Submitted to *Operations Research*
manuscript OPRE-2015-01-039

Pricing and Prioritizing Time-Sensitive Customers with Heterogeneous Demand Rates

Philipp Afèche, Opher Baron, Joseph Milner
Rotman School of Management, University of Toronto, 105 St. George Street, Toronto, Ontario, M5S 3E6
afeche@rotman.utoronto.ca, opher.baron@rotman.utoronto.ca, jmilner@rotman.utoronto.ca

Ricky Roet-Green
Simon Business School, University of Rochester, 305 Schlegel Hall, Rochester, NY 14627 rgricky@gmail.com

We consider the pricing/lead-time menu design problem for a monopoly service where time-sensitive customers have demand on multiple occasions. Customers differ in their demand rates and valuations per use. We assume that customers queue for a finite-capacity service under a general pricing structure. Customers choose a plan from the menu to maximize their expected utility. We compare two models where the demand rate is the private information of the customers to a model where the firm has full information. In the Aggregate Control Model the firm controls the number of plans it sells for each class of service but cannot track each customer's usage level. In the Individual Control Model the firm can track the usage of individual customers but does not control the number of plans sold. In contrast to previous work, we show that, although we assume customers do *not* differ in their waiting cost, prioritizing customers may be optimal as a result of demand rate heterogeneity in the private information case. We provide necessary and sufficient conditions for this result. In particular, we show that for intermediate capacity, more frequent-use customers that hold a lower marginal value per use should be prioritized. Further, less frequent-use customers may receive a consumer surplus. We demonstrate the applicability of these results to relevant examples. The structure of the result implies that in some cases it may be beneficial for the firm to prioritize a customer class with a *lower* marginal waiting cost.

Key words: Capacity Pricing, Heterogeneous Usage Rates, Priority queues

1. Introduction

Service firms sell memberships to frequent users that reduce the price paid by customers, yet raise the revenue received. And membership has its privileges. Season passes to leisure activities such as ski mountains and amusement parks are often accompanied by perks such as access to priority queues at theme parks (e.g., Universal Studios Express Pass) or early admission (e.g., Stratton Mountain Summit Pass). Memberships allow line-jumping for exhibit entrance at cultural institutions and early registration for classes at social organizations. Firms typically offer several different pricing plans, e.g., unlimited access (a season pass), limited access (a multiple use ticket), two-part tariffs (paid-for discount cards such as tastecard), and a pay-per-use price, and these often come with differing benefits. The customer's choice of which price to pay depends on the value expected to be derived from use and the total cost including the cost of waiting. And by inducing different customers to pay different prices, firms can increase their total revenue.

Consider, for example, the choice of whether to purchase a season pass at a ski resort allowing unlimited access. This may be of interest to skiers residing near the mountain (the 'locals'). They are likely to have a higher frequency of use than vacationers coming to the resort (the 'aways'). However, a local also may derive less enjoyment from any particular day of skiing than an away, as they may see multiple opportunities each season, reducing the marginal value. If the mountain's management does not offer a season pass, the locals may reduce their skiing. But if the season pass is priced too low, the aways may purchase one, reducing the mountain's revenue. Thus the mountain's management has the problem of pricing a season pass to attract locals, but not aways. And unless they require proving residency to purchase such a pass, it is difficult to distinguish locals from aways. But as noted above, the firm has another tool, the perks it offers along with the pass. In particular, priority services such as pass-holders' lift lines and early mountain access are of value when the system is congested. We show that these perks are not simply additional benefits of membership, but are necessary to maximize the mountain's revenue.

We focus on the problem of designing price/lead-time menus and the corresponding priority policy for a profit-maximizing service provider serving customers with private information on their preferences. We study the full information case as a benchmark. Customers are risk-neutral and maximize their expected utility by choosing whether to buy service, and if so, which service class (price-leadtime option on the menu). The key novelty is that the paper studies settings where customers have demand for multiple uses, and they are heterogeneous in these demand rates, as is the case for example in ski resorts or amusement parks. Most previous studies restrict attention to the case where customers have *unit* demand, that is, they have identical infinitesimal demand rates. The few papers that do consider heterogeneous demand rates (see Rao and Petersen (1998), Van Mieghem (2000), Masuda and Whang (2006), Randhawa and Kumar (2008), Cachon and Feldman (2011), Plambeck and Wang (2013)) typically restrict attention either to the case in which the provider observes customer preferences, or to undifferentiated First In First Out (FIFO) service which misses the value of differentiated service.

The paper deliberately focuses on the simplest model to understand the *minimal* conditions for *priority service* to be profit-maximizing when customers have heterogeneous demand rates. Specifically, we model the service facility as an M/M/1 queue and consider customer types that differ only in their demand rates and their marginal (per-use) valuations; we do *not* assume the customers have differing marginal costs of waiting. Rather, we show that differences in valuation and demand rate are sufficient in some conditions to make prioritized service optimal for revenue maximization.

Customers repeatedly use the service (if they find it economical to do so) up to a type-dependent rate. We assume a type-dependent marginal value for each use and apply a strict ordering on the marginal value for the types. We allow the marginal value of an additional usage for the more frequent-use type to be either higher than or lower than that of the less frequent-use type. Both cases are possible and represent alternate orderings of the marginal rate of substitution between usage and price depending on the customer's type. (This is the constant sign assumption, cf. Fudenberg and Tirole (1991).) For example, skiing enthusiasts may find higher marginal value than a skiing novice at all frequencies of use. Alternatively, the marginal value of use of a visitor to a ski resort may be higher than that of a local for whom there are multiple opportunities to ski in a season. We investigate the firm's policy for both cases, comparing the firm's optimal policy under full and private information.

We present two models for how the firm controls its service offering in the private information setting. In the Aggregate Control Model, discussed in Section 4, the firm controls the number of plans it sells for each class of service. It cannot distinguish between customers and cannot track the usage rate of individual customers. As such, the firm optimizes over the number of customers to serve, in addition to the price/lead-time menu. Customers are free to use the service (and do so) up to their type-dependent maximum rate. In this model, we allow customers to choose which service class to purchase as well as allow them to purchase multiple (and not necessarily integral) copies of service. As such the aggregate control model may be seen as offering minimal provider control and maximum user flexibility. In Section 5 we discuss the Individual Control Model. Here, we assume the firm can track the usage of individual customers. Such might be the case if customers must present ID on purchasing and all interactions are registered (e.g., by RFID) as is increasingly common at resorts and amusement parks. We assume that the firm will sell to all customers that wish to purchase the service, but restricts each person to a single purchase. The firm determines the price/lead-time menu that may limit the usage rate of a customer to some value less than their preferred usage rate. This model may be seen as having the customers endogenously determine their usage rate based on the menu of offerings. However, as a mechanism design problem, we show that the firm can effectively choose each customer's usage rate, though it is bounded by individual rationality and incentive compatibility. As such, this model may provide greater regulation of the system than the aggregate control model.

This paper makes several contributions on the design of differentiated price-service mechanisms for queueing systems. First, we demonstrate the fundamental point that when customers differ in their demand rates, it may be optimal to offer delay-differentiated services (through priorities) rather than uniform service (e.g., FIFO), even though all customers are equally delay-sensitive. Importantly, we show this holds for both the aggregate control and individual control models. In fact, it follows from our derivation that a firm may prioritize customers that are *less* sensitive to delay (see Section 6.1). This result runs counter to the conventional wisdom, given by the extensive literature on serving customers with equal demand rates, that prioritizing customers has positive value only if they have higher delay costs.

Second, we provide necessary and sufficient conditions, in terms of the demand and capacity characteristics, for priority service to be optimal. In brief, priority service is optimal only if customers with higher demand rates (the 'locals') value the service *less* per use than their low-demand counterparts (the 'aways'). We further distinguish between two types of locals depending on the value per unit time, e.g., value per season, they would be willing to pay for the service. Suppose the locals would not generate high revenue per unit time compared with the aways. Call such customers 'hobbyists' as they would take advantage of the service frequently, but only if the total cost to do so was low. Then if they have sufficient number and there was sufficient capacity, the firm would offer the hobbyists priority service in order to attract them. On the other hand, suppose the locals have a lower marginal value per use, but would go sufficiently often so that the rate they would generate revenue would be higher than the aways. Call such customers 'enthusiasts'. We show that if there is sufficient but not excessive capacity, then offering priority service to the enthusiasts is optimal. However, for an even greater capacity, when there are enthusiasts, the firm can perfectly discriminate between the users and achieve the first best revenue.

Our results also point to potential implementations generating significant value for the firm. When priority service is optimal, the menu is designed such that hobbyists or enthusiasts purchase a subscription, e.g., a season pass. On the other hand, the aways can be offered either a per use fee, or a two part-tariff including a per use fee. We show that offering optimal delay-differentiated services can generate significant profit gains, compared to FIFO service. The result implies that the use of priority queues seen in many environments such as amusement parks and ski resorts is not just a reward for loyal, season-ticket purchasing customers, but part of the mechanism design that allows the firm to differentiate between customer types. Refer to Section 6 for a discussion of our results and how they generalize.

Hassin and Haviv (2003) and Hassin (2016) provide a comprehensive literature review of research into the equilibrium behavior of customers and servers in queueing systems with pricing. The vast majority of pricing studies for queues restrict attention to the case where customers have unit demand, that is, they have identical infinitesimal demand rates. Naor (1969) and Mendelson (1985) consider first-in-first-out (FIFO) service for customers with homogeneous delay costs. Mendelson and Whang (1990), Hassin (1995), and Hsu et al. (2009) characterize the socially optimal price-delay menu and scheduling policy for heterogeneous customers. Motivated by the substatntial waiting times at some U.S.–Canada border crossing stations, Guo and Zhang (2013) study customers' lane choice behavior and congestion-based staffing policies for a multi-server system that offers two options, one free and slower, and the other for a fixed price and faster. Some papers on the revenue-maximization problem for heterogeneous customers restrict the scheduling policy, customers' service class choices, or both (cf. Lederer and Li (1997), Boyaci and Ray (2003), Maglaras and Zeevi (2005), Allon and Federgruen (2009), Zhao et al. (2012), Afèche et al. (2013)). Afèche (2004) initiated a stream of revenue-maximization studies that design jointly optimal prices and scheduling policies in the presence of incentive-compatibility constraints (cf. Katta and Sethuraman (2005), Yahalom et al. (2006), Afèche (2013), Maglaras et al. (2014)). The conventional wisdom that emerges from all of these unit-demand studies is that offering priorities has positive value only if customers have heterogeneous delay costs. In contrast, only a few papers consider customers who have demand for multiple uses and who are heterogeneous in this attribute: some have high, others have low demand. Rao and Petersen (1998) and Van Mieghem (2000) consider the welfare-maximization problem. Rao and Petersen (1998) study a model with pre-specified priority delay functions, which eliminates the scheduling problem. Van Mieghem (2000) considers the menu design question jointly with the optimal scheduling problem under convex increasing waiting cost functions. Papers that consider the revenue-maximization problem under restriction to FIFO service establish the optimality of fixed-up-to tariffs (Masuda and Whang 2006) or compare the performance of simpler tariffs, namely, subscription-only versus pay-per-use pricing (Randhawa and Kumar (2008), and Cachon and Feldman (2011)). Finally, Plambeck and Wang (2013) consider revenue maximization with multiple-use customers whose service valuations are subject to hyperbolic discounting. This model captures the preference structure for unpleasant services. The optimal mechanisms they study are tailored to such settings, which are in marked contrast to the more pleasant services that fit our model.

2. Model

We consider a capacity-constrained monopoly firm that designs a menu of price-service plans for customers that differ based on their maximum demand rate for the service and the value they derive from each usage. Customers are delay-sensitive and prefer faster service. There are two customer types, indexed by $i = 1, 2$. The market for each type consists of a fixed, large number of potential customers, N_i. We assume that all customers have the *same* waiting cost, c, per unit time in the system (including service). This assumption eliminates waiting cost heterogeneity as the driver of delay differentiation, the focus of virtually the entire previous literature on priority pricing. Rather, we focus on identifying the conditions for optimal delay differentiation to arise as a result of demand rate heterogeneity.

We consider a service for which the expected value derived from each use is constant and for which customers have a limited expected number of opportunities to partake in the service. For services where subscriptions or per use sales are available, such as attending the ballet or going to a ski resort, many customers have some notion of the value that a single service will provide. Further, this value is constant, ex ante, in that how rewarding the service will be, depends on a number of factors (such as weather or performer) so that the true value is a random variable. For many such services customers would engage in the activity as much as possible, however, availability of time limits the number of uses. Therefore, we consider a setting where when making a purchasing decision for future use, type i customers expect that they will use the service up to a rate γ_i times per season. Without loss of generality (w.l.o.g.), we assume $\gamma_1 > \gamma_2$. Type-i customers expect to receive a constant marginal value r_i for each service usage. Therefore, $n_i\gamma_i$ is the total value rate that a type-i customer receives per unit time if they use the service at every opportunity. We assume that the firm knows the aggregate demand information (r_i, γ_i, and N_i for $i = 1, 2$, and c). In Section 6.1 we explain how the results and the underlying optimality conditions extend under general decreasing marginal value functions.

The firm operates a service facility with fixed capacity, μ. (Our results characterize the optimal menu as a function of μ.) We assume that the service operates as an $M/M/1$ queue. We further assume $\gamma_i << \mu$, implying that while customers may use the service multiple times during the season, each customer's usage of the capacity is relatively insignificant. We model the total demand rate of each customer type as a Poisson process. This involves two assumptions: that the arrival process is independent of how many customers are in the system and that this arrival process is Poisson. Both assumptions are defendable in our settings as long as the number of customers of each type, N_i, is very large and their usage rate, γ_i is very small. Such settings are consistent with the types of service firms we are modeling (amusement parks, ski resorts, etc.). Specifically, using the Palm-Khintchine Theorem – see Heyman and Sobel (1982, p. 160) – the superposition of infinitely many independent arrival processes each with an infinitesimal arrival rate is a Poisson process. Therefore, under our assumptions and when different user's service opportunities are independent of each other, the total demand process will be (approximately) Poisson. Moreover, with both assumptions and at reasonable capacities, the total number of customers in queue is still relatively small in comparison to N. That is, the total demand rate can be considered practically independent of the number of customers in the system. Further support for approximating a finite population by an infinite one can be found in, e.g., Cachon and Feldman (2011) for a pricing with congestion setting, Green and Savin (2008) for a healthcare setting, Afèche et al. (2017) for a customer relationship management setting, and Lariviere and Mieghem (2004) in the context of competition.

The service provider first designs and announces a static menu of price-service plans for up to two service classes, indexed by $j = 1, 2$. Customers then choose from the menu the class of service to purchase as detailed below. To be clear, "class" refers to the attributes of a service option, "type" refers to those of a customer. We also refer to class-j service as plan j where it is natural to do so. The restriction to two service classes is w.l.o.g. in our model. (The provider cannot generate more revenue by offering more service classes than customer types.) The menu specifies for each class a usage rate-dependent tariff (or price function) and the expected waiting time a customer will encounter at each visit to the facility.

Let $P_j(x)$ be the total per season revenue generated by a customer with usage rate x who chooses class-j service, $j \in \{1, 2\}$. This form is general and can represent any pricing scheme including a service class with unlimited usage at a subscription price, a two-part tariff with or without a maximum usage rate, or a simple per use price. If, for example, $P_j(x)$ were a two-part tariff with subscription fee F_j and price per use of p_j, then $P_j(x) = F_j + p_j x$.

Let W_j be the expected waiting time (or lead time, including service time) for class-j service. We require W_j's to be consistent with the average steady-state wait times that are realized given the provider's scheduling policy and customer usage. This consistency requirement may be enforced by auditors or third party review sites. Practically, for the motivating examples, social media provides a means for customers to learn the expected wait times prior to purchase. See Afèche (2013) for further discussion.

We do not assume a specific scheduling policy but rather let the provider choose any non-anticipative and regenerative policy. This appears to be the most general, easily described restriction of admissible policies that guarantees the existence of long run waiting time averages. We allow preemption, which simplifies the analysis without affecting the results (under priority scheduling with preemption, the waiting time of the prioritized class does not depend on the workload of the non-prioritized class).

Given the menu, customers decide whether to seek service, and if so, choose a plan to maximize their expected total utility. Their decision is based on their value per use and their expected waiting times, noting we allow them to optimize their usage rate (within the bounds set by each service plan). Each customer makes a decision at the start of the season (or when their first service opportunity arises). Customers are risk neutral. They do not observe the queue. This assumption is common in related papers. For the motivating applications, the notion is that the queue cannot be observed by the customer even at the time of purchase as it may be spatially or temporally removed from the ticket window.

Let x_i be the usage rate of a type-i customer and let n_j be the number of class-j plans sold. We assume $0 \le x_i \le \gamma_i$ and $0 \le n_j \le N_j$. (For simplicity, we treat n_i as a continuous variable throughout, rather than as an integer; given large N_i, this is a mild assumption.)

From these definitions, the total expected utility of a type-i customer who buys a class-j plan and uses it at rate x_i is $(r_i - cW_j)x_i - P_j(x_i)$, where $r_i - cW_j$ is her expected net value from every service opportunity. Let $u_i(x_i)$ denote the utility of a type-i customer who buys a class-i plan and uses it at rate x_i. So

$$u_i(x_i) = (r_i - cW_i)x_i - P_i(x_i). \tag{1}$$

As a benchmark, we first investigate the firm's revenue maximization problem under the Full Information setting where the firm can distinguish the customer type. We then turn to the Private Information setting for the Aggregate Control Model in Section 4. In this model the provider cannot track and limit individual usage so that customers choose to use the service at their maximum rate, i.e., $x_i = \gamma_i$. The firm optimizes over n_j. We then consider the Individual Control Model in Section 5 where the firm does not control the number of service plans it sells, but can track and limit the usage of individual customers. Here $n_j = N_j$ and the firm optimizes over x_i.

3. Full Information Setting

In the Full Information (FI) setting the firm can distinguish between the customer types, assign a price for each customer type, and enforce the customers to pay that price if they use the service. Initially, consider the case where the firm can track individual customer usage so that it can dictate x_i. (We relax this assumption at the end of the section.) In the FI setting, the firm maximizes its profit by simultaneously choosing the number of customers, n_i, their usage rate, x_i, the pricing, $P_i(x_i)$, and the prioritization that subsequently defines the waiting time, W_i, for each class. The firm's policy is constrained by the need for customers to receive non-negative utility.

In the FI setting the problem is:

$$\Pi^{\mathrm{FI}} = \max_{n_i, x_i, W_i, P_i(x_i)} \quad \sum_i n_i P_i(x_i) \tag{2a}$$

$$\text{subject to} \quad u_i(x_i) = (r_i - cW_i)x_i - P_i(x_i) \ge 0 \quad \text{for } i = 1,2 \tag{2b}$$

$$W_i \ge \frac{1}{\mu - n_i x_i} \quad \text{for } i = 1,2 \tag{2c}$$

$$\sum_i n_i x_i W_i \ge \frac{\sum_i n_i x_i}{\mu - \sum_i n_i x_i} \tag{2d}$$

$$0 \le n_i \le N_i \quad \text{for } i = 1,2 \tag{2e}$$

$$0 \le x_i \le \gamma_i \quad \text{for } i = 1,2. \tag{2f}$$

The objective function gives the total revenue rate of the firm. Constraint (2b) is the individual rationality (IR) constraint. If type-i customers receive zero utility from buying class-i at their optimal rate, we assume they prefer buying class i over not seeking service. (As the firm controls n_i, the equilibrium participation is unique and so is the resulting revenue.) Constraints (2c) and (2d) define the achievable region for the waiting time. Constraint (2c) ensures for each class that the waiting time is bounded below by the minimum feasible waiting time for class-i under strict preemptive priority in an M/M/1 queue. Constraint (2d) verifies that the (weighted) average wait time for both service classes is bounded below by the minimum achievable, non-idling waiting time (for work-conserving policies (2d) is binding). Constraints (2e) and (2f) enforce the non-negativity and upper bounds for each customer type.

We can simplify the problem by eliminating the pricing, $P_i(x_i)$, and waiting times, W_i, as follows. Observe that in maximizing the objective function, the individual rationality constraints, (2b), are binding. Therefore the price for class-i service in the FI solution must satisfy

$$P_i(x_i) = (r_i - cW_i)x_i.$$

The objective function can then be written as

$$\sum_i n_i x_i r_i - c\sum_i n_i x_i W_i,$$

or, letting $\lambda_i = n_i x_i$ be the demand load from type-i customers,

$$\sum_i \lambda_i r_i - c\sum_i \lambda_i W_i.$$

Rewriting (2d)

$$\sum_i \lambda_i W_i \geq \frac{\sum_i \lambda_i}{\mu - \sum_i \lambda_i}$$

it is evident that, because both customer types have the same delay cost, for any fixed λ_1 and λ_2, every work-conserving policy is optimal. In particular, let W be the waiting time under FIFO service. That is,

$$W_1 = W_2 = W = \frac{1}{\mu - \sum_i \lambda_i}$$

is optimal given λ_i. Therefore, letting $\Lambda_i = N_i \gamma_i$, we can reduce the Full Information problem to

$$\Pi^{\mathrm{FI}} = \max_{\lambda_1,\lambda_2} \quad \sum_i \lambda_i \left(r_i - \frac{c}{\mu - \lambda_1 - \lambda_2} \right) \tag{3a}$$

$$\text{subject to} \quad 0 \leq \lambda_i \leq \Lambda_i \text{ for } i = 1, 2. \tag{3b}$$

For a given μ, let λ_i^{FI} be the optimal usage rate for all type i customers in the FI setting. The solution to (3) depends on which of the customer types has higher marginal value, r_i, and on the capacity of the firm as characterized by the following proposition:

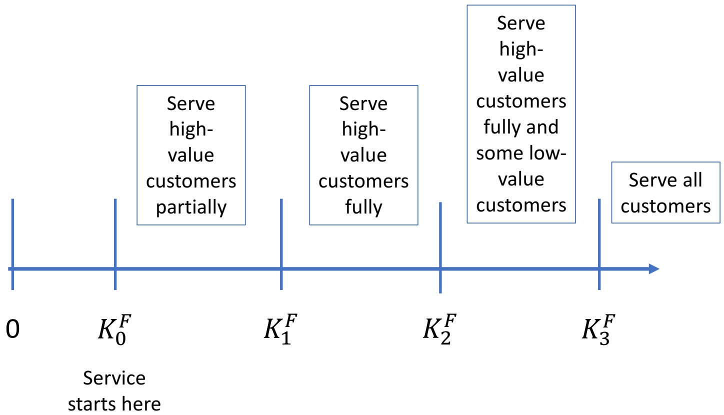

Figure 1: FI Setting Optimal Policy as a Function of Capacity

Proposition 1 *In the FI setting FIFO service is optimal. For $r_i \geq r_j$, $i, j \in \{1, 2\}, i \neq j$, there exist four thresholds over the capacity $K_0^F < K_1^F \leq K_2^F < K_3^F$ such that:*

1. *For $\mu \leq K_0^F$, the provider does not serve any customers, $\lambda_1^{FI} = \lambda_2^{FI} = 0$.*
2. *For $K_0^F < \mu < K_1^F$, the provider serves type-i customers exclusively, but only partially, such that $0 < \lambda_i^{FI} < \Lambda_i$, $\lambda_j^{FI} = 0$.*
3. *For $K_1^F \leq \mu \leq K_2^F$, the provider serves type-i customers exclusively and fully, such that $\lambda_i^{FI} = \Lambda_i$, $\lambda_j^{FI} = 0$.*
4. *For $K_2^F < \mu < K_3^F$, the provider serves type-i customers fully ($\lambda_i^{FI} = \Lambda_i$), and type-$j$ customers partially ($0 < \lambda_j^{FI} < \Lambda_j$).*
5. *For $\mu \geq K_3^F$, the provider serves type-i and type-j customers fully, such that $\lambda_i^{FI} = \Lambda_i$, $\lambda_j^{FI} = \Lambda_j$.*

Proposition 1 implies that as the capacity increases, the higher-value type-i customers are allocated capacity initially, and lower-value type-j customers are served only if there is sufficient residual capacity, as would be expected. In doing so, the firm engages in a revenue skimming policy for its capacity. The policy is illustrated in Figure 1.

Since $\lambda_i = n_i x_i$, the reduced form of (3) implies there is a symmetry between n_i and x_i. This implies that the Aggregate Control Model is equivalent to the Individual Control Model in the Full Information setting. Let n_i^{FI} be the number of type-i customers that are served and let x_i^{FI} be the optimal usage rate of the type-i customers in the FI setting. Under the Aggregate Control Model, discussed next, the firm determines n_i assuming type-i customers use the service at rate γ_i. Thus in the FI case, the solution to the Aggregate Control problem is given by setting $x_i^{FI} = \gamma_i$ and $n_i^{FI} = \lambda_i^{FI}/\gamma_i$. Similarly in the Individual Control Model, the firm offers plans which limit customers' usage rate to x_i, but sells to all customers. In this case, for the Full Information setting, $x_i^{FI} = \lambda_i^{FI}/N_i$ and $n_i^{FI} = N_i$.

Remark 1 For $\mu \le K_2^F$ the Private Information solutions studied in Sections 4 and 5 match the FI solution, i.e., serving only the higher-value type is optimal. Therefore, we focus henceforth on $\mu > K_2^F$.

4. Private Information Setting - Aggregate Control Model

In the Private Information (PI) setting, we assume the firm cannot distinguish between customers' types. To be specific we assume that the firm cannot determine the type of a customer prior to their first purchase of the season or based on their usage during the season. If their type can be identified from usage, we assume that the firm cannot take advantage of that information. For example, the firm does not change their service menu during the season, i.e., as assumed, the menu is static. In particular, for customers that purchase a season pass, the firm has already been paid and differentiating service based on a customer's true type would be difficult as the pass would constitute a contract. For customers that pay per use, they may do so without identifying themselves so that tracking their type may be difficult.

Since the firm cannot distinguish between the types, customers choose from the menu to maximize their expected utility. The firm cannot designate a class of service to a particular type of customer without providing an incentive to ensure they choose one plan over another. We restrict attention w.l.o.g. to menus ensuring incentive compatibility (IC) that target class-i service to type-i customers such that they weakly prefer class i or no service over service in class $j \neq i$. (Based on the revelation principle (e.g., Myerson (1997)), mechanism design problems restrict attention w.l.o.g. to IC direct revelation mechanisms in which each customer directly reveals her type. The mechanism described below is an "indirect mechanism" that is equivalent to a direct mechanism and more naturally describes the purchase process.)

Remark 2 For the sake of analytic simplicity, and without loss of optimality, we can restrict the choice of tariffs to the class of Fixed-Up-To (FUT) tariffs. (We discuss alternative optimal pricing schemes later.) A FUT tariff for class-i service is characterized by two parameters, a maximum usage rate $\overline{x}_i$ per season and a constant (total) price P_i. It is known that any optimal menu of non-linear tariffs can be implemented by a menu of FUT tariffs with the same usage rates and payments (see Masuda and Whang 2006, Theorem 2).

4.1. Aggregate Control Model

In the Aggregate Control model, we assume the firm can track the number of plans sold of each class, however, the firm cannot track the usage of the service by an individual. Thus, even if the firm wanted to restrict the usage rate of an individual, each person could purchase multiple copies of any plan in order to use the service at any rate of their choosing. As all customers can choose to use the service at any rate, and given that a type-i customer sees a constant value r_i for usage, they will choose to use the service at their maximum rate γ_i if they purchase a plan. Therefore, the type-i usage rate is $x_i = \gamma_i$ so we can restrict attention to FUT tariffs for class-i with maximum usage rate γ_i, and optimal fee P_i that remains to be determined. Thus the firm can only control the usage in aggregate by determining n_i, the number of class-i plans to sell. To simplify notation we let $u_i = u_i(\gamma_i)$ in this section.

For the Aggregate Control Model, we need to add an IC constraint eliminating the possibility of a high-use type-1 customer representing himself as a type-2 customer and purchasing multiple copies of class-2 service. In particular, a type-1 customer would require a minimum of γ_1/γ_2 copies of class-2 service assuming each is used as planned at rate γ_2 (we ignore integrality constraints).

Therefore, we require the expected annual cost for a high demand type-1 customer to use class-1 service to be less than the cost to use class-2 service, i.e.,

$$P_1 + cW_1\gamma_1 \le \frac{\gamma_1}{\gamma_2}P_2 + cW_2\gamma_1,$$

or equivalently

$$u_1 \ge \gamma_1(r_1 - cW_2) - \frac{\gamma_1}{\gamma_2}P_2, \text{ if } n_2 > 0.$$

The low demand type-2 customers have no incentive to purchase more than a single class-1 plan, so that IC requires that

$$u_2 \ge \gamma_2(r_2 - cW_1) - P_1 \text{ if } n_1 > 0.$$

The firm's Private Information problem for the Aggregate Control model is as follows (the firm chooses the menu and the number of passes simultaneously):

$$\begin{aligned}
\Pi^A = \max_{n_i, u_i, W_i, P_i} \quad & \sum_i n_i P_i && \text{(4a)}\\
\text{subject to} \quad & u_i = \gamma_i(r_i - cW_i) - P_i \ge 0 \qquad \text{for } i = 1,2 && \text{(4b)}\\
& u_1 \ge \gamma_1(r_1 - cW_2) - \frac{\gamma_1}{\gamma_2}P_2 \text{ if } n_2 > 0 && \text{(4c)}\\
& u_2 \ge \gamma_2(r_2 - cW_1) - P_1 \text{ if } n_1 > 0 && \text{(4d)}\\
& W_i \ge \frac{1}{\mu - n_i\gamma_i} \qquad \text{for } i = 1,2 && \text{(4e)}\\
& \sum_i n_i\gamma_i W_i \ge \frac{\sum_i n_i\gamma_i}{\mu - \sum_i n_i\gamma_i} && \text{(4f)}\\
& 0 \le n_i \le N_i \qquad \text{for } i = 1,2. && \text{(4g)}
\end{aligned}$$

This formulation is almost identical to (2), with the addition of the IC constraints.

We consider two cases. In the first, referred to as the Increasing Order, the marginal value of each use is increasing in the demand rate, i.e., $r_1 > r_2$. In the second case, referred to as the Decreasing Order, $r_1 < r_2$. (Throughout we ignore the trivial case where $r_1 = r_2$.) We show that in any solution, the customer type that values the service less receives zero utility.

Lemma 1 *If $r_1 > r_2$, $u_2 = 0$. Otherwise, if $r_1 < r_2$, $u_1 = 0$.*

4.2. Aggregate Control, Increasing Order

For the Increasing Order case with $r_1 > r_2$, by Lemma 1, $u_2 = 0$, implying $P_2 = \gamma_2(r_2 - cW_2)$. Substituting for P_2 into the high-demand type's IC constraint (4c), we have

$$u_1 \geq \gamma_1/\gamma_2(\gamma_2(r_1 - cW_2) - \gamma_2(r_2 - cW_2)) = \gamma_1(r_1 - r_2) > 0 \text{ if } n_2 > 0. \tag{5}$$

Importantly, the utility of type-1 is *independent* of the class-2 delay. This follows because type-1 have the higher demand rate than type-2. The right-hand-side of the inequality reflects the utility that type-1 can get from γ_1/γ_2 class-2 passes. Since type-1 have the higher demand rate, they would use each class-2 plan at the same rate as type-2, γ_2, and therefore incur the same aggregate delay cost per pass, $c\gamma_2 W_2$. Since this delay cost is reflected in the class-2 price, i.e., $P_2 = \gamma_2(r_2 - cW_2)$, the type-1 utility per type-2 pass is independent of the class-2 delay. Simplifying (4d) results in the problem

$$\Pi^A = \max_{n_i, u_1, W_i} \quad \sum_i n_i \gamma_i (r_i - cW_i) - n_1 u_1 \tag{6a}$$

$$\text{subject to} \quad u_1 \geq 0 \tag{6b}$$

$$u_1 \geq \gamma_1(r_1 - r_2) \text{ if } n_2 > 0 \tag{6c}$$

$$u_1 \leq \gamma_1(r_1 - cW_1) - \gamma_2(r_2 - cW_1) \text{ if } n_1 > 0 \tag{6d}$$

$$W_i \geq \frac{1}{\mu - n_i \gamma_i} \qquad \text{for } i = 1, 2 \tag{6e}$$

$$\sum_i n_i \gamma_i W_i \geq \frac{\sum_i n_i \gamma_i}{\mu - \sum_i n_i \gamma_i} \tag{6f}$$

$$0 \leq n_i \leq N_i \qquad \text{for } i = 1, 2. \tag{6g}$$

We show:

Proposition 2 *(Aggregate Control, Increasing Order) For $r_1 > r_2$, FIFO is optimal. Type 2 is served at some capacity if and only if $(N_1\gamma_1 + N_2\gamma_2)r_2 > N_1\gamma_1 r_1$. In this case there exists a finite threshold $K_2^A > K_2^F$, where K_2^F is defined in Proposition 1, such that both types are served if and only if $\mu > K_2^A$. Moreover, $u_1 = \gamma_1(r_1 - r_2)$, for $\mu > K_2^A$ and $\Pi^A < \Pi^{FI}$ for $\mu > K_2^F$.*

The major point of Proposition 2 is that under the Increasing Order, FIFO service is optimal. (Note, Masuda and Whang (2006) who study a similar model to this case *assume* FIFO service.) In the most natural case where the firm establishes sufficient capacity to serve both types of customers, i.e., $\mu > K_2^A$, we see that the firm must provide the type-1 customers with utility $u_1 = \gamma_1(r_1 - r_2) > 0$, lowering the revenue it receives. This surplus utility is independent of delay, which is key in understanding why FIFO is optimal for the Increasing Order.

Further, when both types are served, (4c) and $u_1 = \gamma_1(r_1 - r_2)$ imply $P_1 = \gamma_1(r_2 - cW)$, so

$$\frac{P_1}{\gamma_1} = \frac{P_2}{\gamma_2} = r_2 - cW.$$

That is, both types pay the *same* amount per use, which implies that charging per use is also an optimal pricing scheme. The condition in Proposition 2, $(N_1\gamma_1 + N_2\gamma_2)r_2 > N_1\gamma_1 r_1$, requires that there are sufficiently many low-marginal-value type-2 customers so that, at ample capacity, the firm can generate more revenue from all customers paying the lower rate, $r_2 - cW$, than can be generated from type-1 customers alone paying $r_1 - cW$.

4.3. Aggregate Control, Decreasing Order

We next consider the Decreasing Order case where $r_1 < r_2$, i.e., the customers with higher demand have lower marginal value per use. Our main result is that it may be optimal to prioritize the type-1 customers ($W_1 < W_2$). Whether to prioritize depends on the capacity and also on the types' total valuation of the service per unit time given by $r_i\gamma_i$. We consider two sub-cases:

- Low value rate: the total (as opposed to marginal) value per unit time of a type-1 customer is less than that of a type-2 customer ($r_1\gamma_1 \leq r_2\gamma_2$). In this case, type-1 customers are not particularly attractive customers for the firm. They neither value the service highly per use nor per unit time. However, if there are a sufficient number of them, and given sufficient capacity, the firm will serve them with strict priority, in order to raise the price they pay and reduce the surplus utility of type-2 customers. Type-2 customers would still receive positive utility ($u_2 > 0$), so that the PI revenue falls short of the FI revenue.
- High value rate: the total value per unit time of a type-1 customer is greater than that of a type-2 customer ($r_1\gamma_1 > r_2\gamma_2$). In this case, the type-1 customers are more attractive to the firm, irrespective of their number. While for low capacity, they are not served, at some intermediate capacity, it is optimal to prioritize them while possibly giving surplus utility to type-2 customers. However, at sufficiently high capacity, the firm can achieve the FI revenue with FIFO service by extracting the full value from both types.

We rewrite the Aggregate Control problem for the Decreasing Order case to simplify its analysis. By Lemma 1, $u_1 = 0$. Substituting into (4c) and simplifying implies $u_2 \le \gamma_2(r_2 - r_1)$. If type-1 customers are served, constraint (4d) implies $u_2 \ge \gamma_2(r_2 - cW_1) - \gamma_1(r_1 - cW_1)$. Combining these simplifications we can write the problem as:

$$\Pi^{\mathrm{A}} = \max_{n_i, W_i} \quad \sum_i (n_i\gamma_i(r_i - cW_i)) - u_2 n_2 \tag{7a}$$

$$\text{subject to} \quad u_2 \ge 0 \tag{7b}$$

$$u_2 \ge \gamma_2 r_2 - \gamma_1 r_1 + (\gamma_1 - \gamma_2)cW_1 \text{ if } n_1 > 0 \tag{7c}$$

$$u_2 \le \gamma_2(r_2 - r_1) \text{ if } n_2 > 0 \tag{7d}$$

$$W_i \ge \frac{1}{\mu - n_i\gamma_i} \qquad \text{for } i = 1,2 \tag{7e}$$

$$\sum_i n_i\gamma_i W_i \ge \frac{\sum_i n_i\gamma_i}{\mu - \sum_i n_i\gamma_i} \tag{7f}$$

$$0 \le n_i \le N_i \qquad \text{for } i = 1,2. \tag{7g}$$

The individual rationality constraint is given by (7b). Constraint (7c) expresses the conditional IC constraint for type-2 which is active if class 1 is offered. The multiple copy purchase constraint is given by (7d).

Definition 1. Let

$$\widetilde{W} = \frac{r_1\gamma_1 - r_2\gamma_2}{c(\gamma_1 - \gamma_2)}.$$

For a service with $W = \widetilde{W}$, $\gamma_1(r_1 - cW) = \gamma_2(r_2 - cW)$ implying both types receive the same utility per unit time. That is, $\widetilde{W}$ is a critical waiting time, dependent only on the model parameters, that determines whether the class-1 delay, W_1, requires surplus utility for the IC constraint (7c) to hold. In particular, the right hand side (RHS) of (7c) is strictly positive iff $W_1 > \widetilde{W}$. We can rewrite (7c) as

$$\frac{u_2}{c(\gamma_1 - \gamma_2)} \ge W_1 - \widetilde{W} \text{ if } n_1 > 0. \tag{8}$$

Inequality (8) is the fundamental constraint governing the solution in the Decreasing Order case. It implies that if, for a given capacity, the FIFO waiting time, W, exceeds $\widetilde{W}$, then the RHS of (8) is strictly positive, so for optimality it must be that either type-1 customers are not served ($n_1 = 0$), or if they are served, then to reduce u_2, they are served with priority ($W_1 < W_2$) or type-2 customers receive some surplus utility ($u_2 > 0$), or both.

Low value rate sub-case With both low marginal valuation for each usage ($r_1 < r_2$) and a low total value rate ($r_1\gamma_1 \le r_2\gamma_2$), $\widetilde{W} < 0$, implying when type-1 customers are served, it must be that the type-2 customers receive positive utility while the type-1 customers receive strict priority service. We show that the firm needs both sufficient capacity and a sufficient number of type-1 customers to find value in serving these customers.

Proposition 3 *(Aggregate Control, Decreasing Order, Low Value Rate) For $r_1 < r_2$ and $\gamma_1 r_1 \le \gamma_2 r_2$, type 1 is served at some capacity if and only if $(N_1+N_2)r_1\gamma_1 > N_2 r_2\gamma_2$. In this case, there exists threshold K_2^A with $K_2^F < K_2^A < \infty$, where K_2^F is defined in Proposition 1, such that $\Pi^A < \Pi^{FI}$ and the following holds for $\mu > K_2^F$:*

1. *For $\mu \le K_2^A$ only type 2 is served.*
2. *For $\mu > K_2^A$ both types are served, with type 1 receiving absolute priority and $u_2 = \gamma_2 r_2 - \gamma_1 r_1 + (\gamma_1 - \gamma_2)cW_1 > 0$.*

The condition in Proposition 3, $(N_1+N_2)r_1\gamma_1 > N_2 r_2\gamma_2$, implies that there is a sufficient number of type-1 customers so that there is more value from all customers paying the lower rate, $r_1\gamma_1$, than can be generated from the type-2 customers alone paying $r_2\gamma_2$. In this case, given sufficient capacity, we find that the type-1 customers are served with absolute priority. The firm uses price discrimination in order to deter type-2 customers from buying class-1 service. Observe, $P_1 - P_2 = (r_1 - cW_1)\gamma_1 - ((r_2 - cW_2)\gamma_2 - u_2) = c\gamma_2(W_2 - W_1) > 0$, where the inequality holds because $W_1 < W_2$ under strict priorities. Note that the price discrimination relies on delay differentiation. As $\mu \to \infty$, $W_i \to 0$, so that $P_1 - P_2 \to 0$. Note, however, that type-2 customers receive utility $u_2 = r_2\gamma_2 - r_1\gamma_1$ as $\mu \to \infty$. As a result, $\Pi^A < \Pi^{FI}$ when both classes are served.

High value rate sub-case We now consider the sub-case where $r_1\gamma_1 > r_2\gamma_2$. Here, the type-1 customers are very attractive if one considers the rate at which they generate value. We show, in contrast to the low value rate sub-case, one might not need to give a surplus or even provide priority to class 1 in order to serve them. That is, with sufficient capacity, the firm can perfectly discriminate between the customer types.

Proposition 4 *(Aggregate Control, Decreasing Order, High Value Rate) For $r_1 < r_2$, and $\gamma_1 r_1 > \gamma_2 r_2$, there exist thresholds $K_2^{A'}$, K_3^A, and K_4^A with $K_2^F \le K_2^{A'} \le K_3^A < K_4^A < \infty$, where K_2^F is defined in Proposition 1 such that:*

1. *For $\mu \le K_2^{A'}$ only type 2 is served and for $\mu > K_2^{A'}$ both types are served.*
2. *For $K_2^{A'} < \mu \le K_3^A$, type 1 is served with absolute priority, $n_1^A < n_1^{FI}$, and $u_2 = \gamma_2 r_2 - \gamma_1 r_1 + c(\gamma_1 - \gamma_2)W_1 \ge 0$.*
3. *For $K_3^A < \mu < K_4^A$, type 1 is served with some priority, $n_1^A = n_1^{FI}$, and $u_2 = 0$.*

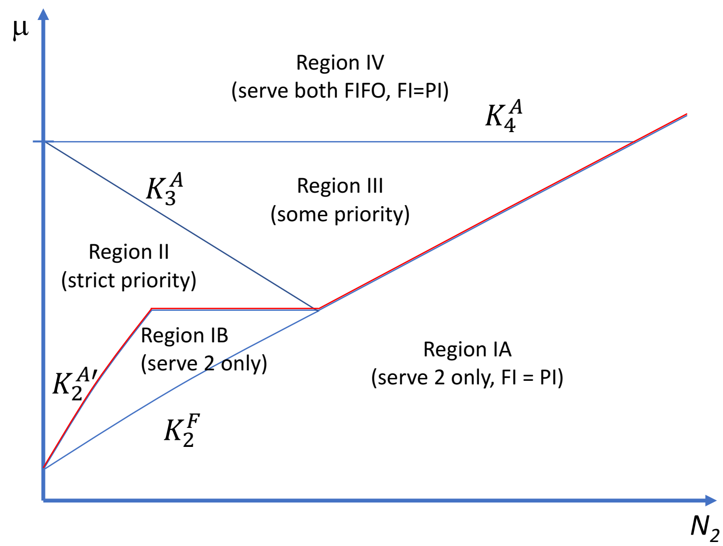

Figure 2: PI, Aggregate Control, High Value Rate Sub-Case Setting Optimal Policy as a Function of Capacity μ and type-2 market size N_2. Region boundaries are given by K_2^F, $K_2^{A'}$,K_3^A, and K_4^A, as functions of N_2.

4. *For* $\mu \geq K_4^A$*, the FI solution with FIFO is optimal,* $n_1^A = n_1^{FI}$*, and* $u_2 = 0$*.*

Moreover $\Pi^A < \Pi^{FI}$ *for* $K_2^F < \mu < K_3^A$ *and* $\Pi^A = \Pi^{FI}$ *for* $\mu \geq K_3^A$*.*

The four cases given in Proposition 4 describe which customers are served, the priority given, and the utility received. To understand what holds, consider the critical waiting time, $\widetilde{W}$ from Definition 1. Here $\widetilde{W} > 0$ so that the fundamental constraint (8) implies if type-1 customers are to be served and if the waiting time under FIFO, $W > \widetilde{W}$ for a given (low) capacity μ, then they must be served with some priority, and possibly $u_2 > 0$. On the other hand, with high capacity, (8) is satisfied with FIFO service, something that could not happen in the low value rate sub-case.

Figure 2 depicts the thresholds as functions of N_2 and μ. (The figure is shown for the case of large $N_1\gamma_1$; slightly different threshold forms hold for smaller values–see the proof of Proposition 4.) In region I, only type 2 is served. When $K_2^F < \mu < K_2^{A'}$ (depicted as Region 1B), both classes are served in the FI solution, but $n_1 = 0$ in the PI solution.) In region II, type 1 is served with priority and class-2 may receive polisitve utility depending on N_2 and μ. In region III, all type-1 customers served in the Full Information case are served ($n_1 = n_1^{FI}$), but with priority; class 2 customers receive no utility. In region IV, all type-1 and type-2 customers are served FIFO as in the Full Information case.

Proposition 4 shows, in contrast to the low value rate case, that type-1 customers need not be served with priority. Because $\widetilde{W} > 0$, when the capacity exceeds K_3^A (and K_2^F), the IC constraint is no longer binding and only the IR constraints are binding. As such, the firm can perfectly discriminate between the two customer types and achieve the Full Information revenue. Effectively, with low waiting times, the firm can offer two FUT tariffs, one for use at rate γ_1 and one at rate γ_2.

5. Private Information Setting - Individual Control Model

In this section we assume, as in Section 4, that the firm cannot distinguish between the customer types before their first purchase of the season. However, in the Individual Control Model the provider can track the usage of individual customers and limit each customer to buying a single plan. Therefore, the firm can limit individual usage rates through the tariff structure, so it optimizes over x_i. If the firm does not offer class-i service, $x_i = 0$. The firm does not control the number of service plans, so $n_i = N_i$. As we show below, the results are similar to the Aggregate Control Model. Most importantly, in the Decreasing Order case, it is optimal to prioritize the high-demand, low-marginal-value type 1 customers for a significant capacity range.

As in the Aggregate Control Model, we restrict attention to FUT tariffs for the analysis (see Remark 2). However, in the Individual Control Model the provider optimizes over both parameters of the class-i tariff, the maximum usage rate x_i and the flat fee P_i.

For the Individual Control model, the objective is to maximize $\sum_i N_i P_i = \sum_i N_i(x_i(r_i - cW_i) - u_i)$ over x_i, u_i, and W_i. The IR constraints remain as $u_i \geq 0$. Incentive compatibility requires that u_i exceed the maximum expected utility of a type-i customer who purchases a class-j plan (if class-j is offered) with delay W_j, fee P_j, and maximum usage rate x_j. That is,

$$u_i \geq \min(\gamma_i, x_j)(r_i - cW_j) - P_j$$

for $i \neq j$. For type-1 customers since $\gamma_1 > \gamma_2 \geq x_2$, this implies

$$u_1 \geq x_2(r_1 - cW_2) - P_2. \tag{9}$$

For type-2 customers

$$u_2 \geq \begin{cases} x_1(r_2 - cW_1) - P_1 & \text{if } x_1 \leq \gamma_2 \\ \gamma_2(r_2 - cW_1) - P_1 & \text{if } \gamma_2 \leq x_1 \leq \gamma_1 \end{cases} \tag{10}$$

As before we specialize the Private Information model to the Increasing and Decreasing Orders.

5.1. Individual Control Model, Increasing Order

In the Increasing Order, $r_1 > r_2$, and by Lemma 1, we have $u_2 = 0$ implying $P_2 = x_2(r_2 - cW_2)$. Then substituting for $P_1 = x_1(r_1 - cW_1) - u_1$ and $u_2 = 0$ in the objective function, (9) and (10), the firm's Private Information problem for the Individual Control model is

$$\Pi^I = \max_{u_1 x_i, W_i} \quad \sum_i N_i x_i(r_i - cW_i) - N_1 u_1 \tag{11a}$$

$$\text{subject to} \quad u_1 \geq 0, \tag{11b}$$

$$u_1 \geq x_2(r_1 - r_2), \tag{11c}$$

$$u_1 \leq \begin{cases} x_1(r_1 - r_2), & \text{if } x_1 \leq \gamma_2, \\ x_1 r_1 - \gamma_2 r_2 + (\gamma_2 - x_1) c W_1, & \text{if } \gamma_2 \leq x_1 \leq \gamma_1, \end{cases} \tag{11d}$$

$$W_i \geq \frac{1}{\mu - N_i x_i} \qquad \text{for } i = 1, 2, \tag{11e}$$

$$\sum_i N_i x_i W_i \geq \frac{\sum_i N_i x_i}{\mu - \sum_i N_i x_i}, \tag{11f}$$

$$0 \leq x_i \leq \gamma_i \qquad \text{for } i = 1, 2. \tag{11g}$$

The problem is similar to (6) with the decision variables x_i replacing the constants γ_i as appropriate. In particular, the utility of the high-marginal-value type 1 is independent of delay in both problems, compare (11c) with (6c). Therefore the result is similar to Proposition 2:

Proposition 5 *(Individual Control, Increasing Order) For $r_1 > r_2$ FIFO is optimal. Type 2 is served at some capacity if and only if $N_2 r_2 > N_1 (r_1 - r_2)$. In this case there exists a finite threshold $K_2^I > K_2^F$, where K_2^F is defined in Proposition 1, such that both types are served if and only if $\mu > K_2^I$. Moreover, $x_1^I = \gamma_1$ and $u_1 = x_2^I (r_1 - r_2) > 0$ for $\mu > K_2^I$, and $\Pi^I < \Pi^{FI}$ for $\mu > K_2^F$.*

Most importantly, FIFO is optimal at all capacity levels, like in the Increasing Order case under Aggregate Control. The driver of this result is the same in both models: The utility of the high-marginal-value type 1 is independent of delay, and this follows because type-1 has a higher demand rate than the usage rate of type 2. To be specific, consider how the type-1 IC constraint (15c) follows from (13). Rewrite (13) and substitute $P_2 = x_2 (r_2 - W_2)$ to obtain

$$u_1 \geq \min(x_2, \gamma_1)(r_1 - cW_2) - P_2 = x_2 (r_1 - cW_2) - x_2 (r_2 - cW_2).$$

The type-1 utility is bounded below by their maximum utility in class-2. This utility equals the difference between type 1's (total) net value rate from class-2, $\min(x_2, \gamma_1)(r_1 - cW_2)$, and type 2's net value rate from class-2, $x_2 (r_2 - W_2) = P_2$. Since the demand rate of type-1 customers exceeds the usage rate of type-2, both types would use a class-2 plan at its maximum rate, x_2, and therefore incur the same aggregate delay cost in that class, $x_2 c W_2$. Therefore, the difference between the types' *net* value rates in class 2 reduces to the difference in their total *gross* value rates, so that $u_1 \geq x_2 (r_1 - r_2)$ as shown in (15c). (In the Aggregate Control Model the same logic applies to each copy of a class-2 plan considered by type-1.)

Note $u_1 = x_2^I (r_1 - r_2)$ in Proposition 5, whereas $u_1 = \gamma_1 (r_1 - r_2)$ in Proposition 2. Under Individual Control type-1 customers receive lower utility compared to Aggregate Control (recall $x_2^I \leq \gamma_2 < \gamma_1$), because the provider can restrict the class-2 usage rate. Therefore, class-1 offers a higher maximum usage-rate than class-2, making class-1 more attractive to the high-demand type 1. As a result, under Individual Control type-1 customers pay a *higher* price per use than type-2:

$$\frac{P_1}{x_1} = \frac{x_1(r_1 - cW) - u_1}{x_1} = \frac{x_1(r_1 - cW) - x_2(r_1 - r_2)}{x_1} > r_2 - cW,$$
$$\frac{P_2}{x_2} = \frac{x_2(r_2 - cW)}{x_2} = r_2 - cW.$$

In other words, the provider can generate more revenue by selling multi-use plans than by charging per use. (In contrast, recall under Aggregate Control charging per use is also optimal.)

5.2. Individual Control, Decreasing Order

In the Decreasing Order, $r_1 \le r_2$, and by Lemma 1, we have $u_1 = 0$ implying $P_1 = x_1(r_1 - cW_1)$. Making similar substitutions as in (11) we have

$$\Pi^I = \max_{x_i, u_2, W_i} \quad \sum_i N_i x_i (r_i - cW_i) - N_2 u_2 \tag{12a}$$
$$\text{subject to} \quad u_2 \ge 0 \tag{12b}$$
$$u_2 \le x_2(r_2 - r_1), \tag{12c}$$
$$u_2 \ge \begin{cases} x_1(r_2 - r_1) & \text{if } x_1 \le \gamma_2, \\ \gamma_2 r_2 - x_1 r_1 + (x_1 - \gamma_2)cW_1 & \text{if } \gamma_2 \le x_1 \le \gamma_1, \end{cases} \tag{12d}$$
$$W_i \ge \frac{1}{\mu - N_i x_i} \qquad \text{for } i = 1, 2, \tag{12e}$$
$$\sum_i N_i x_i W_i \ge \frac{\sum_i N_i x_i}{\mu - \sum_i N_i x_i}, \tag{12f}$$
$$0 \le x_i \le \gamma_i \qquad \text{for } i = 1, 2. \tag{12g}$$

Problem (12) is similar to (7), with one key difference. By the IC constraint (12d), the utility of the high-marginal-value type 2, u_2, is delay-independent if $x_1 \le \gamma_2$. (This case does not arise under Aggregate Control since $x_1 = \gamma_1$, see (7c)). For $x_1 \le \gamma_2$ the logic of the Increasing Order case applies: Since the type-2 demand rate exceeds the class-1 usage limit, type-2 consider class-1 with the same usage rate as type-1, so both types incur the same aggregate delay cost in that class, and the difference between their class-1 net value rates equals $x_1(r_2 - r_1)$. Therefore, if $0 < x_1^I \le \gamma_2$ at optimality, FIFO is optimal and $u_2 = x_1^I(r_2 - r_1) > 0$, see Part 2 of Propositions 6 and 7.

If $x_1 > \gamma_2$, the type-2 utility may be delay-dependent, by the same logic as in the Decreasing Order case under Aggregate Control: Since the type-2 demand rate is lower than the class-1 usage limit, type-2 consider class-1 with lower usage rate than type-1, and therefore incur a *lower* aggregate delay cost in that class. Therefore, prioritizing type-1 customers may be optimal. In this case the solution to (16) again depends on whether type 1 have the lower value rate ($r_1\gamma_1 \le r_2\gamma_2$) or the higher value rate ($r_1\gamma_1 > r_2\gamma_2$). In particular, we can rewrite (12d) for $\gamma_2 \le x_1 \le \gamma_1$ as

$$\frac{u_2}{c\,(x_1-\gamma_2)} \ge W_1 - \widetilde{W}\,(x_1) \tag{13}$$

where $\widetilde{W}(x_1) = (r_1x_1 - r_2\gamma_2)/(c(x_1-\gamma_2))$. In the low value rate case, $\widetilde{W}(x_1) < 0$ so that when $x_1 > \gamma_2$, for (13) to hold for the smallest u_2 would imply decreasing the delay of class-1, as much as possible, i.e., giving strict priority. In the high value rate case for large enough x_1, $\widetilde{W}(x_1) > 0$ implying that priority for class-1 may not be necessary for (13) to hold.

Low value rate sub-case In the low value rate sub-case, the right hand side of (12d) is always positive (when class-1 service is offered). For $x_1 > \gamma_2$ this holds because type-2 customers have both the higher (gross) value rate ($\gamma_2 r_2 - x_1 r_1 \ge 0$) and the lower delay cost ($c(x_1-\gamma_2)W_1 > 0$) in class-1, so that type-2 derive more value from class-1 than type-1. This implies that the type-2 customers receive positive surplus when class 1 is served, and if $x_1 > \gamma_2$, giving priority to class 1 is optimal. We have the following proposition:

Proposition 6 *(Individual Control, Decreasing Order, Low Value Rate) For $r_1 < r_2$ and $\gamma_1 r_1 \le \gamma_2 r_2$, type 1 is served at some capacity if and only if $(N_1+N_2)\,r_1\gamma_1 > N_2 r_2\gamma_2$. In this case, there exist thresholds K_2^I and K_3^I with $K_2^F < K_2^I \le K_3^I < \infty$, where K_2^F is defined in Proposition 1, such that $\Pi^I < \Pi^{FI}$ and the following holds for $\mu > K_2^F$:*

1. *For $\mu \le K_2^I$ only type 2 is served, and for $\mu > K_2^I$ both types are served.*
2. *For $K_2^I < \mu \le K_3^I$, FIFO is optimal, type 1 has the lower usage rate, i.e., $x_1^I < \gamma_2 = x_2^I$, and $u_2 = x_1^I\,(r_2 - r_1) > 0$.*
3. *For $\mu > K_3^I$, type 1 is served with absolute priority and higher usage rate, i.e., $x_1^I > \gamma_2$, and $u_2 = \gamma_2 r_2 - x_1^I r_1 + c\,(x_1^I - \gamma_2)\,W_1 > 0$.*

The key result, Part 3 of Proposition 6, is consistent with Proposition 3 for the Aggregate Control Model: Giving absolute priority to type-1 customers is optimal for all capacity levels above a finite threshold K_3^I, whenever it is optimal to serve them with higher usage rate than the demand rate of the high-value type-2, i.e., $x_1^I > \gamma_2$. In terms of pricing, note that for $\mu > K_3^I$ price discrimination *requires* delay differentiation. Noting that $u_2 = x_2\,(r_2 - cW_2) - P_2 = \gamma_2\,(r_2 - cW_1) - P_1$, where the second equality holds since type-2 are indifferent between the two classes, we have

$$P_1 - P_2 = \gamma_2\,(r_2 - cW_1) - x_2\,(r_2 - cW_2) > 0.$$

The inequality holds since $\gamma_2 \ge x_2$ and $W_1 < W_2$ under strict priorities. However, under ample capacity, as $\mu \to \infty$, we have $W_i \to 0$ and $x_i^I \to \gamma_i$, so that $P_1 - P_2 \to 0$, that is, the provider loses the ability to price discriminate.

The main difference to Proposition 3 is Part 2 of Proposition 6. Under Individual Control, serving both types FIFO may be optimal in some intermediate capacity range, i.e., for $K_2^I < \mu \leq K_3^I$. This capacity range exists (i.e., $K_2^I < K_3^I$) if the number of type-1 customers, N_1, is sufficiently large. In this case, the firm finds it profitable to serve them with low usage rate, and the solution mirrors the Increasing Order case of Proposition 5: FIFO is optimal, as the usage rate of the low-marginal value type 1 customers is smaller than the demand rate of the high-value type-2, i.e., $x_1^I < \gamma_2$. That is, under moderate capacity restricting the class-1 usage rate allows the firm to profit from serving type-1 by keeping class 1 relatively unattractive for type-2, so u_2 is low.

High value rate sub-case In this case, the key result is that priorities are *always* optimal in some intermediate capacity range, as shown by Proposition 7. Indeed, this optimality result is stronger than under Aggregate Control as discussed below.

Proposition 7 *(Individual Control, Decreasing Order, High Value Rate) For $r_1 < r_2$ and $\gamma_1 r_1 > \gamma_2 r_2$, there exist thresholds K_2^I, K_3^I, K_4^I and K_5^I, with $K_2^F < K_2^I \leq K_3^I < K_4^I < K_5^I < \infty$, where K_2^F is defined in Proposition 1, such that the following holds:*

1. *For $\mu \leq K_2^I$ only type 2 is served and for $\mu > K_2^I$ both types are served.*
2. *For $K_2^I < \mu \leq K_3^I$ FIFO is optimal, type 1 has the lower usage rate, i.e., $x_1^I < \gamma_2 = x_2^I$, and $u_2 = x_1^I (r_2 - r_1) > 0$.*
3. *For $K_3^I < \mu \leq K_4^I$, type 1 is served with absolute priority and higher usage rate, i.e., $x_1^I > \gamma_2$, and $u_2 = \gamma_2 r_2 - x_1^I r_1 + c(x_1^I - \gamma_2) W_1 \geq 0$.*
4. *For $K_4^I < \mu < K_5^I$, type 1 is served with some priority and higher usage rate, i.e., $x_1^I > \gamma_2$, $\mathbf{x}^I = \mathbf{x}^{FI}$, and $u_2 = 0$.*
5. *For $\mu \geq K_5^I$ the FI solution with FIFO is optimal, $\mathbf{x}^I = \mathbf{x}^{FI}$ with $x_1^I > \gamma_2$, and $u_2 = 0$.*

Moreover, $\Pi^I < \Pi^{FI}$ for $K_2^F < \mu < K_4^I$ and $\Pi^I = \Pi^{FI}$ for $\mu \geq K_4^I$.

For capacity levels below the threshold K_4^I the results and the underlying logic essentially match those of the low-value rate sub-case. Compare Parts 1-3 of Proposition 7 with Proposition 6.

However, for capacity larger than K_4^I, the firm can achieve the optimal FI revenue by extracting the full value from both types at the FI usage rates, i.e., $x_i^I = x_i^{FI}$, and $x_1^I > \gamma_2$. Notably, though type-2 customers have the higher marginal value they also get zero utility ($u_2 = 0$), because at large enough capacity they derive less *total value* from class-1 than type-1 customers under strict priority. That is, the right-hand side of (12d) is negative: $\gamma_2 r_2 - x_1^F r_1 + (x_1^F - \gamma_2) cW_1 < 0$, or equivalently $W_1 - \widetilde{W}(x_1^F) < 0$, because type-2 have the lower value rate ($\gamma_2 r_2 - x_1^F r_2 < 0$) and the positive delay cost difference $(x_1^F - \gamma_2) cW_1$ is too small to offset this difference. At moderately high capacity, i.e., for $K_4^I < \mu < K_5^I$, extracting all type-2 utility still requires giving type-1 customers some priority (Part 4 of Proposition 7), whereas for $\mu \geq K_5^I$ FIFO is optimal (Part 5 of Proposition 7).

In terms of pricing, in contrast to the low-value rate sub-case where congestion-based delay differentiation is *required* for price discrimination, here congestion is *detrimental* to price discrimination. When delays are low, the provider can eliminate the surplus of both types, by charging type-1 the higher total fee and type-2 the higher per-use fee. However, high delays make class-1 relatively more attractive to the low-demand type-2 customers, so that $u_2 > 0$.

Finally, we observe that the result on the optimality of priorities is stronger under Individual Control compared to Aggregate Control (see Proposition 4). Whereas priorities are *always* optimal in some capacity range under Individual Control, this is not the case under Aggregate Control. Notably, if the number of high-marginal value type-2 customers, N_2, is sufficiently large, then under Aggregate Control FIFO is optimal and the FI solution matches the PI solution at all capacity levels. This discrepancy follows because in the Aggregate Control Model, a given total type-1 usage rate, $\lambda_1 = n_1 x_1$, is allocated to a minimum number of customers with maximum usage rate, i.e., $x_1 = \gamma_1$ and $n_1 = \lambda_1/\gamma_1$. As a result, type-2 customers always have the lower value rate in either class, that is $\gamma_2 r_2 < \gamma_1 r_1$, so from this perspective class-1 is maximally unattractive for type-2 customers. Therefore, the provider only needs to give type 2 positive utility if the class-1 delay is sufficiently high to offset this value deficit. However, for sufficiently large N_2 the capacity threshold where it becomes profitable to open class 1 for service is so large that the class-1 delay is relatively insignificant, so that class-1 is too expensive for type-2 customers. In contrast, under Individual Control, the load λ_1 is allocated by serving all type-1 customers with the smallest corresponding usage rate, i.e., $n_1 = N_1$ and $x_1 = \lambda_1/N_1$. Therefore, the type-1 usage rate gradually increases from $x_1 > \gamma_2$ to $x_1 = \gamma_1$ as capacity increases.

6. Discussion and Implications

6.1. General Decreasing Marginal Value Functions

Our analysis assumes the *simplest* demand model that yields our key results on the optimality of priority service in settings with heterogeneous demand rates. In this section, we show these results and the underlying intuition extend naturally under general decreasing marginal value functions. We assume the provider can limit each customer to purchasing a single plan, as in the Individual Control Model. We provide *sufficient* conditions for priority service to be optimal by considering the properties of the PI solution under restriction to FIFO service. To be clear, we make no attempt to solve the PI problem; we simply *assume* certain PI solution properties hold under restriction to FIFO.

Let $r_i(x_i)$ denote a type-i customer's marginal value, and $R_i(x_i)$ their total value rate, as a function of their usage rate x_i, where $R_i(x_i)$ is strictly increasing and strictly concave. For simplicity we assume $R_i(x_i)$ is differentiable, so $R_i'(x_i) = r_i(x_i) > 0$, and $\lim_{x_i \to \infty} r_i(x_i) = 0$. In this notation our original model with constant marginal valuations has $r_i(x_i) = r_i 1\{0 \le x_i \le \gamma_i\}$ and $R_i(x_i) = r_i \min(x_i, \gamma_i)$.

We note this model also captures situations where the valuations of a customer's service opportunities are i.i.d. random variables whose realizations are observed prior to each use, e.g., tied to the weather when a skiing opportunity arises. Let service valuations be i.i.d. draws from a distribution with c.d.f. F_i. Let $\overline{F}_i = 1 - F_i$, and write r_i for the marginal (or threshold) value realization this customer requires to go skiing. Then the usage rate as a function of the marginal value equals $x_i(r_i) = \gamma_i \overline{F}(x_i)$, where γ_i has the same interpretation as in our model, and conversely, $r_i(x_i) = \overline{F}_i^{-1}(x_i/\gamma_i)$ is the marginal value function.

The PI problem formulation parallels the one in Section 5. We focus again on FUT tariffs (recall Remark 2) with flat rate P_i and usage limit x_i. Setting $W = 1/(\mu - x_1 N_1 - x_2 N_2)$ for FIFO, the objective is to maximize $\sum_i N_i P_i = \sum_i N_i (R_i(x_i) - x_i cW - u_i)$ over x_i and u_i. IR requires

$$u_i = \max_{y \le x_i} R_i(y) - ycW - P_i \ge 0. \tag{14}$$

IC requires that u_i exceed the maximum utility of a type-i in class $j \ne i$ with maximum rate x_j:

$$u_i \ge \max_{y \le x_j} R_i(y) - ycW - P_j. \tag{15}$$

Lemma 2 presents sufficient conditions for optimality of priority service. To parallel our results where priorities are optimal, we suppose that type 1 has zero utility. Observe this is w.l.o.g. as we make no assumptions on how the types differ.

Lemma 2 *For the optimal PI solution under FIFO service, let x_i be the usage rate, P_i the price, and u_i the utility for $i = 1, 2$, and let W be the delay. Suppose both types are served: $x_1 > 0$ and $x_2 > 0$. Assume w.l.o.g. that $u_1 = 0$, so $P_1 = R_1(x_1) - x_1 cW$.*
Let $x_{21} = \arg\max_{y \le x_1} R_2(y) - ycW$ denote the optimal type-2 usage rate in class 1.
Then giving some priority to type-1 generates more revenue if the following two conditions hold:

1. *For class-1 type-2 has the lower optimal usage rate than type-1, i.e., $x_{21} < x_1$. This holds if*

$$r_2(x_1) < cW \le r_1(x_1). \tag{16}$$

In this case type-2 also has the lower usage rate in class 2, i.e., $x_2 < x_1$.

2. *For class-1 type-2 receive the higher net value rate than type-1, so they have positive utility*

$$u_2 = R_2(x_{12}) - x_{21} cW - P_1 = R_2(x_{12}) - R_1(x_1) + c(x_1 - x_{21}) W > 0. \tag{17}$$

We discuss Lemma 2 and revisit our results in its context. In sum, priority service is optimal if, at the FIFO solution, one type has the lower usage rate and aggregate delay cost in the alternate class, i.e., (16) holds, and also the higher net value rate in that class, i.e., (17) holds. Under these conditions the utility of this type is positive and delay-dependent. Together (16) and (17) require for this type both lower marginal value at higher usage and higher total value from lower usage.

With respect to (16), in our model with $\gamma_1 > \gamma_2$, lower marginal value at higher usage can only hold for type-2 and under Decreasing Order ($r_1 < r_2$). Specifically, for $\gamma_2 < x_1$ we have $r_2(x_1) = 0 < r_1(x_1) = r_1$. Otherwise, under Decreasing Order for $x_1 \le \gamma_2$, condition (16) is violated, so both types have the same usage rate and delay cost in class-1, and optimal revenue is achieved by serving customers FIFO, consistent with our results. Similarly, the Increasing Order ($r_1 > r_2$), where type-1 has positive utility, violates condition (16) with the roles of the types reversed, so FIFO is optimal.

When (16) holds, as in our model under Decreasing Order with $\gamma_2 < x_1$, the solution further depends on (17). In the low value rate sub-case ($r_2\gamma_2 \ge r_1\gamma_1$) strict priority service is always optimal when there is sufficient capacity to serve both types. In this case type-2 value class-1 more than type-1 at every delay, because they have not only lower aggregate delay costs but also the higher value rate, so $R_2(x_{12}) - R_1(x_1) > 0$ in terms of (17).

The high value rate sub-case ($r_2\gamma_2 < r_1\gamma_1$) is an intermediate case between the low value rate sub-case, and the Increasing Order case. Here type-2 have the higher marginal value at lower usage, but the high frequency of use of type-1 customers implies they derive a higher value rate at higher usage. Connecting to (17) this case corresponds to $\lim_{x_1\to\infty} R_1(x_1) > \lim_{x_2\to\infty} R_2(x_2)$, so at large capacity, type-2 customers receive lower value from class-1 service. Therefore type-2 do not necessarily receive surplus utility, nor do they necessarily receive priority service. At lower capacity the firm may choose to serve type-1 FIFO, at intermediate capacity it will serve type-1 with strict or partial priority, and for high capacity the firm will simply use the FI solution with FIFO service.

Note also that while our analysis assumes identical sensitivity to waiting for both types, the preceding discussion implies that the firm can potentially benefit from prioritizing type-1 customers even if their delay cost per use is lower than that of the type-2 customers, so long as their aggregate delay cost exceeds that of type-2 customers.

6.2. Price Discrimination: The Interplay of Service Policy and Tariff Structure

Our analysis and results also generate some insights on how the ability to price discriminate depends on the interplay of the service policy and the tariff structure. From this perspective, our paper bridges two literatures on price discrimination. The literature on congestion pricing largely focuses on delay-based service differentiation as a tool for price discrimination in settings with unit demand and heterogeneous delay costs. The literature on quantity-based non-linear pricing focuses by definition on price discrimination in settings with heterogeneous demand rates, and largely ignores quality/service differentiation as an additional discrimination tool.

In our model both the priority policy and the tariff structure play an important role for price discrimination. Whereas the paper emphasizes the role of the priority policy, we briefly summarize some observations concerning the tariff structure.

First, in our model with a common waiting cost, priorities cannot generate more revenue than FIFO service under per-use pricing. If the firm offers two priority classes, then both must offer the same full price (sum of price plus delay cost). Furthermore, when customers have the same waiting cost, the firm cannot gain from reducing the delay of one class and increasing its price, because this revenue gain is exactly offset by the implied delay increase and price drop of the other class.

Second, under multi-use demand it is well known from classic studies without service differentiation that some form of bundling typically dominates per-use pricing. Similarly, in our analysis offering at least one of the types a multi-use tariff outperforms a uniform per-use price, also when FIFO is optimal. One exception is the Increasing Ordering case under Aggregate Control, where the firm cannot limit usage by definition, so cannot prevent the high-marginal-value/high-demand type 1 customers from meeting all their demand at the lowest offered price. However, as discussed in Section 5 the story changes under Individual Control, where the firm can charge type-1 customers more for high usage, by limiting their usage in class-2 targeted to low-demand customers.

Third, whereas we perform the analysis assuming FUT tariffs for analytic simplicity, there is a range of optimal pricing schemes. The tariff menu should typically be designed such that the low-marginal-value/high-demand type-1 customers purchase a subscription, e.g., a season pass. In cases where type-1 are prioritized, this means bundling is necessary for the effectiveness of priority service. Charging type-1 a per-use fee is not effective to deter high-marginal-value type-2 customers. On the other hand, the high-marginal-value customers can also be offered a per-use fee, or a two part-tariff including a per-use fee under Aggregate Control. Under Individual Control a contract with usage limit may be offered for both classes.

Finally, our results also shed light on how the effectiveness of delay-based price discrimination depends on the demand characteristics and the capacity. In the Increasing Order case, FIFO is optimal regardless of capacity. In the Decreasing Order, low value rate sub-case, even with bundling price discrimination requires delay-based differentiation, whereas in the high value rate sub-case congestion is detrimental to price discrimination.

7. The Value of Priority vs. FIFO Service

In this section we numerically compare the revenue received under the Full Information and Private Information solutions to demonstrate the relative value of information and prioritization for a few

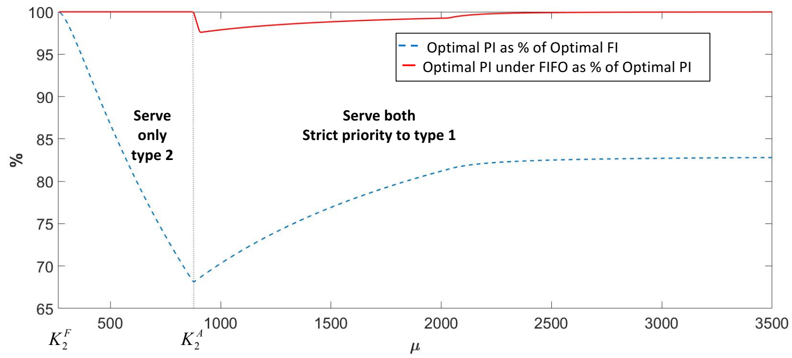

Figure 3: Aggregate Control, Decreasing order, Low Value Rate, with many type-1 customers $N_1 = 150$, $N_2 = 50$. Here $r_1 = 1$, $r_2 = 5$, $\gamma_1 = 11$, $\gamma_2 = 4$, and $c = 15$.

examples. We focus on the decreasing order case ($r_1 < r_2$). We present the optimal PI solution value as a percentage of the optimal FI solution value. We also present the optimal PI solution value when restricted to FIFO service as a percentage of the optimal PI solution value. That is, to evaluate the benefit of prioritization, we compare the PI solution to a suboptimal policy for the private information case where service is restricted to be FIFO. For all of the figures we present results for $\mu > K_2^F$ (recall Remark 1), noting K_2^F, as given by Proposition 1, depends on r_1, N_2, γ_2 and c.

Aggregate Control, Decreasing Order Low Value Rate Sub-case For this case $r_1 < r_2$ and $r_1\gamma_1 < r_2\gamma_2$. We let $r_1 = 1$, $r_2 = 5$, $\gamma_1 = 11$, $\gamma_2 = 4$, $c = 15$, $N_1 = 150$ and $N_2 = 50$. The relative value of the PI solution and the FIFO service solution are presented in Figure 3. As $(N_1 + N_2)r_1\gamma_1 > N_2 r_2\gamma_2$, there are sufficient type-1 customers so that offering class-1 service with absolute priority is beneficial to the firm. That is, Proposition 3 implies $n_1^A > 0$ when there is sufficient capacity. We find that when capacity exceeds $K_2^A \approx 870$, $n_1^A > 0$. The type-1 customers are prioritized and the type-2 customers receive a positive consumer surplus. With sufficient capacity the waiting time tends to zero and, as discussed, $P_1 - P_2 \to 0$. We observe the type-2 customers receive a substantial utility of $r_2\gamma_2 - r_1\gamma_1 = 20 - 11 = 9$ or a 45% of their valuation. This translates into a significant difference between the FI and PI solution. On the other hand, there is little difference (1%–2%) between the profit of the PI solution and that of the suboptimal solution that uses FIFO. Prioritizing the type-1 customers is of little benefit in this example.

Aggregate Control, Decreasing Order High Value Rate Sub-case For this case $r_1 < r_2$ and $r_1\gamma_1 > r_2\gamma_2$. We let $r_1 = 0.5$, $r_2 = 0.8$, $\gamma_1 = 5$, $\gamma_2 = 3$, $c = 50$, $N_1 = N_2 = 100$. In this case, the firm always prefers to serve the type-1 customers given sufficient capacity. Figure 4 illustrates that the PI solution value has a the small loss compared with the FI solution, and a large gain over the suboptimal policy restricting the solution to FIFO service. Following Proposition 4, we

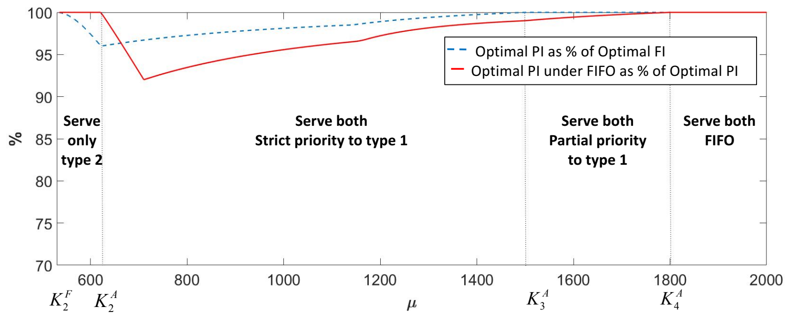

Figure 4: Aggregate Control, Decreasing order, High Value Rate, with $N_1 = 100$, $N_2 = 100$. Here $r_1 = 0.5$, $r_2 = 0.8$, $\gamma_1 = 5$, $\gamma_2 = 3$, and $c = 50$.

find $K_2^F = 530$, $K_2^{A'} \approx 625$. We find for $K_2^F < \mu < K_2^{A'}$ (corresponding to Region IB in Figure 2) that the PI solution performs increasingly worse, as the FI solution serves both customer types on this interval, with the PI solution value approximately 96% of the FI value at $K_2^{A'}$. Then for $K_2^{A'} \le \mu < K_3^A = 1500$ (Region II), an increasing number of type-1 customers are served with strict priority. For $K_3^A \le \mu < K_4^A = 1800$ (Region III), both types of customers are served fully, but class-1 service receives some priority. For $\mu \ge K_4^A$ service is FIFO (Region IV). We observe that the relative value of the PI solution to the FI increases with μ (for $\mu > K_2^{A'}$). On the other hand, under the restriction to FIFO service in the suboptimal solution, we find that fewer type-1 customers are served, lowering the revenue significantly, relative to the PI solution.

Individual Control, Decreasing Order High Value Rate Sub-case For this case $r_1 < r_2$ and $r_1\gamma_1 > r_2\gamma_2$. As above we let $r_1 = 0.5$, $r_2 = 0.8$, $\gamma_1 = 5$, $\gamma_2 = 3$, $c = 50$, $N_1 = N_2 = 100$. Now, however, the firm operating under Individual Control sets the values of x_1 and x_2. We observe in Figure 5 that the PI solution performs significantly worse than the Aggregate Control case for $\mu <\sim 1100$ for this example. Here, the relatively high number of type-1 customers and their relatively low marginal marginal value, following the logic of Proposition 7, imply that the firm must restrict the rate x_1 when they are served FIFO (for $K_2^I < \mu \le K_3^I$). This lowers the revenue from the customers compared with the FI solution. We find that as capacity increases, over the range $K_3^I < \mu \le K_4^I$, that x_1 is still significantly lower than in the FI solution, so that the PI solution does not perform well. The implication is that Individual Control may not always provide a benefit over Aggregate Control. We also observe that the restriction to FIFO in the PI solution has a similar effect as in the Aggregate Control model, with a significant reduction of the reveneu over the unrestricted case.

8. Conclusions

Priority queues have long been studied as a means of discriminating between customers that differ in their cost of waiting. In this paper, we assume a common cost of waiting, and still demonstrate

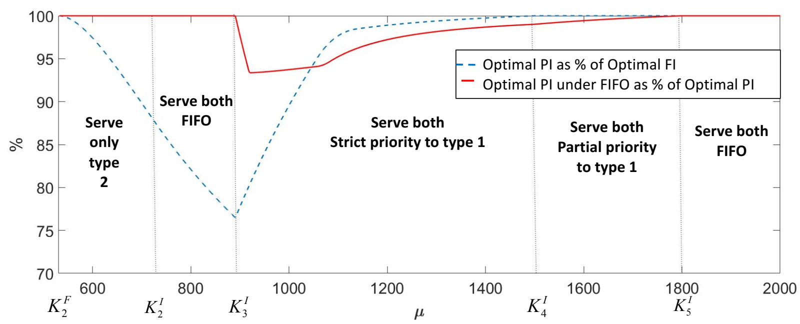

Figure 5: Individual Control, Decreasing order, High Value Rate, with $N_1 = 100$, $N_2 = 100$. Here $r_1 = 0.5$, $r_2 = 0.8$, $\gamma_1 = 5$, $\gamma_2 = 3$, and $c = 50$.

that priority queues may be revenue maximizing. Our insight is that if customers differ in their demand rate and marginal valuation of a service, prioritization provides a means of encouraging high-frequency, low marginal value customers to pay a little more for the service while allowing the firm to reduce the price for lower frequency customers. What may be thought of as a benefit or privilege provided to loyal, frequent-use customers, is demonstrated to be a tool that allows the firm to improve profitability. Moreover, our analysis implies that prioritizing customers with higher demand rate and lower marginal value may also be optimal if they are *more* patient than their low-frequency counterparts.

Using our ski resort example from the introduction, the season pass that is sold to the locals may be accompanied by admission to priority queues at the lifts and early access to the mountain, not just as a perk, but as a means of raising the price of the pass and the revenue of the firm. The higher price discourages the aways from purchasing the season pass. We see that in the low value rate sub-case (the locals are ‘hobbyists’), if the ski resort chooses to sell to them, they receive priority under an optimal policy. The numerical results indicate, however, that the gain over a suboptimal FIFO policy may be small in this sub-case. In the high value rate sub-case (the locals are ‘enthusiasts’), depending on the capacity, the firm may need to prioritize them to optimize its revenues. Here we show that the value of doing so is high vis-a-vis the suboptimal policy.

In either sub-case, when the locals are prioritized, the price of the daily pass sold to the aways is discounted so that they see a consumer surplus. Effectively, the locals subsidize the price paid by the aways, and the total revenue generated grows. Of course, this result depends on the demand and capacity of the firm. With too few locals (in the low value rate sub-case) or too little capacity, only the high-marginal value aways are served. For the high value rate sub-case, if there is high capacity, all customers are served FIFO, while for intermediate capacity the enthusiasts would be prioritized. That is, congestion in the system leads to the value derived from priorities. Interestingly, this would imply that if demand is not uniform over the season, one might expect to see priority given to locals, if in fact they are enthusiasts, precisely when the demand from aways increase such as during the Christmas break or President's day long weekend.

A general guideline that emerges from our results is that providers facing customers who plan for multiple demand use should also consider customers' value functions and usage rates, not just their impatience, in designing price-service policies.

We conclude by outlining a couple of directions for future research. First, extending our analysis to multiple types is challenging but relevant as many service providers typically face a more heterogeneous customer base. Second, we assume that customers have perfect forecasts of their demand rate, γ_i. However, a significant fraction of customers may face some demand rate uncertainty and therefore be reluctant to plan ahead. In this case their menu choice decisions may involve some sort of newsvendor logic, and the provider would want to account for the resulting variability risk, both in designing their service plans.

References

Afèche, P. 2004. Incentive-compatible revenue management in queueing systems: Optimal strategic idleness and other delay tactics. Tech. rep., University of Toronto, Toronto.

Afèche, P. 2013. Incentive-compatible revenue management in queueing systems: optimal strategic delay. *Manufacturing & Service Operations Management* **15**(3) 423–443.

Afèche, P., M. Araghi, O. Baron. 2017. Customer acquisition, retention, and service access quality: Optimal advertising, capacity level, and capacity allocation. *Manufacturing & Service Operations Management* **19**(4) 674–691.

Afèche, P., O. Baron, Y. Kerner. 2013. Pricing time-sensitive services based on realized performance. *Manufacturing & Service Operations Management* **15** 492–506.

Allon, G., A. Federgruen. 2009. Competition in service industries with segmented markets. *Management Science* **55**(4) 619–634.

Boyaci, T., S. Ray. 2003. Product differentiation and capacity cost interaction in time and price sensitive markets. *Manufacturing & Service Operations Management* **5**(1) 18–36.

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

Fudenberg, D., J. Tirole. 1991. *Game Theory*. MIT Press.

Green, L., S. Savin. 2008. Reducing delays for medical appointments: A queueing approach. *Operations Research* **56**(6) 1526–1538.

Guo, P., Z.G. Zhang. 2013. Strategic queueing behavior and its impact on system performance in service systems with the congestion-based staffing policy. *Manufacturing & Service Operations Management* **5**(1) 118–131.

Hassin, R. 1995. Decentralized regulation of a queue. *Management Science* **41**(1) 163–173.

Hassin, R. 2016. *Rational Queueing*. CRC Press.

Hassin, R., M. Haviv. 2003. *To queue or not to queue: Equilibrium behavior in queueing systems*, vol. 59. Springer.

Heyman, D. P., M. J. Sobel. 1982. *Stochastic Models in Operations Research*, vol. 1. McGraw-Hill.

Hsu, V., S. Xu, B. Jukic. 2009. Optimal scheduling and incentive compatible pricing for a service system with quality of service guarantees. *Manufacturing & Service Operations Management* **11**(3) 375–396.

Katta, A., J. Sethuraman. 2005. Pricing strategies and service differentiation in queuesa profit maximization perspective. Tech. rep., Columbia University, New York.

Lariviere, M. A., J. A. Van Mieghem. 2004. Strategically seeking service: How competition can generate poisson arrivals. *Manufacturing & Service Operations Management* **6**(1) 23–40.

Lederer, P., L. Li. 1997. Pricing, production, scheduling and delivery-time competition. *Operations Research* **45**(3) 407–420.

Maglaras, C., J. Yao, A. Zeevi. 2014. Revenue maximization in queues via service differentiation. Tech. rep., Columbia University, New York.

Maglaras, C., A. Zeevi. 2005. Pricing and design of differentiated services: Approximate analysis and structural insights. *Operations Research* **53**(2) 242–262.

Masuda, Y., S. Whang. 2006. On the optimality of fixed-up-to tariff for telecommunications service. *Information System Research* **17** 247–253.

Mendelson, H. 1985. Pricing computer services: Queueing effects. *Communincations of the ACM* **28** 312–321.

Mendelson, H., S. Whang. 1990. Optimal incentive-compatible priority pricing for the m/m/1 queue. *Operations Research* **38** 870–883.

Myerson, R.B. 1997. *Game Theory: Analysis of Conflict*. Harvard University Press.

Naor, P. 1969. On the regulation of queue size by levying tolls. *Econometrica* **37**(1) 15–24.

Plambeck, E.L., Q. Wang. 2013. Implications of hyperbolic discounting for optimal pricing and scheduling of unpleasant services that generate future benefits. *Management Science* **59** 1927–1946.

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

Rao, S., E.R. Petersen. 1998. Optimal pricing of priority services. *Operations Research* **46** 46–56.

Van Mieghem, J.A. 2000. Price and service discrimination in queuing systems: Incentive compatibility of gcμ scheduling. *Management Science* **46** 1249–1267.

Yahalom, T., J.M. Harrison, S. Kumar. 2006. Designing and pricing incentive compatible grades of service in queueing systems. Tech. rep., Stanford University, Stanford, CA.

Zhao, X., K.E. Stecke, A. Prasad. 2012. Lead time and price quotation mode selection: Uniform or differentiated? *Production and Operations Management* **21**(1) 177–193.

Appendix A: Proofs

Proof of Proposition 1 For the FI setting, the solution is found by taking the derivatives of the objective function (3a) with respect to n_i given as

$$\frac{\partial \Pi(n_1,n_2)}{\partial n_i} = \gamma_i \left(r_i - \frac{c\mu}{(\mu - n_1\gamma_1 - n_2\gamma_2)^2} \right), \quad i = 1,2. \tag{A1}$$

(Below we denote $\Pi_{n_i} \equiv \partial\Pi/\partial n_i$.) Define

$$\begin{aligned}
K_0^F &:= \frac{c}{r_i}, \\
K_1^F &:= \arg_{\mu}\{r_i = \frac{c\mu}{(\mu - N_i\gamma_i)^2}\}, \\
K_2^F &:= \arg_{\mu}\{r_j = \frac{c\mu}{(\mu - N_i\gamma_i)^2}\}, \\
K_3^F &:= \arg_{\mu}\{r_j = \frac{c\mu}{(\mu - N_i\gamma_i - N_j\gamma_j)^2}\},
\end{aligned}$$

We consider two cases.

Case 1: $r_1 \geq r_2$.

Because $\gamma_1 > \gamma_2$, it follows from (A1) that

$$\Pi_{n_1} > \Pi_{n_2}.$$

Then we have the following after some algebra:

1. When $\mu \leq \mu_0$, for any $n_1, n_2 \geq 0$, Π_{n_1} is non-positive ($\Pi_{n_1} = 0$ only at $\mu = \mu_0$), and Π_{n_2} is strictly negative. Therefore the provider does not gain any positive revenue from serving any customers.
2. When $\mu_0 < \mu < \mu_1$, for any $0 \leq n_1 \leq N_1$ and $n_2 \geq 0$, Π_{n_2} is strictly negative, and therefore the provider does not serve class-2 customers. But $\Pi_{n_1}(0,0) > 0 > \Pi_{n_1}(N_1, 0)$, and since
$$\frac{\partial^2\Pi}{\partial n_1^2}(n_1,n_2) = -\frac{c\mu\gamma_1}{(\mu - n_1\gamma_1 - n_2\gamma_2)^3} < 0, \tag{A2}$$
for every $n_2 \geq 0$, Π is concave in n_1, and there exists a unique n_1^* between 0 and N_1 that maximizes the revenue.
3. When $\mu_1 \leq \mu \leq K_2^F$, $\Pi_{n_1}(N_1, 0) > 0$ and there exists enough capacity such that the provider exhausts all type-1 customers. Yet, for every $n_2 \geq 0$, $\Pi_{n_2} < 0$, and therefore it is not profitable to serve any type-2 customers.
4. When $K_2^F < \mu < \mu_3$, $\Pi_{n_2}(N_1, 0) > 0 > \Pi_{n_2}(N_1, N_2)$. And since similar to (A2) Π is concave in n_2 for any $n_1 \geq 0$, there exists a unique n_2^* between 0 and N_2 that maximizes the revenue.
5. When $\mu \geq \mu_3$, as in the previous case, it is profitable to serve both classes, and since capacity is high the provider serves all customers from both classes.

Case 2: $r_2 > r_1$

We establish below that over each relevant capacity range we have $\Pi_{n_1} < \Pi_{n_2}$. Then the results in Case 1 hold after switching the subscripts "1" and "2".

1. If $r_1\gamma_1 \le r_2\gamma_2$, the value rate of the type-1 customer is less than the type-2 customer. In this case, from (A1) we have:

$$\begin{aligned}\Pi_{n_1} &= r_1\gamma_1 - \gamma_1 \frac{c\mu}{(\mu - n_1\gamma_1 - n_2\gamma_2)^2} \\ &< r_2\gamma_2 - \gamma_2 \frac{c\mu}{(\mu - n_1\gamma_1 - n_2\gamma_2)^2} = \Pi_{n_2}.\end{aligned}$$

2. If $r_1\gamma_1 > r_2\gamma_2$, the value rate of the type-1 customer is higher than the type-2 customer. Then:
   (a) If $\mu < \widetilde{\mu} = c(\gamma_1 - \gamma_2)/(r_1\gamma_1 - r_2\gamma_2)$, we show that $\Pi_{n_1} < \Pi_{n_2}$:
   Let $\Lambda = n_1\gamma_1 + n_2\gamma_2$. For $0 < \Lambda < \mu$, $\frac{(\mu-\Lambda)^2}{\mu} < \mu$. Therefore

$$\begin{aligned}&\frac{(\mu-\Lambda)^2}{\mu} < \frac{c(\gamma_1-\gamma_2)}{r_1\gamma_1 - r_2\gamma_2} \\ \Rightarrow\quad & \frac{r_1\gamma_1 - r_2\gamma_2}{c(\gamma_1-\gamma_2)} > \frac{\mu}{(\mu-\Lambda)^2} \\ \Rightarrow\quad & r_1\gamma_1 - \gamma_1\frac{c\mu}{(\mu-\Lambda)^2} < r_2\gamma_2 - \gamma_2\frac{c\mu}{(\mu-\Lambda)^2} \\ \Rightarrow\quad & \Pi_{n_1} < \Pi_{n_2}.\end{aligned}$$

   (b) If $\mu \ge \widetilde{\mu}$, observe that

$$\begin{aligned}&\Lambda > \mu - \sqrt{\mu\widetilde{\mu}} \\ \Rightarrow\quad & \frac{(\mu-\Lambda)^2}{\mu} < \widetilde{\mu} = \frac{c(\gamma_1-\gamma_2)}{r_1\gamma_1 - r_2\gamma_2} \\ \Rightarrow\quad & r_1\gamma_1 - \gamma_1\frac{c\mu}{(\mu-\Lambda)^2} < r_2\gamma_2 - \gamma_2\frac{c\mu}{(\mu-\Lambda)^2} \\ \Rightarrow\quad & \Pi_{n_1} < \Pi_{n_2}.\end{aligned}$$

   We show that this condition is either satisfied or redundant for any capacity level. Observe that for any $\mu_0 < \mu < \mu_1$, Λ^* satisfies $r_2 = c\mu/(\mu - \Lambda^*)^2$, or

$$\Lambda^* = \mu - \sqrt{\frac{c\mu}{r_2}} = \mu - \sqrt{\mu\mu_0} > \mu - \sqrt{\mu\widetilde{\mu}}.$$

   Therefore $\Pi_{n_1} < \Pi_{n_2}$ for $\mu < \mu_1$. For $\mu > \mu_1$, $n_2^* = N_2$. To increase Λ one needs to serve type-1 customers as well, and the provider does so only when $\Pi_{n_1} > 0$, which is when $\mu > K_2^F$. For $K_2^F < \mu < \mu_3$, Λ^* satisfies $r_1 = c\mu/(\mu - \Lambda^*)^2$, or equivalently $\Lambda^* = \mu - \sqrt{\frac{c\mu}{r_1}} > \mu - \sqrt{\mu\widetilde{\mu}}$. For $\mu \ge \mu_3$, both types are fully served and therefore the condition is redundant. ■

Proof of Lemma 1 From (4b), $P_i = \gamma_i(r_1 - cW_i) - u_i$. The condition for type-1 is derived from (4c): if $n_2 > 0$, then $u_1 \geq \gamma_1(r_1 - cW_2) + \frac{\gamma_1}{\gamma_2}u_2 - \gamma_1(r_2 - cW_2) \quad \Leftrightarrow \quad u_1 - u_2\frac{\gamma_1}{\gamma_2} \geq \gamma_1(r_1 - r_2)$.

The condition for type-2 is derived from (4d): if $n_1 > 0$, then $u_2 \geq \gamma_2(r_2 - cW_1) + u_1 - \gamma_1(r_1 - cW_1) \quad \Leftrightarrow \quad u_2 - u_1 \geq \gamma_2(r_2 - cW_2) - \gamma_1(r_1 - cW_1)$.

If $r_1 > r_2$, the condition for type-1 implies that if $n_2 > 0$, $u_2 = 0$ and $u_1 = \gamma_1(r_1 - r_2) > 0$. Substituting into the condition for type-1, we get $(\gamma_1 - \gamma_2)(r_2 - cW_1) \geq 0$. The first term is strictly positive from the model assumptions. For $W_1 = W$, the second term is the IR constraint for type-2 customers and therefore non-negative. Since u_1 is independent of the delay of type-2, FIFO is optimal in this case and therefore indeed $W_1 = W$.

If $r_1 < r_2$, the condition for type-2 implies that if $n_1 > 0$, $u_2 = \gamma_2 r_2 - \gamma_1 r_1 + cW_1(\gamma_1 - \gamma_2)$ and $u_1 = 0$. Substituting into the condition for type-2, we get $(\gamma_1 - \gamma_2)(r_1 - cW_1) \geq 0$. Again, the first term is strictly positive. The second term is the IR constraint for type-1 customers and therefore non-negative. ■

Proof of Proposition 2: Observe that the FI solution revenue is an upper bound on the revenue achievable in the PI case. If $\mu \leq K_2^F$, the FI solution can be implemented to the PI case, by setting $n_1 = n_1^{FI}$ and $n_2 = n_2^{FI} = 0$. In this case, in order to maximize the revenue, the firm sets the price for a single plan (plan-1) to $P_1 = (r_1 - cW)\gamma_1$, i.e., $u_1 = 0$. The IR constraint for class-2 (4b) implies that these customers do not purchase and so the FI solution is achieved.

If $\mu > K_2^F$, under the FI setting, $n_1^{FI} = N_1$ and we serve both types. Under the PI setting we need to show that the IC constraints hold.

From the Lemma 1, when serving both types, $u_1 \geq 0$ and $u_2 = 0$. Therefore

$$P_1 = (r_1 - cW_1)\gamma_1 - u_1$$

$$P_2 = (r_2 - cW_2)\gamma_2$$

By (4c), we have

$$(r_1 - cW_1)\gamma_1 - u_1 + cW_1\gamma_1 \leq \frac{\gamma_1}{\gamma_2}(r_2 - cW_2)\gamma_2 + cW_2\gamma_1.$$

Simplifying we have

$$u_1 \geq (r_1 - r_2)\gamma_1$$

Recall that the PI revenue is

$$\Pi^A = ((r_1 - cW_1)\gamma_1 - u_1)n_1 + (r_2 - cW_2)\gamma_2 n_2.$$

Thus, when serving both types, to maximize the revenue we set u_1 to the minimum so that $u_1 = (r_1 - r_2)\gamma_1$.

Note that u_1 is not a function of the delay. Therefore every conserving queueing policy is optimal, including FIFO. Hence, w.l.o.g., we set $W_1 = W_2 = W$. And therefore constraint (4e) is satisfied and (4f) is binding.

So

$$\begin{aligned} P_1 &= (r_1 - cW)\gamma_1 - (r_1 - r_2)\gamma_1 \\ &= (r_2 - cW)\gamma_1. \end{aligned}$$

Because $\gamma_1 \geq \gamma_2$, $P_1 > P_2$ and so (4d) is satisfied.

Let Π_0 be the revenue from serving type-1 fully and exclusively, and let Π_2 be the revenue from serving both types. Let n_2^* be the revenue optimizer for a given capacity. Then

$$\Pi_0 = \left(r_1 - \frac{c}{\mu - \gamma_1 N_1}\right)\gamma_1 N_1,$$
$$\text{and} \quad \Pi_2 = \left(r_2 - \frac{c}{\mu - \gamma_1 N_1 - \gamma_2 n_2^*}\right)(\gamma_1 N_1 + \gamma_2 n_2^*).$$

Define $\Delta\Pi = \Pi_2 - \Pi_0$. Type-2 customers are served only when $\Delta\Pi > 0$. Let $\mathring{\mu}$ solve

$$\left(r_2 - \frac{c}{\mathring{\mu} - \gamma_1 N_1 - \gamma_2 n_2^*}\right)(\gamma_1 N_1 + \gamma_2 n_2^*) - \left(r_1 - \frac{c}{\mathring{\mu} - \gamma_1 N_1}\right)\gamma_1 N_1 = 0.$$

Then if $\mu > \mathring{\mu}$, $\Delta\Pi > 0$. Therefore, n_2^* type-2 customers are served only when $\mu > K_2^A \equiv \max(K_2^F, \mathring{\mu})$, and since $r_1 > r_2$, the PI revenue is strictly lower than the FI revenue because $u_1 > 0$. The PI revenue is also lower than the FI revenue if $K_2^F < \mu \leq \mathring{\mu}$, because the FI revenue from type-2 customers is lost as only the N_1 type-1 customers are served.

It is left to show that the condition

$$r_2(\gamma_1 N_1 + \gamma_2 N_2) > r_1 \gamma_1 N_1 \tag{A3}$$

is necessary and sufficient for the existence of the capacity threshold μ^A. This condition is derived from comparing Π_0 and Π_2 at ample capacity:

$$\begin{aligned} \lim_{\mu \to \infty} \Pi_0 =& r_1 \gamma_1 N_1, \\ \lim_{\mu \to \infty} \Pi_2 =& r_2(\gamma_1 N_1 + \gamma_2 N_2). \end{aligned}$$

Next we show that Π_0 increases less in μ than Π_2, by comparing their derivatives w.r.t. μ:

$$\begin{aligned}\frac{d\Pi_0}{d\mu} &= \frac{c\gamma_1 N_1}{(\mu-\gamma_1 N_1)^2},\\ \frac{d\Pi_2}{d\mu} &= \frac{\partial \Pi_2}{\partial \mu} + \frac{\partial \Pi_2}{\partial n_2^*}\frac{\partial n_2^*}{\partial \mu} = \frac{c(\gamma_1 N_1 + \gamma_2 n_2^*)}{(\mu-\gamma_1 N_1 - \gamma_2 n_2^*)^2}.\end{aligned}$$

Note that $\frac{\partial \Pi_2}{\partial n_2^*} = 0$ because n_2^* is the maximizer of Π_2.

Since for $n_2^* > 0$ we have $\frac{c(\gamma_1 N_1+\gamma_2 n_2^*)}{(\mu-\gamma_1 N_1-\gamma_2 n_2^*)^2} > \frac{c\gamma_1 N_1}{(\mu-\gamma_1 N_1)^2}$, we conclude that $\Delta\Pi$ is monotonically increasing in μ. If condition A3 holds, there exists a capacity level at which we serve type-2 customers, defined as μ^A. If condition A3 does not hold, by monotonicity type-2 customers are not served at any capacity level.

■

Proof of Proposition 3 If $\mu \le K_2^F$, Proposition 1 implies that $n_1^* = 0$. Then (7c) is not active. The remaining problem (7) is identical to problem (2) with $x_i = \gamma_i$. Therefore the solution to the FI setting is feasible and since the PI setting value is bounded by the FI solution, it is optimal.

If $\mu > K_2^F$, for the FI setting $n_1^* > 0$. For the PI setting, if $n_1 > 0$, (7c) is active and since $r_1\gamma_1 \le r_2\gamma_2$, $u_2 > 0$. In any optimal solution (7e) will be tight, i.e.,

$$W_1 = \frac{1}{\mu - \overline{n}_1\gamma_1}, \tag{A4}$$

as increasing W_1 only increases u_2, lowering the objective function value. So

$$u_2 = r_2\gamma_2 - r_1\gamma_1 + \frac{c(\gamma_1-\gamma_2)}{\mu-\overline{n}_1\gamma_1}. \tag{A5}$$

Note (7d) holds since (A5) implies $r_1 - cW_1 \ge 0$, which holds by (4b).

There is no value in increasing the delay for class-2 beyond what is required for (A4) to hold so (7f) holds at equality, implying

$$W_2 = \frac{\mu}{(\mu-\overline{n}_1\gamma_1)(\mu-\bar{n}_1\gamma_1 - n_2\gamma_2)}.$$

Suppose $n_2 = N_2$. Then the objective function value, Π^A, for $n_1 > 0$, say $\Pi_1(n_1)$ is

$$\Pi_1(n_1) = N_2(\gamma_2(r_2 - cW_2) - u_2) + n_1\gamma_1(r_1 - cW_1). \tag{A6}$$

Suppose $\bar{n}_1'$ maximizes (A6). Substituting in W_1 and W_2, and taking derivatives of Π_1 implies $\bar{n}_1'$ solves

$$f(n_1) = r_1 - \frac{c\mu}{(\mu-n_1\gamma_1-N_2\gamma_2)^2} - \frac{N_2 c(\gamma_1-\gamma_2)}{(\mu-n_1\gamma_1)^2} = 0.$$

Let

$$g(n_1) = r_1 - \frac{c\mu}{(\mu-n_1\gamma_1-N_2\gamma_2)^2} \text{ and } h(n_1) = \frac{N_2 c(\gamma_1-\gamma_2)}{(\mu-n_1\gamma_1)^2},$$

so $f(n_1) = g(n_1) - h(n_1)$. From Proposition 1, $g(n_1^*) = 0$ if $n_1^* < N_1$. Further, $g'(n_1) = -2\gamma_1 c\mu/(\mu - n_1\gamma_1 - N_2\gamma_2)^3 < 0$ for n_1 feasible, i.e., $\mu - n_1\gamma_1 - N_2 > 0$. Similarly, $h'(n_1) = 2\gamma_1 N_2 c(\gamma_1 - \gamma_2)/(\mu - n_1\gamma_1)^3 > 0$ for feasible n_1. So $f(\bar{n}_1') = 0$ implies $\bar{n}_1' = N_1$ only if $n_1^* = \bar{n}_1' = N_1$.

For $n_1 = 0$, Π^{PI} given by Π_0, is $\Pi_0 = N_2\gamma_2(r_2 - cW)$, where $W = 1/(\mu - N_2\gamma_2)$. Let

$$\begin{aligned}\Delta\Pi = \Pi_1 - \Pi_0 &= N_2(\gamma_2(r_2 - cW_2) - u_2) + \bar{n}_1'\gamma_1(r_1 - cW_1) - N_2\gamma_2(r_2 - cW)\\ &= (\bar{n}_1' + N_2)r_1\gamma_1 - N_2 r_2\gamma_2 - \frac{c\mu\bar{n}_1'\gamma_1}{(\mu - \bar{n}_1' - N_2\gamma_2)(\mu - N_2\gamma_2)} - \frac{cN_2(\gamma_1 - \gamma_2)}{\mu - \bar{n}_1'\gamma_1}\end{aligned}$$

after substituting in for W_1, W_2, and W. Then

$$\begin{aligned}\frac{\mathrm{d}\Delta\Pi}{\mathrm{d}\mu} = &\frac{c\bar{n}_1'\gamma_1(\mu^2 - N_2\gamma_2(\bar{n}_1'\gamma_1 + N_2\gamma_2))}{(\mu - \bar{n}_1'\gamma_1 - N_2\gamma_2)^2(\mu - N_2\gamma_2)^2} + \frac{cN_2(\gamma_1 - \gamma_2)}{(\mu - \bar{n}_1'\gamma_1)^2}\\ &+ \gamma_1\left(r_1 - \frac{c\mu}{(\mu - \bar{n}_1'\gamma_1 - N_2\gamma_2)^2} - \frac{cN_2(\gamma_1 - \gamma_2)}{(\mu - \bar{n}_1'\gamma_1)^2}\right)\frac{\partial\bar{n}_1'}{\partial\mu} \qquad \text{(A7)}\end{aligned}$$

Noting that the third term in (A7) equals 0 because $f_1(\bar{n}_1') = 0$, as above, we have

$$\frac{\mathrm{d}\Delta\Pi}{\mathrm{d}\mu} = \frac{c\bar{n}_1'\gamma_1(\mu^2 - N_2\gamma_2(\bar{n}_1'\gamma_1 + N_2\gamma_2))}{(\mu - \bar{n}_1'\gamma_1 - N_2\gamma_2)^2(\mu - N_2\gamma_2)^2} + \frac{cN_2(\gamma_1 - \gamma_2)}{(\mu - \bar{n}_1'\gamma_1)^2}$$

Observe,

$$\begin{aligned}\mu^2 - N_2\gamma_2(\bar{n}_1'\gamma_1 + N_2\gamma_2) &> \mu^2 - (\bar{n}_1'\gamma_1 + N_2\gamma_2)^2\\ &= (\mu - \bar{n}_1'\gamma_1 - N_2\gamma_2)(\mu + \bar{n}_1'\gamma_1 + N_2\gamma_2) > 0.\end{aligned}$$

Since $\bar{n}_1' \le n_1^*$, we know $\mu > \bar{n}_1'\gamma_1 + N_2\gamma_2$. So $\mathrm{d}\Delta\Pi/\mathrm{d}\mu > 0$.

Further, we claim $\lim_{\mu\to K_2^{F+}}\Delta\Pi < 0$ and $\lim_{\mu\to\infty}\Delta\Pi > 0$. To see these results, observe that the FI solution implies that when $\mu \to K_2^{F+}$, $n_1^* \to 0^+$ so that $\bar{n}_1' \to 0^+$. Then $W_2 \to W$ and we observe

$$\begin{aligned}\lim_{\mu\to K_2^{F+}}\Delta\Pi &= \lim_{\mu\to K_2^{F+}} -N_2u_2\\ &= \lim_{\mu\to K_2^{F+}} -N_2\left(\frac{c(\gamma_1 - \gamma_2)}{\mu - \bar{n}_1'\gamma_1} - (r_1\gamma_1 - r_2\gamma_2)\right)\\ &= -N_2\left(\frac{c(\gamma_1 - \gamma_2)}{\mu} - (r_1\gamma_1 - r_2\gamma_2)\right) < 0.\end{aligned}$$

Also, as $\mu \to \infty$, $W, W_1,$ and $W_2 \to 0$. Therefore,

$$\lim_{\mu\to\infty}\Delta\Pi = (\bar{n}_1' + N_2)r_1\gamma_1 - N_2r_2\gamma_2.$$

So, if $(N_1 + N_2)(r_1\gamma_1) > N_2r_2\gamma_2$ then $\lim\limits_{\mu\to\infty}\Delta\Pi > 0$. Otherwise $\lim\limits_{\mu\to\infty}\Delta\Pi \le 0$.

Together, the results imply by the intermediate value theorem that if $(N_1 + N_2)(r_1\gamma_1) > N_2r_2\gamma_2$, there exists $\bar{\mu}' > K_2^F$ such that for $\mu > \bar{\mu}'$, $\Delta\Pi > 0$ and for $\mu < \bar{\mu}'$, $\Delta\Pi < 0$. Note that the results were derived for $n_2 = N_2$. Now consider $\overline{\Pi}$, the unrestricted objective function value of (7a), i.e.,

Form	Description	n_1	Class-1 Priority	Surplus Utility
(1A)	Class-2 only served $\Pi^A = \Pi^{FI}$	$n_1 = n_1^* = 0$	NA	No
(1B)	Class-2 only served $\Pi^A < \Pi^{FI}$	$n_1 = 0$ $n_1^* > 0$	NA	No
(2)	Class-1 priority, Class-2 surplus $\Pi^A < \Pi^{FI}$	$0 < n_1 < n_1^*$	Yes	Yes
(3)	Class-1 priority Class-2 no surplus $\Pi^A < \Pi^{FI}$	$0 < n_1 < n_1^*$	Yes	No
(4)	Class-1 priority Class-2 no surplus $\Pi^A = \Pi^{FI}$	$n_1 = n_1^* > 0$	Yes	No
(5)	PI equal to FI solution $\Pi^A = \Pi^{FI}$	$n_1 = n_1^* > 0$	No	No

Table 1: Solution forms for high value rate sub-case.

with $n_2 \geq 0$ as a function of μ. Observe $\overline{\Pi} \geq \Pi_1(n_1)$ for a given μ with $\overline{\Pi} = \Pi_1(n_1)$ at $\mu = K_2^F$, since $n_2 = N_2$ at K_2^F by Proposition 1. Also, there exists a μ' such that $\overline{\Pi} = \Pi_1(n_1)$ for all $\mu \geq \mu'$ since $\lim_{\mu\to\infty} n_2^* = N_2$. We observe that $\overline{\Pi}$ is concave in μ as the feasible set for n_i, W_i, u_2 given by (7) is convex in μ and (7a) is concave in n_i, W_i, and u_2. Therefore by the intermediate value theorem there exists $\bar{\mu} \leq \bar{\mu}'$ such that $\Pi_0 > \overline{\Pi}$ for $\mu < \bar{\mu}$ and $\Pi_0 \leq \overline{\Pi}$ for $\mu \geq \bar{\mu}$. ■

Restatement of Proposition 4 The solution in the high-value sub-case for the fixed usage rate model can be classified as being in one of five forms. Each solution is defined by three elements: n_1, the number of type-1 customers served, whether the type-1 customers are served with priority, and whether the type-1 customers receive a consumer surplus. The five forms are summarized in Table 1. We divide Form 1 into 1A and 1B as in Figure 2 to represent regions where $n_1 = n_1^* = 0$ (1A), and where $n_1 = 0$ for the PI setting but $n_1^* > 0$ in the FI setting (1B). Also, let $\widetilde{N}_2 = (\tilde{\mu}\sqrt{r_1} - \sqrt{\tilde{\mu}c})/(\gamma_2\sqrt{r_1})$; $\widetilde{N}_2$ is the value of N_2 where algebraically $K_2^F = \tilde{\mu}$. Let K_2^F be defined as in Proposition 1. Also let $K_1 = r_1(\widetilde{\mu} - N_2\gamma_2)^2/c$ and $K_2 = r_1\widetilde{\mu}^2/c$. Proposition 4 summarizes all of the possibilities.

Proposition 4 *Suppose $r_1 < r_2$ and $r_1\gamma_1 > r_2\gamma_2$.*

1A. If $\mu \leq K_2^F$, solution 1A holds: $n_1 = 0$, $n_2 = N_2$, $u_2 = 0$, and $\Pi^A = \Pi^{FI}$.

1B. If $K_2^F < \mu \leq \min(\bar{\mu}, \tilde{\mu})$, solution 1B holds: $n_1 = 0$, $n_2 = N_2$, $u_2 = 0$, and $\Pi^A < \Pi^{FI}$.

2. For $N_2 < \tilde{N}_2$, if $\bar{\mu} < \tilde{\mu}$, there exists $\hat{\mu}$ such that for $\bar{\mu} < \mu \leq \hat{\mu}$ solution 2 holds: $n_1 < n_1^*$, $n_2 \leq n_2^*$, class 1 is served with priority, $u_2 = r_1\gamma_1 - r_2\gamma_2 - c(\gamma_1 - \gamma_2)/(\mu - n_1\gamma_1) > 0$ and $\Pi^A < \Pi^{FI}$

3. For $N_2 < \tilde{N}$ and $\max(\hat{\mu}, \tilde{\mu}) < \mu \leq K_1$, solution 3 holds: $n_1 < n_1^*$, $n_2 = N_2$, class 1 is served with priority, $u_2 = 0$ and $\Pi^A < \Pi^{FI}$.

4. *If* $\max(K_2^F, K_1) < \mu \le K_2$ *then solution 4 holds:* $n_1 = n_1^*$, $n_2 = N_2$, *class 1 is served with priority,* $u_2 = 0$ *and* $\Pi^A = \Pi^{FI}$.
5. *If* $\mu > K_2$, *then solution 5 holds:* $n_1 = n_1^*$, $n_2 = N_2$, *both classes are served FIFO,* $u_2 = 0$ *and* $\Pi^A = \Pi^{FI}$.

Proof of Proposition 4 We proceed to show each of the regions holds, in order.

1A. For $\mu \le \mu$, the proof is the same as in Proposition 3.

1B. As in Proposition 3, let $\Pi_0 = N_2\gamma_2(r_2 - cW)$ where $W = 1/(\mu - N_2\gamma_2)$, i.e., the solution with $n_1 = 0$, and let $\overline{\Pi}$ be the (unrestricted) optimal solution to (7) (i.e., $n_1 \ge 0$, $n_2 \ge 0$, and $K_2^F \ge 0$). Then for $K_2^F < \mu \le \widetilde{\mu}$, following the logic in the proof of Proposition 3, there exists K_2^A such that $\Pi_0 > \overline{\Pi}$ and $\Pi_0 \le \overline{\Pi}$ for $\mu \ge K_2^A$. For $\mu \le K_2^A$ only class 2 is served.

2. Let $\hat{\Pi}$ be the optimal solution with $n_1 > 0$, $n_2 > 0$, and $u_2 = 0$, and let $\hat{n}_1$ designate the solution for n_1 in this case. We establish $\hat{n}_1$ in four steps.

   A. Determine $\hat{n}_1$. Forcing $u_2 = 0$ implies $n_2 = N_2$ since (7f) holds at equality and $r_2 > r_1$. So $\hat{\Pi} = N_2\gamma_2(r_2 - cW_2) + \hat{n}_1\gamma_1(r_1 - cW_1)$ with $W_1 = 1/(\mu - \hat{n}_1\gamma_1)$ and $W_2 = \mu/((\mu - \hat{n}_1\gamma_1)(\mu - \hat{n}_1\gamma_1 - N_2\gamma_2))$. From (7f), holding at equality we have

$$\hat{\Pi}_1 = N_2\gamma_2 r_2 + \hat{n}_1\gamma_1 r_2 - \frac{c(\hat{n}_1\gamma_1 + N_2\gamma_2)}{\mu - \hat{n}_1\gamma_1 - N_2\gamma_2}.$$

Taking the derivative with respect to $\hat{n}_1$,

$$\frac{\partial\hat{\Pi}_1}{\partial\hat{n}_1} = \gamma_1\left(r_1 - \frac{c}{(\mu - \hat{n}_1\gamma_1 - N_2\gamma_2)^2}\right).$$

Noting from Proposition 1 for $\mu > K_2^F$,

$$r_1 = \frac{c}{(\mu - n_1^*\gamma_1 - N_2\gamma_2)^2}, \tag{A8}$$

so that $\partial\hat{\Pi}_1/\partial\hat{n}_1 > 0$ if $n_1^* > \hat{n}_1$. This holds for $\mu < K_1$ since $\mu < K_1 = r_1(\widetilde{\mu} - N_2\gamma_2)^2/c$ implies

$$\begin{aligned}\widetilde{\mu} &> N_2\gamma_2 + \sqrt{\frac{c\mu}{r_1}} \\ &= \mu - n_1^*\gamma_1\end{aligned} \tag{A9}$$

where the equality follows from (A8). But $W_1 = 1/(\mu - \hat{n}_1\gamma_1) \le 1/\widetilde{\mu}$ implies $\mu - \hat{n}_1\gamma_1 \ge \widetilde{\mu}$ implying $n_1^* > \tilde{n}_1$, $\partial\hat{\Pi}_1/\partial\hat{n}_1 > 0$, and

$$\hat{n}_1 = \frac{\mu - \widetilde{\mu}}{\gamma_1}. \tag{A10}$$

B. Show $\partial\hat{n}_1/\partial\mu > \partial\bar{n}_1/\partial\mu > 0$. From (A10), $\partial\hat{n}_1/\partial\mu = 1/\gamma_1 > 0$. From (A6), let $\bar{n}_1' < N_1$ solve

$$r_1 - \frac{c\mu}{(\mu - \bar{n}_1'\gamma_1 - N_2\gamma_2)^2} - \frac{N_2 c(\gamma_1-\gamma_2)}{(\mu - \bar{n}_1')^2} = 0.$$

By implicit differentiation

$$\frac{\partial \bar{n}_1'}{\partial\mu} = \frac{1}{\gamma_1}\left(1 - \frac{(\mu-\bar{n}'\gamma_1 - N_2\gamma_2)(\mu-\bar{n}'\gamma_1)^3}{2\mu(\mu-\bar{n}_1'\gamma_1)^3 + 2N_2(\gamma_1-\gamma_2)(\mu-\bar{n}'\gamma_1-N_2\gamma_2)^3}\right)$$

Since

$$0 < \frac{(\mu-\bar{n}'\gamma_1 - N_2\gamma_2)(\mu-\bar{n}'\gamma_1)^3}{2\mu(\mu-\bar{n}_1'\gamma_1)^3 + 2N_2(\gamma_1-\gamma_2)(\mu-\bar{n}'\gamma_1-N_2\gamma_2)^3} < \frac{1}{2}$$

it follows that

$$\frac{\partial\hat{n}_1}{\partial\mu} > \frac{\bar{n}'}{\partial\mu} > 0$$

Noting $\bar{n}_1$ solving

$$r_1 - \frac{c\mu}{(\mu - \bar{n}_1\gamma_1 - n_2\gamma_2)^2} - \frac{n_2 c(\gamma_1-\gamma_2)}{(\mu - \bar{n}_1)^2} = 0.$$

for $n_2 \le N_2$, implies $\bar{n} \ge \bar{n}'$. Also, $\lim_{\mu\to\widetilde{\mu}} \bar{n}_1 = \bar{n}_1'$, since at $\widetilde{\mu}$, $n_2 = N_2$, $\partial\bar{n}/\partial\mu < \partial\hat{n}/\partial\mu$.

C. Next we show $\bar{n}_1 \le \hat{n}_1$ in any feasible solution. Observe for $n_1 = \hat{n}_1 = (\mu - \widetilde{\mu})/\gamma_1$,

$$u_2 = r_2\gamma_2 - r_1\gamma_1 - \frac{c(\gamma_1-\gamma_2)}{\mu - \hat{n}_1\gamma_1} = 0 \quad \text{(as assumed)}$$

while if $n_1 = \bar{n}_1 > \hat{n}_1$,

$$u_2 = r_2\gamma_2 - r_1\gamma_1 - \frac{c(\gamma_1-\gamma_2)}{\mu - \bar{n}_1\gamma_1} < 0$$

which is not feasible.

D. Suppose $\bar{n}_1 > 0$ for $\mu = \widetilde{\mu}$. Then $\overline{\Pi}_1(\widetilde{\mu}) > \Pi_0(\widetilde{\mu})$. Noting $\hat{\Pi}_1(\widetilde{\mu}) = N_2\gamma_2(r_2 - cW(\widetilde{\mu})) = \Pi_0(\widetilde{\mu})$, this implies $\Pi_1(\widetilde{\mu}) > \hat{\Pi}_1(\widetilde{\mu})$ so that it is optimal to serve $\bar{n}_1$ at $\widetilde{\mu}$. By point B above, there exists $\hat{\mu} > \widetilde{\mu}$ such that $\bar{n}_1(\hat{\mu}) = \hat{n}_1(\hat{\mu})$. So if $K_2^A < \widetilde{\mu}$ then for $K_2^A < \mu < \hat{\mu}$, $\bar{n}_1$ is optimal. Note at $\mu = \hat{\mu}$, $\bar{n}_1 = \hat{n}_1$ and $\overline{\Pi}(\hat{\mu}) = \hat{\Pi}(\hat{\mu})$. Further, noting that $\hat{\Pi}(\mu)$ is concave at each (n_1, N_2), then $\hat{\Pi}(\mu)$ is concave in μ_n as is $\overline{\Pi}(\mu)$. Then by point B, $\hat{\mu}$ is the unique intersection of $\overline{\Pi}(\mu)$ and $\hat{\Pi}(\mu)$.

3. Consider the case where $\bar{n}_1\widetilde{\mu} = 0$. As $\Pi_1(\widetilde{\mu}) = \Pi_0(\widetilde{\mu})$, therefore $\Pi_1(\widetilde{\mu}) = \hat{\Pi}_1(\widetilde{\mu})$. By point 2, $\partial\hat{n}_1/\partial\mu > \partial\bar{n}_1/\partial\mu > 0$ and by point 3, it is not feasible to serve $\bar{n}_1$. So for $\widetilde{\mu} < \mu < K_1$ where $\bar{n}_1(\widetilde{\mu}) = 0$, it is optimal to serve $\hat{\mu}_1$ and let $n_2 = N_2$ and $u_2 = 0$, with priority given to class 1. As noted from (A8) and (A9), $\hat{n}_1 < n_1^*$ for $\widetilde{\mu} \le \mu < K_1$. And as n_1^* is the optimal solution to the FI problem, $\Pi^{FI} > \hat{\Pi}$.

4. For $K_1 < \mu \le K_2$, and $\mu \ge K_2^F$, solving $K_1 = K_2^F$ implies $\mu \ge \tilde{\mu}$ as in Figure 2. And $\mu \ge K_1$ implies $\widetilde{\mu} \le N_2\gamma_2 + \sqrt{c\mu/r_1}$. Noting n_1^* solves $r_1 = c\mu/(\mu - n_1^*\gamma_1 - N_2\gamma_2)^2$ or

$$n_1^*\gamma_1 = \mu - N_2\gamma_2 - \sqrt{\frac{c\mu}{r_1}}. \tag{A11}$$

Then

$$W_1 = \frac{1}{\mu - n_1^*\gamma_1} = \frac{1}{N_2\gamma_2 + \sqrt{c\mu/r_1}} \le \frac{1}{\widetilde{\mu}} = \widetilde{W}.$$

So (7c) holds with $n_1 = n_1^*$, when priority is given to class 1. Since FIFO waiting implies from (A11),

$$W = \frac{1}{\mu - n_1^*\gamma_1 - N_2\gamma_2} = \frac{1}{\sqrt{c\mu/r_1}},$$

then for $\mu < K_2 = r_1\widetilde{\mu}^2/c$ or $\widetilde{\mu} > \sqrt{c\mu/r_1}$, $W > \widetilde{W} \ge W_1$ implying prioritizing class 1 is necessary. Note that

$$\begin{aligned}
\Pi^A &= n_1^*\gamma_1(r_1 - cW_1) + N_2\gamma_2(r_2 - cW_2) \\
&= n_1^*\gamma_1 r_1 + N_2\gamma_2 r_2 - c(n_1^*\gamma_1 W_1 + N_2\gamma_2 W_2) \\
&= n_1^*\gamma_1 r_1 + N_2\gamma_2 r_2 - cW \\
&= \Pi^{FI}.
\end{aligned}$$

5. For $\mu \ge K_2$, $\sqrt{c\mu/r_1} \ge \widetilde{\mu}$, so $W \le \widetilde{W}$, implying W satisfies (7c) and no priority is needed.

Note that K_3^A is the minimum capacity for which $u_2 = 0$. If N_1 is relatively small, there exist capacity level for which $u_2(N_1) > 0$. At this capacity, $\overline{n}_1 = n_1^* = N_1$. As we increase capacity, $u_2(N_1)$ decreases. At the minimum capacity for which $u_2 = 0$, $W_1 = \widetilde{W}$ and the FI revenue is achieved. This implies that case 3 of Proposition 4 vanishes, because there is no capacity for which $u_2 = 0$ and $n_1 < n_1^*$. From $W_1 = \widetilde{W}$ we derive the capacity threshold $\mu = \widetilde{\mu} + N_1\gamma_1$. Therefore we define the threshold K_3^A as the minimum of $K_1 = r_1(\widetilde{\mu} - N_2\gamma_2)^2/c$ and $\widetilde{\mu} + N_1\gamma_1$. Observe $K_1 = \widetilde{\mu} + N_1\gamma_1$ if $N_1 = \widetilde{N}_1 = (r_1(\widetilde{\mu} - N_2\gamma_2)^2/c - \widetilde{\mu})/\gamma_1$. Therefore case 3 of Proposition 4 exists only if $N_1 > \widetilde{N}_1$.

In the same way, K_4^A is the minimum capacity for which both types are served and FIFO is optimal. In this case, $W_1 = W$ and $u_2 = 0$. If N_1 is relatively small, $W = \widetilde{W}$ for $\mu = \widetilde{\mu} + N_1\gamma_1 + N_2\gamma_2$. Otherwise, $W = \widetilde{W}$ for $\mu = K_2$. Therefore we define the threshold K_4^A as the minimum of $K_2 = r_1\widetilde{\mu}^2/c$ and $\widetilde{\mu} + N_1\gamma_1 + N_2\gamma_2$. Observe $K_2 > \widetilde{\mu} + N_1\gamma_1 + N_2\gamma_2$ if $N_1 > \frac{1}{\gamma_1}\left(\frac{r_1\widetilde{\mu}^2}{c} - \widetilde{\mu} - N_2\gamma_2\right)$, and $K_2 \le \widetilde{\mu} + N_1\gamma_1 + N_2\gamma_2$ otherwise.

Also note that $\widetilde{N}_2$ is the value of N_2 that solves $K_2^F = K_3^A$. If $N_2 \ge \widetilde{N}_2$, then cases 2 and 3 of Proposition 4 vanish. In the same way, $\widetilde{N}$ is the value of N_2 that solves $K_2^F = K_4^A$. If $N_2 \ge \widetilde{N}$, then cases 2,3 and 4 of Proposition 4 vanish. ■

Proofs for Section 5

The proofs reference the FI problem (3) which we recall for convenience. Substitute $\lambda_i = N_i x_i$ in (3a) and let

$$\Pi^{FI}(\mathbf{x};\mu) := \sum_{i=1}^{2} N_i x_i \left(r_i - c\frac{1}{\mu - N_1 x_1 - N_2 x_2} \right) \tag{A12}$$

denote the FI revenue function, where $\mathbf{x} = (x_1, x_2)$ denotes the usage-rate vector. $\Pi^{FI}(\mathbf{x};\mu)$ is strictly concave in $\mathbf{x}$ (see proof of Proposition 1). Throughout the proofs we write the first partial derivatives of this function as $\Pi^{FI}_{x_i}(\mathbf{x};\mu)$ for $\partial\Pi^{FI}(\mathbf{x};\mu)/\partial x_i$, and as $\Pi^{FI}_{\mu}(\mathbf{x};\mu)$ for $\partial\Pi^{FI}(\mathbf{x};\mu)/\partial\mu$. Let $\Pi^{FI}(\mu)$ denote the optimal FI revenue as a function of capacity. (We use the same conventions for the PI revenue functions defined below.) The FI problem is

$$\Pi^{FI}(\mu) = \max_{\mathbf{x}} \Pi^{FI}(\mathbf{x};\mu) \tag{A13a}$$

subject to

$$0 \le x_i \le \gamma_i, \quad \text{for } i = 1, 2, \tag{A13b}$$

$$N_1 x_1 + N_2 x_2 < \mu. \tag{A13c}$$

Write $\mathbf{x}^{FI}(\mu)$ for the unique solution of (A13a)-(A13c). In arguments where the capacity is fixed, we suppress the dependence of $\mathbf{x}^{FI}(\mu)$ and $\Pi^{FI}(\mu)$ on μ.

Proof of Proposition 5. This proof follows the same line of argument as the proof of Lemma 3 and is therefore relegated to Appendix B.

Proofs of Propositions 6 and 7. For convenience we restate the PI problem (12):

$$\Pi^{I} = \max_{\mathbf{x},\mathbf{W},u_2} \sum_{i=1}^{2} N_i x_i (r_i - cW_i) - N_2 u_2 \tag{A14a}$$

$$\text{subject to} \quad u_2 \ge 0, \tag{A14b}$$

$$u_2 \ge \begin{cases} x_1(r_2 - r_1), & x_1 \le \gamma_2, \\ \gamma_2 r_2 - x_1 r_1 + c(x_1 - \gamma_2)W_1, & \gamma_2 \le x_1 \le \gamma_1, \end{cases} \tag{A14c}$$

$$u_2 \le x_2(r_2 - r_1), \tag{A14d}$$

$$W_i \ge \frac{1}{\mu - N_i x_i}, \quad \text{for } i = 1, 2, \tag{A14e}$$

$$\sum_{i=1}^{2} N_i x_i W_i \ge \frac{N_1 x_1 + N_2 x_2}{\mu - N_1 x_1 - N_2 x_2}, \tag{A14f}$$

$$0 \le x_i \le \gamma_i, \quad \text{for } i = 1, 2, \tag{A14g}$$

$$N_1 x_1 + N_2 x_2 < \mu. \tag{A14h}$$

Due to the IC constraint (A14c) this optimization problem need not be convex. However, we show that each of its two subproblems, for $x_1 \in [0, \gamma_2]$ and for $x_1 \in [\gamma_2, \gamma_1]$, has a unique solution.

The proof proceeds in five steps. (We relegate the proofs of the technical Lemmas 3-8 to Appendix B.)

1. Lemma 3 characterizes the solution of (A14a)-(A14h) for $x_1 \in [0, \gamma_2]$. For this subproblem FIFO is optimal and yields lower revenue than the FI solution whenever both types are served (as in Part 2 of Propositions 6 and 7).

2. Lemmas 4-7 characterize the solution of (A14a)-(A14h) for $x_1 \in [\gamma_2, \gamma_1]$. This subproblem yields one of two solution types. One requires absolute priority to type 1 and yields less revenue than the FI solution (as in Part 3 of Propositions 6 and 7). The other attains the optimal FI revenue, with either some priority to type 1 or FIFO (as in Parts 4 and 5, respectively, of Proposition 7).

3. Lemma 8 compares the FIFO solution for $x_1 \leq \gamma_2$ with the strict priority solution for $x_1 > \gamma_2$.

4. We prove the claims of Proposition 6 using Lemmas 3-8.

5. We prove the claims of Proposition 7 using Lemmas 3-8.

STEP 1: Solution of (A14a)-(A14h) for $x_1 \in [0, \gamma_2]$

We simplify this subproblem based on the following observations:

(i) The constraint (A14b) is redundant as it is implied by (A14c) since $r_2 > r_1$ by hypothesis.

(ii) At optimality (A14c) must be binding, that is, $u_2 = x_1 (r_2 - r_1)$, because by (A14a) the revenue decreases in u_2 and the constraint (A14d) imposes an upper bound on u_2.

(iii) The optimal scheduling policy must be work-conserving, i.e., the constraint (A14f) must bind, because by (A14a) and since $u_2 = x_1 (r_2 - r_1)$ the revenue function is decreasing in $\sum_{i=1}^{2} N_i x_i W_i$. Therefore, defining $\underline{\Pi}^I(\mathbf{x};\mu)$ to be the optimal PI revenue function for $x_1 \in [0, \gamma_2]$, we have

$$\underline{\Pi}^I(\mathbf{x};\mu) := \Pi^{FI}(\mathbf{x};\mu) - N_2 x_1 (r_2 - r_1). \tag{A15}$$

(iv) FIFO is optimal, that is $W_1 = W_2 = 1/(\mu - N_1 x_1 - N_2 x_2)$, because by (A15) the revenue function $\underline{\Pi}^I(\mathbf{x};\mu)$ is invariant under any work-conserving policy.

Observations $(i) - (iv)$ yield the following PI subproblem of (A14a)-(A14h) for $x_1 \in [0, \gamma_2]$. Letting $\underline{\Pi}^I(\mu)$ denote the optimal PI revenue over $x_1 \in [0, \gamma_2]$ as a function of μ, we have

$$\underline{\Pi}^I(\mu) = \max_{\mathbf{x}} \underline{\Pi}^I(\mathbf{x};\mu) \tag{A16a}$$

$$\text{subject to} \quad x_1 \leq x_2, \tag{A16b}$$

$$x_2 \leq \gamma_2, \tag{A16c}$$

$$0 \leq x_1 \leq \gamma_2, \tag{A16d}$$

$$N_1 x_1 + N_2 x_2 < \mu, \tag{A16e}$$

where the constraint $x_1 \leq x_2$ is equivalent to (A14d) since $u_2 = x_1 (r_2 - r_1)$ for $x_1 \in [0, \gamma_2]$.

Write $\underline{\mathbf{x}}^I(\mu)$ for the solution of (A16a)-(A16e). In arguments where the capacity is fixed, we suppress the dependence of $\underline{\mathbf{x}}^I(\mu)$ and $\underline{\Pi}^I(\mu)$ on μ. Lemma 3 characterizes the solution of (A16a)-(A16e).

Lemma 3 *Fix $r_2 > r_1$. The PI subproblem of (A14a)-(A14h) for $x_1 \in [0, \gamma_2]$ specializes to (A16a)-(A16e) and its solution has the following properties:*

1. *For fixed $\mu > 0$ there exists a unique maximizer $\underline{\mathbf{x}}^I(\mu)$ and FIFO is optimal.*
2. *$\underline{\Pi}^I(K_2^F) = \Pi^{FI}(K_2^F)$ and $\underline{\Pi}^I(\mu) < \Pi^{FI}(\mu)$ for $\mu > K_2^F$.*
3. *$\underline{\Pi}^I(\mu)$ and $\underline{\mathbf{x}}^I(\mu)$ are continuous in $\mu \geq K_2^F$.*
4. *If $\frac{N_1}{N_2} \leq \frac{r_2}{r_1} - 1$ then $\underline{x}_1^I = 0$, $\underline{x}_2^I = \gamma_2$, $\underline{\Pi}_{x_1}^I(\underline{\mathbf{x}}^I; \mu) < 0$ and $\underline{\Pi}_{x_2}^I(\underline{\mathbf{x}}^I; \mu) > 0$ for $\mu \geq K_2^F$. If $\frac{N_1}{N_2} > \frac{r_2}{r_1} - 1$ then the solution structure is determined by the capacity thresholds*

$$\mu_2' := \arg\left\{\mu > N_2\gamma_2 : r_1 - \frac{N_2}{N_1}(r_2 - r_1) - c\frac{\mu}{(\mu - N_2\gamma_2)^2}\right\}, \tag{A17}$$

$$\mu_3' := \arg\left\{\mu > N_1\gamma_2 + N_2\gamma_2 : r_1 - \frac{N_2}{N_1}(r_2 - r_1) - c\frac{\mu}{(\mu - N_1\gamma_2 - N_2\gamma_2)^2}\right\}, \tag{A18}$$

and K_2^F (as defined in Proposition 1), where $K_2^F < \mu_2' < \mu_3' < \infty$.

   (a) *For $\mu \in [K_2^F, \mu_2']$ we have $\underline{x}_1^I = 0$, $\underline{x}_2^I = \gamma_2$, $\underline{\Pi}_{x_1}^I(\underline{\mathbf{x}}^I; \mu) \leq 0$ with equality iff $\mu = \mu_2'$, and $\underline{\Pi}_{x_2}^I(\underline{\mathbf{x}}^I; \mu) > 0$.*

   (b) *For $\mu \in (\mu_2', \mu_3')$ we have $\underline{x}_1^I \in (0, \gamma_2)$, $\underline{x}_2^I = \gamma_2$, and $\underline{\Pi}_{x_1}^I(\underline{\mathbf{x}}^I; \mu) = 0 < \underline{\Pi}_{x_2}^I(\underline{\mathbf{x}}^I; \mu)$.*

   (c) *For $\mu \geq \mu_3'$ we have $\underline{x}_1^I = \underline{x}_2^I = \gamma_2$, $\underline{\Pi}_{x_1}^I(\underline{\mathbf{x}}^I; \mu) \geq 0$ with equality iff $\mu = \mu_3'$, and $\underline{\Pi}_{x_2}^I(\underline{\mathbf{x}}^I; \mu) > 0$.*

5. *For $\mu \geq K_2^F$ the right derivative of the optimal revenue with respect to capacity equals*

$$\frac{d_+\underline{\Pi}^I(\mu)}{d\mu} = c\frac{N_1\underline{x}_1^I(\mu) + N_2\underline{x}_2^I(\mu)}{(\mu - N_1\underline{x}_1^I(\mu) - N_2\underline{x}_2^I(\mu))^2}. \tag{A19}$$

STEP 2: Solution of (A14a)-(A14h) for $x_1 \in [\gamma_2, \gamma_1]$

For $x_1 \in [\gamma_2, \gamma_1]$ the PI problem (A14a)-(A14h) simplifies by noting that the constraint (A14f) must bind (the optimal scheduling policy must be work-conserving), because by (A14a), (A14b), and (A14c) the revenue function is decreasing in $\sum_{i=1}^2 N_i x_i W_i$, and non-increasing in W_1.

Let $\overline{\Pi}^I(\mu)$ denote the optimal profit of the PI subproblem of (A14a)-(A14h) for $x_1 \in [\gamma_2, \gamma_1]$. Using the definition (A12) for the FI profit function $\Pi^{FI}(\mathbf{x}; \mu)$, we have:

$$\overline{\Pi}^I(\mu) = \max_{\mathbf{x}, u_2, W_1} \Pi^{FI}(\mathbf{x}; \mu) - N_2 u_2 \tag{A20a}$$

$$\text{subject to} \quad u_2 \geq 0, \tag{A20b}$$

$$u_2 \geq \gamma_2 r_2 - x_1 r_1 + c(x_1 - \gamma_2)W_1, \tag{A20c}$$

$$u_2 \leq x_2 (r_2 - r_1), \tag{A20d}$$

$$W_1 \geq \frac{1}{\mu - N_1 x_1}, \tag{A20e}$$

$$0 \leq x_2 \leq \gamma_2, \tag{A20f}$$

$$\gamma_2 \leq x_1 \leq \gamma_1, \tag{A20g}$$

$$N_1 x_1 + N_2 x_2 < \mu. \tag{A20h}$$

Let $\overline{\mathbf{x}}^I(\mu)$ be the solution of (A20a)-(A20h). In arguments where the capacity is fixed, we suppress the dependence of $\overline{\mathbf{x}}^I(\mu)$ and $\overline{\Pi}^I(\mu)$ on μ. Lemma 4 and Lemma 5 establish that $\overline{\mathbf{x}}^I(\mu)$ is unique. Lemma 4 classifies this solution into one of three "categories", based on three mutually exclusive and exhaustive conditions on the capacity μ and the corresponding FI solution $\mathbf{x}^{FI}$.

Lemma 4 *Fix $r_2 > r_1$ and $\mu \geq K_2^F$. The PI subproblem of (A14a)-(A14h) for $x_1 \in [\gamma_2, \gamma_1]$ specializes to problem (A20a)-(A20h). For $\mathbf{x}$ satisfying (A20f)-(A20h) define the function*

$$u_2^P(x_1;\mu) := \gamma_2 r_2 - x_1 r_1 + c(x_1 - \gamma_2)\frac{1}{\mu - N_1 x_1}. \tag{A21}$$

Then if (A20a)-(A20h) is feasible, its solution depends as follows on three mutually exclusive and exhaustive conditions on the capacity μ and the corresponding unique FI solution $\mathbf{x}^{FI}$.

1. *$\partial u_2^P(\gamma_2;\mu)/\partial x_1 > 0$: Then $\overline{x}_1^I = \overline{x}_2^I = \gamma_2$ is the only feasible solution and FIFO is optimal.*
2. *$\partial u_2^P(\gamma_2;\mu)/\partial x_1 \leq 0$ and the FI solution satisfies one of two conditions, either (a) $x_1^{FI} < \gamma_2$, or (b) $x_1^{FI} \geq \gamma_2$ and $u_2^P(x_1^{FI};\mu) > 0$. Then the optimal solution satisfies:* (1) *if $\overline{x}_1^I > \gamma_2$ then type 1 customers receive absolute priority;* (2) *the constraints (A20c) and (A20e) are binding, so $u_2 = u_2^P(\overline{x}_1^I;\mu) \geq 0$, and* (3) *$\partial u_2^P(\overline{x}_1^I;\mu)/\partial x_1 \leq 0$.*
3. *$u_2^P(x_1^{FI};\mu) \leq 0$. This holds if and only if $\mathbf{x}^{FI} = \overline{\mathbf{x}}^I$ and $\overline{\Pi}^I = \Pi^{FI}$. In this case $x_1^{FI} > \gamma_2$ and $\partial u_2^P(\gamma_2;\mu)/\partial x_1 < 0$.*

We build on Lemma 4 as follows to specify the solution of the PI subproblem (A20a)-(A20h) for $x_1 \in [\gamma_2, \gamma_1]$. We ignore Part 1, because in this case the solution is also feasible in the subproblem for $x_1 \in [0, \gamma_2]$. We use Part 2 to simplify (A20a)-(A20h) and characterize the solution of this simplified problem in Lemma 5. We use Part 3 to identify in Lemma 6 capacity levels μ for which $\overline{\Pi}^I(\mu) = \Pi^{FI}(\mu)$, and in Lemma 7 the optimal scheduling policies corresponding to these capacity levels.

Under the conditions of Lemma 4.2, problem (A20a)-(A20h) has the same solution as its constrained version with (A20c) and (A20e) binding and with the additional constraint $\partial u_2^P(x_1;\mu)/\partial x_1 \leq 0$. In this case, letting

$$\overline{\Pi}^I(\mathbf{x};\mu) := \Pi^{FI}(\mathbf{x};\mu) - N_2 u_2^P(x_1;\mu) \tag{A22}$$

and recalling the definition of $u_2^P(x_1;\mu)$ in (A21), the PI subproblem for $x_1 \in [\gamma_2, \gamma_1]$ specializes to

$$\overline{\Pi}^I(\mu) = \max_{\mathbf{x}} \overline{\Pi}^I(\mathbf{x};\mu) \tag{A23a}$$

$$\text{subject to} \quad u_2^P(x_1;\mu) \geq 0, \tag{A23b}$$

$$\partial u_2^P(x_1;\mu)/\partial x_1 \leq 0, \tag{A23c}$$

$$u_2^P(x_1;\mu) \leq x_2(r_2 - r_1), \tag{A23d}$$

$$0 \leq x_2 \leq \gamma_2, \tag{A23e}$$

$$\gamma_2 \leq x_1 \leq \gamma_1, \tag{A23f}$$

$$N_1 x_1 + N_2 x_2 < \mu. \tag{A23g}$$

Lemma 5 formalizes this implication and characterizes the solution properties of (A23a)-(A23g).

Lemma 5 *Fix $r_2 > r_1$. Under the conditions of Lemma 4.2, i.e., $\mu \geq K_2^F$, $\partial u_2^P(\gamma_2;\mu)/\partial x_1 \leq 0$, and the FI solution satisfies (a) $x_1^{FI} < \gamma_2$, or (b) $x_1^{FI} \geq \gamma_2$ and $u_2^P(x_1^{FI};\mu) > 0$, the PI subproblem of (A14a)-(A14h) for $x_1 \in [\gamma_2, \gamma_1]$ specializes to (A23a)-(A23g), and its solution has the following properties.*

1. *$\overline{\Pi}^I(\mu) < \Pi^{FI}(\mu)$ where $\overline{\Pi}^I(\mu) := -\infty$ if the problem is infeasible at μ.*
2. *There exists a threshold $\underline{\mu} \in [K_2^F, \infty)$ such that (A23a)-(A23g) is infeasible for $\mu \in [K_2^F, \underline{\mu})$ and feasible for $\mu > \underline{\mu}$.*
3. *For fixed $\mu > \underline{\mu}$ there exists a unique maximizer $\overline{\mathbf{x}}^I$. If $\overline{x}_1^I > \gamma_2$ then the unique optimal policy is to give absolute priority to type 1.*
4. *$\overline{\Pi}^I(\mu)$ and $\overline{\mathbf{x}}^I(\mu)$ are continuous in $\mu > \underline{\mu}$.*
5. *For $\mu \in [K_2^F, \underline{\mu}]$ we have $\overline{\Pi}^I(\mu) < \underline{\Pi}^I(\mu)$.*
6. *For $\mu > \underline{\mu}$ the solution satisfies $\overline{x}_2^I = \gamma_2$ or $\overline{\Pi}_{x_2}^I\left(\overline{\mathbf{x}}^I;\mu\right) \leq 0$.*
7. *For $\mu > \underline{\mu}$ the right derivative of the optimal revenue with respect to capacity satisfies*

$$\frac{d_+\overline{\Pi}^I(\mu)}{d\mu} \geq c \frac{N_1 \overline{x}_1^I(\mu) + N_2 \overline{x}_2^I(\mu)}{\left(\mu - N_1 \overline{x}_1^I(\mu) - N_2 \overline{x}_2^I(\mu)\right)^2}. \tag{A24}$$

Lemma 6 *Fix $r_2 > r_1$ and $\gamma_1 r_1 > \gamma_2 r_2$. Consider for $\mu \geq K_2^F$ the PI subproblem of (A14a)-(A14h) for $x_1 \in [\gamma_2, \gamma_1]$ given by (A20a)-(A20h). There exists $\mu^{FI-prio} \in (K_2^F, \infty)$ such that $\overline{\Pi}^I(\mu) = \Pi^{FI}(\mu)$ if and only if $\mu \geq \mu^{FI-prio}$.*

Lemma 7 *Fix $r_2 > r_1$ and $\gamma_1 r_1 > \gamma_2 r_2$. Consider for $\mu \geq K_2^F$ the PI subproblem of (A14a)-(A14h) for $x_1 \in [\gamma_2, \gamma_1]$ given by (A20a)-(A20h). There exists $\mu^{FI-fifo} \in (\mu^{FI-prio}, \infty)$ such that (i) For $\mu \in (\mu^{FI-prio}, \mu^{FI-fifo})$ optimality requires some priority for type-1, and (ii) for $\mu \geq \mu^{FI-fifo}$ FIFO is optimal.*

STEP 3: FIFO solution for $x_1 \leq \gamma_2$ versus strict priority solution for $x_1 > \gamma_2$

Lemma 8 *For $\mu \geq K_2^F$ let $\mathbf{x}^I(\mu)$ denote the optimal solution and $\Pi^I(\mu)$ the optimal revenue of the PI problem (A14a)-(A14h). Recall $\underline{\Pi}^I(\mu)$ is the optimal PI revenue for $x_1 \in [0, \gamma_1]$, given by the solution of (A16a)-(A16e), and $\overline{\Pi}^I(\mu)$ is the optimal PI revenue for $x_1 \in [\gamma_2, \gamma_1]$.*
Suppose there is $\mu_L > K_2^F$ such that the conditions of Lemma 4.2 hold for $\mu \leq \mu_L$, so that by Lemma 5, the problem of (A14a)-(A14h) for $x_1 \in [\gamma_2, \gamma_1]$ specializes to (A23a)-(A23g).
If $\underline{\Pi}^I(\mu_L) < \overline{\Pi}^I(\mu_L)$ then there exists a threshold $\mu_M \in (K_2^F, \mu_L)$ such that the following holds:

1. *For $\mu = \mu_M$, $\underline{\Pi}^I(\mu_M) = \overline{\Pi}^I(\mu_M) = \Pi^I(\mu_M)$ and the PI Problem (A14a)-(A14h) has exactly two solutions, one with $x_1^I(\mu_M) < \gamma_2$ and FIFO is optimal, the other with $x_1^I(\mu_M) > \gamma_2$ and the unique optimal policy is to give absolute priority to type 1.*
2. *For $\mu \in [K_2^F, \mu_M)$, $\Pi^I(\mu_M) = \underline{\Pi}^I(\mu) > \overline{\Pi}^I(\mu)$, the PI-optimal usage rates are unique with $x_1^I(\mu) < \gamma_2$ and FIFO is optimal.*
3. *For $\mu \in (\mu_M, \mu_L]$, $\underline{\Pi}^I(\mu) < \overline{\Pi}^I(\mu) = \Pi^I(\mu_M)$, the PI-optimal usage rates are unique with $x_1^I(\mu) > \gamma_2$ and the unique optimal policy is to give absolute priority to type 1.*

STEP 4: Proof of Proposition 6

First note that any PI subproblem solution with $\overline{x}_1^I > \gamma_2$ must satisfy the conditions of Lemma 4.2. This holds because the three parts of Lemma 4 are mutually exclusive and exhaustive, Lemma 4.1 implies $\overline{x}_1^I = \gamma_2$, and the conditions in Lemma 4.3 cannot hold because $\gamma_1 r_1 \leq \gamma_2 r_2$ implies $u_2^P(x_1, \mu) > 0$ for all x_1 and μ by the definition of u_2^P in (A21). Therefore the properties of the PI subproblem for $x_1 \in [\gamma_2, \gamma_1]$ that are given in Lemma 5 apply for all $\mu \geq K_2^F$.

Therefore we can apply Lemma 8 with $\mu_L = \infty$. Parts 1-3 of Lemma 8 apply if, and only if, $\lim_{\mu\to\infty} \overline{\Pi}^I(\mu) > \lim_{\mu\to\infty} \underline{\Pi}^I(\mu)$. We show $N_1\gamma_1 r_1 > N_2(\gamma_2 r_2 - \gamma_1 r_1)$ is necessary and sufficient, both for $\lim_{\mu\to\infty} \overline{\Pi}^I(\mu) > \lim_{\mu\to\infty} \underline{\Pi}^I(\mu)$ to hold, and for the optimal PI solution to serve type 1 at some capacity. Consider the solutions of the subproblems for $x_1 \leq \gamma_2$ and $x_1 \geq \gamma_2$, as $\mu \to \infty$:

For $x_1 \leq \gamma_2$ by Lemma 3 and (A12) the revenue function satisfies

$$\lim_{\mu\to\infty} \underline{\Pi}^I(\mathbf{x};\mu) = \lim_{\mu\to\infty} \Pi^{FI}(\mathbf{x};\mu) - N_2 x_1 (r_2 - r_1) = ((N_1 + N_2) r_1 - N_2 r_2) x_1 + N_2 r_2 x_2.$$

Therefore, the solution as $\mu \to \infty$ satisfies $\underline{x}_2^I = \gamma_2$, whereas we have two cases for $\underline{x}_1^I$: if $(N_1 + N_2)\, r_1 \leq N_2 r_2$ then $\underline{x}_1^I = 0$ so that $\lim_{\mu\to\infty} \underline{\Pi}^I(\mu) = N_2 r_2 \gamma_2$; however, if $(N_1 + N_2)\, r_1 > N_2 r_2$ then $\underline{x}_1^I = \gamma_2$ and $\lim_{\mu\to\infty} \underline{\Pi}^I(\mu) = (N_1 + N_2)\, r_1 \gamma_2$.

For $x_1 \geq \gamma_2$ by Lemma 5 and (A22) the revenue function satisfies

$$\lim_{\mu\to\infty} \overline{\Pi}^I(\mathbf{x};\mu) = \lim_{\mu\to\infty} \Pi^{FI}(\mathbf{x};\mu) - N_2 \lim_{\mu\to\infty} u_2^P(x_1;\mu) = (N_1+N_2) r_1 x_1 + N_2 r_2 (x_2 - \gamma_2),$$

which is maximized for $\overline{x}_1^I = \gamma_1$ and $\overline{x}_2^I = \gamma_2$, with optimal revenue $\lim_{\mu\to\infty} \overline{\Pi}^I(\mu) = (N_1+N_2) r_1 \gamma_1$.

Now compare the optimal revenues for these subproblems:

If $(N_1 + N_2)\, r_1 \leq N_2 r_2$, so that $\underline{x}_1^I = 0$, we have

$$\lim_{\mu\to\infty} \overline{\Pi}^I(\mu) > \lim_{\mu\to\infty} \underline{\Pi}^I(\mu) \Leftrightarrow (N_1+N_2) r_1 \gamma_1 > N_2 r_2 \gamma_2. \tag{A25}$$

If $(N_1 + N_2)\, r_1 > N_2 r_2$, so that $\underline{x}_1^I = \gamma_2$ then noting that $\gamma_1 > \gamma_2$ we have

$$\lim_{\mu\to\infty} \overline{\Pi}^I(\mu) = (N_1+N_2) r_1 \gamma_1 > \lim_{\mu\to\infty} \underline{\Pi}^I(\mu) = (N_1 + N_2)\, r_1 \gamma_2. \tag{A26}$$

Noting that $(N_1 + N_2)\, r_1 > N_2 r_2 \Rightarrow (N_1+N_2) r_1 \gamma_1 > N_2 r_2 \gamma_2$, the following holds:

If $(N_1 + N_2) r_1 \gamma_1 \leq N_2 r_2 \gamma_2$, then $(N_1 + N_2)\, r_1 \leq N_2 r_2$, so (A25) implies $\lim_{\mu\to\infty} \overline{\Pi}^I(\mu) \leq \lim_{\mu\to\infty} \underline{\Pi}^I(\mu)$. Therefore, $\overline{\Pi}^I(\mu) \leq \underline{\Pi}^I(\mu)$ for all $\mu \geq K_2^F$ by Lemma 8. By Lemma 3.4 serving type-1 with $x_1 \leq \gamma_2$ is not optimal for all μ, so type-1 are not served at all.

If $(N_1 + N_2) r_1 \gamma_1 > N_2 r_2 \gamma_2$, then (A25) and (A26) imply $\lim_{\mu\to\infty} \overline{\Pi}^I(\mu) > \lim_{\mu\to\infty} \underline{\Pi}^I(\mu)$. In this case Parts 1-3 of Lemma 8 apply with $\mu_L = \infty$ for a threshold $\mu_M \in (K_2^F, \infty)$. The capacity thresholds K_2^I and K_3^I and Parts 1-3 of Proposition 6 are obtained as follows.

Part 3 of Proposition 6 follows from Part 3 of Lemma 8 with $K_3^I = \mu_M$.

Parts 1 and 2 follow from Lemma 3 and from Parts 1 and 2 of Lemma 8: If $N_1 r_1 \leq N_2 (r_2 - r_1)$ then by Lemma 3.4 it is not optimal to serve type-1 with $x_1 \leq \gamma_2$ at any capacity, so $K_2^I = \mu_M$. If $N_1 r_1 > N_2 (r_2 - r_1)$ then by Lemma 3.4 it is optimal to serve type-1 with $x_1 \leq \gamma_2$ if, and only if $\mu > \mu_2'$, where $\mu_2' > K_2^F$ is defined in (A17). In this case $K_2^I = \min(\mu_2', \mu_M)$.

STEP 5: Proof of Proposition 7

Since $\gamma_1 r_1 > \gamma_2 r_2$, Part 4 is immediate from Lemma 6 with $K_4^I = \mu^{FI-prio}$ and Part 5 is immediate from Lemma 7 with $K_5^I = \mu^{FI-fifo}$.

Part 3 follows by using Lemma 8 with $\mu_L = K_4^I$: Lemma 4 and Lemma 6 imply that the conditions of Lemma 4.2 hold for $\mu \leq K_4^I$. Furthermore, by Lemma 6 the FI solution is optimal for $\mu = K_4^I$, so that $\overline{\Pi}^I(\mu) > \underline{\Pi}^I(\mu)$ for $\mu = K_4^I$. Therefore, Lemma 8 applies with $\mu_L = K_4^I$, and Part 3 of Proposition 7 follows from Part 3 of Lemma 8 with $K_3^I = \mu_M < K_4^I$.

Parts 1 and 2 follow by the same arguments as in the proof of Proposition 6. ■

Appendix B: Proofs of Proposition 5 and Lemmas 2-8

Proof of Proposition 5. For convenience we restate the PI problem (11):

$$\Pi^I = \max_{\mathbf{x},\mathbf{W},u_2} \sum_{i=1}^{2} N_i x_i \left(r_i - cW_i\right) - N_1 u_1 \tag{A27a}$$

$$\text{subject to} \quad u_1 \geq 0, \tag{A27b}$$

$$u_1 \geq x_2 \left(r_2 - r_1\right), \tag{A27c}$$

$$u_1 \leq \begin{cases} x_1 \left(r_1 - r_2\right), & x_1 \leq \gamma_2, \\ x_1 r_1 - \gamma_2 r_2 - c(x_1 - \gamma_2) W_1, & \gamma_2 \leq x_1 \leq \gamma_1, \end{cases} \tag{A27d}$$

$$W_i \geq \frac{1}{\mu - N_i x_i}, \quad \text{for } i = 1, 2, \tag{A27e}$$

$$\sum_{i=1}^{2} N_i x_i W_i \geq \frac{N_1 x_1 + N_2 x_2}{\mu - N_1 x_1 - N_2 x_2}, \tag{A27f}$$

$$0 \leq x_i \leq \gamma_i, \quad \text{for } i = 1, 2, \tag{A27g}$$

$$N_1 x_1 + N_2 x_2 < \mu. \tag{A27h}$$

Write $\mathbf{x}^I$ for the solution of (A27a)-(A27h). The proof proceeds in three steps, problem simplification, properties of simplified problem, and proof of claims.

STEP 1: Problem Simplification

To characterize $\mathbf{x}^I$ we first simplify the problem as follows. We drop (A27d) and show below in property (vi) that $\mathbf{x}^I$ satisfies (A27d). Next we use three observations:

(i) Constraint (A27b) is implied by (A27c) since $r_1 > r_2$, so at optimality, $u_1 = x_2 \left(r_2 - r_1\right)$.

(ii) The optimal scheduling policy must be work-conserving, i.e., the constraint (A27f) must bind, because by (A27a) and since $u_1 = x_2 \left(r_1 - r_2\right)$ the revenue function is decreasing in $\sum_{i=1}^{2} N_i x_i W_i$. Therefore, defining $\Pi^I \left(\mathbf{x}; \mu\right)$ to be the optimal PI revenue function, we have

$$\Pi^I \left(\mathbf{x}; \mu\right) := \Pi^{FI} \left(\mathbf{x}; \mu\right) - N_1 x_2 \left(r_1 - r_2\right). \tag{A28}$$

(iii) FIFO is optimal, that is $W_1 = W_2 = 1/(\mu - N_1 x_1 - N_2 x_2)$, because by (A28) the revenue function $\Pi^I \left(\mathbf{x}; \mu\right)$ is invariant under any work-conserving policy.

Observations $(i) - (iii)$ simplify the PI problem (A27a)-(A27h) as follows.

$$\Pi^I \left(\mu\right) = \max_{\mathbf{x}} \Pi^I \left(\mathbf{x}; \mu\right) \tag{A29}$$

$$\text{subject to} \quad 0 \leq x_i \leq \gamma_i, \quad \text{for } i = 1, 2, \tag{A30}$$

$$N_1 x_1 + N_2 x_2 < \mu. \tag{A31}$$

STEP 2: Properties of Simplified Problem

(iv) The problem (A29)-(A31) has an unique solution: From (A28) the revenue function $\Pi^I \left(\mathbf{x}; \mu\right)$ is strictly concave in $\mathbf{x}$ because $\Pi^{FI} \left(\mathbf{x}; \mu\right)$ is strictly concave in $\mathbf{x}$. The feasible region is evidently convex, and nonempty for $\mu > 0$. Noting that $\Pi^I(\mathbf{x};\mu)$ is upper-bounded on the closure of the feasible region, it follows that (A29)-(A31) has a unique maximizer for $\mu > 0$, and the first-order necessary optimality conditions are sufficient. By (A28) the partial derivatives of $\Pi^I(\mathbf{x};\mu)$ are

$$\Pi^I_{x_1}(\mathbf{x};\mu) = \Pi^{FI}_{x_1}(\mathbf{x};\mu) = N_1\left(r_1 - c\frac{\mu}{(\mu - N_1x_1 - N_2x_2)^2}\right), \tag{A32}$$

$$\Pi^I_{x_2}(\mathbf{x};\mu) = \Pi^{FI}_{x_2}(\mathbf{x};\mu) - N_1(r_1 - r_2) = N_2\left(r_2 - \frac{N_1}{N_2}(r_1 - r_2) - c\frac{\mu}{(\mu - N_1x_1 - N_2x_2)^2}\right). \tag{A33}$$

(v) If it is optimal to serve type 2, then it is optimal to fully serve type 1, that is, $x_2^I > 0$ implies $x_1^I = \gamma_1$. First, by inspection of the feasible region we have that $x_2^I > 0$ requires $\Pi^I_{x_2}(\mathbf{x}^I;\mu) \geq 0$, and that $\Pi^I_{x_1}(\mathbf{x}^I;\mu) > 0$ implies $x_1^I = \gamma_1$. Second, since $r_1 > r_2$ it follows from (A32) and (A33) that

$$\Pi^I_{x_2}(\mathbf{x};\mu) \geq 0 \Rightarrow \Pi^I_{x_1}(\mathbf{x};\mu) > 0. \tag{A34}$$

(vi) We show that the solution of (A29)-(A31) satisfies the omitted constraint (A27d). Substituting $u_1 = x_2(r_1 - r_2)$ and setting $W_1 = W = 1/(\mu - N_1x_1 - N_2x_2)$ for FIFO service, (A27d) reads

$$x_2(r_1 - r_2) \leq \begin{cases} x_1(r_1 - r_2), & x_1 \leq \gamma_2, \\ x_1(r_1 - cW) - \gamma_2(r_2 - cW), & \gamma_2 \leq x_1 \leq \gamma_1. \end{cases} \tag{A35}$$

First observe that $x_1^I > 0$ implies $r_1 - cW > 0$: since $x_1^I > 0$ implies $\Pi^I_{x_1}(\mathbf{x}^I;\mu) \geq 0$, so by (A32)

$$r_1 \geq c\frac{\mu}{(\mu - N_1x_1^I - N_2x_2^I)^2} > cW,$$

where the second inequality holds since $W_1 = 1/(\mu - N_1x_1^I - N_2x_2^I)$ for FIFO.

Now consider two cases. If the solution has $x_1^I \leq \gamma_2$, then by property (v) we have $x_2^I = 0$ so (A35) holds. If $x_1^I > \gamma_2$ then $r_1 - cW > 0$ implies that $x_1(r_1 - cW) - \gamma_2(r_2 - cW) > \gamma_2(r_1 - cW) - \gamma_2(r_2 - cW) = \gamma_2(r_1 - r_2)$, so (A35) also holds since $x_2 \leq \gamma_2$.

STEP 3: Proof of Claims

For $N_2r_2 \leq N_1(r_1 - r_2)$ serving type-2 customers is not optimal for all μ, because in this case it follows from (A33) that $\Pi^I_{x_2}(\mathbf{x};\mu) < 0$ for all $\mathbf{x}$ and μ.

Next suppose $N_2r_2 > N_1(r_1 - r_2)$. We first establish the finite threshold $K_2^I > K_2^F$. By Proposition 1 and (A32)-(A33), we have for the capacity threshold K_2^F that

$$\Pi^{FI}_{x_1}\left(\mathbf{x};K_2^F\right)\big|_{x_1=\gamma_1,x_2=0} > 0 = \Pi^{FI}_{x_2}\left(\mathbf{x};K_2^F\right)\big|_{x_1=\gamma_1,x_2=0}.$$

Since $\Pi^I_{x_1}(\mathbf{x};\mu) = \Pi^{FI}_{x_1}(\mathbf{x};\mu)$ by (A32) and $\Pi^I_{x_2}(\mathbf{x};\mu) < \Pi^{FI}_{x_2}(\mathbf{x};\mu)$ by (A33), it follows that

$$\Pi^I_{x_1}\left(\mathbf{x};K_2^F\right)\big|_{x_1=\gamma_1,x_2=0} > 0 > \Pi^I_{x_2}\left(\mathbf{x};K_2^F\right)\big|_{x_1=\gamma_1,x_2=0}.$$

Property (v) and the fact that $\Pi^I_{x_1}(\mathbf{x};\mu)$ increases in μ by (A32) imply that $x_1^I = \gamma_1$ for $\mu \geq K_2^F$.

Existence of the finite threshold K_2^I with the claimed properties follows by noting that

$$\lim_{\mu\to\infty} \Pi_{x_2}^I(\mathbf{x};\mu) = N_2\left(r_2 - \frac{N_1}{N_2}(r_1 - r_2)\right) > 0,$$

where the inequality holds since $N_2 r_2 > N_1(r_1 - r_2)$. Since $\Pi_{x_2}^I(\mathbf{x};K_2^F)\big|_{x_1=\gamma_1,x_2=0} < 0$ and $\Pi_{x_2}^I(\mathbf{x};\mu)$ is continuous and increasing in μ by (A33), it follows that there is a unique threshold K_2^I such that

$$\Pi_{x_2}^I\left(\mathbf{x};K_2^F\right)\big|_{x_1=\gamma_1,x_2=0} = 0,$$

and serving type 2 is optimal if, and only if, $\mu > K_2^I$.

To see that $\Pi^I(\mu) < \Pi^{FI}(\mu)$ for $\mu > K_2^I$, note from (A28) that $\Pi^I(\mathbf{x};\mu) < \Pi^{FI}(\mathbf{x};\mu)$ if, and only if, $x_2 > 0$. Since the FI solution satisfies $x_2^{FI} > 0$ for $\mu > K_2^F$ by Proposition 1, and $K_2^I > K_2^F$ as shown, it follows that $\Pi^I(\mu) < \Pi^{FI}(\mu)$ for $\mu > K_2^I$. ■

Proof of Lemma 2.

The proof proceeds in three steps. First, we establish basic properties of the optimal type-i rate in class-i under an optimal FUT tariff. Second, we show Part 1, i.e., (16) implies $x_{21} < x_1$ and $x_2 < x_1$. Third, we show that if Parts 1 and 2 hold, then the firm can strictly improve revenues at the FIFO-optimal usage rates by giving some priority to type-1.

STEP 1: By (14) the optimal type-i usage rate in class-i, for fixed W, P_i and limit $x_i > 0$, solves

$$\max_{y\le x_i} R_i(y) - ycW. \tag{A36}$$

Let y^* denote the maximizer of this problem. We show

$$y^* = x_i \text{ and } r_i(x_i) \ge cW, \text{ for } i = 1,2. \tag{A37}$$

Since $R_i'(y) = r_i(y)$ is strictly decreasing and $\lim_{x\to\infty} r_i(y) = 0$, we have three cases for the solution of (A36). Case 1: If $r_i(0) \le cW$ then $y^* = 0$. This case is ruled out because by hypothesis both types are served. Case 2: If $cW > r_i(x_i)$ then $y^* < x_i$ and $r_i(y^*) = cW$. This case is ruled out for an optimal FUT tariff: reducing the limit x_i to $x_i = y^*$ does not reduce type-i utility but restricts class-i usage for type $j \ne i$. Case 3: The remaining possibility is (A37).

STEP 2: We show Part 1, specifically, (16) implies $x_{21} < x_1$ and $x_2 < x_1$.

By (15) the optimal type-2 usage rate in class-1, for fixed W, P_1 and limit $x_1 > 0$, solves

$$x_{21} = \arg\max_{y\le x_1} R_2(y) - ycW,$$

which matches type-2's problem for class-2, except for the usage limit x_1. As in Case 2 of STEP 1,

$$cW > r_2(x_1) \Rightarrow x_{21} < x_1 \text{ and } r_2(x_{21}) = cW. \tag{A38}$$

Combining $cW > r_2(x_1)$ from (A38) with $r_1(x_1) \geq cW$ from (A37), we have that (16) implies $x_{21} < x_1$. Combining $r_2(x_{21}) = cW$ from (A38) with $r_2(x_2) \geq cW$ from (A37) implies $x_2 \leq x_{21}$ so that $x_2 < x_1$.

STEP 3: Suppose Parts 1 and 2 hold. For clarity we write x_1^o, x_2^o, x_{12}^o, u_2^o and W^o for the optimal values of the these variables under FIFO. Then the optimal FIFO revenue is

$$\sum_{i=1}^{2} N_i(R_i(x_i^o) - x_i^o cW) - N_2 u_2^o$$
$$= \sum_{i=1}^{2} N_i R_i(x_i^o) - c\frac{N_1 x_1^o + N_2 x_2^o}{\mu - N_1 x_1^o - N_2 x_2^o} - N_2\left(R_2(x_{12}^o) - R_1(x_1^o) + c(x_1^o - x_{21}^o)W^o\right), \tag{A39}$$

where the equality follows because $W^o = 1/(\mu - x_1^o N_1 - x_2^o N_2)$ for FIFO and by (17). Fixing usage limits and usage rates at x_i^o we show (i) revenue can be increased by setting $W_1 < W^o < W_2$, and (ii) type-i's optimal usage rate in its own class remains at the limit x_i^o, i.e., $r_i(x_i^o) \geq cW_i$ by (A37).

(i) By (A39), for fixed x_1^o and x_2^o the revenue depends on W only through u_2, and decreases in u_2. Thus we only need to show u_2 is increasing in W_1. Since type-2 optimizes its class-1 rate, x_{21}, in response to delay (see STEP 2), write $x_{21}(W_1) = \arg\max_{y \leq x_1^o} R_2(y) - ycW_1$ for type-2's optimal class-1 rate as a function of W_1, for fixed limit x_1^o. Then using (17) define

$$u_2(x_{21}(W_1), W_1) := R_2(x_{21}(W_1)) - R_1(x_1) + c(x_1 - x_{21}(W_1))W_1. \tag{A40}$$

Since $r_2(x_{21}(W_1)) = cW_1$ from (A38) we have $\partial u_2(x_{21}(W_1), W_1)/\partial x_{21} = 0$, so by (A40),

$$\left.\frac{du_2(x_{21}(W_1), W_1)}{dW_1}\right|_{W_1 = W^o} = \left.\frac{\partial u_2(x_{21}(W_1), W_1)}{\partial W_1}\right|_{W_1 = W^o} = c(x_1^o - x_{21}^o) > 0,$$

where $x_1^o - x_{21}^o > 0$ hypothesis. Therefore, reducing W_1 reduces u_2 and increases the revenue.

(ii) That $r_1(x_1^o) \geq cW_1$ is clear since $r_1(x_1^o) \geq cW^o$ at the FIFO solution and $W_1 < W^o$. For type-2 we show feasibility of $r_2(x_2^o) \geq cW_2$ for some $W_2 > W^o$ by showing $r_2(x_2^o) > cW^o$ holds at the FIFO solution. Write $x_{21}(x_2) = \arg\max_{y \leq x_1} R_2(y) - yc/(\mu - N_1 x_1 - N_2 x_2)$ for type-2's optimal class-1 rate as a function of x_2, under FIFO and fixed limit x_1. Then using (17), define

$$u_2(x_{21}(x_2), x_2) := R_2(x_{21}(x_2)) - R_1(x_1) + c\frac{x_1 - x_{21}(x_2)}{\mu - N_1 x_1 - N_2 x_2}. \tag{A41}$$

Then by the same reasoning as in (i) we have from (A38) that $\partial u_2(x_{21}(x_2), x_2)/\partial x_{21} = 0$, so by (A41),

$$\left.\frac{du_2(x_{21}(x_2), x_2)}{dx_2}\right|_{x_i = x_i^o} = \left.\frac{\partial u_2(x_{21}(x_2), x_2)}{\partial x_2}\right|_{x_i = x_i^o} = cN_2\frac{x_1 - x_{21}(x_2^o)}{(\mu - N_1 x_1^o - N_2 x_2^o)^2} > 0.$$

Finally, note the partial derivatives of the FIFO revenue function (A39) must equal zero at optimality. Substituting the FIFO delay function for W in (A39), relaxing the variable values and setting the partial derivative with respect to x_2 to zero, we have

$$N_2\left(r_2\left(x_2^o\right)-\frac{c}{\mu-N_1x_1^o-N_2x_2^o}\right)=cN_2\frac{N_1x_1^o+N_2x_2^o}{\left(\mu-N_1x_1^o-N_2x_2^o\right)^2}+N_2\left.\frac{du_2\left(x_{21}\left(x_2\right),x_2\right)}{dx_2}\right|_{x_i=x_i^o}.$$

Since the RHS is positive, this establishes $r_2\left(x_2^o\right)>cW^o$. ■

Proof of Lemma 3. By (A15) the partial derivatives of $\underline{\Pi}^I\left(\mathbf{x};\mu\right)$ satisfy

$$\underline{\Pi}_{x_1}^I\left(\mathbf{x};\mu\right)=\Pi_{x_1}^{FI}\left(\mathbf{x};\mu\right)-N_2\left(r_2-r_1\right)=N_1\left(r_1-\frac{N_2}{N_1}\left(r_2-r_1\right)-c\frac{\mu}{\left(\mu-N_1x_1-N_2x_2\right)^2}\right),\quad\text{(A42)}$$

$$\underline{\Pi}_{x_2}^I\left(\mathbf{x};\mu\right)=\Pi_{x_2}^{FI}\left(\mathbf{x};\mu\right)=N_2\left(r_2-c\frac{\mu}{\left(\mu-N_1x_1-N_2x_2\right)^2}\right),\quad\text{(A43)}$$

$$\underline{\Pi}_{\mu}^I\left(\mathbf{x};\mu\right)=\underline{\Pi}_{\mu}^{FI}\left(\mathbf{x};\mu\right)=c\frac{N_1x_1+N_2x_2}{\left(\mu-N_1x_1-N_2x_2\right)^2}.\quad\text{(A44)}$$

Part 1. Optimality of FIFO follows because by observation (iv) FIFO is optimal at every feasible $\mathbf{x}$ with $x_1\in[0,\gamma_2]$. Existence of a unique solution holds because of the following properties. The revenue function $\underline{\Pi}^I\left(\mathbf{x};\mu\right)$ is strictly concave in $\mathbf{x}$ because $\underline{\Pi}^I\left(\mathbf{x};\mu\right)=\Pi^{FI}\left(\mathbf{x};\mu\right)-N_2x_1\left(r_2-r_1\right)$ by (A15) and $\Pi^{FI}\left(\mathbf{x};\mu\right)$ is strictly concave in $\mathbf{x}$ (proof of Proposition 1). The feasible region (A16b)-(A16e) is evidently convex, and nonempty for $\mu>0$. Noting that $\underline{\Pi}^I\left(\mathbf{x};\mu\right)$ is upper-bounded on the closure of (A16b)-(A16e), it follows that (A16a)-(A16e) has a unique maximizer for $\mu>0$, and the first-order necessary optimality conditions are sufficient.

Part 2. By (A16a) we have $\underline{\Pi}^I\left(\mathbf{x};\mu\right)<\Pi^{FI}\left(\mathbf{x};\mu\right)$ if, and only if, $x_1>0$. That $\underline{\Pi}^I\left(K_2^F\right)=\Pi^{FI}\left(K_2^F\right)$ follows because by Proposition 1 and Part 4 of this lemma we have $x_1^{FI}=\underline{x}_1^I=0$, and $x_2^{FI}=\underline{x}_2^I=\gamma_2$ for $\mu=K_2^F$. That $\underline{\Pi}^I\left(\mu\right)<\Pi^{FI}\left(\mu\right)$ follows because by Proposition 1 the FI solution satisfies $x_1^{FI}>0$ for $\mu>K_2^F$.

Part 3. This follows by the maximum theorem because $\underline{\Pi}^I\left(\mathbf{x};\mu\right)$ is strictly concave in $\mathbf{x}$ and continuous in $\mathbf{x}$ and μ satisfying (A16b)-(A16e), and the feasible region (A16b)-(A16e) is convex in $\mathbf{x}$ for fixed μ and nonempty and continuous in $\mu\geq K_2^F$.

Part 4. If $N_1/N_2\leq r_2/r_1-1$ the result follows by noting from (A42) that $\underline{\Pi}_{x_1}^I\left(\mathbf{x};\mu\right)<0$ for all feasible $\mathbf{x}$ and μ. Since $N_1/N_2>r_2/r_1-1\Rightarrow r_1-N_1/N_2\left(r_2-r_1\right)>0$, it follows by inspection of (A17) and (A18) that the inequalities $\mu_2'<\mu_3'<\infty$ hold. The inequality $K_2^F<\mu_2'$ and the claims $(a)-(c)$ hold due to the following properties.

(i) At any fixed capacity, if it is optimal to serve type 1, then it is optimal to fully serve type 2, that is, $\underline{x}_1^I>0$ implies $\underline{x}_2^I=\gamma_2$. This follows from two facts. First, by inspection of the feasible region of (A16a)-(A16e) we have that $\underline{x}_1^I > 0$ requires $\underline{\Pi}_{x_1}^I(\underline{\mathbf{x}}^I;\mu) \geq 0$, and that $\underline{\Pi}_{x_2}^I(\underline{\mathbf{x}}^I;\mu) > 0$ implies $\underline{x}_2^I = \gamma_2$. Second, since $r_2 > r_1$ it follows from (A42) and (A43) that

$$\underline{\Pi}_{x_1}^I(\mathbf{x};\mu) \geq 0 \Rightarrow \underline{\Pi}_{x_2}^I(\mathbf{x};\mu) > 0. \tag{A45}$$

(*ii*) At the capacity threshold K_2^F the solution is $\underline{x}_1^I = 0$, $\underline{x}_2^I = \gamma_2$, because

$$\underline{\Pi}_{x_1}^I\left(\mathbf{x};K_2^F\right)\Big|_{x_1=0,x_2=\gamma_2} < 0 < \underline{\Pi}_{x_2}^I\left(\mathbf{x};K_2^F\right)\Big|_{x_1=0,x_2=\gamma_2}. \tag{A46}$$

The first inequality holds because $\Pi_{x_1}^{FI}(\mathbf{x};K_2^F)\big|_{x_1=0,x_2=\gamma_2} = 0$ by Proposition 1, and since $\underline{\Pi}_{x_1}^I(\mathbf{x};\mu) < \Pi_{x_1}^{FI}(\mathbf{x};\mu)$ by (A42). The second inequality holds by Proposition 1.

(*iii*) At the capacity threshold μ_2' the solution is $\underline{x}_1^I = 0$, $\underline{x}_2^I = \gamma_2$, because

$$\underline{\Pi}_{x_1}^I(\mathbf{x};\mu_2')\big|_{x_1=0,x_2=\gamma_2} = 0 < \underline{\Pi}_{x_2}^I(\mathbf{x};\mu_2')\big|_{x_1=0,x_2=\gamma_2}, \tag{A47}$$

where the equality holds by (A17) and (A42), and the inequality holds by (A45).

(*iv*) The inequality $K_2^F < \mu_2'$ and the solutions in (a) hold since $\underline{\Pi}_{x_1}^I(\mathbf{x};K_2^F)\big|_{x_1=0,x_2=\gamma_2} < 0 = \underline{\Pi}_{x_1}^I(\mathbf{x};\mu_2')\big|_{x_1=0,x_2=\gamma_2}$ by (A46) and (A47), and since $\underline{\Pi}_{x_1}^I(\mathbf{x};\mu)$ increases in μ by (A42).

(*v*) At the capacity threshold μ_3' the solution is $\underline{x}_1^I = \underline{x}_2^I = \gamma_2$, because

$$\underline{\Pi}_{x_1}^I(\mathbf{x};\mu_3')\big|_{x_1=x_2=\gamma_2} = 0 < \underline{\Pi}_{x_2}^I(\mathbf{x};\mu_3')\big|_{x_1=x_2=\gamma_2}, \tag{A48}$$

where the equality holds by (A18) and (A42), and the inequality holds by (A45).

(*vi*) The solutions in (b) and (c) follow from the equalities in (A47) and (A48), and because $\underline{\Pi}_{x_1}^I(\mathbf{x};\mu)$ decreases in x_1 and increases in μ by (A42).

Part 5. Since $\underline{\Pi}^I(\mu) = \underline{\Pi}^I(\underline{\mathbf{x}}^I(\mu);\mu)$ and $\underline{\Pi}^I(\mathbf{x};\mu)$ is differentiable in $\mathbf{x}$ and μ, the right derivative of $\underline{\Pi}^I(\mu)$ equals the right total derivative

$$\frac{d_+\underline{\Pi}^I(\underline{\mathbf{x}}^I(\mu);\mu)}{d\mu} := \sum_{i=1}^{2}\underline{\Pi}_{x_i}^I\left(\underline{\mathbf{x}}^I(\mu);\mu\right)\cdot\frac{d^+\underline{x}_i^I(\mu)}{d\mu} + \underline{\Pi}_{\mu}^I\left(\underline{\mathbf{x}}^I(\mu);\mu\right) = \underline{\Pi}_{\mu}^I\left(\underline{\mathbf{x}}^I(\mu);\mu\right),$$

where the second equality holds because the terms in the summation vanish for $\mu \geq K_2^F$ by Parts 4 and 5: The type-2 usage rate satisfies $\underline{x}_2^I(\mu) = \gamma_2$ so $d^+\underline{x}_2^I(\mu)/d\mu = 0$. For the type-1 usage rate we have for $\mu < \mu_2'$ that $\underline{x}_1^I(\mu) = 0$ and $d^+\underline{x}_1^I(\mu)/d\mu = 0$, for $\mu \geq \mu_3'$ that $\underline{x}_1^I(\mu) = \gamma_2$ and $d^+\underline{x}_1^I(\mu)/d\mu = 0$, and for $\mu \in [\mu_2', \mu_3')$ we have $\underline{\Pi}_{x_1}^I(\underline{\mathbf{x}}^I(\mu);\mu) = 0$ and $\underline{x}_1^I(\mu)$ is right differentiable by the implicit function theorem since $\underline{\Pi}^I(\mathbf{x};\mu)$ is strictly concave in $\mathbf{x}$ with continuous second order partial derivatives with respect to $\mathbf{x}$ and μ.

Then (A19) follows by using (A44) for $\underline{\Pi}_{\mu}^I(\underline{\mathbf{x}}^I(\mu);\mu)$. ■

Proof of Lemma 4.

We start with Part 3 as its proof explains the key role of $u_2^P(x_1;\mu)$.

Part 3. Since the FI solution $\mathbf{x}^{FI}$ is unique by Proposition 1, the revenue functions (A13a) and (A20a) imply that $\overline{\Pi}^I = \Pi^{FI}$ if, and only if, $\mathbf{x}^{FI} = \overline{\mathbf{x}}^I$ and $u_2 = 0$ is feasible at these usage rates. For fixed $\mathbf{x}$ we have that $u_2 = 0$ is feasible (and optimal) if, and only if, there is W_1 that satisfies (A20e), i.e., $W_1 \geq 1/(\mu - N_1 x_1)$, and such that the RHS of the IC constraint (A20c) is non-positive, i.e., $\gamma_2 r_2 - x_1 r_1 + c(x_1 - \gamma_2) W_1 \leq 0$. For $x_1 = \gamma_2$ the RHS of (A20c) is independent of W_1 and equals $\gamma_2(r_2 - r_1) > 0$, whereas for $x_1 > \gamma_2$ it increases in W_1 and is minimized when (A20e) is binding, so $W_1 = 1/(\mu - N_1 x_1)$ and type 1 get absolute priority.

Since $u_2^P(x_1;\mu)$ equals the RHS of (A20c) when (A20e) is binding, it follows that $u_2 = 0$ is feasible if and only if $u_2^P(x_1;\mu) \leq 0$, in which case $x_1 > \gamma_2$ and $\partial u_2^P(\gamma_2;\mu)/\partial x_1 < 0$ also hold. Furthermore, for $\mathbf{x}$ with $u_2^P(x_1;\mu) > 0$ the constraint (A20c) must bind at optimality, so $u_2 = u_2^P(x_1;\mu) > 0$; in this case FIFO is optimal if $x_1 = \gamma_2$, and absolute priority for type 1 is optimal if $x_1 > \gamma_2$.

Part 1. The result follows by inspection of the constraints (A20c)-(A20f), and because by (A21) the function $u_2^P(x_1;\mu)$ is continuous and strictly convex in x_1, with $u_2^P(x_1;\mu) = \gamma_2(r_2 - r_1) > 0$ for $x_1 = \gamma_2$. Specifically, if $\partial u_2^P(\gamma_2;\mu)/\partial x_1 > 0$ then $u_2^P(x_1;\mu) > \gamma_2(r_2 - r_1)$ for $x_1 > \gamma_2$, and noting from (A20e) that $u_2 \geq u_2^P(x_1;\mu)$ for all feasible W_1, it follows that (A20d) cannot hold for $x_1 > \gamma_2$. Furthermore, FIFO is optimal for $\overline{x}_1^I = \gamma_2$ as shown in the proof of Part 3.

Part 2. Suppose $\partial u_2^P(\gamma_2;\mu)/\partial x_1 \leq 0$, the problem (A20a)-(A20h) is feasible, and $\mathbf{x}^{FI}$ satisfies (a) or (b). We need to establish the solution satisfies properties $(1)-(3)$. Note property (1) follows by the proof of Part 3 and by property (2), so we only need to establish properties (2) and (3).

If $\partial u_2^P(\gamma_2;\mu)/\partial x_1 = 0$ then properties (2) and (3) hold trivially, because in this case only $\overline{x}_1^I = \overline{x}_2^I = \gamma_2$ can be feasible, by the same argument as in Part 1.

Next suppose $\partial u_2^P(\gamma_2;\mu)/\partial x_1 < 0$. We show by contradiction that (2) and (3) must hold. Suppose there is a feasible usage rate vector $\mathbf{x}^o = (x_1^o, x_2^o)$ that violates either (2) or (3), that is, either $u_2^P(x_1^o;\mu) < 0$ or $\partial u_2^P(x_1^o;\mu)/\partial x_1 > 0$ (or both). We prove that there exist more profitable usage rates $\widehat{\mathbf{x}} = (\widehat{x}_1, \widehat{x}_2)$ with $u_2^P(\widehat{x}_1;\mu) \geq 0$ and $\partial u_2^P(\widehat{x}_1;\mu)/\partial x_1 \leq 0$.

First note $\max(\gamma_2, x_1^{FI}) < x_1^o$:

(i) We have $\gamma_2 < x_1^o$: Since $u_2^P(\gamma_2;\mu) = \gamma_2(r_2 - r_1) > 0$ by (A21), $\partial u_2^P(\gamma_2;\mu)/\partial x_1 < 0$ by hypothesis and $u_2^P(x_1;\mu)$ is convex, either $u_2^P(x_1^o;\mu) < 0$ or $\partial u_2^P(x_1^o;\mu)/\partial x_1 > 0$ imply $\gamma_2 < x_1^o$.

(ii) We have $x_1^{FI} < x_1^o$: If x_1^{FI} satisfies (a), that is, $x_1^{FI} < \gamma_2$, then this is immediate since $x_1^o > \gamma_2$ by (i). If x_1^{FI} satisfies (b), that is, $x_1^{FI} \geq \gamma_2$ and $u_2^P(x_1^{FI};\mu) > 0$, then $\partial u_2^P(x_1^{FI};\mu)/\partial x_1 < 0$ as we show below. Then, since $u_2^P(x_1;\mu)$ is strictly convex in x_1, it follows from either $u_2^P(x_1^o;\mu) < 0$ or $\partial u_2^P(x_1^o;\mu)/\partial x_1 > 0$ that $x_1^{FI} < x_1^o$. The inequality $\partial u_2^P(x_1^{FI};\mu)/\partial x_1 < 0$ is implied by $\Pi_{x_1}^{FI}(\mathbf{x}^{FI};\mu) \geq 0$ (which must hold at the FI solution since $x_1^{FI} > 0$): By (A12) we have

$$\Pi_{x_1}^{FI}\left(\mathbf{x}^{FI};\mu\right) = N_1\left(r_1 - c\frac{\mu}{\left(\mu - N_1 x_1^{FI} - N_2 x_2^{FI}\right)^2}\right) \geq 0. \tag{A49}$$

Therefore

$$r_1 \geq c\frac{\mu}{\left(\mu - N_1 x_1^{FI} - N_2 x_2^{FI}\right)^2} > c\frac{\mu - N_1\gamma_2}{\left(\mu - N_1 x_1^{FI}\right)^2}, \tag{A50}$$

and using (A21) it follows that

$$\frac{\partial u_2^P\left(x_1^{FI};\mu\right)}{\partial x_1} = -\left(r_1 - c\frac{\mu - N_1\gamma_2}{\left(\mu - N_1 x_1^{FI}\right)^2}\right) < 0. \tag{A51}$$

Next, we use the facts $\gamma_2 < x_1^o$ and $x_1^{FI} < x_1^o$ to construct a feasible $\widehat{\mathbf{x}}$ as the convex combination of $\mathbf{x}^{FI}$ and $\mathbf{x}^o$. We consider the two cases, $u_2^P\left(x_1^o;\mu\right) < 0$ and $\partial u_2^P\left(x_1^o;\mu\right)/\partial x_1 > 0$, in turn:

CASE 1: If $u_2^P\left(x_1^o;\mu\right) < 0$ then choose $\widehat{\mathbf{x}}$ as the convex combination of $\mathbf{x}^{FI}$ and $\mathbf{x}^o$ such that $x_1^{FI} < \widehat{x}_1 < x_1^o$ and $u_2^P\left(\widehat{x}_1;\mu\right) = 0$. Note $\widehat{x}_1$ is the smaller root of u_2^P so $\partial u_2^P\left(\widehat{x}_1;\mu\right)/\partial x_1 < 0$.

Such $\widehat{\mathbf{x}}$ exists and is unique: $\widehat{x}_1$ is the unique root of u_2^P in the interval (γ_2, x_1^o) by convexity and continuity of $u_2^P\left(x_1;\mu\right)$ in x_1, and since $\gamma_2 < x_1^o$ and $u_2^P\left(\gamma_2;\mu\right) > 0$ by (i). Furthermore, if $x_1^{FI} < \gamma_2$ then $x_1^{FI} < \widehat{x}_1$ holds trivially; if $x_1^{FI} \geq \gamma_2$ and $u_2^P\left(x_1^{FI};\mu\right) > 0$ then $x_1^{FI} < \widehat{x}_1$ holds by convexity, since $\partial u_2^P\left(x_1^{FI};\mu\right)/\partial x_1 < 0$ by (ii).

Note that $\widehat{\mathbf{x}}$ is feasible for the FI problem (A13a)-(A13c) since its feasible region is convex and contains $\mathbf{x}^{FI}$ and $\widehat{\mathbf{x}}$. For the PI subproblem (A20a)-(A20h) feasibility holds because $\widehat{\mathbf{x}}$ also satisfies the constraints (A20b)-(A20e) and (A20g): by construction (A20c) and (A20e) bind, so $u_2 = u_2^P\left(\widehat{x}_1;\mu\right) = 0$.

CASE 2: If $\partial u_2^P\left(x_1^o;\mu\right)/\partial x_1 > 0$ then choose $\widehat{\mathbf{x}}$ as in CASE 1 if $\min_{\gamma_2 \leq x_1 \leq x_1^o} u_2^P\left(x_1;\mu\right) < 0$ (as the smaller root of u_2^P) and otherwise choose $\widehat{\mathbf{x}}$ as the convex combination of $\mathbf{x}^{FI}$ and $\mathbf{x}^o$ such that $x_1^{FI} < \widehat{x}_1 < x_1^o$ and $u_2^P\left(\widehat{x}_1;\mu\right) = \min_{\gamma_2 \leq x_1 \leq x_1^o} u_2^P\left(x_1;\mu\right) \geq 0$, so $\partial u_2^P\left(\widehat{x}_1;\mu\right)/\partial x_1 = 0$.

Such $\widehat{\mathbf{x}}$ exists and is unique by convexity and continuity of $u_2^P\left(x_1;\mu\right)$ in x_1, and since $\gamma_2 < x_1^o$ and $\partial u_2^P\left(\gamma_2;\mu\right)/\partial x_1 < 0$ by hypothesis. The argument for $x_1^{FI} < \widehat{x}_1$ is similar as in CASE 1.

That $\widehat{\mathbf{x}}$ is feasible for the FI problem follows from the same argument as in CASE 1. Next note that $\widehat{\mathbf{x}}$ also satisfies the PI subproblem constraints (A20b)-(A20e) and (A20g): By construction we have $u_2 = u_2^P\left(\widehat{x}_1;\mu\right) \geq 0$. Furthermore, $\widehat{x}_2\left(r_2 - r_1\right) \geq u_2^P\left(\widehat{x}_1;\mu\right)$ holds because $x_2^o\left(r_2 - r_1\right) \geq u_2^P\left(x_1^o;\mu\right)$ since $\mathbf{x}^o$ is feasible: we have $\widehat{x}_2 \geq x_2^o$ since $x_2^{FI} = \gamma_2$ by the property of the FI solution for $\mu \geq K_2^F$ (Proposition 1) and because $\widehat{x}_2 \in [x_2^o, x_2^{FI}]$ by construction, and the inequality $u_2^P\left(x_1^o;\mu\right) > u_2^P\left(\widehat{x}_1;\mu\right)$ holds by construction.

Finally, we show $\widehat{\mathbf{x}}$ is more profitable than $\mathbf{x}^o$, that is, $\overline{\Pi}^I\left(\widehat{\mathbf{x}};\mu\right) > \overline{\Pi}^I\left(\mathbf{x}^o;\mu\right)$. Since the FI revenue $\Pi^{FI}\left(\mathbf{x};\mu\right)$ is strictly concave in $\mathbf{x}$ with maximizer $\mathbf{x}^{FI}$, and since $\widehat{\mathbf{x}}$ is a convex combination of $\mathbf{x}^{FI}$ and $\mathbf{x}^o$, we have

$$\Pi^{FI}\left(\mathbf{x}^{FI};\mu\right) > \Pi^{FI}\left(\widehat{\mathbf{x}};\mu\right) > \Pi^{FI}\left(\mathbf{x}^o;\mu\right). \tag{A52}$$

In CASE 1, $\overline{\Pi}^I\left(\widehat{\mathbf{x}};\mu\right) > \overline{\Pi}^I\left(\mathbf{x}^o;\mu\right)$ holds because (A20a), $u_2^P\left(\widehat{x}_1;\mu\right) = 0$ and (A52) imply

$$\overline{\Pi}^{I}(\widehat{\mathbf{x}};\mu)=\Pi^{FI}(\widehat{\mathbf{x}};\mu)>\Pi^{FI}(\mathbf{x}^{o};\mu)\geq\overline{\Pi}^{I}(\mathbf{x}^{o};\mu).$$

In CASE 2, $\overline{\Pi}^{I}(\widehat{\mathbf{x}};\mu)>\overline{\Pi}^{I}(\mathbf{x}^{o};\mu)$ holds because (A20a), (A52) and $u_2^P(\widehat{x}_1;\mu)<u_2^P(x_1^o;\mu)$ imply

$$\overline{\Pi}^{I}(\widehat{\mathbf{x}};\mu)=\Pi^{FI}(\widehat{\mathbf{x}};\mu)-N_2u_2^P(\widehat{x}_1;\mu)>\Pi^{FI}(\mathbf{x}^{o};\mu)-N_2u_2^P(x_1^o;\mu)=\overline{\Pi}^{I}(\mathbf{x}^{o};\mu).$$

■

Proof of Lemma 5.

Part 1. This is immediate from Parts 2 and 3 of Lemma 4.

Part 2. Let $\underline{\mu}$ be the infimum of the set of all $\mu\geq K_2^F$ with a non-empty feasible region (A23b)-(A23g). It remains to prove that if the feasible region is non-empty at some capacity μ_s then the same holds for $\mu>\mu_s$. Only the constraints (A23b)-(A23d) and (A23g) depend on μ. By (A21) the function $u_2^P(x_1;\mu)$ is constant in μ for $x_1=\gamma_2$ and strictly decreases in μ for $x_1>\gamma_2$, and by (A51) the function $\partial u_2^P(x_1;\mu)/\partial x_1$ strictly decreases in μ for $x_1\geq\gamma_2$. Therefore increasing μ tightens only (A23b) but relaxes the other three constraints. However, since $\partial u_2^P(x_1;\mu)/\partial x_1\leq 0$ on the feasible region, it follows that at any μ_s with a non-empty feasible region there exists a feasible $\mathbf{x}$ with $x_2>0$ such that (A23d) binding, that is $u_2^P(x_1;\mu_s)=x_2(r_2-r_1)>0$. Since $u_2^P(x_1;\mu_s)$ is continuous and decreasing in μ, this $\mathbf{x}$ is still feasible for $\mu=\mu_s+\varepsilon$ for sufficiently small $\varepsilon>0$.

Part 3. The revenue function $\overline{\Pi}^{I}(\mathbf{x};\mu)$ is strictly concave in $\mathbf{x}$ because $\Pi^{FI}(\mathbf{x};\mu)$ is strictly concave in $\mathbf{x}$ (Proposition 1) and $\overline{u}_2(x_1;\mu)$ is strictly convex in x_1 (proof of Lemma 4). Convexity of $\overline{u}_2(x_1;\mu)$ in x_1 also implies that the feasible region is convex. Noting that $\overline{\Pi}^{I}(\mathbf{x};\mu)$ is upper-bounded on the closure of (A23b)-(A23g), it follows that (A23a)-(A23g) has a unique maximizer for $\mu>\underline{\mu}$, and the first-order necessary optimality conditions are sufficient. Finally, by Part 2 of Lemma 2 the unique optimal policy is to give absolute priority to type 1 if $\overline{x}_1^I>\gamma_2$.

Part 4. This follows by the maximum theorem because $\overline{\Pi}^{I}(\mathbf{x};\mu)$ is strictly concave in $\mathbf{x}$ and continuous in $\mathbf{x}$ and μ satisfying (A23b)-(A23g), and the feasible region (A23b)-(A23g) is convex in $\mathbf{x}$ for fixed μ and nonempty and continuous in $\mu>\underline{\mu}$.

Part 5. If $\underline{\mu}=K_2^F$ the claim holds because by Part 1 we have $\Pi^{FI}(K_2^F)>\overline{\Pi}^{I}(K_2^F)$ and by Part 2 of Lemma 3 we have $\underline{\Pi}^{I}(K_2^F)=\Pi^{FI}(K_2^F)$.

If $\underline{\mu}>K_2^F$ the claim follows by observing from the feasible region (A23b)-(A23g) that at least one of two conditions hold for $\mu=\underline{\mu}$:

(i) The line $N_1x_1+N_2x_2=\mu$ (the binding version of (A23g)) touches the convex set of $\mathbf{x}$ that satisfy the other constraints at a single point where $u_2^P(x_1;\mu)=x_2(r_2-r_1)$. In this case the problem is infeasible at $\underline{\mu}$ so $\overline{\Pi}^{I}(\mu)=-\infty$ for $\mu\leq\underline{\mu}$ whereas $\underline{\Pi}^{I}(\mu)>0$ for $\mu\geq K_2^F$ by Lemma 3.

(*ii*) $\partial u_2^P(\gamma_2;\mu)/\partial x_1 = 0$. In this case if the problem is feasible then $\overline{x}_1^I = \overline{x}_2^I = \gamma_2$ is the only solution (see proof of Lemma 4), this solution is also feasible in the PI subproblem (A16a)-(A16e) for $x_1 \in [0,\gamma_2]$, and yields the same revenue $\overline{\Pi}^I(\overline{\mathbf{x}}^I;\mu) = \underline{\Pi}^I(\overline{\mathbf{x}}^I;\mu)$ in both problems, because $\overline{x}_1^I = \gamma_2$ implies $u_2^P(\overline{x}_1^I;\mu) = \gamma_2(r_2 - r_1)$.

We show that $\underline{\Pi}_{x_1}^I(\overline{\mathbf{x}}^I;\mu) < 0$ which implies $\underline{\Pi}^I(\underline{\mu}) > \overline{\Pi}^I(\underline{\mu})$. The inequality $\underline{\Pi}_{x_1}^I(\overline{\mathbf{x}}^I;\mu) < 0$ follows because from (A42) we have $\underline{\Pi}_{x_1}^I(\overline{\mathbf{x}}^I;\mu) < \Pi_{x_1}^{FI}(\overline{\mathbf{x}}^I;\mu)$, from (A22) and $\partial u_2^P(\gamma_2;\mu)/\partial x_1 = 0$ we have that $\overline{\Pi}_{x_1}^I(\overline{\mathbf{x}}^I;\mu) = \Pi_{x_1}^{FI}(\overline{\mathbf{x}}^I;\mu) - N_2\partial u_2^P(\gamma_2;\mu)/\partial x_1 = \Pi_{x_1}^{FI}(\overline{\mathbf{x}}^I;\mu)$, and from (A49)-(A51) it follows that $\partial u_2^P(\gamma_2;\mu)/\partial x_1 = 0$ implies $\Pi_{x_1}^{FI}(\overline{\mathbf{x}}^I;\mu) < 0$.

Part 6. This follows by inspection of the feasible region (A23b)-(A23g): In particular, if $\overline{x}_2^{FI} < \gamma_2$ then the direction $(0,1)$ is feasible at $\overline{\mathbf{x}}^I$, so optimality requires $\overline{\Pi}_{x_2}^I(\overline{\mathbf{x}}^I;\mu) \le 0$.

Part 7. Fix μ_s and the corresponding solution $\overline{\mathbf{x}}^I(\mu_s)$. Consider the effect of an infinitesimal increase in capacity on the optimal PI revenue at $\overline{\mathbf{x}}^I(\mu_s)$, without re-optimizing the usage rates. We have two cases with respect to feasibility of $\overline{\mathbf{x}}^I(\mu_s)$ under an infinitesimal increase in μ:

(*i*) If (A23b) is slack at $\overline{\mathbf{x}}^I(\mu_s)$ then by the proof of Part 1 this solution stays feasible and

$$\begin{aligned}\overline{\Pi}_{\mu}^I\left(\overline{\mathbf{x}}^I(\mu_s);\mu_s\right) &= \Pi_{\mu}^{FI}\left(\overline{\mathbf{x}}^I(\mu_s);\mu_s\right) - N_2\frac{\partial u_2^P\left(x_1^I(\mu_s);\mu_s\right)}{\partial\mu} \\ &> \Pi_{\mu}^{FI}\left(\overline{\mathbf{x}}^I(\mu_s);\mu_s\right) = c\frac{N_1\overline{x}_1^I(\mu_s) + N_2\overline{x}_2^I(\mu_s)}{\left(\mu_s - N_1\overline{x}_1^I(\mu_s) - N_2\overline{x}_2^I(\mu_s)\right)^2}. \end{aligned} \tag{A53}$$

where the first equality holds by (A22), the inequality holds since u_2^P decreases in μ by (A21), and the second equality holds by (A44).

(*ii*) If (A23b) is tight, that is, $u_2^P\left(\overline{x}_1^I(\mu_s);\mu_s\right) = 0$, then since $u_2^P(x_1;\mu)$ decreases in μ by (A21) it follows that $\overline{\mathbf{x}}^I(\mu_s)$ violates (A23b) for $\mu > \mu_s$, so the type-2 utility $u_2 = 0 > u_2^P\left(\overline{x}_1^I(\mu_s);\mu\right)$ for $\mu > \mu_s$. Therefore $\overline{\Pi}^I\left(\overline{\mathbf{x}}^I(\mu_s);\mu\right) = \Pi^{FI}\left(\overline{\mathbf{x}}^I(\mu_s);\mu\right)$ for $\mu > \mu_s$ and by (A53)

$$\overline{\Pi}_{\mu}^I\left(\overline{\mathbf{x}}^I(\mu_s);\mu_s\right) = \Pi_{\mu}^{FI}\left(\overline{\mathbf{x}}^I(\mu_s);\mu_s\right) = c\frac{N_1\overline{x}_1^I(\mu_s) + N_2\overline{x}_2^I(\mu_s)}{\left(\mu_s - N_1\overline{x}_1^I(\mu_s) - N_2\overline{x}_2^I(\mu_s)\right)^2}.$$

It remains to show $d_+\underline{\Pi}^I(\mu_s)/d\mu \ge \overline{\Pi}_{\mu}^I\left(\overline{\mathbf{x}}^I(\mu_s);\mu_s\right)$. Note that $d_+\underline{\Pi}^I(\mu_s)/d\mu$ is well defined: Since $\overline{\Pi}^I(\mu) = \overline{\Pi}^I\left(\overline{\mathbf{x}}^I(\mu);\mu\right)$ the right derivative $d_+\underline{\Pi}^I(\mu)/d\mu$ equals the right total derivative $d_+\overline{\Pi}^I\left(\overline{\mathbf{x}}^I(\mu);\mu\right)/d\mu$. It is well defined since $\overline{\Pi}^I(\mathbf{x};\mu)$ is differentiable in $\mathbf{x}$ and μ, and the right derivatives $d^+\underline{x}_i^I(\mu)/d\mu$ exist since $\overline{\Pi}^I(\mathbf{x};\mu)$ is strictly convex in $\mathbf{x}$ and the constraint functions $u_2^P(x_1;\mu)$ and $\partial u_2^P(x_1;\mu)/\partial x_1$ are differentiable in x_1 and μ and strictly convex in x_1.

That $d_+\underline{\Pi}^I(\mu_s)/d\mu \ge \overline{\Pi}_{\mu}^I\left(\overline{\mathbf{x}}^I(\mu_s);\mu_s\right)$ is clear in case (*i*) because $\overline{\mathbf{x}}^I(\mu_s)$ is feasible for $\mu = \mu_s + \varepsilon$ with sufficiently small $\varepsilon > 0$. In case (*ii*) where $\overline{\mathbf{x}}^I(\mu_s)$ violates (A23b) for $\mu > \mu_s$, the inequality $d_+\underline{\Pi}^I(\mu_s)/d\mu \ge \overline{\Pi}_{\mu}^I\left(\overline{\mathbf{x}}^I(\mu_s);\mu_s\right)$ follows from Lemma 4.2: In particular, if the conditions of Lemma 4.2 hold for $\mu = \mu_s$ then they also hold for $\mu_l = \mu_s + \varepsilon$ for sufficiently small $\varepsilon > 0$: Specifically, the conditions of this part are $\partial u_2^P(\gamma_2;\mu)/\partial x_1 \le 0$, and the FI solution satisfies either (a) $x_1^{FI}(\mu) < \gamma_2$ or (b) $x_1^{FI}(\mu) \ge \gamma_2$ and $u_2^P(x_1^{FI}(\mu);\mu) > 0$, and if they hold for $\mu = \mu_s$ then they also hold for $\mu = \mu_s + \varepsilon$ for sufficiently small $\varepsilon > 0$, because $\partial u_2^P(\gamma_2;\mu)/\partial x_1$ decreases in μ and both $\mathbf{x}^{FI}(\mu)$ and $u_2^P(x_1^{FI}(\mu);\mu)$ are continuous in μ. It follows by Part 2 of Lemma 4 that for $\mu_l = \mu_s + \varepsilon$ the optimal solution of (A20a)-(A20h) must satisfy $u_2 = u_2^P\left(\overline{x}_1^I(\mu_l);\mu_l\right) \ge 0$. Since $u_2^P\left(\overline{x}_1^I(\mu_s);\mu_l\right) < 0$ it follows that $d_+\underline{\Pi}^I(\mu_s)/d\mu \ge \overline{\Pi}_\mu^I\left(\overline{\mathbf{x}}^I(\mu_s);\mu_s\right)$ ■

Proof of Lemma 6. Recall that $x_1^{FI}(\mu)$ is the FI-optimal type-1 usage rate as a function of μ. From Part 3 of Lemma 4 the claim is equivalent to

$$\exists\ \mu^{FI-prio} \in (K_2^F,\infty) \text{ such that } u_2^P\left(x_1^{FI}(\mu);\mu\right) \begin{cases} >0, \text{ if } \mu < \mu^{FI-prio} \wedge x_1^{FI}(\mu) \ge \gamma_2, \\ =0, \text{ if } \mu = \mu^{FI-prio}, \\ <0 \text{ if } \mu > \mu^{FI-prio}. \end{cases} \tag{A54}$$

To establish (A54) we first characterize $x_1^{FI}(\mu)$. From Proposition 1 we have for $\mu \ge K_2^F$ that the FI solution satisfies $x_1^{FI} \ge 0$, $x_2^{FI} = \gamma_2$ and

$$\Pi_{x_1}^{FI}\left(\mathbf{x}^{FI};\mu\right) = N_1\left(r_1 - c\frac{\mu}{\left(\mu - N_1 x_1^{FI} - N_2\gamma_2\right)^2}\right) \ge 0, \tag{A55}$$

with equality if $x_1^{FI} < \gamma_1$. It follows that the FI-optimal type-1 usage rate satisfies

$$\mu \ge K_2^F \ \Rightarrow\ x_1^{FI}(\mu) = \min\left(\gamma_1, \frac{1}{N_1}\left(\mu - N_2\gamma_2 - \sqrt{\frac{c\mu}{r_1}}\right)\right) \ge 0 \text{ and } \mu - N_1 x_1^{FI}(\mu) - N_2\gamma_2 > 0. \tag{A56}$$

Write $W_1^P(\mu)$ for the class-1 delay as a function of capacity, under the policy that strictly prioritizes class 1 and given the FI-optimal type-1 usage rate $x_1^{FI}(\mu)$. From (A56) we have

$$W_1^P(\mu) := \frac{1}{\mu - N_1 x_1^{FI}(\mu)} = \min\left(\frac{1}{\mu - N_1\gamma_1}, \frac{1}{N_2\gamma_2 + \sqrt{\frac{c\mu}{r_1}}}\right) > 0 \text{ for } \mu \ge K_2^F, \tag{A57}$$

and it follows from (A21) and (A57) that

$$u_2^P\left(x_1^{FI}(\mu);\mu\right) = \gamma_2 r_2 - x_1^{FI}(\mu) r_1 + c\left(x_1^{FI}(\mu) - \gamma_2\right) W_1^P(\mu). \tag{A58}$$

Below we use these properties of $x_1^{FI}(\mu)$ in characterizing $u_2^P(x_1^{FI}(\mu);\mu)$.

1. $x_1^{FI}(K_2^F) = 0$ by (A56) and the definition of K_2^F in Proposition 1.
2. $x_1^{FI}(\mu)$ is continuous and strictly increases in $\mu \ge K_2^F$ for $x_1^{FI}(\mu) < \gamma_1$:

$$\frac{d}{d\mu}x_1^{FI}(\mu) = \frac{1}{N_1}\left(1 - \frac{1}{2}\sqrt{\frac{c}{r_1\mu}}\right) > 0 \text{ if } x_1^{FI}(\mu) < \gamma_1 \text{ and } \mu \ge K_2^F, \tag{A59}$$

where the equality follows from (A56) and the inequality holds because (A55) implies $r_1/c > 1/\mu$.

3. There is a capacity threshold $\mu_{\gamma_1} < \infty$ such that $x_1^{FI}(\mu) = \gamma_1$ if and only if $\mu \geq \mu_{\gamma_1}$. This follows by (A56) and since $x_1^{FI}(\mu)$ is convex increasing for $x_1^{FI}(\mu) < \gamma_1$ by (A59).
4. There exists an unique capacity threshold $\mu_{\gamma_2} \in (K_2^F, \mu_{\gamma_1})$ such that $x_1^{FI}(\mu_{\gamma_2}) = \gamma_2$, because $x_1^{FI}(\mu)$ strictly increases in $\mu \in [K_2^F, \mu_{\gamma_1}]$ from 0 to γ_1, and since $\gamma_2 < \gamma_1$ by hypothesis.

The following properties of $u_2^P(x_1^{FI}(\mu); \mu)$ establish (A54):

1. For $\mu = \mu_{\gamma_2}$, we have $u_2^P(x_1^{FI}(\mu_{\gamma_2}); \mu_{\gamma_2}) = \gamma_2(r_2 - r_1) > 0$: The equality follows from (A58) and $x_1^{FI}(\mu_{\gamma_2}) = \gamma_2$, and the inequality holds since $r_2 - r_1 > 0$ by hypothesis.
2. We have $\lim_{\mu\to\infty} u_2^P(x_1^{FI}(\mu); \mu) = \gamma_2 r_2 - \gamma_1 r_1 < 0$: The equality holds by (A58) since $\lim_{\mu\to\infty} x_1^{FI}(\mu) = \gamma_1$ as shown above and $\lim_{\mu\to\infty} W_1^P(\mu) = 0$ by (A57), and the inequality holds by hypothesis.
3. $u_2^P(x_1^{FI}(\mu); \mu)$ is continuous, which follows from (A58) since $x_1^{FI}(\mu)$ is continuous as noted above, and $W_1^P(\mu)$ is continuous by (A57).
4. $u_2^P(x_1^{FI}(\mu); \mu)$ strictly decreases in $\mu \geq \mu_{\gamma_2}$: Since $u_2^P(x_1^{FI}(\mu); \mu)$ is continuous, it suffices to prove the claim for $\mu < \mu_{\gamma_1}$ (so $x_1^{FI}(\mu) < \gamma_1$) and $\mu > \mu_{\gamma_1}$ (so $x_1^{FI}(\mu) = \gamma_1$). From (A56)-(A58), we have

$$\frac{d}{d\mu} u_2^P\left(x_1^{FI}(\mu); \mu\right) = -\frac{d}{d\mu} x_1^{FI}(\mu)\left(r_1 - cW_1^P(\mu)\right) + c\left(x_1^{FI}(\mu) - \gamma_2\right)\frac{d}{d\mu} W_1^P(\mu) > 0, \qquad \text{(A60)}$$

where the following facts establish the inequality:

For $\mu < \mu_{\gamma_1}$ we have $x_1^{FI}(\mu) < \gamma_1$. In this case the first product on the right-hand side (RHS) of (A60) is negative, because $dx_1^{FI}(\mu)/d\mu > 0$ by (A59) and

$$r_1 - cW_1^P(\mu) = r_1 - c\frac{1}{\mu - N_1 x_1^{FI}(\mu)} > r_1 - c\frac{\mu}{\left(\mu - N_1 x_1^{FI}(\mu) - N_2\gamma_2\right)^2} \geq 0,$$

where the equality holds by (A57) and the last inequality holds by (A55). The second product on the RHS of (A60) is non-positive because $x_1^{FI}(\mu) - \gamma_2 \geq 0$ for $\mu \geq \mu_{\gamma_2}$ as shown above, and $dW_1^P(\mu)/d\mu < 0$ by (A57).

For $\mu > \mu_{\gamma_1}$ we have $x_1^{FI}(\mu) = \gamma_1$, so the first product on the RHS of (A60) is zero, and the second product is negative because $\gamma_1 - \gamma_2 > 0$ by hypothesis and $dW_1^P(\mu)/d\mu < 0$ by (A57).

■

Proof of Lemma 7. By Part 3 of Lemma 4 and by Lemma 6 and its proof, we have for $\mu \geq \mu^{FI-prio}$ that $\overline{\Pi}^I(\mu) = \Pi^{FI}(\mu)$, $\overline{\mathbf{x}}^I(\mu) = \mathbf{x}^{FI}(\mu)$ and $u_2^P(x_1^{FI}(\mu); \mu) \leq 0$, so giving absolute priority to type-1 is optimal. However, the optimal scheduling policy is not unique for $\mu > \mu^{FI-prio}$. In this case $u_2^P(x_1^{FI}(\mu); \mu) < 0$ by (A54), so for $u_2 = 0$ the constraint (A20c) is slack with $W_1 = 1/(\mu - N_1 x_1^{FI})$ and it is still satisfied if W_1 increases (so (A20e) is slack) by a sufficiently small amount. In particular, FIFO is optimal at $\overline{\mathbf{x}}^I(\mu) = \mathbf{x}^{FI}(\mu)$ if, and only if, (A20c) holds with $u_2 = 0$ and $W_1 = 1/(\mu - N_1 x_1^{FI}(\mu) - N_2 x_2^{FI}(\mu))$.

To prove the result it therefore suffices to show that for $\mu \in [\mu^{FI-prio}, \infty)$, FIFO scheduling is also optimal at the FI-optimal usage rates $\mathbf{x}^{FI}(\mu)$ if, and only if, $\mu \geq \mu^{FI-fifo}$, where $\mu^{FI-fifo} > \mu^{FI-prio}$. For $\mathbf{x}$ satisfying (A20f)-(A20h) define the function

$$u_2^F(\mathbf{x};\mu) := \gamma_2 r_2 - x_1 r_1 + c(x_1 - \gamma_2)\left(r_1 - c\frac{1}{\mu - N_1 x_1 - N_2 x_2}\right). \tag{A61}$$

It follows that FIFO is optimal (i.e. (A20c) holds with $u_2 = 0$) if, and only if, $u_2^F(\mathbf{x}^{FI}(\mu);\mu) \leq 0$. Therefore, establishing the following proves the result:

$$\exists\ \mu^{FI-fifo} \in (\mu^{FI-prio}, \infty) \text{ such that } u_2^F\left(\mathbf{x}^{FI}(\mu);\mu\right) \begin{cases} > 0,\ \mu \in [\mu^{FI-prio}, \mu^{FI-fifo}) \\ = 0\ \ \mu = \mu^{FI-fifo}, \\ < 0\ \ \mu > \mu^{FI-fifo}. \end{cases}. \tag{A62}$$

To prove (A62), we show that (i) for $\mu = \mu^{FI-prio}$ we have $u_2^F(\mathbf{x}^{FI}(\mu);\mu) > 0$, and (ii) the function $u_2^F(\mathbf{x}^{FI}(\mu);\mu)$ is continuous and strictly decreasing in μ with $\lim_{\mu\to\infty} u_2^F(\mathbf{x}^{FI}(\mu);\mu) = \gamma_2 r_2 - \gamma_1 r_1 < 0$. To this end, write $W_1^F(\mu)$ for the class-1 delay as a function of capacity, under FIFO and given the FI-optimal usage rates:

$$W_1^F(\mu) := \frac{1}{\mu - N_1 x_1^{FI}(\mu) - N_2 x_2^{FI}(\mu)} = \min\left(\frac{1}{\mu - N_1\gamma_1 - N_2\gamma_2}, \frac{1}{\sqrt{\frac{c\mu}{r_1}}}\right) > 0 \text{ for } \mu \geq K_2^F, \tag{A63}$$

where $x_1^{FI}(\mu)$ is given by (A56) and $x_2^{FI}(\mu) = \gamma_2$. Then from (A61) and (A63) we have

$$u_2^F\left(\mathbf{x}^{FI}(\mu);\mu\right) = \gamma_2 r_2 - x_1^{FI}(\mu) r_1 + c\left(x_1^{FI}(\mu) - \gamma_2\right) W_1^F(\mu). \tag{A64}$$

To show (i) and (ii) note that $W_1^F(\mu) = W_1^P(\mu - N_2\gamma_2) > W_1^P(\mu)$, where the equality holds by (A57) and (A63), and the inequality holds because $dW_1^P(\mu)/d\mu < 0$ by (A57).

Property (i) holds because by (A54) we have for $\mu = \mu^{FI-prio}$ that $u_2^P(x_1^{FI}(\mu);\mu) = 0$ and $x_1^{FI}(\mu) > \gamma_2$, therefore (A58) and (A64) and $W_1^F(\mu) > W_1^P(\mu)$ imply

$$\begin{aligned} u_2^F\left(\mathbf{x}^{FI}(\mu);\mu\right) &= u_2^F\left(\mathbf{x}^{FI}(\mu);\mu\right) - u_2^P\left(x_1^{FI}(\mu);\mu\right) \\ &= c\left(x_1^{FI}(\mu) - \gamma_2\right)\left(W_1^F(\mu) - W_1^P(\mu)\right) > 0 \text{ for } \mu = \mu^{FI-prio}. \end{aligned}$$

The properties in (ii) are easy to verify because they hold for $u_2^P(x_1^{FI}(\mu);\mu)$ (proof of Lemma 6), and $u_2^F(\mathbf{x}^{FI}(\mu);\mu)$ is obtained from $u_2^P(x_1^{FI}(\mu);\mu)$ by replacing $W_1^P(\mu)$ with $W_1^P(\mu - N_2\gamma_2)$. ■

Proof of Lemma 8.

Part 1. That there exists $\mu_M \in (K_2^F, \mu_L)$ such that $\underline{\Pi}^I(\mu_M) = \overline{\Pi}^I(\mu_M)$ follows from $\underline{\Pi}^I(\mu_L) < \overline{\Pi}^I(\mu_L)$ for $\mu_L > K_2^F$ and three facts: First, $\overline{\Pi}^I(\mu) < \underline{\Pi}^I(\mu)$ for $\mu \in [K_2^F, \underline{\mu}]$ by Part 5 of Lemma

5. Second, $\underline{\Pi}^I(\mu)$ is continuous in $\mu \geq K_2^F$ by Part 3 of Lemma 3. Third, $\overline{\Pi}^I(\mu)$ is continuous in $\mu \in (\underline{\mu}, \mu_L]$ by Part 4 of Lemma 5.

We prove $x_1^I(\mu_M) \neq \gamma_2$ by showing that, for any capacity μ, if the solution of the PI subproblem for $x_1 \in [0, \gamma_1]$ satisfies $\underline{x}_1^I = \gamma_2$, then the solution of the PI subproblem for $x_1 \in [\gamma_2, \gamma_1]$ satisfies $\overline{x}_1^I > \gamma_2$. The following holds for $\mathbf{x} = \underline{\mathbf{x}}^I$ with $\underline{x}_1^I = \gamma_2$. By Part 4(c) of Lemma 3, we have $\underline{\Pi}_{x_1}^I(\mathbf{x};\mu) \geq 0$. This implies $\Pi_{x_1}^{FI}(\mathbf{x};\mu) > 0$ by (A42). However, by (A49)-(A51) we have that $\Pi_{x_1}^{FI}(\mathbf{x};\mu) > 0$ implies $\partial u_2^P(\gamma_2;\mu)/\partial x_1 < 0$, so that from (A42) we have $\overline{\Pi}_{x_1}^I(\mathbf{x};\mu) = \Pi_{x_1}^{FI}(\mathbf{x};\mu) - N_2 \partial u_2^P(\gamma_2;\mu)/\partial x_1 > 0$. Therefore, $x_1 = \gamma_2$ cannot be optimal in the subproblem with $x_1 \in [\gamma_2, \gamma_1]$ so that $\overline{\Pi}^I(\mu) > \underline{\Pi}^I(\mu)$. Therefore $\overline{\Pi}^I(\mu_M) = \Pi^I(\mu_M)$ implies $x_1^I(\mu_M) \neq \gamma_2$.

Parts 2 and 3. The results hold because of the following property: if $\overline{\Pi}^I(\mu_s) \geq \underline{\Pi}^I(\mu_s)$ for some capacity μ_s, then $\overline{\Pi}^I(\mu) > \underline{\Pi}^I(\mu)$ for $\mu > \mu_s$, because an increase in capacity increases $\overline{\Pi}^I(\mu)$ by more than $\underline{\Pi}^I(\mu)$. This property follows from two facts.

(i) We have $N_1 \underline{x}_1^I(\mu) + N_2 \underline{x}_2^I(\mu) < N_1 \overline{x}_1^I(\mu) + N_2 \overline{x}_2^I(\mu)$. First note that $\overline{\Pi}^I(\mu_s) \geq \underline{\Pi}^I(\mu_s)$ implies $\overline{x}_1^I(\mu) > \gamma_2$. The proof of Part 1 shows this if $\overline{\Pi}^I(\mu_s) = \underline{\Pi}^I(\mu_s)$, and if $\overline{\Pi}^I(\mu_s) > \underline{\Pi}^I(\mu_s)$ this follows since $x_1 = \gamma_2$ is feasible for both PI subproblems. Therefore, if $\overline{x}_2^I(\mu) = \gamma_2$ it is immediate that $N_1 \underline{x}_1^I(\mu) + N_2 \underline{x}_2^I(\mu) < N_1 \overline{x}_1^I(\mu) + N_2 \overline{x}_2^I(\mu)$.

Consider the case $\overline{x}_2^I(\mu) < \gamma_2$. In this case we have $\overline{\Pi}_{x_2}^I\left(\overline{\mathbf{x}}^I(\mu);\mu\right) \leq 0$ by Part 6 of Lemma 5, where $\Pi_2^{FI}(\mathbf{x};\mu) = \overline{\Pi}_{x_2}^I(\mathbf{x};\mu)$ by (A22), so by (A43)

$$\frac{r_2}{c} \leq c \frac{\mu}{\left(\mu - N_1 \overline{x}_1^I(\mu) - N_2 \overline{x}_2^I(\mu)\right)^2}.$$

However, Part 4 of Lemma 3 implies $\underline{\Pi}_{x_2}^I(\underline{\mathbf{x}}^I(\mu);\mu) > 0$ for $\mu \geq K_2^F$ so by (A43)

$$\frac{r_2}{c} > c \frac{\mu}{\left(\mu - N_1 \underline{x}_1^I(\mu) - N_2 \underline{x}_2^I(\mu)\right)^2},$$

which implies that $N_1 \underline{x}_1^I(\mu) + N_2 \underline{x}_2^I(\mu) < N_1 \overline{x}_1^I(\mu) + N_2 \overline{x}_2^I(\mu)$.

(ii) Combining (i) with Part 5 of Lemma 3 and Part 7 of Lemma 5 we have

$$\begin{aligned} \frac{d_+ \overline{\Pi}^I(\mu)}{d\mu} &\geq c \frac{N_1 \overline{x}_1^I(\mu) + N_2 \overline{x}_2^I(\mu)}{\left(\mu - N_1 \overline{x}_1^I(\mu) - N_2 \overline{x}_2^I(\mu)\right)^2} \\ &> c \frac{N_1 \underline{x}_1^I(\mu) + N_2 \underline{x}_2^I(\mu)}{\left(\mu - N_1 \underline{x}_1^I(\mu) - N_2 \underline{x}_2^I(\mu)\right)^2} = \frac{d_+^I \underline{\Pi}(\mu)}{d\mu}, \end{aligned}$$

where the strict inequality holds by (i). ■