

Modelling Passengers' Reaction to Dynamic Prices in Ride-on-demand Services: A Search for the Best Fare

SUIMING GUO, The Chinese University of Hong Kong, China

CHAO CHEN*, Chongqing University, China

YAXIAO LIU, Shenzhou UCar Inc., Tsinghua University, China

KE XU, Tsinghua University, China

DAH MING CHIU, The Chinese University of Hong Kong, China

In emerging ride-on-demand (RoD) services such as Uber and Didi (in China), dynamic prices play an important role in regulating supply and demand, trying to improve the service quality for both drivers and passengers. In this paper, we take a new perspective to study RoD services besides the supply or demand, and focus on passengers' reaction to dynamic prices. Passengers' reaction can be regarded as a process of searching for the best price before getting on a car, and the searching process reflects passengers' demand elasticity – “*how eager they are requesting a ride*”. We collect data of passengers' reaction from a real RoD service provider in China, and analyze the patterns of passengers' reaction. The analysis results show that both the dynamic prices and passengers' demand elasticity influence their reaction. We then adopt and extend a previous model for sequential search from a known distribution to understand passengers' reaction, and use our data to obtain the search costs under various circumstances, which could be interpreted as passengers' demand elasticity. Insights on the search cost and other relevant quantities are discussed. Our expectation is that the result of the study should be helpful not only for service providers in designing dynamic pricing algorithms, but also for passengers and policy makers in understanding the effects and implications of dynamic pricing.

CCS Concepts: • **Information systems** → **Location based services**; **Data mining**; • **Human-centered computing** → **User models**; *Empirical studies in ubiquitous and mobile computing*;

Additional Key Words and Phrases: User behavior, ride-on-demand service, dynamic pricing

ACM Reference Format:

Suiming Guo, Chao Chen, Yaxiao Liu, Ke Xu, and Dah Ming Chiu. 2017. Modelling Passengers' Reaction to Dynamic Prices in Ride-on-demand Services: A Search for the Best Fare. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 4, Article 1 (December 2017), 23 pages. <https://doi.org/0000001.0000001>

*This is the corresponding author.

This work is supported by National Key R&D Program of China (No. 2017YFB1002000), the National Science Foundation of China (No. 61602067), and the Fundamental Research Funds for the Central Universities (No. 106112017cdjxy180001). Authors' addresses: Suiming Guo, Department of Information Engineering, The Chinese University of Hong Kong, Hong Kong, China; Chao Chen, cschaochen@cqu.edu.cn, Chongqing University, Chongqing, China; Yaxiao Liu, Shenzhou UCar Inc., Tsinghua University, Beijing, China; Ke Xu, Tsinghua University, Beijing, China; Dah Ming Chiu, Department of Information Engineering, The Chinese University of Hong Kong, Hong Kong, China.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2017 Association for Computing Machinery.

2474-9567/2017/12-ART1 \$15.00

<https://doi.org/0000001.0000001>

1 INTRODUCTION

Emerging ride-on-demand (RoD) services¹ like Uber and Didi² have drawn increasing attention. As a supplement or substitute of the traditional taxi services, it attracts not only customers by convenience, cleanness and (sometimes) lower prices, but also drivers who want to make money using their own cars without applying for licenses. On the other hand, these services also create concerns that sometimes overwhelming prices (as high as 10 times the normal prices) may appear in special situations such as bad weather or big events.

Dynamic pricing is the core and distinctive feature of emerging RoD services, and it reflects the effort to manipulate the supply (i.e., the number of cars/drivers in a specific area) and demand (i.e., the number of passengers): a higher price reduces demand and increases supply in a busy area, and vice versa in a non-busy area. Specifically, the effects of a higher price on the supply include not only bringing more supply onto the roads, but also motivating surplus supply to flow from low- to high-demand areas.

Using dynamic prices creates a new scenario. Traditional taxi service uses (in most cases) fixed prices, i.e., independent with the supply and demand, and passengers rely on past personal experience to decide whether, when and where to get a taxi. By comparison, in emerging RoD services, passengers make decisions based on the prices they see on mobile apps. Specifically, passengers' reaction to dynamic prices before making decisions involves checking the prices for once or multiple times before finally accepting the price and getting on a car, or giving up and turning to other means of transportation. The rationale behind repeated price-checking is that sometimes one considers the price too high, and thus s/he prefers to wait for a few minutes or walk away for hundreds of meters before checking the price again.

The study on passengers' reaction is motivated by two objectives. Firstly, there are concerns among passengers and policy makers about the overly high prices (e.g., in bad weather or big events) in RoD service. A study on passengers' reaction could help to quantify passengers' perception of the dynamic prices and dissipate these concerns. Specifically, if one doesn't estimate the trip fare too many times, mostly s/he could accept the current price; and if the dynamic price is acceptable for most people, it is safer to claim that the price is not overly high. Secondly, to our knowledge, current dynamic pricing algorithms used in RoD services merely consider the supply and demand conditions nearby. But we argue that if a dynamic pricing algorithm could incorporate passengers' reaction as a feedback that in turn influences the dynamic prices, passengers may find it easier to accept the prices. For example, if we find the average passengers leaving from a particular residential area for work during morning rush hours always hesitate for a longer time before making decisions than passengers in other areas, maybe we could lower the dynamic prices (or make it change less radical) in this area during morning rush hours a little bit to make passengers more inclined to take a trip. A deeper understanding of passengers' reaction is thus a necessity before we could improve dynamic pricing algorithms.

Checking prices for multiple times could be regarded as a process of searching for the best prices. Thus, the behavior of checking the price could be considered as performing a search, and a passenger may search once again if the current search cost is smaller than the (expected) price decrease of another potential search. For example, when a passenger is going to work in the morning, the search cost may be high because s/he doesn't want to be late for work. Thus, s/he would not search for many times before getting on a car. Hence, the search cost could be a representation of passengers' demand elasticity – *how eager and necessary they are requesting a ride*.

¹RoD is an analogy to the “video-on-demand (VoD)” that we are familiar with.

²Didi: <http://www.xiaojukeji.com/index/index>

In this paper, we study passengers' reaction to dynamic prices in emerging RoD services from a user-behavior perspective. The study is based on both systematic data analysis and theoretical modelling. We collect real data from a RoD service provider to analyze passengers' reaction, answering questions like:

- *What price multipliers passengers receive for an intended trip?*
- *How many searches passengers perform before actually getting on a car?*
- *Is there any frequent pattern of passengers' behavior during the searching process?*

The goal of answering these questions is to understand the spatial-temporal distribution of dynamic prices in terms of price multipliers, as well as the frequent patterns of passengers' reaction, and the results will be further used in modelling passengers' reaction to dynamic prices. Based on the results, we find that the price multiplier and passengers' demand elasticity both influence their reaction. We then adopt and extend a model from search theory about sequential search from a known distribution, to make it suitable for our RoD service scenario. With the model and the collected data, we obtain the search costs under different circumstances. Finally, we discuss and provide insights on the search costs and other relevant quantities, such as the dependence of search costs on time and locations.

The remainder of the paper is organized as follows. §2 discusses the background and introduces the data we collect. Data analysis of the price multipliers and passengers' reaction patterns are presented in §3. The model, together with the corresponding numerical analysis and insights on search costs, is shown in §4. Finally, §5 presents related work and §6 concludes the paper and discusses future research directions.

2 BACKGROUND AND DATA

2.1 Shenzhou UCar: An Example of How a RoD Service Works

Our data of the RoD service is collected from Shenzhou UCar³, one of the major RoD service providers in China. By the end of 2015, Shenzhou UCar's service covers more than 50 cities in China, with a fleet of more than 30,000 cars, offering more than 300,000 trips per day[20].

Most emerging RoD services (e.g., Uber, Didi, Shenzhou UCar) rely on mobile apps for passengers to request a ride, and the ways in which these apps work are similar. In this subsection, we introduce how the mobile app of Shenzhou UCar works. This not only serves as a generic explanation of emerging RoD services, but also helps in understanding the data analysis and modelling that will be discussed later.

The user interface of Shenzhou UCar's app is shown in Fig. 1. Usually, a passenger first opens the app on his/her mobile phone when s/he wants to travel from a boarding location A to an arriving location B , and types the address of both locations in the app. The passenger could also choose "when to ride (now or several minutes later)" and "using coupon".

After the passenger has specified the locations and chosen all available options, the mobile app sends all the information to the service provider, and obtains in return (a) the estimated trip fare and (b) the current dynamic pricing multiplier. The price multiplier is dynamic and it reflects the current supply and demand condition around the boarding location A . The service provider sets a lower and upper bound on the price multiplier, and in our data the lower and upper bound are 1.0x and 1.6x, respectively. The estimated trip fare is the product of the price multiplier and the normal price from location A to B . The normal price is based on the estimated travelling time and distance between the two locations.

The passenger then chooses to accept the current price (by pressing "Ride a Car!" button) or gives up the current fare estimation if s/he considers the price multiplier too high. As a common practice, s/he may choose to estimate the trip fare once again (i.e., starting a new search), after either waiting for

³Shenzhou UCar: <http://www.10101111.com/>

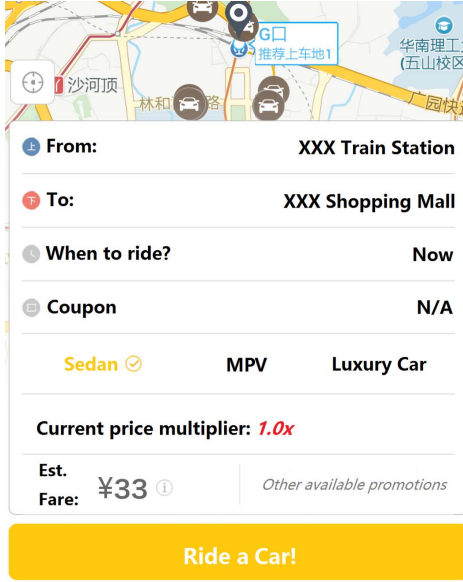


Fig. 1. The user interface of Shenzhou UCar App.

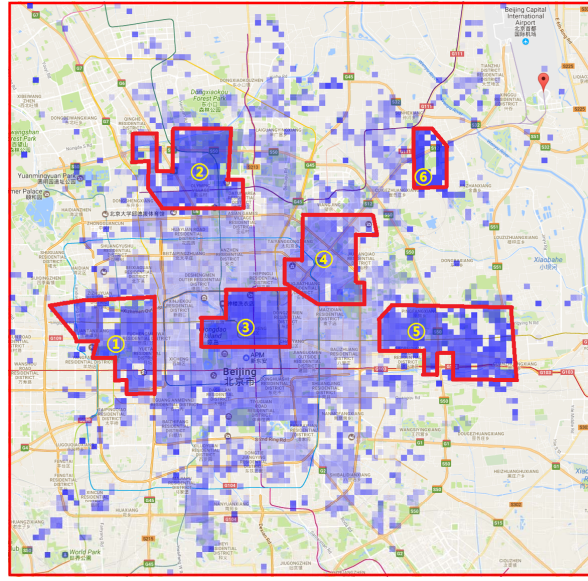


Fig. 2. Price multiplier distribution: low price during [7, 9am]

several minutes, or walking away for one or two blocks, hoping to get a lower price multiplier. Estimating the trip fare for multiple times is an indication of passengers' perception of the dynamic prices and the number of fare estimations (i.e., the number of searches) is the main focus in this paper.

2.2 Passengers' Reaction and the Event-log Dataset

Passengers' reaction to dynamic prices refers to the number of fare estimations passengers perform before finally requesting a ride or giving up. In this paper, we inspect the patterns of passengers' reaction from the event-log dataset collected from Shenzhou UCar. The event-log dataset contains records of passengers' behavior in using the app, represented by different events.

We focus on two major events in this paper. The first is the *EstimateFee* event, triggered when the mobile app sends all the information about a passenger's request (including the two locations, the requested group of cars, the time of the request, etc.) to the service provider. The result returned from an *EstimateFee* event includes the current price multiplier and the estimate trip fare. When one performs multiple fare estimations, the same number of *EstimateFee* events would be generated. The other event is the *CreateOrder* event. This event is triggered when one creates an order (i.e., requesting for a ride), and no event is triggered if s/he gives up. It is true that there are many other events corresponding to different passenger behaviors (e.g., cancelling the order), but we only discuss these two for the focus on passengers' reaction.

The event-log dataset contains the complete record of events from August to November, 2016 for 50 cities in China. Each data entry corresponds to a single event and includes the time it happens, the event code (i.e., *EstimateFee* or *CreateOrder*), the IMEI of the passenger's device, location information, the estimated trip fare and the multiplier from *EstimateFee*, etc. The volume of the dataset is 177 million for

all the 50 cities, and for Beijing alone the volume is 21 million. It should be noted that all data entries are properly anonymized due to privacy considerations.

2.3 Data Pre-processing

We choose Beijing as the target city. One of the reasons is that Beijing is a representative metropolitan city in China, and it has a range of functional areas including business, residential, entertaining, transportation areas, etc. Another reason is that as we mentioned in §2.2, Beijing is the major market of Shenzhou UCar, accounting for about 12% of the number of events. This gives us an ample amount of data.

We define a single *attempt* for a trip as a passenger's reaction from opening the app to finally creating an order or giving up. We use **E** and **C** to denote the *EstimateFee* and *CreateOrder* event, respectively. An attempt is thus represented by a series of events. For example, (EEE) means that a passenger estimates the trip fare for three consecutive times and finally gives up; and (EC) corresponds to the case that s/he estimates the fare once before finally creating an order.

In data pre-processing, one passenger may have a number of events recorded each day, and these events may belong to different attempts. For example, one particular passenger have an attempt (EEE) in the morning when going to work and an attempt (EEC) in the evening when going home. Classifying events into multiple attempts is necessary before performing data analysis. As we have mentioned, one may wait for several minutes or walk for hundreds of meters before estimating the trip fare again. So it is not safe to claim that only events with the exactly same locations belong to the same attempt. For two neighbouring events, if the latter one is either separated by more than 15 minutes from the former one, or having a boarding/arriving location more than 1.0km away from the former one, we regard the latter one as belonging to a new attempt. Using these two parameters (i.e., 15 minutes and 1.0km), we have:

- 90% of the attempts classified in this way have a timespan smaller than 18.7 minutes. The average time difference between neighboring *EstimateFee* events is 4.7 minutes.
- 77.3% of the attempts classified in this way have the exactly same boarding location; for the rest of attempts, 90% have boarding locations within smaller than 517 meters.

In short, using these parameters reasonably groups events into different attempts. The above statistics also verified our earlier claim that one may wait for several minutes or walk for hundreds of meters before estimating the trip fare again.

3 DATA ANALYSIS

In this section, we present the result of data analysis of passengers' reaction to dynamic prices. Specifically, we try to answer two questions:

- *What price multipliers passengers receive for an intended trip?*
- *How many searches passengers perform before actually getting on a car?*

3.1 Identifying Functional Areas

A large city always has a clear partitioning of functional areas including, for example:

- *business area*: the place for working. Different industries (e.g., financial or IT) may have different areas.
- *residential area*: the place for living. In China, some large residential areas accommodate $\geq 10,000$ residents.
- *transportation area*: typical transportation areas include airport terminals, railway stations for inter-city trains, etc.

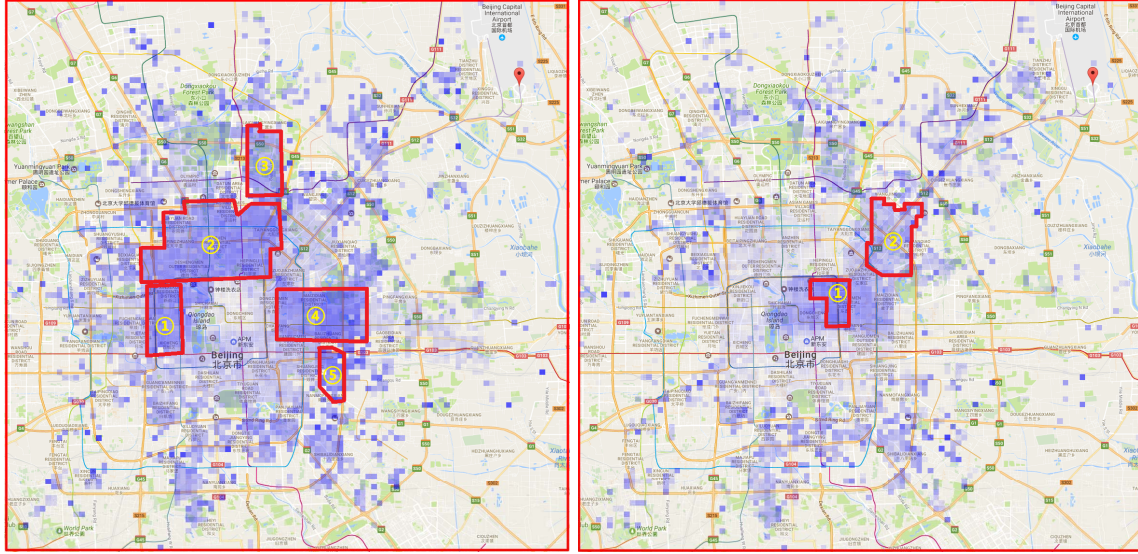


Fig. 3. Price multiplier distribution: high price during [7, 9am]. Fig. 4. Price multiplier distribution: high price during [9, 11pm].

Identifying different functional areas of a large city such as Beijing is a sophisticated task, and it is not the focus of our paper. Consulting the city plan is the easiest way, but the information may be obsolete; other methods use various techniques to analyze different sources of data [9, 25]. Here we propose a method based on the distribution of dynamic price multipliers, and compare its result with that of another method based on the boarding and arriving locations of orders in our previous work [9], also using data from Shenzhou UCar.

We first identify the distribution of functional areas based on the distribution of dynamic price multipliers. As we mentioned in §2.1, the price multiplier is the indication of dynamic pricing and varies from 1.0x to 1.6x in the data we collect. We manually categorize the price multipliers into three groups: *low* (1.0x to 1.1x), *average* (1.2x to 1.3x) and *high* (1.4x to 1.6x) multipliers. In Fig. 2 to Fig. 4, we show the distribution of low and high price multipliers during the morning rush hour [7, 9am], as well as the distribution of high price multipliers during the late evening [9, 11pm]. In these figures, the darker the blue color, the more orders there are that have the specified price multipliers.

The distribution of different groups of price multipliers during particular time period clearly shows the distribution of functional areas. In Fig. 2 to Fig. 4, we use red lines to highlight those areas in which a particular group of price multipliers prevails. For example, in Fig. 2, *low* price multipliers prevail in area 1 to 6 during morning rush hour; in Fig. 3, *high* price multipliers prevail in area 1 to 5 during morning rush hour; and in Fig. 4 *high* multipliers are more frequent in area 1 and 2 during late evening. To our knowledge and based on Beijing's city plan⁴:

- in Fig. 2 area 1 to 6 correspond to residential areas on the outer rings of the city;
- in Fig. 3 area 1 and 4 correspond to the major business areas; other areas correspond to residential areas that are closer to the city center;

⁴Beijing's city plan (in Chinese): <http://www.cityup.org/case/general/20070907/32261.shtml>

- in Fig. 4 area 1 and 2 represent a hybrid combination of business and entertaining area, corresponding to the demand from people leaving the company late at night and those spending the late evening time on entertaining activities.

The above observations are based on the heat-map of the price multiplier distributions, and, though it is not theoretically rigorous enough, it is already a clear representation of some key functional areas used in the remainder of this paper.

We then present briefly the method in [9] based on the boarding and arriving locations of orders. This method only considers frequent passengers – those who have more than 30 rides per month, and we found that in most cases a frequent passenger's rides exhibit strong regularity and are within less than 4 locations. We used the k -means clustering algorithm [18] to cluster the boarding and arriving locations of the orders from frequent passengers, and formed 20 clusters that represent the distribution of functional areas of Beijing. We don't copy the results from [9] here because of limited space, but the comparison between these two works show that:

- The identification result of most residential and business areas match between these two works.
- The method proposed in this paper is able to identify a number of areas that are hybrid combination of business and entertaining areas, corresponding to the demand from those leaving work late at night or those spending the late evening on entertaining activities. The method in [9] failed to do so because these areas are not the regular locations of frequent passengers.

Based on the observations above, in the remainder of this paper, we choose the highlighted areas in Fig. 2 and area 1 & 4 in Fig. 3 as the representative residential and business areas for the study. Additionally, we choose the airport of Beijing and another major train station as the representative transportation areas (not shown here).

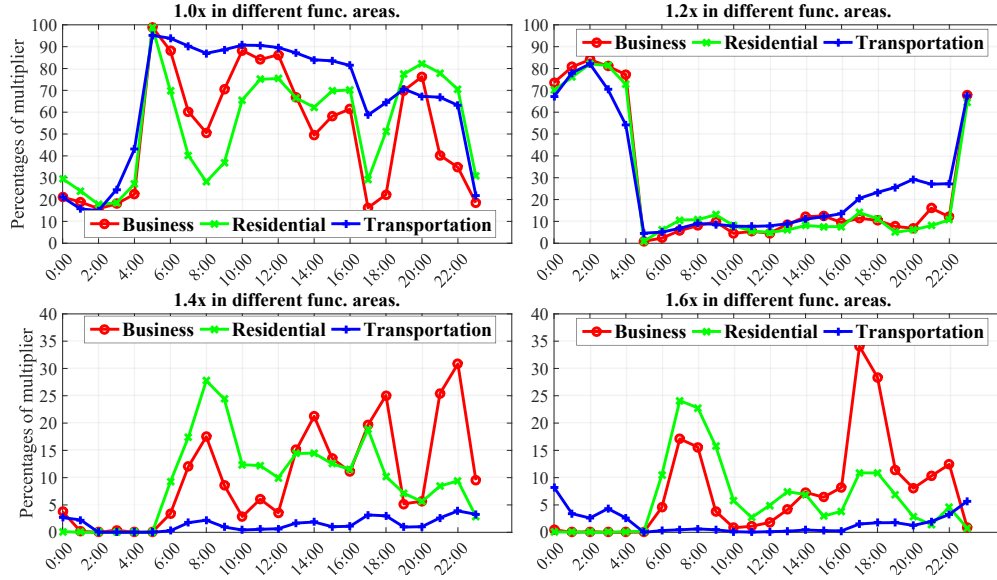
3.2 Price Multipliers: Spatio-temporal Distribution

In this subsection we try to answer the first question proposed in the beginning of §3, by analyzing the distribution of price multipliers that passengers receive during fare estimation. Understanding this distribution lays a foundation for the following data analysis and modelling of passengers' reaction patterns. As mentioned previously, normally a RoD service provider sets a lower bound m and upper bound U for price multipliers. In our data collected from Shenzhou UCar, we have $m = 1.0$ and $U = 1.6$.

In the following analysis, we present the distributions of price multipliers passengers receive in *EstimateFee* events. Specifically, we show the percentage of fare estimations that return a particular price multiplier among all fare estimations, during different hours of the day, in different functional areas. In Fig. 5, we show the corresponding percentages of price multipliers 1.0x, 1.2x, 1.4x and 1.6x in business, residential and transportation area on weekdays. We omit the percentages of multipliers 1.1x, 1.3x and 1.5x because of the limited space, and as the percentage of 1.1x is very similar to that of 1.2x, and for some unknown reason, 1.3x and 1.5x are relatively rare in our data compared to 1.4x and 1.6x, we thus believe the omission here does not lead to the loss of generality.

The distribution of different price multipliers here will be used later in the modelling of passengers' reaction. Here we present some observations from Fig. 5, which is useful and informative for understanding the pattern of price multipliers in a RoD service:

- (1) Comparing among the three functional areas, in transportation area the price multiplier is the most stable and high price multipliers (i.e., 1.4x and 1.6x) are the rarest.
- (2) For the *average* multiplier 1.2x, in all three functional areas it has a high percentage during [11pm, 5am]. Specifically, in transportation area the percentage of 1.2x starts dropping from 84% after

Fig. 5. Price multipliers in *EstimateFee* events.

2am and reaches about 50% during [4am, 5am]; in the other two areas the percentages of 1.2x are more stable and remain higher than 70% until 5am. The explanations are:

- The overall high percentage of 1.2x over the night is due to the lower supply & demand in this period.
 - The relatively lower percentage of 1.2x in transportation area is that even at late night, there are still some drivers going to this kind of area (e.g., airport) to pick up the relatively predictable demand.
- (3) For the *low* multiplier 1.0x, correspondingly, in all functional areas the percentage is low during [11pm, 5am]. In transportation area the percentage is always higher than 60% during [5am, 11pm]. In the other two areas, there are obvious peaks and troughs (the discussions will be presented later based on 1.4x and 1.6x). Numerically, the percentage in business area reaches the lowest (about 17%) during [5pm, 6pm], whereas in residential area the lowest percentage (about 29%) appears during [8am, 9am] and [5pm, 6pm].
- (4) For the *high* multiplier 1.4x, we only consider business and residential areas:
- In residential area, the global peak is during [8am, 9am] with the percentage as 28%, and there are local peaks during [1pm, 2pm] (15%) and during [5pm, 6pm] (20%). The global peak corresponds to people going to work in morning rush hour; the 1st local peak is the result of people going out to have lunch or going to work after noon break; the last local peak is the result of the demand returning from work from those people living near workplaces.
 - In business area, the global peak is during [10pm, 11pm] with the percentage as 31%, and there are three local peaks during [8am, 9am] (16.5%), [2pm, 3pm] (21%) and [6pm, 7pm] (25%). These peaks correspond to the demand from people working overtime, living near workplaces, having lunch and returning to workplaces, and leaving workplaces in the evening rush hour, respectively.
- (5) For the *high* multiplier 1.6x, similarly we only consider business and residential areas:

- In residential area, the global peak is during [7am, 8am] with the percentage as 24.7%, and the second peak is during [5pm, 7pm] (11.3%). This indicates that during the morning rush hour [7am, 9am], 1.6x prevails in the first half and 1.4x prevails in the second half.
- In business area, the global peak is during [5pm, 6pm] with the percentage as 35.3%. For other local peaks, the percentage is over 15% during [7am, 9am] and is about 12.7% during [10pm, 11pm]. Similar to the observation in residential area, here during the evening rush hour [5pm, 7pm], 1.6x prevails in the first half and is still has a percentage slightly higher than 1.4x during the second half. Additionally, in the late evening [10pm, 11pm], the percentage of 1.6x is much smaller than that of 1.4x, because the demand from people working overtime is not as high as the demand in the evening rush hour.

3.3 Patterns of Passengers' Reaction

We have already defined the *attempt* for a trip in §2.3 and use a series of events represented by **E** and **C** to describe attempts. Attempts are the representation of passengers' reaction, and the patterns of attempts are an indication of passengers' perception as well as feedback of the dynamic prices they encounter. In short, we find from our data that the price multiplier and passengers' demand elasticity both have impacts on passengers' reaction. We will show this in this subsection.

Before presenting the patterns of passengers' reaction, we first show some examples to illustrate the need to analyze these patterns:

- Passenger reaction (**EC**) means that one only estimates the trip fare once before creating the order. The reason may be that the passenger is satisfied by and could accept the current price multiplier.
- (**EEC**) means that one estimates the fare twice before creating the order. The reason may be that the passenger is not satisfied with the first fare estimation, and thus s/he wants to try once more to get a lower price. The two estimations may be separated by several minutes, and the passenger may have moved hundreds of meters to perform the second estimation.
- (**EEE**) means that one performs three estimations, but finally gives up. A possible reason is that s/he still considers the price multiplier too high and decides to switch to other means of transportation (e.g., taxi, bus, metro).

We count the frequencies of patterns in the attempts leaving from different functional areas, during different hours of the day, on weekdays. In Fig. 6 we show the frequencies of the top-6 patterns (ranked according to their frequencies) – (**EC**), (**E**), (**EEC**), (**EE**), (**EEE**) and (**EEEC**) – in business, residential and transportation area in different hours. Note that besides these six patterns, there are many possible patterns that exist in our data, including patterns with a lot of **Es** (i.e., involving many times of fare estimation), but these patterns all have frequencies smaller than 2.5%. So in Fig. 6 we only show the top-6 patterns. In the following analysis, we use a two-elements tuple to represent each situation:

- (1) the 1st element: the functional area in which the frequency is counted – business (“B”), residential (“R”), transportation (“T”) or the whole city (“W”);
- (2) the 2nd element: during which hour the frequency is counted. A value t ranging from 0 to 23 means the hour $[t, t + 1]$. The value “A” means that the frequency is counted for the whole day.

Besides, in Tab. 1 we also show the frequencies of the top patterns leaving from different functional areas for the whole day on weekdays. Situations chosen in Tab. 1 include (W,A), (B,A), (R,A) and (T,A), using the two-elements tuple representation.

We have the following observations:

- (1) Patterns (**EC**) and (**E**) are two major patterns, in any situation. In other words, most passengers are not patient enough to estimate the price for more than once.

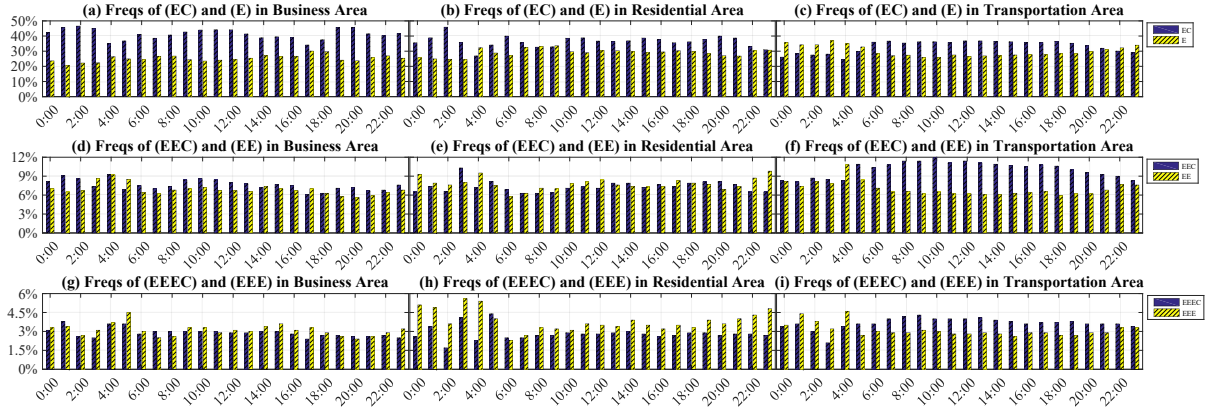


Fig. 6. Frequencies of the top-6 patterns of passengers' reaction in different functional areas.

Table 1. Passengers' reaction: top-6 patterns & their frequencies, averaged for the whole day.

(W,A)		(B,A)		(R,A)		(T,A)	
pattern	freq (%)	pattern	freq (%)	pattern	freq (%)	pattern	freq (%)
(EC)	37.5	(EC)	41.0	(EC)	36.5	(EC)	34.6
(E)	28.9	(E)	26.1	(E)	29.9	(E)	28.8
(EE)	7.3	(EEC)	7.3	(EE)	7.6	(EEC)	10.4
(EEC)	6.9	(EE)	6.6	(EEC)	7.2	(EE)	6.6
(EEE)	3.4	(EEE)	3.0	(EEE)	3.5	(EEEC)	3.8
(EEEC)	2.6	(EEEC)	2.8	(EEEC)	2.8	(EEE)	2.9

- (2) Comparing top patterns and their frequencies averaged for the whole day in Tab. 1:
 - In business area, (**EC**) and (**E**) have the highest and lowest frequency, respectively, compared to other functional areas. This indicates that passengers in business area are the most eager to get a car, maybe because they have smaller demand elasticity, or their companies pay for their trips.
 - In transportation area, (**EEC**) and (**EEEC**) have the highest frequencies, but (**EC**) has the lowest frequency, compared to other functional areas. In other words, when leaving transportation area, those who finally get a car are not that eager and are willing to wait for a longer time.
 - In business area, the demand is the highest: even though passengers estimate the prices for multiple times, more than half finally accept the prices and get a car. Passengers in residential area behave in the opposite way. We could see this from patterns ended with a **C**, i.e., the passenger finally gets a car. For business, residential and transportation area, the frequencies of these patterns are 51.1%, 46.5% and 48.8%, respectively. Contrarily, patterns not ended with a **C** have frequencies of 35.7%, 41.0% and 38.3% for these areas.
- (3) Comparing top patterns and their frequencies in different hours of the day in Fig. 6:
 - Similar to our previous observations in transportation area, the frequency of (**EEC**) (or (**EEEC**)) is significantly higher than that of (**EE**) (or (**EEE**)), for any time during [5am, 11pm], but the frequency of (**EC**) does not have a significant difference with that of (**E**).

- Fig. 6 and Tab. 1 show that most passengers would not choose to estimate too many times. Specifically, in any hour of the day, patterns with one, two and three **E**s have frequencies smaller than 50%, 12% and 6%, respectively. It is thus necessary to consider passengers' reaction in pricing algorithms to make passengers feel better and attract more passengers.
 - In business area, (**EC**) and (**E**) have the lowest and highest frequency during the evening rush hour [5pm, 7pm], compared to any other time during [5am, 11pm]. Similarly, in residential area, (**E**) has the highest frequency during the morning rush hour [7am, 9am]. This shows that passengers in rush hour, either in residential or business area, are more inclined to estimate once and give up, compared to any other time in the day. The reason may be that the price multiplier in rush hour is too high for them to accept.
- (4) Combining all previous observations and those from §3.2, we conclude that both the **dynamic price multipliers** and **passengers' demand elasticity** influence passengers' reaction:
- For price multipliers, when it is high, passengers are more inclined to give up after one or more fare estimations, instead of creating orders, and vice versa.
 - For demand elasticity, passengers in business area not only are the most eager to get a ride, but also have the highest frequency of finally getting a ride instead of giving up, even though sometimes price multipliers in this area are higher. Also in transportation area, the frequency of finally getting a car is in between the corresponding frequencies in the other two areas, but those who finally get a car are not that eager and are willing to wait in order to get a lower multiplier.

4 MODELLING OF PASSENGERS' REACTION

We have already presented in §3 our data analysis on the price multipliers during fare estimations, and the patterns of passengers' reaction. We show the distributions of price multipliers leaving different functional areas in different hours on weekdays. We also show the frequencies of different patterns of passengers' reaction under these circumstances.

With all the data analysis, we observe that both the dynamic prices and passengers' demand elasticity influence their reaction. Next, we go further and model passengers' reaction using search theory [21]. The model builds a relationship between *the distribution of price multipliers*, *passengers' demand elasticity* and *passengers' reaction to dynamic prices*. Given the historical data of price multipliers' distribution and passengers' reaction to prices, we could deduce the spatio-temporal properties of passengers' demand elasticity; and conversely, we could predict passengers' reaction to prices, given the current price multiplier distribution and learned passengers' demand elasticity.

This model has already covered some important aspects, by building a relationship between the three factors mentioned above. For example, *how passengers react to "raising price in rush hours" (or contrarily, "decreasing price in non-rush hours")* is represented by a change of the distribution of price multipliers. In rush hours, the probability of having higher multipliers is higher, so the distribution of price multipliers is shifted towards higher multipliers; and in non-rush hours, things are just the opposite. By considering the price multipliers distribution, we have already considered the aspect of "raising price in rush hours". As another example, *"the type of rides"* is also implicitly considered by the passengers' demand elasticity. For instance, when we consider the demand elasticity of passengers in the transportation area, we focus on those rides "leaving from the airport (or train station)"; when we choose passengers in the residential area during morning rush hours, the type of rides is mostly about "leaving for work in the morning". In other words, the differences in the demand elasticity in different locations and hours represent an inference of the type of rides.

4.1 The Model

We adopt and extend a simple model from search theory to understand passengers' reaction in RoD service. The author in [21] proposes a simple model to understand the process of sequentially searching for the lowest prices when buying a particular product, or a particular category of products. The distribution of the prices of this product is assumed to be known beforehand. Before buying the product, the consumer searches for the lowest price among different distributors. A single search may be going to a shop to get a price quotation, or going to an online shop to search the product's price. Each search incurs a *search cost*, and returns a price. S/he may keep searching when the search cost is less than the expected price decrease if conducting a new search, or stop searching and buy the product otherwise. There are two possibilities when buying the product: s/he may be able to return to the distributor that offers the lowest price among all searches; or could only buy the product at the price offered in the last search. In solving how many searches need to be conducted, [21] concludes that no matter which possibility applies, the consumer should stop searching as soon as a search returns a price lower than a threshold and buy at this price. With this stop condition, the last search always returns the lowest price.

The scenario used in the above model is basically similar to the scenario in RoD service. Firstly, in RoD service a passenger still searches for multiple times, by opening the mobile app and typing the addresses of the intended trip. Secondly, the passenger wants to find the lowest price before getting on a car. So the scenario is still about comparing the search cost with the expected price decrease.

There are also some differences that make it necessary to extend the above model if we want to use it to explain passengers' reaction in RoD service:

- (1) The search cost needs a different explanation. In the original model, the search cost could be regarded as the cost of going to the store, or the cost of using the Internet (for online distributors). In RoD service, a single search only involves using a mobile app and pressing some on-screen buttons, and the cost of doing it is negligible. In fact, *now the search cost should be interpreted as a representation of passengers' elasticity*. The cost is high when the passenger is eager to take a ride and cannot wait to estimate the trip fare for multiple times; the cost is lower when s/he is not in a hurry and could wait longer before taking a ride. An example of higher search cost is going to work in the morning rush hour.
- (2) Instead of searching for the lowest price, now in RoD service a passenger searches for the lowest price multiplier. We have mentioned in §2.2 that the estimated trip fare consists of a base price and a price multiplier, and we could safely assume that the passenger is aware of the base price based on his personal experience, and the price multiplier is the only thing that s/he cares about. The base price is only about the trip distance and the price multiplier is a reflection of the supply and demand condition. This change means that now the search target is a discrete value (i.e., the price multiplier), instead of a continuous value (i.e., price).
- (3) Changing the search target in (2) also changes our focus from absolute to relative value. Correspondingly, the search cost now should be interpreted as the relative search cost. The model should only make use of the relative search cost, and the real (absolute) search cost of an intended trip should be the product of the relative cost and the base price of the trip. Using a relative search cost is tenable, as passengers intended to go for longer trips (in terms of time or distance) usually have more time to hesitate and are thus more likely to wait longer to get lower prices.

We use m and U to denote the minimum and maximum price multiplier, and use c to represent the relative search cost for a single search. In our problem, the price multiplier x in any search can take discrete values in $[m, U]$, separated by Δ . In other words:

$$x \in \{m, m + \Delta, m + 2\Delta, \dots, U - \Delta, U\}. \quad (1)$$

Table 2. Main symbols used in the model.

Symbol	Meaning
m, U	the minimum (or maximum) price multiplier.
Δ	the difference between two possible consecutive price multipliers.
c	the (relative) search cost for a single search.
x	the random variable representing the price multiplier. $x \in \{m, m + \Delta, m + 2\Delta, \dots, U - \Delta, U\}$.
$P(x)$	the probability distribution of price multipliers.
$F(x)$	the cumulative distribution function of price multipliers, $F(x) = \sum_{x_0 \leq x} P(X = x_0)$.
$M(x)$	a function used in modelling, $M(x) = \sum_{x_0 \leq x} x_0 P(X = x_0)$.
$G(x)$	a function used in modelling, $G(x) = xF(x) - M(x)$.
a	the multiplier threshold: the solution to $G(x) = c$.

The distribution of price multipliers in a single search is assumed to be known and remains unchanged during multiple searches. The searches for a single attempt usually take place in a short time, with 90% of attempts having a time span smaller than 18.7 minutes. Moreover, the model will be applied to particular functional areas during every hour in a day. So the assumption that the distribution remains unchanged holds, as we only consider a short time period (≤ 1 hour) and a relatively small functional area. We use $P(X = x_0)$ and F to denote the probability of getting a multiplier x_0 and the corresponding cumulative probability function, i.e., $\sum_{x_0=m}^U P(X = x_0) = 1$ and $F(x) = \sum_{x_0 \leq x} P(X = x_0)$. Tab. 2 summarizes the main symbols.

Let's say the passenger has already searched for t times, and the lowest price multiplier among these searches is x_{1t} . We use the random variable X to denote the price multiplier in the next search. The problem is to determine if one stops searching after t searches or not.

Similar to [21], we first assume that the passenger could return to a previous lower price when taking a ride. We define a loss function $L(X)$ that characterizes the potential loss of the $(t + 1)^{\text{th}}$ search:

$$L(X) = \begin{cases} X - m + c, & \text{if } X < x_{1t} \\ x_{1t} - m + c, & \text{if } X \geq x_{1t} \end{cases}. \quad (2)$$

The expected loss of the $(t + 1)^{\text{th}}$ search is:

$$\begin{aligned} E[L(X)] &= \sum_{x=m}^{x=x_{1t}} (x - m + c)P(X = x) + \sum_{x=x_{1t}+\Delta}^U (x_{1t} - m + c)P(X = x) \\ &= (c - m) + [1 - F(x_{1t})]x_{1t} + \sum_{x \leq x_{1t}} xP(X = x) \\ &= (c - m) + [1 - F(x_{1t})]x_{1t} + M(x_{1t}), \end{aligned}$$

with the definition that $M(x_{1t}) = \sum_{x \leq x_{1t}} xP(X = x)$.

The passenger should stop searching if the loss at the t^{th} search, denoted by $L(x_{1t}) = x_{1t} - m$, is smaller than the expected loss at the $(t + 1)^{\text{th}}$ search:

$$\begin{cases} \text{stop searching,} & \text{if } L(x_{1t}) \leq E[L(X)], \\ \text{continue searching,} & \text{if } L(x_{1t}) > E[L(X)]. \end{cases} \quad (3)$$

With the calculation of $E[L(X)]$, (3) becomes:

$$\begin{cases} \text{stop searching,} & \text{if } x_{1t} \leq c + [1 - F(x_{1t})]x_{1t} + M(x_{1t}), \\ \text{continue searching,} & \text{if } x_{1t} > c + [1 - F(x_{1t})]x_{1t} + M(x_{1t}). \end{cases} \quad (4)$$

If we define $H(x) = xF(x) - M(x) - c$, then $H(x)$ is an increasing function with x , as:

$$\begin{aligned} H(x + \Delta) - H(x) &= (x + \Delta)F(x + \Delta) - M(x + \Delta) - xF(x) + M(x) \\ &= F(x)\Delta, \end{aligned}$$

because $F(x + \Delta) = F(x) + P(x + \Delta)$ and $M(x + \Delta) = M(x) + (x + \Delta)P(x + \Delta)$. Let $G(x) = xF(x) - M(x)$, and $G(x)$ is also an increasing function. We also define a to be the solution to $G(x) = c$. We could then rewrite (4) as follows:

$$\begin{cases} \text{stop searching,} & \text{if } x_{1t} \leq a, \\ \text{continue searching,} & \text{if } x_{1t} > a. \end{cases} \quad (5)$$

As shown in (5), we call a as **the multiplier threshold**, as the passenger should stop searching once getting a price multiplier smaller than this value. The stop condition in (5) also holds when $G(U) \leq c$: the passenger should stop searching no matter what price multiplier s/he gets.

Even though we have the assumption at the beginning that the passenger could return to a previous lower price to request for a ride, (5) shows that the strategy for him/her is to stop searching as soon as the price multiplier falls below a particular threshold. So, the assumption does not matter any more, and it is true that in practice passengers in RoD service are unable to return to a previous lower price.

To sum up, we have the following proposition:

Proposition 4.1 (Stopping strategy). *Given the discrete probability distribution $P(X = x)$ and cumulative distribution function $F(x) = \sum_{x_0 \leq x} P(X = x_0)$ of price multipliers $x \in [m, U]$, as well as the search cost for a single search c , we define:*

$$M(x) = \sum_{x_0 \leq x} x_0 P(X = x_0), \quad (6)$$

$$G(x) = xF(x) - M(x). \quad (7)$$

If $G(U) \leq c$, the passenger should only search once and then take the ride;

*Otherwise, let a be the solution to $G(x) = c$, and the passenger should stop searching as soon as s/he gets a price multiplier smaller than or equal to a . We call a as the **multiplier threshold**.*

We also have the following remarks, concerning the function $G(x)$:

Remark. Concerning the function $G(x)$,

- (1) $G(x)$ is an increasing function with x . The minimum value is $G_{min} = G(m) = m * F(m) - M(m) = 0$. The maximum value is $G_{max} = G(U) = U * F(U) - M(U) = U - E[X]$. $E[X]$ is the expected price multiplier, given its distribution. In other words, for areas or hours that passengers get higher price multipliers, the G_{max} is smaller; and vice versa.
- (2) The difference of the G function values between two consecutive price multipliers, i.e., $G(x)$ and $G(x + \Delta)$, is $G(x + \Delta) - G(x) = F(x)\Delta$. As a result, an abrupt change in the slope of $G(x)$ at some particular point x_c indicates that the probability of getting the price multiplier x_c is high.
- (3) **(The meaning of $G(x)$.)** With the stopping strategy and the fact that $G(x)$ is an increasing function, we claim that $G(x)$ represents the expected decrease of the price multiplier of a potential search. So if the expected decrease is smaller than the search cost, the passenger should stop.

4.1.1 Discussions. We discuss the use and applicability of the model here.

The use of the model. The two quantities – multiplier threshold (i.e., passengers' reaction to the dynamic prices) and search cost (i.e., passengers' demand elasticity) – are not independent with each other: knowing the price multiplier distribution and any one of the two decides the other. How to use the model is dependent on the goals:

- **Learning the search cost.** Based on the historical data of the price multiplier distribution and passengers' reaction to prices, we could use the model to obtain the spatio-temporal distribution of the search cost, which is the representation of passengers' demand elasticity.
- **Predicting passengers' reaction.** Based on the current data of the price multiplier distribution and the learned spatio-temporal distribution of search cost, we could use the model to predict how passengers would react to the dynamic prices in the next hours.

Using the model to perform user-behavior prediction is beneficial to both the service provider and passengers. On one hand, the service provider could use it to improve the acceptance of orders, by incorporating user-behavior prediction in designing dynamic pricing algorithms, as mentioned in §1: if passengers in particular locations or hours are predicted to hesitate for a long time, we should consider lowering the price a little bit or making it change less radical, to lure more passengers to use the service. On the other hand, the service provider could inform passengers the prediction of user-behavior and let them know that they could just request a ride and go without hesitation if they get a price multiplier lower than the threshold. This helps passengers make decisions.

The applicability of the model. The model is applicable in the cases that the service provider does not tweak the dynamic prices based on the reaction of any individual passenger. An example of such a tweak is: if the service provider finds a passenger estimating the trip fare for three times without making a decision, it lowers the price for this specific passenger to lure him/her to the service. In fact, as we learn from Shenzhou UCar at the time of our research, most, if not all, dynamic pricing algorithms used in the industry only consider the supply & demand condition in various forms, and do not consider any factor related to passengers' reaction. As an example, we have mentioned in [9] a generic dynamic pricing algorithm and pointed out that the price multiplier is dependent on a particular decisive factor (calculated based on the supply & demand condition) near the boarding location of an intended trip.

Furthermore, as we mentioned in §1, we hope future dynamic pricing algorithms should consider passengers' reaction in determining price multipliers. This does not mean that the service provider should tweak the prices on the basis of individual passenger; instead, the reaction of *average* passengers (in particular locations or during particular hours) should be considered. In this way, the passengers' reaction is a feedback to the dynamic pricing algorithm, and the equilibrium should be studied, leaving the model of passengers' reaction largely intact.

4.2 Numerical Analysis and Insights

We have already discussed the theoretical model in §4.1, and in this subsection we use our collected data to perform some numerical analysis based on the model. In the following, we present numerical analysis results of the G function, the multiplier threshold and the search cost under different circumstances. The meaning of these quantities have been explained in Proposition 4.1 and the corresponding remark.

We have already presented the data analysis in §3 of the price multipliers that passengers receive, and of the patterns of their reaction under different circumstances. We will put them into our model, and reversely obtain the multiplier threshold and the search cost under these circumstances, i.e., we learn the search cost from our data. Relevant insights are also presented.

From the collected data, we have $m = 1.0$, $U = 1.6$ and $\Delta = 0.1$. So $x \in \{1.0, 1.1, 1.2, \dots, 1.6\}$. The distributions of price multipliers $P(x)$ (during different hours and in different functional areas on weekdays) have already been given in Fig. 5 in §3.2 (with 1.1x, 1.3x and 1.5x are omitted for reasons already explained). Also, from the analysis in §3.3, we are able to calculate the average number of *EstimateFee* events before creating an order. All results together, are enough to reversely calculate the multiplier threshold and the search cost.

Though in this subsection we only present the relevant quantities (i.e., the G function, multiplier threshold and search cost) in different functional areas and during different hours, our model could be extended to incorporate more possible circumstances. For example, if we had the data about the weather condition (e.g., the hourly temperature or precipitation) in different locations, we could group those attempts having close weather condition together and calculate the corresponding search cost. We could then deduce the relationship between the search cost and weather condition by performing curve-fitting or using other data mining techniques.

4.2.1 The G Function. We first show the $G(x)$ function under some representative circumstances in Fig. 7 to Fig. 10. Each value of $G(x)$ is calculated based on (7). The goal of showing the $G(x)$ function is to validate our remarks in §4.1 and to provide a basic understanding of it in our model.

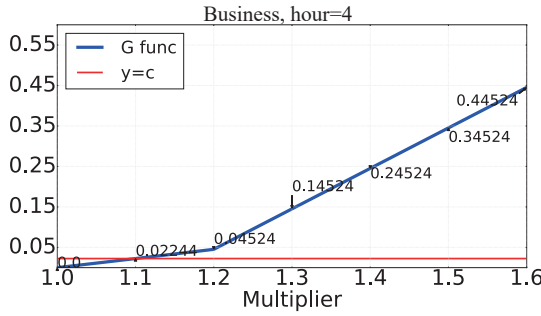


Fig. 7. The G function: leaving business area during [4am, 5am] on weekdays.

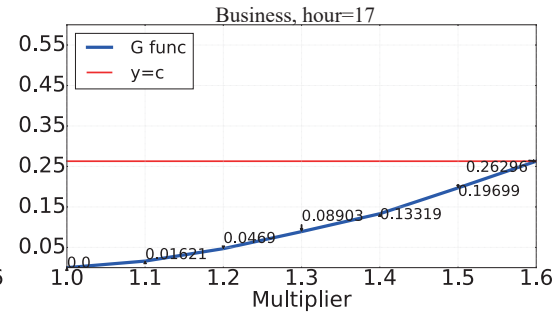


Fig. 8. The G function: leaving business area during [5pm, 6pm] on weekdays.

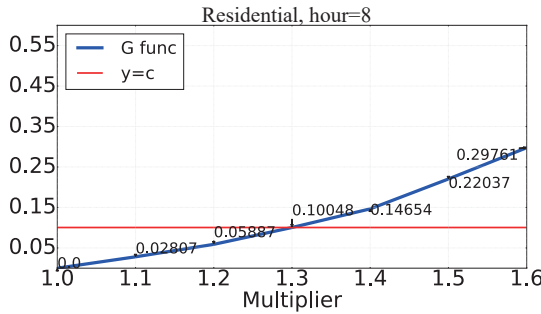


Fig. 9. The G function: leaving residential area during [8am, 9am] on weekdays.

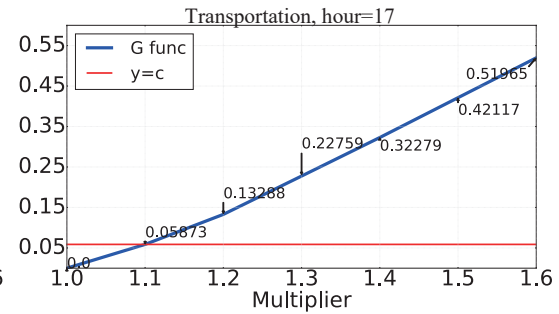


Fig. 10. The G function: leaving transportation area during [5pm, 6pm] on weekdays.

In each figure, the blue line is $y = G(x)$ against x , and each value of $G(x)$ is annotated on the figure. The red line is a constant value $y = c$, showing the search cost, and will be discussed later in §4.2.3.

Fig. 7 and Fig. 8 show the G function describing the price multipliers for fare estimations leaving business area during [4am, 5am] and [5pm, 6pm] on weekdays. These circumstances are selected because they represent a non-rush hour and an evening rush hour. We have the following observations:

- (1) The higher value of $G(1.6)$ in Fig. 7 indicates that the expectation of price multiplier is lower for fare estimations during [4am, 5am], and Fig. 8 shows the opposite. In fact, the expectation of price multiplier is 1.155 and 1.337 for these two circumstances.
- (2) In Fig. 7, the slope change at $x = 1.2$ means that the $1.2x$ price multiplier has the highest probability. Similarly in Fig. 8 the slope changes at $x = 1.2, 1.4$ and 1.6 mean that these three price multipliers have higher probabilities than others.

Similarly, Fig. 9 and Fig. 10 show the G function for price multipliers returned by fare estimations leaving residential area during [8am, 9am] and leaving transportation area during [5pm, 6pm] on weekdays. The circumstance in Fig. 9 corresponds to going to work in the morning, indicating a higher expectation of price multiplier and higher probabilities of having multiplier $1.2x$, $1.4x$ and $1.6x$. Contrarily, the circumstance in Fig. 10 corresponds to leaving the transportation area in the afternoon, a relative busy hour in this area. But the expectation of price multiplier is still the lowest among these 4 figures, and the probabilities of having multiplier $1.1x$ and $1.2x$ are higher than others. All these observations agree with those mentioned in §3.2.

4.2.2 The Multiplier Threshold. We also calculate in reverse the multiplier threshold under different circumstances. Proposition 4.1 states that any passenger should stop searching for the lowest price multiplier when s/he gets an estimated multiplier smaller than a threshold. To obtain the multiplier threshold from our data, we only consider patterns that finally lead to the creation of an order, e.g., (EC), (EEC), (EEEC) and etc. For each attempt belonging to a particular pattern, we record the price multiplier in the last *EstimateFee* event (and call it as the *last multiplier*), and this is the multiplier the corresponding passenger finally accepts. All *last multipliers* belonging to these patterns form a list, and we use the 90th-percentile of the list as the multiplier threshold. The reason of choosing the 90th-percentile instead of the maximum of the list is to handle outliers in the data. If the multiplier threshold obtained in this way is very close to 1.0, it means that the corresponding *last multipliers* are very close to 1.0 and we thus use the average of these *last multipliers* as the multiplier threshold, to avoid getting zero search cost in §4.2.3. In this way, we are able to obtain the multiplier thresholds under different circumstances.

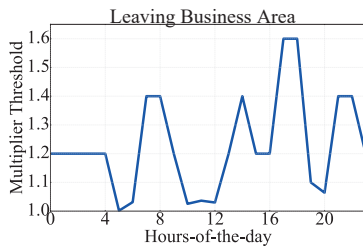


Fig. 11. The multiplier threshold: leaving business area on weekdays.

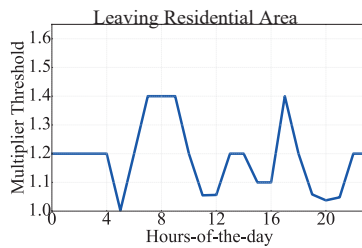


Fig. 12. The multiplier threshold: leaving residential area on weekdays.

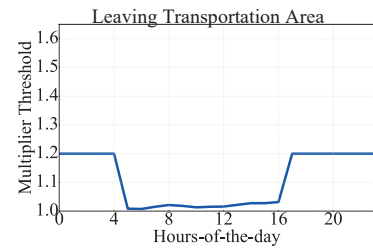


Fig. 13. The multiplier threshold: leaving transportation area on weekdays.

In Fig. 11 to Fig. 13 we show the multiplier threshold for attempts leaving business, residential and transportation area on weekdays, against different hours of a day. Figures for weekends are not shown here because of limited space.

The goal of obtaining the multiplier threshold under different circumstances is to give us an understanding about passengers' reaction to dynamic prices, and to serve as the foundation of calculating the search cost. At this point, we have the following observations concerning the multiplier threshold:

- (1) The peak hour of the multiplier threshold roughly agrees with that of the high price multipliers (i.e., 1.4x to 1.6x), but they are not perfectly coincided. The reason is that the multiplier threshold is the result of not only price multiplier distribution but also passengers' demand elasticity. For example, from Fig. 11 the peak hours of leaving business area on weekdays are [7am, 9am], [2pm, 3pm], [5pm, 7pm] and [9pm, 11pm]. During [5pm, 7pm] a passenger should stop searching as soon as the price multiplier is less than or equal to 1.6; and during other peaks the threshold is 1.4. This roughly agrees with our previous observations in §3.2.
- (2) In transportation area the multiplier threshold is lower than in other two areas. For leaving transportation area, during the daytime ([5am, 4pm]) the threshold remains very close to 1.0. This is because in transportation area there are usually enough cars waiting for passengers: transportation area, especially airport, is always located outside the city center; and passengers leaving the area often have luggage and are more likely to request for rides.
- (3) Specifically, in Fig. 11 the multiplier threshold during [5pm, 7pm] is 1.6, and as we point out in Proposition 4.1, this is the case when $G(U) \leq c$ and hence it is not necessary for the passenger to search. In other words, (EC) is the best reaction.

4.2.3 The Search Cost. We have already pointed out in §4.1 that the search cost in modelling RoD service should be interpreted as a representation of passengers' elasticity. Now with the G function and the multiplier threshold, we are able to calculate the search cost under different circumstances.

In Proposition 4.1, the multiplier threshold a is obtained by solving $G(a) = c$. In other words, if we draw two graphs, $y = G(x)$ and $y = c$, then the x -coordinate of the intersection of these two graphs is the multiplier threshold. Now to calculate the search cost back from the multiplier threshold a and the G function, we could simply use $c = G(a)$. Fig. 7 to Fig. 10 in §4.2.1 show some examples of the intersection between $y = G(x)$ and $y = c$. Fig. 14 to Fig. 16 give the search cost under different circumstances. Similar to §4.2.2, figures for weekends are not shown here because of limited space.

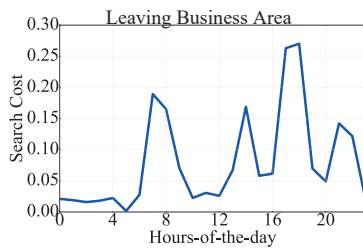


Fig. 14. The search cost: leaving business area on weekdays.

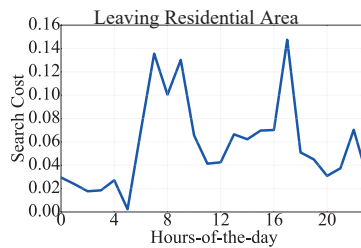


Fig. 15. The search cost: leaving residential area on weekdays.

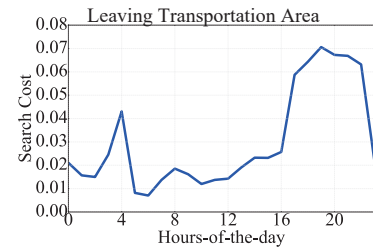


Fig. 16. The search cost: leaving transportation area on weekdays.

Regarding the search cost, we have the following observations:

- (1) Compared to the multipliers' distribution and multiplier threshold, the search cost is a representation with finer granularity, and it quantifies passengers' elasticity in requesting for rides.

- (2) The search cost suggests similar time-of-day patterns, just like those observed from multiplier threshold in §4.2.2 and from price multiplier distribution in §3.2. Moreover, as the search cost is a continuous value, it is a more accurate description than the multiplier threshold.
- (3) Comparing Fig. 14 with Fig. 15, we could see the difference between the information obtained from search cost and multiplier threshold. When the multiplier threshold is the same (e.g., 1.4x), in residential area the search cost is significantly lower. The reason is that either search cost or multiplier threshold is only one perspective relevant to the search model, and they take effects together with the price multiplier distribution.
- (4) In transportation area, the search cost is much lower than in other two functional areas, and it is higher in the evening than during the day. In fact, the peak of search cost in transportation area is only about 25% of the peak in business area. The lower search cost corresponds to our previous observations in §4.2.2 that the multiplier threshold in this area is always lower than in business area. Also, a higher search cost in the evening matches our earlier observations from Fig. 5 & 13 that both the price multiplier and multiplier threshold are higher during evening.
- (5) In business area, the search cost is the highest. Fig. 14 shows that (a) it has a similar time-of-day pattern to Fig. 5 & 11 and (b) passengers in this area have higher search costs (i.e., they are more eager to get a ride).
- (6) We could infer the relationship between the search cost and the trip intention of passengers. Here we only discuss the trip intention related to the highest search costs in each functional area, similar to our discussions in §3.2:
 - For business areas, passengers “leaving workplace during evening rush hour 5pm to 7pm” have the highest search cost.
 - For residential areas, passengers “leaving home during morning rush hours 6am to 9am” and “leaving home during 5pm to 6pm” have the highest search cost.
 - For transportation areas, passengers have lower search costs, and those “leaving transportation areas in the evening” have the highest search cost.

In addition to the learned curves of the search cost, we also perform curve fitting analysis on the search cost. Curve fitting helps us to identify time-of-day variation (peaks and troughs, and their durations) of the search cost, so that we could not only compare quantitatively the search costs under different circumstances, but also build a generative model for passengers' reaction: given the time and location, the search model could predict passengers' reaction based on the search cost and price multiplier distribution. For example, Fig. 17 gives the curve fitting of the search cost for leaving transportation area on weekdays shown in Fig. 16. This is done using two-order and piece-wise linear functions, in the following form:

$$S(t) = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3(t-2)I_{[2,4]} + \beta_4(t-4)I_{[4,5]} + \beta_5(t-6)I_{[6,8]} + \beta_6(t-8)I_{[8,10]} \\ + \beta_7(t-16)I_{[16,19]} + \beta_8(t-17)I_{[17,19]} + \beta_9(t-19)I_{[19,22]} + \beta_{10}(t-22)I_{[22,23]}, \quad (8)$$

in which $I_{[a,b]}$ is 1 when $a \leq t \leq b$, and 0 otherwise. $S(t)$ is the search cost of different hours $t \in [0, 23]$. The curve fitting in Fig. 17 gives an R^2 of 0.9763 and an adjusted R^2 of 0.9546. β_0, β_1 and β_2 characterize the overall shape of the fitted function, and other coefficients characterize the extent of fluctuations of the search cost in different hours.

5 RELATED WORK

There are a number of previous studies on RoD services. Among them, most are about the traditional taxi service that uses fixed prices, and the rest of them are about emerging RoD services using dynamic

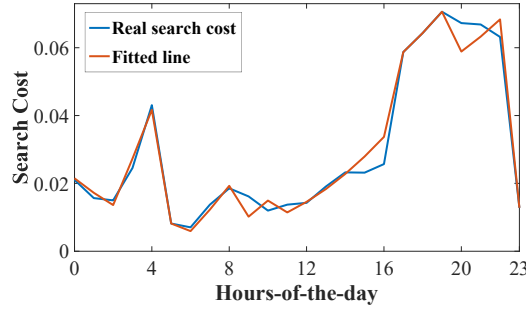


Fig. 17. Curve fitting on the search cost: leaving transportation area on weekdays.

prices such as Uber, Didi or Shenzhou UCar. Currently, there is not any detailed study about passengers' reaction to dynamic prices in emerging RoD services.

5.1 Fixed Prices: Taxi Service

As a traditional service, taxi has already been studied extensively, on the dispatching [12], scheduling [1], pricing [7], driver behavior [13, 15, 30], etc. As to the modelling of the service, there are also a series of studies using a network model to analyze different perspectives of taxi services (e.g., [23, 24]).

Regarding the data used in the study, the taxi trip data (i.e., data about orders) and taxi trajectory data have been used in previous works. An important public dataset of taxi trip data is the NYC trip record data, made available by the NYC Taxi & Limousine Commission. For example, [27] uses this dataset to estimate the travel time between urban places. The taxi trajectory data, on the other hand, is mostly based on GPS records of cars, and has also been studied extensively to mine useful information, in recommending areas for passenger searching [26], identifying efficient taxi service [14], detecting detour in taxi rides [28], probing traffic conditions [3], identifying flaws in city planing [31], predicting traffic speed and exploring congestion sources [22], etc.

Taxi ride-sharing is an important variant of the traditional taxi service, enabling different passengers to share a taxi. In other words, a single taxi is not fully occupied by a ride request, but by several requests that have common road segments. This service has also been a heated topic in recent years, and there are studies on the fare model [29], routing algorithm [11], empirical evaluation [17], large-scale ride-sharing service [16], etc.

However, as in most cases the pricing of taxi service is fixed, studies related to changing the price dynamically based on the supply & demand condition are not prevalent.

5.2 Dynamic Prices: Emerging RoD services

Different from taxi service, most studies on emerging RoD services are centered on its dynamic pricing mechanism. Related topics include measuring the new services, analyzing the effects of dynamic pricing on either supply or demand, measuring passengers' reaction, etc.

For the measurement studies, the authors in [4] conduct the first in-depth investigation of Uber. They try to analyze and evaluate Uber's surge pricing mechanism by treating Uber as a black-box and emulating copies of Uber users in different locations of a city. They also attempt to use the results to predict the prices in the coming minutes, but don't obtain a satisfactory result due to the lack of real and accurate Uber data. In [9] we analyze the demand pattern and the effect of dynamic pricing, using real data from

Shenzhen UCar, one of the largest RoD service providers in China. We also use clustering technique to extract the distribution of functional areas as well as the dynamic pricing areas.

Dynamic pricing also attracts attention from economic researchers. For example, [10] uses a case study to evaluate the effects of Uber's surge pricing, showing that surge pricing could provide better service to users. [19] studies the ridesourcing's usage and impacts by collecting surveys from San Francisco. [2] uses a queueing-theoretic approach to compare the effects of dynamic and static pricing. A recent study [5] discusses the supply elasticity and flexible work in Uber, and shows that drivers could schedule their work more flexibly. Also, [6] tries to estimate the consumer surplus in four main cities of Uber's service, trying to identify consumer's willingness to pay when requesting for a ride.

For passengers' reaction to dynamic prices, there is not any detailed study. As far as we know, the only one is our primitive observation of passengers' reaction to dynamic prices in [8]. But this work is not systematic enough and lacks mathematical modelling of passengers' reaction.

6 CONCLUSION AND FUTURE WORK

In ride-on-demand (RoD) service a passenger uses mobile-app to request for rides, and always uses the app to estimate the fare of the intended trip for multiple times before creating an order. In this paper we focus on passengers' reaction in RoD service – *multiple fare estimations of intended trips*. In the analysis of data from a real RoD service provider in China, we present our observations on the price multipliers passengers receive and the patterns of passengers' reaction. The data analysis indicate that both the price multipliers and the passengers' demand elasticity (i.e., *how eager one needs a ride?*) influence their reaction. A model is then adopted and extended from a previous search theory model, as an attempt to explain how many times a passenger should estimate the trip fare. The result of the model is that a passenger should stop estimating as soon as the price multiplier falls below a threshold. Based on the data and the model, we perform numerical analysis on the multiplier threshold as well as the search cost and show relevant insights. The model could be used in user-behavior model learning as well as prediction.

One future work is to extend the model to handle *drop-out*, to explain passengers' reaction that finally gives up and no order is created. "Drop-out" means that one chooses to stop estimating fare after multiple estimations, either because s/he believes that the price would not go down, or because the search cost is so high that no searching at all is the best reaction. Studying the frequency and reason of drop-out enables researchers and service providers to propose novel solutions to attract more passengers.

Another future work is to express the search cost as a function of multiple features, including the type of functional areas, passengers' demographic properties, weather condition, and etc. In this paper we already use curve fitting technique to express the search cost as a combination of two-order function and several piece-wise linear functions, to represent the time-of-day variation (i.e., peak and troughs and corresponding durations) of the search cost, in an effort to build generative model of passengers' reaction. Expressing it with multiple features would further clarify how different factors, in addition to the time-of-day factor, influence passengers' demand elasticity in RoD service.

ACKNOWLEDGMENTS

This work is supported by National Key R&D Program of China (No. 2017YFB1002000), the National Science Foundation of China (No. 61602067), and the Fundamental Research Funds for the Central Universities (No. 106112017cdjxy180001). Chao Chen is the corresponding author for this paper.

REFERENCES

- [1] Ruibin Bai, Jiawei Li, Jason AD Atkin, and Graham Kendall. 2013. A Novel Approach to Independent Taxi Scheduling Problem Based on Stable Matching. *Journal of the Operational Research Society* 65, 10 (2013), 1501–1510.

- [2] Siddhartha Banerjee, Ramesh Johari, and Carlos Riquelme. 2015. Pricing in Ride-Sharing Platforms: A Queueing-Theoretic Approach. In *Proceedings of the 2015 ACM Conference on Economics and Computation (EC '15)*. ACM, New York, NY, USA, 639–639.
- [3] P.S. Castro, Daqing Zhang, and Shijian Li. 2012. Urban Traffic Modelling and Prediction Using Large Scale Taxi GPS Traces. In *Proceedings of the 10th International Conference on Pervasive Computing*. 57–72.
- [4] Le Chen, Alan Mislove, and Christo Wilson. 2015. Peeking Beneath the Hood of Uber. In *Proceedings of the 2015 ACM Conference on Internet Measurement Conference (IMC '15)*. ACM, New York, NY, USA, 495–508.
- [5] M. Keith Chen. 2016. Dynamic Pricing in a Labor Market: Surge Pricing and Flexible Work on the Uber Platform. In *Proceedings of the 2016 ACM Conference on Economics and Computation (EC '16)*. ACM, New York, NY, USA, 455–455.
- [6] Peter Cohen, Robert Hahn, Jonathan Hall, Steven Levitt, and Robert Metcalfe. 2016. Using Big Data to Estimate Consumer Surplus: The Case of Uber. (2016). Retrieved May 8, 2017 from <http://bit.ly/2pqXiWo>
- [7] Jiarui Gan, Bo An, Haizhong Wang, Xiaoming Sun, and Zhongzhi Shi. 2013. Optimal Pricing for Improving Efficiency of Taxi Systems.. In *Proceedings of the 2013 International Joint Conferences on Artificial Intelligence (IJCAI '13)*. AAAI Press, 2811–2818.
- [8] Suiming Guo, Yaxiao Liu, Ke Xu, and Dah Ming Chiu. 2017. Understanding Passenger Reaction to Dynamic Prices in Ride-on-demand Service. In *Pervasive Computing and Communication Workshops (PerCom Workshops), 2017 IEEE International Conference on*. IEEE, 42–45.
- [9] Suiming Guo, Yaxiao Liu, Ke Xu, and Dah Ming Chiu. 2017. Understanding Ride-on-demand Service: Demand and Dynamic Pricing. In *Pervasive Computing and Communication Workshops (PerCom Workshops), 2017 IEEE International Conference on*. IEEE, 509–514.
- [10] Jonathan Hall, Cory Kendrick, and Chris Nosko. 2015. The effects of Uber’s surge pricing: a case study. (Oct. 2015). Retrieved Feb 10, 2017 from <http://bit.ly/2kayk9O>
- [11] Wen He, Kai Hwang, and Deyi Li. 2014. Intelligent Carpool Routing for Urban Ridesharing by Mining GPS Trajectories. *IEEE Transactions on Intelligent Transportation Systems* 15, 5 (2014), 2286–2296.
- [12] Jaeyoung Jung, R Jayakrishnan, and Ji Young Park. 2013. Design and Modeling of Real-time Shared-taxi Dispatch Algorithms. In *Transportation Research Board 92nd Annual Meeting Compendium of Papers. Report*.
- [13] Biao Leng, Heng Du, Jianyuan Wang, Li Li, and Zhang Xiong. 2016. Analysis of Taxi Drivers’ Behaviors Within a Battle Between Two Taxi Apps. *IEEE Transactions on Intelligent Transportation Systems* 17, 1 (2016), 296–300.
- [14] Bin Li, Daqing Zhang, Lin Sun, Chao Chen, Shijian Li, Guande Qi, and Qiang Yang. 2011. Hunting or Waiting? Discovering Passenger-finding Strategies from a Large-scale Real-world Taxi Dataset. In *Pervasive Computing and Communication Workshops (PerCom Workshops), 2011 IEEE International Conference on*. IEEE, 63–68.
- [15] Liang Liu, Clio Andris, and Carlo Ratti. 2010. Uncovering Cabdrivers’ Behavior Patterns from Their Digital Traces. *Computers, Environment and Urban Systems* 34, 6 (2010), 541–548.
- [16] Shuo Ma, Yu Zheng, and Ouri Wolfson. 2015. Real-Time City-Scale Taxi Ridesharing. *IEEE Transactions on Knowledge and Data Engineering* 27, 7 (2015), 1782–1795.
- [17] Pedro M.d’Orey, Ricardo Fernandes, and Michel Ferreira. 2012. Empirical Evaluation of a Dynamic and Distributed Taxi-sharing System. In *Intelligent Transportation Systems (ITSC), 2012 15th International IEEE Conference on*. 140–146.
- [18] Stuart P.Lloyd. 1982. Least Squares Quantization in PCM. *IEEE Transactions on Information Theory* 28, 2 (1982), 129–137.
- [19] Lisa Rayle, Susan Shaheen, Nelson Chan, Danielle Dai, and Robert Cervero. 2014. App-Based, On-Demand Ride Services: Comparing Taxi and Ridesourcing Trips and User Characteristics in San Francisco. (2014). Retrieved May 8, 2017 from <http://bit.ly/2kVkahg>
- [20] Shenzhen UCar. 2015. Annual results for the year ended 31 Dec 2015. (2015). <http://bit.ly/2cFdL6U>
- [21] Lester G Telser. 1973. Searching for the Lowest Price. *The American Economic Review* 63, 2 (1973), 40–49.
- [22] Jingyuan Wang, Qian Gu, Junjie Wu, Guannan Liu, and Zhang Xiong. 2016. Traffic Speed Prediction and Congestion Source Exploration: A Deep Learning Method. In *Proceedings of the 16th IEEE International Conference on Data Mining (ICDM '16)*. IEEE, 499–508.
- [23] KI Wong, SC Wong, and Hai Yang. 2001. Modeling Urban Taxi Services in Congested Road Networks with Elastic Demand. *Transportation Research Part B: Methodological* 35, 9 (2001), 819–842.
- [24] Hai Yang, Yan Wing Lau, SC Wong, and Hong Kam Lo. 2000. A Macroscopic Taxi Model for Passenger Demand, Taxi Utilization and Level of Services. *Transportation* 27, 3 (2000), 317–340.
- [25] Nicholas Jing Yuan, Yu Zheng, Xing Xie, Yingzi Wang, Kai Zheng, and Hui Xiong. 2015. Discovering Urban Functional Zones Using Latent Activity Trajectories. *IEEE Transactions on Knowledge and Data Engineering* 27, 3 (2015),

- 712–725.
- [26] Nicholas Jing Yuan, Yu Zheng, Liuhan Zhang, and Xing Xie. 2013. T-Finder: A Recommender System for Finding Passengers and Vacant Taxis. *IEEE Transactions on Knowledge and Data Engineering* 25, 10 (2013), 2390–2403.
- [27] Xianyuan Zhan, Samiul Hasan, Satish V Ukkusuri, and Camille Kamga. 2013. Urban Link Travel Time Estimation Using Large-scale Taxi Data with Partial Information. *Transportation Research Part C: Emerging Technologies* 33 (2013), 37–49.
- [28] Daqing Zhang, Nan Li, Zhi-hua Zhou, Chao Chen, Lin Sun, and Shijian Li. 2011. iBAT: detecting anomalous taxi trajectories from GPS traces. In *Proceedings of Ubicomp '11*. 99–108.
- [29] Desheng Zhang, Ye Li, Fan Zhang, Mingming Lu, Yunhuai Liu, and Tian He. 2013. coRide: Carpool Service with a Win-win Fare Model for Large-scale Taxicab Networks. In *Proceedings of the 11th ACM Conference on Embedded Networked Sensor Systems (SenSys '13)*. ACM, Article 9, 14 pages.
- [30] Fuzheng Zhang, David Wilkie, Yu Zheng, and Xing Xie. 2013. Sensing the Pulse of Urban Refueling Behavior. In *Proceedings of Ubicomp '13*. 13–22.
- [31] Yu Zheng, Yanchi Liu, Jing Yuan, and Xing Xie. 2011. Urban computing with taxicabs. In *Proceedings of Ubicomp '11*. 89–98.

Received May 2017; revised XXX 2017; accepted XXX 2017