Optimal Subscription Pricing Design for Ridesharing Platforms

Ben Berger
Blavatnic School of Computer Science, Tel Aviv University., benberger1@tauex.tau.ac.il,

Hongyao Ma
Decision, Risk, and Operations Division, Columbia Business School , hongyao.ma@columbia.edu

David Parkes
John A. Paulson School of Engineering and Applied Sciences, Harvard University, parkes@eecs.harvard.edu

Shreyas Sekar
Rotman School of Management and Department of Management, University of Toronto Scarborough, shreyas.sekar@rotman.utoronto.ca.

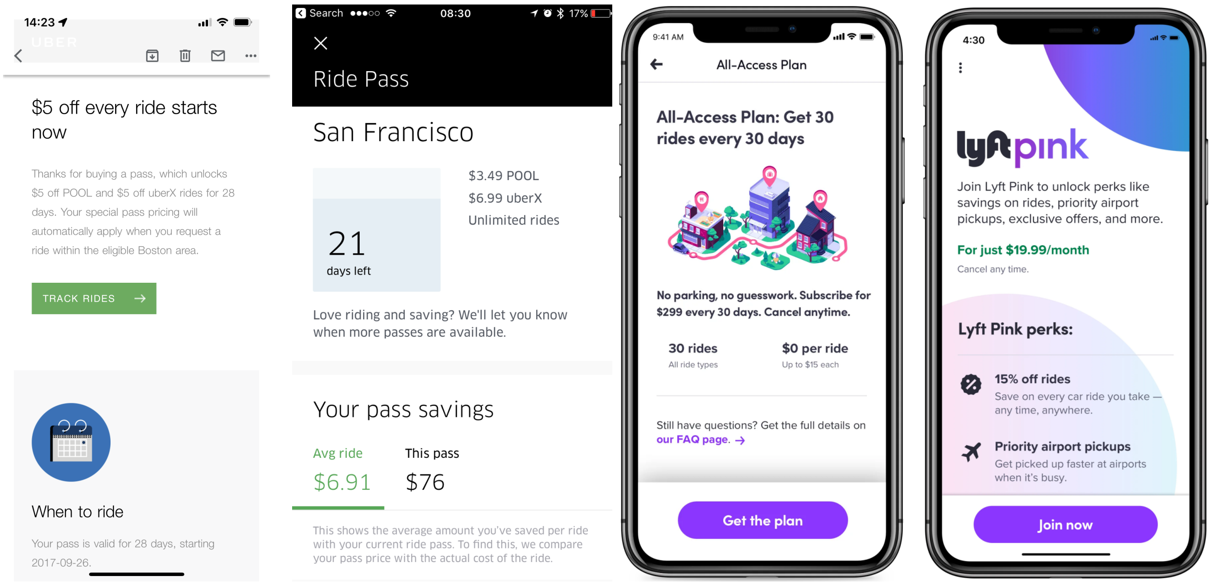
We study the design of subscription pricing in ridesharing platforms, for example pricing such as "pay a $10 fee to get $2 off every ride for a month." Contrary to folk wisdom, any market equilibrium using real-time pricing only is inefficient in ridesharing platforms due to the presence of *network effects*— specifically, a pool of *idle* driver supply is necessary for maintaining a low wait-time for riders and pick-up time for drivers, thus the rider price must be higher than the marginal trip cost to cover the cost of idle drivers. Our first main result establishes that when riders take trips with the same frequency, a *simple subscription* (for example a discount off every ride in a month) can in fact achieve the first-best social welfare by collecting the cost of idle drivers as upfront fees and charging only the marginal cost per trip. Our second main result is that when some riders ride more often than others, the first-best social welfare can still be achieved by offering an additional, *flexible* subscription that (i) charges a smaller fee (compared to the simple subscription), (ii) discounts a limited number of rides and (iii) can be renewed at any time. We also show that this optimal subscription scheme can be designed using only information that is immediately observable from the equilibrium of a non-subscription market, and without assuming a distribution of riders' willingness to pay.

Key words: ridesharing, revenue management, subscription, two-sided market

1. Introduction

Ridesharing platforms have radically changed the way people get around in urban areas. In addition to the use of per trip[1] pricing, which is known to improve operational efficiency (Castillo et al. 2017, Yan et al. 2020), platforms such as Uber and Lyft have also experimented extensively with a variety of subscription programs in the past few years.

[1] For the rest of this work, we use the term per trip price to denote the real time price that riders are charged at the time of booking a ride. When the mechanism that sets prices is clear from the context, we also use terms such as market clearing or Walrasian prices.



(a) Uber $5 off Pass (b) Uber Ride Pass (c) Lyft All-Access (d) Lyft Pink

Figure 1 Examples of subscription programs offered in the US by Uber and Lyft.

To see a few examples, Figure 1a shows a subscription offer from Uber in 2017, wherein purchasing a pass worth $25 gave riders a $5 discount on every ride within a 28-day window. Following this, Uber offered another subscription program (Figure 1b), also at the cost of $25 but where a rider could take an unlimited number of rides at a flat rate of $6.99 per ride.[2] In the case of Lyft, Figure 1c provides a screenshot of Lyft's *All Access Plan* from 2018, where a rider could pay $300 a month and effectively get $15 off each ride for at most 30 rides. Within a year, this was replaced by the *Lyft Pink* program, which charged $20/month and discounted every ride by 15%.[3]

Each time a new subscription program is offered, it is typical for a platform to A/B test (Deb et al. 2018, Dickey 2018) a few versions to explore different features and configurations.[4] Nevertheless, rather than converging to a preferred design, the major platforms have continued to test a large number of different plans (see Appendix A), and the list keeps growing. This speaks to the complexity of finding an optimal subscription plan. At the same time, there is no academic literature that provides operational guidance on designing subscription plans for ridesharing plaforms.

A first challenge in the design of an optimal subscription plan is the complexity of the design space. More precisely, any given subscription offer must optimize over a number of dimensions including but not limited to:

[2] `https://medium.com/@jsbotto/have-you-heard-of-the-uber-ride-pass-3ebf2af24568`, accessed 02/08/2022.

[3] `https://www.theverge.com/2019/10/29/20936982/lyft-pink-subscription-price-discount-perks`, accessed 01/26/2023.

[4] `https://www.macrumors.com/2018/03/16/lyft-monthly-subscription-plans/`, accessed 02/08/2022.

- the time duration of the subscription,
- the “subscription fee” to collect upfront from a rider,
- the number of discounts that are made available to a rider, if not unlimited,
- the size of the discount, and
- whether to discount a percentage of the trip fare, take a fixed number of dollars off the price of a ride, or some hybrid.

Due to marketing considerations on the rider side, and the slow equilibration of the labor market on the driver side (Hall et al. 2017), it is unrealistic to test through this large design space. Further, optimally setting these subscription parameters may require the platform to obtain precise information about riders' willingness to pay, and how they respond to larger wait times. Given that this information is difficult to estimate in practice (Yan et al. 2020, Perakis and Thraves 2022), platforms may prefer to rely on per trip prices, which can be adjusted based on the observed demand and supply.

In terms of the big picture, another challenge is the lack of clarity about the possible goals of a subscription program. For instance, should platforms deploy subscriptions with the goal of locking-in riders, improving trip throughput or gross bookings, attaining a more reliable revenue stream, or some combination thereof? With an *advance pricing* component (e.g. the upfront fee) and a *bundling* component (i.e., multiple services/discounts sold as a single unit), subscription programs in other domains typically improve revenue, sometimes at the expense of operational efficiency, by enticing consumers to commit to a decision before their uncertain values are realized (Cachon 2020). In the case of platforms such as Uber and Lyft, however, attempting to extract a larger profit is unlikely to win over drivers and riders in the presence of tight market.[5] Arguably, a better target in this setting is to improve the operational efficiency, which would allow the platform to deliver higher surplus for riders and drivers without running a larger deficit.

In this work, we design subscription programs with the goal of improving the total economic welfare achieved by the platform. Our main operational insight is that by collecting a higher aggregate payment from riders, subscription programs can in fact substantially improve social welfare in comparison to the best outcome achievable using only per trip pricing. Contrary to folk wisdom, equilibrium outcomes in a market using only per trip pricing are necessarily inefficient due to the presence of *network effects*—i.e, a higher density of idle supply reduces riders' wait time for a driver, which is equivalently the time drivers spend picking up the riders. The network effect

[5] Over a decade after the founding of the two platforms, neither Uber or Lyft have turned a profit. See `https://investor.uber.com/news-events/news/press-release-details/2022/Uber-Announces-Results-for-Third-Quarter-2022/default.aspx` and `https://investor.lyft.com/news-and-events/news/news-details/2022/Lyft-Announces-Third-Quarter-Results/`, accessed 01/26/2023

necessitates a sizeable pool of idle drivers, meaning that extra payment needs to be collected to compensate drivers for their idle time. With per trip pricing only, this extra payment comes at the expense of not picking up riders whose values are above the *marginal trip cost* (i.e., the pick-up and on-trip costs), but below the *average trip cost* (i.e., the pick-up, on-trip, and idle costs).

We show that in many settings, subscription programs can in fact close this gap and achieve the first-best welfare. By collecting upfront—through subscription fees—the cost of idle driver supply, the platform can maintain a larger pool of idle drivers to provide better rider wait time and driver pick-up time, while charging the marginal trip cost as the (discounted) per trip price. When riders take trips with the same frequency, we show that a *simple subscription* that offers a fixed discount on all rides for some period of time (e.g., "pay $25/month for $5 off every ride") optimizes welfare in equilibrium. When some riders ride substantially more than the others, we show that a platform can still achieve the first-best welfare by offering an additional, *flexible subscription* (e.g., "pay a $10 fee for $5 off for up to 8 trips in the next 30 days, and renew anytime"). We further show that an optimal subscription plan can be designed in a data-driven way, using only information that the platform can glean from observing the status-quo market conditions. As mentioned previously, it is particularly desirable that the optimal design does not rely on knowledge of the distribution of riders' willingness to pay, which is very difficult to estimate in practice due to endogeneity and the slow equilibration of the market.

1.1. Contributions

We study a continuous time model for the operation of a ridesharing platform similar to the one analyzed in Yan et al. (2020). In our setting, driver supply and rider demand are modeled as stationary and non-atomic, and a rider who requests a trip is matched to the closest driver in the region. Further, an increase in the number of idle drivers translates to a lower wait time for riders, which in turn improves each rider's value for the trip. Our stationary model abstracts away certain aspects of ridesharing platforms, e.g., time-varying demand. The model, however, captures the trade-offs between (i) maintaining idle supply to reduce rider wait times and (ii) charging a per trip price closer to the marginal cost. This allows us to focus on how subscription mechanisms can overcome this inefficiency. In equilibrium,

1. every rider purchases the subscription plan or chooses to participate in the real-time market only, in order to maximize their expected utility. In either case, riders request a ride based on their realized value, the trip price, and the expected wait time.
2. a supply of drivers join the platform if their expected revenue exceeds the cost per trip. The drivers' cost per trip depends on the average trip time, wait time, and idle time, which in turn depends on the supply of drivers at equilibrium.

3. the total payment collected from riders is equal to the total driver cost.

The objective is to design a pricing policy (which may involve a subscription program) that maximizes social welfare in equilibrium the , i.e., the total rider value minus driver cost. The focus on social welfare is common in the ridesharing literature (Ma et al. 2022, Yan et al. 2020).

Sub-optimality of Market-Clearing Prices. We begin by studying the *status quo* equilibrium solution induced in the absence of subscriptions, when the platform uses only per trip pricing. More precisely, we model this as a competitive (Walrasian) equilibrium, where the per trip price paid by riders is equal to the average trip cost incurred by drivers. Our first result (Proposition 1) establishes that the platform can never achieve the optimum social welfare—referred to as *the first best*—using only per trip pricing. Motivated this inefficiency, we study whether subscriptions can improve upon per trip pricing.

Homogeneous riders. When riders are homogeneous, with values drawn from the same distribution, we propose an optimal subscription program that achieves the first best welfare (Theorem 1). The optimal design charges riders a subscription fee equal to the average cost incurred by drivers due to idling, and offers a fixed discount on the regular per trip price so that the effective trip price equals its marginal cost. In this way, the platform compensates drivers for their idling cost without raising per trip prices above the marginal cost.

We also demonstrate that in this homogeneous setting, the optimal subscription program can be designed without any knowledge of the rider value distribution (Theorem 2). More specifically, both the subscription fee and the discount level can be computed using only the data that is available from observing the the *status quo,* non-subscription market equilibrium, such as drivers' idle time between trips. Furthermore, the subscription program is guaranteed to achieve a strict improvement in welfare even if a fraction of the riders decide (against their own best interest) not to subscribe. These results suggest that the proposed design can not only be implemented in a data-driven manner, but that it is robust to the mistakes on part of the riders and can be rolled out gradually, allowing for slow adoption.

Heterogeneous riders. Our second main result considers the setting in which riders differ in their frequency of riding, but share the same value distribution. Specifically, we consider two types of riders and first show that naively offering a simple subscription can actually lead to outcomes that are substantially worse than the per trip competitive equilibrium. This occurs due to *adverse selection*: the less frequent riders may decide not to subscribe when the subscription fee is sufficiently large to cover the cost of idle drivers, as the benefits of the discounts and the low wait times are disproportionately reaped by the more frequent riders.

Remedying this, we adopt a more general design space, and propose that a platform offer an additional *flexible subscription* which (i) provides the riders a finite number k of discounts that are to be used within a period of T days, and (ii) allows the riders to "renew" the the subscription at any time. A new challenge in the case of a flexible subscription is modeling how riders will apply these discounts. In particular, a rider may not use these discounts in a way that optimizes her expected utility, as the proper timing of when to use discounts is difficult to derive. Instead, we model riders as adopting an empirically well-supported *threshold policy* in deciding when to use a discount (Leider and Şahin 2014). Under minimal assumptions (either on the rider value distribution or the population size), we show that the first best social welfare can be implemented by offering a menu comprising of a simple and a flexible subscription (Theorem 3). Appropriately designed, the lower frequency riders will choose the flexible subscription paying a premium for the ability to 'hold on to their discounts', whereas the high frequency riders would pick the simple subscription since the subscription fee per discount is lower relative to their utilization.

1.2. Related Work

Ridesharing 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, they show that in the presence of network effects, a higher density of idle drivers can improve the waiting time and throughput of riders. Our model leverages the same phenomenon and shows how it can lead to inefficiency under market clearing prices. 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 re-position in space (Lu et al. 2018). Our results complement these works by showing that the introduction of a subscription program that discounts the real-time price can further boost the throughput and in the process, social welfare. Note that the main inefficiency that we address in this paper is caused by a fundamental mismatch between demand and supply, which we correct via subscription pricing. Another inefficiency is pointed out by Freund and van Ryzin (2021), who argue that the rapid fluctuations caused by dynamic pricing can also cause inefficiencies due to strategic behavior by riders, and who propose a queuing-based approach to correct this inefficiency.

There is also work that has studied 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. 2022, Garg and Nazerzadeh 2020, Rheingans-Yoo et al. 2019, Castro et al. 2021). Since our focus is on the use of subscription pricing to address the inefficiencies that arise due to the existence of network effects, we choose to abstract away the strategic behavior of drivers and the real-time stochasticity of ridesharing systems. Finally, recent theoretical work has analyzed the optimal growth of two-sided platforms (Lian and Van Ryzin 2021) and competition between platforms (Lian et al. 2021, Ahmadinejad et al. 2019, Fang et al. 2020) using stylized models.

Subscriptions vs Per-Use Pricing Subscription programs are popular in many settings such as digital media, e-commerce, queuing systems, 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. As far as we are aware, this is the first work that studies the design of subscription mechanisms for ridesharing markets.

Firstly, it is known that mechanisms that charge consumers fully in advance (e.g., see Xie and Shugan (2001) and citations therein) can achieve more welfare 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. 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, Hayes 1987, 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:

$$\eta(O) = \tau O^{-\alpha}, \tag{1}$$

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. Intuitively the above expression captures the idea that as the number of idle drivers (supply) increases, the wait time experienced by riders before pick up drops. For example, holding all else constant, if one assumes that the platform dispatches the nearest driver to each rider, then increasing the supply of idle drivers increases the likelihood of a driver whose location is closer at the time of booking. Empirical estimates using Uber data suggest that this relationship holds in practice with exponent $\alpha \in [\frac{1}{3}, \frac{1}{2}]$ (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:

$$w = O/x, \tag{2}$$

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[6] is weakly above the price.

[6] When 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 is described by Equation (1).

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 (1) and the average idle time w given by Equation (2). 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:

$$y(x,\eta) = x \cdot (d+\eta) + O(\eta). \tag{3}$$

Combining Equations (3) and (1), 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 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 $[0, v^{\max}]$.

For the rest of this paper, we assume that Assumption 1 holds.

2.1. Social Welfare, and the First Best Solutions

We first 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 η to denote the solution (x,y) that maximizes social welfare at the given wait time. 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 the Walrasian equilibria at different wait times typically correspond to different real-time prices. 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 (Yu et al. 2022) 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 total driver cost per unit time, and is formalized below:

$$\mathsf{SW}(\eta,x)=n\cdot\mathbb{E}\left[V^\eta\cdot\mathbb{1}_{\{\bar{F}^\eta(V^\eta)\le x/n\}}\right]-c\cdot y(x,\eta)=n\cdot\mathbb{E}\left[V^\eta\cdot\mathbb{1}_{\{\bar{F}^\eta(V^\eta)\le x/n\}}\right]-cx(d+\eta)-cO,$$

where $\bar{F}^\eta(v)$ denotes $1-F^\eta(v)$ and the second expression comes from Equation (3). The following lemma characterizes the marginal rider valuation (and system cost) at the first best solution.

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 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. Naturally, it makes sense to study non-trivial instances where the social welfare is strictly positive, and we assume implicitly that the platform only operates at wait times η for which $\mathsf{optSW}(\eta)>0$.

2.2. Walrasian Equilibrium

We now discuss the Walrasian or the per trip 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.

In simple terms, the Walrasian mechanism works as follows: *The platform selects a per trip price* p_{wal}*. Each rider arrives at the market, observes the 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}.

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

$$\mathsf{SW}(\eta, x_{\text{wal}}) = n \cdot \mathbb{E}\left[V^\eta \cdot \mathbb{1}_{\{V^\eta \geq p_{\text{wal}}\}}\right] - cx_{\text{wal}}(d+\eta) - cO. \tag{4}$$

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. An outcome $(x_{\text{wal}}, w_{\text{wal}}, p_{\text{wal}}, \eta)$ is a Walrasian equilibrium if:

(i) (rider best-response) $x_{\text{wal}} = n \cdot \bar{F}^\eta(p_{\text{wal}})$,

(ii) (driver best-response) $p_{\text{wal}} = c(d+\eta+w_{\text{wal}})$, and

(iii) (driver idle-time) $w_{\text{wal}} = O(\eta)/x_{\text{wal}}$.

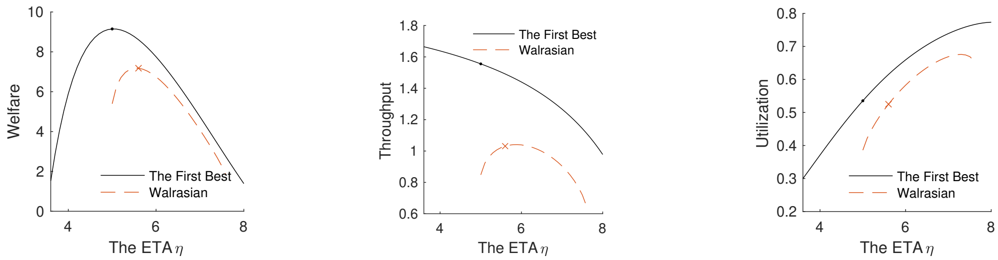
Denote by $\mathsf{walSW}(\eta)$ the social welfare corresponding to the Walrasian equilibrium at η. We are now ready to prove our first main result, which is that the Walrasian equilibrium is sub-optimal.

Proposition 1. *Under any waiting time* η*, the Walrasian equilibrium induced by per trip pricing (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* $\eta > 0$.

Intuitively, the inefficiency of the Walrasian equilibrium stems from the misalignment between the average and marginal system costs in the first best outcome. More precisely, in order to optimize social welfare, it is necessary to accept all riders whose realized value for the trip is at least $c(d+\eta)$; see Lemma 1. However, if we charge a per trip price equal to this marginal value (since there is no price discrimination), the platform cannot compensate drivers for the time spent idling, i.e., the driver cost per trip is $c(d+\eta+w)$. Therefore, at equilibrium, the per trip price must be higher than the above marginal value, which leads to a lower welfare and throughput compared to the first best outcome. We note that such a result has also been observed in other markets with economies of scale (e.g., Mendelson (1985), Cachon and Feldman (2011)).

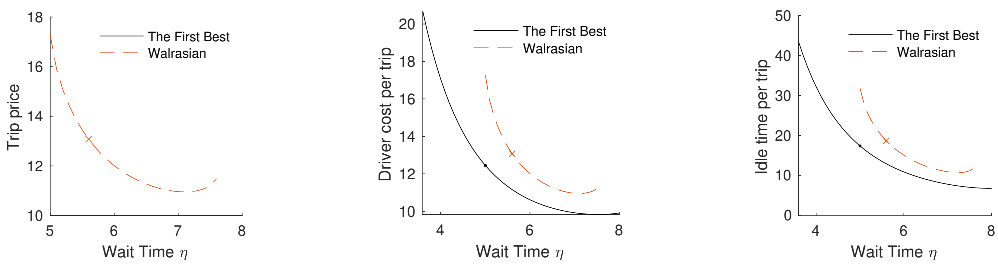
We conclude this section with an example where we compare in a simple economy the first best and the Walrasian equilibrium outcomes at different levels of waiting time.

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,



(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

meaning that 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 highest possible first best welfare 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 very low η's since the cost of providing the 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 (which is equal to the average cost of each trip).

3. Simple Subscriptions and Robust Implementation

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. Appropriately designed, the platform is able to achieve the first best outcome at any desired wait time η, assuming that the riders are homogeneous. Further, we show that the platform can design such a subscription in a data-driven manner using only the information that can be gleaned by observing market conditions (e.g., wait times, idle time, etc.). In other words, an optimal subscription can be designed even when the platform has no knowledge of rider valuations and their sensitivity to wait times.

Definition 2. A *simple subscription* is represented by $S = (s, \delta, T)$ 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. On day zero, the rider decides whether to subscribe and pay s upfront before observing future realized values.
2. *If the rider decides to subscribe*, then: in each of the days 1 through T, the rider requests a trip at the discounted price $p - \delta$, if the realized value V^η under the observed wait time η is at least the discounted price.
3. *If the decision was not to subscribe*, then: the rider requests a trip on any given day if the realized value V^η under the observed wait time η is at least the per trip price p.

We note that on any given day, a rider's best response is to request a trip if the realized value is weakly above the effective price (which is $p - \delta$ if the rider subscribed, and p otherwise).

Equilibrium Outcome Under a Simple Subscription 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 under per trip price p and wait time η is given by:

$$u^S(p, \eta) = \mathbb{E}\left[(V^\eta - (p - \delta))\mathbb{1}_{\{V^\eta \geq p - \delta\}}\right] - s/T. \tag{5}$$

The rider's daily utility from the real-time market (i.e., from not subscribing) is given by

$$u^R(p, \eta) = \mathbb{E}\left[(V^\eta - p)\mathbb{1}_{\{V^\eta \geq p\}}\right]. \tag{6}$$

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.

Denote by n_0, n_1 the arrival rate 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. An outcome $(n_0, n_1, x_0, x_1, y, p, \eta)$, forms an *equilibrium under the simple subscription $S = (s, \delta, T)$,* if the following conditions hold:

(i) Subscriber best-response at subscription stage: $n_0 > 0 \implies u^S(p,\eta) \geq u^R(p,\eta)$.

(ii) Non-subscriber best-response at subscription stage: $n_1 > 0 \implies u^R(p,\eta) \geq u^S(p,\eta)$.

(iii) Subscriber best-response at the real-time market: $x_0 = n_0 \Pr(V^\eta \geq p - \delta)$.

(iv) Non-subscriber best-response at the real-time market: $x_1 = n_1 \Pr(V^\eta \geq p)$.

(v) Total driver supply: $y = (x_0 + x_1)(d + \eta) + O(\eta)$.

(vi) Driver Best-Response: $c \cdot y = n_0 \cdot s/T + x_0 \cdot (p - \delta) + x_1 \cdot p$.

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

THEOREM 1. *Given any waiting time η for which $\mathsf{optSW}(\eta) > 0$, and for any desired duration T, there is a simple subscription with duration T that attains the first best welfare in an equilibrium outcome in which all riders subscribe.*

Recall that the inefficiency of per trip pricing stems from the gap between the marginal rider value and the average driver cost. In other words, the platform needs to extract enough revenue from the riders to pay drivers for any time spent idling, while simultaneously allowing all riders whose realized value is at least $c(d+\eta)$ to take a trip. The issue is further compounded by the higher payment that riders are subject to. If this is not matched by an equivalent increase in their utility, then riders would simply not subscribe, and instead participate in the real-time market paying the per-trip price if their realized value is large enough.

The simple subscription that we design delicately balances all these considerations. In the proposed subscription, the discount value is chosen so that the effective per trip price $(p - \delta)$ for subscribers matches the driver costs corresponding to the on-trip and wait time part of the trip. This ensures that no rides are allocated to low value riders, whose realized value is smaller than $p - \delta$. On the other hand, the subscription fee is carefully chosen to cover the time drivers spend being idle. Once the parameters are fixed, it only remains for us to show that riders prefer subscribing to participating in the real-time market.

Proof: In order to achieve the first best welfare 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 cw_{opt}. Thus, assuming all n riders subscribe, the fee s must satisfy $n \cdot s/T = cw_{\text{opt}} x_{\text{opt}}$ 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. To see this, we first argue that under the above conditions, $u^S(p,\eta)$ is strictly positive. This follows from the fact that 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:

$$\mathsf{optSW}(\eta) = n \cdot u^S(p,\eta). \tag{7}$$

In particular, $u^S(p,\eta) > 0$ as desired. Finally, note that $u^R(p,\eta)$ is monotonically decreasing in p, and there are choices of per trip price (*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 $c(d+\eta)$. □

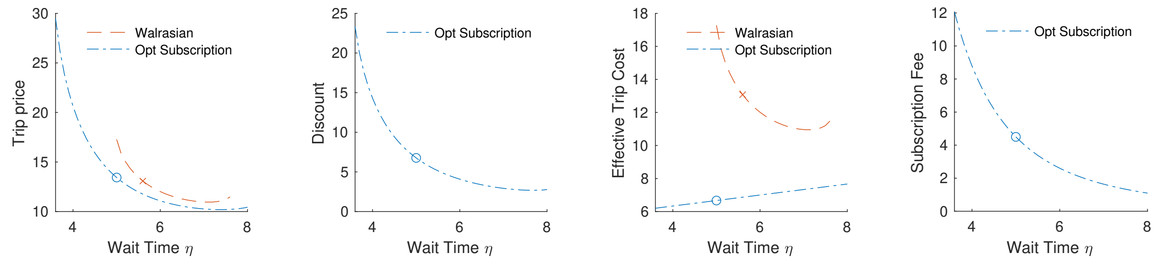
In the simple subscription constructed in the proof of Theorem 1, the per trip price p was set high enough (while keeping the discounted price $p - \delta$ fixed) to ensure riders prefer subscribing over not subscribing. Given that the Walrasian mechanism represents the *status quo*, a natural candidate for this price is the market clearing price p_{wal} under the Walrasian equilibrium at the given wait time η. The following corollary shows that the Walrasian price is sufficient to ensure that the simple subscription reaches an optimal outcome.

Corollary 1. *Given any waiting time η for which* $\mathsf{walSW}(\eta) > 0$*, there exists a simple subscription that achieves the first best outcome even when the per trip price p is equal to the market clearing price under the Walrasian equilibrium at wait time η.*

The proof of the above corollary follows from that of Theorem 2, where a similar idea is used to construct a robust subscription. We now revisit the same economy analyzed in Example 1, and illustrate the outcome under the optimal simple subscriptions.

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.

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



(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.

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. Q.E.D.

3.1. Data-Driven Implementation of Optimal Subscriptions in Practice

In the first half of Section 3, we presented a subscription plan that attains the first best welfare at a given waiting time η. In order to implement the plan, the platform has to be able to compute the subscription fees, which among other things depend on the probability that riders take a trip at the 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 the optimal subscriptions can be implemented in equilibrium *without prior knowledge of the rider value distributions*. Rather, it suffices for the platform to have access to the outcome statistics of a Walrasian equilibrium, *i.e.*, in non-subscription mode. Further, in the first best equilibrium outcome achieved in Theorem 1, all riders subscribe. As a second result, we also show that the subscription plans are robust to riders deciding not to subscribe even if it is in their best interest to do so. Even though these irrational decisions result in sub-optimal welfare (due to the non-subscribers requesting a ride with lower probability), the resulting outcome is still an equilibrium.

We now state the main result in this section. Recall that we denote the rider arrival rate, Walrasian throughput, driver idle time and trip price by $n, x_{\mathrm{wal}}, w_{\mathrm{wal}}, p_{\mathrm{wal}}$ respectively, and note that all these parameters are observable by the platform.

THEOREM 2. *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 based only on $(\eta, x_{\mathrm{wal}}, p_{\mathrm{wal}}, w_{\mathrm{wal}}, c)$ *that achieves the first best welfare at wait time* η *in an equilibrium outcome in which all riders subscribe.*

(ii) This simple subscription achieves a strict improvement in social welfare over the Walrasian equilibrium, even if only $n_0 \in (0, n]$ *riders subscribe and the remaining do not subscribe.*

From a practical perspective, this latter 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. To show this surprising result, we argue that even when only a fraction of riders subscribe (and the rest participate in the real-time market), the platform's revenue from the subscription still suffices to cover the drivers' idle costs.

4. Multiple Rider Types

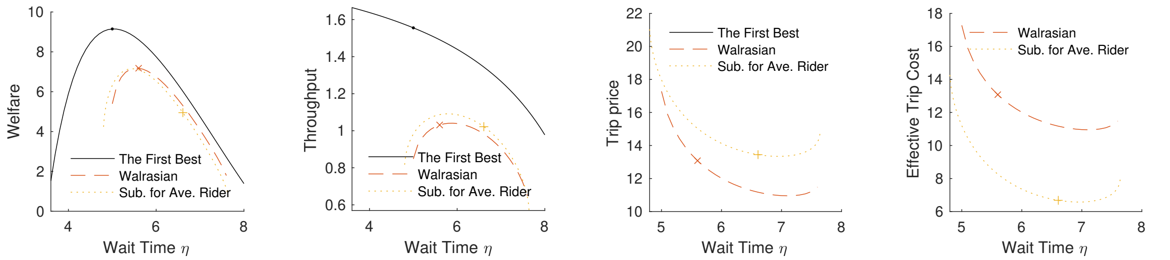
In this section, we study subscription mechanisms in the presence of heterogeneous riders. We show that when some riders ride more frequently than others, the simple subscriptions introduced in Section 3 cannot always attain the first best outcome in equilibrium. We then prove the main result of this 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 under some assumptions on the distribution of rider values or when the arrival rate is sufficiently large.

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.

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). Before proceeding, we illustrate via the following example that when a platform naively treats the rider population as homogeneous and offers a 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 $n_h = 1$



(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.

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

When the platform 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 times 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. As a result, the outcome is sub-optimal for every (feasible) wait time η.

This also creates a large inequity between the two rider types. For example, when $\eta = 6.61$, 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 nine 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. Q.E.D.

Building on Example 2, we show formally that when the rider types are sufficiently different, no simple subscription can attain the 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.

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 is deferred to Appendix D.1.

4.1. Flexible Subscriptions

In this section we introduce a new subscription model that allows us to cope with the different likelihoods of requesting a trip. Intuitively a flexible subscription generalizes the simple subscription by k discounts over a duration of T days, so that riders can utilize the discounts only when their realized value is sufficiently high.

DEFINITION 4. A *flexible subscription mechanism* is represented by $C = (s, \delta, k, T)$ where s is the subscription fee, δ is the discount, k is the number of discounts and T is the maximum duration of the subscription period. Under a flexible subscription C, the timeline works as follows:

1. In the beginning, a rider decides whether or not to subscribe before observing her realized values for future rides. If a rider subscribes to C, then:
   (a) The rider pays the upfront fee s.
   (b) In each of the following T days, the rider may choose to request a trip at the discounted price of $p - \delta$, given the realized value drawn from the distribution V_i^η (at wait time η).
   (c) The subscription ends when the rider uses up all k discounts or when T days have elapsed after subscribing. In the latter case, unused discounts are wasted.
   (d) The rider may choose to renew their subscription after the previous offer ends (even if T days have not elapsed).
2. A rider who does not subscribe participates in the real-time market and request a trip on any given day if the realized value exceeds the per trip price p.

Rider Behavior in the Face of Flexible Subscriptions Given a flexible subscription (e.g., k discounts to be used within T periods), how do riders utilize the discounts once they subscribe? For example, one can deduce that the optimal policy for riders involves dynamic thresholds so that the rider requests a ride during period $t \in [T]$ if the realized valuation V^η is at least the threshold for that period. However, computing these thresholds is a computationally non-trivial task and empirical evidence suggests that consumers use simple decision heuristics as opposed to following an optimal policy (Leider and Şahin 2014, Grubb and Osborne 2015). To incorporate more realistic behavioral patterns in our design, we assume instead that riders follow a static threshold policy in each round, i.e., request a ride if their realized value exceeds some fixed threshold. We consider two natural candidates for this threshold, which leads to the following policies:

1. *Myopic Policy*: A rider applies a myopic policy under a flexible subscription C if they request a ride in each period as long as $V^\eta \geq p - \delta$, where p denotes the per trip price. A myopic threshold implies that riders treat the subscription fee as a sunk cost (Thaler 1985). The myopic policy is optimal when $k = T$.
2. *Fixed Discount Policy* A rider applies a fixed discount policy under a flexible subscription C if they request a ride in each period as long as $V^\eta \geq p - \delta + \frac{s}{k}$, where p denotes the per trip price. Under a fixed discount policy, the rider treats each discount as having an average price of $\frac{s}{k}$. The fixed discount policy is optimal when $k = 1$ and $T \to \infty$.

These two policies come close to matching the performance of the optimal policy under vastly different conditions. Therefore, we take a 'best of both worlds' approach and assume that riders employ either a myopic or fixed discount policy, whichever maximizes their utility.

Definition 5. Straightforward Policy (Best of Both Worlds). We say that a rider employs a straightforward policy if they pick the utility maximizing option among the myopic and fixed discount policies for the given flexible subscription C.

Equilibrium under the Straightforward Policy To capture both the myopic and fixed discount models, we consider an arbitrary rider who requests a trip as long as $V^\eta \geq (p - \delta) + \tau$, where $p - \delta$ denotes the discounted per trip price and $\tau \in \{0, \frac{s}{k}\}$, depending on whether the consumer follows the myopic or fixed discount policy respectively. A rider who purchases a flexible subscription C (as it maximizes her utility) will continue to repurchase the same subscription after the previous one expires. Therefore, in steady-state, the rider's *daily utility* from a flexible subscription $C = (s, \delta, k, T)$ under the realized real time price p and wait time η is given by:

$$u_i^C(p, \eta) = \mathbb{E}\left[(V_i^\eta - (p - \delta))\, \mathbb{1}_{\{V_i^\eta \geq p - \delta + \tau\}}\right] - s \cdot \frac{1}{\mathbb{E}\left[D_i^C(p, \eta)\right]}, \tag{8}$$

where $D_i^C(p, \eta)$ is a random variable that denotes the duration after which the subscription expires, i.e., the minimum between T and the day in which the rider utilizes her k-th discount. Informally, the left-hand side of Equation (8) is the surplus generated when from requesting a trip, which occurs as long as $V_i^\eta \geq p - \delta + \tau$. The right-hand side denotes the rider's daily payment, i.e., the subscription fee s amortized over $\mathbb{E}[D_i^C(p, \eta)]$ days. A formal derivation of Equation (8) is deferred to Appendix D.2. Note that when $k = T$, the flexible subscription is effectively a simple subscription and $D_i^C(p, \eta)$ is deterministic and equals T.

We now define an equilibrium outcome under a general menu $\mathcal{M} = (C^j)_{j=1}^m$ of subscriptions, where $C^j = (s^j, \delta^j, k^j, T^j)$. 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. 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 6. An outcome $\left(\left(n_i^j, x_i^j\right)_{i\in\{h,\ell\},j\in[m]}, y, p, \eta\right)$ is an *equilibrium for h,ℓ type riders under a menu of subscriptions* $\mathcal{M}=(C^j)_{j=1}^m$ if the following conditions hold:

(i) Rider best-response for subscription: $\forall i\in\{h,\ell\}, j, j'\in[m]: n_i^j>0 \implies u_i^{C^j}(p,\eta)\geq u_i^{C^{j'}}(p,\eta)$.

(ii) Rider best-response in real-time: $\forall i\in\{h,\ell\}, j,\in[m]: x_i^j = n_i^j \Pr(V_i^\eta \geq p-\delta^j)$.

(iii) Total driver supply: $y=(\sum_{i\in\{h,\ell\},j\in[m]} x_i^j)(d+\eta)+O(\eta)$.

(iv) Total rider payment equal to driver costs:

$$c\cdot y = \sum_{i\in\{h,\ell\},j\in[m]\setminus\{0\}} n_i^j\cdot s^j/\left(\mathbb{E}\left[D_i^j(p,\eta)\right]\right) + \sum_{i\in\{h,\ell\},j\in[m]} x_i^j(p-\delta^j). \tag{9}$$

4.2. 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 under some conditions described below. To achieve this, we offer the riders a menu of two different subscription plans that they can choose between. One is tailored for the high type riders and the other for the low type riders.

We start by defining a condition on the distribution of rider values.

Definition 7. (Convex Inverse Hazard Rate) For any given wait time η, the distribution of rider values is said to have a convex inverse hazard rate if F^η is differentiable and $\frac{1-F^\eta(v)}{f^\eta(v)}$ is convex where $f^\eta(v)=\frac{\partial F}{\partial v}$.

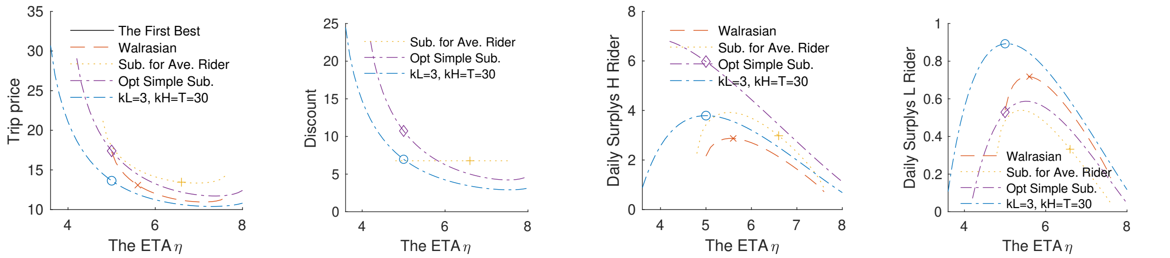
Our assumption is satisfied by several popular distributions such as uniform, exponential, and Pareto (in fact, the inverse hazard rate is linear in these three cases).

We now state our main result for Section 4.

Theorem 3. *For any wait time η at which a Walrasian equilibrium exists, we can design a subscription menu $\mathcal{M}=(H,L)$, where $H=(s^H,\delta,k^H,T)$ and $L=(s^L,\delta,k^L,T)$ with $k^L<k^H=T$ that attains the first best welfare in an equilibrium outcome if riders employ a straightforward policy and one of the following conditions hold:*

1. *$n\geq\bar{n}$, for a suitable threshold $\bar{n}$ on the rider population, or*
2. *The distribution V^η of rider values at wait time η has a convex inverse hazard rate.*

We reiterate that the first condition above is independent of the distribution, whereas the second condition holds for several studied distributions such as Uniform, Exponential, and Pareto (and is independent of the arrival rate n).



(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 probabilities of being interested in a trip, the theorem can also be proved for more general distributions. In particular, the theorem can be shown to hold whenever, for the given η, both types $i \in \{h, \ell\}$ satisfy

$$\mathbb{E}\left[(V_i^\eta - c(d+\eta)) \mid V_i^\eta \geq c(d+\eta)\right] > cw_{\rm opt}(\eta).$$

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 $T = 30$) that 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.

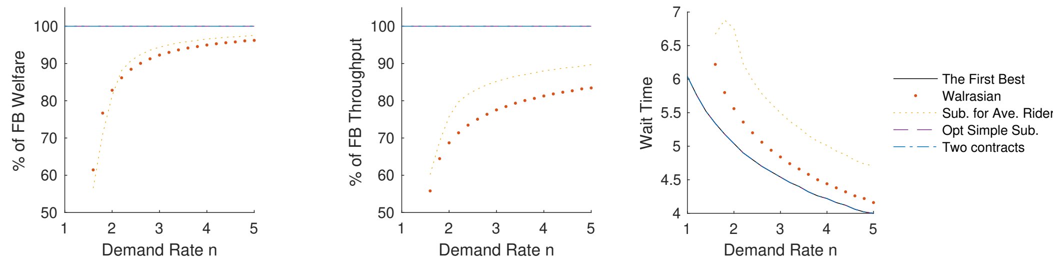
In order to incentivize all riders to subscribe, a platform offering only a simple subscription needs to raise the real-time price substantially. With two plans, on the country, the price is only slightly higher than that under the Walrasian equilibrium, which is desirable in practice.

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. Q.E.D.

5. 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).

Consider economies where the riding frequency and value distribution of each rider type, driver cost, trip duration and the wait time∼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



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

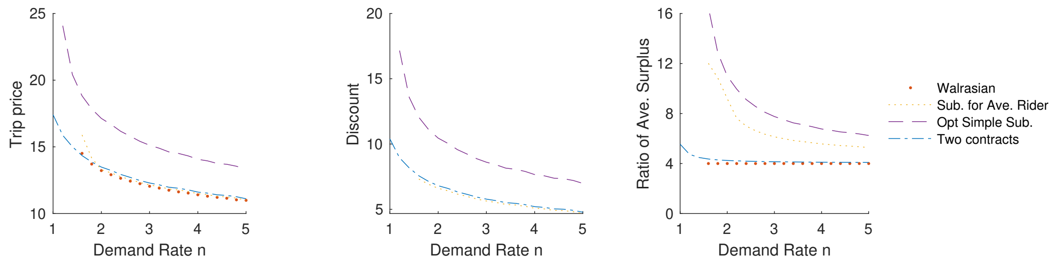
Figure 7 Social welfare (as % of the first best welfare), trip throughput (as % of the first best throughput), and equilibrium wait time under various benchmarks and mechanisms. Varying total demand rate n.

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.

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 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 dense, both the real-time price and the discount level decreases, since the platform is able to operate more efficiently. 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 four times as likely to be interested in a trip, this "daily surplus ratio" is exactly 4 under the Walrasian equilibrium. Two 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.



(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.

6. Discussion and Practical Considerations

Motivated by the challenges faced by ridesharing platforms in designing efficient yet practically appealing subscriptions, we proposed a simple subscription mechanism that charges riders a fixed (upfront) fee and in return, offers discounted rides. When riders are *apriori* homogeneous, we show that simple subscriptions optimize social welfare and can be implemented in a data-driven manner, without access to the distribution of rider values. The parameters of the simple subscription conform to a natural interpretation: a) the fixed fee matches the average costs incurred by drivers due to idling; b) the discounted per trip price matches the marginal cost of a ride. When riders are heterogeneous in their trip frequency, offering a single subscription can be sub-optimal and only benefit high-frequency riders. We argue that platforms can overcome this adverse selection problem by combining the simple subscription with a flexible subscription, which allows riders to utilize their discounts only when their realized value is sufficiently high. In particular, this menu results in the first best welfare by ensuring that low-frequency riders purchase the flexible offer, while high-frequency riders take the simple subscription. Given that the tension between different consumer types is a mainstay of ridesharing platforms, this result provides a simple template for getting low-frequency riders to subscribe (by charging for flexibility) while simultaneously incentivizing high-frequency riders to take more rides (via simple subscriptions).

Our work contributes to the extensive literature on subscription mechanisms in two-sided marketplaces. That said, much of the focus on subscriptions — in both the literature as well as popular media — has centered around (their) extracting more surplus or locking-in customers. While these are important considerations, our work takes a complementary approach and investigates whether subscriptions can improve operation efficiency in ridesharing markets. In particular, our primary objective in designing the above mechanisms was to remedy the gap between the average and marginal cost in ridesharing markets. Arguably, this mismatch can critically undermine the normal operations of a ridesharing market (Castillo et al. 2017), e.g., per trip pricing results in too few drivers, which can have severe downstream consequences. Therefore, to achieve this goal, we considered a stylized model that spotlights the network effects between idle drivers and rider wait time. In what follows, we discuss the performance of the proposed mechanisms as one incorporates other aspects pertaining to ridesharing.

Time-Varying Demand.

In the present work, we model a market with a fixed arrival rate n of riders whose trip value is drawn from a distribution V^η that depends only the wait time η. Naturally, the arrival rate of riders fluctuates in practice leading to higher demand in some intervals (peak hours) and lower demand in other times (off-peak hours). To examine how simple subscriptions perform in this more generalized setting, consider a time interval (e.g., one day), and suppose that the platform observes peak demand for some fraction of time. Mathematically, consider an overall population rate of n such that each rider independently requests a trip with probabilities λ_p and λ_{op} during peak and off-peak hours respectively, with $\lambda_p > \lambda_{op}$ (the respective arrival rates are $\lambda_p n$ and $\lambda_{op} n$). We assume drivers enter the platform if their expected revenue per unit-time (averaged over peak and off-peak hours) exceeds the unit cost, and stay for the entire duration.

Under this model, we argue that simple subscriptions can still implement the first best outcome, albeit as long as the platform can adjust the per trip price from the status quo equilibrium (in the absence of subscriptions). To see why, first note that since the driver supply is fixed, riders would observe a higher wait time (η_p) during peak hours and a lower wait time (η_{op}) during off-peak hours. Although one could charge different per-trip prices based on the demand (i.e., surge pricing), the resulting Walrasian outcome continues to be sub-optimal as the per-trip price would exceed the marginal cost for any given demand (see Proposition 1). Finally, the platform can design a simple subscription where the subscription fee matches the *expected* idle time for drivers, and the discount is calibrated so that the post-discount price in either period equals the corresponding marginal cost.

In Appendix B.4, we highlight via examples that the above design is still capable of maximizing social welfare. However, there are crucial differences in the underlying outcome. First, drivers idle more during off-peak hours while riders derive higher surplus during peak hours. At a high level, the subscription provides a simple pathway to transfer surplus from peak to off-peak hours, i.e., it helps the platform to maintain a higher supply of drivers during periods of low demand, which in turn lowers wait times during peak hours. Second, as the disparity between the two periods (i.e., the ratio $\frac{\lambda_p}{\lambda_{op}}$) increases, it may be necessary to increase the per trip price during peak demand and reduce the off-peak price to ensure that subscribing is still utility-maximizing for riders. In other words, the per trip price upon implementing a simple subscription may no longer match the original Walrasian price; this ensures that the post-discount price is equal to the marginal trip cost. [7] Finally, it is worth acknowledging, that the simple model discussed here may not fully represent the complexities of ridesharing systems with time-varying demand, e.g., in practice driver supply may not remain constant throughout the day. That said, the above discussion hints at the value of simple subscriptions even in a dynamic market, and can serve a starting point for future work.

Continuous Rider types and Driver Costs.

For the sake of tractability and exposition, we assumed that all drivers incur the same marginal cost c per unit time. However, our insights continue to be applicable in cases where the drivers' unit costs follow a convex function $c(y)$, i.e., the marginal costs are increasing. More precisely, simple subscriptions yield the first best outcome as long as the gap between rider value and driver cost in the optimal welfare is sufficiently large. To see why this is necessary, note that since driver payments are uniform, each driver's revenue must be at least the cost incurred by the highest-cost driver. Therefore, when the aggregate rider welfare is smaller than the required payments, it is not possible to implement the first best outcome in equilibrium. Finally, we reiterate that even when marginal costs are uniform, as in our core model, one cannot simply have an infinite pool of drivers as the resulting idling costs would grow proportionally, leading to negative social welfare.

On the rider side, the two type model that we study in Section 4 captures the spirit of the tradeoffs involved in designing subscriptions for heterogeneous consumers. Moreover, assuming two consumer types is a typical modeling choice in the literature on subscriptions (Cachon and Feldman 2011). Generally, in a setting with a continuum of rider types who vary in their propensity to request a trip, offering just two subscriptions may not result in an optimal outcome. In such a scenario, we would expect the frequent riders to take the high-volume, simple subscription, medium-frequency riders to take the flexible subscription, and low-frequency riders to not subscribe at all. Depending on the type distribution, this could still result in substantial welfare improvements compared to per trip pricing. Given the sub-optimality of two subscriptions against a continuum of types, an important open question is to design near-optimal subscription plans that are also practically feasible (e.g., a finite number of subscriptions)

Competition, Ridesharing Priorities, and Economies of Scale

As is the case in almost all of the literature on pricing for ridesharing markets, we consider the subscription design problem faced by a monopolist platform. Although we do not model competing firms explicitly, we make no assumptions on the rider value distribution—one could, for example,

[7] Adjusting the status quo per trip price is necessary to ensure optimality. That said, if the per trip price were unchanged, one could still design approximately optimal subscriptions that improve upon the Walrasian outcome by lowering the subscription fee and the discounts so that subscribing is still utility-maximizing but the post-discount price no longer equals the marginal cost.

interpret the rider demand in our model as the amount of trip requests given the existence of a competitor. Arguably, incorporating price competition more formally may distort our findings. On the one hand, in the presence of competition, subscriptions become more valuable as platforms seek new ways to hold on to customers. Such a viewpoint is well aligned with the present work, where the focus is on maximizing social welfare as well as throughput by ensuring all riders subscribe. On the other hand, competing firms often undercut each other which could undermine the value provided by advance selling (Cachon and Feldman 2017).

More broadly, this work was motivated by the difficulties faced by ridesharing platforms in designing a suitable subscription plan that optimizes across multiple levers (e.g., see Figure 1). Given the current subscription priorities of the major ridesharing plaforms[8], how do the analytical results in this work translate to practice? We point out two key ways in which our results provide operational guidance. First, our work argues in favor of returning to simple subscriptions as in Figure 1a and more importantly, provides a clear prescription on how platforms should set the subscription fee (to compensate drivers for idling) and discount level (to ensure prices align with marginal costs). Even if there is limited appetite for offering larger discounts, our work still provides ridesharing platforms with a novel perspective on subscriptions, namely that they can be leveraged to close the gap between rider valuations and driver costs. This simple insight can be particularly valuable as it provides platforms with a pathway to grow their driver supply and rider base at the same time. Second, in the face of heterogeneous demand, our main contribution is that platforms should offer multiple subscriptions differentiated in terms of their flexibility, i.e., the number of discounts within a fixed period. Designing incentive-compatible pricing strategies for heterogeneous users is considered to be a particularly challenging problem and large body of work has studied how sellers should structure the pricing menu (e.g., multiple subscriptions) (Cachon and Feldman 2011, Afèche et al. 2018). In the context of ridesharing, our central insight is the identification of *flexibility* as the key dimension along which the different offers can be separated. It is worth noting that although the major North American platforms such as Uber and Lyft have not offered flexible subscriptions in the past, other platforms such Meituan have successfully deployed such a flexible subscription, in the context of food delivery (see Figure 10 in Appendix A).

Finally, although the primary focus of this paper was ridesharing, our insights are applicable to other two-sided marketplaces that exhibit *economies of scale*. For example, in online labor markets such as Upwork or Taskrabbit, increasing the supply of idle workers could improve an employer's quality of service as it increases the likelihood of finding the right person for the task. In such scenarios, platforms may find it beneficial to design subscriptions that take idle times into account in

[8] The main subscriptions offered by both Uber and Lyft (Uber One and Lyft Pink) currently charge riders \$9.99 a month and offer a flat 5% discount on all rides along with other perks such as priority pickups.

order to influence more workers to enter the platform. In the same spirit as the present work, Feldman et al. (2023) propose a two-part tariff for food delivery in order to compensate restaurants for the physical externalities (e.g., congestion) caused by online orders. Designing more general subscription mechanisms that account for these type of externalities (e.g., idling, congestion) in two-sided marketplaces is an important direction for future work.

References

Philipp Afèche, Zhe Liu, and Costis Maglaras. 2018. *Ride-hailing networks with strategic drivers: The impact of platform control capabilities on performance.* Technical Report. Columbia Business School. Available at SSRN 3120544.

AmirMahdi Ahmadinejad, Hamid Nazerzadeh, Amin Saberi, Nolan Skochdopole, and Kane Sweeney. 2019. Competition in Ride-Hailing Markets. In *Proceedings of WINE 2019.*

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

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

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

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

Gérard P Cachon. 2020. A research framework for business models: what is common among fast fashion, e-tailing, and ride sharing? *Management science* 66, 3 (2020), 1172–1192.

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

Gérard P Cachon and Pnina Feldman. 2017. Is advance selling desirable with competition? *Marketing Science* 36, 2 (2017), 214–231.

Juan-Camilo Castillo, Dan Knoepfle, and E. Glen Weyl. 2017. Surge Pricing Solves the Wild Goose Chase, Constantinos Daskalakis, Moshe Babaioff, and Hervé Moulin (Eds.).

Francisco Castro, Hongyao Ma, Hamid Nazerzadeh, and Chiwei Yan. 2021. Randomized FIFO Mechanisms. *Technical report, Columbia University* (2021).

M Keith Chen and Michael Sheldon. 2015. Dynamic Pricing in a Labor Market: Surge Pricing and the Supply of Uber Driver-Partners. *University of California (Los Angeles) Working Paper URL http://citeseerx. ist. psu. edu/viewdoc/download* (2015).

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

Megan Rose Dickey. 2018. Lyft is testing monthly subscriptions for riders. `https://techcrunch.com/2018/03/15/lyft-is-testing-monthly-subscriptions-for-riders/` [Online; accessed 26-January-2023].

Zhixuan Fang, Longbo Huang, and Adam Wierman. 2020. Loyalty programs in the sharing economy: Optimality and competition. *Performance Evaluation* (2020).

Pnina Feldman, Andrew E Frazelle, and Robert Swinney. 2023. Managing relationships between restaurants and food delivery platforms: Conflict, contracts, and coordination. *Management Science* 69, 2 (2023), 812–823.

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

Daniel Freund and Garrett van Ryzin. 2021. Pricing Fast and Slow: Limitations of Dynamic Pricing Mechanisms in Ride-Hailing. *Available at SSRN 3931844* (2021).

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

Michael D Grubb and Matthew Osborne. 2015. Cellular service demand: Biased beliefs, learning, and bill shock. *American Economic Review* 105, 1 (2015), 234–271.

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

Jonathan V Hall, John J Horton, and Daniel T Knoepfle. 2017. *Labor Market Equilibration: Evidence from Uber.* Technical Report. New York University Stern School of Business.

Beth Hayes. 1987. Competition and two-part tariffs. *Journal of business* (1987), 41–54.

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

Stephen Leider and Özge Şahin. 2014. Contracts, biases, and consumption of access services. *Management Science* 60, 9 (2014), 2198–2222.

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

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

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

Alice Lu, Peter I Frazier, and Oren Kislev. 2018. Surge Pricing Moves Uber's Driver-Partners. In *Proceedings of the 2018 ACM Conference on Economics and Computation.* 3–3.

Hongyao Ma, Fei Fang, and David C Parkes. 2022. Spatio-temporal pricing for ridesharing platforms. *Operations Research* 70, 2 (2022), 1025–1041.

Haim Mendelson. 1985. Pricing computer services: Queueing effects. *Commun. ACM* 28, 3 (1985), 312–321.

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

Georgia Perakis and Charles Thraves. 2022. On a Variation of Two-Part Tariff Pricing of Services: A Data-Driven Approach. *Manufacturing & Service Operations Management* 24, 3 (2022), 1369–1387.

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

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

Duncan Rheingans-Yoo, Scott Duke Kominers, Hongyao Ma, and David C Parkes. 2019. Ridesharing with driver location preferences. In *Proceedings of IJCAI 2019*.

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

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

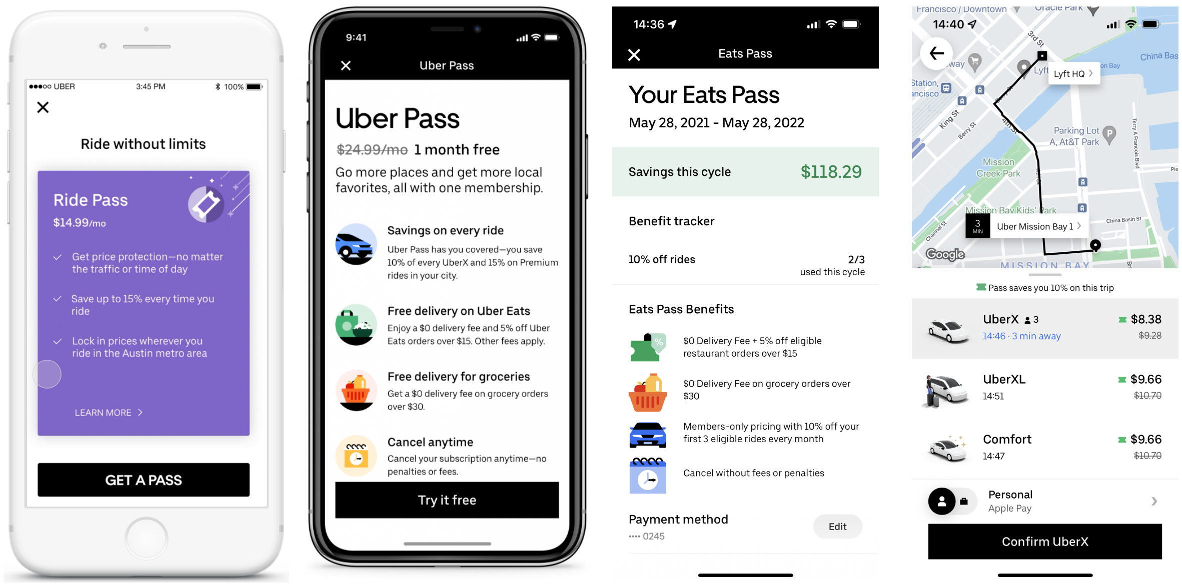
Chiwei Yan, Helin Zhu, Nikita Korolko, and Dawn Woodard. 2020. Dynamic pricing and matching in ride-hailing platforms. *Naval Research Logistics (NRL)* 67, 8 (2020), 705–724.

Qiuping Yu, Yiming Zhang, and Yong-Pin Zhou. 2022. Delay information in virtual queues: A large-scale field experiment on a major ride-sharing platform. *Management Science* 68, 8 (2022), 5745–5757.

Online Appendix

Appendix A: Subscription Offers

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



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

Figure 9 Additional examples of Uber's Subscription Programs.

1. Uber's "Ride Pass" with flat fares, offering unlimited rides at $3.49 for Pool and $6.49 for UberX (Figure 1b). This was offered in a few markets in 2017.[9]
2. Uber pass providing a fixed dollar discount: An email indicating that the rider has purchased a pass that takes $5 off every ride (Figure 1a). This was also offered in fall 2017.
3. Another Uber pass from 2018, offering 15% off every trip as well as "price protection" (Figure 9a), i.e. riders do not pay the dynamic "surge pricing" during times when demand exceeds supply.[10]
4. Guaranteed Price: Locked-in prices between two locations (and 15% off on other rides) for $9.99 every 28 days.[11]
5. Uber pass offering a fixed 10% discount off the fare of every ride, for $24.99 a month (Figure 9b).
6. Uber's Eats Pass offering 10% off for up to three rides each month, in addition to benefits on food delivery orders (Figure 9c). Figure 9d is the shows a screenshot from a subscriber's app, which indicates that the rider's trip price will be discounted by 10%.

[9] `https://medium.com/@jsbotto/have-you-heard-of-the-uber-ride-pass-3ebf2af24568`, accessed 02/08/2022.

[10] `https://www.uber.com/newsroom/ride-pass/`

[11] `https://mashable.com/article/lyft-uber-ride-subscription-plan`, accessed 01/26/2023

Figure 10 **A screenshot from the Chinese delivery platform Meituan for a flexible subscription. Under this subscription, consumers pay ¥15 a month to get ¥5 off for up to six orders (along with some other benefits). This suggests that carefully designed flexible subscriptions can be valuable to platforms.**

7. The UberOne program from 2022: 5% off every ride (together with a number of other benefits).[12]
8. Lyft's "Personal Plan": Lock in consistent pricing on a single preferred route for $7.99 per month.[10]
9. All-Access Plan: 30 rides up to $15 per ride, every 30 days for $299 (Figure 1c), offered in 2018.[13] A more expensive option was also tested, offering 60 rides a month for $399 a month.[14]
10. Lyft Pink: offers subscribes 15% off every ride, among a few other benefits for $20 a month 1d. Offered in 2019.[15]
11. Lyft Pink (updated version): For $9.99 each month, the subscription offers priority pickups and a 5% discount on premium rides.[16]
12. Viapass: A menu of subscriptions with different expiry dates, all of which offer some free rides each day and 10% off on subsequent rides. [17]

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

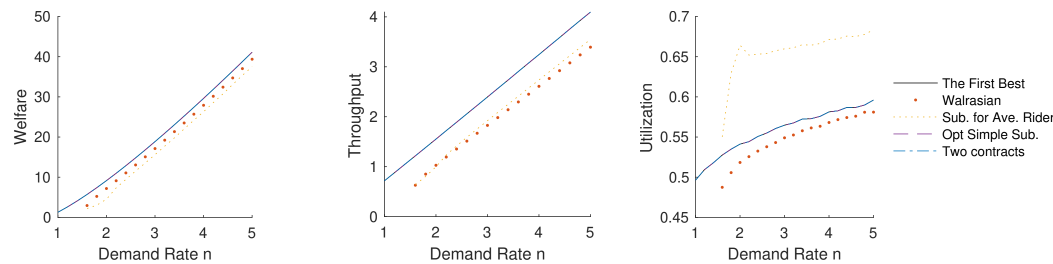
[13] `https://www.theverge.com/2018/10/16/17978626/lyft-monthly-subscription-plan-nationwide`

[14] `https://www.macrumors.com/2018/03/16/lyft-monthly-subscription-plans/`

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

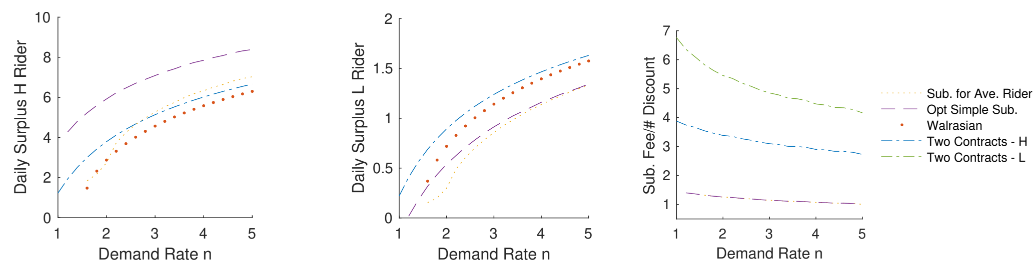
[16] `https://www.lyft.com/blog/posts/introducing-the-new-lyft-pink-membership`

[17] `https://support.ridewithvia.com/hc/en-us/articles/360003097371-What-is-the-ViaPass-`



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

Figure 11 **Social welfare, trip throughput, and driver time utilization under various benchmarks and mechanisms. Varying total demand rate n.**



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

Figure 12 **Average rider surplus for both types, and the subscription fee (per discount) under various mechanisms. Varying total demand rate n.**

Appendix B: Additional Simulation Results

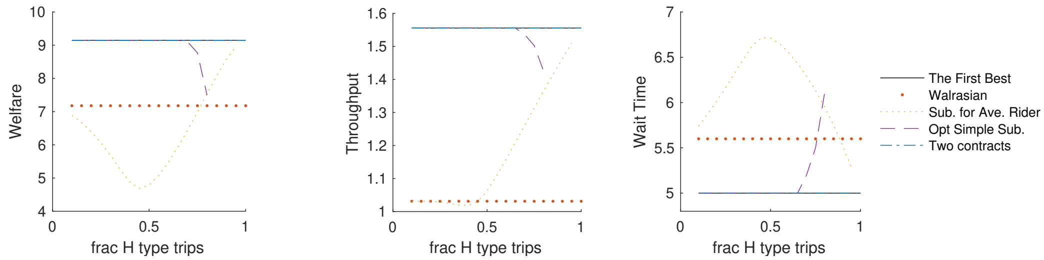
In this section, we include additional simulation results omitted from the body of the paper.

B.1. Varying Market Density

We first provide additional results on the impact of market density, for the settings studied in Section 5. Figures 11a and 11b plot the social welfare and trip throughput for each demand rate n-In addition, Figure 11c 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.

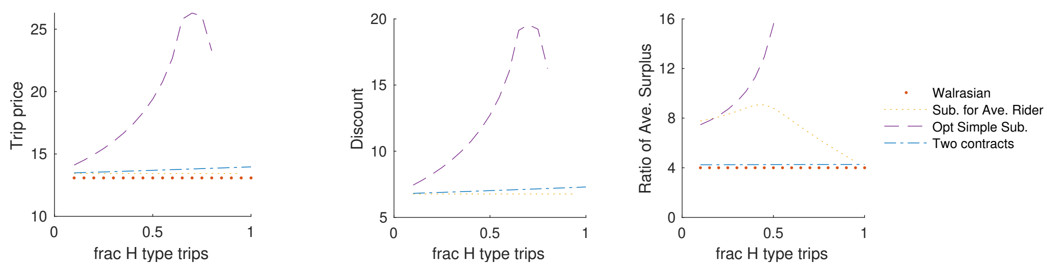
Figures 12a and 12b plot the average surplus per day for each rider type. The ratio between the two is included in Figure 8. Similar to what we have observed in Example 2, the two contract mechanism improves the surplus of both rider types in comparison to the outcome under the Walrasian equilibrium. 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 12c. 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.



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

Figure 13 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.



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

Figure 14 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.

B.2. Varying Rider Composition

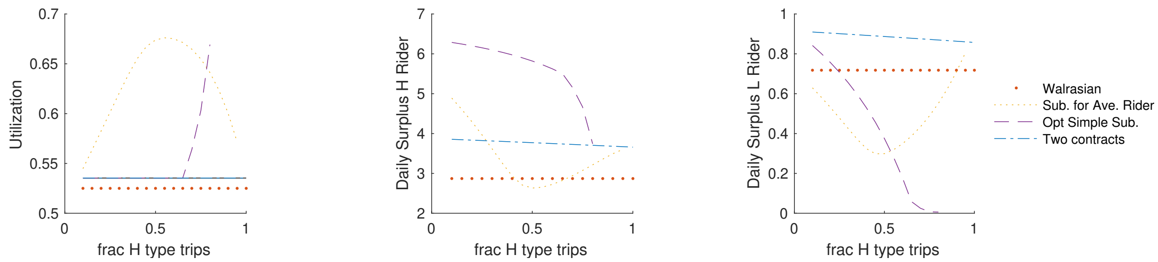
We now examine the impact of rider composition in markets where some riders take trips more often than the others. In this section, we study the same economy as that in Example 2, but vary rider composition such that the fraction of trips taken by the high type riders (under the first best) varies from 0 to 1, while keeping the arrival rate of riders who are interested in a ride fixed at 2 per unit of time.

Figure 13 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 14 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 15 provides driver utilization and the average daily surplus of the two types of riders.

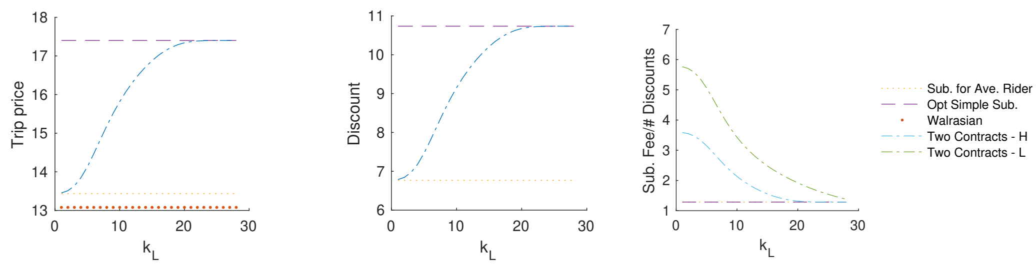
B.3. Varying Flexibility Level

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



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

Figure 15 **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.**



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

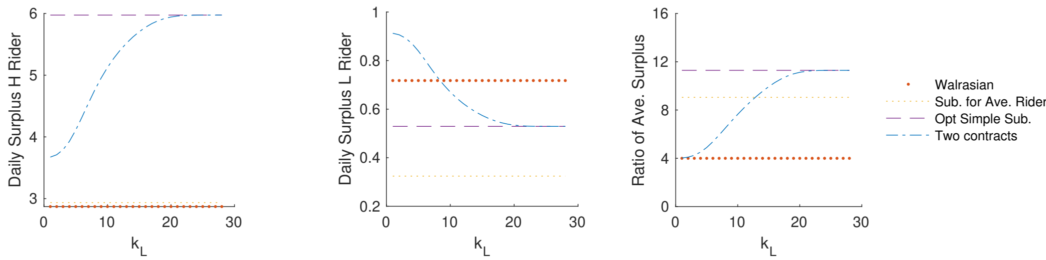
Figure 16 **Prices, discounts, and the subscription fee per discount under various benchmarks and mechanisms. Varying the flexibility level k_L of the two contract mechanism.**

k_L increases, flexibility decreases the L type riders must purchase a larger number of discounts at a time. When $k_L = T = 30$, the flexible subscription reduces to the simple subscription (i.e. there is no longer any flexibility).

Figure 16 presents the trip price, discount, and the subscription fee charged per each discount. Figure 17 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).[18] 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.

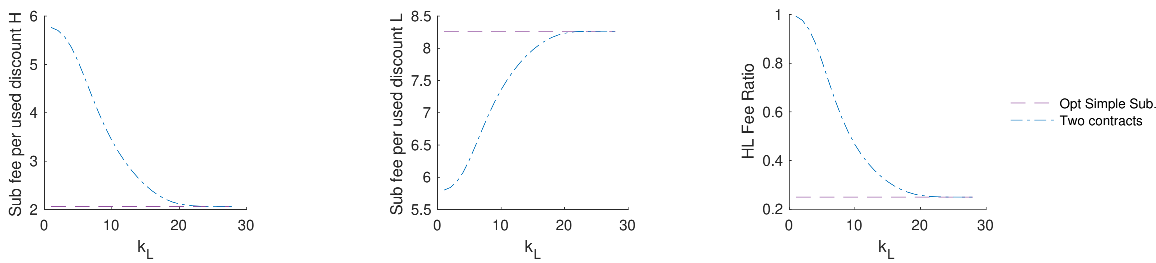
Figures 18a and 18b presents the average subscription fee paid by each rider type for each trip that the riders end up taking (e.g. if a rider pays \$60 a month and takes \$20 trips a month, then the subscription fee per trip is \$3). In other word, this is the fee paid by riders for each "used discount". The ratio of the two is plotted in Figure 18c. As we've discussed in the body of the paper, a smaller k_L i.e. more flexibility leads to better equity between the two types of riders.

[18] 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) Ave Surplus per Day - H. (b) Ave Surplus per Day - L. (c) Ratio of Ave. Surplus per Day.

Figure 17 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 18 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.

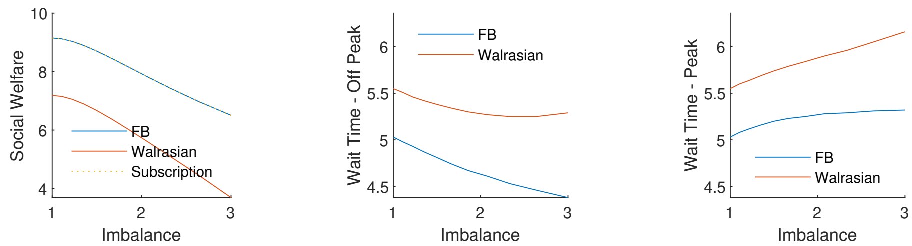
B.4. Varying Demand

Finally, we describe the setup and results for the environment where rider demand can vary over the course of the day. We observe that under some reasonable assumptions, simple subscriptions still maximize social welfare. In this example, we assume a fixed time interval which alternates between off peak and peak demand as described in Section 6.

First of all, we consider the same economy as described in Example 1 with driver costs $c = 1/3$ per minute, trip time $d = 15$ minutes and $\tau = 15$ and $\alpha = 1/3$ such that $\eta(O) = \tau O^{-\alpha}$. We refer the reader to Example 1 for a careful justification of this setup. We also assume that riders are *apriori* homogeneous and their values are uniformly distributed as per: $V^{\eta} \sim \mathrm{U}[0, v_{\max} - \beta\eta]$, with $v_{\max} = 55$ and $\beta = 5$.

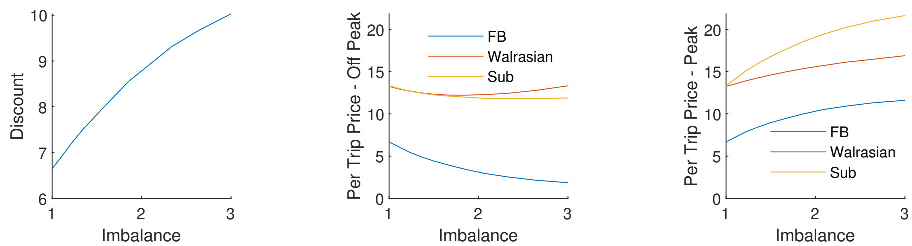
Next, we assume that the overall time horizon is divided into peak and off-peak hours, where each demand rate constitutes one half of the overall time. The aggregate rider mass is given by $n = 2$. Following our notation in Section 6, we assume that each of these riders request a ride in the off-peak hours with probability $\lambda_{op} = 0.25$. Fixing λ_{op}, we vary the demand rate λ_p during the peak hours and illustrate how the social welfare, wait times, and per trip price change as a function of the imbalance parameter given by $\frac{\lambda_p}{\lambda_{op}}$.

Figure 19 depicts how our key metrics vary with the imbalance parameter. First, we observe that more imbalance leads to a reduction in social welfare and a larger (multiplicative) gap between the Walrasian and the first best outcomes. However, simple subscriptions are still able to implement the first best as seen in



(a) Social Welfare. (b) Wait Time Off-Peak. (c) Wait Time Peak.

Figure 19 **Social welfare and wait times as a function of the imbalance, i.e., the ratio between the peak and off-peak demand rate. Simple subscriptions continue to implement the first best outcome even when demand is variable. Since the supply of drivers is fixed, the market experiences larger wait times during peak hours.**



(a) Discount. (b) Price Off-Peak. (c) Price Peak.

Figure 20 **Discount level (δ) for the simple subscription and per trip prices during the two demand periods as a function of the imbalance, i.e., the ratio between the peak and off-peak demand rate. In order to achieve the first best outcome, it is necessary to increase the per trip price from the status quo equilibrium during peak hours. However, this can be compensated by lowering the per trip price during off-peak hours. Since the discount level is independent of the demand, adjusting the per trip price ensures that the post discount price equals the marginal cost of the trip in each period.**

Figure 19a. In Figure 19b and 19c, we observe lower wait times in the off-peak hours compared to the peak hours, and the difference is exacerbated with imbalance. This can be explained by the fact that both periods host the same supply of drivers but the demand is larger during peak hours. Finally, Figure 20 shows the per trip price before and after subscriptions are implemented—we observe that in order to guarantee the optimal outcome, the platform is forced to raise the per trip price in the peak hours, but this is accompanied by a reduction in the off-peak price.

We conclude by reiterating that this is only a first step towards incorporating variable demand. In future work, one may wish to account for more complex features such as varying supply and new riders who only request a ride during peak hours. Naturally, it may not be possible to implement the first best outcome under more sophisticated assumptions. However, our example here illustrates that simple subscriptions can perform well under variable demand patterns.

Appendix C: Omitted Proofs

C.1. Proofs from Section 2 : Proof of Proposition 1

Proof: Let $x_{\text{opt}}, x_{\text{wal}}$ be the throughputs corresponding to the first best and 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

$$x_{\text{opt}} = n \cdot \Pr(V^\eta \geq c(d+\eta)).$$
$$x_{\text{wal}} = n \cdot \Pr(V^\eta \geq c(d+\eta+w_{\text{wal}})).$$

From here we have:

$$\begin{aligned}
&\mathsf{optSW}(\eta) - \mathsf{walSW}(\eta)\\
=&\left(n \cdot \mathbb{E}\left[V^\eta \cdot \mathbb{1}_{\{V^\eta \geq c(d+\eta)\}}\right] - c \cdot (x_{\text{opt}}(d+\eta)+O)\right) - \left(n \cdot \mathbb{E}\left[V^\eta \cdot \mathbb{1}_{\{V^\eta \geq c(d+\eta+w_{\text{wal}})\}}\right] - c \cdot (x_{\text{wal}}(d+\eta)+O)\right)\\
=&n \cdot \mathbb{E}\left[V^\eta \cdot \left(\mathbb{1}_{\{V^\eta \geq c(d+\eta)\}} - \mathbb{1}_{\{V^\eta \geq c(d+\eta+w_{\text{wal}})\}}\right)\right] - (x_{\text{opt}} - x_{\text{wal}})c(d+\eta)\\
=&n \cdot \mathbb{E}\left[V^\eta \cdot \mathbb{1}_{\{c(d+\eta+w_{\text{wal}}) > V^\eta \geq c(d+\eta)\}}\right] - (x_{\text{opt}} - x_{\text{wal}})c(d+\eta)\\
>&n \cdot c \cdot (d+\eta) \Pr\left(c(d+\eta+w_{\text{wal}}) > V^\eta \geq c(d+\eta)\right) - (x_{\text{opt}} - x_{\text{wal}})c(d+\eta)\\
=&n \cdot c \cdot (d+\eta)\left(\Pr\left(V^\eta \geq c(d+\eta)\right) - \Pr\left(V^\eta \geq c(d+\eta+w_{\text{wal}})\right)\right) - (x_{\text{opt}} - x_{\text{wal}})c(d+\eta)\\
=&(x_{\text{opt}} - x_{\text{wal}})c(d+\eta) - (x_{\text{opt}} - x_{\text{wal}})c(d+\eta) = 0
\end{aligned}$$

where the inequality is due to the following property: for any constants $a < b$

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

The intuition behind the above expressions is straight forward. In the first best outcome, all riders whose realized value is at least $c(d+\eta)$ contribute to social welfare. On the other hand, the Walrasian equilibrium only admits riders whose value is at least $c(d+\eta+w_{\text{wal}})$. Moreover, in both cases the aggregate system cost remains the same as the number of open drivers O depends only on the wait time η. This gives us the proposition. □

C.2. Proofs from Section 3: Proof of Equation 7

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

C.3. Proofs from Section 3.1: Proof of Theorem 2

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

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

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

$$n_0 = n \qquad n_1 = 0 \qquad x_0 = n\Pr(V^\eta \ge c(d+\eta)) \qquad x_1 = 0$$
$$y = x_0\cdot(d+\eta) + O(\eta) \qquad p = p_{\text{wal}}$$

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 2. $u^S(p,\eta) \ge u^R(p,\eta)$.

Proof: First, we have $u^R(p,\eta) = \mathbb{E}\left[(V^\eta - p_{\text{wal}})\,\mathbb{1}_{\{V^\eta \ge p_{\text{wal}}\}}\right]$. On the other hand, we have:

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

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

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

Lemma 3. *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 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:

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

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:

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

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. □ □

Appendix D: Heterogeneous Riders: Proofs and Additional Results

D.1. Simple Subscriptions under Heterogeneous Riders: When do they exist?

The purpose of this section is to reason about the welfare that simple subscriptions can achieve in the presence of heterogeneous riders. We will 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 3 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 8. A tuple $\left((n_i^S, n_i^R, x_i^S, x_i^R)_{i\in\{h,\ell\}}, y, p, \eta\right)$ 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 3. *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:*

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

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 3 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.

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

Lemma 4. *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 $n_i^R = 0$.*

We first establish the following claim before proving this lemma.

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

[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.

We now prove Lemma 4.

[Proof of Lemma 4] We shall now show that the claim implies the lemma. 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.

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 $\left((n_i^S, n_i^R, x_i^S, x_i^R)_{i \in \{h,\ell\}}, y, p, \eta\right)$. By Lemma 4, 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

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

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:

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

and, as we did in the proof of Theorem 1, 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

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

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 3 does not hold. The former holds by Observation 1 and since $x_{\text{opt}} = (n_h Q_h + n_\ell Q_\ell)\Pr(V^\eta \geq c(d+\eta)) > 0$, and the latter is given below:

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

D.2. Proofs from Section 4

Before proving our results from Section 4, we state and prove a simple observation.

For any η we let $r^\eta = \mathbb{E}[V^\eta - c(d+\eta) \mid V^\eta \geq c(d+\eta)]$ 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 $\mathsf{optSW}(\eta) = x_{\text{opt}}(r^\eta - cw_{\text{opt}}(\eta))$. In particular, since $\mathsf{optSW}(\eta) > 0$, then we have $r^\eta > cw_{\text{opt}}(\eta)$.*

D.2.1. 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

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

□

D.2.2. Proof of Equation 8

For the sake of exposition, we state a more formal claim on riders' steady state utilities.

Lemma 5. *Consider a ride of type $i \in \{h, \ell\}$ who subscribes to a flexible contract $C = (s, \delta, k, T)$ and requests a ride as long as $V^\eta \geq p - \delta + \tau$, where p denotes the per trip price and τ is some arbitrary threshold. Then, the daily surplus that this rider derives from the flexible contract in steady-state under wait time η is:*

$$u_i^C(p,\eta) = \mathbb{E}\left[(V_i^\eta - (p-\delta))\,\mathbb{1}_{\{V_i^\eta \geq p-\delta+\tau\}}\right] - s\cdot\frac{1}{\mathbb{E}[D_i^C(p,\eta)]}, \tag{10}$$

We note that the above expression implicitly assumes immediate renewals upon expiry of the current subscription.

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 $\mathbb{E}[D_i^C(p,\eta)]=T$. 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

$$\lim_{t\to\infty}\frac{\sum_{j=1}^{t}\mathbb{1}_{\{A_j\}}}{t}=\pi_i(k,T)\tag{11}$$

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

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

Finally, since (11) holds with probability 1, then in particular we have

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

as desired. To conclude, we use the observation that rider's steady state utility can be written as:

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

□

D.3. Proof of Theorem 3

Before proving the result, we review the rider behavior in the face of a flexible subscription. In particular, we assume that riders follow the utility-maximizing option among the myopic and fixed discount policies. Under the myopic policy, a rider who has subscribed to a flexible subscription will use a coupon and request a ride as long as $V^\eta \geq c(d+\eta)$. On the other hand, the fixed discount rider is more conservative and reserves a ride as long as $V^\eta \geq c(d+\eta) + \frac{s}{k}$, where s denotes the subscription fee and k, the number of discounts.

We say that a rider employs a straightforward policy if they pick the utility maximizing option among the myopic and fixed discount policies. Our proof follows.

Proof: We begin by outlining the key steps involved. In Part I of the proof, we first identify a flexible contract (s, δ, k, T) that satisfies the following properties:

(a) The low-type rider derives non-negative utility from behaving myopically.

(b) The low-type rider's utility from employing a fixed discount policy is strictly negative.

(c) Under the conditions specified in the theorem statement, the high-type rider's utility from the flexible subscription is smaller than or equal to their utility in the Walrasian outcome.

After we construct such a flexible subscription, we then impose (in Part II) the budget balance constraint by decreasing the subscription fee s so that the total rider utility matches the social welfare in the first best outcome. Finally, in Part III, we construct a simple subscription so that the high-type rider is indifferent between the simple and the flexible subscriptions. Further, the low-type rider chooses the flexible contract and continues to behave myopically. We conclude by pointing out that this maximizes social welfare and is budget balanced. We note that our proof below holds for $T \to \infty$, but it is possible to use the same approach to construct a contract with a finite (k, T) where the same guarantees hold.

Before proceeding with the proof, we define some simple but useful notation:

- $q_h = Q_h \Pr\left(V^\eta \geq c(d+\eta)\right)$,
- $q_h' = Q_h \Pr\left(V^\eta \geq c(d+\eta+\frac{s}{k})\right)$ when s and k are clear from the context,
- $q_\ell = Q_\ell \Pr\left(V^\eta \geq c(d+\eta)\right)$, and
- $q_\ell' = Q_\ell \Pr\left(V^\eta \geq c(d+\eta+\frac{s}{k})\right)$ when s and k are clear from the context.

Part I - Construction of the Flexible Subscription Consider a flexible subscription C where:

$$\frac{s}{k} = r^\eta = \mathbb{E}\left[(V^\eta - c(d+\eta)) \mid V^\eta \geq c(d+\eta)\right], \quad \text{(and)} \quad k = q^* T.$$

We select $q^* = q_\ell - \epsilon$ for some sufficiently small $\epsilon > 0$ i.e., q is slightly smaller than the probability that a low-type rider requests a ride. We choose a sufficiently high expiry T, and set δ so that the real-time price after applying the discount is $c(d+\eta)$.

As a result of the subscription parameters, the low-type rider uses up all their discounts by the expiry date T when behaving myopically. Formally, let $T \to \infty$ and $u^C_{i,m}(p,\eta)$ and $D^C_{i,m}(p,\eta)$ denote respectively, the expected utility from subscribing to the contract C and the (random) duration between renewals for a rider of type $i \in \{\ell, h\}$ after subscribing to the contract C and behaving myopically. Then, we have that:

$$\lim_{T\to\infty} u^C_{\ell,m}(p,\eta) = \lim_{T\to\infty} \left(Q_\ell \mathbb{E}\left[(V^\eta - c(d+\eta))\, \mathbb{1}_{\{V^\eta \geq c(d+\eta)\}}\right] - \frac{s}{\mathbb{E}[D^C_{i,m}(p,\eta)]} \right)$$

$$
\begin{aligned}
&= \lim_{T\to\infty} \left(Q_\ell \Pr(V^\eta \geq c(d+\eta)) \mathbb{E}\left[V^\eta - c(d+\eta) \mid V^\eta \geq c(d+\eta)\right] - \frac{s}{k/q_\ell} \right) \\
&= \lim_{T\to\infty} \left(q_\ell r^\eta - q_\ell r^\eta\right) \\
&= 0.
\end{aligned}
$$

Therefore, the low-type seller derives exactly zero utility from the myopic subscription while behaving myopically. Note that in the above expressions, we utilized the fact that $\lim_{T\to\infty} E[D^C_{\ell,m}(p,\eta)] = \frac{k}{q_\ell}$, i.e., with high probability the rider uses up all the discounts. This comes from Lemma 6.

We now show that this is myopic rider's preferred strategy as employing the fixed discount policy leads to negative utility. As with our previous notation, we use $u^C_{i,f}(p,\eta)$ and $D^C_{i,f}(p,\eta)$ to denote type i rider's utility and duration under the fixed discount strategy. We note that under this strategy the rider requests a ride only if $V^\eta \geq c(d+\eta) + \frac{s}{k}$. Therefore, we have that:

$$
\begin{aligned}
\lim_{T\to\infty} u^C_{\ell,f}(p,\eta) &= \lim_{T\to\infty} \left(Q_\ell \mathbb{E}\left[(V^\eta - c(d+\eta))\, \mathbb{1}_{\{V^\eta \geq c(d+\eta)+\frac{s}{k}\}}\right] - \frac{s}{\mathbb{E}[D^C_{\ell,f}(p,\eta)]} \right) && (12) \\
&= \lim_{T\to\infty} \left(Q_\ell \mathbb{E}\left[(V^\eta - c(d+\eta))\, \mathbb{1}_{\{V^\eta \geq c(d+\eta)+\frac{s}{k}\}}\right] - \frac{s}{k/(q_\ell-\epsilon)]} \right) && (13) \\
&\leq \lim_{T\to\infty} \left(q' r^\eta - (q_\ell - \epsilon) r^\eta\right) && (14) \\
&< 0.
\end{aligned}
$$

Let us go over the above expressions carefully. First, from Lemma 6, we have that $\lim_{T\to\infty} E[D^C_{\ell,f}(p,\eta)] = T = \frac{k}{q_\ell-\epsilon}$, which we apply in Equation (12) to get Equation (13). Next, in Equation (13), we used the fact that

$$
\mathbb{E}\left[(V^\eta - c(d+\eta))\, \mathbb{1}_{\{V^\eta \geq c(d+\eta)+\frac{s}{k}\}}\right] \leq \mathbb{E}\left[(V^\eta - c(d+\eta))\, \mathbb{1}_{\{V^\eta \geq c(d+\eta)\}}\right].
$$

In this regard, we defined

$$
q' = Q_\ell \frac{\mathbb{E}\left[(V^\eta - c(d+\eta))\, \mathbb{1}_{\{V^\eta \geq c(d+\eta)+\frac{s}{k}\}}\right]}{\mathbb{E}\left[(V^\eta - c(d+\eta)) \mid V^\eta \geq c(d+\eta)\right]} = Q_\ell \Pr(V^\eta \geq c(d+\eta)) \frac{\mathbb{E}\left[(V^\eta - c(d+\eta))\, \mathbb{1}_{\{V^\eta \geq c(d+\eta)+\frac{s}{k}\}}\right]}{\mathbb{E}\left[(V^\eta - c(d+\eta))\, \mathbb{1}_{\{V^\eta \geq c(d+\eta)\}}\right]}.
$$

It is not hard to see that $q' < q_\ell$ and therefore, for a sufficiently small $\epsilon > 0$, it must be the case that $q' < q_\ell - \epsilon$. Going back to Equation (14), we now see why the expression must be strictly negative. Now we have constructed a subscription under which the low-type reader strictly prefers being myopic and derives non-negative utility for doing so.

Next, we move on to the high type rider. We begin by proving that under the two conditions specified in the statement of Theorem 3, the high-type rider's surplus under contract C is smaller than or equal to their surplus under the Walrasian outcome. We consider two cases. First, if the high-type rider behaves myopically this follows trivially since

$$
\begin{aligned}
\lim_{T\to\infty} u^C_{h,m}(p,\eta) &= \lim_{T\to\infty} \left(Q_h \mathbb{E}\left[(V^\eta - c(d+\eta))\, \mathbb{1}_{\{V^\eta \geq c(d+\eta)\}}\right] - \frac{s}{\mathbb{E}[D^C_{h,m}(p,\eta)]} \right) \\
&= \lim_{T\to\infty} \left(Q_h \Pr(V^\eta \geq c(d+\eta)) \mathbb{E}\left[(V^\eta - c(d+\eta)) \mid V^\eta \geq c(d+\eta)\right] - \frac{s}{k/q_h} \right) \\
&= 0,
\end{aligned}
$$

and the rider's utility from the Walrasian is non-negative. Once again, we used the fact that $\lim_{T\to\infty} \mathbb{E}[D^C_{h,m}(p,\eta)] = k/q_h$, where $q_h = Q_h \Pr\left(V^\eta \geq c(d+\eta)\right)$. Given this, we focus on proving the desired property when the high-type rider employs the fixed discount policy. Mathematically, we need to prove that $\lim_{T\to\infty} u^C_{h,f}(p,\eta) < Q_h \mathbb{E}\left[(V^\eta - c(d+\eta+w_{\text{wal}}))\,\mathbb{1}_{\{V^\eta \geq c(d+\eta+w_{\text{wal}})\}}\right]$. The high-type rider's utility under the fixed discount policy can be simplified as follows:

$$\begin{aligned}
\lim_{T\to\infty} u^C_{h,f}(p,\eta) &= \lim_{T\to\infty}\left(Q_h \mathbb{E}\left[(V^\eta - c(d+\eta))\,\mathbb{1}_{\{V^\eta \geq c(d+\eta)+\frac{s}{k}\}}\right] - \frac{s}{\mathbb{E}[D^C_{h,f}(p,\eta)]}\right)\\
&\leq Q_h \mathbb{E}\left[(V^\eta - c(d+\eta))\,\mathbb{1}_{\{V^\eta \geq c(d+\eta)+\frac{s}{k}\}}\right] - \frac{s}{k/q'_h}\\
&= Q_h \mathbb{E}\left[\left(V^\eta - c(d+\eta) - \frac{s}{k}\right)\mathbb{1}_{\{V^\eta \geq c(d+\eta)+\frac{s}{k}\}}\right],
\end{aligned}$$

where w_{wal} denotes drivers' idle time in-between trips at the Walrasian equilibrium given wait time η and $q'_h = Q_h \Pr(V^\eta \geq c(d+\eta) + \frac{s}{k})$. To prove this, it is sufficient to show that $c(d+\eta) + \frac{s}{k}$ is (weakly) greater than the Walrasian equilibrium price $p_{\text{wal}} = c(d+\eta) + cw_{\text{wal}}$, as this would imply $\frac{s}{k} > cw_{\text{wal}}$. We now prove this below under any one of the two conditions in the theorem statement.

1. **Large Rider Population**: First, it is not hard to see that as $n \to \infty$, $w_{\text{wal}} \to 0$. Therefore, there must exist some $\bar{n}$ such that $cw_{\text{wal}} < \frac{s}{k}$ for $n > \bar{n}$.

2. **Convex Inverse Hazard Rate**: Mathematically, the convex inverse hazard rate implies that the function $\frac{1-F^\eta}{f^\eta(v)}$ is convex. Recall that several popular distributions such as uniform, exponential, and Pareto satisfy this condition. Our goal here is to prove that $\frac{s}{k} \geq p_{\text{wal}} - c(d+\eta)$. Before proving the result, it is convenient to show a simple upper bound on the Walrasian price and use this bound as a proxy to prove the desired result. First, let us define

$$\xi(p) \triangleq \mathbb{E}\left[(p - c(d+\eta))\cdot \mathbb{1}_{\{V^\eta \geq p\}}\right]. \tag{15}$$

Intuitively, $\xi(p)$ is the extra revenue the platform collects from the rider (on top of the marginal cost of her trip), when the per trip price is p. At wait time η, the Walrasian equilibrium price is the smallest price p such that the total extra revenue covers the cost of idle drivers, i.e.

$$p_{\text{wal}} = \inf\{p \geq 0 \mid n\xi(p) - cO \geq 0\}.$$

When the Walrasian equilibrium exists, it must be the case that $\sup_{p\geq 0} n\xi(p) - cO \geq 0$. The maximizer of the expression $n\xi(p) - cO$, which is effectively the maximizer of $\xi(p)$ therefore must be weakly above p_{wal}. In other words, we have that:

$$p_{\text{wal}} \leq p^* = \arg\max_{p\geq 0}\{\xi(p)\}.$$

In what follows, we prove that $\frac{s}{k} \geq p^* - c(d+\eta)$, which is sufficient to show the desired property. Consider the following transformation of variable where $z = V^\eta - c(d+\eta)$. Under this new variable, our original condition (i.e., $\frac{s}{k} \geq p^*$) can be rewritten as

$$E[z \mid z \geq 0] \geq \arg\max_x\{x \Pr(z \geq x)\}, \tag{16}$$

where $x = p - c(d+\eta)$ and $z \geq x$ is equilvalent to $V^\eta \geq p$. Let $\bar{h}(x) = \frac{1-F^\eta(x)}{f^\eta(x)}$. Then, the right hand side of Equation (16) can be written in terms of the inverse hazard rate since $z^* \triangleq \arg\max_x\{x \Pr(z \geq x)\}$

also satisfies $h(z^*) = z^*$. This can be obtained by differentiating the expression $x\Pr(z \geq x)$ and setting the derivative to zero. Similarly, we can characterize the left hand side in terms of the hazard rate as follows:

$$\begin{aligned}
E[z \mid z \geq 0] &= \frac{1}{\Pr(z \geq 0)} \int_{z=0}^{z_{max}} (1 - F^\eta(z))dz \\
&= \frac{1}{\Pr(z \geq 0)} \int_{z=0}^{z_{max}} \bar{h}(z) f(z) dz \\
&= \frac{1}{\Pr(z \geq 0)} E[\bar{h}(z) \mathbb{1}_{z \geq 0}] \\
&= E[\bar{h}(z) \mid z \geq 0] \\
&\geq \bar{h}(E[z \mid z \geq 0]),
\end{aligned}$$

where the final inequality is due to Jensen's inequality and our assumption that $\bar{h}(z)$ is convex. Let $\tilde{z} = E[z \mid z \geq 0]$. Then, we have shown that $h(\tilde{z}) \leq \tilde{z}$. On the other hand, we know that $h(z^*) = z^*$. It then immediately follows that $\tilde{z} \geq z^*$, which gives us the desired result. To see why, recall that $\bar{h}(z)$ is convex and $\bar{h}(0) \geq 0$. As a result, the function can only attain a value of zero at no more than two values of z, one of which happens to be z^*. Since z^* is the 'revenue-maximizing' value of z, it is easy to see that $\bar{h}(z) \geq 0$ for all $z \leq z^*$. This completes our argument that $\tilde{z} \geq z^*$.

Finally, we know that the rider utility under Walrasian is strictly smaller than their utility under the first best outcome, i.e.,

$$\mathbb{E}\left[(V^\eta - c(d + \eta + w_{\text{wal}}))\, \mathbb{1}_{\{V^\eta \geq c(d+\eta+w_{\text{wal}})\}}\right] < \mathbb{E}\left[(V^\eta - c(d + \eta + w_{\text{opt}}))\, \mathbb{1}_{\{V^\eta \geq c(d+\eta)\}}\right].$$

To conclude, we have constructed a subscription C with the three desired properties outlined earlier.

Part II - Lowering the Subscription Fee to Ensure Budget Balance Next, we decrease the subscription fee to s_L so that the total utility of the two rider types matches their total surplus under the first best outcome. Specifically, under the original contract C we know that:

$$\begin{aligned}
\lim_{T \to \infty} \left(n_\ell u^C_{\ell,m}(p, \eta) + n_h \max_{j \in \{m,f\}} (u^C_{h,j}(p, \eta) \right) &= n_h \max_{j \in \{m,f\}} (u^C_{h,j}(p, \eta)) && (17) \\
&< n_h Q_h \mathbb{E}\left[V^\eta - c(d + \eta + w_{\text{opt}}) \mathbb{1}_{\{V^\eta \geq c(d+\eta)\}}\right]. && (18)
\end{aligned}$$

Note that $u^C_{\ell,m}(p, \eta) = 0$ under the contract. Next, starting with C, we decrease the subscription fee s so that the above inequality becomes an equality. In other words, let s_L denote the subscription fee at which:

$$\lim_{T \to \infty} \left(n_\ell u^C_{\ell,m}(p, \eta) + n_h \max_{j \in \{m,f\}} (u^C_{h,j}(p, \eta)) \right) = (n_\ell Q_\ell + n_h Q_h) E\left[V^\eta - c(d + \eta + w_{\text{opt}}) \mathbb{1}_{\{V^\eta \geq c(d+\eta)\}}\right] \quad (19)$$

Such a value of s_L must necessarily exist since both $(u^C_{h,j}(p, \eta)$ and $\max_{j \in \{m,f\}}(u^C_{h,j}(p, \eta))$ increases monotonically as we decrease s. We refer to the new contract (s^L, k^L, δ, T) as L. To conclude the second portion of our proof, we now show that the low-type rider continues to employ myopic behavior under the contract C. Towards this goal, note that:

$$\begin{aligned}
\lim_{T \to \infty} u^L_{\ell,m}(p, \eta) &= \left(q_\ell r^\eta - q_\ell \frac{s^L}{k} \right) \\
\lim_{T \to \infty} u^L_{\ell,f}(p, \eta) &= \left(q' r^\eta - (q_\ell - \epsilon) \frac{s^L}{k} \right)
\end{aligned}$$

Since $q' < q_\ell$, this implies that for some sufficiently small $\epsilon > 0$, the low-type rider prefers being myopic over the fixed discount policy.

Part III - Constructing the simple subscription H Unfortunately, this new contract L may not implement the optimal outcome if the high-type rider employs the fixed discount strategy. To compensate for this, we introduce a second, simple subscription $H=(s^H,k^H,\delta,T)$ with $k^H=T$ and s^H chosen carefully so that the high-type rider is indifferent between the two contracts, i.e.,

$$s^H = T\left(Q_h E\left[V^\eta - c(d+\eta)1_{\{V^\eta \geq c(d+\eta)\}}\right] - \max_{j\in\{m,f\}}(u^L_{h,j}(p,\eta))\right).$$

To complete the proof, it is sufficient prove that:

- The low-type rider continues to prefer the contract L (and as a result, behaves myopically upon subscription).
- The contracts are budget balanced. However, this follows immediately from Equation (19) as the utility of both riders remains the same after the introduction of the contract H.

Consider the low-type rider. Their difference in utility between the two contracts is given by:

$$\lim_{T\to\infty}\left(u^L_{\ell,m}(p,\eta) - u^H_{\ell,m}(p,\eta)\right) = \frac{s^H}{T} - q_\ell \frac{s^L}{k}. \tag{20}$$

To show that the quantity above is non-negative, we consider two cases. First, suppose that under the contract L, the high-type rider maximizes their utility with a myopic policy. In that case, we have:

$$\frac{s^H}{T} = q_h \frac{s^L}{k}.$$

Plugging this back into Equation (20), we get that the utility is at least zero since $q_h \geq q_\ell$.

In the second case, suppose that the high-type rider employs the fixed discount policy. Then, we have that:

$$\begin{aligned}
\frac{s^H}{T} &= Q_h E\left[V^\eta - c(d+\eta)1_{\{V^\eta\geq c(d+\eta)\}}\right] - Q_h E\left[V^\eta - c(d+\eta)1_{\{V^\eta\geq c(d+\eta)+\frac{s^L}{k}\}}\right] + \max(q'_h,q^*)\frac{s^L}{k}\\
&= \frac{Q_h}{Q_\ell}\left(Q_\ell E\left[V^\eta - c(d+\eta)1_{\{V^\eta\geq c(d+\eta)\}}\right] - Q_\ell E\left[V^\eta - c(d+\eta) - \frac{s^L}{k}1_{\{V^\eta\geq c(d+\eta)+\frac{s^L}{k}\}}\right]\right) + \max(q'_h,q^*)\frac{s^L}{k}\\
&\geq \frac{Q_h}{Q_\ell}\cdot\frac{s^L}{k}\left(q_\ell - q^*\right)) + \max(q'_h,q^*)\frac{s^L}{k}\\
&\geq q_\ell\frac{s^L}{k}.
\end{aligned}$$

Since $\frac{s^H}{T} \geq q_\ell \frac{s^L}{k}$, it follows that the low-type rider prefers the L-contract. Therefore, the proof follows. □

Lemma 6. *Consider a contract $C=(s,k,\delta,T)$ such that $k=q^*T$ for $q\in(0,1]$. The expected renewal time $E[D]$ for a rider requesting a ride with probability q, conditional on subscribing is given by:*

$$E[D] = \begin{cases} q\frac{s}{k} & \text{if } q > q^* \\ \frac{s}{T} & \text{if } q < q^* \end{cases}$$

Proof: The proof follows from a simple characterization of whether or not a rider gets to use all their discounts within the expiry period T. First, consider the case where $q>q^*$ and let $k^*=\frac{qs}{k}<T$. Let N_t be a random variable denoting the number of discounts the rider is able to use by day $t\in[T]$ after subscription. Then, for any $\epsilon>0$, we have that: $\lim_{T\to\infty} N_{k^*+\epsilon}=k$, which follows from applying a concentration inequality on the sum of k geometric variables. Therefore, the first result follows.

In a similar vein, consider N_T when $q<q^*$. Once again, we have that: $\lim_{T\to\infty} N_T<k$. To see why let G_k be the sum of k geometric random variables with mean $\frac{1}{q}$. Then $\lim_{T\to\infty} Pr(G_k>N_T)=1$. As a result, the second statement also follows. □