BOSTON UNIVERSITY
Questrom School of Business
Working Paper

Surge Pricing on a Service Platform under Spatial Spillovers: Evidence from Uber

Kyungmin (Brad) Lee, Marcus Bellamy, Nitin Joglekar
Questrom School of Business, Boston University, kminlee@bu.edu, bellamym@bu.edu, joglekar@bu.edu,

Shan Jiang, Christo Wilson
College of Computer and Information Science, Northeastern University, sjiang@ccs.neu.edu, cbw@ccs.neu.edu,

Problem definition: Ride-sharing platforms employ zone-specific surge pricing to match capacity with demand. Since the prices are spatially dispersed, labor capacity spills over (i.e., drivers move around) across zones. We explore the questions: Under what conditions are ride-sharing pricing policies effective in the presence of spillovers? And, what are the implications of such pricing policies for managing congestion and the welfare of both consumer and labor?

Academic/Practical Relevance: There is a debate in the academic community about the values of elasticity that drive the efficacy of pricing policies for shared capacity. This debate has yet to consider price and capacity spillovers. We assess the effectiveness of alternative pricing policies to provide guidance in regards to the debate on their efficacy for managing congestion, while assessing consumer and labor surplus for a relevant range of elasticity values.

Methodology: We develop a stylized optimization model to characterize the relationship between price and capacity spillover. We then investigate how a platform accounts for spillovers by estimating a spatial panel model on a dataset from Uber's operations. Alternative pricing policies are assessed through counterfactual simulations.

Results: Our results show that Uber balances price smoothing and capacity spillover effects. Allied counterfactual analysis illustrates the limitations of spot pricing mechanism (i.e., a policy that does not account for spillovers) commonly being used in the extant literature. These results also show that welfare can be improved by spillover-based pricing and capacity decisions.

Managerial Implications: Digital operations based on shared capacity often require yield management choices for manipulating both supply and demand. Our results show that such operations should account for the simultaneous increase in capacity and drop in demand through surge pricing. We also provide guidance in terms of key parameters that are relevant to the algorithmic implantation of surge pricing policies on shared capacity platforms.

Key words: Digital Operations; Platform Driven Operations; Sharing Economy; Spillover Effects; Yield Management
History:

1. Introduction

The sharing economy has altered the way that firms, service providers, and consumers manage their time, money, and resources in a variety of industries ranging from transportation to hospitality (Sundararajan 2016). The use of digital platforms in these settings enables an on-demand service experience for consumers and expedient feedback to service providers. Emergent literature has modeled both the operational aspects that drive stakeholder (firm, provider, and consumer) decisions and the economic ramifications of such decisions (e.g., Cohen et al. 2016, Taylor 2018, Bai et al. 2018). Ride-sharing has received particular attention in this realm (Gurvich et al. 2016, Cachon et al. 2017). Ride-sharing platforms such as Uber, Lyft, Juno, Curb, Gett, Didi Chuxing, and Fasten connect individuals seeking a ride with providers (drivers) who provide a service for a predetermined wage payment. In this setting, drivers are sensitive not only to their wage rates but also to the variation in rider demand. In a given period of time, the platform influences the ultimate demand and available drivers it receives by setting prices. And, since drivers are allowed to dictate their own schedule, the platform's capacity is characterized by the mismatch that must account for variability both in demand and supply. To deal with this mismatch, platforms often incorporate dynamic pricing policies such as surge pricing to increase capacity and decrease demand in a congested area. Prior research on ride-sharing has advanced our understanding of surge pricing as a mechanism to help firms that operate platforms balance capacity and demand. Much of this revenue management literature is analytic. Specific analytic assumptions have lead to results that feature competing findings. According to Benjaafar et al. (2018, p.6):

> *"In settings where workers have discretion over how much they work, there has been some debate regarding the elasticity of labor supply. For example, in a study of New Yorker City taxi drivers, Camerer et al. (1997) find evidence of negative elasticity and argue that this may be due to taxi drivers being income targeters (i.e., drivers tend to stop working once they reach an income target). In contrast, Chen and Sheldon (2015) and Sheldon (2016), using data from the ride-sharing service Uber, find that drivers tend to drive more when earnings are higher (e.g., during price surges)."*

The debate about elasticity of labor supply has relied heavily on spot pricing. There is a gap in the underlying literature, in terms of accounting for labor capacity spillovers and price spillovers that transpire from one ride-sharing service zone to another. Capacity spillover refers to the movement of drivers from one zone to another zone, either to fill an existing demand (e.g., to pick up or drop off existing passengers) or in anticipation of future demand (e.g., to seek riders in surging zones). Price spillover refers to the smoothing or balancing of prices from a focal zone to its neighboring zone. Spot pricing, in contrast, refers to the surging of prices solely in a focal zone. Most of the analytic work, for reasons of tractability, uses spot pricing and ignores capacity spillovers. We help fill this gap in the empirical revenue management literature by incorporating capacity spillovers and price smoothing effects into our model via a spatial econometric specification. In contrast to some research in this domain, we treat demand as endogenous to price changes, and we model capacity as a dynamic construct while we investigate a ride-sharing platforms pricing strategy. Our study addresses two questions: Under what conditions of price elasticity does a ride-sharing platform set a particular level of surge pricing to deal with spatial spillovers so as to provide reliable services (i.e., to reduce congestion expressed by customer waiting time)? And, how effective are such spillover-based pricing policies in managing congestion and the welfare of both the consumer and ride-sharing drivers (labor) in the presence of spillovers?

We start with a stylized capacity management model and find an equilibrium price that is driven by the current state of both capacity and demand. We then specify a spatial econometric model that accounts for both the spatial and temporal effects on price. Using this model, we estimate the pricing decision in the focal zone in terms of various forms of capacity and pricing in the focal zone and its adjacent zones. We then use the estimated parameters to conduct a counterfactual analysis in order to provide guidance in regards to the debate, and to manage congestion, while assessing consumer and driver (labor) welfare.

Our paper contributes to the emergent literature on platform driven revenue management in three unique ways: (i) We contribute to the literature on autonomous capacity management, under two-sided network effects, by showing that depending upon conditions of price elasticity, a ride-sharing platform can improve upon analytic (e.g., spot pricing) results in the operations and information systems literature in terms of managing congestion and welfare of shared capacity systems; (ii) We find that both capacity spillover and price smoothing across zones have a significant association with surge pricing in the focal zone; this finding advances empirical literature on revenue management that treats dynamic pricing and congestion management as means to pool, whereas we also consider the case of congestion due to pooling; (iii) We find that Uber's algorithms update surge prices much more rapidly than the speed of sampling of these data by the Uber drivers and consumers; given this observation and our policy analysis results, we argue that the platform's rapid pricing updating enables Uber to achieve high performance, akin to aggressive pricing, when compared with the spot pricing policies.

2. Literature Review

We draw upon two related literature streams: the empirical studies of revenue management practices and the analytic treatment of allied welfare economics.

Our paper relates to the stream of empirical literature on revenue management that combines capacity management with pricing strategies. We organize this literature into the dimensions of either static versus dynamic (i.e., endogenous) capacity and static versus dynamic (endogenous) demand.

Table 1 Empirical Studies on Demand and Capacity Management

	Static Demand (exogenous to price change)	**Dynamic Demand (endogenous to price change)**
Static (exogenous) Capacity	Olivares and Cachon (2009) Li et al. (2016) Cui and Hu (2018)	Lu et al. (2013) Li et al. (2014) Tereyağoğlu et al. (2018)
Dynamic (endogenous) Capacity	Moreno and Terwiesch (2015) Karacaoglu et al. (2017)	Our paper

The literature on static demand and capacity is abundant. A benchmark paper, Olivares and Cachon (2009), draws upon a data set from the automotive industry to estimate the effect of the number and type of local competitors on inventory holdings. Their results suggest that the service-level effect is strong, nonlinear, and positive in that automobile dealers carry more inventory (controlling for sales) when they face additional competition. More recent work in this static-static stream focuses on pricing mechanisms in on-demand service platforms to help balance capacity and demand. Li et al. (2016) use the context of on-demand service platforms to estimate the effect of behavioral anomalies between amateur and professional service providers on market outcomes. They empirically show that professional providers experience higher occupancy rates and revenues using data from

Airbnb. They also propose a parsimonious model that suggests an incentive for Airbnb's service platform to charge different prices to amateur and professional service providers. In a similar setting using a dataset from Airbnb, Cui and Hu (2018) consider distributed pricing instead of centralized pricing while studying service provider behavior. Their work focuses on mutual benefits of social utility for providers and customers. Given the setup, capacity and price spillovers are not relevant to their work. Moreno and Terwiesch (2015) consider dynamic (endogenous) capacity in a two-sided market setting by matching tasks with service providers. They empirically examine independent contractors' bidding behavior on freelance contractor platforms, allowing for flexibility in production and dynamic pricing over time. Karacaoglu et al. (2017) use data from a large e-hailing company in South America to study reactions of drivers to entry of new competitors in their zone. Just like our paper, they also find that agents are likely to scatter owing to increased capacity. However, they do not examine pricing effects explicitly.

Li et al. (2014) relax the assumption of exogenous demand and instead model demand as endogenous to price. They use airfare and booking data from the air-travel industry to conduct a structural estimation of the proportion of strategic consumers in the population. They identify conditions under which most strategic consumers are found and when such consumers' presence may boost revenues. Lu et al. (2013) develop an econometric framework in a retail context that uses queuing activity along with point-of-sales (POS) information to estimate the impact of queues on consumer behavior, showing that a pooled system can lead to fewer customers joining the system and therefore increase lost sales when customers decide to join a queue based on its queue length. As in Li et al. (2014), they also set demand as endogenous to price and find an indirect cross-elasticity effect, in that lowering the price of one product can increase congestion, which can indirectly affect demand for another product. This effect is magnified by the heterogeneity and the negative relationship between customer waiting and price sensitivity of the customer. Tereyağoğlu et al. (2018) adopt a dynamic pricing model in a concert ticket sales setting to examine consumer purchasing behavior that is dependent on price and capacity sold. They empirically show that the effect of referencing to past experiences is strong and that consumers are loss averse across price and number of seats sold as a fraction of capacity during their past visits. These studies consider service capacity as fixed, whereas our model considers both service capacity and demand as dynamic constructs.

On the analytic side, recent literature in this stream has also examined congestion and welfare implications for pricing and capacity decisions in on-demand service platforms. A bulk of this literature focuses on spot pricing for reasons of tractability (i.e., it considers price surges solely in a focal zone). Cachon et al. (2017) and Taylor (2018) consider agent participation (i.e., the decision of whether to join the platform) under stochastic demand and agent opportunity costs, and they treat price and wage as endogenous. Cachon et al. (2017) presents an analytic model with dynamic prices and wages under self-scheduling capacity (independent agents), while Taylor (2018) considers platforms that commit to prices and wages in advance; they study the effect of agent independence and customer-delay sensitivity on the optimal price and wage. As an exception to spot pricing work, some studies have gone on to examine information spillovers from consumer learning about the quality of a service from past experiences (Musalem et al. 2017) or they explore provider capacity when one service zone spills over to another service zone to meet unfilled demand. Bimpikis et al. (2016) identify possible spillovers by considering ride-sharing platforms that price discriminate based on location to study the network effect of service demand patterns on the platform's pricing policy, profits, and consumer surplus under a stationary environment. However, such models overlook welfare economics. Since the spot pricing dominates a bulk of the literature on welfare economics, we use spot pricing as a benchmark for comparisons in our counterfactual analysis discussed in section 5.

3. Surge Pricing with Spatial Spillovers

3.1. Surge Pricing Model

Drivers in a ride-sharing platform are allowed to autonomously decide whether to drive in a particular zone, and this autonomy leads to variation in capacity levels over time. In order to match demand with the variable capacity over time, the platform conducts a dynamic pricing policy (or a dynamic wage policy) that considers both the current level of capacity and expected demand. For example, a platform sets a higher price (and thus a higher wage) when facing lower capacity levels during a high demand period to increase capacity and decrease unmet demand. In the Uber platform, this price variation is set by a *surge multiplier*, a multiplicative factor offsetting the standard trip fare that is based on distance and time. Throughout this paper, price denotes the surge multiplier since we focus on the dynamic portion of price response to capacity and demand in a zone regardless of trip

Table 2 Variable Descriptions

Variable	Description
$s_{i,t}$	Average surge multiplier in zone i for period t
$FC_{i,t}$	Focal capacity - Number of drivers in zone i for period t
$NC_{i,t}$	Neighbor capacity - Number of drivers in the neighbors of zone i for period t
UC_t	Untapped capacity - Number of non-operating registered drivers for period t
$d_{i,t}$	Aggregate demand - Number of ride requests in zone i for period t
b_{ji}	Proportion of ride requests that transit from zone j to zone i
γ_l	Aggregate price sensitivity of capacity type l ($l \in \{FC, NC, UC\}$)
δ_l	Static portion of capacity type l's participation ($l \in \{FC, NC, UC\}$)
P_l	Capacity type l's participation probability ($l \in \{FC, NC, UC\}$)
α	Aggregate price sensitivity of demand
$\theta_{i,t}$	Exogenous demand states (i.e., levels) in zone i for period t
$\alpha/\theta_{i,t}$	Price elasticity
p	Penalty costs for capacity shortage
h	Holding costs for excess capacity

distance and time. We also assume the driver receives a commission that is proportional to the price.

In addition, the platform deals with spatially distributed capacity since a driver moves from one zone to another for various reasons. A driver's participation decision may be positively (Chen and Sheldon 2015) or negatively (Camerer et al. 1997) linked with price. For example, an idle driver may choose to switch zones based on expectation of earnings in a surging price zone, while an occupied driver may move to another zone while navigating to the current rider's destination. Therefore, capacity levels for the subsequent period depend on capacity spillovers from idle drivers' movements as well as drivers who relocate to new zones due to rider destinations. In this section, we formulate how price encapsulates such capacity spillovers. The degree of spatial price discrimination is determined by how a single zone is defined. As an extreme, a single zone could possibly aggregate all the drivers so that no actual spatial discrimination exists. However, Bimpikis et al. (2016) argues that some degree of spatial discrimination is more beneficial than non-discriminated pricing. In addition, Chen et al. (2015) found that Uber predetermines zones where each driver receives an identical price. We assume that the entire ride-sharing service area is divided into multiple zones and that a distinct price is periodically assigned to each zone.

Price in each zone is determined by the price-dependent demand $d_{i,t}$ and capacity levels. Demand $d_{i,t}$ is characterized by two factors: a price-dependent factor and an exogenous demand state. Throughout our study, we assume a linear demand to capture both factors. Consistent with Uber's demand estimation approach, we set up the aggregate demand in zone i at time t to be:

$$d_{i,t} = -\alpha s_{i,t} + \theta_{i,t} + error \tag{1}$$

Within this setup, two capacity effects are key factors in determining price: the *direct effect* and *spillover effect*. First, the capacity in the focal zone i for the current time period t, denoted by $FC_{i,t}$, has first-order effects on the price. The factors $d_{i,t}$ and $FC_{i,t}$ help characterize the system's operational performance (e.g., capacity shortage and utilization), especially for ride demand from consumers sensitive to wait-time. Once current demand is served by $FC_{i,t}$, the remaining drivers in the focal zone decide whether to stay in the zone based on their opportunity costs, which we assume to be randomly distributed. A driver will stay only if expected earnings exceed the opportunity costs (Cachon et al. 2017). Therefore, assuming the distribution is known with its cumulative density function OC_{FC}, the probability of a driver staying in the same zone P_{FC} is identical to OC_{FC}(*Expected Earnings*), which is a function of price under the commission-based wage scheme. Second, capacity spillovers are also related to price in the focal zone. We characterize two types of price-related capacity spillovers. One is from the spillover of drivers from neighboring zones and the other is from untapped drivers − those not currently driving − who choose to operate in the focal zone because of high expected earnings. Similar to drivers currently in the focal zone, drivers in neighboring zones at time t, denoted by $NC_{i,t}$ and registered drivers not currently in operation at time t, denoted by UC_t, may choose to participate in the focal zone with probabilities P_{NC} and P_{UC}, respectively. These probabilities differ since each of the three capacity types has a distinct distribution of opportunity costs which take earnings from any other alternatives into account. In short, these three forms of capacity (i.e., $FC_{i,t}$, $NC_{i,t}$, UC_t) determine the focal capacity in the subsequent period with different probabilities that are directly proportional to the price. Thus, the focal zone capacity is updated by the specified spillovers with price-dependent probabilities as well as spillovers that arise due to customer destinations ending in zone i. Formally, the updated focal capacity of zone i for time $t+1$ satisfies:

$$\begin{aligned} FC_{i,t+1} = & \sum_{j \neq i} b_{ji} \cdot min(d_{j,t}, FC_{j,t}) + b_{ii} \cdot min(d_{i,t}, FC_{i,t}) \\ & + max(FC_{i,t} - d_{i,t}, 0) P_{FC}(s_{i,t}) + NC_{i,t} P_{NC}(s_{i,t}) + UC_{i,t} P_{UC}(s_{i,t}) \end{aligned} \tag{2}$$

In Equation (2), the first term on the right indicates drivers entering the focal zone from all other zones, the second term on the right indicates drivers who serve demand within the focal zone, and the last three terms indicate drivers who choose to be in the focal zone based on each price-dependent probability (P_{FC}, P_{NC}, P_{UC}). A summary of all variables is shown in Table 2. To characterize the relationship between price and each of the three capacity types, we assume that the opportunity costs for each capacity type are uniformly distributed with distinct ranges, so that the participation probabilities are linearly proportional to the price with different rates. That is, for $l \in \{FC, NC, UC\}$,

$$P_l = \gamma_l s_{i,t} + \delta_l \tag{3}$$

Consumers in the platform are sensitive to waiting time when they have available alternatives. In response, a ride-sharing platform may set a price that minimizes its capacity shortage since the shortage penalizes the platform through a loss of consumers. Moreover, the platform also needs to consider the number of focal drivers in the subsequent period. Having excess drivers in the following period may increase driver competition, which, in turn, reduces drivers' expected earnings. In sum, a high price decreases demand for the current period but increases capacity for the subsequent period. To capture such an temporal trade-off, we model a two-period capacity problem with an equilibrium price set to minimize the penalty costs from capacity shortage and the holding costs associated with the number of drivers in the subsequent period. Formally, we present the following stylized objective function for the platform's capacity management:

$$\underset{s_{i,t}>1}{minimize}\ \mathbf{E}[p \cdot max(0, d_{i,t} - FC_{i,t}) + h \cdot FC_{i,t+1}] \tag{4}$$

With the demand given in Equation (1), the optimization problem characterizes an equilibrium price of the platform as a function of each capacity type, when demand state $\theta_{i,t}$ is greater than $FC_{i,t}$. Formally, an equilibrium price is given by:

$$s_{i,t} = \frac{p - h\delta_{FC}}{2h\gamma_{FC}} - \frac{1}{2\alpha}FC_{i,t} - \frac{\gamma_{NC}}{2\alpha\gamma_{FC}}NC_{i,t} - \frac{\gamma_{UC}}{2\alpha\gamma_{FC}}UC_t + \frac{\theta_{i,t}}{2\alpha} \tag{5}$$

(see Appendix A for the proof)

As a result, in addition to the zone-specific characteristics such as demand state and FC, other external effects such as spillovers from NC and UC are also reflected in the equilibrium price. The equilibrium price is thus based on the current state capacity and demand as well as the price sensitivities of each capacity type (γ_{FC}, γ_{NC}, γ_{UC}) and demand ($-\alpha$).

3.2. Price Sensitivity of Capacity Spillovers

By drawing upon the price sensitivity parameters from the preceding analysis, we build a set of empirically testable hypotheses on the relationship between the current spatial distribution of capacity and price. Neoclassical theory assumes that individuals maximize their own utility. Such an assumption implies that demand is negatively related to an increase in price (i.e., $\alpha > 0$). Negative price sensitivity of demand has been empirically shown in the context of transport (see Oum et al. (1992)). Furthermore, Cohen et al. (2016) find empirical evidence of the negative price elasticity of demand in the ride-sharing context. Nevertheless, a tension remains in regards to the price sensitivity of labor supply. On one hand, Camerer et al. (1997) finds a negative relationship between price and labor hours of New York City taxi drivers, suggesting reference-dependent behavior of drivers. In other words, they argue that drivers tend to stop driving once they reach their desired daily profits. On the other hand, Chen et al. (2015) and Chen and Sheldon (2015) show empirical support for a positive link between price change and driver participation in the Uber platform. These findings are consistent with a common assumption that expected earning must exceed the opportunity costs for a ride-share driver's participation (see e.g., Bai et al. (2018), Benjaafar et al. (2018), Cachon et al. (2017), Gurvich et al. (2016), Ibrahim (2018), Taylor (2018), etc.).

Since prior studies featuring price sensitivity and labor supply have addressed temporal changes, we too look at temporal factors and also focus on the spatial distribution of drivers in the ride-sharing service area. That is, we are interested in two decisions of a driver: when to start and stop work and relocation during their work. The relationship between the price and relocation decision is measured by γ_{FC} and γ_{NC}. Therefore, we claim that these two sensitivities will have the same sign (i.e. $\frac{\gamma_{NC}}{\gamma_{FC}} > 0$) as long as a driver exhibits consistent behavior. Particularly, if a driver has already decided to work in the current period and maximizes expected earnings for the relocation decision, this driver will move to a zone with a higher price. This earning-maximizing behavior suggests that both γ_{FC} and γ_{NC} are positive. On the other hand, γ_{UC} captures both the driver's relocation and work hours decision. If the earning-maximizing assumption applies for both decisions, we can conclude that γ_{UC} is positive. However, if a driver's decision to work a certain number of hours is independent of the current price, then γ_{UC} only measures their relocation decision and the sign may not be positive. However, if a driver's decision to work a certain number of hours is negatively associated with price, the sign of γ_{UC} will be mixed. Farber (2015) finds that a driver's hours of operation decision is generally positively related to price although it is heterogeneous across drivers. In line with their findings, we also assume that γ_{FC} and γ_{UC} have the same sign (i.e., $\frac{\gamma_{UC}}{\gamma_{FC}} > 0$).

The sign of the price sensitivity parameters characterizes the relationship between price and capacity type. First, under a positive γ_{FC} assumption, the price in the focal zone will increase to manage capacity utilization (the consumer-to-driver ratio) when focal capacity is congested (i.e., the platform attempts to maintain the focal capacity by providing relatively high expected earnings). Furthermore, if the price sensitivity of consumers ($-\alpha$) is negative, the equilibrium price will indicate that the platform may set a higher price for a low current focal capacity to match demand with the given capacity by decreasing demand. Second, the current neighbor capacity level is negatively correlated with price. In other words, price will not increase if an adequate number of drivers are in neighboring zones but too few drivers are in the focal zone. Likewise, price will increase when few drivers are in neighbor zones even though the focal capacity was high enough to meet the current demand. The equilibrium price also indicates a negative relationship between price in the focal zone and NC under the assumptions that $\alpha > 0$ and $\frac{\gamma_{NC}}{\gamma_{FC}} > 0$. Finally, we characterize an inverse relationship between the price and UC. When the number of drivers who are expected to potentially participate in the following period is high, the price in the focal zone will increase to reduce the possibility of driver shortage. The equilibrium price also suggests the negative relationship under $\alpha > 0$ and $\frac{\gamma_{NC}}{\gamma_{FC}} > 0$. Therefore, we present a set of hypotheses on the relationship between price and each of the three types of capacity as follows:

HYPOTHESIS 1. *The price in each zone is negatively associated with the number of current drivers in the same zone.*

HYPOTHESIS 2. *The price in each zone is negatively associated with the number of current drivers in its neighboring zones.*

HYPOTHESIS 3. *The price in each zone is negatively associated with the number of current untapped drivers.*

3.3. Price Smoothing

So far, we have assumed that the surge multiplier for a certain zone is independent of pricing characteristics in other zones. However, there may be a set of absolute spatial spillover effects on the price across zones. That is, a price change in the focal zone may directly relate to price changes neighboring zones. In the study of the real estate market, Can (1992) found that housing prices are dependent on nearby prices solely based on their spatial proximity. This suggests that there exists a spatial dependency that translates to neighboring prices. In the ride-sharing context, a pricing policy in the focal zone may capitalize on the spatial proximity of neighboring zones since it directly relates to consumer wait-time. In other words, when a focal zone is highly likely to experience a capacity shortage, the platform may increase prices not only in the focal zone but also in neighboring zones regardless of their capacity and demand levels. In this scenario, a driver outside the focal zone travels into the focal zone only when expected earnings exceed the opportunity costs, which account for the spatial proximity to the focal zone. Thus, the probability of participating in the focal zone decreases with distance from that zone. Therefore, when there is a capacity shortage, the price may not only go up in the congested zone but also gradually rise in neighbor zones around the congested zone, in order to secure drivers nearby. We define such a mechanism as *price smoothing*. Price smoothing can reduce further shortages by attracting drivers into locations in which their participation probabilities increase. In addition, price smoothing can prevent high degrees of price discrimination across zones, which can cause inefficiencies by allowing strategic behavior from drivers and riders. In turn, we hypothesize that the surge price in a focal zone also increases prices in neighboring zones:

HYPOTHESIS 4. *The price in each zone is positively associated with the prices in its neighboring zones.*

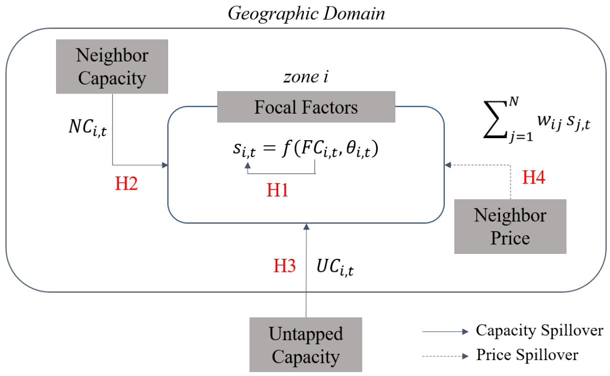
Figure 1 offers a conceptual depiction of the hypothesized internal and external factors that affect price.

4. Data and Methods

4.1. Data Collection and Measures

We collected data on UberX activity to test our hypotheses using the methods developed by Chen et al. (2015) and also adopted by Jiang et al. (2018). UberX is the most popular ride-sharing service in the United States. Although Uber provides various versions of ride-sharing platforms such as UberBlack and UberPool, we focus on UberX since it provides

Figure 1 Conceptual Framework



Note. w_{ij} is a weight of zone j based on spatial proximity between zone i and zone j

a simple one-to-one service and allows for the most autonomy for driver participation and spatial relocation among all Uber-based platforms. We recorded the responses of Uber's server in the passenger app. When a consumer sends a signal for a ride request in the passenger app, Uber's server responds every five seconds with the GPS coordinates of the eight nearest available drivers and the current surge multiplier (i.e., price). We developed a script that sends signals to the Uber server from multiple observation points simultaneously and records each of the responses. We collected the data in San Francisco from 00:00 A.M on November 12, 2016 to 11:59 P.M on November 30, 2016. Aggregate descriptions of the sample along with descriptive statistics are shown in Appendix B.

Our unit of analysis is the *surge zone*, in which a single price is given for each vehicle at each point in time. Consistent with Jiang et al. (2018), we use the *block group* to denote a surge zone. Although variation within the block group is possible, the block group is the most granular unit that is amenable to analysis. Hence we aggregated the collected data into capacity and price levels for block groups. In addition, based on the data shown in Appendix B, we observe that the average travel time from one zone to another zone is about 30 minutes. Therefore, we constructed a panel data set of the average price ($s_{i,t}$) and number of unique drivers ($FC_{i,t}$) for each zone (i) in each 30-minute interval (t) - robustness checks are presented in the Online Appendix.

The other explanatory variables are operationalized from $FC_{i,t}$ with a few assumptions. First, to measure UC_t, we assume a large fixed value as the total number of registered Uber drivers (M) in San Francisco. Thus, we have $UC_t = M - \sum_i FC_{i,t}$. Second, when calculating $NC_{i,t}$, we multiply $FC_{i,t}$ by a spatial weight matrix W. For N zones, the spatial weight matrix is an $N \times N$ nonnegative matrix of which elements (w_{ij}) indicate spatial proximity between any two zones (i and j). Its diagonal elements are zero and the matrix is row-normalized. We adopt the radial distance method (Cliff and Ord 1981, Anselin 1988) for our analysis: $w_{ij} = \begin{cases} 1, & \text{if } 0 < Dist_{ij} \leq 2 \\ 0, & \text{otherwise} \end{cases}$, where $Dist_{ij}$ is the distance between i and j. By assembling $NC_{i,t}$ with W, we were able to incorporate our assumption that opportunity costs increase with distance from the focal zone. Furthermore, this enables us to use a spatial Durbin model, as discussed in section 4.2.

4.2. Estimation Strategy

To begin, we consider a regression model specification with both spatial and temporal effects of the following form:

$$y_{i,t} = \beta x_{i,t} + \mu_i + t + \epsilon_{i,t} \tag{6}$$

where each subscript i refers to the zone and t the time, $y_{i,t}$ represents an observation of the dependent variable, β is a fixed parameter, $x_{i,t}$ is an observation of each explanatory variable, μ_i represents the spatial fixed effect, t represents the vector for temporal effects that control for the time trend, and $\epsilon_{i,t}$ an independently and identically distributed error term with zero mean. The inclusion of μ_i and t should help capture omitted time-invariant factors that characterize each zone and omitted temporal factors, respectively.

In addition to the spatial and time specific effects, our model further accounts for the following interaction effects: (i) pricing decisions in a focal zone that may be influenced by an explanatory variable in nearby zones, (ii) pricing decisions in other zones that may influence pricing decisions in the focal zone, and (iii) pricing decisions in different zones that may be spatially correlated due to unobserved characteristics. To account for such possible interactions among different zones, we develop a spatial panel regression model by extending Equation (6) as follows:

$$y_{i,t} = \lambda \sum_{j=1}^{N} w_{ij} y_{j,t} + \beta x_{i,t} + \beta' \sum_{j=1}^{N} w_{ij} x_{i,t} + \mu_i + t + \upsilon_{i,t} \tag{7}$$

where

$$\upsilon_{i,t} = \rho \sum_{j=1}^{N} w_{ij} \upsilon_{j,t} + \epsilon_{i,t} \tag{8}$$

where $\upsilon_{i,t}$ represents a spatially autocorrelated model and ρ is the spatial autocorrelation coefficient.

The inclusion of a spatially lagged dependent variable term ($\lambda \sum_{j=1}^{N} w_{ij} y_{j,t}$) and a spatially lagged explanatory variable ($\beta' \sum_{j=1}^{N} w_{ij} x_{i,t}$) account for the spatial interaction effects (i) and (ii), respectively, where λ is the spatial autoregressive coefficient and β' is a fixed but unknown parameter. The spatial econometrics literature refers to the model that includes these two effects as the "spatial Durbin model" (SDM), see Anselin (1988) and LeSage and Pace (2009). Moreover, specifying the error term as in equation (8) helps us incorporate the third spatial interaction effect (iii). The literature denotes it as a Manski model, when all three effects are jointly included (Elhorst 2010). However, separately identifying each effect is impossible (Bottasso et al. 2014). Instead, LeSage and Pace (2009) suggest that ignoring the spatial error dependence only reduces efficiency in the estimates, which can be mitigated with a large sample such as ours. Moreover, the authors indicate that the SDM model does not ignore spatial dependence in the disturbances but nests models involving both (i) and (iii). Therefore, LeSage and Pace (2009) argue that the SDM model generates unbiased coefficient estimates even if the true process includes interaction effects (iii). Bottasso et al. (2014) further discuss how unbiased estimates of the SDM model can be obtained even with the possibility of error dependence.

Given these considerations, we use a spatial Durbin specification to estimate our model as follows:

$$s_{i,t} = \lambda \sum_{j=1}^{N} w_{ij} s_{j,t} + \beta_1 FC_{i,t} + \beta_2 NC_{i,t} + \beta_3 UC_t + \mu_i + t + \epsilon_{i,t} \tag{9}$$

where $s_{\cdot,t}$ represents the average surge multiplier in zone $\cdot$ for period t, $NC_{i,t} = \sum_{j=1}^{N} w_{ij} FC_{j,t}$, β_1 represents the degree that price is dependent on focal capacity at a given level of demand, and β_2 and β_3 reflect how much each capacity spillover is factored into the price. λ indicates the extent to which price is affected by spatial proximity. We ran joint Lagrange multiplier (LM) tests that were derived by Baltagi et al. (2003) (i.e., extended LM tests of Breusch and Pagan (1980)) to test for the random effects and spatial error dependence for our model specification. In addition to the main explanatory variables, Chen and Sheldon (2015) found evidence of variation in demand by weather and weekend. Hence, we included t to control for temporal effects such as $Weather_t$, which is 1 if it is rainy and zero otherwise, which controls for the natural variation in demand state due to adverse weather conditions. We also controlled for weekend ($Weekend_t$), and time of day ($Hour_t$) effects, which determine the demand states $\theta_{i,t}$ for a given time period; $Weekend_t$ is equal to one if the observation was made during the weekend and zero otherwise. This controls for variations in demand state by day of the week; $Hour_t$ simply controls for time-of-day effects. Moreover, demand states may also vary with zone. For example, downtown San Francisco, a densely populated commercial area, has a different demand pattern from that of the Sunset District, which is a residential area. We control for any such variations across zones by including the time-invariant spatial effects μ_i. We ran a Hausman test to check whether our model estimates are consistent with a random effects estimation, guided by Mutl and Pfaffermayr (2011). The test statistics indicate that including random effects is inconsistent, thus we consider the fixed effects for our main analysis. In particular, we use the maximum likelihood spatial fixed effects estimator proposed by Lee and Yu (2010), who suggest this as a bias correction procedure for previous maximum likelihood estimators. By adding both temporal factors t and spatial fixed effects μ_i, our estimation procedure controls the exogenous demand state ($\theta_{i,t}$) that is presented in the equilibrium price. For the estimation process, We follow the xsmle routine for STATA proposed by Belotti et al. (2016).

4.3. Estimation Results

The first column of Table 3 summarizes the estimation results of Equation (9). We find that the coefficient estimate (β_1) for the Focal Capacity (FC) is not statistically significant. Thus, Hypothesis 1 is not supported. In other words, the price does not appear to incorporate the number of drivers in the focal zone. This is possibly because the current FC is not independently utilized by itself but integrated with the demand state as a signal to surge. The price may surge when the demand state is expected to exceed the current FC. However, the degree of the price surge may not be dependent on those two factors. Moreover, a driver currently transporting a consumer may pass through other zones quickly to reach the destination. Thus, the pricing does not assume that drivers will stay idle within the same zone in the subsequent time period. Instead, we observe that the median of driver's idle time is only about 3 minutes. The positive sign of β_1 indicates that price may increase with higher FC although it is not statistically significant. This could be explained by driver's forward-looking behavior. That is, a driver may decide where to go based on future price rather than current price. Yılmaz et al. (2017) found that the positive effects of dynamic pricing diminishes with such strategic behavior. To enhance the effectiveness of surge pricing, Uber can discourage driver's strategic behavior. By increasing the price slightly with an additional driver in the focal zone, the focal drivers are encouraged to stay in the zone.

Table 3 Relationship between capacity levels and surge multiplier

	Surge Multiplier ($s_{i,t}$)	
$FC_{i,t}$	1.22×10^{-5}	-
	(4.99×10^{-5})	-
$FC_{i,t-1}$	-	3.23×10^{-5}
	-	(5.16×10^{-5})
$NC_{i,t}$	-4.96×10^{-4}**	-
	(1.03×10^{-4})	-
$NC_{i,t-1}$	-	-4.52×10^{-4}**
	-	(1.08×10^{-4})
UC_t	-1.89×10^{-6}*	-
	(7.31×10^{-7})	-
UC_{t-1}	-	-1.61×10^{-6}*
	-	(7.40×10^{-7})
$\sum_{j=1}^{N} w_{ij} s_{j,t}$	0.89**	0.89**
	(0.013)	(0.013)
Controls	*included*	*included*
Log-likelihood	75369.06	75295.48

Notes. *Controls* include $Hour_t$, $Weekend_t$, and $Weather_t$. Robust standard errors clustered by zone are in parentheses. Number of observation is 156692 for the main model and 156520 for the lagged model. $+p < 0.1$; $*p < 0.05$; $**p < 0.01$.

Hypothesis 2 posits that the price in a focal zone will increase if not enough drivers are close to the zone. We find support for Hypothesis 2 by observing that the number of active drivers in neighboring zones has a significantly negative impact on price. The negative relationship provides two important insights into Uber's pricing policy. One is that Uber assumes a positive price sensitivity for Neighboring Capacity (NC) spillovers. The pricing policy expects neighboring-zone drivers will be stimulated to cross into the focal zone that offers an increased price. The second insight is that Uber assumes that unmet demand in the focal zone can be served by the spilled-over NC. That is, the pricing policy proactively capitalizes on the anticipated NC spillovers. We also observe that the magnitude of the NC's impact is greater than that of the FC's. This suggests that Uber attempts to efficiently address demand in the focal zone by serving relatively less delay-sensitive consumers at a lower price while giving up on serving consumers who are highly sensitive to delay. In other words, as long as consumers are tolerable enough to be served by NC, they can be served at a lower price. On the contrary, if they are highly sensitive to demand, Uber would choose to serve them with higher price. This aligns with the argument of Taylor (2018) that consumers with high delay sensitivity have prompted the platform to raise prices while giving up on serving price-sensitive consumers.

Furthermore, we find that the impact of Untapped Capacity (UC) on price is negative, in line with Hypothesis 3. In other words, an increase in total number of registered Uber drivers across all zones is associated with an increase in price. This illustrates that Uber provides more incentives to currently operating drivers in any zone than to untapped drivers. An untapped driver may not participate in the focal zone only because of its high price. Thus, Uber may facilitate the matching of demand with supply by utilizing existing drivers rather than by attracting untapped drivers. This finding contrasts with a common argument that the untapped driver's participation is motivated by the increased price. Rather, this suggests that working hours of a driver tend to be set ex ante regardless of price level as assumed in Bimpikis et al. (2016).

In our estimation, β_2 and β_3 estimate $\frac{\gamma_{NC}}{2\alpha\gamma_{FC}}$ and $\frac{\gamma_{UC}}{2\alpha\gamma_{FC}}$ respectively, and the ratios are sensitive to the value of α. In other words, despite the statistical significance, the impact of FC and UC on price can be small if consumers are highly sensitive to price. Nonetheless, to further investigate the effects of FC and UC under specific pricing policies, we study outcome performance with different values of α in Section 5.

We next examine the impact of spatially lagged variable $\sum_{j=1}^{N} w_{ij} s_{j,t}$ to assess whether Uber accounts for price smoothing. We find that the focal price is strongly dependent on the prices in the neighboring zones, supporting Hypothesis 4. This indicates that, in addition to exogenous factors such as capacity levels and demand state, price in the focal zone is endogenously affected by itself due to its interrelationship with prices in the neighbor zones. That is, spatial proximity is another significant factor that determines the spatial distribution of prices. For example, even when a price in the focal zone needs to surge based on exogenous factors while prices in neighboring zones do not, Uber increases prices not only in the focal zone but also in neighbor zones. Uber executes price smoothing for two possible reasons. First, it encourages drivers to move or stay near the shortage zone by simultaneously increasing prices in a cluster of zones. By attracting drivers with high prices, Uber can aggressively react to a capacity shortage. Second, by invoking price smoothing, Uber prevents consumers in the focal zone from reneging or moving toward a neighboring zone. In turn, price smoothing allows consumers to trade-off high price for a certain level of delay. In sum, we find that Uber manages the spatial dimension of capacity with a pricing policy that capitalizes on two spatial spillovers: capacity spillovers and price smoothing.

4.4. Robustness Checks

4.4.1. Time Agglomeration: As per our raw data set, Uber updates its states every 5 seconds. However, as shown in Appendix B, the average time that drivers require to learn about the system status, based on their average travel time, is several minutes. We have tested the model by aggregating the data set at 15 and 60 minute intervals as shown in Online Appendix Section 1. These results are structurally similar to the base case shown in Table 3.

4.4.2. Specification of Spatial Weight Matrix (W): We use the radial distance method to specify the W matrix while smoothing prices in our base case (i.e, the result in Table 3). We also checked for the robustness of our results with different Ws. Two other widely-used methods were employed as suggested in Cliff and Ord (1981) and Anselin (1988): the K-nearest and distance decay methods. In the K-nearest method, $w_{ij} = \begin{cases} 1, & \text{if } j \in N_{40}(i) \\ 0, & \text{otherwise} \end{cases}$, where $N_{40}(i)$ is a set of 40 nearest zones of zone i. In the distance decay method, $w_{ij} = 1/Dist_{ij}$, where $Dist_{ij}$ refers to the distance between i and j. The base case results remain robust as shown in Online Appendix Section 2.

4.4.3. Endogeneity: So far, we have assumed that price decisions for a given period are made after the level of focal capacity is observed for that period. That is, the focal capacity might be considered exogenous after controlling for spatial and temporal effects. However, this assumption may not be warranted for a few reasons. For example, an increase in focal capacity for a given period might have resulted from a price increase in the zone for that period. Such simultaneity can bias the coefficient estimates. Although appropriate instrument variables (IVs) would help us to address the issue, such methods with a spatial panel dataset are not well developed in the spatial econometrics literature. Instead, we address the issue in a similar fashion to Bottasso et al. (2014). First, we obtained estimates using lagged explanatory variables ($FC_{i,t-1}$, $NC_{i,t-1}$, UC_{t-1}) in order to account for the difference in time. In this way, endogeneity bias from the current price that could trigger changes in the current focal capacity is likely to be a minor issue. As shown in Table 3, column 2, our results are robust in that NC and UC have a negative impact on price while the impact of FC is insignificant. Secondly, following by Bottasso et al. (2014), we also implemented temporal lags of $FC_{i,t}$ and $NC_{i,t}$ as IVs using a generalized method of moments (GMM) approach. In order for an IV to be valid, it must be correlated with the explanatory variables, $FC_{i,t}$ and $NC_{i,t}$ in our case, and can influence the dependent variable $s_{i,t}$ only through the explanatory variables. Since driver movements can be costly when the driver is idle, the total focal capacity might be correlated over time. Chen and Sheldon (2015) found that the median driving time of Uber drivers was 3.47 hours, which suggests that the number of drivers are correlated within the time interval. Furthermore, the number of drivers is expected to have a certain degree of temporal autocorrelation due to our data aggregation process. However, all lags are not exogenous due to the data aggregation. Among those that satisfy the first condition, the furthest lag could have the least direct impact on the price. Using our data, we find that $FC_{i,t-8}$ and $NC_{i,t-8}$ are not significantly correlated with $s_{i,t}$ but correlated with $FC_{i,t}$ and $NC_{i,t}$. Therefore, we use $FC_{i,t-8}$ and $NC_{i,t-8}$ as our IVs for the GMM approach. As shown in Online Appendix Section 3, empirical results indicate that both price and capacity spillovers are important factors in the pricing decision, suggesting that we have mitigated concerns of this type of endogeneity.

4.4.4. Demand Sensitivity: The coefficients estimated in Table 3 imply relationships among price sensitivities of consumer and labor as suggested in the equilibrium price. We did not have data on demand state $\theta_{i,t}$. Our regression results may be affected by variation in $\theta_{i,t}$. Therefore, we check robustness of the findings by examining the sensitivity of our results to a range of demand levels. The results are presented in Online Appendix Section 4. These results show that the findings reported in Section 5 are robust to choice of level of $\theta_{i,t}$.

5. Counterfactual Analysis

To study the role of spatial surge pricing in addressing congestion and welfare in a ride-sharing platform, we constructed three alternative pricing policies, each featuring estimated parameters, and then we compared their respective performance. Recall that our price elasticity measure is $\alpha/\theta_{i,t}$. Since both these parameters can vary independent of each other, in our simulation, we either fix α (price sensitivity of demand) or $\theta_{i,t}$ (level of demand) and vary the other parameter in a systematic manner.

The first policy (*Case 1*) is a regular spot-surge pricing strategy that disregards both capacity spillovers and price smoothing: $s_{i,t} = \frac{\theta_{i,t} - FC}{2\alpha} + 0.89$; that is, price changes solely based on the demand states ($\theta_{i,t}$) and $FC_{i,t}$. The other variables such as $NC_{i,t}$, UC_t, and W do not affect the price decision. This spot pricing approach is consistent with the analytic literature (Bai et al. 2018, Cachon et al. 2017, Gurvich et al. 2016). The second policy (*Case 2*) mimics Uber's existing spatial surge pricing that accounts for both capacity spillovers and price smoothing: $s_{i,t} = \frac{\theta_{i,t} - FC}{2\alpha} + 0.89 \sum_{j=1}^{N} w_{ij} s_{j,t} - 0.000496 NC_{i,t} - 0.00000189 UC_t$. That is, by using the estimated coefficients (see Table 3), the focal price depends upon NC , UC and the neighboring zone prices. The third policy (*Case 3*) we propose is a surge pricing scheme that considers only price smoothing and ignores capacity spillovers: $s_{i,t} = \frac{\theta_{i,t} - FC}{2\alpha} + 0.89 \sum_{j=1}^{N} w_{ij} s_{j,t}$. Since prices become less elastic to a capacity shortage by considering the anticipated spillovers, Case 3 is the most aggressive pricing policy for reacting to a capacity shortage. We propose Case 3 and compare it with the other two cases since it may be a good candidate for a platform like Uber, aimed to provide reliable services. By comparing these policies, we examine the effects of price smoothing and capacity spillovers on both driver (labor) and consumer welfare.

As mentioned earlier, a ride-sharing platform's pricing policy aims not only to maximize profits but also to increase its market share. To increase its market share, the platform must provide a sufficient number of drivers to quickly serve consumers and it must simultaneously provide sufficient expected earnings for drivers. Consumers may be highly sensitive to delay or price because their cost to switch to another option is relatively low. Therefore, among other factors, utilization ratio (demand-to-supply ratio) and capacity shortage are critical performance metrics for the platform. In addition to utilization and capacity shortage, we also compare Uber's consumer surplus and driver surplus under the three pricing policies. As the platform utilizes autonomous drivers, the platform's pricing policy will affect both the consumer and the driver (labor). Furthermore, the surplus of one side will also affect participants on both sides. For example, an additional rider who joins due to high surplus may attract an additional driver by increasing expected earnings. This two-sided network effect is required for the platform's sustainable growth (Parker and Van Alstyne (2005)). In addition, regulators are also interested in these performance metrics as this industry is growing. In many cities, Uber is now required to limit the daily hours that drivers can work to protect their welfare. Therefore, it is also useful to document the surpluses for a variety of pricing policies.

To compare these policies, we set up a virtual space that consists of 25 identical zones (5×5). In each zone, demands occur randomly based on price, demand state ($\theta_{i,t}$), and exogenous consumer price sensitivity (α). Although the demand state $\theta_{i,t}$ should differ across zone and time, we keep it constant ($\theta_{i,t}$ =100) for the sake of comparison with α throughout our main analysis. Performance comparisons with different $\theta_{i,t}$ are shown in Online Appendix Section 4. $FC_{i,t}$ is initially exogenous but can change based on the price sensitivity of each capacity inferred from our estimation. We ran the simulation 100 times with each pricing policy, and calculated the performance for each run in each of the 25 zones.

5.1. Capacity Shortage and Utilization

Figures 2 and 3 illustrate the operational performance of the three pricing policies where the demand state is greater than current capacity. Overall, these plots clearly show the trade-off between effectiveness (i.e., decreasing the average capacity shortage) and efficiency (i.e., increasing the average capacity utilization). Such trade-offs are commonly seen in the queuing literature (Ou and Wein 1992), but they have not been documented in settings mediated by two-sided platforms. Due to these trade-offs, a platform's pricing decision depends on the weight assigned to effectiveness, instead of efficiency. For example, a platform whose goal is to minimize capacity shortage performs better with Case 3 for consumers while Case 1 is best for a platform primarily concerned with utilization. In addition, the pricing decision also depends on consumer's price sensitivity. Price sensitivity, along with wait-time sensitivity form the consumer's tolerance limit, which determines whether consumer demand is fulfilled by existing drivers. As the price sensitivity increases, the consumer tolerance limit decreases. The number of available drivers is the lowest in Case 2 when the limit is extremely low (i.e., $\alpha > 92$); our results are flipped when this limit increases (i.e., $\alpha \leq 92$).

We initially study the impact of price smoothing by comparing Case 1 (spot pricing) and Case 3. We find that a platform increases its effectiveness (i.e., reducing shortage) by conducting price smoothing. With spot pricing, the shortage increases with consumer's price sensitivity up to the point where no drivers are available. On the other hand, price smoothing induces drivers to capture some of the highly price-sensitive consumer, and in turn the shortage decreases. This is because price smoothing effectively incentivizes drivers to gather near the congested zone so that wait-time in the next period is reduced to within the limits. In turn, more demand may be served. Secondly, we find that as consumers get more sensitive to price, Case 3 efficiently maintains low shortage without employing more drivers. As the price sensitivity increases, price smoothing motivates existing drivers to move closer to the limits of consumers, and thus utilization increases. However, without price smoothing, utilization remains nearly stationary but the shortage decreases as price sensitivity increases, because consumers begin to leave due to lack of available drivers in their tolerance limits.

Figure 2 Average Capacity Shortage (Less is Better)

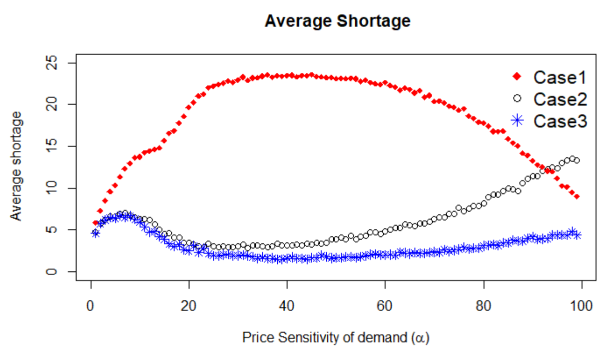
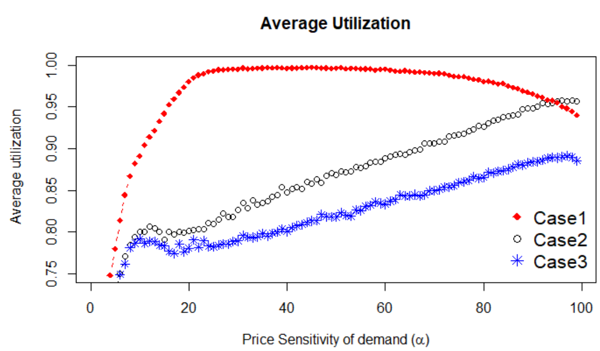


Figure 3 Average Utilization (More is Better)



We next observe the impact of capacity spillovers in price by comparing Cases 2 and 3. Overall, we find that capturing anticipated capacity spillovers not only increases utilization (high efficiency) but also increases shortage (low effectiveness). Case 2 is effective only when consumers are tolerant enough to be served by driver from another zone. However, when a platform faces a wait-time sensitive consumer, the demand can be served only by drivers within the tolerance limit, preventing drivers' chasing behavior. While this increases the drivers' probability of serving demand, the total number of matches may decrease due to loss of demand from wait-time sensitive consumers. In addition, the gap between demand and supply increases with the price sensitivity of consumers. Consumers could be strategic and choose to wait for a low price as long as they can afford the wait. However, high price sensitivity restricts this flexibility as price may not decrease to expected levels and a consumer may be able to afford a longer wait. Such restricted flexibility makes it less effective to utilize the expected capacity spillovers. Therefore, a platform that focuses only on reducing shortage is worse off by incorporating anticipated capacity spillovers.

In summary, in terms of congestion, our results are conditioned upon the price elasticity parameter. We show that (i) price smoothing increases the number of matches for most consumers by allowing drivers to move toward the congested area, and (ii) capitalization of anticipated capacity spillovers increases utilization by attempting to match demand with near-limit drivers while increasing shortage. To understand how these effects affect both the consumer and the driver side, we next analyze the surpluses of both sides in section 5.2.

5.2. Consumer Surplus and Labor Surplus

With a linear demand assumption ($d_{i,t} = -\alpha(s_{i,t} - 1) + \theta_{i,t} + \epsilon_{i,t}$), we calculate consumer surplus (CS) in a similar method to Cachon et al. (2017) and Cohen et al. (2016): $CS = \sum_t \sum_i \frac{1}{2} \cdot (\frac{\theta_{i,t}}{\alpha} + 1 - s_{i,t}) \cdot min(d_{i,t}, FC_{i,t})$ In our case, the willingness-to-pay of a consumer changes with her price sensitivity and the current demand state. For example, in a high demand state such as adverse weather conditions, consumers are willing to take a service at a higher price. Besides, consumer surplus increases with served demand and decreases with prices. We also calculate labor surplus (LS) in a similar fashion to Cachon et al. (2017): $LS = \sum_t \sum_i s_{i,t} \cdot Prob(Serving) \cdot FC_{i,t}$. The expected earning by a single driver conditional on joining is $s_{i,t} \cdot Prob(serving)$ where $Prob(Serving)$ is 1 if $s_{i,t} > 1$ and 0.5 otherwise.

Next, we report on how consumer surplus changes with consumer's price sensitivity under three pricing policies. Counterfactual results are shown in Figure 4. Our first observation is that, as consumers become less tolerant to a rise in price (i.e., α becomes larger), consumer surplus decreases under any pricing. This occurs because a high α decreases the number of matches. As discussed earlier, consumers with lower α are more flexible to choose between high price and longer wait. Thus, our findings indicate that consumers benefit more from dynamic pricing policies with their flexibility. However, the rate of decrease diminishes.

Consumers' willingness-to-pay is $\frac{\theta_{i,t}}{\alpha}+1$, which is inversely related with α. Price is also inversely related with α. These relationships diminish the rate of decrease with α. This suggests that the marginal effect of adding tolerance is salient for more price-sensitive consumers. Furthermore, by comparing policies, we find that consumers are worse off with Case 1 than with the other two policies. Price smoothing enables more consumers to be served by their strategic behavior such as delaying the ride until price drops. However, with too high price sensitivity, the difference diminishes. We also observe that the difference between Case 2 and Case 3 is minimal, which indicates that consumers do not benefit from efficiency generated by adopting anticipated capacity spillovers. This is intuitive because consumers benefit from their transactions no matter how many drivers are nearby, as long as their demand is served.

Figure 4 Consumer Surplus

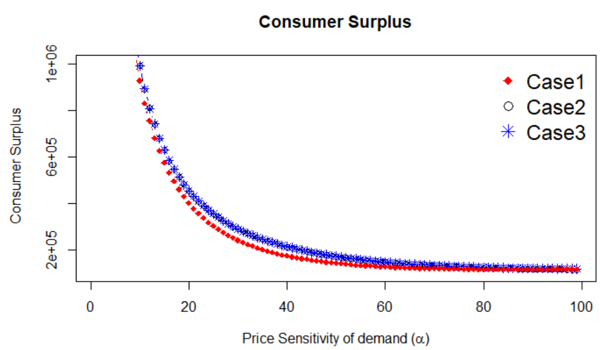
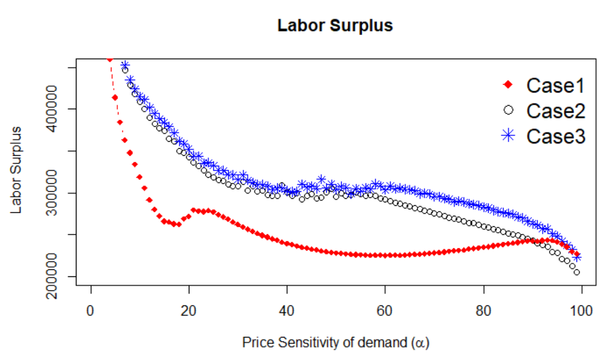


Figure 5 Labor Surplus



We next illustrate the amount of labor surplus generated under various pricing policies. Our results are presented in Figure 5. Overall, drivers benefit from surge pricing policies when consumers have low price sensitivity. Since price inversely relates to α ($s_{i,t} \propto \alpha$), the expected earnings decrease with α. Despite the negative impact of individual driver's expected earning, the total number of drivers in the focal zone has a positive impact on labor surplus. In fact, the price sensitivities of participation from NC and UC in the focal zone positively relate with α (i.e. $\gamma_{NC} = -\beta_2 \cdot 2\alpha\gamma_{FC}, \gamma_{UC} = -\beta_3 \cdot 2\alpha\gamma_{FC}$ and $\beta_2, \beta_3 < 0$). As α increases, FC increases at an increasing rate at a certain price. This diminishes the decreasing rate of labor surplus. However, once α exceeds a certain point, price becomes one (i.e., no congestion in the zone), which decreases the driver's probability of serving. This, in turn, induces a high rate of decrease. The three factors described previously when combined set up a point of inflection for labor surplus; each pricing policy has a distinct level of α at which each curve forms its point of inflection. While the point of inflection under Case 1 is formed when α is about 20, the inflection point appears when α is about 40 under Cases 2 and 3. The different positions indicate that price smoothing ($\sum_{j=1}^{N} w_{ij} s_{j,t}$) dilutes the effects of α on labor surplus. That is, by letting the focal price depend on neighboring prices, drivers' benefits become less sensitive to α. This suggests that, in general, drivers benefit more with less shortage achieved by price smoothing. Nonetheless, unlike consumers, drivers may benefit from both effectiveness and efficiency, since driver's expected earnings may decrease as utilization decreases. Drivers, counterintuitively, are strictly better off with Case 3 that yields higher average utilization than Case 2. The gap is more salient when α is greater than the point of inflection. The counterintuitive phenomenon is explained by the information that price contains. Price is the only information that determines the probability of drivers acquiring riders without utilization level information. When price becomes as low as one, it provides no utilization information. Therefore, in higher alpha ranges that make price become one, the driver's expected earning reacts only to shortage but not to utilization as long as demand is present.

To summarize, and building on the conventional literature (i.e., analytical treatment based on spot pricing, termed as Case 1 in our analyses), both price and capacity spillovers have a direct effect on surpluses. In particular, (i) consumers benefit from the increase of possible matches achieved via price smoothing and (ii) drivers benefit only from effectiveness, regardless of efficiency, when the information that price provides is limited.

6. Discussion and Extensions

6.1. Spillovers Matter

Ride-sharing platforms manage their pricing policies to orchestrate spatially distributed capacity in serving demand. In this paper, by addressing the spatial dimension, we help fill a gap in empirical revenue management literature where most studies rely on the competing results from spot pricing and ignore capacity spillovers. With a spatial econometric specification, we estimate the extent to which spillovers are associated with price and analyze surpluses under different pricing policies. Our analyses provide a variety of insights for designing and managing service platforms. We show that a surge pricing strategy that utilizes price smoothing helps reduce average capacity shortage compared to a spot surge pricing strategy. Through price smoothing, platforms can effectively maintain adequate, spatially distributed capacity. Moreover, such high service rates enable the platform to benefit both consumers (riders) and drivers by increasing the ecosystem in the long-run. In other words, consumer benefits can spread to ride-share drivers (Parker and Van Alstyne 2005). In addition to the long-run positive network effect, drivers benefit from the lower shortage achieved in the short-run, even when drivers have limited access to the utilization information. Our analysis shows that although a strategy that incorporates the anticipated capacity spillovers (Case 2) generates higher average capacity utilization than the strategy without anticipated capacity spillovers (Case 3), drivers fare better in Case 3. This has important implications for platform designers. By limiting the provision of utilization information to drivers, a platform can achieve risk-less growth with a spatial surge pricing strategy.

6.2. Updating Speed Matters

We have shown that the impact of spatial pricing strategies varies with consumer's price sensitivity. Ride-sharing platforms digitize all transactions, which helps the firm to quickly determine demand and capacity. In our data set, Uber adjusts its price as quickly as 5 seconds based on the data regarding demand and capacity data. Our counterfactual analysis indicates the impact of high-frequency data in pricing strategies. Uber currently follows Case 2 as their pricing strategy. Case 3 uses a more aggressive pricing strategy and always outperforms Case 2 in all price elasticity conditions (see Figures 4 and 5). Moreover, it is clear that Case 2's performance nears that of Case 3's performance, and that Cases 2 and 3 outperform Case 1 (spot pricing). Arguably, an explanation for this closeness between Cases 2 and 3 is that the algorithm is updating the state variables (e.g., capacity in neighbouring zones) and customer demand every few seconds. Even though Case 2 is not aggressive enough, Uber's algorithm can catch up and come close to the aggressive performance (i.e., Case 3) well before the drivers and consumers can update their own decision. Typically, drivers take minutes to update their decision to participate based on Uber's algorithm, and this process is an order of magnitude slower than the algorithm. Our results show key parameters for further tuning the algorithms are α, θ, and the γs.

6.3. Limitations and Future Work

Some of the key results are driven by our assumptions. First, we assume that a driver estimates expected earnings based only on the price information. In practice, however, each driver also learns from their previous experiences in estimating the probability of participating. Thus, the results on labor surplus may change with the driver's learning about the relationship between pricing and demand. Such learning could help incorporate efficiency into the long-run labor surplus. Second, we analyze performance under the assumption that no alternatives exist for both consumers and drivers. However, in ride-sharing scenarios where drivers and consumers can easily switch to another platform, it is important to consider how efficiently a match is made since efficiency matters in the long-term participation decision of both drivers and consumers. Although competition among platforms for drivers is beyond the scope for this paper, future work should consider how competition can best be analyzed when determining the optimal pricing policy.

References

Anselin L (1988) *Spatial Econometrics: Methods and Models.* (Boston, MA: Springer Netherlands).

Bai J, So KC, Tang CS, Chen XM, Wang H (2018) Coordinating Supply and Demand on an On-Demand Service Platform with Impatient Customers. *Manufacturing & Service Operations Management* URL `https://doi.org/10.1287/msom.2018.0707`.

Baltagi BH, Song SH, Koh W (2003) Testing panel data regression models with spatial error correlation. *Journal of Econometrics* 117(1):123–150.

Belotti F, Hughes G, Piano Mortari A (2016) Spatial panel data models using Stata. *The Stata Journal* 17(1):139–180.

Benjaafar S, Ding JY, Kong G, Taylor T (2018) Labor Welfare in On-Demand Service Platforms. *Working Paper* URL `https://ssrn.com/abstract=3102736`.

Bimpikis K, Candogan O, Saban D (2016) Spatial Pricing in Ride-Sharing Networks. *Working Paper* URL `https://ssrn.com/abstract=2868080`.

Bottasso A, Conti M, Ferrari C, Tei A (2014) Ports and regional development: A spatial analysis on a panel of European regions. *Transportation Research Part A: Policy and Practice* 65:44–55.

Breusch TS, Pagan AR (1980) The Lagrange Multiplier Test and its Applications to Model Specification in Econometrics. *The Review of Economic Studies* 47(1):239.

Cachon GP, Daniels KM, Lobel R (2017) The Role of Surge Pricing on a Service Platform with Self-Scheduling Capacity. *Manufacturing & Service Operations Management* 19(3):368–384.

Camerer C, Babcock L, Loewenstein G, Thaler R (1997) Labor Supply of New York City Cabdrivers: One Day at a Time. *The Quarterly Journal of Economics* 112(2):407–441.

Can A (1992) Specification and estimation of hedonic housing price models. *Regional science and urban economics* 22(3):453–474.

Chen L, Mislove A, Wilson C (2015) Peeking Beneath the Hood of Uber. *IMC '15 Proceedings of the 2015 Internet Measurement Conference*, 495–508 (ACM Press).

Chen MK, Sheldon M (2015) Dynamic Pricing in a Labor Market: Surge Pricing and Flexible Work on the Uber Platform. *Working Paper* UCLA Anderson School of Management.

Cliff AD, Ord JK (1981) *Spatial processes : models & applications* (London: Pion Ltd).

Cohen P, Hahn R, Hall J, Levitt S, Metcalfe R (2016) Using big data to estimate consumer surplus: The case of uber. Technical report, National Bureau of Economic Research.

Cui Y, Hu M (2018) Under the Same Roof: Value of Shared Living in Airbnb. *Working Paper* URL `https://ssrn.com/abstract=3136138`.

Elhorst JP (2010) Applied Spatial Econometrics: Raising the Bar. *Spatial Economic Analysis* 5(1):9–28.

Farber HS (2015) Why you Cant Find a Taxi in the Rain and Other Labor Supply Lessons from Cab Drivers. *The Quarterly Journal of Economics* 130(4):1975–2026.

Gurvich I, Lariviere M, Moreno A (2016) Operations in the On-Demand Economy: Staffing Services with Self-Scheduling Capacity. *Working Paper* URL `https://ssrn.com/abstract=2336514`.

Ibrahim R (2018) Managing Queueing Systems Where Capacity is Random and Customers are Impatient. *Production and Operations Management* 27(2):234–250.

Jiang S, Chen L, Mislove A, Wilson C (2018) On Ridesharing Competition and Accessibility: Evidence from Uber, Lyft, and Taxi. *WWW '18 Proceedings of the 2018 World Wide Web Conference*, 863–872 (ACM Press).

Karacaoglu N, Moreno A, Ozkan C (2017) Strategically Giving Service: The Effect of Real-Time Information on Service Efficiency. *Working Paper* Harvard Business School.

Lee Lf, Yu J (2010) Estimation of spatial autoregressive panel data models with fixed effects. *Journal of Econometrics* 154(2):165–185.

LeSage J, Pace RK (2009) *Introduction to Spatial Econometrics* (Boca Raton: Taylor & Francis Group, LLC).

Li J, Granados N, Netessine S (2014) Are Consumers Strategic? Structural Estimation from the Air-Travel Industry. *Management Science* 60(9):2114–2137.

Li J, Moreno A, Zhang DJ (2016) Pros vs Joes: Agent Pricing Behavior in the Sharing Economy. *Working Paper* URL `https://ssrn.com/abstract=2708279`.

Lu Y, Musalem A, Olivares M, Schilkrut A (2013) Measuring the Effect of Queues on Customer Purchases. *Management Science* 59(8):1743–1763.

Moreno A, Terwiesch C (2015) Pricing and Production Flexibility: An Empirical Analysis of the U.S. Automotive Industry. *Manufacturing & Service Operations Management* 17(4):428–444.

Musalem A, Shang Y, Song JSJ (2017) An Empirical Study of Customer Spillover Learning About Service Quality. *Working Paper* URL `https://ssrn.com/abstract=2811774`.

Mutl J, Pfaffermayr M (2011) The Hausman test in a Cliff and Ord panel model. *The Econometrics Journal* 14(1):48–76.

Olivares M, Cachon GP (2009) Competing Retailers and Inventory: An Empirical Investigation of General Motors' Dealerships in Isolated U.S. Markets. *Management Science* 55(9):1586–1604.

Ou J, Wein LM (1992) Performance Bounds for Scheduling Queueing Networks. *The Annals of Applied Probability* 2(2):460–480.

Oum TH, Waters WG, Yong JS (1992) Concepts of Price Elasticities of Transport Demand and Recent Empirical Estimates: An Interpretative Survey. *Journal of Transport Economics and Policy* 26(2):139–154.

Parker GG, Van Alstyne MW (2005) Two-Sided Network Effects: A Theory of Information Product Design. *Management Science* 51(10):1494–1504.

Sundararajan A (2016) *The Sharing Economy* (MIT Press).

Taylor T (2018) On-Demand Service Platforms. *Manufacturing & Service Operations Management* 20(4):704–720.

Tereyağoğlu N, Fader PS, Veeraraghavan S (2018) Multiattribute Loss Aversion and Reference Dependence: Evidence from the Performing Arts Industry. *Management Science* 64(1):421–436.

Yılmaz O, Pekgün P, Ferguson M (2017) Would You Like to Upgrade to a Premium Room? Evaluating the Benefit of Offering Standby Upgrades. *Manufacturing & Service Operations Management* 19(1):1–18.

Appendix A: Proof of Equilibrium Price

The surge multiplier does not decrease below 1 even in case of low utilization, but only increases up to s^{max} in the case of congestion ($1 \geq s_{i,t} \geq s^{max}$). Therefore, we are only interested in the equilibrium when the demand state is greater than or equal to the current focal capacity (i.e. $\mathbf{E}[d_{i,t} \mid s_{i,t}=1] \geq FC_{i,t}$). Hence, we can simplify the cost function $V_{i,t}(FC_{i,t})$ as follows:

$$V_{i,t}(FC_{i,t}) = \mathbf{E}[p \cdot max(0, d_{i,t} - FC_{i,t}) + h \cdot FC_{i,t+1}] = p \cdot (\mathbf{E}[d_{i,t}] - FC_{i,t}) + h \cdot \mathbf{E}[FC_{i,t+1}]$$

Then, the first order condition (FOC) is:

$$\frac{\partial V_{i,t}(FC_{i,t})}{\partial s_{i,t}} = p \cdot \frac{\partial \mathbf{E}[d_{i,t}]}{\partial s_{i,t}} + h \cdot \frac{\partial FC_{i,t+1}}{\partial s_{i,t}} = 0$$

(i) If the endogenous demand is still greater than the focal capacity with any price $s^*_{i,t}$ (i.e. $d_{i,t}(s^*_{i,t}) \geq FC_{i,t}$), then $s^*_{i,t} = s^{max}$

(ii) However, if the demand becomes lower than the focal capacity with a certain price $s^*_{i,t}$ (i.e. $d_{i,t}(s^*_{i,t}) < FC_{i,t}$), then the FOC is:

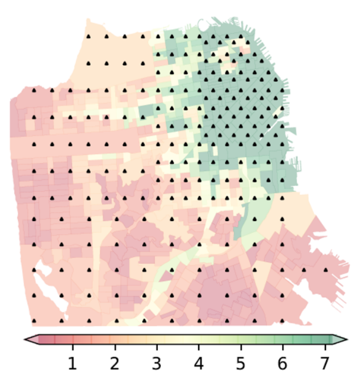
$$-p\alpha + h((FC_{i,t} + \alpha \cdot s^*_{i,t} - \theta_{i,t}) \cdot \gamma_{FC} + \alpha(\gamma_{FC} \cdot s^*_{i,t} + \delta_{FC}) + NC_{i,t} \cdot \gamma_{NC} + UC_t \cdot \gamma_{UC}) = 0$$

$$\begin{aligned} s^*_{i,t} &= \frac{-p\alpha + h\alpha\delta_{FC} - h\theta_{i,t}\gamma_{FC} + h\gamma_{FC}FC_{i,t} + h\gamma_{NC}NC_{i,t} + h\gamma_{UC}UC_t}{-2h\alpha\gamma_{FC}} \\ &= \frac{p - h\delta_{FC}}{2h\gamma_{FC}} - \frac{1}{2\alpha}FC_{i,t} - \frac{\gamma_{NC}}{2\alpha\gamma_{FC}}NC_{i,t} - \frac{\gamma_{UC}}{2\alpha\gamma_{FC}}UC_t + \frac{\theta_{i,t}}{2\alpha} \end{aligned}$$

■

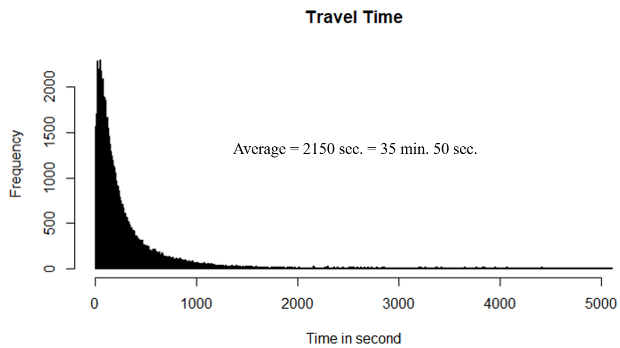
Appendix B: Description of Data

B.1. Multiple Observation Points and Average Focal Capacity Level

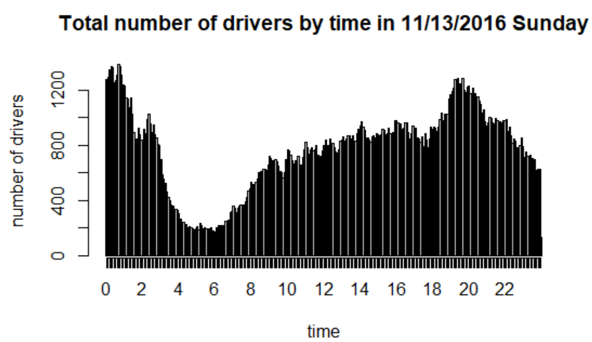


Reproduced from Jiang et al. (2018)

B.2. Travel Time from One Zone to Another Zone



B.3. Example of Trend of Total Drivers



B.4. Descriptive Statistics

	Summary Statistics				Correlation			
	Min.	Max.	Mean	Std. Dev	$s_{i,t}$	$FC_{i,t}$	$NC_{i,t}$	UC_t
$s_{i,t}$	1.0000	3.3702	1.0467	0.1458	1	-0.08	-0.13	0.21
$FC_{i,t}$	0.00	615.00	30.74	41.41		1	0.65	-0.37
$NC_{i,t}$	0.00	155.00	31.63	25.24			1	-0.64
UC_t	0.00	12714.00	7426.02	2660.31				1

BOSTON UNIVERSITY
Questrom School of Business
Working Paper

Online Appendix

---

1. Robustness Check with Different Time-intervals

Table A1 Relationship between capacity levels and surge multiplier with different time intervals

	Dependent Variable: Surge Multiplier	
	15-minute Interval (N = 313,212)	60-minute Interval (N = 78,432)
Spatial Dependency	0.90**	0.87**
	(0.012)	(0.015)
Focal Capacity	6.52×10^{-6}	2.82×10^{-5}
	(8.32×10^{-5})	(3.05×10^{-5})
Neighbor Capacity	-8.80×10^{-4}**	-3.10×10^{-4}**
	(1.82×10^{-4})	(1.41×10^{-4})
Untapped Capacity	-3.49×10^{-6}**	-1.18×10^{-6}**
	(1.35×10^{-6})	(4.08×10^{-7})
Controls	*included*	*included*
Log-likelihood	144053.10	38599.65

Notes. Controls include $Hour_t$, $Weekend_t$, and $Weather_t$. Robust standard errors clustered by zone are in parentheses. $+p < 0.1$; $*p < 0.05$; $**p < 0.01$.

2. Robustness Check with Different Ws

Table A2 Relationship between capacity levels and surge multiplier with different spatial weight matrices

	Dependent Variable: Surge Multiplier					
	W1 (N = 156,692)		W2 (N = 156,692)		W3 (N = 156,692)	
	FE	RE	FE	RE	FE	RE
Spatial Dependency	0.89**	0.89**	0.85**	0.85**	0.87**	0.87**
	(0.013)	(0.013)	(0.014)	(0.014)	(0.015)	(0.015)
Focal Capacity	1.22×10^{-5}	1.06×10^{-5}	8.27×10^{-5}	8.39×10^{-5}	1.11×10^{-4} +	1.10×10^{-4} +
	(4.99×10^{-5})	(4.82×10^{-5})	(5.32×10^{-5})	(5.15×10^{-5})	(5.68×10^{-5})	(5.52×10^{-5})
Neighbor Capacity	-4.96×10^{-4}**	-4.72×10^{-4}**	-3.85×10^{-4}**	-3.76×10^{-4}**	-6.84×10^{-4}**	-6.65×10^{-4}**
	(1.03×10^{-4})	(1.00×10^{-4})	(7.41×10^{-5})	(7.18×10^{-5})	(1.08×10^{-4})	(1.04×10^{-4})
Untapped Capacity	-1.89×10^{-6}*	-1.75×10^{-6}*	-4.90×10^{-7}	-4.30×10^{-7}	-2.42×10^{-6}**	-2.31×10^{-6}**
	(7.31×10^{-7})	(7.18×10^{-7})	(6.16×10^{-7})	(6.07×10^{-7})	(7.32×10^{-7})	(7.16×10^{-7})
Controls	*included*	*included*	*included*	*included*	*included*	*included*
Log-likelihood	75369.06	74919.32	77648.74	77199.68	75132.59	74678.40

Notes. Controls include $Hour_t$, $Weekend_t$, and $Weather_t$. Robust standard errors clustered by zone are in parentheses.
FE: Fixed Effects, RE: Random Effects
$+p < 0.1$; $*p < 0.05$; $**p < 0.01$.

3. Robustness Check with IVs

Table A3 GMM approach with $FC_{i,t-8}$ as IV

	Dependent Variable: Surge Multiplier
	30-minute Interval (N = 156,692)
Spatial Dependency	1.03**
	(0.026)
Focal Capacity	-2.14×10^{-5}**
	(3.05×10^{-5})
Neighbor Capacity	-7.22×10^{-5}**
	(1.36×10^{-5})
Untapped Capacity	-3.67×10^{-7}**
	(3.29×10^{-8})
Controls	*included*

Notes. Controls include $Hour_t$, $Weekend_t$, and $Weather_t$. Robust standard errors clustered by zone are in parentheses.
$+p < 0.1$; $*p < 0.05$; $**p < 0.01$.

4. Demand Sensitivity (Aggregate price sensitivity of demand α set at 50)

Figure A1 Average Capacity Shortage (Less is Better)

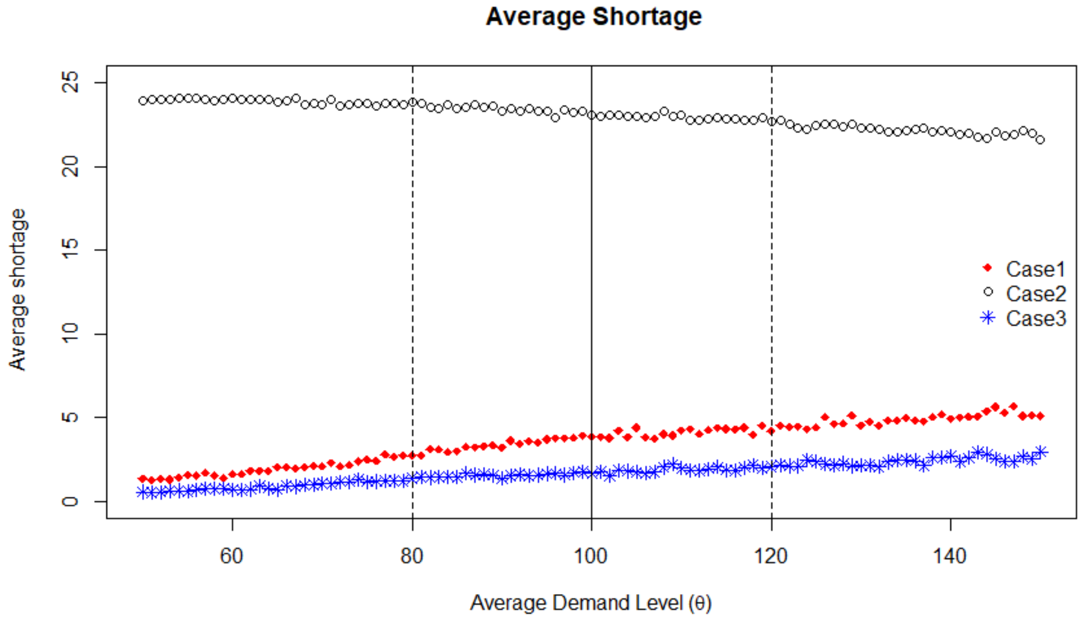


Figure A2 Average Utilization (More is Better)

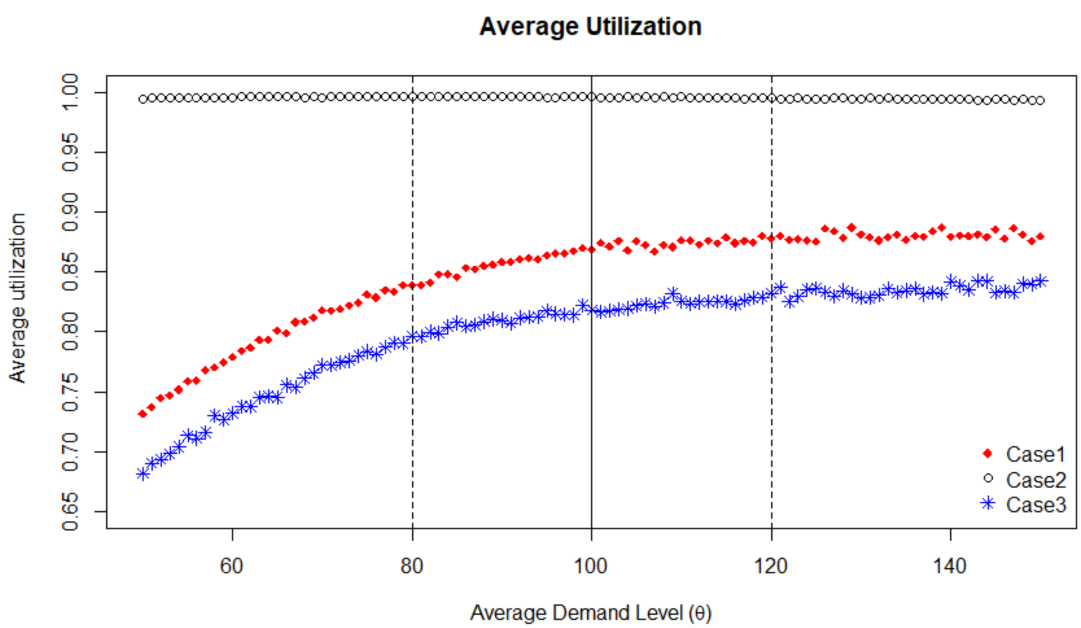


Figure A3 Consumer Surplus

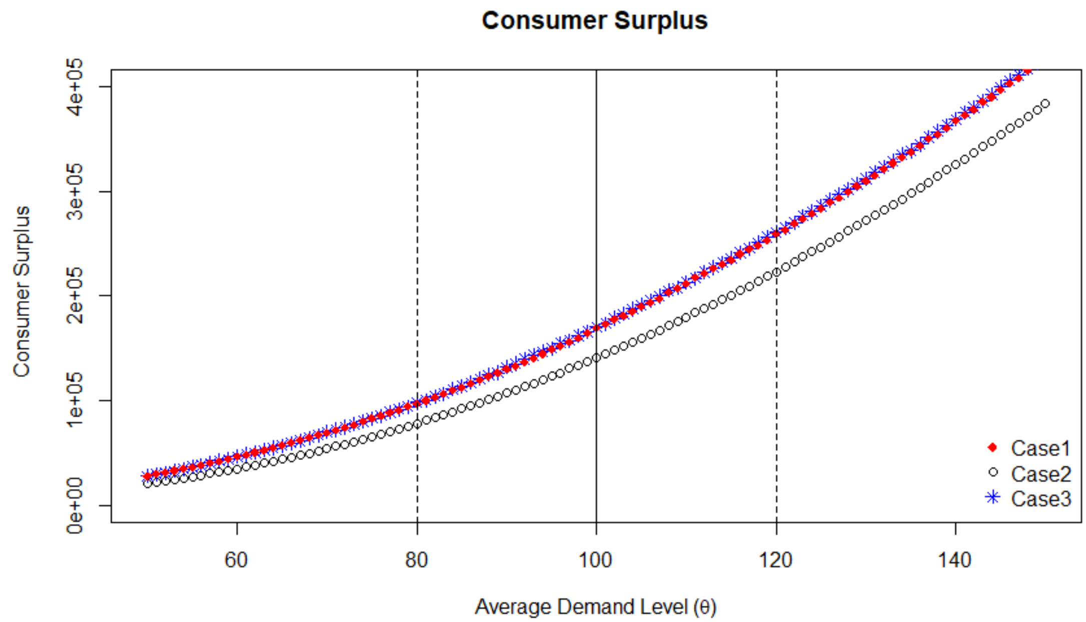


Figure A4 Labor Surplus