.08 |
| Avg. response time (min) | 1.82 | 1.82 | 1.82 |
| Avg. service time (min) | 16.35 | 16.46 | 16.46 |
| Trip/hour | 2.17 | 2.40 | 2.46 |

From the extrapolation in Table 7, we can see that with $\alpha = 0.15$, we could reduce roaming time against non-DGS by 35.6% and increase the number of trips per hour by 13.3%. For plain DGS, these improvements are 28.9% and 10.5% against non-DGS. Finally, by comparing DGS ($\alpha = 0.15$) against plain DGS, we observe that a 9.4% reduction in the vacant roaming time could lead to an increase of 2.6% in the number of trips per hour.

## 7. Conclusion

Our results show that incorporating relative prices in allocating taxi supply improves transportation network efficiency and level of service. Taxi drivers increase their earnings, reduce roaming time, thereby more customers have a fulfilled order. In addition, our reduced-form empirical analyses also show that customer waiting times fall and that they also save on trips on average compared to a counterfactual with zero cross-price elasticity. A back-of-the-envelope calculation suggests that up to 18% of taxi fares can be attributed to the cross-platform substitution.

Our findings are also informative for other markets. Even in a unique market like Singapore, where incumbent taxi companies also have mobile apps for bookings that rival the user experience of those from ride-hailing companies like Uber, we still only estimate an upper bound on the substitution between ride-hailing services and taxi bookings of 0.26. The estimated inelastic demand may be due to marketing tactics in maintaining brand loyalty, inattentive customers, or alternative mechanisms. Other markets where taxi service quality is lower without the ease of a smartphone app would presumably have an even lower cross-price elasticity, as larger differences in service quality imply farther substitutes. For example, if there is no smartphone app available to book taxis, and taxis and ride-hailing companies end up serving different geographies (as in Schaller Consulting (2017)), we would expect even lower substitution. We leave the analyses studying the factors driving the value of cross-price elasticity across different modal transits for future research.

## Acknowledgments

This research is supported in part by the Ministry of Education, Singapore, under its Social Science Research Thematic Grant (Grant Number MOE2020-SSRTG-018). In addition, Sumit Agarwal acknowledges the financial support from the National Natural Science Foundation of China (Grant Number 72150004) for this project.

## Appendix A. Data

Fig. A.1 shows the origins as circles and destinations as squares around Singapore that we sampled. The origin and destination regions are fairly representative of the level of transportation and economic activity in Singapore, which is more concentrated towards the south of Singapore, in the Central Business District.

### *A.1. Adjusting surge factors*

The adjusted surge factors show cyclicality through the day (see Fig. A.3). We adjust the surge factors as the raw surge factors are relative to basic taxi fares only and do not account for peak hour or area pricing on actual metered taxi fares. Therefore, the raw surge factors tend to overstate the actual surge price, which might introduce a negative bias on our estimate of the elasticity of substitution, and we use surge factors adjusted by the actual metered taxi fares.[30]

Since most taxi riders are aware of the surcharge schedule (both multiplicative and lump sum), as mentioned above, we adjust the raw surge factors to get the actual relative surge factor. For the multiplicative surcharges from 9:00 pm to 5:59 am, we scale the taxi rates by the surcharge. For area-based lump-sum surcharges, we first convert reported surge factors to actual prices based on standard taxi fares (both the fixed start-up cost and the variable rate cost) and divide by the taxi fare including the lump-sum charge. This adjustment is possible because we query origin–destination regions. If we only had origin prices and trips, we would not be able to do this adjustment. We also incorporate a lump-sum charge for the booking fee for taxi bookings made through the taxi app.[31] Overall, the adjustment affects 23% of the prices in our sample, with 15% of the observations having a difference in adjusted and unadjusted surge factors by more than 0.3.

[30] For robustness, we also estimate the elasticity of substitution using raw surge factors rather than the adjusted surge factors and find the elasticity of substitution estimate drops from 0.398 to 0.319.

[31] Although there is no shown booking fee in the Grab platform, the minimum price of a JustGrab for any distance is S$6 while the minimum cost of a street-hail taxi is S$3.20.

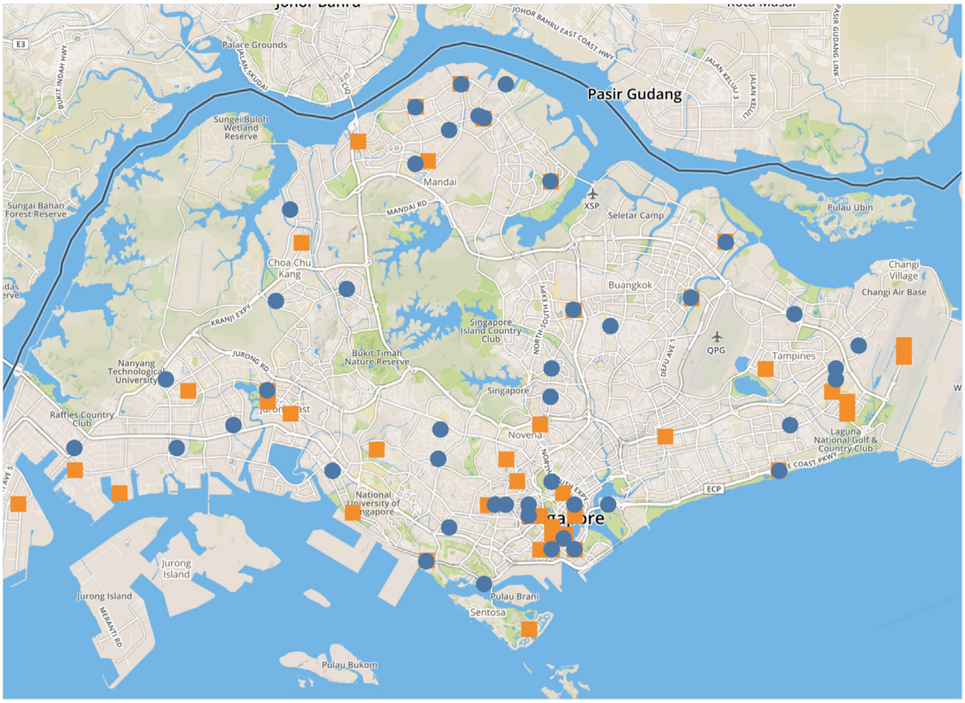



**Fig. A.1.** Origins and Destinations Queried for Prices on the Grab App. Origins and destinations are denoted as circles and squares respectively.

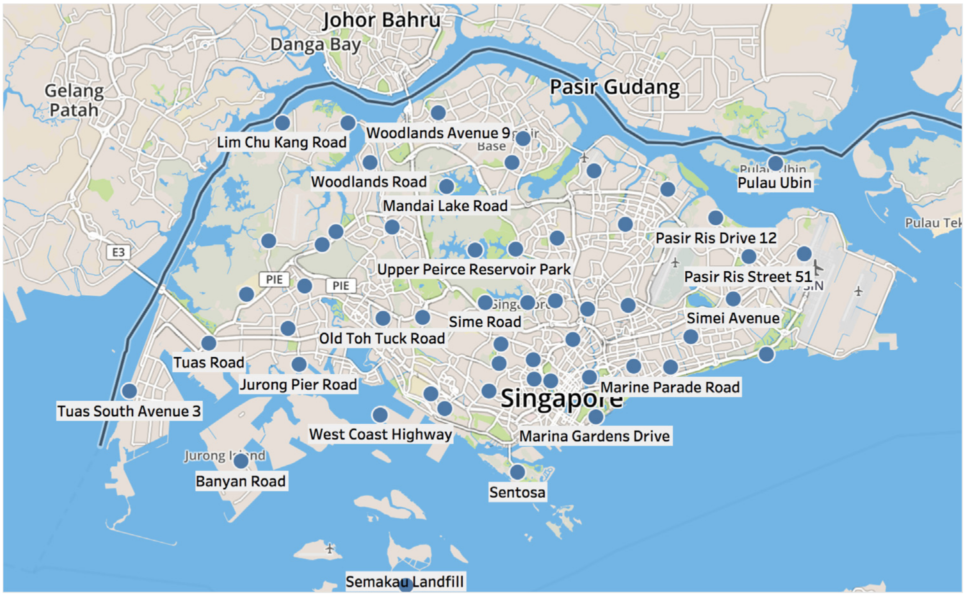



**Fig. A.2.** Locations of all weather stations around Singapore.

Empirically, as shown in Fig. A.4, surge factors are autocorrelated. The gradual decrease in autocorrelation and truncation in partial autocorrelation after one hour (2 lagged half-hour intervals) suggests an AR(2) process for the surge factor.

## Appendix B. Robustness

We run additional analyses for robustness. First, Table A.1 tests whether the relation between surge factors are linear. Second, we also conduct additional falsification tests in Table A.2 by dropping weekends, randomly sampling surge factors, randomly

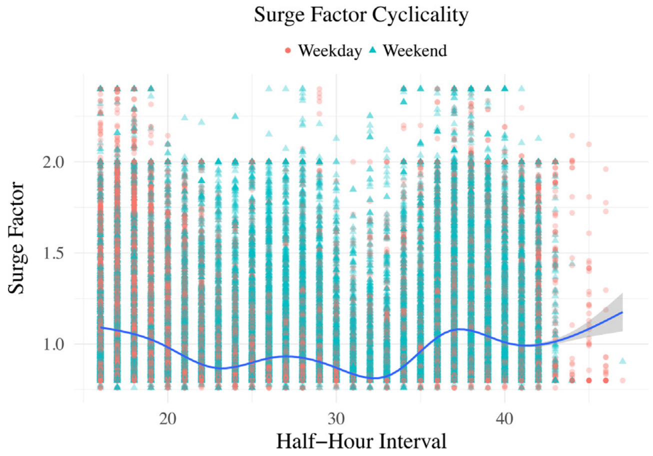



**Fig. A.3.** Surge Factors During the Day. This figure shows a scatter plot of surge factors across different half-hour intervals in the day. The line in the plot is a non-parametric line of best fit estimated using a LOESS algorithm which uses a tri-cube weighting function to categorize local areas. We find some seasonality in surge factors across different hours of the day, corresponding to intuitive notions of peak hours such as morning, lunch, and evening.

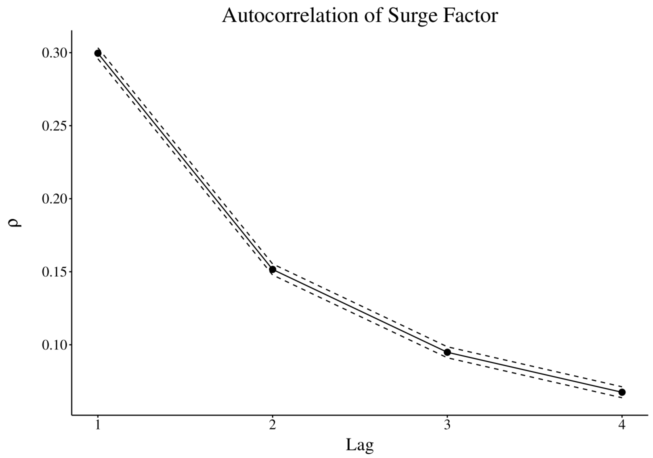



**Fig. A.4.** Autocorrelation of Surge Factors. This figure shows the autocorrelation of surge factors estimated from the equation $SurgeFactor_{o,t,i} = \alpha_{o,i} + Weekend_t + \rho_\tau SurgeFactor_{o,t-\tau,i} + \varepsilon_{o,t,i}$, where $o$ is the origin, $t$ is the date, $i$ is the half-hour interval, and $\tau$ is the lag, from 1 through 4 (corresponding to a half-hour lag through two-hour lag). The dotted lines represent the 95% confidence interval.

sampling from surge factors within the same time interval, and randomly sampling from surge factors within the same route. None of these specifications generate the same results, suggesting that our estimated cross-price elasticity is not generated by confounding correlations across areas or across time intervals.

Despite controlling for taxi supply, if the estimated correlations are due to confounding factors driving both the supply and demand of taxi bookings, a higher surge factor between ride-hailing services and taxis should increase street pick-ups as well. We find that surge factors are actually related to a decrease in street pick-ups. Hence, any confounding factors driving our results would need to consistently impact the supply and demand for taxi bookings but have a slightly negative impact on street-side taxi rides. Since such a confounder is unlikely, we conclude that we are indeed identifying a substitution effect from ride-hailing services to taxi bookings when the relative price of ride-hailing services is high.

In addition, taxi drivers appear to substitute out from street pick-ups, perhaps because they are more likely to accept bookings rather than roam the streets and taxi stands to pick up passengers. This channel, if true, provides even more justification as to why we study taxi bookings rather than the cross-price elasticity of street pick-ups with Grab surge prices, which would be confounded by what we believe is likely a more dominant margin of substitution between Grab and taxi bookings that causes a negative spillover effect to street pick-ups.

During rainy time periods and regions, the surge factor is 16% higher than normal, bringing the surge factor of Grab from an average of 0.9 to 1.05, slightly higher than taxis in the standard pricing period. In periods of heavy rain, defined as rainfall in the top 10% of rainfall of about 10 cm (4 inches), the surge factor increases by about 30%, from 0.9 to 1.2.

We also test whether publicly available signals and market conditions affect the substitution. Table A.3 reports our baseline results using the raw surge factors. Our reduced-form estimate in Column (2) suggests that unconditionally, an increase of 10% in

**Table A.1**
Linearity of price elasticity. This table shows the linearity of the price elasticity as well as tests for differences in elasticities of short and long trips. The observations are at the starting region by half-hour interval by end region level. All regressions include region by time interval by day-of-week fixed effects. Standard errors (in parentheses) are clustered by region and half-hour interval.

| | $y_{o,d,t,i} = \log(\text{Taxi Bookings}_{o,d,t,i})$ | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Surge $\text{Factor}_t$ | 0.181*** | 0.497*** | 0.513*** | | 0.167*** |
| | (0.030) | (0.120) | (0.129) | | (0.033) |
| Surge $\text{Factor}_t^2$ | | −0.121*** | | | |
| | | (0.038) | | | |
| $1\{\text{Surge Factor}_t > 1\}$ | | | 0.382*** | 0.104*** | |
| | | | (0.109) | (0.017) | |
| Surge $\text{Factor}_t \times 1\{\text{Surge Factor}_t > 1\}$ | | | | | −0.046 |
| | | | | | (0.085) |
| Long Distance | | | −0.398*** | | |
| | | | (0.120) | | |
| Surge $\text{Factor}_t \times$ Long Distance | | | | | 0.031 |
| | | | | | (0.040) |
| Taxi $\text{Supply}_t$ | −0.013 | −0.012 | −0.013 | −0.014 | −0.012 |
| | (0.035) | (0.035) | (0.035) | (0.035) | (0.035) |
| Observations | 57,486 | 57,486 | 57,486 | 57,486 | 57,486 |
| $R^2$ | 0.405 | 0.405 | 0.406 | 0.404 | 0.405 |

*p < 0.1; **p < 0.05; ***p < 0.01.

**Table A.2**
Robustness checks. This table shows the response of the log number of taxi bookings in response to the level of the surge factor. Column (1) shows the baseline result, column (2) shows the elasticity of taxi bookings when excluding weekends, column (3) runs the same specification as the baseline result with a randomly assigned surge factor, column (4) repeats the same falsification test sampling from surge factors within the same time interval across routes, column (5) repeats the same falsification test sampling from surge factors within the same route across time, and column (6) tests whether the cross-price elasticity of bookings with respect to surge factor depends on being a weekend. The observations are at the starting region by half-hour interval level. All regressions include region by time interval by day-of-week fixed effects. Standard errors (in parentheses) are clustered by region and half-hour interval.

| Model/Sample: | (1) Baseline | (2) No weekend | (3) Random surge factor | (4) Random surge factor from same time | (5) Random surge factor from same route | (6) Full |
|---|---|---|---|---|---|---|
| Surge Factor | 0.256*** | 0.281*** | −0.015 | 0.013* | −0.015 | 0.279*** |
| | (0.027) | (0.030) | (0.012) | (0.008) | (0.009) | (0.030) |
| Surge Factor × Weekend | | | | | | −0.048 |
| | | | | | | (0.037) |
| Taxi Supply | 0.032 | 0.109* | 0.028 | 0.028 | 0.027 | 0.032 |
| | (0.008) | (0.061) | (0.047) | (0.047) | (0.007) | (0.047) |
| Observations | 57,474 | 34,901 | 57,474 | 57,474 | 57,474 | 57,474 |
| $R^2$ | 0.525 | 0.532 | 0.521 | 0.521 | 0.521 | 0.525 |

*Note:* *p < 0.1; **p < 0.05; ***p < 0.01.

the surge factor leads to a 2.86% increase in taxi bookings. When controlling for the cyclicality of surge factors in a region and whether there is rain in a half-hour interval, we find that the cross-price elasticity increases to 3.06%, shown in Column (1). We also include the interaction of region cyclicality and rain with surge factors, since rain and cyclicality are common knowledge to drivers.[32] Drivers anticipating more riders in an area and time with rain may anticipate higher demand without having information on surge factors. However, we find no statistically significant interaction between region cyclicality or rain with surge factors.

### *B.1. Transitory demand shocks*

Rather than immediately switching from booking a ride-hailing service car to booking a taxi, consumers can also instead wait for surge factors to decrease. In this subsection, we consider two specifications: (*i*) a distributed lag model that isolates past demand shocks by controlling for taxi and grab supplies at different lags within a region, and (*ii*) an impulse response representation that studies how the impact of an isolated demand shock on taxi bookings propagates over time.

In the first analysis, we find that surge factors from half an hour ago still increase current rides but the impact is short-lived beyond that, controlling for the supply factors. Comparing Column (2) to Column (1) in Table A.4, we find the point estimate on contemporaneous surge factors decrease from 0.284 to 0.275 with a positive coefficient on the first lagged interval term of 0.010.

[32] Rain is fairly difficult to predict in Singapore at only about 30 min to 1 h ahead.

**Table A.3**
Public signal results. This table shows the response of the log number of taxi bookings in response to the level of the surge factor. Column (2) shows the baseline result, column (3) shows the interaction of surge factor and cyclical regions (region cyclicality is subsumed by region by time interval fixed effects), column (4) shows the interaction of surge factor with rain, and column (1) shows both. Column (5) shows the effect on street pick-ups. The observations are at the starting region by half-hour interval level. All regressions include region by time interval by day-of-week fixed effects. Standard errors (in parentheses) are clustered by region and half-hour interval.

| | $\log$ Taxi Bookings$_t$ | | | | $\log$ Taxi Street Pick-ups$_t$ |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Surge Factor | 0.306*** | 0.256*** | 0.277*** | 0.313*** | −0.064*** |
| | (0.032) | (0.022) | (0.031) | (0.026) | (0.025) |
| Cyclicality × Surge Factor | 0.032 | | 0.020 | | 0.032 |
| | (0.042) | | (0.043) | | (0.055) |
| Rain | 0.199*** | | | 0.196*** | −0.044 |
| | (0.046) | | | (0.043) | (0.047) |
| Rain × Surge Factor | −0.066 | | | −0.063 | −0.053 |
| | (0.043) | | | (0.042) | (0.043) |
| Observations | 53,760 | 53,760 | 53,760 | 53,760 | 68,309 |
| $R^2$ | 0.400 | 0.412 | 0.385 | 0.400 | 0.606 |

*Note:* $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

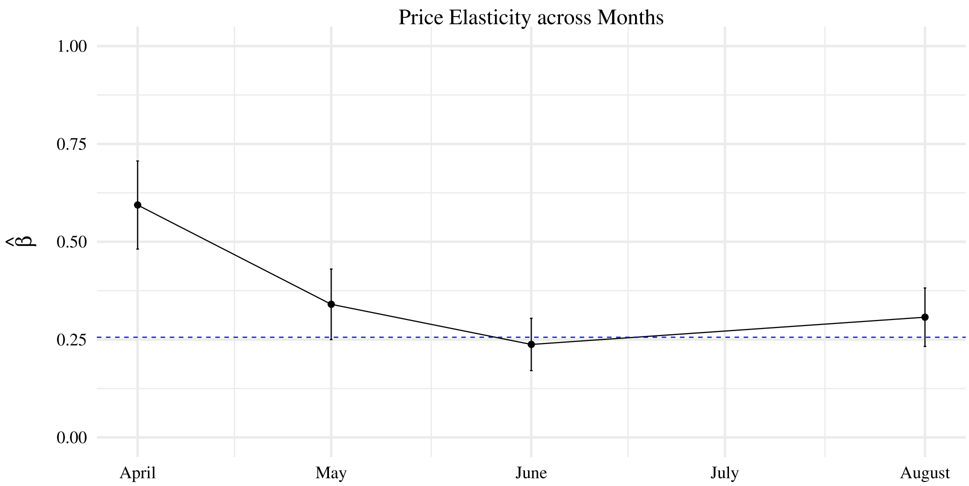



**Fig. A.5.** Price Elasticity across Months. This figure shows elasticity estimates for each month across our sample.

Including additional lags does not change the contemporaneous or the first lagged interval term much. By Table A.4, we conclude that lagged surge factors have weak impact on current taxi bookings.

In the second analysis, we analyze the impact of surge factors on future surge factors, as well as future taxi bookings. Fig. A.6 shows that an increase in the surge factor controlling for taxi supply impacts both future bookings and future surge prices (consistent with Fig. A.4 in the Appendix). However, the lagged impact of the surge factor on taxi bookings decreases to zero after one hour, consistent with the surge factor being a transitory shock. More specifically, the impact of a demand shock on taxi rides decays from 0.4% to zero over the next two and a half hours. The impact of a demand shock on future surge factors also decays from 0.46 half an hour after the shock down to 0.1 over the next two and a half hours.

Our transitory intertemporal results are consistent with consumers who self-select into the taxi or ride-hailing market and who are time inelastic. After all, given the small size of Singapore, any two destinations on the island are reachable with alternative forms of transportation within one and a half hours. Thus, those taking private-hire cars or taxis are likely not willing to wait for a taxi or a Grab for over half an hour. However, partly because current surge factors affect future surge factors, we find that the total effect of a 10% increase in surge factors due to demand on taxi bookings is about 8% over the next two and a half hours.

In these following two subsections, we consider two additional cross-sectional tests to explore whether local average income of residents in an area and the distance of trips affect our estimated cross-price elasticity. Areas with lower incomes should show a higher substitution between ride-hailing services and taxi bookings when surge factors are high. Moreover, since the surge factor is applied uniformly across the whole trip, consumers taking a long trip should be more sensitive to the surge factor since a longer trip, for any given surge factor, means a larger dollar impact of the surge charge.

**Table A.4**
Lagged surge factor. This table shows the response of the log number of taxi bookings in response to the level of the surge factor, distributed up to two-half hour lags. The observations are at the starting region by half-hour interval level. All regressions include region by time interval by day-of-week fixed effects. Standard errors (in parentheses) are clustered by region and half-hour interval.

| | log(Taxi Bookings) | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Surge Factor$_t$ | 0.284$^{***}$ | 0.275$^{***}$ | 0.264$^{***}$ | 0.264$^{***}$ |
| | (0.033) | (0.061) | (0.062) | (0.062) |
| Surge Factor$_{t-1}$ | | 0.010 | −0.039 | −0.040 |
| | | (0.060) | (0.068) | (0.068) |
| Surge Factor$_{t-2}$ | | | 0.071$^{*}$ | 0.048 |
| | | | (0.035) | (0.035) |
| Surge Factor$_{t-3}$ | | | | 0.034 |
| | | | | (0.025) |
| Taxi Supply$_t$ | 0.028 | 0.028 | 0.029 | 0.028 |
| | (0.051) | (0.051) | (0.052) | (0.052) |
| Observations | 36,864 | 36,863 | 36,860 | 36,849 |
| $R^2$ | 0.507 | 0.507 | 0.508 | 0.508 |

$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

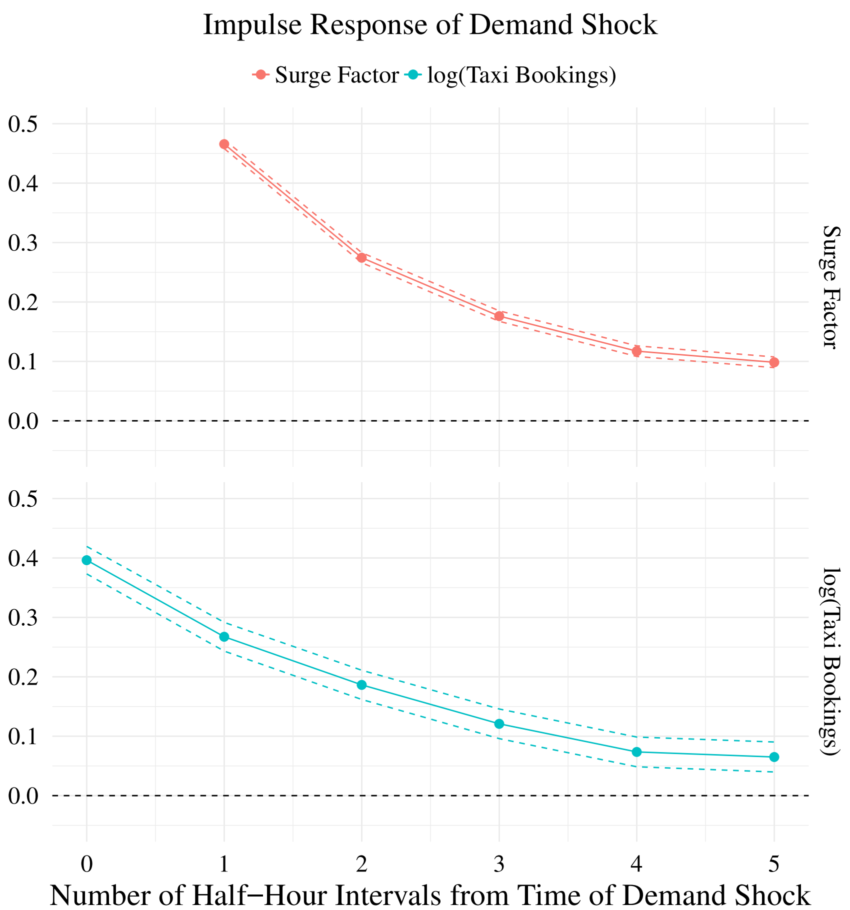



**Fig. A.6.** Impulse Response of Taxi Bookings. This figure shows the impulse response of log taxi bookings for each period from $t$ to $t+5$ due to a surge factor shock in period $t$ controlling for taxi supply at time period $t$. The dotted lines show two-standard error bands.

### *B.2. Distance*

Taxi fares include both a fixed cost and variable cost. Moreover, booking taxis increases the fixed cost component of taking taxis. On the other hand, the marginal benefit of using a taxi meter is that their price per unit of distance is fixed for a given point

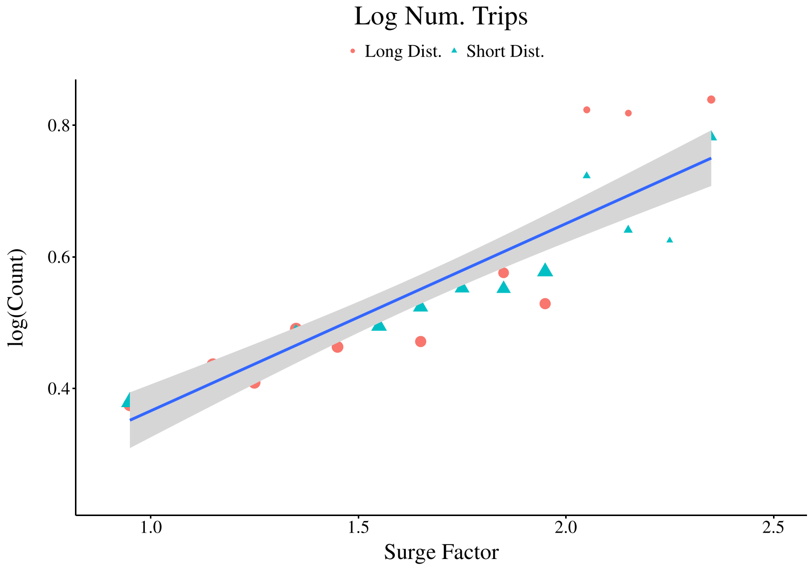



**Fig. A.7.** Taxi Bookings and Surge Factor. The scatter plot for the log number of taxi bookings as a function of surge factor. Each data point represents an average log number of taxi rides where surge factor is split into bins with 0.1 increments. The gray bands represent the 95% confidence interval.

in time.[33] Customers going on longer trips reap a larger absolute gain from using the taxi meter compared to ride-hailing service because surge prices are applied over the whole trip and the fixed cost component of the taxi fare is averaged over a longer distance. Therefore, we expect customers going on longer journeys to be more price elastic.

We test this hypothesis in the data by grouping rides into short and long distances. We define short distance as journeys between two regions that are less than 5 kilometers away from each other. Fig. A.7 shows that there seems to be no difference in the relation between surge factors and taxi bookings based on distance. In untabulated regression results, we find no statistically significant evidence that the price elasticity of longer distances differ from price elasticity of shorter distances. The results persist whether we consider the distance between two regions both as an indicator of a long trip or as a continuous variable.

### *B.3. Impact of income*

Out of the 34 regions for which we have sufficient data, all statistically significant results point to higher surge factors increasing taxi bookings. Although 33 out of 34 regions have a positive point estimate, there is large heterogeneity in cross-price elasticities across regions, from a low of −0.07 to a high of around 0.5 using our baseline specification (see Fig. A.8). A potential explanation of the price elasticity is income. Areas with higher income may have less price sensitivity due to their higher disposable income.

For this analysis, we rely on a proprietary dataset for 2010 income for households with a credit card from a regional bank. We do not find that the heterogeneity in cross-price elasticities is related to income. Table A.5 shows that a region with 10 percentage points more than average income is only 0.02 percentage points less elastic, a less than 1% reduction in the cross-price elasticity compared to a region with the average income. This suggests that whether a consumer shows fickle fingers and switches easily between transportation apps is not related to their income level. This could be because transportation costs are not large compared to the average Singaporean’s overall consumption bundle as well as the easiness to switch between smartphone apps.

## Appendix C. Sensitivity of random forest models

This Appendix shows the sensitivity of the reported random forest models. In the main text, we show the model performance for forests with 500 trees. Fig. A.9 shows the root-mean squared error performance of the models for different numbers of trees where the errors are calculated based on out-of-bag evaluations. In both samples, we see that with over 100 trees, the in-sample predictive power of the full data including surge prices outperforms the data without surge prices.

[33] Recall that taxis in Singapore also have “static” surge factors, either 25% or 50% of metered fare depending on the time of day. We label this static because it does not adjust to changing market conditions on a real-time basis.

**Table A.5**

Cross-price elasticity and income. This table tests whether the cross-price elasticity of taxi rides to surge factors are related to income. Income is defined as the deviation from the average income level in Singapore in 2016. The income variable is absorbed by the origin–destination region by time fixed effect. The observations are at the starting region by half-hour interval level. All regressions include region by time interval by day-of-week fixed effects. Standard errors (in parentheses) are clustered by region and half-hour interval.

| | log Taxi Bookings$_t$<br>Baseline |
|---|---|
| Surge Factor | 0.299*** |
| | (0.027) |
| Surge Factor × Income | −0.0002 |
| | (0.029) |
| Observations | 53,760 |
| $R^2$ | 0.412 |

*Note:* *p < 0.1; **p < 0.05; ***p < 0.01.

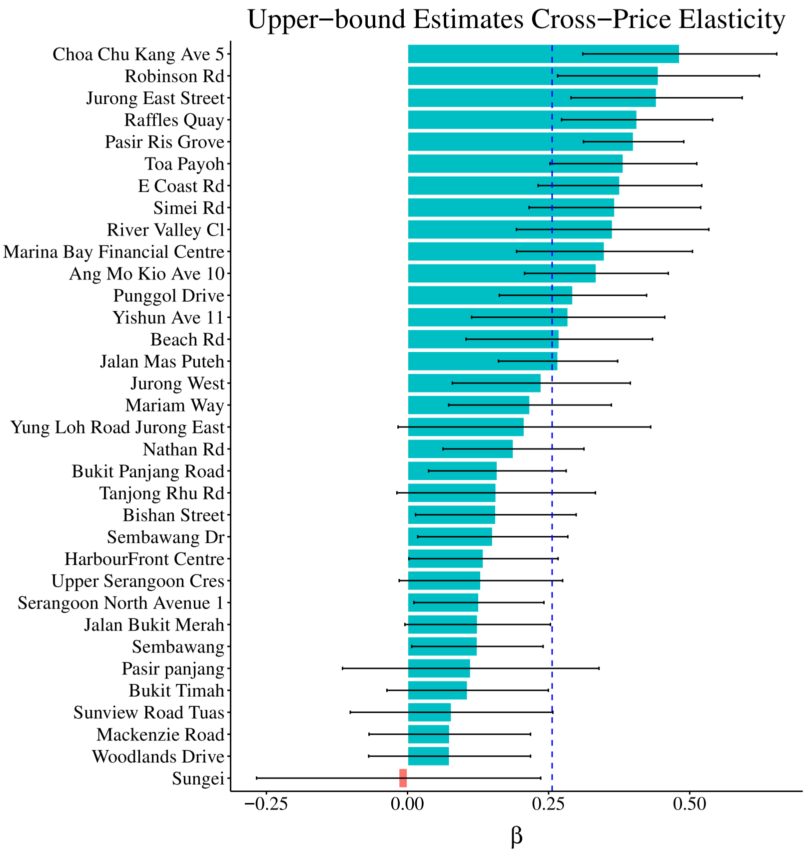
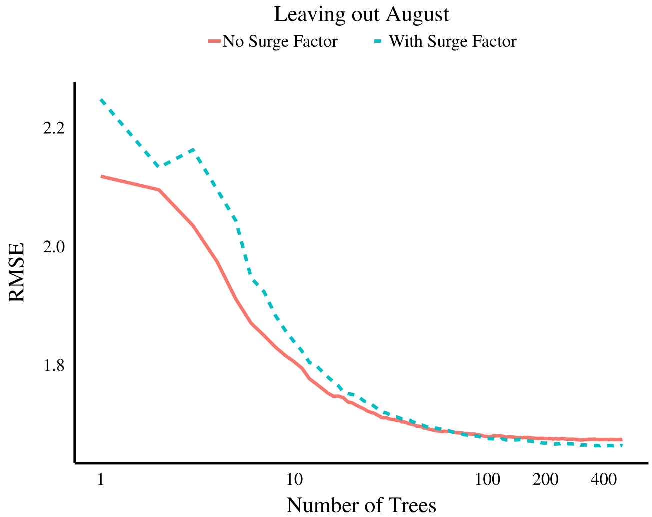



**Fig. A.8.** Elasticity Estimates in Different Regions. This figure shows elasticity estimates across different starting regions in our sample.

### *C.1. Using lagged surge factor in random forest model*

In this subsection, we consider a robustness test of using the surge factor information from the previous half-hour interval in the demand prediction. In this case, we actually find that including the surge factor variable decreases in-sample predictability based on an out-of-bag cross-validated in-sample root mean-squared error. In the leave-August-out specification, we find a decrease in the in-sample performance of 2.4% and in the leave a randomly selected 20% of the sample out of each month, we find a decrease in the in-sample performance of −4.8%.

Meanwhile, as with the contemporaneous surge factor specification, out-of-sample performance improves by around 15% for the leave-August-out specification and 12% for the leave-20%-of-each-month specification. This is likely due to the autocorrelation in the surge factor shown above, and that the autocorrelation in the surge factor suggests a persistent demand–supply mismatch in the Grab ride-hailing market, which means also that they can spillover to the taxi booking market.



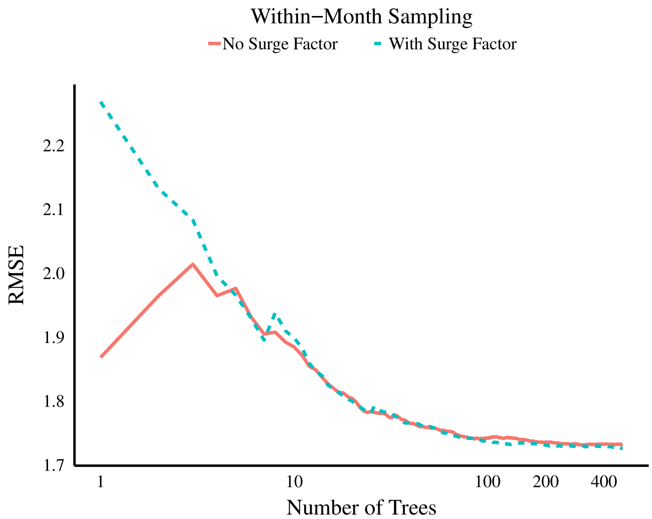




**Fig. A.9.** Root Mean Squared Error Based on the Number of Trees. The figures show the in-sample root-mean squared errors based on the number of trees used for each of the two samples. The "Leaving out August" sample refers to the in-sample using data from April through June. The "Within-Month Sampling" sample refers to the in-sample using a randomly sampled 75% of the full data.

**Table A.6**
Accuracy of taxi bookings predictive model. This table shows the accuracy of the random forest model in predicting the number of taxi bookings with different samples (both in-sample and out-of-sample). RMSE stands for root-mean-squared error.

| Model test sample | In-sample RMSE | | |
|---|---|---|---|
| | No surge factor | With surge factor | Improvement (%) |
| August | 1.417 | 1.451 | −2.399 |
| 20% of each month | 1.420 | 1.448 | −4.789 |
| Model test sample | Out-of-sample RMSE | | |
| | No surge factor | With surge factor | Improvement (%) |
| August | 1.197 | 1.019 | 14.870 |
| 20% of each month | 1.288 | 1.139 | 11.568 |

## References


Acheampong, R.A., Siiba, A., Okyere, D.K., Tuffour, J.P., 2020. Mobility-on-demand: An empirical study of internet-based ride-hailing adoption factors, travel characteristics and mode substitution effects. Transp. Res. C 115, 102638.

Battifarano, M., Qian, Z.S., 2019. Predicting real-time surge pricing of ride-sourcing companies. Transp. Res. C 107, 444–462.

Berger, T., Chen, C., Frey, C.B., 2017. Drivers of Disruption? Estimating the Uber Effect. Working Paper, pp. 1–11.

Bian, Z., Liu, X., 2019a. Mechanism design for first-mile ridesharing based on personalized requirements part I: Theoretical analysis in generalized scenarios. Transp. Res. B 120, 147–171.

Bian, Z., Liu, X., 2019b. Mechanism design for first-mile ridesharing based on personalized requirements part II: Solution algorithm for large-scale problems. Transp. Res. B 120, 172–192.

Bimpikis, K., Candogan, O., Saban, D., 2019. Spatial pricing in ride-sharing networks. Oper. Res. 67 (3), 744–769.

Brown, A., LaValle, W., 2021. Hailing a change: comparing taxi and ridehail service quality in los angeles. Transportation (ISSN: 15729435) 48 (2), 1007–1031. http://dx.doi.org/10.1007/s11116-020-10086-z.

Button, K., 2020. The Ubernomics of ridesourcing: the myths and the reality. Transp. Rev. (ISSN: 14645327) 40 (1), 76–94. http://dx.doi.org/10.1080/01441647.2019.1687605.

Cachon, G.P., Daniels, K.M., Lobel, R., 2017. The role of surge pricing on a service platform with self-scheduling capacity. Manuf. Serv. Oper. Manag. 19 (3), 368–384.

Chakraborty, J., Pandit, D., Chan, F., Xia, J., 2020. A review of Ride-Matching strategies for Ridesourcing and other similar services. Transp. Rev. (ISSN: 14645327) 1–22. http://dx.doi.org/10.1080/01441647.2020.1866096.

Cheng, S.-F., Jha, S.S., Rajendram, R., 2018. Taxis strike back: A field trial of the driver guidance system. In: Seventeenth International Conference on Autonomous Agents and Multiagent Systems. pp. 577–584.

Cheng, S.-F., Nguyen, T.D., 2011. TaxiSim: A Multiagent simulation platform for evaluating taxi fleet operations. In: 2011 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology. pp. 14–21.

Cramer, J., Krueger, A.B., 2016. Disruptive change in the taxi business: The case of uber. Amer. Econ. Rev. Pap. Proc. 106 (5), 177–182.

Garg, N., Nazerzadeh, H., 2021. Driver surge pricing. Manage. Sci. Articles i, 1–19. http://dx.doi.org/10.1287/mnsc.2021.4058.

Geng, X., Li, Y., Wang, L., Zhang, L., Yang, Q., Ye, J., Liu, Y., 2019. Spatiotemporal multi-graph convolution network for ride-hailing demand forecasting. In: Thirty-Third AAAI Conference on Artificial Intelligence.

Glaeser, C.K., Fisher, M., Su, X., 2019. Optimal retail location: Empirical methodology and application to practice. Manuf. Serv. Oper. Manag. 21 (1), 86–102.

Grahn, R., Qian, S., Matthews, H.S., Hendrickson, C., 2021. Are travelers substituting between transportation network companies (TNC) and public buses? A case study in Pittsburgh. Transportation (ISSN: 15729435) 48 (2), 977–1005. http://dx.doi.org/10.1007/s11116-020-10081-4.

Jha, S.S., Cheng, S.-F., Lowalekar, M., Wong, W.H., Rajendram, R.R., Tran, T.K., Varakantham, P., Truong Trong, N., Abd Rahman, F., 2018. Upping the game of taxi driving in the age of Uber. In: Thirtieth Annual Conference on Innovative Applications of Artificial Intelligence. pp. 7779–7785.

Liu, S., He, L., Shen, Z.-J.M., 2020. On-time last mile delivery: Order assignment with travel time predictors. Manage. Sci..

Lu, A., Frazier, P.I., Kislev, O., 2018. Surge pricing moves Uber's driver-partners. In: Nineteenth ACM Conference on Economics and Computation. p. 3.

Moreira-Matias, L., Gama, J., Ferreira, M., Mendes-Moreira, J., Damas, L., 2013. Predicting taxi–passenger demand using streaming data. IEEE Trans. Intell. Transp. Syst. 14 (3), 1393–1402.

Nie, Y.M., 2017. How can the taxi industry survive the tide of ridesourcing? Evidence from Shenzhen, China. Transp. Res. C 79, 242–256.

Posen, H.A., 2015. Ridesharing in the sharing economy: Should regulators impose regulations on Uber? Iowa Law Rev. 101 (1), 405–433.

Ramezani, M., Nourinejad, M., 2018. Dynamic modeling and control of taxi services in large-scale urban networks: A macroscopic approach. Transp. Res. C 94, 203–219.

Rayle, L., Dai, D., Chan, N., Cervero, R., Shaheen, S., 2016. Just a better taxi? A survey-based comparison of taxis, transit, and ridesourcing services in San Francisco. Transp. Policy (ISSN: 1879310X) 45, 168–178. http://dx.doi.org/10.1016/j.tranpol.2015.10.004.

Ross, H., 2015. Ridesharing's house of cards: O'connor V. Uber technologies, inc. and the viability of uber's labor model in washington. Wash. Law Rev. 90, 1431–1469.

Schaller Consulting, 2017. Unsustainable? The Growth of App-Based Ride Services and Traffic, Travel and the Future of New York City. Technical Report, URL www.schallerconsult.com.

Taylor, T.A., 2018. On-demand service platforms. Manuf. Serv. Oper. Manag. 20 (4), 704–720.

Wang, H., Yang, H., 2019. Ridesourcing systems: A framework and review. Transp. Res. B (ISSN: 01912615) 129, 122–155. http://dx.doi.org/10.1016/j.trb.2019.07.009.

Yao, H., Wu, F., Ke, J., Tang, X., Jia, Y., Lu, S., Gong, P., Ye, J., Li, Z., 2018. Deep multi-view spatial–temporal network for taxi demand prediction. In: Thirty-Second AAAI Conference on Artificial Intelligence.

Young, M., Farber, S., 2019. The who, why, and when of Uber and other ride-hailing trips: An examination of a large sample household travel survey. Transp. Res. A 119, 383–392.

Zha, L., Yin, Y., Du, Y., 2018. Surge pricing and labor supply in the ride-sourcing market. Transp. Res. B (ISSN: 01912615) 117, 708–722. http://dx.doi.org/10.1016/j.trb.2017.09.010.

Zuniga-Garcia, N., Tec, M., Scott, J.G., Ruiz-Juri, N., Machemehl, R.B., 2020. Evaluation of ride-sourcing search frictions and driver productivity: A spatial denoising approach. Transp. Res. C (ISSN: 0968090X) 110 (2019), 346–367. http://dx.doi.org/10.1016/j.trc.2019.11.021.