Transportation Research Part C 178 (2025) 105153

Contents lists available at ScienceDirect

Transportation Research Part C

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

On the inefficiencies in the multi-platform e-hailing market with impatient customers

Guipeng Jiao, Mohsen Ramezani *

The University of Sydney, School of Civil Engineering, Sydney, Australia

ARTICLE INFO

Keywords:
Mobility on-demand
Competition
Transportation network companies
Ridesourcing

ABSTRACT

In a competitive e-hailing market, each participating platform can only utilize a portion of the total demand (passengers) and supply (drivers). This fragmentation may lead to market inefficiencies. In this paper, we introduce an equilibrium model of the multi-platform e-hailing market considering exogenous passenger demand and vehicle supply, and addressing two types of passenger cancellation behaviors. A queueing model with reneging is incorporated to allow passenger cancellations during matching due to impatience. A utility choice model is incorporated to allow passenger cancellations if they are dissatisfied with the service being provided. We use the model to quantify the effects of fragmentation, examining varying combinations of demand and supply levels, illustrating different market conditions. We investigate multiple scenarios, which include up to 10 symmetrical platforms and asymmetrical duopolies. Additionally, we construct and compare the supply curves of a symmetrical duopoly and an equivalent monopoly. It can be shown that the duopoly will be in the Wild Goose Chase (WGC) state for a broader range of market conditions compared to the monopoly, which gives rise to inefficiencies in a fragmented market.

1. Introduction

In the e-hailing market, Transport Network Companies (TNCs) provide various digital platforms that match passengers and drivers in real-time. Generally, a platform is able to establish more optimal matches when it has access to more active passengers and/or vehicles. In other words, the real-time matching nature allows the platforms to benefit from network effects (Wu et al., 2020) on both the demand (passengers) and supply (drivers) sides. However, when there are more than one platform in the e-hailing market, passengers and drivers may choose to use only one of the platforms over the others. As each platform can only utilize a portion of the total demand and supply, this fragmentation dilutes the network effect, which would be otherwise enjoyed by a single platform that has access to all the resources. Therefore, assuming the total demand and supply remain the same in the market, the operational efficiency reduces with the increasing number of platforms in the e-hailing market. Consequently, it may lead to increased deadheading, more passenger cancellations, reduced platform profitability, and lowered driver income. In this study, we quantify the inefficiencies due to fragmentation in the e-hailing market.

Inefficiencies in the e-hailing market have been regularly associated with the Wild Goose Chase (WGC) phenomenon. WGC in the e-hailing market happens when drivers are instructed to pick up distant passengers, which is likely to occur when there is a low supply of drivers. As drivers spend more time on empty miles, it further reduces the number of available drivers and may lead to severe inefficiencies in the market. Beojone and Geroliminis (2021) documented the WGC effect for small fleet sizes, where there was

* Corresponding author.
E-mail address: mohsen.ramezani@sydney.edu.au (M. Ramezani).

https://doi.org/10.1016/j.trc.2025.105153
Received 23 February 2024; Received in revised form 10 March 2025; Accepted 27 April 2025
Available online 16 June 2025

a high number of drivers moving to pick up passengers, compared to the number of drivers delivering them, which subsequently led to decreasing driver revenues. Ke et al. (2020) found that monopoly optimum, social optimum, and second-best solutions in e-hailing markets are always in a normal regime rather than the WGC regime. Castillo et al. (2017) and Xu et al. (2020) scrutinized the properties of the supply curve of the e-hailing market (average pickup time vs. boarding rate). They showed that the supply curve is backward bending and has two states, the Good state and the WGC state. They suggested surge pricing and an adaptive search radius to mitigate the WGC phenomenon.

However, few studies have investigated the inefficiencies due to the fragmentation in the multi-platform e-hailing market. Séjourné et al. (2018) considered the case where the demand is exogenously split between multiple platforms. They defined the price of fragmentation as the increase in supply rebalancing cost incurred by the platforms, compared to the cost of a single platform serving the aggregate demand. Similarly, Kondor et al. (2022) measured the cost of market fragmentation or the cost of non-coordination as the additional number of vehicles needed when the platforms only have partial shares of the demand. Zhou et al. (2022) on the other hand took a different approach. They considered the Nash equilibrium solutions of the competitive ride-sourcing market, where no platform can improve its profitability by unilaterally changing its strategy. They quantified the price of competition and fragmentation by comparing the social welfare under the Nash equilibrium to the optimal social welfare.

The above studies provided insight into the cost of fragmentation with different approaches. This paper looks at the problem from another angle. We argue that the level of short-term demand and supply imbalance (due to within-day fluctuations in real-time operations) can be one of the determinants when quantifying the cost of market fragmentation. As an example, consider a presently overall extremely over-supplied market and two identical platforms evenly splitting the demand and supply in such market conditions. Then the effects of fragmentation may not be significant in this case, as the two platforms are likely to still be over-supplied individually. Yet, the same conjecture may not be true when under other market conditions. For example, during peak hours, supply may be scarce, and market fragmentation could lead to each platform having an even more sparse demand and supply, and consequently, there could be an elevated impact on market efficiency. Therefore, in this study, we investigate the effects of fragmentation under varying combinations of demand and supply levels, illustrating different market conditions.

We propose an aggregated equilibrium model for this study. In the proposed equilibrium model, we consider that the potential passenger demand is known (i.e., exogenous) for each of the platforms operating in the market. Note that the realized passenger demand is endogenous to factors such as price, matching time and pickup time, which are considered by our model. This assumption is similar to that from existing equilibrium models of the e-hailing market (e.g., Zha et al., 2016; Bai et al., 2019; Ke et al., 2020; Zhang and Nie, 2021). Furthermore, the model considers both the *matching time* (the time between a passenger requesting a ride and being assigned a driver) and the *pickup time* (the time between driver assignment and their arrival at the passenger's location). We consider the passengers to be impatient during matching. We use the Cobb–Douglas function to estimate the matching rate and utilize a queuing model with reneging to mirror the impatient behavior. The latter method has not been considered in the literature. Additionally, we consider the passengers can also decline services offered by the platforms as they are service quality sensitive, which is achieved through a utility choice model. (Wang et al., 2022 has considered a queueing model that also addresses the two types of cancellation. However, the cancellation rates are exogenous variables in their work, while the cancellations rates are endogenous in our study). From the drivers' side, we assume the operational fleet size for each of the platforms is known. Additionally, we assume the driver does not decline trips that are accepted by the passengers, and they remain in the market after trip completion.

Adjacent to the proposed equilibrium model, we also develop a disaggregated dynamic model, where the actions of individual passengers, the vehicles, and the platform are simulated. Most of the assumptions in the disaggregated dynamic model are equivalent to those made in the equilibrium model, i.e., known potential passenger demand, impatient and quality sensitive passengers, known fleet size, and compliant drivers. *t* the same time, some of the parameters of the equilibrium model are calibrated based on the setup in the dynamic model, for example, the geometry of the road network and the matching algorithm used. The dynamic model acts as a benchmark for the proposed equilibrium model. The calibrated equilibrium model is shown to produce comparable results to those of the dynamic model.

We consider multiple scenarios using the proposed model, which include up to 10 symmetrical platforms and asymmetrical duopolies (we refer to platforms with identical demand and supply volumes, pricing and wage structures, and matching algorithms, as symmetrical platforms). For every case examined, we quantify the inefficiencies due to fragmentation by comparing key performance indicators (e.g., average pickup times, passenger cancellations, total platform profit) to an equivalent monopoly. Additionally, we demonstrate a relationship between inefficiencies due to market fragmentation and inefficiencies due to WGC.

The remainder of the paper is structured as follows. In Section 2, we introduce the aggregated equilibrium model of the e-hailing market. In Section 3, we showcase the disaggregated dynamic model of the e-hailing market. In Section 4, we compare the results obtained from the two models to demonstrate the proposed equilibrium model's applicability of the multi-platform e-hailing market. Then we quantify the inefficiencies in the multi-platform e-hailing market due to fragmentation. Finally, in Section 5, we conclude the study.

2. Equilibrium model of the e-hailing market

2.1. Demand and supply

In this model, we assume the potential passenger demand is known (i.e. exogenous) for each of the platforms operating in the market. Note that the realized passenger demand is endogenous to factors such as price, matching time and pickup time, which are considered by our model. This assumption is similar to that from existing equilibrium models of the e-hailing market (e.g., Zha

Nomenclature

Exogenous Variables

d_p	potential passenger demand
v_t	number of active drivers
t	average trip duration
$\tau \sim TN(\bar{\tau}, \sigma_\tau, a_\tau, b_\tau)$	passenger patience threshold distribution

Endogenous Variables

p_w	number of waiting passengers
v_i	number of idle vehicles
v_m	number of matched vehicles
v_o	number of occupied vehicles
m	matching rate
b	boarding rate
a_v	trip completion rate
c_p^1	Type 1 cancellation rate
c_p^2	Type 2 cancellation rate
w_p	average pickup time
w_m	average matching time for passengers who are matched (observed matching time)
$\bar{w}_m^*$	average time unmatched passengers spent in the queue (observed patience time)
w_v	average time vehicles spend in the idle state before they are matched

et al., 2016; Bai et al., 2019; Ke et al., 2020; Zhang and Nie, 2021). Additionally, we also assume the operational fleet size for each of the platforms is also known. This setup allows us to analyze the effects of fragmentation under varying combinations of demand and supply levels, illustrating different market conditions. Additionally, we assume that passengers and drivers do not multi-home. Investigating the effect of multi-homing is a future research priority that is out of the scope of this paper.

The proposed model is formulated from the perspective of an arbitrary platform and in an enclosed market. The potential passenger demand for this platform is measured by the passenger request arrival rate, which is denoted as d_p. Note that, we may refer to the potential passenger demand simply as 'passenger demand' or 'demand' in this paper. We assume the arrival of passenger requests follows a Poisson process. After a passenger joins the platform, they wait to be matched. We denote the number of waiting passengers as p_w. Furthermore, we denote the number of active vehicles on this platform as v_t. The vehicles can be in any of the three states: idle, matched (dispatched but unoccupied), and occupied. We denote the number of vehicles in each of these three states as v_i, v_m, and v_o respectively, and:

$$v_i + v_m + v_o = v_t. \tag{1}$$

2.2. Matching rate and average passenger pickup time

The waiting passengers and idle vehicles are systematically matched by the platform recurrently. The matching rate, m, and the average passenger pickup time (from time matched to pickup), w_p, are the outcomes of the matching. Logically, these two values are dependent on the number of waiting passengers and idle vehicles in the system, the characteristics and speed of the road network, and also the matching algorithms used by the platform. With other settings equal, a more effective matching algorithm produces a higher matching rate and a lower pickup time than an inferior algorithm. For a given network and a given matching algorithm used by the platform, we have:

$$m = M(p_w, v_i) \tag{2}$$

$$w_p = W_p(p_w, v_i). \tag{3}$$

It is intuitive that with a rational matching algorithm, as the number of waiting passengers (or idle vehicles) increases, the matching rate should not reduce, ceteris paribus. i.e. $\frac{\partial M(p_w, v_i)}{\partial p_w} \geq 0$ and $\frac{\partial M(p_w, v_i)}{\partial v_i} \geq 0$. It has been demonstrated that a Cobb–Douglas production function can capture these relationships (Yang and Yang, 2011), even under non-equilibrium settings (Nourinejad and Ramezani, 2020). Therefore:

$$m = M(p_w, v_i) = A_0 (p_w)^{\alpha_1} (v_i)^{\alpha_2}. \tag{4}$$

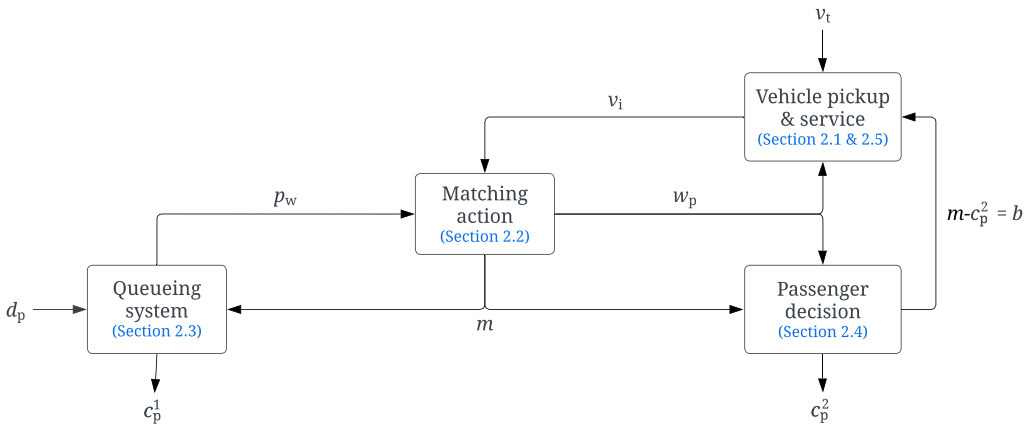


Fig. 1. The schematic diagram of the proposed equilibrium model. It shows the passenger queueing system with reneging as a subpart of the model. $b, c_p^1, c_p^2, d_p, m, p_w, v_i, v_t, w_p$ denote boarding rate, Type I cancellation rate, Type II cancellation rate, passenger demand, matching rate, number of waiting passengers, number of idle vehicles, number of total vehicles, and pickup time, respectively.

Ramezani and Valadkhani (2023) revealed that for a batch matching algorithm, where the waiting passengers and the idle vehicles are accumulated and matched by the platforms periodically over time, the function $W_p(p_w, v_i)$ has the following characteristics:

$$\begin{cases} \frac{\partial W(p_w,v_i)}{\partial p_w} > 0, \frac{\partial W(p_w,v_i)}{\partial v_i} < 0 & \text{if } p_w \le v_i \\ \frac{\partial W(p_w,v_i)}{\partial p_w} < 0, \frac{\partial W(p_w,v_i)}{\partial v_i} > 0 & \text{if } p_w > v_i. \end{cases} \tag{5}$$

They also showed that the following symmetric functional form can be used to estimate the average pickup time:

$$w_p = W_p(p_w, v_i) = \begin{cases} \Gamma_0 (p_w)^{\gamma_1} (v_i)^{\gamma_2} & \text{if } p_w \le v_i \\ \Gamma_0 (p_w)^{\gamma_2} (v_i)^{\gamma_1} & \text{if } p_w > v_i. \end{cases} \tag{6}$$

In Appendix A, we execute a batch matching algorithm (pickup time minimization with finite search radius) in the Manhattan road network. The matching rate and average pickup time of the algorithms are documented. We then log-linearize Eqs. (4) and (6), and estimate the two sets of parameters for the matching algorithm using linear regression. The estimated parameter values are statistically significant at the 1% level, and the R^2 levels are all above 0.89.

2.3. A queueing system with reneging to model impatient passengers

From the perspective of passengers, the process in which they enter the platform, and wait to be matched closely mirrors a queueing system. The passenger demand, d_p, the matching rate, m, and the number of waiting passengers, p_w, mirror the arrival rate, the service rate, and the queue size, respectively.

This queueing system differs from common queueing problems in a number of ways. Firstly, the service rate is dependent on the queue size (Eq. (4)), though under steady state equilibrium of the market, these two values will stabilize. Additionally, under market equilibrium, the service rate must be less than or equal to the arrival rate, since the matching rate cannot be greater than the rate of passenger demand in the market. A utilization factor (arrival rate/service rate) greater than 1 would cause the queue size to grow to infinity in a typical queueing problem. This is dealt with by considering passenger impatience, as they leave the platform if they are not matched within a random patience threshold. This rate of cancellation (Type I), c_p^1, ensures the steady state of the queueing system. Finally, the order of service is not necessarily on a first in first out (FIFO) basis, since the matching algorithm may not prioritize passengers that have waited longer to be matched. In Appendix B, we show that a typical matching algorithm with no prioritization based on waiting time to be matched inclines slightly towards last in first out (LIFO) rather than completely random.

This queueing system models passengers before they are matched, and it is a subpart of the equilibrium model, as shown in Fig. 1. When the market is at equilibrium, the queueing system must also be at equilibrium. Therefore, by considering the steady state equilibrium of this queueing system, the derivations of various endogenous variables are applicable in the formulation of the market equilibrium. Formally, we assume the queueing system is considered to have the following characteristics:

- Customers arriving according to a Poisson process, with arrival rate, d_p;
- The service is provided by a single server, which serves the customers in a random order;
- The successive service times are independent exponentially distributed random variables with rate, m;

- Impatient customers leave the queue if they are not serviced within a certain time. The random patience threshold, τ, is a random variable drawn from a truncated normal distribution. The distribution has mean, standard deviation, and lower and upper bounds denoted as $\bar{\tau}, \sigma_\tau, a_\tau, b_\tau$, respectively, i.e., $\tau \sim TN(\bar{\tau}, \sigma_\tau, a_\tau, b_\tau)$.

The performance metrics of the described queueing system include the following:

- The expected queue size, p_w;
- The rate of cancellation due to impatience (Type I cancellation), c_p^1;
- The average matching time for those who are matched (observed matching time), w_m;
- The average time unmatched passengers spent in the queue (observed patience time), $\bar{w}_m^*$.

A number of studies have examined queueing problems with reneging which involves impatient customers (Barrer, 1957; Haight, 1959; Ancker and Gafarian, 1963; Stanford, 1979; Baccelli et al., 1984; De Kok and Tijms, 1985; Boots and Tijms, 1999; Choudhury, 2008). However, most of those studies consider FIFO, constant customer patience, and/or a utilization factor less than 1. Consequently, none of these studies satisfies all the characteristics of the defined queueing system we aim to analyze. In this study, we do not intend to produce exact analytical solutions to such a queueing system. Therefore, we utilize a series of Monte Carlo simulations to obtain the estimated expressions of the desired measures (see Appendix C).

Nonetheless, we first carry out some initial analysis of this queueing problem. Under the assumption that the arrival rate is greater than or equal to the service rate, $d_p \geq m$, and a large average patience threshold, it is reasonable to assume the probability of an empty server is close to zero. Consequently, all passenger that enters the queue must either be serviced at a rate of m, or leave the queue due to impatience at a rate of c_p^1. In other words, according to flow conservation, the following condition must hold:

$$d_p - c_p^1 - m = 0, \text{ if } d_p \geq m. \tag{7}$$

When a passenger joins the queue, they will eventually either be matched or cancel due to impatience. Therefore, we can conceptualize that there are two distinct pools of waiting passengers, those who will be matched (with matching rate m, and average matching time, w_m), and those who will cancel (with cancellation rate c_p^1, and average time they spent in the queue before leaving, $\bar{w}_m^*$). These two pools constitute the total queue size. Little's law dictates that the long-term average number of customers in a stationary system is equal to the long-term average effective arrival rate multiplied by the average time a customer spends in the system. Then, the expected number of passengers in the queue who will be matched, and the expected number of passengers in the queue who will cancel, are $w_m m$, and $\bar{w}_m^* c_p^1$, respectively. Consequently, the total queue size p_w can be formulated as follows:

$$p_w = w_m m + \bar{w}_m^* c_p^1. \tag{8}$$

Through Monte Carlo simulations shown in Appendix C, we found for each given distribution of τ, the following functional form can be used to estimate the observed matching time and observed patience time:

$$w_m = \begin{cases} \frac{(d_p - m)}{\xi_0 + \xi_1 (d_p - m)} & \text{if } m \leq d_p \\ 0 & \text{if } m > d_p. \end{cases} \tag{9}$$

$$\bar{w}_m^* = \begin{cases} \frac{\phi_0 + (d_p - m)}{\phi_1 + \phi_2 (d_p - m)} & \text{if } m \leq d_p \\ 0 & \text{if } m > d_p. \end{cases} \tag{10}$$

With exogenous d_p and m, solving Eqs. (7)–(10) simultaneously yields the steady state equilibrium of the matching queueing system. These equations can also be used to obtain the market equilibrium where m is endogenous but explicitly defined in Eq. (4).

2.4. Service quality sensitive passengers

We consider passengers to be service quality sensitive. After passengers are matched (at a rate of m), they are presented with the trip details by the platform. Then they have the option to decline the trip if they are not satisfied and choose other modes of travel without penalization (Type II cancellation). We consider that the average utility of the ridesourcing trip, u_s, is a function of the average passenger pickup time, w_p, and the average trip fare, f. Therefore, $u_s = \beta_s - \beta_w w_p - \beta_f f$; where β_s, β_w, and β_f, are the utility constants for ridesourcing trips, for pickup time, and for trip fare, respectively. We also assume that the utility of other modes of travel is a constant, u_o. Note that in this study, we do not consider pricing strategy, therefore it is not necessary to include fare in the utility function, i.e., without loss of generality, $\beta_f = 0$. Though the fare can be readily included with a non-zero β_f and to allow for an upper-level optimization. With the known average utilities of ridesourcing trips and other modes of travel, we use a logit choice model to determine the average probability of passengers choosing other modes of travel, i.e., type II cancellation due to unsatisfactory service for each matching. Therefore, for a given rate of matching, the rate of type II cancellation can be formulated as follows:

$$c_p^2 = m \left(1 - \frac{e^{u_s}}{e^{u_o} + e^{u_s}}\right). \tag{11}$$

2.5. Vehicle pickup and service

In this study, we assume that the drivers do not decline trips that are accepted by the passengers. Subsequently, the rate of vehicle dispatch to pick up passengers is equivalent to the rate of matching minus the rate of passenger cancellation due to unsatisfactory service, i.e., $m - c_{\mathrm{p}}^2$. Moreover, the average time a vehicle spends in this matched (dispatched but unoccupied) state is equivalent to the average passenger pickup time, w_{p}. Therefore, according to Little's law, the number of vehicles in this state under equilibrium can be determined as:

$$v_{\mathrm{m}} = w_{\mathrm{p}}(m - c_{\mathrm{p}}^2). \tag{12}$$

When a vehicle arrives at the designated passenger's location, the passenger boards the vehicle and heads to their destination. We denote the boarding rate in the market, b, and assume the average duration of trips is an exogenous constant, t. The number of vehicles in the occupied state under equilibrium is therefore as follows:

$$v_{\mathrm{o}} = bt. \tag{13}$$

After a vehicle completes the trip, or after the match is rejected by the passenger due to dissatisfaction, the vehicle becomes idle. We denote the rate of trip completion as a_{v} and the average time vehicles spent in the idle state before they are matched as w_{v}. Consequently, the number of idle vehicles at any time under equilibrium is as follows:

$$v_{\mathrm{i}} = w_{\mathrm{v}}(a_{\mathrm{v}} + c_{\mathrm{p}}^2). \tag{14}$$

2.6. Equilibrium condition

Under market equilibrium, the numbers of waiting passengers, idle vehicles, matched vehicles, and occupied vehicles remain unchanged over time. Therefore, we have:

$$\frac{\mathrm{d}(p_{\mathrm{w}})}{\mathrm{dt}} = d_{\mathrm{p}} - c_{\mathrm{p}}^1 - m = 0 \tag{15}$$

(Note here that Eq. (15) is equivalent to Eq. (7))

$$\frac{\mathrm{d}(v_{\mathrm{m}})}{\mathrm{dt}} = m - c_{\mathrm{p}}^2 - b = 0 \tag{16}$$

$$\frac{\mathrm{d}(v_{\mathrm{i}})}{\mathrm{dt}} = a_{\mathrm{v}} + c_{\mathrm{p}}^2 - m = 0 = b - a_{\mathrm{v}} = \frac{\mathrm{d}(v_{\mathrm{o}})}{\mathrm{dt}} \tag{17}$$

In summary, for the proposed equilibrium model of the e-hailing market, we consider exogenous variables which include the potential passenger demand (d_{p}), the total number of active vehicles (v_{t}), passenger matching patience distribution ($\tau \sim TN(\bar{\tau}, \sigma_\tau, a_\tau, b_\tau)$), passenger utility parameters ($u_{\mathrm{o}}, \beta_{\mathrm{s}}, \beta_{\mathrm{w}}$, and β_{f}), and the average trip duration (t). By integrating a queueing system with reneging, and a utility choice model, the proposed equilibrium model is able to handle two types of passenger cancellations. Namely, cancellation during matching due to impatience (Type I), and cancellation after matching due to dissatisfaction with service (Type II). Finally, solving Eqs. (1), (4), (6), and (8)–(17) simultaneously, yields the market equilibrium condition.

3. Dynamic model of the e-hailing market

This section showcases a disaggregated dynamic model of the e-hailing market. In this model, the interactions among market participants, including individual passengers, vehicles, and the platform, are simulated. The dynamic model is developed while ensuring that its underlying assumptions align with and adhere to those of the proposed equilibrium model pertaining to the same subject matter. At the same time, Eqs. (4) and (6) of the equilibrium model would be calibrated based on the geometry of the road network and the matching algorithm used in the dynamic model. The dynamic model would be used as a benchmark for the proposed equilibrium model, and to test the robustness of the proposed equilibrium model.

3.1. The passengers

In this model, we consider the passengers to be heterogeneous. When there is more than one platform in the market, we assume that each passenger randomly uses only one platform without multi-homing. The passengers join the platform and indicate their origins and destinations and wait to be matched. Consider the case of a duopoly, we denote the set of all passengers waiting to be matched who joined platform 1 and 2 as $\mathcal{P}_1 = \{p_{1,1}, \dots, p_{m,1}\}$ and $\mathcal{P}_2 = \{p_{1,2}, \dots, p_{n,2}\}$, respectively. An individual waiting passenger in platform 1 is denoted as $p_{i,1} \in \mathcal{P}_1$, and an individual waiting passenger in platform 2 is denoted as $p_{j,2} \in \mathcal{P}_2$.

The passengers are impatient; if they are not matched by the platform within a certain time frame, they will cancel their orders and use other modes of travel such as public transport. Consider passenger $p_{i,1}$, we denote his/her waiting patience as $\bar{m}_{i,1}$. The passengers are also sensitive to service quality and cost. When passenger $p_{i,1}$ is matched to a vehicle, $v_{a,1}$, by the platform, they will be presented with the trip details, which include the exact waiting time to be picked up, $w_{i,1}^{a,1}$, and fare price, $f_{i,1}$. The passenger then has the option to cancel the trip if they are not satisfied with the service offered by the platform. In practice, some platforms impose penalties on passengers who cancel their trips after being matched. In this study, we assume there is no such penalty for

passengers associated with this cancellation. We use utility-based choice modeling to predict the passenger's behavior in accepting the offered trip by the platform. We assume that the passenger perceives the utility of the trip offered and all other modes of travel as follows:

$$\text{Trip offered}: u^{\text{s}}_{i,1} = \beta^{\text{s}}_{i,1} - \beta^{\text{w}}_{i,1} w^{a,1}_{i,1} - \beta^{\text{f}}_{i,1} f_{i,1} \tag{18}$$

$$\text{Other modes}: u^{\text{o}}_{i,1} = u_{\text{o}} \tag{19}$$

where $\beta^{\text{s}}_{i,1}$ and u_{o} are the utility constants of the two options, while $\beta^{\text{w}}_{i,1}$ and $\beta^{\text{f}}_{i,1}$ are the per unit time and per unit price utility coefficients for passenger $p_{i,1}$, respectively. Similar to the equilibrium model, and without loss of generality, we assume $\beta^{\text{f}}_{i,1} = 0 \quad \forall\, p_{i,1} \in \mathcal{P}_1$ in this study. Then the probability of the passenger accepting the trip offered by the platform, $\Pr^{\text{s}}_{i,1}$, can be calculated:

$$\Pr^{\text{s}}_{i,1} = \frac{e^{u^{\text{s}}_{i,1}}}{e^{u^{\text{s}}_{i,1}} + e^{u^{\text{o}}_{i,1}}}. \tag{20}$$

After accepting the trip, the passenger will be stationary at their origin to wait for the dispatched vehicle for pick up, and they will leave the platform once they have arrived at their destination.

3.2. The vehicles

In this study, we assume that the platforms have full control of their fleet of vehicles (there are numerous studies that focused more on the nuances of drivers' choices in the e-hailing market (e.g., Ramezani et al., 2022; Fielbaum and Tirachini, 2021; Beojone and Geroliminis, 2023), but in this paper, we choose to focus more on passenger decisions), while the wage for the drivers needs to be paid for. That is, the drivers will always accept the trips designated by the platforms, and they follow the routes directed by the platform (Beojone et al., 2024), which is the shortest time path between any two locations. The vehicles are also assumed to be stationary after the completion of each trip as the drivers wait to be matched with the next passenger by the platform. Therefore, the three states that a vehicle could be in are: idle, matched (dispatched but unoccupied), and occupied, similar to the equilibrium model. The idle vehicles are considered for matching by the platform. We denote a single idle vehicle from the set of all idle vehicles working in platforms 1 and 2 as $v_{a,1} \in \mathcal{V}_1 = \{v_{1,1}, \ldots, v_{k,1}\}$ and $v_{b,2} \in \mathcal{V}_2 = \{v_{1,2}, \ldots, v_{l,2}\}$, respectively. Once a vehicle is matched, and the corresponding passenger accepts the trip, then the vehicle is dispatched to pick up the passenger. As the vehicle arrives at the location of the passenger, the vehicle becomes occupied and heads for the passenger's destination.

3.3. The platforms

The platforms are aware of the origins and destinations of their own waiting passengers, as well as the locations of their idle vehicles. Let us consider a waiting passenger $p_{i,1} \in \mathcal{P}_1$, and an idle vehicle $v_{a,1} \in \mathcal{V}_1$. We denote the origin and destination of $p_{i,1}$, and the position of $v_{a,1}$ as $O_{i,1}$, $D_{i,1}$, and $V_{a,1}$, respectively. Furthermore, the travel time of the shortest time path between any two locations, such as from $O_{i,1}$ to $D_{i,1}$, is denoted as $|O_{i,1}D_{i,1}|$.

We assume the fare structure consists of a fixed base price (λ_1) and a trip duration dependent variable price (the price per unit time is λ_2), which is shown as follows:

$$f_{i,1} = \lambda_1 + \lambda_2 |O_{i,1} D_{i,1}|. \tag{21}$$

In this study, we assume that drivers are paid for any time they spent en-route, which includes the states they spent matched and occupied (see Jiao and Ramezani, 2022). We assume the payment per unit time driver spent matched and occupied are λ_3 and λ_4, respectively. Therefore, if the platform dispatches $v_{a,1}$ to pick up $p_{i,1}$ given that the passenger has accepted the trip, then the wage to be paid, $c^{a,1}_{i,1}$, and the profit made by the platform, $\pi^{a,1}_{i,1}$, can be determined respectively as follows:

$$c^{a,1}_{i,1} = \lambda_3 |V_{a,1} O_{i,1}| + \lambda_4 |O_{i,1} D_{i,1}| \tag{22}$$

$$\pi^{a,1}_{i,1} = f_{i,1} - c^{a,1}_{i,1}. \tag{23}$$

There are two reasons for the adoption of this wage structure. Firstly, since we assume the platforms have full control of their fleet and the drivers always accept the designated trips, then it would be unfair to dispatch a vehicle to a far away passenger without paying for the empty miles traveled. Secondly, since the empty miles have to be paid for (i.e., penalized), it would be illogical to dispatch vehicles to passengers that are far away, even if there are extra idle vehicles. Therefore, if a platform chooses to maximize profit, this wage structure non-explicitly curbs the wild goose chase (WGC) problem.

3.4. Matching algorithm

We assume that the platforms utilize a batch matching algorithm, which is used in practice by platforms such as Didi, and considered by a number of studies (e.g. Chen et al., 2021; Yang and Ramezani, 2022; Tafreshian and Masoud, 2020; Alisoltani et al., 2022). With a batch matching algorithm, the idle vehicles and waiting passengers are accumulated and matched by each of the platforms every Δ seconds. There are many ways a platform can match the passengers and the vehicles depending on their

objective. In this paper, we consider the platform uses a matching algorithm with a given search radius that minimize the total passenger pickup time while establishing as many matches as possible. Consider the given sets $\mathcal{P}_1$ and $\mathcal{V}_1$ for platform 1, the cardinality of the sets are denoted $|\mathcal{P}_1|$ and $|\mathcal{V}_1|$, respectively. Let $\mathcal{E}_1$ be the set of edges connecting each element of $\mathcal{P}_1$ and $\mathcal{V}_1$, where an edge $(p_{i,1}, v_{a,1}) \in \mathcal{E}_1$ connects $p_{i,1} \in \mathcal{P}_1$ and $v_{a,1} \in \mathcal{V}_1$. The pickup time minimizing matching for a maximal number of matchings can be formulated by solving the following weighted bipartite matching problem:

$$\min_{x_{i,1}^{a,1}} \sum_{(p_{i,1}, v_{a,1}) \in \mathcal{E}_1} w_{i,1}^{a,1} x_{i,1}^{a,1} \tag{24}$$

$$\text{s.t.} \sum_{a=1}^{k} x_{i,1}^{a,1} \leq 1 \quad \forall\, p_{i,1} \in \mathcal{P}_1 \tag{25}$$

$$\sum_{i=1}^{m} x_{i,1}^{a,1} \leq 1 \quad \forall\, v_{a,1} \in \mathcal{V}_1 \tag{26}$$

$$\sum_{a=1}^{k} \sum_{i=1}^{m} x_{i,1}^{a,1} = \min(|\mathcal{P}_1|, |\mathcal{V}_1|) \tag{27}$$

$$x_{i,1}^{a,1} \in \{0, 1\} \quad \forall\, p_{i,1} \in \mathcal{P}_1,\ v_{a,1} \in \mathcal{V}_1 \tag{28}$$

In Eq. (24), $x_{i,1}^{a,1}$ is the binary decision variable, with $x_{i,1}^{a,1} = 1$ indicating that designating $v_{a,1}$ to pick up $p_{i,1}$ is in the optimal solution. Eqs. (25) and (26) are constraints that ensure each passenger and each vehicle is matched at most once. Eq. (27) dictates the number of matching is the minimum cardinality of set $\mathcal{P}_1$ and $\mathcal{V}_1$. After solving the optimization in Eqs. (24)–(28), the matched pairs whose pickup time is greater than the radius, $\bar{w}_p$, are then dropped to produce the finalized matching.

4. Numerical experiments

So far, we have introduced an equilibrium model and a dynamic model of the e-hailing market. The two models consider similar passenger and driver behaviors at two different aggregation levels. In this section, we will use the dynamic model to validate the results obtained by the equilibrium model. Additionally, we will utilize the equilibrium model to quantify the inefficiencies in various multi-platform markets.

This section first presents the setup of the proposed equilibrium model and the disaggregated dynamic model (Section 4.1). The two models use the same values for corresponding parameters, and some parameter values from the equilibrium model are calibrated based on the dynamic model. Then Section 4.2 shows a comparison between the results obtained by the proposed equilibrium model and the disaggregated dynamic model. The comparison demonstrates the applicability of the proposed equilibrium model to analyze the multi-platform e-hailing market. Following this, we use the equilibrium model to isolate and quantify the inefficiencies due to fragmentation in the multi-platform e-hailing market. We consider multiple scenarios, which include symmetrical duopolies (Sections 4.3 & 4.4), multiple symmetrical platforms (Section 4.5), and asymmetrical duopolies (Section 4.6). We assume that the total demand (passengers) and supply (drivers) in the e-hailing market would be split evenly between the symmetrical platforms. These scenarios of fragmented markets are compared to an equivalent monopoly with the same *total* demand and supply.

4.1. Model setup

For the dynamic model showcased in Section 3, we conduct the experiments in a simulator, where the Manhattan New York road network is transformed into a directed graph. We assume that every link in the road network has a unique and time-invariant speed at which vehicles travel, i.e. the effects of congestion (Alisoltani et al., 2021) are not considered in this study. Vehicles are added at random locations at the beginning of the simulation until the desired fleet size is achieved. The total rate of addition in the market is 2 vehicles per second. The passengers are also added at random origins with random destinations throughout each simulation at a constant rate. The platforms adopt a matching interval, Δ, of 10 s. In this section, we assume the platforms use the pickup time minimization algorithm shown in Eqs. (24)–(28), and a searching radius, $\bar{w}_p$, of 12 min. The simulations run for 2 h, while any measurements are taken after 1 h allowing the simulation to stabilize. For the passenger choice model introduced in Section 3.1, the parameter values for each individual passenger are drawn from bounded normal distributions. The mean, standard deviations, and bounds for those parameters are shown in Table 1.

The parameter values associated with pricing and wage structures are shown in Table 2. Note that since the fare and wage structures are not subjected to optimization in the equilibrium model, these parameters are not present in the equilibrium model. However, we assume the fare and wage structures in the equilibrium model are identical to those in the dynamic model to obtain performance indicators such as platform profits. Also note that to conduct fair comparisons and to isolate the effect of market fragmentation, we assume all platforms in the fragmented market (as well as the only platform under the case of monopoly) adopt the same pricing, wage structure, and matching mechanism.

For the equilibrium model proposed in Section 2, we have introduced a number of exogenous variables and parameters. The descriptions and chosen values of the exogenous variables are summarized in Table 3. These parameter values are the same as their counterparts in the dynamic model.

Table 1
Dynamic model passenger choice parameters. μ, σ, a, and b are the mean, standard deviation, lower bound, and upper bound, respectively for the bounded normal distributions.

Parameter	Description	Unit	μ	σ	a	b
$\bar{m}$	Passenger matching patience	s	60	10	40	80
u^{o}	Utility constant for other travel modes	–	−6	0.5	−7	−5
β^{s}	Utility constant for ridesourcing trip	–	0	0	–	–
β^{w}	Utility coefficient for pickup time	1/min	1	0	–	–

Table 2
Platform pricing and wage structure parameters.

Parameter	Description	Unit	Value
λ_1	Fixed base price	$	2.55
λ_2	Trip duration variable price	$/min	0.6
λ_3	Wage per unit time driver spent matched	$/min	0.48
λ_4	Wage per unit time driver spent occupied	$/min	0.48

Table 3
Equilibrium model exogenous variables.

Variable	Description	Unit	Value
$\bar{\tau}$	Passenger matching patience mean	s	60
σ_τ	Passenger matching patience standard deviation	s	10
a_τ	Passenger matching patience lower bound	s	40
b_τ	Passenger matching patience upper bound	s	80
u_o	Utility for other modes of travel	–	−6
β_s	Utility constant for ridesourcing trip	–	0
β_w	Utility coefficient for pickup time	1/min	1
t	Average Trip duration	s	860

Table 4
Equilibrium model calibrated parameter values.

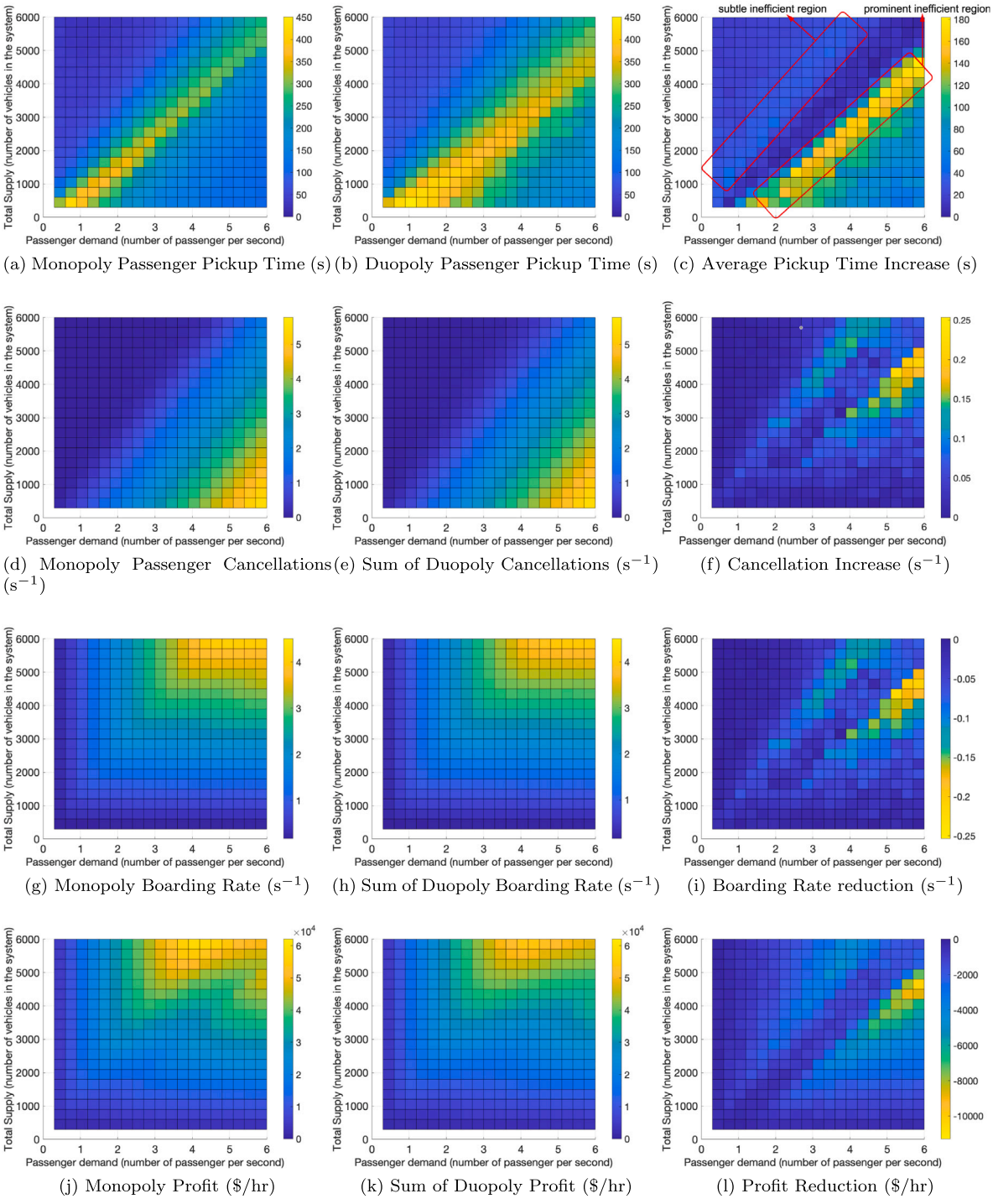
Equation	Parameter values		
Matching rate estimation (Eq. (4))	$A_0 = 0.0111$	$\alpha_1 = 0.7327$	$\alpha_2 = 0.7309$
Pickup time estimation (Eq. (6))	$\Gamma_0 = 972.4$	$\gamma_1 = 0.0891$	$\gamma_2 = -0.3934$
Observed matching time estimation (Eq. (9))	$\xi_0 = 0.0091$	$\xi_1 = 0.0336$	
Observed patience time estimation (Eq. (10))	$\phi_0 = 0.9124$	$\phi_1 = 0.0164$	$\phi_2 = 0.0166$

The matching rate function and pickup time function parameters in Section 2.2 are calibrated in Appendix A. Note that the calibration is based on a pickup time minimization algorithm, which is the same algorithm used by platforms in the dynamic model. The queuing system parameters in Section 2.3 are calibrated in Appendix C. The calibrated parameter values are summarized in Table 4.

4.2. Results comparison between the two models

In this section, we compare the results obtained by the dynamic model and the equilibrium model. The dynamic model can be considered a type of agent-based model, where the interactions between passengers, vehicles, and the platforms are simulated; as opposed to the equilibrium model, which views the e-hailing market on an aggregated level. Here, we use the two models to investigate how a symmetrical duopolistic e-hailing market is different from a monopolistic e-hailing market. The experiments are conducted with varying fleet sizes coupled with varying passenger demand rates to reflect different market conditions. The total passenger demands in the market tested range from 0.3 passengers per second to 6 passengers per second with 0.3 increments. The total fleet sizes in the market tested range from 300 vehicles to 6000 vehicles with 300 increments. We denote the rate of passenger demand for the market as D_p and the number of active vehicles in the market as V_t. That is, $d_p = \frac{1}{2} D_p$, and $v_t = \frac{1}{2} V_t$, for each of the platforms in the duopoly. While $d_p = D_p$, and $v_t = V_t$, for the monopoly. The results of the experiments obtained by the dynamic model and equilibrium model are shown in Figs. 2 and 3, respectively, which include performance indicators of average passenger pickup time, cancellations, boarding rate, and profit. We also construct scatter plots in Fig. 4 to compare the results obtained by the equilibrium model and the dynamic model.

We first study the results in Figs. 2 and 3. It can be observed that the results obtained by the two models have very similar features. For example, we remark that as the demand increases for each level of supply, the average pickup distance increases to a maximum and then decreases to a moderate level. While for the duopoly, the range for which the average pickup time stays high is wider than the monopoly. Furthermore, for the comparison between the duopoly and monopoly, we note the two distinct diagonal regions where the effects of market fragmentation are more significant. One of the diagonal regions is more prominent, and the



(a) Monopoly Passenger Pickup Time (s) (b) Duopoly Passenger Pickup Time (s) (c) Average Pickup Time Increase (s)

(d) Monopoly Passenger Cancellations (s^{-1}) (e) Sum of Duopoly Cancellations (s^{-1}) (f) Cancellation Increase (s^{-1})

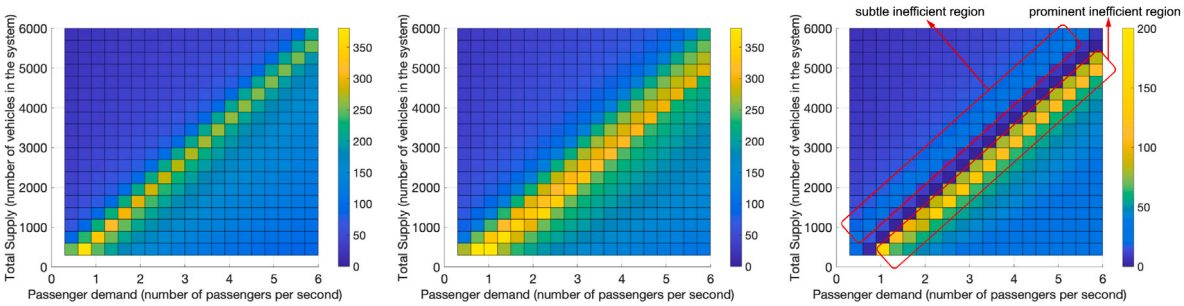
(g) Monopoly Boarding Rate (s^{-1}) (h) Sum of Duopoly Boarding Rate (s^{-1}) (i) Boarding Rate reduction (s^{-1})

(j) Monopoly Profit ($/hr) (k) Sum of Duopoly Profit ($/hr) (l) Profit Reduction ($/hr)

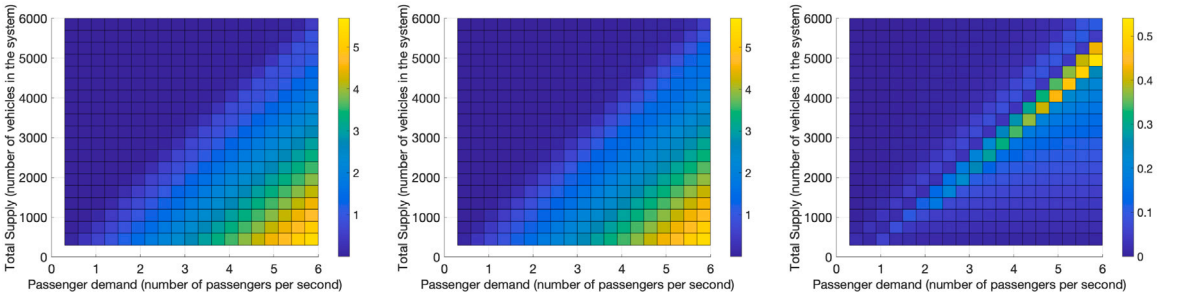
Fig. 2. Performance indicators obtained using the dynamic model. Indicators include average passenger pickup time, cancellations, boarding rate, and profit, for monopoly and duopoly under varying total passenger demand and vehicle supply. Figs. (c), (f), (i), and (l) show the absolute difference in these performance indicators between the duopoly and monopoly.

other is more subtle, while the two are separated by another diagonal region where the effects of market fragmentation are not as considerable.

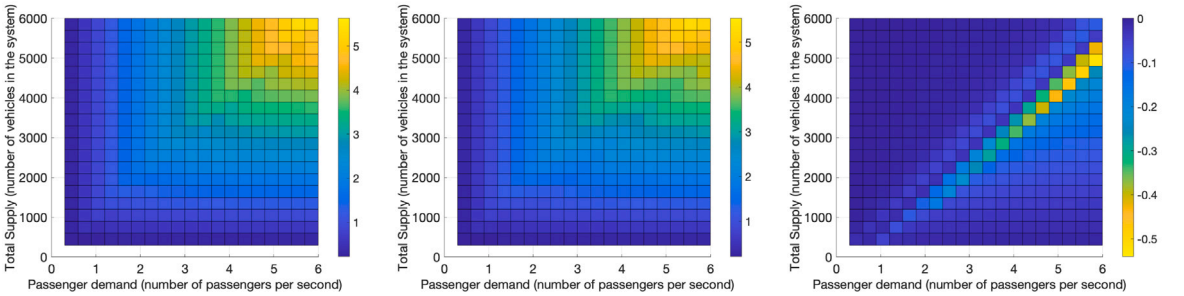
Fig. 4 displays the differences between the results obtained from the two models. In Figs. 4(a), 4(b), 4(e), 4(f), 4(i), and 4(j), we can observe that the r-value between the 2 sets of results ranges 0.88 and 0.98, which implies that they are highly



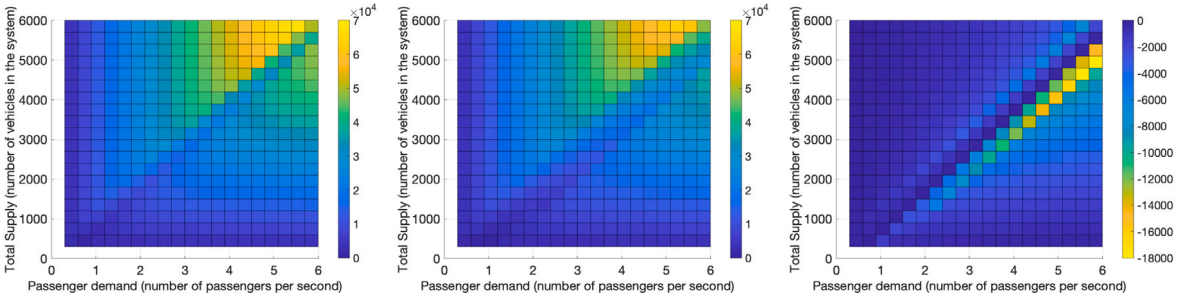
(a) Monopoly Passenger Pickup Time (s) (b) Duopoly Passenger Pickup Time (s) (c) Average Pickup Time Increase (s)



(d) Monopoly Passenger Cancellations (s^{-1}) (e) Sum of Duopoly Cancellations (s^{-1}) (f) Cancellation Increase (s^{-1})



(g) Monopoly Boarding Rate (s^{-1}) (h) Sum of Duopoly Boarding Rate (s^{-1}) (i) Boarding Rate reduction (s^{-1})



(j) Monopoly Profit ($/hr) (k) Sum of Duopoly Profit ($/hr) (l) Profit Reduction ($/hr)

Fig. 3. Performance indicators obtained using the equilibrium model. Indicators include average passenger pickup time, cancellations, boarding rate, and profit, for monopoly and duopoly under varying total passenger demand and vehicle supply. Figs. (c), (f), (i), and (l) show the absolute difference in these performance indicators between the duopoly and monopoly.

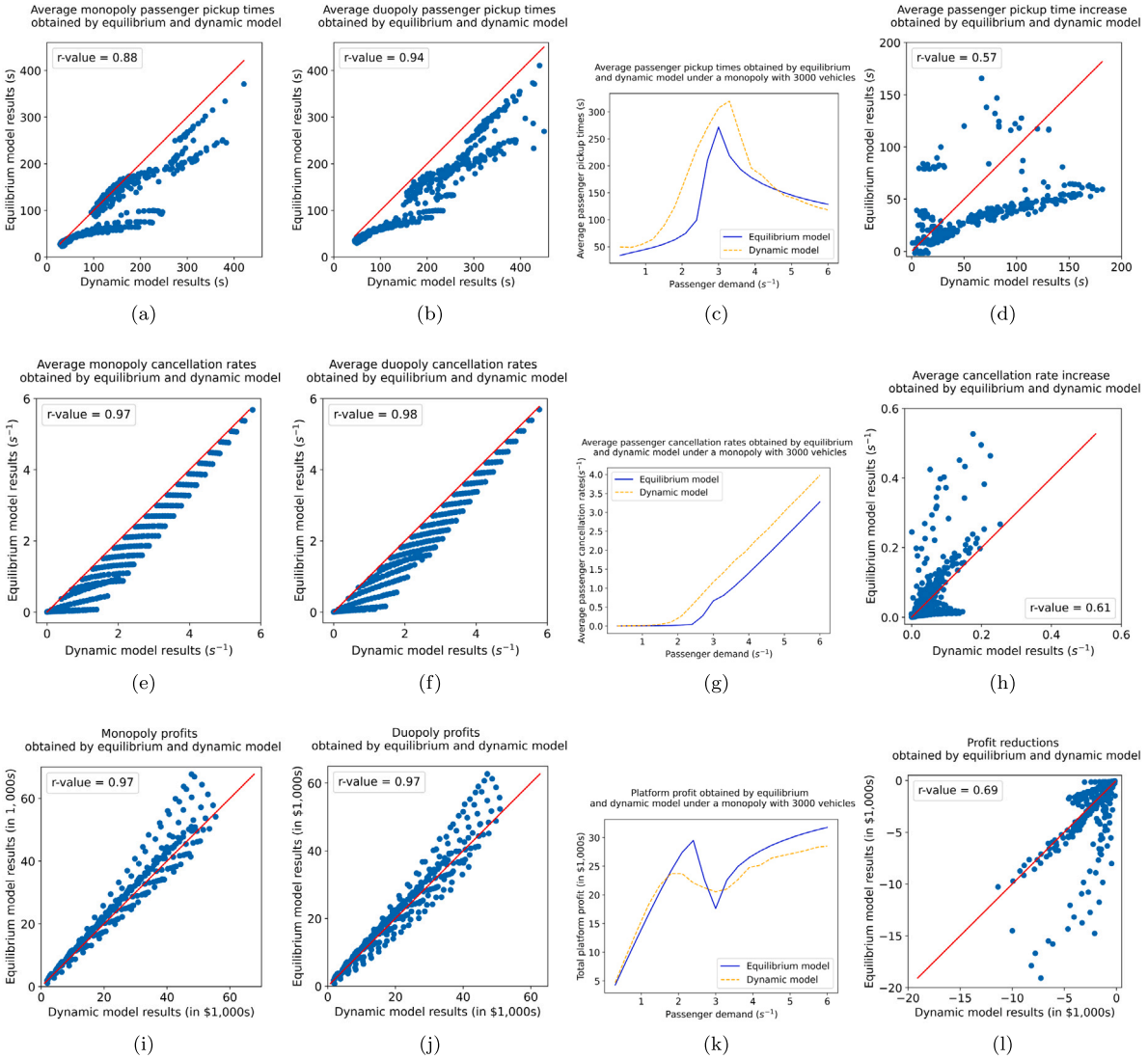


Fig. 4. Results comparison between equilibrium model and dynamic model.

correlated. However, we can also see that the equilibrium model tends to underestimate the average pickup time, and the number of cancellations, while the profit is on par with that obtained by the dynamic model. This observation is reaffirmed in Figs. 4(c), 4(g), and 4(k). On the other hand, the differences between monopoly and duopoly recorded by the two models are seen to have more disparity in Figs. 4(d), 4(h), and 4(l), despite the similarity in trend observed in Figs. 2 and 3.

There are a number of reasons to explain some of the disparities between the results obtained from the two models. Firstly, the calibrated Eq. (6) tends to underestimate average pickup time, as shown in Fig. 11(f), which could contribute to the underestimation observed in Figs. 4(a), 4(b), and 4(c). Consequently, the underestimated pickup time leads to underestimated cancellation. One might expect that the profit should be overestimated in the equilibrium model since the pickup time and cancellation are both underestimated. However, it is not what is observed. This is due to a structural difference between the equilibrium model and the dynamic model. Recall that, after matching, passengers could cancel if the pickup time is too long for them. This cancellation is considered by both models. It is logical that passengers with longer pickup times are more likely to cancel. Therefore, the realized average pickup time must be lower than the average pickup time produced during matching. The equilibrium model does not consider this process and assumes the realized average pickup time is the same as the average pickup time produced during matching. Consequently, the deadheading in the equilibrium model would be overestimated, which reduces profit and counteracts the upward bias. Finally, the increase in disparity between the models when comparing the difference between monopoly and duopoly is likely resulted from the accumulation of discrepancies.

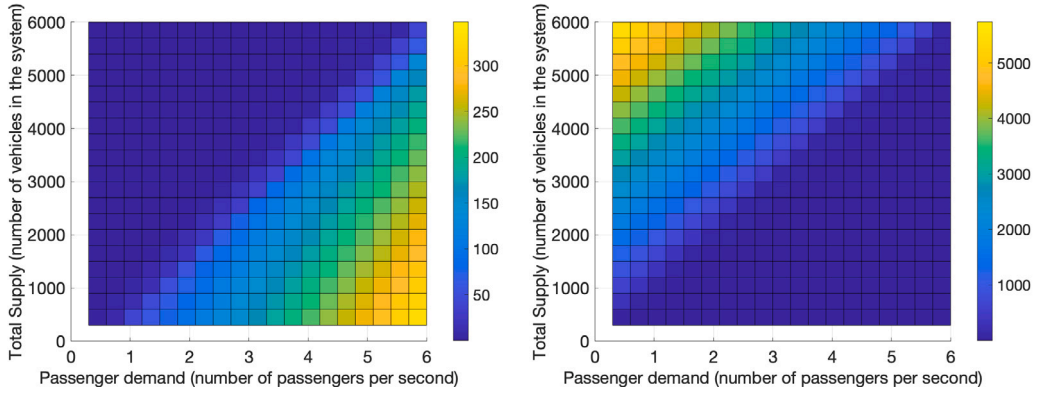


Fig. 5. Equilibrium number of waiting passengers and idle vehicles under varying passenger demand and vehicle supply in a monopoly.

Ultimately, the two models are fundamentally different. That being said, the similarities between the results obtained by the two models, especially the agreement of the key features observed, offer confidence that the proposed equilibrium model can be utilized to provide insight into the multi-platform e-hailing market.

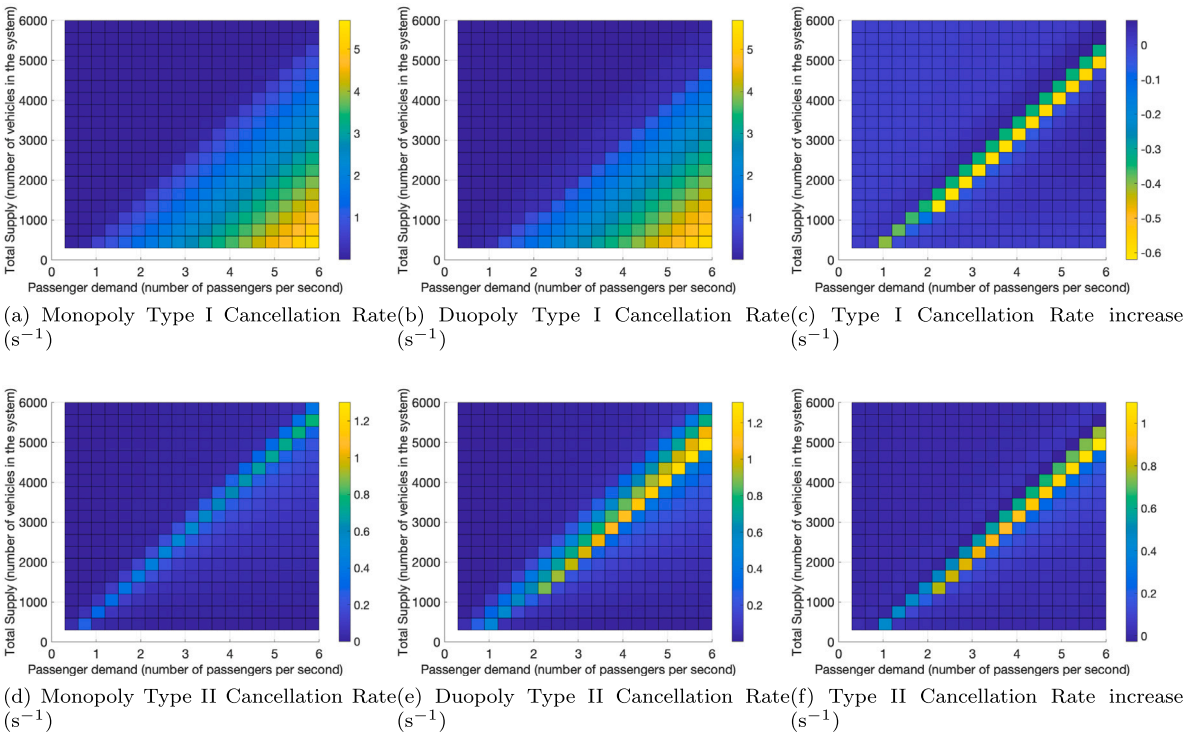
4.3. Symmetrical duopoly vs monopoly

In the previous section, we have demonstrated the applicability of the equilibrium model. In this section, we provide a detailed comparison between a symmetrical duopolistic e-hailing market and a monopolistic e-hailing market looking at Fig. 3.

Figs. 3(a) and 3(b) show the average passenger pickup time for the monopoly and the duopoly, respectively. It can be observed that there are three distinctive market regions where the average pickup time is low, moderate, and high. For each level of supply, when the demand is low, the average passenger pickup time is also low. This is intuitive, as there would be few waiting passengers and many idle vehicles under the static equilibrium (see Fig. 5). Consequently, the waiting passengers can easily be matched with an close-by idle vehicle (see Figs. 11(d) and 11(e) demonstrates the relationship between the number of waiting passengers, the number of idle vehicles and the average pickup time). As demand increases, more vehicles are needed to service the passengers, and the number of idle vehicles reduces, which leads to an increase in pickup time. When the demand increases to a certain threshold for each supply level, the average pickup time abruptly increases to a high level. This implies that the equilibrium number of waiting passengers and idle vehicles are now both at a low level. Therefore it suggests that the market has reached a state where the supply can only just service the demand. As demand continues to increase, the market becomes under-supplied and available vehicles are unable to service all the demand. Subsequently, there will be many waiting passengers and few idle vehicles under equilibrium. This implies that idle vehicles are more likely to be matched to close-by waiting passengers. Thus, the average pickup time reduces to a moderate level for those matched passengers. In Fig. 3(c), we show the average passenger pickup time increase of the duopoly in comparison with the monopoly. It can be observed that there are two diagonal regions where the pickup time of the duopoly is significantly higher than that of the monopoly. One of these inefficient regions is more prominent and the other is more subtle, while the two regions are separated by another diagonal region with minimal pickup time increase. For the market states in the subtle inefficient region, the pickup time increase is around 30 s. Whereas for the market states in the prominent inefficient region, the pickup time increase is steadily above 2 min.

Figs. 3(d) and 3(e) show the total cancellation per second (due to impatience and due to dissatisfaction of service) in the monopoly and the duopoly (sum of two platforms). Intuitively, we can observe that cancellations increase with increasing demand and decreasing supply. Fig. 3(f) shows the increase in total cancellations in the duopoly compared to the monopoly. The observed pattern is similar to that observed in Fig. 3(c), with two distinct regions of cancellation increase. The highest increase in the rate of cancellation can reach over 0.1 and 0.5 passengers per second in the two regions, respectively. Figs. 3(g) and 3(h) show the boarding rate in the monopoly and the duopoly. Boarding rate and cancellations can be considered to be the two sides of the same coin, since boarding rate equates to demand subtracting cancellations. Consequently, Fig. 3(i) is identical to Fig. 3(f). Nonetheless, Figs. 3(g) and 3(h) provide a clearer presentation of an interesting phenomenon. We can observe that for each level of vehicle supply, as the passenger demand increases, the boarding rate increases, reduces, and then increases again. This observed reduction in boarding rate is due to the fact that as demand increases, pickup time increases. As vehicles must travel further to service passengers, the number of idle vehicle reduces, which further increases pickup time, hence creating a vicious cycle. This is precisely the description of the Wild Goose Chase (WGC) problem. Hence, our model shows that due to the presence of the WGC phenomenon, boarding rate could reduce as demand increases, ceteris paribus. This agrees with the theoretical analysis of the WGC problem done by Castillo et al. (2017) and Xu et al. (2020), which we will further elaborate in Section 4.4.

Finally, Figs. 3(j) and 3(k) show the profit of the monopoly and the sum of platform profits of the duopoly, respectively. For each level of vehicle supply, as the passenger demand increases, the profit made by the platform increases, reduces, and then increases



(a) Monopoly Type I Cancellation Rate (s^{-1}) (b) Duopoly Type I Cancellation Rate (s^{-1}) (c) Type I Cancellation Rate increase (s^{-1})

(d) Monopoly Type II Cancellation Rate (s^{-1}) (e) Duopoly Type II Cancellation Rate (s^{-1}) (f) Type II Cancellation Rate increase (s^{-1})

Fig. 6. Type I & II cancellation rates for the monopoly and the duopoly.

again. This is due to the reduction in boarding rate and increase in pickup time (since we assumed drivers are paid during pickup, increasing pickup time reduces the profits earned on each trip). Fig. 3(l) shows the decrease in the total profit generated by both platforms in the duopoly compared to the monopoly profit. Similar to Figs. 3(c), 3(f), and 3(i), there are two separated diagonal regions where the total profit generated by the duopoly is distinctly lower than that by the monopoly. The reduction in total profit is around $5000 per hour in the subtle region. While the highest reduction in total profit generated can reach over $18000 per hour in the prominent region.

The above performance indicators provide sufficient evidence to suggest that the inefficiencies due to market fragmentation are indeed present in the multi-platform e-hailing market. Moreover, it is apparent that there are distinct market states where these indicators concurrently and significantly underperform in the duopoly compared to the monopoly, i.e., the effects of market fragmentation are more prominent in those market states. These market states appear in the demand and supply mesh grid in two separated diagonal regions. In Section 4.4, we attempt to explain this observed phenomenon. Nonetheless, for the within-day operations in the competitive e-hailing market, the demand and supply levels fluctuate, and the market is bound to enter some of those market states, thus it is crucial to tackle these inefficiencies due to market fragmentation.

Now let us scrutinize Type I and Type II cancellations separately. Fig. 6 details how the total cancellations consisted of Type I & II cancellations for both the monopoly and the duopoly. We can observe that the monopoly suffers relatively more from Type I cancellations, as the duopoly generally has fewer Type I cancellations when the market is in the inefficient region. However, the reduction in Type I cancellation rate in the duopoly is overshadowed by the increase in Type II cancellation rate. Consequently, the total cancellation rate in the duopoly is higher than that of a monopoly.

Apart from the difference described above, the observed Type I and II cancellations exhibit similar characteristics for both the monopoly and the duopoly. Let us examine the figures considering any constant level of supply and increasing demand (i.e., moving from left to right). We start with a very low demand and the market is over-supplied, there are almost no cancellations at this stage. It is expected as there would be very few waiting passengers and sufficient idle vehicles.

As demand increases, passengers would first cancel due to dissatisfaction with service (Type II). At this stage, vehicles start being used up, which leads to a relatively low number of idle vehicles. However, the passengers can still be readily matched. Thus, there would still be a relatively low number of waiting passengers. These relatively low numbers of both idle vehicles and waiting passengers result in increasing average pickup times, and consequently, the observed Type II cancellations. We point out that as passengers cancel (Type II) after they are matched, the corresponding vehicles immediately become idle again. Therefore, the relatively low number of idle vehicles is replenished by both trip completions and those failed matches.

Notably, as demand continues to increase, there is a sudden transition from predominantly Type II cancellations to predominantly Type I cancellations. As the transition occurs, the supply can no longer sustain the demand, which suggests the waiting passengers

start being queued up. This leads to a reduction in average pickup time for matched passengers. Consequently, Type II cancellation reduces, which in turn limits the number of vehicles becoming idle due to this cancellation. The increase in the number of waiting passengers coupled with the reduction in idle vehicles further reduces the average pickup times. This process repeats until a new equilibrium is reached. This new equilibrium is characterized by a large number of waiting passengers, a very low number of idle vehicles that primarily come from trip completions, long matching times, and low pickup times, which results in the observed predominantly Type I cancellations.

We point out that, near the narrow corridor where the transition from predominantly Type II to Type I cancellations occurs, the proposed equilibrium model may yield non-unique solutions. The multiple equilibria suggest that there exists a critical state where both scenarios (predominantly Type I or Type II cancellations) could materialize for the same demand and supply. We detail this non-uniqueness in Appendix D. Also, a priority research direction is to scrutinize the cancellation behavior of customers more in detail using field data (Wang et al., 2019; Xu et al., 2022).

4.4. Supply curves of the symmetrical duopoly and monopoly

Castillo et al. (2017) and Xu et al. (2020) scrutinized the supply curve of the e-hailing market. The curve defines a relationship between the number of trips that can be serviced by a given number of drivers and the pickup time of the passengers being serviced. They suggested that the supply curve is backward bending as shown in Fig. 7(a). They viewed the supply curve to have two states, the Good state, and the WGC state. The WGC state is characterized by increased passenger waiting time and reduced trips conducted compared to the Good state.

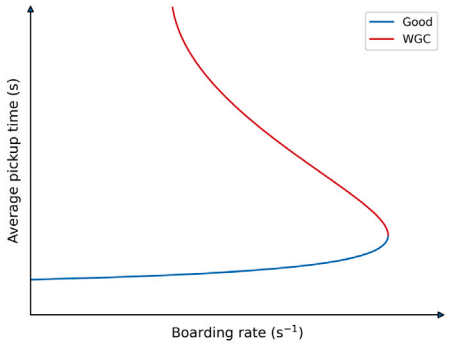
We attempt to recreate the supply curve using the proposed equilibrium model. Since supply and demand are exogenous in the proposed model, for each market state represented by different combinations of supply and demand, we can record its average pickup time and boarding rate. Hence, for each fixed level of total supply, we can vary the total demand in the market to construct a supply curve. In Fig. 7(b), we construct such supply curves for a monopoly and a single platform in the duopoly. Note that the supply curve of the entire market for the duopoly has double the boarding rate comparing a single platform in the duopoly, which is shown in Fig. 7(c). We consider the case where the total supply (V_t) in the market is 4200 vehicles, and the total demand (D_p) is varied between 0.15 to 6 (with increments of 0.15) passengers per second in the market. Note that the solid lines are constructed by connecting the observations for each demand increment. We only present the key data points in the Figure while omitting the rest. In Fig. 7(d), we construct the supply curves for platforms with varying fleet sizes ranging from 300 vehicles to 6000 vehicles with 300 increments.

In the observed supply curves shown in Figs. 7(a) to 7(d), the Good state and the WGC state are distinctly present. However in Figs. 7(b) to 7(d), in addition to the Good and WGC states, we observe that as the demand rate continues to increase, the market can exit the WGC state (we call it the Post WGC state). As the market exits the WGC state, the average pickup time reduces, and the boarding rate increases. The deviation from the supply curve in Fig. 7(a) is due to the average pickup time being used instead of the average waiting time (pickup time and matching time) on the vertical axis. The reason that we did not include matching time is due to the differences in passenger behavior models. Previous studies did not consider passenger cancellations due to impatience, thus the matching time would continue to increase if the fleet size is not sufficient. However, when Type I cancellation is introduced in our model, passengers would cancel when their patience runs out, thus the average matching time would be capped. Therefore, it is not suitable to include the matching time in our analysis. Given our setup, the observation of the Post WGC state is intuitive, as we can imagine when the market is flooded by demand, the vehicles that become idle are likely to be close by to a waiting passenger. Furthermore, as demand increases to infinity, it is logical that the pickup time reduces to zero, since any vehicle that becomes idle after finishing a ride will immediately find a waiting passenger at its position. Additionally, the boarding rate would not increase to infinity, as it is capped by the finite fleet size.

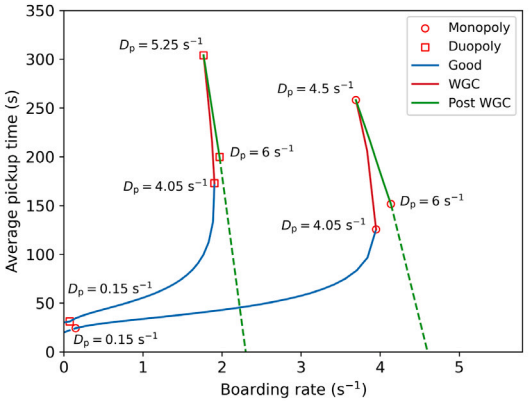
In Fig. 7(b), as the demand rate increases from 0.15 passengers per second, the average pickup time and boarding rate increase simultaneously; this represents the Good state. As the demand rate crosses 4.05 passengers per second for both the monopoly and duopoly, the average pickup time increases while the boarding rate reduces, this represents the WGC state. For the monopoly, we observe that it exits the WGC state after the demand exceeds 4.5 passengers per second. Whereas the duopoly exits the WGC state later when the total demand exceeds 5.25 passengers per second. Therefore, there exist market conditions where the duopoly is still in the WGC state while the monopoly has exited the WGC state.

In Fig. 8, we combine Fig. 7(b) with Fig. 3(c), and we aim to explain the two observed diagonal inefficient regions. For the considered case with 4200 vehicles, as demand increases from 0, when both the monopoly and the duopoly are in the Good state, the market will gradually enter the subtle inefficient region. As demand reaches 4.05 passenger requests per second, both the monopoly and the duopoly enter the WGC state, this coincides with the ending of the subtle inefficient region. As demand continues to increase to 4.5 passenger requests per second, the monopoly first leaves the WGC state, and we observe it corresponds to the beginning of the prominent inefficient region. Finally, when the duopoly leaves the WGC state as the demand reaches 5.25 passenger requests per second, the prominent inefficient region diminishes. Therefore, from these observations, we can deduce the following:

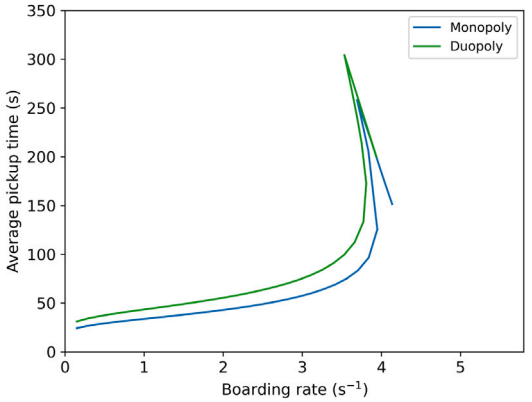
- When both the monopoly and the duopoly are in the Good state, the monopoly slightly outperforms the duopoly which leads to the subtle inefficient region.
- When both the monopoly and the duopoly enter the WGC state, both their performances deteriorate, and the differences in performance between them become minimal.
- Depending on the supply and demand in the market, the monopoly may be able to leave the WGC state while the duopoly is still in the WGC state. This causes the monopoly to significantly outperform the duopoly, which leads to the prominent inefficient region.
- When both the monopoly and the duopoly leave the WGC state, the differences in performance between them ease but remain.



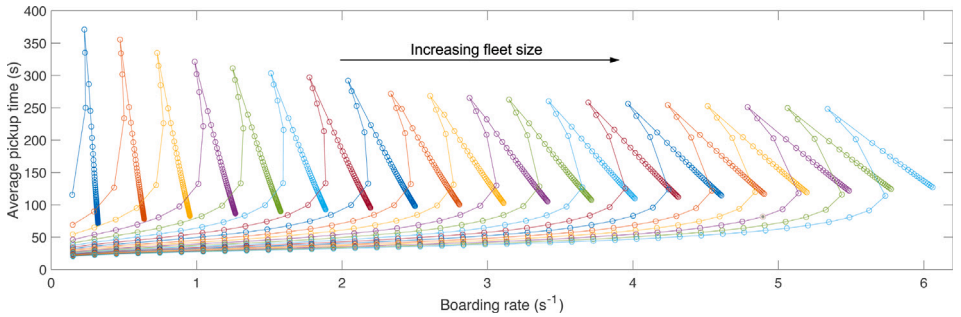
(a) Theoretical backward bending supply curve



(b) Observed supply curves, duopoly (one platform) vs monopoly



(c) Observed supply curves, duopoly (two platforms) vs monopoly



(d) Observed supply curves for platforms with varying fleet sizes

Fig. 7. Theoretical and observed supply curves. (a) shows the theoretical supply curve first introduced by Castillo et al. (2017). (b)–(d) are the observed supply curves by numerical experiments conducted using the proposed model. (b) shows the observed supply curves for a single platform in the duopoly vs monopoly (dotted lines are projections), (c) shows the observed supply curves for the two platforms in the duopoly vs monopoly, and (d) shows the supply curves for platforms with varying fleet sizes.

4.5. Multiple symmetrical platforms

We further utilize the equilibrium model to analyze the effects of having more than two platforms in the e-hailing market. As it would be computationally too expensive, we did not use the dynamic model to replicate the results. We compare the same performance indicators, i.e., average passenger pickup time, total cancellations, and the sum of platform profit, for up to 10 symmetrical platforms in the market. We choose five different combinations of total passenger demand and vehicle supply levels for conciseness. The results are shown in Fig. 9.

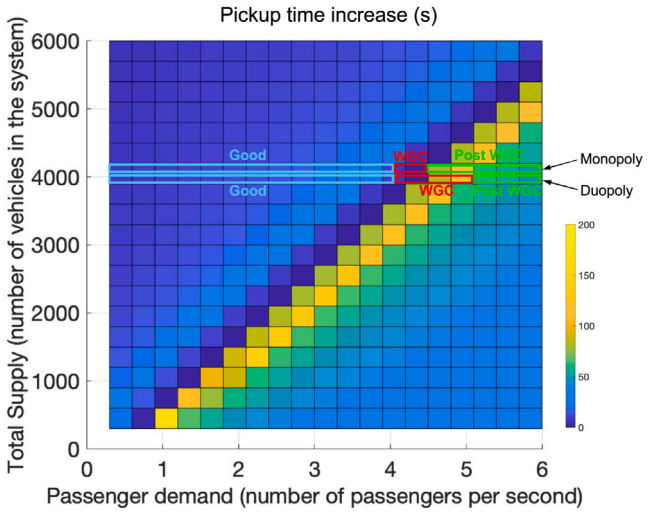
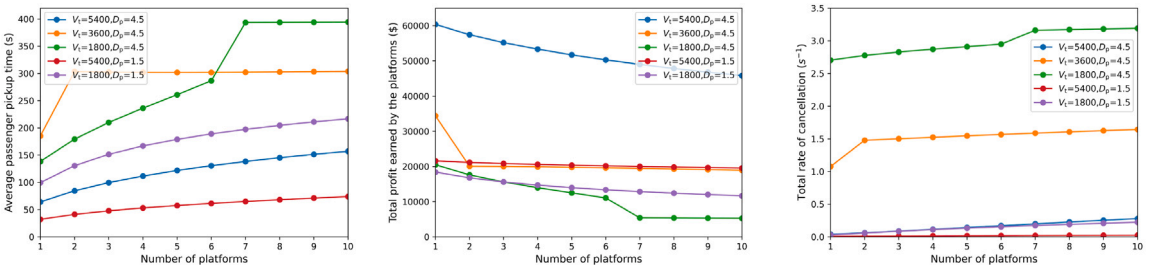


Fig. 8. Relationship between the supply curve and the inefficient regions.



(a) Average Passenger Pickup Time (s) (b) Total Profit ($/hr) (c) Total Rate of Cancellation (s^{-1})

Fig. 9. Average passenger pickup time, total profit, and cancellation rate for up to 10 symmetrical platforms in the market, under 5 different total passenger demand and vehicle supply levels.

It can be observed that there are certain market states where the effects of market fragmentation are not as significant. This occurs when the market is in an over-supplied state, e.g., when $V_t = 5400$ and $D_p = 1.5\ s^{-1}$. The total profit generated by the platforms and the total rate of cancellation worsens only slightly with the increasing number of platforms. The increase in average passenger pickup time is more noticeable but not very substantial, with about a 10 second increase for every doubling of the number of platforms. Therefore when there is sufficient supply in the market, even though the average passenger pickup time increases with the number of platforms, the passengers' decisions may be unaffected, which does not lead to the deterioration of the other performance indicators.

When the market is in the under-supplied state, e.g., when $V_t = 1800$ and $D_p = 4.5\ s^{-1}$, or $V_t = 3600$ and $D_p = 4.5\ s^{-1}$, it can be observed that there is a sudden worsening in the performance indicators when the number of platforms reaches a certain threshold. After the threshold is exceeded, the performance indicators remain roughly the same with a continued increase in the number of platforms. We deduce that when the market is under-supplied, given that the market initially (the monopoly) is not in the WGC state, each increase in the number of platforms has the possibility of plunging all the platforms into the WGC state. The sudden deterioration signals such an occurrence. Additionally, once the platforms are in the WGC state, further market fragmentation does not significantly amplify the inefficiencies.

Finally, when the market is neither over-supplied nor under-supplied, e.g., when $V_t = 1800$ and $D_p = 1.5\ s^{-1}$, or $V_t = 5400$ and $D_p = 4.5\ s^{-1}$, we do not observe sudden worsening in the performance indicators as the number of platforms increases. The reduction in performance is more significant when the number of platforms is lower, as the market becomes more fragmented, the performance gradually declines further.

4.6. Asymmetrical duopoly vs monopoly

In this section, we utilize the equilibrium model to consider asymmetrical duopolies. The asymmetry can be present in both passenger demand and vehicle supply. We measure the asymmetries using the ratios of vehicle supply and the ratios of passenger

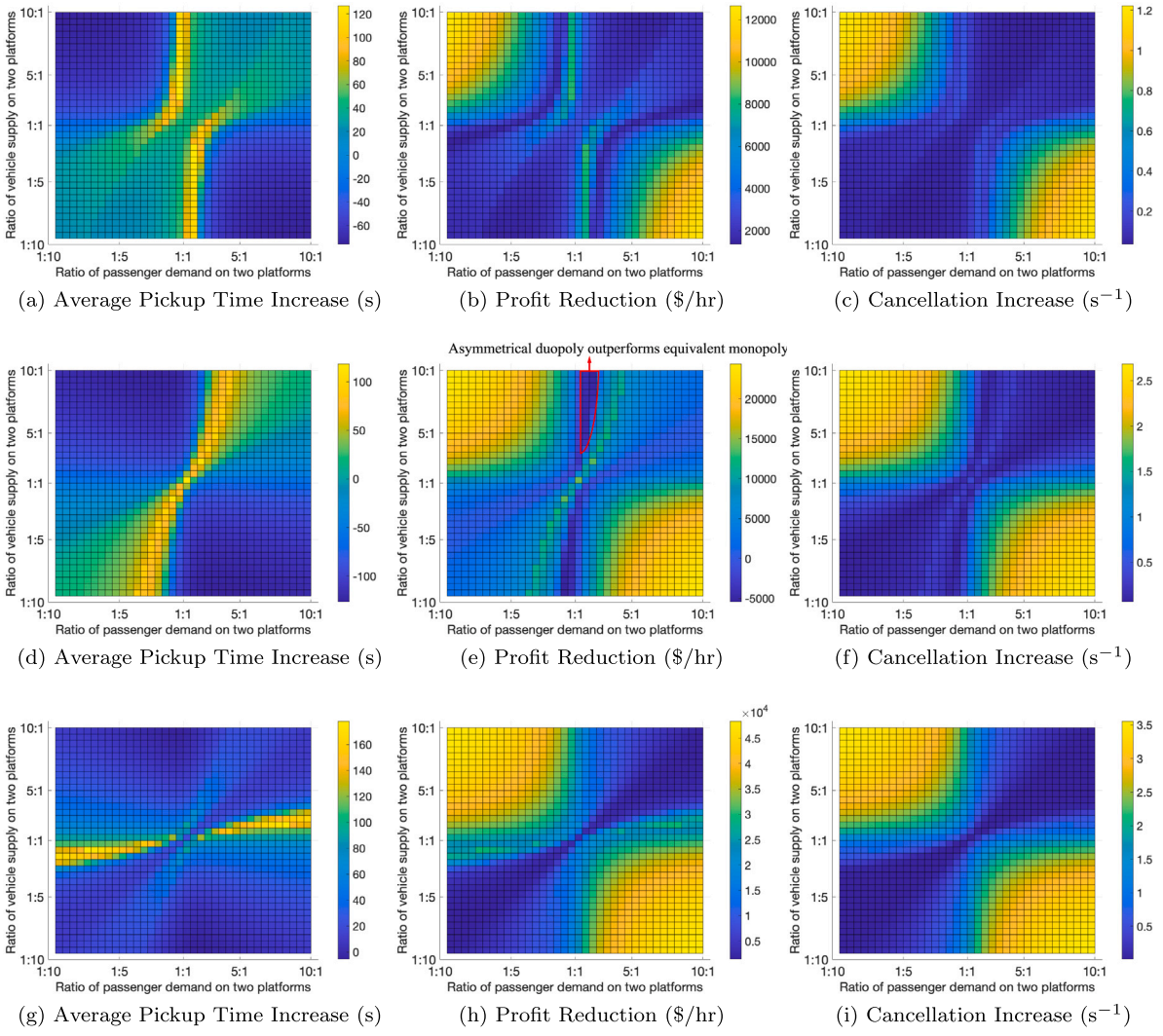


Fig. 10. Asymmetrical duopoly performance indicators compared to equivalent monopolies. (a)–(c): $V_t = 1800$, $D_p = 4.5$ (s^{-1}); (d)–(f): $V_t = 3600$, $D_p = 4.5$ (s^{-1}); and (g)–(i): $V_t = 5400$, $D_p = 4.5$ (s^{-1}).

demand from the two platforms. E.g., with $V_t = 300$, and a vehicle supply ratio of 1:2, the two platforms would have fleets of 100 vehicles and 200 vehicles, respectively. For conciseness, we consider three scenarios with different combinations of total passenger demand and vehicle supply levels. For each of the total demand and supply combinations, we further test different combinations of vehicle supply ratios and passenger demand ratios. For each test conducted, the performance indicator of the asymmetrical duopoly is compared to an equivalent monopoly. The results are shown in Fig. 10. Note that the average pickup time of the asymmetrical duopoly is weighted based on each platform's matching rate. Additionally, each figure in Fig. 10 is symmetrical along the diagonals, as a ratio of 1:2 is equivalent to a ratio of 2:1 from the point of view of the other platform. The top right and bottom left parts in each figure represent asymmetrical duopolies where one platform has greater shares of both demand and supply, while the other platform has smaller shares of both demand and supply. On the other hand, the top left and bottom right sectors represent asymmetrical duopolies where one platform has a greater share of demand, but a smaller share of supply, while the other platform is the opposite. Furthermore, the center of the figure is equivalent to the symmetrical duopoly for the given total demand and supply.

Figs. 10(a)–10(c) showcase an overall extremely under-supplied scenario. It can be observed that in the top left and bottom right parts, the asymmetrical duopolies significantly underperform compared to the equivalent monopoly or symmetrical duopolies. The cancellations and profit reduction are substantially increased. Though the average pickup times for matched passengers are reduced as there are much fewer matched passengers. This is expected as these parts represent cases where the asymmetry leads to a further imbalance of supply and demand for both platforms in the duopolies. Another interesting observation is that when the demand is shared close to evenly, as the asymmetry in supply increases, the duopolies also significantly underperform with increased pickup times, profit reduction, and cancellations increase. We reason that under this overall under-supplied scenario, a symmetrical duopoly

would imply both platforms to be in the Post WGC state. However, if the supply is not shared symmetrically, the platform that has the larger fleet may enter the WGC state, which leads to the overall underperformance observed.

Figs. 10(d)–10(f) showcase an overall slightly under-supplied scenario. In these figures, we can draw the same observation that the asymmetrical duopolies significantly underperform in the top left and bottom right sectors. However, it is interesting to point out that, under this scenario, when the demand is shared evenly, as the asymmetry in supply increases, the asymmetrical duopoly can outperform an equivalent monopoly. This is because, with the total demand and supply under this scenario, a monopoly would be in the Post WGC state. While the two platforms in a symmetrical duopoly would be in the WGC state. Therefore, as the supply increases for one platform and decreases for the other platform, that is the duopoly becomes asymmetrical, one platform may enter the Good state and the other leave the WGC state. Consequently, the overall performance of the asymmetrical duopoly is improved and can exceed the equivalent monopoly. Another implication of this observation is that, when a monopoly finds itself in the situation described by this scenario (with the given demand, supply, and matching algorithm), there must exist certain strategies that would improve its efficiency.

Figs. 10(g)–10(i) showcase an overall over-supplied scenario. A monopoly and the two platforms in a symmetrical duopoly would all be in the Good state. We notice that when either the asymmetry in supply or demand increases (while the other is shared close to evenly), the duopolies significantly underperform compared to an equivalent monopoly. Again, this is due to one of the platforms entering the WGC state.

There are many scenarios and variables to consider when analyzing the inefficiencies in asymmetrical duopolies. However, it is helpful to investigate where each platform is on its respective supply curves. In summary, allowing as many platforms in the Good state, and as few platforms in the WGC state, would lead to an overall improvement in the efficiency of asymmetrical duopolies.

5. Summary and future work

This paper has presented an equilibrium model of the multi-platform e-hailing market. The proposed model considers exogenous total market supply and demand, which allows us to analyze the e-hailing market under varying combinations of demand and supply levels illustrating different market states. Furthermore, the model allows two types of passenger cancellations. It utilizes a queueing system with reneging to model passenger cancellation due to impatience, and it utilizes a utility choice model to consider passenger cancellation due to unsatisfactory service. We also introduced a disaggregated dynamic model of the same multi-platform e-hailing market. The two models produced comparable results, which validates the proposed equilibrium model. Hence, using the equilibrium model, we achieved the following:

- We first isolate and quantify the inefficiencies due to fragmentation in the symmetrical duopoly e-hailing market. We identify two regions in the demand and supply domain, where a duopoly significantly underperforms compared to a monopoly in these market states, i.e., the effects of market fragmentation are more prominent in the market states represented by these regions.
- We then show that the inefficiencies in a fragmented market can be explained by analyzing whether the platforms are in the Good, WGC, or Post WGC states. For example, it can be shown that the symmetrical duopoly will be in the Wild Goose Chase (WGC) state for a broader range of market conditions compared to the monopoly, which gives rise to inefficiencies in a fragmented market.
- We further extend the analysis to more than two platforms. We show that there is a probability of sudden deterioration in market performance occurring with each additional platform in the market. The number of platforms when the sudden deterioration occurs, or if it occurs at all, depends on the total demand and supply in the market.
- Finally, we investigated the performance of asymmetrical duopolies. We show that for some given total demand and supply, certain asymmetrical duopolies may outperform an equivalent monopoly. It implies when a monopoly finds itself in this situation where it can be outperformed, there must exist certain strategies that would improve its efficiency.

This study considers the isolated effects of market fragmentation. However, a real e-hailing market can be much more dynamic. Therefore, future studies can investigate the effects of competition in the e-hailing market by relaxing some of the assumptions in this study. Some examples include allowing for driver's decision making (Yang et al., 2024) based on factors such as wage, pricing (Fayed et al., 2024), considering passengers' and drivers' multi-homing behaviors, examining the effects of traffic congestion (Alisoltani et al., 2021; Zhang and Zhang, 2022), investigating walking integration in e-hailing services (Zhang et al., 2025), etc. Furthermore, in this study, we utilize a queueing system with reneging, it would be worthwhile to obtain an analytical solution to such a queuing system, as it may well be applicable in other fields. Given the identified inefficiencies in the fragmented e-hailing market, a future research direction is to explore various strategies that may address these inefficiencies (Jiao and Ramezani, 2024). Finally, a priority research direction is to use real data to model the cancellation behavior of customers (Type I and Type II), and to scrutinize the relationship between these cancellation types and market conditions.

CRediT authorship contribution statement

Guipeng Jiao: Writing – review & editing, Writing – original draft, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Mohsen Ramezani:** Writing – review & editing, Writing – original draft, Validation, Supervision, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization.

Acknowledgment

This research was partially funded by the Australian Research Council (ARC) Discovery Early Career Researcher Award (DECRA) DE210100602.

Appendix A. Matching rate and average pickup time estimation

The passengers and vehicles are systematically matched over time by the platform in an e-hailing system. In practice, most platforms use batch-matching algorithms. These algorithms allow the waiting passengers and idle vehicles to accumulate over a small period of time before establishing matchings between the two groups. On an aggregated level, the matching rate, m, and the average passenger pickup time, w_p, are the outcomes of the matching algorithm.

For a given road network (geometry and speed) and a given matching algorithm, the matching rate and the average pickup time are dependent primarily on the number of waiting passengers and idle vehicles. In the dynamic model showcased in Section 3, we considered the platforms utilizes a pickup time minimization algorithm with finite search radius (Eqs. (24)–(28)). In this section, we aim estimate the outcome of this algorithm when employed in the Manhattan New York road network. We proceed to estimate the outcome of the pickup time minimization algorithm using Eqs. (4) and (6). The matching function in Eq. (4) can be readily transformed into a linear function as follows:

$$\ln(m) = \ln(A_0) + \alpha_1 \ln(p_w) + \alpha_2 \ln(v_i). \tag{29}$$

Similarly, Eq. (6) can also be transformed:

$$\ln(w_p) = \ln(\Gamma_0) + \gamma_1 \ln(\min(v_i, p_w)) + \gamma_2 \ln(\max(v_i, p_w)). \tag{30}$$

We execute the matching algorithms for varying numbers (2:2:100) of waiting passengers and idle vehicles, where the origins and destinations of the passengers and the positions of the vehicles are random on the network. For each combination of the number of waiting passengers and idle vehicles, each matching algorithm is executed 20 times. Therefore a total of 50,000 executions are performed for each matching algorithm. The number of matches and the average pickup distance for each execution are documented. We divide the number of matches by the matching interval, $\Delta = 10$ s, to obtain the matching rate. We then use the linear best fit to estimate the parameter values, which are listed in Table 5.

The estimated parameter values are statistically significant at the 1% level, and the R^2 levels are all above 0.89. We further illustrate the comparison between the simulated and estimated matching rate and pickup time in Fig. 11. It can be observed that the estimations from the two calibrated functions are able to capture the characteristics of the simulated results.

Appendix B. Order of service for waiting passengers

A number of studies modeled the e-hailing market utilizing queueing theory (Banerjee et al., 2015; Bai et al., 2019; Zhang et al., 2020). The queueing systems considered in those studies all assume the order of service is FIFO. However, this assumption may not be accurate in practice when platforms utilize a batch-matching algorithm. Note that for typical queueing models, e.g., M/M/1, the order of service does not affect the expected values of the measurements such as queue length, but only the distributions (Flatto, 1997). Therefore, relaxing the FIFO assumption does not impact the analytical solutions of the previous studies. However, as our proposed model considers a queue system with reneging, the order of service does have an effect on the solutions. Hence in this section, we show that the order of service for passengers in the e-hailing system is not necessarily FIFO, and we allow this assumption to be relaxed in our model.

When a platform utilizes a batch matching algorithm, where the waiting passengers and idle vehicles are accumulated for a brief period of time before they are matched, then it is not guaranteed that the passengers who enter the platform earlier are prioritized. To test this hypothesis, we document the mean entrance time (trip order time) of matched passengers, and the mean entrance time of all passengers in the queue just prior to matching, for each matching instance in a 1 hour simulation experiment (with the pickup time minimization algorithm). If the order of service inclines towards FIFO, then we would expect the mean entrance time of matched passengers to be earlier (lower) than the mean entrance time of all passengers in the queue (i.e., waiting passengers in the matching process). We plot the histogram of the difference between the mean entrance time of matched passengers and the mean entrance time of all passengers in the queue in Fig. 12.

It can be observed in Fig. 12 that the center of the distribution is clearly above zero. It shows that on average, the mean entrance time of matched passengers is greater (later) than the mean entrance time of all passengers in the queue. Therefore, it implies that for any given group of waiting passengers, the matching algorithm is more likely to match passengers who joined the platform

Table 5
Parameter values for Eqs. (4) and (6) calibrated based on the pickup time minimization algorithm.

	Parameter values			R^2
Eq. (4)	$A_0 = 0.0111$	$\alpha_1 = 0.7327$	$\alpha_2 = 0.7309$	0.916
Eq. (6)	$\Gamma_0 = 972.4$	$\gamma_1 = 0.0891$	$\gamma_2 = -0.3934$	0.894

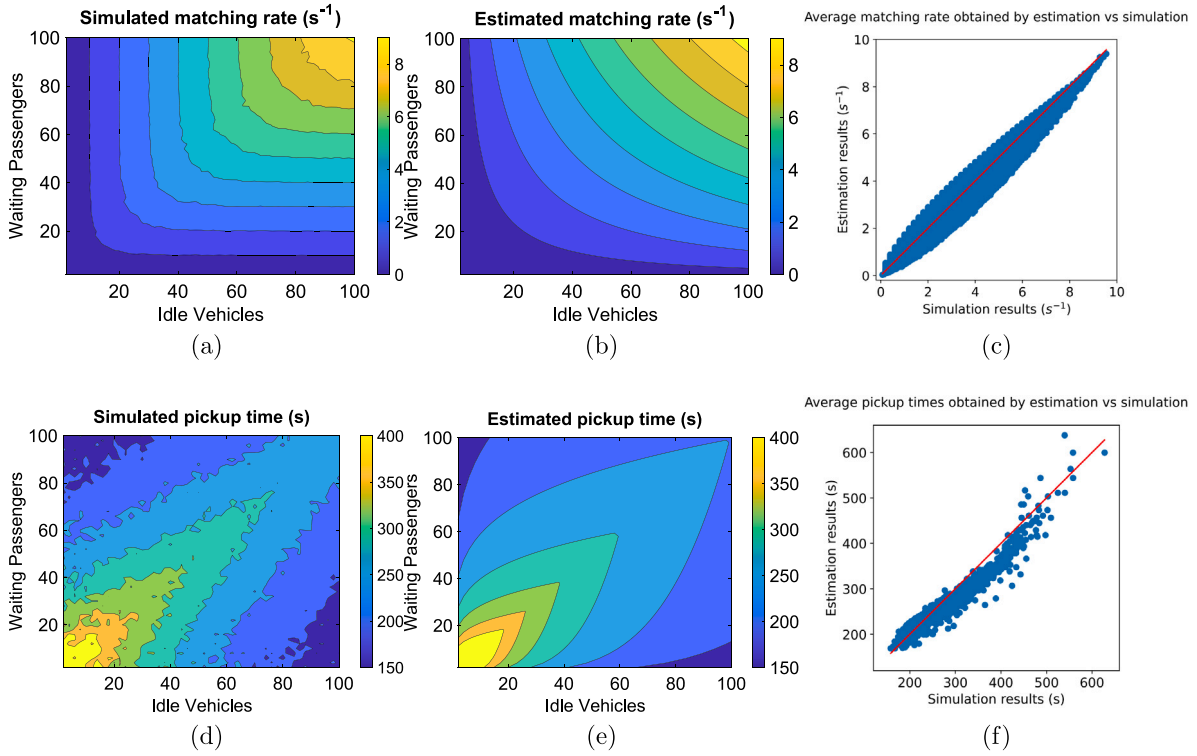


Fig. 11. Simulated vs estimated matching rates and pickup times for the pickup time minimization algorithm.

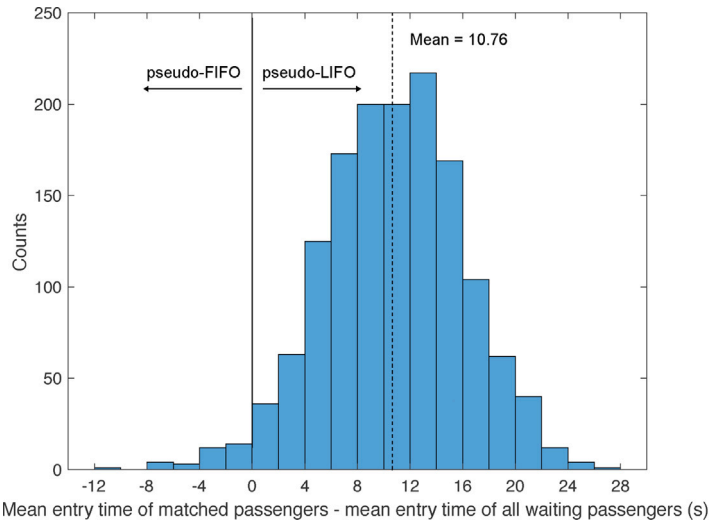


Fig. 12. Histogram of the mean entry time of matched passengers minus mean entry time of all waiting passengers. A distribution with a mean time difference less than 0 implies pseudo-FIFO service, and a mean time difference greater than 0 implies pseudo-LIFO service.

later rather than earlier. Consequently, it suggests that the order of service inclines towards LIFO rather than FIFO. This might be considered intuitive, since those who have been waiting longer indicate they might be further away from any vehicles, which means they are also less likely to be matched in the future.

Since the simulation experiment was performed under only one supply and demand setting, the observation may be different under different settings. However, some thought experiments can be conducted to show the order of service to be non-FIFO. Consider an extremely over-supplied market, such that any passenger would always get a match in their first matching instance (i.e. the queue is always cleared by each matching). Also consider an extremely under-supplied market, such that there would be at most 1 match

per matching instance. It is straightforward to show that the order of service for those two extreme scenarios based on current definitions would be random.

Appendix C. Monte Carlo simulation of queue with reneging

In this section, we perform Monte Carlo simulations of a queueing system with reneging described in Section 2.3. We consider the queueing system with the following characteristics: Poisson arrival process, with arrival rate d_p; exponentially distributed service times, with service rate m; random order of service; and impatient customers who leave the queue if they are not serviced within a certain time. The random patience threshold, τ, is a random variable drawn from a truncated normal distribution. The distribution has mean, standard deviation, and lower and upper bounds denoted as $\bar{\tau}, \sigma_\tau, a_\tau, b_\tau$, respectively, i.e., $\tau \sim TN(\bar{\tau}, \sigma_\tau, a_\tau, b_\tau)$.

We aim to obtain the expressions for the following endogenous variables: the expected queue size, p_w; the rate of cancellation, c_p^1; the average matching time for those who are matched (observed matching time), w_m; and the average time unmatched passengers spent in the queue (observed patience time), $\bar{w}_m^*$.

For a given distribution of customer impatience, the endogenous variables of interest are dependent only on the arrival rate and the service rate. Therefore, we ran simulations of the queueing system with different combinations of arrival and service rates, for different impatience distributions. The pseudo-code for the simulation is presented in Algorithm 1.

Algorithm 1: Simulation of queue with reneging

for $i \leftarrow 1$ **to** N **do**
- $t_i^a \leftarrow t_{i-1}^a - \frac{\ln(u_i^a)}{d_p}$, where $t_0^a = 0$ and $u_i^a \sim$ Uniform(0,1) // Customer i arrival time
- $t_i^r \leftarrow t_i^a + \tau$, where $\tau \sim TN(\bar{\tau}, \sigma_\tau, a_\tau, b_\tau)$ // Customer i patience limit
- $t_i^s \leftarrow -\frac{\ln(u_i^s)}{m}$, where $u_i^s \sim$ Uniform(0,1) // Customer i service time

end

initialize:
- $Q \leftarrow \emptyset$ // Empty queue
- $t^a \leftarrow 0$ // Time next customer arrives
- $t^s \leftarrow 0$ // Time next service begins
- $t^r \leftarrow 0$ // Time next customer reneges
- $n \leftarrow 1$ // Next arrival number

while $n \leq N$ **do**
- $t^a \leftarrow t_n^a$ // Update time next customer arrives
- **if** Q is *empty* **then**
  - $t^s \leftarrow \max(t^s, t_n^a)$ // Update time next service begins
  - $t^r \leftarrow \infty$
- **else**
  - $t^r \leftarrow \min(Q)$ // Update time next customer reneges
- **end**
- **if** $t^a \leq t^s$ and $t^a \leq t^r$ **then**
  - append t_n^r to Q // New customer joins queue (record renege time)
  - $n \leftarrow n + 1$ // Update next arrival number
- **else if** $t^s \leq t^a$ and $t^s \leq t^r$ **then**
  - service random customer j in Q
  - $t^s \leftarrow t^s + t_j^s$ // Update server busy time
  - remove t_j^r from Q // Remove serviced passenger from queue
- **else**
  - the customer k with $t_k^r = t^r$ reneges
  - remove t_k^r from Q // Remove reneged passenger from queue
- **end**

end

For each simulation, we let N = 30,000. The simulation is ended once every customer has joined the queue. We record the average queue size (p_w) and average rate of cancellation (c_p^1) in the final 1000 s of the simulation. We record the average matching time (w_m) for the last 1000 customers who have been serviced. If the total number of cancellations exceeds 1000, we record the average patience ($\bar{w}_m^*$) of the last 1000 customers who renege. If the total number of cancellations is between 100 and 1000, we record the average patience of all customers who renege. Finally, if the total number of cancellations is below 100 (around 0.3%), we assume there are no cancellations and do not record the average patience. This is to avoid the low sample size that could skew the measurement. We consider five customer patience distributions as shown in Table 6. For each patience distribution, we perform

Table 6
Customer patience distributions tested.

$\bar{\tau}$ (s)	σ_τ (s)	a_τ (s)	b_τ (s)
20	10	0	40
30	10	10	50
40	10	20	60
50	10	30	70
60	10	40	80

Table 7
Estimated parameter values and the Mean Squared Errors of the regressions for different customer patience distributions.

$\bar{\tau}$ (s)	ξ_0	ξ_1	MSE	ϕ_0	ϕ_1	ϕ_2	MSE
20	0.0565	0.0825	1.9026	0.5488	0.0684	0.0459	2.7845
30	0.0255	0.0625	3.2300	0.9257	0.0418	0.0324	1.0761
40	0.0165	0.0487	5.3566	0.9879	0.0291	0.0246	0.6088
50	0.0117	0.0398	7.5469	0.9408	0.0209	0.0198	0.3971
60	0.0091	0.0336	10.657	0.9124	0.0164	0.0166	0.2698

the simulation under varying combinations of arrival rates, $d_p = (0.3 : 0.3 : 9)$, and service rates, $m = (0.3 : 0.3 : 9)$. For each of the 900 total combinations, 5 simulations are performed and the results are documented.

The simulation results are visualized in Fig. 13. It can be observed that for all customer patience distributions, the endogenous variables of interest exhibit similar patterns with variations in arrival and service rate.

It can also be observed that for the average matching times, their values remain roughly the same on every line parallel to $d_p = m$. Therefore, we can assume w_m is a function of $(d_p - m)$. Additionally, we can observe that as $(d_p - m)$ reduces to zero, the average matching time also declines to roughly zero. When the average matching time is non-zero, the following functional form can describe the relationships between w_m and $(d_p - m)$:

$$w_m = \frac{(d_p - m)}{\xi_0 + \xi_1 (d_p - m)}. \tag{31}$$

As m exceeds d_p, the average matching time stays at zero. Hence:

$$w_m = \begin{cases} \frac{(d_p - m)}{\xi_0 + \xi_1 (d_p - m)} & \text{if } m \le d_p \\ 0 & \text{if } m > d_p. \end{cases} \tag{32}$$

The plotted average observed patience times exhibit similar patterns as the average matching times. Therefore, it is sufficient to assume $\bar{w}^*_m$ is also a function of $(d_p - m)$. However as $(d_p - m)$ reduces to zero, the observed patience remains positive, depending on the patience distribution, it may never reach zero. Therefore, we assume the following functional form for $\bar{w}^*_m$:

$$\bar{w}^*_m = \frac{\phi_0 + (d_p - m)}{\phi_1 + \phi_2 (d_p - m)}. \tag{33}$$

When m becomes greater than d_p and exceeds a certain threshold, there might not be enough cancellation to record the observed patience. We have plotted those points as zero in the figures. However, since for $m > d_p$, the cancellation rate is reduced to essentially zero, such that we can also consider the observed patience as zero analytically. Consequently, we have:

$$\bar{w}^*_m = \begin{cases} \frac{\phi_0 + (d_p - m)}{\phi_1 + \phi_2 (d_p - m)} & \text{if } m \le d_p \\ 0 & \text{if } m > d_p. \end{cases} \tag{34}$$

We then use the simulated results for $m \le d_p$ and Gauss–Newton method to obtain the parameter values of Eqs. (31) and (33) that minimize the respective squared errors. We show the estimated parameter values in Table 7.

It is tempting to use the simulated results to also create expressions for p_w and c^1_p. However, Eqs. (7) and (8) already provide an analytical expression for these two variables with given w_m and $\bar{w}^*_m$. Therefore, with the two estimated functions for w_m and $\bar{w}^*_m$, and the two analytical equations, we can estimate the expected values of the endogenous variables of the queueing system for any given d_p and m and a given customer patience distribution in Table 6. We compare the estimated results to the simulated results for the case of $\bar{\tau} = 60$ in Fig. 14. It can be observed that the estimated w_m and $\bar{w}^*_m$ provided by the calibrated functions closely replicate the simulated results. Furthermore, the estimated average cancellation rates dictated by Eq. (7) are almost identical to the simulated average cancellation rates. Finally, the estimated queue size also captures the nuances of the simulated results. Therefore, Eqs. (7) and (10) can describe the characteristics of the simulated queueing system accurately.

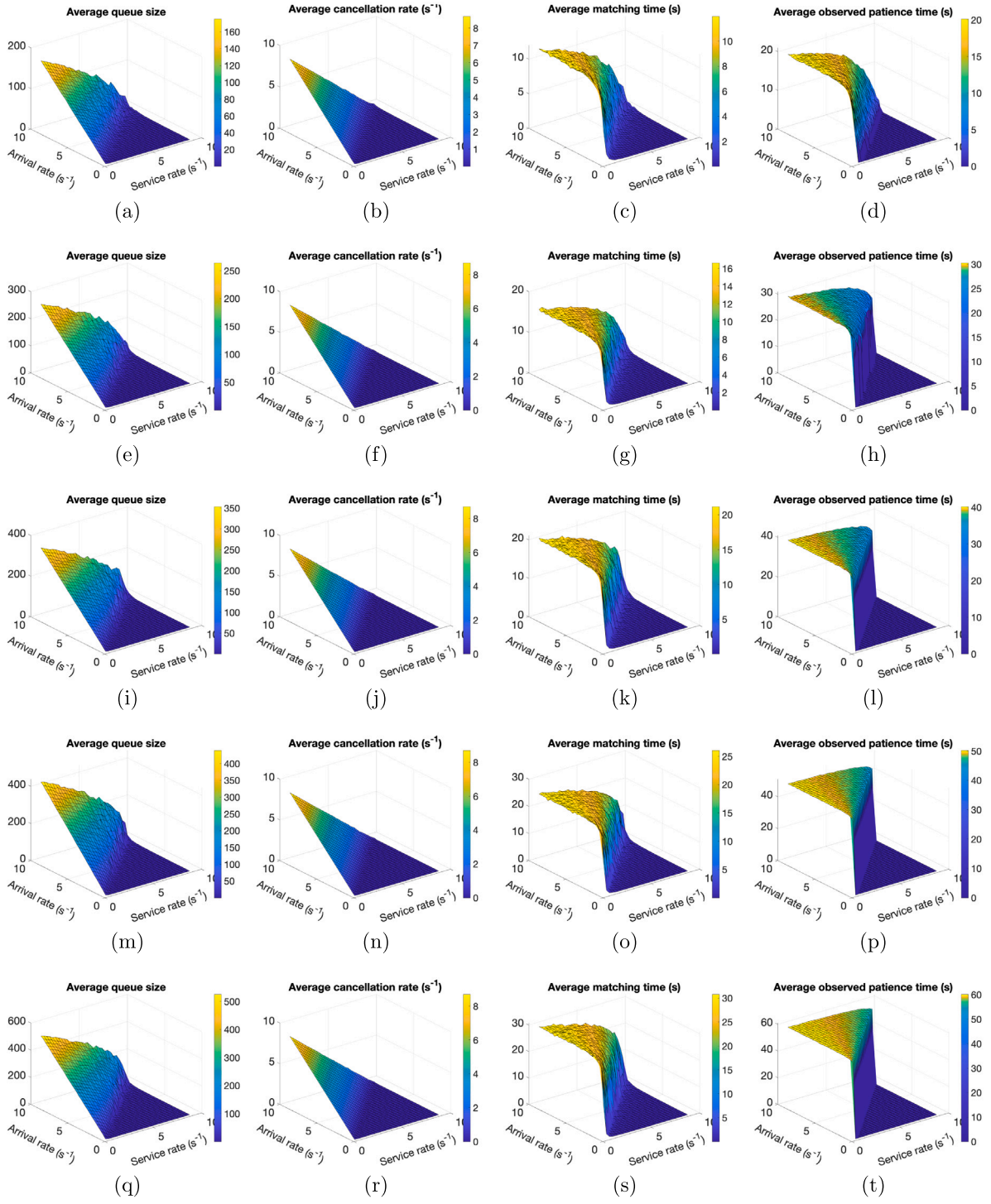


Fig. 13. Simulation results of the queueing system with reneging for different customer patience distributions. (a)–(d): $\bar{\tau} = 20$; (e)–(h): $\bar{\tau} = 30$; (i)–(l): $\bar{\tau} = 40$; (m)–(p): $\bar{\tau} = 50$; and (q)–(t): $\bar{\tau} = 60$.

Appendix D. Non-uniqueness in the solutions of the equilibrium model

It has been observed that there could be multiple equilibria in the proposed equilibrium model. With the exogenous variables and parameters set out in Tables 3 and 4, the non-uniqueness in the solutions of the equilibrium model are observed to occur in

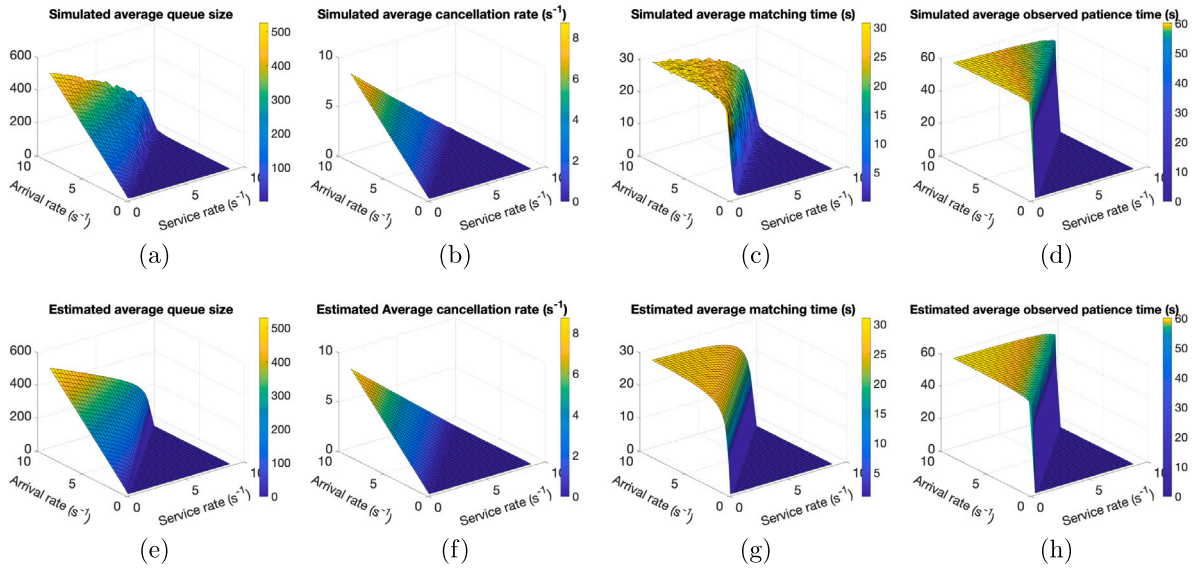


Fig. 14. Estimated vs simulated results of the queueing system for the case of $\bar{\tau} = 60$.

Table 8
Three non-unique solutions obtained for the case of a monopoly with $D_p = 1.2$ and $V_t = 600$.

No.	p_w	v_m	v_i	v_o	w_p	w_m	w_v	$\bar{w}_w^*$	m	b	c_p^1	c_p^2	a_v
1	21.2	174	23	403	371	11.4	22.4	56.3	1.03	0.47	0.17	0.56	0.47
2	23.5	174	20	406	367	12.4	20.0	56.4	1.01	0.47	0.19	0.53	0.47
3	45.7	134	5	461	251	20.3	8.59	57.3	0.62	0.54	0.58	0.09	0.54

Table 9
Three non-unique solutions obtained for the case of a monopoly with $D_p = 3.3$ and $V_t = 2700$.

No.	p_w	v_m	v_i	v_o	w_p	w_m	w_v	$\bar{w}_w^*$	m	b	c_p^1	c_p^2	a_v
1	46.7	685	48	1967	300	11.8	15.3	56.3	3.12	2.29	0.18	0.84	2.29
2	47.8	684	47	1969	299	12.0	14.9	56.4	3.12	2.29	0.18	0.83	2.29
3	88.5	543	21	2136	218	20.3	7.63	57.3	2.72	2.48	0.58	0.23	2.48

Table 10
Three non-unique solutions obtained for the case of a symmetric duopoly with $D_p = 3$ and $V_t = 1800$.

No.	p_w	v_m	v_i	v_o	w_p	w_m	w_v	$\bar{w}_w^*$	m	b	c_p^1	c_p^2	a_v
1	24.9	252	28	620	350	11.5	21.0	56.3	1.33	0.72	0.17	0.61	0.72
2	28.5	251	23	626	345	13.0	18.1	56.4	1.29	0.73	0.21	0.56	0.73
3	51.4	200	8	692	249	20.2	8.86	57.3	0.93	0.80	0.57	0.13	0.80

only certain regions in the supply and demand mesh grid. In Fig. 15, the supply and demand combinations that yield non-unique solutions are highlighted in red. The figure also shows the Type I & II cancellation rates for the monopoly and the duopoly, with randomly chosen solutions when multiple equilibria exist.

It can be seen in Fig. 15 that the supply and demand combinations with non-unique solutions clearly correlate with the Type I & II cancellations. As demand increases for each level of supply, most of the cancellations are due to dissatisfaction of service (Type II). When demand continues to increase, we observe that there is a sudden transition to cancellation due to impatience (Type I). It is clear that the supply and demand combinations with non-unique solutions coincide with the occurrence of the transition.

Additionally, in these supply and demand combinations that yield non-unique solutions, we documented that there are apparently always 3 solutions. For example, for the case of a monopoly with $D_p = 1.2$ and $V_t = 600$, and another case of a monopoly with $D_p = 3.3$ and $V_t = 2700$, as well as the case of a symmetric duopoly with $D_p = 3$ and $V_t = 1800$, we show the obtained solutions in Tables 8, 9, and 10, respectively.

In Tables 8–10, it is apparent that the first two solutions are rather similar, though they are very different compared to the third solution. The third solution illustrates a scenario where the transition from Type II to Type I cancellation has occurred, whereas the transition has not occurred for the first two solutions. This suggests that as the market transitions from predominantly Type II cancellations to predominantly Type I cancellations, there is a window in which both scenarios can be plausible for the same given

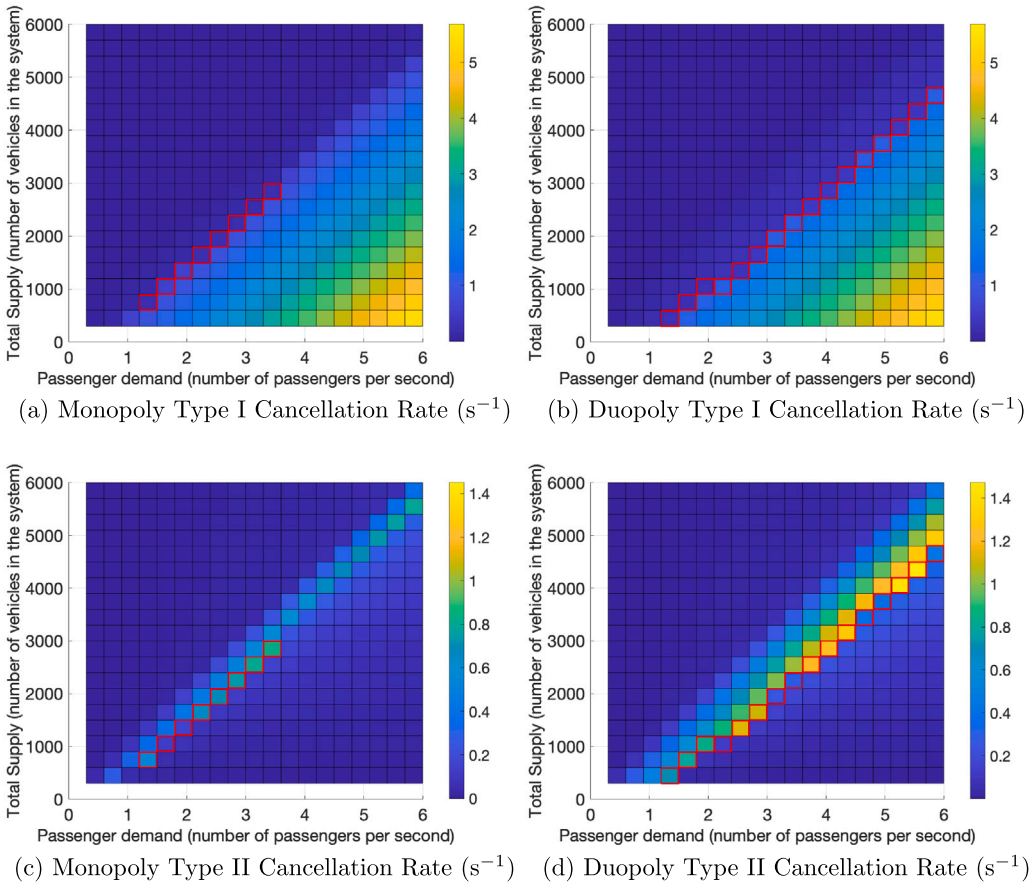


Fig. 15. Type I & II cancellation rates for the monopoly and the duopoly. The supply and demand combinations that yield non-unique solutions are highlighted in red.

demand and supply. This observed existence of non-unique solutions for the proposed model is worth investigating. However, it does not affect the obtained result's interpretability, its similarities to that of the dynamic model, and consequently, its usefulness to provide insight into the e-hailing market.

References

Alisoltani, N., Ameli, M., Zargayouna, M., Leclercq, L., 2022. Space-time clustering-based method to optimize shareability in real-time ride-sharing. PLoS One 17 (1), e0262499.

Alisoltani, N., Leclercq, L., Zargayouna, M., 2021. Can dynamic ride-sharing reduce traffic congestion? Transp. Res. Part B: Methodol. 145, 212–246.

Ancker, Jr., C., Gafarian, A., 1963. Some queuing problems with balking and reneging. I. Oper. Res. 11 (1), 88–100.

Baccelli, F., Boyer, P., Hebuterne, G., 1984. Single-server queues with impatient customers. Adv. in Appl. Probab. 16 (4), 887–905.

Bai, J., So, K.C., Tang, C.S., Chen, X., Wang, H., 2019. Coordinating supply and demand on an on-demand service platform with impatient customers. Manuf. Serv. Oper. Manag. 21 (3), 556–570.

Banerjee, S., Riquelme, C., Johari, R., 2015. Pricing in ride-share platforms: A queueing-theoretic approach. Available At SSRN 2568258.

Barrer, D.Y., 1957. Queuing with impatient customers and indifferent clerks. Oper. Res. 5 (5), 644–649.

Beojone, C.V., Geroliminis, N., 2021. On the inefficiency of ride-sourcing services towards urban congestion. Transp. Res. Part C: Emerg. Technol. 124, 102890.

Beojone, C.V., Geroliminis, N., 2023. Relocation incentives for ride-sourcing drivers with path-oriented revenue forecasting based on a Markov chain model. Transp. Res. Part C: Emerg. Technol. 157, 104375.

Beojone, C.V., Zhu, P., Sirmatel, I.I., Geroliminis, N., 2024. A hierarchical control framework for vehicle repositioning in ride-hailing systems. Transp. Res. Part C: Emerg. Technol. 168, 104717, Collected Papers from the 25th International Symposium on Transportation and Traffic Theory (ISTTT25).

Boots, N.K., Tijms, H., 1999. A multiserver queueing system with impatient customers. Manag. Sci. 45 (3), 444–448.

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. pp. 241–242.

Chen, L., Valadkhani, A.H., Ramezani, M., 2021. Decentralised cooperative cruising of autonomous ride-sourcing fleets. Transp. Res. Part C: Emerg. Technol. 131, 103336.

Choudhury, A., 2008. Impatience in single server queueing model. Amer. J. Math. Management Sci. 28 (1–2), 177–211.

De Kok, A.G., Tijms, H.C., 1985. A queueing system with impatient customers. J. Appl. Probab. 22 (3), 688–696.

Fayed, L., Nilsson, G., Geroliminis, N., 2024. A dynamic macroscopic framework for pricing of ride-hailing services with an optional bus lane access for pool vehicles. Transp. Res. Part C: Emerg. Technol. 169, 104854, URL https://www.sciencedirect.com/science/article/pii/S0968090X24003759.

Fielbaum, A., Tirachini, A., 2021. The sharing economy and the job market: The case of ride-hailing drivers in Chile. Transportation 48 (5), 2235–2261.
Flatto, L., 1997. The waiting time distribution for the random order service $M/M/1$ queue. Ann. Appl. Probab. 7 (2), 382–409.
Haight, F.A., 1959. Queueing with reneging. Metrika 2 (1), 186–197.
Jiao, G., Ramezani, M., 2022. Incentivizing shared rides in e-hailing markets: Dynamic discounting. Transp. Res. Part C: Emerg. Technol. 144, 103879.
Jiao, G., Ramezani, M., 2024. A real-time cooperation mechanism in duopoly e-hailing markets. Transp. Res. Part C: Emerg. Technol. 162, 104598.
Ke, J., Yang, H., Li, X., Wang, H., Ye, J., 2020. Pricing and equilibrium in on-demand ride-pooling markets. Transp. Res. Part B: Methodol. 139, 411–431.
Kondor, D., Bojic, I., Resta, G., Duarte, F., Santi, P., Ratti, C., 2022. The cost of non-coordination in urban on-demand mobility. Sci. Rep. 12 (1), 1–10.
Nourinejad, M., Ramezani, M., 2020. Ride-sourcing modeling and pricing in non-equilibrium two-sided markets. Transp. Res. Part B: Methodol. 132, 340–357.
Ramezani, M., Valadkhani, A.H., 2023. Dynamic ride-sourcing systems for city-scale networks-part I: Matching design and model formulation and validation. Transp. Res. Part C: Emerg. Technol. 152, 104158.
Ramezani, M., Yang, Y., Elmasry, J., Tang, P., 2022. An empirical study on characteristics of supply in e-hailing markets: A clustering approach. Transp. Lett. 1–14.
Séjourné, T., Samaranayake, S., Banerjee, S., 2018. The price of fragmentation in mobility-on-demand services. Proc. ACM Meas. Anal. Comput. Syst. 2 (2), 1–26.
Stanford, R.E., 1979. Reneging phenomena in single channel queues. Math. Oper. Res. 4 (2), 162–178.
Tafreshian, A., Masoud, N., 2020. Trip-based graph partitioning in dynamic ridesharing. Transp. Res. Part C: Emerg. Technol. 114, 532–553.
Wang, X., Liu, W., Yang, H., Wang, D., Ye, J., 2019. Customer behavioural modelling of order cancellation in coupled ride-sourcing and taxi markets. Transp. Res. Procedia 38, 853–873.
Wang, G., Zhang, H., Zhang, J., 2022. On-demand ride-matching in a spatial model with abandonment and cancellation. Oper. Res..
Wu, T., Zhang, M., Tian, X., Wang, S., Hua, G., 2020. Spatial differentiation and network externality in pricing mechanism of online car hailing platform. Int. J. Prod. Econ. 219, 275–283.
Xu, K., Saberi, M., Liu, W., 2022. Dynamic pricing and penalty strategies in a coupled market with ridesourcing service and taxi considering time-dependent order cancellation behaviour. Transp. Res. Part C: Emerg. Technol. 138, 103621.
Xu, Z., Yin, Y., Ye, J., 2020. On the supply curve of ride-hailing systems. Transp. Res. Part B: Methodol. 132, 29–43.
Yang, Y., Ramezani, M., 2022. A learning method for real-time repositioning in E-hailing services. IEEE Trans. Intell. Transp. Syst..
Yang, Y., Umboh, S.W., Ramezani, M., 2024. Freelance drivers with a decline choice: Dispatch menus in on-demand mobility services for assortment optimization. Transp. Res. Part B: Methodol. 190, 103082.
Yang, H., Yang, T., 2011. Equilibrium properties of taxi markets with search frictions. Transp. Res. Part B: Methodol. 45 (4), 696–713.
Zha, L., Yin, Y., Yang, H., 2016. Economic analysis of ride-sourcing markets. Transp. Res. Part C: Emerg. Technol. 71, 249–266.
Zhang, K., Alonso-Mora, J., Fielbaum, A., 2025. What do walking and e-hailing bring to scale economies in on-demand mobility? Transp. Res. Part B: Methodol. 192, 103156.
Zhang, W., Honnappa, H., Ukkusuri, S.V., 2020. Modeling urban taxi services with e-hailings: A queueing network approach. Transp. Res. Part C: Emerg. Technol. 113, 332–349.
Zhang, K., Nie, Y.M., 2021. To pool or not to pool: Equilibrium, pricing and regulation. Transp. Res. Part B: Methodol. 151, 59–90.
Zhang, Z., Zhang, F., 2022. Ride-pooling services with differentiated pooling sizes under endogenous congestion effect. Transp. Res. Part C: Emerg. Technol. 144, 103883.
Zhou, Y., Yang, H., Ke, J., 2022. Price of competition and fragmentation in ride-sourcing markets. Transp. Res. Part C: Emerg. Technol. 143, 103851.