Optimal Subscriptions for Ridesharing Platforms

Ben Berger* Hongyao Ma† David C. Parkes‡ Shreyas Sekar§

Abstract

In this paper, we seek to understand the use of subscription pricing in rideshaing platforms such as Uber and Lyft. For this, we adopt a model for the operation of two-sided markets, specifically ridesharing markets, under *network effects*— a higher spatial density of idle supply implies a lower pickup time for riders and a lower idle time between trips for the drivers. As a first result we demonstrate that contrary to folk wisdom, real time market-clearing prices do not maximize social welfare. This leads to inefficiencies when a platform is limited to real-time pricing, and yet the first-best can be achieved through the use of *subscription pricing* that combines a fixed fee with a real-time price. In particular, we show how subscription pricing can be used to maximize social welfare for markets with homogeneous riders as well as with multiple-rider types, each differing in their frequency of riding. We also provide a model-free implementation, such that a platform can design the optimal subscription mechanisms using only information immediately observable in the equilibrium of the real-time market.

1 Introduction

Ridesharing platforms such as Uber and Lyft use dynamic pricing to match rider demand with driver supply in real-time. It has been empirically established that such real-time pricing can improve the throughput of completed rides on the platform [Chen and Sheldon, 2015] and the service reliability [Hall et al., 2015]. At the same time, it is crucial that a ridesharing platform also consider *network effects* in ensuring efficient operations, as a higher spatial density of drivers reduces

*Blavatnic School of Computer Science, Tel Aviv University. Email: benberger1@tauex.tau.ac.il.

†Decision, Risk, and Operations Division, Columbia Business School. 1113 Kravis Hall, 655 W130 Street, New York, NY, 10027, USA. Email: hongyao.ma@columbia.edu.

‡John A. Paulson School of Engineering and Applied Sciences, Harvard University, 150 Western Ave, Boston, MA 02138, USA. Email: parkes@eecs.harvard.edu.

§Rotman School of Management, University of Toronto and Department of Management, University of Toronto, Scarborough. Email: shreyas.sekar@rotman.utoronto.ca

the wait time for riders before pickup. In the presence of network effects, it is known that real-time pricing can improve a platform's operational efficiency by maintaining appropriate levels of open driver supply and balancing drivers' idle time between trips versus their pickup time [Castillo et al., 2017, Yan et al., 2020]. However, our understanding of whether such mechanisms can implement socially optimal outcomes, also considering network effects, remains incomplete.

In this paper, we will be interested to understand the role of subscription pricing mechanisms in promoting better outcomes. Indeed, it is an empirical fact that subscription mechanisms have over the years also become popular among all major ridesharing platforms (see Figure 1). Typically, these subscription offers comprise of two parts: a) an *advance pricing* component involving a fixed fee paid before consumers realize their valuation for the service, and b) a *bundling* component where multiple services are sold as a single unit.

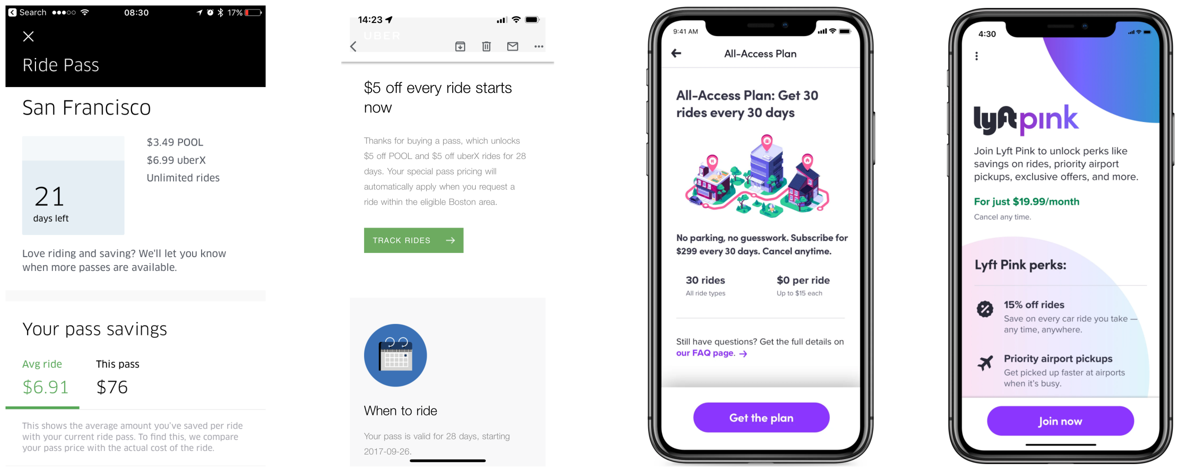
Practitioners designing such mechanisms are often faced with a multitude of levers that have to be simultaneously optimized. These include but are not limited to subscription fees, the type of discount (percentage or fixed), the number of trips bundled together, and the expiration time. As a result, ridesharing platforms have experimented across a number of subscription programs in the recent past. Although these programs all charge riders an advance subscription fee, the exact offer provided can vary wildly and may involve:

- *Flat Rates and Price Protection*[1]: Subscribers are offered specific rides at a flat rate, thereby shielding them from price surges (Figure 1a)
- *Fixed (Dollar) Discounts*: A fixed discount (e.g., $5) is applied on every ride (Figure 1b).
- *Percentage Discounts*: A fixed percentage discount (e.g., 10%) is offered on the real-time price of every ride. The current *UberOne* program[2] provides such a discount along with other benefits.
- *Capacity Limits*: A variant of the above programs, where discounts are offered only for a fixed number of trips within a given time frame (e.g, 20 rides in a month; see Figure 1c).

Given the range of options, platforms have resorted to A/B testing and randomized trials to identify advantageous features [Deb et al., 2018]. Although experimentation can lead to small

[1] `https://web.archive.org/web/20190801010342/https://www.uber.com/us/en/ride/how-it-works/uber-pass/`, accessed 12/09/2021.

[2] `https://www.uber.com/us/en/u/uber-one/`, accessed 12/09/2021.



(a) Uber Ride Pass, 2017. (b) Uber $5 off Pass, 2017. (c) Lyft All-Access, 2018. (d) Lyft Pink, 2019.

Figure 1: Examples of subscription programs offered by Uber and Lyft over the years.

improvements, this is not a sustainable approach for designing globally optimal solutions because of the complexity of the design space, the long-term nature of network effects, and marketing considerations. This paper will advance new theory to guide the design of optimal subscription mechanisms.

From a business perspective, one reason that subscription schemes hold interest is that they lock-in existing customers and provide a more predictable revenue stream. Little is known, however, about whether or not they result in efficient outcomes in the case of ridesharing. Motivated by the need to bridge the gap between theory and practice, *we study the design of rider-centric subscription pricing in the present work*. For this problem, we provide a model for the operation of two-sided markets, specifically ridesharing markets, under network effects, and as a first result we demonstrate that in this model, and contrary to folk wisdom, *market-clearing prices do not maximize social welfare*.

Along the way, we spotlight two fundamental barriers to the design of efficient pricing mechanisms in our setting:

1. In order for riders to garner the benefits of network effects, it is necessary to have a pool of idle drivers close by—but at the same time, this requires the average price to be higher than the marginal system cost, which leads to inefficiencies when limited to market-clearing prices.

2. In the presence of heterogeneity in riding frequencies among individual riders, there can be a

problem with adverse selection and this complicates the setting of subscription prices.

To address the first challenge, we develop a *subscription pricing mechanism* that combines an advance fee with a real-time price and in the process, is able to compensate drivers for time spent idling. Towards the second challenge, we extend this mechanism to design a *flexible subscription policy* that offers riders a menu of subscriptions, trading price for flexibility (i.e., ability to use the rides over a larger time window). In both settings, and including in the presence of rider heterogeneity, we prove that our *subscription mechanisms maximize social welfare*. Finally, with a view towards practical implementation, we show that platforms can implement these subscriptions in a data-driven manner without access to any distribution on rider values, and still improve social welfare compared to a market-clearing mechanism.

1.1 Contributions

We consider a stylized model that is consistent with recent work on spatially separated markets [Castillo et al., 2017, Yan et al., 2020]. In this model, a continuum of riders request trips that are uniformly distributed within a single region, with independently drawn values that depend on the wait time before pickup (denoted η). A monopolist ridesharing platform selects a pricing policy (advance and/or real-time prices), following which: (i) a fraction of riders request trips based on the expected price and waiting time; (ii) a supply of drivers enter the market if their expected revenue exceeds their cost. At equilibrium, the demand rate (x), driver supply (y), and wait time (η) are connected by the number of idle drivers (O) according to the following relationship,

$$O = y - (d + \eta)x = (\frac{\tau}{\eta})^{\frac{1}{\alpha}}, \tag{1}$$

where $\tau > 0$ and $\alpha > 0$ are constants, and d denotes the average time per trip. In simple terms, this expression captures the notion that as the gap between supply and demand widens, the number of idle drivers also increases. This leads to a reduction in the average waiting time perceived by the riders. Such a relationship between the number of idle drivers and wait time is often taken for granted in the literature [Arnott, 1996, Yan et al., 2020, Kleywegt and Shao, 2021] and closely approximates the behavior of actual ridesharing sytems when the demand and supply are uniformly distributed across a region.

The platform's objective in our setting is to design a pricing mechanism that maximizes social welfare, i.e., the riders' value minus the costs incurred by drivers. The focus on social welfare is common in the literature [Ma et al., 2021, Yan et al., 2020] and is typically motivated by the fact that many ridesharing platforms aspire to maximize growth (and retain existing customers). We also show that our subscription mechanisms simultaneously improve throughput and lower the average waiting time compared to the status-quo.

Sub-optimality of Walrasian Equilibrium We begin our analysis with the widely used Walrasian market-clearing mechanism, under which the platform charges each rider a real-time trip price equal to the average trip cost incurred by drivers at equilibrium. We treat the Walrasian equilibrium as the status-quo, as it closely simulates the pricing policy used by many ridesharing platforms in the absence of subscriptions. In Theorem 1 we show that the equilibrium resulting from Walrasian pricing does not maximize social welfare. We supplement this result by showing (Section 6 and Appendix D) that the Walrasian equilibrium also achieves strictly lower throughput and higher wait times compared to the socially optimal outcome, which we refer to as the highest-welfare first best.

Subscriptions Maximize Social Welfare Any optimal mechanism must charge riders a price larger than the marginal cost of the drivers in order to compensate them for idling. To achieve this goal, we present a *simple subscription* mechanism that charges riders (*i*) an advance price (fixed fee) equal to the costs incurred by drivers due to idling, (*ii*) a real time price (via a fixed discount) equal to the marginal cost of the trip. Under this mechanism, we show the following results:

1. *Homogeneous Riders*: When rider values are drawn from the same distributions, the socially optimal outcome can be implemented through the simple subscription mechanism in equilibrium (Theorem 2).

2. *Multiple Rider Types*: When riders differ in their probability of riding, but share the same value distribution conditional on wanting a trip, we first show that naively offering a simple subscription can lead to outcomes that are substantially worse than the Walrasian equilibrium—this occurs due to adverse selection. Remedying this, we generalize the simple subscription above and design a *flexible subscription* mechanism that offers riders a menu of subscriptions. The

options in this menu differ in the advance price as well as the number of discounts applicable during the subscription period, thereby allowing riders the option of trading price for flexibility. When riders are *myopic*—i.e., upon subscribing, they request a trip every time their realized value exceeds the real-time price—we prove (Theorem 3) that this menu implements the welfare-maximizing outcome. Furthermore, we present simulations showing that the proposed mechanism is robust to non-myopic riders and achieves higher welfare than Walrasian pricing in this case as well.

Robust Subscriptions for Model-Free Environments In Yan et al. [2020], the authors pose the open question of whether it is possible to design efficient pricing mechanisms in a model-free fashion to ensure that the outcomes are robust to misspecification. More generally, practitioners seeking to implement the proposed subscription mechanism (and many of the other pricing policies in the literature) may require careful knowledge of the distribution from which riders' valuations are drawn as well as their sensitivity to increasing wait times. Based on this, we study whether a platform can improve upon the *status-quo*, i.e., the Walrasian mechanism, using only the data that is available from observing the conditions at the market equilibrium, i.e., the average wait time, idle time, etc. We answer the open question from Yan et al. [2020] in the affirmative for the case of homogenous riders and prove in Theorem 4 that the platform can still design a simple subscription mechanism that is guaranteed to maximize social welfare under the current system wait time in equilibrium. Furthermore, it is also guaranteed to achieve a strict improvement in social welfare with respect to the Walrasian equilibrium *even if a fraction of the riders decide not to subscribe against their interest.* These results suggest that the proposed design—(i) is robust to the information available to the designer, (ii) can be implemented in a data-driven manner, and (iii) can be rolled out gradually over time.

1.2 Related Work

Ridesharing platforms. Pricing and matching in ridesharing platforms have been extensively studied in recent years. Castillo et al. [2017] and Yan et al. [2020] demonstrate the importance of dynamic "surge" pricing in maintaining sufficient open driver supply in space, and improving the operational efficiency of the platform. In particular, it has been shown that in the presence of network effects, a higher density of idle drivers can improve the waiting time and throughput of riders. Empirical studies have also established the effectiveness of dynamic pricing in improving reliability and efficiency [Hall et al., 2015], driver supply [Chen and Sheldon, 2015], and the incentives for drivers to relocate to higher surge areas [Lu et al., 2018].

Existing works also study the design and operation of ridesharing platforms, with potentially strategic drivers, and in the presence of spatial imbalance and temporal variation of supply and demand [Bimpikis et al., 2019, Besbes et al., 2020, Ma et al., 2021, Garg and Nazerzadeh, 2020]. On the matching side, Ashlagi et al. [2018], Aouad and Saritaç [2020] focus on matching between riders and drivers and the pooling of shared rides, and Kanoria and Qian [2020], Qin et al. [2020], Özkan and Ward [2020] design dispatching policies that take the availability of supply in space into consideration. Queueing-theoretical models have also been used to analyze dynamic pricing [Banerjee et al., 2015], state-dependent dispatching [Banerjee et al., 2018, Castro et al., 2020], driver admission control [Afèche et al., 2018] and the capacity planning in space [Besbes et al., 2018]. Our work abstracts away the strategic behavior of drivers and the stochasticity of the system in real time. Our focus is on the use of subscription pricing to address the inefficiencies that arise due to the existence of network effects.

A large body of empirical work has studied a variety of topics, e.g.,consumer surplus [Cohen et al., 2016], the labor market of drivers [Hall et al., 2017], and the flexible work arrangements [Chen et al., 2019, 2020]. Recent theoretical work also analyzed the optimal growth of two-sided platforms Lian and Van Ryzin [2021], competition between platforms Lian et al. [2021], Ahmadinejad et al., and utilization-based minimum wage regulations Asadpour et al. [2019].

Subscriptions vs Per-Use Pricing Looking beyond ridesharing, subscription programs are also popular in a variety of other settings such as digital media, e-commerce, queuing systems, and shared facilities (e.g., gyms, theme parks), and have been the subject of a rich literature. At a high level, the prescriptions from these papers typically concern two-sided markets with digital goods or one-sided markets with limited supply and no network effects, and as such are not directly applicable to ridesharing.

Firstly, it is well known that mechanisms that charge consumers fully in advance (e.g., see Xie and Shugan [2001] and citations therein) can achieve more welfare and extract higher surplus compared to real-time prices as long as supply is inexpensive. However, when supply costs are non-trivial (as in our setting), advance pricing can be inefficient as it may result in allocations to consumers whose realized value is rather low. More generally, there is an extensive literature on subscriptions in markets with digital goods [Alaei et al., 2019, Bakos and Brynjolfsson, 1999, DeValve and Pekeč, 2021]. In this context, there is an assortment of heterogeneous goods, and the benefits of subscription stem from bundling, i.e., the platform can offer the assortment to heterogeneous consumers as a single bundle. Naturally, ridesharing platforms offer a commoditized service and the benefits of bundling are quite different in our setting.

The present work is perhaps more similar to the study of subscriptions in one-sided markets with limited supply or queuing systems with congestion [Belavina et al., 2017, Randhawa and Kumar, 2008, Cachon and Feldman, 2011]. All of these works analyze a *pure subscription* program where there is only an upfront price—this can increase consumption leading to more congestion or supply shortages. For example, Randhawa and Kumar [2008] impose capacity limits on subscriptions to circumvent this problem. In Cachon and Feldman [2011], the decision maker can alleviate congestion by purchasing supply at a fixed cost. However, their insights do not transfer to our setting due to the network effects and interaction between demand and supply, e.g., increasing the supply of drivers can improve wait times for riders but also result in lower revenue for the drivers due to increased idle time, leading drivers to drop out. As mentioned above, *pure subscriptions* can be inefficient in our setting as it results in allocations to buyers whose realized value is lower than the marginal cost. Two-part tariffs, conceptually similar to our simple subscription mechanism, have also received considerable attention Oi [1971], Png and Wang [2010] in economics. However, in these works, the advance price serves to extract rents from the consumer or offer insurance to risk-averse buyers. In contrast, the advance fee in our setting is essential to reaching social optimality in two-sided markets with network effects.

2 Preliminaries

We consider a monopolist ridesharing platform that operates in a single region. Both driver supply and rider demand within this region are non-atomic and stationary over time. Further, we model a continuous-time system where riders arrive at a specified rate and request a trip as long as their valuation for the ride is not smaller than the price. Specifically, each rider requests a trip between two randomly chosen locations within the region. Following this, the platform dispatches the closest available (i.e., idle) driver to that rider. Drivers in the system cycle between the following three states: (*i*) *idle*: waiting for dispatch, (*ii*) *en-route*: the time between dispatch and rider pick-up, which equals the riders' wait time, and (*iii*) *on-trip*: the time between pick-up and trip completion. We use w, η, and d to denote the idle time, en-route or wait time, and on-trip time respectively, per trip. We abstract away any stochasticity in the underlying wait times and idle times, and assume that the system operates at the average wait time (η) and average idle time (w). For simplicity, we also assume that the on-trip time, denoted d, is the same across all trips.

After requesting a trip, the rider incurs an average wait time η before pickup that depends inversely [Yan et al., 2020] on the distance between the origin location and the driver. We follow the example of recent papers on ridesharing Castillo et al. [2017], Yan et al. [2020] as well as classic transportation models Arnott [1996], Larson and Odoni [1981], and capture the relationship between the wait time η and the number of idle drivers O as follows:

$$\eta(O) = \tau O^{-\alpha}, \tag{2}$$

where $\tau > 0$ corresponds to the wait time when there is a unit mass of drivers in the region, and $\alpha > 0$ is a constant that depends on the system being analyzed. For example, Larson and Odoni [1981] prove that when supply and demand are distributed uniformly at random within a 2-dimensional space, Equation (2) is satisfied with an exponent $\alpha = 1/2$. Empirical estimates using Uber data suggest that this exponent holds in practice as well Yan et al. [2020]. In order to model a variety of topologies, all of our results are applicable for a general $\alpha > 0$.

Similarly, the relationship between drivers' idle time per trip w (i.e., the time spent idling between successive trips) and the number of open drivers O can be represented using Little's law as follows:

$$w = O/x, \tag{3}$$

where x is the throughput, i.e., the mass of riders requesting a trip per unit of time.

Rider Valuation A mass of riders arrive at any given time at an arrival rate specified by $n > 0$. Each rider's value for their (potential) trip is drawn independently from a distribution V^{η}, which is parameterized by wait time η. Lower wait times are preferable and associated with higher valuations. Therefore, for any $\eta_1 < \eta_2$, we assume that distribution V^{η_1} (first-order) stochastically dominates V^{η_2}. We assume that riders know their value distributions ahead of time, but discover the realization only upon arrival and just before requesting a trip. Riders act as price takers, and in particular, request a trip if and only if their value given the wait time is weakly above the price.[3]

Driver Costs There is an infinitely large supply of potential drivers available, all of whom have a uniform cost of $c > 0$ per unit of time. One could also interpret this as an opportunity cost, e.g., the average wage that can be earned through alternative employment. As with the riders, we assume every driver's actions (e.g., whether or not to enter the system) is based on their average wage, which in turn depends on the average wait time η given by Equation (2) and the average idle time w given by Equation (3). Once drivers decide to enter the system, we assume that they accept all trips dispatched by the platform and do not selectively filter them. Recall that η and w are the average wait time and idle time observed by drivers. Therefore, the time that drivers expect to spend for a single trip equals $(d + \eta + w)$.

While we take the wait time and the idle time to be uniform across the system, it is worth reiterating that these are not constants. Instead, these times depend on the number of idle drivers O at equilibrium, which in turn is a function of the throughput (demand) and supply. Mathematically, we can represent the relationship between these quantities using the following equation, where $y(x, \eta)$ denotes the supply of drivers in the system:

$$y(x, \eta) = x \cdot (d + \eta) + O(\eta) \tag{4}$$

Combining Equations (4) and (2), we observe that as driver supply increases, either the wait time increases or the number of idle drivers increases. However, both of these consequences adversely affect drivers' costs per trip, and this may lead to some drivers dropping out. The tension between

[3] At the time of making a decision, both riders and drivers are not aware of the precise wait time for any given trip. Therefore, all actions depend only on the average wait time η, which depends only on the number of open drivers according to Equation (2).

these forces keeps the system balanced. Example 1 illustrates these issues in greater detail.

We conclude this subsection with some common sense assumptions on the cumulative distribution function (CDF) F^η of the rider value distribution V^η.

Assumption 1. We make the following assumptions on the CDF of the rider value distribution F^η:

(a) Finite support, (i) for any η and $v < 0$, we have $F^\eta(v) = 0$ and (ii) there exists some upper-bound $v_{\max} > 0$ such that for any η we have $F^\eta(v_{\max}) = 1$.

(b) Strictly increasing CDF: for any η, F^η is strictly increasing in its support.

The finite support assumption implies that the supports of all value distributions are subsets of $[0, v_{\max}]$. Further, for every possible wait time η, the value distribution has non-zero density at each point within its support. For the rest of this paper, we assume that Assumption 1 is true.

2.1 Social Welfare, and the First Best Solutions

Our objective is to characterize the social welfare of various pricing mechanisms and compare it to optimal, welfare-maximizing outcomes. In particular, we use the term *first best* with respect to a wait time[4] η to denote the solution that maximizes social welfare at a particular wait time. Further, we use *highest-welfare first best* to refer to the solution that maximizes social welfare over all first-best solutions, i.e., all choices of (η, x, y). We remark that our interest in the first best solutions (at a fixed η) is driven by practical considerations. More specifically, ridesharing platforms may seek to implement solutions at a fixed—typically lower—wait time due to behavioral reasons or to distinguish themselves from their competitors. Our framework provides the platforms with the flexibility to select a benchmark that corresponds to their business needs. In addition to social welfare, we also compare solutions in their terms of throughput, and *utilization*, which is defined as $1 - w/(d + \eta + w)$, i.e., the fraction of time that drivers spend on trips.

We begin by formally defining social welfare, following which we characterize the first-best solutions at every value of η. The social welfare per unit time is the total rider value minus the

[4]In practice, the platform cannot simply select a wait time. However, it is possible to implement a specific wait time at equilibrium by carefully designing the pricing mechanism. For example, Figure 3a illustrates that Walrasian equilibria at different wait times typically correspond to different real-time prices.

total driver cost per unit time, and is formalized below:

$$\mathsf{SW}(\eta, x) = n \cdot \mathbb{E}\left[V^\eta \cdot \mathbb{1}_{\{\bar{F}^\eta(V^\eta) \leq x/n\}}\right] - c \cdot y(x, \eta) \tag{5}$$

where $\bar{F}^\eta(v)$ denotes $1 - F^\eta(v)$ and the second expression comes from Equation (4).

Lemma 1. The first best outcome under waiting time η is attained whenever the riders who take a trip are exactly those whose value is above or equal to $c(d + \eta)$.

Proof. At throughput x and waiting time η, the total cost incurred by the drivers per unit of time is given by

$$c \cdot y(x, \eta) = c \cdot (x(d + \eta) + O(\eta)).$$

Fixing η, it follows that each extra trip adds $c(d + \eta)$ to the total cost. Since the social welfare is given by the total rider value minus the total driver cost per unit of time, the marginal welfare of an extra trip is the value of the rider taking that extra trip minus $c(d + \eta)$. □

We denote by $\mathsf{optSW}(\eta)$ the welfare of the first best outcome attainable under η. The trip throughput, driver supply level and idle driver time per trip in the first best outcome are denoted by $x_{\text{opt}}(\eta), y_{\text{opt}}(\eta)$ and $w_{\text{opt}}(\eta)$, respectively.

2.2 Walrasian Equilibrium

We now discuss the Walrasian or the real-time pricing mechanism under our model. Following our earlier discussion, we treat the equilibrium resulting from this mechanism as the status quo and design subscriptions that outperform this solution. As with the first best, Walrasian equilibria may exist at multiple wait times so we concern ourselves with both Walrasian equilibria at fixed values of η and the solution that maximizes social welfare among all such equilibria which we refer to as the *highest-welfare Walrasian equilibrium.*

In simple terms, the Walrasian mechanism works as follows: *The platform selects a real-time price per trip p_{wal}. Each rider arrives at the market, observes the average wait time η, and decides to take a trip as long as the realized value (drawn from distribution V^η) is at least p_{wal}.*

We note that this mechanism results in a throughput $x_{\text{wal}} = n \cdot \bar{F}^\eta(p_{\text{wal}})$ and its social welfare is given by:

$$\mathsf{SW}(\eta, x_{\text{wal}}) = n \cdot \mathbb{E}\left[V^\eta \cdot \mathbb{1}_{\{V^\eta \geq p_{\text{wal}}\}}\right] - c x_{\text{wal}}(d+\eta) - cO \tag{6}$$

In the context of the Walrasian equilibrium, we simply use (η, x_{wal}) as arguments for the social welfare as the price p_{wal} and supply $y(x,\eta)$ follow directly. From a single driver's perspective, a trip costs $c(d+\eta+w) = c(d+\eta+O/x)$. At equilibrium, the trip price must equal the marginal driver cost per trip due to *budget balance*, which leads to the following formal definition.

Definition 1. A tuple $(x_{\text{wal}}, w_{\text{wal}}, p_{\text{wal}}, \eta)$ is a Walrasian equilibrium if:

$$\begin{aligned} x_{\text{wal}} &= n \cdot \bar{F}^\eta(p_{\text{wal}}), && \text{(Rider Best Response)} \\ p_{\text{wal}} &= c(d+\eta+w_{\text{wal}}), && \text{(Driver Best Response)} \\ w_{\text{wal}} &= O(\eta)/x_{\text{wal}} && \text{(Driver Idle Time)} \end{aligned}$$

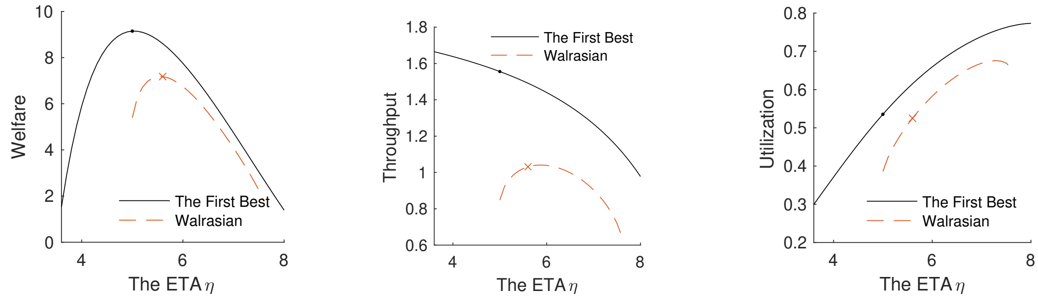
Denote by $\mathsf{walSW}(\eta)$ the social welfare corresponding to the Walrasian equilibrium at η. We are now ready to prove our first main result, which is that the Walrasian equilibrium is sub-optimal.

Theorem 1. *Under any waiting time η, the Walrasian equilibrium in the real-time market (if one exists) achieves a strictly lower social welfare than that of the first best outcome under the same wait time, i.e.* $\mathsf{walSW}(\eta) < \mathsf{optSW}(\eta)$, *for all $\eta > 0$.*

Theorem 1 also implies that the *highest-welfare first best outcome is strictly better than the highest-welfare Walrasian equilibrium.* We complement this result in Appendix D (also see Figure 2) by proving that the highest-welfare first best also achieves more throughput at a smaller wait time compared to the highest-welfare Walrasian equilibrium.

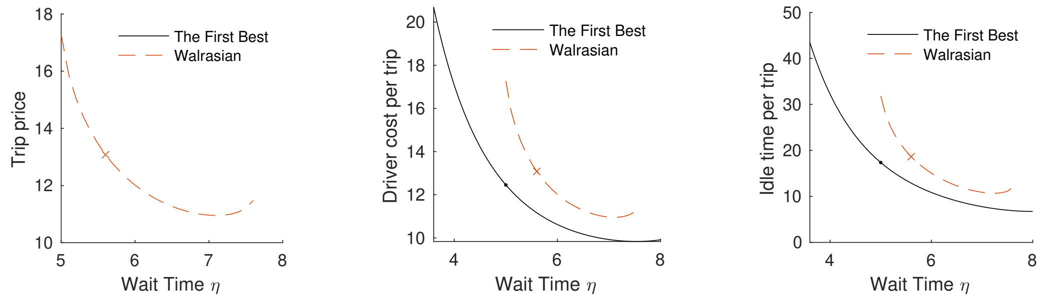
We conclude this section with an example where we compare in a simple economy the first best and the Walrasian equilibrium outcomes at different levels of waiting time.

Example 1. Consider a region where riders arrive at a rate of $n = 2$ per unit of time (e.g., minutes). Rider values are uniformly distributed: $V^\eta \sim \mathrm{U}[0, v_{\max} - \beta\eta]$, with $v_{\max} = 55$ and $\beta = 5$. This corresponds to the scenario where riders are willing to pay an average of \$20 per trip when the wait time η is around 3 minutes. Drivers incur an opportunity cost of $c = 1/3$ per minute, meaning that



(a) Social Welfare. (b) Trip Throughput. (c) Driver Time Utilization.

Figure 2: Social welfare, trip throughput, and driver time utilization under the first best and the Walrasian equilibrium at different wait times, for the economy in Example 1.



(a) Trip Price. (b) Driver Cost per Trip. (c) Idle Time per Trip.

Figure 3: Real-time trip price, the average driver cost per trip, and drivers' average waiting time for the next dispatch, under the first best and the Walrasian equilibrium at different wait times in Example 1

the drivers make an average of \$20 per hour. Each rider trip takes $d = 15$ minutes to complete. The wait time as a function of open supply level $\eta(O) = \tau O^{-\alpha}$ is specified by $\tau = 15$ and $\alpha = 1/3$.

Figures 2 and 3 compare the first best outcome and the Walrasian equilibrium at each wait time η. We can see that at any η, the first best has higher welfare, trip throughput, and driver time utilization, and lower driver cost per trip as well as driver idle time per trip. The Walrasian equilibrium with highest welfare is achieved at $\eta = 5.6$, whereas the highest possible first best is achieved at $\eta = 5$. Comparing the two outcomes, the Walrasian equilibrium achieves 78.5% of the first best welfare and 66.3% of the first best throughput.

Welfare is lower at small wait times since the cost of providing the required number of idle drivers is too high; welfare is also lower at very high η's since high wait times negatively affect rider values. Also observe that the welfare gap at lower wait times are larger— intuitively, a higher idle time per trip at these lower wait times creates a larger gap between the marginal cost of each trip and the amount riders are charged for each trip under Walrasian.

The crucial issue is that no anonymous, real-time trip price can implement the first best outcome, for any wait time. The reason is that the marginal cost of a trip is lower than the average cost of a trip, and thus the highest total payment a platform can collect from the riders is lower than the total payment that needs to be made to drivers. □

Intuitively, the first best outcome is achieved when the rider payment covers only the on-trip and wait time part of the trip (see Lemma 1), whereas the drivers also need to be compensated for the time they spend being idle. Consequently, the first best outcome cannot be implemented via only real-time pricing.

Remark Although we treat riders valuations as being drawn from a single distribution V^η, all of our claims in this section are applicable to the case where riders belong to multiple populations with heterogeneous value distributions. In such a scenario, one can simply treat F^η as the distribution of the average rider. In Section 4, we specifically model the case with multiple types of riders and design subscriptions for such a setting.

3 Simple Subscriptions

In this section, we design subscription mechanisms under which riders decide on whether to subscribe *before* they arrive at the real-time market and realize their values. Under an appropriately designed subscription, the platform is able to achieve the first best outcome at any desired wait time η, assuming that the riders are homogeneous.

Definition 2. A *simple subscription* is a tuple $S = (s, \delta, T)$ where s is the subscription fee, δ is a discount level and T is the duration of the subscription. The mechanism works as follows:

1. At the morning of day 1, with knowledge of the plan's parameters, the rider decides whether to subscribe and pay s upfront before observing the realized value.

2. *If the decision was not to subscribe*, then the rider arrives at the real-time market , observes the realized average wait time η and chooses whether to take a trip at the real time price p, given her realized value from the distribution V^η.

3. *If the rider decides to subscribe*, then: In each of the days 1 through T, the rider arrives at the real-time market, observes the realized wait time η and chooses whether to use a discount and take a trip at the reduced price $p - \delta$, given her daily realized value from the distribution V^η.

We note that at the real-time market stage, a rider's best response is to request a trip if her realized value is weakly above the effective real-time price she faces (whether that is p in case she did not subscribe, or $p - \delta$ in the case she did).

Equilibrium Outcome Under a Simple Subscription In what follows we prove that at a given wait time η, simple subscriptions can attain the first best welfare under the assumption that the riders are homogeneous. First we reason about a rider's derived utility from a simple subscription, and then we present our solution concept for this setting. Given a simple subscription $S = (s, \delta, T)$, the average daily subscription fee is s/T. Thus, a rider's daily utility from subscribing to S given the real-time price p and wait time η is given by:

$$u^S(p, \eta) = \mathbb{E}\left[(V^\eta - (p - \delta))\mathbb{1}_{\{V^\eta \geq p - \delta\}}\right] - s/T. \tag{7}$$

The daily utility from participating in the real-time market under price p and wait time η without subscribing is given by

$$u^R(p, \eta) = \mathbb{E}\left[(V^\eta - p)\mathbb{1}_{\{V^\eta \geq p\}}\right]. \tag{8}$$

If the latter is strictly greater than the former, then the rider derives higher utility by participating in the real-time market T days in a row as opposed to the total utility derived from the subscription, assuming that the wait time and real time price do not change in that time period.

Denote by n_0, n_1 the mass of subscribers and non-subscribers, respectively, such that $n_0 + n_1 = n$. Further denote by x_0, x_1 the respective throughputs, *e.g*, x_0 is the mass of non-subscribers that decide to take a trip at the discounted price.

Definition 3. A tuple $(n_0, n_1, x_0, x_1, y, p, \eta)$, forms an *equilibrium under the simple subscription*

$S = (s, \delta, T)$, if the following conditions hold:

$$
\begin{aligned}
n_0 > 0 &\implies u^S(p,\eta) \geq u^R(p,\eta), && \text{Subscriber Best Response at Subscription Stage} \\
n_1 > 0 &\implies u^R(p,\eta) \geq u^S(p,\eta), && \text{Non-Subscriber Best Response at Subscription Stage} \\
x_0 &= n_0 \Pr(V^\eta \geq p - \delta), && \text{Subscriber Best Response at Real-Time Market Stage} \\
x_1 &= n_1 \Pr(V^\eta \geq p), && \text{Non-Subscriber Best Response at Real-Time Market Stage} \\
y &= (x_0 + x_1)(d + \eta) + O(\eta), && \text{Driver Supply Level} \\
c \cdot y &= n_0 \cdot s/T + x_0 \cdot (p - \delta) + x_1 \cdot p. && \text{Total Rider Payments Equals Total Driver Costs}
\end{aligned}
$$

Note that if a strict subset of riders subscribe, then we must have $u^S(p,\eta) = u^R(p,\eta)$.

Theorem 2. *Given any waiting time η for which $\mathsf{optSW}(\eta) > 0$, for any desired duration T, there is a simple subscription with duration T that attains first best welfare in an equilibrium outcome in which all riders subscribe.*

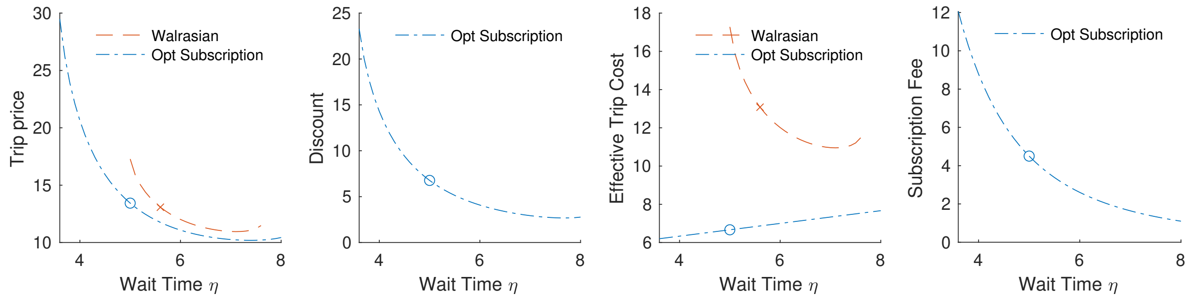
Intuitively, the subscription we design has subscribers face the desired effective real-time price corresponding to the on-trip and wait time part of the trip, while the subscription fees cover the time drivers spend being idle.

In the simple subscription constructed in the proof of Theorem 2, the real time price p was set high enough (while keeping the discounted price $p - \delta$ fixed) to ensure riders prefer subscribing over not subscribing. A natural candidate for such a price p is the actual amount paid to a driver per trip in the corresponding first-best outcome. In the following we show that this price is not high enough.

Proposition 1. If drivers must be paid the non-discounted real-time price, then simple subscriptions cannot achieve the first best outcome in equilibrium.

We now revisit the same economy analyzed in Example 1, and illustrate the outcome under the optimal simple subscription at different levels of ETA.

Example 1 (Continued)**.** The optimal simple subscription achieves in equilibrium the first best outcome at every wait time η. As a result, the welfare, throughput, utilization, idle time per trip and driver cost per trip are the same as those under the first best, as illustrated in Figures 2 and 3.



(a) Trip Price. (b) Discount. (c) Effective Trip Cost. (d) Subscription Fee.

Figure 4: The trip price, discount, the effective cost of trips for riders (i.e. price - discount), and the subscription fee at different wait time levels, for the economy in Example 1.

Trip prices and discounts at each η are as shown in Figures 4a and 4b. In comparison to the best Walrasian equilibrium (at $\eta = 5.6$), the best outcome under subscription (at $\eta = 5$) slightly increases the trip price, from 13.08 to 13.43. The effective costs of a trip a rider needs to pay in real-time (i.e. the trip price under Walrasian, and the trip price minus the discount under subscription) are shown in Figure 4c. The substantially lowered cost under subscription allows all riders whose value is above the marginal driver cost for providing an additional trip (i.e. $c(d+\eta)$) to take a ride.

Figure 4d illustrates the subscription fee each rider is charged (per day). To achieve a lower wait time, which corresponds to a higher driver idle time per trip, the optimal subscription uses a higher real-time price, higher discount, and charges a higher subscription fee. $\square$

4 Multiple Rider Types

In this section we study subscription mechanisms in the presence of heterogeneous riders. We show that when some riders ride more frequently than the others, the simple subscriptions introduced in Section 3 cannot always attain the first best outcome in equilibrium. We then prove the main result of this section, that by offering a *flexible subscription* in addition to a simple subscription, a platform is able to incentivize all riders to subscribe and achieve the first best welfare.

Model for Heterogeneous Riders We focus on the case of two rider types: *high type riders* whose values are drawn from distribution V_h^η, and *low type riders* with value distribution V_ℓ^η. For $i \in \{h, \ell\}$, n_i denotes the arrival rate of the i type rider, and $n := n_h + n_\ell$ is the total rider arrival rate.

We assume that the two value distributions only differ in their *initial interest to take a trip*: For each $i \in \{h, \ell\}$ we have a probability Q_i denoting this interest, and the value conditioned on not being interested is 0. Furthermore, conditioned on being interested in a trip, the rider value is drawn from the same value distribution V^η for both types. In other words, each individual rider belonging to type $i \in \{h, \ell\}$ decides on whether or not to take a trip according to a Bernoulli random variable with probability Q_i. If this random variable evaluates to zero, then the rider's value for the trip is also zero; otherwise the rider value is drawn from V^η.

Remark We reiterate that all claims in Section 2 which assumed homogeneous riders are also applicable to our setting when applied to the "average rider" population. In this section we shall mainly make use of Lemma 1, whose statement and proof carry over to our setting in their current formulations exactly.

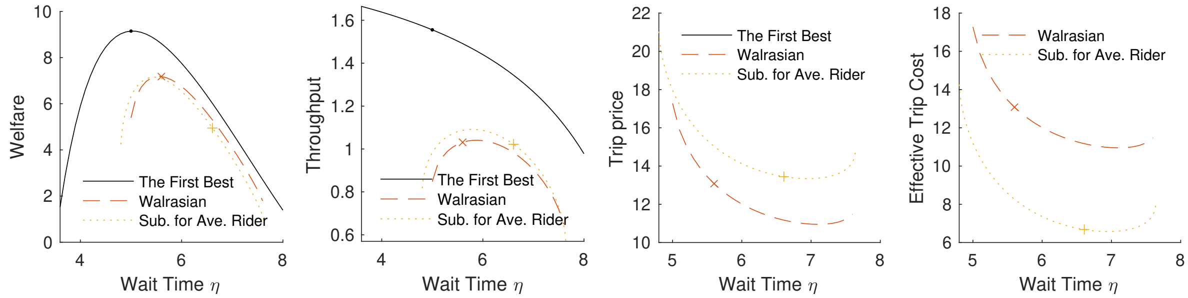
Naturally we assume that $Q_h > Q_\ell > 0$. This allows for several interpretations. For instance, the high type riders can be thought of as commuting riders who tend to request more trips, and the low type riders can be thought of as casual riders who use the platform occasionally (*e.g.*, only when the weather is bad).

For any η we let $r^\eta = \mathbb{E}\left[V^\eta - c(d+\eta) \mid V^\eta \geq c(d+\eta)\right]$ be the real time rider surplus conditioned on taking a trip in the first best outcome under η. The following is a useful observation.

Observation 1. For any wait time η, we have $\mathsf{optSW}(\eta) = x_{\mathrm{opt}}(r^\eta - cw_{\mathrm{opt}}(\eta))$. In particular, since $\mathsf{optSW}(\eta) > 0$, then we have $r^\eta > cw_{\mathrm{opt}}(\eta)$.

Before proceeding, we illustrate via the following example that when a platform naively treats the rider population as homogeneous and offers a seemingly-optimal simple subscription, the resulting equilibrium outcome can be substantially worse than the Walrasian equilibrium.

Example 2. Consider a setting that is otherwise identical to the economy in Example 1, but has two types of riders with different riding frequencies. High type riders arrive at a rate of $n_h = 1$ per unit of time, and the low type riders' arrival rate is $n_\ell = 6$. Each day, the two types of riders are interested in take a trip with probabilities $Q_h = 0.8$ and $Q_\ell = 0.2$, respectively. In this way, the arrival rate of riders who are interested in riding is $n_h Q_h + n_\ell Q_\ell = 2$ per unit of time. Conditioned on being interested in a trip, we assume that the value of a rider follows the same



(a) Social Welfare. (b) Trip Throughput. (c) Trip Price. (d) Effective Trip Cost.

Figure 5: The social welfare, trip throughput, trip price, and the effective real-time trip cost for riders at different wait times, for the economy in Example 2.

uniform distribution $V^\eta \sim \mathrm{U}[0, 55 - 5\eta]$ regardless of the rider's type.

When the platform assumes homogeneous riders, and naively offers a simple subscription that is optimal for a rider population who is interested in riding with probability $(n_h Q_h + n_\ell Q_\ell)/(n_h + n_\ell) = 0.2$, the equilibrium outcomes at different wait time η are as shown in Figure 5. Since the subscription fee is determined for the "average rider", it is to the low type rider's best interest to only participate in the real-time market. When the platform targets the real-time price that is appropriate for the average rider, the market equilibrates at $\eta = 6.61$, achieving a social welfare of only 4.95.

This also created a large inequity between the two rider types. In real-time, the low type riders are charged 13.45 per trip, whereas after the discount, high type riders pay only 6.68 per trip. Even after paying the subscription fee, high type riders get an average surplus of 2.97 per day. This is 9 times the average daily surplus of 0.33 obtained by the low type riders, despite the fact that the high type riders are only 4 times as likely to be interested in a ride. □

In addition to Example 2, we show formally that when the different rider types' probability of being interested in a trip is different enough, no simple subscription can attain first best welfare for a given waiting time η in an equilibrium outcome. Intuitively, since the subscription fee is the same for all riders, it must be targeted at the average rider. As a consequence, the high type riders get a much better deal relative to the low type riders, since their likelihood to take a trip is higher than that of the average rider, which in turn is higher than that of the low type rider. Simple subscriptions lack *flexibility*, in the sense that the each day that a rider does not take a trip results in a wasted discount.

Proposition 2. Simple subscriptions cannot always attain the first best outcome in equilibrium when riders are heterogeneous.

In addition to Proposition 2, we also derive a precise condition under which implementing the first best welfare with a simple subscription is possible in equilibrium. A formal definition of an equilibrium outcome for two rider types under a simple subscription and proofs of these claims are deferred to the appendix.

4.1 Flexible Subscriptions

In this section we introduce a new subscription model that allows us to cope with the different rider types' likelihood of requesting a trip. As before, we assume throughout that the system waiting time is some fixed η.

Definition 4. A *flexible subscription mechanism* C is a tuple $C = (s, \delta, k, T)$ where s is the subscription fee, δ is the discount level, k is the number of applicable discounts and T is the maximum duration of the subscription period. Given a menu $\mathcal{M} = \left(C^j\right)_{j=1}^m$ of m such subscriptions, where $C^j = \left(s^j, \delta^j, k^j, T^j\right)$, the mechanism works as follows:

1. At the morning of day 1, with knowledge of all the different plans' parameters, and before observing her daily realized value, the (type i) rider decides whether to subscribe to any of the offered subscriptions and pay its subscription fee upfront, before observing her realized value.

2. If the decision was not to subscribe, then the rider arrives at the real-time market, observes the realized average wait time η and chooses whether to take a trip at the realized real time price p, given her realized value from the distribution V_i^η. Otherwise,

3. Starting from the day in which the current subscription ,denoted C^j, was purchased, the rider arrives at the real-time market every day, observes the realized average wait time η and can choose to use a discount and take a trip at the reduced price $p - \delta^j$, given her realized daily value from the distribution V_i^η.

4. The subscription period ends in the day that she used all k^j discounts or at the T^j'th day since purchasing the current subscription, the earlier of the two. In the latter case any unused discounts are wasted.

5. The rider can choose to renew the same subscription immediately in the morning of the day that follows the last day of the previous subscription period. Alternatively she can immediately purchase an alternative subscription. In any of these cases, she loops back to step (3).

Myopic Assumption For this section, we assume that all riders are *myopic*, in the sense that they always use a discount if their daily realized value is above the discounted trip price. This is rational for $k = T$, when the subscription plan is really a simple subscription. In this case, any day in which the subscriber does not take a trip results in a wasted discount. Thus, whenever the realized value is above the effective trip price, the rider should take the trip. For $k < T$, the rider may be incentivized to behave in a strategic fashion, and save the discount for a later day if her realized value is very close to the discounted price. However, we argue that it is reasonable for riders to treat the subscription fee as a sunk cost Thaler [1985] and exhibit bounded rationality (myopic behavior) on a day-to-day basis. For instance, a rider may not wish to save their discount for a later day when it is essential to take a trip on the present day, the realized value is larger than the discounted price, and future valuations are unknown. This is particularly salient when the value of k is close to T. There is considerable evidence for the prevalence of sunk cost bias when consumers are faced with two part tariffs Kong et al. [2018].

In what follows we analyze the surplus that a rider derives from a flexible subscription plan, assuming myopia. When a rider considers whether to subscribe or not, she should compare the daily surplus from the different subscription plans being offered and from participating in the real-time market only.

It is challenging to reason about the daily surplus from a subscription plan, since the durations of these plans are random, implying that the days in which the rider pays the subscription fee are random. These depend on the daily random trip values that determine whether the rider uses a discount or not.

To analyze the daily utility from a subscription plan C, we assume that the riders join the system and purchase the plan on day 1, and thereafter they always immediately repurchase the same subscription upon completion of the previous one. This assumption is reasonable assuming that the other alternatives faced by the rider exhibit the same utility over time (*e.g.*, if the non discounted real time price p and wait time η stay the same). Thus, since riders are myopic, the *(daily) surplus* that a type i rider derives from a flexible subscription $C = (s, \delta, k, T)$ given the realized real time price p and wait time η is given by

$$u_i^C(p,\eta) = \mathbb{E}\left[(V_i^\eta - (p-\delta))\,\mathbb{1}_{\{V_i^\eta \geq p-\delta\}}\right] - s \cdot \mathbb{E}\left[\lim_{t\to\infty} \frac{1}{t}\sum_{j=1}^{t} \mathbb{1}_{\{i \text{ purchases } C \text{ on day } j\}}\right]$$

The term on the left is the real-time surplus generated from the daily discounted trip and the term on the right is the amortized daily subscription fee payment, assuming immediate renewal of the same subscription plan upon completion of the previous one.

For a given flexible subscription mechanism C, fixed realized real time price p, fixed realized wait time η and a type i rider, we denote by $D_i^C(p,\eta)$ the corresponding (random) duration of a single subscription period. Formally, this is the number of days between purchase and completion of the plan, which is the minimum between T and, assuming myopia, the day since purchase which marks the k'th time that the daily realized value from the distribution V_i^η exceeded $p - \delta$ (*i.e.*, the day in which the last discount was applied).

The following lemma captures the intuition that, on average, a rider pays the subscription fee s once every $\mathbb{E}\left[D_i^C(p,\eta)\right]$ days.

Lemma 2 (Subscription fee payments per day)**.** Let C be a flexible subscription. Then a type i rider who subscribes to C with repeated immediate renewals, given fixed real time price p and wait time η, satisfies

$$\mathbb{E}\left[\lim_{t\to\infty} \frac{1}{t}\sum_{j=1}^{t} \mathbb{1}_{\{i \text{ purchases } C \text{ on day } j\}}\right] = \frac{1}{\mathbb{E}\left[D_i^C(p,\eta)\right]}$$

As a corollary, the daily surplus of a type i rider from subscribing to $C = (s, \delta, k, T)$, given fixed realized real time price p and wait time η, is given by

$$u_i^C(p,\eta) = \mathbb{E}\left[(V_i^\eta - (p-\delta))\,\mathbb{1}_{\{V_i^\eta \geq p-\delta\}}\right] - \frac{s}{\mathbb{E}\left[D_i^C(p,\eta)\right]}. \tag{9}$$

Remark When $k = T$, the flexible subscription is effectively a simple subscription. Note that in this case $D_i^C(p,\eta)$ is deterministic and equals T, and accordingly the corresponding utility expressions in Equation (9) (for a flexible subscription) and Equation (7) (for a simple subscription) coincide.

Optimal Flexible Subscriptions for Multiple Rider Types In this section we prove that at a given wait time η, flexible subscriptions can achieve the first best welfare when there are high type and low type riders. To achieve this, we offer the riders a menu of two different subscription plans that they can choose to subscribe to. One is tailored for the high type riders and the other for the low type riders.

We start by defining an equilibrium outcome under a general menu $\mathcal{M} = \left(C^j\right)_{j=1}^m$, where $C^j = (s^j, \delta^j, k^j, T^j)$, for any number of subscriptions $m \in \mathbb{N}_{>0}$. For $i \in \{h, \ell\}$ and $j \in \{1, \ldots, m\}$, we denote by n_i^j the mass of type i riders who subscribe to C^j, and we denote by n_i^0 the mass of type i riders who choose to participate in the real time market only. These must satisfy $\sum_{j=0}^m n_i^j = n_i$. Further denote by x_i^j the respective throughputs, *e.g.*, x_ℓ^0 is the mass of non-subscribing low type riders who decide to take a trip given the realized non-discounted real time price and wait time η. In the definition below, δ^0 is interpreted as 0 (*i.e.* there is zero discount for non-subscribers), and $[m]$ denotes the set $\{0, 1, \ldots, m\}$.

Definition 5. A tuple $\left(\left(n_i^j, x_i^j\right)_{i\in\{h,\ell\}, j\in[m]}, y, p, \eta\right)$ is an *equilibrium outcome for h, ℓ type riders under the menu of subscriptions* $\mathcal{M} = \left(C^j\right)_{j=1}^m$ if the following conditions hold:

$$\forall i \in \{h, \ell\}, j, j' \in [m] : n_i^j > 0 \implies u_i^j(p, \eta) \geq u_i^{j'}(p, \eta), \tag{10}$$

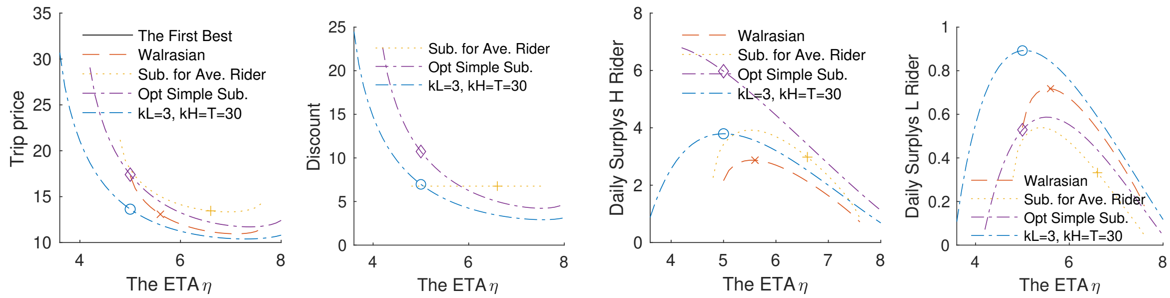
$$\forall i \in \{h, \ell\}, j, \in [m] : x_i^j = n_i^j \Pr(V_i^\eta \geq p - \delta^j), \tag{11}$$

$$y = (\sum_{i\in\{h,\ell\}, j\in[m]} x_i^j)(d + \eta) + O(\eta), \tag{12}$$

$$c \cdot y = \sum_{i\in\{h,\ell\}, j\in[m]\setminus\{0\}} n_i^j \cdot s^j / \left(\mathbb{E}\left[D_i^j(p, \eta)\right]\right) + \sum_{i\in\{h,\ell\}, j\in[m]} x_i^j (p - \delta^j) \tag{13}$$

where (10) is rider best response at subscription stage, (11) is rider best response at the real time market stage, (12) is the correct driver supply level, and (13) requires that total rider payments equals total driver costs.

Theorem 3. *For any wait time η for which* $\mathsf{optSW}(\eta) > 0$ *, we can design a subscription menu $\mathcal{M} = (H, L)$, where $H = \left(s^H, \delta, k^H, T\right)$ and $L = \left(s^L, \delta, k^L, T\right)$ with $k^L < k^H = T$ that attains the first best welfare in an equilibrium outcome in which all high type (low type) riders subscribe to the H (L) plan.*



(a) Trip Price. (b) Discount. (c) Surplus per Day - H. (d) Surplus per Day - L.

Figure 6: Trip prices, discounts, and the expected surplus per day for the two rider types at different wait times, for the economy in Example 2.

Remark While we prove Theorem 3 assuming that the rider types only differ in their initial interest of taking a trip, we mention that the theorem can also be proved for more general distributions. Namely, the theorem can be shown to hold whenever, for the given η, both types $i \in \{h, \ell\}$ satisfy $r_i^\eta = \mathbb{E}\left[(V_i^\eta - c(d+\eta)) \mid V_i^\eta \geq c(d+\eta)\right] > cw_{\text{opt}}(\eta)$. and note that this indeed holds in our setting by Observation 1 and since $r_i^\eta = r^\eta$ for both rider types.

Example 2 (Continued)**.** We now revisit the economy studied at the beginning of this section, and compare in Figure 6 two additional mechanisms: (i) a single simple subscription that implements the first best outcome, and (ii) a flexible subscription (with $k^L = 3$ and $T = 30$) is offered in addition to a simple subscription. Both settings implement the first best outcome at all wait times, thus we refer readers to Figure 5 for the equilibrium welfare and throughput at different wait times.

In order to incentivize all riders to subscribe, a platform offering only a simple subscription needs to raise the real-time price substantially. With two plans, on the country, the price is only slightly higher than that under the Walrasian equilibrium. This is desirable in practice, since there are always newer riders and out-of-town travelers who have not yet oped-in any subscription plan.

Similar to naively catering to the "average rider", the use of an optimal simple contract also leads to a large inequity— the high type riders enjoy an average surplus of 5.97 per day, whereas that of the low type rider's daily surplus is only 0.53. In contrast, when both a simple and a flexible contract are offered, both types of riders are better off than under Walrasian equilibrium. □

5 Robust Implementation of Optimal Subscriptions in Practice

In Sections 3 and 4, we presented different subscription plans that attain first best welfare at a given waiting time η. In order to implement these, the platform has to be able to compute the different subscription fees. However, among other things, the fees depend on the probability that riders take a trip at the first best effective trip price $c(d+\eta)$. If the platform is not currently running at the first best outcome, how can it measure these probabilities? It is not clear what existing data a platform can use to perform these calculations.

In this section we show that, under the assumption that riders are homogeneous, the platform can implement a simple subscription that attains a first best outcome in equilibrium, *without prior knowledge of the rider distributions.* Rather, the platform only has access to the outcome statistics of a Walrasian equilibrium, *i.e.*, in non-subscription mode.

In this first best equilibrium outcome, all riders subscribe. However, We also show that the subscription plan is robust to riders deciding not to subscribe against their interest. Even though these irrational decisions result in sub-optimal welfare (due to the non-subscribers facing a real-time price higher than $c(d+\eta)$), the resulting outcome is still an equilibrium:

In our previous subscription plans, the subscription fees were charged in order to cover for the time the drivers spend being idle (recall that subscribers pay a discounted price of $c(d+\eta)$ per trip, whereas the drivers get paid $c(d+\eta+w_{\text{opt}})$ per trip). It may seem that if some riders do not subscribe, the platform would not have enough money to cover this idle time. Here, we show that the platform collects the exact amount of money needed to cover for the idle drivers, no matter how many riders subscribe. Consequently, the system remains in equilibrium.

From a practical perspective, this property is important, since it allows the platform to roll out the subscription gradually: the system will still continue to function, and the more riders subscribe, the higher the attained welfare.

Robust Optimal Subscriptions for Homogeneous Riders We show how to implement a simple subscription assuming homogeneous riders with (unknown) value distribution V^η, under the assumption that the system is currently running at a Walrasian equilibrium with wait time η. As before, we denote the rider arrival rate, Walrasian throughput, driver idle time and trip price by

$n, x_{\text{wal}}, w_{\text{wal}}, p_{\text{wal}}$ respectively, and note that all these parameters are observable by the platform.

Theorem 4. *Let the throughput x_{wal}, driver idle time w_{wal} and trip price p_{wal} specify a Walrasian equilibrium outcome, under wait time η. Then, assuming that riders are homogeneous:*

(i) We can design a simple subscription with parameters based only on $(\eta, x_{\text{wal}}, p_{\text{wal}}, w_{\text{wal}}, c)$ that achieves the first best welfare at wait time η in an equilibrium outcome in which all riders subscribe.

(ii) For any $n_0 \in (0, n]$, the exact same subscription as above achieves a strict improvement in social welfare over the Walrasian equilibrium, even if only n_0 riders subscribe and the remaining do not subscribe.

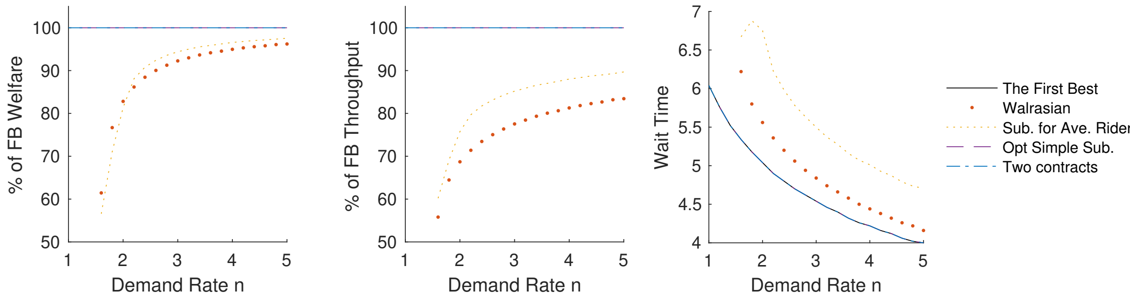
The second part of the theorem is motivated by practical considerations as some riders may not opt to subscribe even if it is in their best interest to do so. Under such an event, our result indicates that the outcome from the same simple subscription in the first part still satisfies all of the conditions in Definition 3 except for non-subscriber best response at the subscription stage.

6 Simulation Results

In this section, we compare in simulation various mechanisms and benchmarks as we vary the density of the market (i.e. the arrival rate of riders). Additional simulation results are provided in Appendix B, where we vary (i) the mix of high and low frequency riders, and (ii) the level of flexibility provided by the flexible subscription plan.

Consider economies where the riding frequency and value distribution of each rider type, driver cost, trip duration and the wait time$\sim$open supply relation are the same as in the economy analyzed in Example 2. We vary the rider arrival rate from $n = 1$ to $n = 5$ per unit of time, but keep the fraction of high type riders fixed at $1/7$ and the fraction of low type riders at $6/7$. Figures 7 and 8 compare the equilibrium outcomes under various mechanism and benchmarks. Additional performance metrics are provided in Figures 10 and 11 in Appendix B.1.

Sub. for Ave. Rider (i.e. the optimal simple subscription for the average rider) naively offers a single simple subscription that is optimal for a homogeneous rider population who are interested in riding with probability $(n_h Q_h + n_\ell Q_\ell)/(n_h + n_\ell) = 0.2$. Opt Simple Sub. (i.e. the optimal simple



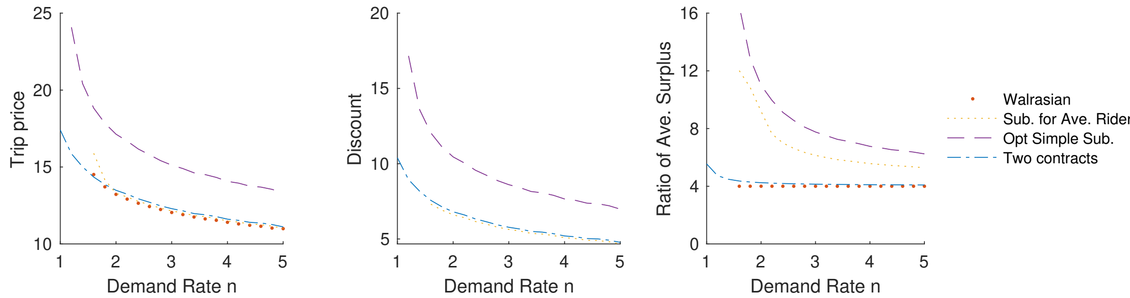
(a) % of FB Welfare. (b) % of FB Throughput. (c) Wait Time.

Figure 7: Social welfare (as % of the first best welfare), trip throughput (as % of the first best throughput), and equilibrium wait time under various benchmarks and mechanisms. Varying total demand rate n.

subscription) takes rider heterogeneity into consideration, and raises both trip prices and discount levels in order to incentivize all riders to subscribe. Two contracts, on the other hand, offers a simple subscription as well as a flexible subscription (with $k^L = 3$ and $T = 30$), which are priced in a way that (i) the high and low type riders will opt-in for the simple and flexible plans respectively, and (ii) the platform is budget balanced overall.

Figure 7 compares for each mechanism and benchmark (a) the highest equilibrium welfare, (b) trip throughput under the welfare-optimal outcome, and (c) the wait time that optimizes welfare. The (relative) inefficiency of the Walrasian equilibrium outcome decreases as the market becomes more sense. The optimal simple contract and the two contracts both implement the first best outcome with highest welfare at any demand rate. The simple subscription targeting the “average rider” is able to achieve some improvements over the Walrasian equilibrium for sense markets, however, the equilibrium wait time is substantially higher due to the fact that the low type riders choose not to subscribe, and as a result the platform fails to collect sufficient payment from riders to maintain a larger pool of open driver supply.

The equilibrium prices and discount levels are presented in Figures 8a and 8b. As the market becomes more sense, both the real-time price and the discount level decreases, since the platform is able to operate in a more efficient manner. See Appendix B.1 for additional results illustrating that the subscription fee decreases whereas the the utilization of drivers increases for denser markets. Figure 8c plots the ratio between the average surplus per day of the high type riders, and that of the low type riders. Since the high type riders are 4 times as likely to be interested in a trip, this “daily surplus ratio” is exactly 4 under the Walrasian equilibrium for any demand rate n. Two



(a) Trip Price. (b) Discount. (c) Ratio of Ave. Surplus per Day.

Figure 8: Prices, discounts, and ratio between the average surplus per day for high type and the low type riders under various benchmarks and mechanisms. Varying total demand rate n.

contracts achieve a surplus ratio very close to 4 (hence is more "fair"), whereas the ratio under the optimal simple contract is substantially higher, especially for sparse markets.

7 Discussion and Concluding Remarks

In many markets, the optimality of the Walrasian market clearing mechanism is taken for granted, and subscriptions are typically considered a business necessity. As is evident from Figure 1 and Appendix A, there is also no consensus among the ridesharing community on what a good subscription program looks like. Against this backdrop, we establish the suboptimality of real-time pricing in the presence of network effects and present a subscription mechanism that is:

1. *Easy to implement*: The proposed mechanism is simple, intuitive, and follows the structure of a fixed discount subscription. At the same time, the prices lend themselves to a natural interpretation as the advance fee allows the platform to close the gap between the average and marginal costs, and compensate drivers for time spent idling.

2. *Optimizes social welfare and throughput*: In the case of rider heterogeneity, we design a flexible subscription that offers a menu of options. The mechanism achieves the optimum welfare without any targeting as long as riders behave myopically. Further, our mechanism partially overcomes adverse selection and ensures higher surplus for all rider types (see Example 2).

3. *Amenable to a data-driven rollout*: We do not need access to riders' precise valuations. Surprisingly, the subscription can improve social welfare even the platform rolls out the program gradually or only a fraction of riders decide to subscribe.

References

Philipp Afèche, Zhe Liu, and Costis Maglaras. Ride-hailing networks with strategic drivers: The impact of platform control capabilities on performance. Technical report, SSRN:3120544, 2018.

AmirMahdi Ahmadinejad, Hamid Nazerzadeh, Amin Saberi, Nolan Skochdopole, and Kane Sweeney. Competition in ride-hailing markets. In *Proceedings of WINE 2019*.

Saeed Alaei, Ali Makhdoumi, and Azarakhsh Malekian. Optimal subscription planning for digital goods. *Available at SSRN 3476296*, 2019.

Ali Aouad and Ömer Saritaç. Dynamic stochastic matching under limited time. In *Proceedings of the 21st ACM Conference on Economics and Computation*, pages 789–790, 2020.

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

Arash Asadpour, Ilan Lobel, and Garrett van Ryzin. Minimum earnings regulation and the stability of marketplaces. *Available at SSRN 3502607*, 2019.

Itai Ashlagi, Maximilien Burq, Chinmoy Dutta, Patrick Jaillet, Amin Saberi, and Chris Sholley. Maximum weight online matching with deadlines. *arXiv preprint arXiv:1808.03526*, 2018.

Yannis Bakos and Erik Brynjolfsson. Bundling information goods: Pricing, profits, and efficiency. *Management science*, 45(12):1613–1630, 1999.

Siddhartha Banerjee, Ramesh Johari, and Carlos Riquelme. Pricing in ride-sharing platforms: A queueing-theoretic approach. In *ACM Conference on Economics and Computation (EC)*, 2015.

Siddhartha Banerjee, Yash Kanoria, and Pengyu Qian. State dependent control of closed queueing networks. In *Abstracts of the 2018 ACM International Conference on Measurement and Modeling of Computer Systems*, 2018.

Elena Belavina, Karan Girotra, and Ashish Kabra. Online grocery retail: Revenue models and environmental impact. *Management Science*, 63(6):1781–1799, 2017.

Omar Besbes, Francisco Castro, and Ilan Lobel. Spatial capacity planning. *Available at SSRN 3292651*, 2018.

Omar Besbes, Francisco Castro, and Ilan Lobel. Surge pricing and its spatial supply response. *Management Science*, 2020.

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

Gérard P Cachon and Pnina Feldman. Pricing services subject to congestion: Charge per-use fees or sell subscriptions? *Manufacturing & Service Operations Management*, 13(2):244–260, 2011.

Juan Camilo Castillo, Dan Knoepfle, and Glen Weyl. Surge pricing solves the wild goose chase. In *Proceedings EC 2017*, pages 241–242. ACM, 2017.

Francisco Castro, Peter Frazier, Hongyao Ma, Hamid Nazerzadeh, and Chiwei Yan. Matching queues, flexibility and incentives. *arXiv preprint arXiv:2006.08863*, 2020.

Kuan-Ming Chen, Claire Ding, John A List, and Magne Mogstad. Reservation wages and workers' valuation of job flexibility: Evidence from a natural field experiment. Technical report, National Bureau of Economic Research, 2020.

M Keith Chen and Michael Sheldon. Dynamic pricing in a labor market: Surge pricing and the supply of uber driver-partners. *University of California (Los Angeles) Working Paper*, 2015.

M Keith Chen, Peter E Rossi, Judith A Chevalier, and Emily Oehlsen. The value of flexible work: Evidence from uber drivers. *Journal of Political Economy*, 127(6):2735–2794, 2019.

Peter Cohen, Robert Hahn, Jonathan Hall, Steven Levitt, and Robert Metcalfe. Using big data to estimate consumer surplus: The case of uber. *NBER Working Paper No. 22627*, 2016.

Anirban Deb, Suman Bhattacharya, Jeremey Gu, T Zhou, Eva Feng, and Mandie Liu. Under the hood of uber's experimentation platform. *Uber engineering.*, 2018.

Levi DeValve and Saša Pekeč. Optimal price/advertising menus for two-sided media platforms. *Available at SSRN 2715531*, 2021.

Ari Freedman. Convergence theorem for finite markov chains. *Proc. REU*, 2017.

Nikhil Garg and Hamid Nazerzadeh. Driver surge pricing. *Proceedings of the 21st ACM Conference on Economics and Computation*, 2020.

Jonathan Hall, Cory Kendrick, and Chris Nosko. The effects of uber's surge pricing: A case study. Technical report, The University of Chicago Booth School of Business, 2015.

Jonathan V Hall, John J Horton, and Daniel T Knoepfle. Labor market equilibration: Evidence from uber. Technical report, New York University Stern School of Business, 2017.

Yash Kanoria and Pengyu Qian. Blind dynamic resource allocation in closed networks via mirror backpressure. In *Proceedings of EC 2020*, 2020.

Anton J Kleywegt and Hongzhang Shao. Optimizing pricing, repositioning, en-route time, and idle time in ride-hailing systems. *arXiv preprint arXiv:2111.11551*, 2021.

Guangwen Kong, Sampath Rajagopalan, and Chunyang Tong. Pricing diagnosis-based services when customers exhibit sunk cost bias. *Production and Operations Management*, 27(7):1303–1319, 2018.

Richard C Larson and Amedeo R Odoni. *Urban operations research*. Number Monograph. 1981.

David A Levin and Yuval Peres. *Markov chains and mixing times*, volume 107. American Mathematical Soc., 2017.

Zhen Lian and Garrett Van Ryzin. Optimal growth in two-sided markets. *Management Science*, 2021.

Zhen Lian, Sebastien Martin, and Garrett van Ryzin. Larger firms pay more in the gig economy. *Available at SSRN*, 2021.

Alice Lu, Peter I Frazier, and Oren Kislev. Surge pricing moves uber's driver-partners. In *Proceedings of the 2018 ACM Conference on Economics and Computation*, pages 3–3, 2018.

Hongyao Ma, Fei Fang, and David C Parkes. Spatio-temporal pricing for ridesharing platforms. *Operations Research*, 2021.

Walter Y Oi. A disneyland dilemma: Two-part tariffs for a mickey mouse monopoly. *The Quarterly Journal of Economics*, 85(1):77–96, 1971.

Erhun Özkan and Amy R Ward. Dynamic matching for real-time ride sharing. *Stochastic Systems*, 10(1):29–70, 2020.

Ivan PL Png and Hao Wang. Buyer uncertainty and two-part pricing: theory and applications. *Management Science*, 56(2):334–342, 2010.

Zhiwei Qin, Xiaocheng Tang, Yan Jiao, Fan Zhang, Zhe Xu, Hongtu Zhu, and Jieping Ye. Ride-hailing order dispatching at DiDi via reinforcement learning. *INFORMS Journal on Applied Analytics*, 50(5):272–286, 2020.

Ramandeep S Randhawa and Sunil Kumar. Usage restriction and subscription services: Operational benefits with rational users. *Manufacturing & Service Operations Management*, 10(3):429–447, 2008.

Richard Thaler. Mental accounting and consumer choice. *Marketing science*, 4(3):199–214, 1985.

Jinhong Xie and Steven M Shugan. Electronic tickets, smart cards, and online prepayments: When and how to advance sell. *Marketing Science*, 20(3):219–243, 2001.

Chiwei Yan, Helin Zhu, Nikita Korolko, and Dawn Woodard. Dynamic pricing and matching in ride-hailing platforms. *Naval Research Logistics (NRL)*, 67(8):705–724, 2020.

Appendix

A Subscription Offers

In this section, we briefly describe some more example subscription programs offered by the ridesharing platforms Uber and Lyft over the past few years, as well as the typical app interface for riders.

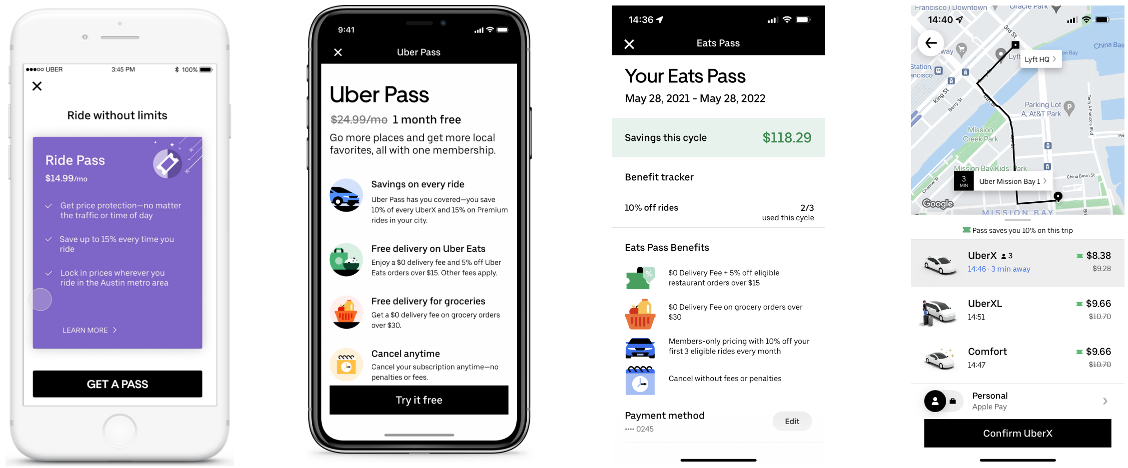
Uber

- Ride Pass with "flat fares", unlimited rides at \$3.49 for Pool and \$6.49 for UberX (Figure 1a). Offered in 2017.[5]
- Fixed dollar discount: An email indicating that a rider has purchased a pass that takes \$5 off every ride (Figure 1b). Offered in fall 2017.
- Ride Pass offering 15% off as well as "price protection" (Figure 9a), i.e. riders do not pay the dynamic "surge pricing" during times when demand exceeds supply. Offered in 2018.[6]
- Uber pass offering a fixed 10% discount off the fare of every ride, for \$24.99 a month (Figure 9b).
- Eats Pass offering 10% off for up to three rides each month, in addition to benefits on food delivery orders (Figure 9c). Figure 9d indicates that the rider's trip price will be discounted by 10%.
- The current UberOne program: 5% off every ride (together with a number of other benefits).[7]

Lyft

- All-Access Plan: 30 rides up to \$15 per ride, every 30 days for \$299 (Figure 1c), offered in 2018.[8] A more expensive option was also tested, offering 60 rides a month for \$399 a month.[9]

[5] https://medium.com/@jsbotto/have-you-heard-of-the-uber-ride-pass-3ebf2af24568
[6] https://www.uber.com/newsroom/ride-pass/
[7] https://www.uber.com/us/en/u/uber-one/, accessed 02/08/2022.
[8] https://www.theverge.com/2018/10/16/17978626/lyft-monthly-subscription-plan-nationwide
[9] https://www.macrumors.com/2018/03/16/lyft-monthly-subscription-plans/



(a) Price protection, 2018. (b) 10% off pass, 2020. (c) Savings this cycle. (d) 10% off trip prices.

Figure 9: Additional examples of Uber's Subscription Programs, and the rider App interface indicating a rider's saving per cycle and the trip discount.

- Lyft Pink: offers subscribes 15% off every ride, among a few other benefits for \$20 a month 1d. Offered in 2019.[10]

B Additional Simulation Results

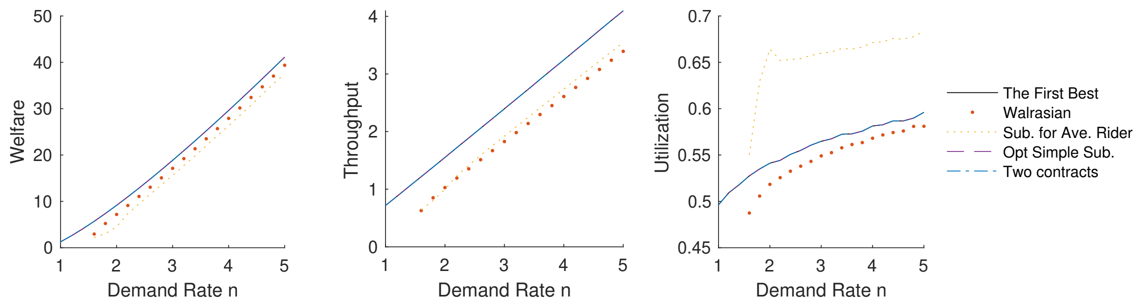
In this section, we include additional simulation results omitted from the body of the paper.

B.1 Varying Market Density

We first provide additional results on the impact of market density, for the settings studied in Section 6. Figures 10a and 10b plot the social welfare and trip throughput for each demand rate n. In addition, Figure 10c provides the average utilization of drivers' time, i.e. the fraction of drivers' time spent on either driving towards the riders or driving riders from their origins to their destinations. We can see that driver utilization improves under all mechanisms and mechanisms as the market becomes more dense.

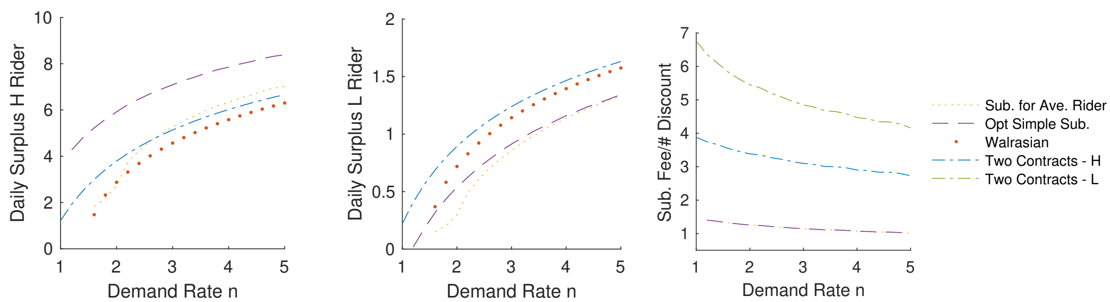
Figures 11a and 11b plot the average surplus per day for each rider type. The ratio between the two is included in Figure 8. Similar to what we have observed in Example 2, the two contract mechanism improves the surplus of both rider types in comparison to the outcome under the

[10] https://www.theverge.com/2019/10/29/20936982/lyft-pink-subscription-price-discount-perks



(a) Social Welfare. (b) Trip Throughput. (c) Utilization.

Figure 10: Social welfare, trip throughput, and driver time utilization under various benchmarks and mechanisms. Varying total demand rate n.



(a) Ave Surplus per Day - H. (b) Ave Surplus per Day - L. (c) Subscription Fee per Discount.

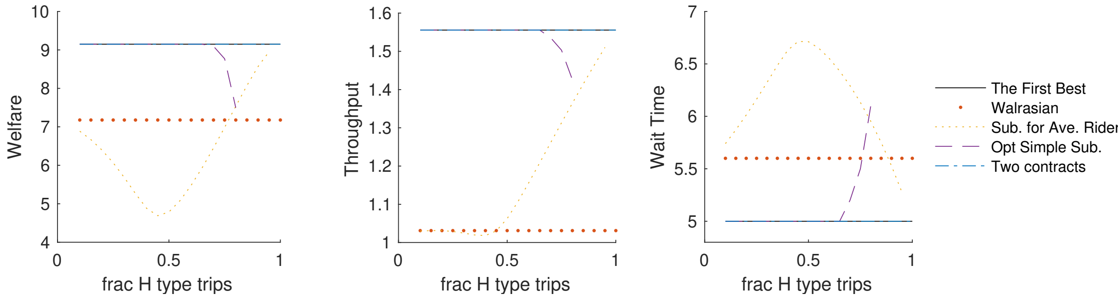
Figure 11: Average rider surplus for both types, and the subscription fee (per discount) under various mechanisms. Varying total demand rate n.

Walrasian equilibrium. On the other hand, raising the price in order to incentivize all riders to accept the same simple subscription makes the low type rider substantially worse off.

Finally, we present the subscription fees per discount charged by the various mechanisms in Figure 11c. As an example, if the L plan charges a fee of \$12 and provides three discounts, then the fee per discount is \$4. The H plan charges a lower fee per discount in comparison to the flexible subscription in order to incentivize the high type riders to subscribe, whereas the L type riders are willing to pay more for each discount for the flexibility allowed by the flexible plan.

B.2 Varying Rider Composition

We now examine the impact of rider composition in markets where some riders take trips more often than the others. Observe that in Example 2, the total arrival rate of riders who are interested in a ride is $n_h Q_h + n_\ell Q_\ell = 2$. Out of these riders, $n_h Q_h / 2 = 40\%$ are of the H type, thus the fraction of trips taken by the H type riders under the first best outcome is also 40%. In this section, we



(a) Social Welfare. (b) Trip Throughput. (c) Wait Time.

Figure 12: Welfare, throughput, and equilibrium wait time under various benchmarks and mechanisms. Varying the fraction of high type riders among all riders who are interested in a trip on a given day.

study the same economy as that in Example 2, but vary rider composition such that the fraction of trips taken by the high type riders (under the first best) varies from 0 to 1, while keeping the arrival rate of riders who are interested in a ride fixed at 2 per unit of time.

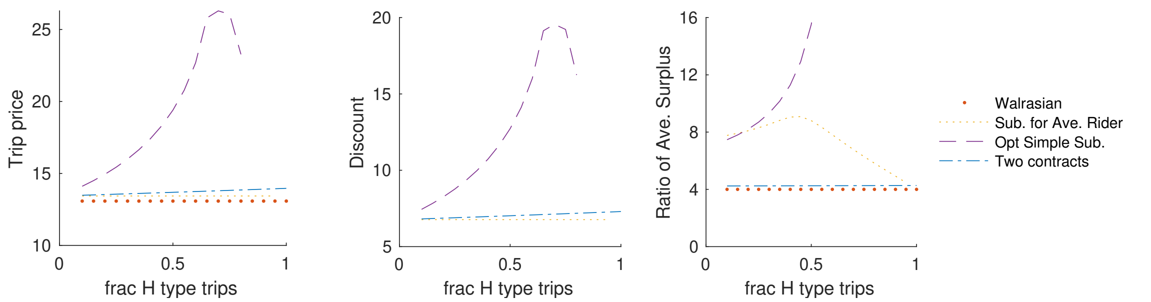
Figure 12 shows that despite being able improve trip throughput most of the time, the simple subscription for the "average rider" perform poorly in terms of welfare and wait time when there is a close-to-even mix of high and low type rider trips. When the fraction of H type riders is small, a platform is able to implement the first best outcome using a single simple subscription, but the equilibrium outcome is increasingly inequitable as the fraction of H type rider trips increases. The outcome under Two contracts, on the other hand,

Figure 13 presents the trip price, discount level, and the ratio between the average surplus per day for the H and the L type riders. Again, a single simple subscription uses very high price and discount levels, and leads to substantial inequity in the surplus of the two rider types. Figure 14 provides driver utilization and the average daily surplus of the two types of riders.

B.3 Varying Flexibility Level

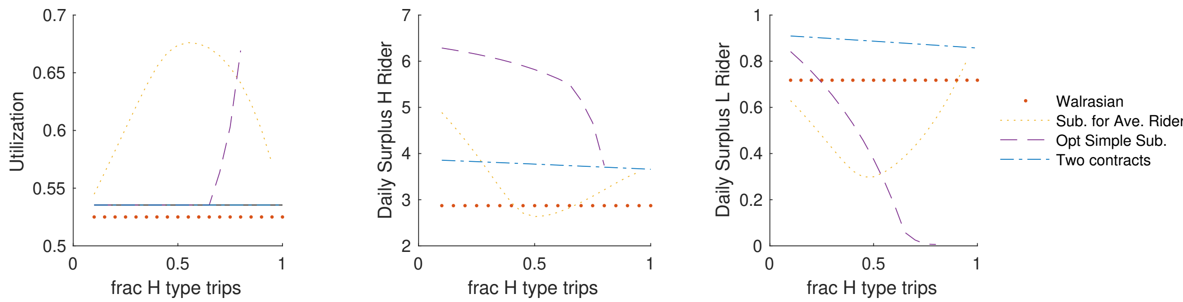
In this section, we fix the economy as the one in Example 2, but vary the level of flexibility provided by the flexible subscription part of the two contract plan by varying k^L from 1 to 30 (while fixing $k^H = T = 30$). As k^L increases, flexibility decreases the L type riders must purchase a larger number of discounts at a time. When $k^H = T = 30$, the flexible subscription reduces to the simple subscription (i.e. there is no longer any flexibility).

Figure 15 presents the trip price, discount, and the subscription fee charged per each discount.



(a) Trip Price. (b) Discount. (c) Ratio of Ave. Surplus per Day.

Figure 13: Prices, discounts, and ratio between the average surplus per day for high type and the low type riders under various benchmarks and mechanisms. Varying the fraction of high type riders among all riders who are interested in a trip on a given day.



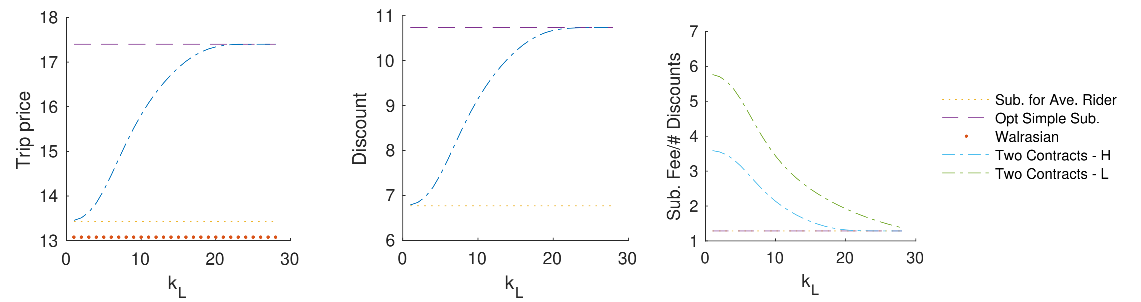
(a) Utilization. (b) Ave Surplus per Day - H. (c) Ave Surplus per Day - L.

Figure 14: Driver utilization and the average surplus per day for the two types of riders under various benchmarks and mechanisms. Varying the fraction of the high type riders among all riders who are interested in a trip on a given day.

Figure 16 plots the average surplus per day for the two rider types, and the "surplus ratio" of the two rider types (a ratio of 4 is "fair", as discussed earlier).[11] Unsurprisingly, as k^L increases and the flexibility for the L type rider decreases, the equilibrium outcome converges to that under single simple subscription. More flexibility corresponds to a smaller inequity in rider surpluses, however, when there is a large number of discounts that needs to be used up in $T = 30$ days, the amount a L type rider can gain from rational instead of myopic decision making is smaller.

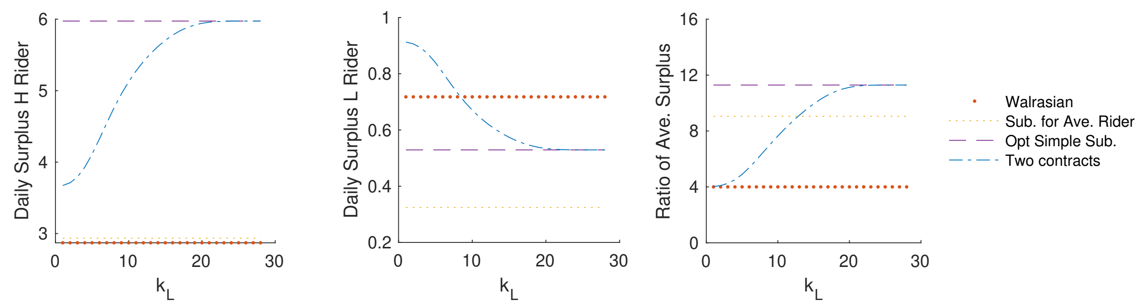
Figures 17a and 17b presents the average subscription fee paid by each rider type for each trip that the riders end up taking (e.g. if a rider pays \$60 a month and takes \$20 trips a month, then the subscription fee per trip is \$3). In other word, this is the fee paid by riders for each "used discount". The ratio of the two is plotted in Figure 17c. As we've discussed in the body of the

[11]The outcome under every other mechanism does not depend on k^L, and are included here for comparison purposes only. Also note that the two contract plan implements the first best outcome under all k^L's, thus we omitted the performance metrics such as welfare, throughput, and wait time.



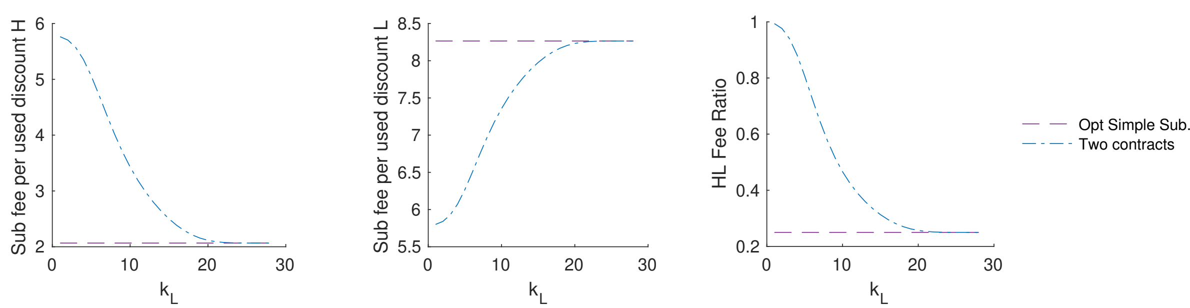
(a) Trip Price. (b) Discount. (c) Sub. Fee / # Discounts.

Figure 15: Prices, discounts, and the subscription fee per discount under various benchmarks and mechanisms. Varying the flexibility level k^L of the two contract mechanism.



(a) Ave Surplus per Day - H. (b) Ave Surplus per Day - L. (c) Ratio of Ave. Surplus per Day.

Figure 16: Average surplus per day for the two rider types, and the ratio of the two under various benchmarks and mechanisms. Varying the flexibility level k^L of the two contract mechanism.



(a) Sub. Fee per Trip - H. (b) Sub. Fee per Trip - L. (c) Ratio of Fee/Trip.

Figure 17: Social welfare, trip throughput, and driver time utilization under various benchmarks and mechanisms. Fraction of trips coming from H type riders under the first best outcome.

paper, a smaller k^L i.e. more flexibility leads to better equity between the two types of riders.

C Mathematical Properties

Lemma 3. For any η and constants $a < b$, if $[a, b] \subseteq \mathsf{supp}\,(V^\eta)$, then

$$a \Pr(a \leq V^\eta \leq b) < \mathbb{E}\left[V^\eta \cdot \mathbb{1}_{\{a \leq V^\eta \leq b\}}\right] < b \Pr(a \leq V^\eta \leq b)$$

and the claim remains true when making any of the inequalities within the probabilities strict.

Proof. We prove that $\mathbb{E}\left[V^\eta \cdot \mathbb{1}_{\{a \leq V^\eta \leq b\}}\right] < b \Pr(a \leq V^\eta \leq b)$. The other inequality is analogous. Let c be some number such that $a < c < b$. Then:

$$\begin{aligned}
\mathbb{E}\left[V^\eta \cdot \mathbb{1}_{\{a \leq V^\eta \leq b\}}\right] &= \mathbb{E}\left[V^\eta \cdot \mathbb{1}_{\{a \leq V^\eta \leq c\}}\right] + \mathbb{E}\left[V^\eta \cdot \mathbb{1}_{\{c < V^\eta \leq b\}}\right] \\
&\leq c \Pr(a \leq V^\eta \leq c) + b \Pr(c < V^\eta \leq b) \\
&< b \Pr(a \leq V^\eta \leq c) + b \Pr(c < V^\eta \leq b) \\
&= b \Pr(a \leq V^\eta \leq b)
\end{aligned}$$

where the last inequality holds since $\Pr(a \leq V^\eta \leq c) > 0$ by Assumption 1. $\square$

Lemma 4. Suppose that $F(v)$ and $f(v)$ denote the cumulative distribution function (CDF) and probability density function (PDF) for some distribution of consumer valuations such that $F(v_{\max}) = 1$ for $v_{\max} > 0$. Then, the expected value of a unit mass of consumers whose valuation is at least $p^* < v_{\max}$ is given by:

$$\int_{v=p^*}^{v_{\max}} v f(v) dv = p^*(1 - F(p^*)) + \int_{v=p^*}^{v_{\max}} (1 - F(v)) dv.$$

Proof. The proof proceeds along typical lines used to write the expected value of a random variable in terms of its CDF. In particular, we have that:

$$\begin{aligned}
\int_{v=p^*}^{v_{\max}} v f(v) dv &= \int_{p^*}^{v_{\max}} f(v) \int_0^v dt dv \\
&= \int_0^{p^*} \int_{p^*}^{v_{\max}} f(v) dv dt + \int_{p^*}^{v_{\max}} \int_t^{v_{\max}} f(v) dv dt \\
&= p^*(1 - F(p^*)) + \int_{p^*}^{v_{\max}} (1 - F(t)) dt
\end{aligned}$$

$$= p^*(1 - F(p^*)) + \int_{v=p^*}^{v_{\max}} (1 - F(v))dv.$$

We note that the second line above is the result of changing the order of the integration. This completes the proof of the lemma. □

Assumption 2. (*Monotone Hazard Rate (MHR)*) A distribution with cumulative density function $F(v)$ and probability density function $f(v)$ is said to have a monotone hazard rate if Equation (14) is non-increasing in v.

$$\frac{1 - F(v)}{f(v)} \tag{14}$$

Lemma 5. Given a monotone hazard rate distribution with CDF $F(v)$, the expression in Equation (15) is non-increasing in v.

$$\frac{\int_{t=v}^{\infty} (1 - F(t))dt}{1 - F(v)}. \tag{15}$$

Proof. Intuitively, the expression in Equation (15) captures the average hazard rate for $t \geq v$ and as such must be non-increasing due to the MHR assumption. Formally, suppose we use $h(v)$ to denote the hazard rate at v, i.e., $h(v) = \frac{1-F(v)}{f(v)}$. Then, we have that:

$$\int_{t=v}^{\infty} (1 - F(t))dt \leq \int_{t=v}^{\infty} h(v)f(t)dt = h(v)(1 - F(v)). \tag{16}$$

Consider the expression in Equation (15) at some arbitrary v_1, v_2 such that $v_1 < v_2$. Our goal is to show that:

$$\frac{\int_{t=v_1}^{\infty} (1 - F(t))dt}{1 - F(v_1)} \geq \frac{\int_{t=v_2}^{\infty} (1 - F(t))dt}{1 - F(v_2)}.$$

Towards this goal, we have that:

$$\frac{\int_{t=v_1}^{\infty} (1 - F(t))dt}{1 - F(v_1)} = \frac{\int_{t=v_1}^{v_2} (1 - F(t))dt + \int_{t=v_2}^{\infty} (1 - F(t))dt}{(F(v_2) - F(v_1)) + 1 - F(v_2)}.$$

Consider the second terms in the numerator and denominator. We have from (16) that:

$$\frac{\int_{t=v_2}^{\infty} (1 - F(t))dt}{1 - F(v_2)} \leq \frac{h(v_2)(1 - F(v_2))}{1 - F(v_2)} = \frac{\int_{t=v_1}^{v_2} h(v)f(t)dt}{F(v_2) - F(v_1)} \leq \frac{\int_{t=v_1}^{v_2} (1 - F(t))dt}{F(v_2) - F(v_1)}.$$

In the final expression, we used the fact that $h(v_2)f(t) \leq h(t)f(t) = 1 - F(t)$. Using the above inequality and the identity that $\frac{a+b}{c+d} \geq \frac{c}{d}$ when $\frac{a}{b} > \frac{c}{d}$, we can conclude that:

$$\frac{\int_{t=v_1}^{\infty}(1-F(t))dt}{1-F(v_1)} = \frac{\int_{t=v_1}^{v_2}(1-F(t))dt + \int_{t=v_2}^{\infty}(1-F(t))dt}{(F(v_2)-F(v_1))+1-F(v_2)}$$
$$\geq \frac{\int_{t=v_2}^{\infty}(1-F(t))dt}{1-F(v_1)}$$

□

D Properties of First Best and Walrasian Outcomes

Next, we prove that under some mild assumptions, the highest-welfare first best outcome outperforms the highest-welfare Walrasian equilibrium in terms of both throughput and waiting time, i.e. the first best achieves more throughput at a lower wait time. Before stating our result, we introduce some structure on how wait times affect a rider's valuation.

Assumption 3. (*Additive Separable Valuations*) We say that the cumulative distribution function F^η is *additively separable* if there exist non-decreasing, differentiable functions F and g such that: (i) for all (v, η), we have $F^\eta(v) = F(v + g(\eta))$, (ii) $\eta g'(\eta)$ is non-decreasing in η.

This assumption is consistent with the literature, e.g., using data, Yan et al. [2020] estimate that the CDF has the following logit form $F^\eta(v) = 1 - \frac{e^{\beta \cdot v + \delta \cdot \eta}}{1+e^{\beta \cdot v + \delta \cdot \eta}}$ where $\beta, \delta < 0$ denote some suitable constants. Further, the function $g(\eta)$ denotes riders' sensitivity to wait times. The second part of Assumption (3)—i.e., $\eta g'(\eta)$ is non-decreasing in η—implies that as the wait time increases, a rider's sensitivity to wait times must also increase at a sufficient rate.

We begin by characterizing the wait time η^* at which the highest-welfare first best solution occurs.

Proposition 3. Suppose that $(\eta^*_{\text{opt}}, x^*_{\text{opt}})$ denotes the highest-welfare first best solution. Then, under Assumption 3, we have that:

$$(c + g'(\eta^*_{\text{opt}}))x^*_{\text{opt}} = c\frac{1}{\alpha\eta^*_{\text{opt}}}(\frac{\tau}{\eta^*_{\text{opt}}})^{\frac{1}{\alpha}} \tag{17}$$

Further, as long as the distribution $F(v + g(\eta))$ has a monotone hazard rate, the unique highest welfare first best solution occurs at the smallest value of η^*_{opt} that satisfies Equation (17).

Proof. We know from Lemma 1 that the first best solution at any fixed η is characterized by $x_{\text{opt}} = n\bar{F}^{\eta}(d+\eta)$. Therefore, we can write the social welfare at the first best solutions as a function of η using Lemma 4 as follows:

$$\begin{aligned} SW^*(\eta) &= c(d+\eta)x_{\text{opt}} + n\int_{v=c(d+\eta)}^{v_{\max}} (1-F(v+g(\eta)))dv - c(d+\eta)x_{\text{opt}} - c(\frac{\tau}{\eta})^{\frac{1}{\alpha}} \\ &= n\int_{v=c(d+\eta)}^{v_{\max}} (1-F(v+g(\eta)))dv - c(\frac{\tau}{\eta})^{\frac{1}{\alpha}}. \end{aligned} \tag{18}$$

We note that while applying Lemma 4 for a specific η, we used $p^* = c(d+\eta)$ and $1-F^{\eta}(p^*) = \frac{x}{n}$. Differentiating Equation (18) with respect to η (via the Leibniz rule), we get that:

$$\begin{aligned} \frac{d}{d\eta}SW^*(\eta) &= -nc\left(1-F(c(d+\eta)+g(\eta))\right) - ng'(\eta)\int_{v=c(d+\eta)}^{v_{\max}} f(v+g(\eta))dv + c\frac{1}{\alpha\eta}(\frac{\tau}{\eta})^{\frac{1}{\alpha}} \\ &= -cx_{\text{opt}} - g'(\eta)x_{\text{opt}} + c\frac{1}{\alpha\eta}(\frac{\tau}{\eta})^{\frac{1}{\alpha}}. \end{aligned}$$

Note that in the above expression x_{opt} is also a function of η. At the highest-welfare first best solution $(\eta^*_{\text{opt}}, x^*_{\text{opt}}, y^*)$, the right hand side of the above expression must equal zero. Substituting that, we get:

$$(c+g'(\eta^*_{\text{opt}}))x^*_{\text{opt}} = c\frac{1}{\alpha\eta^*_{\text{opt}}}(\frac{\tau}{\eta^*_{\text{opt}}})^{\frac{1}{\alpha}}$$

Uniqueness: Now we argue about uniqueness. Since $\frac{d}{d\eta}SW^*(\eta)$ is continuous in η, it is easy to see that the highest welfare first best solution cannot occur at any η that does not satisfy Equation (17); if that were the case, one could either increase or decrease η slightly to improve upon the social welfare. Therefore, we restrict our attention to solutions that do satisfy Equation (17). In particular, let us denote by $(\eta^*_{\text{opt}}, x^*_{\text{opt}})$ the solution with the smallest wait time satisfying Equation (17) and let (η', x'_{opt}) denote any other solution satisfying Equation (17). We need to show that $SW^*(\eta^*_{\text{opt}}) > SW^*(\eta')$. We begin with two simple observations:

- $\eta^*_{\text{opt}} < \eta'$: This is by definition.
- $x^*_{\text{opt}} > x'_{\text{opt}}$: This follows from Equation (17) according to which x decreases as η increases. Note that $\eta g'(\eta)$ is also non-decreasing as η increases according to Definition (3).

We begin by writing out the expression for social welfare in terms of Equation (18) and then replacing $c(\frac{\tau}{\eta^*_{\text{opt}}})^{\frac{1}{\alpha}} = n(c + g'(\eta^*_{\text{opt}}))x^*_{\text{opt}} \cdot \alpha\eta^*_{\text{opt}}$ as per Equation (17).

$$\begin{aligned}
SW^*(\eta^*_{\text{opt}}) &= n\int_{v=c(d+\eta^*_{\text{opt}})}^{v_{\max}} (1 - F(v + g(\eta^*_{\text{opt}})))dv - c(\frac{\tau}{\eta^*_{\text{opt}}})^{\frac{1}{\alpha}} \\
&= n\int_{v=c(d+\eta^*_{\text{opt}})}^{v_{\max}} (1 - F(v + g(\eta^*_{\text{opt}})))dv - n(c + g'(\eta^*_{\text{opt}}))x^*_{\text{opt}} \cdot \alpha\eta^*_{\text{opt}} && (19) \\
&= nx^*_{\text{opt}}\left(\frac{\int_{v=c(d+\eta^*_{\text{opt}})}^{v_{\max}}(1 - F(v + g(\eta^*_{\text{opt}})))dv}{x^*_{\text{opt}}} - (c + g'(\eta^*_{\text{opt}}))\alpha\eta^*_{\text{opt}}\right) \\
&= nx^*_{\text{opt}}\left(\frac{\int_{v=c(d+\eta^*_{\text{opt}})}^{v_{\max}}(1 - F(v + g(\eta^*_{\text{opt}})))dv}{1 - F(c(d + \eta^*_{\text{opt}}) + g(\eta^*_{\text{opt}}))} - (c + g'(\eta^*_{\text{opt}}))\alpha\eta^*_{\text{opt}}\right) \\
&\geq nx^*_{\text{opt}}\left(\frac{\int_{v=c(d+\eta')}^{v_{\max}}(1 - F(v + g(\eta')))dv}{1 - F(c(d + \eta') + g(\eta'))} - (c + g'(\eta^*_{\text{opt}}))\alpha\eta^*_{\text{opt}}\right) && (20) \\
&> nx'_{\text{opt}}\left(\frac{\int_{v=c(d+\eta^*_{\text{opt}})}^{v_{\max}}(1 - F(v + g(\eta^*_{\text{opt}})))dv}{1 - F()} - (c + g'(\eta'))\alpha\eta^*_{\text{opt}}\right). \\
&= SW^*(\eta'). && (21)
\end{aligned}$$

Since F has a monotone hazard rate and since $c(d + \eta^*_{\text{opt}}) + g(\eta^*_{\text{opt}}) < c(d + \eta') + g(\eta')$, we can apply Lemma 5 and infer that

$$\frac{\int_{v=c(d+\eta^*_{\text{opt}})}^{v_{\max}}(1 - F(v + g(\eta^*_{\text{opt}})))dv}{1 - F(c(d + \eta^*_{\text{opt}}) + g(\eta^*_{\text{opt}}))} \geq \frac{\int_{v=c(d+\eta')}^{v_{\max}}(1 - F(v + g(\eta')))dv}{1 - F(c(d + \eta') + g(\eta'))}.$$

Moreover, since $\eta^*_{\text{opt}} < \eta'$ and $x^*_{\text{opt}} > x'_{\text{opt}}$, we get the strict inequality that $SW(\eta^*_{\text{opt}}) > SW(\eta')$. This completes the uniqueness proof. □

We are now ready to prove that the highest welfare first best solution outperforms the (highest welfare) Walrasian outcome in terms of both throughput and wait time.

Proposition 4. Given a problem instance, suppose that $(\eta^*_{\text{opt}}, x^*_{\text{opt}})$ and $(\eta^*_{\text{wal}}, x^*_{\text{wal}})$ denote the highest-welfare first best and Walrasian outcomes respectively. Under Assumption 3 and as long as the distribution $F(v + g(\eta))$ has a monotone hazard rate, we have that:

$$\eta^*_{\text{opt}} < \eta^*_{\text{wal}} \qquad \text{(and)} \qquad x^*_{\text{opt}} > x^*_{\text{wal}}$$

We note here that we make the MHR assumption strictly for convenience as this class of distributions has been extensively studied in the literature. In fact, our results hold under a much weaker assumption on the distribution $F(v+g(\eta))$, namely that the expression in Equation (22) is non-increasing in η.

$$\frac{\int_{v=c(d+\eta)}^{v_{\max}}(1-F(v+g(\eta)))dv}{1-F(c(d+\eta)+g(\eta))} \tag{22}$$

Proof. We begin by pointing out that given wait time $\eta > 0$, the Walrasian outcome (η, x_{wal}) at this wait time (if one exists) is characterized by the real-time price $p_{\text{wal}} = c(d+\eta+w_{\text{wal}})$, where $w_{\text{wal}} = \frac{1}{x_{\text{wal}}}(\frac{\tau}{\eta})^{\frac{1}{\alpha}}$. Using Lemma 4, we can write the social welfare of any Walrasian outcome (η, x_{wal}) as follows:

$$\begin{aligned}\mathsf{walSW}(\eta) &= c(d+\eta+w_{\text{wal}})x_{\text{wal}} + n\int_{v=c(d+\eta+w_{\text{wal}})}^{v_{\max}}(1-F(v+g(\eta)))dv - c(d+\eta)x_{\text{wal}} - cx_{\text{wal}}w_{\text{wal}} \\ &= n\int_{v=c(d+\eta+w_{\text{wal}})}^{v_{\max}}(1-F(v+g(\eta)))dv.\end{aligned}$$

It is worth noting here that as per the above expression, since the rider distribution is additively separable, social welfare is maximized when the expression $p' \triangleq c(d+\eta+w_{\text{wal}})+g(\eta)$ is minimized. Equivalently, one can argue that social welfare is maximized for Walrasian equilibria only when the throughput $x_{\text{wal}} = n(1-F(p'))$ is maximized. Therefore, to complete the proof, it suffices for us to show that the throughput is not maximized at any $\eta \leq \eta^*_{\text{opt}}$.

Towards this end, consider any $\eta \leq \eta_{\text{opt}}$ and let $x_{\text{wal}} > 0$ be such that (η, x_{wal}) is a Walrasian equilibrium at wait time η. Also, let $\eta' = \eta + \epsilon$ for some sufficiently small ϵ—we will show that for any Walrasian equilibrium (η', x'_{wal}), it must be the case that $x_{\text{wal}} < x'_{\text{wal}}$, thereby completing the required claim. Define:

$$h(\eta_1, x_1) = n\left(1 - F(c(d+\eta_1+\frac{1}{x_1}(\frac{\tau}{\eta_1})^{\frac{1}{\alpha}}) + g(\eta_1))\right),$$

and note that at the Walrasian equilibria (η, x_{wal}), it must be true that $x_{\text{wal}} = h(\eta, x_{\text{wal}})$. Next, we claim (and prove later) that fixing $x = x_{\text{wal}}$, $\frac{\partial h(\eta, x_{\text{wal}})}{\partial \eta} > 0$, for $\eta \leq \eta^*$. This implies that as long as

ϵ is sufficiently small:

$$h(\eta', x_{\text{wal}}) > x_{\text{wal}}$$
$$h(\eta', n) < n.$$

The first inequality above stems from the fact that $\eta' = \eta + \epsilon$ and that the partial derivative of $h(\eta, x_{\text{wal}})$ is strictly increasing at η. The second inequality is a simple consequence of the fact that $F(c(d + \eta' + \frac{1}{n}(\frac{\tau}{\eta'})^{\frac{1}{\alpha}}) + g(\eta')) > 0$. Since $h(\eta', x)$ is continuous in x, there must exist some $x'_{\text{wal}} \in (x_{\text{wal}}, n)$ at which $h(\eta', x_{\text{wal}}) = x'_{\text{wal}}$. This completes our proof that the throughput is strictly increasing for $\eta \leq \eta^*_{\text{opt}}$, and so the highest welfare Walrasian equilibrium must occur at $\eta > \eta^*_{\text{opt}}$. Further, it is not hard to see that this solution must have a throughput that is strictly smaller than x^*_{opt}. That is because according to Theorem (1), x^*_{wal} must be strictly smaller than the throughput of the first best solution at $\eta = \eta^*_{\text{wal}}$. Since $\eta^*_{\text{wal}} < \eta^*_{\text{opt}}$, the throughput of the first best at η^*_{wal} must in turn be smaller than or equal to x^*_{opt}.

All that remains is for us to prove that $\frac{\partial h(\eta, x_{\text{wal}})}{\partial \eta} > 0$, for $\eta \leq \eta^*$. For conciseness, let $p' = c(d + \eta_1 + \frac{1}{x_1}(\frac{\tau}{\eta_1})^{\frac{1}{\alpha}}) + g(\eta_1)$. Then, have that:

$$\frac{\partial h(\eta, x_{\text{wal}})}{\partial \eta} = -nf(p')\left(c + g'(\eta) - \frac{1}{\alpha \eta x_{\text{wal}}}(\frac{\tau}{\eta})^{\frac{1}{\alpha}})\right)$$

From Proposition 3, we know that under the conditions mentioned in this proposition (i.e., additive separability and MHR distribution) that the unique optimum solution occurs at η^*_{opt} and so, for any $\eta_{\text{opt}} < \eta^*_{\text{opt}}$, the derivative of the social welfare of the first best cannot be positive. Mathematically, this can be expressed as: $c + g'(\eta) \leq \frac{1}{\alpha \eta x_{\text{opt}}}(\frac{\tau}{\eta})^{\frac{1}{\alpha}}$, where x_{opt} denotes the throughput of the first best solution at η. Since $x_{\text{opt}} > x_{\text{wal}}$, we can conclude that:

$$c + g'(\eta) - \frac{1}{\alpha \eta x_{\text{wal}}}(\frac{\tau}{\eta})^{\frac{1}{\alpha}} < c + g'(\eta) - \frac{1}{\alpha \eta x_{\text{opt}}}(\frac{\tau}{\eta})^{\frac{1}{\alpha}} \leq 0.$$

This completes our proof.

□

We conclude this section by proving that the throughput monotonically increases with the arrival rate n while the wait time decreases.

Proposition 5. Given a problem instance, suppose that $\mathrm{opt}(n)$ denotes the set of highest-welfare first best solutions when the rider arrival rate is n. Consider $n^2 > n^1$. For every $(\eta^1, x^1_{\mathrm{opt}}) \in \mathrm{opt}(n^1)$, there exists $(\eta^2, x^2_{\mathrm{opt}}) \in \mathrm{opt}(n^2)$ such that:

$$x^2_{\mathrm{opt}} \geq x^1_{\mathrm{opt}} \quad \eta^2 \leq \eta^1.$$

Proof. Consider any arbitrary $(x^1_{\mathrm{opt}}, \eta^1) \in \mathrm{opt}(n^1)$, i.e., a welfare-maximizing solution when the arrival rate is n^1. We know based on our characterization that $x^1_{\mathrm{opt}} = n^1 \Pr\left(V^{\eta^1} \geq c(d+\eta^1)\right)$. Now, given the arrival rate n^2, consider the solution whose wait time is still η^1 and whose throughput is

$$x = n^2 \Pr\left(V^{\eta^1} \geq c(d+\eta^1)\right) = x^1_{\mathrm{opt}} \frac{n^2}{n^1}.$$

Given an arrival rate of n^2, we prove that the social welfare at the solution (η^1, x) is greater than or equal to the social welfare of any solution whose wait time $\eta' > \eta^1$. This implies that there exists a socially optimal solution at n^2 whose wait time is at most η^1, i.e., there exists $(x^2_{\mathrm{opt}}, \eta^2) \in \mathrm{opt}(n^2)$ such that $\eta^2 \leq \eta^1$. Further, since $x^2_{\mathrm{opt}} = n^2 \Pr\left(V^{\eta^2} \geq c(d+\eta^2)\right)$, we can also infer that $x^2_{\mathrm{opt}} \geq x^1_{\mathrm{opt}}$ since smaller wait times are more efficient (note that $c(d+\eta^2) \leq c(d+\eta^1)$).

For the remainder of this proof, let us fix n^2 as the arrival rate. Consider the solution (x, η^1) and for the purpose of comparison, another feasible solution (x', η') such that $\eta' > \eta^1$. Without loss of generality it is sufficient to consider $x' = n^2 \Pr\left(V^{\eta'} \geq c(d+\eta')\right)$. We now prove that $\mathsf{SW}_{n^2}(x, \eta^1) \geq \mathsf{SW}_{n^2}(x', \eta')$, where the subscript indicates the arrival rate. First, we have:

$$\mathsf{SW}_{n^2}(x, \eta^1) = n^2 \mathbb{E}\left[V^{\eta^1} \cdot \mathbb{1}_{\{V^{\eta^1} \geq c(d+\eta^1)\}}\right] - c \cdot \left(x_{\mathrm{opt}}(d+\eta^2) + (\tau/\eta^1)^{\frac{1}{\alpha}}\right).$$

For the sake of brevity, let us use Z^η as shorthand for the condition $V^\eta \geq c(d+\eta)$. Then, we can rearrange the above expression as follows:

$$\begin{aligned}\mathsf{SW}_{n^2}(x,\eta^1) &= n^2\mathbb{E}\left[V^{\eta^1} \mid Z^{\eta^1}\right]\Pr(Z^{\eta^1}) - cn^2\Pr(Z^{\eta^1})(d+\eta^1) - c(\tau/\eta^1)^{\frac{1}{\alpha}} \\ &= n^2\Pr(Z^{\eta^1})\left(\mathbb{E}\left[V^{\eta^1} \mid Z^{\eta^1}\right] - c(d+\eta^1)\right) - c(\tau/\eta^1)^{\frac{1}{\alpha}} \\ &= n^2\Pr(Z^{\eta^1})\left(\mathbb{E}\left[V^{\eta^1} - c(d+\eta^1) \mid Z^{\eta^1}\right]\right) - c(\tau/\eta^1)^{\frac{1}{\alpha}}. \end{aligned} \tag{23}$$

Once again for brevity, define $\Delta_{n_2} = \mathsf{SW}_{n^2}(x,\eta^1) - \mathsf{SW}_{n^2}(x',\eta')$. Then, using Equation (23), we can write the difference in welfare as:

$$\begin{aligned}\Delta_{n_2} &= n^2\big(\Pr(Z^{\eta^1})\mathbb{E}[V^{\eta^1} - c(d+\eta^1) \mid Z^{\eta^1}] - \Pr(Z^{\eta'})\mathbb{E}[V^{\eta'} - c(d+\eta') \mid Z^{\eta'}]\big) - c\big((\frac{\tau}{\eta^1})^{\frac{1}{\alpha}} - (\frac{\tau}{\eta'})^{\frac{1}{\alpha}}\big) \\ &\geq n^1\big(\Pr(Z^{\eta^1})\mathbb{E}[V^{\eta^1} - c(d+\eta^1) \mid Z^{\eta^1}] - \Pr(Z^{\eta'})\mathbb{E}[V^{\eta'} - c(d+\eta') \mid Z^{\eta'}]\big) - c\big((\frac{\tau}{\eta^1})^{\frac{1}{\alpha}} - (\frac{\tau}{\eta'})^{\frac{1}{\alpha}}\big) \end{aligned} \tag{24}$$

$$\begin{aligned} &= \mathsf{SW}_{n^1}(x^1_{\text{opt}},\eta^1) - \mathsf{SW}_{n^1}(\frac{n_1}{n_2}x',\eta') \\ &\geq 0. \end{aligned}$$

The final step is due to the fact that $(x^1_{\text{opt}},\eta^1) \in \text{opt}(n^1)$ due to which the social welfare of this solution, i.e., $\mathsf{SW}_{n^1}(x^1_{\text{opt}},\eta^1)$ must be at least the welfare of any other solution at arrival rate n^1, including $\mathsf{SW}_{n^1}(\frac{n_1}{n_2}x',\eta')$. More importantly, the crux of this proof lies in Equation 24, which stems from the fact that:

$$\Pr(Z^{\eta^1})\mathbb{E}[V^{\eta^1} - c(d+\eta^1) \mid Z^{\eta^1}] - \Pr(Z^{\eta'})\mathbb{E}[V^{\eta'} - c(d+\eta') \mid Z^{\eta'}] \geq 0.$$

To see why this is true, we note that since $\mathsf{SW}_{n^1}(x^1_{\text{opt}},\eta^1) - \mathsf{SW}_{n^1}(\frac{n_1}{n_2}x',\eta') \geq 0$, this in turn implies that:

$$n^1\big(\Pr(Z^{\eta^1})\mathbb{E}[V^{\eta^1} - c(d+\eta^1) \mid Z^{\eta^1}] - \Pr(Z^{\eta'})\mathbb{E}[V^{\eta'} - c(d+\eta') \mid Z^{\eta'}]\big) \geq c\big((\frac{\tau}{\eta^1})^{\frac{1}{\alpha}} - (\frac{\tau}{\eta'})^{\frac{1}{\alpha}}\big).$$

Since $\eta^1 < \eta'$, the right hand side of the above expression must be strictly positive. This in turn implies that the left hand side must also be strictly positive. Therefore, we have just proved that if the solution $(x^1_{\text{opt}},\eta^1)$ is socially optimal at arrival rate n^1, then for any $\eta' > \eta^1$, the best social welfare at wait time η' cannot be more than the best social welfare at η^1 for a larger arrival rate n^2.

This concludes our proof of the proposition. $\square$

E Simple Subscriptions in the Presence of Heterogeneous Riders

In this section we reason about the welfare that simple subscriptions can achieve in the presence of heterogeneous riders. We show that while these subscriptions are capable of attaining first best welfare in the case of rider homogeneity, this is not always the case for heterogeneous riders.

We adopt the model for heterogeneous riders of Section 4, namely, we assume to have *high type riders* and *low type riders* (see Section 4 for an elaborate presentation of this model). In Proposition 6 below we present a condition on this model under which the first best welfare is attainable by simple subscriptions in equilibrium. Then we show an instance for which the condition does not hold, thus proving Proposition 2.

We begin by formally defining an equilibrium in our model. For a simple subscription S, and for $i \in \{h, \ell\}$, we denote by n_i^S the arrival rate of type i riders who subscribe to S and we denote by n_i^R the arrival rate of type i riders who choose to participate in the real time market only — these satisfy $n_i^S + n_i^R = n_i$. Furthermore, x_i^S, x_i^R denote the respective throughputs. In the definition below, δ^R is interpreted as 0 (*i.e*, there is zero discount for non-subscribers).

Definition 6. A tuple $\left(\left(n_i^S, n_i^R, x_i^S, x_i^R\right)_{i\in\{h,\ell\}}, y, p, \eta\right)$ is an *equilibrium outcome for h, ℓ type riders under the simple subscription $S = (s, \delta, T)$* if the following conditions hold:

- Rider best response at subscription stage:

$$\forall i \in \{h, \ell\}, C \in \{S, R\} : n_i^C > 0 \implies \forall C' \in \{S, R\} : \quad u_i^C(p, \eta) \geq u_i^{C'}(p, \eta).$$

- Rider best response in the real time market stage:

$$\forall i \in \{h, \ell\}, C \in \{S, R\} : x_i^C = n_i^C \Pr(V_i^\eta \geq p - \delta^C).$$

- Driver supply level:

$$y = (\sum_{i \in \{h,\ell\}, C \in \{S,R\}} x_i^C)(d+\eta) + O(\eta).$$

- Total rider payments equals total driver costs:

$$c \cdot y = (n_h^S + n_\ell^S)\frac{s}{T} + \sum_{i \in \{h,\ell\}, C \in \{S,R\}} x_i^C (p - \delta^C)$$

Proposition 6. Let η be a wait time such that $\mathsf{optSW}(\eta) > 0$. The following condition is necessary and sufficient for the existence of a simple subscription S and a corresponding equilibrium outcome that achieves first-best welfare under η:

$$Q_\ell \cdot r^\eta - \frac{n_h Q_h + n_\ell Q_\ell}{n} c w_{\text{opt}}(\eta) \geq 0$$

Remark Note that $\frac{n_h Q_h + n_\ell Q_\ell}{n}$ is the average probability of being interested in a trip. Furthermore, $\mathsf{optSW}(\eta)$ implies by Observation 1 that $r^\eta \geq c w_{\text{opt}}(\eta)$. Thus, Proposition 6 claims intuitively that simple subscriptions can achieve first best welfare in equilibrium as long as the probabilities for being interested in a trip, Q_h and Q_ℓ, are not too different.

Proof. We will need the following lemma that shows that in order to achieve first best welfare in equilibrium, all riders must subscribe.

Lemma 6. Let η satisfy $\mathsf{optSW}(\eta) > 0$, and let $\left((n_i^S, n_i^R, x_i^S, x_i^R)_{i \in \{h,\ell\}}, y, p, \eta\right)$ be an equilibrium outcome under the simple subscription $S = (s, \delta^S, T)$ that achieves first best welfare under η. Then all riders subscribe to S in this outcome. In other words, for both $i \in \{h, \ell\}$ we have $n_i^S = n_i$ and $n_i^R = 0$.

Proof. To prove the lemma, we shall need the following claim:

Claim 1. In the given equilibrium outcome, all riders face the same effective real time trip price which equals $c(d + \eta)$.

We shall now show that the claim implies the lemma, and then we prove the claim. The claim implies that either all riders subscribe, or all riders do not subscribe (since subscribers and non-subscribers face a different effective real time price). Thus we need to show that that latter is impossible. If all riders do not subscribe then in particular the subscriber throughput is $x_i^C = 0$ for both $i \in \{h, \ell\}$. Driver supply level then requires that

$$y = (x_h^R + x_\ell^R)(d + \eta) + O(\eta)$$

implying (since total rider payments equal total driver costs) that

$$c \cdot ((x_h^R + x_\ell^R)(d + \eta) + O(\eta)) = c \cdot y = (x_h^R + x_\ell^R)p.$$

Since all riders do not subscribe, the effective real time price they face is p, which equals $c(d + \eta)$ by the claim. Plugging that in the above equation, we get $c \cdot O(\eta) = 0$, a contradiction.

Proof of Claim 1. By Lemma 1 we have

$$\begin{aligned}\mathsf{optSW}(\eta) &= n_h \cdot \mathbb{E}\left[V_h^\eta \cdot \mathbb{1}_{\{V_h^\eta \geq c(d+\eta)\}}\right] - n_\ell \cdot \mathbb{E}\left[V_\ell^\eta \cdot \mathbb{1}_{\{V_\ell^\eta \geq c(d+\eta)\}}\right] - cy_{\text{opt}} \\ &= (n_h Q_h + n_\ell Q_\ell) \cdot \mathbb{E}\left[V^\eta \cdot \mathbb{1}_{\{V^\eta \geq c(d+\eta)\}}\right] - cy_{\text{opt}}\end{aligned}$$

where y_{opt} is the optimal driver supply level. In particular, we must have

$$\mathbb{E}\left[V^\eta \cdot \mathbb{1}_{\{V^\eta \geq c(d+\eta)\}}\right] > 0$$

implying that $\Pr(V^\eta \geq c(d + \eta)) > 0$. Now, let us assume that a number $\tilde{n} > 0$ of riders face an effective real time price $p > c(d + \eta)$ (the proof for the case $p < c(d + \eta)$ is analogous). Since $\Pr(V^\eta \geq c(d+\eta)) > 0$, Assumption 1 implies that there is some number $c(d+\eta) \leq g < p$ such that $\Pr(V^\eta \in [c(d + \eta), g]) > 0$. It follows that there are $\tilde{n} \Pr(V^\eta \in [c(d + \eta), g]) > 0$ riders with value at least $c(d + \eta)$ who do not take a trip when faced with the price p. Consequently, by Lemma 1, the outcome does not attain first-best welfare. □

□

Let η be a wait time such that $\mathsf{optSW}(\eta) > 0$. We first show that the condition in the proposition statement is necessary. Let $S = (s, \delta, T)$ be a simple subscription that achieves first best welfare under η in an equilibrium outcome $\left(\left(n_i^S, n_i^R, x_i^S, x_i^R\right)_{i\in\{h,\ell\}}, y, p, \eta\right)$. By Lemma 6, all riders subscribe in this outcome, *i.e.*, $n_i^S = n_i$ and $n_i^R = 0$ for both $i \in \{h, \ell\}$, and this in turn implies (due to rider best response in the real time market stage) that

$$x_i^S = n_i^S \Pr(V_i^\eta \geq p - \delta) = n_i Q_i \Pr(V^\eta \geq p - \delta)$$

By Claim 1 the discounted price satisfies $p - \delta = c(d + \eta)$, and this in turn implies (due to the requirements on driver supply level and that total rider payments equals total driver costs):

$$\begin{aligned}(n_h + n_\ell)\frac{s}{T} + (x_h^S + x_\ell^S) \cdot c(d + \eta) &= c \cdot y \\ &= c((x_h^S + x_\ell^S)(d + \eta) + O(\eta))\end{aligned}$$

Solving the above for s gives us

$$\begin{aligned}s &= \frac{cTO(\eta)}{n_h + n_\ell} \\ &= cTw_{\text{opt}}\frac{x_{\text{opt}}}{n}\cdot \\ &= cTw_{\text{opt}}\frac{(n_h Q_h + n_\ell Q_\ell)}{n} \Pr(V \geq c(d + \eta))\end{aligned}$$

Finally, the low type riders' best response at the subscription stage gives us:

$$\begin{aligned}0 &\leq u_\ell^R(p, \eta) \\ &\leq u_\ell^S(p, \eta) \\ &= \mathbb{E}\left[(V_i^\eta - c(d + \eta))\mathbb{1}_{\{V_i^\eta \geq c(d+\eta))\}}\right] - \frac{s}{T} \\ &= Q_\ell \mathbb{E}\left[V^\eta - c(d + \eta)\mathbb{1}_{\{V^\eta \geq p - \delta\}}\right] - \frac{s}{T} \\ &= Q_\ell r^\eta \Pr(V^\eta \geq c(d + \eta)) - \frac{s}{T} \\ &= Q_\ell r^\eta \Pr(V^\eta \geq c(d + \eta)) - cw_{\text{opt}}\frac{(n_h Q_h + n_\ell Q_\ell)}{n} \Pr(V \geq c(d + \eta)) \\ &= \Pr(V \geq c(d + \eta))\left(Q_\ell r^\eta - cw_{\text{opt}}\frac{(n_h Q_h + n_\ell Q_\ell)}{n}\right)\end{aligned}$$

and the required condition follows since $\Pr(V \geq c(d+\eta)) > 0$ (which in turn follows from Lemma 1 and the fact that $\mathsf{optSW}(\eta) > 0$).

To see that the condition is sufficient, we can choose the desired simple subscription and equilibrium outcome with the same parameters given by the analysis above. That is, we choose some maximum duration T, and we also set:

$$
\begin{aligned}
s &= cTw_{\text{opt}}\frac{(n_hQ_h + n_\ell Q_\ell)}{n}\Pr(V \geq c(d+\eta)),\\
\forall i \in \{h,\ell\} : n_i^S &= n_i,\\
\forall i \in \{h,\ell\} : n_i^R &= 0,\\
\forall i \in \{h,\ell\} : x_i^S &= n_iQ_i\Pr(V^\eta \geq c(d+\eta)),\\
\forall i \in \{h,\ell\} : x_i^R &= 0,\\
y &= (x_h^S + x_\ell^S)(d+\eta) + O(\eta),
\end{aligned}
$$

and, as we did in the proof of Theorem 2, we set p and δ high enough so that $p - \delta = c(d+\eta)$ but $u_i^R(p,\eta) = 0$.

All requirements are given for free by the above analysis, except for the riders' best response at the subscription stage. To see that this final requirement also holds, note that $u_\ell^S(p,\eta) \geq 0$ due to the assumed condition. In order to show that $u_h^S(p,\eta) \geq 0$, an analogous calculation shows that

$$
u_h^S(p,\eta) = \Pr(V \geq c(d+\eta))\left(Q_hr^\eta - cw_{\text{opt}}\frac{(n_hQ_h + n_\ell Q_\ell)}{n}\right)
$$

and the fact that this is non-negative follows immediately from the assumed condition and the fact that $Q_h > Q_\ell$.

□

We end this section by proving Proposition 2. To this end, we set $n_h = n_\ell = n/2$ for some $n > 0$, $Q_h = 0.8, Q_\ell = 0.2$ and we choose any distribution V^η for which $\Pr(V \geq c(d+\eta)) = 0.5$, and $r^\eta = cw_{\text{opt}} + \epsilon$ for some small enough ϵ.

We need to show that for this instance we have $\mathsf{optSW}(\eta) > 0$ and that the condition from Proposition 6 does not hold. The former holds by Observation 1 and since $x_{\text{opt}} = (n_hQ_h +$

$n_\ell Q_\ell) \Pr(V^\eta \geq c(d+\eta)) > 0$, and the latter is given below:

$$
\begin{aligned}
Q_\ell r^\eta - cw_{\text{opt}} \frac{(n_h Q_h + n_\ell Q_\ell)}{n} &= 0.2 \cdot (cw_{\text{opt}} + \epsilon) - cw_{\text{opt}} \frac{0.8 \cdot n/2 + 0.2 \cdot n/2}{n} \\
&= 0.2 \cdot (cw_{\text{opt}} + \epsilon) - 0.5 \cdot cw_{\text{opt}} \\
&< 0
\end{aligned}
$$

F Omitted Proofs

F.1 Proofs from Section 2

F.1.1 Proof of Theorem 1

Proof. Let $x_{\text{opt}}, x_{\text{wal}}$ be the throughputs corresponding to the first best and best Walrasian outcomes under η, respectively. Let $w_{\text{wal}} = O(\eta)/x_{\text{wal}}$ be the idle driver time in the Walrasian outcome. By Lemma 1 and Definition 1 we have

$$
\begin{aligned}
x_{\text{opt}} &= n \cdot \Pr(V^\eta \geq c(d+\eta)). \\
x_{\text{wal}} &= n \cdot \Pr(V^\eta \geq c(d+\eta+\tilde{w})).
\end{aligned}
$$

From here we have:

$$
\begin{aligned}
&\mathsf{optSW}(\eta) - \mathsf{walSW}(\eta) \\
=&\left(n \cdot \mathbb{E}\left[V^\eta \cdot \mathbb{1}_{\{V^\eta \geq c(d+\eta)\}}\right] - c \cdot (x_{\text{opt}}(d+\eta) + O)\right) \\
&\quad - \left(n \cdot \mathbb{E}\left[V^\eta \cdot \mathbb{1}_{\{V^\eta \geq c(d+\eta+w_{\text{wal}})\}}\right] - c \cdot (x_{\text{wal}}(d+\eta) + O)\right) \\
=&n \cdot \mathbb{E}\left[V^\eta \cdot \left(\mathbb{1}_{\{V^\eta \geq c(d+\eta)\}} - \mathbb{1}_{\{V^\eta \geq c(d+\eta+w_{\text{wal}})\}}\right)\right] - (x_{\text{opt}} - x_{\text{wal}})c(d+\eta) \\
=&n \cdot \mathbb{E}\left[V^\eta \cdot \mathbb{1}_{\{c(d+\eta+w_{\text{wal}}) > V^\eta \geq c(d+\eta)\}}\right] - (x_{\text{opt}} - x_{\text{wal}})c(d+\eta) \\
>&n \cdot c \cdot (d+\eta) \Pr\left(c(d+\eta+w_{\text{wal}}) > V^\eta \geq c(d+\eta)\right) - (x_{\text{opt}} - x_{\text{wal}})c(d+\eta) \\
=&n \cdot c \cdot (d+\eta) \left(\Pr\left(V^\eta \geq c(d+\eta)\right) - \Pr\left(V^\eta \geq c(d+\eta+w_{\text{wal}})\right)\right) - (x_{\text{opt}} - x_{\text{wal}})c(d+\eta) \\
=&(x_{\text{opt}} - x_{\text{wal}})c(d+\eta) - (x_{\text{opt}} - x_{\text{wal}})c(d+\eta)
\end{aligned}
$$

$=0$

where the inequality holds by Lemma 3.

∎

F.2 Proofs from Section 3

F.2.1 Proof of Equation 25

Proof.

$$
\begin{aligned}
\mathsf{optSW}(\eta) &= n \cdot \mathbb{E}\left[V^\eta \cdot \mathbb{1}_{\{V^\eta \geq c(d+\eta)\}}\right] - c \cdot (x_{\text{opt}}(d+\eta) + O(\eta)) \\
&= n \cdot \mathbb{E}\left[V^\eta \cdot \mathbb{1}_{\{V^\eta \geq c(d+\eta)\}}\right] - c x_{\text{opt}} \left(d + \eta + w_{\text{opt}}\right) \\
&= n \cdot \left(\mathbb{E}\left[V^\eta \cdot \mathbb{1}_{\{V^\eta \geq c(d+\eta)\}}\right] - c \frac{x_{\text{opt}}}{n} \left(d + \eta + w_{\text{opt}}\right)\right) \\
&= n \cdot \left(\mathbb{E}\left[V^\eta \cdot \mathbb{1}_{\{V^\eta \geq c(d+\eta)\}}\right] - c \Pr(V^\eta \geq c(d+\eta)) \left(d + \eta + w_{\text{opt}}\right)\right) \\
&= n \cdot \left(\mathbb{E}\left[V^\eta \cdot \mathbb{1}_{\{V^\eta \geq c(d+\eta)\}}\right] - c \left((d+\eta)\mathbb{E}\left[\mathbb{1}_{\{V^\eta \geq c(d+\eta)\}}\right] - w_{\text{opt}} \Pr(V^\eta \geq c(d+\eta))\right)\right) \\
&= n \cdot \left(\mathbb{E}\left[(V^\eta - c(d+\eta)) \cdot \mathbb{1}_{\{V^\eta \geq c(d+\eta)\}}\right] - \frac{1}{T} \cdot T c w_{\text{opt}} \Pr(V^\eta \geq c(d+\eta))\right) \\
&= n \cdot \left(\mathbb{E}\left[(V^\eta - c(d+\eta)) \cdot \mathbb{1}_{\{V^\eta \geq c(d+\eta)\}}\right] - \frac{s}{T}\right) \\
&= n \cdot u^S(p, \eta)
\end{aligned}
$$

∎

F.2.2 Proof of Theorem 2

Proof. Let η be as in the theorem statement. In order to achieve first best under waiting time η, it is enough to have riders face an effective real time trip price of $c(d+\eta)$ (see Lemma 1). Thus, we require $p - \delta = c(d+\eta)$. Assuming all riders subscribe, the resulting throughput is $x_{\text{opt}} = n \cdot \Pr(V^\eta \geq c(d+\eta))$. Under this throughput, the driver idle time is $w_{\text{opt}} = O(\eta)/x_{\text{opt}}$, implying that the driver cost per trip is $c(d+\eta+w_{\text{opt}})$. Assuming that drivers get paid exactly their cost, every rider who takes a trip gets an effective discount of $c w_{\text{opt}}$. Thus, assuming all n riders subscribe, the fee s must satisfy $n \cdot s/T = c w_{\text{opt}} x_{\text{opt}}$ in order to cover the discounts. We thus set $s = Tcw_{\text{opt}} \cdot x_{\text{opt}}/n = Tcw_{\text{opt}} \Pr(V^\eta \geq c(d+\eta))$.

To summarize, the necessary conditions for a simple subscription to achieve first best welfare under η in an equilibrium outcome in which all riders subscribe, are: (i) $p - \delta = c(d+\eta)$ and (ii) $s = Tcw_{\text{opt}} \Pr(V^\eta \geq c(d+\eta))$.

It is left to show that we can set p in a way that also pushes the riders to subscribe while still satisfying the above conditions (note that such a p determines δ). To see this, we first argue that under the above conditions, $u^S(p,\eta)$ is strictly positive. This follows from the fact that under the above conditions drivers get zero surplus, implying that the first best welfare (which is strictly positive) equals the total rider surplus. The claim then follows from rider homogeneity, which implies that all riders get the same surplus. Formally, we can show the following relation, whose proof is deferred to the appendix:

$$\mathsf{optSW}(\eta) = n \cdot u^S(p,\eta). \tag{25}$$

In particular, $u^S(p,\eta) > 0$ as desired. Finally, note that $u^R(p,\eta)$ is monotonically decreasing in p, and there are p (*e.g.*, $p = v_{\max}$) for which $u^R(s,\eta) = 0$. Thus there is some threshold price for which riders are indifferent between subscribing and not. For any price p above this threshold, we are guaranteed that all riders subscribe, and the outcome still achieves first-best welfare in equilibrium as long as the difference $p - \delta$ is fixed at $c(d+\eta)$. In fact, in Section 5 we show that the Walrasian price under η is above this threshold.

□

F.2.3 Proof of Proposition 1

Proof. As shown in the proof of Theorem 2, a simple subscription $S = (s, \delta, T)$ that achieves the first best outcome in an equilibrium outcome $(n_0, n_1, x_0, x_1, y, p, \eta)$ must satisfy:

- Every driver gets paid $c(d + \eta + w_{\text{opt}})$ per trip.
- $s = T \cdot cw_{\text{opt}} \Pr(V^\eta \geq c(d+\eta))$.
- $p - \delta = c(d+\eta)$.

We show that if we set $p = c(d + \eta + w_{\text{opt}})$, then riders strictly prefer to participate in the real time market only. Consequently, the equilibrium outcome is in effect a Walrasian equilibrium, which we know does not achieve first best welfare by Theorem 1. Note that under the above choice of p, we have $\delta = cw_{\text{opt}}$. We analyze $u^S(p, \eta)$ under this p:

$$\begin{aligned}
u^S(p, \eta) &= \mathbb{E}\left[(V^\eta - (p - \delta))\mathbb{1}_{\{V^\eta \geq p-\delta\}}\right] - cw_{\text{opt}} \Pr(V^\eta \geq c(d + \eta)) \\
&= \mathbb{E}\left[(V^\eta - (p - \delta))\mathbb{1}_{\{V^\eta \geq p-\delta\}}\right] - \delta \Pr(V^\eta \geq c(d + \eta)) \\
&= \mathbb{E}\left[(V^\eta - (p - \delta))\mathbb{1}_{\{V^\eta \geq p-\delta\}}\right] - \mathbb{E}\left[\delta \mathbb{1}_{\{V^\eta \geq p-\delta\}}\right] \\
&= \mathbb{E}\left[(V^\eta - p)\mathbb{1}_{\{V^\eta \geq p-\delta\}}\right]
\end{aligned}$$

On the other, the utility from participating in the real time market only is

$$u^R(p, \eta) = \mathbb{E}\left[(V^\eta - p)\mathbb{1}_{\{V^\eta \geq p\}}\right].$$

Thus we get:

$$\begin{aligned}
u^S(p, \eta) - u^R(p, \eta) &= \mathbb{E}\left[(V^\eta - p)\mathbb{1}_{\{p-\delta \leq V^\eta < p\}}\right] \\
&= \mathbb{E}\left[V^\eta \mathbb{1}_{\{p-\delta \leq V^\eta < p\}}\right] - p \Pr(p - \delta \leq V^\eta < p) \\
&< p \Pr(p - \delta \leq V^\eta < p) - p \Pr(p - \delta \leq V^\eta < p) \\
&= 0
\end{aligned}$$

where the inequality holds by Lemma 3. Thus, when p equals the rider payment per trip, riders prefer not to subscribe.

We finish by recalling that for any p, δ such that $p - \delta = c(d + \eta)$, we have $u^S(p, \eta) > 0$ (see proof of Theorem 2). In particular this implies that if the platform restricts real-time access only to subscribers, then all riders will still subscribe. □

F.3 Proofs from Section 4

F.3.1 Proof of Observation 1

Proof. By Lemma 1, the first best outcome is achieved when the riders who take a trip are exactly those whose value is weakly above $c(d+\eta)$. Thus we have

$$
\begin{aligned}
\mathsf{optSW}(\eta) &= n_h \cdot \mathbb{E}\left[V_h^\eta \cdot \mathbb{1}_{\{V_h^\eta \geq c(d+\eta)\}}\right] + n_\ell \cdot \mathbb{E}\left[V_\ell^\eta \cdot \mathbb{1}_{\{V_\ell^\eta \geq c(d+\eta)\}}\right] - c \cdot y_{\text{opt}} \\
&= (n_h Q_h + n_\ell Q_\ell)\mathbb{E}\left[V^\eta \cdot \mathbb{1}_{\{V^\eta \geq c(d+\eta)\}}\right] - c \cdot y_{\text{opt}} \\
&= (n_h Q_h + n_\ell Q_\ell)\Pr(V^\eta \geq c(d+\eta))\mathbb{E}\left[V^\eta \mid V^\eta \geq c(d+\eta)\right] - c \cdot y_{\text{opt}} \\
&= (n_h \Pr(V_h^\eta \geq c(d+\eta)) + n_\ell \Pr(V_\ell^\eta \geq c(d+\eta)))\mathbb{E}\left[V^\eta \mid V^\eta \geq c(d+\eta)\right] - c \cdot (x_{\text{opt}}(d+\eta) + O(\eta)) \\
&= x_{\text{opt}}\mathbb{E}\left[V^\eta \mid V^\eta \geq c(d+\eta)\right] - c \cdot (x_{\text{opt}}(d+\eta) + O(\eta)) \\
&= x_{\text{opt}}(\mathbb{E}\left[V^\eta \mid V^\eta \geq c(d+\eta)\right] - c(d+\eta) - cO(\eta)/x_{\text{opt}}) \\
&= x_{\text{opt}}(r^\eta - c w_{\text{opt}})
\end{aligned}
$$

□

F.3.2 Proof of Lemma 2

Proof. Denote the event that the type i rider renews C on day j by A_j. We first assume that $k = T$. In this case $D_i^C(p,\eta) = T$ with probability 1, and in particular $\mathbb{E}\left[D_i^C(p,\eta)\right] = T$. Furthermore, the rider renews the discount every T days in this case, implying that

$$
\lim_{t \to \infty} \frac{\sum_{j=1}^t \mathbb{1}_{\{A_j\}}}{t} = \frac{1}{T}
$$

with probability 1, and the claim follows. It is left to prove the claim assuming that $k < T$, and to this end we use the ergodic theorem for Markov chains.

Assuming immediate renewal once the current plan expires, we can associate the type i rider with a stochastic process, *i.e.*, a sequence of random variables $\{X_j\}_{j=1}^\infty$, where X_j is the *state* of the rider's subscription plan on the j'th day. Each state is a pair (k', T') where k' is the number of unused discounts left, and T' is the number of days left until expiry.

The *state space*, denoted by Ω_i, is the set of states that are reachable with positive probability from the state corresponding to the first day of the subscription cycle (k, T), and M_i denotes the sequence $X_1, X_2, \ldots$. This is Markovian and M_i is a *finite state Markov-Chain.* For example, if on the j'th day the rider has 4 discounts left and 10 days left until expiry, then X_{j+1} can be one of two states: either 3 discounts left and 9 days left until expiry, or 4 discounts left and 9 days left until expiry. The process transitions to the former state with probability $\Pr(V_i^\eta \geq p - \delta)$ and it transitions to the latter with probability $1 - \Pr(V_i^\eta \geq p - \delta)$. Note that the state corresponding to the first day of a subscription cycle, namely (k, T), also corresponds to a day in which the rider pays the subscription fee (and this is the only such state).

Now, by definition M_i is an irreducible Markov chain, and as such it admits a unique stationary distribution π_i. We apply the ergodic theorem for Markov Chains (see Theorem C.1 in [Levin and Peres, 2017]), to obtain

$$\lim_{t \to \infty} \frac{\sum_{j=1}^{t} \mathbb{1}_{\{A_j\}}}{t} = \pi_i(k, T) \tag{26}$$

with probability 1. Furthermore, since π_i is stationary, $\pi_i(k, T)$ is the inverse of the expected hitting time of the state (k, T), when starting from (k, T) (see Theorem 3.3 in [Freedman, 2017]). But this hitting time is exactly $D_i^C(p, \eta)$, implying that

$$\pi_i(k, T) = \frac{1}{\mathbb{E}\left[D_i^C(p, \eta)\right]}$$

Finally, since (26) holds with probability 1, then in particular we have

$$\mathbb{E}\left[\lim_{t \to \infty} \frac{\sum_{j=1}^{t} \mathbb{1}_{\{A_j\}}}{t}\right] = \frac{1}{\mathbb{E}\left[D_i^C(p, \eta)\right]}$$

as desired.

□

F.3.3 Proof of Theorem 3

Proof. Let η be a wait time for which $\mathsf{optSW}(\eta) > 0$. We start by showing that first best welfare can also be attained by a single flexible subscription plan in an all-subscribe equilibrium.

Lemma 7. For any given k, there exists a flexible subscription $C = (s, \delta, k, T)$ that attains the first best welfare under η in an equilibrium outcome in which all riders (both high and low type) subscribe to C.

Proof. Let k be any desirable amount of discounts. We describe below how we construct the flexible subscription $C = (s, \delta, k, T)$ along with the desired equilibrium outcome

$$((n_i^C, x_i^C)_{i\in\{h,\ell\}}, (n_i^R, x_i^R)_{i\in\{h,\ell\}}, y, p, \eta)$$

that achieves first best welfare. Here n_i^R and x_i^R correspond to the arrival rate and throughput of non-subscribers, respectively. We would like all riders to subscribe, thus we set, for $i \in \{h, \ell\}$:

$$n_i^C = n_i, \quad n_i^R = 0$$

In order to achieve first-best welfare, the riders who take a trip at the real time market stage must be exactly those with value greater or equal to $c(d + \eta)$ (by Lemma 1). Since all riders subscribe (this assumption shall be justified later), this implies that the discounted trip price that every rider faces must satisfy $p - \delta = c(d + \eta)$. Fix such p and δ (as in the proof of Theorem 2, in the end of the proof these shall be set high enough while keeping $p - \delta$ fixed to ensure that riders indeed prefer subscribing over not subscribing).

In order to satisfy the requirement of rider best response at the real time market stage, the corresponding throughput x_i^C for $i \in \{h, \ell\}$ must satisfy:

$$x_i^C = n_i \Pr(V_i^\eta \geq c(d + \eta)) = n_i Q_i \Pr(V^\eta \geq c(d + \eta)).$$

Furthermore, the driver supply level needed to sustain the above throughput must satisfy

$$y = (x_h^C + x_\ell^C)(d + \eta) + O(\eta).$$

Consequently, in order to have total rider payment equal total driver cost, we must have:

$$\begin{aligned}
c \cdot \left(\left(x_h^C + x_\ell^C \right) (d+\eta) + O(\eta) \right) &= c \cdot y \\
&= s \cdot \left(\frac{n_h^C}{\mathbb{E}\left[D_h^C(p,\eta) \right]} + \frac{n_\ell^C}{\mathbb{E}\left[D_\ell^C(p,\eta) \right]} \right) + (x_h^C + x_\ell^C)(p-\delta) \\
&= s \cdot \left(\frac{n_h^C}{\mathbb{E}\left[D_h^C(p,\eta) \right]} + \frac{n_\ell^C}{\mathbb{E}\left[D_\ell^C(p,\eta) \right]} \right) + (x_h^C + x_\ell^C)c(d+\eta)
\end{aligned}$$

which simplifies to

$$s = \frac{cO(\eta)}{\frac{n_h}{\mathbb{E}\left[D_h^C(p,\eta)\right]} + \frac{n_\ell}{\mathbb{E}\left[D_\ell^C(p,\eta)\right]}}. \tag{27}$$

Note that the terms $\mathbb{E}\left[D_i^C(p,\eta)\right]$ depend also on the maximum duration of the subscription, T (but not on s), which is the only parameter that is yet to be defined. In order to set T, we use the following lemma, which captures the intuition that as T grows larger, the effective duration of the subscription period for a type i rider is approximated by a sum of k geometric random variables with parameter $\Pr(V_i^\eta \geq c(d+\eta))$.

Lemma 8. For any T, denote the corresponding flexible subscription by $C(T) = (s, \delta, k, T)$. Then:

$$\lim_{T\to\infty} \mathbb{E}\left[D_i^{C(T)}(p,\eta)\right] = \frac{k}{\Pr(V_i^\eta \geq c(d+\eta))}.$$

Proof. In the following we abuse notation and write D instead of $D_i^{C(T)}(p,\eta)$. We further denote $q = \Pr(V_i^\eta \geq c(d+\eta))$.

Recall that a type i rider takes a trip each day with probability q, and that the subscription plan $C(T)$ provides a rider with k discounts she can use before T days have passed. Thus, if the subscription plan had no expiry date, then D would simply have been a sum of k i.i.d geometric random variable with parameter q.

Let G_i, $i = 1, \ldots, k$ be such geometric random variables. Then we can write D as $D = \min\{\sum_{i=1}^k G_i, T\}$. Let $S(T) = \sum_{i=1}^k G_i - D$, the amount of days "cut off" due to D not really being a sum of geometric random variables. Then

$$\mathbb{E}\left[D\right] = \mathbb{E}\left[\sum_{i=1}^{k} G_i\right] - \mathbb{E}\left[S\right] = \frac{k}{q} - E\left[S(T)\right].$$

Therefore, it is enough to prove that $\lim_{T\to\infty} \mathbb{E}\left[S(T)\right] = 0$.

To this end, we split to cases according to how many discounts were actually applied before expiry. For any $j = 0, \ldots, k$, let A_j be the event that only the first j discounts were applied by the expiry date. In other words, A_j is the event that

$$\sum_{i=1}^{j} G_i \leq T \wedge \sum_{i=1}^{j+1} G_i > T$$

(where we vacously define $\sum_{i=1}^{0} G_i = 0$ and $\sum_{i=1}^{k+1} G_i = \infty$). In the event A_j, $S(T)$ accounts for the $(j+1)$'th discount whose application attempts started (weakly) before the expiry and ended after it, and the discounts $j+2, \ldots, k$ whose application attempts never even started before the expiry.

By the memorylessness property of the ($(j+1)$'th) geometric random variable, we have

$$\mathbb{E}\left[S(T) \mid A_j\right] = \frac{k-j}{q} \tag{28}$$

We now turn to analyze $\Pr(A_j)$. In order for A_j to occur, exactly j discounts have to be applied in T days. Thus we get

$$\Pr\left(A_j\right) = \binom{T}{j} (q)^j \left(1-q\right)^{T-j} \tag{29}$$

Combining (28) and (29) we get

$$\mathbb{E}\left[S(T)\right] = \sum_{j=0}^{k} \mathbb{E}\left[S(T) \mid A_j\right] \Pr\left(A_j\right) = \sum_{j=0}^{k} \binom{T}{j} (k-j) \left(q\right)^{j-1} \left(1-q\right)^{T-j}$$

and this expression indeed tends to 0 as $T \to \infty$ since $0 < q$, implying that for every j, the factor $(1-q)^{T-j}$ dominates the factor $\binom{T}{j}$.

$\square$

By Lemma 8, and equation (27) we have

$$
\begin{aligned}
\lim_{T\to\infty} s &= \frac{cO(\eta)}{\frac{n_h}{k/\Pr(V_h^\eta \geq c(d+\eta))} + \frac{n_\ell}{k/\Pr(V_\ell^\eta \geq c(d+\eta))}} \\
&= kcO(\eta)/\left(n_h \Pr(V_h^\eta \geq c(d+\eta)) + n_\ell \Pr(V_\ell^\eta \geq c(d+\eta))\right) \\
&= kcO(\eta)/\left(x_h^C + x_\ell^C\right) \\
&= kcO(\eta)/x_{\text{opt}}.
\end{aligned}
$$

Applying Lemma 8 again, we get:

$$
\begin{aligned}
\lim_{T\to\infty} u_i^{C(T)}(p,\eta) &= \lim_{T\to\infty} \mathbb{E}\left[(V_i^\eta - c(d+\eta))\mathbb{1}_{\{V_i^\eta \geq c(d+\eta)\}}\right] - \frac{s}{\mathbb{E}\left[D_i^{C(T)}(p)\right]} \\
&= Q_i \mathbb{E}\left[(V^\eta - c(d+\eta))\mathbb{1}_{\{V^\eta \geq c(d+\eta)\}}\right] - \frac{kcO(\eta)\Pr(V_i^\eta \geq c(d+\eta))}{kx_{\text{opt}}} \\
&= Q_i \Pr(V^\eta \geq c(d+\eta)) r^\eta - cw_{\text{opt}} \Pr(V_i^\eta \geq c(d+\eta)) \\
&= \Pr(V_i^\eta \geq c(d+\eta))(r^\eta - cw_{\text{opt}}) \\
&> 0
\end{aligned}
$$

where the first equality holds by Equation (9) (since $p-\delta = c(d+\eta)$), the third equality holds since $w_{\text{opt}} = O(w_{\text{opt}})/x_{\text{opt}}$ and the inequality holds by Observation 1 and since $\mathsf{optSW}(\eta) > 0$ (which implies that $\Pr(V_i^\eta \geq c(d+\eta)) = Q_i \Pr(V^\eta \geq c(d+\eta)) > 0$).

We conclude that there is some large enough $T > k$ such that $u_i^{C(T)}(p,\eta) > 0$ for both $i \in \{h,\ell\}$. In other words, for this T the subscription plan $C = C(T)$ is individually rational. We still need to show that the riders' best response at the subscription stage is indeed to subscribe to C. As we did in Section 3, this can be satisfied by setting the non-discounted real-time price p high enough, while keeping the discounted price $p-\delta$ fixed at $c(d+\eta)$. $\square$

In order to prove Theorem 3, we need to present two different subscription plans $H = \left(s^H, \delta^H, k^H, T^H\right)$ and $L = \left(s^L, \delta^L, k^L, T^L\right)$, and a corresponding equilibrium outcome

$$
\left(\left(n_i^C, x_i^C\right)_{i\in\{h,\ell\}, C\in\{H,L,R\}}, y, p, \eta\right)
$$

in which all the high type (low type) riders to subscribe to the H (L) plan (in the above, R represents the alternative of participating in the real time market only).

We show how to obtain these given the subscription plan $C = (s, \delta, k, T)$ and the equilibrium outcome denoted $((n_i^C, x_i^C)_{i\in\{h,\ell\}}, (n_i^R, x_i^R)_{i\in\{h,\ell\}}, y, p, \eta)$ that we have from Lemma 7. We essentially aim for the same equilibrium outcome. Formally, we set, for $i \in \{h, \ell\}$:

$$\begin{aligned}
n_\ell^L = n_\ell^C \quad &, \quad n_\ell^H = n_\ell^R = 0 \\
n_h^H = n_h^C \quad &, \quad n_h^L = n_h^R = 0 \\
x_\ell^L = x_\ell^C \quad &, \quad x_\ell^H = x_\ell^R = 0 \\
x_h^H = x_h^C \quad &, \quad x_h^L = x_h^R = 0
\end{aligned}$$

and we set the same driver supply level, real time price and wait time. In words, this is the same outcome as in Lemma 7, only that here we require all high type (low type) riders to subscribe to the H (L) plan instead of the single plan C. Furthermore we set $\delta^H = \delta^L = \delta$.

As in the proof of Lemma 7, setting these parameters for the outcome immediately implies rider best response at the real time market stage, correct driver supply level to sustain the trip throughput, and that the outcome achieves first-best welfare as desired.

We now turn to setting the parameters of the subscription plans, and show that under these, the outcome we defined also satisfies rider best response at subscription stage and that total rider payments equals total driver costs. We start by setting $L = C$, *i.e.*, L is the same plan we have from Lemma 7. Note that from the proof of Lemma 7 we can immediate conclude that for $i \in \{h, \ell\}$:

$$u_i^L(p, \eta) = u_i^C(p, \eta) > u_i^R(p, \eta)$$

We now set the remaining undefined parameters of H, namely s^H, k^H and T^H. First we set $k^H = T^H = T$, and note that in this case, H is a simple subscription, implying that

$$D_h^H(p, \eta) = D_\ell^H(p, \eta) = T \tag{30}$$

with probability 1.

It is left to set the subscription fee s^H. This is set in a way that makes a high type rider indifferent between subscribing to H or L. In other words, s^H is the solution to the equation $u_h^H(p,\eta) = u_h^L(p,\eta)$, which by Equation (9), the fact that $p - \delta^H = p - \delta^L = c(d+\eta)$ and equation (30), translates to

$$\mathbb{E}\left[(V_h^\eta - c(d+\eta))\mathbb{1}_{\{V_h^\eta \geq c(d+\eta)\}}\right] - \frac{s^H}{T} = \mathbb{E}\left[(V_h^\eta - c(d+\eta))\mathbb{1}_{\{V_h^\eta \geq c(d+\eta)\}}\right] - \frac{s^L}{\mathbb{E}\left[D_h^L(p,\eta)\right]}.$$

We thus set $s^H = \frac{s^L \cdot T}{\mathbb{E}\left[D_h^L(p,\eta)\right]}$, the solution to the equation. In order to obtain rider best response at the subscription stage under these parameters, it is left to show that the low type riders prefer subscribing to L over H. To this end, we have

$$\begin{aligned}
u_\ell^H(p,\eta) \leq u_\ell^L(p,\eta) & \quad\iff \\
\mathbb{E}\left[(V_\ell^\eta - c(d+\eta))\mathbb{1}_{\{V_\ell^\eta \geq c(d+\eta)\}}\right] - \frac{s^H}{T} \leq \mathbb{E}\left[(V_\ell^\eta - c(d+\eta))\mathbb{1}_{\{V_\ell^\eta \geq c(d+\eta)\}}\right] - \frac{s^L}{\mathbb{E}\left[D_\ell^L(p,\eta)\right]} & \quad\iff \\
\mathbb{E}\left[D_h^L(p,\eta)\right] \leq \mathbb{E}\left[D_\ell^L(p,\eta)\right],
\end{aligned}$$

and we note that the last inequality indeed holds since $Q_h > Q_\ell$ which implies that a high type rider is expected to complete any subscription plan earlier than a low type rider.

We still have to show that the outcome we defined has total rider payments equal total driver costs. To see this, observe that in the transition from the equilibrium outcome of Lemma 7 to the outcome we constructed here:

- The social welfare stayed the same (both achieving the first best).
- The low type riders' surplus stayed the same (since $C = L$).
- The high type riders' surplus stayed the same (H was defined to precisely satisfy this).

Therefore, since social welfare equals the total sum of riders' and drivers' surplus, we conclude that the total drivers' surplus also did not change in the transition, *i.e.*, the total driver surplus remains 0. Since the total driver surplus is precisely the difference between total rider payments and total driver costs, we are done.

$\square$

F.4 Proofs from Section 5

F.4.1 Proof of Theorem 4

Proof. Denote the desired simple subscription by $S = (s, \delta, T)$. We need to set these parameters using only the observed parameters $x_{\text{wal}}, w_{\text{wal}}, p_{\text{wal}}, \eta$ from a Walrasian equilibrium in which the system is presumably currently operating. We then need to show that there is a corresponding equilibrium outcome $E = (n_0, n_1, x_0, x_1, y, p, \eta)$ in which all riders subscribe.

To this end, we set

$$s = T \cdot cw_{\text{wal}} \frac{x_{\text{wal}}}{n}, \delta = cw_{\text{wal}}.$$

and T can be any number (*e.g.*, a month). As for E, we set (recall that we want all riders to subscribe)

$$\begin{aligned}
n_0 &= n \\
n_1 &= 0 \\
x_0 &= n \Pr(V^\eta \geq c(d+\eta)) \\
x_1 &= 0 \\
y &= x_0 \cdot (d+\eta) + O(\eta) \\
p &= p_{\text{wal}}
\end{aligned}$$

Note that

$$p - \delta = p_{\text{wal}} - cw_{\text{wal}} = c(d + \eta + w_{\text{wal}}) - cw_{\text{wal}} = c(d+\eta)$$

and thus, by Lemma 1, E achieves first best welfare. To show that total rider payments equal total driver costs, note that we have:

$$\begin{aligned}
c \cdot y &= n \cdot s/T + x_0(p - \delta) & \iff \\
c \cdot (x_0 \cdot (d+\eta) + O(\eta)) &= ncw_{\text{wal}} \frac{x_{\text{wal}}}{n} + x_0 c(d+\eta) & \iff \\
c \cdot (x_0 \cdot (d+\eta) + O(\eta)) &= c \frac{O(\eta)}{x_{\text{wal}}} x_{\text{wal}} + x_0 c(d+\eta)
\end{aligned}$$

and the last equality indeed holds. It remains to show best subscriber response at the subscription stage. This is given by the following lemma.

Lemma 9. $u^S(p,\eta) \geq u^R(p,\eta)$.

Proof. First, we have

$$\begin{aligned} u^R(p,\eta) &= \mathbb{E}\left[(V^\eta - p)\,\mathbb{1}_{\{V^\eta \geq p\}}\right] \\ &= \mathbb{E}\left[(V^\eta - p_{\text{wal}})\,\mathbb{1}_{\{V^\eta \geq p_{\text{wal}}\}}\right]. \end{aligned}$$

On the other hand, we have:

$$\begin{aligned} u^S(p,\eta) = \quad & \mathbb{E}\left[(V^\eta - (p-\delta))\,\mathbb{1}_{\{V^\eta \geq p-\delta\}}\right] - s/T \\ = \quad & \mathbb{E}\left[(V^\eta - (p_{\text{wal}}-\delta))\,\mathbb{1}_{\{V^\eta \geq p_{\text{wal}}-\delta\}}\right] - cw_{\text{wal}}\frac{x_{\text{wal}}}{n} \\ = \quad & \mathbb{E}\left[(V^\eta - (p_{\text{wal}}-\delta))\left(\mathbb{1}_{\{V^\eta \geq p_{\text{wal}}\}} + \mathbb{1}_{\{p_{\text{wal}} > V^\eta \geq p_{\text{wal}}-\delta\}}\right)\right] - \delta\,\Pr\left(V^\eta \geq p_{\text{wal}}\right) \\ = \quad & \mathbb{E}\left[(V^\eta - p_{\text{wal}})\,\mathbb{1}_{\{V^\eta \geq p_{\text{wal}}\}}\right] + \mathbb{E}\left[\delta\mathbb{1}_{\{V^\eta \geq p_{\text{wal}}\}}\right] + \\ & \mathbb{E}\left[(V^\eta - (p_{\text{wal}}-\delta))\,\mathbb{1}_{\{p_{\text{wal}} > V^\eta \geq p_{\text{wal}}-\delta\}}\right] - \delta\,\Pr\left(V^\eta \geq p_{\text{wal}}\right) \\ = \quad & u^R(p,\eta) + \mathbb{E}\left[(V^\eta - (p_{\text{wal}}-\delta))\,\mathbb{1}_{\{p_{\text{wal}} > V^\eta \geq p_{\text{wal}}-\delta\}}\right], \end{aligned}$$

and we conclude that $u^C(p,\eta) \geq u^R(p,\eta)$, since the second term in the last expression is non-negative. $\square$

The following lemma settles the robustness of the subscription plan to varying amounts of subscribers.

Lemma 10. If, in the equilibrium outcome E above, we modify n_0 and n_1 to any choice that satisfies $n_0 + n_1 = n$ —and consequently also modify x_0, x_1, y to satisfy rider best response at the real market stage and the correct driver supply level — then the resulting outcome still satisfies the property that total rider payments equals total rider costs. In other words, the only violated equilibrium condition is the rider best response at the subscription stage. Furthermore, as long as some riders subscribe (*i.e.*, $n_0 > 0$), the outcome obtains strictly better welfare than the Walrasian equilibrium outcome.

Proof. Let n_0, n_1 be the number of subscribers and non-subscribers, respectively, such that $n_0+n_1 = n$. Denote by x_0, x_1 the throughput for subscribers and non-subscribers, respectively, such that $x = x_0 + x_1$. Consequently, the modified driver supply level is $y = x(d+\eta) + O(\eta)$. Thus the total driver cost per unit of time is $c \cdot y = cx(d+\eta) + cO$, and we need to show that this equals the total amount of money collected from the riders. We first calculate the money collected from the subscribers. This is given by:

$$\begin{aligned} n_0 \cdot s/T + x_0(p-\delta) &= n_0 \cdot c w_{\text{wal}} \frac{x_{\text{wal}}}{n} + x_0 \cdot c(d+\eta) \\ &= n_0 \cdot c \frac{O}{x_{\text{wal}}} \cdot \frac{x_{\text{wal}}}{n} + x_0 \cdot c(d+\eta) \\ &= x_0 \cdot c(d+\eta) + \frac{n_0}{n} cO, \end{aligned}$$

and we observe that the fraction of the idle drivers' cost that the subscribers cover equals their share of the entire rider population. We now turn to calculate the money collected from the non-subscribers. This is given by:

$$\begin{aligned} x_1 \cdot p = x_1 \cdot p_{\text{wal}} & \\ &= x_1 \cdot c\left(d + \eta + w_{\text{wal}}\right) \\ &= x_1 \cdot c(d+\eta) + x_1 \cdot c \frac{O}{x_{\text{wal}}} \\ &= x_1 \cdot c(d+\eta) + n_1 \Pr\left(V^\eta \geq p\right) \cdot c \frac{O}{n \Pr\left(V^\eta \geq p_{\text{wal}}\right)} \\ &= x_1 \cdot c(d+\eta) + n_1 \Pr\left(V^\eta \geq p_{\text{wal}}\right) \cdot c \frac{O}{n \Pr\left(V^\eta \geq p_{\text{wal}}\right)} \\ &= x_1 \cdot c(d+\eta) + \frac{n_1}{n} cO. \end{aligned}$$

Summing up the total subscriber and non-subscriber payments, we have that the total rider payment equals the total driver cost. We end by showing that the outcome achieves better welfare than the Walrasian outcome. The difference in welfare is given by:

$$
\begin{aligned}
&\mathsf{SW}(x_0+x_1,\eta)-\mathsf{walSW}(\eta)\\
=&\big(n_0\cdot\mathbb{E}\left[V^\eta\cdot\mathbb{1}_{\{V^\eta\geq c(d+\eta)\}}\right]+n_1\cdot\mathbb{E}\left[V^\eta\cdot\mathbb{1}_{\{V^\eta\geq p_{\mathrm{wal}}\}}\right]-c\cdot((x_0+x_1)(d+\eta)+O)\big)\\
-&\big(n\cdot\mathbb{E}\left[V^\eta\cdot\mathbb{1}_{\{V^\eta\geq p_{\mathrm{wal}}\}}\right]-c\cdot(x_{\mathrm{wal}}(d+\eta)+O)\big)\\
=&\big(n_0\cdot\mathbb{E}\left[V^\eta\cdot\mathbb{1}_{\{V^\eta\geq c(d+\eta)\}}\right]-c\cdot(x_0(d+\eta)+O)\big)\\
-&\big(n_0\cdot\mathbb{E}\left[V^\eta\cdot\mathbb{1}_{\{V^\eta\geq p_{\mathrm{wal}}\}}\right]-cO\big)-c(d+\eta)\cdot(x_1-x_{\mathrm{wal}})\\
=&\big(n_0\cdot\mathbb{E}\left[V^\eta\cdot\mathbb{1}_{\{V^\eta\geq c(d+\eta)\}}\right]-c\cdot(x_0(d+\eta)+O)\big)\\
-&\big(n_0\cdot\mathbb{E}\left[V^\eta\cdot\mathbb{1}_{\{V^\eta\geq p_{\mathrm{wal}}\}}\right]-cO\big)-c(d+\eta)\cdot((n_1-n)\Pr(V^\eta\geq p_{\mathrm{wal}}))\\
=&\big(n_0\cdot\mathbb{E}\left[V^\eta\cdot\mathbb{1}_{\{V^\eta\geq c(d+\eta)\}}\right]-c\cdot(x_0(d+\eta)+O)\big)\\
-&\big(n_0\cdot\mathbb{E}\left[V^\eta\cdot\mathbb{1}_{\{V^\eta\geq p_{\mathrm{wal}}\}}\right]-c\cdot(n_0\Pr(V^\eta\geq p_{\mathrm{wal}})(d+\eta)+O)\big)\\
=&\mathsf{optSW}_{n_0}(\eta)-\mathsf{walSW}_{n_0}(\eta)
\end{aligned}
$$

where $\mathsf{optSW}_{n_0}(\eta)$ and $\mathsf{walSW}_{n_0}(\eta)$ denote the first best welfare at η and the highest Walrasian welfare at η, respectively, when the arrival rate of riders is n_0 (instead of n). Thus the above difference is positive by Theorem 1. □

□

G A List of Notations

For the readers' convenience, we provide in this section a list of notations used in this paper.

- O: the open driver supply, i.e. the mass of drivers in the region of interest available to dispatch
- d: the average trip duration, i.e. the units of time it takes a driver to complete an average trip
- $\eta(O) = \tau O^{-\alpha}$: the ETA for riders / *en-route time* for drivers (i.e. the average units of time it takes for a closest driver to pick up a rider) when there are O available drivers
- w: drivers' average waiting time for the next dispatch
- V^η: the random variable representing drivers' value distribution when the ETA is η
- F_η: The CDF of V^η
- The *uniform model* for rider value distribution
  - $v_{\max}$: the highest possible rider value for a trip, when the ETA is $\eta = 0$
  - β
- y: the total supply of drivers
- SW: the social welfare
- optSW: the highest achievable social welfare in an economy (i.e. the first best welfare)
- x_{opt}: the trip throughput under the first best outcome
- w_{opt}: the waiting time for the next dispatch for dirvers under the first best outcome
- q^η: the probability for a rider to take a trip on an average day under the first best outcome corresponding to the ETA η
- walSW: the social welfare under the Walrasian equilibrium
- x_{wal}: the trip throughput under the Walrasian equilibrium