https://dx.doi.org/10.1016/j.trc.2020.102922

Data-driven analysis on matching probability, routing distance and detour distance in ride-pooling services

Jintao Ke[a], Zhengfei Zheng[b,*], Hai Yang[b], Jieping Ye[c]

[a]*Department of Logistics and Maritime Studies, the Hong Kong Polytechnic University, Hong Kong, China*
[b]*Department of Civil and Environmental Engineering, the Hong Kong University of Science and Technology, Hong Kong, China*
[c]*Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, United States*

Abstract

By serving two or more passenger requests in each ride in ride-sourcing markets, ride-pooling service is now becoming an important component of shared smart mobility. It is generally expected to improve vehicle utilization rate, and therefore alleviate traffic congestion and reduce carbon dioxide emissions. A few recent theoretical studies are conducted, mainly focusing on the equilibrium analysis of the ride-sourcing markets with ride-pooling services and the impacts of ride-pooling services on transit ridership and traffic congestion. In these studies, there are three key measures that distinguish ride-pooling service analysis from the non-pooling ride-sourcing market analysis. The first is the proportion of passengers who are pool-matched(referred to as pool-matching probability), the second is passengers' average detour distance, and the third is average vehicle routing distance to pick up and drop off all passengers with different origins and destinations in one specific ride. These three measures are determined by passenger demand for ride-pooling and matching strategies. However, due to the complex nature of ride-resourcing market, it is difficult to analytically determine the relationships between these measures and passenger demand. To fill this research gap, this paper attempts to empirically ascertain these relationships through extensive experiments based on the actual on-demand mobility data obtained from Chengdu, Haikou, and Manhattan. We are surprised to find that the relationships between the three measures (pool-matching probability, passengers' average detour distance, average vehicle routing distance) and number of passengers in the matching pool (which reflects passenger demand) can be fitted by some simple curves (with fairly high goodness-of-fit) or there exist elegant empirical laws on these relationships. Our findings are insightful and useful to theoretical modeling and applications in ride-resourcing markets, such as evaluation of the impacts of ride-pooling on transit usage and traffic congestion.

Keywords: ride-sourcing, ride-pooling, detour distance, routing distance, matching probability

1. Introduction

With the increasing pervasiveness of mobile devices, on-demand ride-sourcing services provided by transportation network companies (TNCs) such as Uber, Lyft and Didi, have experienced explosive growth since its emergence in 2009. It is reported that Uber has served cumulative more than 10 billion passenger requests in over 80 countries (Iqbal, 2020) while Didi Chuxing is now serving over 30 million rides per day in China (Smith, 2020). Due to 24-hour-a-day availability and capacity to serve door-to-door on-demand requests, ride-sourcing services have played a vital role of urban transportation and largely revolutionized the way people travel. Most transportation

*Corresponding author: zhengfei.zheng@connect.ust.hk

network companies provide both non-pooling (each passenger/request is served with one vehicle) and ride-pooling (two or more passengers/requests with similar itinerary and schedule are served with one vehicle) services. The ride-pooling service is also referred to non-shared or ride-splitting service in some recent literature (Shaheen et al., 2016, Chen et al., 2017, Li et al., 2019, Wang and Yang, 2019); typical examples are UberPool, Didi Express Pool and Lyft Share. Ride-pooling service is expected to improve vehicle utilization rate and achieve societally beneficial objectives such as traffic congestion alleviation and emission reductions (Li et al., 2016, Alexander and González, 2015), and are generally favored by both ride-sourcing platforms (increasing vehicle utilization rate) and traffic regulators (benefiting society as a whole). The platform has opportunity to increase vehicle utilization rate through ride-pooling, so as to attract more passengers by reducing the waiting time. Although the platform generally offers to discounted trip fare to passengers opting for ride-pooling service, the gains from the increased passenger demand may offset the loss from the reduced trip fare, and the platform can earn higher revenue as a whole. Meanwhile, the traffic regulator who cares about the social welfare may benefit from ride-pooling, which may reduce traffic congestion by vehicle sharing across passengers. Indeed, ride-pooling service is now becoming an important component in the shared mobility system. For example, it is reported that Lyft aimed to have 50 percent of rides being shared by 2022 (Schaller, 2018).

Nonetheless, there are some doubts regarding the effectiveness and efficiencies of ride-pooling service, particularly their impacts on traffic congestion. For example, Schaller (2018) claimed that the ride-pooling services have little ability to offset the traffic congestion impacts brought by non-shared ride-sourcing services. They found that, in major cities of United States, the non-pooling service add 2.8 new vehicle miles onto the road network by taking off each mile of auto-driving (indicating that ride-sourcing deteriorates traffic congestion). The implementation of ride-pooling service only brings marginal reduction on mileage increase–2.6 new vehicle miles for each mile of auto-driving taken off, which implies that ride-pooling service is unable to significantly reduce traffic congestion as expected. There are various possible reasons. First, if an inefficient pool-matching strategy is implemented, the pool-matching probability is low, such that most of the passengers opting for ride-pooling may eventually get unmatched. Second, passengers experience very long detour distance/time in ride-pooling, and thus are unwilling to choose ride-pooling service. Third, the vehicle routing distances to pick-up and drop-off passengers in ride-pooling are too long, consequently worsening traffic congestion. Last, the generally cheaper trip fare of ride-pooling service (compared to non-pooling ride-sourcing service or regular taxi trip) may encourage and intake passengers to switch their travel mode from other space-efficient modes, such as walking, biking and public transits. These negative effects warrant a need for a better understanding and efficient operations of the emerging ride-pooling service, such as pricing and matching strategies.

While most of the previous studies focus on the modelling of ride-sourcing markets with non -pooling service, only a few studies are directed towards modelling of the markets with ride-pooling service. For example, Yan et al. (2019) examined the steady-state of a market with ride-pooling service and showed that the joint optimization of dynamic pricing and matching can help improve vehicle utilization and reduce passenger waiting time. Jacob and Roet-Green (2019) studied a platform providing both non-pooling and ride-pooling services and investigated the optimal prices for each type of services. Ke et al. (2020) studied the complex relations among matching probability, passengers' average detour time, and demand for ride-pooling services in a two-sided market with matching frictions. Ke et al. (2019c) compared non-pooling and ride-pooling services and their impacts on traffic congestion in a framework of fundamental diagrams.

There are three important measures in the modelling for ride-sourcing markets with ride-pooling services: the pool-matching probability, passengers' average detour distance, drivers' average rout-

ing distance. On one hand, these three measures affect the generalized costs of passengers, passenger demand, the platform profit and social welfare. An increase of pool-matching probability indicates a reduction on the number of vehicles occupied by passengers and thus decreases passengers' waiting time, but makes more passengers endure the detour time caused by ride-pooling. A decrease of passengers' average detour distance will reduce the generalized cost of passengers and thus attract more passengers to use ride-pooling services. A decrease of drivers' average routing distance results in an increase of the idle time of drivers, which further reduces passengers' waiting time and thus increases passenger demand. On the other hand, these three measures are dependent on passenger demand. As we can imagine, if passenger demand is larger or the density of passengers in the matching pool is higher, then passengers are more easily pool-matched, which indicates a larger pool-matching probability. Meanwhile, with a larger number of passengers waiting in the pool, the platform is able to pool-match passengers with closer origins and destinations; as a result, passengers can enjoy a shorter average detour distance while drivers spend less time to pick-up and drop-off the pool-matched passengers (indicating a shorter vehicle routing distance). This indicates the three measures are endogenously interacted with passenger demand, which makes the ride-sourcing market with ride-pooling service much more complex than the market with only non-pooling service. Therefore, the modelling of ride-pooling services requires a precise understanding of the relationship between the three key measures and passenger demand, in particular, how to express the three measures as functions of passenger demand.

Unfortunately, due to the complex nature of the pool-matching process in ride-pooling service, it is very difficult to theoretically spell out the relationship among the pool-matching probability, passenger detour distances, vehicle routing distances, passenger demand and matching strategies. For analytical tractability, some simplifications and strong assumptions are adopted in the above mentioned theoretical studies. For example, Yan et al. (2019) and Ke et al. (2019c) assumed that the two pool-matched passengers are required to walk to a meeting point midway between their origins, at which they are picked-up by a nearby vehicle. This assumption avoids the estimation of the passenger detour distances and vehicle routing distances since a driver picks up and/or drops off two passengers at the same point without any extra routing distance. However, in a real ride-pooling system, the origins and/or destinations of those pool-matched trips may not be within walkable range, and thus drivers need to move from one origin to another and/or from one destination to another, which generally incur extra routing distance. Ke et al. (2020) proposed an aggregate function to describe the relationship between passengers' detour distance, pool-matching probability, vehicle routing distance, and passenger demand. However, this paper assumed an idealized detour-unconstrained scenario where the platform pool-matches all passengers in the pool no matter how long the passengers' detour distances are. Jacob and Roet-Green (2019) used a hypothesized queuing model to approximate the pool-matching probability, but did not verify their assumptions with actual trip data.

To tackle this challenge, this paper makes a first attempt to empirically examine the relationships between the three key measures (pool-matching probability, passenger detour distance and vehicle routing distance) and passenger demand in ride-pooling service. Through extensive experiments on the basis of the actual mobility data in three cities (Chengdu, Haikou, and Manhattan), we interestingly find out some empirical laws can well approximate these relationships. To be specific, (1) the ratio of passenger detour distance to average non-pooling trip distance is inversely proportional to passenger demand (defined as the number of passengers in the matching pool in this paper) with a positive intercept; (2) the ratio of driver detour distance to average non-pooling trip distance is inversely proportional to the logarithm of passenger demand with a multiplier; (3) the pool-matching probability exhibits a negative-exponential saturation curve with respect to

passenger demand. Surprisingly, although these simple empirical laws contain only with one or two parameters, they can fit the relationship between the measures and passenger demand very well, with fairly high goodness-of-fit. In addition, the empirical laws adapt to different settings of matching strategies (reflected by the matching radii).

It is worth noting that these empirical laws do not have theoretical foundations. Nevertheless, due to their simple forms and high goodness-of-fit, they can be used in the modelling of ride-sourcing markets with ride-pooling service, and help researchers better understand the complex interactions between the three key measures and passenger demand for ride-pooling service. Furthermore, by using the formula expressing vehicle routing distance as a function of passenger demand, we can quantify the aggregate effects of ride-pooling on vehicle-miles-traveled as well as traffic congestion, under different passenger demand level and different matching strategies. In all, these findings will have far-reaching fundamental implications on the modelling, optimization and analyses of the ride-pooling service.

The remainder of the paper is organized as follows. Section 2 reviews the related literature, from modelling and optimization to data mining in on-demand ride-sourcing markets. Section 3 describes the experimental settings and the mobility data used for experiments. Section 4 presents the major findings, i.e. the empirical laws characterizing how the three measures vary with the passenger demand. Section 5 summarizes the potential implications of the discovered empirical laws and outlook future studies.

2. LITERATURE REVIEW

In recent years, on-demand ride-sourcing services have become increasingly popular and greatly reshaped the way we travel. There is a growing body of research focused on the analysis and modeling of equilibrium of the ride-sourcing markets. For example, Zha et al. (2016) examine the optimal pricing strategies in a monopoly market, competitive market and duopoly market. Zha et al. (2018a) and Bai et al. (2019) investigate surge pricing and found that the platform should charge passengers higher prices as demand increases. Castillo et al. (2018) observe a market failure called "wild goose chase", which happens when only very few idle drivers are available to spend substantial time on picking up far away passengers. They find that the wild goose chase can be avoided by using surge pricing to suppress the high demand. In addition to time-based pricing, Zha et al. (2018b) and Bimpikis et al. (2019) examine the use of spatial pricing strategies to balance supply and demand across space. Many other issues pertaining to the ride-sourcing markets are examined, including government regulations (Yu et al., 2019), modelling of non-equilibrium markets (Nourinejad and Ramezani, 2020), effects of parking provisions (Xu et al., 2017), the supply curves of ride-hailing markets (Xu et al., 2020), working and recharging schedules of electric vehicle drivers for ride sourcing services (Ke et al., 2019a), impacts of ride-sourcing services on transit usage and congestion (Ke et al., 2019c, Zhu et al., 2020), order cancellation behaviors of customers (He et al., 2018, Wang et al., 2020), and optimization of matching time interval and matching radius (Yang et al., 2020, Ke et al., 2019b). Some of these studies are in the spirits of earlier studies on traditional taxi markets (Arnott, 1996, Yang et al., 2005, 2010, Yang and Yang, 2011).

As a special type of ride-sourcing services, ride-pooling emerges in various online platforms (Didi, Uber, Lyft, etc.) and has drawn attentions from researchers. Ke et al. (2020) study the equilibrium properties of a ride-pooling market, in which the platform charges an up-font discounted trip fare from each passenger opting for ride-pooling, who may end up with either matched or unmatched. A matched passenger experiences an extra detour time for picking up and dropping off a

second passenger, while an unmatched passenger does not encounter extra inconvenience. Therefore, the average trip time of passengers (an important component in passengers' generalized cost that affects passengers' willingness to opt for ride-pooling) depends on the pool-matching probability and average extra detour time. Conversely, the mass of passengers opting for ride-pooling will in turn affect the pool-matching probability and average detour time. To spell out such a complex system, Ke et al. (2020) propose some aggregate functions to depict the relationships between the pool-matching probability, detour time and arrival rate of passengers opting for ride-pooling. In a similar fashion, Ke et al. (2019c) examines the impacts of ride-pooling on traffic congestion and the time costs of different users including normal private car drivers. They find that, replacing normal non-shared services with ride-pooling services leads to a win-win situation which make both passengers and private car users better off. As aforementioned, to simplify the models, Yan et al. (2019) proposes a "dynamic waiting" mechanism that matches two passengers whose origins are within a walkable range. In this setting, passengers do not experience additional detour time and vehicle routing distance is roughly the same as that in non-shared services. However, none of these preliminary studies provides explicit formulations to characterize how pool-matching probability, passenger detour distance and vehicle routing distance change with passenger demand and other factors, in a general case.

Moreover, the tremendous data generated by on-demand rides facilitate some data-driven research works, which help understand passengers' and drivers' behaviors and guide operational strategy designs. For example, Chen et al. (2017) develop an ensemble learning approach to investigate the key factors that affect passengers' choices between non-ride-pooling and ride-pooling services. By analyzing the ride-sourcing data published by Didi in Chengdu, China, Li et al. (2019) find that the proportion of passengers using ride-pooling services is still low (6-7%) due to long extra detour and degraded travel time reliability. Sun et al. (2019) propose an econometric framework to estimate the elasticity of labor supply in ride-sourcing markets, and find that both the participation and working-hour elasticities of labor supply are positive and significant. Using the trajectory data of the mobility trips in Didi Chuxing, Yang et al. (2018) discover a universal law of detour ratios which states that the detour ratio (the ratio of average actual road distance to straight-line distance) is a constant plus a term inversely proportional to the straight-line distance. Recently, with the rapid development of artificial intelligence, some advanced machine learning and reinforcement learning based approaches are proposed to address some important research issues related to ride-sourcing systems, including real-time supply demand prediction (Ke et al., 2017, 2018, Yao et al., 2018, Geng et al., 2019), order dispatching (Xu et al., 2018, Wang et al., 2018), and idle vehicle allocation (Lin et al., 2018), etc. To the best of our knowledge, this paper is the first empirical study to explore the relationships among pool-matching probability, detour/routing distance, and passenger demand in ride-pooling markets.

3. Optimization framework and data descriptions

In this section, we first give definitions to the three key measures, i.e. pool-matching probability, passenger detour distance and vehicle routing distance, in batch matching. Then we design an optimization algorithm to pool-match passengers with two distinctive objectives, subject to constraints in terms of matching radius. Finally, we give a brief introduction to the actual mobility data in three cities as well as the details of the experimental settings.

A few points are worth mentioning here. First, our optimization framework considers the pool-matching between passengers, instead of the matching between passengers and drivers (such as the model in Yang et al. (2020)). Second, the optimization framework considers an instant

with a number of available/waiting passengers, and pool-matches ride-pooling passengers with similar origins and destinations. The paired or unpaired passengers can be further matched to an available nearby driver shortly afterwards. Therefore, at any instant, the three key measures (as the outcomes of the pool-matching) depend on the number of waiting passengers (reflecting passenger demand or density of passengers) and the matching strategy in terms of the matching radius that restrains the pool-matching between passengers with distant origins and destinations. Third, this optimization model can be seamlessly integrated into an online pool-matching strategy. For example, the platform can execute this optimization model repeatedly at the end of each time interval as passengers arrive continuously. In our experiments, we investigate the trends of the three key measures with the increase of the number of passengers in the matching pool. As expected, the empirical laws obtained from our experiments can also be adapted to a market with online pool-matching strategy for ride-pooling service, as long as we look into a stationary equilibrium state with steady number of waiting passengers opting for ride-pooling service at any given instant. Fourth, we show that the empirical laws are robust with respect to the distinct objective functions. Fifth, we do not consider the effects of the level of supply or the number of drivers, but only focus on the pool-matching between passengers and simply assume the supply is sufficient to serve the paired and unpaired passengers after the ride-pooling. For convenience, Table 1 provides a summary of notation.

Table 1: List of main notation

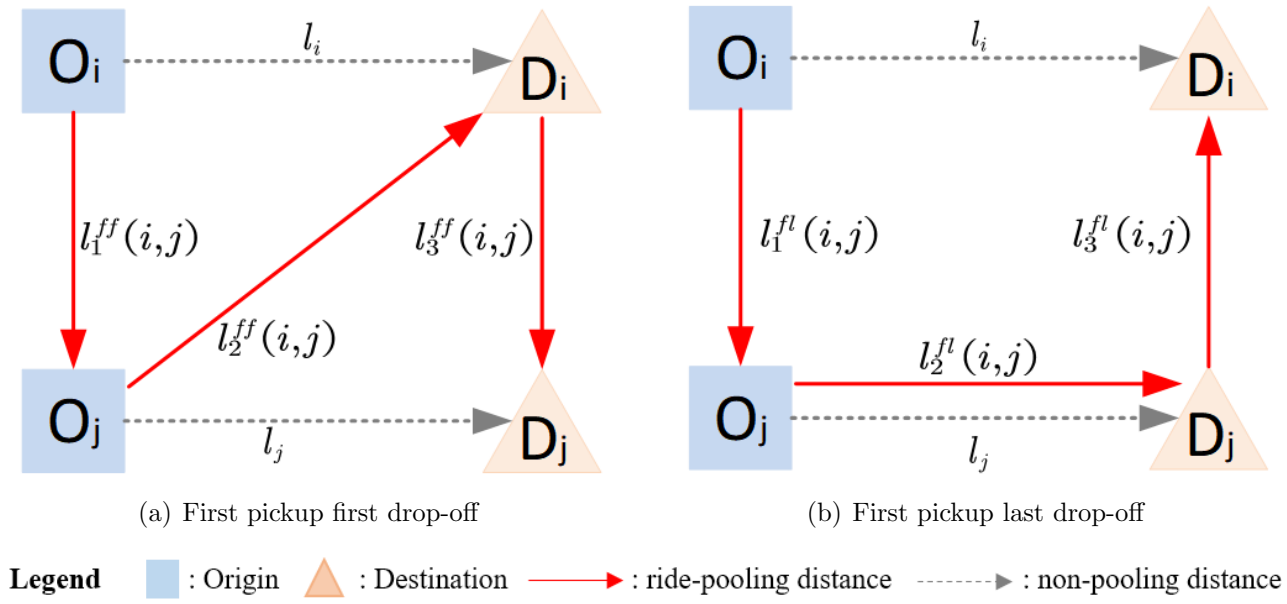
Symbol	Description
N	Passengers(interchangeable with trips) in the pool to be matched per time unit
p	Pool-matching probability, the proportion of successful pool-matched passengers among all passengers
$\bar{l}$	Average of non-pooling trip distance
$\Delta\bar{l}$	Average detour distance experienced by the successfully pool-matched passengers
$\bar{l_d}$	Average vehicle routing distance for serving each two pool-matched passengers
l_i	Distance of non-pooling trip i
$l_{min}\,(i,j)$	Minimum vehicle routing distance of four possible pickup drop-off sequences between trip i and trip j
$\Delta l_i\,(i,j)$	Detour distance experienced by passengers i associated with the minimum vehicle routing distance $l_{min}\,(i,j)$
$\Delta l_j\,(i,j)$	Detour distance experienced by passengers j associated with the minimum vehicle routing distance $l_{min}\,(i,j)$
$l^O\,(i,j)$	Distance between origins of trip i and trip j
$l^D\,(i,j)$	Distance between destinations of trip i and trip j
x_{ij}	Binary variable to indicate if trip i and trip j are matched

3.1. Definitions of the key measures

Let p denote the pool-matching probability, the proportion of successful pool-matched passengers among all passengers, let $\bar{l}$ denote the average of non-shared trip distances, let $\Delta\bar{l}$ denote the average of detour distance experienced by the successfully pool-matched passengers, and let denote $\bar{l_d}$ the average routing distance of drivers for serving each two pool-matched passengers[1]. As

[1]For simplicity, we assume that each vehicle can at most serve two passengers on each ride.

illustrated in Figure 1, suppose the vehicle first pickups passenger i (corresponding to trip i with a non-pooling trip length l_i) and then deviates to pickup passenger j (corresponding to trip j with a non-shared trip length l_j). Then the vehicle has two sequences to serve these two passengers: $O_i \to O_j \to D_i \to D_i$ (first pickup first drop-off sequence, Figure 1(a)) and $O_i \to O_j \to D_j \to D_i$ (first pickup last drop-off sequence, Figure 1(b)). For the former, the detour distance experienced by passenger i and j are $l_1^{fl}(i,j) + l_2^{fl}(i,j) + l_3^{fl}(i,j) - l_i$ and $l_2^{fl}(i,j) - l_j$, respectively, with the superscript fl referring to first pickup last drop-off sequence. For the latter, the detour distance experienced by passenger i and j are $l_1^{ff}(i,j) + l_2^{ff}(i,j) - l_i$ and $l_2^{ff}(i,j) + l_3^{ff}(i,j) - l_j$, respectively, with the superscript ff referring to first pickup first drop-off sequence. The routing distance of drivers for the two sequences are $l_1^{fl}(i,j) + l_2^{fl}(i,j) + l_3^{fl}(i,j)$ and$l_1^{ff}(i,j) + l_2^{ff}(i,j) + l_3^{ff}(i,j)$ respectively. Here the routing distance of driver only begins when the pickup of the first pool-matched passenger and ends when the drop-off or of the second pool-matched passenger. By taking the average of the detour distance experienced by passengers and the routing distance of drivers in all successfully paired rides (across one day, one month or even one year), we can obtain the average passenger detour distance $\Delta\bar{l}$ and average drivers routing distance $\bar{l}_d$ in one shared ride. Note that the trip length between any two nodes (e.g.$l_i, l_j, l_1^{fl}(i,j)$, $l_2^{ff}(i,j)$, $\cdots$) is the minimum route distance along the shortest path in the road networks.



(a) First pickup first drop-off

(b) First pickup last drop-off

Figure 1: Two en-route pool-matching sequences

3.2. Optimization algorithms

Suppose the matching pool has N accumulated passengers (i.e. passenger demand in each batch matching) with heterogeneous origins and destinations: (O_i, D_i), $i = 1, 2, \cdots, N$. For any two passengers, $i, j \in \{1, 2, \cdots, N\}$, there are four types of sequences to serve them: $O_i \to O_j \to D_i \to D_j$, $O_i \to O_j \to D_j \to D_i$, $O_j \to O_i \to D_j \to D_i$, and $O_j \to O_i \to D_i \to D_j$. Let $l_{min}(i,j)$ denote the minimum vehicle routing distance of these four sequences. As mentioned above, l_i, l_j refer to the direct travel distance from $O_i \to D_i$ and $O_j \to D_j$, respectively, in normal ride-sourcing service without ride-pooling. Let $\Delta l_i(i,j)$,$\Delta l_j(i,j)$ denote the detour distance experienced by passengers i and j in the sequence associated with the minimum vehicle routing distance $l_{min}(i,j)$.

Let $l^O(i,j)$ and $l^D(i,j)$ denote the travel distance between the origins of passenger i and j, and the travel distance between the destinations of passenger i and j, respectively.

At the decision point of a batch matching, the platform executes an optimal matching with an objective to pool-match as many passengers as possible through a minimum overall vehicle routing distance, subject to a constraint that the origins and destinations of each pair of pool-matched passengers are within a matching radius R (a decision variable). Let x_{ij} be the binary decision variable such that: if passengers i, j are pool-matched, then $x_{ij} = 1$, otherwise, $x_{ij} = 0$. Notice that pool-matching passenger i with passenger j is the same as pool-matching passenger j with passenger i, thus $x_{ij} = x_{ji}$, indicating that the matrix of decision variables is symmetric. To reduce computational burden when solving this optimization problem, we only need to consider the decision variables on the upper right of the leading diagonal of the matrix, i.e. x_{ij} where $i > j$. Now we offer a demo to readily illustrate the meanings of the constraints, Table 2 for decision variables and Table 3 for costs in the objective. Taking $k = 2$ as an example, Constraint (2) forces the summation of decision variables in the shadow area to be smaller than or equal to one, which implicitly indicates that Passenger 2 can at most be pool-matched with one other passenger from $\{1, 3, ..., N\}$. Notice that the lower left part of the decision table (namely, for all $i < j$ and $i = j$) is intentionally set to zero, as indicated in Constraint (4), which would not cause hindrances to the optimal solution. By doing this, we can avoid (1) duplicated pool matching between passengers, (2) self-pool-matching (i.e. pool matching a passenger with himself/herself). Then the optimization problem for pool-matching is given by

$$(P1) \quad \max_{x_{ij}} \sum_{i=1}^{N} \sum_{j=i}^{N} \left[V - \ l_{min}(i,j)\right] x_{ij} \tag{1}$$

$$\text{s.t.} \quad \begin{cases} \sum_{i=k}^{N} x_{ki} + \sum_{j=1}^{k-1} x_{jk} \leq 1, \, \forall k \in \{2, \cdots, N\} \\ \sum_{i=1}^{N} x_{1i} \leq 1 \\ x \in \{0, \ 1\}, \ \forall \ i, j \ \in \{1, 2, \cdots, N\} \end{cases} \quad \textit{and} \tag{2}$$

$$\max\left\{l^O(i,j), l^D(i,j)\right\} x_{ij} \leq R, \ \forall i, j \in \{1, 2, \cdots, N\} \tag{3}$$

$$x_{ij} = 0, \forall i \leq j, \quad i, j \in \{1, 2, \cdots, N\} \tag{4}$$

where V indicates the benefit of pool-matching one pair of passengers. Note that we set V to be a value that is larger than all $l_{min}(i,j)$, to let the algorithm pool-matches as many passenger requests as possible. Clearly, Objective (1) seeks to maximize the pairs of pool-matched passengers while minimize the overall vehicle routing distance required for all pool-matched passengers. Constraints (2) guarantees that each passenger $k, \forall k \in \{1, 2, \cdots, N\}$ can be pool-matched with at most one passenger. If the k^{th} constraint is binding, then passenger k is successfully pool-matched. Otherwise, he or she is not pool-matched. Note that, Constraint (3) ensures that the origins and destinations of the pool-matched passengers are within a matching radius; in other words, the distances between the origins (destinations) passengers i and j should be smaller than the matching radius R (also can be interpreted as the maximum allowable matching distance), if they are pool-matched. Constraints (4) states two scenarios. When $i = j$, it makes sure that

Table 2: Decision table - x_{ij}

Passenger	1	2	3	$\cdots$	j	$\cdots$	N
1	0	x_{12}	x_{13}	$\cdots$	x_{1j}	$\cdots$	x_{1N}
2	0	0	x_{23}	$\cdots$	x_{2j}	$\cdots$	x_{2N}
$\cdots$	0	0	0	$\cdots$	$\cdots$	$\cdots$	$\cdots$
i	0	0	0	$\cdots$	x_{ij}	$\cdots$	x_{iN}
$\cdots$	0	0	0	$\cdots$	$\cdots$	$\cdots$	$\cdots$
N	0	0	0	$\cdots$	0	$\cdots$	0

Taking $k = 2$ for an example, the shadow area describes the Constraint (2), which indicates among passenger $1, 3, \cdots, N$, only one can be paired with passenger 2.

Table 3: Cost table - $l_{min}(i, j)$

Passenger	1	2	$\cdots$	j	$\cdots$	N
1	0	$l_{min}(1,2)$	$\cdots$	$l_{min}(1,j)$	$\cdots$	$l_{min}(1,N)$
2	0	0	$\cdots$	$l_{min}(2,j)$	$\cdots$	$l_{min}(2,N)$
$\cdots$	0	0	$\cdots$	$\cdots$	$\cdots$	$\cdots$
i	0	0	$\cdots$	$l_{min}(i,j)$	$\cdots$	$l_{min}(i,N)$
$\cdots$	0	0	$\cdots$	$\cdots$	$\cdots$	$\cdots$
N	0	0	$\cdots$	0	$\cdots$	0

each passenger cannot be pool-matched with himself/herself, while $i < j$ means that we only need to consider half of the decision variable matrix as aforementioned. Note that this pool-matching strategy is one simple strategy that does not consider the pick-up and drop-off events en-route. This simplifies the experimental settings and helps discover the laws through extensive repeated experiments. Moreover, as pointed out by Yan et al. (2019), this type of pool-matching strategies is already used in Uber ExpressPool[2].

Additionally, to evaluate the robustness of the empirical laws, we conduct experiments with an alternative objective function that aims to pool-match as many passengers as possible while minimizing passengers' total detour distance. Mathematically, the optimization problem is formulated as follows:

$$(P2) \quad \max_{x_{ij}} \sum_{i=1}^{N} \sum_{j=i}^{N} \left[V - \Delta l_i\,(i,j) - \Delta l_j\,(i,j) \right] x_{ij} \tag{5}$$

$$\text{s.t.} \quad Eq.\ (2) \sim (4)$$

where $\Delta l_i\,(i,j) + \Delta l_j\,(i,j)$ refers to the total detour distance experienced by passengers i and j if they are pool-matched by the algorithm; V indicates the benefit of pool-matching one pair of passengers, and is set to be a value that is larger than all $\Delta l_i\,(i,j) + \Delta l_j\,(i,j)$, to let the algorithm pool-matches as many passenger requests as possible. Clearly, Objective (5) seeks to maximize the pairs of pool-matched passengers while minimize the overall detour distance of the pool-matched passengers. Same as optimization $P1$, Constraints (2) guarantees that each passenger can be pool-matched with at most one passenger, while Constraints (3) ensures that the origins and destinations of the pool-matched passengers are within a matching radius.

[2]Description to Uber ExpressPool:https://www.uber.com/us/en/ride/express-pool/

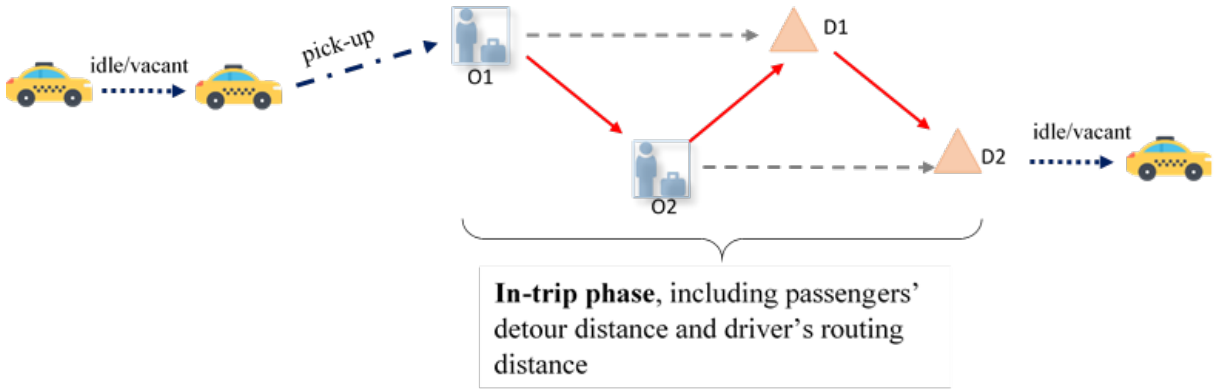


Figure 2: In-trip phase of ride-pooling

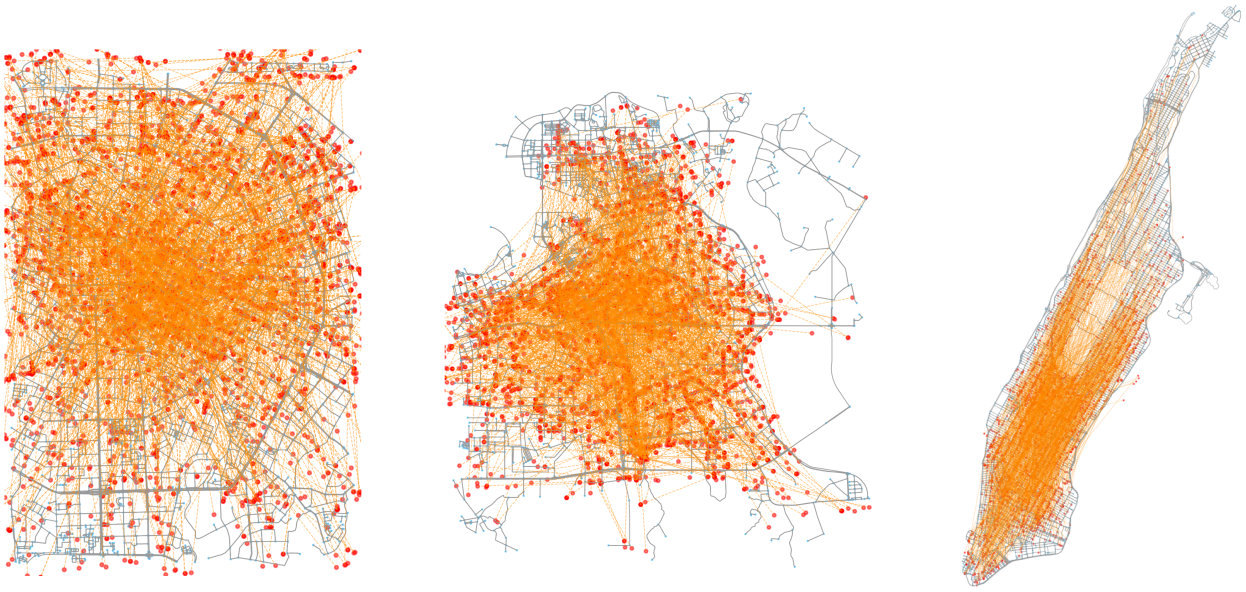
Notably, we investigate how the passenger detour distance and vehicle routing distance change with passenger demand or density, which involves the process after a driver picks up the paired passengers, i.e., the in-trip phase shown in Figure 2. Nonetheless, the pick-up distance of a vehicle dispatched to meet the first passenger of a pooled ride, which depends on the spatial density of both passengers and idles vehicles, is of practical importance. This is, however, a separate issue that has been examined in a few recent studies on ride-sourcing markets (Ke et al., 2020, Castillo et al., 2018, Yang et al., 2020). Undoubtedly, the empirical laws for detour distances identified in this study, together with the pick-up distance distribution formulas available in the literature, constitute a toolkit useful for ride-pooling studies.

3.3. Random matching without optimization objective

To verify the generality of our findings, we also run our experiments with a random pooling strategy (scenario denoted by $P3$). Namely, without using a predefined optimization objective function, two passengers are paired randomly under the condition that both their origins and their destinations are within the maximum matching radius.

3.4. Experimental settings and data description

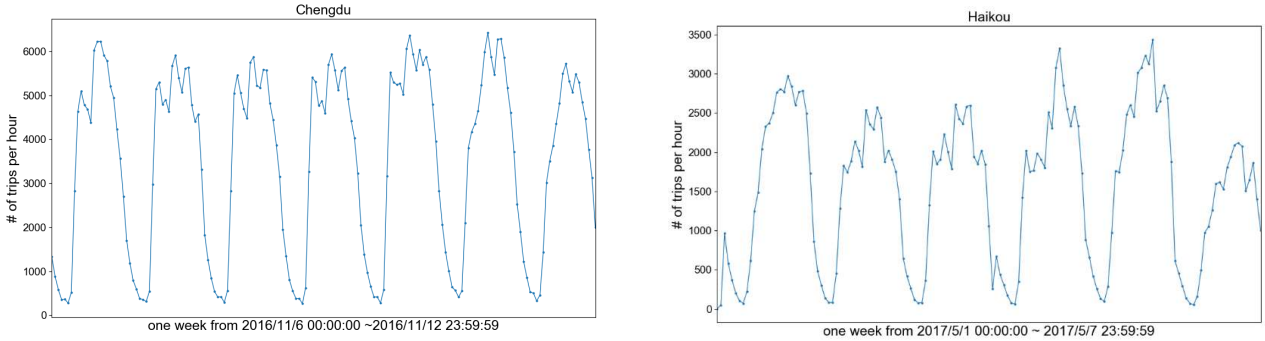
Now we examine the relationships between the three key measures (pool-matching probability, passenger detour distance and vehicle routing distance) and passenger demand N through the following sensitivity analysis. First, we set several groups of matching radii R, from $2\,km$ to $5\,km$ with a step of $1km$. Note that we do not start from $1\,km$ due to two reasons. First, according to the experience from the actual data analysis, the distance between two successful matched trips in general is larger than $1\,km$. Second, if the matching radius is set by too small value, such as $1\,km$, it would lead to an unrealistic result with limited pairs of pool-matched passengers, which cannot produce meaningful laws. For each group of matching radii, we range passengers in the pool N from 10 to 200 with a step of 10 in each experiment. Second, for each combination of R and N, we repeat the following experiment: (1) randomly sample N passenger requests from the mobility dataset of historical trip records with origins and destinations; (2) find the optimal pool-matching strategy by solving $P1$ and $P2$, under a constraint of matching radius R; (3) store the information (such as vehicle routing distance, passenger detour distance) of the pool-matched pairs of passengers in the optimal strategy. After the steps (1)-(3) are repeated by sufficient times (4,000 in our experiments), the average passenger detour distance $\Delta\bar{l}$ and average vehicle routing distance $\bar{l}_d$ can be calculated by taking the average of detour and routing distances of the pool-matched passengers in all experiments, while the pool-matching probability p is estimated by the ratio of the number of all pool-matched passengers to the number of sampled passengers.



(a) Chengdu (b) Haikou (c) Manhattan

(Note that only 1000 trips are selected in each city for illustration)

Figure 3: Spatial distribution of trip origins and destinations in different road networks [3]



(a) Chengdu (b) Haikou

Figure 4: Hourly variation of trip generation in one week

The above mentioned experimental procedures are conducted on three dataset: on-demand ride-sourcing trip records in Chengdu and Haikou in China, which are obtained from Didi Gaia Open Data[4]; Yellow taxi trip records[5] in Manhattan, New York City (NYC) in the United States, which are retrieved from NYC Taxi and Limousine Commission. Each trip request record in the three data sets contains: the pick-up locations (origins with lat, lon), drop-off locations (destinations), the time at which the request is picked up and dropped off. The basic information of three dataset is presented in Table 4, while the corresponding network structure of the three examined cities are illustrated in Figure 3. Note that only trips with both origins and destinations located in the downtown regions of the three cities are utilized in the sampling process. It is worth mentioning

[3]Road networks from Open street map: www.openstreetmap.org

[4]Website of Didi Gaiya Open data: https://outreach.didichuxing.com/research/opendata/

[5]Website of New York Taxi Dataset: https://data.cityofnewyork.us/Transportation/

that ride-sourcing dataset in Chengdu and Haikou do not include the ride-pooling trips, and the trends of the trip requests during one week are demonstrated in Figure 4. In addition, we removed some abnormal trip records from the dataset. A trip data is excluded if it hits one of the following conditions: (a) the shortest distance between its origin and destination is less than $0.1\,km$ or larger than $50\,km$; (b) the trip duration time, from its pick-up location to drop-off location, is shorter than $10s$ or longer than $2h$ (only consider trips, of which both origins and destination are within the targeted area); (c) average speed of the trip is greater than $100\,km/s$, (the speed limits in the urban area of most China's cities are usually $30-60\,km/h$ and the speed limit in Manhattan, NYC is not higher than $50\,mph$.). For computational efficiencies, the origins and destinations of the trips are assigned to their nearest nodes in the network. Then the shortest path from one point to another (from origin i to destination j, or from origin i to origin j, etc.) can be quickly calculated by the Dijkstra algorithm. The spatial distributions of the filtered trips with their origins and destinations in the three cities are displayed in Figure 3. It can be seen that the network typologies and origin-destination distributions of the mobility trips are different across the three cities.

Table 4: Dataset description

City name	Total trips	Time period	Area
Chengdu, China	595151	2016/11/6 00:00:00 - 2016/11/12 23:59:59	$\sim$85 km^2
Haikou, China	227883	2017/5/1 00:00:00 - 2017/5/7 23:59:59	$\sim$70 km^2
Manhattan NYC, US	2827464	2016/3/2 00:00:00 - 2016/3/8 23:59:59	$\sim$59.1 km^2

4. Empirical laws

By conducting the aforementioned experiments separately on these three cities, we can obtain the curves of the three measures v.s. passenger demand, under different settings of matching radii in different cities. It is generally expected that the pool-matching probability increases but the average passenger detour distance and vehicle routing distance decreases with passenger demand. However, few of previous studies investigated the properties of these ascending or decaying curves. Our experiments by solving $P1$, $P2$ and $P3$ (as presented in the last section) discover that the average passenger detour distance decreases with passenger demand at a linear rate, the vehicle routing distance decreases with passenger demand at a logarithm rate, and the pool-matching probability exhibits a negative-exponential saturation curve with respect to passenger demand. Since the outcomes of $P1$, $P2$, $P3$ are similar, we present the results of $P1$ in this section while the results of $P2$, $P3$ are relayed to Appendix F - Appendix J and Appendix K - Appendix O, respectively, to avoid redundancy. This also indicates that the found empirical laws are robust with respect to optimization algorithms with different objective functions. The details are presented below.

4.1. Law of passenger detour distance

Figure 5 illustrates the probabilistic density distribution of (a) vehicle routing distance $\bar{l}_d$ and (b) passengers' detour distance $\Delta\bar{l}$ in the simulations with four groups of passenger demand $N = 10$, $N = 50$, $N = 100$ and $N = 150$ given the $R = 3\,km$. It can be showed that, as passenger demand increases, the mean values of both passenger detour distance and vehicle routing distance becomes smaller and the right tails become shorter (which indicates that long passenger detour and vehicle

Table 5: Empirical law - passenger detour distance given $R = 4\,km$ in Manhattan

Name	Expression	Parameter	r-square
Inverse proportional form, proposed in this paper *	$\frac{\Delta\bar{l}}{\bar{l}} = \frac{1}{\alpha_t N + \beta_t}$	$\alpha_t,\ \beta_t$	0.9525
Inverse Log	$\frac{\Delta\bar{l}}{\bar{l}} = \frac{1}{\beta_d Log(\alpha_d N) + C}$	$\beta_d,\ \alpha_d,\ C$	0.7565
Exponential	$\frac{\Delta\bar{l}}{\bar{l}} = Ae^{-\lambda_e N}$	$A,\ \lambda_e$	0.3245
Power law	$\frac{\Delta\bar{l}}{\bar{l}} = B\,(N)^{-\kappa}$	$B,\ \kappa$	0.6154

* In fact, an inverse relationship is also a power law when its exponent equals 1, because a change in one quantity results in a negative change in another.

routing distances become less). Intuitively, by the pool-matching optimization algorithm as in $P1$, a shorter detour distances can be found, if the number of passengers in the matching pool becomes larger.

It is worth noting that, as reported in the literature, human mobility and many other natural, social, and cognitive phenomena generally exhibit Levy flight characteristics and behavior following distributions of a power-law or its variants (Brockmann et al., 2006, Batty, 2008, Song et al., 2010, Liang et al., 2012), but the fundamental mechanisms underlying such behavior are yet to be fully explored. Likewise, we acknowledge that the empirical laws identified here are useful but lack rigorous theoretical foundations. After substantial trial-and-error efforts, a handful of common functions are left for further examination in our study; Table 5 lists a few examples for $R = 4\,km$ in Manhattan as an illustration, where the inverse proportional function is found to be the best to characterize the relations between passenger detour distance and demand, in terms of the *r-square* value.

Next, we look at the mean values of passengers' detour distance $\Delta\bar{l}$ with different passenger demand N as inputs. It is surprising to find that the relationship between $\Delta\bar{l}$ and N can be well fitted by the following simple equation

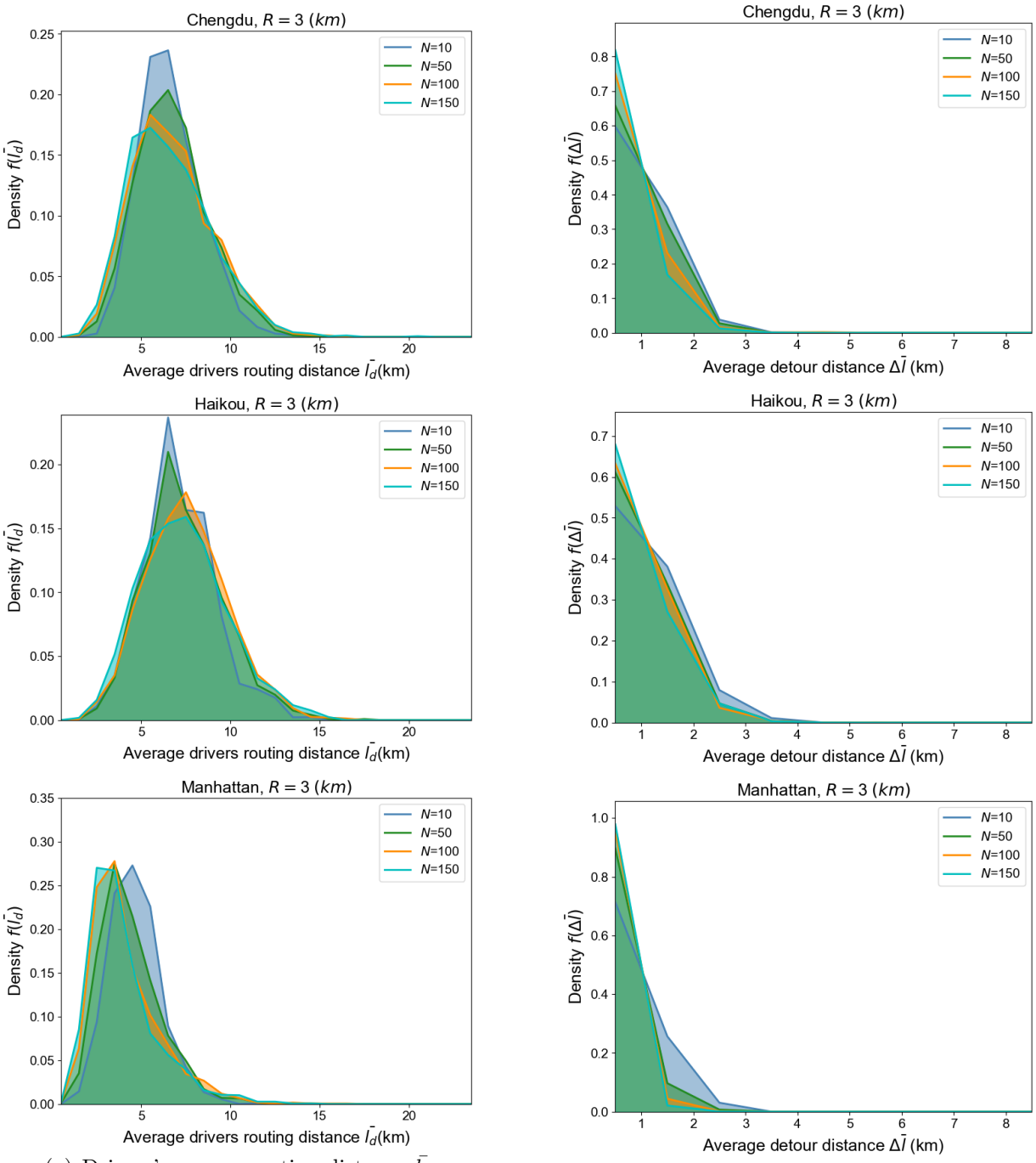
$$\frac{\Delta\bar{l}}{\bar{l}} = \frac{1}{\alpha_t N + \beta_t} \tag{6}$$

where $\bar{l}$ is the average non-pooling trip distance as aforementioned, parameter α_t has a dimension of time (hour/unit), β_t is a dimensionless parameter. Figure 5 shows the curve fitting results: $\frac{\Delta l}{\bar{l}}$ (denoted as detour distance ratio) v.s. N in (a), and $\frac{\bar{l}}{\Delta\bar{l}}$ (the reciprocal of the detour distance ratio) v.s. N in (b), in Manhattan, NYC under a matching radius of $R = 3\,km$. It can be shown that $\frac{\bar{l}}{\Delta\bar{l}}$ is linear with N, with a positive intercept. See Appendix A and Appendix C for the curve-fitting figures of experiments in other cities with other settings of matching radii.

We further show that the empirical law Eq. (6) is universal across different cities and different settings of matching radius. Table 6 shows the fitted parameters and goodness-of-fit, i.e. r-squared of Eq. (6) in experiments with different matching radius R in the three cities. It is found that, in most cases (especially when the matching radius is large), the empirical law fit the relationship between the average passenger detour distance and passenger demand well, with reasonably high r-squared (over 90%).

4.2. Law of average vehicle routing distance

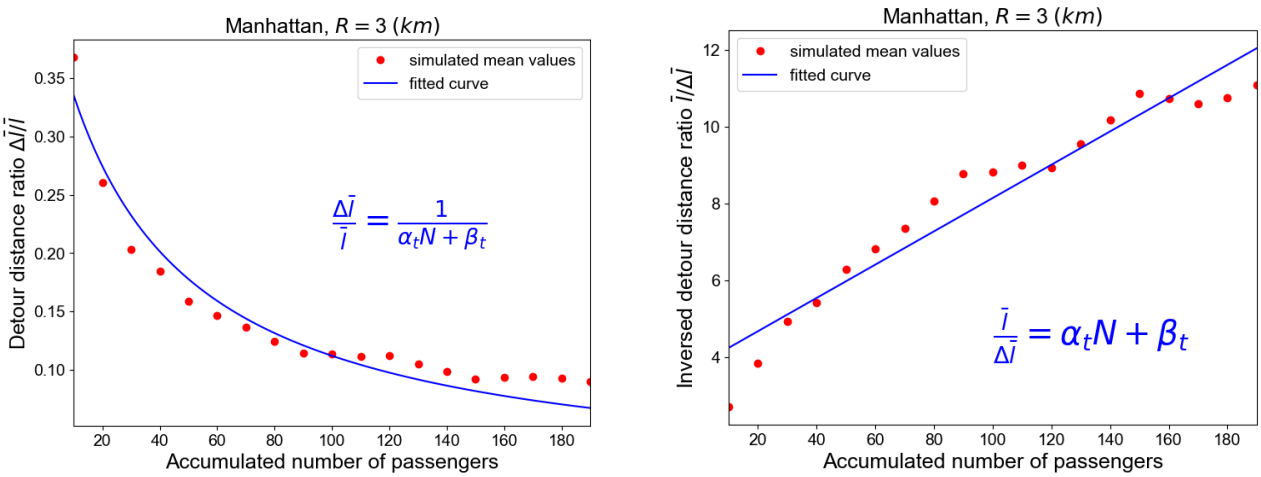
To find out the sound applicable empirical law to describe the relationship between average vehicle routing distance with respect to demand, we again examine the several common function



(a) Drivers' average routing distance $\bar{l_d}$

(b) Passengers' average detour distance $\Delta\bar{l}$

Figure 5: Probabilistic density distribution of $\Delta\bar{l}$ and $\bar{l_d}$ at different passenger demand



(a) $\Delta\bar{l}/\bar{l}$ v.s. N (b) $\bar{l}/\Delta\bar{l}$ v.s. N

Figure 6: $(P1)$ Curve fitting for average passenger detour distance $\Delta\bar{l}$

Table 6: Empirical law $\Delta\bar{l}/\bar{l} \sim N$- values of parameters in curve fitting

R (km)	Chengdu			Haikou			Manhattan		
	α_t	β_t	r-square	α_t	β_t	r-square	α_t	β_t	r-square
2	0.0089	5.2765	0.7986	0.0059	6.6671	0.5609	0.0309	3.5343	0.9944
3	0.0219	3.8355	0.9884	0.0141	4.1537	0.9354	0.0531	2.5723	0.9744
4	0.0336	3.3114	0.9559	0.0243	3.4480	0.9932	0.0646	2.2271	0.9525
5	0.0386	3.2476	0.9103	0.0347	2.9205	0.9697	0.0661	2.3196	0.9395

Table 7: Empirical law - vehicle routing distance given $R = 4\,km$ in Manhattan

Name	Expression	Parameter	r-square
Inverse proportional function	$\frac{\bar{l_d}}{\bar{l}} = \frac{1}{\alpha_t N + \beta_t}$	$\alpha_t,\ \beta_t$	0.6147
Inverse Log, proposed in this paper	$\frac{\bar{l_d}}{\bar{l}} = \frac{1}{\beta_d Log(\alpha_d N)} + 1$	$\beta_d,\ \alpha_d$	0.9968
Exponential	$\frac{\bar{l_d}}{\bar{l}} = Ae^{-\lambda_e N}$	$A,\ \lambda_e$	0.2457
Power law	$\frac{\bar{l_d}}{\bar{l}} = B\,(N)^{-\kappa}$	$B,\ \kappa$	0.5146

Table 8: Empirical law $\bar{l_d}/\bar{l} \sim N$ - values of parameters in curve fitting

R (km)	Chengdu			Haikou			Manhattan		
	α_d	β_d	r-square	α_d	β_d	r-square	α_d	β_d	r-square
2	5E+12	0.0197	0.9524	2E+25	0.0112	0.6838	3817.07	0.0497	0.9548
3	2E+04	0.0445	0.9671	5E+06	0.0321	0.9770	80.8712	0.0713	0.9933
4	530.25	0.0596	0.9873	7357	0.0471	0.9465	31.1548	0.0802	0.9968
5	267.03	0.0639	0.9985	319	0.0615	0.9873	28.8668	0.0812	0.9972

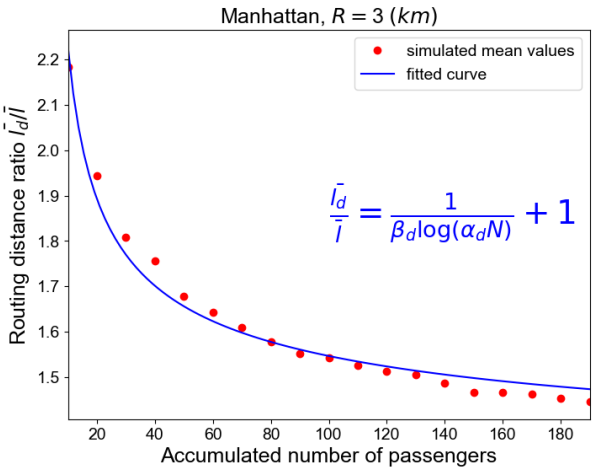
forms in Table 7. The inverse log formula is selected to describe the change of vehicle routing distance with demand by virtue of the highest *r-square* and concise form.

Then we look into the average vehicle routing distance $\bar{l_d}$ in experiments with different passenger demand N as inputs. We find that the following empirical law can well fit the relationship between $\bar{l_d}$ and N,

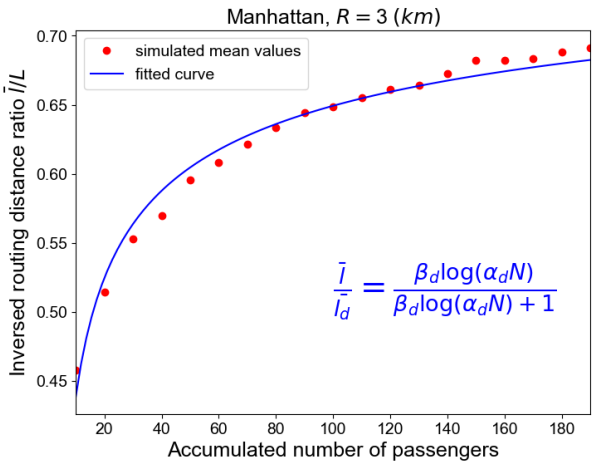
$$\frac{\bar{l_d}}{\bar{l}} = 1 + \frac{1}{\beta_d \log(\alpha_d N)} \tag{7}$$

where parameter α_d and β_d are dimensionless parameters, $\bar{l_d}/\bar{l}$ is referred to as the routing distance ratio. Figure 7 demonstrates the means of the routing distance ratio $\bar{l_d}/\bar{l}$ in (a) and the inverse of routing distance ratio $\bar{l}/\bar{l_d}$ in (b), at different passenger demand N and the curves fitted by Eq. (7). It is interesting to find that, unlike the relationship between average passenger detour distance and passenger demand, the inverse of the routing distance ratio $\bar{l_d}/\bar{l}$ is not linear with passenger demand N. Instead, the ratio increases with the logarithm of passenger demand N. See Appendix B and Appendix D for the figures of the relationships between $\bar{l_d}$ and N in different cities with different settings of matching radii. As shown by the calibrated formula, the ratio $\bar{l_d}/\bar{l}$ approaches 1, as N approaches infinity. This is consistent with the intuition that the routing distance approaches the trip distance as the number of waiting passengers in the matching pool approaches infinity such that the platform can pool-match passengers with extremely close origins and destinations and drivers hardly need to detour for picking up and dropping off a second passenger.

Through extensive experiments, we further show that the empirical law Eq. (7) can well fit the relationship between $\bar{l_d}$ and N under different settings of matching radii in various cities. Table 8 shows the fitted parameters and r-squared of Eq. (7) in experiments with different matching radius in three cities. We find that, in most cases (especially when the matching radius is large), the empirical law Eq. (7) well fit the relationship between the average vehicle routing distance and passenger demand, with fairly high r-squared.



(a) $\bar{l}_d/\bar{l}$ v.s. N



(b) $\bar{l}/\bar{l}_d$ v.s. N

Figure 7: ($P1$) Curve fitting for drivers' average routing distance $\bar{l}_d$

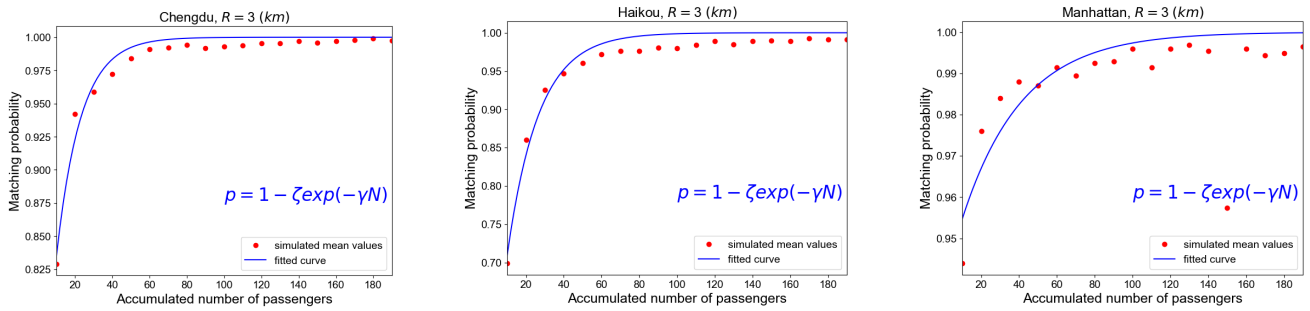
4.3. Law of pool-matching probability

Finally, we investigate the pool-matching probability p with different passenger demand N as inputs. We find that the following empirical law can well fit the relationship between p and N,

$$p = 1 - \zeta exp\left(-\gamma N\right) \tag{8}$$

where γ is a parameter with a dimension of $1/unit$ and ζ is a positive parameter($0 < \zeta \leq 1$) related to the network structure, matching radius and passengers' and drivers' distribution. Particularly, if $\zeta = 1$, Eq. (8) reduces to the formula identified in the numerical studies of Yan et al. (2019), i.e. $p = 1 - exp\left(-\gamma q \phi\right)$, where q is the arrival rate of passengers, ϕ is the length of the batch window. In a batch matching setting, suppose that the platform assigns vehicles to all waiting passengers at the end of each batch, $q\phi$ should be the number of waiting passengers in the matching pool when the platform executes a matching. Moreover, they assume that vehicles are required to pick-up two passengers at the middle point of their origins, and to drop-off them at the middle point of their destinations. It is noteworthy that our formula is a more general case of their formula, without the strict assumptions of pick-up and drop-off at the middle points of the two passengers. Compared our proposed formula with the similar formula in Yan et al. (2019), Eq. (8) is more persuasive in low demand situation where the passengers still have certain possibility to get matched with others, and the possibility depends on ζ. Different cities may give diverse ζ while the formula in Yan et al. (2019) presumes all the cities have the same pattern. Figure 8 demonstrates the trends of pool-matching probability in Chengdu, Haikou, Manhattan NYC, respectively, given different passenger demand N (with a matching radius equal to 3km), and the fitted curve with the form of Eq. (8). See Appendix E for figures of experiments in other cities with other settings of matching radii. It is shown that the ascending curve of pool-matching probability with passenger demand looks like a saturation curve that first increases quickly and then slowly.

It is further shown that the empirical law Eq. (8) well fit the relationship between p and N under different settings of matching radii in various cities. Table 9 shows the fitted parameters and r-squared of Eq. (8) in experiments with different matching radius in different cities. It is found that, when the matching radius is not extremely large, the relationship between pool-matching probability and passenger demand can be well fitted by Eq. (8), with very high r-squared (over



(a) Chengdu - p v.s. N (b) Haikou - p v.s. N (c) Manhattan - p v.s. N

Figure 8: ($P1$) Pool-matching probability v.s. passenger demand

Table 9: Empirical law of matching probability v.s. passengers' arrival rate in different cities

R (km)	Chengdu			Haikou			Manhattan		
	γ	ζ	r-square	γ	ζ	r-square	γ	ζ	r-square
2	0.0074	0.8923	0.9800	0.0042	0.9429	0.9835	0.0189	0.7053	0.9451
3	0.0233	0.7360	0.9689	0.0135	0.8000	0.9730	0.0453	0.4873	0.9073
4	0.0537	0.5697	0.9645	0.0343	0.7166	0.9689	0.0624	0.2541	0.9057
5	0.0766	0.3555	0.9664	0.0600	0.5282	0.9611	0.0316	0.0620	0.8749

90%). The reason for the relatively low R squared in experiments with large matching radius (for example 5 km) is that the matching radius is too large such that the pool-matching probability is always approaching to 1 as passenger demand increases from its minimum (10 in our experiment) to its maximum (200 in our experiment). It is also worth noting that the proposed empirical laws for the three measures are adapted to both downtown and suburb regions. To verify it, in Appendix Q, we conduct some experiments based on the mobility data in the downtown area of Manhattan, and show that the proposed empirical laws Eq. (6), Eq. (7), Eq. (8) can also well fit the relationships between the three measures and passenger demand respectively, expect that the parameters are different.

4.3.1. Discussions on matching probability

Our findings have far-reaching implications on the understanding of ride-pooling services. First, from Eq. (6) and Eq. (7), we know that the ratio of passenger detour distance Δl to non-shared trip distance is inversely proportional to a linear form of passenger demand N, while the ratio of vehicle routing distance $\bar{l}_d$ to non-shared trip distance $\bar{l}$ is inversely proportional to a logarithm form of passenger demand N. Second, these simple empirical laws have nice properties that are useful to the theoretical modellings. For example, taking the partial derivative of pool-matching probability p with respect to passenger demand N gives rise to

$$\frac{\partial p}{\partial N} = \zeta\gamma exp\left(-\gamma N\right) = \gamma\left(1-p\right) \tag{9}$$

which indicates that the partial derivative $\partial p/\partial N$ is always positive and decreases with p (saturation curve). Moreover, a key concept in the modelling of ride-pooling service is the number of vehicles required to serve N passengers,

$$M = \frac{1}{2}pN + \left(1-p\right)N = \left(1-\frac{1}{2}p\right)N \tag{10}$$

where the first term indicates that only one vehicle is required for two pool-matched passengers, while the second term shows that one vehicle is required for each unmatched passenger. Thus M is the total number of vehicles required to serve N passengers. Taking the partial derivative of M with respect to N yields

$$\frac{\partial M}{\partial N} = \frac{1}{2}\left[1 + \zeta exp\left(-\gamma N\right) - \zeta\gamma N exp\left(-\gamma N\right)\right] \tag{11}$$

It can be proved that $\partial M/\partial N > 0$ and $\partial M/\partial N \leq 1$ (as shown in Appendix P). The former shows that the number of vehicles required to serve all ride-pooling passengers increases with the number of passengers. This implies that, as passenger demand increases, the additional vehicles required to serve new passengers cannot be offset by the decrease in the number of required vehicles due to the increased pool-matching probability. The latter shows that, as passenger demand increases by one unit, the additional number of vehicles required is less than one unit due to pool-matching.

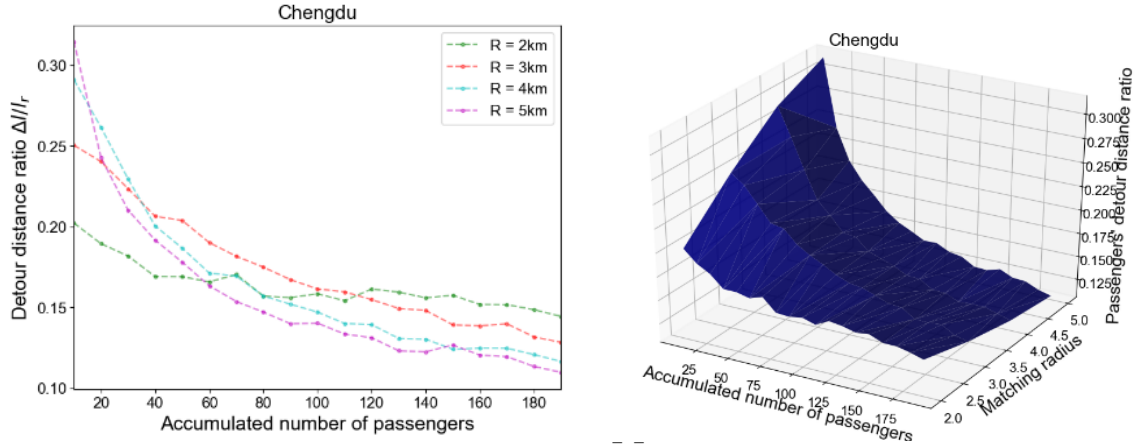
Moreover, one open question in the field of ride-sourcing services is the impacts of ride-pooling on transit usage, congestion. As mentioned in the introduction, the pool-matching probability, passenger detour distances and vehicle routing distances are the important factors that govern the aggregate impacts of ride-pooling services. The discovered empirical laws can help estimate the pool-matching probability, passenger detour distances, vehicle routing distances at different levels of passenger demand. As a result, the critical passenger demand that leads to a win-win situation where all participants, including the platform, passengers, drivers and the government (who aims to reduce traffic congestion) can benefit from the activation of ride-pooling services(Ke et al., 2019c).

4.4. Effect of matching radius

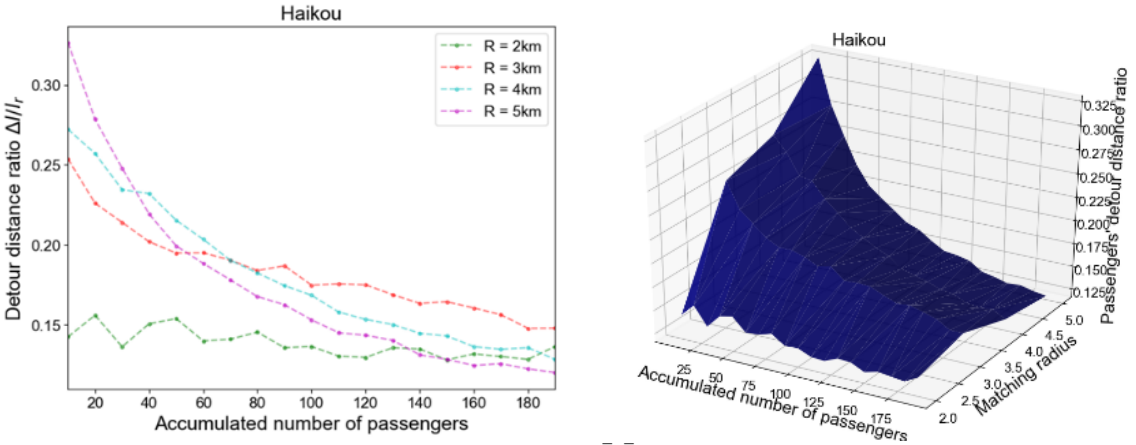
As discussed in Section 4.1-4.3, the maximum allowable matching radius R greatly affects the parameters of the empirical laws that describe the relationships between the three key measures and passenger demand. It can be easily found from Table 6 that α_t increases while β_t decreases with R in the empirical law for passengers' average detour distance. By contrast, from Table 8 we show that α_d decreases and β_d increases with in the empirical law for vehicle routing distance. These monotonic properties inspire us to incorporate the matching radius R into the empirical laws. First, we show that the following formula can well describe how the average detour distance $\Delta\bar{l}$ depends on passenger demand N and the matching radius R:

$$\frac{\Delta\bar{l}}{\bar{l}} = \frac{1}{N\left(\frac{\bar{l}}{R}\right)^{\mu_t} + \omega_t} \tag{12}$$

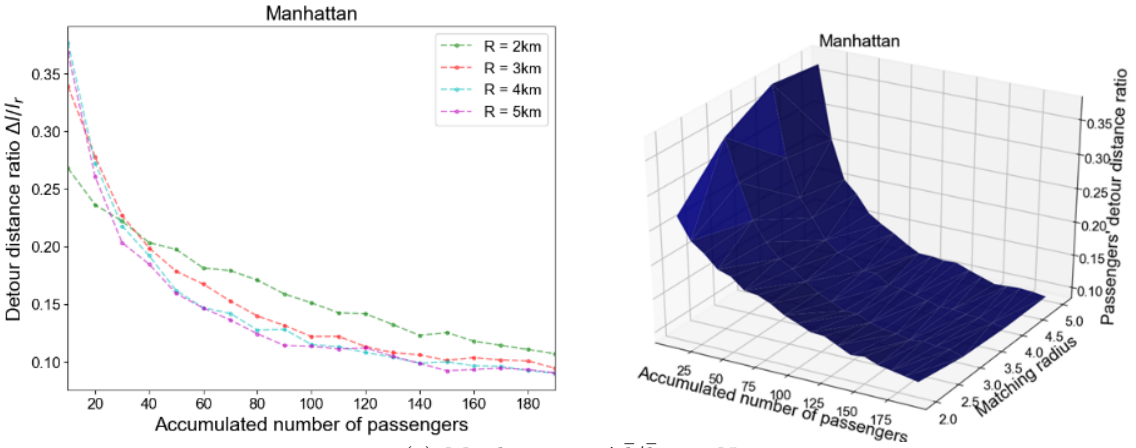
where μ_t and ω_t are two dimensionless parameters. It is clear that Eq. (12) has the same form as Eq. (6) when R takes a specific value. In this sense, Eq. (12) is a generalized formula of Eq. (6) by integrating the matching radius R. Figure 9 demonstrates the trends of the detour distance ratio $\Delta\bar{l}/\bar{l}$ with respect to the number of passengers N in the systems with different matching radii. Table 10 demonstrates the parameters and goodness-of-fit of calibrated formula specified by Eq. (12). Although the r-squared for the formula with matching radius is not as high as the formula without matching radius, it is still acceptable. It indicates that the formula in Eq. (12) can well spell out the relationship among passengers' average detour distance, passenger demand, and matching radius. It is also consistent with the intuition that the larger the matching radius, the larger the average detour distance of passengers.



(a) Chengdu - $\Delta\bar{l}/\bar{l}$ v.s. N



(b) Haikou - $\Delta\bar{l}/\bar{l}$ v.s. N



(c) Manhattan - $\Delta\bar{l}/\bar{l}$ v.s. N

Figure 9: ($P1$) Empirical Law of detour distance integrating R

Table 10: values of parameters in curve fitting of average passengers' detour distance when integrate R

City name	μ_t	ω_t	r-square
Chengdu	0.0020	9.1015	0.7053
Haikou	0.0013	10.4842	0.7460
Manhattan	0.0036	6.7223	0.6909

Table 11: values of parameters in curve fitting of average passengers' detour distance when integrate R

City name	μ_d	ω_d	r-square
Chengdu	24917.5869	0.0487	0.6605
Haikou	4170.7138	0.0471	0.6847
Manhattan	534.6560	0.0670	0.7140

Second, we show that the following formula can approximate how the divers' average routing distance $\bar{l_d}$ depend on the passenger demand N and matching radius R:

$$\frac{\bar{l_d}}{\bar{l}} = 1 + \frac{1}{\omega_d \log\left(\frac{\mu_d N \bar{l}}{R}\right)} \tag{13}$$

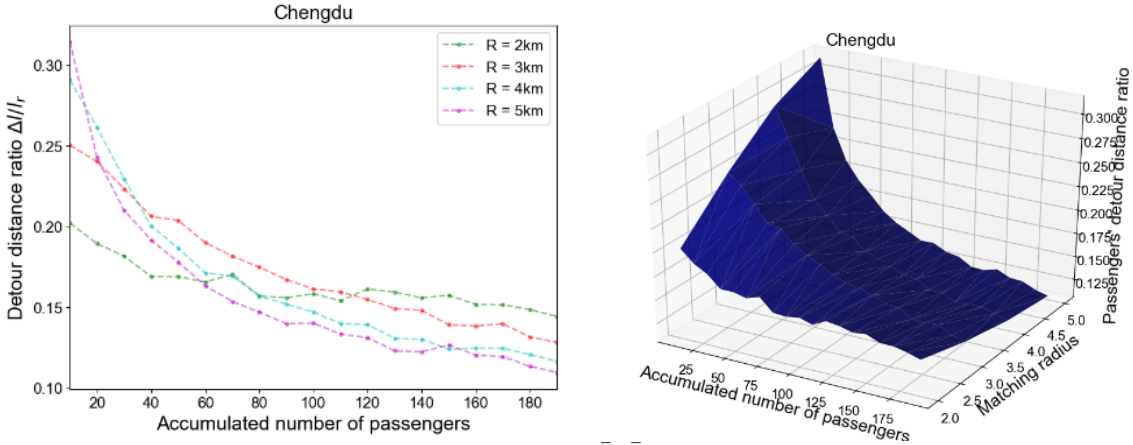
where ω_d, μ_d are two dimensionless parameters. Clearly, the formula in Eq. (13) can be regarded as a generalized form of the formula in Eq. (7), as the matching radius takes a specific value. Figure 10 displays the curves of the routing distance ratio $\bar{l_d}/\bar{l}$ v.s. the number of passengers N in the matching systems with distinctive matching radii. Table 11 shows the parameters and goodness-of-fit of calibrated formula specified by Eq. (13). We can see that r-squared is in an acceptable range, which indicates that this simple formula can characterize the relationship among vehicle routing distance, number of passengers and the matching radius. From the formula, we can infer that the larger the matching radius, the lager the average vehicle routing distance, which is intuitive.

Third, we construct the following formula to delineate how the passenger demand N and matching radius R affect the pool-matching probability p:

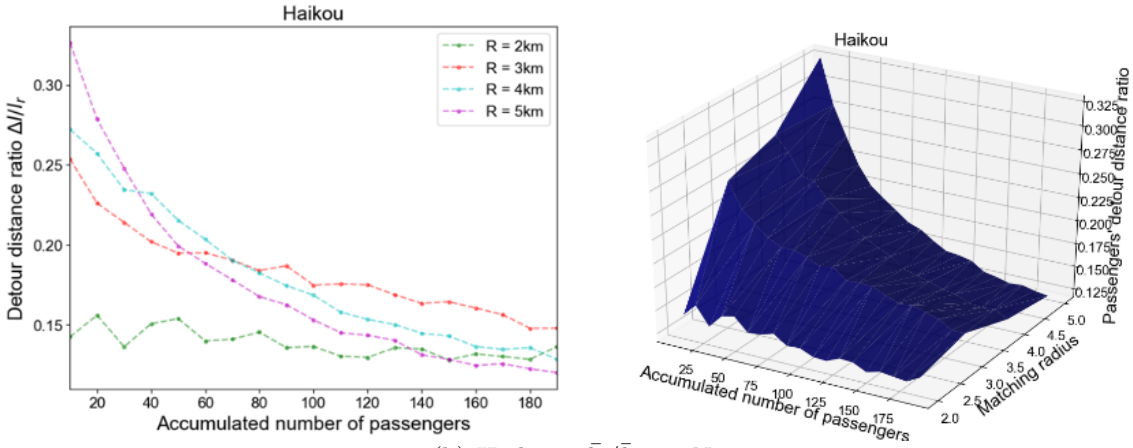
$$p = 1 - \omega \exp\left(-N\left(\frac{R}{\bar{l}}\right)^{\mu}\right) \tag{14}$$

where ω, μ are two dimensionless parameters. Clearly, the generalized formula in Eq. (14) will be reduced to Eq. (8), when the matching radius R is given. Figure 11 illustrates the pool-matching probability with the increase of the number of passengers in the matching pool, under different matching radii. We calibrate Eq. (14) with the actual mobility data in the three cities, the results of which are shown in Table 12. It can be found that the r-squared is still high, although it is lower than that in Eq. (14). This indicates that this formula can well fit the relationship among the pool-matching probability, the passenger demand and the matching radius. In addition, we can see that the larger the matching radius, the higher the pool-matching probability. This is intuitive since a larger allowable maximum matching radius R is able to pool-match more passengers in each matching time interval.

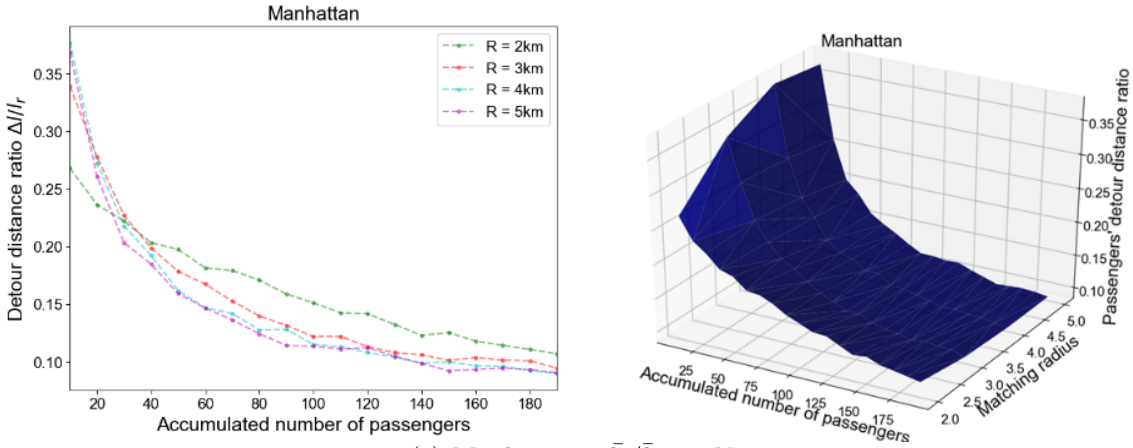
To summarize, this subsection discusses three generalized empirical laws that integrate the key decision variable – matching radius R, for the three key measures – the passengers' detour distance,



(a) Chengdu - $\bar{l_d}/\bar{l}$ v.s. N

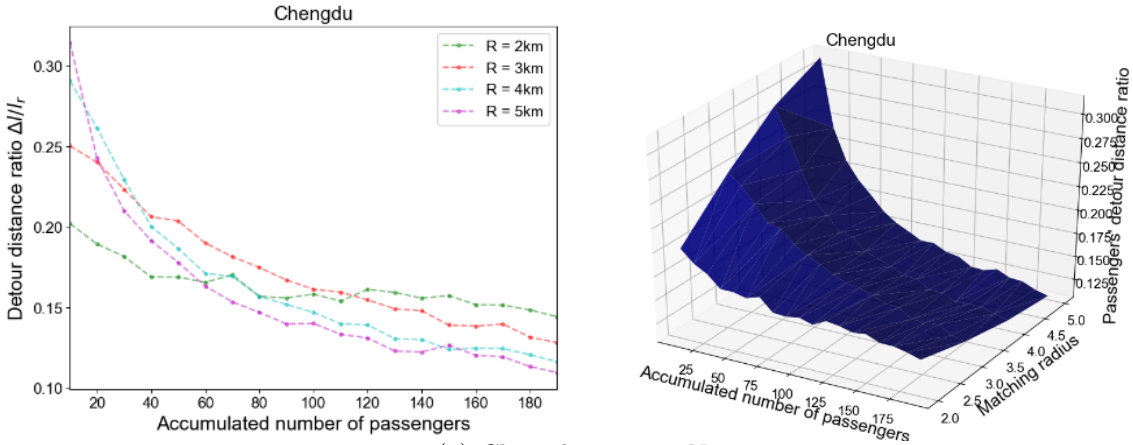


(b) Haikou - $\bar{l_d}/\bar{l}$ v.s. N

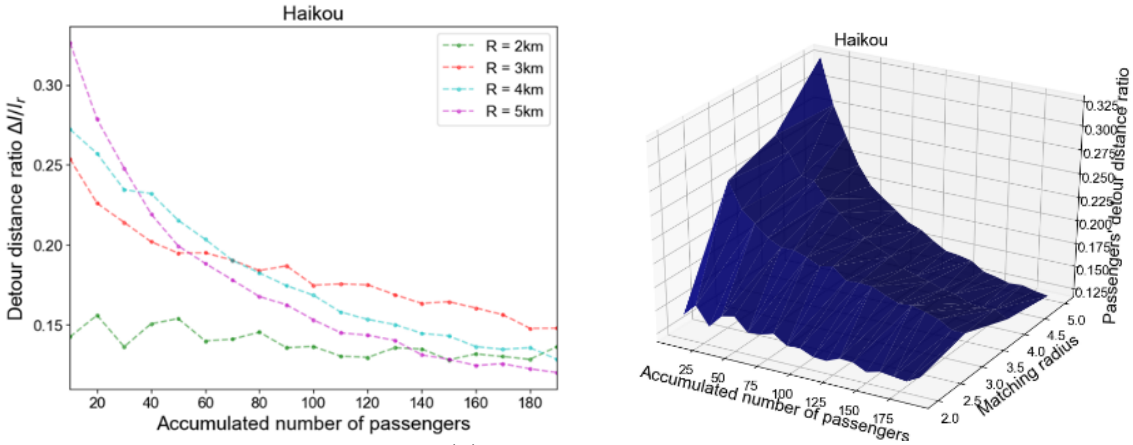


(c) Manhattan - $\bar{l_d}/\bar{l}$ v.s. N

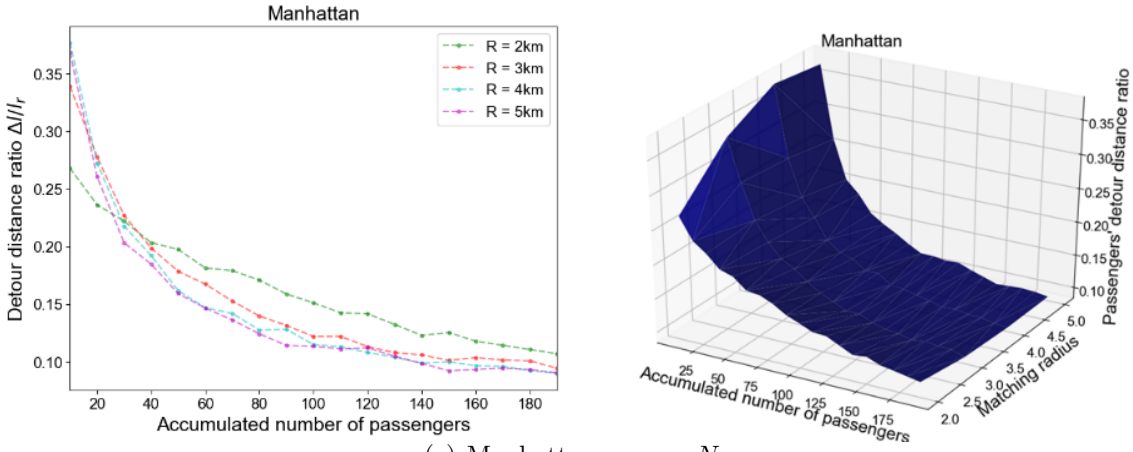
Figure 10: ($P1$) Empirical Law of vehicle routing distance integrating R



(a) Chengdu - p v.s. N



(b) Haikou - p v.s. N



(c) Manhattan - p v.s. N

Figure 11: ($P1$) Empirical Law of matching probability integrating R

Table 12: values of parameters in curve fitting of matching probability when integrate R

City name	μ	ω	r-square
Chengdu	0.0020	9.1015	0.7053
Haikou	0.0013	10.4842	0.7460
Manhattan	0.0036	6.7223	0.6909

vehicle routing distance and pool-matching probability. These empirical laws are neat and well fit the relationships by having only a few parameters. In addition, they spell out the effects of the matching radius on the three measures, which are also consistent with our intuitions. Although the matching radius is incorporated into the empirical laws, we can see that the calibrated parameters are different across cities. These parameters in a way reflect the geographical characteristics of the cities, such as network structure, density of roads, and the spatial distributions of the trip requests. It is interesting to examine how these exogenous factors influence the empirical laws or the relationships among the key measures, passenger demand and matching radius. However, since we have not collected data from a sufficient number of cities for the inference, this exploration will be relayed to future studies.

4.5. Discussion on empirical laws

Despite the proposed empirical laws perform well with a relative high goodness-of-fit, there are still some limitations in the experiments under certain settings. For example, when the matching radius is extremely small, i.e. less than $1\,km$, we can observe that matching probability drops dramatically, and the pattern of vehicle routing distance and passengers' detour distance becomes ambiguous and unstable. This is mainly due to the fact that very few passengers are pool-matched as the allowable matching distance is rather short, resulting in large variance of vehicle routing distance and passenger detour distance, and thus the unstable estimates in curve fitting. Another limitation in our study is related to the network topology. In our experiment settings, We attach all the trips' origins and destinations into the road network's intersection for convenience and only consider the arterial roads, which may lead to the deviation in the calculation of average trip distance. To what exact extent will the network topology influence our findings remains unclear. Moreover, due to data limitation, we only investigate three cities, which are insufficient to conduct some statistical analysis on how the values of the curve parameters are related to the city characteristics, such as density of demand, area of the network, density of road segments and shape of the network. We leave such an analysis for our future studies as mobility data of more cities becomes available.

5. Conclusions

This paper investigates how the three key measures (pool-matching probability, passenger detour distance and vehicle routing distance) increase or decrease with passenger demand. Through extensive experiments, we find that (1) the ratio of passenger detour distance to non-shared trip distance is inversely proportional to a linear form of passenger demand; (2) the ratio of vehicle routing distance to non-shared trip distance is inversely proportional to a logarithm form of passenger demand; (3) the pool-matching probability increases with passenger demand in a negative-exponential formula. It is found that these empirical laws can well fit the relationships between the three measures and passenger demand under different settings of matching radii in various cities. The empirical laws exhibit some nice properties and can be adopted to serve as the fundamental building blocks in the modelling of the on-demand ride-pooling markets such as assessment of the impacts of ride-pooling services on traffic congestio.

There are several possible extensions. First, the optimization models only consider the pool-matching for two passengers for simplicity, while the pool-matching for more than two passengers can be investigated in future studies. Second, it is of immerse interest to investigate the effects of the geographical features and demand's spatial patterns on passengers' detour distance, vehicle routing distance and pool-matching probability. Third, despite the neat and simple empirical laws

can well fit the curves of the key measures v.s. passenger demand, they lack theoretical foundations and explanations. Fourth, it is interesting to investigate the empirical laws under a continuous ride-pooling strategy, where a driver may pick up a passenger before dropping off all passengers in the vehicle. A challenging but important research direction is to evaluate the empirical laws (or derive some new laws) by theoretically analyzing the physical matching process.

ACKNOWLEDGEMENT

The authors wish to express their thanks to three anonymous reviewers for their valuable comments on an early version of this paper. The work described in this paper was supported by a grant from Hong Kong Research Grants Council under project HKUST16208619, and by NSFC/RGC Joint Research grants: N_HKUST627/18 and 71861167001. This work was also supported by Didi Chuxing.

References

L. P. Alexander and M. C. González. Assessing the impact of real-time ridesharing on urban traffic using mobile phone data. *Proc. UrbComp*, pages 1–9, 2015.

R. Arnott. Taxi travel should be subsidized. *Journal of Urban Economics*, 40(3):316–333, 1996.

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

M. Batty. The size, scale, and shape of cities. *science*, 319(5864):769–771, 2008.

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

D. Brockmann, L. Hufnagel, and T. Geisel. The scaling laws of human travel. *Nature*, 439(7075): 462–465, 2006.

J. C. Castillo, D. T. Knoepfle, and E. G. Weyl. Surge pricing solves the wild goose chase. *Available at SSRN:id=2890666*, 2018.

X. M. Chen, M. Zahiri, and S. Zhang. Understanding ridesplitting behavior of on-demand ride services: An ensemble learning approach. *Transportation Research Part C: Emerging Technologies*, 76:51–70, 2017.

X. Geng, Y. Li, L. Wang, L. Zhang, Q. Yang, J. Ye, and Y. Liu. Spatiotemporal multi-graph convolution network for ride-hailing demand forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 3656–3663, 2019.

F. He, X. Wang, X. Lin, and X. Tang. Pricing and penalty/compensation strategies of a taxi-hailing platform. *Transportation Research Part C: Emerging Technologies*, 86:263–279, 2018.

M. Iqbal. Uber revenue and usage statistics. https://www.businessofapps.com/data/uber-statistics/, 2020.

J. Jacob and R. Roet-Green. Ride solo or pool: Designing price-service menus for a ride-sharing platform. *Available at SSRN :id=3008136*, 2019.

J. Ke, H. Zheng, H. Yang, and X. M. Chen. Short-term forecasting of passenger demand under on-demand ride services: A spatio-temporal deep learning approach. *Transportation Research Part C: Emerging Technologies*, 85:591–608, 2017.

J. Ke, H. Yang, H. Zheng, X. Chen, Y. Jia, P. Gong, and J. Ye. Hexagon-based convolutional neural network for supply-demand forecasting of ride-sourcing services. *IEEE Transactions on Intelligent Transportation Systems*, 20(11):4160–4173, 2018.

J. Ke, X. Cen, H. Yang, X. Chen, and J. Ye. Modelling drivers' working and recharging schedules in a ride-sourcing market with electric vehicles and gasoline vehicles. *Transportation Research Part E: Logistics and Transportation Review*, 125:160–180, 2019a.

J. Ke, F. Xiao, H. Yang, and J. Ye. Optimizing online matching for ride-sourcing services with multi-agent deep reinforcement learning. *IEEE Transactions on Knowledge and Data Engineering (submitted, https://arxiv.org/abs/1902.06228)*, 2019b.

J. Ke, H. Yang, and Z. Zheng. On the benefits of ride-splitting in the presence of traffic congestion. *Transportation Research Part B: Methodological(submitted)*, 2019c.

J. Ke, H. Yang, X. Li, H. Wang, and J. Ye. Pricing and equilibrium in on-demand ride-pooling markets. *Transportation Research Part B: Methodological*, page (in press), 2020.

W. Li, Z. Pu, Y. Li, and X. J. Ban. Characterization of ridesplitting based on observed data: a case study of chengdu, china. *Transportation Research Part C: Emerging Technologies*, 100: 330–353, 2019.

Z. Li, Y. Hong, and Z. Zhang. Do ride-sharing services affect traffic congestion? an empirical study of uber entry. *Social Science Research Network*, 2002:1–29, 2016.

X. Liang, X. Zheng, W. Lv, T. Zhu, and K. Xu. The scaling of human mobility by taxis is exponential. *Physica A: Statistical Mechanics and its Applications*, 391(5):2135–2144, 2012.

K. Lin, R. Zhao, Z. Xu, and J. Zhou. Efficient large-scale fleet management via multi-agent deep reinforcement learning. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 1774–1783, 2018.

M. Nourinejad and M. Ramezani. Ride-sourcing modeling and pricing in non-equilibrium two-sided markets. *Transportation Research Part B: Methodological*, 132:340–357, 2020.

B. Schaller. The new automobility: Lyft, uber and the future of american cities. http://www.schallerconsult.com/rideservices/automobility.htm, 2018.

S. Shaheen, A. Cohen, I. Zohdy, et al. Shared mobility: current practices and guiding principles. Technical Report FHWA-HOP-16-022, United States. Federal Highway Administration, 2016.

C. Smith. 25 amazing didi facts and statistics. https://expandedramblings.com/index.php/didi-chuxing-facts-statistics/, 2020.

C. Song, T. Koren, P. Wang, and A.-L. Barabási. Modelling the scaling properties of human mobility. *Nature Physics*, 6(10):818–823, 2010.

H. Sun, H. Wang, and Z. Wan. Model and analysis of labor supply for ride-sharing platforms in the presence of sample self-selection and endogeneity. *Transportation Research Part B: Methodological*, 125:76–93, 2019.

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

X. Wang, W. Liu, H. Yang, D. Wang, and J. Ye. Customer behavioural modelling of order cancellation in coupled ride-sourcing and taxi markets. *Transportation Research Part B: Methodological*, 132:358–378, 2020.

Z. Wang, Z. Qin, X. Tang, J. Ye, and H. Zhu. Deep reinforcement learning with knowledge transfer for online rides order dispatching. In *2018 IEEE International Conference on Data Mining (ICDM)*, pages 617–626. IEEE, 2018.

Z. Xu, Y. Yin, and L. Zha. Optimal parking provision for ride-sourcing services. *Transportation Research Part B: Methodological*, 105:559–578, 2017.

Z. Xu, Z. Li, Q. Guan, D. Zhang, Q. Li, J. Nan, C. Liu, W. Bian, and J. Ye. Large-scale order dispatch in on-demand ride-hailing platforms: A learning and planning approach. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 905–913, 2018.

Z. Xu, Y. Yin, and J. Ye. On the supply curve of ride-hailing systems. *Transportation Research Part B: Methodological*, 132:29–43, 2020.

C. Yan, H. Zhu, N. Korolko, and D. Woodard. Dynamic pricing and matching in ride-hailing platforms. *Naval Research Logistics (NRL)*, 2019.

H. Yang and T. Yang. Equilibrium properties of taxi markets with search frictions. *Transportation Research Part B: Methodological*, 45(4):696–713, 2011.

H. Yang, M. Ye, W. H. Tang, and S. C. Wong. Regulating taxi services in the presence of congestion externality. *Transportation Research Part A: Policy and Practice*, 39(1):17–40, 2005.

H. Yang, C. W. Leung, S. C. Wong, and M. G. Bell. Equilibria of bilateral taxi–customer searching and meeting on networks. *Transportation Research Part B: Methodological*, 44(8-9):1067–1083, 2010.

H. Yang, J. Ke, and J. Ye. A universal distribution law of network detour ratios. *Transportation Research Part C: Emerging Technologies*, 96:22–37, 2018.

H. Yang, X. Qin, J. Ke, and J. Ye. Optimizing matching time interval and matching radius in on-demand ride-sourcing markets. *Transportation Research Part B: Methodological*, 131:84–105, 2020.

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

J. J. Yu, C. S. Tang, Z.-J. Max Shen, and X. M. Chen. A balancing act of regulating on-demand ride services. *Management Science*, 2019.

L. Zha, Y. Yin, and H. Yang. Economic analysis of ride-sourcing markets. *Transportation Research Part C: Emerging Technologies*, 71:249–266, 2016.

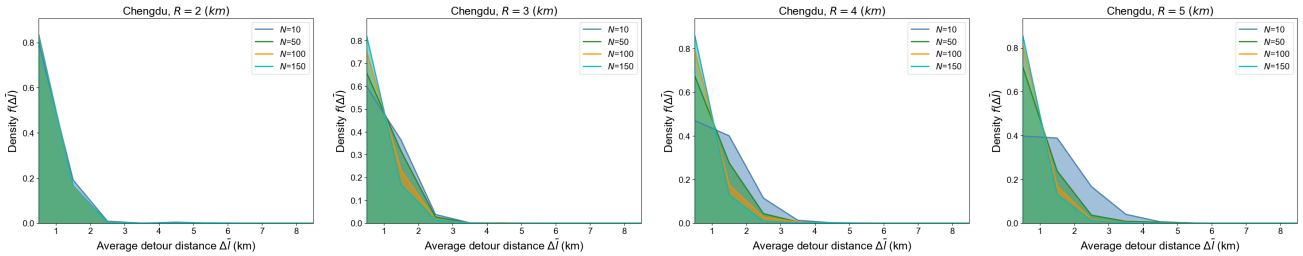
L. Zha, Y. Yin, and Y. Du. Surge pricing and labor supply in the ride-sourcing market. *Transportation Research Part B: Methodological*, 117:708–722, 2018a.

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

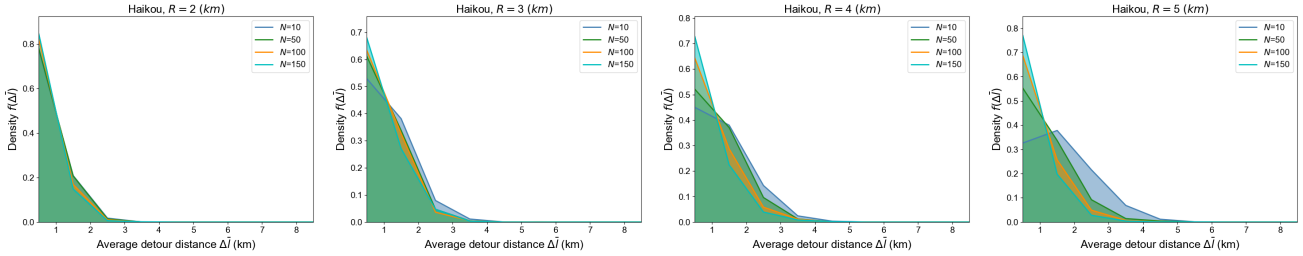
Z. Zhu, X. Qin, J. Ke, Z. Zheng, and H. Yang. Analysis of multi-modal commute behavior with feeding and competing ridesplitting services. *Transportation Research Part A: Policy and Practice*, 132:713–727, 2020.

Appendix A. Probabilistic density distribution of average passengers' detour distance under objective $P1$

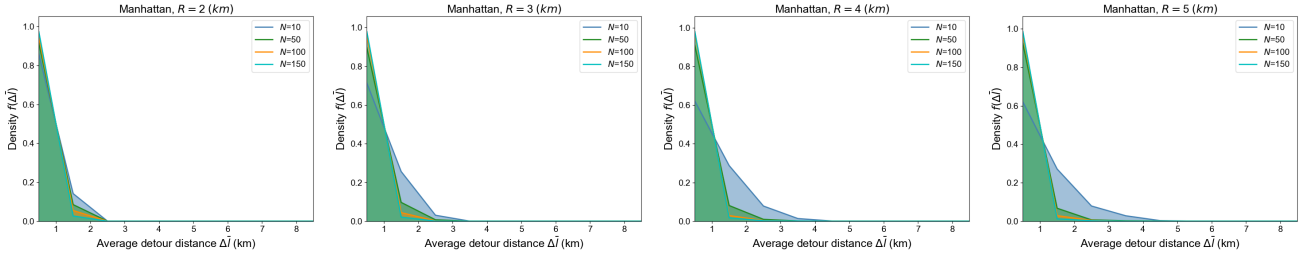
($P1$) Probabilistic density distribution of average passengers' detour distance $\Delta\bar{l}$ in the simulations with four groups of passenger demand $N = 10$, $N = 50$, $N = 100$ and $N = 150$ given the R, (a), $R = 2$km, (b), $R = 3$km, (c), $R = 4$km, (d), $R = 5$km, respectively.



(a) Chengdu - *pdf* of $\Delta\bar{l}$



(b) Haikou - *pdf* of $\Delta\bar{l}$

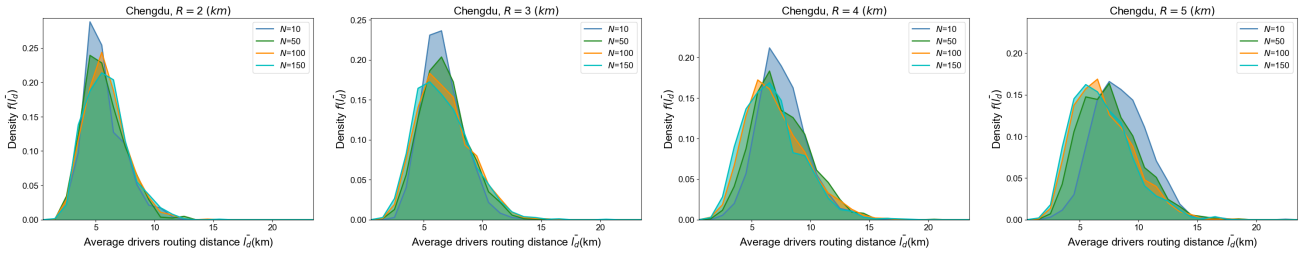


(c) Manhattan - *pdf* of $\Delta\bar{l}$

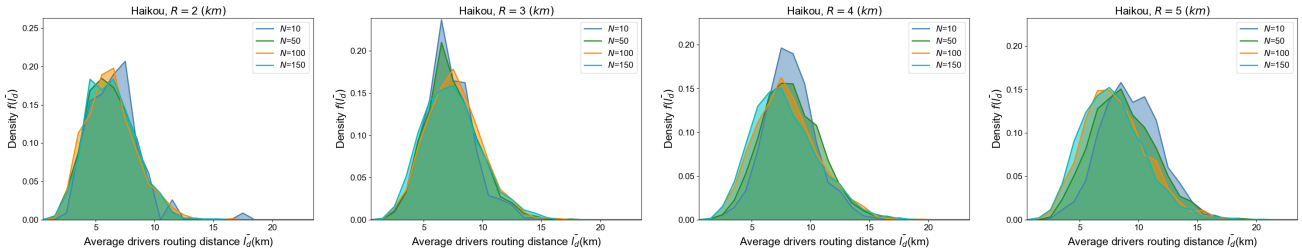
Figure A.12: ($P1$) Probabilistic density distribution of average passengers' detour distance $\Delta\bar{l}$ with different R

Appendix B. Probabilistic density distribution of average vehicle routing distance under objective $P1$

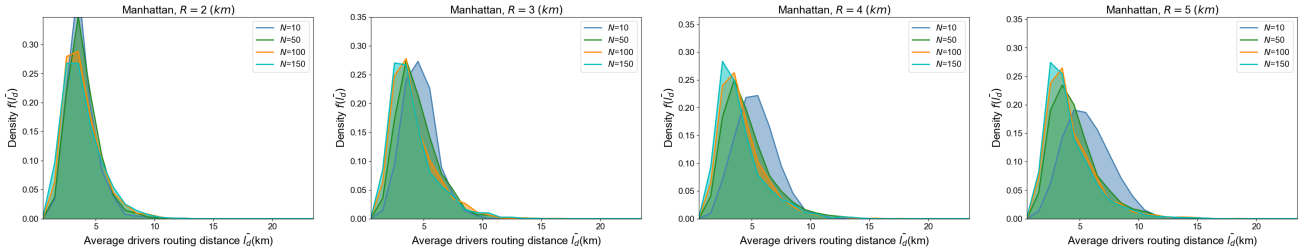
($P1$) Probabilistic density distribution of vehicle routing distance $\bar{l_d}$ in the simulations with four groups of passenger demand $N = 10$, $N = 50$, $N = 100$ and $N = 150$ given the , $R = 2$km, $R = 3$km, $R = 4$km, $R = 5$km for three cities, respectively.



(a) Chengdu - *pdf* of $\bar{l_d}$



(b) Haikou - *pdf* of $\bar{l_d}$

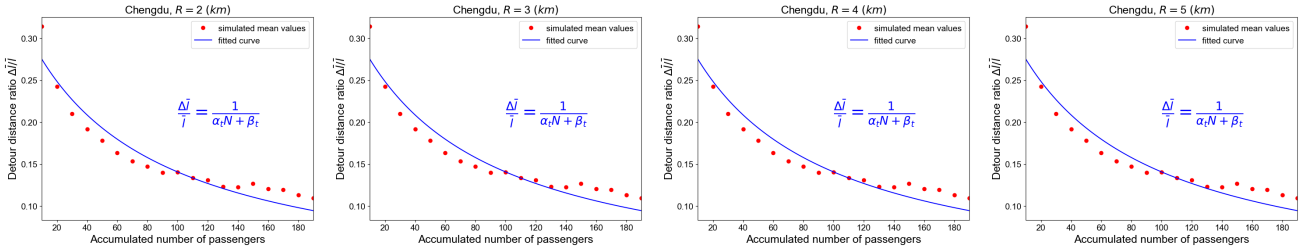


(c) Manhattan - *pdf* of $\bar{l_d}$

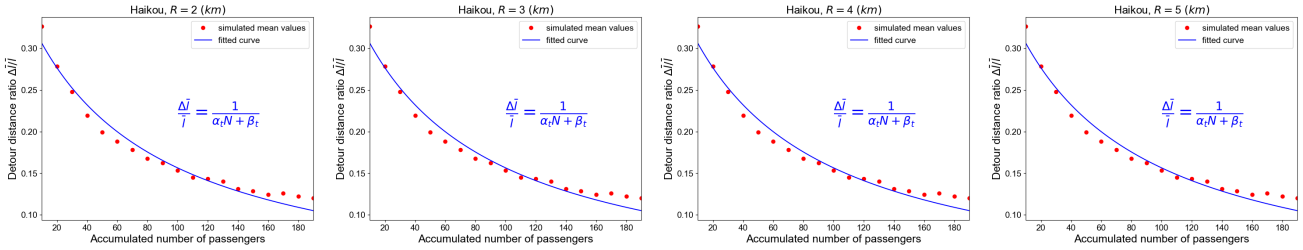
Figure B.13: ($P1$) Probabilistic density distribution of vehicle routing distance $\bar{l_d}$ with different R

Appendix C. Empirical law of passenger detour distance under objective $P1$

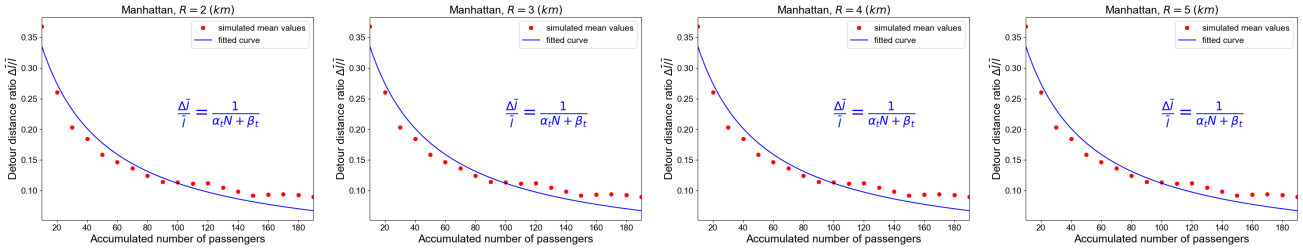
($P1$) Passenger detour distance v.s. passenger demand in various matching radii, $R = 2$km, $R = 3$km, $R = 4$km, $R = 5$km for three cities, respectively.



(a) Chengdu - Empirical fitting between $\Delta\bar{l}$ v.s. N



(b) Haikou - Empirical fitting between $\Delta\bar{l}$ v.s. N

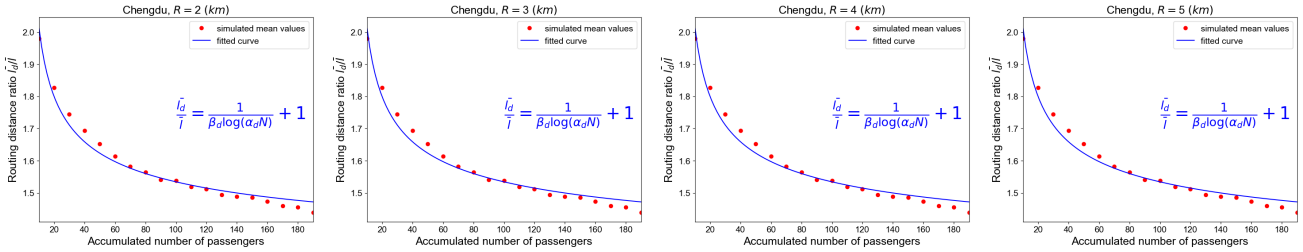


(c) Manhattan - Empirical fitting between $\Delta\bar{l}$ v.s. N

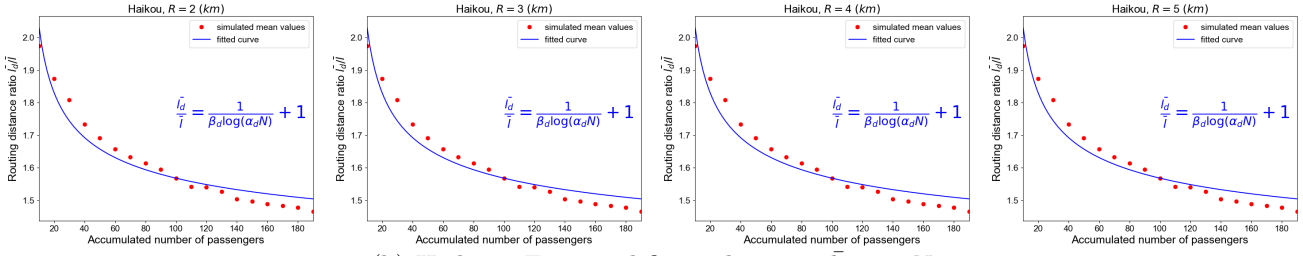
Figure C.14: ($P1$) Passenger detour distance v.s. passenger demand in various matching radii R

Appendix D. Empirical law of vehicle routing distance under objective $P1$

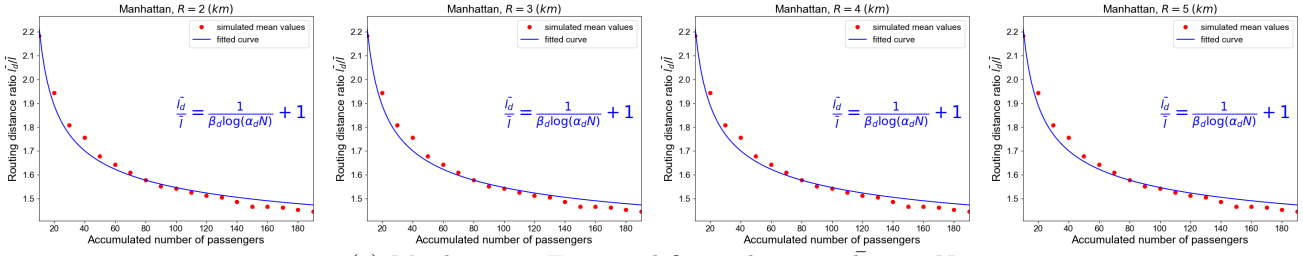
($P1$) vehicle routing distance v.s. passenger demand in various cities when given different matching radii, $R = 2$km, $R = 3$km, $R = 4$km, $R = 5$km for three cities, respectively.



(a) Chengdu - Empirical fitting between $\bar{l}_d$ v.s. N



(b) Haikou - Empirical fitting between $\bar{l}_d$ v.s. N

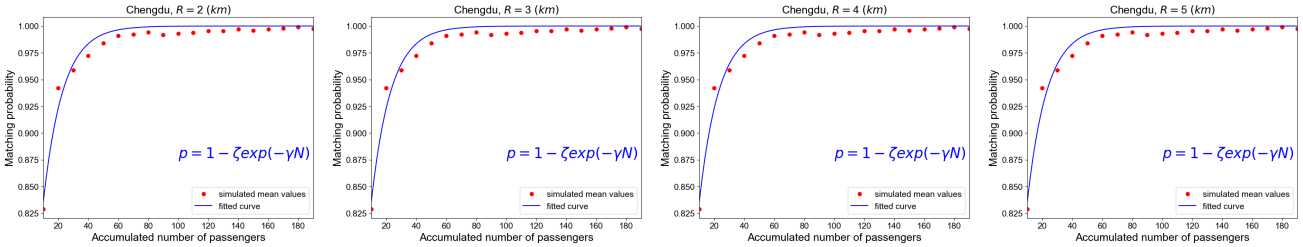


(c) Manhattan - Empirical fitting between $\bar{l}_d$ v.s. N

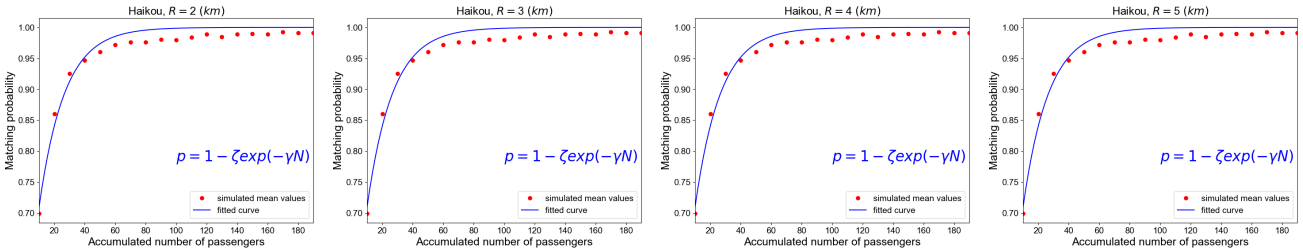
Figure D.15: ($P1$) vehicle routing distance under different arrival rate and matching radii R

Appendix E. Empirical fitting of pool-matching probability under objective $P1$

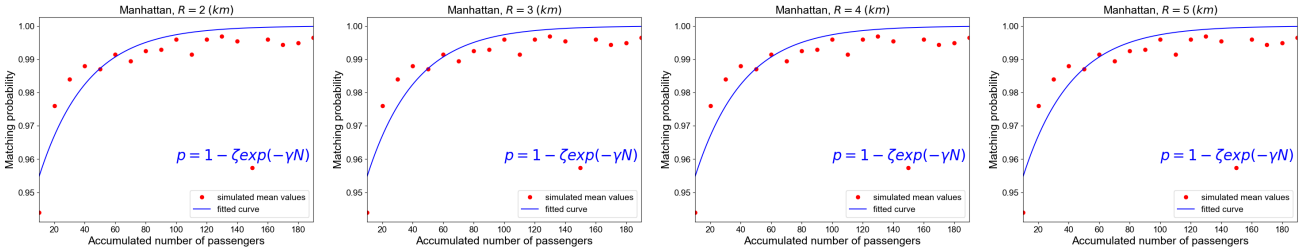
($P1$) Pool-matching probability v.s. passenger demand in various cities when given different matching radius, $R = 2$km, $R = 3$km, $R = 4$km, $R = 5$km for three cities, respectively.



(a) Chengdu - Empirical fitting between p v.s. N



(b) Haikou - Empirical fitting between p v.s. N

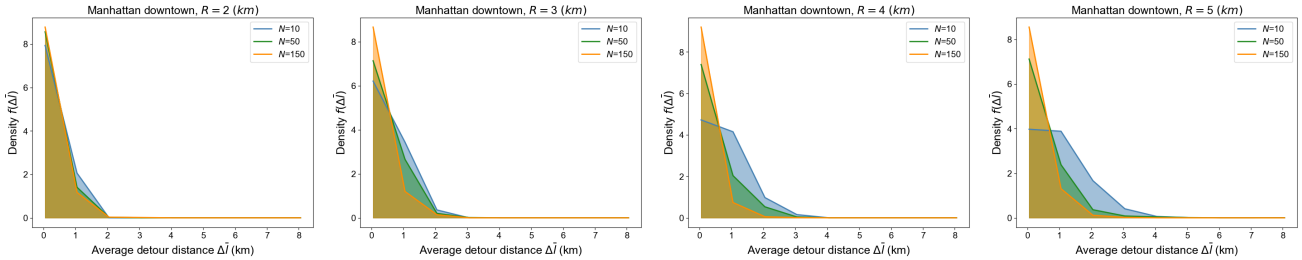


(c) Manhattan - Empirical fitting between p v.s. N

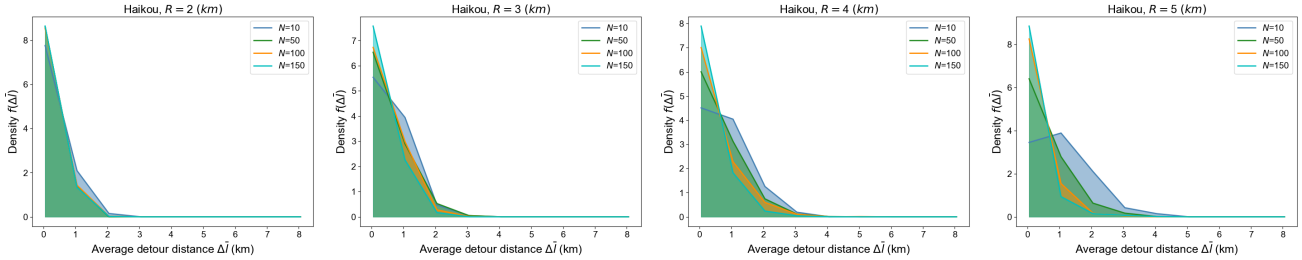
Figure E.16: ($P1$) Passenger-passenger matching probability under different arrival rate and matching radii R

Appendix F. Probabilistic density distribution of average passengers' detour distance under objective $P2$

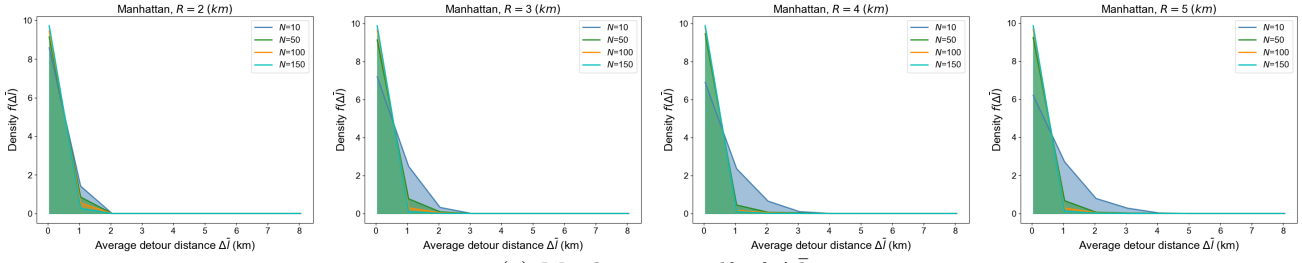
($P2$) Probabilistic density distribution of average passengers' detour distance $\Delta\bar{l}$ in the simulations with four groups of passenger demand $N = 10$, $N = 50$, $N = 100$ and $N = 150$ given the R, (a), $R = 2$km, (b), $R = 3$km, (c), $R = 4$km, (d), $R = 5$km, respectively.



(a) Chengdu - *pdf* of $\Delta\bar{l}$



(b) Haikou - *pdf* of $\Delta\bar{l}$

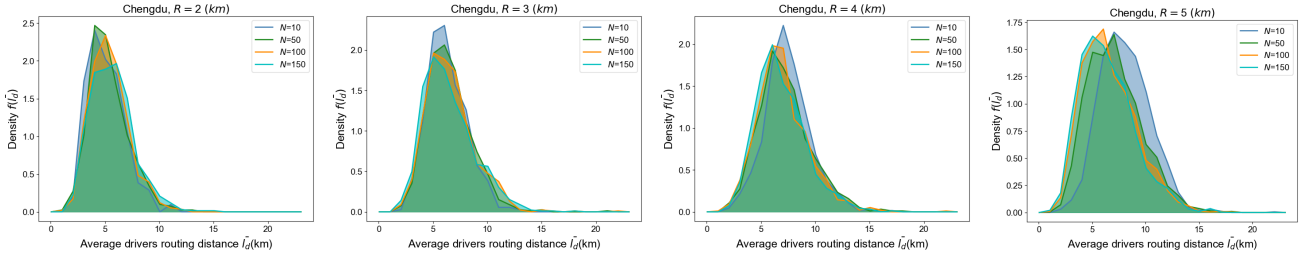


(c) Manhattan - *pdf* of $\Delta\bar{l}$

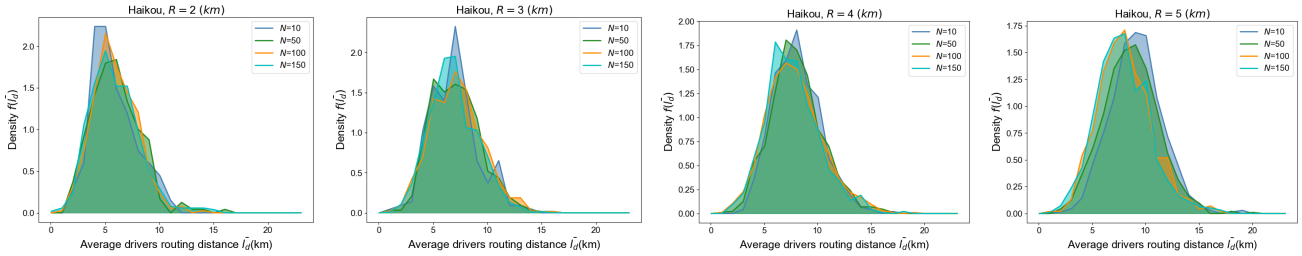
Figure F.17: ($P2$) Probabilistic density distribution of average passengers' detour distance $\Delta\bar{l}$ with different R

Appendix G. Probabilistic density distribution of vehicle routing distance under objective *P2*

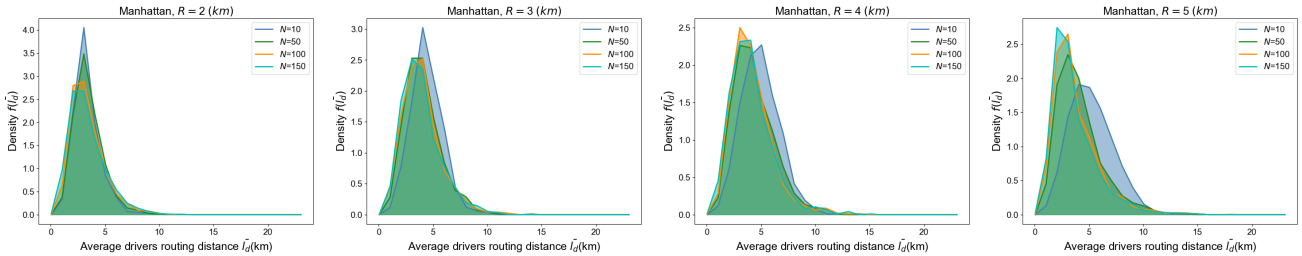
($P2$) Probabilistic density distribution of vehicle routing distance $\bar{l_d}$ in the simulations with four groups of passenger demand $N = 10$, $N = 50$, $N = 100$ and $N = 150$ given the , $R = 2$km, $R = 3$km, $R = 4$km, $R = 5$km for three cities, respectively.



(a) Chengdu - *pdf* of $\bar{l_d}$



(b) Haikou - *pdf* of $\bar{l_d}$

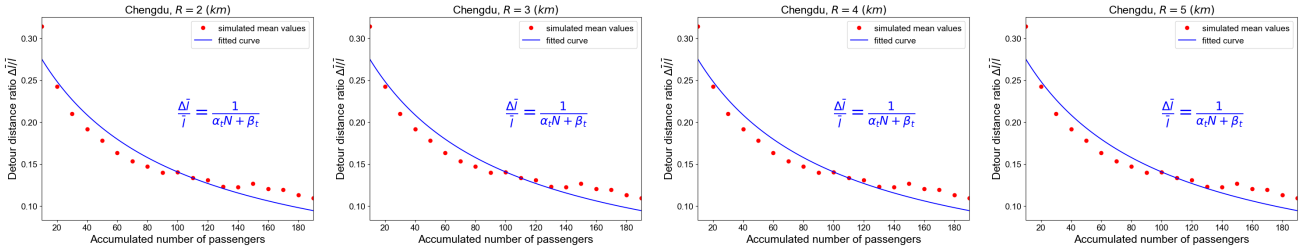


(c) Manhattan - *pdf* of $\bar{l_d}$

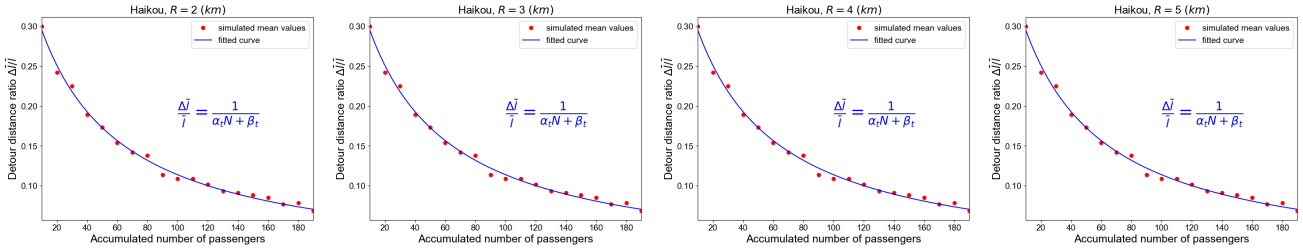
Figure G.18: ($P2$) Probabilistic density distribution of vehicle routing distance $\bar{l_d}$ with different R

Appendix H. Empirical law of passenger detour distance under objective *P2*

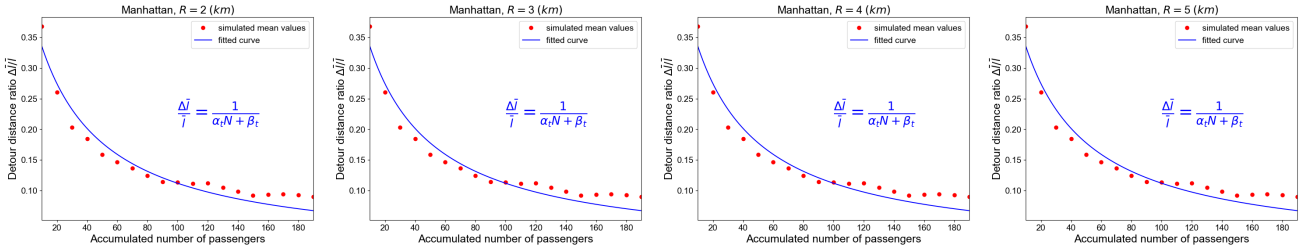
($P2$) Passenger detour distance v.s. passenger demand in various matching radii, $R = 2$km, $R = 3$km, $R = 4$km, $R = 5$km for three cities, respectively.



(a) Chengdu - Empirical fitting between $\Delta\bar{l}$ v.s. N



(b) Haikou - Empirical fitting between $\Delta\bar{l}$ v.s. N

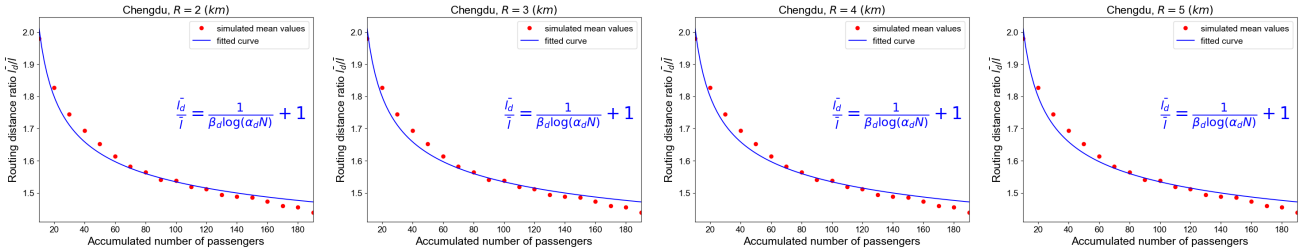


(c) Manhattan - Empirical fitting between $\Delta\bar{l}$ v.s. N

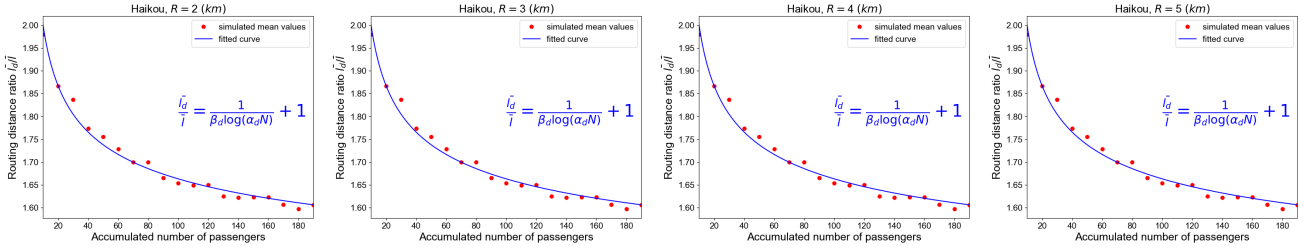
Figure H.19: ($P2$) Passenger detour distance v.s. passenger demand in various matching radii R

Appendix I. Empirical law of vehicle routing distance under objective $P2$

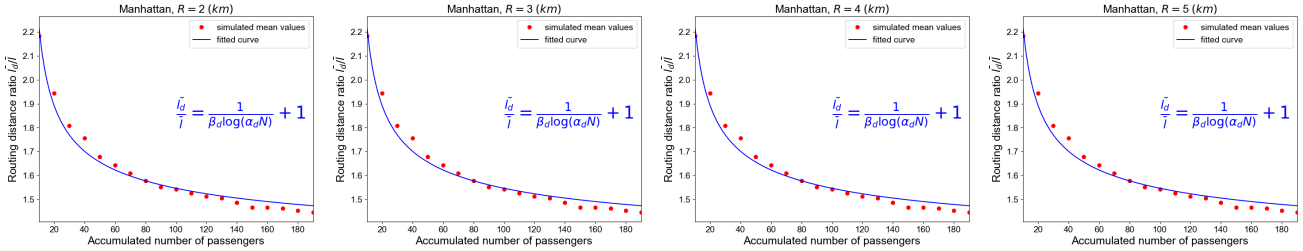
($P2$) vehicle routing distance v.s. passenger demand in various cities when given different matching radii, $R = 2$km, $R = 3$km, $R = 4$km, $R = 5$km for three cities, respectively.



(a) Chengdu - Empirical fitting between $\bar{l}_d$ v.s. N



(b) Haikou - Empirical fitting between $\bar{l}_d$ v.s. N

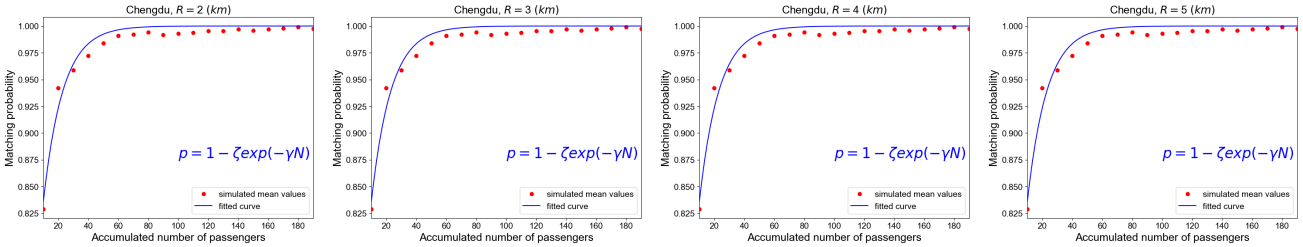


(c) Manhattan - Empirical fitting between $\bar{l}_d$ v.s. N

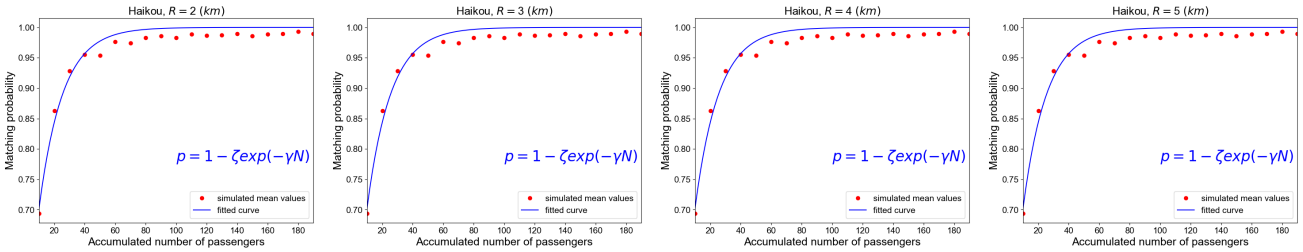
Figure I.20: ($P2$) vehicle routing distance under different arrival rate and matching radii R

Appendix J. Empirical fitting of pool-matching probability under objective $P2$

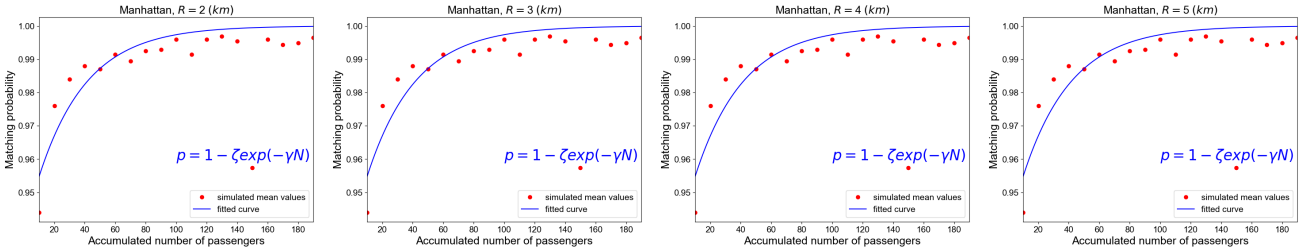
($P2$) Pool-matching probability v.s. passenger demand in various cities when given different matching radius, $R = 2$km, $R = 3$km, $R = 4$km, $R = 5$km for three cities, respectively.



(a) Chengdu - Empirical fitting between p v.s. N



(b) Haikou - Empirical fitting between p v.s. N

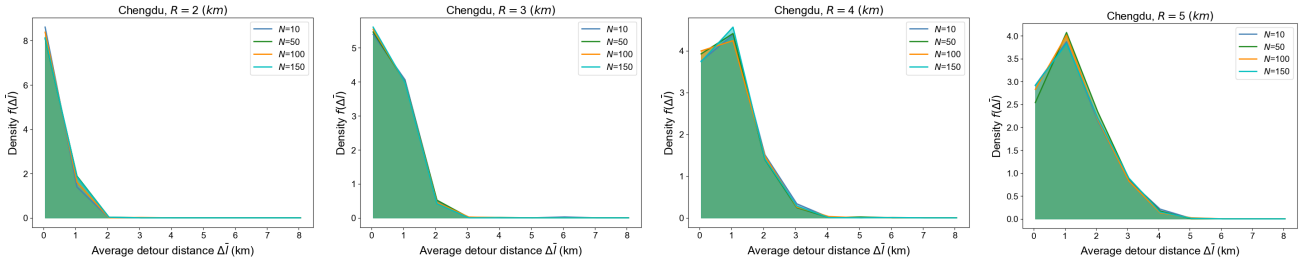


(c) Manhattan - Empirical fitting between p v.s. N

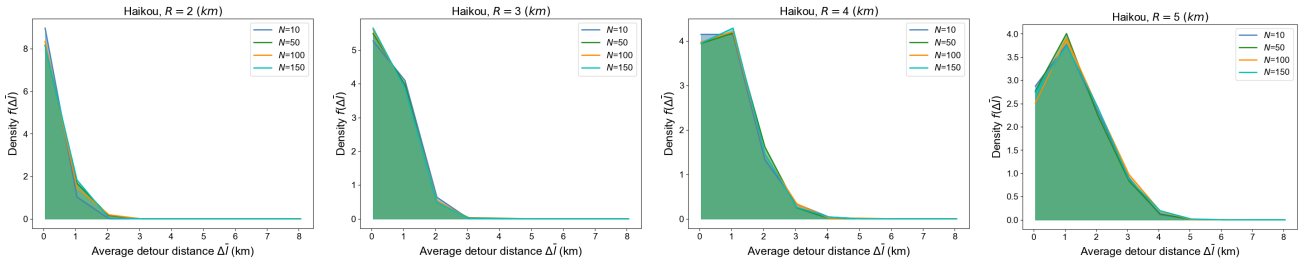
Figure J.21: ($P2$) Passenger-passenger matching probability under different arrival rate and matching radii R

Appendix K. Probabilistic density distribution of average passengers' detour distance under scenario $P3$

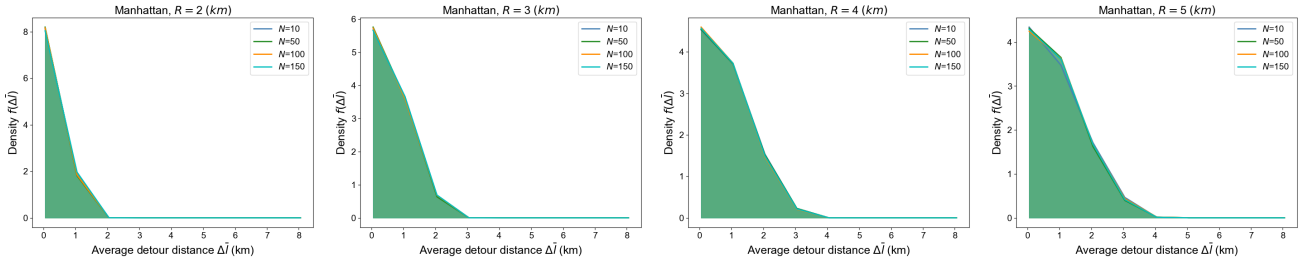
($P3$) Probabilistic density distribution of average passengers' detour distance $\Delta\bar{l}$ in the simulations with four groups of passenger demand $N = 10$, $N = 50$, $N = 100$ and $N = 150$ given the R, (a), $R = 2$km, (b), $R = 3$km, (c), $R = 4$km, (d), $R = 5$km, respectively.



(a) Chengdu - *pdf* of $\Delta\bar{l}$



(b) Haikou - *pdf* of $\Delta\bar{l}$

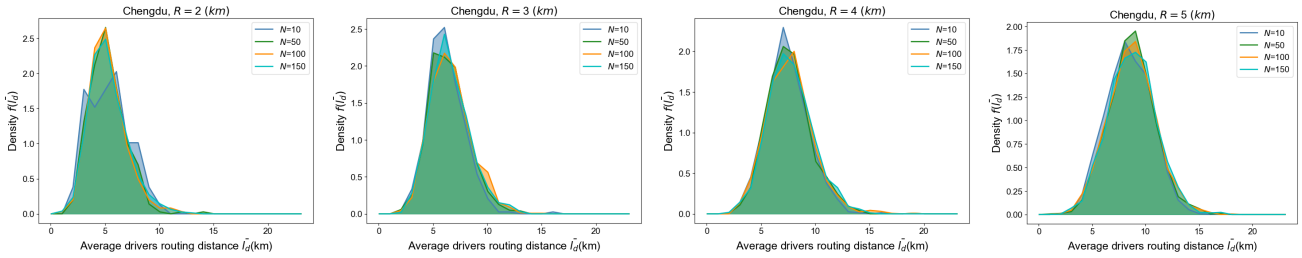


(c) Manhattan - *pdf* of $\Delta\bar{l}$

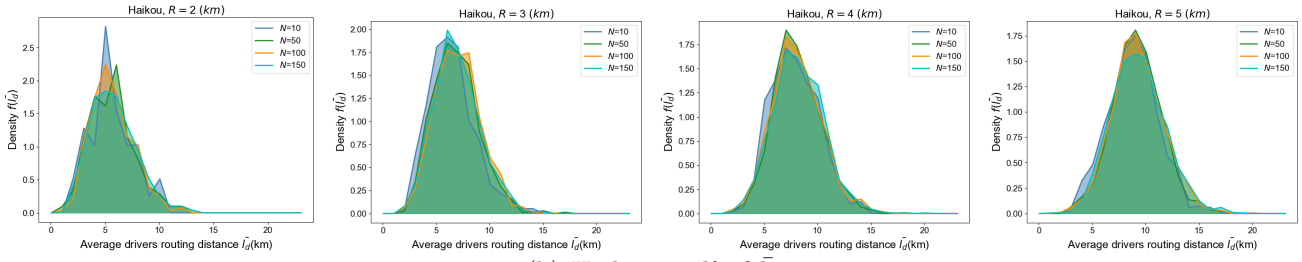
Figure K.22: ($P3$) Probabilistic density distribution of average passengers' detour distance $\Delta\bar{l}$ with different R

Appendix L. Probabilistic density distribution of vehicle routing distance under scenario *P*3

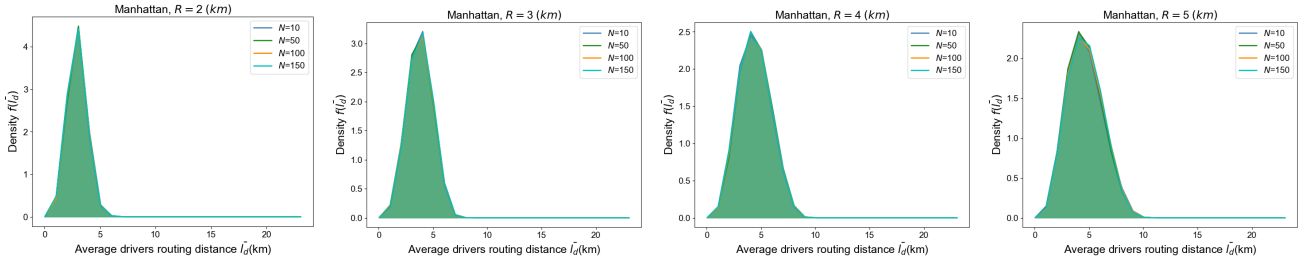
(*P*3) Probabilistic density distribution of vehicle routing distance $\bar{l_d}$ in the simulations with four groups of passenger demand $N = 10$, $N = 50$, $N = 100$ and $N = 150$ given the , $R = 2$km, $R = 3$km, $R = 4$km, $R = 5$km for three cities, respectively.



(a) Chengdu - *pdf* of $\bar{l_d}$



(b) Haikou - *pdf* of $\bar{l_d}$

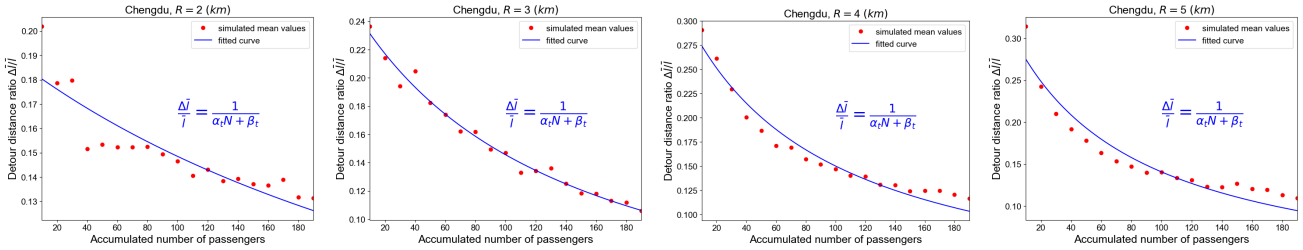


(c) Manhattan - *pdf* of $\bar{l_d}$

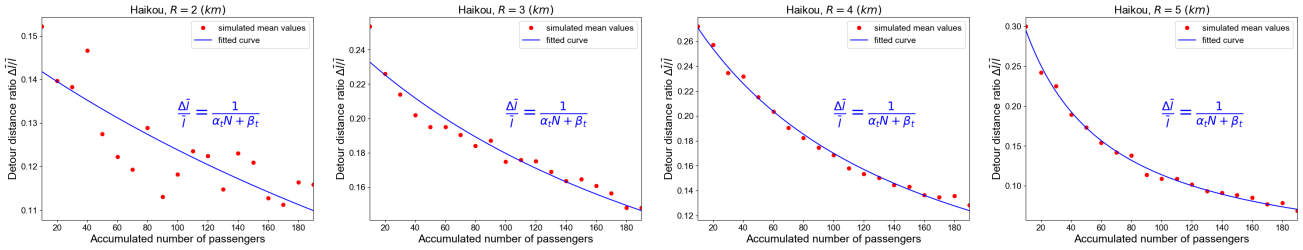
Figure L.23: (*P*3) Probabilistic density distribution of vehicle routing distance $\bar{l_d}$ with different R

Appendix M. Empirical law of passenger detour distance under scenario $P3$

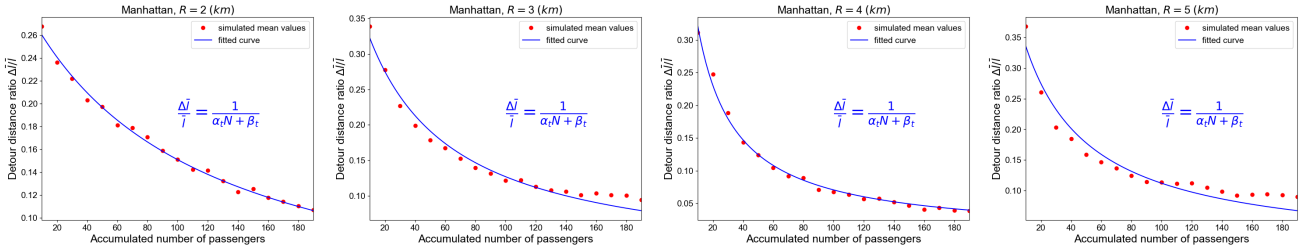
($P3$) Passenger detour distance v.s. passenger demand in various matching radii, $R = 2$km, $R = 3$km, $R = 4$km, $R = 5$km for three cities, respectively.



(a) Chengdu - Empirical fitting between $\Delta\bar{l}$ v.s. N



(b) Haikou - Empirical fitting between $\Delta\bar{l}$ v.s. N

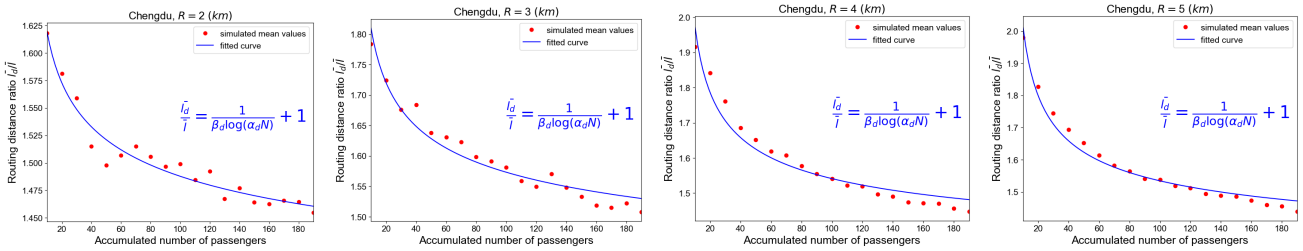


(c) Manhattan - Empirical fitting between $\Delta\bar{l}$ v.s. N

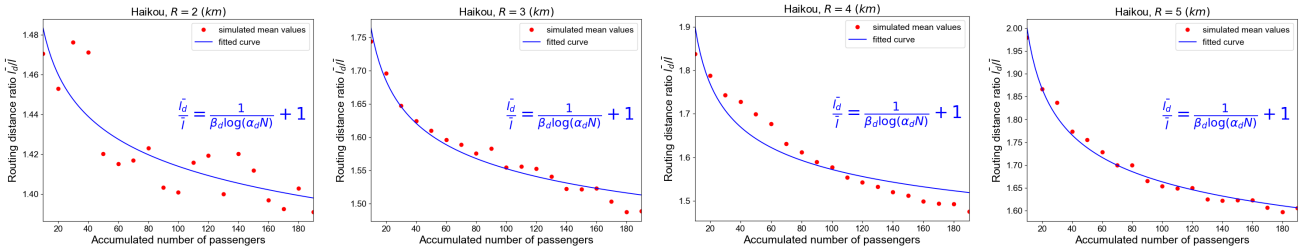
Figure M.24: ($P3$) Passenger detour distance v.s. passenger demand in various matching radii R

Appendix N. Empirical law of vehicle routing distance under objective $P3$

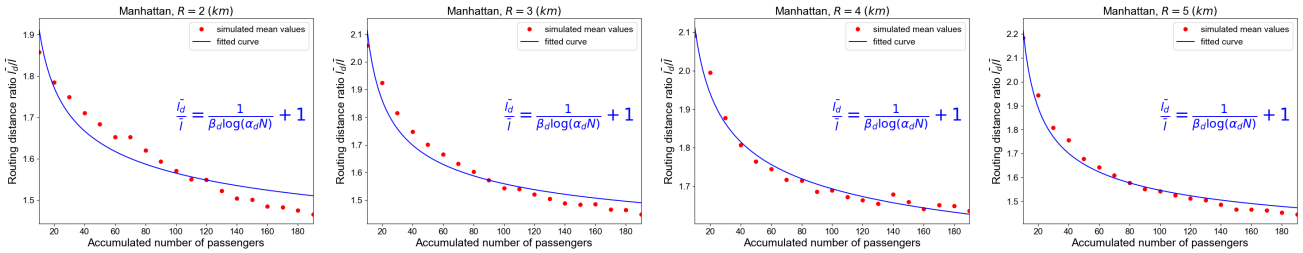
($P3$) vehicle routing distance v.s. passenger demand in various cities when given different matching radii, $R = 2$km, $R = 3$km, $R = 4$km, $R = 5$km for three cities, respectively.



(a) Chengdu - Empirical fitting between $\bar{l_d}$ v.s. N



(b) Haikou - Empirical fitting between $\bar{l_d}$ v.s. N

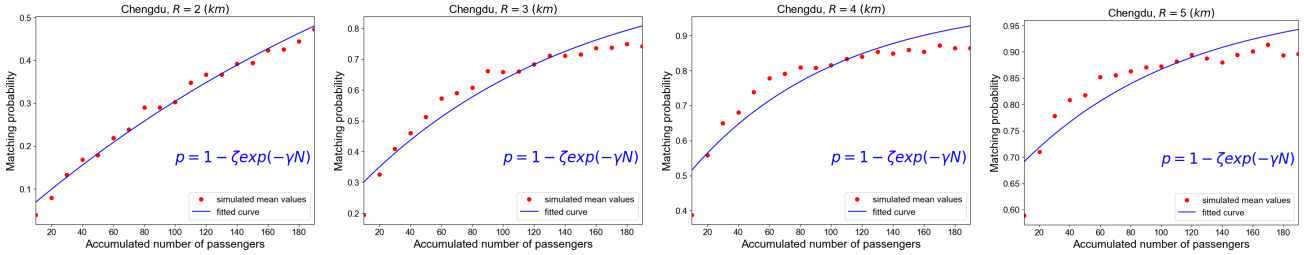


(c) Manhattan - Empirical fitting between $\bar{l_d}$ v.s. N

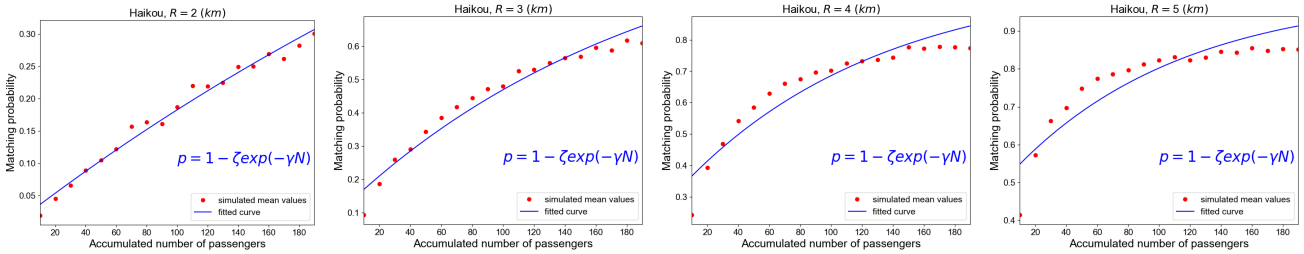
Figure N.25: ($P3$) vehicle routing distance under different arrival rate and matching radii R

Appendix O. Empirical fitting of pool-matching probability under objective $P3$

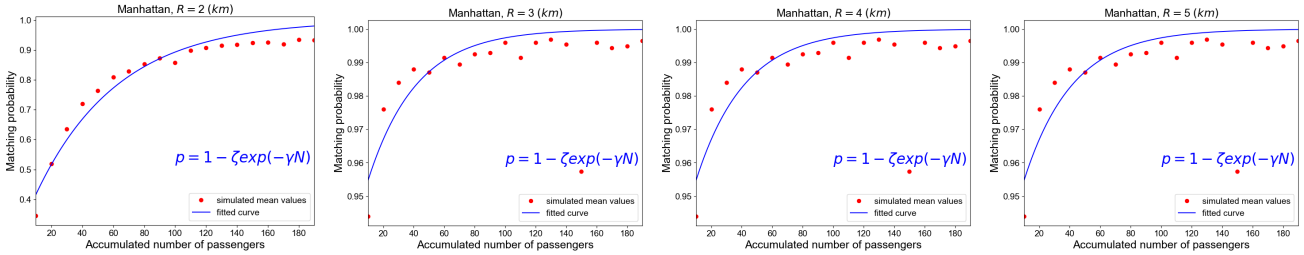
($P3$) Pool-matching probability v.s. passenger demand in various cities when given different matching radius, $R = 2$km, $R = 3$km, $R = 4$km, $R = 5$km for three cities, respectively.



(a) Chengdu - Empirical fitting between p v.s. N



(b) Haikou - Empirical fitting between p v.s. N



(c) Manhattan - Empirical fitting between p v.s. N

Figure O.26: ($P3$) Passenger-passenger matching probability under different arrival rate and matching radii R

Table Q.13: Empirical law- values of parameters in curve fitting

R (km)	$\Delta\bar{l}/\bar{l}$ v.s. N			$\bar{l}_d/\bar{l}$ v.s. N			p v.s. N		
	α_t	β_t	r-square	α_d	β_d	r-square	γ	ζ	r-square
1	0.0753	2.8695	0.8577	0.7808	0.4658	0.9879	0.0962	0.0187	0.9566
4	0.0886	2.7028	0.9773	0.7328	0.4754	0.9832	0.1003	0.0331	0.8826

Appendix P. Proof of $\partial M/\partial N > 0$ and $\partial M/\partial N \leq 1$

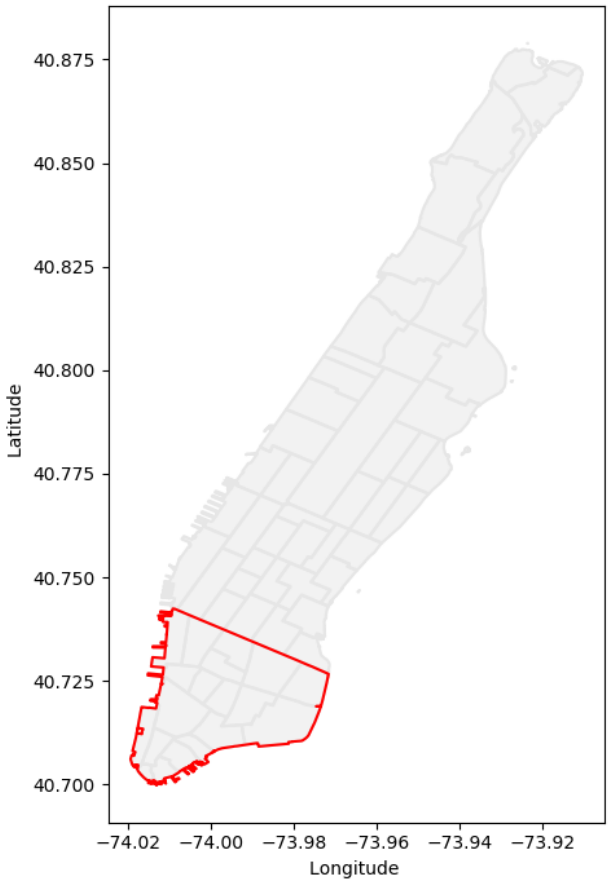
Proof. Let $f(x) = 1 + \zeta(\exp(-x) - x\exp(-x))$, where $x > 0$, $1 \geq \zeta > 0$. We can show that

$$f'(x) = \zeta \exp(-x)(x-2) \tag{P.1}$$

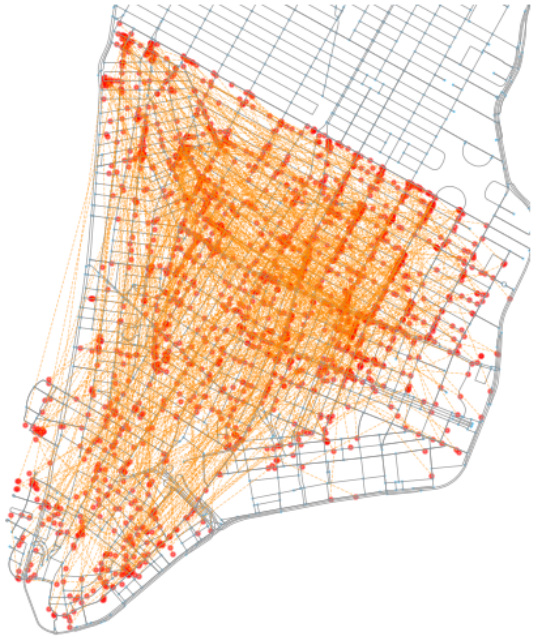
which implies that $f'(x) > 0$ if $x > 2$, and $f'(x) < 0$ if $x < 2$. Therefore, we have $f(x) \geq f(2) > 0$, and $f(x) \leq \max(f(0), f(\infty)) = 1 + \zeta \leq 2$. This implies that $\partial M/\partial N > 0$ and $\frac{\partial M}{\partial N} \leq 1$. This completes the proof. □

Appendix Q. Empirical laws from the downtown area of Manhattan.

To investigate the impact of demand spatial distribution on our empirical law, we conduct our experiments in Manhattan downtown area. The downtown area is shown in Figure 27(a). According to our records, there are about 8.2% trips, namely, 231972 trips in the whole dataset with its origins and destinations in the downtown area of Manhattan. Spatial distribution of these trips is displayed in Figure 27(b). And we re-run our experiments based on the new dataset with similar settings and algorithm in Section 3. Notably, the targeted area is relatively small, we accordingly set a smaller matching radius, i.e. $R = 1$ km, to start our experiment. To examine the validity of empirical law in a more general scope, we also set a larger matching radius, i.e. $R = 4$ km. The results can be found from Figure Q.28 - Q.30, which demonstrate the results from both cases in the downtown area are in line with our main findings. Table Q.13 gives the fitting values of parameters with respect to three corresponding empirical laws.

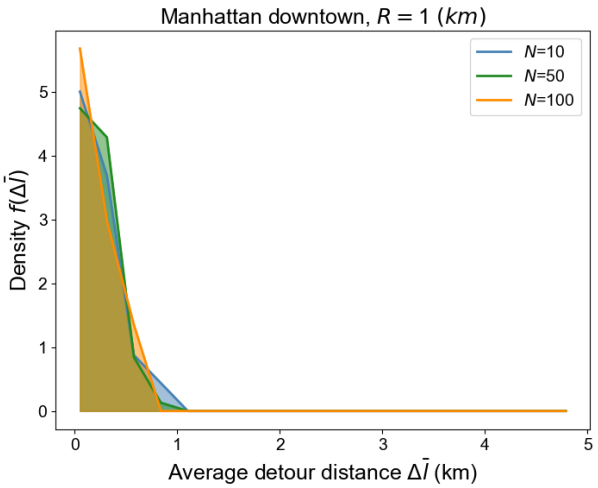


(a) Selected Manhattan downtown area

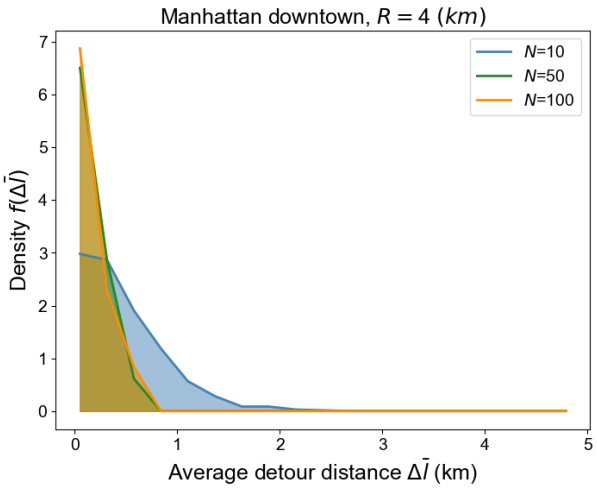


(b) Trips spatio distributions (1000 trips as representative)

Figure Q.27: Downtown area in Manhattan and Spatial distribution of trip origins and destinations

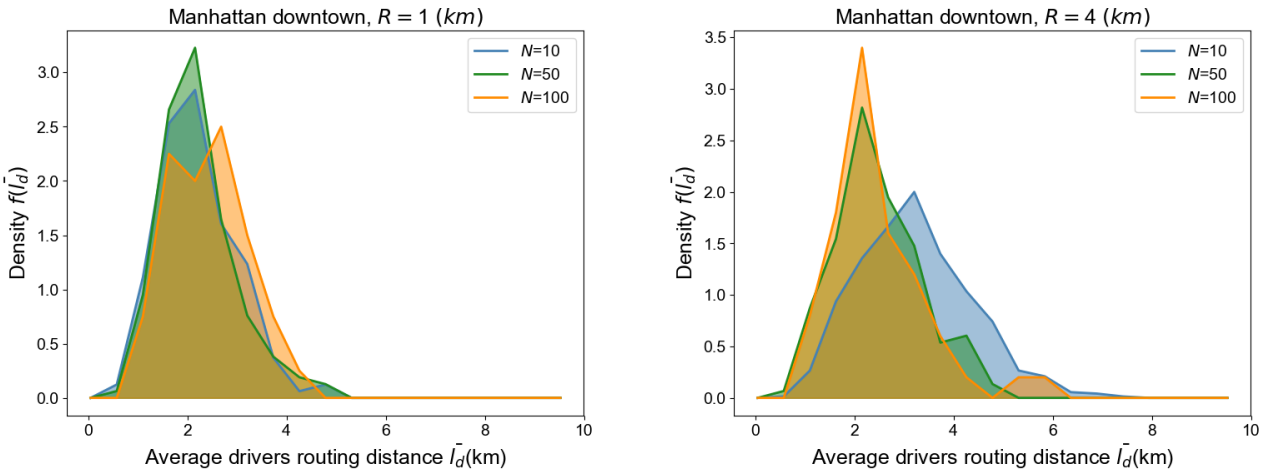


(a) *pdf* of $\Delta\bar{l}$, $R = 1$ km



(b) *pdf* of $\Delta\bar{l}$ km

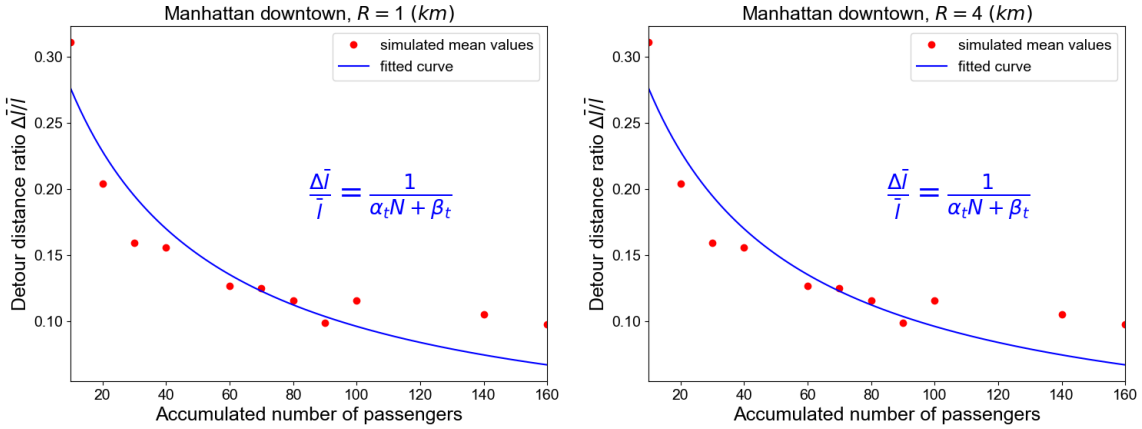
Figure Q.28: Probabilistic density distribution of average passengers' detour distance



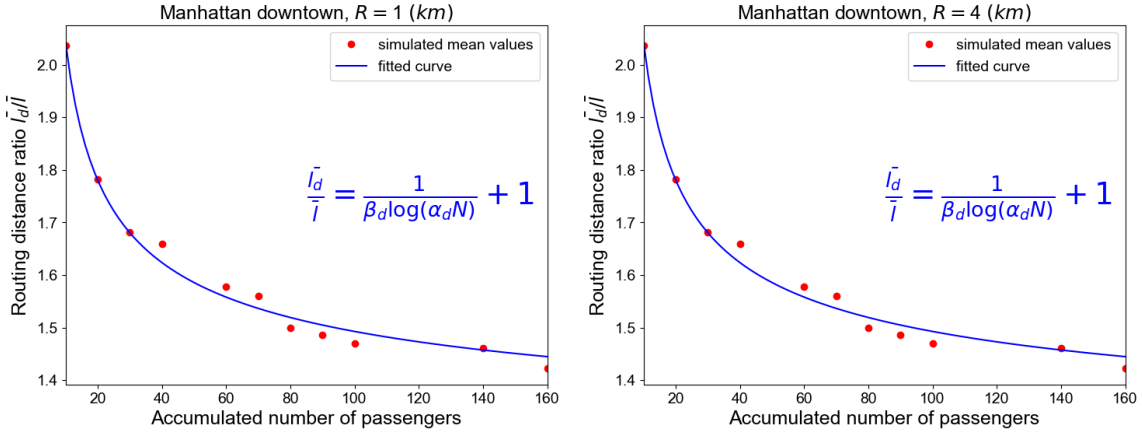
(a) *pdf* of $\bar{l}_d$, $R = 1$ km

(b) *pdf* of $\bar{l}_d$, $R = 4$ km

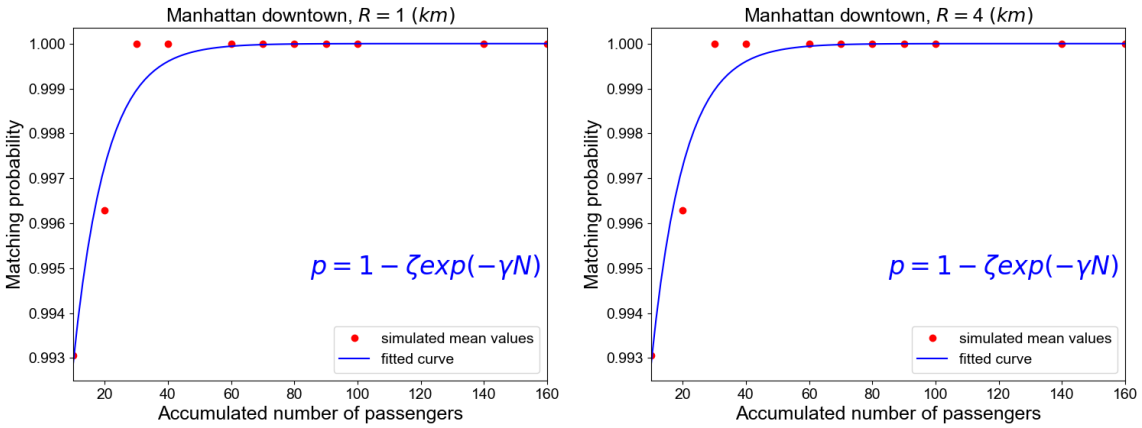
Figure Q.29: Probabilistic density distribution of vehicle routing distance



(a) Empirical fitting between $\Delta\bar{l}/\bar{l}$ v.s. N when $R = 1$km and $R = 4$ km



(b) Empirical fitting between $\bar{l_d}/\bar{l}$ v.s. N when $R = 1$km and $R = 4$ km



(c) Empirical fitting between p v.s. N when $R = 1$km and $R = 4$ km

Figure Q.30: ($P1$) Manhattan downtown area – Empirical fitting under different scenarios with $R = 1$km and $R = 4$km