Transportation Research Part B 192 (2025) 103156

Contents lists available at ScienceDirect

Transportation Research Part B

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

What do walking and e-hailing bring to scale economies in on-demand mobility?

Kenan Zhang [a,*], Javier Alonso-Mora [b], Andres Fielbaum [c]

[a] School of Architecture, Civil and Environmental Engineering, EPFL, Switzerland
[b] Department of Cognitive Robotics, Delft University of Technology, The Netherlands
[c] School of Civil Engineering, University of Sydney, Australia

ARTICLE INFO

Keywords:
Passenger–vehicle matching
On-demand mobility service
Scale economies
Walking

ABSTRACT

This study investigates the impact of walking and e-hailing on the scale economies of on-demand mobility services. An analytical framework is developed to i) explicitly characterize the physical interactions between passengers and vehicles in the matching and pickup processes, and ii) derive the closed-form degree of scale economies (DSE) to quantify scale economies. The general model is then specified for conventional street-hailing and e-hailing, with and without walking before pickup and after dropoff. We show that, under a system-optimum fleet size, the market always exhibits economies of scale regardless of the matching mechanism and the walking behaviors, though the scale effect diminishes as passenger demand increases. Yet, street-hailing and e-hailing show different scale economies in their matching process. While street-hailing matching shows a constant DSE of two, e-hailing matching is more sensitive to demand and its DSE diminishes to one when passenger competition emerges. Walking, on the other hand, has mixed effects on the scale economies: while the reduced pickup and in-vehicle times bring a positive scale effect, the extra walking time and possible concentration of vacant vehicles and waiting passengers on streets negatively affect scale economies. All these analytical results are validated through agent-based simulations on Manhattan with real-life demand patterns.

1. Introduction

The competition between taxis and ride-hailing companies (e.g., Uber) has been a heated debate in both academia and industry in recent years. After commencing their operations, ride-hailing services grew at an impressive pace, both in terms of passengers and drivers, which posed the question of whether traditional taxis would cease to exist (Cramer and Krueger, 2016; Nie, 2017). However, ten years later, both systems still co-exist in many cities, especially those with high demand densities. To well explain this phenomenon, one needs a clear understanding of the physical process underlying both services and more importantly, their economic implications. This is also the key motivation of the present study.

One major difference between taxi and ride-hailing services is how they match their vehicles with passengers. For decades, the taxi industry has been struggling with the so-called *search friction*, i.e., the spatial mismatch between vehicle supply and passenger demand. Typically, the hailing process in conventional taxi services relies on either passengers and drivers making visual contact on the streets, known as *street-hailing*, or manual dispatch. On the other hand, ride-hailing companies use advanced mobile communication technologies to match passengers and drivers online. Such a hailing process is referred to as *e-hailing* and has been celebrated for significantly reducing search friction (Frechette et al., 2019; Buchholz, 2022). Another considerable difference

* Corresponding author.
E-mail address: kenan.zhang@epfl.ch (K. Zhang).

https://doi.org/10.1016/j.trb.2025.103156
Received 28 June 2024; Received in revised form 13 January 2025; Accepted 14 January 2025
Available online 28 January 2025

between taxis and ride-hailing, yet often overlooked, is *walking*. Most ride-hailing trips are served on a door-to-door basis, where passengers are picked up and dropped off at their very origins and destinations, respectively. For street-hailing taxis, this is however not the case. Instead, passengers usually walk towards where they expect taxis to appear (e.g., major streets and points of interest). While less implemented in practice, recent studies have shown that walking can also greatly benefit on-demand mobility as it prevents vehicles from traversing local streets (Fielbaum et al., 2021; Fielbaum, 2022; Gurumurthy and Kockelman, 2022).

In this paper, we aim to establish the relationship between the above two fundamental features and the scale economies in on-demand mobility, i.e., how the system output (e.g., trip throughput and service quality) evolves with the input (e.g., demand level and fleet size). Although it is widely acknowledged that conventional taxi markets display strong scale economies, also known as increasing returns to scale (e.g., Douglas, 1972; Arnott, 1996), there is a lack of convincing theory for e-hailing, and the role of walking has not been previously analyzed. Meanwhile, empirical evidence has suggested a loss of scale economies in e-hailing when demand and supply densities are both high (Zhang et al., 2019). These seemingly contradictory results motivate us to conduct a comprehensive analysis of the scale economies in on-demand mobility services. Along with its theoretical contributions, our results are of great significance for practice. In general, the existence of scale economies indicates if monopoly or competition shall be expected in markets at different demand levels. Furthermore, in-depth scrutiny of the scale effect of e-hailing and walking provides practical insights on whether and where they shall be encouraged to improve service quality and system efficiency (Jiao and Ramezani, 2024).

Formally, this paper sets out to theoretically quantify scale economies in on-demand mobility with e-hailing and walking. To this end, we develop an analytical framework consisting of a physical matching model and an aggregated economic model. The former explicitly characterizes the *matching* (i.e., how are users and vehicles assigned to each other) and *pickup* (i.e., what happens after the matching and before the passenger boards the vehicle) processes in street-hailing and e-hailing, with and without walking. Its key outputs are then integrated into the economic model to solve the system-optimum fleet size given the demand rate. With mild approximations, we derive close-form solutions to the system-optimal fleet size and accordingly, the *degree of scale economies* (DSE), a traditional quantitative measure of scale economies whose formal definition is provided later in the paper. This enables us to rigorously analyze the scale effects of e-hailing and walking, and consequently, give insights into the design and operations of ride-hailing services at various market scales. The derived analytical results are further demonstrated by simulations on a realistic case study with more realistic operations constructed on the Manhattan network.

The main findings and practical insights of this paper are summarized as follows:

- The system-level DSE is always greater than one but decreases with demand, regardless of the matching mechanism and walking. This theoretical result is validated through simulations. It implies that, without external interventions (e.g., control on market entry), a monopoly is expected in low-density areas while competition could be observed in high-demand areas.
- DSE in street-hailing matching is constant and equal to two, insensitive to demand level. In contrast, DSE in e-hailing matching equals 1.5 in low-demand areas, while it decreases with the demand level, approaching one in high-demand areas. The loss of scale effect is largely due to the unmatched passengers competing for vehicles in the same region. Consequently, introducing additional vehicles often saves more waiting time in street-hailing than e-hailing. This also explains why street-hailing is more competitive in a high-density market.
- Walking does not change the fundamental properties of the DSE and has mixed impacts on its value. On the one hand, the reduced pickup and in-vehicle times due to walking bring a positive scale effect. On the other hand, the extra walking time and the concentration of vacant vehicles and waiting passengers on certain streets negatively affect scale economies. We derive a necessary condition under which the overall impact of walking on DSE is negative, while our simulation results suggest that the influence of walking on DSE tends to be negative in street-hailing and positive in e-hailing.

The remainder of this paper is organized as follows. Section 2 reviews the related literature. Section 3 describes the overall setting and the service schemes studied in this paper. Section 4 describes in detail the matching mechanisms and derives the closed-form expressions of matching outputs for each scheme. Section 5 presents the economic model and derives the DSE in each scheme with e-hailing and/or walking. Section 6 details the simulations run over the Manhattan network and compares their results to the theoretical findings. Section 7 summarizes the practical implications derived from this study, and finally, Section 8 concludes and proposes directions for future research.

2. Related work

In this section, we review two streams of literature. The first part summarizes different approaches to modeling the matching process in on-demand mobility services, and the second reviews existing studies on the scale economies in shared transport services, including on-demand mobility.

2.1. Matching in on-demand mobility

The matching process in on-demand mobility services refers to the time interval from passengers' arrival in the system to their pickups. A matching model then maps from the matching input, which is dictated by passenger demand and vehicle supply, to the matching output, which is often expressed by metrics like passenger waiting time, vehicle searching time, pickup rate, etc. The simplest model assumes matching is frictionless (e.g., Lagos, 2000; Bimpikis et al., 2019), that is, the number of matched trips (or pickups) equals the minimum between the number of waiting passengers and the number of vacant vehicles. While this model

produces reasonable estimates of the matching outcomes, particularly in the extremely over- and under-supplied system, it fails to capture the characteristics of different matching mechanisms (e.g., street-hailing vs. e-hailing). A more widely used matching model in the transportation literature is the Cobb–Douglas function (e.g., Yang et al., 2010; Yang and Yang, 2011; Zha et al., 2016), which draws an analogy between matching and production. Accordingly, the matching mechanism does not change the model structure but only affects its parameters. Due to the lack of a physical model, these parameters are often set based on assumptions (e.g., He and Shen, 2015; Wang et al., 2016). On the other hand, the matching model is often built on queuing theory in the literature of operations research. For instance, Banerjee et al. (2015) models the single-region ride-hailing system as an $M/M/1$ queue, where the vacant vehicles form the queue and the service rate is given by the passenger arrival rate. Other studies consider waiting passengers to form an $M/M/k$ queue, where k denotes the number of vehicles (Taylor, 2018; Bai et al., 2019; Feng et al., 2021). This modeling framework is further extended in Besbes et al. (2022) to have the service rate — the inverse of total passenger travel time that includes both waiting time and in-vehicle time — dependent on the demand-supply relationship in the system.

Although Cobb–Douglas and queuing-based models capture the matching friction in on-demand mobility, they have ignored the physical interactions between passengers and vehicles. In contrast, another array of research builds the matching model from detailed cruising and searching behaviors of vehicles and passengers on streets. Douglas (1972) considered a street-hailing taxi market where vacant vehicles are uniformly distributed on roads and perform random cruising. Accordingly, the passenger waiting time is derived to be inversely proportional to the line density of vacant vehicles. With a different setting of radio-dispatch taxi service, Arnott (1996) showed the passenger waiting time is inversely proportional to the square root of the spatial density of vacant vehicles. The same model is further extended to model the matching process in e-hailing with more operational details such as matching radius and matching interval (Zha et al., 2016; Xu et al., 2017; Yang et al., 2020). However, one critical factor that distinguishes e-hailing from radio-dispatch taxis — the competition among waiting passengers — has been largely omitted in the literature. While radio-dispatch services are often operated in low-density regions, e-hailing is more often found in high-density urban areas. Consequently, e-hailing passengers are likely to compete for the same pool of vehicles against each other due to the sufficiently large matching radius. The negative impact of passenger competition was first recognized in Zhang et al. (2019) and later integrated into the matching model proposed in Yang et al. (2020). Yet, neither of them differentiated the virtual matching and physical pickup processes. Hence, the influence of passenger competition on each subprocess is still unclear.

Recently, Li et al. (2024b) proposed an aggregate matching and pickup model for mobility-on-demand services. Similar to this study, they separate the matching and pickup processes and derive the expected matching and pickup times under the assumption of uniformly distributed vacant vehicles and waiting passengers. Li et al. (2024b) also discussed returns to scale in the matching process and showed the model reduces to several previous studies (e.g., Yang and Yang, 2011; Castillo et al., 2017) at corner cases (e.g., sufficiently small or large matching radius). Differently, this paper focuses on the scale economies at the market level and aims to derive close-form quantitative measures under general market conditions and matching mechanisms.

2.2. Scale economies in shared transport

In the literature on transportation economics, it is widely accepted that public transit (e.g., bus) enjoys economies of scale. A well-known example is the *Mohring effect*: an increase in demand would induce a higher transit frequency, which, in turn, reduces the passenger waiting time and improves the service quality (Mohring, 1972). Scale effects have also been demonstrated in vehicle sizing (Jansson, 1980) and transit network design (Basso and Jara-Díaz, 2006; Fielbaum et al., 2020). Another common finding in transit systems is that their scale economies tend to diminish as demand continues to increase.

By an informal definition, a service system is said to exhibit economies of scale if the average cost decreases with the service throughput. Douglas (1972) showed such property also holds in a taxi market when the service quality, dictated by the vacant vehicle density, is fixed. Besides taking the entire market as a whole, researchers also believe the matching process in on-demand mobility displays economies of scale, or equivalently, increasing returns to scale (e.g., Arnott, 1996; Yang and Yang, 2011). Empirical evidence, however, only supports this hypothesis to be true in conventional taxi services, whereas in e-hailing services, it is found the matching process is prone to constant returns to scale (Schroeter, 1983; Frechette et al., 2019; Zhang et al., 2019). Nevertheless, it has been shown that ride-pooling has a great potential for enhancing the scale economies in e-hailing (Kaddoura and Schlenther, 2021; Fielbaum et al., 2023; Liu et al., 2023). Walking, however, has not been much touched in the literature. Although it has been implemented in practice to improve the efficiency of matching and pickup (Yan et al., 2020; Gurumurthy and Kockelman, 2022; Fielbaum et al., 2023), its impact on the system's scale economies is still not well understood.

3. Model setup and service schemes

In this study, we consider a grid network with two types of streets, as shown in Fig. 1. The *major* streets form the skeleton of the network. Between every two major streets, there are L *local* streets with equal spacing s. Accordingly, the intersections are classified into three types: (i) Type-1: between major streets; (ii) Type-2: between local streets; and (iii) Type-3: between local and major streets.

Based on the matching mechanism and the existence of walking, we define four service schemes: (i) street-hailing without walking (DS), (ii) street-hailing with walking (WS), (iii) e-hailing without walking (DE), and (iv) e-hailing with walking (WE).[1]

[1] The letter D in DS and DE stands for "door-to-door", while W denotes "walking".

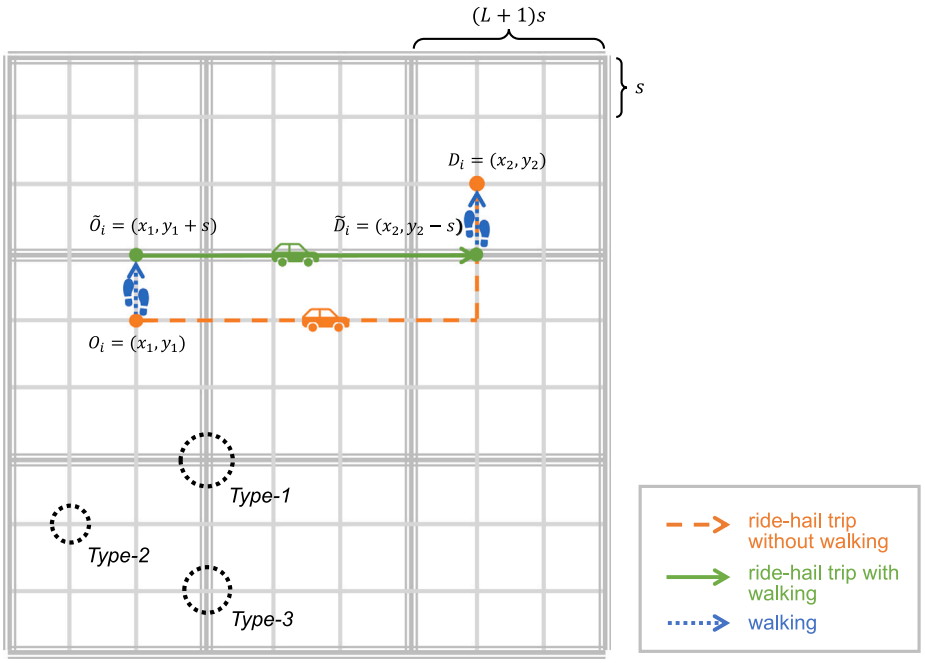


Fig. 1. Illustration of studies grid network and trips. O_i (D_i) is the trip origin (destination) and $\tilde{O}_i$ ($\tilde{D}_i$) is the pickup (dropoff) location in the schemes with walking.

In all schemes, origins and destinations are all situated at Type-2 intersections (i.e., passengers travel between local blocks). When walking is considered, passengers are picked up and dropped off on major streets that are closest to their origins and destinations, respectively, thus pickups and dropoffs happen at Type-3 intersections. It is worth noting that although WS (i.e., taxis pick up and drop off passengers at major streets) and DE (i.e., e-hailing vehicles serve door-to-door trips) are more commonly observed in real life, all four combinations are studied in this paper to identify the specific effects of e-hailing and walking.

Throughout this paper, we use one road segment of length s, referred to as an *arc*, as the geographic unit (see Fig. 1). Hence, an arc corresponds to one block on local streets and one block between two major streets measures $L+1$ arcs. Hereafter, all density, length, and area variables are measured in arcs. Let v be the average travel speed on major streets, then vehicles take $\delta = s/v$ to traverse an arc. We introduce two factors $\alpha^l, \alpha^a > 1$ to denote the ratios of v to local street travel speed and walking speed, respectively. Hence, the time for vehicles to traverse an arc is $\alpha^l\delta$ and for passengers, it takes $\alpha^a\delta$ to walk over an arc.

To facilitate the analysis, we introduce the following assumptions, which are also used in previous studies to obtain insights at the macroscopic level (e.g., Yang and Yang, 2011; Zhang et al., 2019), which is also the main objective of this study.

Assumption 1 (*Stationary Market*)**.** The market has reached a steady state where passenger demand and vehicle supply are both uniformly distributed in space. Accordingly, the market conditions can be described by the average vacant vehicle density $V > 0$ (veh/arc), the average unmatched passenger density $W \geq 0$ (pax/arc), and the demand rate Q (pax/hr/arc), i.e., the passenger arrival rate per unit area.

Assumption 2 (*Passenger Behaviors*)**.** The trip origins and destinations are uniformly distributed over Type-2 intersections. Passengers are patient, i.e., they will remain waiting until picked up. When walking is allowed, passengers walk to the closest major street, breaking ties randomly.

Assumption 3 (*Vehicle Behaviors*)**.** When being vacant, vehicles randomly cruise in the network. In DS (WS), vehicles only cruise on local (major) streets to maximize the probability of finding a passenger, whereas in e-hailing (DE and WE), vehicles cruise on both types of streets.

While the above assumptions are introduced to simplify the analytical model, they are relaxed in the simulations (Section 6) to capture more realistic demand and supply patterns.

4. Physical matching model

In this section, we present a general matching model and use $k \in \{DS, WS, DE, WE\}$ to denote the service scheme. Our main objective is to derive the closed-form estimates of the matching output, i.e., the passenger waiting time that consists of three parts:

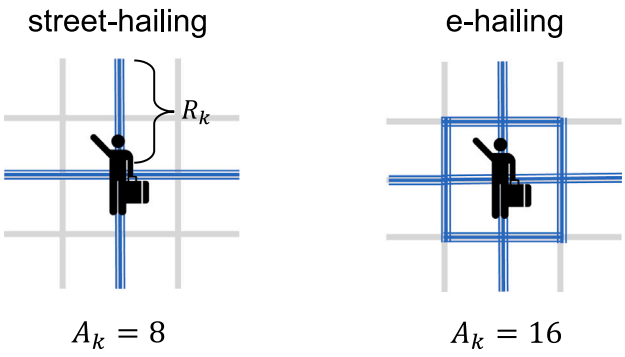


Fig. 2. Matching areas in street-hailing and e-hailing with the same matching radius $R_k = 2$.

(i) matching time w_k^m, from the passengers' emergence to their engagement with a pickup vehicle, (ii) pickup time w_k^p, from the engagement to the pickup, and (iii) walking time w_k^q when applicable.

4.1. Components of passengers waiting time

4.1.1. Matching time

The matching time refers to the time elapsed from when a passenger starts *searching* for a vehicle until when it is *assigned*, either because of visual contact in street-hailing (DS and WS) or via the centralized dispatching in e-hailing (DE and WE).

To begin with, we define the *matching radius* R_k in units of arc as the maximum distance from which a passenger can be matched with a vacant vehicle. In practice, the matching radius in street-hailing is determined by physical constraints (e.g., how far can users and drivers make visual contact), whereas that in e-hailing is normally bounded by the maximum acceptable pickup time. Accordingly, the *matching area* A_k, also in units of arc, defines the area within which vacant vehicles can be hailed by the passenger. Due to the underlying matching mechanism, the same matching radius yields different matching areas in street-hailing and e-hailing, as illustrated in Fig. 2. Specifically, we have $A_k = 4R_k,\ k \in \{DS, WS\}$ and $A_k = 4R_k^2,\ k \in \{DE, WE\}$. The detailed derivation is provided in Appendix B.1.

Note that when vehicle supply is excessive, passengers can be guaranteed to be matched with vehicles within the matching area A_k. However, when demand is relatively high, a vehicle within the matching area may be dispatched to passengers because the vehicle is closer to them. In this case, unmatched passengers in close proximity are essentially competing with others for vehicles. To capture such a *passenger competition*, we further introduce the notion of *effective matching area*, denoted by $\widetilde{A}$, to represent the area within which any vacant vehicle is for sure matched to the passenger.[2] Let $\gamma_k = \max(A_k W, 1)$ be the number of waiting passengers in the matching area, including the studied passenger, then the effective matching area is $\widetilde{A}_k = A_k/\gamma_k$, which implies the matching area is evenly distributed to the unmatched passengers. Similarly, we define the effective vacant vehicle density as $\widetilde{V} = \theta_k V$, where θ_k is a density correction factor that reflects the accumulation of vehicles on a certain type of street. As per Assumption 3, vacant vehicles accumulate on different types of streets in DS and WS, which yields different values of θ_k. On the other hand, vacant vehicles are uniformly distributed in DE and WE and thus θ_k is simply 1 in the two schemes. The detailed derivation of θ_k is provided in Appendix B.2.

Let δ_k denote the matching interval of scheme k, which dictates how frequently matching is performed. In practice, the matching interval in street-hailing and e-hailing with instant matching can be considered approaching zero because passengers are constantly searching for vehicles. However, a small matching interval would largely complicate the model. Thus, the following assumption is introduced:

Assumption 4 (*Minimum Matching Interval*)**.** The matching interval is always equal to or larger than the travel time of a single arc, i.e., $\delta_k \geq \delta$.

We remark that although this assumption might seem strong, especially in the case of street-hailing, its impact is mild and limited to a few seconds as detailed in Appendix C. With Assumption 4 and the others introduced in Section 3, we can derive the matching time in a general form, as detailed in the following proposition.

Proposition 1 (*Matching Time*)**.** *Under Assumptions 1 and 4, the expected matching time can be approximated as*

$$w_k^m \approx \left(\frac{1}{\widetilde{A}_k \widetilde{V}_k} + \frac{1}{2}\right)\delta_k = \left(\frac{\gamma_k}{\theta_k A_k V} + \frac{1}{2}\right)\delta_k. \tag{1}$$

Proof. See Appendix E.1. □

[2] A similar definition of "dominant area" is introduced in Yang et al. (2020).

4.1.2. Pickup time

Following the literature (e.g., Besbes et al., 2022), we assume the passenger is matched with the closest vacant vehicle in the effective matching area and derive the pickup time as the time interval from the passenger being matched to its pickup.

The derivation of pickup time in street-hailing is straightforward thanks to its simple matching mechanism (i.e., passengers can only see vehicles on streets crossing the intersection). For e-hailing, it is considerably more complicated because the probability varies among possible pickup distances and a fraction of the pickup route is taken on local streets. Further, passenger competition shall be considered in both street-hailing and e-hailing when the waiting passenger density is sufficiently large ($\gamma_k > 1$), though its impact is more significant in e-hailing due to the much larger matching area. With mild approximations, we obtain a closed-form pickup time summarized in the following proposition, which generalizes the extreme cases of the demand-supply relationship considered in Besbes et al. (2022) and Li et al. (2024b).

Proposition 2 (*Pickup Time*). *Under Assumptions 1–3, the expected pickup time can be approximated as*

$$w_k^p = \rho_k d_k^p \delta, \tag{2}$$

For street-hailing schemes $k \in \{DS, WS\}$, the approximation is also exact and

$$d_k^p = \frac{R_k}{2\gamma_k}, \qquad \rho_k = \begin{cases} \alpha^l, & k = DS, \\ 1, & k = WS. \end{cases} \tag{3}$$

For e-hailing schemes $k \in \{DE, WE\}$,

$$d_k^p \approx \frac{c_1}{\sqrt{V}} \left(1 + \frac{W}{c_2 V}\right)^{-1}, \qquad \rho_k = \begin{cases} 1 + (\alpha^l - 1)\left(1 + \frac{1}{\theta_{DS}}\right)\frac{D^a}{R_k}, & k = DE, \\ 1 + (\alpha^l - 1)\frac{D^a}{\theta_{DS} R_k}, & k = WE, \end{cases} \tag{4}$$

where $c_1 = \sqrt{\pi}/4$, $c_2 = e\sqrt{\pi}$ are constant parameters, and $D^a = \frac{(L+1)(L+2)}{6L}$ is the average walking distance.

Proof. See Appendix E.2. □

Several corner cases can be checked to validate Eq. (2) for e-hailing: (i) when there is sufficient vehicle supply ($V \to \infty$), d_k^p reduces to zero and so does w_k^p; (ii) when there is no passenger competition ($W \to 0$), $d_k^p \propto 1/\sqrt{V}$, in line with the literature (e.g., Arnott, 1996); (iii) when there is severe passenger competition ($W \to \infty$), d_k^p again reduces to zero, also reflecting the reality (each passenger has an extremely small dominant area and thus, if matched, the vehicle must be close to the passenger). One exception is the case of scarce vehicle supply ($V \to 0$), where $d_k^p \to \infty$ but the pickup distance in reality is bounded by the matching radius R_k. We note that the matching time in this case also goes to infinity and thus the overall estimate of waiting time is still valid.

4.1.3. Walking and in-vehicle times

Finally, with the average walking distance D^a derived in Appendix B.3, we can easily derive the walking time as

$$w_k^a = \begin{cases} 0, & k \in \{DS, DE\}, \\ \alpha^a D^a \delta, & k \in \{WS, WE\}. \end{cases} \tag{5}$$

Let $\bar{\tau}$ be the average door-to-door trip duration, then the in-vehicle time is adjusted as

$$\tau_k = \begin{cases} \bar{\tau}, & k \in \{DS, DE\}, \\ \bar{\tau} - 2\alpha^l D^a \delta, & k \in \{WS, WE\}. \end{cases} \tag{6}$$

4.2. Impact of matching and walking on passenger waiting time

Let $w_k = w_k^m + w_k^p + w_k^a$ be the expected total passenger waiting time for scheme k. To simplify the notation and generate main insights, we set $\delta_k = \delta$, $\forall k$ and $R_k = 1$, $k \in \{DS, WS\}$. As demonstrated in Appendix B.4, passenger competition rarely happens in street-hailing. Hence, we assume $\gamma_k = 1$, $k \in \{DS, WS\}$, which yields the total waiting time as follows:

$$w_k = \begin{cases} \left(\frac{1}{\theta_k A_k V} + 1\right)\delta + \frac{\alpha^l - 1}{2}\delta, & k = DS, \\ \left(\frac{1}{\theta_k A_k V} + 1\right)\delta + \alpha^a D^a \delta, & k = WS, \\ \left(\frac{1}{A_k V} + \frac{c_1 \rho_k}{\sqrt{V}}\right)\delta, & k = DE \text{ w/o competition} \\ \left(\frac{1}{A_k V} + \frac{c_1 \rho_k}{\sqrt{V}}\right)\delta + \alpha^a D^a \delta, & k = WE \text{ w/o competition} \\ \left[\frac{W}{V} + \frac{c_1 \rho_k}{\sqrt{V}}\left(1 + \frac{W}{c_2 V}\right)^{-1}\right]\delta, & k = DE \text{ w/ competition}, \\ \left[\frac{W}{V} + \frac{c_1 \rho_k}{\sqrt{V}}\left(1 + \frac{W}{c_2 V}\right)^{-1}\right]\delta + \alpha^a D^a \delta, & k = WE \text{ w/ competition}. \end{cases} \tag{7}$$

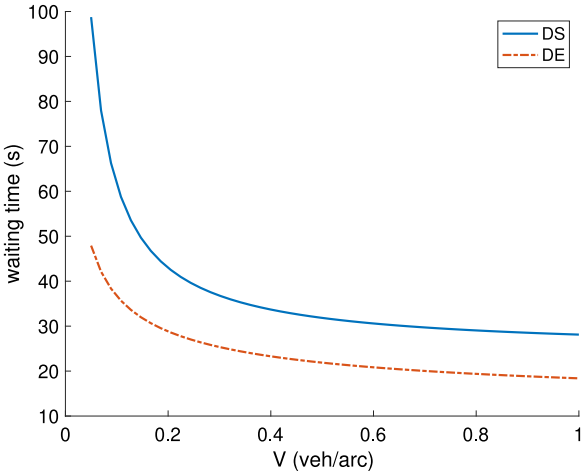


Fig. 3. Comparison of passenger waiting time between street-hailing and e-hailing without competition.

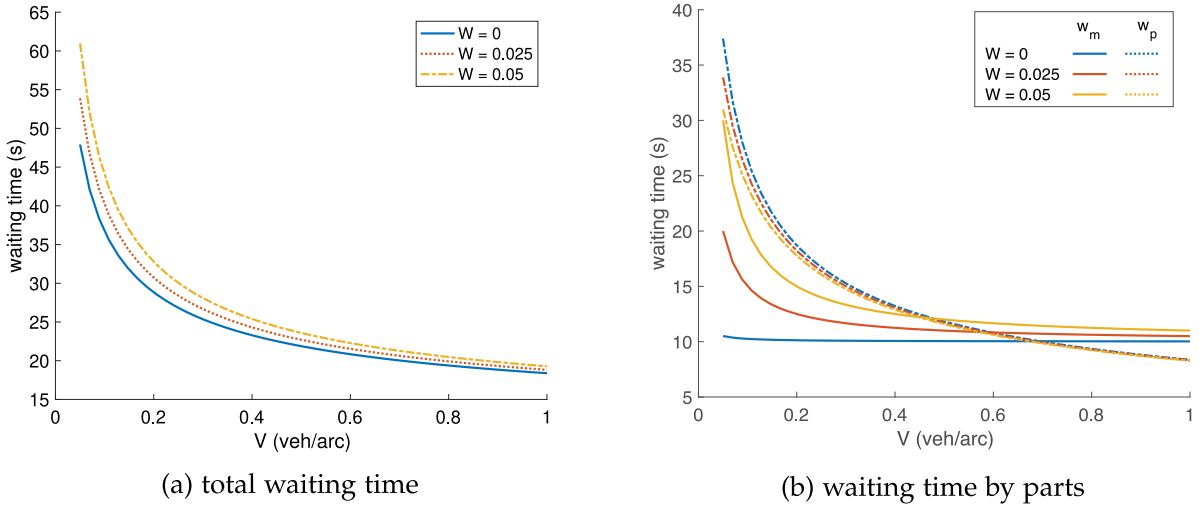


Fig. 4. Impact of passenger competition in e-hailing.

Several observations are made from Eq. (7):

- If there is no distinction between major and local streets (i.e., $\alpha^l = 1$), the passenger waiting time in DS is in line with (Douglas, 1972) (i.e., $w_k \propto 1/V$) while that in DE reduces to the result in Arnott (1996) (i.e., $w_k \propto 1/\sqrt{V}$) when A_k is sufficiently large.
- In street-hailing, walking brings two effects. On the one hand, it influences the matching time as it introduces vehicle concentration reflected in the parameter θ_k. On the other hand, it replaces part of the pickup time $(\alpha^l - 1)\delta/2$ on local streets into walking time $\alpha^a D^a \delta$. In e-hailing, however, only the second effect persists, which is captured in ρ_k.
- The impact of passenger competition in e-hailing is also two-fold. On the one hand, it extends the matching time as $W \gg 1/A_k$. On the other hand, it reduces the pickup time due to the factor $\left(1 + \frac{W}{c_2 V}\right)^{-1}$. In the extreme case where the market is crowded by waiting passengers, the pickup distance would approach zero. Such a phenomenon is also discussed in Besbes et al. (2022), though only for extreme cases. While Zhang et al. (2019) only recognizes and studies the joint influence using historical e-hailing data.

Fig. 3 plots the total waiting time in DS and DE without passenger competition against the vacant vehicle density (with default parameter values in Table A.1). As per Eq. (7), street-hailing waiting time decreases with V faster than e-hailing, though it is unable to surpass. On the other hand, e-hailing shows a much stronger robustness in its level of service when the vehicle supply is scarce.

Fig. 4 illustrates the passenger waiting time in DE under different levels of passenger competition. As shown in Fig. 4(a), the total waiting time increases with W but the gap is not significant unless V drops below 0.2 veh/arc. Yet, it does not mean passenger competition has no impact on waiting time. Instead, its two opposite influences on matching and pickup times cancel each out; see

Fig. 4(b). When the vehicle supply is depleted, the negative impact of passenger competition on matching is dominant, which yields an extensively long wait. Such a phenomenon has already been observed in the e-hailing market during a demand peak (Castillo et al., 2017).

5. Aggregated economic model

In this section, we derive the closed-form expressions of scale economies in the four service schemes. To this end, we solve the optimal fleet size N^* that ensures a desired demand rate Q is served. Accordingly, the system cost can be evaluated as a function of Q and, as usual, the *degree of scale economies* (DSE) is derived as

$$DSE(Q) = \frac{AC(Q)}{MC(Q)}, \tag{8}$$

where AC and MC denote the average and marginal costs, respectively.

In general, a system enjoys scale economies if its DSE is greater than 1, and the greater the DSE, the stronger the scale effect. For a thorough discussion about DSE in transport theory, we refer the reader to the paper by Jara-Diaz and Cortes (1996). In Section 5.1, we will first describe the optimal fleet sizing problem and its solution procedure. The key impacts of the matching mechanism and walking on scale economies will then be summarized in Sections 5.2 and 5.3, respectively. Besides scale economies at the system level, we are also interested in the results of the matching process, which is often discussed in literature (e.g., Yang and Yang, 2011; Zhang et al., 2019). Therefore, we also derive the DSE in the endogenous matching process for each service scheme and denote it as DSE^{en} hereafter.

5.1. Optimal fleet sizing

The optimal fleet sizing problem for a given demand rate Q is formulated as follows:

$$\min_{N} \quad TC(N,Q), \tag{9a}$$

$$s.t. \quad N = V + Q(w_k^p + \tau_k), \tag{9b}$$

$$W = Qw_k^m, \tag{9c}$$

$$w_k^p = f_k^p(V,W), \tag{9d}$$

$$w_k^m = f_k^m(V,W), \tag{9e}$$

where Constraint (9b) describes the vehicle flow conservation, Constraint (9c) adopts Little's law to connect the waiting passenger density W and the matching time w_k^m, and functions $f_k^p(\cdot)$ and $f_k^m(\cdot)$ in Constraints (9d) and (9e) are simplified expressions of the matching model derived in Section 4. The objective of (9) is defined as the sum of vehicle operating cost and passenger travel cost:

$$TC(N,Q) = c_0 N + Q(\beta^m w_k^m + \beta^p w_k^p + 2\beta^a w_k^a + \beta^t \tau_k), \tag{10}$$

where c_0 is the operation cost of each vehicle and $\beta^j,\ j \in \{m,p,a,t\}$ is the value of time specified for different legs of a trip. Such an objective is usually adopted in the analysis of public transport (e.g., Fielbaum et al., 2020) and has also been used to analyze the taxi market (Arnott, 1996). It essentially addresses the problem of finding the least costly way to serve a certain demand.

As preliminaries, we first derive the necessary conditions under which the steady state exists, i.e., Problem (9) is feasible.

Lemma 1. *Given the fleet size N and demand rate Q such that $N - Q(\bar{\tau} + \varepsilon_k) > 0$, where*

$$\varepsilon_k = \begin{cases} \frac{\rho_k^p}{2} R_k \delta, & k \in \{DS, WS\}, \\ \delta_k, & k \in \{DE, WE\}, \end{cases} \tag{11}$$

there must exist $V, W > 0$ that satisfy Eqs. (9b)–(9e).

Proof. See Appendix E.3. □

Due to Lemma 1, we can express V and W as solutions to the following implicit functions:

$$W = Qf_k^m(V,W) \quad \Rightarrow \quad W = h_k^W(V,Q), \tag{12}$$

$$N = V + Q(f_k^p(V, f_k^W(V,Q)) + \tau_k) \quad \Rightarrow \quad V = h_k^V(N,Q). \tag{13}$$

Accordingly, Problem (9) turns into an unconstrained optimization problem and its first-order condition reads

$$0 = \frac{\partial TC}{\partial N} = c_0 + Q\left[\beta^m\left(\partial_V f_k^m + \partial_W f_k^m \partial_V h_k^W\right) + \beta^p\left(\partial_V f_k^p + \partial_W f_k^p \partial_V h_k^W\right)\right]\partial_N h_k^V. \tag{14}$$

For street-hailing, it can be easily proved that there is a unique solution to Eq. (14) that coincides with the optimal fleet size N^*. For e-hailing, this result is not guaranteed, though our extensive numerical analyses show it indeed holds in general. All results presented in the following implicitly assume Eq. (14) to have a unique solution.

Note that Problem (9) leads to the system-optimal fleet size that does not necessarily yield the maximum profit, which is, however, the common objective of on-demand mobility services in practice. Yet, as shown in Appendix D, Problem (9) can also be transformed into a fleet-sizing problem for a profit-driven operator. Therefore, the analytical results presented in the following offer general implications for ride-hailing services.

5.2. Impacts of the matching mechanism on scale economies

Recall that the four schemes result from combining two matching mechanisms and the binary chance of walking. While the influences of walking are well captured in several model parameters (e.g., ρ_k and τ_k), the matching mechanism leads to different expressions of passenger waiting time. Consequently, the degree of scale economies (DSE) is largely shaped by matching. The main results are presented in the following propositions, categorized by the existence of passenger competition.

5.2.1. Without passenger competition

Proposition 3 (*Scale Economies in Street-hailing Without Passenger Competition*). *The system-level DSE in street-hailing without passenger competition is expressed as*

$$DSE = \frac{2B_1 + B_2 Q^{\frac{1}{2}}}{B_1 + B_2 Q^{\frac{1}{2}}}, \tag{15}$$

where $B_1 = \sqrt{\frac{c_0 \beta^m \delta_k}{\theta_k R_k}}$, *and* $B_2 = \frac{\beta^m \delta_k}{2} + \frac{\rho_k R_k \delta}{2}(c_0 + \beta^p) + 2\beta^a w_k^a + (c_0 + \beta^t)\tau_k$. *Hence, its value falls in the range of (1,2], monotonically decreases with demand Q, and finally converges to 1. Besides, the DSE in the endogenous matching process is constant and equals 2, i.e.,* $DSE^{en} = 2$.

Proof. See Appendix E.4. □

Proposition 4 (*Scale Economies in E-hailing Without Passenger Competition*). *With mild approximation, the system-level DSE in e-hailing without passenger competition is expressed as*

$$DSE \approx \frac{3B_1 + B_2 Q^{\frac{1}{3}}}{2B_1 + B_2 Q^{\frac{1}{3}}}, \tag{16}$$

where $B_1 = c_0 \left[\frac{c_1 \rho_k \delta}{2}\left(1 + \frac{\beta^p}{c_0}\right)\right]^{\frac{2}{3}}$, *and* $B_2 = \frac{\beta^m \delta_k}{2} + 2\beta^a w_k^a + (c_0 + \beta^t)\tau_k$. *Hence, its value falls in the range of (1,1.5], monotonically decreases with demand Q, and finally converges to 1. Besides, the DSE in the endogenous matching process is constant and equals 1.5, i.e.,* $DSE^{en} = 1.5$.

Proof. See Appendix E.5. □

5.2.2. With passenger competition

Proposition 5 (*Scale Economies in Street-hailing with Passenger Competition*). *With mild approximation, the system-level DSE in street-hailing with passenger competition is expressed as*

$$DSE \approx 1 + \frac{B_1}{(B_2 + B_3)Q}, \tag{17}$$

and the DSE in the endogenous matching process is

$$DSE^{en} \approx 1 + \frac{B_1}{B_3 Q}, \tag{18}$$

where $B_1 = \frac{\rho_k R_k \delta}{\delta_k A_k}(c_0 + \beta^p)\left(1 + \sqrt{\frac{2c_0}{\beta^m \theta_k}}\right)^{-1}$, $B_2 = \frac{\beta^m \delta_k}{2} + \frac{\rho_k R_k \delta}{2}(c_0 + \beta^p) + 2\beta^a w_k^a + (c_0 + \beta^t)\tau_k$, *and* $B_3 = \frac{c_0 \delta_k}{\theta_k}\left(1 + \sqrt{\frac{2\beta^m \theta_k}{c_0}}\right)$. *Hence, both values are greater than 1, monotonically decrease with demand Q, and finally converge to 1.*

Proof. See Appendix E.6. □

Table 1
Impacts of walking on DSE.

Factor	$DS \to WS$	$DE \to WE$	Cause of change
Reduced matching time	–	NA	Accumulation of vacant vehicles on major streets in street-hailing results in an increase in θ_k.
Reduced pickup time	+	+	The reduced pickup distance and/or increased vehicle speed results in a decrease in ρ_k.
Extra walking time	–	–	The additional walking time yields a positive w_k^a.
Reduced in-vehicle time	+	+	The reduced in-vehicle trip distance and increased travel speed result in a decrease in τ_k.

Proposition 6 (*Scale Economies in E-hailing with Passenger Competition*). *With mild approximation, the system-level DSE in e-hailing with passenger competition is expressed as*

$$DSE \approx \frac{2B_1 + (B_2 + B_3)Q^{\frac{1}{2}}}{B_1 + (B_2 + B_3)Q^{\frac{1}{2}}}, \tag{19}$$

and the DSE in the endogenous matching process is

$$DSE^{en} \approx \frac{2B_1 + B_3 Q^{\frac{1}{2}}}{B_1 + B_3 Q^{\frac{1}{2}}}, \tag{20}$$

where $B_1 = \frac{c_1 \rho_k \delta \varphi^*}{2}(c_0 + \beta^p)\left[\delta_k\left(1 + \sqrt{\frac{\beta^m}{2c_0}}\right)\right]^{-\frac{1}{2}}$, $B_2 = \frac{\beta^m \delta_k}{2} + 2\beta^a w_k^a + (c_0 + \beta^t)\tau_k$, $B_3 = c_0\delta_k\left(1 + \sqrt{\frac{2\beta^m}{c_0}}\right)$, *and* $\varphi^* = \left(1 + \frac{1}{c_1}\sqrt{\frac{c_0}{2\beta^m}}\right)^{-1}$. *Hence, both values are greater than 1, upper bounded by 2, monotonically decrease with demand Q, and finally converge to 1.*

Proof. See Appendix E.7. □

5.2.3. Summary

By definition, a service system with DSE greater than 1 exhibits increasing returns to scale while it shows constant (decreasing) returns to scale if its DSE equals (less than) 1 (Tone and Sahoo, 2004). Therefore, the propositions presented in this section can be summarized as follows:

- The on-demand mobility market exhibits increasing returns to scale in every scheme, but the scale effects eventually vanish as DSE converges to 1.
- The scale effects in the matching process are strongly dependent on the hailing mechanism and the eventual existence of competition. Specifically, without passenger competition, the DSE in street-hailing matching is constant and equals 2 while that in e-hailing matching is also constant but reduces to 1.5.
- When there exists passenger competition, the DSE in matching for both services diminishes to 1 with the demand rate.

Although the impact of passenger competition affects both street-hailing and e-hailing, competition can only occur in street-hailing at extremely large demand rates whereas it is rather common in e-hailing. The key reason is that street-hailing has a much smaller matching radius so waiting passengers are unlikely to have overlapping matching areas with others. Suppose the matching radius is $R_k = 1$, then competition only happens when there is another passenger waiting at the same or adjacent intersection. In contrast, e-hailing waiting passengers are likely to have overlapping matching areas due to the much larger matching radius. This difference in the existence of passenger competition is numerically demonstrated in Appendix B.4. Due to the third point, in what follows, we only present the analysis for the case without passenger competition for street-hailing.

5.3. Impacts of walking on scale economies

Although walking does not change the fundamental properties of scale economies, it does affect the values of DSE in both street-hailing and e-hailing. In particular, walking affects scale economies via two opposing forces. Besides, most factors are shared between e-hailing and street-hailing except for one that occurs specifically in street-hailing. These factors and their impacts on DSE are formally stated below.

Proposition 7. *Walking influences the system-level DSE in both street-hailing and e-hailing, while only affecting DSE in the street-hailing matching process. The sources of change in DSE along with the directions are summarized in Table 1.*

Proof. See Appendix E.8. □

We further observe that, under certain conditions, walking brings a common negative scale effect, as formally stated in the following proposition.

Proposition 8. *Walking leads to a lower system-level DSE in both street-hailing and e-hailing when* $\alpha^l = 1$ *and* $\alpha^a > (c_0 + \beta^t)/\beta^a$.

Proof. See Appendix E.9. □

The result of Proposition 8 may be counter-intuitive at first glance because it seems to suggest walking does not benefit the system. Yet, we note that DSE quantifies the potential for improving system efficiency by scaling up the market. Since walking pushes passengers and vehicles to the major streets (particularly in street-hailing), it helps increase demand and supply densities without scaling up the market. As a result, the benefit of scale becomes less significant, leading to a lower DSE.

As detailed in Appendix E.9, besides the negative scale effect due to the conditions introduced in Proposition 8, walking introduces another negative effect to street-hailing due to the change in parameter θ_k, as detailed in the first row of Table 1. Therefore, walking is in general more likely to lower the DSE in street-hailing compared to e-hailing. This is also observed in the simulation results reported in Section 6.

Additionally, as the vehicle travel speed is usually much higher than walking speed, even if on local streets, it is safe to assume $\alpha^a \geq 2$. Then, the second condition further reduces to $\beta^a > (c_0+\beta^t)/2$ and the main trade-off comes to be among the values of walking time, in-vehicle time, and unit vehicle operation time. Specifically, walking is more likely to reduce the market scale economies if walking is more costly to passengers.

6. Simulation experiments

In this section, we run simulations over a subnetwork of Manhattan, New York, based on historical demand profiles. For each setting of passenger demand and vehicle supply, we simulate a one-hour operation for the four studied schemes $\{DS, WS, DE, WE\}$. The main purpose of this section is to assess the robustness of our conclusions under more realistic circumstances, both network and demand present heterogeneity that is not captured in the model. While we report results for a number of relevant indices, the most important ones are those in Section 6.3.5, where we compare the DSE obtained in the simulations and the ones predicted by our analytical model.

6.1. Network and demand

We leverage the publicly available taxi data in Manhattan, New York, and consider the peak period 7–8 AM on 15/01/2013. We selected the data from 2013 because they include detailed coordinates of trip origins and destinations, along with the trip starting and ending times. To be consistent with our model, the origin and destination of each trip are assigned to the closest intersection. Such a mapping has also been done in previous studies (e.g., Alonso-Mora et al., 2017; Simonetto et al., 2019; Fielbaum et al., 2022). It is worth noting that the observed trip origin is essentially the pickup location, not necessarily the true origin of the passenger (e.g., the passenger might have walked a bit before hailing the taxi). Nevertheless, as the exact trip origins are not traceable, we directly use the observed trip origins, acknowledging that our results might underestimate the amount of walking.

Since the taxi demand in Manhattan is strongly unbalanced, with most trips starting and ending in the midtown and downtown areas, we crop the Manhattan network developed in Fielbaum et al. (2021) and only consider the subnetwork to the south of Central Park, which contains 1966 nodes (intersections) and 4253 edges (road arcs). This provides us with a more uniform demand pattern, which is assumed in our analytical model. We prune unreasonable trips whose in-vehicle times are shorter than 5 min and finally end up with a demand reference with a total of 14,213 trips. For each simulation, we sample trips from the demand reference at different demand rates. Each scheme is simulated five times to reduce the impact of randomness. All results reported in this section are the average values.

To analyze the WS schemes, we further identify "major streets" that form a skeleton of the cropped Manhattan network. The procedure to construct this connected subnetwork is described in Appendix F. Fig. 5 plots the network used for simulations, where the blue nodes represent intersections where passengers are picked up and dropped off in WS.

6.2. Implementation details

6.2.1. Matching and vehicle cruising in e-hailing

In the simulations of e-hailing, we apply the matching algorithm developed by Fielbaum et al. (2021). While the algorithm was designed for on-demand ride-pooling, it is easily adapted to our case by setting the vehicle capacity equal to 1. In all simulations, we set the matching interval as $\delta_{DE} = \delta_{WE} = 1$ min. By the end of each matching interval, all feasible passenger–vehicle assignments are first found that do not violate an upper bound of total waiting time (5 min). An integer-linear program is then solved to obtain the assignments that optimize a customized objective (e.g., service rate). In the case of WE, the pickup-dropoff (PUDO) points are jointly selected in the first step that minimizes a weighted sum of walking time and in-vehicle time. Specifically, the weights are set according to the model parameters reported in Table A.1. Besides, a maximum walking time of 4 min is included as a constraint. Same as in Fielbaum et al. (2021), trips with different lengths are treated equally in the assignment. If a passenger fails to be picked up and starts the trip within the maximum waiting time (5 min), it will leave the system, and the trip is marked as unserved.

Similar to matching, vehicles are rebalanced in a centralized way in e-hailing. Specifically, vacant vehicles will be dispatched to undersupplied areas following the same procedure as Alonso-Mora et al. (2017) and Fielbaum et al. (2021).

(a) Whole network (b) A zoom-in area

Fig. 5. Network used in simulations. Blue nodes are kept in the subnetwork for WS. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

6.2.2. Matching and vehicle cruising in street-hailing

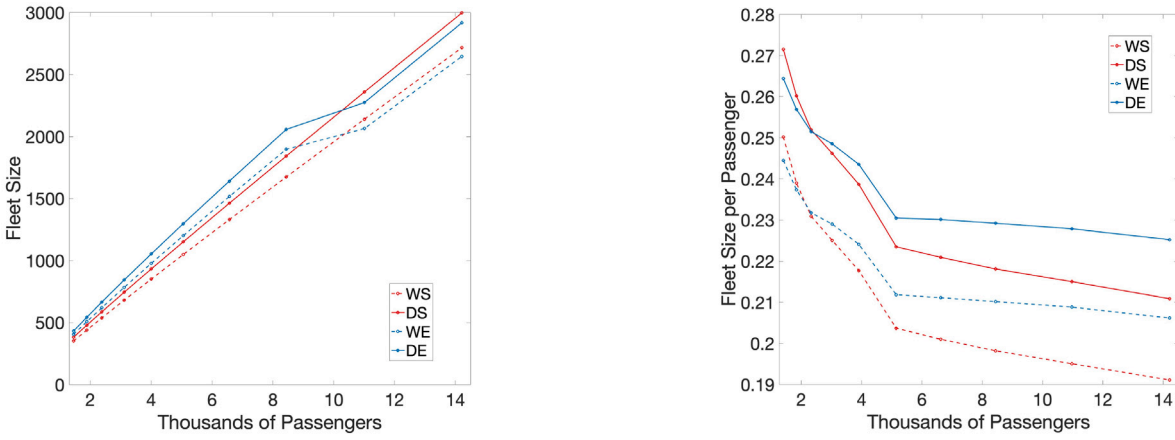
The simulations of DS and WS are the same except that they run on different networks. For WS, we first map trip origins and destinations from the full network to the subnetwork of major streets, ensuring the maximum walking time (4 min) is not violated meanwhile minimizing the in-vehicle time. In all simulations of street-hailing, the matching interval is set to be $\delta_{DS} = \delta_{WS} = 20$ seconds. By the end of each matching interval, we check, for each waiting passenger, whether there is a vacant vehicle that can reach the passenger's location within 20 s (approximately the time to traverse one street segment). If there are multiple vehicles, we assign the closest vehicle to the passenger. Same as e-hailing, a passenger would leave the system if it is not engaged with any vehicle within 5 min.

In line with the matching model, we simulate vehicle cruising behaviors differently with and without walking. In WS, vehicles only cruise on major streets. Thus we simply run simulations on the subnetwork. As for DS, vehicles cruise on all streets, instead of only on local streets as assumed in the matching model, because trips originate from any type of intersection. To implement random cruising, we assign trips to vacant vehicles with randomly generated artificial destinations, and vehicles take the shortest path to their destinations. Each vehicle travels along its paths for 20 s (at the same frequency as street-hailing matching), after which either it is matched with a passenger or continues its cruising by selecting a new artificial destination.

6.2.3. Choice of fleet size

For each scheme and demand rate, we run simulations with two fleet sizes. The first one is the optimal fleet size derived in Propositions 3–6, where the model parameters are calibrated according to the simulation environment (see Table F.1). Because this fleet size is solved as a function of trip demand, we call it the *endogenous fleet*. The resulting fleet sizes are depicted in Fig. 6. It can be observed from Fig. 6(a) that the fleet size increases almost linearly with demand in all schemes. This is expected as the in-vehicle time, which is linear with demand, contributes most of the vehicle time. The nonlinear part, on the other hand, is better illustrated in Fig. 6(b). For all schemes, the number of vehicles per passenger decreases with demand. Another expected finding is that the schemes with walking (WS and WE) require fewer vehicles compared to their non-walking counterparts. At a relatively low demand level, e-hailing requires more vehicles than street-hailing. The difference gradually expands as demand increases. As the trip number goes beyond a certain threshold, however, the e-hailing fleet size suddenly drops and then remains slightly lower than street-hailing. A closer look at the model outputs reveals that the sudden change corresponds to the emergence of passenger competition in e-hailing.

If the analytical model were perfect, the endogenous fleet would lead to a 100% service rate (i.e., all trips are served with the maximum waiting time constraint). This is, however, never the case in simulations. In fact, several model assumptions can hardly hold in simulations, including (i) the network is not a perfect grid network with alternative local and major streets, (ii) travel demand



(a) Fleet size (b) Number of vehicles per passenger

Fig. 6. Vehicle fleet for each scheme and demand set.

is not homogeneous over space, and (iii) the simulation may not reach a stationary state. Hence, we introduce a second set of fleet sizes that are computed as a fraction of the total trip numbers (15% in this study), which is called *exogenous fleet* hereafter. We wish to note that the simulation results from both types of fleet size are insightful. While results of exogenous fleets can better indicate the economies of scale, those of endogenous fleets provide insight into whether or not the model prediction error is consistent over demand rates and across schemes.

6.2.4. Performance metrics

We use the following four metrics to identify and compare scale economies in different schemes:

- Waiting time: Average passenger waiting time that includes matching and pickup times (walking time is not counted).
- Service rate: The fraction of trips served under the service quality constraint (i.e., maximum waiting time is 5 min).
- Utilization rate: The fraction of occupied vehicle time.
- Degree of scale economies (DSE): The average system cost divided by a numerical marginal cost.

6.3. Results

Recall that one factor that has been largely ignored in e-hailing matching is the competition among waiting passengers. We have analytically shown that it can fundamentally change the scale economies of e-hailing matching, as well as that of the whole market. Therefore, in what follows, we will first discuss the existence of passenger competition and then present the simulation results of each metric.

6.3.1. Passenger competition

To evaluate the intensity of passenger competition, we investigate how a passenger's dominance in the matching area evolves as demand increases. Specifically, at each e-hailing matching instance, we randomly draw several requests and compute the following: (i) the number of arcs within the matching area, and (ii) among the arcs identified in (i), the number of arcs to which the target request is the closest one, compared to all other unmatched requests. The former provides an estimate of the matching area, while the latter is considered a surrogate of the effective matching area. We define their ratio as the fraction of the dominance area preserved and plot the results against the demand rate in Fig. 7. It can be clearly seen that, with both fleet sizes, e-hailing matching is always subject to passenger competition—the effective matching area is far from 100% and plunges with demand.

It is worth noting that street-hailing is almost free from passenger competition in the simulations due to the matching mechanism described in Section 6.2.2. Specifically, each passenger can only hail vacant vehicles moving towards the intersection at which it is waiting. Hence, there is no competition unless some other passengers are waiting at the same intersection, which is rarely observed in the simulations.

6.3.2. Waiting time

Passenger waiting time is a widely used indicator of economies of scale in public transportation systems (e.g., Mohring, 1972; Zhang et al., 2019). If the average waiting time is observed to decrease with demand, then the system is likely to possess economies of scale. As shown in Fig. 8, in all schemes, the average waiting time indeed decreases with demand. As per Fig. 6, the vehicle supply increases almost proportionally with passenger demand. Hence, the observation in Fig. 8 can be translated into a statement that waiting time decreases as the market scales up, which directly implies economies of scale. Another observation from Fig. 8

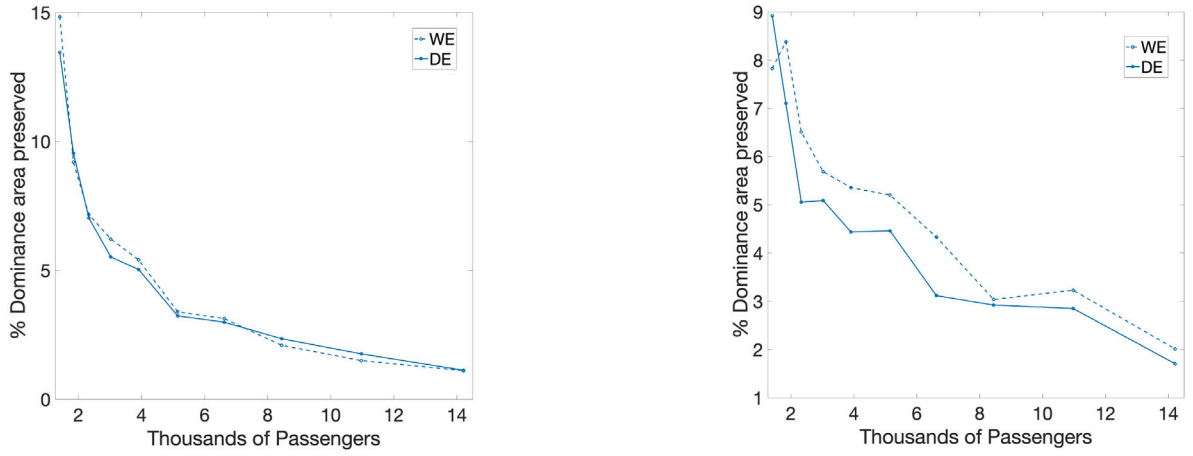


Fig. 7. Fraction of the dominance area preserved.

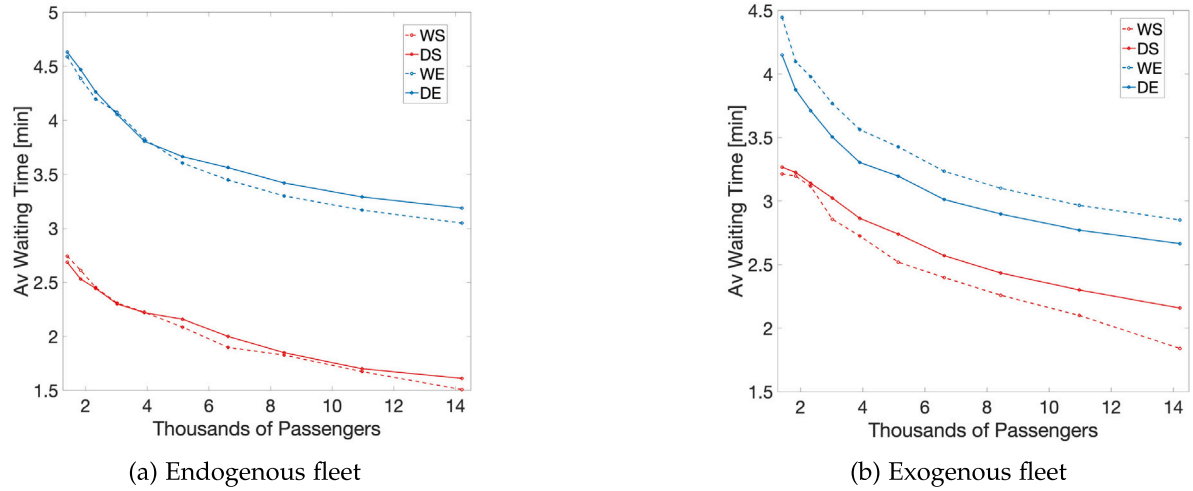


Fig. 8. Average waiting time.

is that the decreasing rate of waiting time reduces as demand grows. Such a diminishing scale effect is in line with our analytical results presented in Section 5.

Additionally, it is found the benefit of walking is less significant in the simulations with endogenous fleets compared to those with exogenous fleets. This is also expected because the impact of walking has already been incorporated into the model when optimizing the fleet size. As a result, WS (WE) achieves a similar service quality as DS (DE) with a smaller fleet size (see Fig. 6). On the other hand, with exogenous fleets, street-hailing is found to benefit a lot from walking in terms of reducing the waiting time, whereas e-hailing observes an increase in waiting time when walking is allowed. We note that the longer waiting time in WE is due to the specialized matching algorithm that prioritizes maximizing the service rate (i.e., the number of matched trips) in our simulations. This is also confirmed by Fig. 9(b), where the service rate of WE is 10% higher than DE.

6.3.3. Service rate

The service rate is another indicator of scale economies as it links between the system inputs and outputs. In particular, if the service rate remains stable as the market scales up, a property of constant returns to scale is expected. Instead, the market is likely to exhibit increasing returns to scale if one observes an increasing service rate.

As discussed above, the theoretical service rate with endogenous fleets is 100%. The gap between this ideal outcome and the realized service rate in simulations thus reflects on the prediction error due to the violation of model assumptions. Several conclusions are drawn from Fig. 9(a). First, the model in general overestimates the matching efficiency in street-hailing, which leads to a much lower service rate (lower than 75%). On the other hand, the model tends to underestimate the scale effect in

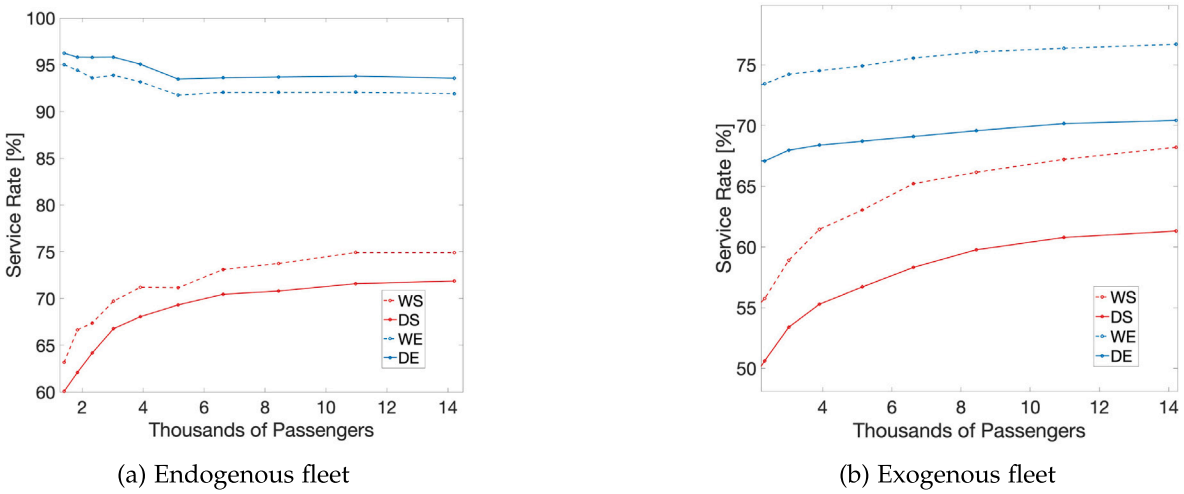


Fig. 9. Service rate.

street-hailing, which explains the fast growth in service rate as demand first increases. In contrast, our model correctly predicts the matching efficiency and scale economies of e-hailing as the trip loss is well-bounded and stable (consistently below 10%). The model also correctly captures the influence of walking as the gap between walking and non-walking schemes is also consistent across demand levels.

Fig. 9(b) further demonstrates our theoretical findings of scale economies. Compared to e-hailing, street-hailing shows a strong scale effect with a faster-increasing service rate. On the contrary, the growth of service rate in e-hailing is rather minor, which implies the market is prone to constant returns to scale. Walking significantly improves the service rate in both street-hailing and e-hailing, but the improvement varies little with the market scale. This finding confirms our analysis that walking does not fundamentally change the property of scale economies but only affects its exact value. Finally, in all schemes, the service rate gradually converges as the trip number continues to increase, which is in line with our conclusion that the scale effect diminishes with demand.

It is worth noting that the service rate of WE is slightly lower than DE with the endogenous fleet, in contrast to the substantial improvement with the exogenous fleet. The reason is two-fold. On the one hand, walking is the most beneficial when there is insufficient vehicle supply. As per Fig. 6(b), the endogenous fleet size is greater than 20% of demand while the exogenous fleet size is 15% of demand. Hence, the improvement generated by walking is less in the case of endogenous fleets. On the other hand, as also shown in Fig. 6(b), the endogenous fleet in WE is smaller than that in DE. As a result, the benefit of walking is not sufficient to offset the negative influence due to the loss of supply and finally, yields a decline in service rate.

6.3.4. Utilization rate

The utilization rate measures the fraction of occupied time per vehicle and thus indicates the system performance from the driver's perspective. As shown in Fig. 10, vehicle utilization in general increases with demand but gradually saturates. This result again confirms the scale effect exists but diminishes with demand. Regardless of the type of fleet, e-hailing drivers always enjoy a higher utilization rate meanwhile walking is always beneficial. Nevertheless, the gap between street-hailing and e-hailing decreases with the demand rate whereas that between walking and non-walking remains stable and even slightly expands in the case of street-hailing.

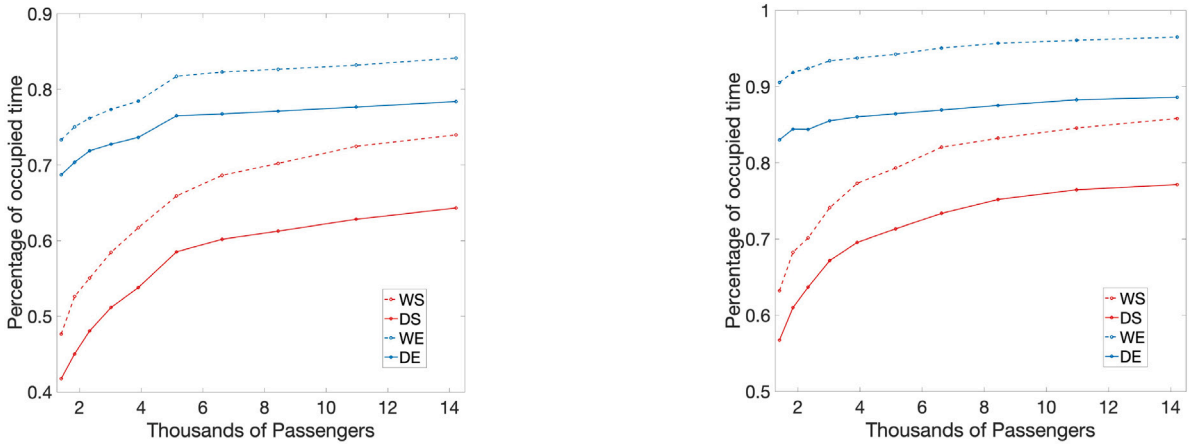
6.3.5. Degree of scale economies

Finally, we numerically evaluate the DSE of each scheme and compare the results with our theoretical findings (Propositions 3–6). However, there remains a gap in the computing of total cost. While the economic model assumes all trips are well served, in simulations some trips are not. Hence, we need to define the penalty for each unserved trip and include the cost of unserved trips in the total cost. Specifically, we set the penalty to be the passenger cost for the maximum walking and walking times (because they are the constraints corresponding to trip rejections).

DSE is defined as the ratio of average cost to marginal cost. While the former can be easily obtained from the simulations, the latter is not directly available. We compute the marginal costs via a linear approximation. Let Q_i, Q_{i+1} be two consecutive demand rate and TC_i, TC_{i+1} be their corresponding total system cost. The marginal cost at Q_i is then computed as $MC(Q_i) = \frac{TC_i - TC_{i+1}}{Q_i - Q_{i+1}}$.

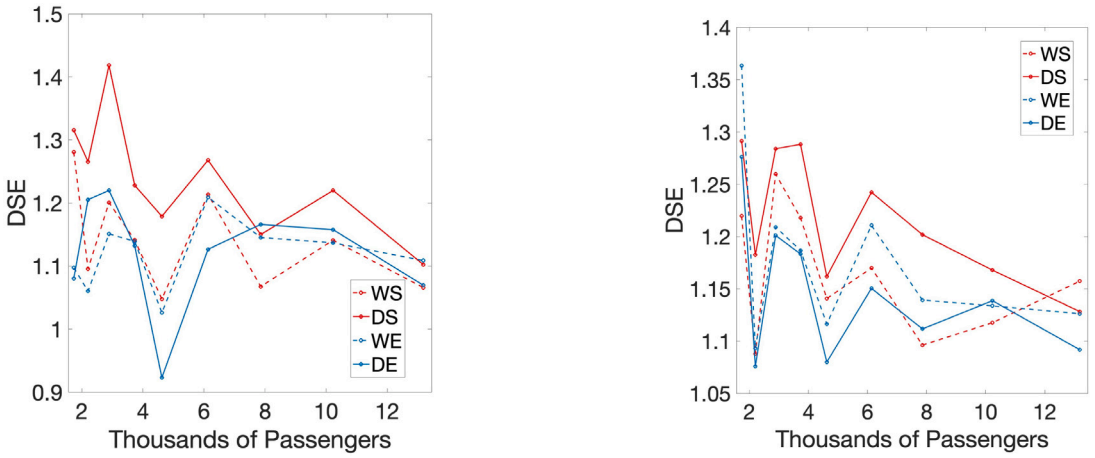
The numerical DSE values in different schemes are plotted in Fig. 11. While oscillating a lot due to the limited simulation rounds, several general patterns can be recognized:

- DSE in all combinations of matching mechanism and fleet size decreases with demand while remaining in the range of (1,2) except for the case of DE with the endogenous fleet, which drops to around 0.9. We expect such an exception is largely due to the randomness in the simulations.



(a) Endogenous fleet (b) Exogenous fleet

Fig. 10. Vehicle utilization rate.



(a) Endogenous fleet (b) Exogenous fleet

Fig. 11. Numerical degree of scale economies (DSE).

- Street-hailing always archives a higher DSE compared to its e-hailing counterpart (e.g., DS vs. DE).
- The difference in DSE between walking and non-walking schemes is ambiguous and does not vary with demand. Yet, walking is shown to have a negative impact on the DSE of street-hailing.

All these observations are consistent with the theoretical results presented in Propositions 3–6 and thus provide strong support to our proposed model. It is worth noting that the demand profile used in the simulations is constructed from historical taxi demand in New York City and shows a high spatial heterogeneity (e.g., hotspots). Therefore, the simulation outcomes also suggest the analytical model built upon a uniform market is sufficient to capture the key trade-off in scale economies and produce reliable estimates.

7. Practical implications

Our findings offer important implications for the practice of on-demand mobility that go beyond a monopoly market with a single operator. In general, with DSE greater than one, a market is prone to a *natural monopoly* that exploits the scale economies as much as

possible. However, monopoly does not necessarily lead to system efficiency and thus regulation becomes justifiable (Douglas, 1972; Tirachini, 2020). Particularly, a monopoly market can be risky in the case of e-hailing, where the central platform can manipulate vehicle supply and create artificial scarcity in order to raise prices (Joskow, 2007).

Regulation in an on-demand mobility market can be either introducing competition or enforcing constraints on operations (e.g., pricing). While the specific instruments go beyond the scope of this paper, the greater DSE in low-demand areas suggests that introducing competition is less beneficial or even infeasible. On the other hand, competition is preferable in high-demand areas where the DSE of e-hailing converges to one. In other words, little loss in service quality would be induced when demand is divided and served by several e-hailing platforms.

In the following, we summarize the practical implications of our findings for several other frequently discussed problems in on-demand mobility:

- Matching friction between supply and demand: The existence of matching friction and its variations with demand and supply partially determine DSE. In real practice, the overall matching friction could be lower than the model prediction based on the assumption of uniform distribution of supply and demand along with random vehicle cruising. The two main reasons are: (i) the existence of hotspots with a higher concentration of trip origins and destinations (Liu et al., 2010), and (ii) the strategic cruising of experienced drivers (Urata et al., 2021; Zhang et al., 2023). These factors would push the market closer to the scenario with higher demand, which is more efficient but has fewer gains from a demand increase due to the lower DSE.
  Besides, we have also assumed the trip origin and destination follow the same distribution, which is not necessarily satisfied in real-world practice either. When demand is highly imbalanced, much of the vehicle operation time would be dedicated to repositioning instead of cruising. Hence, the optimal fleet size would be larger than the one derived in this paper. These phenomena are expected to share among all four schemes studied. Since the ratio of repositioning time to travel demand is likely to be constant, it would not decrease as the market scales up. As a result, the DSE value would also decrease and become closer to 1.
- Platform competition and multi-homing drivers: Although the model established in this paper considers a single operator in the market, the results provide some insights into a market with competing operators, which is the case of e-hailing in most cities. Specifically, the scale economies highly depend on the market structure. If the market is fully segmented, i.e., each passenger or driver participates in just one platform, then all of our analyses remain valid for each platform separately. Yet, in real practice, drivers often join several platforms. Such a so-called *multi-homing* behavior essentially pools waiting passengers and vacant vehicles in the matching process (Zhang and Nie, 2021; Jiao and Ramezani, 2024). According to our findings, it leads to higher efficiency in low-demand areas but hardly brings additional benefits to a high-demand market. Nevertheless, other factors, such as the wage structure (Guo et al., 2023), could also largely affect the total vehicle supply, and accordingly, DSE.
- Service integration: Recently, *platform integrator* that consolidates trip requests and assigns them to different e-hailing platforms emerges to be a hot research topic (e.g., Li et al., 2024a; Zhou et al., 2024). We note that, when no matching priority or price adjustment is introduced into the matching process, the integration tends to have a similar impact on DSE as multi-homing. In other words, the benefit of integrating small platforms is much larger than that of integrating large platforms. This result also well explains why in real practice, the platforms joining in the integrator are mostly rather small (e.g., Gaode Map in China). However, matching and pricing are the primary operational strategies of the integrator and they both have significant influences on the market. Hence, the overall impact on DSE is not conclusive and this is indeed an intriguing direction for future research.
- Ride-pooling: Another matching mechanism that has not been characterized in the model is pooling, i.e., two or more passengers served by a single vehicle. Yet, our findings provide hints on how DSE would change when pooling is integrated. In essence, effective pooling reduces the necessary fleet for a given demand. Hence, the market is expected to exhibit a higher DSE. Besides, pooling is expected to reduce passenger competition as some passengers can now share the same vehicle. Since passenger competition is a major cause of the compromised scale effect in e-hailing, pooling may further strengthens the scale economies in e-hailing. However, pooling also induces extra detours in the pickup and dropoff phases, which potentially compromise the scale economies. Accordingly, the pooling and routing algorithms might become more influential factors (Liu et al., 2023; Fielbaum et al., 2023; Lehe et al., 2021). Yet, obtaining a closed-form expression of detours in ride-pooling has proved to be particularly challenging unless strong assumptions and simplification are introduced (Wang et al., 2021; Mühle, 2023; Fielbaum and Pudāne, 2024). Therefore, extending this study to a reliable approximation of DSE for ride-pooling requires further research efforts.

8. Conclusions

This study investigates the impact of walking and e-hailing on the scale economies of on-demand mobility services. To this end, we develop a physical matching model that characterizes the detailed matching process between waiting passengers and cruising vehicles in a grid network. The expected matching time, pickup time, and walking time when applicable at the stationary state are derived for four particular service schemes: street-hailing without walking (DS), street-hailing with walking (WS), e-hailing without walking (DE), and e-hailing with walking (WE). These results are then plugged into a system-optimum fleet sizing problem to derive the optimal fleet as a function of demand rate. Accordingly, the scale economies in each scheme are evaluated as the degree of scale economies (DSE) under the optimal fleet size.

We analytically prove that the system exhibits economies of scale in all studied schemes with DSE falling in the range of (1,2) and monotonically decreasing with demand. The DSE in the endogenous matching process, however, is distinct between street-hailing and e-hailing, and further, highly related to the existence of passenger competition. In particular, if there is no passenger competition, the DSE of the matching process is constant and equals 2 in street-hailing and 1.5 in e-hailing, respectively. These findings are consistent with the findings of previous studies (Douglas, 1972; Arnott, 1996). While passengers are unlikely to compete with each other in street-hailing, it is quite often the case in e-hailing. We show that the DSE of e-hailing matching with passenger competition is no longer a constant but also decreases with demand. This finding is also in line with existing empirical evidence (Zhang et al., 2019) but has never been proved analytically. Finally, we show that, although walking does not fundamentally change the property of DSE, it does affect its value with opposing forces. On the one hand, the reduced pickup and in-vehicle time improve the scale effect. On the other hand, the extra walking time compromises the scale economies. Additionally, the vehicle concentration induced by the passenger walking behaviors in street-hailing tends to further lower its DSE.

To validate these theoretical findings, we conduct a series of simulations on a subnetwork of Manhattan based on historical taxi trip demand. The results demonstrate the economics of scale in on-demand mobility, showcase the benefits of walking, prove the existence of passenger competition in e-hailing, and generally agree with the DSE values predicted in our theory. It is worth noting that several assumptions introduced to the model (e.g., uniform demand distribution, regular network structure, and no rejections) are relaxed in the simulation, which thus further verifies that our proposed model successfully captures the main trade-off and key factors of scale economies.

This study has several limitations that also open directions for future research. First, the fleet sizing problem has assumed a constant cost per unit vehicle. When the labor supply is elastic, it shall be replaced with an inverse function of fleet size, which could yield a different optimal fleet size and accordingly, a different DSE. We leave the study on DSE under various labor elasticities to future research. On the other hand, we have assumed all passengers either walk or not, though, in reality, they may be heterogeneous. The vehicle cruising may also adapt to passengers' walking behaviors. Hence, another future direction is to extend the current model to capture scenarios with mixed walking behaviors. Other heterogeneity among passengers and drivers (e.g., trip origin and destination, cruising strategies, willingness to wait) also potentially affects DSE. A comprehensive sensitivity analysis of various factors is regarded as a relevant direction for future work. Besides, when the demand is large enough, taxis and e-hailing can contribute significantly to congestion (Erhardt et al., 2019; Yang et al., 2005), which is a source of scale diseconomies. Such an impact, however, has not been captured in the current model as the vehicle speeds are assumed to be constant. Therefore, a future direction is to investigate the congestion effect on the system scale economies. Finally, the simulations conducted in this study only provide a preliminary validation. More sophisticated numerical analyses are still needed to verify whether or not and in which condition our theories hold.

CRediT authorship contribution statement

Kenan Zhang: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization. **Javier Alonso-Mora:** Writing – review & editing, Validation, Supervision, Resources, Project administration, Funding acquisition, Conceptualization. **Andres Fielbaum:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization.

Acknowledgments

The work was partially funded by the Swiss National Science Foundation (219232). We would also like to thank Dr. Yu Nie from Northwestern University for providing insightful feedback on this work.

Appendix A. Notations

See Table A.1.

Appendix B. Key parameters in the matching model

B.1. Matching area

As shown in Fig. B.1, there are 4 arcs at a distance i from the passenger in street-hailing and $4(2i-1)$ arcs in e-hailing. Hence, the matching area, i.e., the number of arcs within the matching radius R, is $A = \sum_{i=1}^{R} 4 = 4R$ for street-hailing and $A = \sum_{i=1}^{R} 4(2i-1) = 4R^2$ for e-hailing.

Table A.1
Notations and default values.

Variable	Description	Unit	Value
L	Number of local streets between every two major streets		3
s	Length of each street segment	m	120
v	Vehicle travel speed on major street	km/h	25
α^l	Speed scaling factor for local street		1.67
α^a	Speed scaling factor for walking		5
V ($\tilde{V}$)	(effective) idle vehicle density	veh/arc	
W	Waiting passenger density	pax/arc	
Q	Demand rate	pax/arc/s	
w^m	Matching time	s	
w^p	Pickup time	s	
w^a	Walking time	s	
A_k ($\widetilde{A}_k$)	(effective) matching area in scheme k	arc	
γ_k	Passengers competition in scheme k		
$\theta_{DS}(\theta_{WS})$	Vehicle density correction factor in street-hailing		1.33 (2)
R_{DS}, R_{WS}	Matching radius in street-hailing	arc	1
R_{DE}, R_{WE}	Matching radius in e-hailing	arc	14
δ_{DS}, δ_{WS}	Matching interval in street-hailing	s	20
δ_{DE}, δ_{WE}	Matching interval in e-hailing	s	20
δ	Time to drive through one segment of major street	s	17.28
ρ_k	Pickup time correction factor in scheme k		
D^p (d^p)	Expected (approximated) e-hailing pickup distance	arc	
D^a	Expected walking distance	arc	
c_1	Approximation parameters in e-hailing pickup distance		$\sqrt{\pi}/4$
c_2			$e\sqrt{\pi}$
$\bar{\tau}$	Average door-to-door trip duration	s	656.5
τ_k	Average trip duration in scheme k	s	
w_k	Total waiting time in scheme k	s	
c_0	Operation cost per human-driven vehicle	\$/h	20
β^m	Value of time for matching	\$/h	15.00
β^p	Value of time for pickup	\$/h	12.51
β^a	Value of time for walking	\$/h	14.51
β^t	Value of time for in-vehicle time	\$/h	10.00

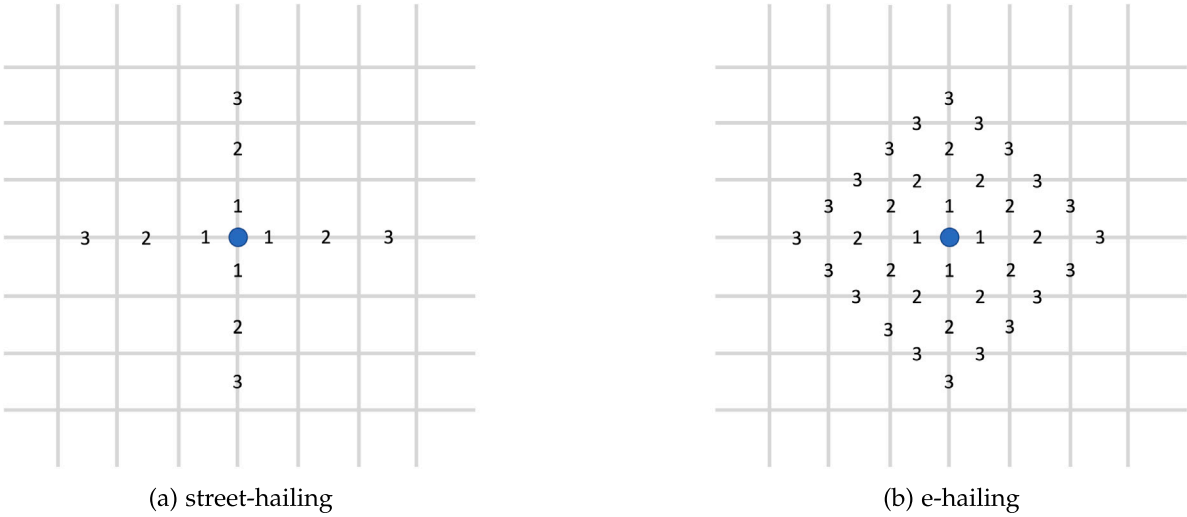


Fig. B.1. Illustration of matching radius and area.

B.2. Vacant vehicle density factor

As per Assumption 3, the distribution of vacant vehicles in street-hailing depends on the existence of walking. Specifically, vehicles only cruise on local streets in DS whereas only on major streets in WS. To derive θ_k in DS and WS, we first consider the network shown in Fig. 1 with K major streets per side and L local streets between every two major streets. In this case, the number

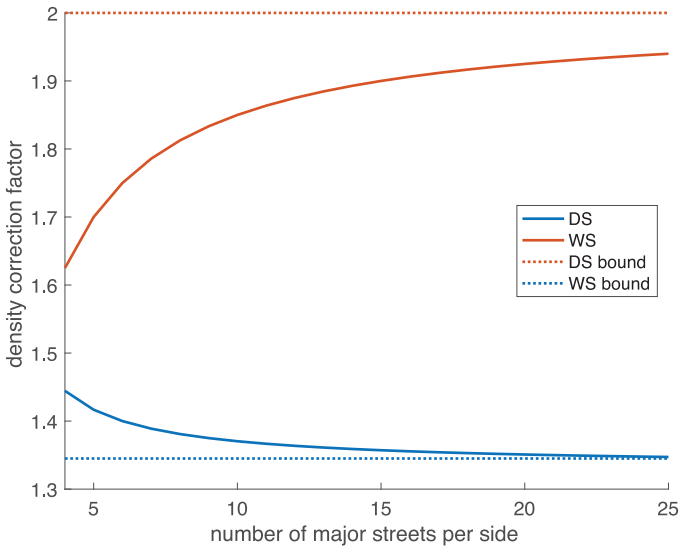


Fig. B.2. Density correction factor against the number of major streets per side (with $L = 3$).

of local and major arcs can be derived as

$$N_{\text{local}} = 2L(K-1)^2(L+1), \tag{B.1}$$

$$N_{\text{major}} = 2K(K-1)(L+1). \tag{B.2}$$

The density factors are then computed as

$$\theta_{DS} = \frac{N_{\text{local}} + N_{\text{major}}}{N_{\text{local}}} = 1 + \frac{1}{L(1-1/K)}, \tag{B.3}$$

$$\theta_{WS} = \frac{N_{\text{local}} + N_{\text{major}}}{2N_{\text{major}}} = \frac{1+L}{2} - \frac{L}{2K}. \tag{B.4}$$

In Eq. (B.4), θ_{WS} is further divided by two because the passenger is waiting at a Type-3 intersection (see Fig. 1). As the study region scales up, i.e., $K \to \infty$, the factors have limits $\theta_{DS} \to 1 + \frac{1}{L}$ and $\theta_{WS} = \frac{1+L}{2}$, as illustrated in Fig. B.2 with $L = 3$. These limits are then used as the default values in the numerical experiments.

B.3. Walking distance

Recall there are L local streets between every two major streets. Hence, the number of Type-2 intersections inside each major block is L^2. The derivation of expected walking distance depends on whether L is odd or even, though it leads to the same result.

When L is even, there are 4 local intersections with walking distance $L/2$, 12 local intersections with walking distance $L/2-1$, and so on. The general expression of expected walking distance is derived as

$$\begin{aligned} D^a &= \frac{1}{L^2}\left[\sum_{i=1}^{L/2}(8i-4)\left(\frac{L}{2}-i+1\right)\right] = \frac{1}{L^2}\left[4\left(\frac{L}{2}+1\right)\sum_{i=1}^{L/2}(2i-1) - 8\sum_{i=1}^{L/2}i^2 + 4\sum_{i=1}^{L/2}i\right] \\ &= \frac{(L+1)(L+2)}{6L}. \end{aligned} \tag{B.5}$$

On the other hand, when L is odd, the central local intersection has walking distance $(L+1)/2$, surrounded by 8 local intersections with walking distance $(L+1)/2-1$, and so on. In this case, the expected walking distance is given by

$$\begin{aligned} D^a &= \frac{1}{L^2}\left[\frac{L+1}{2} + \sum_{i=2}^{\frac{L+1}{2}} 8(i-1)\left(\frac{L+1}{2}+1-i\right)\right] \\ &= \frac{1}{L^2}\left[\frac{L+1}{2} + 4(L+1)\sum_{i=2}^{\frac{L+1}{2}}(i-1) - 8\sum_{i=2}^{\frac{L+1}{2}}(i-1)^2\right] = \frac{(L+1)(L+2)}{6L}. \end{aligned} \tag{B.6}$$

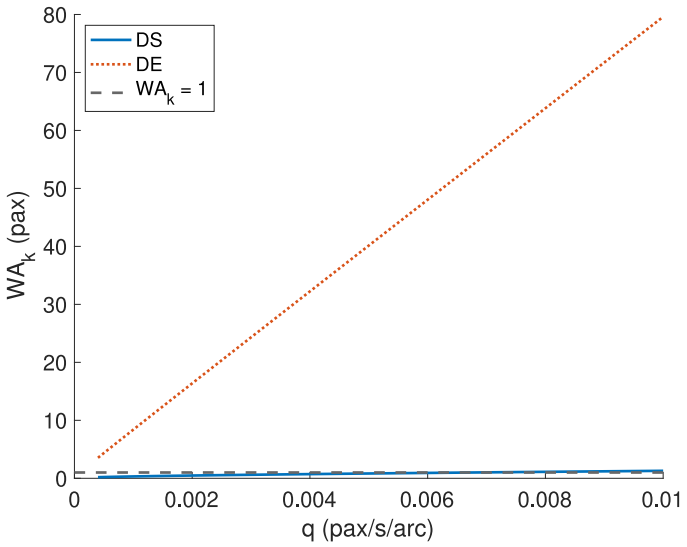


Fig. B.3. Existence of passenger competition in DS and DE (with default parameter values in Table A.1).

B.4. Passenger competition in the analytical model

In Section 4, we introduce $\gamma_k = \max(A_k W, 1)$ to capture passenger competition in the matching process. Later in the analysis, we argue it has a negligible impact on street-hailing due to the limited matching radius. Here, we validate this state meanwhile demonstrating the prevalence of passenger competition in e-hailing. Specifically, we solve the steady-state waiting passenger density W using the matching models of street-hailing and e-hailing without passenger competition (without presuming the existence of passenger competition). Fig. B.3 plots the number of waiting passengers within a matching area $A_k W$ against the demand rate Q. It can be observed that the assumption of no passenger competition is violated in e-hailing at a fairly low demand rate, whereas it is generally satisfied in street-hailing. While the increased matching radius in e-hailing substantially reduces the matching friction between passengers and vehicles, it also induces considerable competition among passengers. This phenomenon has also been recognized in some previous work (e.g., Zhang et al., 2019), but unfortunately not yet been widely adopted in recent studies on e-hailing services.

Appendix C. Remarks on Assumption 4

To simplify the derivation of matching and pickup times, we have assumed $\delta_k \geq \delta$ (see Assumption 4). One could argue that in street-hailing, this assumption does not hold as passengers are constantly looking for vacant vehicles and thus the matching interval δ_k is infinitesimal. As discussed below, such a deviation, however, does not cause much error in the analysis of passenger waiting time.

Suppose the matching interval is infinitesimal. Then the pickup time is $R_k\delta$ for all passengers and the matching time corresponds to the moment when the first vacant vehicle enters the matching area. Let $\tilde{w}_k^m$ denote the matching time in this case, then the total waiting time is $\tilde{w}_k^m + R_k\delta$.

Now let us consider the case with $\delta_k \geq \delta$ and suppose the pickup vehicle is $R_k s - r$ from the passenger when matched with the passenger with $r \leq \delta_k v$. Accordingly, the matching time in this case is $w_k^m = \tilde{w}_k^m + r/v$ and the pickup time is $w_k^p = (R_k s - r)/v = R_k\delta - r/v$. Therefore, the total waiting time remains the same as the true scenario despite that the matching (pickup) time is overestimated (underestimated) by $r/v \leq \delta_k$ (less than 20 s in the simulation experiments).

Appendix D. Profit-maximizing fleet size

The general pricing and fleet sizing problem of a profit-driven ride-hailing platform is formulated as follows:

$$\max_{p,N} \quad \Pi(p,N) = pQ - c_0 N, \tag{D.1a}$$

$$s.t. \quad u = p + \beta^m w^m + \beta^p w^p + 2\beta^a w^a + \beta^t \tau, \tag{D.1b}$$

$$Q = q(u), \tag{D.1c}$$

$$N = V + Q(w_k^p + \tau_k), \tag{D.1d}$$

$$W = Q w_k^m, \tag{D.1e}$$

$$w_k^p = f_k^p(V, W), \tag{D.1f}$$

$$w_k^m = f_k^m(V, W), \tag{D.1g}$$

where u is the generalized cost that consists of trip fare p and the monetary cost of total travel time $\beta^m w^m + \beta^p w^p + 2\beta^a w^a + \beta^t \tau$, $q(\cdot)$ is the demand function that decreases with the generalized cost, and all other constraints remained the same as per (9). The subscript of service scheme k is dropped for notation simplicity.

The formulation (D.1) has been widely used in the literature (e.g., Douglas, 1972). Now consider the platform aims to serve a desired demand level Q as per the setting in Section 5. This is achieved by sustaining a reasonable service quality measure by endogenous travel times $\check{w}^m, \check{w}^p, \check{w}^a$ for each price $\check{p}$ within a certain range. Accordingly, Problem (D.1) is reduced to

$$\max_{N \geq 0} \quad \Pi(\check{q}, N) = \check{p}Q - c_0 N, \tag{D.2a}$$

$$s.t. \quad \beta^m w^m + \beta^p w^p + 2\beta^a w^a + \beta^t \tau = q^{-1}(Q) - \check{p}, \tag{D.2b}$$

$$N = V + Q(w_k^p + \tau_k), \tag{D.2c}$$

$$W = Q w_k^m, \tag{D.2d}$$

$$w_k^p = f_k^p(V, W), \tag{D.2e}$$

$$w_k^m = f_k^m(V, W), \tag{D.2f}$$

where $q^{-1}(\cdot)$ denotes the inverse demand function.

Problem (D.2) is equivalent to the following with additional dual variable λ:

$$\max_{\lambda \geq 0} \min_{N \geq 0} \quad \mathcal{L}(N, \lambda) = c_0 N + \lambda(\beta^m w^m + \beta^p w^p + 2\beta^a w^a + \beta^t \tau), \tag{D.3a}$$

$$s.t. \quad N = V + Q(w_k^p + \tau_k), \tag{D.3b}$$

$$W = Q w_k^m, \tag{D.3c}$$

$$w_k^p = f_k^p(V, W), \tag{D.3d}$$

$$w_k^m = f_k^m(V, W). \tag{D.3e}$$

Accordingly, for each feasible price p, we can solve a pair of optimal solutions N, λ. Let λ^* be the dual solution at the profit-maximizing price p^*. Then, the profit-maximizing fleet size is the same as the solution to the optimal fleet sizing problem with a revised objective:

$$\widetilde{TC}(N, Q) = c_0 N + \mu^* Q(\beta^m w^m + \beta^p w^p + 2\beta^a w^a + \beta^t \tau), \tag{D.4}$$

where the scale factor $\mu^* = \lambda^*/Q$.

It can be easily verified that the properties of DSE derived from this revised system-optimal fleet sizing problem remain the same as per those presented in Sections 5.2 and 5.3, though its absolute value changes due to the scaled user cost.

Appendix E. Proofs

E.1. *Proposition* 1

Due to Assumption 1, the number of vacant vehicles N_k^v within the effective matching area follows a spatial Poisson distribution, where the effective vehicle density $\widetilde{V}_k$ is the rate parameter and the effective matching area $\widetilde{A}_k$ gives the area of the bounded region.[3] Hence, at one matching instance, the probability that at least one vehicle appears in the effective matching area is given by $p_k^m = 1 - Pr(N_k^v = 0) = 1 - \exp(-\widetilde{A}_k \widetilde{V}_k)$. Thanks to Assumption 4, the consecutive matching instances are independent Bernoulli trials with a success rate p_k^m. Hence, the expected number of matching instances until a successful match is $1/p_k^m$, which yields the expected matching time

$$w_k^m = \left(\frac{1}{p_k^m} - \frac{1}{2}\right)\delta_k = \left[\frac{1}{1 - \exp(-\widetilde{A}_k \widetilde{V}_k)} - \frac{1}{2}\right]\delta_k \approx \left(\frac{1}{\widetilde{A}_k \widetilde{V}_k} + \frac{1}{2}\right)\delta_k. \tag{E.1}$$

In Eq. (E.1), the first term in the parentheses gives the expected number of matching instances while the second accounts for the passenger arrival time within the first matching interval. Here we introduce an approximate $\frac{1}{1-\exp(-x)} \approx \frac{1}{x} + 1$ as illustrated in Fig. E.1, where the error is computed as the approximation minus the exact value. Accordingly, Eq. (E.1) overestimates matching time with an approximate error bounded by $0.5\delta_k$ when $\widetilde{A}_k \widetilde{V}_k \to 0$.

[3] See https://web.mit.edu/urban_or_book/www/book/chapter3/3.8.html.

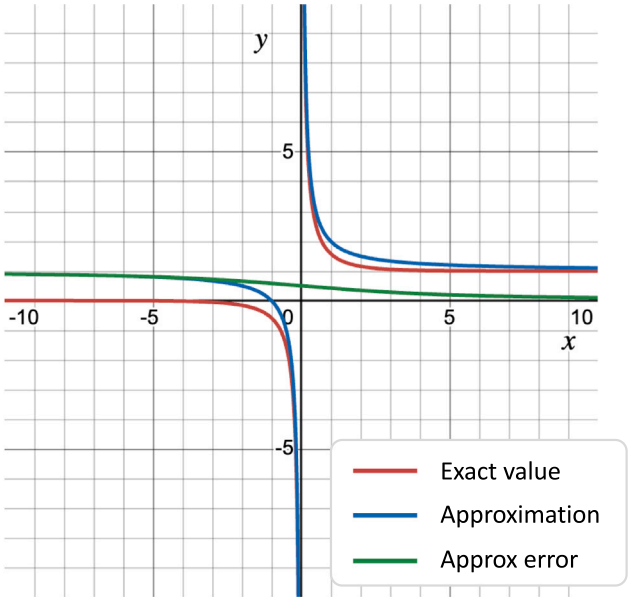


Fig. E.1. Approximation introduced in Eq. (E.1).

E.2. Proposition 2

For both street-hailing and e-hailing, we first derive the pickup distance and then compute the pickup time according to the fraction of the pickup route on local streets.

Suppose there are N vacant vehicles inside the effective matching area and each vehicle v is at a distance d_v from the passenger. Hence, the pickup distance is given by $d_{\min} = \min_v d_v$, i.e., the passenger is picked up by the closest vehicle within the effective matching area. Then, the expected pickup distance is given by $D_p = \mathbb{E}[\mathbb{E}[d_{\min}|N]]$.

For street-hailing, it is safe to assume at most one vacant vehicle enters the effective matching area in each matching interval. Therefore, the expected pickup distance is reduced to

$$d_k^p = \mathbb{E}[d_v] = \sum_{i=1}^{\widetilde{R}} i\Pr(d_v = i) = \sum_{i=1}^{\widetilde{R}} \frac{i}{\widetilde{R}} = \frac{1+\widetilde{R}}{2},\ k \in \{DS, WS\}, \tag{E.2}$$

where $\widetilde{R}$ is the effective matching radius corresponding to $\widetilde{A}$. Eq. (E.2) finally yields the expression in Proposition 2 by replacing $\widetilde{R}$ with R_k/γ_k and subtracting the overestimated half arc.

As for e-hailing, we first introduce the approximation $D_p = \mathbb{E}[\mathbb{E}[d_{\min}|N]] \approx \mathbb{E}[d_{\min}|\mathbb{E}[N]]$ and replace $\mathbb{E}[N] = \widetilde{V}\widetilde{A}$. It then yields the approximate pickup distance

$$\begin{aligned} d_k^p &= \mathbb{E}[d_{\min}|\widetilde{V}\widetilde{A}] = \sum_{i=1}^{\widetilde{R}} \Pr(d_{\min} \geq i|\widetilde{V}\widetilde{A}) \\ &= \sum_{i=1}^{\widetilde{R}} \Big(\Pr(d_v > i)\Big)^{\widetilde{V}\widetilde{A}} = \sum_{i=1}^{\widetilde{R}} \left(\sum_{k=i+1}^{\widetilde{R}} \Pr(d_v = k)\right)^{4\widetilde{V}\widetilde{R}^2} = \sum_{i=1}^{\widetilde{R}} \left(1 - \frac{i^2}{\widetilde{R}^2}\right)^{4\widetilde{V}\widetilde{R}^2},\ k \in \{DE, WE\}, \end{aligned} \tag{E.3}$$

which can be further approximated as follows:

$$\sum_{i=1}^{\widetilde{R}} \left(1 - \frac{i^2}{\widetilde{R}^2}\right)^{4V\widetilde{R}^2} \approx \sum_{i=1}^{\widetilde{R}} \exp(-4Vi^2) \approx \int_0^{\widetilde{R}} \exp(-4Vx^2)\mathrm{d}x = \frac{1}{4}\sqrt{\frac{\pi}{V}}\mathrm{erf}(2\sqrt{V}\widetilde{R}), \tag{E.4}$$

where the first approximation is due to the product limit formula of e^x and the second smooths the sum over $\widetilde{R}$ with an integral over range $[0, \widetilde{R}]$.

The performance of approximation Eq. (E.4) is illustrated in Fig. E.2. Specifically, we plot the approximation error computed as the approximation minus the exact value. In general, the approximation tends to overestimate the effective matching radius, though the error is well bounded by 0.5 arc except when the vehicle supply is scarce ($V < 0.1$ veh/arc) and the effective matching radius is small (e.g., $\widetilde{R} = 5$ arc).

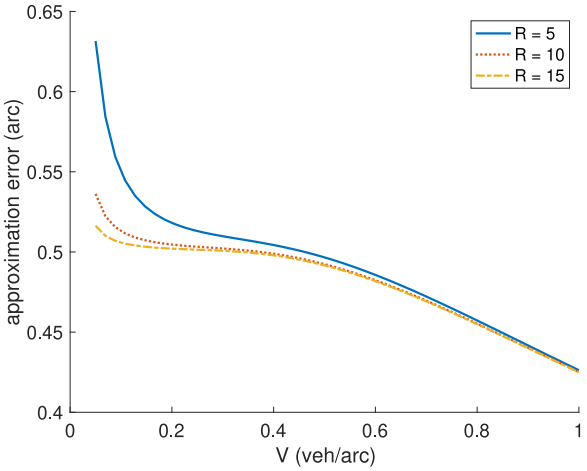


Fig. E.2. Approximation error of Eq. (E.4) at different effective matching radius.

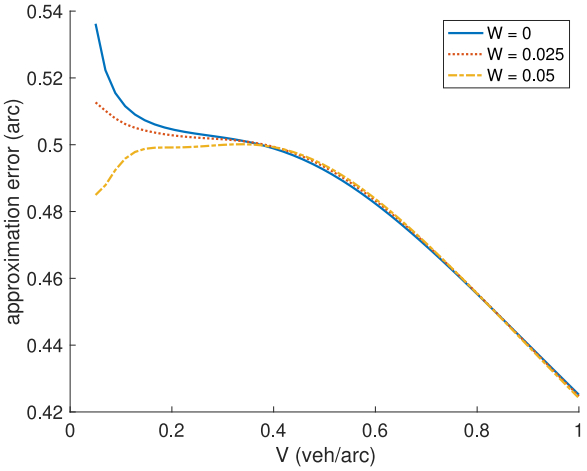


Fig. E.3. Approximation error of pickup distance at different passenger competition levels.

In practice, the matching radius in e-hailing is usually set sufficiently large, e.g., $R > 10$. When there is no passenger competition, we have $\widetilde{R} = R$ and $\mathrm{erf}(2\sqrt{V}R) \approx 1$. When there exists passenger competition, the error function is further expanded as

$$\begin{aligned}
\mathrm{erf}(2\sqrt{V}\widetilde{R}) &= 1 - \frac{2}{\sqrt{\pi}} \int_{2\sqrt{V}\widetilde{R}}^{\infty} \exp(-x^2)\mathrm{d}x \\
&\approx 1 - \frac{1}{\sqrt{\pi}} \exp(-4V\widetilde{R}^2) = 1 - \frac{1}{\sqrt{\pi}} \exp\left(-\frac{V}{W}\right) \\
&\approx 1 - \frac{1}{\sqrt{\pi}} (1 + eV/W)^{-1} = \frac{(\sqrt{\pi}-1)W + e\sqrt{\pi}V}{(\sqrt{\pi}-1)W + (W + e\sqrt{\pi}V)} \\
&\approx \frac{e\sqrt{\pi}V}{W + e\sqrt{\pi}V} = \left(1 + \frac{W}{e\sqrt{\pi}V}\right)^{-1},
\end{aligned} \tag{E.5}$$

where the first approximation is due to linear interpolation, the second is due to Taylor approximation, and the third is due to the much smaller value of $(\sqrt{\pi}-1)W$ compared to the other terms. Plugging Eq. (E.5) back into Eq. (E.4) then yields the expression of d^p in Eq. (4).

Fig. E.3 plots the error of pickup distance approximation Eq. (4), again, computed as the approximation minus the exact value. Similar to Fig. E.2, it shows that the approximation tends to overestimate the pickup distance with an error consistently below 0.5 arc when $V > 0.1$ veh/arc. Otherwise, it shows different behaviors, depending on the passenger competition levels.

Finally, we derive the factor ρ_k that accounts for the pickup route on local streets. For street-hailing, it simply concludes $\rho_k = \alpha^l$ for $k = DS$ and $\rho_k = 1$ for $k = WS$ as vacant vehicles are cruising on one type of streets in each scheme (see Assumption 3). In

contrast, in e-hailing, only part of the pickup route happens on local streets. Let $d^p_{k,\text{local}}$ be the length of the pickup route on local streets, then the pickup time is given by

$$w^p = \left[\alpha^l d^p_{k,\text{local}} + (d^p_k - d^p_{k,\text{local}})\right]\delta = \left[1 + (\alpha^l - 1)\frac{d^p_{k,\text{local}}}{d^p_k}\right] d^p_k \delta = \rho_k d^p_k \delta. \tag{E.6}$$

When the pickup vehicle is very close to the passenger ($d^p_k \to 0$), the pickup route is fully on the same type of street: $d^p_{k,\text{local}} = d^p_k$ in DE and $d^p_{k,\text{local}} = 0$ in WE. Yet, the difference due to walking in this case is almost negligible. Hence, we focus on the approximation of ρ_k when the pickup distance is relatively large ($d^p_k \to R_k$). In this case, it is reasonable to assume the vehicle makes the majority of its pickup trip on major streets and the length of the pickup route on local streets equals the sum of distances from the vehicle location and from the trip origin to the closest major street, respectively. Depending on where the vehicle is cruising, the first leg is either zero (i.e., on a major street) or with an average of D^a (i.e., on a local street). The probability of each case depends on the ratio of local and major street arcs in the network, i.e., θ_k derived in Appendix B.2. The second leg is simply D^a for DE and zero for WE. With these considerations, the correction factor is finally specified as

$$\rho_k = \begin{cases} 1 + (\alpha^l - 1)\left(1 + \frac{1}{\theta_{DS}}\right)\frac{D^a}{R_k}, & k = DE, \\ 1 + (\alpha^l - 1)\frac{D^a}{\theta_{DS} R_k}, & k = WE. \end{cases} \tag{E.7}$$

In Eq. (E.7), $1/\theta_{DS}$ gives the probability of the vehicle cruising on local streets and d^p_k is replaced by R_k.

E.3. *Lemma* 1

For street-hailing, the market equilibrium constraints are reduced to

$$\begin{cases} N = V + Q\left(\frac{\rho^p_k}{2} R_k \delta + \tau_k\right), \\ W = Q\left(\frac{1}{\theta_k A_k V} + \frac{1}{2}\right)\delta, \end{cases} \quad \Rightarrow \quad \begin{cases} V = N - Q\left(\frac{\rho^p_k}{2} R_k \delta + \tau_k\right), \\ W = \left(\frac{Q\delta}{\theta_k A_k}\right)\frac{1}{V} + \frac{\delta}{2}. \end{cases} \tag{E.8}$$

Hence, a sufficient condition for positive solutions of V and W is $N - Q\left(\frac{\rho^p_k}{2} R_k \delta + \bar{\tau}\right) > 0$.

For e-hailing without passenger competition, the market equilibrium constraints reduce to

$$\begin{cases} N = V + Q\left(\frac{c_1 \rho^p_k \delta_k}{\sqrt{V}} + \tau_k\right), \\ W = Q\left(\frac{1}{A_k V} + \frac{1}{2}\right)\delta, \end{cases} \quad \Rightarrow \quad \begin{cases} \frac{(c_1 \rho^p_k \delta_k Q)^2}{V} = [V - (N - Q\tau_k)]^2, \\ W = \left(\frac{Q\delta}{A_k}\right)\frac{1}{V} + \frac{\delta}{2}. \end{cases} \tag{E.9}$$

In this case, the sufficient condition for positive V and W is $N - Q\bar{\tau} > 0$.

Finally, for e-hailing with passenger competition, we first solve the implicit function Eq. (9e) to derive W as a function of V:

$$W = \left(\frac{W}{V} + \frac{1}{2}\right)\delta_k Q \quad \Rightarrow \quad W = \frac{\delta_k Q V}{2(V - \delta_k Q)}. \tag{E.10}$$

Thus, a necessary and sufficient condition for $W > 0$ is $V > \delta_k Q$.

On the other hand, Eq. (9b) is evaluated as

$$N = V + Q\left[\frac{c_1 \rho^p_k \delta_k}{\sqrt{V}}\left(1 + \frac{W}{c_2 V}\right)^{-1} + \tau_k\right] \quad \Rightarrow \quad \frac{(c_1 \rho^p_k \delta_k Q)^2}{V} = [V - (N - Q\tau_k)]^2 \left(1 + \frac{W}{c_2 V}\right)^2 \tag{E.11}$$

From Eq. (E.10), we can further derive $\mathrm{d}W/\mathrm{d}V = -2(W/V)^2 < 0$. Therefore, $\left(1 + \frac{W}{c_2 V}\right)$ decreases with V but remains larger than 1. Accordingly, the solution of V falls in the range of $(N - Q\tau_k, \infty)$ and a sufficient condition for positive V and W is $N - Q(\bar{\tau} + \delta_k) > 0$.

E.4. *Proposition* 3

In street-hailing, the matching time is independent of W while the pickup time is independent of both V and W. Then, plugging $V = N - \left(\frac{\rho_k}{2} R_k \delta + \tau_k\right) Q$ into Eq. (14) yields

$$0 = c_0 - \frac{\beta^m \delta_k Q}{\theta_k A_k V_k^2} \quad \Rightarrow \quad V^* = \sqrt{\frac{\beta^m \delta_k Q}{c_0 \theta_k A_k}}. \tag{E.12}$$

Let $B_0 = \sqrt{\frac{\beta^m \delta_k}{c_0 \theta_k A_k}}$, then we have $V^* = B_0\sqrt{Q}$ and the optimal fleet size is

$$N^* = B_0\sqrt{Q} + \left(\frac{\rho_k R_k \delta}{2} + \tau_k\right) Q, \tag{E.13}$$

which further yields the system cost as

$$TC^* = \left(c_0 B_0 + \frac{\beta^m \delta_k}{\theta_k A_k B_0}\right) Q^{\frac{1}{2}} + \left[\frac{\beta^m \delta_k}{2} + \frac{\rho_k R_k \delta}{2}(c_0+\beta^p) + 2\beta^a w_k^a + (c_0+\beta^t)\tau_k\right] Q \tag{E.14}$$
$$= 2\sqrt{\frac{c_0 \beta^m \delta_k}{\theta_k A_k}} Q^{\frac{1}{2}} + \left[\frac{\beta^m \delta_k}{2} + \frac{\rho_k R_k \delta}{2}(c_0+\beta^p) + 2\beta^a w_k^a + (c_0+\beta^t)\tau_k\right] Q$$

The average and marginal costs are then given by

$$AC^* = \left(2\sqrt{\frac{c_0 \beta^m \delta_k}{\theta_k A_k}}\right) Q^{-\frac{1}{2}} + \left[\frac{\beta^m \delta_k}{2} + \frac{\rho_k R_k \delta}{2}(c_0+\beta^p) + 2\beta^a w_k^a + (c_0+\beta^t)\tau_k\right] \tag{E.15}$$

$$MC^* = \left(\sqrt{\frac{c_0 \beta^m \delta_k}{\theta_k A_k}}\right) Q^{-\frac{1}{2}} + \left[\frac{\beta^m \delta_k}{2} + \frac{\rho_k R_k \delta}{2}(c_0+\beta^p) + 2\beta^a w_k^a + (c_0+\beta^t)\tau_k\right] \tag{E.16}$$

Note that Eqs. (E.15) and (E.16) can be expressed as

$$AC^* = 2B_1 Q^{-\frac{1}{2}} + B_2, \tag{E.17}$$

$$MC^* = B_1 Q^{-\frac{1}{2}} + B_2, \tag{E.18}$$

where $B_1 = \sqrt{\frac{c_0 \beta^m \delta_k}{\theta_k A_k}}$ and $B_2 = \frac{\beta^m \delta_k}{2} + \frac{\rho_k R_k \delta}{2}(c_0+\beta^p) + 2\beta^a w_k^a + (c_0+\beta^t)\tau_k$ are both positive and only depend on model parameters. Accordingly, we have

$$DSE = \frac{AC^*}{MC^*} = \frac{2B_1 + B_2 Q^{\frac{1}{2}}}{B_1 + B_2 Q^{\frac{1}{2}}} \tag{E.19}$$

which falls in the range of (1,2] with the maximum value of 2 at $Q = 0$ and decreases with Q. Note that all the costs associated with the endogenous matching process are included in the factor B_1. It thus concludes that

$$DSE^{\text{en}} = \frac{2B_1 Q^{-\frac{1}{2}}}{B_1 Q^{-\frac{1}{2}}} = 2. \tag{E.20}$$

E.5. *Proposition 4*

In e-hailing without passenger competition, the matching and pickup times are still independent of W while the pickup time becomes a function of V. Their partial derivatives in this case are

$$f_k^m = \left(\frac{1}{A_k V} + \frac{1}{2}\right)\delta_k \quad \Rightarrow \quad \begin{cases} \partial_V f_k^m = -\frac{\delta_k}{A_k V^2} \\ \partial_W f_k^m = 0 \end{cases} \tag{E.21}$$

$$f_k^p = \frac{c_1 \rho_k \delta}{\sqrt{V}} \quad \Rightarrow \quad \begin{cases} \partial_V f_k^p = -\frac{c_1 \rho_k \delta}{2V^{\frac{3}{2}}} \\ \partial_W f_k^p = 0 \end{cases} \tag{E.22}$$

To derive $\partial_N h_k^V$, we take derivative with respect to N on both sides of Eq. (9b), which yields

$$1 = \partial_N h_k^V + Q\left(-\frac{c_1 \rho_k \delta}{2V^{\frac{3}{2}}}\right)\partial_N h_k^V, \quad \Rightarrow \quad \partial_N h_k^V = \frac{V^{\frac{3}{2}}}{V^{\frac{3}{2}} - \frac{c_1 \rho_k \delta Q}{2}}. \tag{E.23}$$

Plugging Eqs. (E.21), (E.22) and (E.23) into Eq. (14) yields

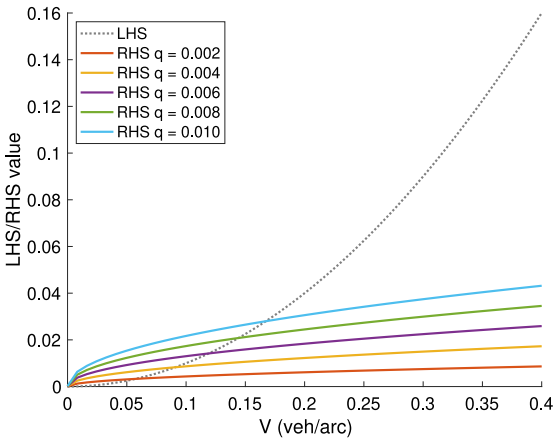
$$0 = c_0 + Q\left[\beta^m\left(-\frac{\delta_k}{A_k V^2} + 0\right) + \beta^p\left(-\frac{c_1 \rho_k \delta}{2V^{\frac{3}{2}}} + 0\right)\right]\frac{V^{\frac{3}{2}}}{V^{\frac{3}{2}} - \frac{c_1 \rho_k \delta Q}{2}} \tag{E.24}$$

$$\Rightarrow \; V^2 = \left[\frac{\beta_m \delta_k}{c_0 A_k} + \frac{c_1 \rho_k \delta}{2}\left(1 + \frac{\beta^p}{c_0}\right)\sqrt{V}\right] Q \tag{E.25}$$

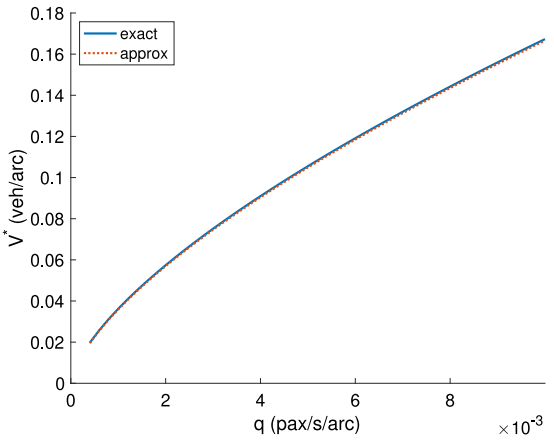
V^* can be solved numerically by searching the intersection of both sides of Eq. (E.25) at different demand levels, as illustrated in Fig. E.4(a). Further, as the first term in the right-hand-side (RHS) of Eq. (E.25) is close to zero due to the large value of A_k ($\gg 100$ arc), we can drop it and introduce the approximation

$$V^* \approx \left[\frac{c_1 \rho_k \delta}{2}\left(1 + \frac{\beta^p}{c_0}\right) Q\right]^{\frac{2}{3}} = (B_0 Q)^{\frac{2}{3}}, \tag{E.26}$$

where $B_0 = \frac{c_1 \rho_k \delta}{2}\left(1 + \frac{\beta^p}{c_0}\right)$. The approximation performance is demonstrated in Fig. E.4(b). It can be seen that the approximation in general underestimates V^* but the error can be safely ignored.



(a) numerical solution to V^*



(b) exact and approximated V^*

Fig. E.4. V^* in DE without passenger competition (with default parameter values in Table A.1).

We then derive the matching and pickup times as

$$w_k^{m*} = \left[\frac{(B_0 Q)^{-\frac{2}{3}}}{A_k} + \frac{1}{2}\right]\delta_k, \tag{E.27}$$

$$w_k^{p*} = c_1\rho_k\delta(V^*)^{-\frac{1}{2}} = c_1\rho_k\delta(B_0 Q)^{-\frac{1}{3}}, \tag{E.28}$$

Plugging them into the fleet size formula gives

$$N^* = B_0^{\frac{2}{3}}\left(1 + \frac{c_1\rho_k\delta}{B_0}\right)Q^{\frac{2}{3}} + \tau_k Q. \tag{E.29}$$

It finally gives the system cost as

$$\begin{aligned} TC^* &= B_0^{\frac{2}{3}}\left[c_0 + \frac{c_1\rho_k\delta}{B_0}(c_0+\beta^p)\right]Q^{\frac{2}{3}} + \frac{\beta^m\delta_k}{A_k}B_0^{-\frac{2}{3}}Q^{\frac{1}{3}} + \left[\frac{\beta^m\delta_k}{2} + 2\beta^a w_k^a + (c_0+\beta^t)\tau_k\right]Q, \\ &= 3c_0\left[\frac{c_1\rho_k\delta}{2}\left(1+\frac{\beta^p}{c_0}\right)\right]^{\frac{2}{3}}Q^{\frac{2}{3}} + \frac{\beta^m\delta_k}{A_k}\left[\frac{c_1\rho_k\delta}{2}\left(1+\frac{\beta^p}{c_0}\right)\right]^{-\frac{2}{3}}Q^{\frac{1}{3}} + \left[\frac{\beta^m\delta_k}{2} + 2\beta^a w_k^a + (c_0+\beta^t)\tau_k\right]Q. \end{aligned} \tag{E.30}$$

Note that the first two terms in Eq. (E.30) have a lower bound $2\sqrt{\frac{3c_0\beta^m\delta_k}{A_k}}Q^{\frac{1}{2}}$ due to Cauchy inequality, which shares the similar form as street-hailing as per Eq. (E.14). It thus implies that, without passenger competition, the matching process in e-hailing potentially has DSE equal to 2. However, this only happens when A_k is small, i.e., the matching radius is small. In the more general case with large A_k, the second term in Eq. (E.30) is close to zero and thus the system cost can be approximated as

$$TC^* \approx 3c_0\left[\frac{c_1\rho_k\delta}{2}\left(1+\frac{\beta^p}{c_0}\right)\right]^{\frac{2}{3}}Q^{\frac{2}{3}} + \left[\frac{\beta^m\delta_k}{2} + 2\beta^a w_k^a + (c_0+\beta^t)\tau_k\right]Q. \tag{E.31}$$

Accordingly, the average and marginal costs are given by

$$AC^* = 3c_0\left[\frac{c_1\rho_k\delta}{2}\left(1+\frac{\beta^p}{c_0}\right)\right]^{\frac{2}{3}}Q^{-\frac{1}{3}} + \left[\frac{\beta^m\delta_k}{2} + 2\beta^a w_k^a + (c_0+\beta^t)\tau_k\right], \tag{E.32}$$

$$MC^* = 2c_0\left[\frac{c_1\rho_k\delta}{2}\left(1+\frac{\beta^p}{c_0}\right)\right]^{\frac{2}{3}}Q^{-\frac{1}{3}} + \left[\frac{\beta^m\delta_k}{2} + 2\beta^a w_k^a + (c_0+\beta^t)\tau_k\right]. \tag{E.33}$$

These can be similarly expressed as

$$AC^* = 3B_1 Q^{-\frac{1}{3}} + B_2, \tag{E.34}$$

$$MC^* = 2B_1 Q^{-\frac{1}{3}} + B_2, \tag{E.35}$$

where $B_1 = c_0\left[\frac{c_1\rho_k\delta}{2}\left(1+\frac{\beta^p}{c_0}\right)\right]^{\frac{2}{3}}$ and $B_2 = \frac{\beta^m\delta_k}{2} + 2\beta^a w_k^a + (c_0+\beta^t)\tau_k$. It then yields

$$DSE = \frac{AC^*}{MC^*} = \frac{3B_1 + B_2 Q^{\frac{1}{3}}}{2B_1 + B_2 Q^{\frac{1}{3}}}. \tag{E.36}$$

The conclusions are drawn similarly to Proposition 3.

E.6. Proposition 5

In street-hailing with passenger competition, the matching time depends on both V and W, while the pickup time only depends on W. The partial derivatives are derived as

$$f_k^m = \left(\frac{W}{\theta_k V} + \frac{1}{2}\right)\delta_k \quad \Rightarrow \quad \begin{cases} \partial_V f_k^m = -\frac{\delta_k W}{\theta_k V^2}, \\ \partial_W f_k^m = \frac{\delta_k}{\theta_k V}, \end{cases} \tag{E.37}$$

$$f_k^p = \frac{\rho_k R_k \delta}{2W A_k} \quad \Rightarrow \quad \begin{cases} \partial_V f_p^m = 0, \\ \partial_W f_p^m = -\frac{\rho_k R_k \delta}{2A_k W^2}. \end{cases} \tag{E.38}$$

From Eq. (9c), we can solve $W = \frac{\theta_k \delta_k Q V}{2(\theta_k V - \delta_k Q)}$, which yields

$$\partial_V h_k^W = -\frac{1}{2\theta_k}\left(\frac{\theta_k \delta_k Q}{\theta_k V - \delta_k Q}\right)^2 = -\frac{2}{\theta_k}\left(\frac{W}{V}\right)^2. \tag{E.39}$$

We then take derivatives with respect to N on both sides of Eq. (9b), which yields

$$1 = \partial_N h_k^V + Q\left(-\frac{\rho_k R_k \delta}{2A_k W^2}\right)\left(-\frac{2W^2}{\theta_k V^2}\right)\partial_N h_k^V, \tag{E.40}$$

$$\Rightarrow \ \partial_N h_k^V = \frac{V^2}{V^2 + \frac{\rho_k R_k \delta}{\theta_k A_k} Q}. \tag{E.41}$$

Plugging all above equations into Eq. (14) yields

$$\begin{aligned} 0 = c_0 + Q\Bigg\{ & \beta^m \left[\left(-\frac{\delta_k W}{V^2}\right) + \left(\frac{\delta_k}{\theta_k V}\right)\left(-\frac{2W^2}{\theta_k V^2}\right)\right] \\ & + \beta^p \left[0 + \left(-\frac{\rho_k R_k \delta}{2A_k W^2}\right)\left(-\frac{2W^2}{\theta_k V^2}\right)\right]\Bigg\} \frac{V^2}{V^2 + \frac{\rho_k R_k \delta}{\theta_k A_k} Q} \end{aligned} \tag{E.42}$$

$$\Rightarrow \ V^2 = \left[\frac{\beta^m \delta_k}{c_0 \theta_k}\left(1 + \frac{2W}{\theta_k V}\right) W - \frac{\rho_k R_k \delta}{\theta_k A_k}\left(1 + \frac{\beta^p}{c_0}\right)\right] Q. \tag{E.43}$$

Same as the cases without competition, V^* can be solved numerically. Further, as shown in Fig. E.5(a), the first term in the parentheses on RHS of Eq. (E.66) increases with Q while the second remains constant. To derive a close-formed solution to V^*, we drop the second term and plug in the ratio $\frac{W}{V} = \frac{\delta_k Q}{2(\theta_k V - \delta_k Q)}$. This finally yields

$$V^* \approx \frac{\delta_k}{\theta_k}\left(1 + \sqrt{\frac{\beta^m \theta_k}{2c_0}}\right) Q \tag{E.44}$$

As shown in Fig. E.5(b), the approximation is not perfect with a stable overestimate around 0.02 veh/arc. Nevertheless, it well captures the linear relationship between Q and V^*, which is fundamentally different from the case without passenger competition, i.e., $V^* \propto \sqrt{Q}$ as per Eq. (E.12).

Let $B_0 = \sqrt{\frac{\beta^m \theta_k}{2c_0}}$, then the waiting passenger density can be represented as

$$W^* = \frac{\delta_k}{2}\left(1 + \frac{1}{B_0}\right) Q, \quad \frac{W^*}{V^*} = \frac{\theta_k}{2B_0}. \tag{E.45}$$

We then derive the matching and pickup times as

$$w_k^{m*} = \frac{\delta_k}{2}\left(1 + \frac{1}{B_0}\right), \tag{E.46}$$

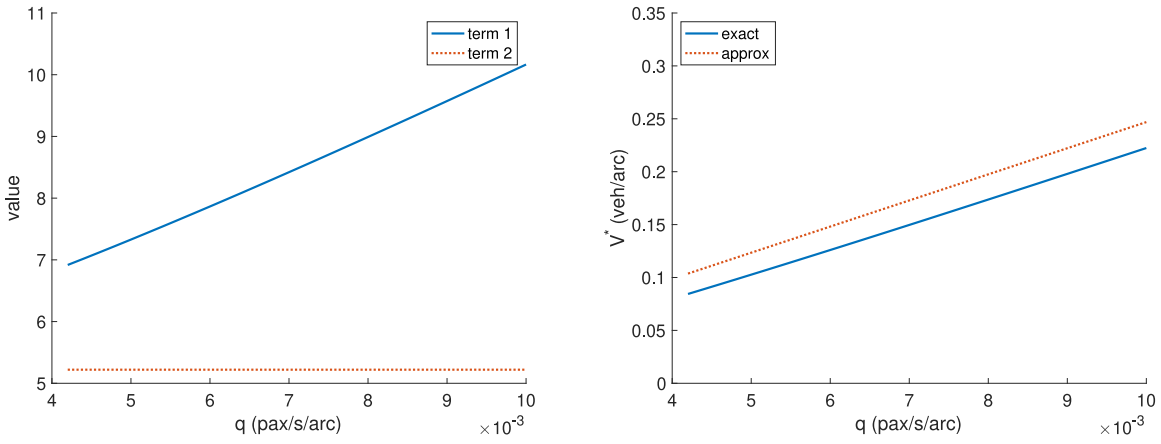
$$w_k^{p*} = \frac{\rho_k R_k \delta}{\delta_k A_k}\left[\left(1 + \frac{1}{B_0}\right) Q\right]^{-1}. \tag{E.47}$$

These yield the optimal fleet size

$$N^* = \frac{\rho_k R_k \delta}{\delta_k A_k}\left(1 + \frac{1}{B_0}\right)^{-1} + \left[\frac{\delta_k}{\theta_k}(1 + B_0) + \tau_k\right] Q, \tag{E.48}$$

and the system cost

$$TC^* = \frac{\rho_k R_k \delta}{\delta_k A_k}(c_0 + \beta^p)\left(1 + \frac{1}{B_0}\right)^{-1} + \left[\frac{c_0 \delta_k}{\theta_k}(1 + B_0) + \frac{\beta^m \delta_k}{2B_0} + \frac{\beta^m \delta_k}{2} + 2\beta^a w_k^a + (c_0 + \beta^t)\tau_k\right] Q, \tag{E.49}$$



(a) two terms in RHS of Eq. (E.43) (b) exact and approximated V^*

Fig. E.5. V^* in DS with passenger competition (with default parameter values in Table A.1).

$$= \frac{\rho_k R_k \delta}{\delta_k A_k}(c_0+\beta^p)\left(1+\sqrt{\frac{2c_0}{\beta^m\theta_k}}\right)^{-1} + \left[\frac{c_0\delta_k}{\theta_k}\left(1+\sqrt{\frac{2\beta^m\theta_k}{c_0}}\right)+\frac{\beta^m\delta_k}{2}+2\beta^a w_k^a+(c_0+\beta^t)\tau_k\right]Q.$$

The average and marginal costs in this case are

$$AC^* = \frac{\rho_k R_k \delta}{\delta_k A_k}(c_0+\beta^p)\left(1+\sqrt{\frac{2c_0}{\beta^m\theta_k}}\right)^{-1} Q^{-1} + \left[\frac{c_0\delta_k}{\theta_k}\left(1+\sqrt{\frac{2\beta^m\theta_k}{c_0}}\right)+\frac{\beta^m\delta_k}{2}+2\beta^a w_k^a+(c_0+\beta^t)\tau_k\right], \tag{E.50}$$

$$MC^* = \frac{c_0\delta_k}{\theta_k}\left(1+\sqrt{\frac{2\beta^m\theta_k}{c_0}}\right)+\frac{\beta^m\delta_k}{2}+2\beta^a w_k^a+(c_0+\beta^t)\tau_k,. \tag{E.51}$$

The simplified expression is

$$AC^* = B_1 Q^{-1} + B_2 + B_3, \tag{E.52}$$

$$MC^* = B_2 + B_3, \tag{E.53}$$

where $B_1 = \frac{\rho_k R_k \delta}{\delta_k A_k}(c_0+\beta^p)\left(1+\sqrt{\frac{2c_0}{\beta^m\theta_k}}\right)^{-1}$, $B_2 = \frac{\beta^m\delta_k}{2}+2\beta^a w_k^a+(c_0+\beta^t)\tau_k$, and $B_3 = \frac{c_0\delta_k}{\theta_k}\left(1+\sqrt{\frac{2\beta^m\theta_k}{c_0}}\right)$.

We then obtain the system-level DSE as

$$DSE = \frac{AC^*}{MC^*} = \frac{B_1}{(B_2+B_3)Q}+1, \tag{E.54}$$

and the DSE in the endogenous matching process is given by

$$DSE^{\text{en}} = \frac{B_1 Q^{-1}+B_3}{B_3} = \frac{B_1}{B_3 Q}+1, \tag{E.55}$$

due to the observation that all the associated costs are included in B_1 and B_3.

Both DSE monotonically decrease with Q and converge to 1, though their values are unbounded as per the formulas. Nevertheless, we note that the demand rate Q has some lower bound to ensure the passenger competition indeed exists. Specifically, we have

$$\gamma_k = W^* A_k = \frac{\delta_k A_k}{2}\left(1+\frac{1}{B_0}\right)Q > 1 \tag{E.56}$$

$$\Rightarrow \quad Q > \frac{2}{\delta_k A_k}\left(1+\frac{1}{B_0}\right)^{-1} = \frac{2}{\delta_k A_k}\left[1+\sqrt{\frac{2c_0}{\beta^m\theta_k}}\right]^{-1}. \tag{E.57}$$

Therefore, the DSE upper bounds are derived as

$$DSE < 1+\frac{B_1}{B_2+B_3}\left[\frac{\delta_k A_k}{2}\left(1+\sqrt{\frac{2c_0}{\beta^m\theta_k}}\right)\right] = 1+\frac{\rho_k R_k \delta(c_0+\beta^p)}{2(B_2+B_3)}, \tag{E.58}$$

$$DSE^{\text{en}} < 1 + \frac{B_1}{B_3}\left[\frac{\delta_k A_k}{2}\left(1+\sqrt{\frac{2c_0}{\beta^m\theta_k}}\right)\right] = 1 + \frac{\rho_k R_k \delta(c_0+\beta^p)}{2B_3}. \tag{E.59}$$

E.7. Proposition 6

In e-hailing with passenger competition, the matching and pickup times depend both on V and W, and the partial derivatives are given by

$$f_k^m = \left(\frac{W}{V}+\frac{1}{2}\right)\delta_k \quad \Rightarrow \quad \begin{cases} \partial_V f_k^m = -\frac{\delta_k W}{V^2}, \\ \partial_W f_k^m = \frac{\delta_k}{V}, \end{cases} \tag{E.60}$$

$$f_k^p = \frac{c_1\rho_k\delta}{\sqrt{V}}\left(1+\frac{W}{c_2V}\right)^{-1} \quad \Rightarrow \quad \begin{cases} \partial_V f_p^m = -\frac{c_1\rho_k\delta\varphi}{2V^{\frac{3}{2}}}\left(1-\frac{2\varphi W}{c_2V}\right), \\ \partial_W f_p^m = -\frac{c_1\rho_k\delta\varphi^2}{c_2V^{\frac{3}{2}}}, \end{cases} \tag{E.61}$$

where $\varphi = \left(1+\frac{W}{c_2V}\right)^{-1}$.

Similarly, we can solve W from Eq. (9c) as $W = \frac{\delta_k QV}{2(V-\delta_k Q)}$, which yields

$$\partial_V h_k^W = -\frac{1}{2}\left(\frac{\delta_k Q}{V-\delta_k Q}\right)^2 = -2\left(\frac{W}{V}\right)^2. \tag{E.62}$$

Again, we take derivatives with respect to N on both sides of Eq. (9b), which yields

$$1 = \partial_N h_k^V + Q\left\{\left[-\frac{c_1\rho_k\delta\varphi}{2V^{\frac{3}{2}}}\left(1-\frac{2\varphi W}{c_2V}\right)\right] + \left(-\frac{c_1\rho_k\delta\varphi^2}{c_2V^{\frac{3}{2}}}\right)\left[-2\left(\frac{W}{V}\right)^2\right]\right\}\partial_N h_k^V, \tag{E.63}$$

$$\Rightarrow \ \partial_N h_k^V = \frac{V^{\frac{3}{2}}}{V^{\frac{3}{2}} - \frac{c_1\rho_k\delta\varphi Q}{2}\left[1-\frac{2\varphi W}{c_2V}\left(1+\frac{2W}{V}\right)\right]} \tag{E.64}$$

Plugging all above equations into Eq. (14) yields

$$0 = c_0 + Q\left\{\beta^m\left\{\left(-\frac{\delta_k W}{V^2}\right)+\left[-\frac{2\delta_k}{V}\left(\frac{W}{V}\right)^2\right]\right\} + \beta^p\left\{\left[-\frac{c_1\rho_k\delta\varphi}{2V^{\frac{3}{2}}}\left(1-\frac{2\varphi W}{c_2V}\right)\right]\right.\right. \tag{E.65}$$

$$\left.\left. + \left[\frac{2c_1\rho_k\delta\varphi^2}{c_2V^{\frac{3}{2}}}\left(\frac{W}{V}\right)^2\right]\right\}\right\}\frac{V^{\frac{3}{2}}}{V^{\frac{3}{2}} - \frac{c_1\rho_k\delta\varphi Q}{2}\left[1-\frac{2\varphi W}{c_2V}\left(1+\frac{2W}{V}\right)\right]}$$

$$\Rightarrow \ V = \left\{\frac{\beta^m\delta_k}{c_0}\left(\frac{W}{V}\right)\left(1+\frac{2W}{V}\right) + \left(1+\frac{\beta^p}{c_0}\right)\frac{c_1\rho_k\delta\varphi}{2\sqrt{V}}\left[1-\frac{2\varphi}{c_2}\left(\frac{W}{V}\right)\left(1+\frac{2W}{V}\right)\right]\right\}Q. \tag{E.66}$$

Following the same procedure, we first solve V^* numerically and then plot the two terms in the parentheses on RHS of Eq. (E.66) in Fig. E.6(a). In this case, both terms converge to constants when Q is relatively large and the second is much closer to zero. Like the case of street-hailing, we drop the second term and derive the following approximation by exploiting the ratio $\frac{W}{V} = \frac{\delta_k Q}{2(V-\delta_k Q)}$:

$$V^* \approx \delta_k\left(1+\sqrt{\frac{\beta^m}{2c_0}}\right)Q. \tag{E.67}$$

The approximation performance is illustrated in Fig. E.6(b). Similar to the case without passenger competition, our approximation tends to underestimate V^* but the gap is almost negligible.

Note that V^* above shares a similar form as per Eq. (E.44). Following the same procedure, we let $B_0 = \sqrt{\frac{\beta^m}{2c_0}}$, then the waiting passenger density and its ratio to the vacant vehicle density are obtained as

$$W^* = \frac{1}{2}\left(1+\frac{1}{B_0}\right)\delta_k Q, \quad \frac{W^*}{V^*} = \frac{1}{2B_0}. \tag{E.68}$$

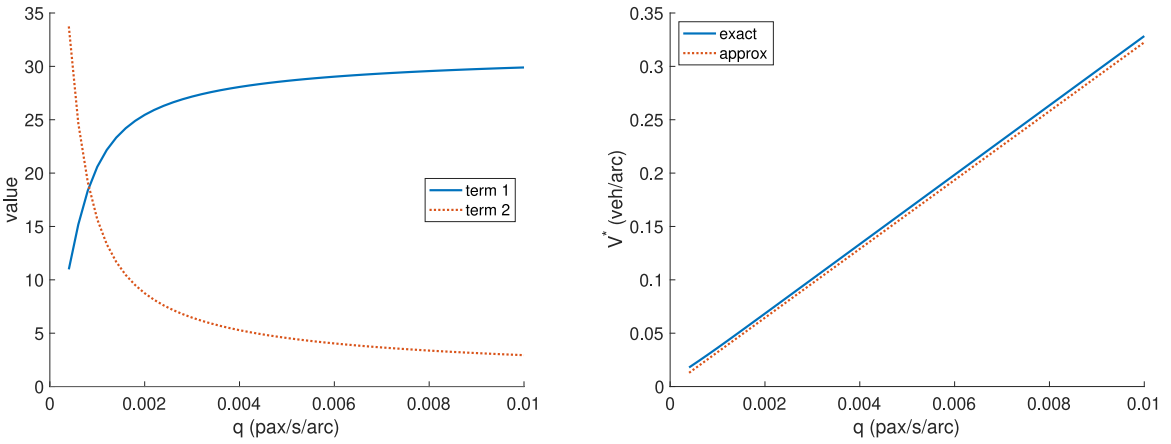
Furthermore, the matching and pickup times are obtained as

$$w_k^{m*} = \frac{\delta_k}{2}\left(1+\frac{1}{B_0}\right), \tag{E.69}$$

$$w_k^{p*} = c_1\rho_k\delta\varphi^*\left[\delta_k(1+B_0)Q\right]^{-\frac{1}{2}}, \tag{E.70}$$

where $\varphi^* = \left(1+\frac{1}{2c_1B_0}\right)^{-1}$. We then derive the optimal fleet size

$$N^* = c_1\rho_k\delta\varphi^*\left[\delta_k(1+B_0)\right]^{-\frac{1}{2}}Q^{\frac{1}{2}} + \left[\delta_k(1+B_0)+\tau_k\right]Q, \tag{E.71}$$



(a) two terms in RHS of Eq. (E.66) (b) exact and approximated V^*

Fig. E.6. V^* in DE with passenger competition (with default parameter values in Table A.1).

and the system cost

$$\begin{aligned} TC^* &= c_1\rho_k\delta\varphi^*(c_0+\beta^p)\left[\delta_k(1+B_0)\right]^{-\frac{1}{2}}Q^{\frac{1}{2}} \\ &\quad+\left[\left(c_0\delta_k(1+B_0)+\frac{\beta^m\delta_k}{2B_0}\right)+\left(\frac{\beta^m\delta_k}{2}+2\beta^a w_k^a+(c_0+\beta^t)\tau_k\right)\right]Q \\ &= c_1\rho_k\delta\varphi^*(c_0+\beta^p)\left[\delta_k\left(1+\sqrt{\frac{\beta^m}{2c_0}}\right)\right]^{-\frac{1}{2}}Q^{\frac{1}{2}}+\left[c_0\delta_k\left(1+\sqrt{\frac{2\beta^m}{c_0}}\right)+\frac{\beta^m\delta_k}{2}+2\beta^a w_k^a+(c_0+\beta^t)\tau_k\right]Q. \end{aligned} \tag{E.72}$$

Accordingly, the average and marginal costs are

$$AC^* = c_1\rho_k\delta\varphi^*(c_0+\beta^p)\left[\delta_k\left(1+\sqrt{\frac{\beta^m}{2c_0}}\right)\right]^{-\frac{1}{2}}Q^{-\frac{1}{2}}+\left[c_0\delta_k\left(1+\sqrt{\frac{2\beta^m}{c_0}}\right)+\frac{\beta^m\delta_k}{2}+2\beta^a w_k^a+(c_0+\beta^t)\tau_k\right], \tag{E.73}$$

$$MC^* = \frac{c_1\rho_k\delta\varphi^*}{2}(c_0+\beta^p)\left[\delta_k\left(1+\sqrt{\frac{\beta^m}{2c_0}}\right)\right]^{-\frac{1}{2}}Q^{-\frac{1}{2}}+\left[c_0\delta_k\left(1+\sqrt{\frac{2\beta^m}{c_0}}\right)+\frac{\beta^m\delta_k}{2}+2\beta^a w_k^a+(c_0+\beta^t)\tau_k\right]. \tag{E.74}$$

The simplified expression is

$$AC^* = 2B_1Q^{-\frac{1}{2}}+B_2+B_3, \tag{E.75}$$

$$MC^* = B_1Q^{-\frac{1}{2}}+B_2+B_3, \tag{E.76}$$

where $B_1=\frac{c_1\rho_k\delta\varphi^*}{2}(c_0+\beta^p)\left[\delta_k\left(1+\sqrt{\frac{\beta^m}{2c_0}}\right)\right]^{-\frac{1}{2}}$, $B_2=\frac{\beta^m\delta_k}{2}+2\beta^a w_k^a+(c_0+\beta^t)\tau_k$, and $B_3=c_0\delta_k\left(1+\sqrt{\frac{2\beta^m}{c_0}}\right)$.

This finally yields the system-level DSE

$$DSE=\frac{AC^*}{MC^*}=\frac{2B_1+(B_2+B_3)Q^{\frac{1}{2}}}{B_1+(B_2+B_3)Q^{\frac{1}{2}}}, \tag{E.77}$$

and the endogenous DSE

$$DSE^{\text{en}}=\frac{2B_1+B_3Q^{\frac{1}{2}}}{B_1+B_3Q^{\frac{1}{2}}}. \tag{E.78}$$

Different from street-hailing, there is an explicit upper bound for both DSE as 2, though we can also derive a tighter upper bound using the minimum demand that induces passenger competition:

$$\gamma_k = W^*A_k = 2\delta_k R_k^2\left(1+\frac{1}{B_0}\right)Q>1 \tag{E.79}$$

$$\Rightarrow\quad Q>\frac{1}{2\delta_k R_k^2}\left(1+\frac{1}{B_0}\right)^{-1}=\frac{1}{2\delta_k R_k^2}\left(1+\sqrt{\frac{2c_0}{\beta^m}}\right)^{-1}, \tag{E.80}$$

which yields

$$DSE < \frac{2B_1 + (B_2+B_3)\left[2\delta_k R_k^2\left(1+\sqrt{\frac{2c_0}{\beta^m}}\right)\right]^{-\frac{1}{2}}}{B_1 + (B_2+B_3)\left[2\delta_k R_k^2\left(1+\sqrt{\frac{2c_0}{\beta^m}}\right)\right]^{-\frac{1}{2}}} = 1 + \frac{c_1\rho_k R_k\delta\varphi^*(c_0+\beta^p)}{c_1\rho_k R_k\delta\varphi^*(c_0+\beta^p)+\sqrt{2}(B_2+B_3)}, \tag{E.81}$$

$$DSE^{\text{en}} < \frac{2B_1 + B_3\left[2\delta_k R_k^2\left(1+\sqrt{\frac{2c_0}{\beta^m}}\right)\right]^{-\frac{1}{2}}}{B_1 + B_3\left[2\delta_k R_k^2\left(1+\sqrt{\frac{2c_0}{\beta^m}}\right)\right]^{-\frac{1}{2}}} = 1 + \frac{c_1\rho_k R_k\delta\varphi^*(c_0+\beta^p)}{c_1\rho_k R_k\delta\varphi^*(c_0+\beta^p)+\sqrt{2}B_3}. \tag{E.82}$$

E.8. Proposition 7

Recall that walking directly affects the parameter values in the following way:

- Increase vacant vehicle density factor θ_k in street-hailing if $L(K-1) > K$, where L is the number of local streets between each pair of major streets and K is the number of major streets along one side of the service region. See Appendix B.2 for more details.
- Reduce the speed factor ρ_k; see Eqs. (3) and (4).
- Increase walking time w_a; see Eq. (5).
- Reduce average in-vehicle time τ_k; see Eq. (6).

Recall that the DSE in street-hailing without passenger competition is given by

$$DSE = \frac{2B_1 + B_2 Q^{\frac{1}{2}}}{B_1 + B_2 Q^{\frac{1}{2}}}, \tag{E.83}$$

where $B_1 = \sqrt{\frac{c_0\beta^m\delta_k}{\theta_k A_k}}$ and $B_2 = \frac{\beta^m\delta_k}{2} + \frac{(c_0+\beta^p)\rho_k}{2}R_k\delta + 2\beta^a w_k^a + (c_0+\beta^t)\tau_k$.

Accordingly, a larger B_1 leads to a larger DSE while a larger B_2 leads to a lower DSE. Note that θ_k only appears in B_1 and a larger θ_k yields a smaller B_1. As the shift from DS to WS increases θ_k, it thus concludes that the increased vacant vehicle density due to walking compromises the DSE in street-hailing. On the other hand, ρ_k, w_a, τ_k only appear in B_2 and an increase in their values all leads to a larger B_2. Since walking induces a smaller ρ_k, a positive w_a, and a reduced τ_k, their corresponding impacts are as reported in Table 1.

For e-hailing without passenger competition, we have

$$DSE = \frac{3B_1 + B_2 Q^{\frac{1}{3}}}{2B_1 + B_2 Q^{\frac{1}{3}}}, \tag{E.84}$$

where $B_1 = c_0\left[\left(1+\frac{\beta^p}{c_0}\right)\frac{c_1\rho_k\delta}{2}\right]^{\frac{2}{3}}$ and $B_2 = \frac{\beta^m\delta_k}{2} + 2\beta^a w_k^a + (c_0+\beta^t)\tau_k$. In this case, the impacts of ρ_k, w_a, τ_k are consistent with the case of street-hailing.

Finally, e-hailing with passenger competition yields

$$DSE = \frac{2B_1 + (B_2+B_3)Q^{\frac{1}{2}}}{B_1 + (B_2+B_3)Q^{\frac{1}{2}}}, \tag{E.85}$$

where $B_1 = \frac{c_1\rho_k\delta\varphi^*}{2}(c_0+\beta^p)\left[\delta_k\left(1+\sqrt{\frac{\beta^m}{2c_0}}\right)\right]^{-\frac{1}{2}}$, $B_2 = \frac{\beta^m\delta_k}{2} + 2\beta^a w_k^a + (c_0+\beta^t)\tau_k$, $B_3 = c_0\delta_k\left(1+\sqrt{\frac{2\beta^m}{c_0}}\right)$, and $\varphi^* = \left(1+\frac{1}{c_1}\sqrt{\frac{c_0}{2\beta^m}}\right)^{-1}$. It is easy to verify that the same impacts of the three factors on DSE apply in this case.

E.9. Proposition 8

Following the results derived in Appendix E.8, we first compute the change in B_2 in street-hailing due to walking:

$$B_{2,WS} - B_{2,DS} = -\frac{c_0+\beta^p}{2}(\alpha^l - 1)R_k\delta + 2D^a\delta\left[\beta^a\alpha^a - (c_0+\beta^t)\alpha^l\right]. \tag{E.86}$$

Plugging $\alpha^l = 1$ into Eq. (E.86), it thus yields $B_{2,WS} > B_{2,DS}$ when $\beta^a\alpha^a > c_0+\beta^t$. Additionally, as walking leads to a larger value of θ_{WS}, we have $B_{1,WS} < B_{1,DS}$ and it yields another negative scale effect. Therefore, the result in Proposition 8 for street-hailing holds.

The result for e-hailing can be proved in a similar way. Specifically, under the condition $\alpha^l = 1$ and $\beta^a\alpha^a > c_0+\beta^t$, one can easily show $B_{2,WE} > B_{2,DE}$ and $B_{1,WE} = B_{2,DE}$. It thus completes the proof.

Table F.1
Main features of the network.

Feature	Value	Unit
Av. Door-to-door trip	11.3	min
Av. Arc length	44	s
Av. Major arc length	40	s
Av. Local arc length	50.3	s
Av. Walking time in WS	5.3	min
Av. WS in-vehicle trip	9.5	min

Appendix F. Network in simulations

The whole Manhattan dataset is strongly unbalanced, with most origins and destinations occurring south of Central Park (Fielbaum et al., 2021). To obtain a setting that is closer to our analytical model, we crop the network to consider only the south of Central Park. The resulting network contains 1966 nodes and 4253 edges.

As for travel demand, we first select trips with both origin and destination in the cropped network. We also prune all requests with an in-vehicle distance between origin and destination shorter than 5 min. This finally gives a total of 14,213 requests. In the simulations, we random subsets from this pool of requests. To reduce the impact of randomness, each scenario is simulated 5 times and the average results are reported.

To simulate the WS scheme, we need to identify the connected subnetwork of the "major streets". Let us denote $\mathbb{G} = (\mathbb{V}, \mathbb{E})$ the original network, and $d(u_1, u_2)$ the length of the shortest walking path between any pair of nodes u_1, u_2. How to identify a continuous street within a city graph is a complex task per se (Lin and Ban, 2013) and beyond the scope of this paper. Instead, we use a simple method that consists of two steps:

1. Select a set of "major nodes" $\mathbb{M} \subseteq \mathbb{V}$. We first define distance threshold d_M (e.g., maximum walking distance), then sort all the edges in $\mathbb{E}$ in the descending order of travel speed and exclude those longer than d_M. Starting from $\mathbb{M} = \emptyset$, we sequentially take $e_i = (u_i, v_i)$ for $i = 1, \dots, \ell$ and compute the distance of both edge nodes $\mathbb{M}$. If there exist $w_1, w_2 \in \mathbb{M}$ such that $d(u_i, w_1) \leq d_M, d(v_i, w_2) \leq d_M$, we regard nodes u_i, v_i as already *covered*, i.e., they can be reached from some major nodes within the threshold distance. Otherwise, we add u_i in $\mathbb{M}$ and move to the next edge.
2. Define the set of major streets. An edge $e \in \mathbb{E}$ is labeled as a major street if there exist $u_1, u_2 \in \mathbb{M}$ such that the shortest path between u_1 and u_2 contains edge e.

In this study, we use the maximum walking time of 2 min to define the threshold distance d_M, which yields a subnetwork with 3991 arcs.

References

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

Arnott, R., 1996. Taxi travel should be subsidized. J. Urban Econ. 40 (3), 316–333.

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.

Basso, L.J., Jara-Díaz, S.R., 2006. Distinguishing multiproduct economies of scale from economies of density on a fixed-size transport network. Netw. Spat. Econ. 6, 149–162.

Besbes, O., Castro, F., Lobel, I., 2022. Spatial capacity planning. Oper. Res. 70 (2), 1271–1291.

Bimpikis, K., Candogan, O., Saban, D., 2019. Spatial pricing in ride-sharing networks. Oper. Res. 67 (3), 744–769.

Buchholz, N., 2022. Spatial equilibrium, search frictions, and dynamic efficiency in the taxi industry. Rev. Econ. Stud. 89 (2), 556–591.

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.

Cramer, J., Krueger, A.B., 2016. Disruptive change in the taxi business: The case of Uber. Amer. Econ. Rev. 106 (5), 177–182.

Douglas, G.W., 1972. Price regulation and optimal service standards: The taxicab industry. J. Transp. Econ. Policy 116–127.

Erhardt, G.D., Roy, S., Cooper, D., Sana, B., Chen, M., Castiglione, J., 2019. Do transportation network companies decrease or increase congestion? Sci. Adv. 5 (5), eaau2670.

Feng, G., Kong, G., Wang, Z., 2021. We are on the way: Analysis of on-demand ride-hailing systems. Manuf. Serv. Oper. Manag. 23 (5), 1237–1256.

Fielbaum, A., 2022. Optimizing a vehicle's route in an on-demand ridesharing system in which users might walk. J. Intell. Transp. Syst. 26 (4), 432–447.

Fielbaum, A., Bai, X., Alonso-Mora, J., 2021. On-demand ridesharing with optimized pick-up and drop-off walking locations. Transp. Res. C 126, 103061.

Fielbaum, A., Jara-Diaz, S., Gschwender, A., 2020. Beyond the Mohring effect: Scale economies induced by transit lines structures design. Econ. Transp. 22, 100163.

Fielbaum, A., Kronmueller, M., Alonso-Mora, J., 2022. Anticipatory routing methods for an on-demand ridepooling mobility system. Transportation 49 (6), 1921–1962.

Fielbaum, A., Pudāne, B., 2024. Are shared automated vehicles good for public-or private-transport-oriented cities (or neither)? Transp. Res. Part D: Transp. Environ. 136, 104373.

Fielbaum, A., Tirachini, A., Alonso-Mora, J., 2023. Economies and diseconomies of scale in on-demand ridepooling systems. Econ. Transp. 34, 100313.

Frechette, G.R., Lizzeri, A., Salz, T., 2019. Frictions in a competitive, regulated market: Evidence from taxis. Amer. Econ. Rev. 109 (8), 2954–2992.

Guo, X., Haupt, A., Wang, H., Qadri, R., Zhao, J., 2023. Understanding multi-homing and switching by platform drivers. Transp. Res. C 154, 104233.
Gurumurthy, K.M., Kockelman, K.M., 2022. Dynamic ride-sharing impacts of greater trip demand and aggregation at stops in shared autonomous vehicle systems. Transp. Res. Part A: Policy Pr. 160, 114–125.
He, F., Shen, Z.-J.M., 2015. Modeling taxi services with smartphone-based e-hailing applications. Transp. Res. C 58, 93–106.
Jansson, J.O., 1980. A simple bus line model for optimisation of service frequency and bus size. J. Transp. Econ. Policy 53–80.
Jara-Diaz, S.R., Cortes, C.E., 1996. On the calculation of scale economies from transport cost functions. J. Transp. Econ. Policy 157–170.
Jiao, G., Ramezani, M., 2024. A real-time cooperation mechanism in duopoly e-hailing markets. Transp. Res. C 162, 104598.
Joskow, P.L., 2007. Regulation of natural monopoly. Handb. Law Econ. 2, 1227–1348.
Kaddoura, I., Schlenther, T., 2021. The impact of trip density on the fleet size and pooling rate of ride-hailing services: A simulation study. Procedia Comput. Sci. 184, 674–679.
Lagos, R., 2000. An alternative approach to search frictions. J. Polit. Econ. 108 (5), 851–873.
Lehe, L., Gayah, V.V., Pandey, A., 2021. Increasing returns to scale in carpool matching: Evidence from Scoop. Transp. Find..
Li, X., Ke, J., Yang, H., Wang, H., Zhou, Y., 2024b. An aggregate matching and pick-up model for mobility-on-demand services. Transp. Res. B 190, 103070.
Li, R., Liu, Y., Liu, X., Nie, Y.M., 2024a. Allocation problem in cross-platform ride-hail integration. Transp. Res. B 188, 103056.
Lin, J., Ban, Y., 2013. Complex network topology of transportation systems. Transp. Rev. 33 (6), 658–685.
Liu, L., Andris, C., Ratti, C., 2010. Uncovering cabdrivers' behavior patterns from their digital traces. Comput. Environ. Urban Syst. 34 (6), 541–548.
Liu, H., Devunuri, S., Lehe, L., Gayah, V.V., 2023. Scale effects in ridesplitting: A case study of the City of Chicago. Transp. Res. Part A: Policy Pr. 173, 103690.
Mohring, H., 1972. Optimization and scale economies in urban bus transportation. Am. Econ. Rev. 62 (4), 591–604.
Mühle, S., 2023. An analytical framework for modeling ride pooling efficiency and minimum fleet size. Multimodal Transp. 2 (2), 100080.
Nie, Y.M., 2017. How can the taxi industry survive the tide of ridesourcing? Evidence from Shenzhen, China. Transp. Res. C 79, 242–256.
Schroeter, J.R., 1983. A model of taxi service under fare structure and fleet size regulation. Bell J. Econ. 81–96.
Simonetto, A., Monteil, J., Gambella, C., 2019. Real-time city-scale ridesharing via linear assignment problems. Transp. Res. C 101, 208–232.
Taylor, T.A., 2018. On-demand service platforms. Manuf. Serv. Oper. Manag. 20 (4), 704–720.
Tirachini, A., 2020. Ride-hailing, travel behaviour and sustainable mobility: an international review. Transportation 47 (4), 2011–2047.
Tone, K., Sahoo, B.K., 2004. Degree of scale economies and congestion: A unified DEA approach. European J. Oper. Res. 158 (3), 755–772.
Urata, J., Xu, Z., Ke, J., Yin, Y., Wu, G., Yang, H., Ye, J., 2021. Learning ride-sourcing drivers' customer-searching behavior: A dynamic discrete choice approach. Transp. Res. C 130, 103293.
Wang, X., He, F., Yang, H., Gao, H.O., 2016. Pricing strategies for a taxi-hailing platform. Transp. Res. Part E: Logist. Transp. Rev. 93, 212–231.
Wang, J., Wang, X., Yang, S., Yang, H., Zhang, X., Gao, Z., 2021. Predicting the matching probability and the expected ride/shared distance for each dynamic ridepooling order: A mathematical modeling approach. Transp. Res. B 154, 125–146.
Xu, Z., Yin, Y., Zha, L., 2017. Optimal parking provision for ride-sourcing services. Transp. Res. B 105, 559–578.
Yan, C., Zhu, H., Korolko, N., Woodard, D., 2020. Dynamic pricing and matching in ride-hailing platforms. Naval Res. Logist. 67 (8), 705–724.
Yang, H., Leung, C.W., Wong, S., Bell, M.G., 2010. Equilibria of bilateral taxi–customer searching and meeting on networks. Transp. Res. B 44 (8), 1067–1083.
Yang, H., Qin, X., Ke, J., Ye, J., 2020. Optimizing matching time interval and matching radius in on-demand ride-sourcing markets. Transp. Res. B 131, 84–105.
Yang, H., Yang, T., 2011. Equilibrium properties of taxi markets with search frictions. Transp. Res. B 45 (4), 696–713.
Yang, H., Ye, M., Tang, W.H., Wong, S.C., 2005. Regulating taxi services in the presence of congestion externality. Transp. Res. Part A: Policy Pr. 39 (1), 17–40.
Zha, L., Yin, Y., Yang, H., 2016. Economic analysis of ride-sourcing markets. Transp. Res. C 71, 249–266.
Zhang, K., Chen, H., Yao, S., Xu, L., Ge, J., Liu, X., Nie, M., 2019. An efficiency paradox of uberization. Available at SSRN 3462912.
Zhang, K., Mittal, A., Djavadian, S., Twumasi-Boakye, R., Nie, Y.M., 2023. RIde-hail vehicle routing (RIVER) as a congestion game. Transp. Res. B 177, 102819.
Zhang, K., Nie, Y.M., 2021. Inter-platform competition in a regulated ride-hail market with pooling. Transp. Res. Part E: Logist. Transp. Rev. 151, 102327.
Zhou, Y., Ke, J., Yang, H., Guo, P., 2024. Platform integration in ride-sourcing markets with heterogeneous passengers. Transp. Res. B 188, 103041.