Time-of-Day Pricing in Taxi Markets

Xinwu Qian and Satish V. Ukkusuri

***Abstract*— For a regular weekday in New York City, the number of taxi trips at 8 P.M. may be 10 times greater than that at 5 A.M., while passengers are charged under the same pricing scheme. Motivated by temporally non-stationary demand and supply in the taxi market, the time-of-day (TOD) pricing scheme for taxi industry is framed to vary trip cost dynamically over time, so that total market revenue is maximized. Temporal market dynamics is modeled as a semi-Markov process, which captures leftover of drivers, spillover of passengers, and restoration of drivers in service along the time horizon. The TOD pricing scheme is therefore formulated as discrete time stochastic dynamic programming with the goal to find the optimal sequence of price multipliers. The approximate dynamic programming (ADP) approach is introduced to solve the curse of dimensionality. Numerical experiments are conducted using New York City taxi trip data to illustrate the effectiveness of TOD price in real-world taxi market. The results suggest that TOD price may increase daily market revenue by over 10% using the ADP approach. Our experiments also show that TOD price may be even more effective if sudden surges in demand take place in the market.**

***Index Terms*— Time-of-day pricing, market dynamics, revenue maximization, value function approximation, daily operation, surge demand.**

I. INTRODUCTION

TAXICAB is an indispensable transportation mode in urban areas. Within the past few decades, large cities have witnessed the rapid expansion of the taxi market, both in terms of the level of demand as well as the number of taxicabs. By the end of 2011, there were 47,236 taxicabs in Tokyo, which corresponded to one taxicab for every hundred population. Similar population to taxi ratio was also reported for Paris, London and Seoul [1]. As for New York City (NYC), in 2014, there were 13, 437 yellow cabs serving over 175 million trips over the year [2]. There is no doubt that large cities are benefited from the extraordinary capacity emerged from the growing market and people are merited from the 7-24 door-to-door service provided by cab drivers. Nevertheless, the take-off of the taxi industry also results in negative externalities such as excessive congestion and emission [3]. Consequently, in the literature, market regulations were investigated [3]–[5] and mathematical models were proposed to understand macroscopic relationship between drivers and passengers, [6]–[9], for improving the sustainable operation of the taxi market.

Manuscript received August 10, 2016; accepted September 28, 2016. Date of publication January 16, 2017; date of current version May 29, 2017. The Associate Editor for this paper was N. Geroliminis.

The authors are with the Department of Civil Engineering, Purdue University, West Lafayette, IN 47906 USA (e-mail: qian39@purdue.edu; sukkusur@purdue.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TITS.2016.2614621

Recent advances in the development of data sensing and ubiquitous computing technologies offer whole new perspectives to reexamine the downsides of the taxi market. Especially, since taxicabs in large cities are equipped with on-board GPS devices, the streaming data obtained from real-time taxi trips improve our understanding of market operations in urban areas. The aggregated trip data can be used to infer underlying land use [10], [11], estimate taxi ridership [12], understand taxi drivers' searching behavior [13], [14], and explore spatiotemporal dynamics of urban activities [15]. From disaggregated level, trip data were used to show that present taxi market is considerably inefficient due to lack of perfect information [16]. Consequently, besides market regulations, taxi recommendation system [17], [18], dispatching algorithms [19] and ride-sharing strategy [20], [21] were developed to match individual passengers and drivers using real-world trip information.

It is suggested from aforementioned studies that taxi market is non-stationary both spatially and temporally. The non-stationary nature of taxi market ends up with spatio-temporal mismatch between drivers and passengers. While entry regulations help to balance overall demand and supply, daily operations still suffer oversupply at off-peak hours and excessive demand during peak hours. On the other hand, recommendation strategies mitigate the impact of spatial mismatch, however, the temporal non-stationary demand may significantly weaken the effectiveness of the approach. Consequently, temporal unbalance of the number of taxi drivers and passengers remains an unsolved issue in the market which degrades the efficiency of the taxi market. In light of the issue, Uber introduced the idea of surge price to their products [22]. While echnical details of the pricing policy are confidential, the main idea behind Uber's surge price is that riders will be charged a premium if there are excessive demand, and the premium is determined based on the gap between the levels of supply and demand. However, many users were frustrated by the expensive premiums and drivers were found to be tactical under the scheme, where many of them show up only if the premiums are high enough [23]. While the approach aims to address the unbalance between supply and demand, apparently, such scheme is not well-suited for taxi industry. Devising an effective and applicable pricing scheme for time-varying demand in taxi industry is the first motivation of the study.

The second motivation is in light of the situation where taxi drivers are overloaded in an increasingly competitive market, and it has long been suggested that taxi drivers should be subsidized for their vacant time [24]. According to New York Taxi and Limousine Commission (NYCTLC), the medallion price in 2014 was almost four times higher than that back in 2002, which becomes a heavy burden for taxi drivers. On the other hand, new companies such as Uber and Lyft started a revolution in the taxi industry. In less than three years, traditional taxi industry has been deprived of over two-thirds of the revenue due to the entry of the new start-ups [25]. The consequence is that taxi drivers may have to make more trips than before to earn a living, which leads to a dead end of an ecosystem with heavier congestion and emission. Taken in composition with the temporal non-stationary nature of the taxi market, it will be extremely beneficial if trip cost can be altered dynamically so that demand and supply are balanced and extra revenue can be created as subsidies for taxi drivers.

In this study, we introduce the time-of-day (TOD) taxi pricing framework which aims to maximize total system revenue by taking advantage of the temporal non-stationary nature of taxi market. The price multiplier is used to alter the price dynamically, where value less than 1 represents discounts and value greater than 1 stands for premiums. This suggests an important difference between the TOD pricing and the existing surge price polices on the market, where only premium is charged and discount is not applicable. The policy is potentially more acceptable for passengers, since the premium charged during peak hours may be compensated by discount during off-peaks. The NYC taxi trip data are used to understand the effectiveness of such scheme for real world taxi market. The main philosophy of the approach is that more passengers may be attracted by lowering price when supply is abundant, and surcharges are placed for excessive demand. We use the semi-Markov process to model the temporal dynamics of the taxi market and formulate the dynamic programming (DP) to find the optimal sequences of price multipliers to maximize the total system revenue. Solving the DP is non-trivial, as the dimensionality of the problem scales up exponentially over the time horizon. The approximate dynamic programming (ADP) approach is therefore introduced which significantly reduces the complexity of the problem through value function approximation, policy iteration and Monte Carlo simulation. The feasibility of the TOD pricing framework is evaluated through two experiments, which account for market daily operation and surges in demand resulted from special events. The results suggest that ADP approach may contribute to over 10% increase in daily revenue by serving similar level of demand. The increase in revenue during demand surge may even reach 78.6%.

The rest of the paper is organized as follows: the next section presents an overview of the data used in the study; section III gives the mathematical notation and describes the dynamics of taxi market; the mathematical formulation and solution approach are discussed in section IV; experiment settings and result discussion are presented in section V; finally, in section VI, we summarize key findings and future works related to the study.

II. Data

The data used in this study were released by NYCTLC and are publicly available [26]. The data were collected from January 1st to December 31st in 2013 and contained trip information related to yellow cabs in NYC. Trip records vary from 300,000 to 600,000 pieces daily. Each piece of record consists of the geo-coordinates and time stamps for trip origin and destination, a unique medallion ID, number of passengers onboard, trip distance (miles), and trip cost (USD, including base cost, tax, and gratuity). As a result, it is possible to track the number of trips and capacity of taxi drivers for any given time periods.

The one-year data are processed and aggregated for further analysis. Records are first screened based on trip distance, trip duration, trip cost, and resultant average trip speed. Holidays such as thanksgiving and Christmas week are removed to restore the supply and demand level of common daily operations. Weekday and weekend data are separated considering their repeatable but distinct trip patterns [15]. For each individual day, passenger data are aggregated at 5-minutes interval and the availability of drivers, trip cost and trip duration are summarized on hourly basis. This approach ensures the flexibility to model the taxi market at a fine level while at the same time reduces errors and noises in the aggregated data.

III. Modeling the Taxi Market: Preliminaries

In this study, the taxi market is modeled as a discrete time finite horizon system at the aggregated level. The finite horizon sets up the boundary for the operation time of the taxi market and the discrete time setting suggests that the time horizon is divided into consecutive time intervals. Moreover, market stochasticity is also incorporated in light of external uncertainties such as weather and road traffic condition. In this section, the market dynamics is first described based on the finite horizon and discrete time settings of the problem. Understanding market dynamics is essential to formulate the problem as mathematical programming. TABLE I presents the notations that are used throughout the study.

A. Passenger

Passengers are the demand side of the taxi market and are price sensitive. We consider the value of price elasticity being negative, which suggests that passenger demand will increase monotonically as the trip cost decreases, and vice versa. In addition, we make a further rational assumption that passenger behavior is myopic with respect to the trip cost. That is to say, current passengers will continue riding taxis as long as the trip cost is below the present value. They may bring forward their trip when they perceived lower cost. However, they will either leave the market or postpone their travel time when the trip cost is higher than expected.

The arrival of passengers is assumed to follow the Poisson process and individual arrival is therefore independent from the others. Note that when Poisson mean is sufficiently large, according to central limit theorem, value of independent Poisson variables can be well approximated using normal distribution. As a result, the normal distribution is applied to model the number of passengers that arrive in time interval t. If the length of the time horizon is T, we have T independent normal distributions to model the temporally non-stationary arrival of passengers in the market. The parameter values of normal distributions are obtained by using one year of trip data. For each time interval, in general, there are 262 samples for weekdays and 103 samples for weekend.

TABLE I
NOTATION USED IN THE PAPER

Variables	Descriptions
t	Time interval, where $t \in \{1, 2, ..., T-1, T\}$
$\mathcal{X}_t$	The feasible set at time t
x_t	Decision variable (price multiplier) at time t, $x_t \in \mathcal{X}_t$
X_{min}, X_{max}	The upper and lower bounds for the decision variable
D_t	The number of passengers served at time t
$\hat{D}_t$	The number of passengers arrived at the beginning of time t
$g^{-1}(\hat{D}_t, x_t)$	The inverse demand function which gives the induced demand based on the price multiplier
γ	The spillover rate for unsatisfied passengers at time t
$\tilde{D}_t$	Induced demand at time t
S_t	The amount of drivers that are available for time t
$\hat{S}_t$	The number of drivers that are observed to leave or join the market at time t
ΔS_t	The amount of drivers who are in service before time t but become available at the beginning of time t
λ_t	The observed mean service time per taxi trip at time t
$F_t(k)$	The cumulative distribution function for trip service time at time t
$\bar{\lambda}$	The upper bound for trip service time.
a_t	The observed mean trip cost for time t
ω_i	Sample path i, $i \in 0, 1, ..., N-1, N$
$\mathcal{P}$	The set of candidate policies
π	Policy selected, $\pi \in \mathcal{P}$
$\mathcal{S}_t$	Market state at the beginning of time interval t
$\mathcal{W}_t$	The collection of all random variates at time t

B. Driver

Taxi drivers are the supply side of the market. In this study, the total number of drivers is considered as bounded from above, which corresponds to entry regulation of taxi industry. The behavior of taxi drivers is assumed to be revenue maximizing. They will make the trip no matter the revenue is \$1 or \$10, otherwise they will have to run the risk of being empty.

Being different from passenger demand, it is inappropriate to assume that the availability of drivers at different time intervals are independent. Instead, provided that the number of drivers is upper-bounded and the length of time interval is short (e.g., 5 minutes), a large number of passengers at current time interval may imply limited driver availability in the following time step, due to most drivers being occupied. Moreover, this also implies that there are likely to have more available drivers 3 to 4 time steps ahead, upon the completion of passenger trips. Therefore, the availability dynamics of drivers at time t can be modeled as a function of the demand level and the amount of drivers in service in previous time intervals. Nevertheless, for each time interval, there are also drivers that leave or join the market due to different time schedules. This portion of drivers are assumed to be independent across time and the value is observed from the data to follow the normal distribution.

C. Trip Service Time

Knowing the amount of drivers in service in a particular time interval, the trip service time distribution can be used to identify the proportion of drivers that will become available in subsequent time steps. In this study, we define trip service time as the time span between the start of successive trips. The trip service time can be further decomposed into the searching time and the trip time. For each time interval, the trip time distribution mainly depends on road traffic condition, which is barely affected by the change of the number of passengers. On the contrary, we consider the searching time as the time needed for taxi drivers to successfully find a passenger, which is sensitive to the demand level. As suggested by Yang *et al.* [27], the taxi-customer searching behavior can be characterized by using the meeting function and the expected driver searching time at steady-state market equilibrium follows

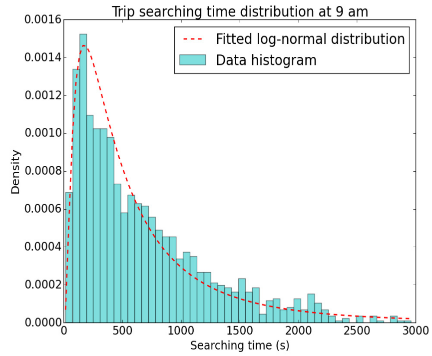
$$w_k^t = \frac{1}{A_k w_k^c Q_k^c} \tag{1}$$

where w_k^t is the expected driver waiting time of location k, A_k is the location related coefficient, w_k^c is the customer waiting time at location c, and Q_k^c is the passenger trip rate at location k. Equation 1 tells that searching time is inversely proportional to the average number of unserved passenger trips. While the dynamic pricing scheme is deemed to alter the demand profile, it is important to model the impact on trip searching time accordingly. In light of this issue, trip sequence data for individual drivers are processed to obtain trip time and searching time distributions separately, and their correlation is also calculated. It is observed from the data that trip time and searching time can be well approximated by the log-normal distribution, and an example for the distributions at 9 AM is given in FIGURE 1(a)-(b).

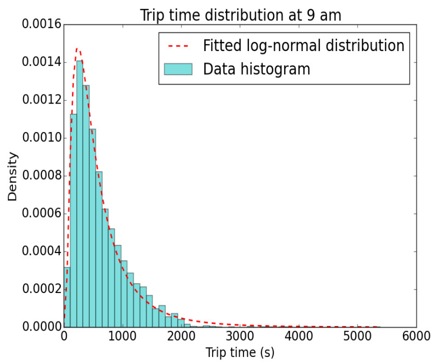
Given the trip time and baseline searching time distributions, the next step is to adjust the searching time distribution based on demand level and combine them into one single trip service time distribution. The modified searching time distribution can be obtained by scaling the mean value with the change in demand level as:

$$u_t' = u_t \frac{\bar{D}_t}{D_t} \tag{2}$$

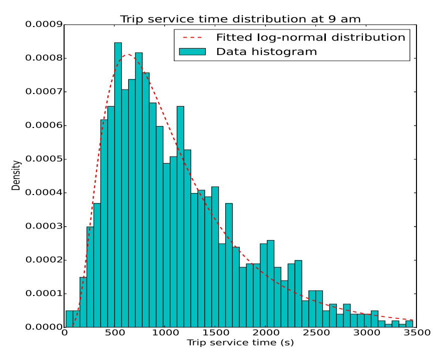
where $\bar{D}_t$ is the observed mean demand level at time t from data, and u_t' and u_t are modified mean value and baseline



(a)



(b)



(c)

Fig. 1. Histograms and fitted log-normal distributions for trip time and searching time at 9 AM. (a) Distribution of searching time. (b) Distribution of trip time. (c) Distribution of trip service time.

mean value for searching time log-normal distribution respectively. The intuition behind Equation (2) is the principle of proportionality, which is derived from the relationship between searching time and passenger rate in Equation (1).

The trip service time distribution can be viewed as the summation of two correlated log-normal random variables, and can be well approximated using the Schwartz-Yeh approximation [28]. The Schwartz-Yeh approximation assumes that the resulting distribution is also log-normal, which matches with our empirical observations as shown in FIGURE 1(c), and is reported to provide accurate estimation for the body part of the cumulative density function (CDF) [29]. Finally, the restoration of taxi drivers in service can therefore be calculated by using the cumulative density function (CDF) of the log-normal distribution at different time intervals.

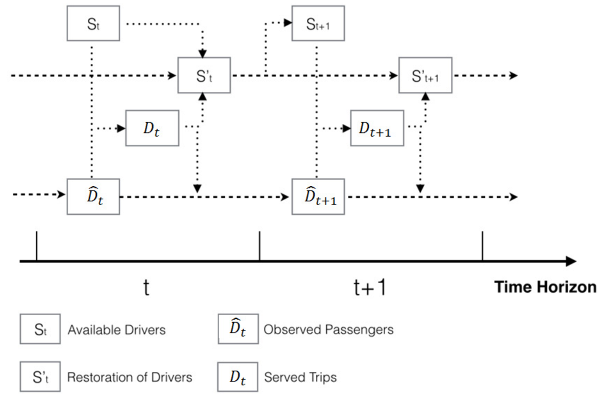


Fig. 2. Illustration of the semi-Markov process in the taxi market.

D. Market Dynamics

The dynamics of the market is driven by the arrival of passengers, the availability of drivers and also the change of trip cost. An illustration of the market dynamics is given in FIGURE 2. As indicated by the dot-dashed lines, the number of served trips at time t is determined by the arrival of passengers $\hat{D}_t$ and the number of available drivers S_t. Meanwhile, there is a semi-Markov process with three transition functions which governs the changes in market states over time. First, the availability of drivers at time t relies on the restoration of drivers that are in service at time $1, 2, \ldots, t-1$. Similarly, the served number of trips T_t will contribute to $S'_t, S'_{t+1}, \ldots, S'_T$, which capture the number of restored drivers at the end of the time interval. Second, the amount of leftover drivers from time $t-1$, which is determined by T_{t-1} and S_{t-1}, also contributes to the availability of drivers at time t. Finally, the amount of passengers at time t is also affected by the spillover of unserved passengers from previous time intervals.

E. Unconstraining Taxi Demand

It can be justified from the market dynamics that the TOD price is largely affected by the taxi demand. Nevertheless, the trip data carries incomplete demand information due to limited supply. In particular, trip observations are right-censored during peak hours and it is of great importance to come up with a method to reconstruct the true distribution of taxi demand. Similar problems are well studied in multiple fields, including incomplete medical follow-ups [30] and incomplete booking data in hotel [31] and airline industry [32].

In this study, the expectation maximization (EM) algorithm [33] is implemented to unconstrain the taxi demand for both weekday and weekend data. The EM algorithm is reported to be one of the most effective demand-unconstraining methods [34], and readers may refer to [35] for detailed implementation directions. One main assumption of the EM algorithm is that the full data are normally distributed. By truncating trips during midnight, as shown in FIGURE 3, the distribution for the rest of the trips resembles the bell shape. Consequently, the normality assumption is satisfied and

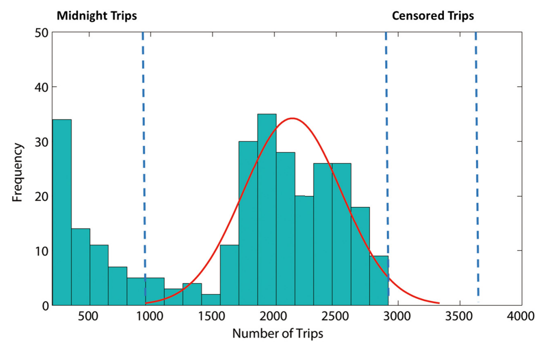


Fig. 3. Histogram of taxi demand aggregated at 5 minutes interval.

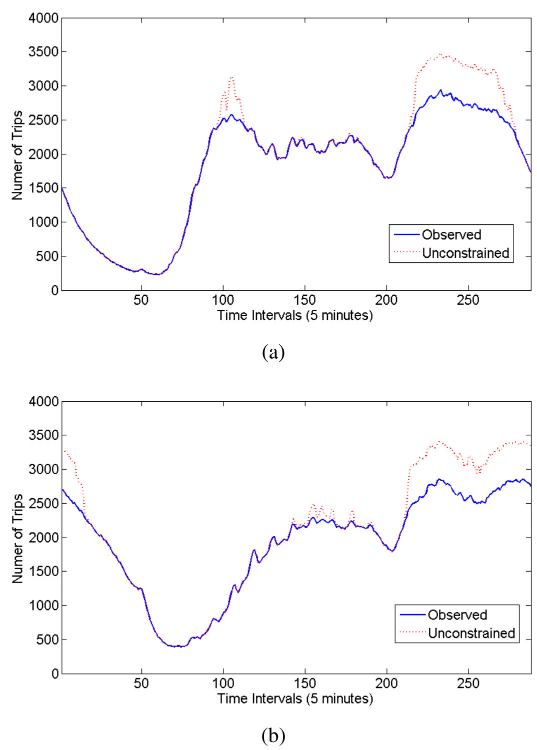


Fig. 4. Results of unconstrained weekday and weekend taxi demand. (a) Weekday. (b) Weekend.

the right tail of the true distribution serves as the unconstrained demand. Demand over 2,400 for weekday and 2,100 for weekend within 5 minutes time interval are considered as censored. The resulting unconstrained demand is presented in FIGURE 4.

IV. Formulation

A. Dynamic Programming Formulation

One main objective of the study is to propose a mathematical framework for the taxi industry to maximize the total system revenue. Due to the dynamic nature of the taxi market and the existence of an underlying semi-Markov process, it is appropriate to model the problem as a stochastic DP which takes the form:

$$\underset{x}{\text{maximize}}\ \mathbb{E}[\sum_{t=1}^{T} a_t x_t D_t] \tag{3}$$

subject to

$$X_{min} \le x_t \le X_{max} \tag{4}$$
$$\tilde{D}_t = g^{-1}[\hat{D}_t + \gamma \max(0, D_{t-1} - S_{t-1}), x_t] \tag{5}$$
$$D_t = \min(S_t, \tilde{D}) \tag{6}$$
$$S_t = \hat{S}_t + \Delta S_t \tag{7}$$
$$\Delta S_t = \sum_{k=1}^{t-1} [D_k[F_k(t-k) - F_k(t-k-1)]] \tag{8}$$

The objective function is set to maximize the expected total system revenue. Equation (4) restraints the price multiplier in the predefined range. Equation (5) suggests that the number of passengers served at each time interval is determined by the arrival and spillover of passengers and the price multiplier at the time interval. Moreover, the inverse demand function $g^{-1}(x_t)$ decreases monotonically as x_t increases, which characterizes price elasticity of passengers. Equation (6) indicates that the number of passengers served should not exceed the number of available drivers. Equations (7) and (8) define the transition process of the system. In particular, the availability of drivers is determined by the join or leave of taxi drivers for the time interval and the restoration of drivers that were in service from previous time intervals.

It is easy to verify from the DP formulations that the price multiplier at each time interval is a function of the availability of drivers and the number of passengers. To maximize the system revenue, one has to take into account the resulting number of drivers and passengers for subsequent time intervals when select decision x_t at time t. Assuming that the cardinality of $\mathcal{X}_t$ is k, even k being small, possible combinations of decisions over the time horizon will scale up exponentially to an extremely large value k^T which is infeasible to evaluate. Moreover, this is exclusive of the space of possible system states due to the stochastic nature of the system. Such phenomenon is the well-known *"curse of dimensionality"* for DP problems [36]. In order to address the dimensionality issue, the approximate dynamic programming (ADP) framework is introduced which reduces the problem complexity significantly by approximating future function values.

B. ADP Formulation

1) State Space: The state space $\mathcal{S}_t$ refers to the collection of variables that characterizes the state of the system for time interval t. In this study, the state of the system is determined by the availability of drivers and the number of passengers to serve. In addition, the state can be further divided into two parts. One is the time-dependent exogenous information $\mathcal{W}_t$, which is the set of random variates that are revealed at the beginning of time interval t and independent of the system state of other time intervals. To elaborate, the exogenous information includes the arrival of passengers, the number of drivers that leave or join the market, average trip cost, and the mean trip service time. The other part of the system state is associated with the transition process. This compromises the availability of drivers, which depends on the amount of drivers in service in previous time intervals, and the total number of passengers, which is affected by the amount of spillover passengers. As a result, the state vector can be written as

$$\mathcal{S}_t = (D_t, S_t, \mathcal{W}_t) \tag{9}$$

By taking action x_t, the resulting system state at the following time interval can be written as

$$\mathcal{S}_{t+1} = \mathcal{S}^M(\mathcal{S}_t, x_t, \mathcal{W}_{t+1}) \tag{10}$$

Where $\mathcal{S}^M$ is the function that governs the transition of the system state.

2) Value Function Approximation: By using the state space notation for the taxi market, the DP defined by Equation (1)-(6) can thus be rewritten as

$$\underset{x}{\text{maximize}}\ \mathbb{E}[\sum_{t=1}^{T} C(\mathcal{S}_t, x_t)] \tag{11}$$

Where $C(\mathcal{S}_t, x_t)$ is the contribution function which evaluates the system revenue when the state is $\mathcal{S}_t$ and the decision taken is x_t. Equation (9) simply indicates that we are seeking the best pricing policy so that total system rewards is maximized. For a certain time interval t, the DP can be further rephrased using the Bellman's equation

$$V_t(\mathcal{S}_t) = \underset{x_t}{\text{maximize}}\ C(\mathcal{S}_t, x_t) + \mathbb{E}\{V_{t+1}[\mathcal{S}^M(\mathcal{S}_t, x_t, \mathcal{W}_{t+1})]\} \tag{12}$$

Where $V_t(\mathcal{S}_t)$ is the reward value function at time t, which is the summation of the contribution value at time t and the reward value of subsequent time intervals. Computing the expectation of future rewards may be immensely difficult as it involves intensive integrations over future time intervals. Instead, the Monte Carlo simulation approach is implemented which generates repeated samples and compute the expectation by averaging over simulation outcomes of the Markov process [37].

Nevertheless, subject to dimensionality issue, the total system state may scale exponentially over time and expectation calculation may require infinite sample paths so that all state combinations are visited. This may result in an inefficient usage of the simulation approach. Alternatively, one may construct appropriate functions to approximate future contributions with much lower complexity. This approach falls within the scope of approximate dynamic programming (ADP) [38]. In particular, we use the linear parametric model to approximate system rewards of being in state $\mathcal{S}_t$ as

$$\mathbb{E}\{V_{t+1}[\mathcal{S}^M(\mathcal{S}_t, x_t, \mathcal{W}_{t+1})]\} = \bar{V}(\mathcal{S}_t|\mathbf{w_t}, \mathbf{x_t}) = \mathbf{w_t^T}\phi(\mathcal{S}_\mathbf{t}, \mathbf{x_t}) \tag{13}$$

Where $\phi(\mathcal{S}_t, x_t)$ is the vector of basis functions constructed for state $\mathcal{S}_t$ and decision x_t, and $\mathbf{w_t}$ is the vector of coefficients for each basis function. As a consequence, at time t, we approximate system rewards of taking decision x_t given that the system state is $\mathcal{S}_t$. Moreover, basis functions should be capable of accounting for trade-offs between immediate rewards and future impacts. For instance, lower price rate at time t may generate more revenue at the moment but limit income in the next few time intervals due to limited availability of drivers. Obviously, the main challenge is to devise effective basis functions, which requires fundamental understanding of the system and trials and errors are also necessary to determine the proper function form (e.g., linear versus quadratic). We discuss the set of well-suited basis functions for taxi market in the next section. By fixing basis functions, the optimal decision can be obtained via

$$x_t^i = \underset{x_t \in \mathcal{X}_t^i}{\text{argmax}}\ C(\mathcal{S}_t^i, x_t) + \mathbf{w_t^T}\phi(\mathcal{S}_\mathbf{t}^\mathbf{i}, \mathbf{x_t}) \tag{14}$$

3) Basis Function: Nine basis functions are introduced for each state. As a result, the huge state space is approximated by the 9-dimensional parametric function. In general, the 9 basis functions can be divided into four groups as presented below.

a) Constant term: the first group involves the constant term, where $\phi_1 = 1$. It helps to calibrate unobserved heterogeneity for each state and offsets the parametric model for better fitness.

b) Current market states: the second group mainly describes the system state at current stage, which constitutes of the observed amount of passenger arrivals and the availability of taxi drivers:

$$\phi_2 = \hat{D}_t, \phi_3 = S_t$$

This provides a direct proxy for the immediate impact of making a decision.

c) Short-term impact: the third group of functions focuses on evaluating the short-term impact of the current decision on the taxi market, including

$$\phi_4 = \frac{a_{t+1}}{a_t}\tilde{D}_t, \phi_5 = S_t - \tilde{D}_t$$

where ϕ_4 captures the differences of the average trip revenue, which has distinct temporal characteristics and can be easily referred from historical trip information. ϕ_5 denotes the residual number of drivers after completing the demand at current time. Both functions contribute to adjusting the number of served passengers by inferring potential gains in the next time interval. Note that drivers in service at current time interval are likely to be unavailable in the next time interval. Consequently, the short-term impact is of great importance to optimally assign the limited drivers available over time.

d) Long-term impact: the last group of basis function is developed to measure the long-term impact of current decision on the taxi market, which includes

$$\begin{aligned}
\phi_6 &= \tilde{D}_t[F_t(t+1) - F_t(t)] \\
\phi_7 &= \tilde{D}_t[F_t(t+2) - F_t(t+1)] \\
\phi_8 &= \tilde{D}_t[F_t(t+3) - F_t(t+2)] \\
\phi_9 &= \tilde{D}_t[F_t(t+4) - F_t(t+3)]
\end{aligned}$$

Due to complex dynamic nature of the market, it is typically hard to track the long-term impact on system state. In our study, a series of basis functions from ϕ_6 to ϕ_9 are designed to probe future rewards by measuring subsequent restoration of taxi drivers for up to 4 time intervals. This establishes the connection between current served trips and future availability of taxi drivers. If trips in further time steps are profitable,

TABLE II
APPROXIMATE POLICY ITERATION ALGORITHM FOR THE TOD PRICING PROBLEM

1 Initialize $\mathbf{w}_t\ \forall t$
2 Fix the set of basis functions $\phi_f(\mathcal{S})$
3 **for** *i=1:N* **do**
4 Initialize system state $\mathcal{S}_0^i$
5 **for** *j=1:M* **do**
6 Select a sample path ω_j
7 **for** *t=1:T* **do**
8 Compute: $x_t^{i,j} = \underset{x\in\mathcal{X}_t^{i,j}}{\text{argmax}}\ C(\mathcal{S}_t^{i,j},x) + \sum_{k=1}^{f} w_t^k \phi_k(\mathcal{S}_t^{i,j})$
9 State update: $\mathcal{S}_{t+1}^{i,j} = \mathcal{S}^M(\mathcal{S}_t^{i,j}, x_t^{i,j}, W_{t+1}(\omega_j))$
10 **end**
11 Set $V_t^{i,j} = 0\ \forall t$
12 **for** *t=T:1* **do**
13 Compute: $V_t^{i,j} = C(\mathcal{S}_t^{i,j}, x_t^{i,j}) + V_{t+1}^{i,j}$
14 **end**
15 Update regression coefficients $w_t\ \forall t$
16 **end**
17 **end**
18 **Output** w_t

these functions will motivate ADP to serve more passengers at current time so that more drivers will be able to meet the demand at both time intervals and the revenue is therefore maximized.

C. Solution Algorithm

We relaxed the DP by using value function approximation and the solution for the ADP is therefore given by Equation (14). The remaining question is how to estimate regression coefficients so that the system revenue is maximized. As discussed in the previous section, the Monte Carlo methods can be used to evaluate the expectation of future rewards through simulation observations. Moreover, if a series of decisions x_t^i is made while simulating through the Markov process based on Equation (14), the exact rewards can be observed at time t by tracing backwards and stacking contribution values:

$$V_t^i = \sum_{k=t+1}^{T} C(\mathcal{S}_t^i, x_t^i) \tag{15}$$

Repeating the process for N sample paths will give a collection of simulation observations and the regression coefficients can therefore be calibrated by minimizing the mean-square error over all observations

$$\mathbf{w_t} = \underset{\mathbf{w}}{\text{argmin}}\ \frac{\mathbf{1}}{\mathbf{N}} \sum_{\mathbf{i=1}}^{\mathbf{N}} (\mathbf{w^T}\phi(\mathcal{S}_\mathbf{t}^\mathbf{i}) - \mathbf{V_t^i})^2 \tag{16}$$

The procedure is known as *policy evaluation*. After the regression coefficients are calculated, the series of decisions can be improved by solving Equation (14) with updated parameters. This in return results in a new series of decisions and the procedure is known as *policy improvement*. Both policy evaluation and policy improvement constitute the core of the approximate policy iteration algorithm which is illustrated in detail in TABLE II

Line 15 of the algorithm requires to recalibrate the regression coefficient and batch update given in Equation (16) is a common solution. However, the loop structure of the algorithm makes it possible to utilize the streaming data within simulation iterations. It is more efficient to update the coefficients recursively after each sample path is simulated, so that more representative observation can be obtained through more frequent policy improvement and faster convergence rate is achieved. The recursive least square update of regression coefficient [36] takes the form:

$$\mathbf{w_t^i} = \mathbf{w_t^{i-1}} + \mathbf{H_t^i}\phi(\mathcal{S}_\mathbf{t}^\mathbf{i})\epsilon^\mathbf{i} \tag{17}$$

The matrix H^n is computed as

$$H^i = \frac{1}{\delta^i}\mathbf{B^{i-1}} \tag{18}$$

The error ϵ^i is measured as the difference between observed reward value and the predicted value

$$\epsilon^i = V_t^i - \mathbf{w_t^{i-1}}\phi(\mathcal{S}_\mathbf{t}^\mathbf{i}) \tag{19}$$

The matrix $\mathbf{B^i}$ is calculated following the Sherman-Morrison formula for updating the inverse of a matrix:

$$\mathbf{B^i} = \mathbf{B^{i-1}} - \frac{\mathbf{1}}{\delta^\mathbf{i}}(\mathbf{B^{i-1}}\phi(\mathcal{S}_\mathbf{t}^\mathbf{i})\phi(\mathcal{S}_\mathbf{t}^\mathbf{i})^\mathbf{T}\mathbf{B^{i-1}}) \tag{20}$$

and

$$\delta^i = 1 + \phi(\mathcal{S}_t^i)^T\mathbf{B^{i-1}}\phi(\mathcal{S}_\mathbf{t}^\mathbf{i}) \tag{21}$$

The initial choice of $\mathbf{B_t^0}$ could be cI, where I is the $|f|\times|f|$ identify matrix and c is a small constant. The initial regression coefficient $\mathbf{w_t^0}$ can simply be $\vec{0}$. Both $\mathbf{w_t}$ and $\mathbf{B_t}$ will converge to stable values after moderate realizations of sample paths.

D. Benchmark Algorithm

To validate the performance of the ADP approach, and to comprehensively investigate the effect of TOD pricing scheme for taxi market, three benchmark algorithms are also tested in this study.

The first algorithm is the static approach, which corresponds to the baseline taxi market where no TOD pricing is implemented. Consequently, we have

$$x_t = 1, \quad \forall t = 1, 2, \ldots, T \tag{22}$$

The second algorithm is the greedy revenue (GR) approach, where the objective is to maximize the revenue in each time interval myopically. For each time interval t, the pricing policy is calculated as:

$$\begin{aligned} x_t = \underset{x}{\text{argmax}}&\ a_t x D_t \\ \text{subject to}& \\ &X_{min} \le x_t \le X_{max} \\ &D_t \le S_t \end{aligned} \tag{23}$$

Finally, the third algorithm applies the greedy trip (GT) approach, where the objective is to serve the maximum number of passengers at each time step. For each time interval t, the pricing policy is derived as:

$$\begin{aligned} x_t = \underset{x}{\text{argmax}}&\ g^{-1}(\hat{D}_t, x) \\ \text{subject to}& \\ &X_{min} \le x_t \le X_{max} \\ &g^{-1}(\hat{D}_t, x) \le S_t \end{aligned} \tag{24}$$

V. Experiment Result

A. Experiment Setting

Two distinct scenarios are tested to evaluate the effectiveness of ADP in improving the market revenue. The first scenario is for the daily operation of taxi market, where data are obtained from preprocessed taxi trip records which reconstruct the market environment in real world. The second scenario is artificially designed to model sudden demand surges in the market, which may be resulted from the end of large events or the break-down of other transportation systems.

The time interval is set to 5 minutes in order to model the detailed dynamics of the taxi market. 2,000 simulations with identical initial random seed are performed for the static, greedy revenue, greedy trip, and ADP approaches. Identical random seed ensures that all algorithms are performed under the exactly same system environment. All simulation results are carried out using a 3.4 GHz six-core processor and 12 GB RAM. The average computation time for individual simulation is 1.32 second and the 2,000 simulation runs are completed within 45 minutes. Finally, results are obtained by averaging over the outcomes of simulation observations.

The price elasticity of taxi demand is considered as the percentage change in demand in response to the percentage change in price. Let η_t denote the price elasticity at time t, the induced demand with respect to the price multiplier is given by

$$\tilde{D}_t = \hat{D}_t[1 + \eta_t(x_t - 1)] \tag{25}$$

where η_t is set to -0.6 [39] for moderate demand during off-peak hours where supply and demand level are more balanced compared with other time periods. Additionally, we alter the coefficient value to mimic the change of price elasticity with respect to market condition, where passengers are assumed to be inelastic in shortage of supply and elastic with excessive supply. Consequently, the elasticity coefficient can be computed as

$$\eta_t = -0.6(\frac{S_t}{\hat{D}_t})^{\theta} \tag{26}$$

where the exponential term θ is used to control the scale of change in price elasticity and is set to 1 in the study.

B. Performance of Basis Function

The Monte Carlo simulation is used to iteratively generate sample paths and observations of contribution function values are collected to update regression coefficients for the set of basis functions. The performance of the basis function is measured using the prediction error, which is calculated as

$$error_t^i = \frac{|\mathbf{w_t^T}\phi(\mathcal{S}_\mathbf{t}^\mathbf{i}) - \mathbf{V_t^i}|}{V_t^i} \tag{27}$$

FIGURE 5 presents the prediction error at the beginning of the time horizon. The value is more difficult to estimate since the observed future rewards are cumulated over the rest of the time and may contain a lot of noise. The prediction error is observed to drop drastically in the first few iterations and fluctuations are observed due to the stochastic nature of the taxi market.

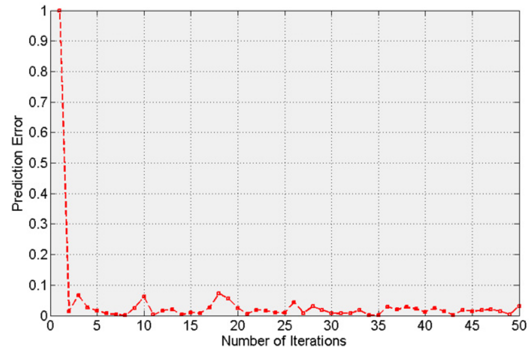


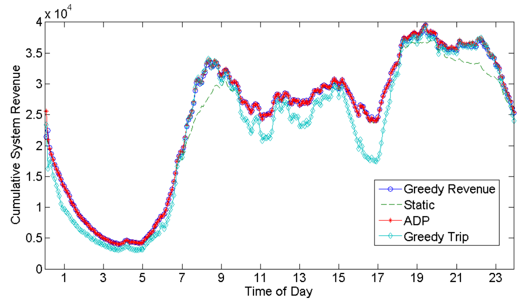
Fig. 5. Basis function performance at $t = 1$.

After 30 iterations, the error pattern becomes much more stable and the error rate is observed to be small, indicating that the set of basis functions fits future rewards quite well. Consequently, the coefficient vector $\mathbf{w_t}$ trained after 30 iterations is used to derive TOD price multipliers under the ADP framework.

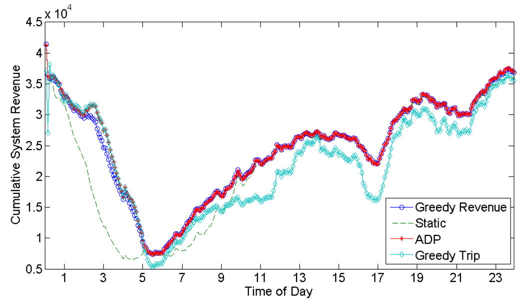
C. Discussion

1) Market Daily Operation: Both weekday and weekend taxi market are examined for the daily operation scenario. With static price, the average daily market revenue takes the value of \$6.71 million on weekday and \$6.43 million during weekend. For GT approach, the revenues are \$6.51 million and \$6.46 million, which correspond to -2.98% and 0.47% improvement. Consequently, simply maximizing fleet utilization of each time interval may fail to improve the total system revenue. On the other hand, by maximizing the revenue per time interval, the GR approach reaches \$6.99 million and \$7.05 million total system revenue, and the improvements are 4.17% and 9.6% respectively. The results of ADP approach are more promising, with weekday revenue of \$7.02 million and weekend revenue of \$7.11 million, which corresponds to 4.62% and 10.57% increase in total system revenue. Judging from the performance on total system revenue, we can see that ADP approach outperforms all other benchmark algorithms for both weekday and weekend scenarios. Specifically, there is a good potential to optimize the market revenue given the same fleet setting, and more significant improvement can be expected for weekend operations. We next examine how these approaches differ with different performance metrics.

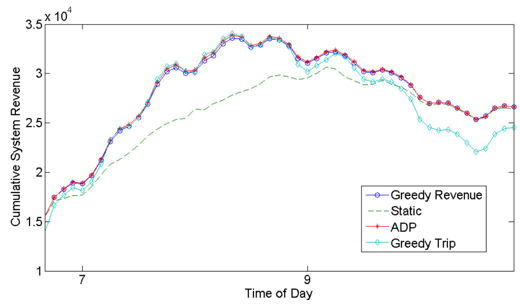
FIGURE 6 presents the total system revenue per time interval. It can be observed that GT approach has similar performance compared with ADP and GR approaches during peak hours, and gets consistently lower revenue during off-peak time. This can be explained from the market nature that we discussed in the preliminary section. Passengers are more price-sensitive during off-peak hours as there is excessive amount of supply, and to maximize the served number of passengers will require higher marginal lost in revenue (e.g., it may require 20% reduction in price to attract 10% more passengers). On the contrary, since there is much more demand than supply during peak hours, passengers are more likely to tolerate high premium for a ride, which leads to comparable performance between GT and GR approaches. This explanation is also applicable to the taxi market around



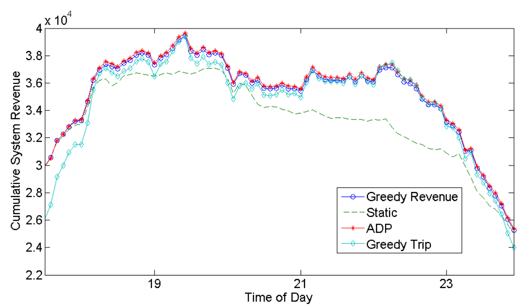
(a)



(b)

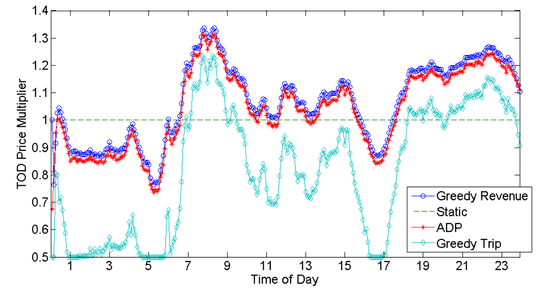


(c)

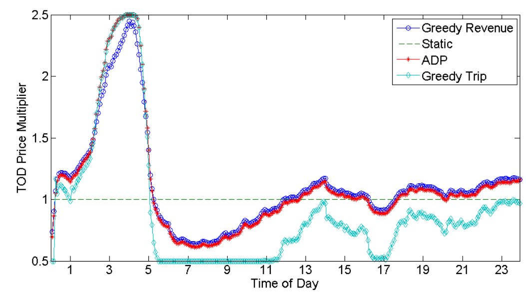


(d)

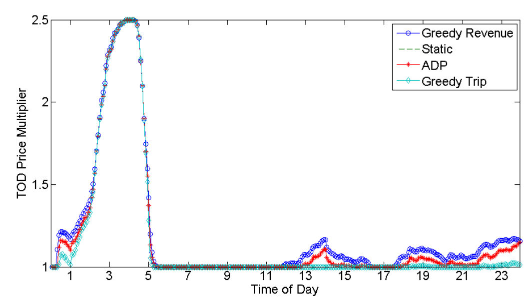
Fig. 6. Total system revenue per time interval. (a) Weekday hourly revenue. (b) Weekend hourly revenue. (c) Snapshot of weekday hourly revenue during morning peak. (d) Snapshot of weekday hourly revenue during evening peak.



(a)



(b)

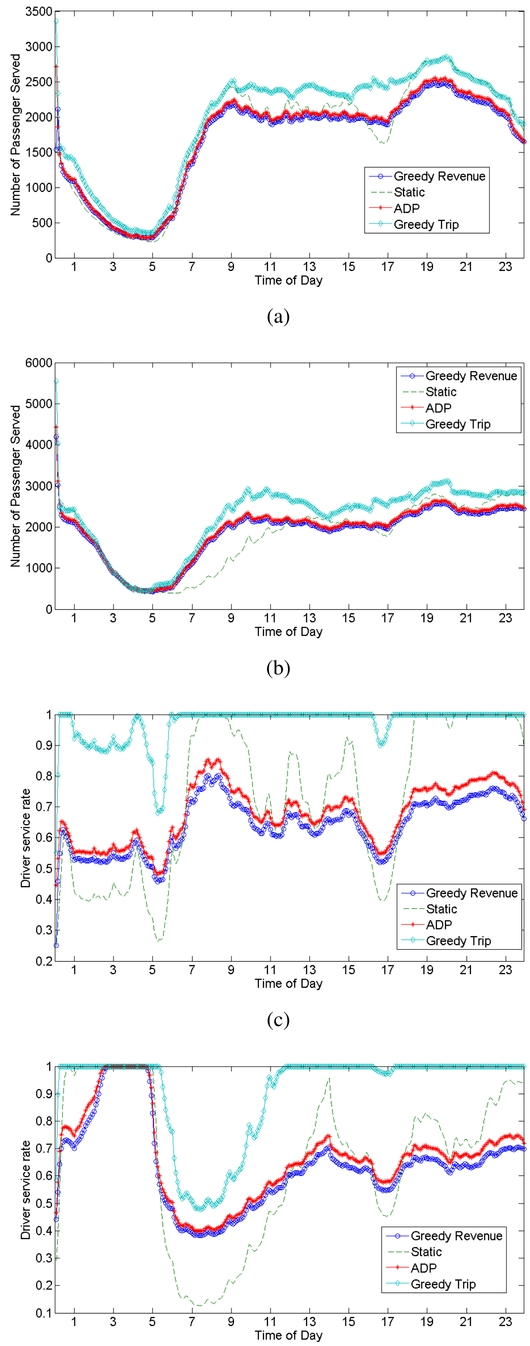


(c)

Fig. 7. Distribution of TOD price multiplier. (a) Weekday price multiplier distribution. (b) Weekend price multiplier distribution. (c) Weekend price multiplier distribution, with lower bound=1

3AM on weekend, where there is high level of demand due to the end of nightlife activities and only limited number of drivers are still running on the road. More subtle differences among these approaches can be identified through the snapshots for weekday morning peak and evening peak time as shown in FIGURE 6(c)-(d). The most significant revenue gaps between GT, GR, and ADP approaches and static price are observed during these peak hours, which correspond to the time periods which have the highest demand-supply ratio, and consequently passengers are less price sensitive and drivers spend less time in searching for passengers. Moreover, the ADP approach is consistently better than GR approach, and GT approach is found to even outperform GR and ADP approaches at the first half of morning peaks.

While the market condition provides general ideas in understanding the revenue differences in different time intervals, the detailed pricing policy show exactly how change in price multipliers results in such differences, as shown in FIGURE 7(a)-(b). First, for weekdays, the GT approach generates significantly lower price multipliers than GR approach, which is expected as the objective of GT approach is to serve the maximum number of passengers. Second, the pricing policies produced by ADP and GR are very similar for weekday scenario, with ADP giving slightly smaller price multiplier in majority of the time. The reason is two-fold: 1) while there are more passengers than drivers during weekday peak hours, the gap may not be that significant, which can be observed from the value of price multiplier (with maximum value around 1.35); 2) the ADP approach has the looking ahead property which seeks to maximize the global system revenue. The philosophy under such market condition is that



(a)

(b)

(c)

(d)

Fig. 8. Served number of passengers and driver service rate. (a) Weekday passenger served. (b) Weekend passenger served. (c) Weekday driver service rate. (d) Weekend driver service rate.

it offers lower trip cost than GR approach in order to increase the utilization rate of taxi fleet, as shown in FIGURE 8(c). Note that majority of drivers in service at current time will become available after 2 to 3 time steps. This leads to the case where each driver may make less revenue per trip compared with GR approach, however, they are able to make more trips during the time period, thus winning higher total revenue. Whereas for the case observed around 3AM on weekend, there is a much more significant gap between demand and supply which is reflected by the value of price multiplier. Under such market condition, the ADP beats GR by offering lower cost initially to keep most of the drivers busy. When it comes to the time step with the highest demand, there are fewer available drivers of ADP approach than that of GR approach, which enlargers the demand-supply gap. This is in contrast to the weekday scenario, where in this case ADP approach wins the total system revenue by making fewer trips, but generating much higher revenue per trip. We can also conclude from the pattern that GP approach will grant higher revenue when there is a large gap between demand and supply, GR is more favorable for other cases, and a mixed use of GR and GP may help to generate higher system revenue. However, it is non-trivial to determine what is the ratio of demand to supply that should be considered as sufficient large for implementing GP approach, and when to start implementing GR or GP strategies, while ADP approach is observed to provide the most favorable solution.

We make a further comparison between the proposed TOD pricing policy and the existing surge pricing policy on the taxi market. As discussed before, the main difference between TOD pricing and surge pricing is that TOD pricing applies both discount and premium, depending on the market condition, while surge pricing only charge premiums when there is excessive demand than supply. The resulting price multiplier distribution during weekend is shown in FIGURE 7(c). The TOD pricing policy is modified to model the surge pricing behavior by setting the price lower bound to 1. The overall shape looks similar to the distribution of the TOD pricing policy, except that there is no discount during off-peak hours. Several subtle differences are observed, such as GT, GP, and AP have almost the same strategy during 3-5 AM, GT approach is unable to maximize the passenger trips in most of the time, and the gap between GR and ADP is larger, due to the change of price lower bound. However, based on the experiment results, we observe that surge pricing may have poorer performance compared with policy adopting both price surge and discount. The total system revenues for GT, GR, and ADP policy with only price surge are \$6.94 million, \$6.97 million, and \$6.98 million respectively. This suggests that there is potentially over \$ 0.1 million loss in daily revenue under the best scenario. As a consequence, pricing policy with only surges may be inferior than that with both discounts and surges, considering that the system revenue is potentially higher, and applying price discount during off-peaks to compensate the surges in peak hours may be better accepted by passengers.

Finally, changes in price multipliers also lead to changes in served number of trips, as well as the driver service rate, which are presented in FIGURE 8. The market with static price is observed to serve 5.08×10^5 and 5.18×10^5 daily passengers on weekdays and weekends respectively. The corresponding values for GT approach are 5.63×10^5 and 6.39×10^5, for GR approach are 4.77×10^5 and 5.34×10^5, and for ADP approach are 4.95×10^5 and 5.51×10^5. The driver service rate is calculated as the ratio of served number of passengers over the number of available drivers for each time interval. While GT approach serves more passengers than all other approaches and results in higher driver service rate as consequence, it should

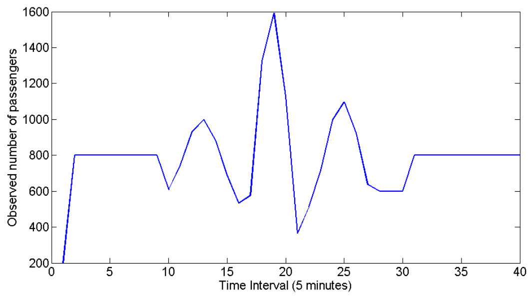


Fig. 9. Surge demand distribution.

not be considered as a stable solution since it fails to gain enough improvement in total system revenue. On the other hand, the ADP approach is found to produce higher total revenue, serve more passengers, and reach higher driver service rate compared with greedy revenue approach. Moreover, in contrast to the static price, pricing policy from ADP results in comparable served amount of passengers during weekdays and even higher driver service rate on weekend. This proves the effectiveness and validity of ADP approach, where higher system is achieved while the level of service is maintained.

D. Demand Surge

Despite the regular demand level from daily operations, there are special events or break-down of transportation systems that happen occasionally in urban areas, which may introduce a sudden surge in demand for taxi market. As a result, it is also important to investigate the applicability of TOD pricing scheme and performance of the different pricing algorithms under such condition. An artificial demand profile is created which mimics the pattern of demand surge, as shown in FIGURE 9. The time interval is set to 5 minutes. The surge in demand takes place between time interval 10 and time interval 30, and the demand peak locates around time step 19. We assume there are in average 3,000 drivers at the beginning of the scenario, with a standard deviation of 50. As a consequence, the system has insufficient amount of drivers to serve all passengers during the demand surge. All other system parameters are kept the same as in the daily operation scenario, except otherwise stated.

FIGURE 10 summarizes the results for the demand surge scenario. The total market revenue with static price is $\$2.85 \times 10^5$, the GR approach generates a total revenue of $\$4.75 \times 10^5$, and the total revenue for GT approach is $\$5.04 \times 10^5$. The ADP approach again achieves the highest total market revenue, which takes the value of $\$5.09 \times 10^5$ and is 78.6% higher than that of the static case. It can be verified that TOD pricing has a huge advantage over the static pricing scheme when the surge in demand happens.

However, different from the daily operation cases, where GR is in general better compared with GT, the GT outperforms GR approach when there is a large temporal variation in demand level. Similar pattern is also observed for the 3AM case on weekend. And the reason is that large temporal variation of demand also results in significant fluctuation in the distribution of available drivers across time. While GT approach serves more passengers before the surge takes place,

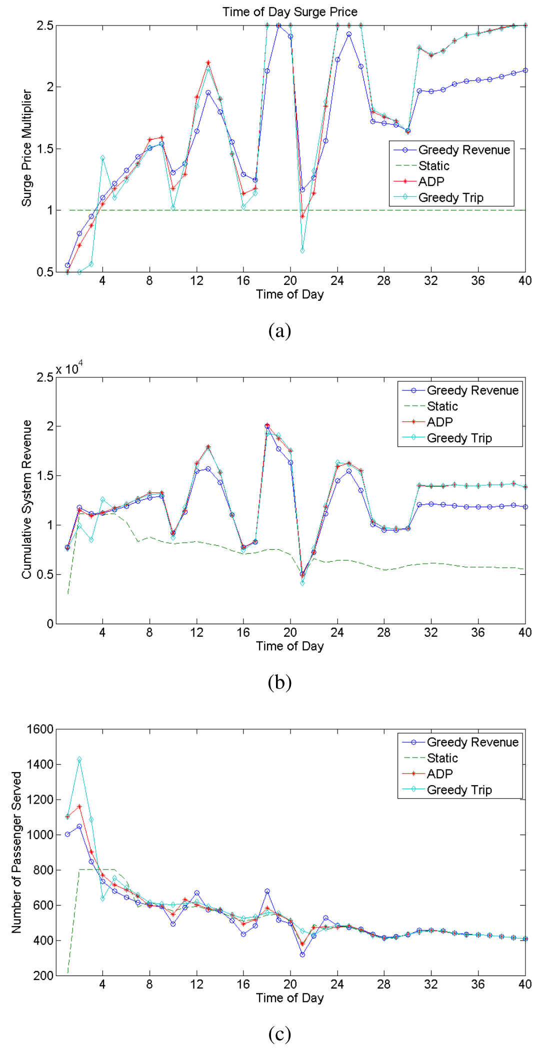


Fig. 10. Performance of TOD pricing scheme under the demand surge scenario. (a) Price multiplier. (b) Revenue per time interval. (c) Passenger served.

it also leads to fewer drivers available and a huge gap between demand and supply, which gives rise to the winning strategy of fewer trips but higher revenue per trip. But maximizing served number of passengers myopically may not lead to the highest system revenue, where ADP is seen to be more advantageous in improving total system revenue as a global optimization scheme. A subtle difference between ADP and GT can be observed from FIGURE 10(a), where ADP places a higher price multiplier than GT at time 10 and a lower price multiplier at time step 11. Consequently, it serves more passengers right before the surge peak and therefore having fewer drivers available during the surge peak. Similar behavior is also observed for the other two surge peaks, and this subtle difference contributes to the additional revenue generated by ADP approach. While there are insufficient amount of drivers, the TOD pricing is also observed to serve more passengers compared with the static case. In particular, the static case serves 2.06×10^4 passengers, the GR approach serves 2.15×10^4 passengers, the GT approach serves 2.25×10^4 passengers, and the ADP approach serves 2.21×10^4 passengers. While the objective of the ADP approach is to maximize the total system revenue, it can be seen that GT approach also produces a reasonable solution, where it is able to serve more passengers during the demand surge with slightly less system revenue.

VI. Conclusion

The distance and travel time based taxi pricing scheme has been prevalent for decades. One major drawback of the current taxi price is that it fails to take the time of day into consideration while the demand in the market is time sensitive. As a consequence, customers are seen as overcharged when the demand is low and undercharged when the demand is high. With technology advances and seeing the successes of the dynamic pricing in industries, there is an emerging need to assess if the similar mechanism is also applicable to the taxi market. This may help to provide more accurate price policy based on market condition, increase the income level of taxi drivers, and consequently improve the overall performance of the taxi industry.

In this study, we investigate the time-of-day pricing scheme which introduces price multipliers to dynamically alter trip cost. The temporal dynamics of the taxi market is discussed in detail and is modeled as the semi-Markov process with three transition functions. The issue related to incomplete trip observations is addressed by applying the EM algorithm. The optimal sequence of price multiplier is obtained by formulating market dynamics as discrete stochastic DP. Considering the complexity of solving high-dimensional DP problems, the ADP approach is introduced by approximating future values using a linear combination of basis functions, and computing function coefficients through iterative policy evaluations and updates. Current market states, resulting states for subsequent time intervals, and potential impact on future time steps are used to construct the basis functions.

Two numerical experiments are designed to validate the effectiveness of the TOD price. The first experiment is obtained from the real-world taxi trip data, which captures common daily operations of taxi market. The second experiment models sudden surges in demand caused by special events. In addition, three benchmark pricing algorithms are introduced to provide side-by-side comparison as to evaluate the performance of the ADP approach. In all cases, the ADP approach is observed to reach the highest total market revenue, and TOD pricing scheme is found to be especially efficient when there is a sudden surge in demand. Moreover, the look ahead property of the ADP approach finds a balance between the price and the served number of passengers, and generates a sequence of price multipliers that is more suitable for the taxi industry. As a result, the TOD price scheme is shown to have great potential to improve the market performance and the ADP is shown to give the best solution for the taxi industry.

The study also have several drawbacks which will be addressed in future studies. First, the study proposes an aggregated model without explicitly considering spatial heterogeneity. In real world, the level of demand, supply, and road condition may vary spatially. While the proposed model can be applied to individual area separately, future study may look into modeling multivariate distribution with mean and covariance matrices to account for the spatial interactions. This will contribute to multi-region modeling and framing more accurate pricing policies. Second, while the study only focuses on historical data, it is worthwhile modifying the model for real-time implementations. This will require the use of streaming taxi trip data, and real-time information from Google Maps can serve as the input for road congestion. Further, reliable short-term forecasting techniques are needed to estimate the market condition in the immediate future. Finally, as only homogeneous users are considered in the study, the future work may also consider different user groups with distinct price elasticity and establish the TOD pricing scheme for heterogeneous users.

References

[1] *Taxi Market in India—A Case Study of Delhi*, Valoriser Consultants, New Delhi, India, 2014.

[2] M. R. Bloomberg and D. Yassky, NYC Taxi and Limousine Commission, "The New York City taxicab fact book," NYCTLC, New York, NY, USA, Tech. Rep., 2014.

[3] T. Çetin and K. Y. Eryigit, "Estimating the effects of entry regulation in the Istanbul taxicab market," *Transp. Res. A, Policy Pract.*, vol. 45, no. 6, pp. 476–484, Jul. 2011.

[4] R. B. Coffman and C. Shreiber, "The economic reasons for price and entry regulation of taxicabs," *J. Transp. Econ. Policy*, vol. 11, no. 3, pp. 81–83, 1981.

[5] P. S. Dempsey, "Taxi industry regulation, deregulation, and reregulation: The paradox of market failure," *Univ. Denver College Law, Transp. Law J.*, vol. 24, no. 1, pp. 73–120, 1996.

[6] C. F. Daganzo, "An approximate analytic model of many-to-many demand responsive transportation systems," *Transp. Res.*, vol. 12, no. 5, pp. 325–333, Oct. 1978.

[7] H. Yang, Y. W. Lau, S. C. Wong, and H. K. Lo, "A macroscopic taxi model for passenger demand, taxi utilization and level of services," *Transportation*, vol. 27, no. 3, pp. 317–340, Jun. 2000.

[8] H. Yang, S. C. Wong, and K. Wong, "Demand–supply equilibrium of taxi services in a network under competition and regulation," *Transp. Res. B, Methodol.*, vol. 36, no. 9, pp. 799–819, Nov. 2002.

[9] H. Yang, M. Ye, W. H.-C. Tang, and S. C. Wong, "A multiperiod dynamic model of taxi services with endogenous service intensity," *Oper. Res.*, vol. 53, no. 3, pp. 501–515, 2005.

[10] Y. Liu, F. Wang, Y. Xiao, and S. Gao, "Urban land uses and traffic 'source-sink areas': Evidence from GPS-enabled taxi data in Shanghai," *Landscape Urban Planning*, vol. 106, no. 1, pp. 73–87, 2012.

[11] G. Pan, G. Qi, Z. Wu, D. Zhang, and S. Li, "Land-use classification using taxi GPS traces," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 1, pp. 113–123, Mar. 2013.

[12] X. Qian and S. V. Ukkusuri, "Spatial variation of the urban taxi ridership using GPS data," *Appl. Geogr.*, vol. 59, pp. 31–42, May 2015.

[13] R. C. P. Wong, W. Y. Szeto, S. Wong, and H. Yang, "Modelling multi-period customer-searching behaviour of taxi drivers," *Transportmetrica B, Transp. Dyn.*, vol. 2, no. 1, pp. 40–59, 2014.

[14] R. C. P. Wong, W. Y. Szeto, and S. C. Wong, "A cell-based logit-opportunity taxi customer-search model," *Transp. Res. C, Emerg. Technol.*, vol. 48, pp. 84–96, Nov. 2014.

[15] X. Qian, X. Zhan, and S. V. Ukkusuri, "Characterizing urban dynamics using large scale taxicab data," in *Engineering and Applied Sciences Optimization*. Springer, 2015, pp. 17–32.

[16] X. Zhan, X. Qian, and S. V. Ukkusuri, "Graph-based approach to measuring efficiency of urban taxi service system," in *Proc. Transp. Res. Board 94th Annu. Meeting*, 2015, p. 19.

[17] J. Yuan, Y. Zheng, L. Zhang, X. Xie, and G. Sun, "Where to find my next passenger," in *Proc. 13th Int. Conf. Ubiquitous Comput.*, Sep. 2011, pp. 109–118.

[18] N. J. Yuan, Y. Zheng, L. Zhang, and X. Xie, "T-finder: A recommender system for finding passengers and vacant taxis," *IEEE Trans. Knowl. Data Eng.*, vol. 25, no. 10, pp. 2390–2403, Oct. 2013.

[19] Z. Liao, "Real-time taxi dispatching using global positioning systems," *Commun. ACM*, vol. 46, no. 5, pp. 81–83, May 2003.

[20] S. Ma, Y. Zheng, and O. Wolfson, "T-share: A large-scale dynamic taxi ridesharing service," in *Proc. IEEE 29th Int. Conf. Data Eng. (ICDE)*, Apr. 2013, pp. 410–421.

[21] P. Santi, G. Resta, M. Szell, S. Sobolevsky, S. H. Strogatz, and C. Ratti, "Quantifying the benefits of vehicle pooling with shareability networks," *Proc. Nat. Acad. Sci. USA*, vol. 111, no. 37, pp. 13290–13294, 2014.

[22] J. Rempel, "A review of Uber, the growing alternative to traditional taxi service," Amer. Found. Blind, *AccessWorld Mag.*, vol. 15, no. 6, 2014.

[23] D. Kerr, *Detest Uber's Surge Pricing? Some Drivers Don't Like it Either*. accessed on Sep. 10, 2015. [Online]. Available: http://www.cnet.com/news/detest-ubers-surge-pricing-some-drivers-dont-like-it-either/

[24] R. Arnott, "Taxi travel should be subsidized," *J. Urban Econ.*, vol. 40, no. 3, pp. 316–333, Nov. 1996.

[25] R. Botsman. (2014). *Collaborative Economy: A Transformative Lens, Not a Start-Up Trend*. [Online]. Available: http://www.collaborativeconsumption.com/2014/11/27/collaborative-economy-a-transformative-lens-not-a-start-up-trend/

[26] NYCTLC. *TLC Trip Record Data*. accessed on May 20, 2015. [Online]. Available: http://www.nyc.gov/html/tlc/html/about/trip_record_data.shtml

[27] H. Yang, C. W. Y. Leung, S. Wong, and M. G. H. Bell, "Equilibria of bilateral taxi–customer searching and meeting on networks," *Transp. Res. B, Methodol.*, vol. 44, nos. 8–9, pp. 1067–1083, 2010.

[28] S. C. Schwartz and Y. S. Yeh, "On the distribution function and moments of power sums with log-normal components," *Bell Syst. Tech. J.*, vol. 61, no. 7, pp. 1441–1462, Sep. 1982.

[29] P. Cardieri and T. S. Rappaport, "Statistics of the sum of lognormal variables in wireless communications," in *Proc. IEEE 51st Veh. Technol. Conf. (VTC)*, vol. 3. Tokyo, Japan, May 2000, pp. 1823–1827.

[30] E. L. Kaplan and P. Meier, "Nonparametric estimation from incomplete observations," *J. Amer. Statist. Assoc.*, vol. 53, no. 282, pp. 457–481, 1958.

[31] L. R. Weatherford and S. E. Kimes, "A comparison of forecasting methods for hotel revenue management," *Int. J. Forecast.*, vol. 19, no. 3, pp. 401–415, 2003.

[32] R. H. Zeni and K. D. Lawrence, "Unconstraining demand data at US airways," in *Revenue Management and Pricing: Case Studies and Applications*. Cengage Learning EMEA, 2004, pp. 124–136.

[33] T. L. Bailey and C. Elkan, "Fitting a mixture model by expectation maximization to discover motifs in biopolymers," in *Proc. 2nd Int. Conf. Intell. Syst. Molecular Biol.*, 1994, pp. 28–36.

[34] C. C. Queenan, M. Ferguson, J. Higbie, and R. Kapoor, "A comparison of unconstraining methods to improve revenue management systems," *Prod. Oper. Manage.*, vol. 16, no. 6, pp. 729–746, Nov./Dec. 2007.

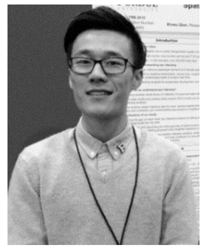
[35] R. H. Zeni, *Improved Forecast Accuracy in Airline Revenue Management by Unconstraining Demand Estimates From Censored Data*. Boca Raton, FL, USA: Universal-Publishers, 2001.

[36] W. B. Powell, *Approximate Dynamic Programming: Solving the Curses of Dimensionality*, vol. 703. Hoboken, NJ, USA: Wiley, 2007.

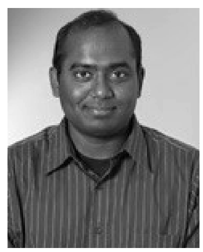
[37] W. K. Hastings, "Monte Carlo sampling methods using Markov chains and their applications," *Biometrika*, vol. 57, no. 1, pp. 97–109, Apr. 1970.

[38] D. P. Bertsekas, *Dynamic Programming and Optimal Control*, vol. 1. Belmont, MA, USA: Athena Scientific, 1995.

[39] P. L. Anderson, R. D. McLellan, J. P. Overton, and G. L. Wolfram, *Price Elasticity of Demand*, vol. 13. Mackinac Center for Public Policy, 1997, p. 2010, accessed on Jul. 15, 2015. [Online]. available: http://www.mackinac.org/1247



Xinwu Qian received the B.S. degree in transportation engineering from Tongji University, Shanghai, China, and the M.S. degree in transportation engineering from Purdue University, West Lafayette, IN, USA, where he is currently pursuing the Ph.D. degree with the School of Civil Engineering. His research interests include big data analytics, complex network analysis, and transportation network modeling.



Satish V. Ukkusuri is currently a Professor with the Lyles School of Civil Engineering, Purdue University, and is recognized nationally and internationally in the area of transportation network modeling and disaster management. He is also the Director of the Interdisciplinary Transportation Modeling and Analytics Laboratory, Purdue University. He has authored or co-authored extensively on these topics in peer-reviewed journals and conferences. His current areas of interest include: dynamic network modeling, large-scale data analytics, disaster management issues, and freight transportation and logistics.