9-2020

Pricing and equilibrium in on-demand ride-pooling markets

Jintao KE
Hong Kong University of Science and Technology

Hai YANG
Hong Kong University of Science and Technology

Xinwei LI
Beijing University of Aeronautics and Astronautics (Beihang University)

Hai WANG
Singapore Management University, haiwang@smu.edu.sg

Jieping YE
University of Michigan - Ann Arbor

Transportation Research Part B 139 (2020) 411–431

Contents lists available at ScienceDirect

Transportation Research Part B

journal homepage: www.elsevier.com/locate/trb

Pricing and equilibrium in on-demand ride-pooling markets

Jintao Ke[a], Hai Yang[a], Xinwei Li[b], Hai Wang[c,d,*], Jieping Ye[e,f]

[a] Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Hong Kong, China
[b] School of Economics and Management, Beihang University, Beijing, China
[c] School of Information Systems, Singapore Management University, Singapore
[d] Heinz College of Information Systems and Public Policy, Carnegie Mellon University, Pittsburgh, United States
[e] Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, United States
[f] AI Labs, Didi Chuxing, Beijing, China

ARTICLE INFO

Article history:
Received 4 March 2019
Revised 27 June 2020
Accepted 8 July 2020
Available online 24 July 2020

Keywords:
Ride-sourcing
Ride-pooling
Average detour time
Pricing and equilibrium

ABSTRACT

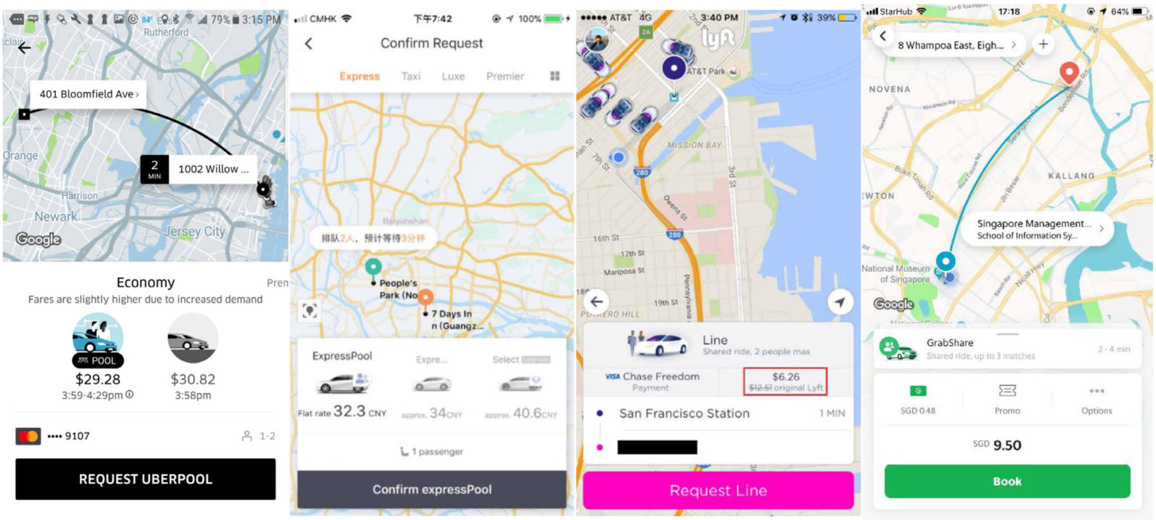
With the recent rapid growth of technology-enabled mobility services, ride-sourcing platforms, such as Uber and DiDi, have launched commercial on-demand ride-pooling programs that allow drivers to serve more than one passenger request in each ride. Without requiring the prearrangement of trip schedules, these programs match on-demand passenger requests with vehicles that have vacant seats. Ride-pooling programs are expected to offer benefits for both individual passengers in the form of cost savings and for society in the form of traffic alleviation and emission reduction. In addition to some exogenous variables and environments for ride-sourcing market, such as city size and population density, three key decisions govern a platform's efficiency for ride-pooling services: trip fare, vehicle fleet size, and allowable detour time. An appropriate discounted fare attracts an adequate number of passengers for ride-pooling, and thus increases the successful pairing rate, while an appropriate allowable detour time prevents passengers from giving up ride-pooling service. This paper develops a mathematical model to elucidate the complex relationships between the variables and decisions involved in a ride-pooling market. We find that the monopoly optimum, social optimum and second-best solutions in both ride-pooling and non-pooling markets are always in a normal regime rather than the *wild goose chase* (WGC) regime—an inefficient equilibrium in which drivers spend substantial time on picking up passengers. Besides, in general, a unit decrease in trip fare in a ride-pooling market attracts more passengers than would a non-pooling market, because it not only directly increases passenger demand due to the negative price elasticity, but also reduces actual detour time, which in turn indirectly increases ride-pooling passenger demand. As a result, we prove that monopoly optimum, social optimum and second-best solution trip fares in a ride-pooling market are lower than that in a non-pooling market under certain conditions. These theoretical findings are further verified by a set of numerical studies.

1. Introduction

Ride-sharing programs offer numerous advantages, including reduced travel costs, energy savings, less traffic congestion, and lower carbon dioxide emissions (Chan and Shaheen, 2012). Ride-sharing dates to as early as World War II, when the US

* Corresponding author at: School of Information Systems, Singapore Management University, Singapore.
E-mail address: haiwang@smu.edu.sg (H. Wang).

https://doi.org/10.1016/j.trb.2020.07.001
0191-2615/



(a) UberPool (b) DiDi ExpressPool (c) Lyft Line (d) GrabShare

Fig. 1. Examples of on-demand ride-pooling programs.

government established a Car-Sharing Club for fuel conservation. In general, ride-sharing requires prearrangement, in which agencies can pair requests that are announced in advance. Traditional methods for ride-sharing include carpooling, vanpooling, dial-a-ride, etc. For example, carpooling was initially introduced by large companies to encourage ride-sharing among their employers during trips to and from work, and has been extensively studied (Ferguson, 1997; Yang and Huang, 1999; Huang et al., 2000; Konishi and Mun, 2010). Dial-a-ride programs employ dedicated drivers to serve prearranged passenger ride requests with diverse origins and destinations (Cordeau and Laporte, 2007). Comprehensive reviews of the ride-sharing have been conducted, for example, by Furuhata et al. (2013) and Ho et al. (2018).

Recent breakthroughs in mobile internet technologies have made on-demand dynamic (and real-time) ride-sharing services possible. In these applications, on-demand ride requests can be matched en route with vehicles that have vacant seats. On-demand dynamic ride-sharing can be provided by a fleet of dedicated drivers (such as taxi drivers or drivers affiliated with transportation network companies). These types of on-demand dynamic ride-sharing programs provided by for-hire dedicated drivers, termed as ride-pooling (or ridesplitting or ridepooling) services in the literature (Shaheen et al., 2015; Chen et al., 2017; Li et al., 2019a; Wang and Yang, 2019), are already available on the major commercial ride-sourcing platforms, such as UberPool, DiDi Express Pool, Lyft Line, and GrabShare (as shown in Fig. 1).

When a passenger launches a ride-sourcing platform application, he or she can select an on-demand ride-pooling service or a non-pooling ride-sourcing service. Normally, a passenger choosing a ride-pooling service pays an up-front discounted fare, which is predetermined and lower than the fare for a non-pooling service. A key concern for platform operators is the probability of en route pairing (successful pairing rate), i.e., the proportion of matched/paired passengers among those who select the ride-pooling option. If successfully paired, passengers may experience a longer trip time than they would with non-pooling service. If the pairing is unsuccessful, the platform suffers a loss of revenue due to the lower predetermined fare with the up-front discount for passengers who opted for ride-pooling. Fig. 2 displays the empirical probability density functions of trip time for ride-pooling service and non-pooling service of for-hire-vehicle ride-sourcing services in New York City.[1] The average trip time of passengers opting for ride-pooling and non-pooling service are 20.63 min and 18.71 min respectively. Clearly, the ride-pooling service has a slightly longer average trip time.

The relationships between the variables and decisions involved in ride-pooling services are complicated: (1) the successful pairing rate depends on the number of passengers opting for on-demand ride-pooling services (i.e., passenger demand) and the allowable detour time; (2) the discounted fare directly affects platform revenue and passenger demand; (3) passenger demand affects the successful pairing rate and, in turn, platform revenue; and (4) the allowable detour time in the ride-pooling service also directly governs the successful pairing rate and affects passenger demand. Intuitively, a larger allowable detour time will increase the system's successful pairing rate on one hand, but increase a passenger's actual detour time and thus decrease passenger demand for ride-pooling on the other hand—which, in turn, decreases the successful pairing rate. A precise understanding of the intricate relationships between the platform decision variables (i.e., trip fare,

[1] The open-source dataset that contains trip records of for-hire-vehicle services (including Uber and Lyft) in New York City is obtained from the New York City Taxi & Limousine Commission. Link: https://www1.nyc.gov/site/tlc/about/tlc-trip-record-data.page. Unfortunately, this dataset does not provide vehicle trajectories and detailed information of the shared rides, thus we cannot calculate the actual detour distance and time for each ride-pooling trip.

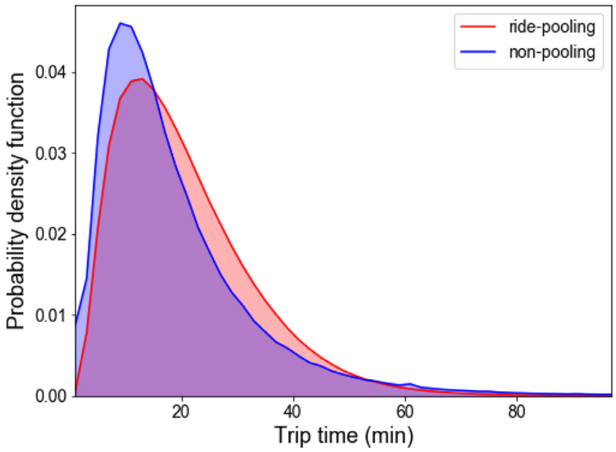


Fig. 2. Distribution of passengers' trip times of ride-pooling and non-pooling for-hire-vehicle services in New York City.

vehicle fleet size and allowable detour time) and the system's endogenous variables (e.g., pick-up time, passenger demand, successful pairing rate and actual detour time) is critical for optimal operating strategy designs.

In this paper, we establish a mathematical model to elucidate the complex relationships between the system's decision variables and endogenous variables in a ride-sourcing market with on-demand ride-pooling services. The reciprocal interactions between passenger demand, successful pairing rate, and actual detour time in equilibrium under certain platform operating strategies are characterized by a system of simultaneous equations. Based on the model, we compare the ride-sourcing markets with ride-pooling service and non-pooling service and examine the impacts of operating strategies on the platform's profit and social welfare. The major contributions of this paper are summarized below:

- We propose a modeling framework to characterize the equilibrium in ride-sourcing markets, in which the vehicles are in one of the three statuses: vacant, picking up, and occupied. The picking up status is what distinguishes ride-sourcing market from the conventional street-hailing taxi market. By spelling out the intriguing relationship among passenger demand, successful pairing rate, and actual detour time, we use the model to describe the equilibrium in an on-demand ride-pooling market.
- We identify and compare the monopoly optimum, social optimum and second-best solutions in the two markets — a non-pooling market and a ride-pooling market and obtain some managerial insights. We investigate the joint impacts of platform decision variables (i.e., trip fare, vehicle fleet size and allowable detour time) on the platform's profit and social welfare, analytically and numerically.
- We find that the monopoly optimum, social optimum and second-best solutions in the two markets are always in the normal regime rather than the *wild goose chase* (WGC) regime, and show that the monopoly optimum, social optimum and second-best solution trip fares in a ride-pooling market are lower than those in a non-pooling market under certain conditions. The reason is that a unit decrease in trip fare in a ride-pooling market attracts more passengers than would in a non-pooling ride-sourcing market due to a reduced actual detour time. These observations are also verified by the numerical studies.

The rest of the paper is organized as follows. Section 2 reviews the relevant studies and differentiates this study from the previous ones. Section 3 establishes a model to describe the stationary equilibrium of a non-pooling ride-sourcing market. Section 4 extends the model to delineate the equilibrium of an on-demand ride-pooling market, in which two passengers can be paired up with a certain probability and experience certain detour time. Section 5 analytically examines the properties of the monopoly and social optimum solutions of both the non-pooling and ride-pooling markets. Section 6 conducts numerical studies to investigate how the platform leverages the key decision variables to achieve maximum platform profit or social welfare. Section 7 summarizes the paper and discusses future research directions.

2. Literature review

This section reviews relevant studies on the ride-sourcing market from the following perspectives: (1) general research on ride-sourcing markets; (2) optimization algorithms for ride-sharing programs served by drivers with their own trip plans; and (3) trip fare and cost-sharing strategies for dynamic ride-sharing services provided by drivers with their own trip plans.

As a typical business model in a sharing economy, ride-sourcing service has been reshaping our mobility and sparked heated discussion since its emergence in 2009. Wang and Yang (2019) provide a general framework and comprehensive review on research problems in ride-sourcing markets. Due to the similarities between the ride-sourcing market and conventional taxi market, supply-demand properties in equilibrium have their roots in research on street-hailing taxi services

(Yang and Yang, 2011; Yang et al., 2010) and e-hailing taxi services (He and Shen, 2015; Wang et al., 2016; He et al., 2018). In contrast to conventional taxi markets, however, which are generally subject to strict entry restriction and price regulation, there is fewer entry restriction for drivers in ride-sourcing markets and less strict regulation on service pricing—i.e., registered private car owners can flexibly decide whether, when, and where to provide ride-sourcing services with dynamically adjusted trip fares (Sun et al. 2019a, 2019b). Other specific research includes the coordination of demand and supply using price and wage (Bai et al., 2018; Taylor, 2018); pricing and surge-pricing strategies (Cachon et al., 2017; Castillo et al., 2017; Zha et al., 2016; Yang et al., 2020b; Chen et al., 2020); government regulations and policies (Yu et al., 2019; Li et al., 2019b); impacts on conventional taxi markets (Nie, 2017; Wallsten, 2015); geometrical matching and order dispatching (Xu et al., 2015, 2018; Zha et al., 2018; Zhang et al., 2017; Lyu et al., 2019; Ke et al., 2020; Yang et al., 2018, 2020a); driver labor supply (Zha et al., 2017); supply and demand predictions (Ke et al., 2017, 2019c; Tong et al., 2017); electrified ride-sourcing vehicles (Ke et al., 2019a); and multi-modal transportation with ride-sourcing and public transit services (Zhu et al., 2020).

Of the research issues above, surge pricing is of particular interest, as it is considered to be an efficient method for dynamically coordinating supply-demand balance. For example, Castillo et al. (2017) point out that a static price scheme may lead to a “wild goose chase” (WGC) during periods when the platform is depleted of available nearby vehicles and forced to match drivers with distant passengers. Based on both theoretical analysis and real-world data, they argue that WGC could be avoided by implementing surge pricing. Chen and Sheldon (2016) and Sun et al. (2019a) find significant evidence using real data that surge pricing/wage incentivizes drivers to adjust their work schedules to align with periods of high demand (as indicated by surge prices). By establishing a time-expanded network to coordinate surge pricing and classical labor supply hypotheses, Zha et al. (2017) show that both the platform and its drivers benefit from surge pricing, while passengers may be worse off during high surge periods. Based on a queueing model with endogenous supply and demand, Bai et al. (2018) find that if potential customer demand is large, the platform should charge passengers a high price, pay drivers a high wage, and implement a high payout ratio (the ratio of wage over price). Taylor (2018) examines the impacts of two important features of a ride-sourcing market—delay sensitivity and agent independence—on a platform’s optimal strategies in terms of price and wage. Yang et al. (2020b) propose a novel reward scheme integrated with surge pricing and find that in some situations, passengers, drivers, and the platform will be better off under the reward scheme.

Thanks to the rapid growth of mobile technologies, ride-sharing services are now able to accommodate on-demand dynamic requests and no longer require users to schedule their routes in advance (Furuhata et al., 2013). Primary efforts thus far have been directed towards designing algorithms to efficiently match drivers and riders on short notice in a dynamic ride-sharing environment. Agatz et al. (2011) develop optimization methods to minimize total system travel distance and individual riders’ travel cost in an on-demand ride-sharing program. Verified by a simulation in metropolitan Atlanta, they show that sophisticated optimization approaches significantly improve the performance of ride-sharing systems when compared to simple greedy matching rules. With appropriate matching algorithms, an on-demand ride-sharing program could even be successfully implemented in relatively sprawling urban areas. Agatz et al. (2012) systematically outline major concerns and challenges that on-demand ride-sharing programs face. One of the most important components is the real time matching. A good matching strategy reduces system-wide vehicle miles and travel times and increases the number of participants to provide the most in terms of societal and environmental benefits. They also note that a good understanding of riders’ behavioral preferences and mode choices is essential to the success of an on-demand ride-sharing program. More recently, Wang et al. (2018a) develop a stable matching algorithm to minimize the total travel distance of all potential participants, either in a successfully paired ride or an unsuccessfully paired ride. The method can greatly increase the stability of ride-sharing at the cost of only slightly reducing system-wide performance. Stiglic et al. (2015) assess the potential benefits of meeting points in on-demand ride-sharing systems through extensive simulation studies, and Stiglic et al. (2016) quantify the effects of flexibility for different participants on the performance of an on-demand ride-sharing program. Lee and Savelsbergh (2015) point out that the introduction of meeting points can significantly improve the number of matched participants and reduce the total driving distance in an on-demand ride-sharing system.

Several recent studies also assess the impacts of on-demand ride-sharing on the minimum fleet size required to serve passengers. Alonso-Mora et al. (2017) propose a general mathematical model that enables real-time high-capacity ride-sharing on the shareability network or “vehicle-share-networks” that have been examined by Santi et al. (2014), Sagarra et al. (2015), Tachet et al. (2017), and Vazifeh et al. (2018). Based on simulation experiments using New York City taxi data, the authors show that only 3000 high-capacity taxis were needed to serve 98% of the taxi rides that were currently being served by over 13,000 taxis.

Some recent studies have investigated the impact of trip fare for on-demand ride-sharing programs and the possibility of cost sharing between riders and drivers. Xu et al. (2015) study the endogenous interactions between traffic congestion, ride-sharing trip fare, and passengers’ route choices in a framework that combines classical Wardrop network equilibrium model and ride-sharing passenger demand. Di et al. (2017, 2018) examine ride-sharing in the context of an equilibrium-based network design problem. Notably, Wang et al. (2018b) develop an equilibrium model to describe the interactions between riders’ and drivers’ mode choices, costs, and matching probability. They consider a single-corridor network in which car owners choose to be solo or share cars with other riders, and non-car owners have two alternatives: ride-sharing or public transit. They investigate the properties of cost-sharing strategies to prevent mode shifts among transit users to autos and/or reduce vehicular traffic, and find that a suitable cost-sharing strategy is crucial for encouraging riders to use ride-sharing modes. There is much relevant research into the pricing and trip fare in shared transportaton under diverse settings, such as pricing for on-demand last-mile transportation (Chen and Wang, 2018a, 2018b) and dial-a-ride system (Sayarshad and Chow, 2015).

As described previously, however, most prior studies have focused on ride-sharing services provided by individual drivers who have their own trip plans. Few efforts have been made to understand the emerging dynamics of on-demand ride-pooling services provided by dedicated drivers affiliated with ride-sourcing companies, and in particular, the impacts of key platform decision variables (i.e., trip fare, vehicle fleet size, and allowable detour time) on the platform's profit and social welfare.

3. Equilibrium in a non-pooling ride-sourcing market

Ride-sourcing companies provide two major types of services: on-demand ride-pooling service, denoted as RP service; on-demand non-pooling service, denoted as NP service. For analytical tractability to obtain managerial insights, we delineate and compare the equilibrium of two markets: (a) a non-pooling ride-sourcing market (abbreviated as non-pooling market) in which a ride-sourcing platform provides NP service and (b) a ride-pooling market in which a ride-sourcing platform provides RP service. In this section, we propose a modeling framework to characterize the equilibrium of the non-pooling market. Notation in this paper is summarized in Appendix A for the convenience of the readers.

A few basic assumptions are worth noting here. First, we adopt the assumption in Castillo et al. (2017) that the platform matches passengers and vehicles based on a First-Come-First-Serve (FCFS) strategy. Second, the congestion externality (Yang et al., 2005) caused by both ride-sourcing vehicles and background traffic is not considered. Third, the model investigates the stationary equilibrium of the ride-sourcing markets in an aggregate context without considering network structures and dynamic time-varying operations.

We first consider a stationary equilibrium in which each ride-sourcing vehicle serves one passenger (one ride request) in the non-pooling market. As mentioned above, we assume that passengers are matched in sequence with the closest vacant vehicle according to the FCFS mechanism. Let w and t_{np} denote the average pick-up time (i.e., waiting time from being matched to being picked-up) and the average trip time (i.e., riding time from being picked-up to being dropped-off, which is assumed to be a constant in non-pooling services). Let F denote the average trip fare, then the generalized cost of a non-pooling ride-sourcing trip is $F + \beta(w + t_{np})$, where β is the value of time. We assume that passenger demand (i.e., the number of passengers per unit time) for non-pooling ride-sourcing services, denoted by Q, is a strictly decreasing function with respect to the generalized cost:

$$Q = f(F + \beta \cdot (w + t_{np})) \tag{1}$$

where $f'(\bullet) < 0$. Let N denote the total number of vehicles (i.e., vehicle fleet size) on the platform. Each vehicle can be in one of three statuses (Castillo et al., 2017): vacant (idle to be matched), picking up (on the way to pick up a passenger), and occupied (with passenger(s) onboard). Let N^v denote the number of vacant vehicles in stationary equilibrium, then the conservation equation of vehicles is given by

$$N = N^v + Q \cdot w\left(N^v\right) + Q \cdot t_{np} \tag{2}$$

where the average pick-up time w is inversely proportional to the number of vacant vehicles N^v, i.e., $w = w(N^v)$ with $w' = \partial w / \partial N^v < 0$. Then Eq. (2) can be written as:

$$Q = \frac{N - N^v}{w(N^v) + t_{np}} \tag{3}$$

which shows that passenger demand Q can be written as an explicit function of the number of vacant vehicles N^v. Taking the partial derivative of Q with respect to N^v gives rise to:

$$\frac{\partial Q}{\partial N^v} = \frac{-\left(Qw' + 1\right)}{w + t_{np}} \tag{4}$$

Conversely, N^v is also an explicit function of Q. By taking the partial derivative of both sides of Eq. (2) with respect to Q, we can obtain:

$$\frac{\partial N^v}{\partial Q} = \frac{-(w + t_{np})}{Qw' + 1} \tag{5}$$

where $Qw' < 0$, which indicates that the sign of $\partial Q / \partial N^v$ and $\partial Q / \partial N^v$ are undetermined. If $Qw' + 1 < 0$, then Q strictly increases with N^v and also N^v strictly increases with Q, which indicates a WGC regime. Initially identified by Castillo et al. (2017), the WGC is an inefficient outcome of the ride-sourcing system with extremely low density of vacant vehicles and a large proportion of vehicles wasted in the picking-up phase. If $Qw' + 1 > 0$, then Q strictly decreases with N^v and also N^v strictly decreases with Q, which indicates a normal regime. It is noteworthy to mention that, in the conventional street-hailing taxi market, the pick-up phase can be ignored in the vehicle conservation equation, which then becomes $N = N^v + Q \cdot t_{np}$. In this case, $\partial N^v / \partial Q = -t_{np} < 0$, which implies that the conventional street-hailing taxi market always falls into the normal regime. Combining Eqs. (1) and (2), we can obtain the market equilibrium of the non-pooling market as follows:

$$Q = \frac{N - N^v}{w(N^v) + t_{np}} = f\Big(F + \beta \cdot \Big(w\Big(N^v\Big) + t_{np}\Big)\Big) \tag{6}$$

which is an implicit function of N^v. Taking the partial derivative of both sides of Eq. (6) with respect to the two decision variables F and N gives rise to:

$$\frac{\partial N^v}{\partial F} = \frac{-f'(w + t_{np})}{(Qw' + 1) + f'\beta w'(w + t_{np})} \tag{7}$$

$$\frac{\partial N^v}{\partial N} = \frac{1}{(Qw' + 1) + f'\beta w'(w + t_{np})} \tag{8}$$

As mentioned above, passenger demand Q can be written as an explicit function of N^v in Eq. (3), and thus the partial derivatives of Q with respect to the two decision variables F and N are given by

$$\frac{\partial Q}{\partial F} = \frac{\partial Q}{\partial N^v} \cdot \frac{\partial N^v}{\partial F} = \frac{f'\Big(Qw' + 1\Big)}{(Qw' + 1) + f'\beta w'(w + t_{np})} \tag{9}$$

$$\frac{\partial Q}{\partial N} = \frac{1}{w + t_{np}} + \frac{\partial Q}{\partial N^v} \cdot \frac{\partial N^v}{\partial N} = \frac{f'\beta w'}{(Qw' + 1) + f'\beta w'(w + t_{np})} \tag{10}$$

In addition, since the average pick-up time w is a decreasing function of N^v, the partial derivatives of w with respect to the two decision variables F and N are given by

$$\frac{\partial w}{\partial F} = \frac{\partial w}{\partial N^v} \cdot \frac{\partial N^v}{\partial F} = \frac{-f'w'(w + t_{np})}{(Qw' + 1) + f'\beta w'(w + t_{np})} \tag{11}$$

$$\frac{\partial w}{\partial N} = \frac{\partial w}{\partial N^v} \cdot \frac{\partial N^v}{\partial N} = \frac{w'}{(Qw' + 1) + f'\beta w'(w + t_{np})} \tag{12}$$

In the normal regime with $Qw' + 1 > 0$, passenger demand Q increases with vehicle fleet size and decreases with trip fare (i.e., $\partial Q/\partial N > 0$, $\partial Q/\partial F < 0$), the number of vacant vehicles N^v increases with both vehicle fleet size and trip fare (i.e., $\partial N^v/\partial N > 0$, $\partial N^v/\partial F > 0$), and the average pick-up time w decreases with both vehicle fleet size and trip fare (i.e., $\partial w/\partial N < 0$, $\partial w/\partial F < 0$). Yet, these monotonic properties do not necessarily hold in the WGC regime with $Qw' + 1 < 0$.

4. Equilibrium in a ride-pooling market

In this section, we extend the model in Section 3 to delineate the stationary equilibrium in a ride-pooling market in which ride-sourcing vehicles provide ride-pooling services. Due to the complexity of urban spatial topology and randomness of passengers' origins and destinations, a passenger opting for ride-pooling service may end up paired or unpaired with a second passenger. The major difference between ride-pooling market and non-pooling ride-sourcing market lies in the successful pairing rate (i.e., the fraction of passengers who are successfully paired with other passengers who opt for ride-pooling services) and the actual detour time (i.e., the actual extra trip time experienced by passengers who choose ride-pooling services and are successfully paired). These two factors are endogenously correlated with the average trip time and pick-up time, and affect passenger demand, the platform's profit and social welfare. To simplify the model, we consider pairing at most two passengers (two ride requests) although three or more passengers can be pooled, and for analytical tractability, all passengers are assumed to opt for RP service, although they may end up with unpaired.

4.1. Average trip time for ride-pooling

Let p denote the successful pairing rate, Δt denote the average actual detour time, and Δt_A denote the allowable detour time (i.e., the allowable maximum extra detour time experienced by passengers who choose ride-pooling services and are successfully paired). Note that Δt_A is a decision variable of the platform while Δt endogenously depends on many other factors. In this paper, if the allowable detour time is unlimited, i.e., $\Delta t_A \to \infty$, the situation is termed as a *detour-unconstrained scenario*; otherwise, it is termed as a *detour-constrained scenario.*

If successfully paired, passengers who opt for RP service will experience an average trip time that equals the sum of the average trip time for NP service (i.e., t_{np}) and the average actual detour time (i.e., Δt); otherwise, as in NP service, it equals the average trip time only. Thus, the expected average trip time of passengers opting for the ride-pooling mode is

$$t_{rp} = (t_{np} + \Delta t) \cdot p + t_{np} \cdot (1 - p), \tag{13}$$

where t_{np} and $t_{np} + \Delta t$ are the average trip times of unsuccessfully and successfully paired passengers, respectively, and p is the successful pairing rate. It is also worth mentioning that, Li et al. (2019a) point out that trip time reliability of ride-pooling services is much worse than that of non-pooling services due to uncertain extra detour and waiting. For tractability, we ignore the impact of the standard deviation of trip time on the passenger demand in this paper and leave it for future studies.

Rate p depends on two key factors: passenger demand for RP service Q and allowable detour time Δt_A. Intuitively, the more the passengers who opt for RP services, the higher the success rate for pairing two requests with similar routes and schedules, i.e., $p \to 1$ as $Q \to \infty$ and $p \to 0$ as $Q \to 0$. Also, the longer the allowable detour time Δt_A, the more requests that can be paired, and hence the higher the pairing rate, i.e., $p \to 0$ as $\Delta t_A \to 0$ and $p \to 1$ as $\Delta t_A \to \infty$. For a given allowable detour time, Santi et al. (2014) demonstrate that the curve of the successful pairing rate against ride-pooling passenger demand resembles a "fast" saturation process. In other words, the successful pairing rate first increases quickly and then slowly approaches 1.0, as passenger demand for sharing increases.

The average actual detour time, Δt, is also an endogenous variable that depends on Q and Δt_A, and in general increases with Δt_A and decreases with Q. Moreover, $\Delta t \to 0$ as $\Delta t_A \to 0$ and $\Delta t \to 0$ as $Q \to \infty$. Specifically, in the detour-unconstrained scenario with $\Delta t_A \to \infty$, the successful paring rate $p \to 1$ and the average detour time Δt should be monotonically decreasing with Q (i.e., more ride-pooling passengers imply better pairings with shorter actual detour time). In this case, with a slight abuse of notation, the average detour time can be written as a function on passenger demand, i.e., $\Delta t = \Delta t(Q)$, where $\partial \Delta t / \partial Q < 0$. To summarize, p and Δt can be written as functions of Q and Δt_A, i.e., $p = p(Q, \Delta t_A)$ and $\Delta t = \Delta t(Q, \Delta t_A)$, with the following mild assumptions:

Assumption 1. *The successful pairing rate $p(Q, \Delta t_A)$ and average detour time $\Delta t(Q, \Delta t_A)$ satisfy:*

1 *For all $Q \geq 0$, $p(Q, \Delta t_A)$ strictly increases in Δt_A with $p(Q, 0) = 0$ and $\lim_{\Delta t_A \to \infty} p(Q, \Delta t_A) = 1$; for all $\Delta t_A \geq 0$, $p(Q, \Delta t_A)$ strictly increases in Q with $p(0, \Delta t_A) = 0$ and $\lim_{Q \to \infty} p(Q, \Delta t_A) = 1$.*

2 *For all $Q \geq 0$, $\Delta t(Q, \Delta t_A)$ strictly increases in Δt_A with $\Delta t(Q, 0) = 0$; for all $\Delta t_A \geq 0$, $\Delta t(Q, \Delta t_A)$ strictly decreases in Q with $\lim_{Q \to \infty} \Delta t(Q, \Delta t_A) = 0$; if $\Delta t_A \to \infty$, Δt is a decreasing function of Q.*

With some mild assumptions, we propose a probabilistic model in Appendix B to investigate how the passenger demand Q and allowable detour time Δt_A jointly affect the successful pairing rate p and average actual detour time Δt. The probabilistic model is then used for numerical studies in Section 6.

4.2. Market equilibrium

In the ride-pooling market, the generalized cost of passengers opting for ride-pooling services is $F + \beta \cdot (w + t_{rp})$, where $t_{rp} = t_{np} + p\Delta t$ given by Eq. (13). Similar to the non-pooling market, passenger demand Q can be written as a decreasing function of the generalized cost:

$$Q = f(F + \beta \cdot (w + t_{np} + p\Delta t)) \tag{14}$$

where $f' < 0$. Particularly, in the detour-unconstrained scenario, the average trip time for ride-pooling $t_{np} + p\Delta t$ becomes $t_{np} + \Delta t$. Different from the non-pooling market, a vehicle in a ride-pooling market can be dispatched to an unpaired passenger or two paired passengers. In the former case, the number of vehicles in the pick-up phase and in-trip phase is $(1 - p)Q[t_{np} + w(N^v)]$. In the latter case, each vehicle corresponds to two passengers, and thus the number of vehicles in the pick-up phase and in-trip phase can be estimated by $\frac{1}{2}pQ[t_{np} + \Delta t_d + w(N^v)]$, where Δt_d is the average driver detour time, i.e., the extra detour time a driver experiences in a shared ride serving two requests in comparison with a normal ride serving one request. Therefore, the vehicle conservation function in the ride-pooling market is given by

$$N = N^v + \frac{1}{2}pQ\left[t_{np} + \Delta t_d + w\left(N^v\right)\right] + (1 - p)Q\left[t_{np} + w\left(N^v\right)\right] \tag{15}$$

Next, we use realistic examples to discuss Δt_d and Δt and how they are correlated. As shown in Fig. 3, there are two possible route sequences in a shared ride. Suppose a driver picks-up passenger i first and passenger j second, then he/she has two route sequences: (1) *first-pickup-last-dropoff*, i.e., first drops off passenger j and then passenger i in Fig. 3(a); (2) *first-pickup-first-dropoff*, i.e., first drops off passenger i and then passenger j in Fig. 3(b). Let t_i, t_j denote the normal trip time for passengers i and j without ride-pooling, and let $t_{1(i,j)}$, $t_{2(i,j)}$, $t_{3(i,j)}$ denote the trip time for the three consecutive segments in the shared ride. In the *first-pickup-last-dropoff* case, the detour time experienced by passengers i and j is $t_{1(i,j)} + t_{2(i,j)} + t_{3(i,j)} - t_i$ and $t_{2(i,j)} - t_j = 0$ respectively, while the driver detour time is the difference between the total trip time $t_{1(i,j)} + t_{2(i,j)} + t_{3(i,j)}$ and the average normal trip time $(t_i + t_j)/2$. In the *first-pickup-first-dropoff* case, the detour time experienced by passengers i and j is $t_{1(i,j)} + t_{2(i,j)} - t_i$ and $t_{2(i,j)} + t_{3(i,j)} - t_j$ respectively, while the driver detour time is still $t_{1(i,j)} + t_{2(i,j)} + t_{3(i,j)} - (t_i + t_j)/2$. Then the average detour time Δt experienced by passenger and the average detour time Δt_d experienced by driver can be estimated by:

$$\Delta t = \frac{t_{1(i,j)} + t_{2(i,j)} - t_i + t_{2(i,j)} + t_{3(i,j)} - t_j}{2} \tag{16}$$

$$\Delta t_d = t_{1(i,j)} + t_{2(i,j)} + t_{3(i,j)} - \frac{t_i + t_j}{2} \tag{17}$$

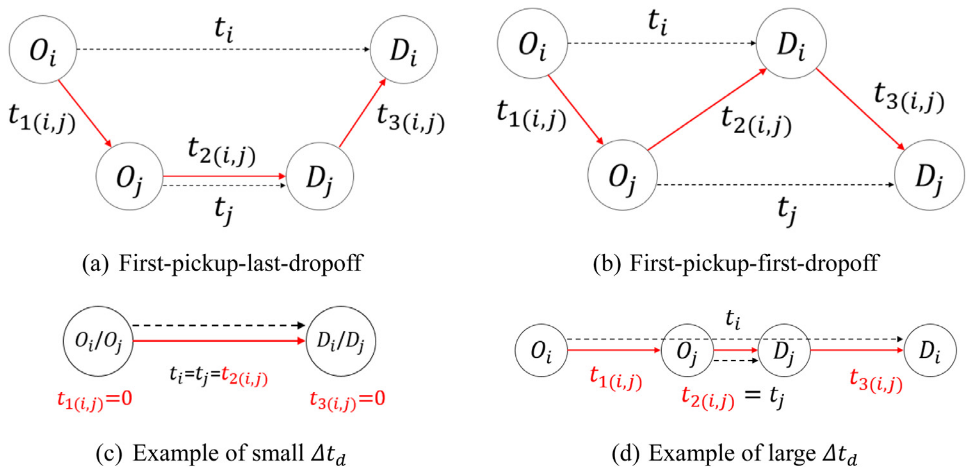


Fig. 3. Ride-pooling route sequences.

From the two equations above, we have:

$$\Delta t_d - 2\Delta t = \frac{t_i + t_j}{2} - t_{2(i,j)} \tag{18}$$

where the first element on the RHS is an estimate for the average normal trip time, and the second element on the RHS refers to the average trip time for the second segment, i.e., the "shared" segment. This indicates that $\Delta t_d - 2\Delta t$ is proportional to the difference between the average normal trip time and the average trip time for the shared segment. The average trip time for the shared segment is hard to ascertain and depends on many factors, including the matching algorithms and network structures, which make the exact identification of the relationship between Δt_d and Δt intractable. For example, in simplified cases where the travel directions of passengers i and j are exactly the same in Fig. 3(c) and (d), passengers do not experience detour time, while the driver detour time is given by $t_{1(i,j)} + t_{2(i,j)} + t_{3(i,j)} - (t_i + t_j)/2$. If both the origins and destinations of the two passengers are close, the driver detour time is almost zero as in Fig. 3(c); if the origins and destinations of the two passengers are far from each other, the shared segment is short and the driver detour time will be substantially larger than zero as in Fig. 3(d). To summarize, the exact relationship between Δt_d and Δt is difficult to determine, which requires further explorations from both theoretical and empirical perspectives. Yet, we can generally expect that Δt_d increases with Δt. For simplicity in this paper, we assume $\Delta t_d = \gamma \Delta t$ in the analysis, where γ is an exogenous positive parameter.

The equilibrium of the ride-pooling market can be given by solving a system of simultaneous equations that consists of Eqs. (14) and (15). Particularly, in the detour-unconstrained scenario in which $\Delta t_A \to \infty$, $p \to 1$, and $\Delta t = \Delta t(Q)$ with $\partial \Delta t / \partial Q < 0$, all passengers can be successfully paired, thus the stationary equilibrium of the ride-pooling market can be simplified in the following system of nonlinear equations:

$$Q = f(F + \beta \cdot (w + t_{np} + \Delta t)) \tag{19}$$

$$N = N^v + \frac{1}{2} Q \left[t_{np} + \gamma \Delta t + w\left(N^v\right) \right] \tag{20}$$

where Eq. (19) depicts the demand curve and Eq. (20) describes the supply curve. The intersection of the demand and supply curves gives the equilibrium. As we can see from the supply curve, without a specific form of the average detour time Δt, we cannot obtain the passenger demand Q as an explicit function of the number of vacant vehicles N^v, which makes it intractable to investigate the effects of decision variables on the endogenous variables, such as Q, N^v and w. A recent paper, Ke et al., (2020a), find that the average detour time Δt is inversely proportional to the passenger demand through extensive experiments using actual data from Manhattan, Chengdu and Haikou. For analytical tractability, we follow their findings and assume $\Delta t = A/Q$, where A is a parameter. Note that this formula satisfies the properties of Δt in the probabilistic model described in Appendix B. Combing $\Delta t = A/Q$ and Eq. (20), we can obtain Q as an explicit function of N^v as follows:

$$N = N^v + \frac{1}{2} Q \left[t_{np} + w\left(N^v\right) \right] + \frac{1}{2} \gamma A \tag{21}$$

or equivalently,

$$Q = \frac{2\left(N - \frac{1}{2}\gamma A - N^v\right)}{t_{np} + w(N^v)} \tag{22}$$

The partial derivative of Q with respect to N^v is given by

$$\frac{\partial Q}{\partial N^v} = \frac{-\left(Qw' + 2\right)}{w + t_{np}} \tag{23}$$

Conversely, the partial derivative of N^v with respect to Q is given by

$$\frac{\partial N^v}{\partial Q} = \frac{-(w + t_{np})}{Qw' + 2} \tag{24}$$

Clearly, the signs of $\partial Q/\partial N^v$ and $\partial N^v/\partial Q$ are undetermined. If $Qw' + 2 < 0$, the market is in the WGC regime and the number of vacant vehicles N^v increases with passenger demand Q; otherwise, the market is in the normal regime and N^v decreases with Q. Note that the condition for the WGC regime in the non-pooling and ride-pooling markets is $Qw' < -1$ and $Qw' < -2$, respectively. Next we look into the equilibrium in the ride-pooling market by combining Eqs. (19) and (21):

$$Q = \frac{2\left(N - \frac{1}{2}\gamma A - N^v\right)}{w(N^v) + t_{np}} = f\left(F + \beta \cdot \left(w\left(N^v\right) + t_{np} + \Delta t\right)\right) \tag{25}$$

which is an implicit function of N^v. Taking the partial derivative of both sides of Eq. (25) with respect to the two decision variables F and N gives rise to:

$$\frac{\partial N^v}{\partial F} = \frac{-f'(w + t_{np})}{(Qw' + 2)\left(1 - \beta f' \frac{\partial \Delta t}{\partial Q}\right) + f'\beta w'(w + t_{np})} \tag{26}$$

$$\frac{\partial N^v}{\partial N} = \frac{-2\left(1 - \beta f' \frac{\partial \Delta t}{\partial Q}\right)}{(w + t_{np})\left[\frac{\partial Q}{\partial N^v}\left(1 - \beta f' \frac{\partial \Delta t}{\partial Q}\right) - f'\beta w'\right]} \tag{27}$$

Since passenger demand Q can be written as an explicit function of N^v in Eq. (21), we can derive the partial derivatives of Q with respect to the two decision variables F and N as follows:

$$\frac{\partial Q}{\partial F} = \frac{\partial Q}{\partial N^v} \cdot \frac{\partial N^v}{\partial F} = \frac{f'\left(Qw' + 2\right)}{(Qw' + 2)\left(1 - \beta f' \frac{\partial \Delta t}{\partial Q}\right) + f'\beta w'(w + t_{np})} \tag{28}$$

$$\frac{\partial Q}{\partial N} = \frac{2}{t_{np} + w} + \frac{\partial Q}{\partial N^v} \cdot \frac{\partial N^v}{\partial N} = \frac{2f'\beta w'}{(Qw' + 2)\left(1 - \beta f' \frac{\partial \Delta t}{\partial Q}\right) + f'\beta w'(w + t_{np})} \tag{29}$$

Since the average pick-up time w is a decreasing function of N^v, the partial derivatives of w with respect to the two decision variables F and N are given by

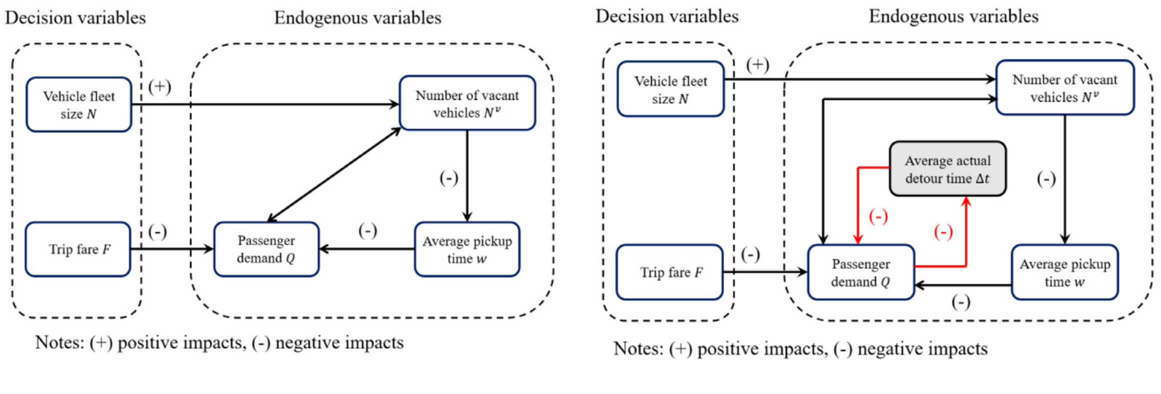
$$\frac{\partial w}{\partial F} = \frac{\partial w}{\partial N^v} \cdot \frac{\partial N^v}{\partial F} = \frac{-f'w'(w + t_{np})}{(Qw' + 2)\left(1 - \beta f' \frac{\partial \Delta t}{\partial Q}\right) + f'\beta w'(w + t_{np})} \tag{30}$$

$$\frac{\partial w}{\partial N} = \frac{\partial w}{\partial N^v} \cdot \frac{\partial N^v}{\partial N} = \frac{2\left(1 - \beta f' \frac{\partial \Delta t}{\partial Q}\right)w'}{(Qw' + 2)\left(1 - \beta f' \frac{\partial \Delta t}{\partial Q}\right) + f'\beta w'(w + t_{np})} \tag{31}$$

In contrast to the non-pooling market, the signs of $\partial Q/\partial N$, $\partial Q/\partial F$, $\partial N^v/\partial N$, $\partial N^v/\partial F$, $\partial w/\partial N$ and $\partial w/\partial F$ are undetermined and dependent on the signs of $Qw' + 2$ and $1 - \beta f' \frac{\partial \Delta t}{\partial Q}$. Fig. 4 illustrates the complicated relationships among decisions and endogenous variables in the two markets, i.e., non-pooling market and ride-pooling market. In the non-pooling market as shown in Fig. 4(a), passenger demand Q and the number of vacant vehicles N^v interacts with each other. Q decreases with N^v in the normal regime but increases with N^v in the WGC regime. The average pick-up time w decreases with N^v, and Q decreases with w. Therefore, the three endogenous variables Q, N^v and w form a cycle leading to a market equilibrium. The decision trip fare F and vehicle fleet size N influence the equilibrium through passenger demand Q and vacant vehicles N^v. In the ride-pooling market as shown in Fig. 4(b), the three endogenous variables Q, N^v and w also form a similar cycle. Besides, a unit increase in passenger demand Q brings an additional indirect effect on itself: it reduces the average actual detour time Δt, which in turn increases Q. Therefore, intuitively, reducing the trip fare in the ride-pooling market brings greater marginal increase in passenger demand, and thus the platform is more willing to charge a lower trip fare.

It is interesting to find that the formulations of Eqs. (26)-(31) and Eqs. (7)-(12) are similar except that: a) $Qw' + 2$ replaces $Qw' + 1$ in the ride-pooling market, which indicates that ride-pooling market has some passengers sharing vehicles; b) the ride-pooling market has an additional term $-\beta f' \frac{\partial \Delta t}{\partial Q}$, which represents the additional indirect effect on passenger demand Q through average actual detour time Δt and corresponds to the red cycle in Fig. 4.

Our model is different from Castillo et al. (2017)'s model in the following aspects. First, their model does not consider the extra detour time experienced by passengers and drivers, and thus its drivers' service rate in the ride-pooling market is assumed to be exactly twice of that in the non-pooling market. In our model, drivers' service rate is lower and more realistic



(a) Non-pooling market (b) Ride-pooling market

Fig. 4. Relationships between decisions and endogenous variables in two markets.

by considering a driver detour time $\gamma\Delta t$ in Eq. (20). Second, their model assumes that the average pick-up time in the ride-pooling market depends on the number of available vehicles (including vacant vehicles and vehicles serving 1 passenger), while our model assumes that the average pick-up time depends on the number of vacant vehicles. We then derive the monopoly and social optimum conditions in the detour-unconstrained scenario and obtain some theoretical insights (as shown in the next section).

5. Properties of optimal solutions

In this section we compare the two markets described in Sections 3 and 4 by examining the properties of their optimal solutions under three scenarios: (1) a monopoly scenario in which a monopoly platform aims to maximize its profit; (2) a social optimum scenario in which the platform aims to maximize social welfare without profit constraint; and (3) a second-best scenario in which a second-best solution is sought out to maximize social welfare while guaranteeing a certain level of platform profit.

5.1. Monopoly optimum (MO) in the non-pooling market

In the monopoly scenario in the non-pooling market, the ride-sourcing platform aims to maximize its profit by determining trip fare F and vehicle fleet size N. This is a typical market that has been examined by, for example, Zha et al. (2016) and Yang and Yang (2011). Specifically, Zha et al. (2016) argue that the ride-sourcing platform behaves like a conventional taxi company in terms of having an objective of maximizing its revenue without possessing any vehicles, under the following conditions: drivers' entry to the market is free, drivers' reservation rates are homogenous, labor supply is sufficient such that drivers will enter the ride-sourcing market until reaching zero net earnings. Under these conditions, the revenue-maximizing problem for a ride-sourcing platform is the same as the revenue-maximizing problem for a monopoly street-hailing taxi market examined in Yang and Yang (2011). Therefore, the problem is formulated as follows:

$$(\text{P1})\quad \max \Pi(F,N) = FQ - cN \tag{32}$$

where Π is the profit of the ride-sourcing platform and c is the unit time operating cost per ride-sourcing vehicle. The profit equals the total revenue, FQ, minus the total payment to the drivers, which equals cN since drivers' net profit is zero in a free entry market. The first-order conditions of P1 are:

$$\frac{\partial \Pi}{\partial F} = Q + F\frac{\partial Q}{\partial F} = 0 \tag{33}$$

$$\frac{\partial \Pi}{\partial N} = F\frac{\partial Q}{\partial N} - c = 0 \tag{34}$$

where $\partial Q/\partial F$ and $\partial Q/\partial N$ are given by Eqs. (9) and (10). Combining Eqs. (33) and (34), together with Eqs. (9) and (10), we obtain:

$$c\left(Q_{np}^{*}w_{np}^{'*} + 1\right) = -\beta Q_{np}^{*}w_{np}^{'*} \tag{35}$$

$$F_{np}^{*} = c\left(w_{np}^{*} + t_{np}\right) - \frac{Q_{np}^{*}}{f_{np}^{'*}} \tag{36}$$

where we use the subscript "np" to specify the non-pooling market, and "*" to indicate the optimality. Eq. (36) follows the form of the Lerner formula (Lerner, 1934), in which the RHS consists of two terms: the average cost for a driver to serve a passenger in pick-up and occupied phases, i.e., $c(w_{np}^{*} + t_{np})$, and the monopoly mark-up, $-Q_{np}^{*}/f_{np}^{'*} > 0$. Moreover, in view of $w_{np}^{'*} = dw_{np}^{*}/\mathrm{d}N_{np}^{v*}$, Eq. (35) can be re-written as:

$$c\left(Q_{np}^{*}\frac{\mathrm{d}w_{np}^{*}}{\mathrm{d}N_{np}^{v*}} + 1\right) = -\beta Q_{np}^{*}\frac{\mathrm{d}w_{np}^{*}}{\mathrm{d}N_{np}^{v*}} \Leftrightarrow c\cdot\left(\mathrm{d}N_{np}^{v*} + Q_{np}^{*}\mathrm{d}w_{np}^{*}\right) = -\beta Q_{np}^{*}\mathrm{d}w_{np}^{*} \tag{37}$$

where the LHS indicates the marginal operating cost of a ride-sourcing vehicle in the vacant phase, i.e., $c\cdot \mathrm{d}N_{np}^{v*}$, minus the marginal operating cost reduction of vehicles in the pick-up phase, i.e., $-cQ_{np}^{*}\mathrm{d}w_{np}^{*}$, while the RHS indicates the marginal cost reduction of passengers in the pick-up phase, i.e., $-\beta Q_{np}^{*}\mathrm{d}w_{np}^{*}$. This implies that the total marginal cost of operating the vehicles in the vacant and pick-up phases equals the marginal pick-up time cost reduction of passengers at the monopoly optimum. This is different from the traditional street-hailing taxi market, in which the marginal cost of operating vacant vehicles equals the marginal pick-up time cost reduction of passengers. In addition, in view of Eq. (35) and the fact that $w_{np}^{'*} < 0$, we have $Q_{np}^{*}w_{np}^{'*} + 1 > 0$, which indicates that,

Lemma 1. *The monopoly optimum in the non-pooling market always locates in the normal regime rather than the WGC regime.*

5.2. Monopoly optimum (MO) in the ride-pooling market

In the monopoly scenario in a ride-pooling market, the optimization problem is similar to that in a non-pooling market. The ride-sourcing platform receives trip fare F from Q passengers and pays unit time operating cost c for a total of N drivers.

$$\text{(P2)}\quad \max \Pi(F,N) = FQ - cN \tag{38}$$

where Q is the solution of the market equilibrium in Eq. (25). The first-order conditions of P2 are:

$$\frac{1}{2}c\left(Q_{rp}^{*}w_{rp}^{'*} + 2\right) = -\beta Q_{rp}^{*}w_{rp}^{'*} \tag{39}$$

$$F_{rp}^{*} = \frac{1}{2}c\left(w_{rp}^{*} + t_{np}\right) + \beta Q_{rp}^{*}\frac{\partial \Delta t^{*}}{\partial Q_{rp}^{*}} - \frac{Q_{rp}^{*}}{f_{rp}^{'*}} \tag{40}$$

where we use the subscript "rp" to specify the ride-pooling market, and "*" to indicate the optimality. It is interesting to see that the optimal pricing formula given by Eq. (40) in a ride-pooling market also follows the Lerner formula. The RHS consists of three terms: the average cost for a driver to serve a passenger in pick-up and occupied phases (half of a vehicle is required to serve each passenger), i.e., $\frac{1}{2}c(w_{rp}^{*} + t_{np})$, an additional term $\beta Q_{rp}^{*}\frac{\partial\Delta t^{*}}{\partial Q_{rp}^{*}}$ associated with actual detour time, and a monopoly mark-up, $-Q_{rp}^{*}/f_{rp}^{'*}$. Since $\frac{\partial\Delta t^{*}}{\partial Q_{rp}^{*}} < 0$, the additional term $\beta Q_{rp}^{*}\frac{\partial\Delta t^{*}}{\partial Q_{rp}^{*}} < 0$. This implies that a decrease in trip fare increases passenger demand, and then reduces the average actual detour time Δt, which in turn increases passenger demand. In other words, a unit decrease in trip fare in a ride-pooling market can in general attract more passengers than would a non-pooling market, due to the reduced actual detour time. Therefore, the platform operating ride-pooling services has a stronger incentive to reduce trip fare than the platform operating non-pooling services. In addition, substituting $w_{rp}^{'*} = dw_{rp}^{*}/\mathrm{d}N_{rp}^{v*}$ into Eq. (39) leads to:

$$\frac{1}{2}c\left(Q_{rp}^{*}\frac{\mathrm{d}w_{rp}^{*}}{\mathrm{d}N_{rp}^{v*}} + 2\right) = -\beta Q_{rp}^{*}\frac{\mathrm{d}w_{rp}^{*}}{\mathrm{d}N_{rp}^{v*}} \Leftrightarrow c\cdot\left(\mathrm{d}N_{rp}^{v*} + \frac{1}{2}Q_{rp}^{*}\mathrm{d}w_{rp}^{*}\right) = -\beta Q_{rp}^{*}\mathrm{d}w_{rp}^{*} \tag{41}$$

where the LHS indicates the marginal cost of operating vehicles in the vacant phase, i.e., $c\cdot \mathrm{d}N_{rp}^{v*}$, minus the marginal operating cost reduction of vehicles in the pick-up phase (each vehicle corresponds to two passengers), i.e., $-c\cdot\frac{1}{2}Q_{rp}^{*}\mathrm{d}w_{rp}^{*}$, while the RHS indicates the marginal cost reduction of passengers in the pick-up phase, i.e., $-\beta Q_{rp}^{*}\mathrm{d}w_{rp}^{*}$. In addition, in view of Eq. (39) and the fact that $w_{rp}^{'*} < 0$, we have $Q_{rp}^{*}w_{rp}^{'*} + 2 > 0$, which indicates that:

Lemma 2. *The monopoly optimum in the ride-pooling market always locates in the normal regime rather than the WGC regime.*

5.3. Social optimum (SO) in the non-pooling market

Next we discuss the first-best social optimum (SO) solution in the non-pooling market. The following problem (P3) aims to maximize social welfare $S(F, N)$ as a function of trip fare F and vehicle fleet size N.

$$\text{(P3)}\quad \max S(F,N) = \int_0^Q f^{-1}(z)dz - \beta\cdot(w + t_{np})Q - cN \tag{42}$$

The first-order conditions of P3 are:

$$\frac{\partial S}{\partial F} = 0 \Rightarrow F\frac{\partial Q}{\partial F} = \beta Q\cdot\frac{\partial w}{\partial F} \tag{43}$$

$$\frac{\partial S}{\partial N} = 0 \Rightarrow c = F\frac{\partial Q}{\partial N} - \beta Q \cdot \frac{\partial w}{\partial F} \tag{44}$$

Combining Eq. (43) and (44) together with Eqs. (9)-(12) yields:

$$c\left(Q_{np}^{*}w_{np}^{'*} + 1\right) = -\beta Q_{np}^{*}w_{np}^{'*} \tag{45}$$

$$F_{np}^{*} = c\left(w_{np}^{*} + t_{np}\right) \tag{46}$$

While Eq. (45) for the social optimum is the same as Eq. (35) for the monopoly optimum, the passenger demand and pick-up time in the two equations are different, because the social optimum trip fare in Eq. (46) does not contain a term of the monopoly markup. Similar to the analysis in Section 5.1, substituting $w_{np}^{'*} = dw_{np}^{*}/\mathrm{d}N_{np}^{v*}$ into Eq. (45) yields:

$$c\left(Q_{np}^{*}\frac{\mathrm{d}w_{np}^{*}}{\mathrm{d}N_{np}^{v*}} + 1\right) = -\beta Q_{np}^{*}\frac{\mathrm{d}w_{np}^{*}}{\mathrm{d}N_{np}^{v*}} \Leftrightarrow c \cdot \left(\mathrm{d}N_{np}^{v*} + Q_{np}^{*}\mathrm{d}w_{np}^{*}\right) = -\beta Q_{np}^{*}\mathrm{d}w_{np}^{*} \tag{47}$$

where the LHS indicates the marginal operating cost of a vehicle in the vacant and pickup phases, i.e., $c \cdot \mathrm{d}N_{np}^{v*} + cQ_{np}^{*}\mathrm{d}w_{np}^{*}$, equals the marginal pick-up time cost reduction of passengers, i.e., $-\beta Q_{np}^{*}\mathrm{d}w_{np}^{*}$, at the social optimum. Moreover, using Eq. (46), we show that the joint profit of the platform and its affiliated drivers at the social optimum is given by

$$\Pi_{np}^{so} = F_{np}^{*}Q_{np}^{*} - cN = -cN_{np}^{v*} < 0 \tag{48}$$

Clearly, Π_{np}^{so} is always negative, and the social optimum is unsustainable unless a certain amount of government subsidy is paid to the platform in the non-pooling market. Moreover, similar to Lemma 1, from Eq. (45), we find that:

Lemma 3. *The social optimum in the non-pooling market always locates in the normal regime rather than the WGC regime.*

5.4. Social optimum (SO) in the ride-pooling market

In a ride-pooling market (under the detour-unconstrained scenario), the SO solution can be found from the following problem (P4):

$$\text{(P4)} \quad \max S(F, N) = \int_0^Q f^{-1}(z)dz - \beta \cdot (w + t_{np} + \Delta t)Q - cN \tag{49}$$

The first-order conditions of P4 are:

$$\frac{1}{2}c\left(Q_{rp}^{*}w_{rp}^{'*} + 2\right) = -\beta Q_{rp}^{*}w_{rp}^{'*} \tag{50}$$

$$F_{rp}^{*} = \frac{1}{2}c\left(w_{rp}^{*} + t_{np}\right) + \beta Q_{rp}^{*}\frac{\partial \Delta t^{*}}{\partial Q_{rp}^{*}} \tag{51}$$

Eq. (50) for the social optimum is the same as Eq. (39) for the monopoly optimum. Eq. (51) states that the social optimum trip fare in the ride-pooling market includes two terms: the average cost for a driver to serve a passenger and an additional term associated with the actual detour time Δt. It is the same as Eq. (40) except for the monopoly mark-up. As mentioned above, the additional term $\beta Q_{rp}^{*}\frac{\partial \Delta t^{*}}{\partial Q_{rp}^{*}}$ associated with Δt is negative, which shows that the reduction of Δt due to an increase of passenger demand will pull down the social optimum trip fare. In other words, a unit decrease in trip fare in the ride-pooling market attracts more passengers than would a non-pooling market, and thus the social optimum trip fare in the ride-pooling market is generally lower than that in the non-pooling market. In addition, substituting $w_{rp}^{'*} = dw_{rp}^{*}/\mathrm{d}N_{rp}^{v*}$ into Eq. (50) gives rise to:

$$\frac{1}{2}c\left(Q_{rp}^{*}\frac{\mathrm{d}w_{rp}^{*}}{\mathrm{d}N_{rp}^{v*}} + 2\right) = -\beta Q_{rp}^{*}\frac{\mathrm{d}w_{rp}^{*}}{\mathrm{d}N_{rp}^{v*}} \Leftrightarrow c \cdot \left(\mathrm{d}N_{rp}^{v*} + \frac{1}{2}Q_{rp}^{*}\mathrm{d}w_{rp}^{*}\right) = -\beta Q_{rp}^{*}\mathrm{d}w_{rp}^{*} \tag{52}$$

which indicates that the marginal operating cost of a vehicle in the vacant and pickup phases, i.e., $c \cdot \mathrm{d}N_{rp}^{v*} + c \cdot \frac{1}{2}Q_{rp}^{*}\mathrm{d}w_{rp}^{*}$, equals the marginal pick-up time cost reduction of passengers, i.e., $-\beta Q_{rp}^{*}\mathrm{d}w_{rp}^{*}$, at the social optimum. Moreover, at the social optimum, the joint profit of the platform and its affiliated drivers is given by

$$\Pi_{rp}^{so} = F_{rp}^{*}Q_{rp}^{*} - cN = -cN_{rp}^{v*} + \beta\left(Q_{rp}^{*}\right)^{2}\frac{\partial \Delta t^{*}}{\partial Q_{rp}^{*}} < 0 \tag{53}$$

which indicates the profit of the platform operating ride-pooling service at the social optimum is always negative, and therefore a government subsidy is needed. In addition, similar to Lemma 2, from Eq. (50), we find that:

Lemma 4. *The social optimum in the ride-pooling market always locates in the normal regime rather than the WGC regime.*

In addition, we prove that:

Proposition 1. *Under a mild condition that* $w(N_{rp}^{\nu*}) + t_{np} \geq \Delta t_d^*$*, the social optimum trip fare in the ride-pooling market is lower than that in the non-pooling market.*

Condition $w(N_{rp}^{\nu*}) + t_{np} \geq \Delta t_d^*$ indicates that the sum of the average pick-up time and normal trip time is greater than the average detour time experienced by drivers in a shared ride, which generally holds in actual operations. It is also worth to mention that $w(N_{rp}^{\nu*}) + t_{np} \geq \Delta t_d^*$ is a sufficient condition but not a necessary condition, and therefore, we can expect that the social optimum trip fare in the ride-pooling market is lower than that in the non-pooling market in most cases.

Proposition 2. *Under conditions that* $w(N_{rp}^{\nu*}) + t_{np} \geq \Delta t_d^*$ *and certain relationship between passenger demand and generalized cost,* e.g.*, a negative exponential demand function* $Q = f(C) = \bar{Q}\exp(-\kappa C)$*, the monopoly optimum trip fare in the ride-pooling market is lower than that in the non-pooling market.*

Condition $w(N_{rp}^{\nu*}) + t_{np} \geq \Delta t_d^*$ is the same as the condition in Proposition 1, which generally holds as aforementioned. The second condition, i.e., passenger demand is proportional to a negative exponential function of the generalized cost, is widely used in literature (such as Yang and Yang, 2011). Therefore, we generally expect that both the monopoly and social optimum trip fares in the ride-pooling market are lower than those in the non-pooling market. The reason behind is intuitive. A platform operating ride-pooling service is able to attract more passengers through a unit decrease in trip fare than the platform operating non-pooling service, and thus is more prone to decrease trip fare to improve the platform's profit and social welfare. The proofs for these two propositions are shown in Appendix C.

5.5. Second-best solution with a profit constraint in the non-pooling market

Since the profit of a ride-sourcing platform at social optimum is negative, next we consider a second-best solution whose objective is to maximize social welfare subject to a nonnegative profit constraint, given in the following problem (P5):

$$\text{(P5)} \quad \max S(F, N) = \int_0^Q f^{-1}(z)dz - \beta \cdot (w + t_{np})Q - cN \tag{54}$$

$$\text{subject to}: \ \Pi(F, N) = FQ - cN \geq \Pi^o \tag{55}$$

where Π^o is a nonnegative reservation profit. To solve this problem, we form the following Lagrangian function:

$$L(F, N) = \int_0^Q f^{-1}(z)dz - \beta \cdot (w + t_{np})Q - cN + \xi \cdot [(FQ - cN) - \Pi^o] \tag{56}$$

where ξ is a Lagrange multiplier. The first-order conditions are:

$$c\left(Q_{np}^* w_{np}^{'*} + 1\right) = -\beta Q_{np}^* w_{np}^{'*} \tag{57}$$

$$F_{np}^* = c\left(w_{np}^* + t_{np}\right) - \frac{\xi Q_{np}^*}{(1+\xi) f_{np}^{'*}} \tag{58}$$

Clearly, the pricing formula in Eq. (58) is a linear combination of the pricing formulas for the monopoly optimum in Eq. (36) and the social optimum in Eq. (46). Then we can infer that:

Corollary 1. *The second-best solution in the non-pooling market always locates in the normal regime rather than the WGC regime.*

5.6. Second-best solution with a profit constraint in the ride-pooling market

In the ride-pooling market, a second-best solution can be found from the following problem (P6):

$$\text{(P6)} \quad \max S(F, N) = \textstyle\int_0^Q f^{-1}(z)dz - \beta \cdot (w + t_{np} + \Delta t)Q - cN \tag{59}$$

$$\text{subject to}: \ \Pi(F, N) = FQ - cN \geq \Pi^o \tag{60}$$

where Π^o is a nonnegative reservation profit. To solve this problem, we form the following Lagrangian function:

$$L(F, N) = \int_0^Q f^{-1}(z)dz - \beta \cdot (w + t_{np} + \Delta t)Q - cN + \xi \cdot [(FQ - cN) - \Pi^o] \tag{61}$$

where ξ is a Lagrange multiplier. The first order conditions are:

$$\frac{1}{2}c\left(Q_{rp}^* w_{rp}^{'*} + 2\right) = -\beta Q_{rp}^* w_{rp}^{'*} \tag{62}$$

$$F_{rp}^{*} = \frac{1}{2}c\left(w_{rp}^{*} + t_{np}\right) + \beta Q_{rp}^{*}\frac{\partial \Delta t^{*}}{\partial Q_{rp}^{*}} - \frac{\xi Q_{rp}^{*}}{(1+\xi)f_{rp}^{\prime *}} \tag{63}$$

Clearly, the pricing formula in Eq. (63) is a linear combination of the pricing formulas for the monopoly optimum in Eq. (40) and the social optimum in Eq. (51). Then we can infer that:

Corollary 2. *The second-best solution in the ride-pooling market always locates in the normal regime rather than the WGC regime.*

Corollary 3. *Under conditions that $w(N_{rp}^{v*}) + t_{np} \geq \Delta t_d^{*}$ and certain relationship between passenger demand and generalized cost,* e.g., *a negative exponential demand function $Q = f(C) = \bar{Q}\exp(-\kappa C)$, the second-best solution optimum trip fare in the ride-pooling market is lower than that in the non-pooling market.*

6. Numerical studies

In this section, a set of numerical experiments is conducted to evaluate the performance of the ride-pooling markets. Specifically, we discuss the impacts of decision variables (i.e., trip fare, vehicle fleet size and allowable detour time) on the key endogenous variables (e.g. pick-up time, passenger demand, successful pairing rate and actual detour time), the platform's profit and social welfare. Both detour-unconstrained and constrained scenarios are examined.

6.1. Experimental settings

The demand function in Eq. (1) is assumed to be of the following negative exponential form:

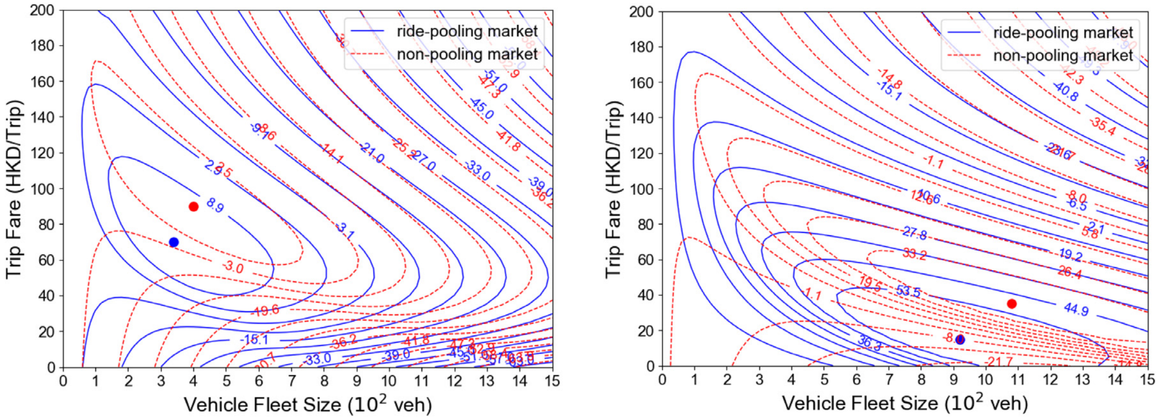
$$Q = f(F + \beta \cdot (w + t_{np})) = \bar{Q}\exp\{-\kappa \cdot [F + \beta \cdot (w + t_{np})]\} \tag{64}$$

where $\bar{Q}$ is the potential passenger demand and κ is a parameter representing the demand sensitivity with respect to the generalized cost. Throughout the numerical studies, we assume $\bar{Q} = 5.0 \times 10^3$ (trips/h), $\kappa = 0.02$ (1/HKD), $\beta = 60$ (HKD/h), $t_{np} = 0.4$ (h), and the unit time operating cost of a vehicle $c = 50$ (HKD/h). The average pick-up time is assumed to be inversely proportional to the square root of the number of vacant vehicles, i.e., $w = H/\sqrt{N^v}$, where parameter H is set to be 5 h. In this paper we assume that the actual detour time Δt follows an exponential distribution with a parameter λ (based on the model presented in Appendix B) which is proportional to passenger demand Q. This implies that the average actual detour time Δt is inversely proportional to passenger demand Q in the detour-unconstrained scenario, which is consistent with the theoretical discussions above. Let $\lambda = 0.2Q$ and thus $\Delta t = 5/Q$ under the detour-unconstrained scenario; let $\gamma = 2$, i.e., $\Delta t_d \cong 2\Delta t$. Note that these parameter values are chosen with partial references to previous studies (e.g., Yang and Yang, 2011), and just for illustrative purposes. In actual operations, one may calibrate the parameters of the proposed functions (e.g. average actual detour time versus passenger demand) and identify their properties with real data.

6.2. Detour-unconstrained scenario

This section verifies the theoretical findings described in Sections 5 with numerical examples in a detour-unconstrained scenario in which allowable detour time $\Delta t_A \to +\infty$. Under this scenario, a platform has two key decision variables, i.e., trip fare and vehicle fleet size, in both ride-pooling and non-pooling markets. The platform's profit and social welfare are evaluated with different combinations of the two decision variables and illustrated by contour maps in a two-dimensional space in Fig. 5. The optimal values in the contour maps in Fig. 5(a) and (b) correspond to the monopoly optimum (MO) and social optimum (SO) solutions, respectively.

Fig. 5(a) shows the iso-profit contours together with the MO solutions of the two markets in a two-dimensional space with vehicle fleet size on the *X*-axis and trip fare on the *Y*-axis. Clearly, both the optimal trip fare and the optimal vehicle fleet size for a monopoly in the ride-pooling market are lower than those in the non-pooling market. This is because a decrease in trip fare not only directly increases passenger demand due to the negative price elasticity, but also reduces actual detour time, which in turn indirectly increases ride-pooling passenger demand (an additional indirect effect). This implies that in general, a decrease in trip fare in the ride-pooling market brings more benefits and thus attracts more passengers than would in the non-pooling market. Therefore, the platform has stronger incentives to reduce trip fare (and thus has a lower optimal trip fare) in the ride-pooling market than the non-pooling market. Fig. 5(b) shows the iso-social-welfare contours together with the social optimum solutions of the two markets in a two-dimensional space with vehicle fleet size on the *X*-axis and trip fare on the *Y*-axis. It is clearly shown that both trip fare and vehicle fleet size at the social optimum in the ride-pooling market are lower than those in the non-pooling market. This is also attributed to the additional indirect effect through the actual detour time in ride-pooling, which yields a larger consumer surplus with a unit decrease in trip fare in the ride-pooling market than would in the non-pooling market.



(a) Profit (10^3 HKD) (b) Social welfare (10^3 HKD)

Fig. 5. Profit and social welfare in a two-dimensional space of vehicle fleet size and trip fare.

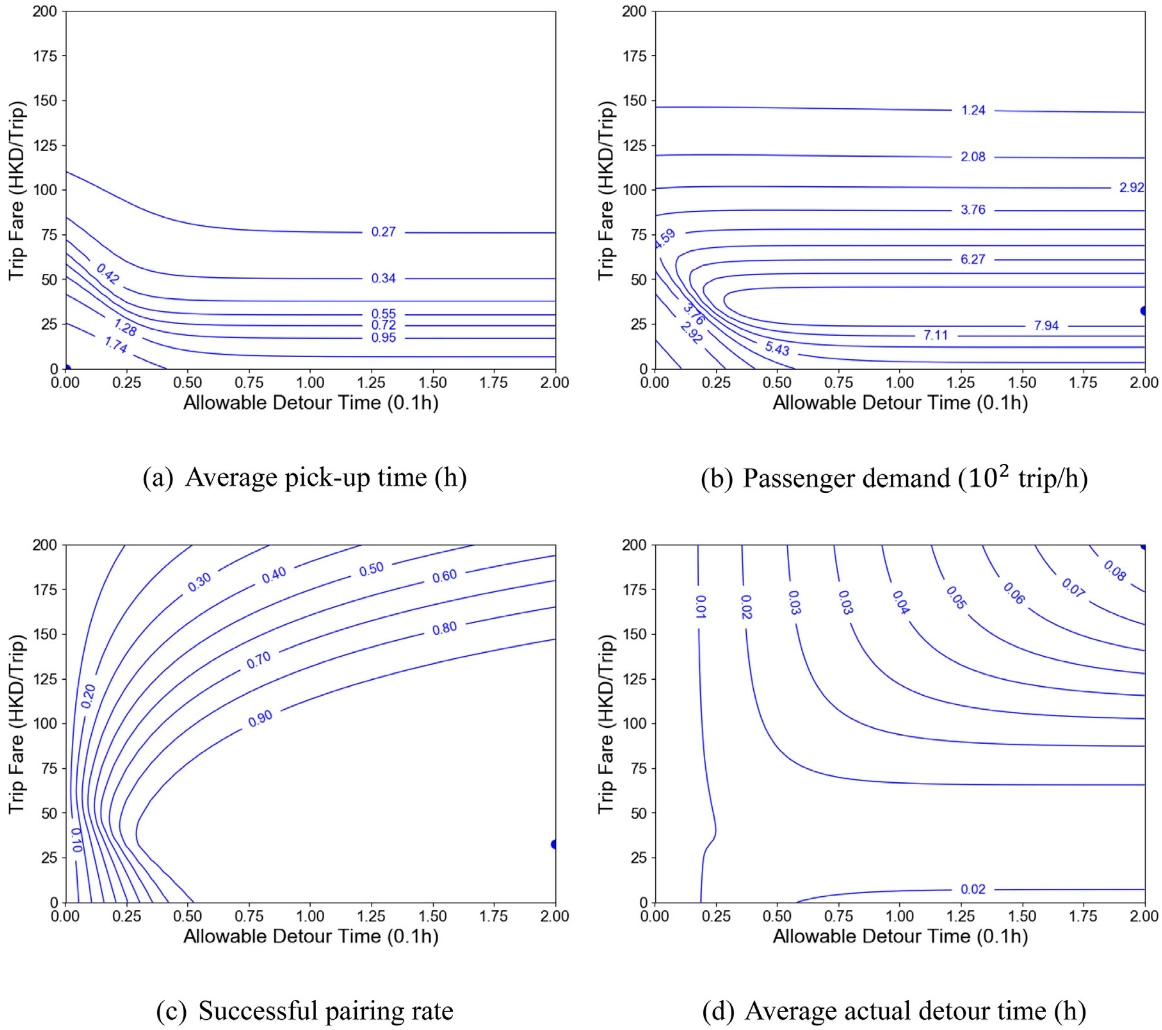
6.3. Detour-constrained scenario

Although it is difficult to theoretically identify the exact impacts of allowable detour time Δt_A in the detour-constrained scenario, this section provides numerical examples to investigate the operating strategies of the ride-sourcing platform in the detour-constrained scenario. In this scenario, the platform has three decision variables: trip fare, vehicle fleet size, and allowable detour time. For illustrative purposes, we fix the vehicle fleet size and explore the contours of key endogenous variables (average pick-up time, passenger demand, successful pairing rate and average actual detour time), platform's profit and social welfare in a two-dimensional space with allowable detour time on the X-axis and trip fare on the Y-axis. N is set to be 500veh (a relatively low level of supply).

Fig. 6(a)–(d) show the contours of average pick-up time w, passenger demand Q, successful pairing rate p and average actual detour time Δt, respectively. Given a fixed vehicle fleet size, it is interesting to find that passenger demand first increases and then decreases with trip fare, which is in contrast to the traditional wisdom that passenger demand always monotonically decreases with trip fare. This is due to the fact that the market will collapse into a WGC regime when the supply is insufficient and/or demand is extremely large due to a low trip fare. It can also be seen from Fig. 6(a) that the average pick-up time may become extremely large at a very low trip fare, which indicates that vehicles spend substantial time for picking up passengers and leads to the WGC regime. Therefore, as pointed out by Castillo et al. (2017), when trip fare is extremely low, increasing the trip fare can drag the market out of the WGC regime by reducing the pick-up time significantly, which then increases the passenger demand.

In addition, while successful paring rate p increases with allowable detour time Δt_A, it also shows a non-monotonic trend (i.e., first increasing and then decreasing) with respect to trip fare. The reason is that, in the normal regime (e.g., when trip fare is high), an increase in trip fare will reduce passenger demand and then reduce successful pairing rate. Conversely, in the WGC regime (e.g., when trip fare is very low), an increase in trip fare increases passenger demand by reducing the pick-up time significantly, and thus increases successful pairing rate. Meanwhile, the average actual detour time Δt monotonically increases with Δt_A because a larger allowable detour time Δt_A tends to pair more passengers with long detours. It is also observed that Δt increases with trip fare in the normal regime since an increase in trip fare reduces passenger demand and thus leads to pairings of larger actual detour time.

Next, we discuss the monopoly optimum (MO) and social optimum (SO) solutions of allowable detour time and trip fare under different vehicle supply levels: a low supply level with vehicle fleet size $N = 500$ veh and a high supply level with vehicle fleet size $N = 5,000$ veh. Fig. 7 shows platform's profit Π and social welfare S under the low supply level in a two-dimensional space with allowable detour time on the X-axis and trip fare on the Y-axis, together with the MO and SO solutions ("MO for RP" and "SO for RP" in the figures). Note that when allowable detour time $\Delta t_A = 0$, the ride-pooling market is reduced to the non-pooling market. The MO and SO trip fares in a non-pooling market with $\Delta t_A = 0$ are calculated and denoted as "MO for NP" and "SO for NP" (on the Y-axis with $\Delta t_A = 0$). Clearly, the optimal trip fare in the ride-pooling market is lower than that in the non-pooling market, with either a profit- or social-welfare-maximizing objective. When the supply level is low, the marginal decrease in pick-up time in response to a unit increase in the number of vacant vehicles is large (imagining that the average pick-up time is convex with the number of vacant vehicles). In this case, the ride-pooling program can greatly reduce pick-up time by having more passengers sharing vehicles to release more vacant vehicles. To this end, the platform will set a relatively large allowable detour time for ride-pooling services to increase the successful pairing rate.



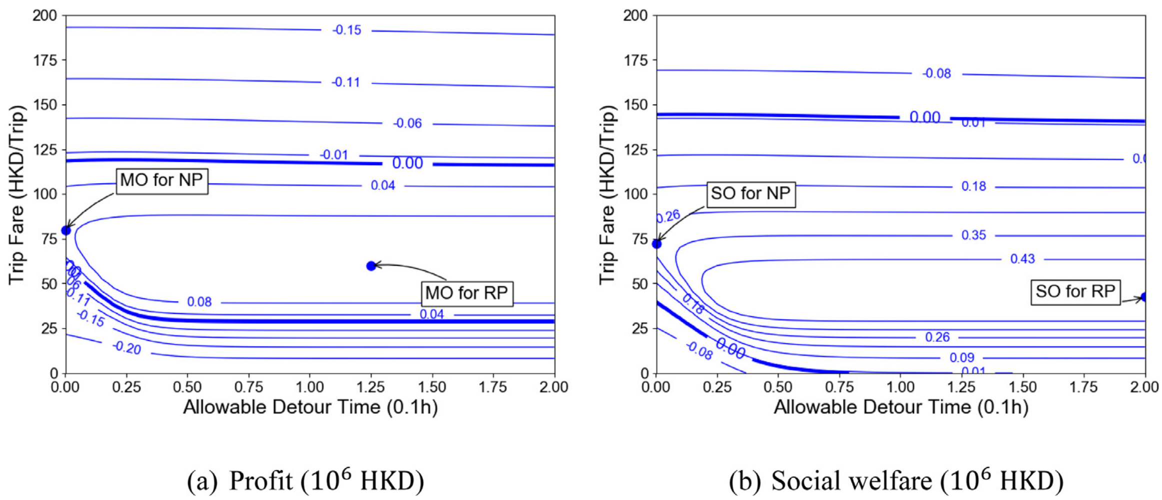
(a) Average pick-up time (h)

(b) Passenger demand (10^2 trip/h)

(c) Successful pairing rate

(d) Average actual detour time (h)

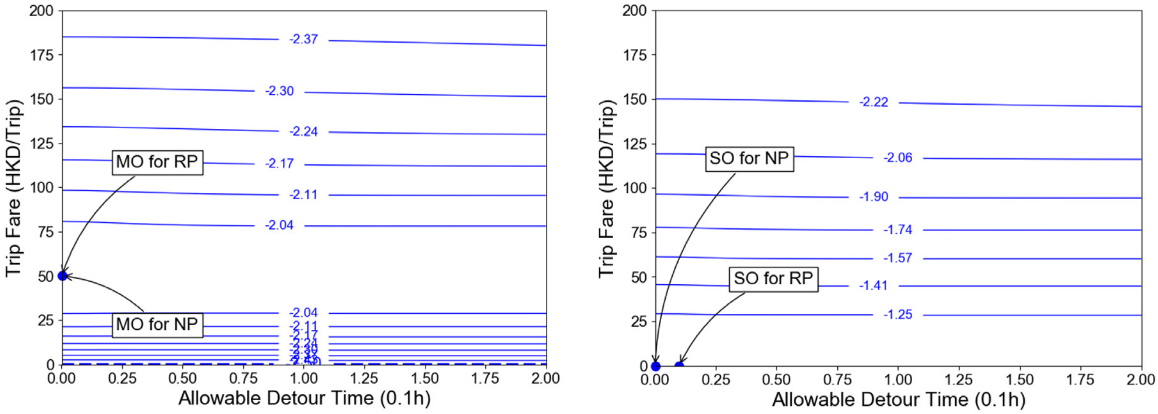
Fig. 6. Endogenous variables in a two-dimensional space of allowable detour time and trip fare.



(a) Profit (10^6 HKD)

(b) Social welfare (10^6 HKD)

Fig. 7. Profit and social welfare in a two-dimensional space of allowable detour time and trip fare with a low supply level ($N = 500$veh).



(a) Profit (10^6 HKD) (b) Social welfare (10^6 HKD)

Fig. 8. Profit and social welfare in a two-dimensional space of allowable detour time and trip fare with a high supply level ($N = 5,000$veh).

Fig. 8 shows the iso-profit contours and iso-social-welfare contours under the high supply level, together with MO and SO solutions, in a two-dimensional space with allowable detour time on the *X*-axis and trip fare on the *Y*-axis. Interestingly, the MO solutions in the ride-pooling market are very close to those in the non-pooling market; in other words, the MO solutions in the ride-pooling market are achieved when $\Delta t_A = 0$. This is because the marginal decrease in pick-up time in response to a unit increase in the number of vacant vehicles is small and even negligible when the supply level is high. On the other hand, pairing passenger requests will increase the actual detour time, which will reduce the passenger demand. In this case, the gain by introducing ride-pooling services is limited and the platform has less incentive to pair passenger requests, thereby setting a small allowable detour time.

7. Conclusion

This paper investigates the emerging on-demand ride-pooling services provided by a fleet of dedicated drivers affiliated with ride-sourcing platforms. A system of nonlinear equations is established to elucidate the complicated relationships between the platform decision variables (i.e., trip fare, vehicle fleet size and allowable detour time) and the system's key endogenous variables (e.g., pick-up time, passenger demand, successful pairing rate and actual detour time) in ride-sourcing markets with and without on-demand ride-pooling services. Based on the modeling framework, the impacts of two decision variables—trip fare and vehicle fleet size—on the platform's profit and social welfare in the detour-unconstrained scenario are analyzed theoretically. We prove that the monopoly optimum, social optimum, and second-best solutions in the ride-pooling and non-pooling markets are located in the normal regime rather than the WGC regime. We also show that monopoly optimum, social optimum and second-best optimum trip fares in the ride-pooling market are in general lower than those in the non-pooling market. This is because a decrease in trip fare not only directly increases passenger demand due to negative price elasticity, but also brings some additional indirect effects—i.e., the increase in demand itself will reduce actual detour time, which in turn increases passenger demand. With numerical experiments, we further examine the impacts of allowable detour time and trip fare on platform's profit and social welfare under different supply levels.

Our study opens other avenues that merit further exploration. To name a few, (1) extending aggregate models to network-based equilibrium models to evaluate network effects; (2) extending stationary models to dynamic models to capture multi-period non-stationary operations; (3) examining market equilibrium and operating strategies for multi-shared rides (ride service to accommodate more than two requests); (4) taking into account the impact of travel time reliability on passenger demand of ride-pooling services (Li et al., 2019a; Long et al., 2018); (5) examining impacts of ride-pooling services on traffic congestion, private car usage, and transit ridership; and (6) calibrating functions of actual detour time experienced by passengers and drivers and the successful pairing rate with real data.

CRediT authorship contribution statement

Jintao Ke: Conceptualization, Methodology, Writing - original draft. **Hai Yang:** Conceptualization, Methodology, Writing - review & editing. **Xinwei Li:** Conceptualization, Methodology, Writing - original draft. **Hai Wang:** Conceptualization, Methodology, Writing - review & editing. **Jieping Ye:** Conceptualization.

Acknowledgments

The work is supported by a grant from Hong Kong Research Grants Council under project HKUST16208619 and a NSFC/RGC Joint Research grant N_HKUST627/18. The corresponding author gratefully acknowledges the support by the Lee Kong Chian (LKC) Fellowship awarded by Singapore Management University.

Appendix A. Nomenclature

Input variables

Notation	Interpretation
$\bar{Q}$	Total potential passenger demand (requests/hour)
t_{np}	Average trip time of non-pooling ride-sourcing services (hour)
κ	Demand sensitivity parameter (1/HKD)
c	Unit time operating cost of each vehicle (HKD/hour)
β	Value of time (HKD/hour)
A	Parameter in the actual detour time function
γ	The ratio of the average driver detour time to the average passenger detour time
Π^0	A nonnegative reservation profit

Decision variables

Notation	Interpretation
Δt_A	Maximum allowable detour time (hour)
N	Vehicle fleet size (i.e., total number of vehicles)
F	Average trip fare (HKD/trip)

System endogenous variables

Notation	Interpretation
t_{rp}	Average trip time of ride-pooling service (hour)
Q	Arrival rate of passengers (i.e., passenger demand) (requests/hour)
w	Average pick-up time (hour)
Δt	Average actual passenger detour time (hour)
Δt_d	Average driver detour time(hour)
p	Successful pairing rate
N^v	Number of vacant vehicles in stationary equilibrium
Π	Profit of ride-sourcing platform (HKD/hour)
S	Social welfare (HKD/hour)
$\tilde{t}$	Random variable—actual detour time between a pair of requests (hour)
λ	Parameter in the exponential distribution of $\tilde{t}$ that is governed by passenger demand Q

System endogenous variables in optimality (monopoly optimum, social optimum, second-best optimum)

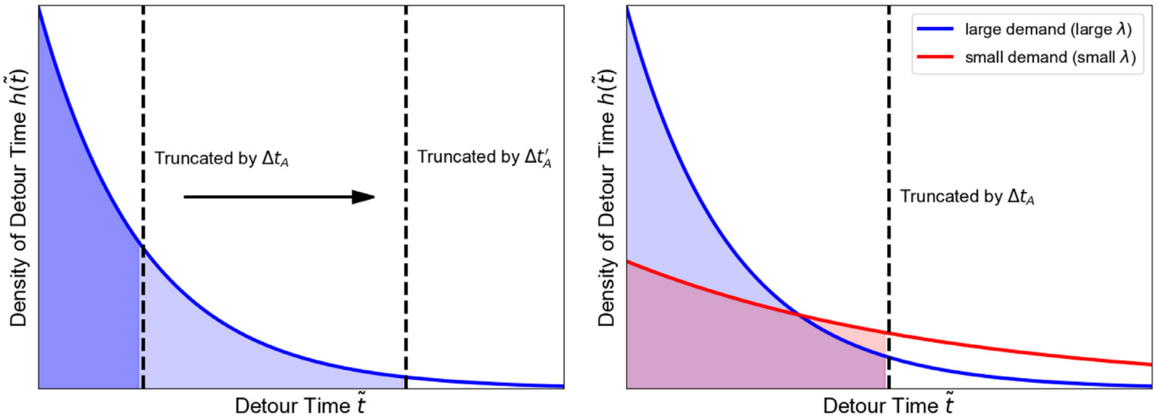
Notation	Interpretation
Q^*_{rp}, Q^*_{np}	Passenger demand in ride-pooling and non-pooling markets (requests/hour)
w^*_{rp}, w^*_{np}	Average pick-up time in ride-pooling and non-pooling markets (hour)
Δt^*	Average actual passenger detour time (hour)
N^{v*}_{rp}, N^{v*}_{np}	Number of vacant vehicles in ride-pooling and non-pooling markets
f'^*_{rp}, f'^*_{np}	Derivative of demand with respect to generalized cost in ride-pooling and non-pooling markets
w'^*_{rp}, w'^*_{np}	Derivative of average pick-up time with respect to number of vacant vehicles in ride-pooling and non-pooling markets
Π^{so}_{rp}, Π^{so}_{np}	Profit of ride-sourcing platform in ride-pooling and non-pooling markets (HKD/hour)
F^*_{rp}, F^*_{np}	Average trip fare in ride-pooling and non-pooling markets (HKD/trip)

Appendix B. A probabilistic model for successful pairing rate and average actual detour time

In what follows, we propose a probabilistic model to characterize how passenger demand Q and allowable detour time Δt_A jointly determine successful paring rate p and average actual detour time Δt. Suppose that the actual detour time between a pair of requests in the detour-unconstrained scenario is a random variable denoted by $\tilde{t}$, which follows a distribution over the range $(0, \infty)$. For analytical tractability to obtain managerial insights, we assume that $\tilde{t}$ follows an exponential distribution with a parameter λ that depends on Q. Probability density function $h(t)$ and cumulative density function $H(t)$ for $\tilde{t}$ are then given by

$$H(t) = \Pr\left(\tilde{t} \le t\right) = 1 - e^{-\lambda t} \tag{65}$$

$$h(t) = \frac{\mathrm{d}}{\mathrm{d}t} H(t) = \lambda e^{-\lambda t} \tag{66}$$



(a) Impacts of Δt_A (b) Impacts of Q

Fig. A1. A representation of successful pairing rate and actual detour time.

In the detour-constrained scenario with allowable detour time Δt_A, the successful pairing rate p can be approximated by the probability that the actual detour time is in the range $(0, \Delta t_A)$. Meanwhile, as shown in Fig. A1(a), the average actual detour time Δt can be approximated by the conditional expectation of the actual detour time in the range $(0, \Delta t_A)$. In other words, the distribution of the actual detour time for successfully paired passengers follows an exponential distribution that is truncated at Δt_A. Formally, successful pairing rate p and average actual detour time Δt are given by

$$p = \Pr\left(\tilde{t} \le \Delta t_A\right) = H(\Delta t_A) = 1 - e^{-\lambda \Delta t_A} \tag{67}$$

$$\Delta t = E\left[\tilde{t} | \tilde{t} < \Delta t_A\right] = \int_0^{\Delta t_A} \frac{h(t)}{p} t \mathrm{d}t = \frac{1}{\lambda} \cdot \left[\frac{1 - (\lambda \Delta t_A + 1)e^{-\lambda \Delta t_A}}{1 - e^{-\lambda \Delta t_A}}\right] \tag{68}$$

where the parameter λ determines the mean of the random variable $\tilde{t}$ (which equals $1/\lambda$).

Furthermore, in the detour-unconstrained scenario (where $\Delta t_A \to \infty$ and all passengers opting for RP service are paired), one can intuitively expect that the larger the passenger demand for RP service Q, the closer the itineraries of paired requests, and thus the smaller the value of $\tilde{t}$. Therefore, it is reasonable to assume that the mean of $\tilde{t}$, i.e., $1/\lambda$, is a decreasing function of Q (as shown in Fig. A1(b)), which indicates that $\partial\lambda/\partial Q > 0$. With the assumptions above, we take partial derivatives of successful pairing rate p with respect to Δt_A and Q:

$$\frac{\partial p}{\partial \Delta t_A} = \lambda e^{-\lambda \Delta t_A} > 0 \tag{69}$$

$$\frac{\partial p}{\partial Q} = \Delta t_A e^{-\lambda \Delta t_A} \frac{\partial \lambda}{\partial Q} > 0 \tag{70}$$

which meets our anticipation that p increases with both Δt_A and Q. Furthermore, from Eq. (67), $p \to 0$ if $\Delta t_A \to 0$, and $p \to 1$ if $\Delta t_A \to \infty$, indicating that the expected boundary conditions are also satisfied. Moreover, taking partial derivative of average actual detour time Δt with respect to Δt_A yields:

$$\frac{\partial \Delta t}{\partial \Delta t_A} = \frac{e^{-\lambda \Delta t_A} \cdot \left(\lambda \Delta t_A - 1 + e^{-\lambda \Delta t_A}\right)}{\left(1 - e^{-\lambda \Delta t_A}\right)^2} \tag{71}$$

Clearly, the sign of $\partial \Delta t / \partial \Delta t_A$ depends on the sign of the term $\lambda \Delta t_A - 1 + e^{-\lambda \Delta t_A}$. Let $y(x) = e^{-x} + x - 1$, for $x > 0$. We have $dy(x)/dx = 1 - e^{-x} > 0$ and $y(0) = 0$, and thus $y(x) > 0$ for all $x > 0$. Therefore, we have $\lambda \Delta t_A - 1 + e^{-\lambda \Delta t_A} > 0$ and obtain that Δt monotonically increases with Δt_A. Moreover, in the detour-unconstrained scenario with $\Delta t_A \to \infty$, we have $\Delta t = 1/\lambda$, and thus $\partial \Delta t / \partial Q = -(\partial \lambda / \partial Q)/\lambda^2 < 0$, which implies that Δt only depends on Q and monotonically decreases with Q when $\Delta t_A \to \infty$. One can show that most of the properties presented still hold for other assumptions on the distributions $\tilde{t}$ (such as normal distribution and log-normal distribution).

Appendix C. Proofs for the propositions

Proof for proposition 1.

We first assume $F^*_{rp} \geq F^*_{np}$. Then by comparing Eq. (46) and Eq. (51), we have $w^*_{rp} > w^*_{np}$. On the demand side, since $F^*_{rp} \geq F^*_{np}$, $w^*_{rp} > w^*_{np}$ and $\Delta t^* > 0$, we shall have $Q^*_{rp} < Q^*_{np}$. On the supply side, in view of the fact that w is a decreasing and convex function with respect to N^v, we have $N^{v*}_{rp} < N^{v*}_{np}$. Since the social optimum locates in the normal regime (i.e., Q decreases with N^v on the supply curve as Eq. (2)), we shall have:

$$\frac{N - N^{v*}_{rp}}{w\left(N^{v*}_{rp}\right) + t_{np}} > \frac{N - N^{v*}_{np}}{w\left(N^{v*}_{np}\right) + t_{np}}$$

If $w(N^{v*}_{rp}) + t_{np} \geq \Delta t^*_d = \gamma \Delta t^*$, then we have:

$$Q^*_{rp}\left[w\left(N^{v*}_{rp}\right) + t_{np}\right] \geq \gamma Q^*_{rp}\Delta t^* \Leftrightarrow N - N^{v*}_{rp} = \frac{1}{2}Q^*_{rp}\left[w\left(N^{v*}_{rp}\right) + \gamma\Delta t^* + t_{np}\right] \geq \gamma Q^*_{rp}\Delta t^*$$

Then we shall have:

$$Q^*_{rp} = \frac{2\left(N - \frac{1}{2}\gamma A - N^v\right)}{t_{np} + w\left(N^{v*}_{rp}\right)} = \frac{N - N^{v*}_{rp} + \left(N - N^{v*}_{rp} - \gamma Q^*_{rp}\Delta t\right)}{w\left(N^{v*}_{rp}\right) + t_{np}} > \frac{N - N^{v*}_{np}}{w\left(N^{v*}_{np}\right) + t_{np}} = Q^*_{np}$$

which is contradictory with the results on the demand side, i.e., $Q^*_{rp} < Q^*_{np}$. This implies that the initial assumption $F^*_{rp} \geq F^*_{np}$ is false, and thus we have $F^*_{rp} < F^*_{np}$.

Proof for proposition 2.

The derivative of passenger demand Q with respect to generalized cost C is given by

$$f' = -\kappa\bar{Q}\exp\left(-\kappa C\right) = -\kappa Q$$

Hence, $Q/f' = -1/\kappa$ is a constant for any given Q.

We first assume $F^*_{rp} \geq F^*_{np}$. Then by comparing Eq. (36) and Eq. (40), in view of that $-Q^*_{rp}/f'^*_{rp} = -Q^*_{np}/f'^*_{np} = -1/\kappa$, we have $w^*_{rp} > w^*_{np}$. On the demand side, since $F^*_{rp} \geq F^*_{np}$, $w^*_{rp} > w^*_{np}$ and $\Delta t^* > 0$, we shall have $Q^*_{rp} < Q^*_{np}$. On the supply side, in view of the fact that w is a decreasing and convex function with respect to N^v, we have $N^{v*}_{rp} < N^{v*}_{np}$. Since the social optimum locates in the normal regime (i.e., Q decreases with N^v on the supply curve as Eq. (2)), we shall have:

$$\frac{N - N^{v*}_{rp}}{w\left(N^{v*}_{rp}\right) + t_{np}} > \frac{N - N^{v*}_{np}}{w\left(N^{v*}_{np}\right) + t_{np}}$$

If $w(N^{v*}_{rp}) + t_{np} \geq \Delta t^*_d = \gamma \Delta t^*$, then we have:

$$Q^*_{rp}\left[w\left(N^{v*}_{rp}\right) + t_{np}\right] \geq \gamma Q^*_{rp}\Delta t^* \Leftrightarrow N - N^{v*}_{rp} = \frac{1}{2}Q^*_{rp}\left[w\left(N^{v*}_{rp}\right) + \gamma\Delta t^* + t_{np}\right] \geq \gamma Q^*_{rp}\Delta t^*$$

Then we shall have:

$$Q^*_{rp} = \frac{2\left(N - \frac{1}{2}\gamma A - N^v\right)}{t_{np} + w\left(N^{v*}_{rp}\right)} = \frac{N - N^{v*}_{rp} + \left(N - N^{v*}_{rp} - \gamma Q^*_{rp}\Delta t\right)}{w\left(N^{v*}_{rp}\right) + t_{np}} > \frac{N - N^{v*}_{np}}{w\left(N^{v*}_{np}\right) + t_{np}} = Q^*_{np}$$

which is contradictory with the results on the demand side, i.e., $Q^*_{rp} < Q^*_{np}$. This implies that the initial assumption $F^*_{rp} \geq F^*_{np}$ is false, and thus we have $F^*_{rp} < F^*_{np}$.

References

Agatz, N., Erera, A., Savelsbergh, M., Wang, X., 2012. Optimization for dynamic ride-sharing: a review. Eur J Oper Res 223 (2), 295–303.

Agatz, N.A.H., Erera, A.L., Savelsbergh, M.W.P., Wang, X., 2011. Dynamic ride-sharing: a simulation study in metro Atlanta. Transp. Res. Part B 45 (9), 1450–1464.

Alonso-Mora, J., Samaranayake, S., Wallar, A., Frazzoli, E., Rus, D., 2017. On-demand high-capacity ride-sharing via dynamic trip-vehicle assignment. Proc. Natl. Acad. Sci. 114 (3), 462–467.

Bai, J., So, K.C., Tang, C.S., Chen, X., Wang, H., 2018. Coordinating supply and demand on an on-demand service platform with impatient customers. Manuf. Serv. Oper. Manage. 21 (3), 556–570.

Cachon, G.P., Daniels, K.M., Lobel, R., 2017. The role of surge pricing on a service platform with self-scheduling capacity. Manuf. Serv. Oper. Manage. 19 (3), 368–384.

Castillo, J.C., Knoepfle, D., Weyl, G., 2017. Surge pricing solves the wild goose chase. In: Proceedings of the 2017 ACM Conference on Economics and Computation. ACM, pp. 241–242.

Chan, N.D., Shaheen, S.A., 2012. Ridesharing in North America: past, present, and future. Transp. Rev. 32 (1), 93–112.

Chen, M.K., Sheldon, M., 2016. Dynamic pricing in a labor market: surge pricing and flexible work on the Uber platform. In: EC, p. 455.

Chen, X.M., Zahiri, M., Zhang, S., 2017. Understanding ridesplitting behavior of on-demand ride services: an ensemble learning approach. Transp. Res. Part C 76, 51–70.

Chen, X.M., Zheng, H., Ke, J., Yang, H., 2020. Dynamic optimization strategies for on-demand ride services platform: surge pricing, commission rate, and incentives. Transp. Res. Part B 138, 23–45.

Chen, X.M., Wang, H., 2018a. Pricing for a last-mile transportation system. Transp. Res. Part B 107, 57–69.

Chen, Y., Wang, H., 2018b. Why are fairness concerns so important? Lessons from a shared last-mile transportation system. Available at SSRN: https://ssrn.com/abstract=3168324.

Cordeau, J.F., Laporte, G., 2007. The dial-a-ride problem: models and algorithms. Ann. Oper. Res. 153 (1), 29–46.

Di, X., Liu, H.X., Ban, X.J., Yang, H., 2017. Ridesharing user equilibrium and its implications for high-occupancy toll lane pricing. Transp. Res. Record 2667, 39–50.

Di, X., Ma, R., Liu, H.X., Ban, X., 2018. A link-node reformulation of ridesharing user equilibrium with network design. Transp. Res. Part B 112, 230–255.

Ferguson, E., 1997. The rise and fall of the American carpool: 1970–1990. Transportation 24 (4), 349–376.

Furuhata, M., Dessouky, M., Ordóñez, F., Brunet, M.E., Wang, X., Koenig, S., 2013. Ridesharing: the state-of-the-art and future directions. Transp. Res. Part B 57, 28–46.

He, F., Shen, Z.J.M., 2015. Modeling taxi services with smartphone-based e-hailing applications. Transp. Res. Part C 58, 93–106.

He, F., Wang, X., Lin, X., Tang, X., 2018. Pricing and penalty/compensation strategies of a taxi-hailing platform. Transp. Res. Part C 86, 263–279.

Ho, S.C., Szeto, W.Y., Kuo, Y.H., Leung, J.M.Y., Petering, M., Tou, T.W.H., 2018. A survey of dial-a-ride problems: literature review and recent developments. Transp. Res. Part B 111, 395–421.

Huang, H.J., Yang, H., Bell, M.G.H., 2000. The models and economics of carpools. Ann. Reg. Sci. 34 (1), 55–68.

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

Ke, J., Yang, H., Zheng, H., Chen, X., Jia, Y., Gong, P., Ye, J., 2019c. Hexagon-based convolutional neural network for supply-demand forecasting of ride-sourcing services. IEEE Trans. Intell. Transp. Syst. 20 (11), 4160–4173.

Ke, J., Zheng, H., Yang, H., Chen, X., 2017. Short-term forecasting of passenger demand under on-demand ride services: a spatio-temporal deep learning approach. Transp. Res. Part C 85, 591–608.

Ke, J., Zheng, Z., Yang, H. and Ye, J., 2020a. Data-driven analysis of matching probability, routing distance and detour distance in on-demand ride-pooling services (June 13, 2020). Available at SSRN: https://ssrn.com/abstract=3626041.

Ke, J., Xiao, F., Yang, H. and Ye, J., 2020b. Learning to delay in ride-sourcing systems: a multi-agent deep reinforcement learning framework. IEEE Transactions on Knowledge and Data Engineering (in press). DOI:10.1109/TKDE.2020.3006084.

Konishi, H., Mun, S., 2010. Carpooling and congestion pricing: HOV and HOT lanes. Reg. Sci. Urban Econ. 40 (4), 173–186.

Lee, A., Savelsbergh, M., 2015. Dynamic ridesharing: is there a role for dedicated drivers? Transp. Res. Part B 81, 483–497.

Lerner, A.P., 1934. The concept of monopoly and the measurement of monopoly power. Rev. Econ. Stud. 1, 157–175.

Li, W., Cui, Z., Li, Y., Ban, X., 2019a. Characterization of ridesplitting based on observed data: a case study of Chendu, China. Transp. Res. Part C 100, 330–353.

Li, S., Tavafoghi, H., Poolla, K., Varaiya, P., 2019b. Regulating TNCs: should Uber and Lyft set their own rules. Transp. Res. Part B 129, 193–225.

Long, J., Tan, W., Szeto, W.Y., Li, Y., 2018. Ride-sharing with travel time uncertainty. Transp. Res. Part B 118, 143–171.

Lyu, G., Cheung, W., Teo, C. and Wang, H., 2019. Multi-objective online ride-matching. Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3356823.

Nie, Y., 2017. How can the taxi industry survive the tide of ridesourcing? Evidence from Shenzhen, China. Transp. Res. Part C 79, 242–256.

Sagarra, O., Szell, M., Santi, P., Díaz-Guilera, A., Ratti, C., 2015. Super sampling and network reconstruction of urban mobility. PLoS ONE 10 (8), 1–15.

Santi, P., Resta, G., Szell, M., Sobolevsky, S., Strogatz, S.H., Ratti, C., 2014. Quantifying the benefits of vehicle pooling with shareability networks. Proc. Natl. Acad. Sci. 111 (37), 13290–13294.

Sayarshad, H.R., Chow J.Y., G., 2015. A scalable non-myopic dynamic dial-a-ride and pricing problem. Transp. Res. Part B. 81, 539–554.

Shaheen, S., Chan, N., Bansal, A., Cohen, A., 2015. Shared Mobility: Definitions, Industry Developments, and Early Understanding. University of California Berkeley Transportation Sustainability Research Center, Berkeley.

Stiglic, M., Agatz, N., Savelsbergh, M., Gradisar, M., 2015. The benefits of meeting points in ride-sharing systems. Transp. Res. Part B 82, 36–53.

Stiglic, M., Agatz, N., Savelsbergh, M., Gradisar, M., 2016. Making dynamic ride-sharing work: the impact of driver and rider flexibility. Transp. Res. Part E 91, 190–207.

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

Sun, H., Wang, H. and Wan, Z., 2019b. Flexible labor supply behavior on ride-sourcing platforms. Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3357365.

Tachet, R., Sagarra, O., Santi, P., Resta, G., Szell, M., Strogatz, S.H., Ratti, C., 2017. Scaling law of urban ride sharing. Sci. Rep. 7, 1–6.

Taylor, T.A., 2018. On-demand service platforms. Manuf. Serv. Oper. Manage. 20 (4), 704–720.

Tong, Y., Chen, Y., Zhou, Z., Chen, L., Wang, J., Yang, Q., ... Lv, W., 2017. The simpler the better: a unified approach to predicting original taxi demands based on large-scale online platforms. In: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, pp. 1653–1662.

Vazifeh, M.M., Santi, P., Resta, G., Strogatz, S.H., Ratti, C., 2018. Addressing the minimum fleet problem in on-demand urban mobility. Nature 557 (7706), 534–538.

Wallsten, S., 2015. The competitive effects of the sharing economy: how is Uber changing taxis. Technol. Policy Inst. 1–22 (June).

Wang, H., Yang, H., 2019. Ridesourcing systems: a framework and review. Transp. Res. Part B 129, 122–155.

Wang, X., Agatz, N., Erera, A., 2018a. Stable matching for dynamic ride-sharing systems. Transp. Sci. 52 (4), 850–867.

Wang, X., He, F., Yang, H., Gao, H.O., 2016. Pricing strategies for a taxi-hailing platform. Transp. Res. Part E 93, 212–231.

Wang, X., Yang, H., Zhu, D., 2018b. Driver-rider cost sharing strategies and equilibrium in a ridesharing program. Transp. Sci. 52 (4), 868–881.

Xu, H., Pang, J.S., Ordóñez, F., Dessouky, M., 2015. Complementarity models for traffic equilibrium with ridesharing. Transp. Res. Part B 81, 161–182.

Xu, Z., Li, Z., Guan, Q., Zhang, D., Li, Q., Nan, J., ... Ye, J., 2018. 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. ACM, pp. 905–913.

Yang, H., Huang, H.J., 1999. Carpooling and congestion pricing in a multilane highway with high-occupancy-vehicle lanes. Transp. Res. Part A 33 (2), 139–155.

Yang, H., Yang, T., 2011. Equilibrium properties of taxi markets with search frictions. Transp. Res. Part B 45 (4), 696–713.

Yang, H., Ke, J., Ye, J., 2018. A universal distribution law of network detour ratios. Transp. Res. Part C 96, 22–37.

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

Yang, H., Qin, X., Ke, J., Ye, J., 2020a. Optimizing matching time interval and matching radius in on-demand ride-sourcing markets. Transp. Res. Part B 131, 84–105 (accepted).

Yang, H., Shao, C., Wang, H., Ye, J., 2020b. Integrated reward scheme and surge pricing in a ridesourcing market. Transp. Res. Part B 134, 126–142.

Yang, H., Ye, M., Tang, W.H., Wong, S.C., 2005. Regulating taxi services in the presence of congestion externalities. Transp. Res. Part A 39 (1), 17–40.

Yu, J., Tang, C.S., Shen, Z.M., Chen, X., 2019. A balancing act of regulating on-demand ride services. Manage. Sci.. (in press) 10.1287/mnsc.2019.3351.

Zha, L., Yin, Y., Du, Y., 2017. Surge pricing and labor supply in the ride-sourcing market. Transp. Res. Procedia 23, 2–21.

Zha, L., Yin, Y., Xu, Z., 2018. Geometric matching and spatial pricing in ride-sourcing markets. Transp. Res. Part C 92, 58–75.

Zha, L., Yin, Y., Yang, H., 2016. Economic analysis of ride-sourcing markets. Transp. Res. Part C 71, 249–266.

Zhang, L., Hu, T., Min, Y., Wu, G., Zhang, J., Feng, P., ... Ye, J., 2017. A taxi order dispatch model based on combinatorial optimization. In: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, pp. 2151–2159.

Zhu, Z., Qin, X., Ke, J., Zheng, Z., Yang, H., 2020. Analysis of multi-modal commute behavior with feeding and competing ridesplitting services. Transp. Res. Part A 132, 713–727.