Peak-Load Pricing and Demand Management for Ridesourcing Platforms

Cesar N. Yahia
Department of Civil, Architectural and Environmental Engineering
The University of Texas at Austin
Austin, TX 78712-1172
Email: cesaryahia@utexas.edu
(*Corresponding author.*)

Stephen D. Boyles
Department of Civil, Architectural and Environmental Engineering
The University of Texas at Austin
Austin, TX 78712-1172
Email: sboyles@austin.utexas.edu

Word count:
5220 words text+
137 words abstract+
428 words references+
0 tables $\times$250 words (each) = 5785 words

Submission date: August 1st, 2020

ABSTRACT

We investigate dynamic pricing mechanisms in ridesourcing systems where the platform offers users multiple trip alternatives. The offered alternatives consist of the trip cost at delayed departure times, and we seek to determine the cost associated with each offered departure time. The objective of the pricing strategy is to maximize revenue while limiting the anticipated (future) demand profile to the available driver supply. Thus, the pricing strategy depends on a probabilistic characterization of predicted spatial and temporal demand patterns. In contrast to equilibrium-based methods, we use transient analysis to evaluate state-dependent pricing policies. Simulation results using Lyft rides illustrate the trade-off between maximizing revenue and limiting future demand peaks to the available supply. In addition, we show that as the user's value of time increases, the ability of the pricing mechanism to minimize peaks in demand decreases.

Keywords— network modeling, ridesourcing, pricing, demand management

1 INTRODUCTION

In ridesourcing systems, the spatiotemporal mismatch between supply and demand results in operational inefficiencies and reduced quality of service. To limit the adverse impact of the supply-demand mismatch, platforms use control levers such as surge pricing, rewards for drivers, and reservations. During peak hours, surge pricing reduces the supply-demand mismatch by inhibiting passenger demand and attracting additional drivers to the surge location. However, surge pricing is controversial *(12, 18)*. For example, drivers chasing the surge may reach the surge location after demand subsides while leaving passengers in other locations without service. To address such surge pricing drawbacks, we investigate alternative pricing policies where passengers in areas with high demand are offered the option to delay their trip in exchange for a reduced cost.

Recent research on pricing in ridesourcing systems focuses on evaluating the optimal trip cost under supply-demand equilibrium *(1, 3, 5, 13, 16, 17)*, analyzing operational inefficiencies attributed to pricing *(14, 18)*, and determining pricing strategies in transient (non-equilibrium) systems *(9)*. The majority of existing studies analyze equilibrium conditions within time periods where driver supply, passenger demand, or trip costs are time invariant. However, since supply and demand patterns vary rapidly across time, ridesourcing systems may never attain equilibrium *(6)*. The proposed pricing strategy focuses on the transient nature of ridesourcing dynamics *(15)*. We predict a time-dependent probabilistic characterization of anticipated ride requests. Then, when users request a ride, we use the demand predictions to compute the cost of each offered departure time alternative.

In particular, we consider that the platform dynamically provides users with multiple ride options, where each ride alternative consists of the trip cost at a delayed departure time. In turn, the passengers evaluate the utility of offered alternatives, and a multinomial logit model (MNL) is used to represent the probability that a passenger selects a specific alternative given its cost and associated delay. To determine future demand peaks, we use a probabilistic characterization of anticipated spatiotemporal demand patterns. Then, given a model for passenger choice and the anticipated demand, we evaluate the trip cost for each offered departure time using an optimization problem that maximizes platform revenue subject to constraints that stagger demand peaks. The pricing policy is state-dependent, and it is successively implemented as ride requests appear across time.

The remainder of this article proceeds as follows: In Section 2 we review related work on pricing in ridesourcing systems. Section 3 presents the system model and the demand processes. Section 4 discusses departure time choice and its impact on anticipated demand. Section 5 discusses the platform pricing policy. Section 6 demonstrates the impact of the proposed pricing strategy using Lyft rides in Manhattan. Section 7 concludes the article.

2 RELATED WORK

The majority of existing literature on pricing in ridesourcing systems investigates the role of surge pricing in alleviating or worsening operational inefficiencies. In general, these studies can be classified as either equilibrium-based evaluation of optimal prices or data-driven investigation of pricing inefficiencies.

Modeling ridesourcing systems as two-sided markets, researchers examined the impact of prices on the equilibrium between earning-sensitive drivers and price-sensitive passengers *(1)*. In this approach, the prices, demand rate, and expected supply are fixed across different time-periods. Thus, the steady-state equilibrium is assumed to hold within each time period where the optimal price is determined. Alternative steady-state equilibrium methods include: the analysis of threshold-based dynamic pricing strategies, where the prices are determined by the number of idle drivers *(3)*; spatial pricing across a network of regions *(5, 17)*; and the use of pricing to alleviate system inefficiencies such as matching drivers to distant passengers at high demand levels *(7, 14, 17)*.

While equilibrium-based methods provide valuable strategic-level insights into supply and demand management *(2)*, their value may be limited in operational analysis where the system parameters vary rapidly. In the context of driver rebalancing, it was shown that the time needed to converge to a steady-state equilibrium is on the order of 10 hours *(6)*. Thus, since parameters and system characteristics vary over a much shorter time scale, transient (non-equilibrium) methods are needed for operational decisions. Recently, transient analysis of ridesourcing systems resulted in novel pricing strategies where the platform may incur losses over short time periods *(9)*; the authors emphasize that such policies can not be evaluated using time-invariant steady-state methods.

In addition to model-based analysis, pricing was further examined using data-driven approaches. Notably, by analyzing the spatial variation in the mismatch between supply and demand (search frictions), it was shown that the future earnings of drivers starting at the same location differ significantly based on the assigned destination *(18)*; consequently, there is a need for "destination invariant" pricing mechanisms where drivers starting their trip at the same location and the same time have equal expected future income *(8)*. Other data-driven methods include the prediction of future surge pricing patterns to inform driver and rider decisions *(4)*.

In this research, we propose a pricing mechanism that induces users to travel during time periods when the predicted demand is low relative to the available supply. Instead of surge pricing, we offer users multiple departure time alternatives such that the price of each alternative depends on the future spatiotemporal demand patterns. In particular, given a multinomial logit model representing user choice, the prices are determined using an optimization problem that maximizes platform revenue subject to constraints that staggers future peaks in demand. As opposed to existing equilibrium-based methods, we focus on state-dependent pricing using *transient* analysis of ridesourcing dynamics. In other words, instead of assuming steady-state conditions within successive time periods, we implement real-time pricing that reacts to the current and predicted stochastic system state. Moreover, in contrast to origin-based pricing strategies, the proposed mechanism depends on both spatial and temporal components of the predicted demand.

3 SYSTEM MODEL

The ridesourcing platform aims to price trip alternatives for ride requests that initiate in a geographic region $r \in \mathcal{R}$ (where $\mathcal{R}$ is the set of regions). As illustrated in Figure 1, we assume that the platform dynamically updates the offered alternatives at the beginning of regular time intervals $\mathcal{U} = \{[u_l, u_{l+1}), [u_{l+1}, u_{l+2}), \ldots\}$. For example, at time u_l, the platform evaluates alternatives that will be offered to ride requests that will initiate during $[u_l, u_{l+1})$. Each alter-

native consists of a delayed departure time $\tau \in \mathcal{T} = \{\tau_1, \tau_2, \ldots, \tau_M\}$ within the time horizon $\mathsf{T}_{l+1} = [u_{l+1}, u_{l+1} + T]$ and an associated trip cost. The offered departure times $\tau \in \mathcal{T}$ do not have to coincide with end points of time intervals in $\mathcal{U}$.

Then, after the ride requests that initiate during $[u_l, u_{l+1})$ choose their trip departure time and cost, the platform generates a new set of alternatives (at time u_{l+1}) for ride requests that will initiate during $[u_{l+1}, u_{l+2})$. Similar to ride requests that previously initiated, the ride requests that initiate during $[u_{l+1}, u_{l+2})$ will be offered a new set of departure times $\tau \in \mathcal{T}'$ within the time horizon $\mathsf{T}_{l+2} = [u_{l+2}, u_{l+2} + T]$ and an associated trip cost for each departure time. The alternatives offered to ride requests that initiate during $[u_l, u_{l+1})$ are different from those offered to requests that initiate during $[u_{l+1}, u_{l+2})$, where this difference reflects variation of the system state between the time horizons T_{l+1} and T_{l+2}.

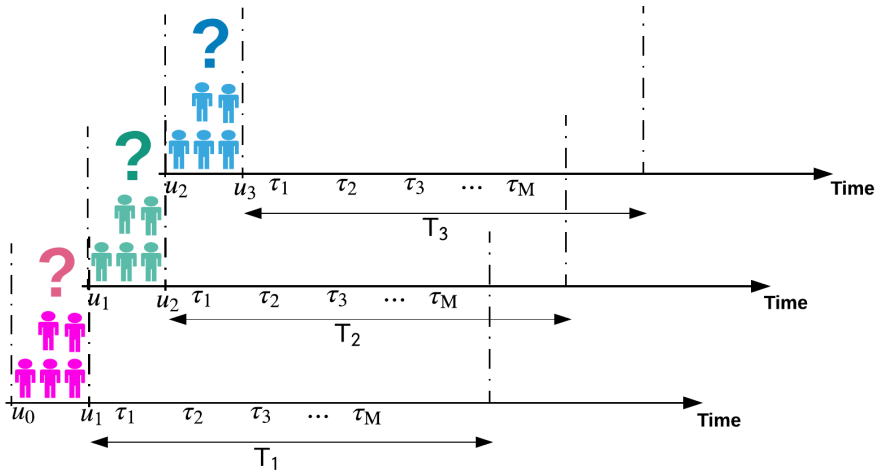

FIGURE 1 Time-dependent rolling horizon pricing mechanism

Since the same pricing procedure is repeatedly used for ride requests that initiate in any time interval $[u_l, u_{l+1}) \in \mathcal{U}$, we restrict our analysis to requests that initiate during $[u_0, u_1)$. For those rides, the offered departure time alternatives $\tau \in \mathcal{T} = \{\tau_1, \tau_2, \ldots, \tau_M\}$ are within the time horizon $\mathsf{T} = [u_1, u_1 + T]$.

In the subsequent analysis, we determine the trip cost of each departure time alternative $\tau \in \mathcal{T}$ based on the anticipated system state during T. We start by describing the predicted demand in Section 3.1. Then, in Section 3.2 we analyze the impact of the demand patterns on the shortage in supply (change in idle drivers), and we define the *load* in a region as the process describing lost idle drivers. The resulting variation in idle drivers informs pricing strategies in Section 5.

In more detail, the *load process* that informs pricing decisions is derived from the anticipated trips that start or end in region r within T. We categorize those trips into *future* or *past* depending on whether the ride request is received prior to u_0 (past) or within T (future). Note that requests received prior to u_0 may start their trip within T due to users delaying their ride. Section 3.1 discusses past and future processes. Moreover, the users for which we are currently evaluating departure time alternatives (i.e., the users that will appear during $[u_0, u_1)$) are referred to as *now* users; Section 4 describes their choices and their impact on the load process.

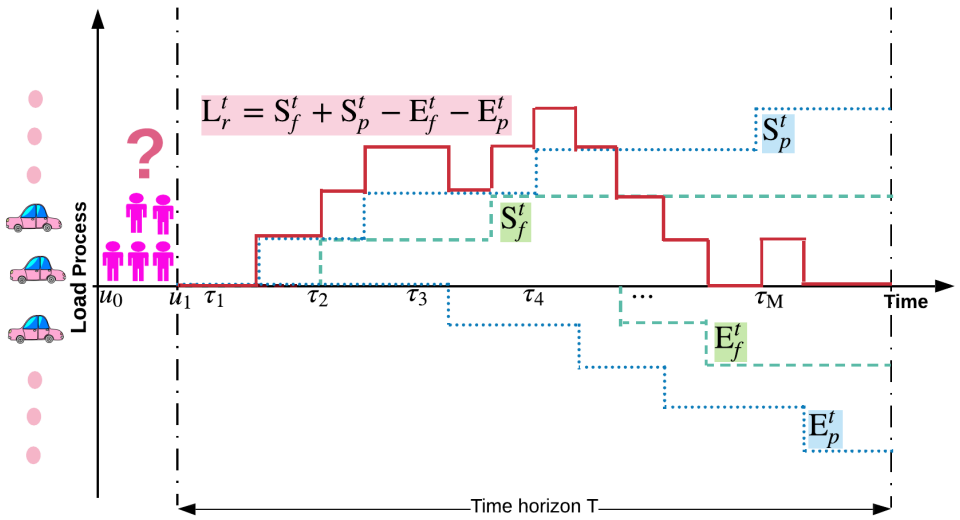

FIGURE 2 System model characterizing time-dependent ridesourcing dynamics in a region (zone) $r \in \mathcal{R}$. S_f^t represents the cumulative number of trips that start in r by time t and correspond to ride requests received in the *future* within T. E_f^t represents the cumulative number of trips that end in r by time t and correspond to ride requests received in the *future* within T. S_p^t represents the cumulative number of trips that start in r by time t and correspond to *past* ride requests that are received prior to u_0 (those rides start within T even though the request is received prior to u_0). E_p^t represents the cumulative number of trips that end in r by time t and correspond to *past* ride requests that are received prior to u_0 (those rides end within T). The load process is $L_r^t = S_f^t + S_p^t - E_f^t - E_p^t$

3.1 Prediction of Demand Processes

We proceed by describing further the system state and the spatiotemporal demand during the time horizon T. The predicted demand processes dictate the supply-demand mismatch and the resulting shortage in idle drivers (high load). In Section 5, we use these predictions to compute the trip price at each offered departure time.

As previously discussed, apart from *now* users, the demand during T has two components: (1) *past* demand that corresponds to ride requests received before u_0, and (2) *future* demand that corresponds to ride requests that will be received during T. In the following demand characterization, we assume future ride requests that appears in T do not delay their trip start time; this assumption ensures computational tractability and it is conservative in that it represents a worst case future demand scenario from the perspective of users that request a ride between $[u_0, u_1)$.

3.1.1 Future Demand

First, we focus on future demand corresponding to requests received within T. For any pair of regions $i, j \in \mathcal{R}$, we consider that future ride requests for users traveling between i and j will appear according to a Poisson process. In addition, we assume that the platform can estimate the rate of ride requests $\{\lambda_{ij}^t : t \in \mathsf{T}\}$. For simplicity, we consider that the rate λ_{ij} is fixed within the horizon T; however, the proposed mechanism can be easily generalized to cases with a time-dependent ride request rate. We also assume that the ride duration will be generally distributed

such that the service time distribution for rides between i, j is denoted by $g_{ij}(\cdot)$ and its cumulative density function is $G_{ij}(\cdot)$.

Thus, at time u_0 and for any region $r \in \mathcal{R}$, the platform can characterize two different predicted demand process $\{\mathrm{S}_f^t, \mathrm{E}_f^t : t \in \mathsf{T}\}$ associated with *future* ride requests that will appear during T. These processes are stochastic since they are determined by ride requests that appear according to a Poisson process and ride durations that are generally distributed. Moreover, these processes depend on the spatial distribution of demand across the regions in $\mathcal{R}$. The process S_f^t represents the *cumulative* number of future rides that will start in region $r \in \mathcal{R}$ by time $t \in \mathsf{T}$. A ride *starts* when the driver is assigned to fulfill the trip (i.e., the driver is no longer idle). On the other hand, E_f^t represents the *cumulative* number of rides that end in region $r \in \mathcal{R}$ by time $t \in \mathsf{T}$ (once a trip ends, region r would gain an idle driver). The processes $\{\mathrm{S}_f^t, \mathrm{E}_f^t : t \in \mathsf{T}\}$ are illustrated in Figure 2.

We assume that the processes $\{\mathrm{S}_f^t, \mathrm{E}_f^t : t \in \mathsf{T}\}$ are unbounded. An equivalent assumption is that all ride requests can be served. Thus, we may observe that the predicted number of trips that start in r is greater than the number of available idle drivers throughout the time horizon; in practice, this would correspond to distant drivers from an external region being dispatched to serve requests that start in r. In other words, the demand processes corresponds to trips starting in r or trips ending in r (even if the driver had to be dispatched from an external region to serve requests in r).

3.1.2 Previously Observed Demand

In addition to the future demand, we assume that the platform can accurately determine the trip start time and duration for *previously* observed ride requests (i.e., the platform has full information on ride requests received prior to time u_0). Thus, for each region $r \in \mathcal{R}$, the platform can characterize *deterministic* processes $\{\mathrm{S}_p^t, \mathrm{E}_p^t : t \in \mathsf{T}\}$ corresponding to the cumulative number of starts/ends that occur during T given that the request was received prior to time u_0. S_p^t represents prior ride requests that start in region r by time $t \in \mathsf{T}$. Similarly, E_p^t represents prior ride requests that end in region r by time $t \in \mathsf{T}$. We emphasize that $\{\mathrm{S}_p^t, \mathrm{E}_p^t : t \in \mathsf{T}\}$ are restricted to starts/ends within T and that requests received prior to u_0 may start their trip within T due to users delaying their ride

3.2 Predicted Load Process

Given these demand processes, we can define the *load process* L_r^t, where L_r^t corresponds to the change in the number of idle drivers between u_1 and $t \in \mathsf{T}$. In particular, we can express L_r^t in Equation 1 as the number of trips ending in r subtracted from the number of trips that start in r. Observe that if the trips starting in region r is greater than the trips ending in region r the load will increase; thus, large load values indicate that there is a net decrease in idle driver that results from the spatiotemporal demand patterns. Note that L_r^t is independent of the users that appear between $[u_0, u_1)$. In other words, L_r^t is either caused by future demand or prior demand such that it is independent of the users we seek to price. The load process L_r^t is illustrated in Figure 2.

$$\mathrm{L}_r^t = \mathrm{S}_f^t + \mathrm{S}_p^t - \mathrm{E}_f^t - \mathrm{E}_p^t \tag{1}$$

In Section 5, we will use the expected value of L_r^t to compute the price of each offered departure time alternative. In particular, the pricing strategy aims to disperse users that initiate between $[u_0, u_1)$ away from periods with high expected load $\mathbb{E}\left[\mathrm{L}_r^t\right]$. Thus, we proceed by evaluating $\mathbb{E}\left[\mathrm{L}_r^t\right]$ given in Equation 2.

$$\mathbb{E}\left[\mathrm{L}_r^t\right] = \mathbb{E}\left[\mathrm{S}_f^t\right] + \mathrm{S}_p^t - \mathbb{E}\left[\mathrm{E}_f^t\right] - \mathrm{E}_p^t \tag{2}$$

3.2.1 Expected Number of Future Ride Requests that Start in Region $r \in \mathcal{R}$

We start by deriving $\mathbb{E}\left[\mathrm{S}_f^t\right]$, the expected number of future ride requests that start in r by time t. Since future ride requests are received according to a Poisson process, the expected number of trips starting in r by time t is given in Equation 3. Observe that $\mathbb{E}\left[\mathrm{S}_f^t\right]$ is time-dependent indicating lost idle drivers as time progresses.

$$\mathbb{E}\left[\mathrm{S}_f^t\right] = \sum_{j \in \mathcal{R}} \lambda_{rj}(t - u_1) \tag{3}$$

3.2.2 Expected Number of Future Ride Requests that End in Region $r \in \mathcal{R}$

Next, we derive $\mathbb{E}\left[\mathrm{E}_f^t\right]$, the expected number of future ride requests that end in r by time t. Recall that we assume all requests could be served. In addition, for demand traveling from $j \in \mathcal{R}$ to r, we assume that future ride requests will be received according to a Poisson process with rate λ_{jr} and that the ride duration has a general distribution $g_{jr}(\cdot)$ with the CDF being $G_{jr}(\cdot)$. In the following, we use a graphical approach to show that E_f^t has a time-dependent Poisson distribution and $\mathbb{E}\left[\mathrm{E}_f^t\right]$ is its time-dependent mean *(11)*.

In Figure 3, let x_j denote the trip start time (ride request initiation) of the j^{th} future user that appears according to the Poisson process. Note that the trip may start in an external region (provided it starts during T as a future ride request). Moreover, let s_j denote the corresponding service time for the j^{th} future user. As shown in Figure 3, we observe that (x_j, s_j) is a random point in the two-dimensional plane $[u_1, u_1 + T] \times [0, \infty)$ that represents the trip start time and service duration. Thus, for any two-dimensional set S in $[u_1, u_1 + T] \times [0, \infty)$, the number of points in the set represents random sampling of the ride requests Poisson process; therefore, the number of points in the set S is *Poisson distributed*. We also know that disjoint two-dimensional sets correspond to independent sampling of a Poisson process; this implies that the number of points in each set is independent of other disjoint sets.

Furthermore, if we isolate an infinitesimal two-dimensional square set with an area $ds(dx)$, we can see that the mean number of points in that set is $\lambda_{jr}dx\left(g_{jr}(s)(ds)\right)$. Thus, for any two-dimensional set S, the intensity of the two-dimensional Poisson distribution is $\lambda_{jr}g_{jr}(s)$. In other words, the distribution of points defined as (arrival time, service duration) is Poisson over the two-dimensional space, and the *expected* number of points for any set S is given by $\int_S \lambda_{jr}g_{jr}(s)dsdx$.

As a result, to determine the *expected* number of arrivals from region j, we evaluate the integral $\int_S \lambda_{jr}g_{jr}(s)dsdx$ over the shaded area illustrated in Figure 3. This shaded area represents

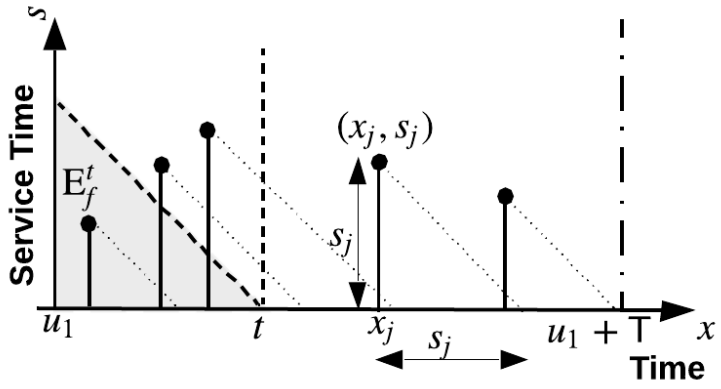

FIGURE 3 Service time vs. arrival time for future rides that are received after time u_1. The dotted diagonal lines represent the decrease in remaining service time as the user is being served. For any time t, the number of users that have completed service is the number of points in the shaded area. For all such points, the intersection of the associated dotted diagonal line with the x-axis is less than t. The shaded area also corresponds to users that are served by time t in a transient $M/GI/\infty$ queue that starts empty at time u_1.

trips that started in j and have completed in region r prior to time $t \in \mathsf{T}$. In particular, for each (arrival time, service time) pair associated with a specific user, the diagonal line represents the decrease in remaining service time as the user is being served. Note that for all users in the shaded area, the corresponding diagonal line intersects the x-axis prior to time t; this indicates that the trip terminates in region r before time t.

In addition, observe that evaluating the integral $\int_S \lambda_{jr} g_{jr}(s) ds dx$ over the shaded area is equivalent to calculating the expected number of served users in a transient $M/GI/\infty$ queue that starts empty at u_1, where the $M/GI/\infty$ queue has an arrival rate λ_{jr} and a general service distribution g_{jr} (the queue has infinite servers since all users can be served).

Then, to compute $\mathbb{E}\left[\mathrm{E}_f^t\right]$, the expected number of total trips that start in *any region* and end in r by time t, we sum the integral $\int_S \lambda_{jr} g_{jr}(s) ds dx$ across all regions $j \in \mathcal{R}$ (where the integral is evaluated using the bounds of the shaded area). The resulting expression for $\mathbb{E}\left[\mathrm{E}_f^t\right]$ is given in Equation 4. Similar to $\mathbb{E}\left[\mathrm{S}_f^t\right]$, we observe that $\mathbb{E}\left[\mathrm{E}_f^t\right]$ is time-dependent indicating the change in load across time.

$$\mathbb{E}\left[\mathrm{E}_f^t\right] = \sum_{j \in \mathcal{R}} \int_{u_1}^{t} \int_{0}^{t-x} \lambda_{jr} g_{jr}(s) ds dx = \sum_{j \in \mathcal{R}} \lambda_{jr} \int_{0}^{t-u_1} G_{jr}(x) dx \tag{4}$$

4 PASSENGER PRICE AND DEPARTURE TIME CHOICE

In this section, we discuss the departure time choice of users that initiate *now* between u_0 and u_1. We also examine the impact of the choices on the future mismatch between supply and demand. In Section 5, we use the passenger choices and their impact on future supply-demand to determine the price of each offered departure time alternative.

4.1 The Multinomial Logit Model

The probability $p_k(c_k, \tau_k)$ of a passenger choosing a particular departure time alternative $\tau \in \mathcal{T} = \{\tau_1, \tau_2, \ldots, \tau_M\}$ is determined by random utility theory. In particular, we use a multinomial logit (MNL) model. The MNL model and the subsequent pricing optimization problem in Section 5 use the time-dependent surcharge c_k instead of the full trip cost, where the full trip cost consists of the surcharge added to a base fare. However, since the base fare for each user is time-invariant (depending on factors such as the length of the trip, operational costs, type of service etc.), it does not factor into the departure time choice. In other words, the base fare would be the same for different departure time alternatives and the difference in cost is determined solely by the time-dependent surcharge. Moreover, while the base fare differs across users, the same surcharge is assigned for all users that choose a specific departure time alternative.

In the MNL model, the utility of departing at time τ_1 and paying a surcharge c_1 is $V_1 = 0$. On the other hand, the utility of departing at time τ_k is $V_k = \beta_c a_k + \beta_d d_k$. The term $d_k = \tau_k - \tau_1$ is the delay that results from departing at τ_k instead of τ_1. Thus, the parameter β_d is negative and it corresponds to the sensitivity of a user to the delay. Similarly, the term $a_k = c_1 - c_k$ is the difference in surcharge between the alternative at τ_1 and the alternative at τ_k. In the subsequent pricing optimization formulation, we restrict a_k to be greater than zero such that a user only delays their trip if the future surcharge c_k is less than the current surcharge c_1. In other words, a_k corresponds to user savings, and the parameter β_c is positive indicating an increase in utility with savings. We assume that the platform can estimate β_c and β_d. Then, the MNL probabilities $p_k(c_k, \tau_k)$ are given in Equations 5 and 6. For brevity, we denote $p_k(c_k, \tau_k)$ as p_k.

$$p_1 = \frac{1}{1 + \sum_{\tau_j \in \mathcal{T} \setminus \{\tau_1\}} e^{\beta_c a_j + \beta_d d_j}} \tag{5}$$

$$p_k = \frac{e^{\beta_c a_k + \beta_d d_k}}{1 + \sum_{\tau_j \in \mathcal{T} \setminus \{\tau_1\}} e^{\beta_c a_j + \beta_d d_j}} \qquad \forall \tau \in \mathcal{T} \setminus \{\tau_1\} \tag{6}$$

4.2 Impact of Choices on the Load Process

Given the probability p_k that users arriving *now* between $[u_0, u_1)$ select to depart at time τ_k, we determine the impact of these choices on the load process. Similar to the analysis approach of future rides in Section 3, we determine the number of trips that start/end in T given that the ride request will be received during $[u_0, u_1)$ and the departure time choice follows from the MNL model. In particular, the additional load δ_r^t at time $t \in \mathsf{T}$ that is associated with users that appear between $[u_0, u_1)$ is shown in Equation 7. The term S_n^t represents the cumulative number of trips that start by time t and correspond to users requesting service between $[u_0, u_1)$. Similarly, the term E_n^t represents the cumulative number of trips that end by time t and correspond to users requesting service between $[u_0, u_1)$. The expected additional load $\mathbb{E}\left[\delta_r^t\right]$ is given in Equation 8. In what follows, we evaluate $\mathbb{E}\left[\delta_r^t\right]$.

$$\delta_r^t = \mathrm{S}_n^t - \mathrm{E}_n^t \tag{7}$$

$$\mathbb{E}\left[\delta_r^t\right] = \mathbb{E}\left[\mathrm{S}_n^t\right] - \mathbb{E}\left[\mathrm{E}_n^t\right] \tag{8}$$

Note that the departure time alternatives and their prices are generated at time u_0. Thus, the platform needs to characterize the anticipated ride request rate and trip duration for requests that appear in $[u_0, u_1)$. Similar to the future ride requests, the platform would estimate the arrival rate λ_{rj} for trips between regions $r, j \in \mathcal{R}$. In addition, the ride duration also follows a general distribution $g_{rj}(\cdot)$ with CDF $G_{rj}(\cdot)$.

4.2.1 *Expected Number Ride Requests that Start in Region $r \in \mathcal{R}$ for Users Requesting Service between $[u_0, u_1)$*

Then, we derive $\mathbb{E}\left[S_n^t\right]$, the expected cumulative number of trips that start before time t in r and are associated with requests that will be received during $[u_0, u_1)$. The expression for $\mathbb{E}\left[S_n^t\right]$ is given in Equation 9; we arrive at this expression by calculating the *total* expected number of requests received between $[u_0, u_1)$ and multiplying by the probability that those requests choose to depart prior to time $t \in \mathsf{T}$ (i.e., they choose a departure time $\tau \in \mathcal{T} = \{\tau_1, \tau_2, ..\tau_M\}$ that is less than t).

$$\mathbb{E}\left[S_n^t\right] = \left[\sum_{j \in \mathcal{R}} \lambda_{rj}(u_1 - u_0)\right] \sum_{\tau_k \in \mathcal{T}: \tau_k \leq t} p_k \tag{9}$$

4.2.2 *Expected Number Ride Requests that End in Region $r \in \mathcal{R}$ for Users Requesting Service between $[u_0, u_1)$*

Similarly, we derive $\mathbb{E}\left[E_n^t\right]$, the expected number of trips that end by time t in r and are associated with requests that will be received during $[u_0, u_1)$. The expression for $\mathbb{E}\left[E_n^t\right]$ is given in Equation 10. To obtain Equation 10, we multiply the expected *total* number of users that would be received between $[u_0, u_1)$ and are destined to r by the probability that their trip ends before time $t \in \mathsf{T}$; in turn, the probability that the trip ends before time t is the product of the probability that the trip starts prior to time t and the probability that the ride duration is less than the difference between t and the start time.

$$\mathbb{E}\left[E_n^t\right] = \lambda_{rr}(u_1 - u_0) \sum_{\tau_k \in \mathcal{T}: \tau_k \leq t} p_k G_{rr}(t - \tau_k) \tag{10}$$

5 PEAK-LOAD PRICING

In this section, we describe the platform pricing strategy. In particular, we determine the trip cost associated with each departure time alternative $\tau \in \mathcal{T} = \{\tau_1, \tau_2, \ldots, \tau_M\}$ offered to users that initiate between $[u_0, u_1)$. The pricing mechanism aims to maximize platform revenue while reducing peaks in the load process.

5.1 Platform Revenue Maximization

The platform optimization problem is shown in formulation 11–16. The objective of the platform pricing strategy is to maximize revenue. In particular, the expected revenue *per ride* associated

with the pricing scheme, as given in the maximization objective, is the sum across alternatives of the surcharge multiplied by the choice probability (i.e., $\sum_{\tau_k \in \mathcal{T}} c_k p_k$).

The pricing strategy also aims to restrict the load process, where an increase in the load process indicates lost idle drivers (i.e., the cumulative number of trips starting in the region is greater than the cumulative trips ending in the region). Thus, by restricting peaks in the load process, the platform can limit the need for excess drivers that may require dispatching drivers from external regions to serve requests initiating within the pricing region. To that end, the component of the objective given by wz along with constraints 12 and 16 minimize the *increase* in load across departure time alternatives. The term w is a constant that weights the two components of the objective. Constraints 13 and 14 represent the MNL model that relates the surcharge to the choice probabilities. Constraint 15 guarantees that the savings are positive (i.e., that the future departure time alternatives have a lower surcharge). Thus, formulation 11–16 finds the optimal surcharges $\{c_k \in \mathbb{R} : \tau_k \in \mathcal{T}\}$ that maximize platform revenue, minimize peaks in the load process, and reflect user choices.

$$\max_{\substack{c_k, p_k : \tau_k \in \mathcal{T}, \\ a_k : \tau_k \in \mathcal{T} \setminus \{\tau_1\}, z}} \quad \sum_{\tau_k \in \mathcal{T}} c_k p_k - wz \tag{11}$$

$$\text{s.t.} \quad \left(\mathbb{E}\left[\mathrm{L}_r^{\tau_{k+1}}\right] + \mathbb{E}\left[\delta_r^{\tau_{k+1}}\right]\right) - \left(\mathbb{E}\left[\mathrm{L}_r^{\tau_k}\right] + \mathbb{E}\left[\delta_r^{\tau_k}\right]\right) \leq z \qquad \forall \tau_k \in \mathcal{T} \setminus \{\tau_M\} \tag{12}$$

$$p_1 = \frac{1}{1 + \sum_{\tau_j \in \mathcal{T} \setminus \{\tau_1\}} e^{\beta_c a_j + \beta_d d_j}} \tag{13}$$

$$p_k = \frac{e^{\beta_c a_k + \beta_d d_k}}{1 + \sum_{\tau_j \in \mathcal{T} \setminus \{\tau_1\}} e^{\beta_c a_j + \beta_d d_j}} \qquad \forall \tau \in \mathcal{T} \setminus \{\tau_1\} \tag{14}$$

$$a_k \geq 0 \qquad \forall \tau \in \mathcal{T} \setminus \{\tau_1\} \tag{15}$$

$$z \geq 0 \tag{16}$$

The formulation in 11–16 has a nonlinear objective function and nonlinear constraints. Specifically, the choice probabilities p_k are a nonlinear function of the surcharge decision variables c_k. In Section 5.2, we reformulate the optimization problem to arrive at a convex program. First, we reduce formulation 11–16 to an equivalent formulation where the decision variables are only p_k and z. Then, we show that (in terms of p_k and z) the objective is convex and the constraints form a convex set.

5.2 Convex Revenue Maximization Formulation Given Passenger Choice

To reformulate the problem into a convex program, we start by replacing the maximization in 11–16 with the *minimization* in 17–22. As shown below, the revised objective is in terms of a_k. Note that since the utility of alternatives with departure time $\tau_k \in \mathcal{T} \setminus \{\tau_1\}$ depends on the relative difference in cost between c_1 and $\{c_k : t \in \mathcal{T} \setminus \{t_1\}\}$, we fix c_1 as a constant. Thus, by solving for a_k in the revised formulation, we seek to find the optimal surcharge $\{c_k : \tau_k \in \mathcal{T} \setminus \{\tau_1\}\}$ relative to c_1.

The revised formulation is as follows:

$$\min_{\substack{p_k:\tau_k\in\mathcal{T},a_k:\tau_k\in\mathcal{T}\setminus\{\tau_1\},\\ z_k:\tau_k\in\mathcal{T}\setminus\{\tau_M\}}} \quad \sum_{\tau_k\in\mathcal{T}\setminus\{\tau_1\}} a_k p_k + wz \tag{17}$$

$$\text{s.t.} \quad \left(\mathbb{E}\left[\mathrm{L}_r^{\tau_{k+1}}\right] + \mathbb{E}\left[\delta_r^{\tau_{k+1}}\right]\right) - \left(\mathbb{E}\left[\mathrm{L}_r^{\tau_k}\right] + \mathbb{E}\left[\delta_r^{\tau_k}\right]\right) \le z \qquad \forall \tau_k \in \mathcal{T}\setminus\{\tau_M\} \tag{18}$$

$$p_1 = \frac{1}{1+\sum_{\tau_j\in\mathcal{T}\setminus\{\tau_1\}} e^{\beta_c a_j+\beta_d d_j}} \tag{19}$$

$$p_k = \frac{e^{\beta_c a_k+\beta_d d_k}}{1+\sum_{\tau_j\in\mathcal{T}\setminus\{\tau_1\}} e^{\beta_c a_j+\beta_d d_j}} \qquad \forall \tau \in \mathcal{T}\setminus\{\tau_1\} \tag{20}$$

$$a_k \ge 0 \qquad \forall \tau \in \mathcal{T}\setminus\{\tau_1\} \tag{21}$$

$$z \ge 0 \tag{22}$$

Claim 1. *Solving for an optimal solution to the maximization formulation 11–16 is equivalent to solving for the minimum of formulation 17–22*

Proof. We show that the objectives of the two formulations are equivalent as follows:

$$\max_{\substack{c_k,p_k:\tau_k\in\mathcal{T},a_k:\tau_k\in\mathcal{T}\setminus\{\tau_1\},\\ z_k:\tau_k\in\mathcal{T}\setminus\{\tau_M\}}} \quad \sum_{\tau_k\in\mathcal{T}} c_k p_k - wz \tag{23}$$

$$\Leftrightarrow \max_{\substack{c_k,p_k:\tau_k\in\mathcal{T},a_k:\tau_k\in\mathcal{T}\setminus\{\tau_1\},\\ z_k:\tau_k\in\mathcal{T}\setminus\{\tau_M\}}} \quad \sum_{\tau_k\in\mathcal{T}\setminus\{\tau_1\}} (c_1 - a_k)\, p_k + c_1 p_1 - wz \tag{24}$$

$$\Leftrightarrow \max_{\substack{c_k,p_k:\tau_k\in\mathcal{T},a_k:\tau_k\in\mathcal{T}\setminus\{\tau_1\},\\ z_k:\tau_k\in\mathcal{T}\setminus\{\tau_M\}}} \quad -\sum_{\tau_k\in\mathcal{T}\setminus\{\tau_1\}} a_k p_k + c_1 \cancel{\sum_{\tau_k\in\mathcal{T}} p_k}^{\,1} - wz \tag{25}$$

$$\Leftrightarrow \min_{\substack{p_k:\tau_k\in\mathcal{T},a_k:\tau_k\in\mathcal{T}\setminus\{\tau_1\},\\ z_k:\tau_k\in\mathcal{T}\setminus\{\tau_M\}}} \quad \sum_{\tau_k\in\mathcal{T}\setminus\{\tau_1\}} a_k p_k + wz \tag{26}$$

□

The revised formulation is still a non-convex optimization problem since p_k is a nonlinear function of a_k. Thus, in the subsequent reformulation 27–32, we reduce the optimization problem 17–22 into a *convex program* in terms of p_k and z. We start by rewriting the objective as a convex function. Then, we drop constraints 19–20 and add instead constraints that restrict the probabilities. Specifically, the MNL model in 19–20 ensures that the probabilities add up to 1 and are between $(0, 1)$, we add those constraints upon dropping 19–20. Then, we replace 21 with an equivalent constraint that is linear in p_k. Thus, in the space of p_k and z, we obtain a convex program.

The revised formulation is as follows:

$$\min_{\substack{p_k:\tau_k\in\mathcal{T},\\ z_k:\tau_k\in\mathcal{T}\setminus\{\tau_M\}}} \quad \frac{1}{\beta_c}\left[\sum_{\tau_k\in\mathcal{T}\setminus\{\tau_1\}} p_k\log(p_k)-\beta_d d_k p_k\right]-\frac{1}{\beta_c}(1-p_1)\log(p_1)+wz \tag{27}$$

$$\text{s.t.} \quad \left(\mathbb{E}\left[\mathrm{L}_r^{\tau_{k+1}}\right]+\mathbb{E}\left[\delta_r^{\tau_{k+1}}\right]\right)-\left(\mathbb{E}\left[\mathrm{L}_r^{\tau_k}\right]+\mathbb{E}\left[\delta_r^{\tau_k}\right]\right)\le z \qquad \forall\tau_k\in\mathcal{T}\setminus\{\tau_M\} \tag{28}$$

$$\sum_{\tau_k\in\mathcal{T}} p_k = 1 \tag{29}$$

$$0\le p_k\le 1 \qquad \forall\tau\in\mathcal{T} \tag{30}$$

$$p_k\ge e^{\beta_d d_k}p_1 \qquad \forall\tau\in\mathcal{T}\setminus\{\tau_1\} \tag{31}$$

$$z\ge 0 \tag{32}$$

Claim 2. *Solving for an optimal solution to 17–22 is equivalent to solving for the minimum of formulation 27–32*

Proof. First, we show that the two objectives are equivalent.
From Equation 19, we know that $\log(p_1) = -\log\left(1+\sum_{\tau_j\in\mathcal{T}\setminus\{\tau_1\}} e^{\beta_c a_j+\beta_d d_j}\right)$
From Equation 20, we know that $\log(p_k) = \beta_c a_k+\beta_d d_k-\log\left(1+\sum_{\tau_j\in\mathcal{T}\setminus\{\tau_1\}} e^{\beta_c a_j+\beta_d d_j}\right)$
Thus, $\log(p_k) = \beta_c a_k+\beta_d d_k+\log(p_1)$
Rearranging, we can write a_k as follows:

$$a_k = \frac{1}{\beta_c}\left[\log(p_k)-\log(p_1)-\beta_d d_k\right] \tag{33}$$

Thus, $a_k p_k = \frac{1}{\beta_c}\left[p_k\log(p_k)-p_k\log(p_1)-\beta_d d_k p_k\right]$
Then,

$$\sum_{\tau_k\in\mathcal{T}\setminus\{\tau_1\}} a_k p_k = \frac{1}{\beta_c}\sum_{\tau_k\in\mathcal{T}\setminus\{\tau_1\}}\left[p_k\log(p_k)-\beta_d d_k p_k\right]-\frac{1}{\beta_c}\log(p_1)\sum_{\tau_k\in\mathcal{T}\setminus\{\tau_1\}} p_k \;\nearrow\; 1-p_1 \tag{34}$$

$$= \frac{1}{\beta_c}\sum_{\tau_k\in\mathcal{T}\setminus\{\tau_1\}}\left[p_k\log(p_k)-\beta_d d_k p_k\right]-\frac{1}{\beta_c}(1-p_1)\log(p_1) \tag{35}$$

where equation 34 follows from the requirement that the probabilities sum to one.
This implies that:

$$\min_{\substack{p_k:\tau_k\in\mathcal{T},a_k:\tau_k\in\mathcal{T}\setminus\{\tau_1\},\\ z_k:\tau_k\in\mathcal{T}\setminus\{\tau_M\}}} \quad \sum_{\tau_k\in\mathcal{T}\setminus\{\tau_1\}} a_k p_k + wz$$

is equivalent to:

$$\min_{\substack{p_k:\tau_k\in\mathcal{T},\\ z_k:\tau_k\in\mathcal{T}\setminus\{\tau_M\}}} \quad \frac{1}{\beta_c}\left[\sum_{\tau_k\in\mathcal{T}\setminus\{\tau_1\}} p_k\log(p_k)-\beta_d d_k p_k\right]-\frac{1}{\beta_c}(1-p_1)\log(p_1)+wz$$

Furthermore, we show that constraint 21 can be expressed in terms of p_k as follows:
From Equation 33, since β_c is positive (sensitivity to savings), we know that $a_k \geq 0$ if $\log(p_k) - \log(p_1) - \beta_d d_k \geq 0$.
Thus, $a_k \geq 0$ if $\log(\frac{p_k}{p_1}) \geq \log\left(e^{\beta_d d_k}\right)$, and this implies that $a_k \geq 0$ if $p_k \geq e^{\beta_d d_k} p_1$ □

Thus, since the constraints form a convex set in p_k and z, we can show that the formulation 27–32 is a convex program by verifying that the objective is a convex function. In particular, we show that the Hessian matrix associated with the objective function is positive semidefinite.

Claim 3. *The objective function*
$\mathrm{F} = \frac{1}{\beta_c}\left[\sum_{\tau_k \in \mathcal{T}\setminus\{\tau_1\}} p_k \log(p_k) - \beta_d d_k p_k\right] - \frac{1}{\beta_c}(1-p_1)\log(p_1) + wz$
is convex

Proof. Observe that the objective is separable with respect to each decision variable $\{p_k : \tau_k \in \mathcal{T}\}, z$
Then, we can easily determine the second derivative with respect to each variable and construct the corresponding diagonal Hessian matrix as follows:
$\frac{\partial^2 \mathrm{F}}{\partial p_1^2} = \frac{1}{\beta_c}\left[\frac{1}{p_1^2} + \frac{1}{p_1}\right]$
$\frac{\partial^2 \mathrm{F}}{\partial p_k^2} = \frac{1}{\beta_c}\left(\frac{1}{p_k}\right) \qquad$ for all $\tau_k \in \mathcal{T}$
$\frac{\partial^2 \mathrm{F}}{\partial z^2} = 0$
Then the Hessian is an $(M+1)\times(M+1)$ diagonal matrix with the entries given by $\frac{\partial^2 \mathrm{F}}{\partial p_1^2}, \frac{\partial^2 \mathrm{F}}{\partial p_k^2}, \frac{\partial^2 \mathrm{F}}{\partial z^2}$.
Recall that β_c is positive since greater savings a_k (i.e., lower c_k) correspond to higher utility.
In addition, for more than one departure time alternative, all the multinomial choice probabilities $\{p_k : \tau_k \in \mathcal{T}\}$ are between $(0,1)$.
Thus, all the diagonal entries of the Hessian are non-negative; this implies that the Hessian is positive semidefinite.
Since the Hessian is positive semidefinite, the objective is convex. □

Then, after solving the convex program 27–32 for the optimal $\{p_k : \tau_k \in \mathcal{T}\}$, we can determine the associated optimal surcharge $\{c_k : \tau_k \in \mathcal{T}\setminus\{\tau_1\}\}$ using Equation 36, where Equation 36 follows from Equation 33.

$$c_k^\star = c_1 - \frac{1}{\beta_c}\left[\log(p_k^\star) - \beta_d d_k - \log(p_1^\star)\right] \qquad \forall t_k \in \mathcal{T}\setminus\{t_1\} \tag{36}$$

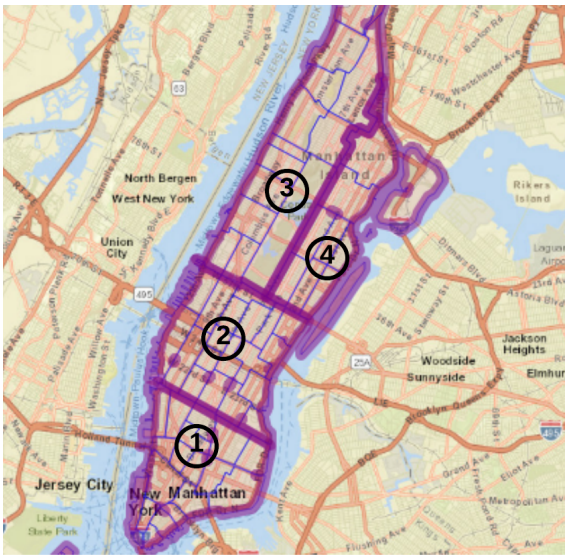

FIGURE 4 Manhattan divided into four regions.

6 DEMONSTRATIONS & NETWORK ANALYSIS

In this section, we present experimental results using data from Lyft operations in Manhattan, NYC. We use data from rides that occurred on Friday December 14th, 2018 *(10)* to estimate the model parameters. In particular, we consider trips that started between 16:00-19:00 (local time) in four regions that are shown in Figure 4, where all rides in a zone are offered the same departure time alternatives and corresponding time-dependent surcharges. We use a rolling time horizon T of 50 minutes, and the users are offered five departure times that are evenly spaced out within the horizon (i.e., there is a difference of 10 minutes between successive departure time offers). Note that we consider the pricing intervals $[u_0, u_1)$ to be 10 minutes as well.

Our primary findings suggest that as the users value of time increases, the effectiveness of the peak-load pricing strategy decreases. In addition, to control lost revenue, the platform can adjust the weight parameter w. In other words, as w increases, the platform loses more revenue in favor of shaving peaks in the load process.

6.1 System model specification and rolling horizon implementation

At any pricing interval and associated future time horizon T, we use ride request received prior to $[u_0, u_1)$ to generate S_p^t and E_p^t. Then, we use the Manhattan ride request data to determine the maximum likelihood estimator of the upcoming arrival rates λ_{ij} between regions $i, j \in \mathcal{R}$. In addition, we use the ride duration of Manhattan trips to estimate an empirical service distribution $g_{ij}(\cdot)$ with CDF $G_{ij}(\cdot)$. The arrival rate and empirical service distribution are used to evaluate cumulative starts/ends $\mathrm{S}_f^t/\mathrm{E}_f^t$ associated with ride requests that will be received within the time horizon T.

In each region, after we evaluate the load process and determine the optimal prices that will be offered to users, we consider that the Manhattan ride requests that subsequently appear in $[u_0, u_1)$ to be ground truth observed data. Then, we probabilistically delay the associated start time

of each ride based on the optimal MNL probabilities.

This process is successively repeated by first updating S_p^t and E_p^t to account for the choices of observed users $[u_0, u_1)$. Then, we analyze the subsequent pricing interval $[u_1, u_2)$ and generate a new time horizon T that begins at u_2. Note that we also discount trips that start/end at u_1 from S_p^t/E_p^t since we are now only concerned with cumulative starts/ends in the new horizon $[u_2, u_2 + T]$.

6.2 Value of time and lost revenue

We analyze the impact of parameters β_c and β_d on the pricing strategy. For any specific departure time alternative τ_k, the change in utility with savings and delay is given as $\Delta V_k = \beta_c \Delta a_k + \beta_d \Delta d_k$. Setting ΔV_k to zero, we can evaluate the trade-off between savings and delay. In particular, $\Delta V_k = 0$ implies that $\Delta a_k = -(\beta_d/\beta_c)\,\Delta d_k$. Thus, in terms of the impact on utility, a unit increase in delay is equivalent to $-(\beta_d/\beta_c)$ in additional savings (recall that β_d is negative representing sensitivity to increased delay and β_c is positive representing sensitivity to greater savings). In other words, we can consider the value of time to be VOT $= -(\beta_d/\beta_c)$.

Maximizing the platforms revenue is equivalent to minimizing the expected user savings (See Claim 1), where the expected user savings per region is $\sum_{\tau_k \in \mathcal{T}\setminus\{\tau_1\}} a_k p_k$. In particular, the expected user savings $\sum_{\tau_k \in \mathcal{T}\setminus\{\tau_1\}} a_k p_k$ corresponds to the average lost revenue *per ride* based on the choices of the users (the choice probabilities are given by $\{p_k : \tau_k \in \mathcal{T}\setminus\{\tau_1\}\}$).

In Figure 5, we evaluate $\sum_{\tau_k \in \mathcal{T}\setminus\{\tau_1\}} a_k p_k$ for each region and then average the resulting sum across regions. Thus, we compute the average lost revenue per ride given in Equation 37 at every pricing time interval. Then, we plot the lost revenue against time for different VOT instances.

$$\text{Lost Revenue} = \frac{1}{|\mathcal{R}|} \sum_{r \in \mathcal{R}} \sum_{\tau_k \in \mathcal{T}\setminus\{\tau_1\}} a_k p_k \tag{37}$$

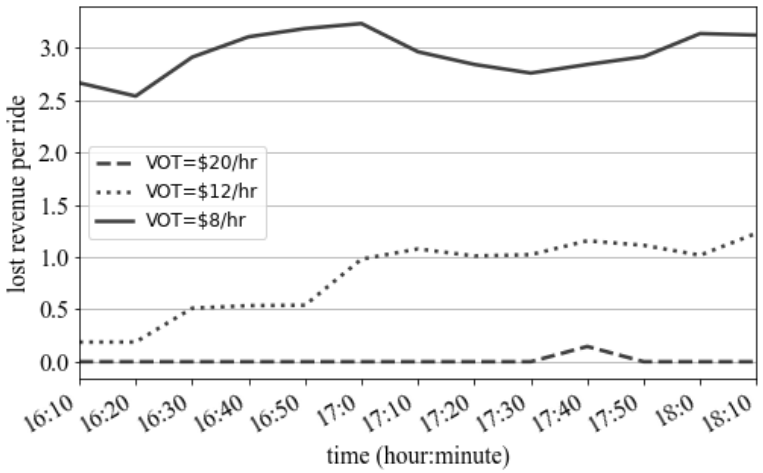

FIGURE 5 Lost revenue across time for different VOT values. The weight parameter w is set to one.

We observe that as the VOT decreases the average lost revenue increases. This indicates that users with a lower value of VOT are more likely to delay their departure time. In turn, delayed departure times result in losses to the platform. In contrast, users with high VOT choose to depart at earlier times and forgo the savings. To further incentivize high VOT users to delay the trip, the platform may increase the weight w to place greater emphasis on minimizing peaks in the load process as opposed to maximizing revenue.

In Figure 6, we illustrate the impact of the weight w on the lost revenue for a fixed VOT of \$12 per hour. We show that as the weight parameter increases in the optimization objective, the losses to the platform increase as well; this indicates that the platform prioritizes restricting peaks in load the process over generating revenue. On the other hand, when the weight is low, the lost revenue is negligible; this indicates that the platform does not offer users low cost departure time alternatives to avoid a decrease in its revenue.

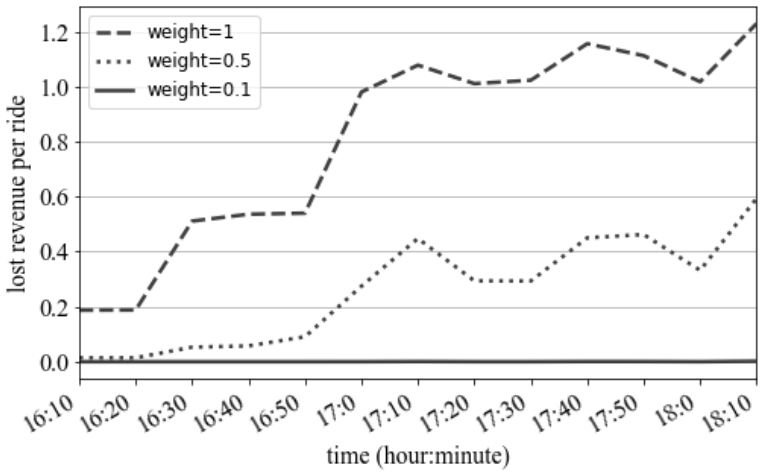

FIGURE 6 Lost revenue across time for different weight values. VOT is \$12 per hour.

7 CONCLUSION

In this article, we propose a pricing mechanism that limits peaks in demand to the available supply. In contrast to surge pricing, we offer user the option to delay their trip departure time in exchange for a reduced trip cost. Thus, by pricing different departure time alternatives, we aim to disperse users away from peaks in the load process, where an increase in the load process represents lost idle drivers.

As opposed to equilibrium-based methods that assume steady-state conditions, the proposed pricing mechanism focuses on the time-dependent system state and the associated transient probabilistic demand processes. In particular, we use a probabilistic characterization of future spatio-temporal demand to determine time periods with increased load. Then, we use the resulting load process to implement real-time pricing that reacts to the current and predicted system state.

In addition to restricting the load process, the pricing strategy aims to maximize platform revenue while representing user choices using a multinomial logit model. Simulation results using

data from Lyft rides observed in Manhattan highlight the trade-off between maximizing revenue and restricting the load process. The results also exhibit the impact of user characteristics on the performance of the pricing strategy; specifically, we observe that as the users value of time increases, the effectiveness of the pricing strategy in terms of restricting the load process is diminished.

ACKNOWLEDGMENTS

This research was supported by the National Science Foundation under the following grants: CMMI-1562291, NSF-1636154, and NSF-1826320.

AUTHOR CONTRIBUTION STATEMENT

All authors reviewed the results and approved the final version of the manuscript.

REFERENCES

[1] Bai, J., So, K., Tang, C., Chen, X., and Wang, H. (2019). Coordinating supply and demand on an on-demand service platform with impatient customers. *Manufacturing & Service Operations Management*, 21(3):556–570.

[2] Ban, X., Dessouky, M., Pang, J., and Fan, R. (2019). A general equilibrium model for transportation systems with e-hailing services and flow congestion. *Transportation Research Part B: Methodological*, 129:273–304.

[3] Banerjee, S., Riquelme, C., and Johari, R. (2016). Pricing in ride-share platforms: A queueing-theoretic approach. *SSRN 2568258*.

[4] Battifarano, M. and Qian, Z. (2019). Predicting real-time surge pricing of ride-sourcing companies. *Transportation Research Part C: Emerging Technologies*, 107:444–462.

[5] Bimpikis, K., Candogan, O., and Saban, D. (2019). Spatial pricing in ride-sharing networks. *Operations Research*, 67(3):744–769.

[6] Braverman, A., Dai, J. G., Liu, X., and Ying, L. (2019). Empty-car routing in ridesharing systems. *Operations Research*, 67(5):1437–1452.

[7] Castillo, J., Knoepfle, D., and Weyl, G. (2017). Surge pricing solves the wild goose chase. In *Proceedings of the 2017 ACM Conference on Economics and Computation*, pages 241–242.

[8] Ma, H., Fang, F., and Parkes, D. (2018). Spatio-temporal pricing for ridesharing platforms. *arXiv preprint arXiv:1801.04015*.

[9] Nourinejad, M. and Ramezani, M. (2019). Ride-Sourcing modeling and pricing in non-equilibrium two-sided markets. *Transportation Research Part B: Methodological*, (forthcoming).

[10] NYCTLC (2019). TLC Trip Record Data. https://www1.nyc.gov/site/tlc/about/tlc-trip-record-data.page.

[11] Prékopa, A. (1958). On secondary processes generated by a random point distribution of Poisson type. *Annales Univ. Sci. Budapest de Eötvös Nom. Sectio Math*, 1:153–170.

[12] Wang, H. and Yang, H. (2019). Ridesourcing systems: A framework and review. *Transportation Research Part B: Methodological*, 129:122–155.

[13] Wang, X., He, F., Yang, H., and Gao, H. (2016). Pricing strategies for a taxi-hailing platform. *Transportation Research Part E: Logistics and Transportation Review*, 93:212–231.

[14] Xu, Z., Yin, Y., and Ye, J. (2019). On the supply curve of ride-hailing systems. *Transportation Research Part B: Methodological*, (forthcoming).

[15] Yahia, C. N., de Veciana, G., Boyles, S. D., Rahal, J. A., and Stecklein, M. (2020). Book-Ahead & Supply Management for Ridesourcing Platforms. *arXiv preprint*.

[16] Zha, L., Yin, Y., and Du, Y. (2018a). Surge pricing and labor supply in the ride-sourcing market. *Transportation Research Part B: Methodological*, 117:708–722.

[17] Zha, L., Yin, Y., and Xu, Z. (2018b). Geometric matching and spatial pricing in ride-sourcing markets. *Transportation Research Part C: Emerging Technologies*, 92:58–75.

[18] Zuniga-Garcia, N., Tec, M., Scott, J. G., Ruiz-Juri, N., and Machemehl, R. B. (2020). Evaluation of ride-sourcing search frictions and driver productivity: A spatial denoising approach. *Transportation Research Part C: Emerging Technologies*, 110:346–367.