Pricing in Ride-share Platforms: A Queueing-Theoretic Approach

Carlos Riquelme, Siddhartha Banerjee, Ramesh Johari
Department of Management Science and Engineering, Stanford University, Stanford, CA 94025
rikel@stanford.edu, sidb@stanford.edu, ramesh.johari@stanford.edu

We study optimal pricing strategies for ride-sharing platforms, such as Lyft, Sidecar, and Uber. Analysis of pricing in such settings is complex: On one hand these platforms are two-sided – this requires economic models that capture the incentives of both drivers and passengers. On the other hand, these platforms support high temporal-resolution for data collection and pricing – this requires stochastic models that capture the dynamics of drivers and passengers in the system.

In this paper we build a queueing-theoretic economic model to study optimal platform pricing. In particular, we focus our attention on the value of *dynamic pricing*: where prices can react to instantaneous imbalances between available supply and incoming demand. We find two main results: We first show that performance (throughput and revenue) under any dynamic pricing strategy cannot exceed that under the optimal *static* pricing policy (i.e., one which is agnostic of stochastic fluctuations in the system load). This result belies the prevalence of dynamic pricing in practice. Our second result explains the apparent paradox: we show that dynamic pricing is much more *robust* to fluctuations in system parameters compared to static pricing. Thus dynamic pricing does not necessarily yield higher performance than static pricing – however, it lets platforms realize the benefits of optimal static pricing, even with imperfect knowledge of system parameters.

Key words: Ride-Sharing, Dynamic Pricing, Matching Markets, Queueing Networks

1. Introduction

In this paper we study *ride-sharing platforms* such as Lyft, Sidecar, and Uber. Since their founding in the last several years, these platforms have experienced extraordinary growth. At their core, the platforms reduce the friction in matching and dispatch for transportation. A typical transaction on these platforms is as follows: a potential rider opens the app on her phone and requests a ride; the system matches her to a nearby driver if one is available, else blocks the ride request. These platforms typically do not employ drivers, but deliver a share of the earnings per ride to the driver, to incentivize driver participation.

Ride-sharing platforms are thus two-sided markets: drivers on one side, and passengers on the other. As a consequence, a central goal of the platform's intermediation is to calibrate supply and demand relative to each other, while ensuring relatively high satisfaction to both sides. A key tool used by these platforms to manage supply and demand is *dynamic pricing* – the platform can adjust ride prices in real-time, to react to changes in ride requests and available drivers. The central focus of this paper is in understanding how these two features of ride-sharing platforms – their two-sided nature, and the ability to price based on real-time state – influence the volume of trade and the revenue of the platform.

To capture the fast-timescale dynamics of ride-sharing platforms, we employ a queueing theoretic approach. Our primary modeling contribution lies in combining this queueing model for the underlying stochastic dynamics, with an equilibrium analysis that captures incentives of both drivers and passengers, as well as throughput/revenue maximization by the platform. The general model we consider is one where a geographic area is divided into *regions*. Each ride involves a driver picking up a passenger in one region, and dropping her off in another. For simplicity, we analyze this model first for a single region; subsequently, using tools from classical queueing theory, we generalize some of our results to networks of regions.

Key to our model is an intrinsic timescale separation in the behavior of drivers and passengers. In ride-sharing platforms driven by mobile apps, passengers are typically sensitive to the *immediate* price of the ride that they request. If this price is too high, price-sensitive customers will abandon. On the other hand, drivers are typically sensitive to the *average* wages they earn over a longer period of time, usually several days or a week. Thus, drivers often select certain periods of time when they are "active" during the week, and adjust their activity levels based on their assessment of earnings during the last week. This timescale separation has two implications on agent dynamics in our model:

1. We model passengers as living for one ride, while drivers (probabilistically) re-join the platform after each ride. Thus, drivers have "long" lifetimes in our model, while passengers have "short" lifetimes.
2. We model passengers' entry decisions as being based on the immediate price they are shown, which is a function of the current state of the region where the passenger opens her app; if the price is below a passenger-specific reservation value, then the passenger requests a ride. By contrast, we assume that drivers' entry decisions are made by comparing their expected lifetime earnings with their expected lifetime; if this wage rate exceeds a driver-specific reservation wage rate, the driver chooses to enter the platform.

Our model allows us to concisely describe the endogenously-determined equilibrium arrival-rates of drivers and passengers, for a given pricing policy of the platform operator. Using this specification, we then study the behavior of revenue under different pricing policies employed by the platform. We begin with a seemingly basic question: *is there value in dynamic pricing*? In other words, does the platform benefit by being able to adjust prices in response to changes in system state?

We focus in this paper on a class of dynamic pricing policies commonly used in practice: *threshold* policies, where the platform raises the price whenever the number of available drivers in a region falls below a threshold. Somewhat surprisingly, in our first main result, we find in a broad range of conditions that *the platform cannot increase throughput or revenue by employing threshold dynamic pricing.*

This result is counterintuitive, and belies the fact that dynamic pricing is prevalent in practice across most ride-sharing platforms. Our second main result reveals a significant benefit that dynamic pricing holds over static pricing: *robustness*. Specifically, suppose the system operator chooses the optimal threshold dynamic (resp., static) pricing policy given fixed system parameters (arrival rates, service rates, and preference distributions of passengers and drivers). Now suppose the true parameters deviate from the original postulated environment. We show that *dynamic pricing maintains a much higher share of the optimal throughput relative to the optimal static pricing*; in other words, it is much less brittle to lack of knowledge of system parameters.

Intuitively, threshold dynamic pricing helps discover the "correct" static price, by mixing between the high price (i.e., low driver availability) and low price (i.e., low driver availability) regimes. A similar robustness phenomenon also holds for revenue in a *supply-constrained* regime, which arguably is the case for ride-sharing platforms. Moreover, given a surfeit of drivers, we show that the platform can maximize its revenue by charging a (static) marked-up price relative to the market-clearing price; dynamic pricing however may perform poorly in this setting.

Finally, we generalize much of our analysis and results to networks of regions. Our model allows us to capture geographical variations in driver availability/demand. Nevertheless, we demonstrate that the optimal region-dependent threshold dynamic pricing policy cannot yield higher throughput than the optimal region-dependent static pricing policy. We conjecture that a similar robustness result to that derived in the single region setting holds for networks as well.

The remainder of the paper is organized as follows. We introduce our queueing model; the strategic model of drivers and passengers; the pricing model of the platform; and our equilibrium definition in Section 2. Next, in Section 3, we study the optimal static pricing policy of the platform for a single region. In Section 4, we show that the optimal threshold dynamic pricing policy cannot exceed the performance of the optimal static pricing policy, whether we are optimizing for throughput or revenue. In Section 5, we establish that dynamic pricing is much more robust to changes in system parameters than static pricing. In Section 6, we extend our analysis to general networks. For ease of exposition, we defer the proofs of our theorems, along with some supporting results, to the Appendix.

1.1. Related Work

Our paper sits at the intersection of several active areas of research; we briefly describe each of these below.

Matching queues. In most queueing models, servers are fixed, and jobs arrive and depart. By contrast, in our model, both passengers (jobs) and drivers (servers) experience queueing phenomena. A recent literature on *matching queues* has studied systems with this feature; see, e.g., [1, 15] for an example of this modeling approach. However, these works do not consider any strategic behavior by the agents.

Strategic queueing models. On the other hand, a long line of research in queueing theory has studied strategic behavior in queueing systems; see [11] for early work in this area and [8] for an overview of these models. Typically, these works consider systems with a fixed number of servers, who serve arriving customers who are sensitive to price and delay. In such settings, dynamic pricing affects the instantaneous arrival rates of customers; moreover, the Our work studies strategic behavior in a queueing system, but where the servers (i.e., the drivers) are also strategic. In this respect our work is closest to the recent paper of [7].

Large-scale matching markets. The combination of queueing and strategic behavior can be challenging to analyze, and so in this paper we employ a *large market limit* to yield a tractable yet insightful model. This approach has grown in importance in recent years; see, e.g., [10, 3] for early examples in the matching market literature, and [2] for an example of this approach for dynamic matching markets.

Revenue management. The platform plays a key role in our paper: we analyze optimal pricing strategies both for a monopolist, and to optimize social welfare. The comparison of static and dynamic pricing policies is similar in spirit to revenue management; see, e.g., [14, 4] for an overview of pricing approaches, and [6] for a study of dynamic pricing based on current inventory levels. Surprisingly, several of these works obtain results which are similar in spirit to ours – in particular, the fact that static pricing policies are optimal in a large-system limit. However, a critical feature of our work is in considering two-sided platforms, where both sides react to the pricing, albeit on different timescales – thus, instead of a dynamic programming formulation, we need to analyze the equilibrium of the resulting system.

Two-sided platforms. From an economic standpoint, our paper is closest to the literature on two-sided platforms. See [12, 13] for a summary of this literature, and [16] for a unified approach to price structure for two-sided platforms. Our paper contributes to this literature by studying optimal pricing to match dynamically changing supply and demand.

2. A Single Region Model

In this section we introduce our model of a ride-sharing platform. We begin by describing an underlying queueing model that captures the stochastic dynamics of the system, assuming exogenously given behavior of drivers and passengers. We then describe the platform's choice of pricing policy, including both static and dynamic policies. Next, we describe the strategic decisions of drivers and passengers. We conclude by describing equilibrium in our system, taking into account incentives of the agents and the pricing policy of the platform. For pedagogical reasons, the model and the initial analysis focusses on the special case of a single geographic region; this approximates a setting where the platform provider does not set different prices based on location within a city. In Section 6, we generalize our model and results to a *network* of regions.

2.1. Queueing Dynamics: A Single Region

Our queueing model requires three key components: (1) a network of *regions* (e.g., of a city); (2) dynamics of *drivers* in the system; and (3) dynamics of *passengers* in the system. As noted above, we focus in this section on a *single region system*. In this subsection we describe the queueing dynamics assuming exogenously specified arrival-rates for drivers and passengers.

Drivers, passengers, and rides. The agents on a ride-sharing platform are *drivers* and *passengers*. The fundamental atomic unit of service in our system is a *ride*, which consists of a driver transporting a passenger. We assume that drivers live in the system for a relatively large contiguous block of time (e.g., several hours or a day), giving rides to several passengers. On the other hand, we assume each passenger is unique and only lives in the system for the lifetime of their ride; therefore ride requests and passengers are equivalent in our model. [1]

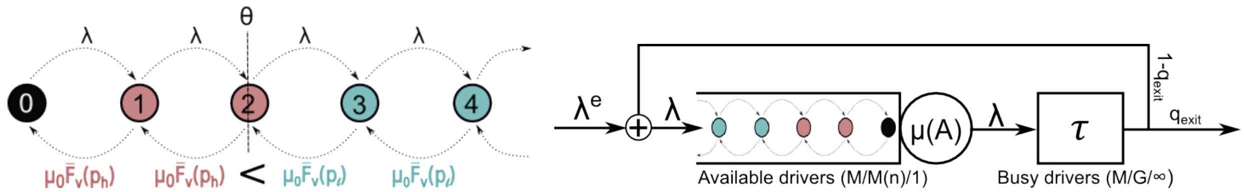
The queue. We model the system as a continuous-time queueing process evolving over $t \geq 0$, which keeps track of the number of drivers in the system. More specifically, at any time, each driver can be in one of two states – *available*, i.e., free to be matched to a passenger, or *busy*, i.e., occupied in transporting a passenger on a ride. We use $A(t)$ to denote the number of available drivers at time t, while $B(t)$ tracks the number of busy drivers. Our analysis employs steady-state performance criteria of the queueing system describing the platform. For convenience, we suppress dependence on time when clear from context.

Driver and passenger arrivals. We assume the arrival process of drivers to the available-drivers queue is Poisson with rate $\lambda > 0$. New drivers immediately join the queue of available drivers at an exogenous rate λ^e; drivers currently on the platform may also rejoin the queue after completing rides, as described below. In addition, we assume that the arrival process of new ride requests is Poisson with rate $\mu(A) > 0$ when the number of available drivers $A \geq 0$. A passenger requesting a ride at time t is *matched* if $A(t) > 0$ (reducing the number of available drivers by one), and is *blocked* if $A(t) = 0$. With respect to rides, all available drivers within a region are assumed equal, and any of them can be matched to any ride-request originating in that region. Below we discuss how both λ and μ arise from strategic interactions between the passengers, drivers, and the platform.

Ride completion and driver departures. We assume that each ride lasts an exponential length of time, with mean τ. When drivers complete service on a given ride, they can either exit the system, or become available. We assume that after each ride, a driver signs out and departs from the system with probability $q_{\text{exit}} > 0$, else the driver returns to the queue of available drivers.

Analysis. The queueing system described in this way is fairly straightforward to analyze: the queue of available drivers functions as an $M/M/1$ queue (more specifically, an $M/M(k)/1$ queue, where the service rate can be state-dependent – cf. Section 2.2), and the queue of busy drivers functions as an $M/M/\infty$ queue.

[1] A potentially interesting extension involves modeling incentives for passengers to return to the platform, based on past experience.



(a) Birth-death chain for available drivers (b) Flow of drivers in single-region model

Figure 1 Queueing model for (single-region) ride-sharing platform:

Figure 1(a) shows the birth-death chain for the number of available drivers in the region. In particular, we have shown a *single-threshold* pricing policy (p_ℓ, p_h, θ), wherein the platform uses a 'base' fare-multiplier p_ℓ when the number of drivers is greater than a threshold θ (here $\theta = 2$), else charges a 'primetime' price-multiplier $p_h > p_\ell$ (hence the queue drains slower when there are $\leq \theta$ drivers. cf. Section 4 for an analysis of this policy).

Figure 1(b) shows the flow of drivers in the network: exogenous drivers arrive to the platform at rate λ and join the available-drivers queue; when matched with a passenger, they transition from the $M/M(k)/1$ available-drivers queue to an $M/G/\infty$ busy-drivers queue; after completing a ride, they either exit the platform, or return to the available-drivers queue.

All external arrivals (new drivers) enter the $M/M(k)/1$ queue. Departures from the $M/M(k)/1$ queue (requested rides being picked up) enter the $M/M/\infty$ queue. Departures from the $M/M/\infty$ queue (drivers completing service on a ride) either leave the system (with probability q_{exit}) or return to the $M/M(k)/1$ queue (with probability $1 - q_{\text{exit}}$).

This is a simple two-queue example of an *open Jackson network* [9]. Note that from standard flow-balance constraints (cf. Figure 1), we have that the arrival-rate of drivers to the queue is related to their exogenous arrival-rate as $\lambda^e = \lambda q_{\text{exit}}$. Moreover, the steady-state distribution of open Jackson networks follows a product-form distribution: in steady-state, the two queue sizes are independent. We exploit this simple characterization in our analysis below; moreover, this also allows us to extend the results to networks of regions in Section 6.

2.2. Platform pricing

To model dynamic pricing, we allow the platform to choose prices that can vary based on the number of available drivers A. Formally, a pricing policy for the platform is a function $P(A)$ that maps the number of available drivers A to a price. The function $P(A)$ can be chosen depending on the system parameters. As done in practice, we assume that the platforms define a *base price*, based on the distance/time, and implements dynamic pricing by means of a *multiplier* for the base price. In particular, the *base* price is sufficient to cover any per-ride costs to the platform and driver (which we can then subtract from it); thus, for the remainder of the paper, we assume that both the platform's and drivers' costs per ride are zero. Finally, throughout this work, we assume that the prices are non-increasing with number of available drivers (i.e., $p(A) \geq p(A') \,\forall\, A < A'$). This mirrors the economic intuition that prices rise when available drivers drop, to better match incoming ride requests to available supply.

Throughout the paper, we assume that from each ride at price p, the platform gives a fraction γp to the driver, and retains $(1-\gamma)p$ for itself. Crucially, we analyze the pricing behavior of the platform *holding* γ *constant*. This assumption is motivated by the fact that most ride-sharing platforms manipulate their revenues through the pricing policy itself, rather than by changing the percentage they share with drivers. The latter quantity is fixed on very long timescales, to ensure transparency in communication with drivers about the benefits of participation on the platform. However, our model can accommodate variations in γ, and understanding the optimal revenue-sharing structure is an interesting avenue for future work.

Our work focusses on two important special cases of the platform's pricing policy:

Static pricing. The first case we consider is where the platform sets a single price for all A. We refer to this case as *static* pricing, because the price does not change based on instantaneous available service capacity. In reality, this case might be viewed as "quasi-static", in the sense that the price remains fixed as long as the average platform parameters (rates of arrival, demand/supply elasticities) remain fixed. As these parameters change at different times of day, the platform will likely change even fixed prices across the day. (Even most taxicab services price evenings differently from daytime hours.) However, the important property of static prices are that they only react to such coarse changes, and not to instantaneous state; we use static pricing to understand the platform's incentives during a temporal block where the exogenous system parameters are constant. These policies are analyzed in detail in Section 3.

Threshold dynamic pricing. The second case we consider is a class of *dynamic* pricing policies, where the platform does in fact set the price based on the number of available drivers. In Section 4, we consider a particular class of dynamic pricing policies, that we refer to as *single-threshold* pricing. These policies are characterized by three parameters: a low price p_ℓ, a high price $p_h > p_\ell$, and a threshold $\theta > 0$. The platform charges $P(A) = p_\ell$ when $A > \theta$, and $P(A) = p_h$ when $A \leq \theta$. In Appendix 10, we extend the analysis to consider more general threshold policies, and show that in large-markets, under some mild scaling assumptions, multiple threshold policies in fact reduce to some equivalent single-threshold policies.

2.3. Driver and passenger incentives

Given the space of pricing policies, we next want to model the strategic incentives of both drivers and passengers. A key assumption we make is that drivers and passengers are sensitive to price on *different timescales*. Informally, passengers respond instantaneously to the posted price they are shown for a ride (e.g., when they open a ride-sharing app on their phone). On the other hand (cf. the Introduction), we presume that drivers are sensitive to their earnings within a given time period (e.g., several hours, a day, or a week). This is motivated by the fact that most drivers in ride-sharing platforms approximately commit to their schedules on these longer timescales, and moderate their level of activity on the basis of the overall earnings they expect to receive while actively driving.

We take the perspective that steady-state behavior in our queueing model is representative of the longer timescale that drivers use in making their participation decisions. In particular, we assume that drivers are sensitive to their expected earnings on the platform over their lifetime, assuming they arrive to the system in steady-state.

The second component of the specification of passenger and driver utility is by considering their sensitivity to system performance. Passengers obtain disutility if their ride request goes unserved (i.e., if they are *blocked*). Drivers obtain disutility if they spend time *idling* in the system, because this decreases the overall earning rate. Both passengers and drivers are heterogeneous in their reservation utility for participation on the platform. Formally, we specify agents' utility as follows.

Passengers. Recall that passengers only live for at most one ride request in our model of the system. Each passenger has a reservation value V, drawn independently from a distribution $\mathbf{F}_V$. Confronted with a price p for a ride (potentially dependent on the number of available drivers), an arriving passenger requests a ride if $V \geq p$, and abandons otherwise. If she requests a ride, she is matched to an available driver – however, if none are present, then she remains unmatched and exits the system (blocking). The utility to the passenger is $V - p$ if she is served, and zero if she is blocked.

We assume the rate of *potential* ride requests is μ_0. Thus if the price is currently p_i, then the arrival process of actual ride requests is Poisson with rate $\mu_0 \overline{F}_V(p) = (1 - F_V(p))\mu_0$. Recall that the platform only sets prices on the basis of available drivers in the system. In particular, when the number of available drivers is A, the price is $P(A)$, and therefore the arrival rate of passengers is:

$$\mu(A) = \mu_0 \overline{F}_V(P(A)), \quad A \geq 0. \tag{1}$$

This yields the passenger arrival process μ as a function of the number of available drivers. Such a queueing model with state-dependent service rates is often referred to as a $M/M(k)/1$ queue [5].

Drivers. Drivers have reservation wages (measured in earnings per unit time) drawn independently from a distribution $\mathbf{F}_C$. A driver with reservation wage C will only enter and participate in the platform if their expected earnings exceeds C times their expected life in system. Drivers evaluate their earnings and life in the system assuming they arrive to the system (with fixed parameters λ, μ) in steady-state. This assumption is justified by the PASTA property (Poisson arrivals see time averages).

Note that since we have fixed q_{exit}; for fixed λ and μ the system functions as an open Jackson network; and the platform uses a pricing policy that depends only on the state of the network, the dynamics of a single driver are easy to describe. The driver arrives to find the $M/M(k)/1$ queue describing available drivers in steady-state. His idle time is equivalent to the waiting time in this queue; then, after picking up a passenger, he is busy for an exponential length of time with mean τ. After completing a ride, with probability $1 - q_{\mathsf{exit}}$ the driver returns to the available queue and *again finds it in steady-state* (by standard results for open Jackson networks); and with probability q_{exit}, the driver leaves the system.

From this discussion, it follows that the number of rides given by an arriving driver is *independent* of either the earnings on each ride, or the length of each ride. In particular, for a given driver, let K denote the number of rides she gives while in the system; let $P_1,\dots,P_K$ denote the prices charged to passengers on each of these rides (note that these are i.i.d. random variables, and depend on the state of the available queue at the time the driver is matched); let $T_1,\dots,T_K$ denote the ride times of each of the K rides (these are i.i.d. exponential random variables); and let $I_1,\dots,I_K$ denote the idle time prior to each ride (these are also i.i.d. random variables, distributed as the waiting time of an arriving driver into the available queue in steady-state). All these quantities are independent of each other.

It follows by Wald's lemma that the expected total earnings for the driver is $\mathbb{E}[K]\mathbb{E}[P_1]$; the expected total idle time is $\mathbb{E}[K]\mathbb{E}[I_1]$; and the expected total time spent driving is $\mathbb{E}[K]\mathbb{E}[T_1]=\mathbb{E}[K]\tau$. A driver with reservation wage C participates on the platform if $\mathbb{E}[K]\mathbb{E}[p_1]$ exceeds $\mathbb{E}[K](\mathbb{E}[I_1]+\tau)\ C$. For later reference, for a driver arriving to the system in steady-state, we define $\eta=\mathbb{E}[p_1]$ as the *expected earnings per ride*, and define $\iota=\mathbb{E}[I_1]$ as the *expected idle time per ride*. Drivers participate if $\eta/(\iota+\tau)\geq C$.

We let the rate of *potential* new drivers entering the system be Λ_0. Thus the rate of exogenous driver-arrivals to the queue (assuming the queue is stable) is:

$$\lambda^e=\lambda q_{\mathsf{exit}}=\Lambda_0\mathbf{F}_C\left(\frac{\eta}{\iota+\tau}\right). \tag{2}$$

Note that η and ι each depend in turn on λ and μ; for ease of notation, we suppress this dependence. The following section discusses how equilibrium conditions ensure consistency of these system parameters.

We conclude with an observation: note that in our model, q_{exit} is independent of the earnings of the driver. This is the key modeling point that captures our assumptions that drivers do not react to earnings on a per-ride basis. Instead they determine participation on a longer timescale, averaging over the length of their time in system. This modeling choice is what leads to Equation (2).

2.4. Equilibrium

In this section we combine our model of driver and passenger incentives with the platform's pricing strategy to produce a set of equilibrium conditions that determine λ and μ, for a fixed pricing policy $P(A)$.

First we define a quartet $\lambda,\mu=\{\mu(A)\}_{A\geq 0},\eta,\iota$ to be an *equilibrium* for a fixed $P(A)$ if the following conditions hold:

1. Given the pricing policy $P(A)$, μ is determined by Equation (1).
2. η is the expected earnings per ride, and ι is the expected idle time of a driver arriving in steady-state to a system with fixed λ and μ.
3. λ is determined by Equation (2), given μ,η and ι.

To specify an equilibrium, therefore, we need expressions for η and ι in terms of λ and $\mu(A)$. We now provide these in this section assuming a general pricing policy $P(A)$ (with prices which are non-increasing with A). In subsequent sections, we specialize it to static pricing (cf. Section 3), single-threshold dynamic pricing (cf. Section 4) and multi-threshold dynamic pricing (cf. Appendix 10).

Given exogenous λ and μ_0, and a pricing policy $P(A)$, this results in an $M/M(k)/1$ available-driver queue. For the associated Markov chain to be positive recurrent, we require $\exists A^* \geq 0$ such that:

$$\mu(A^*) = \mu_0 \overline{\mathbf{F}}_V(P(A^*)) > \lambda. \tag{3}$$

Since $P(A)$ is non-increasing, we have $\mu(A) > \lambda \forall A > A^*$. Assuming Equation (3) holds, the resulting Markov chain has a steady-state distribution $\pi(A)$, given by:

$$\pi(0) = \left(\sum_{i=0}^{\infty} \frac{(\lambda/\mu)^i}{\Pi_{j=1}^{i-1} \overline{\mathbf{F}}_V(P(j))} \right)^{-1} \quad , \quad \pi(i) = \frac{\pi(0)(\lambda/\mu)^i}{\Pi_{j=1}^{i-1} \overline{\mathbf{F}}_V(P(j))}. \tag{4}$$

We can now compute the performance metrics under steady-state as follows:

Proposition 1 *Given exogenous λ, μ and pricing policy $\{P(i)\}_{i\geq 1}$ obeying Equation (3), and assuming the $M/M(k)/1$ queue has steady-state as in Equation (4), the average earnings and idle-time per ride obey:*

$$\eta = \gamma \sum_{i=0}^{\infty} \pi(i) \cdot P(i+1) \quad , \quad \iota = \frac{\sum_{i=1}^{\infty} \pi(i) \cdot i}{\lambda}. \tag{5}$$

Proof. From the PASTA property [9] we have that a typical driver (i.e., arriving to the queue according to a Poisson process) sees the queue in steady-state. Now, the formula for the average idle-time is a direct consequence of Little's Law, i.e., $\mathbb{E}[\iota] = \mathbb{E}[A]/\lambda$. To calculate the average per-ride earnings, we want to find the amount earned by a typical departing driver from the queue (i.e., one who was just matched to a passenger). Note that the $M/M(k)/1$ queue is reversible, and a typical departing driver is an arrival to the reverse queue. Moreover if an arrival to the reverse queue sees it in state $i \geq 0$, then the corresponding departing driver must have received $P(i+1)$ (as there are i drivers plus the departing driver at the moment when the match was made). Thus $\mathbb{E}[\eta] = \sum_{i=0}^{\infty} \pi(i) P(i+1)$. $\square$

Equations (1), (2), (3), (4) and (5) thus together define the equilibrium of the system.

Lemma 2 *Given $(\Lambda_0, \mu_0) \in \mathbb{R}_+^2$, $(\gamma, q_{exit}) \in (0,1)^2$, continuous distributions $\mathbf{F}_C, \mathbf{F}_V$ and a non-increasing pricing policy $P(A)$, the equilibrium defined by Equations (1), (2), (3), (4) and (5) always has a solution $\lambda \geq 0$. Moreover, the equilibrium is unique if ι is non-decreasing in λ.*

Proof. Fix $\Lambda_0, \mu_0 \in \mathbb{R}_+^2$, continuous distributions $\mathbf{F}_C, \mathbf{F}_V$ and a pricing policy $P(A)$. Now the existence of the equilibrium defined by Equation 2 along with Equations (1), (3), (4) and (5) follows from the Brouwer fixed-point theorem, since $\mathbf{F}_C$ is continuous and λ is constrained to lie in the compact set $[0, \Lambda_0]$.

For uniqueness, let $f(\lambda) = \lambda q_{\text{exit}} - \Lambda_0 \mathbf{F}_C(\gamma\eta/(\tau+\iota))$. Then for the given p we have $f(0) \leq 0$ and also $\lim_{\lambda \to (\mu_0 \overline{\mathbf{F}}_V(p))^-} f(\lambda) = \mu_0 \overline{\mathbf{F}}_V(p) q_{\text{exit}} > 0$. Provided $P(A)$ is non-increasing with A, we have that the per-ride earning $\eta(\lambda)$ is continuous and non-increasing w.r.t. λ, keeping all other parameters fixed. This follows from a standard coupling argument to show that the number of drivers in queue is sample-pathwise higher when λ is higher. Now if in addition we have that $\iota(\lambda)$ is non-decreasing w.r.t. λ, then the average earning rate $\gamma\eta/(\tau+\iota)$ is overall non-increasing with λ. Finally, since $\mathbf{F}_C$ is continuous and non-decreasing, we have that $f(\lambda)$ is continuous and strictly increasing in λ – thus $f(\lambda) = 0$ has a unique solution $\lambda \geq 0$. □

2.5. Throughput and Revenue

Given the space of pricing policies and the associated equilibrium, as described in the previous sections, we are now in a position to design pricing policies for specific objectives. A first order target of pricing is to maximize the *volume of trade*, i.e., the rate of successful matches in steady-state. Provided the queue is stable under pricing policy $P(A)$, the rate of matches is given by the equilibrium $\lambda(P)$.

An alternate target for the platform is to maximize its own *revenue*. Under any pricing policy $P(A)$, the platform earns a $(1-\gamma)$ fraction of each ride's value. Thus, the revenue-rate $\Pi(P)$ is given by:

$$\mathbb{E}[\Pi(P)] = (1-\gamma)\lambda(P)\eta(P) \tag{6}$$

This follows from the fact that $\eta(P)$ is the per-ride expected payment, while, assuming queue stability, the rate of successful matches is $\lambda(P)$. Note again that we assume γ is held fixed by the platform; the pricing policy alone is used to optimize the performance.

2.6. The Large-Market Limit

Though Equations (1), (2), (3), (4) and (5) fully define the equilibrium of the ride-sharing platform, and Equations (6) the associated platform revenue, analyzing the system presents two challenges. First, although Lemma 2 guarantees the existence of the equilibrium, it does not guarantee its uniqueness. More significantly, though we can numerically solve for the equilibrium, it is difficult to use these solutions to get qualitative insights into throughput/revenue maximization. To circumvent this, we study the ride-sharing platform under an appropriate *large-market limit*. The scaling we consider is identical to the fluid scaling in queueing systems. However, we require the scaling not to simplify the queueing dynamics (as we have closed-form expressions for the queueing metrics), but rather, to simplify the equilibrium computations.

We henceforth assume that all prices lie in the compact interval $[0, p_{max}]$ (for some given maximum price p_{max}). For static pricing policies (i.e., $P(A) = p$), we define the large-market limit as follows: We consider a sequence of systems parametrized by n, wherein $\Lambda_0(n) = \Lambda_0 n$ and $\mu_0(n) = \mu_0 n$, and all other parameters $(\tau, \gamma, q_{\text{exit}}, \mathbf{F}_C, \mathbf{F}_V, p)$ are held fixed. We then let n go to ∞, and study the *normalized* equilibrium state, i.e. $\lim_{n\to\infty} \lambda(n)/n$, of the limiting system. For dynamic pricing policies, in addition to scaling Λ_0 and μ_0, we also let $P(A)$ to scale with n. In particular, for threshold dynamic pricing under large-market scaling, we keep the price vector fixed, but allow the thresholds to scale with n.

The large-market scaling allows us to get insights into first-order effects on the platform's performance metrics. For a system parametrized by 'size' n (i.e., with $\Lambda_0(n) = \Lambda_0 n, \mu_0(n) = \mu_0 n$ and pricing policy P chosen appropriately), we can define normalized versions of the above quantities as $\lambda(P,n)/n, \mathbb{E}[\Pi(P,n)]/n$ and $\mathbb{E}\left[W(P,n)\right]/n$. In the large-market limit, we are thus interested in the limiting normalized rates, given by:

$$\widehat{\lambda}(P) = \lim_{n\to\infty} \frac{\lambda(P,n)}{n}, \qquad \mathbb{E}[\widehat{\Pi}(P)] = \lim_{n\to\infty} \frac{\mathbb{E}[\Pi(P,n)]}{n}, \qquad \mathbb{E}[\widehat{W}(P)] = \lim_{n\to\infty} \frac{\mathbb{E}\left[W(P,n)\right]}{n}.$$

The large-market scaling corresponds to studying a ride-sharing platform where the potential pool of drivers and passengers scale up together. An alternate viewpoint is that scaling Λ_0 and μ_0 together is equivalent to fixing Λ_0, μ_0, but compressing the timescale on which the processes operate. In the case of static pricing, as long as the queue is stable, it is thus clear that the idle-time between rides goes to 0 in the limit. A similar property can be shown to hold for threshold pricing policies, provided the thresholds scale in an appropriate manner (cf. Section 4). It is this vanishing idle-time property which makes the large-market regime more amenable to analysis.

3. The Single Region under Static Pricing

We first consider the ride-sharing platform operating in a single region under *static pricing* with price p. For ease of notation, we parametrize all quantities in this section by price p, suppressing the dependence on other system parameters (Λ_0, μ_0, etc.).

To summarize this setting in brief: we have a single available-driver queue, where the potential rate of exogenous driver-arrivals is Λ_0, the actual realized rate of driver-arrivals at equilibrium is denoted $\lambda^e(p)$ (where $0 \leq \lambda^e(p) \leq \Lambda_0$). Each driver waits in the queue until she is matched to a passenger. After each ride, a driver departs from the platform with probability q_{exit}, else returns to the queue of available drivers – thus the effective arrival-rate of drivers to the queue is $\lambda(p) = \lambda^e(p)/q_{\text{exit}}$. On the other hand, we assume the potential rate of passenger arrivals (or 'unique app-opens') is μ_0. Now, solving Equations (4) and (5), we get that $\eta(p) = \gamma p$ (where γ is the fraction of the price going to the drivers), and $\iota(p) = 1/(\mu_0 \overline{\mathbf{F}}_V(p) - \lambda(p))$ if $\mu_0 \overline{\mathbf{F}}_V(p) > \lambda(p)$, and ∞ otherwise. Substituting in Equation (2), we get the equilibrium condition as:

$$\lambda^e(p) = \lambda(p) q_{\text{exit}} = \Lambda_0 \mathbf{F}_C \left(\frac{\gamma p}{\tau + (\mu_0 \overline{\mathbf{F}}_V(p) - \lambda(p))^{-1}} \right) \mathbb{1}_{\{\mu_0 \overline{\mathbf{F}}_V(p) > \lambda(p)\}}, \tag{7}$$

where $\mathbb{1}_A$ is the indicator function of event A, and we define $\lambda(p) = 0$ if $\overline{\mathbf{F}}_V(p) = 0$.

Now note that the available-drivers queue is $M/M/1$ in this setting – thus, for fixed μ, the average idle-time ι is increasing in the arrival-rate λ. Lemma 2 now guarantees both existence and uniqueness of the equilibrium $\lambda(p)$; further, under the large-market scaling, we get the following equilibrium expressions:

Theorem 3 *We are given* $(\Lambda_0, \mu_0) \in \mathbb{R}^2_+$, $(\gamma, q_{exit}) \in (0,1)^2$ *and continuous distributions* $\mathbf{F}_C, \mathbf{F}_V$. *Let* $\lambda(p,n)$ *denote the equilibrium driver arrival-rate at fixed price* p *in* n^{th} *system (i.e., with* $\Lambda_0(n) = n\Lambda_0, \mu_0(n) = n\mu_0$*). Then:*

$$\lambda(p,n) = n \cdot \min\left[\Lambda_0\, \mathbf{F}_C(\gamma p/\tau)/q_{exit},\ \mu_0\, \overline{\mathbf{F}}_V(p)\right] - o(n).$$

Moreover, we have $\frac{\lambda(p,n)}{n} \to \min\left[\Lambda_0\, \mathbf{F}_C(\gamma p/\tau)/q_{exit},\ \mu_0\, \overline{\mathbf{F}}_V(p)\right] \triangleq \widehat{\lambda}(p)$ *uniformly as* $n \nearrow \infty$.

Theorem 3 characterizes the first-order equilibrium behavior under a fixed price as the market grows large. Moreover, it shows that the limiting normalized driver-arrival rate is given by $\widehat{\lambda}(p) = \min\left[\Lambda_0\, \mathbf{F}_C(\gamma p/\tau)/q_{\text{exit}},\ \mu_0\, \overline{\mathbf{F}}_V(p)\right]$. Due to space constraints, the proof is deferred to Appendix 7. However, we provide a visual depiction of Theorem 3 in Figure 2, where we have shown the normalized rates $\lambda(p,n)/n$ converge to $\widehat{\lambda}(p)$ in the large-market limit. Note how $\widehat{\lambda}(p)$ decomposes into two parts:

i. *Supply-bottleneck:* For low prices, $\widehat{\lambda}(p)$ is determined by supply considerations – the potential driver arrival-rate Λ_0 and reservation-earning distribution $\mathbf{F}_C$.

ii. *Demand-bottleneck:* For high prices, $\widehat{\lambda}(p)$ is determined by demand considerations – the potential passenger arrival-rate μ_0 and ride-value distribution $\mathbf{F}_V$.

Next, we want to use the large-market limit to understand the platform's choice of price p in order to maximize throughput/revenue. Given the sequence of systems parametrized by n under the large-market scaling, we would ideally like to perform this optimization for a given parameter n – this though appears difficult as we do not have closed form expressions for the equilibrium state. However, the fact that $\frac{\lambda(p,n)}{n}$ converges to $\widehat{\lambda}(p)$ uniformly also allows us to interchange the limits and optimization, and therefore study performance optimization under the large-market limit. This is encoded in the following result:

Corollary 4 *Given* $(\Lambda_0, \mu_0) \in \mathbb{R}^2_+$, $(\gamma, q_{exit}) \in (0,1)^2$ *and continuous distributions* $\mathbf{F}_C, \mathbf{F}_V$, *let* $\lambda^*(n) = \max_{p\in[0,p_{\max}]} \lambda(p,n)$ *denote the maximum equilibrium driver arrival-rate in the* n^{th} *system, and* $p^*(n)$ *denote the corresponding price. Moreover, let* $\widehat{\lambda}^* = \max_{p\in[0,p_{\max}]} \widehat{\lambda}(p)$, *and* $p^* = \arg\max_{p\in[0,p_{\max}]} \widehat{\lambda}(p)$. *Then* $\lim_{n\to\infty} \frac{\lambda^*(n)}{n} = \widehat{\lambda}^*$ *and* $\lim_{n\to\infty} p^*(n) = p^*$. *Furthermore, the operations commute even if we choose prices to maximize revenue (i.e.,* $p^*(n) = \arg\max_{p\in[0,p_{\max}]} \mathbb{E}[\widehat{\Pi}(p,n)]$*).*

Getting an explicit characterization for $\lambda(p)$, the normalized driver arrival-rate at equilibrium in the large-market limit, allows us to set a price to maximize the platform's throughput and/or revenue.From Section 2.5, we have that in the large-market limit, the platform's normalized average revenue-rate is given by $\mathbb{E}\left[\Pi(p)\right] = \lambda(p)p(1-\gamma)$.Before stating our results, we need two additional definitions. We define the *balance-price* p_{bal} to be the one satisfying the following implicit equation:

$$\mu_0\, \overline{\mathbf{F}}_V(p_{bal}) = \frac{\Lambda_0}{q_{\text{exit}}}\, \mathbf{F}_C\left(\frac{\gamma}{\tau}.p_{bal}\right). \tag{8}$$

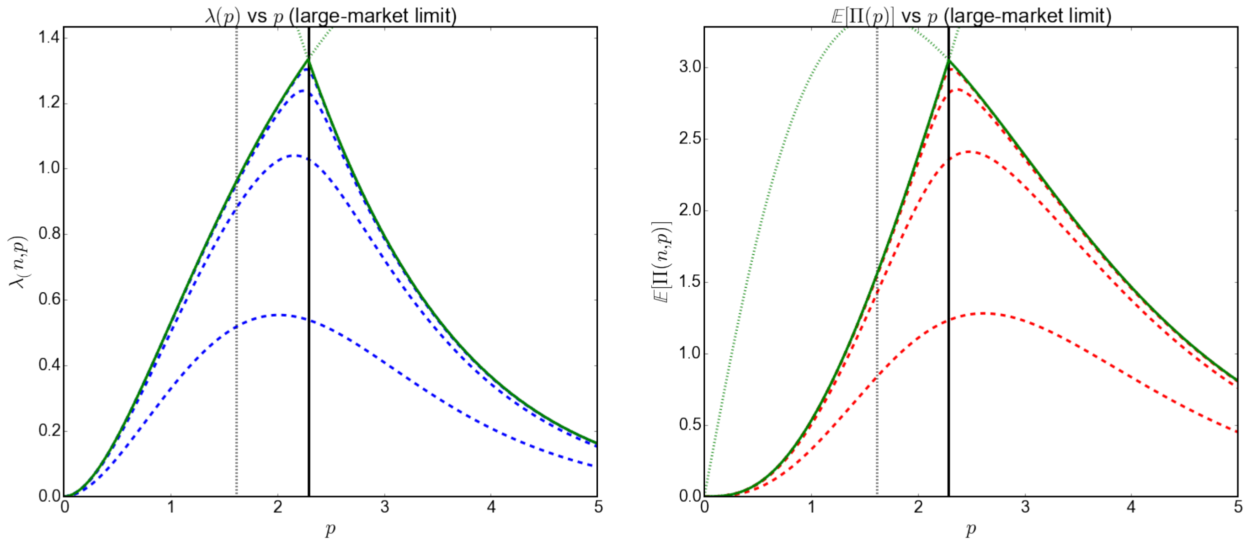


Figure 2 Scaling behavior of the normalized equilibrium driver-arrival rate and revenue rate (i.e., $\lambda(p,n)/n$ and $\mathbb{E}[\Pi(p,n)]/n$) vs. static price p, for $n=1$ (the bottom-most dashed curve in either plot), $10, 100$ and 1000 (the topmost dashed curve). The solid green curves depict the normalized driver-arrival rate $\widehat{\lambda}(p)$ and normalized revenue-rate $\mathbb{E}[\widehat{\Pi}(p)]$ in the large-market limit (c.f. Theorem 3). The solid black vertical line marks out the balance price p_{bal}, while the dotted black vertical line marks the demand-optimal price p_{d-opt}. We use $\Lambda_0/q_{\text{exit}}=2, \mu_0=4, \tau=1$ and distributions $\mathbf{F}_C \sim Gamma(2,1)$ and $\mathbf{F}_V \sim Gamma(2,1)$

Intuitively, p_{bal} is the price at which the effective demand for rides equals the effective supply of drivers assuming the idle time between rides is zero [2]. Assuming $\mathbf{F}_V, \mathbf{F}_C$ have continuous CDFs, it is easy to check that p_{bal} exists and is unique.

In addition, we also define the *demand-optimal price* p_{d-opt} as:

$$p_{d-opt} = \arg\max_{p>0} \left\{ p \cdot \overline{\mathbf{F}}_V(p) \right\}, \tag{9}$$

As the name suggests, p_{d-opt} is the price that maximizes the platform's revenue rate if considering the demand profile alone, i.e., if the drivers were not strategic. We assume that this maximum exists (this can be guaranteed by assuming $p \in [0, p_{\max}]$ for some chosen $p_{\max}$). Note though that the above optimization need not yield a unique solution – moreover, there could be multiple local maxima. To make our subsequent results more concise and transparent, *we assume throughout this work that* $p\overline{\mathbf{F}}_V$ *has a unique maxima* p_{d-opt} *and also that* $p\overline{\mathbf{F}}_V(p)$ *is decreasing for* $p \geq p_{d-opt}$ [3].

We now show that *in the large-market limit, the revenue is maximized at the greater of* p_{bal} *and* p_{d-opt}. Formally, we have the following result:

[2] This intuition is exact in the large-market limit. To see this, note that in any equilibrium state, the queue must be stable (as otherwise the expected idle-time blows up); moreover, in a stable queue, scaling Λ_0 and μ_0 together is equivalent to fixing Λ_0, μ_0, but speeding up time, thereby driving the idle-time to 0.

[3] Note that this is not a very restrictive assumption – it holds if $\mathbf{F}_V$ has a monotone (increasing or decreasing) hazard-rate $h_F(x) = f_V(x)/\overline{\mathbf{F}}_V(x)$ (where p_{d-opt} is the unique price satisfying $p_{d-opt} h_F(p_{d-opt}) = 1$), or $\mathbf{F}_V$ is Pareto distributed, i.e., $F_V(x) = 1 - \left(\frac{x_{min}}{x}\right)^{\alpha}, x \geq x_{min} > 0$, with $\alpha > 1$.

Theorem 5 *Let $p_{rev-opt}(n) = \arg\max_p \mathbb{E}[\Pi(p,n)]$ be the revenue-optimal price. Then, in the large-market regime, we have:* $\lim_{n\to\infty} p_{rev-opt}(n) = \max[p_{bal}, p_{d-opt}]$

The proof follows from analyzing the revenue under the equilibrium $\lambda(p)$ in large-market limit. Due to lack of space, we defer the details to Appendix 7.

To summarize: the limiting normalized throughput (match-rate) $\widehat{\lambda}(p)$ is maximized at the balance price – moreover, as long as $p_{bal} \geq p_{d-opt}$, static-pricing with p_{bal} also maximizes the platform's revenue. Note that this condition corresponds to a *supply-limited regime*, where the potential request-rate for rides is much greater than the potential driver pool. We return to this assumption in later results.

4. The Single Ride-Share Queue under Dynamic Pricing

We next consider the single-region ride-share platform under *dynamic pricing*. In particular, we consider the single-threshold pricing policy (p_ℓ, p_h, θ), as described in Section 2.2. We consider the case of generalized threshold policies in Appendix 10. Note that the queueing model for available drivers in this setting is the same as that in Section 3, except that we now have state-dependent service rates (cf. Section 2.1).

As before, the potential rate of drivers is Λ_0, the actual exogenous rate of driver-arrivals is λ^e, and the effective arrival-rate of drivers to the queue is $\lambda = \lambda^e/q_{exit}$ (conditional on the queue stability). The potential rate of passenger arrivals is μ_0; the actual rate of matches now depends on the price. In particular, the rate is $\mu_0\overline{\mathbf{F}}_V(p_\ell)$ when there are more than θ available drivers, but falls to $\mu_0\overline{\mathbf{F}}_V(p_h)$ when there are less or equal. At equilibrium, the arrival-rate of drivers again must satisfy $\lambda q_{\text{exit}} = \Lambda_0\mathbf{F}_C(\gamma\eta/(\iota+\tau))$ (cf. Equation (2)). As before, the average ride-time τ is independent of pricing; the average per-ride earning η and the average idle-time ι are however different from the static pricing setting, as is the blocking probability $p_{block} \triangleq \pi[0]$.

Suppose the net arrival-rate of drivers to the queue is fixed at λ. For notational convenience, we define $\phi_h = 1/\overline{\mathbf{F}}_V(p_h), \phi_\ell = 1/\overline{\mathbf{F}}_V(p_\ell)$ and $\rho = \lambda/\mu_0$. For stability (cf. Equation (3)), we now need $\lambda < \mu_0\overline{\mathbf{F}}_V(p_\ell)$ (i.e. $\rho\phi_\ell < 1$) – if this holds, then substituting the steady-state probabilities in Equation (5) and simplifying, we get:

$$p_{block} = \frac{(\rho\phi_h - 1)(1-\rho\phi_\ell)}{(\rho\phi_h - \rho\phi_\ell)(\rho\phi_h)^\theta - 1}, \quad \eta = \frac{((\rho\phi_h)^\theta - 1)(1-\rho\phi_\ell)p_h + (\rho\phi_h - 1)(\rho\phi_h)^\theta p_\ell}{(\rho\phi_h - \rho\phi_\ell)(\rho\phi_h)^\theta - (1-\rho\phi_\ell)},$$
$$\iota = \left(\frac{p_{block}}{\lambda}\right)\left(\frac{\phi_h\rho(1+(\rho\phi_h)^\theta(\theta(\rho\phi_h - 1) - 1))}{(\phi_h\rho - 1)^2} + \frac{\rho\phi_\ell(\phi_h\rho)^\theta(1+\theta(1-\rho\phi_\ell))}{(1-\rho\phi_\ell)^2}\right). \tag{10}$$

On the other hand, if $\lambda > \mu_0\overline{\mathbf{F}}_V(p_\ell)$ (i.e., $\rho\phi_\ell > 1$), we have $p_{block} = 0$, $\eta = p_\ell$ and $\iota = \infty$.

To get qualitative insights into different dynamic-pricing policies, we consider the scaled system parametrized by n (with $\Lambda_0(n) = \Lambda_0 n$ and $\mu_0(n) = \mu_0 n$), and consider the large-market limit $n \to \infty$. We assume all other exogenous system parameters remain fixed. We first consider the large-market limit for fixed values of p_ℓ and p_h; we have the following proposition.

Proposition 6 *Fix system parameters $\Lambda_0, \mu_0, \tau, \gamma, q_{exit}$, distributions $\mathbf{F}_C, \mathbf{F}_V$, and prices p_ℓ, p_h. For each n, let $\theta^*(n, p_\ell, p_h)$ be the throughput-optimal choice of θ. Then $\theta^*(n) \nearrow \infty$ as $n \nearrow \infty$, with $\theta^*(n) = o(\sqrt{n})$. Furthermore, the same scaling holds if $\theta^*(n)$ is chosen as the revenue-optimal choice of θ for each n with fixed (p_ℓ, p_h).*

The preceding proposition demonstrates that regardless of our objective of interest, the optimal $\theta^*(n)$ scales as $\omega(1)$ but $o(\sqrt{n})$. The lower bound follows from the observation that the average per-ride earnings are monotone in θ; moreover, as $n \to \infty$, the per-ride earnings converge to a constant independent of θ. However this is not sufficient to maximize the performance metrics as the average idle-time also increases with θ. Optimality requires the average idle-time goes to 0 as $n \to \infty$, and this is guaranteed by the upper bound on $\theta^*(n)$. (See Theorem 13 for details).

Motivated by the preceding result, we consider a limiting regime where p_ℓ and p_h are fixed, and the platform's choice of $\theta(n)$ behaves as $\omega(1)$ and $o(\sqrt{n})$. Given these assumptions, we have the following characterization of the large-market limit of the platform under dynamic pricing:

Theorem 7 *Consider a system with $(\Lambda_0, \mu_0) \in \mathbb{R}_+^2$, $(\gamma, q_{exit}) \in (0,1)^2$ and continuous distributions $\mathbf{F}_C, \mathbf{F}_V$. Let $\lambda(n, p_\ell, p_h)$ denote the equilibrium arrival-rate of drivers in the n^{th} system under dynamic pricing with parameters (p_ℓ, p_h) and $\theta = \theta^*(n, p_\ell, p_h)$. Then we have:*

$$\frac{\lambda(n, p_\ell, p_h)}{n} \to \widehat{\lambda}(p_\ell, p_h) \text{ uniformly as } n \nearrow \infty$$

Moreover, the limiting normalized driver arrival-rate $\widehat{\lambda}(p_\ell, p_h)$ obeys:

(1) If $p_\ell > p_{bal}$: $\widehat{\lambda}(p_\ell, p_h) = \mu_0 \overline{\mathbf{F}}_V(p_\ell)$; *(2) If $p_h < p_{bal}$: $\widehat{\lambda}(p_\ell, p_h) = \frac{\Lambda_0}{q_{exit}} \mathbf{F}_C\left(\frac{\gamma p_h}{\tau}\right)$;*

(3) If $p_\ell \leq p_{bal} \leq p_h$: $\widehat{\lambda}(p_\ell, p_h)$ satisfies the fixed-point equation:

$$\widehat{\lambda}(p_\ell, p_h) = \frac{\Lambda_0}{q_{exit}} \mathbf{F}_C\left(\frac{\gamma}{\tau}\left(\frac{p_\ell(\phi_h - \mu_0/\widehat{\lambda}(p_\ell, p_h)) + p_h(\mu_0/\widehat{\lambda}(p_\ell, p_h) - \phi_\ell)}{\phi_h - \phi_\ell}\right)\right). \tag{11}$$

Due to lack of space, the proof is deferred to Appendix 8. Figure 3(a) shows an example of the convergence to large-market limits under dynamic pricing policies. We fix one price below p_{bal}, and vary the other one, while scaling $\theta(n)$ as $\theta_0 \log n$ (note that $\theta(n)$ is $\omega(1)$ but $o(\sqrt{n})$). As long as both prices are below p_{bal}, the platform's performance is dictated by the higher price (as in case (1) in Theorem 7). The interesting scenario is when the second price is greater than p_{bal}, in which case we are in case (3) of Theorem 7. In this setting, in the large-market limit, the number of available drivers in the n'th system concentrates around the threshold $\theta(n)$. The resultant per-ride earning is thus a convex combination of (p_ℓ, p_h), with the coefficients corresponding to the probability of the number of drivers being above and below (and equal to) θ respectively. Note though that the resulting normalized arrival-rate $\widehat{\lambda}(p_\ell, p_h)$ of drivers *is not a convex combination of the extreme values* $(\widehat{\lambda}(p_\ell), \widehat{\lambda}(p_h))$ – rather, it is determined by the (non-linear) fixed-point condition arising from equilibrium considerations (Equation (2)).

Now we can turn to optimizing revenue under single-threshold dynamic pricing. First, as in the static pricing setting, we again show that the limit and maximization operations commute:

Corollary 8 *Given* $(\Lambda_0,\mu_0)\in\mathbb{R}_+^2$, $(\gamma,q_{exit})\in(0,1)^2$ *and continuous distributions* $\mathbf{F}_C,\mathbf{F}_V$, *let* $\lambda^*(n)=\max_{0\le p_\ell\le p_h\le p_{\max}}\lambda(n,p_\ell,p_h)$ *denote the maximum equilibrium driver arrival-rate in the* n^{th} *system, and let* $\widehat{\lambda}^*=\max_{0\le p_\ell\le p_h\le p_{\max}}\widehat{\lambda}(p_\ell,p_h)$. *Then* $\lim_{n\to\infty}\frac{\lambda^*(n)}{n}=\widehat{\lambda}^*$. *Moreover, this commutativity holds if we choose prices to maximize revenue.*

We focus on the setting where we scale $\theta(n)=\omega(1)$ and $\theta(n)=o(\sqrt{n})$. The platform's average revenue-rate is now given by $\mathbb{E}[\widehat{\Pi}(p_\ell,p_h)]=(1-\gamma)\widehat{\lambda}(p_\ell,p_h)\eta(p_\ell,p_h)$.[4] Also, from the discussion on static pricing (Section 3), we have $\widehat{\lambda}(p_{bal})=(\Lambda_0/q_{exit})\mathbf{F}_C(\gamma p_{bal}/\tau)$ and $\mathbb{E}[\widehat{\Pi}(p_{bal})]=\widehat{\lambda}(p_{bal})p_{bal}$ as the normalized throughput and revenue under static price p_{bal}. We now have the following characterization of system under single-threshold dynamic pricing:

Theorem 9 *Given* $(\Lambda_0,\mu_0)\in\mathbb{R}_+^2$, $(\gamma,q_{exit})\in(0,1)^2$ *and continuous distributions* $\mathbf{F}_C,\mathbf{F}_V$. *We also choose* $p_\ell<p_{bal}$ *and* $\theta(n)=o(\sqrt{n})$. *Then we have:*

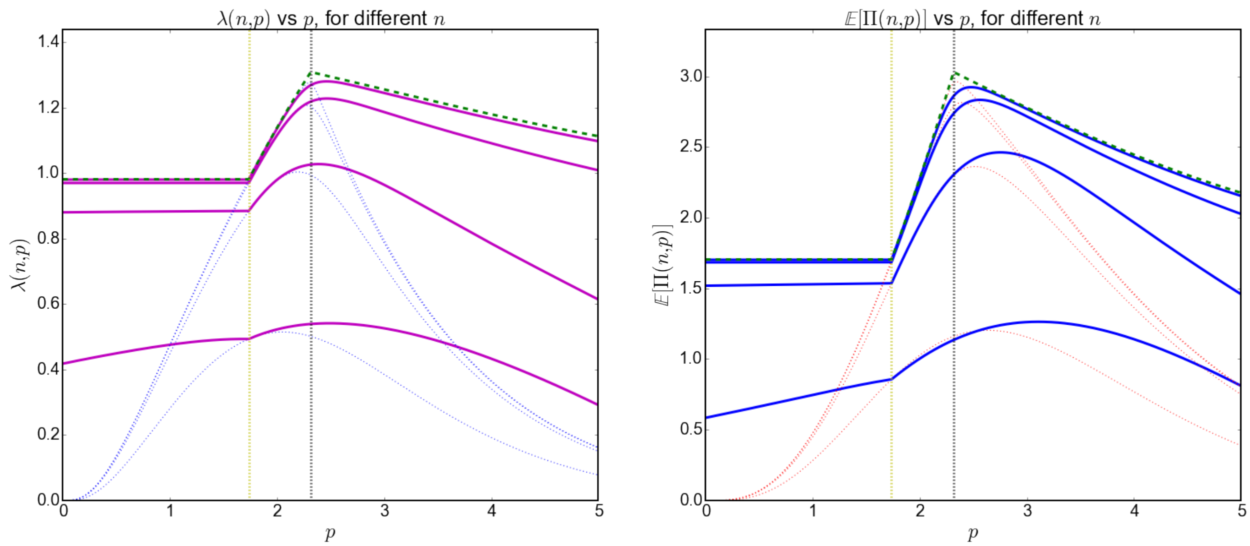
1. *If* $p_h=p_{bal}$, *then* $\widehat{\lambda}(p_\ell,p_h)=\widehat{\lambda}(p_{bal})$, *and* $\mathbb{E}\left[\widehat{\Pi}(p_{bal})\right]=\lambda_{bal}(1-\gamma)p_{bal}$.
2. *For any* $p\ge p_{bal}>p_\ell$ *with* $\overline{\mathbf{F}}_V(p)>0$, *suppose* $\mathbf{F}_V(\cdot)$ *satisfies:*
$$\frac{\mathbf{F}_V(p)-\mathbf{F}_V(p_\ell)}{f_V(p)(p-p_\ell)}<\frac{1-\mathbf{F}_V(p_\ell)}{1-\mathbf{F}_V(p)}. \tag{12}$$
Then $\widehat{\lambda}(p_\ell,p_h)\le\widehat{\lambda}(p_{bal})$ *and* $\mathbb{E}\left[\widehat{\Pi}(p_\ell,p_h)\right]\le\mathbb{E}\left[\widehat{\Pi}(p_{bal})\right]$.

Theorem 9 shows that as long as $\mathbf{F}_V$ satisfies Equation (12), then, in the large-market limit, *no dynamic pricing is strictly better than the optimal static pricing policy* in terms of volume of trade (i.e., rate of successful matchings).Moreover, in the case of revenue, Theorem 9 shows that the platform revenue under dynamic pricing is always bounded by the revenue under static pricing with p_{bal}. In case $p_{bal}\ge p_{d-opt}$, this is the optimal revenue. If on the other hand $p_{bal}<p_{d-opt}$, then static pricing can get a higher revenue than any single-threshold

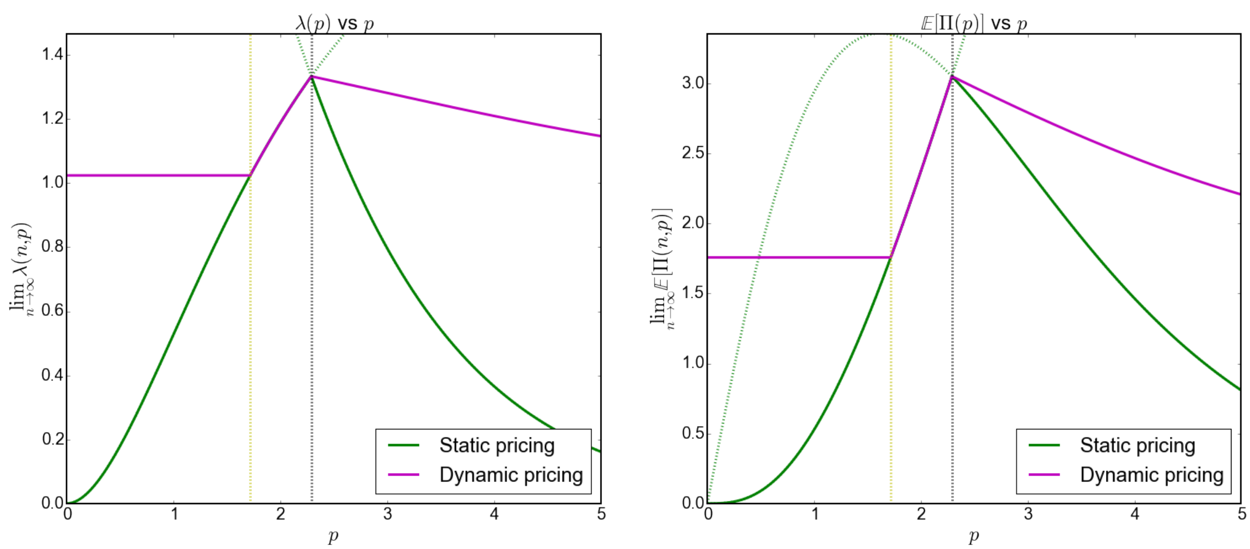
The proof of Theorem 9 involves differentiating the implicit expression for $\widehat{\lambda}(p_\ell,p_h)$ (Equation (11)) w.r.t. p_h while keeping p_ℓ fixed, and studying its behavior around p_{bal} – the revenue characterization is handled similarly. The formal proof is provided in Appendix 8. Note that the condition only depends on the distribution of passengers' ride values, and not on the drivers' reservation values. Moreover, we prove that it is satisfied by the two canonical distribution classes we referred to before (cf. Appendix 8 for details):

Corollary 10 *The condition in Equation* (12) *is satisfied if:*

[4] Note that the limiting function is computed by taking the limit of the revenue-rate in the n'th system as $n\to\infty$, cf. Section 2.6; $\theta(n)$ implicitly scales to ∞ in this limit



(a) Scaling behavior of Dynamic-Pricing Policies



(b) The Large-Market Limits under Static and Dynamic Pricing

Figure 3 Figure 3(a) shows the scaling behavior of $\lambda(p_\ell, p_h, \theta(n), n)/n$ and $\mathbb{E}[\Pi(p_\ell, p_h, \theta(n), n)]/n$ with n, for $n = 1$ (the bottom-most solid curve in either plot), $10, 100$ and 1000 (the topmost solid curve); Figure 3(b) plots the large-market limiting functions. In both, θ is scaled as $\theta(n) = \theta_0 \log n$. We keep one price fixed (indicated by the dotted yellow vertical line), while the second is varied as p; the dotted blue curves indicate the equilibrium values for the corresponding static-pricing policy at price p. The dashed green curves indicate the normalized large-market limiting values. The dashed black vertical line marks out the balance price p_{bal}. We use $\Lambda_0/q_{exit} = 2, \mu_0 = 4, \gamma/\tau = 1$ and distributions $\mathbf{F}_C \sim Gamma(2,1), \mathbf{F}_V \sim Gamma(2,1)$, choose the fixed price as $0.75 \cdot p_{bal}$ and use $\theta_0 = 3$.

1. F_V *is an MHR distribution, i.e., with increasing hazard rate* $h_V(x) = \frac{f_V(x)}{\overline{\mathbf{F}}_V(x)}$.
2. F_V *is a Pareto distribution, i.e.,* $F_V(x) = 1 - \left(\frac{x_{min}}{x}\right)^{\alpha}, x \geq x_{min} > 0$, *with* $\alpha \geq 1$.

Dynamic Pricing with Multiple Thresholds: A more general dynamic-pricing policy can allow multiple thresholds $\theta_1 > \theta_2 > \ldots > \theta_k$ (for some $k \in \mathbb{N}_+$) and prices $p_1 < p_2 < \ldots < p_{k+1}$. Characterizing the performance for a general threshold policy can be challenging even in the large-market limit. However, in Theorem 14 in Appendix 7, we now show that the behavior of a wide range of pricing policies in the large-market limit reduces to that of an appropriate single-threshold policy. In particular, we show that in the large-market limit, the only states with non-zero mass involve only the two prices (p_{j^*}, p_{j^*+1}) which bracket p_{bal}.

5. Robustness of Dynamic Pricing

The results in the previous section, in particular, Theorem 9, suggest that in the large-market limit, there are no benefits to using dynamic pricing policies over static pricing policies. This seems to run counter to the perception that dynamic pricing policies perform very well in practice. One potential reason is that the effect of dynamic pricing gets washed out under the large-market scaling. In particular, though our results expose first-order scaling effects, there could be significant second-order benefits of dynamic pricing. Simulating the system for small values of n provides some support for this hypothesis: For example, in Figure 3(a), we can see that for smaller n, the optimal value for the normalized revenue-rate under dynamic pricing (the solid lines in the plot) are higher than the corresponding optima for the static pricing curves (the dotted lines). In this section, however, we advance an alternate hypothesis for the success of dynamic pricing – that *dynamic-pricing policies are significantly more robust to uncertainty in the underlying parameters*.

A natural way to characterize the robustness of different pricing policies is to consider a setting where the platform does not have exact knowledge of the underlying passenger and/or driver arrival rates, but knows it lies in some *uncertainty set*. In this setting, we can then compare the equilibrium performance under a fixed policy (static/dynamic) to the optimal pricing-policy (i.e., with perfect knowledge of system parameters). For example, in Figure 4, we compare the performance of *fixed* static and dynamic pricing policies when the underlying parameters (in this case, μ_0) exhibit some uncertainty. Visually, it appears that the performance of the static price falls of sharply with changes in Λ_0 and μ_0, while dynamic pricing seems to perform well when compared to the *optimal* static price [5]. We now develop a novel geometric way to formalize this robustness property of dynamic-pricing policies.

As in previous sections, we restrict ourselves to static and two-price dynamic pricing policies (with $\theta(n) = \omega(1)$), and consider the system performance in the large-market limit. For notational convenience, henceforth in this section we use Λ^* and μ^* to denote the *true underlying* normalized potential driver/passenger arrival rates in the large-market limit. We also restrict ourselves to supply-constrained

[5] Note that we choose $\mathbf{F}_V$ to be a $Gamma$ distribution, which is MHR; hence, from Theorem 9, we know that the static balance price is optimal even when compared to any single threshold dynamic-pricing policy.

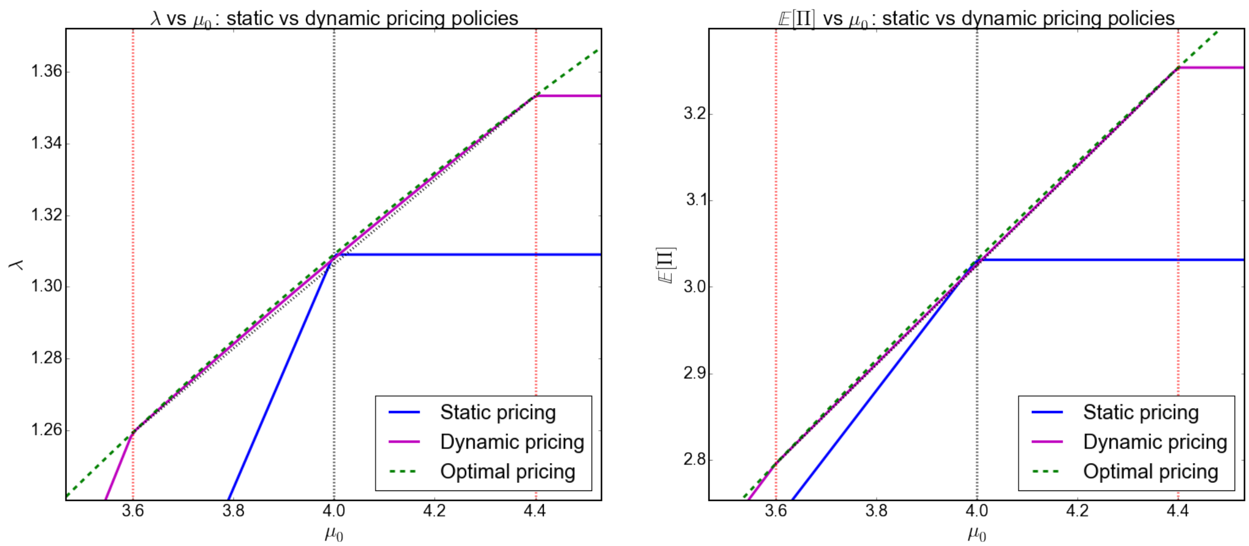


Figure 4 Performance (in large-market limit) of static and dynamic pricing under uncertainty in μ_0. We fix $\Lambda_0 = 3$, and vary the potential arrival-rate of passengers μ_0 as $4 \pm 10\%$. We then compare a static-pricing with p_{bal} based on $(\Lambda_0, \mu_0) = (3, 4)$, and a dynamic-pricing policy with p_ℓ set as the p_{bal} for $(\Lambda_0, \mu_0) = (3, 3.6)$, and p_h set as the the p_{bal} for $(\Lambda_0, \mu_0) = (3, 4.4)$. We plot the normalized metrics $(\widehat{\lambda}, \mathbb{E}[\widehat{\Pi}])$, using distributions $\mathbf{F}_V \sim Gamma(2,1), \mathbf{F}_C \sim Lognormal(1,1)$. The dashed green curve shows the performance of the system under static-pricing with the correct p_{bal} corresponding to the actual Λ_0, μ_0. The the dotted black vertical line indicates the μ_0 which was used to fix the static-pricing policy, while the dotted red vertical lines mark the μ_0 for which the balance price is p_ℓ and p_h.

regimes, wherein $p_{bal} > p_{d-opt}$, and assume throughout that $\mathbf{F}_V$ obeys Equation (12). Under these assumptions, Theorem (9) shows that using static pricing with the appropriate p_{bal} (corresponding to system parameters) is optimal. Before establishing our robustness result, we first present a simple characterization of this benchmark optimal policy:

Lemma 11 *Let $p_{bal}(\Lambda, \mu)$ denote the balance price of the system with parameters Λ and μ. Similarly, let $\widehat{\lambda}_{bal}(\Lambda, \mu)$ denote the corresponding equilibrium value of $\widehat{\lambda}$ under static pricing with $p_{bal}(\Lambda, \mu)$. Then, for every $\beta \geq 0$, we have:*

$$p_{bal}(\Lambda, \mu) = p_{bal}(\beta\Lambda, \beta\mu), \qquad \widehat{\lambda}_{bal}(\beta\Lambda, \beta\mu) = \beta\widehat{\lambda}_{bal}(\Lambda, \mu).$$

Lemma 11 captures two important geometric facts regarding the system. First, we see that the locus of points $(\Lambda, \mu) \in \mathbb{R}_+^2$ that share the same p_{bal} are lines through the origin. Moreover, we see that the manifold $(\Lambda, \mu, \widehat{\lambda}_{bal}(\Lambda, \mu))$ is a cone – given any point on the manifold, all points on the line passing through the origin and the given point also lie on the manifold.

Suppose now we are given two points $\Gamma_1 = (\Lambda_1, \mu_1), \Gamma_2 = (\Lambda_2, \mu_2)$ in the (Λ, μ) parameter space, and we know that the true system parameters lie on the line connecting these two points. More formally, given $\Gamma_1, \Gamma_2 \in \mathbb{R}_+^2$, we know that (Λ^*, μ^*) lies on the line $\Gamma(\alpha) = (\alpha\Lambda_1 + (1-\alpha)\Lambda_2, \alpha\mu_1 + (1-\alpha)\mu_2) \triangleq$

$(\Lambda(\alpha), \mu(\alpha))$, for some $\alpha \in [0,1]$. Now we have the following result (whose proof we defer to Appendix 9):

Theorem 12 *Consider the dynamic-pricing policy $(p_1, p_2, \theta(n))$, where $p_1 := p_{bal}(\Gamma_1)$, $p_2 := p_{bal}(\Gamma_2)$, and $\theta(n) = \omega(1)$ but $o(n)$. Let the true system parameters $(\Lambda^*, \mu^*) = (\Lambda(\alpha^*), \mu(\alpha^*))$ for some $\alpha \in [0,1]$, and define $\widehat{\lambda}_1$ and $\mathbb{E}[\widehat{\Pi}_1]$ to be the normalized throughput and revenue in the large-market limit under static pricing with $p_{bal}(\Gamma_1)$ (similarly for $p_{bal}(\Gamma_2)$). Now, if $\mathbf{F}_C$ is a log-concave distribution, then:*

$$\widehat{\lambda}(\Lambda^*, \mu^*) \geq \alpha^* \widehat{\lambda}_1 + (1-\alpha^*)\widehat{\lambda}_2,$$

Further, if $p_1, p_2 > p_{d-opt}$, we have $\mathbb{E}[\widehat{\Pi}(\Lambda^*, \mu^*)] \geq \alpha^* \mathbb{E}[\widehat{\Pi}_1] + (1-\alpha^*)\mathbb{E}[\widehat{\Pi}_2]$.

Theorem 12 provides a geometric characterization of the robustness of dynamic pricing policies. Recall that from Lemma 11, we have that the manifold $(\Lambda, \mu, \widehat{\lambda}_{bal}(\Lambda, \mu))$ is a cone. Moreover, any fixed static price corresponds to a ray of the cone. Theorem 12 shows that given a dynamic pricing policy with two prices p_1, p_2, the resulting throughput is bounded below by the hyperplane defined by the rays corresponding to the static policies using each of the two prices. As a special case, when Λ_0 is fixed, but μ_0 varies, then the throughput due to the dynamic pricing policy is at least as much as the linear interpolation between the two static pricing policies; this is the case in Figure 4. Moreover, we get a similar result for revenue.

6. Networks

This section presents a generalization of our model to a network of regions. We now generalize the queueing model to a network of regions. We present only the salient features of our model here – cf. Appendix 12 for details.

Queueing model. We assume the platform operates in a geographic area which is partitioned into a number of non-overlapping regions. Each region corresponds to a node in a graph $G(V, E)$, with edges between nodes representing direct roads between two regions. As before, a driver who is currently driving is either *available* (i.e., free to be matched to a passenger), or *busy* (i.e., giving a ride to a passenger) – however, now the driver is associated with the available/busy queue (denoted A_i and B_i) of his current region. Passengers arrive at each region i according to a Poisson process at rate μ_i. As before, a passenger lives for at most one ride – she requests a ride if $V > P_i(A_i)$ (where $P_i(A_i)$ denotes the current price in her region i), and is successfully matched if $A_i > 0$.

Driver dynamics. The main difference in the network model is in how we handle driver dynamics. We assume that the potential rate of driver-arrivals is Λ_0, while the actual exogenous rate is Λ. Next, when a new driver enters the system, we assume he chooses an initial region according to distribution $\sigma = (\sigma_i)_{i \in G}$ – thus the exogenous arrivals to region i is $\lambda_i^e = \sigma_i \Lambda^e$. To determine the destination of a ride and the ride time, we assume busy drivers perform a random walk on graph G in order to serve the ride; we discuss the technical details of this approach in Appendix 12. This random walk gives rise to $t_{ij} = \mathbb{P}[\text{dest} = j|\text{source} = i]$; the matrix $T = \{t_{ij}\}$ is henceforth referred to as the *traffic* matrix. When a driver completes a service at node i, we assume that the driver signs-out and departs from the system with probability $q_i^{exit} > 0$; alternatively, he chooses to stay in the system, and becomes available at j as an available driver with probability q_{ij}.

Analysis. The queueing model described above is an open Jackson network consisting of $M/M(k)/1$ queues for the available drivers, and $M/M/\infty$ queues for the busy drivers. Since all the queues are reversible, the resulting steady-state distribution is product-form (i.e., the queue-length distributions are independent; cf. Appendix 12). This allows us to generalize most of our results for the single region to networks (modulo some differences arising from the variation in demand/supply across regions). Due to lack of space, we only sketch our results here; the precise statements and their proofs can be found in Appendix 11.

First, we consider pricing policies where the platform charges a different static price in each region. We define an equivalent concept to the static balance-price policy studied for a single region (cf. (8)). Theorem 15 shows the extended characterization for this *localized* static policy and that an appropriate balance-price vector $\mathbf{p}_{bal}$ that maximizes Λ among all localized static policies. Moreover, analogously to Theorem 5 for a single region, Theorem 16 shows that revenue is maximized among localized static policies either at the balance-price vector $\mathbf{p}_{bal}$ or – when the system is not supply-constrained – at a component-wise marked-up price $\mathbf{p}_{d-opt}$.

Next, in the single region setting, Theorem 9 showed that no dynamic pricing policy outperforms static pricing with p_{bal}. We get an equivalent result for the network setting in Theorem 17. Note that, as in Theorem 9, we require $\mathbf{F}_V$ at each region to satisfy Equation (12). Moreover, we show that pricing policies with multiple thresholds also reduce to single-threshold policies (Theorems 18 and 19).

Given that the steady state distribution of the network model obeys a product-form distribution, it is reasonable to conjecture that our robustness result (cf. Theorem 12) generalizes to networks as well. This is an important direction of ongoing work.

Acknowledgments

The authors would like to thank the data science team at Lyft, particularly Adam Greenhall, Chris Pouliot, and Chris Sholley; many of the ideas in this work arose from discussions with them (and with Lyft drivers), and they played an important role in the formulation of the model we studied. (Part of this work was carried out when SB was a technical consultant at Lyft.)

We gratefully acknowledge support from the National Science Foundation, the Defense Advanced Research Projects Agency, and the Air Force Office of Scientific Research.

References

[1] I. Adan and G. Weiss. Exact fcfs matching rates for two infinite muti-type sequences. *Operations Research*, 60:475–489, 2012.

[2] N. Arnosti, R. Johari, and Y. Kanoria. Managing congestion in decentralized matching markets. *Available at SSRN 2427960*, 2014.

[3] E. M. Azevedo and E. B. Budish. Strategy-proofness in the large. *Chicago Booth Research Paper*, (13-35), 2013.

[4] G. Bitran and R. Caldentey. An overview of pricing models for revenue management. *Manufacturing & Service Operations Management*, 5(3):203–229, 2003.

[5] H. Chen and D. D. Yao. *Fundamentals of queueing networks: Performance, asymptotics, and optimization*, volume 46. Springer Science & Business Media, 2001.

[6] G. Gallego and G. Van Ryzin. Optimal dynamic pricing of inventories with stochastic demand over finite horizons. *Management science*, 40(8):999–1020, 1994.

[7] I. Gurvich, M. Lariviere, and A. Moreno. Staffing service systems when capacity has a mind of its own. *Available at SSRN 2336514*, 2014.

[8] R. Hassin and M. Haviv. *To queue or not to queue: Equilibrium behavior in queueing systems*, volume 59. Springer Science & Business Media, 2003.

[9] F. P. Kelly. *Reversibility and stochastic networks*. Cambridge University Press, 1979.

[10] F. Kojima and P. A. Pathak. Incentives and stability in large two-sided matching markets. *The American Economic Review*, pages 608–627, 2009.

[11] P. Naor. The regulation of queue size by levying tolls. *Econometrica: journal of the Econometric Society*, pages 15–24, 1969.

[12] J.-C. Rochet and J. Tirole. Two-sided markets: a progress report. *The RAND Journal of Economics*, 37(3):645–667, 2006.

[13] M. Rysman. The economics of two-sided markets. *The Journal of Economic Perspectives*, pages 125–143, 2009.

[14] K. T. Talluri and G. J. Van Ryzin. *The theory and practice of revenue management*, volume 68. Springer Science & Business Media, 2006.

[15] J. Visschers, I. Adan, and G. Weiss. A product form solution to a system with multi-type jobs and multi-type servers. *Queueing Systems*, 70(3):269–298, 2012.

[16] E. G. Weyl. A price theory of multi-sided platforms. *The American Economic Review*, pages 1642–1672, 2010.

7. Additional proofs for single queue with static pricing

In this appendix, we provide complete proofs for the results in Section 3. We start with the proof for Theorem 3:

Proof of Theorem 3. From Equation (7), we have that:

$$\lambda = \lambda(p,n) = \frac{n\Lambda_0}{q_{\text{exit}}}\,\mathbf{F}_C\left(\frac{\gamma p}{\tau + \frac{1}{n\mu_0\overline{\mathbf{F}}_V(p)-\lambda}}\right)\mathbb{1}_{\{n\mu_0\overline{\mathbf{F}}_V(p)>\lambda\}},$$

Note that $\lambda(p,n) \le \frac{n\Lambda_0}{q_{\text{exit}}}\,\mathbf{F}_C\left(\frac{\gamma p}{\tau}\right)$ and also, at equilibrium, we must have $\lambda(p,n) \le n\mu_0\overline{\mathbf{F}}_V(p)$. Furthermore, we can assume w.l.o.g that $\overline{\mathbf{F}}_V(p) \ge 0$ (else the only equilibrium solution is $\lambda(p,n) = 0$). Hence we have that either $\lambda(p,n) = n\mu_0\overline{\mathbf{F}}_V(p) - o(n)$, or $\lim_{n\to\infty}\lambda/n < \mu_0\overline{\mathbf{F}}_V(p)$.

If $\lambda(p,n) = n\mu_0\overline{\mathbf{F}}_V(p) - o(n)$, then as $n\to\infty$, we have $\frac{\lambda}{n} \to \mu_0\overline{\mathbf{F}}_V(p)$. However, since $\lambda(p,n) \le \frac{n\Lambda_0}{q_{\text{exit}}}\,\mathbf{F}_C\left(\frac{\gamma p}{\tau}\right)$, so we have that $\Lambda_0\,\mathbf{F}_C(\gamma p/\tau)/q_{\text{exit}} \ge\ \mu_0\,\overline{\mathbf{F}}_V(p)$.

If on the other hand $\lim_{n\to\infty}\lambda/n < \mu_0\overline{\mathbf{F}}_V(p)$, then for any $g(n) > 1$ such that $g(n) = \omega(1)$ and $g(n) = o(n)$ (i.e., $g(n)$ goes to ∞ at a sublinear rate), we have:

$$\begin{aligned}
\lambda &> \frac{n\Lambda_0}{q_{\text{exit}}}\,\mathbf{F}_C\left(\frac{\gamma p}{\tau + \frac{g(n)}{n\mu_0\overline{\mathbf{F}}_V(p)-\lambda}}\right)\\
&= \frac{n\Lambda_0}{q_{\text{exit}}}\mathbf{F}_C\left(\frac{\gamma p}{\tau}\right) - \frac{n\Lambda_0}{q_{\text{exit}}}\left[\left(\mathbf{F}_C\left(\frac{\gamma p}{\tau}\right) - \mathbf{F}_C\left(\frac{\gamma p}{\tau}\cdot\frac{1}{1+\frac{\tau^{-1}\,g(n)}{n\mu_0\overline{\mathbf{F}}_V(p)-\lambda}}\right)\right)\right]\\
&= \frac{n\Lambda_0}{q_{\text{exit}}}\mathbf{F}_C\left(\frac{\gamma p}{\tau}\right) - o(n).
\end{aligned}$$

For the last equality, we use the fact that as $n\to\infty$, we have:

$$\mathbf{F}_C\left(\frac{\gamma p}{\tau}\right) - \mathbf{F}_C\left(\frac{\gamma p}{\tau}\cdot\frac{1}{1+\frac{\tau^{-1}\,g(n)}{n\mu_0\overline{\mathbf{F}}_V(p)-\lambda}}\right) \qquad\to\qquad 0,$$

which follows from the continuity of $\mathbf{F}_C$, and the fact that $\frac{\tau^{-1}\,g(n)/n}{\mu_0\overline{\mathbf{F}}_V(p)-\lambda/n} \to 0$. Thus, we have that $\lambda(p,n) = \frac{n\Lambda_0}{q_{\text{exit}}}\mathbf{F}_C\left(\frac{\gamma p}{\tau}\right) - o(n)$ when $\Lambda_0\,\mathbf{F}_C(\gamma p/\tau)/q_{\text{exit}} <\ \mu_0\,\overline{\mathbf{F}}_V(p)$.

Combining both cases, we get the proposed result. $\square$

We also provide the full proof for Theorem 5:

Proof of Theorem 5. The normalized platform revenue is given by:

$$\begin{aligned}
\frac{\mathbb{E}_p[\Pi(p,n)]}{n} &= p\,\left(\min\left[\Lambda_0\,\mathbf{F}_C(\gamma p/\tau)/q_{\text{exit}},\ \mu_0\,\overline{\mathbf{F}}_V(p)\right] - \frac{o(n)}{n}\right)\\
\Rightarrow \lim_{n\to\infty}\frac{\mathbb{E}_p[\Pi(p,n)]}{n} &= p\cdot\min\left[\Lambda_0\mathbf{F}_C(\gamma p/\tau)/q_{\text{exit}}, \mu_0\overline{\mathbf{F}}_V(p)\right].
\end{aligned}$$

For prices $p \le p_{bal}$, we have from definition that $\Lambda_0\mathbf{F}_C(\gamma p/\tau)/q_{\text{exit}} \le \mu_0\overline{\mathbf{F}}_V(p)$. Thus, to maximize revenue, we need to maximize $p\mathbf{F}_C(\gamma p/\tau)$. This however is clearly increasing, and so the maximum is reached at the balance price p_{bal}.

On the other hand, for prices $p > p_{bal}$, we have $\Lambda_0 \mathbf{F}_C(\gamma p/\tau)/q_{\text{exit}} \geq \mu_0 \overline{\mathbf{F}}_V(p)$, and so we now want to maximize $p\overline{\mathbf{F}}_V(p)$. Recall though that we defined p_{d-opt} be the (highest) price that maximizes $p\overline{\mathbf{F}}_V(p)$. If $p_{d-opt} < p_{bal}$, then for $p = p_{d-opt}$ the objective function is actually $p\mathbf{F}_C(\gamma p/\tau)$, as $\Lambda_0 \mathbf{F}_C(\gamma p_{d-opt}/\tau)/q_{\text{exit}} \leq \mu_0 \overline{\mathbf{F}}_V(p_{d-opt})$. Further, $p\overline{\mathbf{F}}_V(p)$ is decreasing for $p > p_{d-opt}$ (since we defined it to be the highest revenue-maximizing price), so we conclude the global maximum is reached at p_{bal}. On the other hand, if $p_{d-opt} \geq p_{bal}$, it follows that the price maximizing revenue is precisely p_{d-opt}. Hence, $\lim_{n\to\infty} p_{opt}(n) = \max[p_{bal}, p_{d-opt}]$.

□

8. Proofs for the single region with dynamic pricing

In this appendix, we provide complete proofs, along with some additional results, for the results in Section 4.

To get qualitative insights into different dynamic-pricing policies, we again consider the scaled system parametrized by n (with $\Lambda_0(n) = \Lambda_0 n$ and $\mu_0(n) = \mu_0 n$), and consider the large-market limit $n \to \infty$. We assume all other parameters remain fixed (including prices p_ℓ and p_h), but allow the threshold θ to scale as $\theta(n) = g(n)$ (for some suitable function $g(n)$). In particular, suppose $g(n) = \omega(1)$ and invertible – then we can re-parametrize the system in terms of θ (i.e., $\Lambda_0(\theta) = \Lambda_0 g^{-1}(\theta), \mu_0(\theta) = \mu_0 g^{-1}(\theta)$). This lets us characterize the dependence of the platform metrics (cf. Equation (10)) on θ:

Theorem 13 *Consider a system with fixed λ, μ_0, distributions $\mathbf{F}_C, \mathbf{F}_V$ and prices p_ℓ, p_h satisfying $\rho\phi_l < 1$. Then the blocking probability $p_{block}(\theta)$ is monotonically decreasing, while the average earning-per-ride η and the average idle-time between rides ι are monotonically increasing in θ. Moreover, suppose we scale θ to ∞, and $\lambda(\theta), \mu(\theta)$ also scale to ∞ with θ but in a manner such that $\rho \triangleq \lambda(\theta)/\mu(\theta)$ stays constant. Then:*

1. *If $\rho\phi_\ell < \rho\phi_h < 1$:*
$$p_{block}(\theta) = (1-\rho\phi_h)(1+o(1)),\, \eta(\theta) = p_h - o(1),\, \iota(\theta) = \frac{1}{\mu(\theta)}\left(\frac{\phi_h}{1-\rho\phi_h}\right)(1+o(1)).$$
2. *If $\rho\phi_\ell < 1 < \rho\phi_h$:*
$$p_{block}(\theta) = e^{-\theta\log(\rho\phi_h)+o(\theta)}, \eta(\theta) = \frac{(1-\rho\phi_\ell)p_h + (\rho\phi_h - 1)p_\ell}{(\rho\phi_h - \rho\phi_\ell)} - o(1), \iota(\theta) = \frac{\theta}{\lambda}(1+o(\theta)).$$

Proof of Theorem 13. The change in blocking probability and idle time can be derived via a coupling argument. Consider the queue with the original θ, and couple it with a queue with threshold $\theta+1$ by modifying the departure process when there are θ available drivers – it is easy to see that the queue length in the latter queue is sample-pathwise greater, and hence the blocking probability is stochastically lesser and idle time stochastically greater. For the earnings-per-ride, we can re-write the expression as:

$$\mathbb{E}[\eta] = p_\ell + (p_h - p_\ell)\left(\frac{(\rho\phi_h)^\theta - 1}{\left(\frac{\rho\phi_h - \rho\phi_\ell}{1-\rho\phi_\ell}\right)(\rho\phi_h)^\theta - 1}\right)$$

Note that $p_h \geq p_\ell$ by definition, and also $\phi_h = (\overline{\mathbf{F}}_V(p_h))^{-1} \geq (\overline{\mathbf{F}}_V(p_\ell))^{-1} = \phi_\ell$. Let $a = \rho\phi_h, b = 1 - \rho\phi_\ell > 0$, and define $g(\theta) = \frac{a^\theta - 1}{(1+(a-1)/b)a^\theta - 1}$. If $a > 1$, we have $g'(\theta) = \frac{(a-1)a^\theta \log a}{b((1+(a-1)/b)a^\theta - 1)^2}$ which is positive $\forall\, \theta > 0$. Similarly, if $a < 1$, we have $g'(\theta) = \frac{a^{-\theta}(-\log a)(1-a)}{b((1-(1-a)/b) - a^{-\theta})^2}$, which again is positive $\forall\, \theta > 0$. This completes the first claim in the Theorem.

For the scaling behavior of the metrics with θ, first note that the second case corresponds to an unstable queue for all θ, and hence the corresponding metrics follow immediately. Similarly, in the first case, observe that for large values of θ, the queue essentially behaves as an $M/M/1$ queue with arrival-rate λ and departure-rate μ_0/ϕ_h. For the third case, the limits can be obtained via straightforward algebraic manipulations of the expressions in Equation (10). As an example, we derive the expression for the idle-time.

First recall that for fixed values of $\lambda, \mu_0, \phi_h, \phi_\ell$ and θ, we have:

$$\mathbb{E}[\iota] = \left(\frac{\pi^{dyn}[0]}{\lambda}\right)\left(\frac{\rho\phi_h(1+(\rho\phi_h)^\theta(\theta(\rho\phi_h - 1) - 1))}{(\rho\phi_h - 1)^2} + \frac{\rho\phi_\ell(\rho\phi_h)^\theta(1+\theta(1-\rho\phi_\ell))}{(1-\rho\phi_\ell)^2}\right)$$

Let us denote $\alpha = \rho\phi_h$ and $\beta = \rho\phi_\ell$, and consider the case where $\beta < 1 < \alpha$. Then, using the expression for $\pi^{dyn}[0]$, we can write the idle-time $\mathbb{E}[\iota]$ as a function $I(\theta)$ of θ as:

$$\begin{aligned}
\lambda I(\theta) &= \frac{(\alpha-1)(1-\beta)}{(\alpha-\beta)\alpha^\theta - (1-\beta)}\left(\frac{\alpha(1-\alpha^\theta + \theta\alpha^\theta(\alpha-1))}{(\alpha-1)^2} + \frac{\beta\alpha^\theta(1+\theta(1-\beta))}{(1-\beta)^2}\right) \\
\Rightarrow \frac{\lambda I(\theta)}{\theta} &= \frac{\alpha(1-\beta)((\alpha-1)+\theta^{-1}\alpha^{-\theta} - \theta^{-1})}{(\alpha-1)((\alpha-\beta)-(1-\beta)\alpha^{-\theta})} + \frac{\beta(\alpha-1)((1-\beta)+\theta^{-1})}{(1-\beta)((\alpha-\beta)-(1-\beta)\alpha^{-\theta})} \\
&= \frac{\alpha(1-\beta)(\alpha-1)}{(\alpha-1)(\alpha-\beta)} + \frac{\beta(\alpha-1)(1-\beta)}{(1-\beta)(\alpha-\beta)} + o(\theta) = 1 + o(\theta)
\end{aligned}$$

The expressions for the blocking probability and earning-per-ride follow in a similar manner. □

We are now in a position to understand the large-market limit of the ride-share platform under dynamic pricing. First, note that in both cases, average per-ride earnings are monotone in θ – moreover, as $n \to \infty$, the per-ride earnings converge to a constant independent of θ. On the other hand, although the average idle-time increases with increase in θ, if we scale θ with n, then ι goes to 0 as $n \to \infty$ for any choice of $\theta(n)$ in the first case, and as long as $\theta(n) = o(\sqrt{n})$ in the second case. Thus, to maximize either the throughput or the revenue, the platform needs to scale θ as $\omega(1)$, but in a manner such that ι goes to 0 (i.e., $\theta(n) = o(\sqrt{n})$). Moreover, Theorem 13 also shows that under such a scaling, the resulting limit is independent of θ. In light of this, *we henceforth consider dynamic-pricing policies* $(p_\ell, p_h, \theta(n))$*, where* $\theta(n) = \omega(1)$ *but* $o(\sqrt{n})$. Also, since the effect of θ disappears in the limit, we parametrize the limiting expressions for $\widehat{\lambda}, \mathbb{E}[\widehat{\Pi}], \mathbb{E}[\widehat{W}]$ in terms of (p_ℓ, p_h).

Next, we have the proofs for Theorem 9 and Corollary 10:

Proof of Theorem (9). To characterize the performance of dynamic pricing when $p_\ell < p_{bal} \leq p_h$, we need to characterize the solution to the fixed-point equation:

$$\lambda = \frac{\Lambda_0}{q_{\mathsf{exit}}}\mathbf{F}_C\left(\frac{\gamma}{\tau}\left(p_\ell + \frac{p_h - p_\ell}{\phi_h - \phi_\ell}\left(\frac{\mu_0}{\lambda} - \phi_\ell\right)\right)\right) \tag{13}$$

First, note that for $\lambda \in [0, \Lambda_0/q_{\text{exit}}]$, Equation (13) always has a *unique fixed-point* – to see this, observe that the LHS is strictly increasing in λ, with range $[0, \Lambda_0/q_{\text{exit}}]$, while the RHS is non-increasing, with range $\subseteq [0, \Lambda_0/q_{\text{exit}}]$.

Next, for $p_\ell < p_h = p_{bal}$, we claim that $\lambda_{bal} = \frac{\Lambda_0}{q_{\text{exit}}} \mathbf{F}_C\left(\frac{\gamma p_{bal}}{\tau}\right) = \mu_0 \overline{\mathbf{F}}_V(p_{bal})$ is a fixed-point of Equation (13). To see this, setting $\phi_h \triangleq \phi_{bal} = 1/\overline{\mathbf{F}}_V(p_{bal})$, we get:

$$\begin{aligned}\frac{\Lambda_0}{q_{\text{exit}}}\mathbf{F}_C\left(\frac{\gamma}{\tau}\left(p_\ell + \frac{p_{bal}-p_\ell}{\phi_{bal}-\phi_\ell}\left(\frac{\mu_0}{\lambda_{bal}} - \phi_\ell\right)\right)\right) &= \frac{\Lambda_0}{q_{\text{exit}}}\mathbf{F}_C\left(\frac{\gamma}{\tau}\left(p_\ell + \left(\frac{p_{bal}-p_\ell}{\phi_{bal}-\phi_\ell}\right)(\phi_{bal}-\phi_\ell)\right)\right) \\ &= \frac{\Lambda_0}{q_{\text{exit}}}\mathbf{F}_C\left(\frac{\gamma}{\tau}(p_\ell + p_{bal} - p_\ell)\right) = \lambda_{bal}\end{aligned}$$

In addition, the above proof also shows that $\mathbb{E}[\eta] = p_{bal}$ whenever $p_\ell \le p_{bal}$ and $p_h = p_{bal}$ – this gives us the claim regarding the platforms revenue in the large-market limit.

Next, we want to show that if $p_\ell < p_{bal} < p_h$, then $\lambda \le \lambda_{bal}$. Note that we can always define $\mathbf{F}_C^{-1}(y), y \in [0,1]$ as the inverse of $\mathbf{F}_C(x)$. Now, setting $p_h = p$ and $\phi(p) = 1/\overline{\mathbf{F}}_V(p)$, we can re-write Equation (13) as:

$$\frac{\tau}{\gamma}\mathbf{F}_C^{-1}\left(\frac{q_{\text{exit}}\lambda(p)}{\Lambda_0}\right) = p_\ell + \left(\frac{p-p_\ell}{\phi(p)-\phi_\ell}\right)\left(\frac{\mu_0}{\lambda(p)} - \phi_\ell\right)$$

Assuming that $\mathbf{F}_C^{-1}$ is differentiable, we can differentiate both sides w.r.t p to get:

$$\begin{aligned}&\lambda'(p)\frac{q_{\text{exit}}}{\Lambda_0}\frac{d}{d\lambda}\left(\frac{\tau}{\gamma}\mathbf{F}_C^{-1}\left(\frac{q_{\text{exit}}\lambda(p)}{\Lambda_0}\right)\right) \\ &\quad= \left(\frac{\mu_0}{\lambda(p)} - \phi_\ell\right)\left(\frac{\phi_p - \phi_\ell - (\phi_p)^2 f_V(p)(p-p_\ell)}{(\phi(p)-\phi_\ell)^2}\right) - \frac{\mu_0}{\lambda(p)^2}\left(\frac{p-p_\ell}{\phi(p)-\phi_\ell}\right)\lambda'(p) \\ &\Rightarrow \lambda'(p) = \left(\left(\frac{q_{\text{exit}}\tau}{\Lambda_0\gamma}\right)\frac{d\mathbf{F}_C^{-1}\left(\frac{q_{\text{exit}}\lambda(p)}{\Lambda_0}\right)}{d\lambda} + \frac{\mu_0}{\lambda(p)^2}\left(\frac{p-p_\ell}{\phi(p)-\phi_\ell}\right)\right)^{-1}. \\ &\qquad\cdots\left(\frac{\mu_0}{\lambda(p)} - \phi_\ell\right)\left(\frac{\phi_p - \phi_\ell - (\phi_p)^2 f_V(p)(p-p_\ell)}{(\phi(p)-\phi_\ell)^2}\right)\end{aligned} \tag{14}$$

Now note that $\frac{d\mathbf{F}_C^{-1}\left(\frac{q_{\text{exit}}\lambda(p)}{\Lambda_0}\right)}{d\lambda} \ge 0$ as $\mathbf{F}_C^{-1}(x)$ is non-decreasing for $x \in [0,1]$. Also, if $p > p_\ell$, we have $\phi(p) \ge \phi_\ell$ and so the coefficient of $\lambda'(p)$ in the LHS is positive. Next, by stability requirements, we have $\lambda(p) \le \mu_0/\phi_\ell$, which implies $\left(\frac{\mu_0}{\lambda(p)} - \phi_\ell\right) \ge 0$. Finally, consider the function $g(p) = \phi_p - \phi_\ell - (\phi_p)^2 f_V(p)(p-p_\ell)$. We can write:

$$\begin{aligned}g(p) &= \phi(p) - \phi_\ell - \phi(p)^2 f_V(p)(p-p_\ell) \\ &= \phi(p)^2\phi_\ell\Big[\overline{\mathbf{F}}_V(p_\ell)\overline{\mathbf{F}}_V(p) - \overline{\mathbf{F}}_V(p)^2 - \overline{\mathbf{F}}_V(p_\ell) f_V(p)(p-p_\ell)\Big] \\ &= \phi(p)^2\phi_\ell\Big[\overline{\mathbf{F}}_V(p)\left(\mathbf{F}_V(p) - \mathbf{F}_V(p_\ell)\right) - f_V(p)(p-p_\ell)\overline{\mathbf{F}}_V(p_\ell)\Big]\end{aligned}$$

Thus, for $g(p) < 0$, we need $\overline{\mathbf{F}}_V(p)\left(\mathbf{F}_V(p) - \mathbf{F}_V(p_\ell)\right) - f_V(p)(p-p_\ell)\overline{\mathbf{F}}_V(p_\ell) < 0$. This however is precisely the condition we stated in Equation (12). Now, substituting back in Equation (14), we have that $\lambda'(p) \le 0\, \forall\, p > p_{bal}$. Since we already have that $\lambda(p_{bal}) = \lambda_{bal}$, thus we get that $\lambda(p) \le \lambda_{bal}\, \forall\, p_h > p_{bal}$

Revenue: Next, we want to show that the platform revenue under dynamic pricing is also bounded by that under static pricing with $p = p_{bal}$. Now we have $\mathbb{E}[\Pi^{dyn}(p)] = (1-\gamma)\lambda(p)\mathbb{E}[\eta(p)]$. From above, we already know that $\lambda(p)$ is decreasing for $p \geq p_{bal}$ – thus, it is sufficient to show that $\mathbb{E}[\eta(p)]$ is also decreasing in p in the same range. Now we have:

$$\frac{d\mathbb{E}[\eta(p)]}{dp} = \frac{d}{dp}\left(p_\ell + \left(\frac{p-p_\ell}{\phi(p)-\phi_\ell}\right)\left(\frac{\mu_0}{\lambda(p)} - \phi_\ell\right)\right) = \lambda'(p)\frac{q_{\text{exit}}\tau}{\Lambda_0\gamma}\frac{d}{d\lambda}\left(\mathbf{F}_C^{-1}\left(\frac{q_{\text{exit}}\lambda(p)}{\Lambda_0}\right)\right),$$

where the last line follows from the derivation of Equation (14). Now, as before, we have that $\frac{d}{d\lambda}\left(\mathbf{F}_C^{-1}\left(\frac{q_{\text{exit}}\lambda(p)}{\Lambda_0}\right)\right) > 0$, and also from Equation (14), we know that $d\lambda(p)/dp < 0$ – thus we have that $\mathbb{E}[\eta(p)]$ is decreasing in p, and hence so is $\mathbb{E}[\Pi^{dyn}(p)]$

□

Proof of Corollary 10. Rewriting Equation (12), we have that we want to show $g(p) \triangleq \overline{\mathbf{F}}_V(p)\left(\mathbf{F}_V(p) - \mathbf{F}_V(p_\ell)\right) - f_V(p)(p-p_\ell)\overline{\mathbf{F}}_V(p_\ell) < 0$:

1. Suppose the hazard-rate of $\mathbf{F}_V$ is given by $h_V(x)$ – this implies $\overline{\mathbf{F}}_V(x) = \exp\left(-\int_0^x h_V(t)dt\right)$ and $f_V(x) = h_V(x)\overline{\mathbf{F}}_V(x)$. Substituting in the above condition, we get:

$$\begin{aligned}
g(p) &= \overline{\mathbf{F}}_V(p)\left(\mathbf{F}_V(p) - \mathbf{F}_V(p_\ell)\right) - f_V(p)(p-p_\ell)\overline{\mathbf{F}}_V(p_\ell) \\
&= \overline{\mathbf{F}}_V(p)\overline{\mathbf{F}}_V(p_\ell)\left(1 - \frac{\overline{\mathbf{F}}_V(p)}{\overline{\mathbf{F}}_V(p_\ell)} - \frac{f_V(p)(p-p_\ell)}{\overline{\mathbf{F}}_V(p)}\right) \\
&\leq 1 - \exp\left(-\int_{p_\ell}^{p} h_V(t)dt\right) - (p-p_\ell)h_V(p) \\
&\leq 1 - (p-p_\ell)h_V(p) - \exp\left(-(p-p_\ell)h_V(p)\right) \quad \text{(By the MHR condition)} \\
&\leq 0 \quad (\text{Since } (1-x) < e^{-x}\,\forall\, x > 0)
\end{aligned}$$

2. Let $\overline{\mathbf{F}}_V(p) = \left(\frac{p_{min}}{p}\right)^\alpha$, with $p_\ell \geq p_{min} > 0$ and $\alpha \geq 1$. Consider $p = p_\ell + \Delta_p$, for some $\Delta_p > 0$. Substituting in $g(p)$, we get:

$$\begin{aligned}
g(p) &= \overline{\mathbf{F}}_V(p)\overline{\mathbf{F}}_V(p_\ell)\left(1 - \frac{\overline{\mathbf{F}}_V(p)}{\overline{\mathbf{F}}_V(p_\ell)} - \frac{f_V(p)(p-p_\ell)}{\overline{\mathbf{F}}_V(p)}\right) \\
&= \overline{\mathbf{F}}_V(p)\overline{\mathbf{F}}_V(p_\ell)\left(1 - \left(\frac{p_\ell}{p}\right)^\alpha - \frac{\alpha(p-p_\ell)}{p}\right) \\
&= \overline{\mathbf{F}}_V(p)\overline{\mathbf{F}}_V(p_\ell)\left(\left(1 - \alpha\frac{\Delta_p}{p}\right) - \left(1 - \frac{\Delta_p}{p}\right)^\alpha\right) \\
&\leq 0 \quad (\text{Since } (1-x)^\alpha \geq (1-\alpha x)\;\forall\, x \in [0,1], \alpha > 1)
\end{aligned}$$

This completes verifying Equation (12) in both the cases. □

9. Additional Details on Robustness of Dynamic Pricing

In this appendix, we provide proofs for the results in Section 5, and some additional numerical demonstrations of the robustness of dynamic pricing policies. We start with the proof for Lemma 11

Proof of Lemma 11 The proof directly follows from the definition of balance price p_{bal}: $\Lambda \mathbf{F}_C\left(\gamma p_{bal}/\tau\right) = \mu \overline{\mathbf{F}}_V(p_{bal}) = \lambda_{bal}(\Lambda,\mu)$. From the first equality, we get the locus of (Λ,μ) for a given p_{bal}; the second shows that $\lambda_{bal}(\Lambda,\mu)$ scales linearly with (Λ,μ). $\square$

Finally, we have the proof of Theorem 12:

Proof of Theorem 12 For ease of notation, in this proof we use λ to denote the normalized large-market limit, i.e., $\lambda = \widehat{\lambda} := \lim_{n\to\infty} \lambda(n)/n$, where $\lambda(n)$ corresponds to the rate of incoming drivers to the $n-$th system (with $\Lambda(n) = n\Lambda$ and $\mu(n) = n\mu$).

First, note that since (Λ^*,μ^*) lies on the line connecting Γ_1 and Γ_2, then we have that $p_{bal}(\Lambda^*,\mu^*) \in [p_1,p_2]$ – this follows from the geometric characterization in Lemma 11. Next w.l.o.g, we assume $p_1 < p_2$; furthermore, we can also set $q_{\mathsf{exit}} = 1$ and $\gamma/\tau = 1$ (since otherwise we can re-scale the Λ,μ axes). Finally recall that $\phi_i = 1/\overline{\mathbf{F}}_v(p_i)$, $i \in \{1,2\}$, and also, from Equation (8), we have that $\lambda_1 = \mu\overline{\mathbf{F}}_V(p_1) = \Lambda F_C(p_1)$.

Our goal is to show that, under dynamic-pricing policy $(p_1,p_2,\theta(n))$, the normalized driver arrival-rate in the large-market limit $\lambda(\Lambda^*,\mu^*) > \alpha^*\lambda_1 + (1-\alpha^*)\lambda_2$. Let $\lambda = \lambda(\Lambda^*,\mu^*)$; modifying Equation (11) using the above assumptions, we have:

$$\lambda = \Lambda\ \mathbf{F}_C\left(p_\ell\left(\frac{\phi_2-\mu_0/\lambda}{\phi_2-\phi_\ell}\right) + p_h\left(\frac{\mu_0/\lambda-\phi_1}{\phi_h-\phi_1}\right)\right) = \Lambda\ \mathbf{F}_C\left(\frac{p_1\left(\frac{\mu_2}{\lambda_2}-\frac{\mu}{\lambda}\right)+p_2\left(\frac{\mu}{\lambda}-\frac{\mu_1}{\lambda_1}\right)}{\frac{\mu_2}{\lambda_2}-\frac{\mu_1}{\lambda_1}}\right).$$

To characterize the properties of the equilibrium λ under dynamic pricing, we define the gap function g as $g(\lambda,\Lambda,\mu) = \lambda - \Lambda\ \mathbf{F}_C\left(\frac{p_1\left(\frac{\mu_2}{\lambda_2}-\frac{\mu}{\lambda}\right)+p_2\left(\frac{\mu}{\lambda}-\frac{\mu_1}{\lambda_1}\right)}{\frac{\mu_2}{\lambda_2}-\frac{\mu_1}{\lambda_1}}\right)$. Now we have:

$$\begin{aligned}
\frac{g(\lambda(\alpha),\Lambda(\alpha),\mu(\alpha))}{\Lambda(\alpha)} &= \frac{\lambda(\alpha)}{\Lambda(\alpha)} - \mathbf{F}_C\left(\frac{p_1\left(\frac{\mu_2}{\lambda_2}-\frac{\mu(\alpha)}{\lambda(\alpha)}\right)+p_2\left(\frac{\mu(\alpha)}{\lambda(\alpha)}-\frac{\mu_1}{\lambda_1}\right)}{\frac{\mu_2}{\lambda_2}-\frac{\mu_1}{\lambda_1}}\right)\\
&= \frac{\lambda(\alpha)}{\Lambda(\alpha)} - \mathbf{F}_C\left(\frac{p_1\left(\frac{\mu_2}{\lambda_2}-\frac{t\mu_1+(1-\alpha)\mu_2}{\alpha\lambda_1+(1-\alpha)\lambda_2}\right)+p_2\left(\frac{t\mu_1+(1-\alpha)\mu_2}{\alpha\lambda_1+(1-\alpha)\lambda_2}-\frac{\mu_1}{\lambda_1}\right)}{\frac{\lambda_1\mu_2-\lambda_2\mu_1}{\lambda_1\lambda_2}}\right)\\
&= \frac{\alpha\lambda_1+(1-\alpha)\lambda_2}{\alpha\Lambda_1+(1-\alpha)\Lambda_2} - \mathbf{F}_C\left(\frac{\alpha\lambda_1}{\alpha\lambda_1+(1-\alpha)\lambda_2}p_1 + \frac{(1-\alpha)\lambda_2}{\alpha\lambda_1+(1-\alpha)\lambda_2}p_2\right).
\end{aligned}$$

Note that for a fixed value of (Λ,μ), the second term of $g(\lambda,\Lambda,\mu)$ is non-increasing in λ, hence, it follows that $g(\lambda,\Lambda,\mu)$ is non-decreasing in λ. Also, by definition, we know that $g(\lambda(\Lambda,\mu),\Lambda,\mu) = 0$. We conclude that $\lambda(\alpha) < \lambda_\pi(\Lambda(\alpha),\mu(\alpha))$ if and only if $g(\lambda(\alpha),\Lambda(\alpha),\mu(\alpha)) < 0$ or, equivalently, if and only if:

$$\begin{aligned}
\frac{t\alpha\Lambda_1\mathbf{F}_C(p_1)+(1-\alpha)\Lambda_2\mathbf{F}_C(p_2)}{t\Lambda_1+(1-\alpha)\Lambda_2} &= \frac{\alpha\lambda_1+(1-\alpha)\lambda_2}{\alpha\Lambda_1+(1-\alpha)\Lambda_2}\\
&\le \mathbf{F}_C\left(\frac{\alpha\lambda_1}{\alpha\lambda_1+(1-\alpha)\lambda_2}p_1 + \frac{(1-\alpha)\lambda_2}{\alpha\lambda_1+(1-\alpha)\lambda_2}p_2\right)
\end{aligned}$$

Since both sides are positive, proving the above identity is equivalent to showing:

$$\log\left(\frac{t\Lambda_1\mathbf{F}_C(p_1)+(1-\alpha)\Lambda_2\mathbf{F}_C(p_2)}{t\Lambda_1+(1-\alpha)\Lambda_2}\right) \le \log\left(\mathbf{F}_C\left(\frac{\alpha\lambda_1}{\alpha\lambda_1+(1-\alpha)\lambda_2}p_1 + \frac{(1-\alpha)\lambda_2}{\alpha\lambda_1+(1-\alpha)\lambda_2}p_2\right)\right). \tag{15}$$

Note that by the log-concavity of $\mathbf{F}_C$, the RHS satisfies:

$$\log\left(\mathbf{F}_C\left(\frac{\alpha\lambda_1}{\alpha\lambda_1+(1-\alpha)\lambda_2}p_1+\frac{(1-\alpha)\lambda_2}{\alpha\lambda_1+(1-\alpha)\lambda_2}p_2\right)\right)$$
$$\geq \frac{\alpha\lambda_1}{\alpha\lambda_1+(1-\alpha)\lambda_2}\log(\mathbf{F}_C(p_1))+\frac{(1-\alpha)\lambda_2}{\alpha\lambda_1+(1-\alpha)\lambda_2}\log(\mathbf{F}_C(p_2))$$
$$=\frac{\alpha\Lambda_1\mathbf{F}_C(p_1)\log(\mathbf{F}_C(p_1))+(1-\alpha)\Lambda_2\mathbf{F}_C(p_2)\log(\mathbf{F}_C(p_2))}{\alpha\Lambda_1\mathbf{F}_C(p_1)+(1-\alpha)\Lambda_2\mathbf{F}_C(p_2)}.$$

Finally, applying the log-sum inequality [6], with $a=[\alpha\Lambda_1\mathbf{F}_C(p_1),(1-\alpha)\Lambda_2\mathbf{F}_C(p_2)]$, $b=[\alpha\Lambda_1,(1-\alpha)\Lambda_2]$, we get:

$$\frac{\alpha\Lambda_1\mathbf{F}_C(p_1)\log(\mathbf{F}_C(p_1))+(1-\alpha)\Lambda_2\mathbf{F}_C(p_2)\log(\mathbf{F}_C(p_2))}{\alpha\Lambda_1\mathbf{F}_C(p_1)+(1-\alpha)\Lambda_2\mathbf{F}_C(p_2)}$$
$$\geq \log\left(\frac{\alpha\Lambda_1\mathbf{F}_C(p_1)+(1-\alpha)\Lambda_2\mathbf{F}_C(p_2)}{\alpha\Lambda_1+(1-\alpha)\Lambda_2}\right),$$

which is the LHS of (15). This completes the proof of robustness for $\hat{\lambda}(\Lambda^*,\mu^*)$.

The above robustness argument also extends to revenue robustness, under the additional assumption that $p_2>p_1>p_{d-opt}$. First, recall that $\Pi(\widehat{\lambda}_i)=p_i\widehat{\lambda}_i$ and $\Pi(\widehat{\lambda})=\mathbb{E}_\pi[P]\widehat{\lambda}$. For *fixed* μ^*, assume that $\Lambda^*\in[\Lambda_2,\Lambda_1]$. Again, let π be the dynamic-pricing policy $\pi=(p_1,p_2,\theta(n))$, where $(p_1,\widehat{\lambda}_1):=(p_{bal}(\Lambda_1,\mu),\widehat{\lambda}_{bal}(\Lambda_1,\mu))$, $(p_2,\widehat{\lambda}_2):=(p_{bal}(\Lambda_2,\mu),\widehat{\lambda}_{bal}(\Lambda_2,\mu))$, and $\theta(n)=\omega(1)$ but $o(n)$. Note that $p_1<p_2$, so that $\widehat{\lambda}_1>\widehat{\lambda}_2$. We write $\bar{\alpha}^*=1-\alpha^*$. Let $\widehat{\lambda}=\widehat{\lambda}(\Lambda^*,\mu^*)$ be the rate of incoming drivers induced by policy π under parameters Λ^*,μ^*. By the robustness result with respect to $\widehat{\lambda}$, we know that:

$$\widehat{\lambda}(\Lambda^*,\mu^*)>\alpha^*\widehat{\lambda}_1+\bar{\alpha}^*\widehat{\lambda}_2. \tag{16}$$

We want to show a similar statement for revenue, i.e.,

$$\mathbb{E}[\Pi(\widehat{\lambda})]>\alpha^*\mathbb{E}[\Pi(\widehat{\lambda}_1)]+\bar{\alpha}^*\mathbb{E}[\Pi(\widehat{\lambda}_2)].$$

Equivalently, by the definition of revenue, we want to show:

$$\widehat{\lambda}\mathbb{E}_\pi[P]>\alpha^*\widehat{\lambda}_1p_1+\bar{\alpha}^*\widehat{\lambda}_2p_2.$$

In the limit $n\to\infty$, $\pi(p_1)=\frac{\rho\phi_2-1}{\rho(\phi_2-\phi_1)}$, $\pi(p_2)=\frac{1-\rho\phi_1}{\rho(\phi_2-\phi_1)}$. Therefore, we expand the expressions to find that:

$$\mathbb{E}_\pi[P]=\pi(p_1)p_1+\pi(p_2)p_2=\frac{\rho\phi_2-1}{\rho(\phi_2-\phi_1)}\cdot p_1+\frac{1-\rho\phi_1}{\rho(\phi_2-\phi_1)}\cdot p_2$$

[6] Let $a_i,b_i>0$ for $i=1,\dots,n$ and define $a=\sum_i a_i, b=\sum_i b_i$. Then $\sum_i a_i\log(a_i/b_i)\geq a\log a/b$.

Recall that $\widehat{\lambda}_i = \mu/\phi_i$. It follows that:

$$
\begin{aligned}
&\mathbb{E}[\Pi(\widehat{\lambda})] - \alpha^*\mathbb{E}[\Pi(\widehat{\lambda}_1)] - \bar{\alpha}^*\mathbb{E}[\Pi(\widehat{\lambda}_2)] \\
&= \left(\widehat{\lambda}\frac{\rho\phi_2 - 1}{\rho(\phi_2-\phi_1)} - \alpha^*\widehat{\lambda}_1\right)p_1 + \left(\widehat{\lambda}\frac{1-\rho\phi_1}{\rho(\phi_2-\phi_1)} - \bar{\alpha}^*\widehat{\lambda}_2\right)p_2 \\
&= \left(\frac{\widehat{\lambda}\phi_2 - \mu}{\phi_2-\phi_1} - \alpha^*\widehat{\lambda}_1\right)p_1 + \left(\frac{\mu - \widehat{\lambda}\phi_1}{\phi_2-\phi_1} - \bar{\alpha}^*\widehat{\lambda}_2\right)p_2 \\
&= \left(\frac{\widehat{\lambda}/\widehat{\lambda}_2 - 1}{1/\widehat{\lambda}_2 - 1/\widehat{\lambda}_1} - \alpha^*\widehat{\lambda}_1\right)p_1 + \left(\frac{1-\widehat{\lambda}/\widehat{\lambda}_1}{1/\widehat{\lambda}_2 - 1/\widehat{\lambda}_1} - \bar{\alpha}^*\widehat{\lambda}_2\right)p_2 \\
&= \frac{1}{1/\widehat{\lambda}_2 - 1/\widehat{\lambda}_1}\left[\left(\frac{\widehat{\lambda}}{\widehat{\lambda}_2} - 1 - \alpha^*\left(\frac{\widehat{\lambda}_1}{\widehat{\lambda}_2} - 1\right)\right)p_1 + \left(1 - \frac{\widehat{\lambda}}{\widehat{\lambda}_1} - \bar{\alpha}^*\left(1 - \frac{\widehat{\lambda}_2}{\widehat{\lambda}_1}\right)\right)p_2\right] \\
&= \frac{1}{1/\widehat{\lambda}_2 - 1/\widehat{\lambda}_1}\left[\left(\frac{\widehat{\lambda} - \alpha^*\widehat{\lambda}_1}{\widehat{\lambda}_2} - 1 + \alpha^*\right)p_1 + \left(\frac{\bar{\alpha}^*\widehat{\lambda}_2 - \widehat{\lambda}}{\widehat{\lambda}_1} + 1 - \bar{\alpha}^*\right)p_2\right] \\
&= \frac{1}{1/\widehat{\lambda}_2 - 1/\widehat{\lambda}_1}\left[\left(\frac{\widehat{\lambda} - \alpha^*\widehat{\lambda}_1 - \bar{\alpha}^*\widehat{\lambda}_2}{\widehat{\lambda}_2}\right)p_1 + \left(\frac{\alpha^*\widehat{\lambda}_1 + \bar{\alpha}^*\widehat{\lambda}_2 - \widehat{\lambda}}{\widehat{\lambda}_1}\right)p_2\right] \\
&= \frac{\widehat{\lambda} - \alpha^*\widehat{\lambda}_1 - \bar{\alpha}^*\widehat{\lambda}_2}{1/\widehat{\lambda}_2 - 1/\widehat{\lambda}_1}\left[\frac{p_1}{\widehat{\lambda}_2} - \frac{p_2}{\widehat{\lambda}_1}\right] = \frac{\widehat{\lambda} - \alpha^*\widehat{\lambda}_1 - \bar{\alpha}^*\widehat{\lambda}_2}{1/\widehat{\lambda}_2 - 1/\widehat{\lambda}_1}\mu\left[\frac{p_1/\phi_1 - p_2/\phi_2}{\widehat{\lambda}_1\widehat{\lambda}_2}\right].
\end{aligned}
$$

By (16), we know that $\widehat{\lambda} - \alpha^*\widehat{\lambda}_1 - \bar{\alpha}^*\widehat{\lambda}_2 > 0$ and, as $\widehat{\lambda}_1 > \widehat{\lambda}_2$, it follows that $1/\widehat{\lambda}_2 - 1/\widehat{\lambda}_1 > 0$. Moreover, the value of $p\bar{F}_V(p) = p/\phi_p$ decreases for values of p larger than $p = p_{d-opt}$, which implies that $p_1/\phi_1 - p_2/\phi_2 > 0$. We conclude that $\mathbb{E}[\Pi(\widehat{\lambda})] > \alpha^*\mathbb{E}[\Pi(\widehat{\lambda}_1)] + \bar{\alpha}^*\mathbb{E}[\Pi(\widehat{\lambda}_2)]$. $\square$

10. Dynamic Pricing with Multiple Thresholds

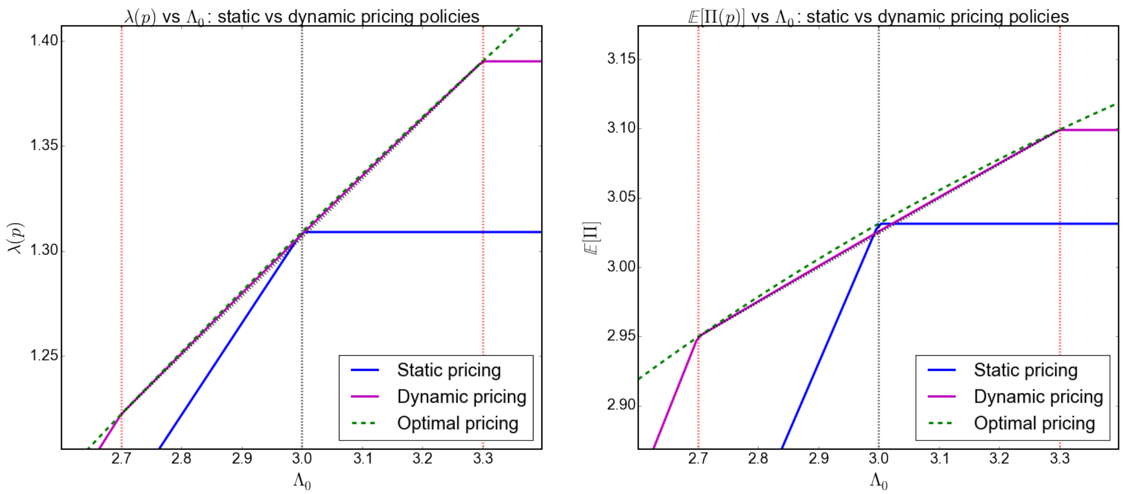
In Section 4, we discussed a dynamic-pricing policy with two prices (p_ℓ, p_h) and a threshold θ. A more general dynamic-pricing policy can allow multiple thresholds $\theta_1 > \theta_2 > \ldots > \theta_k$ (for some $k \in \mathbb{N}_+$) and prices $p_1 < p_2 < \ldots < p_{k+1}$. The highest price p_{k+1} is charged if the number of available drivers is less than or equal to the minimum threshold θ_k, the lowest price p_1 charged if the number of available drivers (strictly) exceeds the maximum threshold θ_1, and price p_i, $1 < i < k+1$ charged if the number of available drivers lies in $(\theta_i, \theta_{i-1}]$. Characterizing the performance for a general threshold policy can be challenging even in the large-market limit. However, we now show that the behavior of a wide range of pricing policies in the large-market limit reduces to that of an appropriate single-threshold policy.

Theorem 14 *Given a pricing schedule $(\underline{\theta}, \underline{\mathbf{p}})$ such that:*

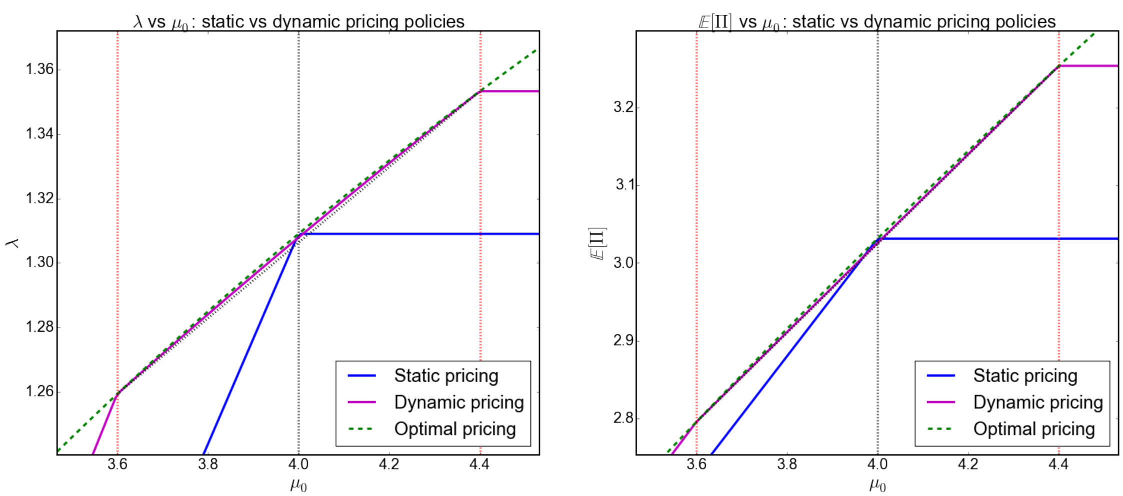
$$p_i \le p_{i+1} \,\forall\, i \in \{1,2,\ldots,k\}, \qquad \theta_i(n) - \theta_{i+1}(n) = \omega(1) \,\forall\, i \in \{1,2,\ldots,k-1\}$$

Suppose $p_{bal} \le p_{k+1}$. Then there exists a unique $j^* \in \{1,\ldots,k\}$ s.t.:

1. $p_{bal} \in [p_{j^*-1}, p_{j^*}]$.



(a) Robustness to Supply Variability



(b) Robustness to Demand Variability

Figure 5 Performance (in large-market limit) of static and dynamic pricing under changes in Λ_0 and μ_0. In Figure 5(a), we fix $\mu_0 = 4$, and vary the potential arrival-rate of drivers Λ_0 as $3 \pm 10\%$. We then compare a static-pricing with p_{bal} based on $(\Lambda_0, \mu_0) = (3,4)$, and a dynamic-pricing policy with p_ℓ set as the p_{bal} for $(\Lambda_0, \mu_0) = (3.3, 4)$, and p_h set as the the p_{bal} for $(\Lambda_0, \mu_0) = (2.7, 4)$. Similarly, in Figure 5(a), we set $\Lambda_0 = 3$, vary the potential arrival-rate of passengers μ_0 as $4 \pm 10\%$, and design the static and dynamic pricing policies as before. In both plots, we plot the normalized metrics $(\widehat{\lambda}, \mathbb{E}[\widehat{\Pi}])$, using distributions $\mathbf{F}_V \sim Gamma(2,1), \mathbf{F}_C \sim Lognormal(1,1)$. The dashed green curve shows the performance of the system under static-pricing with the correct p_{bal} corresponding to the actual Λ_0, μ_0. The the dotted black vertical line indicates the Λ_0 (respectively, μ_0) which was used to fix the static-pricing policy. The dotted red vertical lines mark the Λ_0 (correspondingly, μ_0) for which the balance price is p_ℓ and p_h.

2. *For any function $c(n) = \omega(1)$, we have:*

$$\pi\left([\theta_{j^*} - c(n), \theta_{j^*} + c(n)]\right) \to 1, \qquad \text{as } n \to \infty.$$

[of Theorem 14]

We first prove the second statement of the theorem, and then we show the first one. So we start showing that, under the assumptions of the theorem, for any function $c(n) = \omega(1)$, we have:

$$\pi\left([\theta_{j^*} - c(n), \theta_{j^*} + c(n)]\right) \to 1, \qquad \text{as } n \to \infty.$$

We start by noting that due to the monotonicity of $\bar{F}_V$ we have that

$$p_1 < p_2 < \cdots < p_k, \qquad \phi_1 < \phi_2 < \cdots < \phi_k.$$

For convenience, we define $m(i) = \theta_i - \theta_{i+1}$, and $m(k-1) = \theta_{k-1}$.

After fixing the pricing schedule, the rate at which new drivers enter the system is determined and, say, equal to λ. We define ρ to be the ratio λ/μ, where μ is the rate at which potential passengers arrive to the system. The number of available drivers at time t is then distributed as a birth–death process where the birth rate is independent of the state —and equal to λ— but the death rate depends on the price that was charged to passengers arriving at that state of the system. In particular, at those states charging p, the death rate is precisely μ/ϕ_p. We can now compute the steady-state distribution π to find that it is given by

$$\begin{aligned}
\pi(i) &= (\rho\phi_k)^i\, \pi(0), && i \le \theta_{k-1},\\
\pi(i) &= \rho^i \left[\prod_{s=j+1}^{k} \phi_s^{m(s-1)}\right] \phi_j^{i-\theta_j}\, \pi(0), && \theta_j < i \le \theta_{j-1},\\
\pi(i) &= \rho^i \left[\prod_{s=2}^{k} \phi_s^{m(s-1)}\right] \phi_1^{i-\theta_1}\, \pi(0), && i > \theta_1.
\end{aligned}$$

We can safely assume that $1 > \rho\phi_1$, as otherwise the queue would be unstable. Further, if $1 > \rho\phi_k$ then $\pi(0)$ is the most likely state, and the system will tend to spend all the time in the highest price regime. So we assume $\rho\phi_k > 1$. We are mainly interested in finding the pair of prices (p_{j^*-1}, p_{j^*}) such that

$$\rho\phi_{j^*-1} < 1 < \rho\phi_{j^*}. \tag{17}$$

The reason is that, in order to see how π behaves as a function of the state, it is useful to compute the ratio among consecutive states. Formally, if $\theta_j < i \le \theta_{j-1}$ —think of $\theta_k = 0$ and $\theta_0 = \infty$—, then

$$\frac{\pi(i)}{\pi(i-1)} = \rho\phi_j = \begin{cases} > 1, & j \ge j^*,\\ < 1, & j < j^*. \end{cases} \tag{18}$$

We conclude that π reaches its maximum value at $\pi(\theta_{j^*-1})$ and it exhibits an exponential decay of mass for states greater and smaller than θ_{j^*-1}. In particular,

$$\pi(\theta_{j^*-1}) = \rho^{\theta_{j^*-1}} \left[\prod_{s=j^*+1}^{k} \phi_s^{m(s-1)}\right] \phi_{j^*}^{\theta_{j^*-1}-\theta_{j^*}}\, \pi(0) = \rho^{\theta_{j^*-1}} \left[\prod_{s=j^*}^{k} \phi_s^{m(s-1)}\right] \pi(0).$$

Let us compute the normalizing constant $\pi(0)$. In this case, we need to do some laborious work:

$$
\begin{aligned}
\pi(0)^{-1} &= \pi(0)^{-1}\sum_i \pi(i) = \pi(0)^{-1}\left(\sum_{i=0}^{\theta_{k-1}}\pi(i) + \sum_{j=2}^{k-1}\sum_{i=\theta_j+1}^{\theta_{j-1}}\pi(i) + \sum_{i=\theta_1+1}^{\infty}\pi(i)\right)\\
&= \sum_{i=0}^{\theta_{k-1}}(\rho\phi_k)^i + \sum_{j=2}^{k-1}\sum_{i=\theta_j+1}^{\theta_{j-1}}\rho^i\left[\prod_{s=j+1}^{k}\phi_s^{m(s-1)}\right]\phi_j^{i-\theta_j} + \sum_{i=\theta_1+1}^{\infty}\rho^i\left[\prod_{s=2}^{k}\phi_s^{m(s-1)}\right]\phi_1^{i-\theta_1}\\
&= \frac{1-(\rho\phi_k)^{\theta_{k-1}+1}}{1-\rho\phi_k} + \sum_{j=2}^{k-1}\left[\prod_{s=j+1}^{k}\phi_s^{m(s-1)}\right]\phi_j^{-\theta_j}\sum_{i=\theta_j+1}^{\theta_{j-1}}(\rho\phi_j)^i + \left[\prod_{s=2}^{k}\phi_s^{m(s-1)}\right]\phi_1^{-\theta_1}\sum_{i=\theta_1+1}^{\infty}(\rho\phi_1)^i\\
&= \frac{1-(\rho\phi_k)^{\theta_{k-1}+1}}{1-\rho\phi_k} + \sum_{j=2}^{k-1}\left[\prod_{s=j+1}^{k}\phi_s^{m(s-1)}\right]\phi_j^{-\theta_j}(\rho\phi_j)^{\theta_j+1}\frac{1-(\rho\phi_j)^{\theta_{j-1}-\theta_j}}{1-\rho\phi_j} + \left[\prod_{s=2}^{k}\phi_s^{m(s-1)}\right]\phi_1^{-\theta_1}\frac{(\rho\phi_1)^{\theta_1+1}}{1-\rho\phi_1}\\
&= \frac{1-(\rho\phi_k)^{\theta_{k-1}+1}}{1-\rho\phi_k} + \sum_{j=2}^{k-1}\left[\prod_{s=j+1}^{k}\phi_s^{m(s-1)}\right]\rho^{\theta_j}(\rho\phi_j)\frac{1-(\rho\phi_j)^{\theta_{j-1}-\theta_j}}{1-\rho\phi_j} + \left[\prod_{s=2}^{k}\phi_s^{m(s-1)}\right]\frac{\rho^{\theta_1}(\rho\phi_1)}{1-\rho\phi_1}\\
&= \frac{(\rho\phi_k)^{\theta_{k-1}+1}-1}{\rho\phi_k-1} + \sum_{j=2}^{j^*-1}w(j)\rho^{\theta_j}(\rho\phi_j)\frac{1-(\rho\phi_j)^{\theta_{j-1}-\theta_j}}{1-\rho\phi_j} + \sum_{j=j^*}^{k-1}w(j)\rho^{\theta_j}(\rho\phi_j)\frac{(\rho\phi_j)^{\theta_{j-1}-\theta_j}-1}{\rho\phi_j-1} + w(1)\frac{\rho^{\theta_1}(\rho\phi_1)}{1-\rho\phi_1}.
\end{aligned}
$$

where we defined $w(j)=\prod_{s=j+1}^{k}\phi_s^{m(s-1)}$. As $\phi_s>1$ for all s, we see that $w(j)$ is decreasing in j. Further,

$$
\frac{w(j)}{w(j+t)} = \frac{\prod_{s=j+1}^{k}\phi_s^{m(s-1)}}{\prod_{s=j+t+1}^{k}\phi_s^{m(s-1)}} = \prod_{s=j+1}^{j+t}\phi_s^{m(s-1)}.
$$

Note that $\theta_j=\sum_{t=j}^{k-1}m(t)$. It follows that $w(j)\rho^{\theta_j}=\prod_{s=j+1}^{k}(\rho\phi_s)^{m(s-1)}$. We are interested in exploring the asymptotic behavior of $\pi(\theta_{j^*-1})=w(j^*-1)\ \rho^{\theta_{j^*-1}}\ \pi(0)=1/(\pi(0)^{-1}/w(j^*-1)\rho^{\theta_{j^*-1}})$, and to do so we consider the terms of $\pi(0)^{-1}$ separately, according to whether they correspond to $j<j^*$ or $j\geq j^*$.

Firstly, for terms $j<j^*$, we see that, as $m(j-1,n)=\theta_{j-1}(n)-\theta_j(n)\to\infty$ when $n\to\infty$,

$$
\begin{aligned}
&\frac{1}{w(j^*-1)\ \rho^{\theta_{j^*-1}}}\left(\sum_{j=2}^{j^*-1}w(j)\rho^{\theta_j}(\rho\phi_j)\frac{1-(\rho\phi_j)^{\theta_{j-1}-\theta_j}}{1-\rho\phi_j} + w(1)\frac{\rho^{\theta_1}(\rho\phi_1)}{1-\rho\phi_1}\right)\\
&\qquad\sim \frac{1}{w(j^*-1)\ \rho^{\theta_{j^*-1}}}\sum_{j=1}^{j^*-1}w(j)\rho^{\theta_j}\frac{\rho\phi_j}{1-\rho\phi_j}\\
&\qquad= \sum_{j=1}^{j^*-1}\frac{\prod_{s=j+1}^{k}(\rho\phi_s)^{m(s-1)}}{\prod_{s=j^*}^{k}(\rho\phi_s)^{m(s-1)}}\frac{\rho\phi_j}{1-\rho\phi_j}\\
&\qquad= \frac{\rho\phi_{j^*-1}}{1-\rho\phi_{j^*-1}} + \sum_{j=1}^{j^*-2}\prod_{s=j+1}^{j^*-1}(\rho\phi_s)^{m(s-1)}\frac{\rho\phi_j}{1-\rho\phi_j} \quad\to\quad \frac{\rho\phi_{j^*-1}}{1-\rho\phi_{j^*-1}},
\end{aligned}
$$

when $n\to\infty$, as $\rho\phi_s<1$ when $s<j^*$.

On the other hand, when $j\geq j^*$,

$$
\frac{1}{w(j^*-1)\ \rho^{\theta_{j^*-1}}}\left(\sum_{j=j^*}^{k-1}w(j)\rho^{\theta_j}(\rho\phi_j)\frac{(\rho\phi_j)^{m(j-1)}-1}{\rho\phi_j-1} + \frac{(\rho\phi_k)^{\theta_{k-1}+1}-1}{\rho\phi_k-1}\right)
$$

$$
\begin{aligned}
&= \sum_{j=j^*}^{k-1} \frac{w(j)\rho^{\theta_j}}{w(j^*-1)\ \rho^{\theta_{j^*-1}}}(\rho\phi_j)\frac{(\rho\phi_j)^{m(j-1)}-1}{\rho\phi_j-1} + \frac{(\rho\phi_k)^{m(k-1)+1}-1}{w(j^*-1)\ \rho^{\theta_{j^*-1}}(\rho\phi_k-1)} \\
&= \sum_{j=j^*}^{k-1} \frac{\prod_{s=j+1}^{k}(\rho\phi_s)^{m(s-1)}}{\prod_{s=j^*}^{k}(\rho\phi_s)^{m(s-1)}}(\rho\phi_j)\frac{(\rho\phi_j)^{m(j-1)}-1}{\rho\phi_j-1} + \frac{(\rho\phi_k)^{m(k-1)+1}-1}{\prod_{s=j^*}^{k}(\rho\phi_s)^{m(s-1)}(\rho\phi_k-1)} \\
&= \sum_{j=j^*}^{k-1} \frac{(\rho\phi_j)^{m(j-1)}-1}{\prod_{s=j^*}^{j}(\rho\phi_s)^{m(s-1)}}\frac{\rho\phi_j}{\rho\phi_j-1} + \frac{\rho\phi_k}{\rho\phi_k-1}\frac{(\rho\phi_k)^{m(k-1)}-1/(\rho\phi_k)}{\prod_{s=j^*}^{k}(\rho\phi_s)^{m(s-1)}} \\
&= \left(1-\frac{1}{(\rho\phi_{j^*})^{m(j^*-1)}}\right)\frac{\rho\phi_{j^*}}{\rho\phi_{j^*}-1} + \sum_{j=j^*+1}^{k-1} \frac{(\rho\phi_j)^{m(j-1)}-1}{\prod_{s=j^*}^{j}(\rho\phi_s)^{m(s-1)}}\frac{\rho\phi_j}{\rho\phi_j-1} + \frac{\rho\phi_k}{\rho\phi_k-1}\frac{(\rho\phi_k)^{m(k-1)}-1/(\rho\phi_k)}{\prod_{s=j^*}^{k}(\rho\phi_s)^{m(s-1)}} \\
&\to \frac{\rho\phi_{j^*}}{\rho\phi_{j^*}-1},
\end{aligned}
$$

when $n\to\infty$, as $\rho\phi_s>1$ when $s\geq j^*$. It follows that as $n\to\infty$

$$
\pi(\theta_{j^*-1}) \to \frac{1}{\frac{\rho\phi_{j^*-1}}{1-\rho\phi_{j^*-1}}+\frac{\rho\phi_{j^*}}{\rho\phi_{j^*}-1}} > 0. \tag{19}
$$

Our goal is to show that when n grows the measure π strongly concentrates around $\theta_{j^*-1}=\theta_{j^*-1}(n)$. Moreover, π does concentrate around states on which passengers are charged either p_{j^*-1} or p_{j^*}. In order to prove such result, we will consider intervals of the form $[\theta_{j^*-1}-c,\theta_{j^*-1}+c]$ for some suitable c. Let us choose any positive integer function $c=c(n)$ such that $c(n)\to\infty$ as $n\to\infty$ and $c(n)=o(m(j^*-1))$ and $c(n)=o(m(j^*-2))$, that is, both pricing gaps grow faster. Actually, c could grow in an arbitrarily slow fashion.

Note the interval dependance on n. Then, by (18), we see that

$$
\pi([\theta_{j^*-1}-c,\theta_{j^*-1}]) = \pi(\theta_{j^*-1})\sum_{i=0}^{c}(\rho\phi_{j^*})^{-i} = \pi(\theta_{j^*-1})\frac{1-(\rho\phi_{j^*})^{-(c+1)}}{1-(\rho\phi_{j^*})^{-1}}. \tag{20}
$$

Similarly,

$$
\pi([\theta_{j^*-1},\theta_{j^*-1}+c]) = \pi(\theta_{j^*-1})\sum_{i=0}^{c}(\rho\phi_{j^*-1})^{i} = \pi(\theta_{j^*-1})\frac{1-(\rho\phi_{j^*-1})^{c+1}}{1-\rho\phi_{j^*-1}}.
$$

Taking into account the double counting of θ_{j^*-1}, it follows that the mass of $[\theta_{j^*-1}-c,\theta_{j^*-1}+c]$ is

$$
\begin{aligned}
\pi([\theta_{j^*-1}-c,\theta_{j^*-1}+c]) &= \pi(\theta_{j^*-1})\left[\frac{1-(\rho\phi_{j^*})^{-(c+1)}}{1-(\rho\phi_{j^*})^{-1}}+\frac{1-(\rho\phi_{j^*-1})^{c+1}}{1-\rho\phi_{j^*-1}}-1\right] \\
&= \pi(\theta_{j^*-1})\left[\frac{\rho\phi_{j^*}-(\rho\phi_{j^*})^{-c}}{\rho\phi_{j^*}-1}+\frac{\rho\phi_{j^*-1}-(\rho\phi_{j^*-1})^{c+1}}{1-\rho\phi_{j^*-1}}\right].
\end{aligned} \tag{21}
$$

We conclude by (17), (19) and (21) that as $n\to\infty$

$$
\pi([\theta_{j^*-1}-c,\theta_{j^*-1}+c]) \to 1.
$$

Now, let us consider the first statement of the theorem. We claim that

$$
p_{bal} \in [p_{j^*-1}, p_{j^*}].
$$

Recall that by the second part of the theorem, we know that there exists a unique $j^* \in \{2, \dots, k\}$ such that for any sequence $c = c(n) \to \infty$ as $n \to \infty$, we have that

$$\pi([\theta_{j^*-1} - c, \theta_{j^*-1} + c]) \to 1, \qquad \text{as } n \to \infty.$$

Further, j^* satisfies $\rho\phi_{j^*-1} < 1 < \rho\phi_{j^*}$. By taking $c(n) = o(m(j^*-1))$ and $c(n) = o(m(j^*))$, we conclude that the system is always around θ_{j^*-1} and only two prices (p_{j^*-1}, p_{j^*}) are charged to passengers.

Now we consider two *different* systems, while both share $\Lambda_0(n)$ and $\mu_0(n)$. In the first one, say system A, we apply the k-price policy $\pi_A = (\underline{\theta}, \underline{\mathbf{p}})$. System B uses a two-price policy, namely $\pi_B = (p_{j^*-1}, p_{j^*}, \theta_{j^*-1})$. Note that the Markov Chains induced by π_A and π_B are not equal, however, we claim that in the limit $(n \to \infty)$ their behavior is identical in a way that we now formally describe.

We claim that

$$\lambda_A = \lim_{n\to\infty} \lambda_A(n) = \lim_{n\to\infty} \lambda_B(n) = \lambda_B. \tag{22}$$

By definition, λ_A and λ_B are the solutions to the following fixed-point equations:

$$\lambda_A = \Lambda_0 \, \mathbf{F}_C \left(\frac{\mathbb{E}_{\lambda_A}[R]}{\tau} \right), \qquad \lambda_B = \Lambda_0 \, \mathbf{F}_C \left(\frac{\mathbb{E}_{\lambda_B}[R]}{\tau} \right).$$

Let us denote by $p_\pi(i)$ the price at state i when using policy π. We also define $\pi_n(p)$ to be the sum of π over all states with price p for fixed n. Then, $\pi(p) = \lim_{n\to\infty} \pi_n(p)$. For system A, we have that

$$\begin{aligned}\lambda_A &= \Lambda_0 \, \mathbf{F}_C \left(\frac{\mathbb{E}_{\lambda_A}[R]}{\tau} \right) = \Lambda_0 \, \mathbf{F}_C \left(\frac{\sum_{i=0}^{\infty} p_{\pi_A}(i)\pi_A(i)}{\tau} \right) \\ &= \Lambda_0 \, \mathbf{F}_C \left(\frac{\sum_{i=1}^{k} p_i \, \pi_A(p_i)}{\tau} \right) = \Lambda_0 \, \mathbf{F}_C \left(\frac{p_{j^*-1} \, \pi_A(p_{j^*-1}) + p_{j^*} \, \pi_A(p_{j^*})}{\tau} \right),\end{aligned} \tag{23}$$

as for $i \notin \{j^*-1, j^*\}$ we have that $\pi_A(i) = 0$. On the other hand, system B only uses two prices.

It follows that

$$\lambda_B = \Lambda_0 \, \mathbf{F}_C \left(\frac{\mathbb{E}_{\lambda_B}[R]}{\tau} \right) = \Lambda_0 \, \mathbf{F}_C \left(\frac{p_{j^*-1} \, \pi_B(p_{j^*-1}) + p_{j^*} \, \pi_B(p_{j^*})}{\tau} \right). \tag{24}$$

We want to show that λ_A is, in fact, a fixed-point solution for equation (24). We do the following: we first consider system B assuming $\lambda = \lambda_A$, we compute $\pi_B(p_{j^*-1})$ and $\pi_B(p_{j^*})$ in such a system, and then we show that the right-hand side of (24) actually equals λ_A. In order to complete the last step, we will use equation (23) —note that under system A. Finally, by uniqueness it follows that $\lambda_A = \lambda_B$.

The key point of using λ_A for system B is that we actually know that $\rho\phi_{j^*-1} < 1 < \rho\phi_{j^*}$, where $\rho = \lambda_A/\mu_0$, and that also holds when applied to system B. Recall that $\pi_B = (p_{j^*-1}, p_{j^*}, \theta_{j^*-1})$.

Therefore, the steady-state distribution is given by

$$\begin{aligned}\pi_B(i) &= (\rho\phi_{j^*})^i \, \pi_B(0), && \text{for } i \le \theta_{j^*-1}, \\ \pi_B(i) &= \rho^i \phi_{j^*-1}^{\theta_{j^*-1}} \phi_{j^*}^{i-\theta_{j^*-1}} \, \pi_B(0), && \text{for } i > \theta_{j^*-1}.\end{aligned}$$

Let us find $\pi_B(0)$,

$$1 = \sum_{i=0}^{\infty} \pi_B(i) = \pi_B(0) \sum_{i=0}^{\theta_{j^*-1}} (\rho\phi_{j^*})^i + \pi_B(0) \sum_{i=\theta_{j^*-1}+1}^{\infty} \rho^i \phi_{j^*-1}^{\theta_{j^*-1}} \phi_{j^*}^{i-\theta_{j^*-1}}$$
$$= \pi_B(0) \left[\frac{(\rho\phi_{j^*})^{\theta_{j^*-1}+1} - 1}{\rho\phi_{j^*} - 1} + \frac{(\rho\phi_{j^*-1})(\rho\phi_{j^*})^{\theta_{j^*-1}}}{1 - \rho\phi_{j^*-1}} \right].$$

Hence,

$$\pi_B(0) = \left[\frac{(\rho\phi_{j^*})^{\theta_{j^*-1}+1} - 1}{\rho\phi_{j^*} - 1} + \frac{(\rho\phi_{j^*-1})(\rho\phi_{j^*})^{\theta_{j^*-1}}}{1 - \rho\phi_{j^*-1}} \right]^{-1}.$$

We conclude that

$$\pi_B(p_{j^*-1}) = \lim_{n\to\infty} \frac{\frac{(\rho\phi_{j^*-1})(\rho\phi_{j^*})^{\theta_{j^*-1}}}{1-\rho\phi_{j^*-1}}}{\frac{(\rho\phi_{j^*})^{\theta_{j^*-1}+1}-1}{\rho\phi_{j^*}-1} + \frac{(\rho\phi_{j^*-1})(\rho\phi_{j^*})^{\theta_{j^*-1}}}{1-\rho\phi_{j^*-1}}}$$
$$= \lim_{n\to\infty} \frac{\rho\phi_{j^*-1}}{(1-\rho\phi_{j^*-1})\frac{\rho\phi_{j^*} - (\rho\phi_{j^*})^{-\theta_{j^*-1}}}{\rho\phi_{j^*}-1} + \rho\phi_{j^*-1}}$$
$$= \frac{\phi_{j^*-1}}{\frac{1-\rho\phi_{j^*-1}}{\rho\phi_{j^*}-1}\phi_{j^*} + \phi_{j^*-1}}.$$

Similarly, we have that

$$\pi_B(p_{j^*}) = \frac{\frac{1-\rho\phi_{j^*-1}}{\rho\phi_{j^*}-1}\phi_{j^*}}{\frac{1-\rho\phi_{j^*-1}}{\rho\phi_{j^*}-1}\phi_{j^*} + \phi_{j^*-1}}.$$

Now, let us go back to System A. By equation (19), we know that

$$\pi_A(\theta_{j^*-1}) \to \frac{1}{\frac{\rho\phi_{j^*-1}}{1-\rho\phi_{j^*-1}} + \frac{\rho\phi_{j^*}}{\rho\phi_{j^*}-1}} > 0. \tag{25}$$

If we take $c(n) = m(j^*-1) = \theta_{j^*}(n) - \theta_{j^*-1}(n)$ in (10), we conclude that

$$\pi_A(p_{j^*-1}) = \lim_{n\to\infty} \pi_A(\theta_{j^*-1}) \sum_{i=1}^{c(n)} (\rho\phi_{j^*-1})^i$$
$$= \lim_{n\to\infty} \pi_A(\theta_{j^*-1}) \left(\frac{1 - (\rho\phi_{j^*-1})^{c(n)+1}}{1-\rho\phi_{j^*-1}} - 1 \right)$$
$$= \lim_{n\to\infty} \pi_A(\theta_{j^*-1}) \frac{\rho\phi_{j^*-1} - (\rho\phi_{j^*-1})^{c(n)+1}}{1-\rho\phi_{j^*-1}}$$
$$= \frac{\frac{\rho\phi_{j^*-1}}{1-\rho\phi_{j^*-1}}}{\frac{\rho\phi_{j^*-1}}{1-\rho\phi_{j^*-1}} + \frac{\rho\phi_{j^*}}{\rho\phi_{j^*}-1}}$$
$$= \frac{\phi_{j^*-1}}{\phi_{j^*-1} + \frac{1-\rho\phi_{j^*-1}}{\rho\phi_{j^*}-1}\phi_{j^*}} = \pi_B(p_{j^*-1}).$$

Similarly, we have that

$$\pi_A(p_{j^*}) = 1 - \pi_A(p_{j^*-1}) = \frac{\frac{1-\rho\phi_{j^*-1}}{\rho\phi_{j^*-1}}\phi_{j^*}}{\phi_{j^*-1} + \frac{1-\rho\phi_{j^*-1}}{\rho\phi_{j^*-1}}\phi_{j^*}} = \pi_B(p_{j^*}).$$

Therefore, it directly follows that λ_A satisfies the fixed-point equation for system B:

$$\begin{aligned}\lambda_B &= \Lambda_0\, \mathbf{F}_C\left(\frac{p_{j^*-1}\ \pi_B(p_{j^*-1}) + p_{j^*}\ \pi_B(p_{j^*})}{\tau}\right)\\ &= \Lambda_0\, \mathbf{F}_C\left(\frac{p_{j^*-1}\ \pi_A(p_{j^*-1}) + p_{j^*}\ \pi_A(p_{j^*})}{\tau}\right) = \lambda_A.\end{aligned}$$

Note that Markov Chains A and B are not equal. However, informally speaking, their behavior in equilibrium is effectively identical: they always remain in regions where prices are p_{j^*-1} and p_{j^*}, and they do so with equal probability in both systems.

Recall that in system B, we have that $\rho\phi_{j^*-1} < 1 < \rho\phi_{j^*}$. Suppose now that $p_{bal} \notin [p_{j^*-1}, p_{j^*}]$. Then there are two possible cases, either $p_{bal} > p_{j^*} > p_{j^*-1}$ or $p_{bal} < p_{j^*-1} < p_{j^*}$.

By Theorem 7, in the first case, System B would lead to $\lambda_B = \Lambda_0\, \mathbf{F}_C(p_{j^*}/\tau)$, which directly implies that $\mathbb{E}_B[R] = p_{j^*}$. However, we know that the expected price —in equilibrium— of System B is precisely

$$\mathbb{E}_B[R] = p_{j^*-1}\ \pi_B(p_{j^*-1}) + p_{j^*}\ \pi_B(p_{j^*}) < p_{j^*}.$$

In the second case, again by Theorem 7, system B would lead to $\lambda_B = \mu_0 \bar{\mathbf{F}}_V(p_{j^*-1}) = \mu_0/\phi_{j^*-1}$. But then,

$$\rho\phi_{j^*-1} = \frac{\lambda_B}{\mu_0}\phi_{j^*-1} = \frac{\mu_0/\phi_{j^*-1}}{\mu_0}\phi_{j^*-1} = 1,$$

which is a contradiction with the fact that $\rho\phi_{j^*-1} < 1 < \rho\phi_{j^*}$ for system B. It follows that $p_{bal} \in [p_{j^*-1}, p_{j^*}]$.

11. Results and proofs for Networks of Ride-share Queues

In this appendix, we provide complete proofs for the results in Section 6. We start with the proof for Theorem 15:

Theorem 15 *The static localized policy* $\mathbf{p}_{bal}$ *maximizing* $\lim_{n\to\infty} \Lambda(n)/n$ *satisfies*

$$\Lambda_0\, \mathbf{F}_C\left(\frac{\sigma^T}{\tau}\sum_{k=0}^{\infty} W^k\, \mathbf{p}\right) = \min_i \beta_i(p_i) = \beta_j(p_j), \qquad \textit{for all } j \in G, \tag{26}$$

where $\beta_i(p_i) = \frac{\mu_0(i)}{(\sigma^T B)_i}\, \bar{\mathbf{F}}_V(p_i)$ *and* $W = TQ$.

[of Theorem 15]

For ease of notation, in this proof we use λ to denote the normalized large-market limit $\lambda^e = \lim_{n\to\infty} \lambda^e(n)/n$, where $\lambda^e(n)$ corresponds to the vector of rates of incoming drivers to the $n-$th system, with $\Lambda_0(n) = n\Lambda_0$ and $\mu(n) = n\mu_0$. Similarly, we define $\Lambda^e = \lim_{n\to\infty} \Lambda(n)/n$, where $\Lambda(n)$ is the total rate of incoming drivers to the $n-$th system. We add subscripts and parameters to λ and Λ when several different systems are being considered.

Suppose that we fix some *static* localized prices $(p_i)_{i\in G}$. Then, the rates at which passengers arrive at each region i are totally determined:

$$\mu(i) = \mu_0(i)\ \bar{\mathbf{F}}_V(p_i), \qquad i \in G.$$

For simplicity, we assume that the distribution of V does not change from one region to another. One could think of users from different neighborhoods having distinct utility functions, or being willing to pay different amounts of money for the same ride. However, as long as distributions V_i satisfy the same conditions as V does, the analysis could be easily extended.

Let $\lambda^e \in \mathbb{R}^n$ be the rate at which *new* drivers arrive at each region, and let $\Lambda^e \in \mathbb{R}$ be the total rate at which drivers enter the system. We have that $\lambda^e = \Lambda^e \sigma$. Further, we know that the *effective* rate of available drivers is given by —see appendix 12—

$$\lambda = \lambda^{eT}(I - QT)^{-1} = \lambda^{eT} B,$$

where we defined $B := (I - QT)^{-1} = \sum_{k=0}^{\infty}(QT)^k$. Note that $b_{ij} \geq 0$.

We would like to find the value of λ as a function of the pricing policy $\mathbf{p} = (p_i)_{i\in G}$. Firstly, we need to satisfy the stability conditions at *each* queue $i \in G$: $\lambda(i) \leq \mu(i)$. Equivalently, for all $i \in G$, we need

$$(\lambda^{eT} B)_i = \Lambda^e(\sigma^T B)_i \leq \mu_0(i)\bar{\mathbf{F}}_V(p_i),$$

which implies that

$$\Lambda^e \leq \min_{i\in G} \frac{\mu_0(i)}{(\sigma^T B)_i}\ \bar{\mathbf{F}}_V(p_i) = \min_{i\in G} \beta_i(p_i),$$

where we defined $\beta_i(p_i) = \frac{\mu_0(i)}{(\sigma^T B)_i}\ \bar{\mathbf{F}}_V(p_i)$. Further, we also define $C_i = \frac{\mu_0(i)}{(\sigma^T B)_i}$, an exogenous constant —as long as no control is introduced on Q. So, we have that $\beta_i(p_i) = C_i\ \bar{\mathbf{F}}_V(p_i)$.

On the other hand, each potential driver will sample his reservation price C so that, *if* the system is supply constrained, Λ will be the fixed-point solution to

$$\Lambda^e = \Lambda_0\ \mathbf{F}_C\left(\frac{\mathbb{E}_{\mathbf{p},\Lambda}[R]}{\mathbb{E}_{\mathbf{p},\Lambda}[I] + \tau}\right),$$

where τ is the expected time taken by a ride and R is the revenue per ride.

Let us compute $\mathbb{E}_{\mathbf{p},\Lambda}[R]$ now. Denote by w_{ij} the probability that a driver starts a ride from j given that he *started* his last ride from i. We see that $w_{ij} = \sum_{k\in G} t_{ik} q_{kj}$, and it follows that $W = TQ$. Consider the Markov Chain M over G with transition matrix W and initial distribution σ. We add a new state s_{m+1} which represents when a driver has already exited the system. Note that for any $i \in G$, we have $w_{i,m+1} = \sum_{k\in G} t_{ik} q^k_{\mathsf{exit}} > 0$ and $w_{m+1,i} = 0$. Finally, $w_{m+1,m+1} = 1$. Assume we add the new row and column to W.

If prices p_i are fixed, then a random walk on M will determine the amount of money earned by a new driver —more formally, the amount of money earned by a driver follows the same distribution as the total reward on M of a random walk. Then,

$$\mathbb{E}_{\mathbf{p},\Lambda}[R] = \sigma^T \sum_{k=0}^{\infty} W^k\ \mathbf{p}.$$

When scaling the system with n, if we meet the stability conditions, we conclude that $\mathbb{E}[I] \to 0$ —note that $\mathbb{E}_{\mathbf{p},\Lambda}[I]$ can be computed in a similar fashion to the way we computed $\mathbb{E}_{\mathbf{p},\Lambda}[R]$. Therefore, we find that Λ in the supply constrained setting satisfies

$$\Lambda^e = \Lambda_0\ \mathbf{F}_C \left(\frac{\sigma^T}{\tau} \sum_{k=0}^{\infty} W^k\ \mathbf{p} \right).$$

It follows that

$$\Lambda^e = \min \left(\min_i \beta_i(p_i), \Lambda_0\ \mathbf{F}_C \left(\frac{\sigma^T}{\tau} \sum_{k=0}^{\infty} W^k\ \mathbf{p} \right) \right).$$

Let us find $\mathbf{p}$ maximizing $\Lambda^e = \Lambda(\mathbf{p})$. Note that $\beta_i(p_i)$ is a decreasing function of p_i. On the other hand, $\Lambda_0\ \mathbf{F}_C \left(\frac{\sigma^T}{\tau} \sum_{k=0}^{\infty} W^k\ \mathbf{p} \right)$ is increasing in p_i. Let $j = \arg\min_i \beta_i(p_i)$, and suppose there is some i such that $\beta_j(p_j) < \beta_i(p_i)$. Clearly, by increasing p_i, Λ will not decrease, but it could increase in case $\Lambda_0\ \mathbf{F}_C \left(\frac{\sigma^T}{\tau} \sum_{k=0}^{\infty} W^k\ \mathbf{p} \right) < \beta_j(p_j)$. It follows that for each vector $\mathbf{p}$, there exists $\mathbf{p}^*$ such that $\Lambda(\mathbf{p}) \le \Lambda(\mathbf{p}^*)$ and $\beta_j(p_j^*) = \beta_i(p_i^*)$ for all $i, j \in G$, under the assumption that prices p_i can take any value and $\mathbf{F}_V, \mathbf{F}_C$ are well-behaved (strictly monotone, for example).

If $\Lambda_0\ \mathbf{F}_C \left(\frac{\sigma^T}{\tau} \sum_{k=0}^{\infty} W^k\ \mathbf{p} \right) < \min_i \beta_i(p_i)$, by increasing $\mathbf{p}$ we will increase Λ. In this case, the system was supply constrained. Also, if $\min_i \beta_i(p_i) < \Lambda_0\ \mathbf{F}_C \left(\frac{\sigma^T}{\tau} \sum_{k=0}^{\infty} W^k\ \mathbf{p} \right)$, then by decreasing $\mathbf{p}$ we will increase Λ, as the system was demand constrained. We conclude that there exists $\mathbf{p} \in \arg\max_p \Lambda(p)$ such that

$$\Lambda_0\ \mathbf{F}_C \left(\frac{\sigma^T}{\tau} \sum_{k=0}^{\infty} W^k\ \mathbf{p} \right) = \min_i \beta_i(p_i) = \beta_j(p_j), \qquad \text{for all } j \in G. \tag{27}$$

By fixing p_1, we write all other prices p_i as a function of p_1 as follows ($C_i > 0$ for all i):

$$\beta_i(p_i) = \beta_1(p_1), \qquad C_i \bar{\mathbf{F}}_V(p_i) = C_1 \bar{\mathbf{F}}_V(p_1), \qquad p_i = \bar{\mathbf{F}}_V^{-1} \left(\frac{C_1}{C_i}\ \bar{\mathbf{F}}_V(p_1) \right), \tag{28}$$

and $p_1 \in \mathbb{R}^+$ satisfies

$$\beta_1(p_1) = C_1 \bar{\mathbf{F}}_V(p_1) = \Lambda_0\ \mathbf{F}_C(Z\ \mathbf{p}),$$

where $Z = \frac{\sigma^T}{\tau} \sum_{k=0}^{\infty} W^k\ \mathbf{p}$ and $p_i = \bar{F}_V^{-1} \left(\frac{C_1}{C_i}\ \bar{\mathbf{F}}_V(p_1) \right)$.

With respect to revenue maximization, we proved that each region behaves as an *independent* queue in equilibrium, hence, for a static policy $\mathbf{p} = (p_i)_{i \in G}$,

$$\mathbb{E}[R] = \sum_{i \in G} \mathbb{E}[R(i)] = \sum_{i \in G} p_i \lambda(i) = \Lambda^e\ \sum_{i \in G} p_i (\sigma^T B)_i,$$

where we defined $B := (I - QT)^{-1} = \sum_{k=0}^{\infty}(QT)^k$ (see Appendix 12).

In order to compare two policies, we say $\mathbf{p}_1 > \mathbf{p}_2$ if $\mathbf{p}_1(i) \geq \mathbf{p}_2(i)$ for all $i \in G$ and the inequality is strict at least for one region. Then we have the following result,

Theorem 16 *The optimal localized static pricing policy* $\mathbf{p}$ *with respect to revenue in the large-market limit is given by*

$$\mathbf{p} = \max(\mathbf{p}_{bal}, \mathbf{p}_{d-opt}),$$

where $\mathbf{p}_{bal}$ *satisfies* (26) *and* $\mathbf{p}_{d-opt}$ *is the maximum policy satisfying for every* $i \in G$

$$C_i\, \bar{\mathbf{F}}_V(p_i) = \frac{\sum_{j\in G} p_j(\sigma^T B)_j}{\sum_{j\in G} \frac{(\sigma^T B)_j}{C_j f_V(p_j)}}.$$

Proof of Theorem 16 For ease of notation, in this proof we use λ to denote the normalized large-market limit $\lambda^e = \lim_{n\to\infty} \lambda^e(n)/n$, where $\lambda^e(n)$ corresponds to the vector of rates of incoming drivers to the $n-$th system, with $\Lambda_0(n) = n\Lambda_0$ and $\mu(n) = n\mu_0$. Similarly, we define $\Lambda^e = \lim_{n\to\infty} \Lambda(n)/n$, where $\Lambda(n)$ is the total rate of incoming drivers to the $n-$th system. We add subscripts and parameters to λ and Λ when several different systems are being considered.

As $\sum_{i\in G} p_i(\sigma^T B)_i$ is non-decreasing in every p_i —note that $(\sigma^T B)_i \geq 0$— we directly conclude that we can restrict our search of optimal prices $\mathbf{p}$ to those satisfying

$$\beta_i(p_i) = \beta_j(p_j), \qquad \text{for all } i, j \in G. \tag{29}$$

Let $\mathbf{p} = \arg\max_{\mathbf{p}} \mathbb{E}_{\mathbf{p}}[R] = \arg\max_{\mathbf{p}} \Lambda^e\ \sum_{i\in G} p_i(\sigma^T B)_i$. We take the maximum over the set of policies satisfying (29). Suppose $\Lambda^e = \Lambda_0\ \mathbf{F}_C(Z\ \mathbf{p}) < \min_{i\in G} \beta_i(p_i)$, then by increasing the prices, both Λ and $\mathbb{E}_{\mathbf{p}}[R]$ will increase. So it follows that $\Lambda_0\ \mathbf{F}_C(R\ \mathbf{p}) \geq \min_{i\in G} \beta_i(p_i) = \Lambda$. If they are equal, we recover the balance prices $\mathbf{p}_{bal}$.

On the other hand, we could have $\Lambda_0\ \mathbf{F}_C(Z\ \mathbf{p}) > \min_{i\in G} \beta_i(p_i) = \Lambda$, with $\mathbf{p}$ still leading to the maximum expected revenue $\max_{\mathbf{p}} \mathbb{E}_{\mathbf{p}}[R]$. Let us denote this vector of prices —which may *not* be unique— by $\mathbf{p}_{d-opt}$. Note that, in such a case, we have $\mathbf{p}_{d-opt}(i) > \mathbf{p}_{bal}(i)$ for all i, as $\Lambda_{d-opt} < \Lambda_{bal}$ by Theorem 15, implying $\beta_i(\mathbf{p}_{d-opt}(i)) < \beta_i(\mathbf{p}_{bal}(i))$ by (29). Recall that β_i is strictly decreasing in p_i as long as $\mathbf{F}_V$ is strictly monotone, which is assumed.

In particular, for each p_i, by (29) we need

$$\begin{aligned}\frac{\partial \mathbb{E}[R]}{\partial p_i} &= \frac{\partial \Lambda^e}{\partial p_i}\sum_{j\in G} p_j(\sigma^T B)_j + \Lambda^e \sum_{j\in G}\frac{\partial p_j}{\partial p_i}(\sigma^T B)_j \\ &= \frac{\partial \beta_i(p_i)}{\partial p_i}\sum_{j\in G} p_j(\sigma^T B)_j + \beta_i(p_i)\sum_{j\in G}\frac{\partial \bar{F}_V^{-1}\left(\frac{C_i}{C_j}\ \bar{F}_V(p_i)\right)}{\partial p_i}(\sigma^T B)_j\end{aligned}$$

$$= C_i \frac{\partial \bar{\mathbf{F}}_V(p_i)}{\partial p_i} \sum_{j \in G} p_j (\sigma^T B)_j - C_i \, \bar{\mathbf{F}}_V(p_i) \sum_{j \in G} (\bar{F}_V^{-1})' \left(\frac{C_i}{C_j} \, \bar{F}_V(p_i) \right) \frac{C_i}{C_j} f_V(p_i) (\sigma^T B)_j$$
$$= -C_i f_V(p_i) \sum_{j \in G} p_j (\sigma^T B)_j + C_i \, \bar{\mathbf{F}}_V(p_i) \sum_{j \in G} \frac{1}{f_V(p_j)} \frac{C_i}{C_j} f_V(p_i) (\sigma^T B)_j.$$

Hence, $\frac{\partial \mathbb{E}[R]}{\partial p_i} = 0$ if and only if

$$\sum_{j \in G} p_j (\sigma^T B)_j = C_i \, \bar{\mathbf{F}}_V(p_i) \sum_{j \in G} \frac{(\sigma^T B)_j}{C_j f_V(p_j)}.$$

Or equivalently,

$$\Lambda^e = C_i \, \bar{\mathbf{F}}_V(p_i) = \frac{\sum_{j \in G} p_j (\sigma^T B)_j}{\sum_{j \in G} \frac{(\sigma^T B)_j}{C_j f_V(p_j)}}.$$

Note that when G has only one region, we recover the characterization we found in Section 3.

□

Theorem 17 *Let G be a network of regions and $\mathbf{p} = (p_\ell^i, p_h^i, \theta^i)_{i \in G}$ a localized dynamic pricing policy. Assume that F_V is common to all regions, and that it satisfies (12). Denote by $\mathbf{p}_b$ the balance localized static policy, which satisfies (26). Then, we have that in the large-market limit $\Lambda_d^e \leq \Lambda_b^e$, where Λ_d^e denotes the rate at which drivers sign-in into the system under the dynamic pricing policy.*

In particular, if $p_h^i = p_b^i$ for all $i \in G$, then $\Lambda_d^e = \Lambda_b^e$.

[of Theorem 17]

For ease of notation, in this proof we use λ to denote the normalized large-market limit $\lambda^e = \lim_{n \to \infty} \lambda^e(n)/n$, where $\lambda^e(n)$ corresponds to the vector of rates of incoming drivers to the $n-$th system, with $\Lambda_0(n) = n\Lambda_0$ and $\mu(n) = n\mu_0$. Similarly, we define $\Lambda^e = \lim_{n \to \infty} \Lambda(n)/n$, where $\Lambda(n)$ is the total rate of incoming drivers to the $n-$th system. We add subscripts and parameters to λ^e and Λ^e when several different systems are being considered.

The structure of the proof is actually very similar to that of Theorem 9. Let $\mathbf{p} = (p_\ell^i, p_h^i, \theta^i)_{i \in G}$ be any localized dynamic pricing policy. Once the policy is fixed, both $(\lambda_i)_{i \in G}$ and $(\lambda_i)_{i \in G}$ are determined. We also consider the static balance policy, which we denote by $\mathbf{p}_b$. Recall that $\mathbf{p}_b$ satisfies (26). We denote by Λ_d^e and Λ_s^e the values of Λ^e for $\mathbf{p}$ and $\mathbf{p}_b$ respectively.

For stability, $\mathbf{p}$ needs to satisfy at each location $\lambda_i \leq \mu_0(i) \bar{\mathbf{F}}_V(p_\ell^i)$. Equivalently, $({\lambda^e}^T B)_i = \Lambda_d^e (\sigma^T B)_i \leq \mu_0(i) \bar{\mathbf{F}}_V(p_\ell^i)$. As the above inequality holds at each $i \in G$, we conclude that

$$\Lambda_d^e \leq \min_{i \in G} \frac{\mu_0(i)}{(\sigma^T B)_i} \bar{\mathbf{F}}_V(p_\ell^i) = \min_{i \in G} \beta_i(p_\ell^i).$$

So, we either assume that $\Lambda_d^e \leq \Lambda_b^e$ —and the proof is complete— or we assume that $\Lambda_d^e > \Lambda_b^e$. Hence, suppose the latter. By (26), we need that

$$\beta_j(p_\ell^j) \geq \min_{i\in G} \beta_i(p_\ell^i) \geq \Lambda_d^e > \Lambda_b^e = \min_{i\in G} \beta_i(p_b^i) = \beta_j(p_b^j), \qquad \text{for all } j \in G. \tag{30}$$

We conclude that $p_\ell^j < p_b^j$ for every $j \in G$. In words, the low price at each region has to be lower than the balance price for the region. Let us compute the expected revenue R per ride under $\mathbf{p}$. Recall that, by independence of the queues, $\mathbb{E}_{\mathbf{p}}[R] = \mathbb{E}_{\mathbf{p}}\left[\sum_{i\in G} R_i\right] = \sum_{i\in G} \mathbb{E}_{\mathbf{p}}[R_i]$. As derived in Theorem 7, the expected revenue per ride at region $i \in G$ is given by

$$\begin{aligned} v_i := \mathbb{E}_{\mathbf{p}}[R(i)] &= p_\ell^i + \frac{p_h^i - p_\ell^i}{\phi_h^i - \phi_\ell^i}\left(\frac{\mu_0(i)}{\lambda_i} - \phi_\ell^i\right) \\ &= p_\ell^i + \frac{p_h^i - p_\ell^i}{\phi_h^i - \phi_\ell^i}\left(\frac{\mu_0(i)}{\Lambda_d^e(\sigma^T B)_i} - \phi_\ell^i\right) \\ &= p_\ell^i + \frac{p_h^i - p_\ell^i}{\phi_h^i - \phi_\ell^i}\left(\frac{C_i}{\Lambda_d^e} - \phi_\ell^i\right). \end{aligned} \tag{31}$$

Write $\mathbf{v} = (v_i)_{i\in G}$. Recall that we assumed $\Lambda_d^e > \Lambda_b^e = \Lambda_0 \mathbf{F}_C(Z\,\mathbf{p}_b)$, where $Z = \frac{\sigma^T}{\tau}\sum_{k=0}^\infty W^k = \frac{\sigma^T}{\tau} S$. We defined $S = (I - W)^{-1}$, and it follows that $s_{ij} \geq 0$ as $w_{ij} \geq 0$. In particular, we need $\Lambda_0 \mathbf{F}_C(Z\,\mathbf{v}) \geq \Lambda_d^e > \Lambda_0 \mathbf{F}_C(Z\,\mathbf{p}_b)$, which implies $\mathbf{F}_C(Z\,\mathbf{v}) > \mathbf{F}_C(Z\,\mathbf{p}_b)$. We denote by s_i the i-th row of S. Then,

$$\begin{aligned} Z\mathbf{v} = \frac{\sigma^T}{\tau} S\mathbf{v} = \frac{\sigma^T}{\tau}\begin{pmatrix} s_1^T\mathbf{v} \\ \vdots \\ s_m^T\mathbf{v}\end{pmatrix} &= \frac{1}{\tau}\sum_{i=1}^m \sigma_i s_i^T \mathbf{v} \\ &= \frac{1}{\tau}\sum_{i=1}^m \sigma_i \sum_{j=1}^m s_{ij} v_j = \frac{1}{\tau}\sum_{j=1}^m\left(\sum_{i=1}^m \sigma_i s_{ij}\right) v_j = \frac{1}{\tau}\sum_{j=1}^m \alpha_j v_j. \end{aligned} \tag{32}$$

Similarly, we have that $Z\mathbf{p}_b = \frac{1}{\tau}\sum_{j=1}^m \alpha_j p_b^j$. Note that $\alpha_j \geq 0$ for all j.

Assume that $p_h^i \leq p_b^i$ for all $i \in G$ with at least one j such that $p_h^j < p_b^j$. Then, we see that $v_i \leq p_b^i$ for all i, with strict inequality for some j:

$$v_i = p_\ell^i\left(\frac{\phi_h^i - C_i/\Lambda_d^e}{\phi_h^i - \phi_l^i}\right) + p_h^i\left(\frac{C_i/\Lambda_d^e - \phi_l^i}{\phi_h^i - \phi_l^i}\right) \leq p_h^i \leq p_b^i.$$

so that $Z\mathbf{v} < Z\mathbf{p}_b$ and it directly follows that $\Lambda_d^e < \Lambda_s^e$.

Now, we assume that for some $i \in G$ we have $p_b^i < p_h^i$. Let Y be the set of those indices and assume $Y \neq \emptyset$. Analogously to the proof of Theorem 9, we proceed as follows: we start by assuming $\Lambda_d^e = \Lambda_0 \mathbf{F}_C(Z\mathbf{v})$.

Recall that $\Lambda_d^e = \min(\Lambda_0 \mathbf{F}_C(Z\mathbf{v}), \min_{i\in G}\beta_i(p_\ell^i))$, but if $\Lambda_d^e = \min_{i\in G}\beta_i(p_\ell^i)$, then $\partial\Lambda_d^e/\partial p_h^i = 0$ for $i \in Y$ until we, eventually, have that $\Lambda_d^e = \Lambda_0\mathbf{F}_C(Z\mathbf{v})$. By taking $\mathbf{F}_C^{-1}$ on both sides of ${\Lambda^e}_d$, we see that

$$\mathbf{F}_C^{-1}\left(\frac{{\Lambda^e}_d}{\Lambda_0}\right) = Z\mathbf{v} = \frac{1}{\tau}\sum_{j=1}^m \alpha_j v_j.$$

Let $i \in Y$ and take the partial derivative of the above equation with respect to p_h^i

$$\begin{aligned}
\frac{\tau}{\Lambda_0}\frac{\partial {\Lambda^e}_d}{\partial p_h^i}\frac{\partial \mathbf{F}_C^{-1}}{\partial {\Lambda^e}_d}\left(\frac{\Lambda_d^e}{\Lambda_0}\right) &= \sum_{j\neq i}\alpha_j C_j\left(\frac{p_h^j - p_\ell^j}{\phi_h^j - \phi_\ell^j}\right)\frac{\partial}{\partial p_h^i}\left(\frac{1}{{\Lambda^e}_d}\right) + \alpha_i\frac{\partial}{\partial p_h^i}\left[\left(\frac{p_h^i - p_\ell^i}{\phi_h^i - \phi_\ell^i}\right)\left(\frac{C_i}{{\Lambda^e}_d} - \phi_\ell^i\right)\right] \\
&= \frac{\partial}{\partial p_h^i}\left(\frac{1}{{\Lambda^e}_d}\right)\sum_{j\neq i}\alpha_j C_j\left(\frac{p_h^j - p_\ell^j}{\phi_h^j - \phi_\ell^j}\right) + \alpha_i\left(\frac{C_i}{{\Lambda^e}_d} - \phi_\ell^i\right)\frac{\partial}{\partial p_h^i}\left(\frac{p_h^i - p_\ell^i}{\phi_h^i - \phi_\ell^i}\right) + \alpha_i C_i\left(\frac{p_h^i - p_\ell^i}{\phi_h^i - \phi_\ell^i}\right)\frac{\partial}{\partial p_h^i}\left(\frac{1}{{\Lambda^e}_d}\right) \\
&= -\frac{1}{{\Lambda^e}_d^2}\frac{\partial {\Lambda^e}_d}{\partial p_h^i}\sum_{j\in G}\alpha_j C_j\left(\frac{p_h^j - p_\ell^j}{\phi_h^j - \phi_\ell^j}\right) + \alpha_i\left(\frac{C_i}{{\Lambda^e}_d} - \phi_\ell^i\right)\frac{\phi_h^i - \phi_\ell^i - (p_h^i - p_\ell^i)\partial\phi_h^i/\partial p_h^i}{(\phi_h^i - \phi_\ell^i)^2} \\
&= -\frac{1}{{\Lambda^e}_d^2}\frac{\partial {\Lambda^e}_d}{\partial p_h^i}\sum_{j\in G}\alpha_j C_j\left(\frac{p_h^j - p_\ell^j}{\phi_h^j - \phi_\ell^j}\right) + \alpha_i\left(\frac{C_i}{{\Lambda^e}_d} - \phi_\ell^i\right)\frac{\phi_h^i - \phi_\ell^i - (p_h^i - p_\ell^i)f_V(p_h^i){\phi_h^i}^2}{(\phi_h^i - \phi_\ell^i)^2}.
\end{aligned}$$

Therefore,

$$\frac{\partial {\Lambda^e}_d}{\partial p_h^i} = \left(\frac{\tau}{\Lambda_0}\frac{\partial \mathbf{F}_C^{-1}}{\partial {\Lambda^e}_d}\left(\frac{{\Lambda^e}_d}{\Lambda_0}\right) + \frac{1}{{\Lambda^e}_d^2}\sum_{j\in G}\alpha_j C_j\left(\frac{p_h^j - p_\ell^j}{\phi_h^j - \phi_\ell^j}\right)\right)^{-1}\left(\alpha_i\left(\frac{C_i}{{\Lambda^e}_d} - \phi_\ell^i\right)\frac{\phi_h^i - \phi_\ell^i - (p_h^i - p_\ell^i)f_V(p_h^i){\phi_h^i}^2}{(\phi_h^i - \phi_\ell^i)^2}\right).$$

We see that $\frac{1}{{\Lambda^e}_d^2}\sum_{j\in G}\alpha_j C_j\left(\frac{p_h^j - p_\ell^j}{\phi_h^j - \phi_\ell^j}\right) > 0$. Further, $\mathbf{F}_C^{-1}(x)$ is increasing, by the Inverse Function Theorem. By stability, we know that $C_i/\phi_\ell^i > \Lambda_d^e$. Hence, by assumption, as $\mathbf{F}_V$ is such that

$$\phi_h^i - \phi_\ell^i - (p_h^i - p_\ell^i)f_V(p_h^i){\phi_h^i}^2 < 0, \tag{33}$$

we conclude that $\frac{\partial {\Lambda^e}_d}{\partial p_h^i} < 0$. Clearly, as we said before, as long as $p_\ell^i < p_h^i$, changing the value of p_h^i does not affect the value of $\min_{i\in G}\beta_i(p_\ell^i)$, so for $i \in Y$, we always have $\frac{\partial {\Lambda^e}_d}{\partial p_h^i} \leq 0$.

We now consider the policy $\mathbf{p}'$ where $p_h^i = p_b^i$ for all $i \in G$. We show that in this case ${\Lambda^{e\prime}}_d = \Lambda_s^e$. We start by noticing that for all $i \in G$

$$\Lambda_s^e = \beta_i(p_b^i) = \frac{C_i}{\phi_{p_b^i}} = \frac{C_i}{\phi_h^i}.$$

Let us show that Λ_s^e is a fixed-point solution of $\Lambda^e = \Lambda_0\mathbf{F}_C(Z\mathbf{v}')$. Take ${\Lambda_d^e}' = \Lambda_s^e$. In this case, the expected revenue per ride at location $i \in G$ for the dynamic pricing policy is

$$\begin{aligned}
v_i' = \mathbb{E}_{\mathbf{p}'}[R(i)] = p_\ell^i + \frac{p_h^i - p_\ell^i}{\phi_h^i - \phi_\ell^i}\left(\frac{C_i}{{\Lambda^{e\prime}}_d} - \phi_\ell^i\right) &= p_\ell^i + \frac{p_h^i - p_\ell^i}{\phi_h^i - \phi_\ell^i}\left(\frac{C_i}{\Lambda_s^e} - \phi_\ell^i\right) \\
&= p_\ell^i + \frac{p_h^i - p_\ell^i}{\phi_h^i - \phi_\ell^i}\left(\phi_h^i - \phi_\ell^i\right) = p_h^i = p_b^i.
\end{aligned}$$

It follows that $\mathbf{v}' = \mathbf{p}_b$, so that $Z\mathbf{v}' = Z\mathbf{p}_b$ and $\Lambda_0 \mathbf{F}_C(Z\mathbf{v}') = \Lambda_0 \mathbf{F}_C(Z\mathbf{p}_b)$. As $p^i_\ell \leq p^i_h$ —with at least one strict—,

$$\min_{i\in G} \beta_i(p^i_\ell) > \min_{i\in G} \beta_i(p^i_b) = \Lambda^e_s = \Lambda_0 \mathbf{F}_C(Z\mathbf{p}_b) = \Lambda_0 \mathbf{F}_C(Z\mathbf{v}'),$$

then ${\Lambda^e}'_d = \min(\Lambda_0 \mathbf{F}_C(Z\mathbf{v}'), \min_{i\in G} \beta_i(p^i_\ell)) = \Lambda_0 \mathbf{F}_C(Z\mathbf{v}')$ so we conclude that ${\Lambda^e}'_d = \Lambda^e_s$.

But at this point, we can go from $\mathbf{p}'$ back to our original policy $\mathbf{p}$. So that we increase the value of those p^i_h for which $i \in Y$—, and as $\frac{\partial {\Lambda^e}_d}{\partial p^i_h} \leq 0$ it follows that taking $p^i_h > p^i_b$ decreases —not necessarily strictly— the value of Λ^e for all $i \in Y$. Similarly, for those regions where $p^i_h < p^i_b$, we showed that $v_i < p^i_b$, so Λ^e also decreases.

We conclude that $\Lambda^e_d \leq \Lambda^e_s$.

Theorem 18 *Given a localized pricing schedule $\pi = \{(\underline{\theta}^i, \underline{\mathbf{p}}^i)_i : i \in G\}$ such that:*

$$p^i_j \leq p^i_{j+1}\, \forall j \in \{1, 2, \ldots, k_i\}, \qquad \theta^i_j(n) - \theta^i_{j+1}(n) = \omega(1)\, \forall j \in \{1, 2, \ldots, k_i - 1\},$$

for all regions $i \in G$. Then, for each region $i \in G$, there exists a unique $j^*_i \in \{1, \ldots, k_i - 1\}$ s.t. for any sequence $c(n) = \omega(1)$, we have:

$$\pi_i\left([\theta^i_{j^*_i} - c(n), \theta^i_{j^*_i} + c(n)]\right) \to 1, \qquad \text{as } n \to \infty,$$

where π_i represents the marginal steady state distribution of region i.

Moreover, the equilibrium λ_i satisfies $\mu_0(i)\, \overline{\mathbf{F}}_V(p_{j^*_i+1}) < \lambda_i < \mu_0(i)\, \overline{\mathbf{F}}_V(p_{j^*_i})$.

[of Theorem 18] For ease of notation, in this proof we use λ to denote the normalized large-market limit $\lambda^e = \lim_{n\to\infty} \lambda^e(n)/n$, where $\lambda^e(n)$ corresponds to the vector of rates of incoming drivers to the $n-$th system, with $\Lambda_0(n) = n\Lambda_0$ and $\mu(n) = n\mu_0$. Similarly, we define $\Lambda^e = \lim_{n\to\infty} \Lambda(n)/n$, where $\Lambda(n)$ is the total rate of incoming drivers to the $n-$th system. We add subscripts and parameters to λ and Λ^e when several different systems are being considered.

Assume we fix policy π and let the system run. Then the policy induces some $(\lambda_i)_{i\in G}$. Further, we know that $\lambda = {\lambda^e}^T B$ or, equivalently, $\lambda_i = \Lambda^e(\sigma^T B)_i$. Hence, π induces some value of Λ^e. But once these λ_is are fixed, each region behaves as an *independent* queue, as the system is an open Jackson Network. For each $i \in G$, we have a system identical to the one studied in the proof of Theorem 14.

We can apply the exact same reasoning now: for each i compute $\pi_i(j+1)/\pi_i(j)$ to see that it is unimodal as a function of j as prices are monotone. It then follows that the queue will concentrate around the two prices $p_{j^*_i}, p_{j^*_i+1}$ satisfying

$$\hat{\rho}\phi_{p_{j^*_i}} = \frac{\lambda_i}{\mu_0(i)} \phi_{p_{j^*_i}} < 1 < \frac{\lambda_i}{\mu_0(i)} \phi_{p_{j^*_i+1}} = \hat{\rho}\phi_{p_{j^*_i}}.$$

Theorem 19 *Under the assumptions of Theorem 18 and assuming that F_V is common to all regions and that it satisfies (12), if $p^i_{bal} \in (p_1, p_{k_i})$ for all $i \in G$, then at every region i*

$$p^i_{bal} \in [p_{j^*_i}, p_{j^*_i+1}].$$

[of Theorem 19] For ease of notation, in this proof we use λ to denote the normalized large-market limit $\lambda^e = \lim_{n\to\infty} \lambda^e(n)/n$, where $\lambda^e(n)$ corresponds to the vector of rates of incoming drivers to the $n-$th system, with $\Lambda_0(n) = n\Lambda_0$ and $\mu(n) = n\mu_0$. Similarly, we define $\Lambda^e = \lim_{n\to\infty} \Lambda(n)/n$, where $\Lambda(n)$ is the total rate of incoming drivers to the $n-$th system. We add subscripts and parameters to λ and Λ^e when several different systems are being considered.

Analogously to the proof of Theorem 14, we consider two different policies. On the one hand, the original policy $\pi_A = \{(\underline{\theta}^i, \underline{\mathbf{p}}^i)_i : i \in G\}$. On the other, we consider policy $\pi_B = \{(p_{j^*_i}, p_{j^*_i+1}, \theta^i_{j^*_i})_i : i \in G\}$, where we reduced the original vector of prices to the two prices around which π_A will oscillate according to Theorem 18. We want to show that $\lambda_A(i) = \lambda_B(i)$ for every $i \in G$. As $\lambda_A(i) = \Lambda^e_A(\sigma^T B)_i$ and $\lambda_B(i) = \Lambda^e_B(\sigma^T B)_i$, that will be the case if and only if $\Lambda^e_A = \Lambda^e_B$. We know that in the limit $n \to \infty$,

$$\begin{aligned}
\Lambda^e_A = \Lambda_0\, F_C\left(\frac{\mathbb{E}_{\Lambda^e_A}[R]}{\tau}\right) &= \Lambda_0\, F_C\left(\frac{\sum_{i\in G} \mathbb{E}_{\Lambda^e_A}[R_i]}{\tau}\right) \\
&= \Lambda_0\, F_C\left(\frac{\sum_{i\in G}\sum_{j=0}^{\infty} p_i(j+1)\pi^A_i(j)}{\tau}\right) \\
&= \Lambda_0\, F_C\left(\frac{\sum_{i\in G}\sum_{j=0}^{k_i} p^i_j \pi^A_i(p^i_j)}{\tau}\right) \\
&= \Lambda_0\, F_C\left(\frac{\sum_{i\in G}\ p_{j^*_i}\pi^A_i(p_{j^*_i}) + p_{j^*_i+1}\pi^A_i(p_{j^*_i+1})}{\tau}\right),
\end{aligned}$$

as $\pi^A_i(p_j) = 0$ for all $j \neq j^*_i, j^*_i + 1$.

On the other hand, for system B, we see that

$$\begin{aligned}
\Lambda^e_B = \Lambda_0\, F_C\left(\frac{\mathbb{E}_{\Lambda^e_B}[R]}{\tau}\right) &= \Lambda_0\, F_C\left(\frac{\sum_{i\in G} \mathbb{E}_{\Lambda^e_B}[R_i]}{\tau}\right) \\
&= \Lambda_0\, F_C\left(\frac{\sum_{i\in G}\sum_{j=0}^{\infty} p_i(j+1)\pi^B_i(j)}{\tau}\right) \\
&= \Lambda_0\, F_C\left(\frac{\sum_{i\in G}\ p_{j^*_i}\pi^B_i(p_{j^*_i}) + p_{j^*_i+1}\pi^B_i(p_{j^*_i+1})}{\tau}\right).
\end{aligned}$$

We want to show that Λ_A^e is a fixed-point solution of the above equation for System B. But, at this point, this follows by computations identical to those in the proof of Theorem 14. Further, for system A, by Theorem 18, we know that for each $i \in G$,

$$\Lambda_A^e \frac{(\sigma^T B)_i}{\mu_0(i)} \phi_{p_{j_i^*}} = \frac{\lambda_i}{\mu_0(i)} \phi_{p_{j_i^*}} < 1 < \frac{\lambda_i}{\mu_0(i)} \phi_{p_{j_i^*+1}} = \Lambda_A^e \frac{(\sigma^T B)_i}{\mu_0(i)} \phi_{p_{j_i^*+1}}.$$

As $\Lambda_A^e = \Lambda_B^e$, it directly follows that for every $i \in G$,

$$\Lambda_B^e \frac{(\sigma^T B)_i}{\mu_0(i)} \phi_{p_{j_i^*}} < 1 < \Lambda_B^e \frac{(\sigma^T B)_i}{\mu_0(i)} \phi_{p_{j_i^*+1}}. \tag{34}$$

We can rewrite the equation as

$$\beta_{j_i^*+1}(p_{j_i^*+1}) < \Lambda_B^e < \beta_{j_i^*}(p_{j_i^*}).$$

Hence,

$$\max_{i \in G} \beta_{j_i^*+1}(p_{j_i^*+1}) < \Lambda_B^e < \min_{i \in G} \beta_{j_i^*}(p_{j_i^*}). \tag{35}$$

We argument by contradiction. Assume there exists some region $i \in G$ for which $p_{bal}^i \notin [p_{j_i^*}, p_{j_i^*+1}]$. There are two possible cases, as in Theorem 14.

Either $p_{bal}^i > p_{j_i^*+1} > p_{j_i^*}$ or $p_{bal}^i < p_{j_i^*} < p_{j_i^*+1}$.

By (26), we know that $\Lambda_b^e = \beta_i(p_{bal}^i) = C_i \bar{F}_V(p_{bal}^i) = C_i / \phi_{p_{bal}^i}$. It follows that

$$\Lambda_b^e \frac{(\sigma^T B)_i}{\mu_0(i)} \phi_{p_{bal}^i} = 1.$$

We apply Theorem 17 to System B to conclude that $\Lambda_B^e \leq \Lambda_b^e$, which implies

$$\Lambda_B^e \frac{(\sigma^T B)_i}{\mu_0(i)} \phi_{p_{bal}^i} \leq \Lambda_b^e \frac{(\sigma^T B)_i}{\mu_0(i)} \phi_{p_{bal}^i} = 1 < \Lambda_B^e \frac{(\sigma^T B)_i}{\mu_0(i)} \phi_{p_{j_i^*+1}}.$$

It immediately follows that $p_{bal}^i < p_{j_i^*+1}$. So we discard the case $p_{bal}^i > p_{j_i^*+1} > p_{j_i^*}$.

Assume then that $p_{bal}^i < p_{j_i^*} < p_{j_i^*+1}$.

For $i \in G$ such that $p_\ell^i < p_{bal}(i) < p_h^i$, we have that

$$\beta_i(p_\ell^i) > \Lambda_b^e = \beta_i(p_{bal}(i)) > \beta_i(p_h^i).$$

For $i \in G$ such that $p_{bal}(i) < p_\ell^i < p_h^i$ —call this set Y— we have that

$$\Lambda_b^e = \beta_i(p_{bal}(i)) > \beta_i(p_\ell^i) > \beta_i(p_h^i).$$

We conclude by (35) that

$$\max_{i \in G} \beta_{j_i^*+1}(p_{j_i^*+1}) < \Lambda_B^e < \min_{i \in Y} \beta_{j_i^*}(p_{j_i^*}) < \max_{i \in Y} \beta_{j_i^*}(p_{j_i^*}) < \Lambda_b^e < \min_{i \in G/Y} \beta_{j_i^*}(p_{j_i^*}).$$

We know that

$$\Lambda_B^e = \min\left(\Lambda_0 \mathbf{F}_C\left(\frac{\mathbb{E}[R]}{\tau}\right), \min_{j\in G}\beta_j(p_\ell^B(j))\right). \tag{36}$$

Let us define $\bar{j}$ such that $\bar{j} = \arg\min_{j\in G}\beta_j(p_\ell^B(j))$. Note that $\bar{j}\in Y$.

So, it follows that

$$\Lambda_B^e = \Lambda_0 \mathbf{F}_C\left(\frac{\mathbb{E}[R]}{\tau}\right) < \beta_{\bar{j}}(p_\ell^B(\bar{j})). \tag{37}$$

In other words, the system is supply constrained while, at least in one region, the prices seem to be too high. Let us now consider a third system, say, system C. In this case, the policy we consider is a variation of B.

Let $U\subseteq Y$ be the set of regions j for which $\beta_j(p_\ell^B(j)) = \beta_{\bar{j}}(p_\ell^B(\bar{j}))$. Note that $\bar{j}\in U$.

For $i\in U$, we take $p_\ell^C(i) = p_h^C(i) = p_\ell^B(i)$. We do not change the rest of prices, that is, for $i\in G/U$, we have that $p_\ell^C(i) = p_\ell^B(i)$ and $p_h^C(i) = p_h^B(i)$. Note that we have *not* decreased any low price, but only some high prices —those in U.

It then follows that

$$\begin{aligned}\Lambda_C^e &= \min\left(\Lambda_0 \mathbf{F}_C\left(\frac{\mathbb{E}[R]}{\tau}\right), \min_{j\in G}\beta_j(p_\ell^C(j))\right)\\ &= \min\left(\Lambda_0 \mathbf{F}_C\left(\frac{\mathbb{E}[R]}{\tau}\right), \beta_{\bar{j}}(p_\ell^C(\bar{j}))\right) \le \beta_{\bar{j}}(p_\ell^C(\bar{j})) = \beta_{\bar{j}}(p_\ell^B(\bar{j})).\end{aligned}$$

Recall that, by (31) and (32), we have that

$$v_i = p_\ell^i + \frac{p_h^i - p_\ell^i}{\phi_h^i - \phi_\ell^i}\left(\frac{C_i}{\Lambda_d^e} - \phi_\ell^i\right), \qquad \frac{\mathbb{E}[R]}{\tau} = \frac{1}{\tau}\sum_{j=1}^m \alpha_j v_j.$$

We first claim that $\Lambda_C^e \ge \Lambda_B^e$. This follows by the proof of Theorem 17 as $\partial\Lambda^e/\partial p_h \le 0$. Note that we can take any policy in-between C and B, that is $p_h(j)\in[p_\ell^B(j), p_h^B(j)]$ for $j\in U$ and keeping p_ℓ fixed, as $p_\ell^B(i) = p_\ell^C(i) > p_{bal}(i)$ for $i\in U$, and the stability condition holds by assumption for B.

For regions $k\in G/U$, the prices are identical, so we conclude that

$$\begin{aligned} v_k^C &= p_\ell^k + \frac{p_h^k - p_\ell^k}{\phi_h^k - \phi_\ell^k}\left(\frac{C_k}{\Lambda_C^e} - \phi_\ell^k\right)\\ &\le p_\ell^k + \frac{p_h^k - p_\ell^k}{\phi_h^k - \phi_\ell^k}\left(\frac{C_k}{\Lambda_B^e} - \phi_\ell^k\right) = v_k^B.\end{aligned}$$

On the other hand, for $j\in U$, we see that

$$v_j^C = p_\ell^C(j) = p_h^C(j) < v_j^B,$$

as v_j^B is a convex combination of $p_\ell^B(j)$ and $p_h^B(j)$, given that the system spends a positive fraction of time at each pricing state.

As $\mathbb{E}[R] = \sum_{j=1}^m \alpha_j v_j$, we conclude that

$$\Lambda_C^e \le \Lambda_0 \mathbf{F}_C\left(\frac{\mathbb{E}_C[R]}{\tau}\right) < \Lambda_0 \mathbf{F}_C\left(\frac{\mathbb{E}_B[R]}{\tau}\right) = \Lambda_B^e, \tag{38}$$

which is a contradiction with the fact that $\Lambda_C^e \ge \Lambda_B^e$.

12. Analyzing our Model: Steady-State and Performance Metrics

To complete a service, a driver naturally follows the shortest-delay route to the destination. To make these dynamics more tractable, we assume instead that a busy driver follows a random walk through the nodes of G, with a different random delay at each node. More formally, if a busy driver s is at some node i at time t (i.e., in queue $B_i(t)$), then s incurs a random delay with an $Exponential(\eta_i)$ distribution. Subsequently, the driver either finishes the service with probability p_{ii} (i.e., the passenger gets dropped off at vertex i), or remains busy and transits to vertex $j \neq i$ with probability p_{ij}, while continuing serving the same passenger. Naturally, we impose $\sum_{j \in V} p_{ij} = 1$. We call P the *routing* matrix. Furthermore, the routing matrix P can be related to the traffic matrix T via the following fixed-point relations:

$$t_{ij} = \begin{cases} \sum_{k \in V, k \neq i} p_{ik} \cdot t_{kj} & \text{if } i \neq j, \\ \sum_{k \in V, k \neq i} p_{ik} \cdot t_{ki} + p_{ii} & \text{if } i = j. \end{cases}$$

We can write the above equations in matrix form as $T = (P - D)T + D$, where $D = \text{Diag}(p_{ii})$. Hence, we have that $T = (I - (P - D))^{-1} D$. Note that $(I - (P - D))$ is diagonally-dominated, and hence non-singular.

Summarizing the model described above, the state space $\mathcal{S}$ is defined as:

$$\mathcal{S} = \{a_i \geq 0, b_i \geq 0 : i \in V\}, \tag{39}$$

where a_i and b_i are the number of available and busy drivers at vertex i. Further, owing to our model specification, we can represent the system as a collection of $M/M(k)/1$ queues of available drivers (i.e., $A_i(t)$) and $M/M/\infty$ queues of busy drivers (i.e., $B_i(t)$) at each node, with appropriate (probabilistic) routing between the queues. Note that the routing is *state-independent* by our assumptions.

12.1. Flow-Balance Equations

Suppose that the Markov-chain describing the system is ergodic – then this would correspond to an effective rate of arrival $\{\lambda_i^A, \lambda_i^B\}$ to the available and busy queues. Furthermore, the effective rates must satisfy the following flow-balance equations: for each $i \in V$

$$\lambda_i^A = \lambda_i^e + \sum_{j \neq i} p_{jj} \cdot q_{ji} \cdot \lambda_j^B,$$
$$\lambda_i^B = \lambda_i^A + \sum_{j \neq i} p_{ji} \cdot \lambda_j^B.$$

Let $\lambda^e = \{\lambda_i^e\}_{i \in V}$ be the $1 \times n$ vector of driver sign-in rates at different nodes, and similarly we define effective-rates vectors $\lambda^A = \{\lambda_i^A\}_{i \in V}, \lambda^B = \{\lambda_i^B\}_{i \in V}$. As before, we have $D = Diag(p_{ii})$; now we can rewrite the above equations as:

$$\lambda^A = \lambda^e + \lambda^B DQ, \quad \lambda^B = \lambda^A + \lambda^B (P - D).$$

Recall that the traffic matrix can be written in terms of the routing matrix as $T = (I - (P - D))^{-1}D$. Solving the above equations, and simplifying via the matrix inversion lemma [7], we get:

$$\lambda^A = \lambda^e + \lambda^e T(I - QT)^{-1}Q$$

Equivalently, and again by the matrix inversion lemma, we find that $\lambda^A = \lambda^e(I - QT)^{-1}$. Note that if $q_i^{exit} > 0\,\forall\, i \in V$, then $(I - QT)$ is diagonally-dominant (since QT is a sub-stochastic matrix), and hence non-singular.

12.2. Steady-State Distribution

The previous section characterized the flow-balance equations assuming that the underlying system is ergodic. We now have the following characterization of the conditions under which the system is ergodic, as well as the corresponding steady-state distribution:

Theorem 20 *If for every $i \in V$, we have $\lambda_i^A < \mu_i$ – then the Markov-chain $\mathbf{S}(t)$ is ergodic. Further, for state $\mathbf{x} = \{a_i, b_i\}_{i=1}^n \in \mathcal{S}$, the steady-state probability is given by:*

$$\pi(\mathbf{x}) = \prod_{i \in V} \frac{1}{b_i!} \left(\frac{\lambda_i^A}{\mu_i}\right)^{a_i} \left(\frac{\lambda_i^B}{\eta_i}\right)^{b_i} \pi(\emptyset), \tag{40}$$

where $\emptyset$ is the state where all queues are 0*, and $\pi(\emptyset) = \prod_{i=1}^n \left(1 - \frac{\lambda_i^A}{\mu_i}\right) \exp\left(-\frac{\lambda_i^B}{\eta_i}\right)$.*

Note that the steady-state distribution has *product-form* – each vertex $i \in G$ behaves as an *independent* system of two consecutive queues, an $M/M/1$ queue of parameter μ_i and an $M/M/\infty$ queue of parameter η_i. Further, the arrivals to the first queue (of available drivers) follow a Poisson process of parameter λ_i^A, while the arrivals to the second queue (of busy drivers) follow a Poisson process of parameter λ_i^B.

One way to understand this is via *Burke's Theorem*, which shows that the output from an $M/M/1$ or an $M/M/\infty$ queue is a Poisson process with appropriate rate, and which is independent of the input. In our system, we are essentially taking these outputs, and then probabilistically splitting and feeding them back as inputs to other queues. Such a system of queues is known as a *Jackson network*.

To prove the above result, we use the following result (from Kelly [9], Theorem 1.13):

Theorem 21 *Let $X(t)$ be a stationary Markov process with transition rates $q(j,k)$, $j,k \in \mathcal{S}$. If we can find a collection of numbers $q'(j,k)$, $j,k \in \mathcal{S}$, such that*

$$q'(j) = q(j), \qquad j \in \mathcal{S}$$

[7] For matrices A, B, C, we have $(I - BC)^{-1} = I + B(I - CB)^{-1}C$

and a collection of positive numbers $\pi(j)$, $j \in \mathcal{S}$, summing to unity, such that

$$\pi(j)\; q(j,k) = \pi(k)\; q'(k,j), \qquad j,k \in \mathcal{S} \tag{41}$$

then $q'(j,k)$, $j,k \in \mathcal{S}$, are the transition rates of the reversed process $X(\tau - t)$ and $\pi(j)$, $j \in \mathcal{S}$, is the equilibrium distribution of both processes.

We now return to the proof of Theorem 20:

T o apply Theorem 21, we define the rates of the reversed process for each $i,j \in V$:

$$\begin{aligned}
q\left[(a_i,b_i),(a_i+1,b_i)\right] &= \lambda_i, & q'\left[(a_i+1,b_i),(a_i,b_i)\right] &= \frac{\lambda_i}{\lambda_i^A}\,\mu_i,\\
q\left[(a_i,b_i),(a_i-1,b_i+1)\right] &= \mu_i, & q'\left[(a_i-1,b_i+1),(a_i,b_i)\right] &= \frac{\lambda_i^A}{\lambda_i^B}\,(b_i+1)\eta_i,\\
q\left[(a_i,b_i),(b_i-1,b_j+1)\right] &= b_i\,\eta_i\,p_{ij}, & q'\left[(b_i-1,b_j+1),(a_i,b_i)\right] &= \frac{p_{ij}\,\lambda_i^B}{\lambda_j^B}\,(b_j+1)\eta_j,\\
q\left[(a_i,b_i),(a_i+1,b_i-1)\right] &= b_i\,\eta_i\,p_{ii}, & q'\left[(a_i+1,b_i-1),(a_i,b_i)\right] &= \frac{p_{ii}\,\lambda_i^B}{\lambda_i^A}\,\mu_i,\\
q\left[(a_i,b_i),(a_i,b_i-1)\right] &= b_i\,\eta_i\,p_{i0}, & q'\left[(a_i,b_i-1),(a_i,b_i)\right] &= p_{i0}\,\lambda_i^B.
\end{aligned}$$

It can now be easily checked that for any two states $\mathbf{x},\mathbf{y} \in \mathcal{S}$, Eqn. (41) is satisfied.

12.3. Performance Metrics

Given the steady-state distribution, we can now study various performance metrics of the system. First, we consider the *blocking* rate, i.e., the rate at which passengers are denied service when the system is in equilibrium.

Corollary 22 *The blocking rate R_{blk} in equilibrium is given by:*

$$R_{blk} = \sum_{i \in V} \mu_i - \lambda_i^A.$$

F rom Theorem 20, assuming $\lambda_i^A < \mu_i$, we have:

$$\begin{aligned}
\mathbb{P}[A_i(t) = 0] &= \sum_{s \in \mathcal{S}: a_i = 0} \pi(s)\\
&= \frac{1}{\sum_{a_i=0}^{\infty} \left(\frac{\lambda_i^A}{\mu_i}\right)^{a_i}} = 1 - \frac{\lambda_i^A}{\mu_i}.
\end{aligned}$$

Furthermore, requests arrive at region i at rate μ_i. Thus, the overall blocking probability R_{blk} is

$$R_{blk} = \sum_{i \in V} \mu_i - \lambda_i^A.$$

Blocking results in a loss of welfare for passengers. On the other hand, drivers experience a loss of welfare due to idle-time – when they are not busy, and hence not earning. Two relevant metrics for this are the number of available drivers in the system, and the average delay in different regions.

Corollary 23 *The expected number of available drivers* $\mathbb{E}[A(t)]$ *in equilibrium is given by:*

$$\mathbb{E}[A(t)] = \sum_{i \in V} \frac{\lambda_i^A}{\mu_i - \lambda_i^A}.$$

Furthermore, in region i*, the expected delay experienced by a driver is:*

$$\mathbb{E}[D_i(t)] = \frac{1}{\mu_i - \lambda_i^A}.$$

Proof. First, we have $\mathbb{E}[A(t)] = \sum_{i \in V} \mathbb{E}[A_i(t)]$. Next, using the product-form of the stationary distribution, we have:

$$\begin{aligned}\mathbb{E}[A_i(t)] &= \sum_{s \in \mathcal{S}} a_i(s)\, \pi(s) \\ &= \left(1 - \frac{\lambda_i^A}{\mu_i}\right) \sum_{a_i=0}^{\infty} a_i \left(\frac{\lambda_i^A}{\mu_i}\right)^{a_i} = \frac{\lambda_i^A}{\mu_i - \lambda_i^A}.\end{aligned}$$

Finally, the expression for the expected delay follows from Little's Law. □

Finally, to estimate the welfare of agents, we need to know the rate of requests served between any source and destination nodes (which we denote ϕ_{ij}), and also the rate of idle-driver traffic between any two nodes (which we denote as ψ_{ij}). From Theorem 20, we have:

$$\phi_{ij} = \lambda_i^A t_{ij}, \quad \psi_{ij} = q_{ij} \left(\sum_{k \in V} \phi_{ki} \right).$$

Further, we define $\phi_i^{in} = \sum_{k \in V} \phi_{ki}$ to be the net rate of busy drivers finishing service at node i (and similar for ψ_i^{in}) – note also that $\sum_{j \in V} \phi_{ij} = \lambda_i^A, \sum_{j \in V} \psi_{ij} = (1 - q_i^{exit})\phi_i^{in}$. This completes the description of the system dynamics given (λ, μ, B, Q). We now consider how these parameters can be made endogenous as a function of prices and information-displays (i.e., *heat-maps*), and how they affect the welfare of the agents.

13. Counter-examples

In this section we show that when the conditions of Theorems (9) and (12) on distributions $\mathbf{F}_V$ and $\mathbf{F}_C$ are not satisfied, the results may not hold.

13.1. Optimality of best static policy

We consider the case where the reservation value of users V is distributed as a mixture of independent Gaussian random variables. In particular, we take

$$\mathbf{F}_V(x) = \sum_{t=1}^{5} \pi(t)\mathbf{F}_t(x),$$

where $\pi(t) = 1/5$ for all t and

$$\mathbf{F}_t \sim \mathcal{N}\left(\mu(t), \sigma^2(t)\right), \qquad \mu = (0.1, 0.15, 0.3, 0.5, 0.85), \qquad \sigma^2 = (0.01, 0.05, 0.02, 0.01, 0.02).$$

Clearly, $\mathbf{F}_V$ is not MHR and it does not satisfy equation (12).

Further, we take $\lambda_0 = 7$, $\mu_0 = 4$, $\tau = 1$, $\theta = 3$ and prices ranging in $[0,1]$. The reservation value for drivers is given by $C \sim Exp(1/4)$.

We show in Figure 6 that dynamic pricing policies lead to higher λ and revenue.

13.2. Robustness

In this case, consider the situation where $C \sim \mathbf{F}_C$ and $\mathbf{F}_C$ is given by the following mixture of Gaussians:

$$\mathbf{F}_C(x) = \pi_1\mathbf{F}_1(x) + \pi_2\mathbf{F}_2(x) + \pi_3\mathbf{F}_3(x), \tag{42}$$

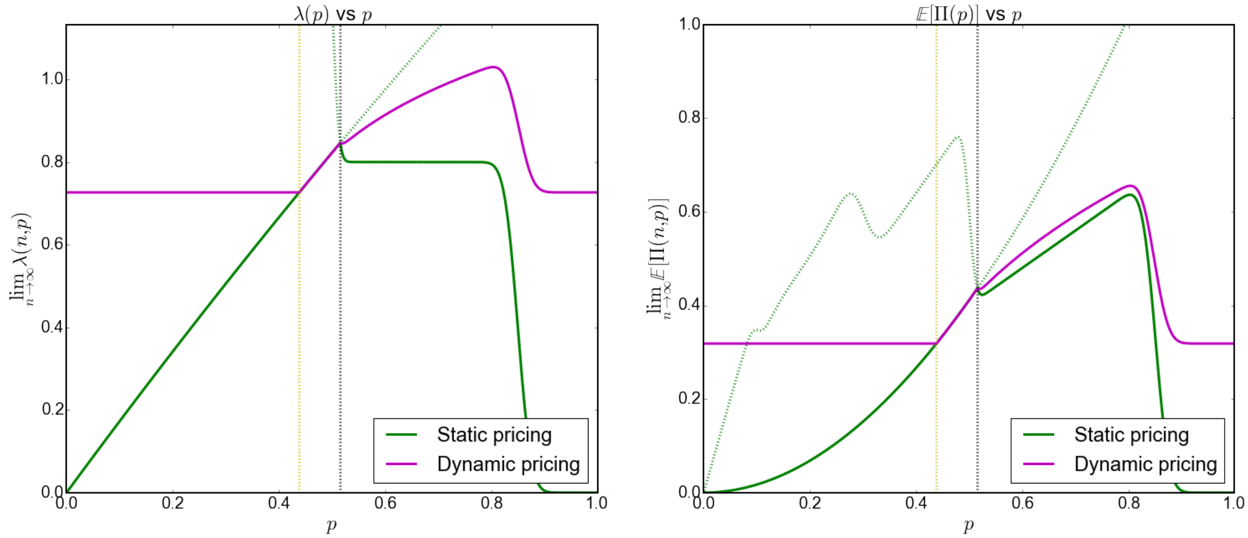
where

$$\mathbf{F}_1 \sim \mathcal{N}(0.2, 0.1), \qquad \mathbf{F}_2 \sim \mathcal{N}(0.5, 0.2), \qquad \mathbf{F}_3 \sim \mathcal{N}(0.7, 0.05) \tag{43}$$

and $\pi(1) = 0.3$ while $\pi(2) = \pi(3) = 0.35$. One can easily check that $\mathbf{F}_C$ is not log-concave.

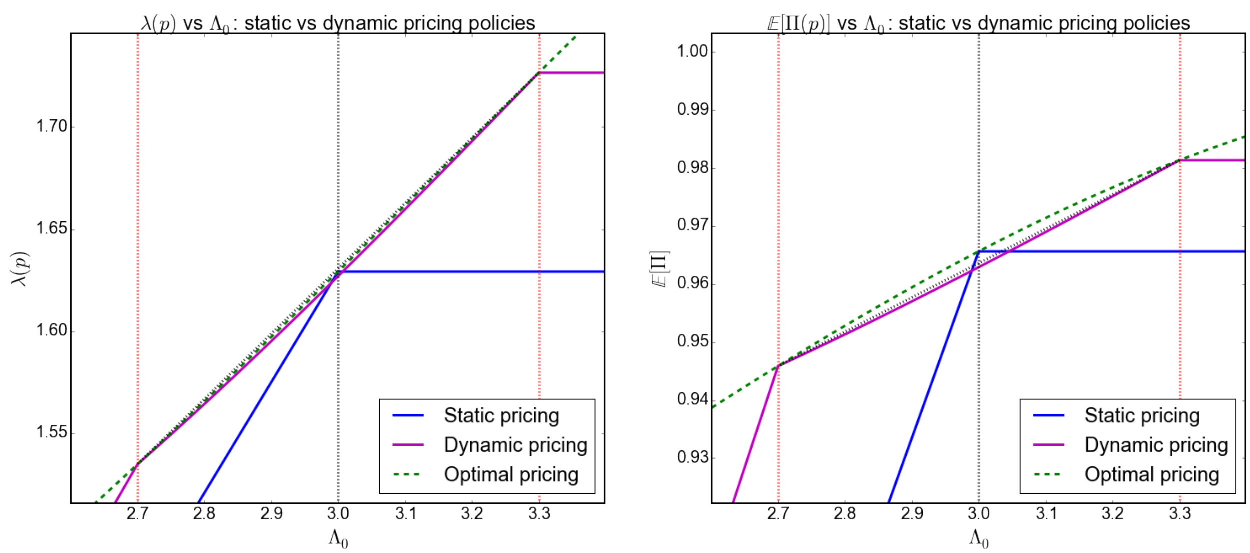
We take $\lambda_0 = 3$, $\mu_0 = 4$, $\tau = 1$, $\theta = 3$ and prices range in $[0,1]$. The reservation value for drivers is given by $V \sim U(0,1)$. The balance price is $p_{\text{bal}} = 0.593$ and the dynamic pricing policy is $\pi = (p_\ell, p_h) = (0.568, 0.616)$.

We show in Figure 7 that the performance of dynamic pricing policy π lies below the linear interpolation between the performance of the static policies of their extreme points p_ℓ and p_h when they are optimal (in other words, when these prices are actually the balance price).



(a) Static optimality does not hold when $\mathbf{F}_V$ violates the assumptions of Theorem 9.

Figure 6 When $\mathbf{F}_V$ is not well-behaved, the results of Theorem 9 need not to be true. We see that —in the large-market limit— we can find dynamic pricing policies leading to values of λ higher than λ_{bal} —left picture— and also higher revenue than the optimal one in the static case —right picture.



(a) Robustness does not hold when $\mathbf{F}_C$ violates the assumptions of Theorem 12.

Figure 7 When $\mathbf{F}_C$ is not well-behaved, the results of Theorem 12 need not to be true. We see that —in the large-market limit— dynamic pricing policies $\pi = (p_1, p_2, \theta)$ enclosing the balance price may not be robust, that is, the performance of these policies lie strictly below the linear interpolation between that of optimal static policies corresponding to p_1 and p_2 when these prices correspond to the balance price of the system. In the left, we show λ and, in the right, we show revenue.