Pareto-Improving Pricing: Why 3 Is Better Than 2*

Zi Yang Kang† ⓡ Piotr Dworczak‡

August 2026

Abstract

We study the design of priority pricing systems with heterogeneous agents in environments in which improving quality for some agents reduces the average quality that can be provided. Contrary to the equity–efficiency tradeoff emphasized in public debates, we show that under economically natural conditions priority pricing can Pareto-improve on an equal-allocation benchmark. Such an improvement requires at least three priority tiers, combining higher quality for a fee, lower quality with compensation, and an intermediate tier at the benchmark quality; two tiers are never enough. Our results provide a framework for overcoming equity–efficiency tensions in applications such as lane pricing, waiting-line design, public provision, and insurance.

JEL classification: D82, D47, D63

Keywords: Pareto improvements, mechanism design, inequality-aware market design

* We thank Axel Ockenfels, Paula Onuchic, Agathe Pernoud, Bruno Strulovici, Filip Tokarski, Shosh Vasserman, Frank Yang, and several seminar audiences for helpful discussions. We gratefully acknowledge the support received under the ERC Starting Grant IMD-101040122.

† Department of Economics, University of Toronto; zy.kang@utoronto.ca.

‡ Department of Economics, Northwestern University, and Group for Research in Applied Economics (GRAPE); piotr.dworczak@northwestern.edu.

1 Introduction

Many scarce resources—such as space on highways, bandwidth in communication networks, and access to public services—are allocated through waiting rather than explicit prices. Since at least Pigou (1920), economists have emphasized the inefficiency of such time-based allocation: waiting is socially costly, and many individuals would pay to avoid it, creating scope for gains from trade. Yet pricing is often resisted on equity grounds because it is seen as replacing time-based rationing with income-based rationing.

Public debates surrounding "pay-to-skip-the-line" systems vividly reflect this concern. On roads, critics deride priority toll lanes as "Lexus lanes" that allow the rich to buy their way out of congestion while others bear the cost of greater delays (Anderson, 2005). At airports, paid access to expedited security lines—such as CLEAR—has been portrayed as allowing some passengers to cut in line (Mull, 2023) and as "splitting travelers into haves and have-nots" (Stewart, 2022). Similar concerns arise when governments offer fast-track treatment for payment. Critics of investor-citizenship programs, for example, describe them as allowing wealth to purchase accelerated access to a public status for which ordinary applicants must wait (Shachar, 2018).

Our main insight in this paper is that this equity–efficiency tradeoff need not be as stark as these debates suggest. We define the *equity benchmark* as an allocation in which all individuals receive the same quality (e.g., face the same wait time). We assume that departures from this benchmark are costly: giving some individuals higher quality requires both giving others lower quality and lowering quality on average. Even so, prices and priority can be introduced in a way that *Pareto-improves* on the equity benchmark, but only by moving beyond the two-tier systems that dominate practice. With only two tiers—a paid "fast lane" and a free regular option—no Pareto improvement is possible: even if the revenue from selling priority is redistributed, some individuals must be worse off than under the equity benchmark. With three tiers, by contrast, every individual can be made strictly better off. Our construction combines a paid high-priority tier, a compensated low-priority tier, and an intermediate tier that preserves the benchmark allocation. Together, these results help explain why simple priority systems generate backlash: without a third tier, unequal priority necessarily harms some individuals. At the same time, our framework identifies designs under which everyone benefits from the introduction of prices and priority.

Our formal model is stylized but flexible enough to cover a wide range of applications. We consider a designer who allocates potentially different qualities of a good to a unit mass of agents with different marginal rates of substitution between money and quality. In lane pricing, for example, quality decreases with travel time, and agents vary in their distaste for waiting. Under the equity benchmark, every agent receives the same quality and makes no monetary transfer. The designer seeks to replace this benchmark with a pricing mechanism—a menu of qualities and monetary transfers—that introduces dispersion in quality. Application-specific feasibility constraints determine which quality distributions can be implemented. Rather than model these constraints directly, we impose three assumptions on the feasible set of quality distributions.

First, we assume there is a tradeoff—what we call *friction*—between dispersion and average quality. Specifically, the degenerate quality distribution associated with the equity benchmark uniquely maximizes average quality. This assumption captures the idea that reallocating quality to create dispersion is costly. For example, if travel time is convex in lane traffic, unequal lane loads create dispersion in travel times while raising average travel time. In other applications, friction may arise from curvature in preferences, such as risk aversion, or from heterogeneity in costs, as under adverse selection.

Second, we assume that this mean–dispersion tradeoff is locally mild. More precisely, the loss in average quality from creating dispersion can be made arbitrarily small relative to the aggregate quality gain received by agents allocated qualities above the equity benchmark. Small departures from equal lane loads, for instance, generate first-order dispersion but only a higher-order increase in average travel time.

Third, we make the standard assumption that the designer can randomize assignments. Formally, if a quality distribution is feasible, then every mean-preserving contraction of that distribution is also feasible.[1]

Under these assumptions, we prove two main results. On the one hand, no mechanism with only two tiers—that is, two quality levels—can Pareto-improve on the equity benchmark. Indeed, the agent who is indifferent between the high- and low-quality options is necessarily worse off. On the other hand, there exists a three-tier mechanism that strictly Pareto-improves on the equity benchmark. Thus, three tiers are the minimum required to make everyone better off—a modest but practically feasible level of complexity.

[1] In our applications, we also establish when the Pareto improvement can be implemented without randomizing assignments.

Our applications illustrate the breadth and flexibility of the framework. Our assumptions, and hence the main results, apply across environments with different allocation technologies and sources of friction. We also show how both the framework and the results can be generalized beyond these assumptions—for example, to settings in which the resource cost of quality depends on the recipient.

The remainder of this paper is organized as follows. We conclude this section by reviewing the related literature. Section 2 introduces the formal model, and Section 3 presents the main results. Section 4 applies and generalizes these results, and Section 5 concludes.

1.1 Related Literature

The pricing of priority in access to goods and services—and its distributional consequences—has received sustained attention in economics. As we discuss below, several papers show that priority pricing can Pareto-improve on equal access or laissez-faire in particular environments.

Our contribution to this literature is twofold. First, we develop a novel framework for modeling the supply side through the set of feasible quality distributions. This approach accommodates a wide range of environments while isolating the mean–dispersion tradeoff at the heart of our results. Second, to the best of our knowledge, our paper is the first to show that *three*, rather than two, priority tiers are necessary to achieve a Pareto improvement in environments characterized by such a tradeoff.

Chao and Wilson (1987) are perhaps the first to observe that selling priority to agents with higher willingness to pay and redistributing the resulting revenue may generate a Pareto improvement. Gershkov and Schweinzer (2010) study a related model in which agents differ in marginal utility of time and hold property rights to specific positions in a queue. They characterize when these positions can be efficiently reordered through a Vickrey–Clarke–Groves mechanism while leaving every agent weakly better off, extending the classical analysis of Cramton, Gibbons and Klemperer (1987) to settings with multiple goods. Most relevant for us, they show that a random queue order—which corresponds to our equity benchmark—can be Pareto-improved upon.[2] These analyses, however, do not feature the mean–dispersion tradeoff that drives our distinction between two and three tiers. With linear disutility from waiting and

[2] Gershkov and Winter (2023) make the complementary point that priority service can uniformly harm consumers when access is controlled by a profit-maximizing monopolist.

no supply-side friction, priority can be reassigned without reducing average quality.

A sizable literature studies the distributional consequences of congestion pricing in the bottleneck model of Vickrey (1969). Arnott, De Palma and Lindsey (1994), van den Berg and Verhoef (2011), Hall (2018), and Bobbio, Cramton and Ockenfels (2021) identify conditions under which congestion pricing yields a Pareto improvement. These models differ from ours in two respects. First, agents choose their departure times, a margin absent from our framework that is often central to obtaining a Pareto improvement. Second, they do not generally satisfy our friction assumption: pricing may reduce congestion enough to improve outcomes for all drivers even before revenue is redistributed. For example, Hall (2018) shows that pricing can reduce travel times for all drivers by eliminating "hypercongestion," which arises when too many vehicles use the road at the same time. Our friction assumption, by contrast, requires any dispersion in quality to reduce average quality relative to the equity benchmark. Pareto improvements must therefore rely on the sorting gains generated by pricing and the redistribution of the resulting revenue.[3]

More broadly, our paper contributes to a growing market-design literature on congestion pricing, including work by Ostrovsky and Schwarz (2018), Cramton, Geddes and Ockenfels (2019), Beheshtian, Geddes, Rouhani, Kockelman, Ockenfels, Cramton and Do (2020), and Ostrovsky and Yang (2024).[4]

Finally, our paper relates to the recent literature on *inequality-aware market design*. Building on classical work by Weitzman (1977), Spence (1977), and Nichols and Zeckhauser (1982), this literature studies mechanism design under redistributive objectives; contributions include work by Condorelli (2013), Dworczak ⓡ Kominers ⓡ Akbarpour (2021), Kang (2023), and Akbarpour ⓡ Dworczak ⓡ Kominers (2024), among many others. The paper most closely related to ours within this literature is Kang and Watt (2026). They study public provision when agents also have access to a private market. The private option imposes a constraint mathematically similar to Pareto improvement. While this literature characterizes optimal mechanisms for a specified social welfare function, we instead identify conditions under which a Pareto improvement exists without specifying interpersonal welfare comparisons.

[3] Because our positive result is strict, it remains valid when transfers are imperfect and a sufficiently small share of revenue is dissipated.

[4] An active empirical literature also quantifies the welfare effects of congestion pricing; see, for example, Hall (2021), Kreindler (2024), Cook and Li (2025), Cook, Kreidieh, Vasserman, Allcott, Arora, Tomkins, Turkel and van Sambeek (2026), Ater, Ross, Shany and Vasserman (2026).

2 Framework

We develop an abstract framework for the design of Pareto-improving pricing systems. The framework is intentionally parsimonious: it is broad enough to encompass a wide range of applications, yet sufficiently structured to isolate the key economic forces that drive our results.

2.1 Setup

There is a unit mass of agents, each of whom demands one unit of a good. Units of the good may differ in their physical quality $q \in \mathbf{R}_+$. Each agent privately observes his type r. The distribution of types in the population is G, which has positive continuous density g on a compact interval $[\underline{r}, \bar{r}] \subset \mathbf{R}_+$. An agent of type r who is assigned a unit of physical quality q and makes a payment $p \in \mathbf{R}$ obtains utility

$$rv(q) - p,$$

where a negative payment is interpreted as a rebate to the agent. The function $v : \mathbf{R}_+ \to [0, 1]$ is assumed to be continuously differentiable, increasing,[5] and concave, with $v(0) = 0$.

We refer to $Q := v(q)$ as the good's *effective quality* (or simply *quality*, when there is no ambiguity). Agent utility is $rQ - p$, so the agent's type is his marginal rate of substitution between effective quality and money. The distinction between physical and effective quality separates the objective characteristics of the good from their utility consequences for agents. In particular, agents are risk-neutral over lotteries in effective quality, but may be risk-averse over lotteries in physical quality (when v is strictly concave).

We use the following running example throughout the paper to illustrate the abstract framework.

Example. *Agents are commuters traveling from a common origin to a common destination. A trip entails time $t \in [\underline{t}, \bar{t}]$ spent in traffic, where $0 \leq \underline{t} < \bar{t} < 1$. Normalizing the physical quality of a trip with no time spent in traffic to 1, define the physical quality of a trip with travel time t as $q = 1 - t$. Thus, shorter travel time corresponds to higher physical quality. When v is strictly concave, agents are risk-averse over travel time: they strictly prefer a certain travel time to any nondegenerate lottery with the same mean.*

[5] Throughout, we use "increasing" and "decreasing" to mean strictly increasing and strictly decreasing, respectively; weak monotonicity is denoted by "nondecreasing" and "nonincreasing."

2.2 Outcomes

A designer assigns each agent a possibly degenerate lottery over physical qualities. Because utility is linear in effective quality Q, an agent's payoff depends on his assigned lottery only through its expected quality. We can therefore summarize the physical allocation by a distribution $F \in \Delta([0,1])$ of expected qualities available for assignment, where $1 - F(Q)$ is the mass of agents who can be assigned expected quality strictly greater than Q.[6]

Different applications restrict the feasible distributions of expected qualities in different ways. We represent these restrictions by a set $\mathcal{F} \subseteq \Delta([0,1])$ from which the designer must select F.[7] The set $\mathcal{F}$ is the central object of our framework; Section 2.4 introduces assumptions on $\mathcal{F}$ that generate the tradeoff between dispersion and average quality underlying our results.

Each feasible distribution $F \in \mathcal{F}$ uniquely determines, up to a G-null set, a budget-balanced and incentive-compatible deterministic reduced-form mechanism in expected qualities and payments.[8] Incentive compatibility requires expected quality to be nondecreasing in type (Myerson, 1981). Hence, the allocation rule is

$$Q(r) = F^{-1}(G(r)),$$

the unique nondecreasing allocation rule whose distribution is F, up to G-null sets. Given this allocation rule, the envelope formula determines payments up to an additive constant (Milgrom and Segal, 2002):

$$p(r) = rQ(r) - U - \int_{\underline{r}}^{r} Q(s)\,\mathrm{d}s,$$

where U is uniquely chosen to satisfy budget balance:

$$\int_{\underline{r}}^{\overline{r}} p(r)\,\mathrm{d}G(r) = 0.$$

[6] This representation remains without loss of generality when goods are scarce: an agent who receives no good can be treated as receiving physical quality zero.

[7] For simplicity, we abstract from variable production costs. Section 4.4 shows how our results accommodate such costs.

[8] Here, "deterministic" refers to the reduced-form pair (Q, p) assigned to each type and does not rule out randomization over physical qualities, which can be incorporated into $\mathcal{F}$. This restriction is without loss because utility is linear in expected quality and payments.

We say that the resulting mechanism $(Q(\cdot), p(\cdot))$ *implements* F.

2.3 Pareto Improvements

We define the *equity benchmark* as the degenerate quality distribution δ_{Q_0}, where

$$Q_0 \coloneqq \sup_{F \in \mathcal{F}} \mathbf{E}_F[Q].$$

We assume that $\delta_{Q_0} \in \mathcal{F}$.[9] Under the equity benchmark, average quality is maximized while every agent receives expected quality Q_0 and pays zero; the resulting direct mechanism is budget-balanced and incentive-compatible. As we show in Section 4, this benchmark often coincides with the laissez-faire non-price allocation—that is, the equilibrium allocation in a system without monetary transfers.

The designer seeks to construct a Pareto improvement over the equity benchmark. For any distribution $F \in \mathcal{F}$, we say that F is a *Pareto improvement over the equity benchmark* (or simply a *Pareto improvement*) if its implementing mechanism $(Q(\cdot), p(\cdot))$ satisfies

$$rQ(r) - p(r) \geq rQ_0, \qquad \text{for every } r \in [\underline{r}, \overline{r}], \tag{PI}$$

with strict inequality for a positive mass of agents. We say that F is a *strict Pareto improvement* if every agent strictly prefers its implementing mechanism to the equity benchmark—that is, if the inequality in (PI) is strict for every type $r \in [\underline{r}, \overline{r}]$.

Our focus on Pareto improvements reflects an intentionally conservative welfare criterion. We make no interpersonal comparisons of utility and posit no social welfare function. Unlike much of the existing literature, which studies optimal allocations under explicit welfare objectives, we ask only whether prices can realize Pareto gains, subject to budget balance and incentive compatibility, without invoking distributional judgments.

[9] This assumption is without loss of generality if $\mathcal{F}$ is compact and the designer can randomize.

2.4 Feasible Policies

We complete the description of our framework by imposing three economically motivated restrictions on the set of feasible expected-quality distributions $\mathcal{F} \subseteq \Delta([0,1])$. We first motivate these restrictions using our running example.

Example. *There are $N \geq 2$ lanes on a highway. If a mass m_i of agents is assigned to lane i, each agent in that lane experiences travel time $w(m_i)$, where $w : [0,1] \to [\underline{t}, \bar{t}]$ is continuously differentiable, increasing, and convex. These assumptions capture congestion effects: travel time increases with lane usage at a nondecreasing rate.*

Given a vector of lane loads $\mathbf{m} = (m_1, \ldots, m_N) \in \Delta^{N-1}$, the distribution of travel times is $\sum_{i=1}^{N} m_i \delta_{w(m_i)}$. Because travel time t corresponds to effective quality $v(1-t)$, the induced distribution of effective qualities is

$$F^{\mathbf{m}} := \sum_{i=1}^{N} m_i \delta_{v(1-w(m_i))}.$$

Starting from any lane-load vector $\mathbf{m}$, the designer can randomize agents' assignments across lanes while preserving the marginal distribution of realized physical qualities. The distributions of expected qualities that can be generated this way are precisely the mean-preserving contractions of $F^{\mathbf{m}}$, which we denote by $\mathrm{MPC}(F^{\mathbf{m}})$.[10] *Consequently, the feasible set of expected quality distributions is*

$$\mathcal{F} = \left\{ F \in \Delta([0,1]) : F \in \mathrm{MPC}(F^{\mathbf{m}}) \text{ for some } \mathbf{m} \in \Delta^{N-1} \right\}.$$

Rather than committing to a particular technology, we impose three axioms on the feasible set $\mathcal{F}$.

Axiom 1 (Randomization)**.** *The feasible set $\mathcal{F}$ is closed under mean-preserving contractions: if $F \in \mathcal{F}$, then* $\mathrm{MPC}(F) \subseteq \mathcal{F}$.

[10] Strassen's theorem (Müller and Stoyan, 2002, Theorem 3.4.2) implies that there exist random variables X and $\bar{X}$ such that $X \sim F$, $\bar{X} \sim \bar{F}$, and $\mathbf{E}[X \mid \bar{X}] = \bar{X}$ if and only if $\bar{F}$ is a mean-preserving contraction of F. Here, X is realized effective quality and $\bar{X}$ is expected effective quality. For any $F \in \Delta([0,1])$, define

$$\mathrm{MPC}(F) := \left\{ \bar{F} \in \Delta([0,1]) : \int_0^x \bar{F}^{-1}(u)\,\mathrm{d}u \geq \int_0^x F^{-1}(u)\,\mathrm{d}u \text{ for every } x \in [0,1], \text{ with equality at } x = 1 \right\}.$$

Axiom 1 formalizes the designer's ability to randomize assignments. Starting from any feasible distribution, the designer can reduce the dispersion in agents' expected qualities without changing their average. The axiom holds by construction in the running example.

Axiom 2 (Friction)**.** *The equity benchmark δ_{Q_0} uniquely maximizes average quality: for every $F \in \mathcal{F}$, if $F \neq \delta_{Q_0}$, then $\mathbf{E}_F[Q] < Q_0$.*

Axiom 2 captures a friction in the allocation technology: any feasible departure from the equity benchmark lowers average quality. The axiom therefore imposes a local tradeoff between dispersion and average quality around the benchmark, thereby making improvements over the equity benchmark more difficult to achieve.

The axiom holds in the running example under the maintained assumptions. To see this, define the function $\phi(m) := mv(1 - w(m))$, and note that the average quality generated by a lane-load vector $\mathbf{m}$ is $\sum_{i=1}^{N} \phi(m_i)$. Since w is increasing and convex and v is increasing and concave, ϕ is strictly concave. By Jensen's inequality, for any $\mathbf{m} \in \Delta^{N-1}$,

$$\sum_{i=1}^{N} \phi(m_i) \leq N\phi\left(\frac{1}{N}\right) = v\left(1 - w\left(\frac{1}{N}\right)\right),$$

with equality if and only if $m_i = 1/N$ for every $i \in \{1, \ldots, N\}$. Hence, $Q_0 = v(1 - w(1/N))$ satisfies the requirements of the axiom, and the equal assignment of agents to lanes induces the unique feasible distribution with average quality Q_0. Intuitively, because travel times increase with lane usage at a nondecreasing rate and agents are weakly risk-averse over travel time, equalizing lane usage uniquely maximizes average effective quality.[11]

Axiom 3 (Smoothness)**.** *The tradeoff between dispersion and average quality around the equity benchmark δ_{Q_0} is locally mild: for every $\varepsilon > 0$, there exists $F \in \mathcal{F}$ such that*

$$\mathbf{E}_F[(Q - Q_0)_+] > 0 \quad \textit{and} \quad \frac{Q_0 - \mathbf{E}_F[Q]}{\mathbf{E}_F[(Q - Q_0)_+]} \leq \varepsilon. \tag{S}$$

[11] Note that neither strict risk aversion nor strictly convex congestion effects are required for Axiom 2 to hold.

Axiom 3 provides a counterpoint to Axiom 2. Whereas friction requires every feasible departure from the equity benchmark to lower average quality, smoothness requires that this loss, $Q_0 - \mathbf{E}_F[Q]$, can be made arbitrarily small relative to the cumulative improvement $\mathbf{E}_F[(Q - Q_0)_+]$ above Q_0. If the uniqueness requirement in Axiom 2 were violated, so that a nondegenerate feasible distribution had average quality Q_0, condition (S) would hold trivially.

Axiom 3 also holds in the running example. Starting from equal lane loads, transfer a mass η of agents from one lane to another. The first-order effects on average quality cancel because w and v are continuously differentiable. The resulting loss in average quality is therefore $o(\eta)$, whereas the cumulative improvement over Q_0 is of order η because effective quality is locally strictly decreasing in lane usage. Hence, the ratio in condition (S) converges to zero as $\eta \downarrow 0$. Section 4.1 provides a formal proof.

3 Main Results

Our main results characterize the minimum number of tiers required to construct a Pareto improvement. To formalize simplicity in a policy-relevant way, we count the number of distinct expected-quality levels offered by the mechanism.

Definition 1. *A distribution $F \in \mathcal{F}$ offers n tiers if $|\mathrm{supp}(F)| = n$.*

Any Pareto improvement must offer at least two tiers: a one-tier mechanism assigns the same quality $Q \leq Q_0$ to every agent, while incentive compatibility and budget balance require payments to be zero. We therefore characterize the minimum number of tiers required for a Pareto improvement.

3.1 Statement and Discussion of Main Results

Our first result shows that two tiers do not suffice.

Theorem 1. *Under Axiom 2, there does not exist a Pareto improvement with two tiers.*

Our second result shows that three tiers suffice.

Theorem 2. *Under Axioms 1 and 3, there exists a strict Pareto improvement with three tiers.*

Together, Theorems 1 and 2 establish that, under our three axioms, three is the minimum number of tiers required for a Pareto improvement.

These results are particularly relevant because two-tier priority systems are widespread in practice. Examples include express lanes on roads and fast-track options in airport security, visa processing, and health care. Such systems are appealing for their simplicity and are often defended as a way to generate allocative gains by allowing agents who value priority more highly to pay for it.

In this light, Theorem 1 is a negative result: any nontrivial two-tier priority system must make some agents strictly worse off than under the equity benchmark. Intuitively, the high tier benefits high types by providing higher quality, while payments collected from its users can be rebated to low types. With only two tiers, however, the same threshold type governs both the difference in quality and the difference in payments. The mechanism therefore cannot compensate types near this threshold for the friction-induced loss in average quality. Our proof below formalizes this intuition. It also shows that even without Axiom 2—when dispersion is free—a two-tier priority system cannot yield a *strict* Pareto improvement.

The negative result of Theorem 1 therefore helps explain why many two-tier priority systems have generated public backlash. When congestion or other frictions create a tradeoff between dispersion and average quality, allowing some users to purchase priority necessarily harms others, even if the resulting revenue is redistributed. Our result formalizes this distributional concern. Moreover, if type measures the marginal value of reduced waiting time relative to the marginal value of money (e.g., Dworczak ℗ al., 2021), the intermediate types harmed by a two-tier system may include low- and middle-income users who place a high value on time but also have a high marginal value of money.

Theorem 2 provides the positive counterpart: three tiers are sufficient to make every agent *strictly* better off. We construct a mechanism with an intermediate tier that provides expected quality exactly equal to Q_0. This tier separates the threshold between the low and intermediate tiers from the threshold between the intermediate and high tiers. Axiom 3 ensures that the loss in average quality can be made small relative to the cumulative improvement above Q_0. The separation between the two thresholds then converts these quality improvements into sorting gains that are unavailable with only two tiers. In turn, these gains can be redistributed through payments so that every agent, including those assigned to the low and intermediate tiers, is strictly better off.

This positive result has direct implications for policy. Under our three axioms, achieving a Pareto improvement does not require a complex or finely targeted pricing system: three appropriately designed priority tiers suffice. Policy debates framed as a binary choice between uniform service and a two-tier priority system are therefore unnecessarily restrictive. An intermediate tier that preserves the benchmark quality can allow the gains from priority pricing to be shared by all agents while remaining simple.

Example. *Because our running example satisfies Axioms 1–3, Theorems 1 and 2 apply. A conventional two-tier lane-pricing system cannot Pareto-improve on equal access: creating one faster lane necessarily leaves some commuters strictly worse off, even if toll revenues are redistributed. By contrast, a three-tier system can make every commuter strictly better off by combining a fast lane, a slow lane, and an intermediate lane that preserves the laissez-faire travel time. Section 4.1 returns to this application and studies its implementation in more detail.*

3.2 Proofs of Main Results

The idea underlying the proof of Theorem 1 is illustrated in Figure 1, which depicts agents' indirect utility as a function of their type r. Consider a two-tier priority system whose higher quality is above Q_0 (the second quality level must then be below Q_0). Let r_0 denote the threshold type who is indifferent between the two tiers. Incentive compatibility requires the difference in prices between the two tiers to equal r_0 times the difference in quality. Budget balance then determines the levels of the two prices so that, for type r_0, the price adjustment on either tier exactly offsets the difference between that tier's quality and the mechanism's average quality. Hence, r_0 is as well off as if she received the average quality at zero price. By Axiom 2, that average quality must be strictly below Q_0, so type r_0 is strictly worse off than under the equity benchmark.

Proof of Theorem 1. Consider any feasible two-tier distribution that assigns mass $\alpha \in (0, 1)$ to a low quality Q_L and mass $1 - \alpha$ to a high quality Q_H:

$$F = \alpha \delta_{Q_L} + (1 - \alpha)\, \delta_{Q_H}.$$

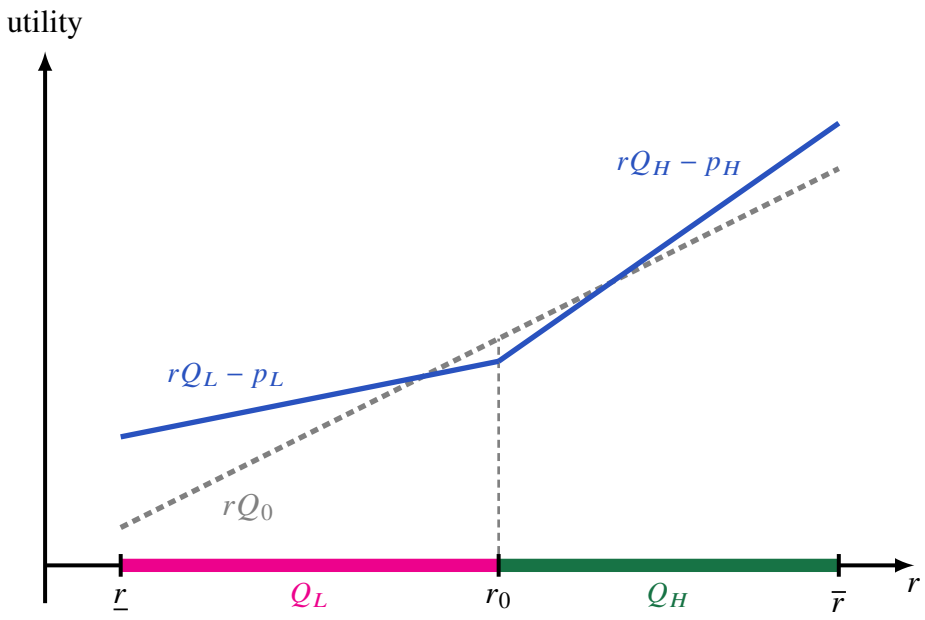

Figure 1: Agent utility in a two-tier priority system. The utility schedule lies below benchmark utility at r_0.

Let p_L and p_H denote the corresponding payments, and let r_0 denote the threshold type, so that $G(r_0) = \alpha$. Incentive compatibility requires this type to be indifferent between the two tiers:

$$r_0 Q_L - p_L = r_0 Q_H - p_H \iff p_H = p_L + r_0 (Q_H - Q_L) .$$

Combining this equality with budget balance yields

$$0 = \alpha p_L + (1 - \alpha) p_H = p_L + (1 - \alpha) r_0 (Q_H - Q_L) \implies -p_L = (1 - \alpha) r_0 (Q_H - Q_L) .$$

Consequently, the threshold type's utility is

$$r_0 Q_L - p_L = r_0 [\alpha Q_L + (1 - \alpha) Q_H] < r_0 Q_0. \tag{1}$$

The strict inequality in equation (1) follows from Axiom 2 and $r_0 > 0$ since F is nondegenerate. Hence, the threshold type (and by continuity, a positive mass of types in the neighborhood of r_0) is strictly worse off than under the equity benchmark. □

Even without Axiom 2, the definition of Q_0 implies that the inequality in (1) holds weakly. The threshold type therefore cannot be strictly better off, yielding the following corollary.

Corollary 1. *There does not exist a strict Pareto improvement with two tiers.*

The idea underlying the proof of Theorem 2 is illustrated in Figure 2. We first use Axiom 3 to select a feasible distribution $F \in \mathcal{F}$ that creates dispersion around the equity benchmark with a small loss in average quality relative to the cumulative improvement above Q_0. We then use Axiom 1 to contract F into a three-tier distribution whose intermediate quality is exactly Q_0. Let r_L and r_H denote the threshold types between the low and intermediate tiers and between the intermediate and high tiers, respectively. Relative to the equity benchmark, utility decreases with type below r_L, remains constant between r_L and r_H, and increases above r_H, so the intermediate-tier agents are the worst off. It therefore suffices to show that they receive a subsidy. Because $r_H > r_L$, quality improvements at the top can be priced at a higher marginal willingness to pay than is required to compensate for quality reductions at the bottom—a separation unavailable in the two-tier case. As we show below, these sorting gains are large enough to finance a subsidy for the intermediate-tier agents.[12]

Proof of Theorem 2. Let $\varepsilon \in (0, 1]$, to be specified below. By Axiom 3, there exists a feasible distribution $F \in \mathcal{F}$ such that $\mathbf{E}_F[(Q - Q_0)_+] > 0$ and

$$\frac{Q_0 - \mathbf{E}_F[Q]}{\mathbf{E}_F[(Q - Q_0)_+]} \leq \varepsilon. \tag{2}$$

By definition, $Q_0 \geq \mathbf{E}_F[Q]$. It follows that

$$\mathbf{E}_F[(Q_0 - Q)_+] = \underbrace{\mathbf{E}_F[(Q - Q_0)_+]}_{>0} + \underbrace{Q_0 - \mathbf{E}_F[Q]}_{\geq 0} > 0.$$

We now explicitly construct a three-tier Pareto improvement via mean-preserving contractions of F.

[12] The key technical challenge is to use only Axioms 1 and 3 to construct a three-tier distribution whose intermediate tier has a mass bounded away from zero, ensuring that $r_H - r_L$ is also bounded away from zero and can generate the required gains.

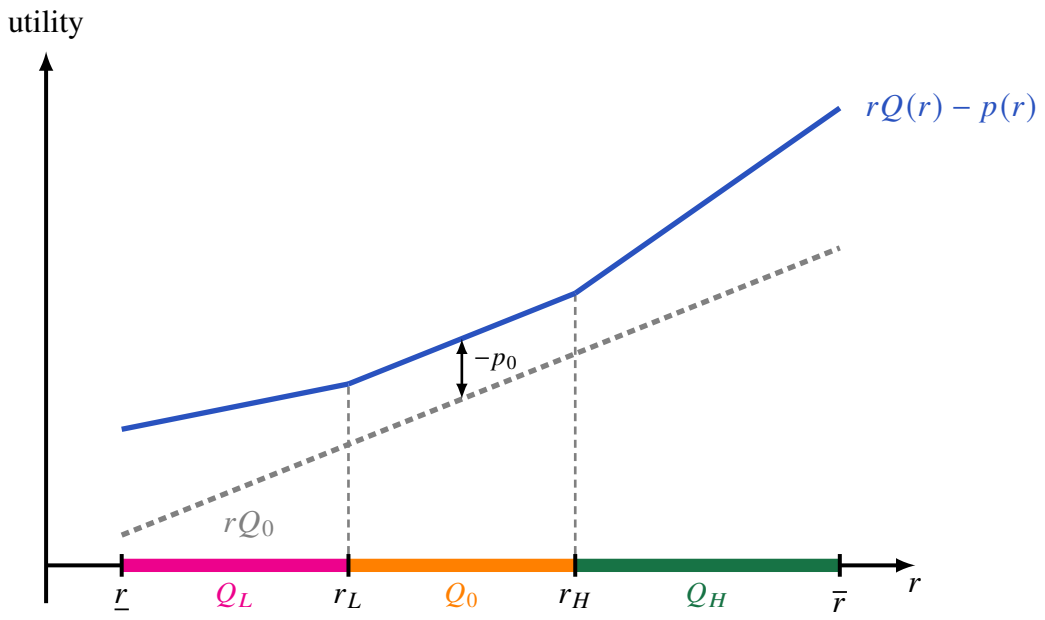

Figure 2: Agent utility in a three-tier Pareto improvement.

Define $Q_L \coloneqq \mathbf{E}_F[Q \mid Q < Q_0]$ and $Q_H \coloneqq \mathbf{E}_F[Q \mid Q > Q_0]$, and let

$$\begin{cases} \pi_H \coloneqq \dfrac{1}{2} \Pr_F[Q > Q_0], \\ \pi_L \coloneqq \Pr_F[Q < Q_0] \left[1 - \dfrac{\mathbf{E}_F[(Q - Q_0)_+]}{2\,\mathbf{E}_F[(Q_0 - Q)_+]}\right]. \end{cases}$$

Consider the distributions

$$\begin{cases} \hat{F}^{(3)} \coloneqq \Pr_F[Q < Q_0]\delta_{Q_L} + \Pr_F[Q = Q_0]\delta_{Q_0} + 2\pi_H \delta_{Q_H}, \\ F^{(3)} \coloneqq \pi_L \delta_{Q_L} + (1 - \pi_L - \pi_H)\,\delta_{Q_0} + \pi_H \delta_{Q_H}. \end{cases}$$

Since $\mathbf{E}_F[(Q_0 - Q)_+] = \mathbf{E}_F[(Q - Q_0)_+] + Q_0 - \mathbf{E}_F[Q]$ and $\mathbf{E}_F[Q] \le Q_0$,

$$0 < \frac{\mathbf{E}_F[(Q - Q_0)_+]}{2\,\mathbf{E}_F[(Q_0 - Q)_+]} = \frac{1}{2} - \frac{Q_0 - \mathbf{E}_F[Q]}{2\,\mathbf{E}_F[(Q_0 - Q)_+]} \le \frac{1}{2}.$$

Hence, $\pi_L \in (0, \Pr_F[Q < Q_0])$ and $\pi_H \in (0, \Pr_F[Q > Q_0])$, while $1 - \pi_L - \pi_H \in (\Pr_F[Q = Q_0], 1) \subset (0, 1)$. Thus, $F^{(3)}$ is well-defined.

Claim 1. *$F^{(3)}$ is a mean-preserving contraction of $\hat{F}^{(3)}$.*

Proof of Claim 1. Observe that $F^{(3)}$ and $\hat{F}^{(3)}$ have the same mean:

$$\begin{aligned}
\mathbf{E}_{\hat{F}^{(3)}}[Q] - \mathbf{E}_{F^{(3)}}[Q] &= [\Pr_F[Q < Q_0] - \pi_L]\, Q_L + [\Pr_F[Q = Q_0] - (1 - \pi_L - \pi_H)]\, Q_0 + \pi_H Q_H \\
&= [\Pr_F[Q < Q_0] - \pi_L]\, (Q_L - Q_0) + \pi_H\, (Q_H - Q_0) \\
&= \underbrace{\Pr_F[Q < Q_0] \frac{\mathbf{E}_F[(Q - Q_0)_+]}{2\,\mathbf{E}_F[(Q_0 - Q)_+]}\, (Q_L - Q_0)}_{= -\frac{1}{2}\mathbf{E}_F[(Q - Q_0)_+]} + \underbrace{\frac{1}{2}\Pr_F[Q > Q_0]\, (Q_H - Q_0)}_{= \frac{1}{2}\mathbf{E}_F[(Q - Q_0)_+]} = 0.
\end{aligned}$$

Consider pooling a mass $\Pr_F[Q < Q_0] - \pi_L$ at Q_L with a mass π_H at Q_H. The preceding calculation shows that the pooled mass has mean Q_0. Replacing it by an atom of the same mass at Q_0 is therefore a mean-preserving contraction. The remaining masses at Q_L and Q_H are π_L and π_H, respectively. Moreover, the resulting mass at Q_0 is

$$\Pr_F[Q = Q_0] + \Pr_F[Q < Q_0] - \pi_L + \pi_H = 1 - \pi_L - \pi_H > \Pr_F[Q = Q_0] \geq 0.$$

The resulting distribution is therefore $F^{(3)}$; hence, $F^{(3)} \in \mathrm{MPC}(\hat{F}^{(3)})$, as claimed. □

Note that because $\hat{F}^{(3)}$ is obtained from F by pooling all the mass below Q_0 at Q_L and all the mass above Q_0 at Q_H, $\hat{F}^{(3)}$ is a mean-preserving contraction of F. Thus, by Claim 1, $F^{(3)}$ is also a mean-preserving contraction of F. Hence, $F^{(3)}$ is feasible by Axiom 1.

We next obtain a uniform lower bound on the mass assigned to the middle tier. Using the identity $\mathbf{E}_F[(Q_0 - Q)_+] = \mathbf{E}_F[(Q - Q_0)_+] + Q_0 - \mathbf{E}_F[Q]$ again, we compute that

$$1 - \frac{\mathbf{E}_F[(Q - Q_0)_+]}{2\,\mathbf{E}_F[(Q_0 - Q)_+]} = 1 - \frac{1}{2\left[1 + \frac{Q_0 - \mathbf{E}_F[Q]}{\mathbf{E}_F[(Q - Q_0)_+]}\right]} \leq 1 - \frac{1}{2\,(1 + \varepsilon)}.$$

Since $\varepsilon \leq 1$ by construction, we obtain an upper bound of $3/4$. Consequently,

$$\pi_L + \pi_H \leq \frac{3}{4}\Pr_F[Q < Q_0] + \frac{1}{2}\Pr_F[Q > Q_0] \leq \frac{3}{4}.$$

Thus, the middle tier has mass $1 - \pi_L - \pi_H \geq 1/4$.

We now consider the mechanism that implements $F^{(3)}$. Let $r_L := G^{-1}(\pi_L)$ and $r_H := G^{-1}(1 - \pi_H)$ denote the threshold types between the low and middle tiers and between the middle and high tiers, respectively. Since $1 - \pi_H - \pi_L \geq 1/4$, it follows that

$$r_H - r_L \geq \min_{x \in [0,3/4]} \left[G^{-1}\left(x + \frac{1}{4}\right) - G^{-1}(x) \right] > 0. \tag{3}$$

The minimum is strictly positive because G^{-1} is continuous and strictly increasing. Thus, the mechanism implementing $F^{(3)}$ assigns the low tier to types $r \leq r_L$, the middle tier to types $r_L < r \leq r_H$, and the high tier to types $r > r_H$. Let p_L, p_0, and p_H denote the corresponding payments, which satisfy

$$p_0 - p_L = r_L(Q_0 - Q_L) \quad \text{and} \quad p_H - p_0 = r_H(Q_H - Q_0). \tag{4}$$

Budget balance then implies that

$$\pi_L p_L + (1 - \pi_L - \pi_H)\, p_0 + \pi_H p_H = 0 \iff p_0 = \pi_L r_L\,(Q_0 - Q_L) - \pi_H r_H\,(Q_H - Q_0)\,. \tag{5}$$

Finally, we show that the mechanism implementing $F^{(3)}$ leads to a strict Pareto improvement. By equation (4), the utility gain relative to the equity benchmark is therefore

$$\begin{cases} -p_0 + (r - r_H)(Q_H - Q_0), & \text{if } r > r_H, \\ -p_0, & \text{if } r_L < r \leq r_H, \\ -p_0 + (r_L - r)(Q_0 - Q_L), & \text{if } r \leq r_L. \end{cases}$$

Since $Q_L < Q_0 < Q_H$, agents in the middle tier receive the smallest utility gain, as shown in Figure 2.

Using the definitions of π_L and π_H, equation (5) implies

$$\begin{aligned}
-p_0 &= -r_L \cdot \left[\mathbf{E}_F[(Q_0 - Q)_+] - \frac{1}{2}\mathbf{E}_F[(Q - Q_0)_+]\right] + r_H \cdot \frac{1}{2}\mathbf{E}_F[(Q - Q_0)_+] \\
&= -r_L \cdot \left[\frac{1}{2}\mathbf{E}_F[(Q - Q_0)_+] + Q_0 - \mathbf{E}_F[Q]\right] + r_H \cdot \frac{1}{2}\mathbf{E}_F[(Q - Q_0)_+] \\
&= \frac{1}{2}\mathbf{E}_F[(Q - Q_0)_+]\,(r_H - r_L) \;-\; \underbrace{[Q_0 - \mathbf{E}_F[Q]]}_{\leq \varepsilon\,\mathbf{E}_F[(Q-Q_0)_+]}\, r_L \\
&\geq \frac{1}{2}\mathbf{E}_F[(Q - Q_0)_+]\left[\min_{x\in[0,3/4]}\left[G^{-1}\left(x + \frac{1}{4}\right) - G^{-1}(x)\right] - 2\varepsilon\bar{r}\right].
\end{aligned}$$

The inequality follows from equations (2) and (3). To guarantee that $-p_0 > 0$, we choose ε in equation (2) to satisfy

$$0 < \varepsilon < \min\left\{1, \frac{1}{2\bar{r}}\min_{x\in[0,3/4]}\left[G^{-1}\left(x + \frac{1}{4}\right) - G^{-1}(x)\right]\right\}.$$

Under this choice of ε, every agent is strictly better off under the feasible three-tier mechanism than under the equity benchmark. Thus, $F^{(3)}$ results in a strict Pareto improvement, as claimed. □

Our construction uses an intermediate tier that replicates the quality of the equity benchmark. This feature alone is not sufficient for a Pareto improvement, but the preceding argument yields a simple necessary and sufficient condition.

Corollary 2. *A three-tier distribution $F = \pi_L\delta_{Q_L} + (1 - \pi_L - \pi_H)\,\delta_{Q_0} + \pi_H\delta_{Q_H}$, where $Q_L < Q_0 < Q_H$, $\pi_L, \pi_H > 0$, and $\pi_L + \pi_H < 1$, yields a strict Pareto improvement over the equity benchmark if and only if*

$$\frac{Q_H - Q_0}{Q_0 - Q_L} > \frac{\pi_L}{\pi_H}\frac{G^{-1}(\pi_L)}{G^{-1}(1 - \pi_H)}.$$

Corollary 2 shows that merely adding an arbitrarily small intermediate tier to a two-tier system is not enough. As the mass assigned to the intermediate tier approaches zero, the two threshold types converge, and the condition in Corollary 2 requires that $\pi_H Q_H + \pi_L Q_L \geq Q_0$ in the limit, which Axiom 2 rules out. A sufficiently large mass must therefore be assigned to the intermediate tier. Conversely, as $\pi_L = \pi_H \downarrow 0$, the right-hand side of the condition converges to $\underline{r}/\bar{r}$, a simple measure of heterogeneity

in rates of substitution—and hence of gains from trade—across agents.[13] We exploit this observation in Section 4.4 to relax Axiom 3 by comparing the loss in average quality directly with these gains from trade.

4 Applications

This section applies and generalizes the main results of Section 3 to four settings: lane pricing, waiting in line, public provision of goods with heterogeneous quality, and insurance under adverse selection.

4.1 Lane Pricing

We begin with lane pricing, our leading application and primary motivation for the framework. Recall that there are $N \geq 2$ lanes; if a mass m_i of agents is assigned to lane $i \in \{1, 2, \ldots, N\}$, each agent in that lane experiences travel time $w(m_i)$. Congestion effects are captured by $w : [0, 1] \to [\underline{t}, \bar{t}]$, which is continuously differentiable, increasing, and convex, with $0 \leq \underline{t} < \bar{t} < 1$. A travel time t corresponds to effective quality $v(1 - t)$, where v is continuously differentiable, increasing, and concave.

The laissez-faire allocation in this setting provides a natural candidate for the equity benchmark. In the absence of pricing, agents distribute themselves evenly across lanes, resulting in the lane assignment

$$\mathbf{m}_0 = \left(\frac{1}{N}, \frac{1}{N}, \ldots, \frac{1}{N}\right).$$

The corresponding laissez-faire distribution assigns full probability to the quality $Q_0 := v(1 - w(1/N))$.

In this setting, the laissez-faire allocation uniquely maximizes average quality across all feasible allocations, as shown in Section 2. The convexity and strict monotonicity of the congestion technology induce a mean–dispersion tradeoff: any dispersion in qualities requires a departure from equal lane usage, which necessarily decreases average quality. At the same time, this tradeoff is locally mild. Specifically, the decrease in average quality induced by small amounts of dispersion is of higher order, as formally

[13] This observation is a point of contact with Ater et al. (2026) who emphasize that driver heterogeneity is crucial for assessing the welfare gains of congestion pricing.

demonstrated in Appendix A.1.

Given that Axioms 1 to 3 hold in this setting, our main results apply: Theorems 1 and 2 imply that there exists a strict Pareto improvement with three tiers, but no Pareto improvement with two tiers. As shown in Appendix A.1, our main results extend to deterministic lane assignments:

Proposition 1. *When $N \geq 3$, there exists a mechanism using only deterministic lane assignments that strictly Pareto-improves on the laissez-faire outcome. When $N = 2$, no mechanism using only deterministic lane assignments can be a Pareto improvement.*

Proposition 1 has a simple practical interpretation. With at least three physical lanes, the Pareto improvement can be implemented deterministically: commuters can be assigned to fast, regular, and slow lanes, with no need to randomize their access. Prices can then be calibrated so that the regular lane has the same level of congestion as in the laissez-faire allocation.

With only two physical lanes, some randomization is necessary to generate a Pareto improvement, but this need not require literal lotteries. It could instead be implemented through restrictions on access over time—for example, by allowing commuters choosing the intermediate tier to use the fast lane only on certain days or for a limited amount of time.

Our positive results rely on the ability to redistribute toll revenues to commuters. As emphasized in the congestion-pricing literature (e.g., Hall, 2018), such redistribution may be difficult in practice. Importantly, however, our proof of Theorem 2 implies that redistribution need not be perfect. Fixing the distribution G of commuters' values, there exists a cutoff $\overline{\alpha} < 1$ such that a Pareto improvement remains possible whenever a fraction $\alpha > \overline{\alpha}$ of toll revenues can be returned to commuters. Such redistribution could be implemented indirectly, for example through reductions in other taxes or fees paid by drivers, or through credits deposited into electronic toll accounts that can ultimately be redeemed for cash.[14]

[14] What our framework abstracts from is the possible extensive-margin responses to the introduction of priority pricing. By definition, a strict Pareto improvement among existing commuters increases the attractiveness of driving and may therefore induce additional traffic.

4.2 Waiting in Line

We next consider a designer who can *endogenously create lines* by partitioning agents and dedicating processing capacity to each line, as in airport security screening or access to public services. Unlike our lane-pricing application, where congestion depends only on the mass assigned to each lane, here congestion depends on demand relative to capacity in each line.

In this setting, agents arrive at a constant rate $\lambda > 0$, and the designer has total processing capacity $\mu > 0$. Denote the demand–capacity ratio by $\rho := \lambda/\mu \in (0, 1)$, which we refer to as the *load* of the entire system. A deterministic line design consists of a finite number of lines $K \in \{1, 2, \ldots\}$ together with routing shares $\mathbf{m} = (m_1, \ldots, m_K) \in \Delta^{K-1}$ and capacity shares $\mathbf{s} = (s_1, \ldots, s_K) \in \Delta^{K-1}$. Given such a design $(K, \mathbf{m}, \mathbf{s})$, line $i \in \{1, 2, \ldots, K\}$ is assigned arrival rate λm_i and capacity μs_i, so its load is

$$\rho_i := \frac{\lambda m_i}{\mu s_i} = \rho \frac{m_i}{s_i}.$$

We restrict attention to designs in which $m_i, s_i > 0$ and $\rho_i < 1$ for every i, so expected delays are finite.

We model expected delay using a reduced-form congestion technology. Each agent assigned to line i spends time $w(\rho_i)$ waiting, where $w : (0, 1) \to [\underline{t}, \bar{t}]$ is continuously differentiable, increasing, and convex, with $0 \leq \underline{t} < \bar{t} < 1$. The key property of this technology is *scale-freeness*: if both demand and capacity in a line are scaled by the same factor, the load ρ_i is unchanged and so is the expected wait time. As in the lane-pricing application, the quality associated with wait time $w(\rho_i)$ is $v(1 - w(\rho_i))$.

Next, we specify the feasible set of expected quality distributions. We identify each deterministic line design $(K, \mathbf{m}, \mathbf{s})$ with the quality distribution that it induces,

$$F_{(K,\mathbf{m},\mathbf{s})} := \sum_{i=1}^{K} m_i \delta_{v(1-w(\rho_i))}.$$

The designer can randomize agents' assignments within a deterministic line design, so the feasible set is

$$\mathcal{F} = \left\{ F \in \Delta([0, 1]) : F \in \mathrm{MPC}(F_{(K,\mathbf{m},\mathbf{s})}) \text{ for some deterministic line design } (K, \mathbf{m}, \mathbf{s}) \right\}.$$

Under the laissez-faire outcome, all agents use a single line. This corresponds to a design with $K = 1$ and $m_1 = s_1 = 1$. Each agent receives quality $Q_0 = v(1 - w(\rho))$, so the induced distribution is δ_{Q_0}.

Axiom 1 holds by construction. We show in Appendix A.2 that Axioms 2 and 3 also hold in this setting, with the laissez-faire expected quality distribution as the equity benchmark, by an argument similar to the lane-pricing application. Intuitively, making one line faster requires allocating more capacity to it relative to demand, leaving another line with less capacity. Because waiting time is convex in line load, this dispersion increases average wait time. Moreover, small departures from proportional capacity allocation create first-order dispersion at a higher-order cost to the average.

Since Axioms 1 to 3 hold in this setting, Theorems 1 and 2 apply. As in the lane-pricing application, the results extend to deterministic line designs:

Proposition 2. *There exists a mechanism based on a deterministic three-line design that strictly Pareto-improves on the laissez-faire outcome. No mechanism based on a deterministic line design with $K \leq 2$ can be a Pareto improvement.*

The implementation suggested by Proposition 2 is again simple. The designer can partition existing capacity across three lines, assigning relatively more capacity per user to the fast line, relatively less to the slow line, and preserving the laissez-faire demand–capacity ratio in the middle line. In applications, this can be implemented through separate queues or service windows with different capacity allocations.

4.3 Public Provision With Heterogeneous Quality

Our third application considers environments in which the designer allocates goods of heterogeneous physical quality. It illustrates how the mean–dispersion tradeoff underlying our results can arise from economic frictions unrelated to congestion.

We consider a unit mass of agents who demand a good of heterogeneous physical quality, $q \in [0, 1]$. For example, q might represent the size of a public housing unit. Each agent privately observes his type $r \in [\underline{r}, \bar{r}] \subset \mathbf{R}_+$. An agent of type r who receives physical quality q and pays price p obtains utility $rv(q) - p$, where $v : [0, 1] \to [0, 1]$ is continuously differentiable, increasing, and strictly concave.

A capacity-constrained designer has $q_0 \in (0,1)$ units of physical quality to allocate per capita. For instance, q_0 might represent the total amount of space that a public housing authority can subdivide into individual units. Given an allocation function $q : [\underline{r}, \bar{r}] \to [0,1]$, feasibility requires that the total physical quality allocated not exceed available capacity:

$$\int_{\underline{r}}^{\bar{r}} q(r)\, \mathrm{d}G(r) \leq q_0.$$

To apply the framework of Section 2, we express this constraint in terms of effective quality. Because $Q = v(q)$ and v is increasing, providing effective quality Q requires physical quality $v^{-1}(Q)$. Accordingly, the feasible set of expected effective quality distributions is

$$\mathcal{F} = \left\{ F \in \Delta([0, v(1)]) : \mathbf{E}_F\left[v^{-1}(Q)\right] \leq q_0 \right\}.$$

Because v is concave, any lottery delivering expected effective quality Q uses at least $v^{-1}(Q)$ units of expected physical quality; assigning physical quality $v^{-1}(Q)$ with certainty attains this bound.

In the absence of pricing, the designer cannot use quality to screen agents; a natural candidate for the equity benchmark is therefore uniform provision at physical quality q_0. This allocation provides every agent with effective quality $Q_0 \coloneqq v(q_0)$.

Because the capacity constraint is preserved under mean-preserving contractions (since v^{-1} is convex), the feasible set satisfies Axiom 1. We show that Axiom 2 holds as well, leaving the verification of Axiom 3 to Appendix A.3. For any feasible distribution F, Jensen's inequality gives

$$\mathbf{E}_F[Q] = \mathbf{E}_F\left[v\left(v^{-1}(Q)\right)\right] \leq v\left(\mathbf{E}_F\left[v^{-1}(Q)\right]\right) \leq v(q_0) = Q_0.$$

Because v is strictly concave, equality holds if and only if physical quality is almost surely constant at q_0. Thus, the unique maximizing feasible distribution is δ_{Q_0}. Any dispersion in physical quality around the equity benchmark necessarily decreases average *effective* quality due to decreasing marginal utility from physical quality, yielding the key mean–dispersion tradeoff formalized by Axiom 2.

Since Axioms 1 to 3 hold in this public-provision environment, our main results apply:

Proposition 3. *Any Pareto improvement over the equity benchmark must offer at least three levels of expected quality. Moreover, there exists a strict Pareto improvement that offers three levels of expected quality.*

Public housing provides a natural interpretation of this result. Public housing programs often restrict households to a relatively limited set of options, making uniform provision a natural benchmark (e.g., Olsen, 2003; Sitaraman and Alstott, 2019). Singapore's public housing system provides a useful contrast: the public housing authority chooses among different types of apartments to build and allocate, allowing households to express preferences over these alternatives (e.g., Ferdowsian, Lee and Yap, 2026). Our results suggest that differentiation—in particular, in size—need not come at the expense of equity. If prices are appropriately adjusted, offering at least three quality levels can generate a Pareto improvement over uniform provision, even if dispersion in physical quality lowers average effective quality.

4.4 Insurance Under Adverse Selection

Our final application shows how our main results generalize beyond the axioms introduced in Section 2. We consider an insurance program with a fixed budget, where the main friction generating the mean–dispersion tradeoff arises from adverse selection. Employer-provided health insurance (e.g., Einav, Finkelstein and Cullen, 2010) provides a natural example: a self-insured employer with a fixed benefits budget may offer plans of different generosity to employees whose willingness to pay for coverage is positively correlated with their expected medical costs.

Unlike the preceding applications, the resource cost of coverage depends on the type of the agent who receives it due to adverse selection. We first argue that the feasible set satisfies Axioms 1 and 2. We then show that Axiom 3 does not hold. Finally, we provide a more general condition under which a three-tier Pareto improvement nevertheless exists.

Let $Q \in [0, 1]$ denote the generosity of insurance coverage. A type-r agent who receives coverage Q and makes a premium adjustment p obtains utility $rQ - p$. Providing one unit of coverage to this agent generates expected cost $c(r)$ to the designer, where $c : [\underline{r}, \overline{r}] \to \mathbf{R}_{++}$ is continuous and increasing. Thus, willingness to pay and expected claims are positively associated, as in a standard adverse-selection

environment. A simple example is when r reflects the probability of using the insurance, while $c(r)$ also incorporates the expected expenditure conditional on use.

The equity benchmark provides uniform coverage $Q_0 \in (0,1)$ with no premium adjustments. An allocation $Q : [\underline{r}, \overline{r}] \to [0,1]$ is feasible if the aggregate expected cost does not exceed that under the equity benchmark:

$$\int_{\underline{r}}^{\overline{r}} c(r)Q(r)\,\mathrm{d}G(r) \le Q_0 \int_{\underline{r}}^{\overline{r}} c(r)\,\mathrm{d}G(r).$$

Premium adjustments must satisfy budget balance. To express the feasibility constraint in terms of distributions of expected quality, define $C(u) := c(G^{-1}(u))$ for every $u \in [0,1]$. Because incentive compatibility assigns higher coverage to higher types, an expected quality distribution F results in

$$\int_{\underline{r}}^{\overline{r}} c(r)Q(r)\,\mathrm{d}G(r) = \int_0^1 C(u)F^{-1}(u)\,\mathrm{d}u$$

units of expected cost. Accordingly, the feasible set is

$$\mathcal{F} = \left\{ F \in \Delta([0,1]) : \int_0^1 C(u)F^{-1}(u)\,\mathrm{d}u \le Q_0 \int_0^1 C(u)\,\mathrm{d}u \right\}.$$

We first verify Axiom 1. For any $\bar{F} \in \mathrm{MPC}(F)$, since $\mathrm{d}C$ is a nonnegative measure under our maintained assumptions, integration by parts yields

$$\begin{aligned}\int_0^1 C(u)\bar{F}^{-1}(u)\,\mathrm{d}u &= C(1)\int_0^1 \bar{F}^{-1}(z)\,\mathrm{d}z - \int_0^1 \int_0^u \bar{F}^{-1}(z)\,\mathrm{d}z\,\mathrm{d}C(u) \\ &\le C(1)\int_0^1 F^{-1}(z)\,\mathrm{d}z - \int_0^1 \int_0^u F^{-1}(z)\,\mathrm{d}z\,\mathrm{d}C(u) = \int_0^1 C(u)F^{-1}(u)\,\mathrm{d}u.\end{aligned}$$

Thus, every mean-preserving contraction of a feasible distribution remains feasible, and Axiom 1 holds.

Next, we verify Axiom 2. Since both C and F^{-1} are nondecreasing, Chebyshev's integral inequality gives

$$\int_0^1 C(u)F^{-1}(u)\,\mathrm{d}u \ge \left[\int_0^1 C(u)\,\mathrm{d}u\right]\left[\int_0^1 F^{-1}(u)\,\mathrm{d}u\right] = \mathbf{E}_F[Q]\int_0^1 C(u)\,\mathrm{d}u.$$

Feasibility therefore implies that $\mathbf{E}_F[Q] \le Q_0$. Because C is increasing, this inequality is strict whenever

F is nondegenerate. Hence, the distribution δ_{Q_0} uniquely maximizes average quality, and Axiom 2 holds.

Unlike in the preceding applications, however, the mean–dispersion tradeoff is not locally mild, due to adverse selection. Increasing coverage for high types is more expensive than reducing the same amount of coverage for low types saves, so feasibility requires a non-negligible additional reduction in coverage at the bottom. As a result, the loss in average quality remains *proportional* to the quality gains at the top, rather than being of lower order. In Appendix A.4, we show that Axiom 3 fails.

Even without Axiom 3, however, a three-tier strict Pareto improvement remains possible if high types value coverage sufficiently highly relative to their expected cost. We consider the condition

$$\frac{\overline{r}}{c(\overline{r})} > \frac{\underline{r}}{c(\underline{r})}. \tag{B}$$

The "bang-for-the-buck" ratio $r/c(r)$ measures a type's willingness to pay for coverage per unit of expected cost. While this ratio may vary nonmonotonically with type, Condition (B) requires this ratio to be higher at the upper endpoint than at the lower endpoint.

Proposition 4. *There does not exist a two-tier Pareto improvement over uniform coverage. However, under condition* (B)*, there exists a three-tier strict Pareto improvement.*

In the context of employer-provided health insurance, Proposition 4 implies that—conditional on the fixed benefits budget—offering a choice between a low-coverage and a high-coverage plan cannot make all employees better off compared to a simple uniform-coverage scheme. A third, intermediate-coverage plan can achieve a Pareto improvement when employees with the highest willingness to pay for coverage have sufficiently high willingness to pay relative to their expected claims.

Without some restriction on the relationship between willingness to pay and expected cost, a Pareto improvement need not exist. Suppose, for example, that $\underline{r} > 0$ and $c(r) = r$. For any feasible, budget-balanced mechanism,

$$\int_{\underline{r}}^{\overline{r}} \left[rQ(r) - p(r) - rQ_0\right] \, \mathrm{d}G(r) = \int_{\underline{r}}^{\overline{r}} r \left[Q(r) - Q_0\right] \, \mathrm{d}G(r) = \int_{\underline{r}}^{\overline{r}} c(r) \left[Q(r) - Q_0\right] \, \mathrm{d}G(r) \leq 0.$$

A Pareto improvement would make the first integrand nonnegative for every type and strictly positive for

a positive mass of types, contradicting this inequality. Thus, no Pareto improvement exists in this case, regardless of the number of tiers.

Condition (B) motivates a generalization of Axiom 3. While our framework in Section 2 deliberately abstracted from variable production costs for simplicity, such costs can be accommodated:

Axiom 3* (Generalized Smoothness). *There exists a continuous function* $\kappa : [0,1] \to \mathbf{R}_{++}$ *such that:*

(i) *For every bounded nondecreasing function* $h : [0,1] \to \mathbf{R}$ *satisfying* $\int_0^1 \kappa(u)h(u)\,\mathrm{d}u < 0$*, there exists a sequence* $\{(\varepsilon_n, F_n)\}_{n=1}^{\infty}$ *such that* $\varepsilon_n \downarrow 0$*,* $F_n \in \mathcal{F}$*, and*

$$\int_0^1 \left| \frac{F_n^{-1}(u) - Q_0}{\varepsilon_n} - h(u) \right| \mathrm{d}u \to 0.$$

(ii) *The extreme types satisfy*

$$\frac{G^{-1}(1)}{\kappa(1)} > \frac{G^{-1}(0)}{\kappa(0)}.$$

Theorem 2*. *Under Axioms 1 and 3*, there exists a strict Pareto improvement with three tiers.*

Proof. See Appendix B. □

Intuitively, Axiom 3* replaces the requirement that the loss in average quality be of higher order with a condition on the local resource tradeoff. The function $\kappa(u)$ can be interpreted as the marginal resource cost of increasing expected quality at quantile u. Condition (i) requires every bounded, nondecreasing perturbation that strictly reduces resource use to be approximated by arbitrarily small feasible departures from the equity benchmark. Condition (ii) requires the highest type to have a higher willingness to pay per unit of marginal resource cost than the lowest type. Together, these conditions make it possible to reduce quality for a small mass of low types and increase quality for a small mass of high types, while retaining Q_0 as the middle tier. The resulting gains can then be redistributed through payments so that every type is strictly better off.

Axiom 3* is satisfied in our insurance application by choosing $\kappa = C$. With this choice, condition (ii) is equivalent to condition (B). To see that condition (i) is also satisfied, fix any bounded nondecreasing

function $h : [0,1] \to \mathbf{R}$ such that

$$\int_0^1 C(u)h(u)\,\mathrm{d}u < 0.$$

Since $Q_0 \in (0,1)$ and h is bounded, choose any sequence $\varepsilon_n \downarrow 0$ such that $Q_0 + \varepsilon_n h(u) \in [0,1]$ for every $u \in [0,1]$ and every n. Let F_n be a distribution whose quantile function satisfies $F_n^{-1}(u) = Q_0 + \varepsilon_n h(u)$ almost everywhere; such a distribution exists because h is nondecreasing. Moreover,

$$\int_0^1 C(u)F_n^{-1}(u)\,\mathrm{d}u = Q_0 \int_0^1 C(u)\,\mathrm{d}u + \varepsilon_n \int_0^1 C(u)h(u)\,\mathrm{d}u < Q_0 \int_0^1 C(u)\,\mathrm{d}u.$$

Hence, $F_n \in \mathcal{F}$ for every n. Finally, condition (i) holds, since for every n,

$$\int_0^1 \left| \frac{F_n^{-1}(u) - Q_0}{\varepsilon_n} - h(u) \right| \mathrm{d}u = 0.$$

Our insurance application shows that Axiom 3* can hold even when Axiom 3 fails. In Appendix B, we show that under Axioms 1 and 2, Axiom 3 implies Axiom 3*.

5 Concluding Remarks

This paper studies when prices and priority can be introduced without making anyone worse off than under the equity benchmark. In environments characterized by a mean–dispersion tradeoff, two tiers are not enough. With only two tiers, the same threshold type governs both the quality difference and the associated payment difference, and some agents must be worse off. A third tier separates the threshold between the low and middle tiers from the threshold between the middle and high tiers. This separation allows the gains from sorting to be redistributed so that every agent is strictly better off.

Our framework isolates this logic without specifying interpersonal welfare comparisons or committing to a particular allocation technology. It applies to lane pricing, waiting-line design, and public provision, and extends to insurance under adverse selection, where the resource cost of quality depends on the recipient. More broadly, our results show that the distributional consequences of priority pricing depend not only on whether prices are introduced, but also on how the priority tiers are designed. Policy debates

framed as a choice between uniform access and a two-tier priority system therefore overlook a simple alternative: a third tier can allow the gains from priority pricing to be shared by all agents.

References

Akbarpour, M. ® P. Dworczak ® S. D. Kominers (2024): "Redistributive Allocation Mechanisms," *Journal of Political Economy*, 132, 1831–1875.

Anderson, M. G. (2005): "The Bias Built Into 'Lexus Lanes'," *The Washington Post*, https://www.washingtonpost.com/archive/opinions/2005/07/01/the-bias-built-into-lexus-lanes/8fa5656e-83b4-47c8-af11-f6d3fb08c661/, Letter to the Editor.

Arnott, R., A. De Palma, and R. Lindsey (1994): "The Welfare Effects of Congestion Tolls With Heterogeneous Commuters," *Journal of Transport Economics and Policy*, 28, 139–161.

Ater, I., B. Ross, A. Shany, and S. Vasserman (2026): "Road Pricing Under Heterogeneity," *Working paper*.

Beheshtian, A., R. R. Geddes, O. M. Rouhani, K. M. Kockelman, A. Ockenfels, P. Cramton, and W. Do (2020): "Bringing the Efficiency of Electricity Market Mechanisms to Multimodal Mobility Across Congested Transportation Systems," *Transportation Research Part A: Policy and Practice*, 131, 58–69.

van den Berg, V., and E. T. Verhoef (2011): "Congestion Tolling in the Bottleneck Model With Heterogeneous Values of Time," *Transportation Research Part B: Methodological*, 45, 60–78.

Bobbio, E., P. Cramton, and A. Ockenfels (2021): "Welfare Effects of Efficient Road Pricing in a Simple Model of a City With Heterogeneous Agents," *Working paper*.

Chao, H.-P., and R. B. Wilson (1987): "Priority Service: Pricing, Investment, and Market Organization," *American Economic Review*, 77, 899–916.

Condorelli, D. (2013): "Market and Non-Market Mechanisms for the Optimal Allocation of Scarce Resources," *Games and Economic Behavior*, 82, 582–591.

Cook, C., A. Kreidieh, S. Vasserman, H. Allcott, N. Arora, A. Tomkins, E. Turkel, and F. van Sambeek (2026): "The Network-Wide Effects of Congestion Pricing: Evidence From New York City," *Working paper*.

Cook, C., and P. Z. Li (2025): "Value Pricing or Lexus Lanes? The Distributional Effects of Dynamic Tolling," *Working paper*.

Cramton, P., R. R. Geddes, and A. Ockenfels (2019): "Markets for Road Use: Eliminating Congestion Through Scheduling, Routing, and Real-Time Road Pricing," *Working paper*.

Cramton, P., R. Gibbons, and P. Klemperer (1987): "Dissolving a Partnership Efficiently," *Econometrica*, 55, 615–632.

Dworczak, P. ® S. D. Kominers ® M. Akbarpour (2021): "Redistribution Through Markets," *Econometrica*, 89, 1665–1698.

Einav, L., A. Finkelstein, and M. R. Cullen (2010): "Estimating Welfare in Insurance Markets Using Variation in Prices," *Quarterly Journal of Economics*, 125, 877–921.

Ferdowsian, A., K. H. Lee, and L. Yap (2026): "Build-To-Order: Endogenous Supply in Centralized Mechanisms," *Working paper*.

Gershkov, A., and P. Schweinzer (2010): "When Queueing Is Better Than Push and Shove," *International Journal of Game Theory*, 39, 409–430.

Gershkov, A., and E. Winter (2023): "Gainers and Losers in Priority Services," *Journal of Political Economy*, 131, 3103–3155.

Hall, J. D. (2018): "Pareto Improvements From Lexus Lanes: The Effects of Pricing a Portion of the Lanes on Congested Highways," *Journal of Public Economics*, 158, 113–125.

——— (2021): "Can Tolling Help Everyone? Estimating the Aggregate and Distributional Consequences of Congestion Pricing," *Journal of the European Economic Association*, 19, 441–474.

Kang, Z. Y. (2023): "The Public Option and Optimal Redistribution," *Working paper*.

Kang, Z. Y., and M. Watt (2026): "Topping Up and Optimal Redistribution," *Working paper*.

Kreindler, G. (2024): "Peak-Hour Road Congestion Pricing: Experimental Evidence and Equilibrium Implications," *Econometrica*, 92, 1233–1268.

Milgrom, P., and I. Segal (2002): "Envelope Theorems for Arbitrary Choice Sets," *Econometrica*, 70, 583–601.

Mull, A. (2023): "The Perfect Service to Make Everyone at the Airport Hate You," *The Atlantic*, https://www.theatlantic.com/technology/archive/2023/07/clear-airport-security-lines-tsa-infrastructure/674809/.

Müller, A., and D. Stoyan (2002): *Comparison Methods for Stochastic Models and Risks*, Volume 389, West Sussex, UK: John Wiley & Sons.

Myerson, R. B. (1981): "Optimal Auction Design," *Mathematics of Operations Research*, 6, 58–73.

Nichols, A. L., and R. J. Zeckhauser (1982): "Targeting Transfers through Restrictions on Recipients," *American Economic Review*, 72, 372–377.

Olsen, E. O. (2003): "Housing Programs for Low-Income Households," in *Means-Tested Transfer Programs in the United States* ed. by Moffitt, R. A.: University of Chicago Press.

Ostrovsky, M., and M. Schwarz (2018): "Carpooling and the Economics of Self-Driving Cars," *Working paper*.

Ostrovsky, M., and F. Yang (2024): "Effective and Equitable Congestion Pricing: New York City and Beyond," *Working paper*.

Pigou, A. C. (1920): *The Economics of Welfare*, London, UK: Macmillan.

Shachar, A. (2018): "Dangerous Liaisons: Money and Citizenship," in *Debating Transformations of National Citizenship* ed. by Bauböck, R. Cham: Springer, 7–15, 10.1007/978-3-319-92719-0_2.

Sitaraman, G., and A. L. Alstott (2019): *The Public Option: How to Expand Freedom, Increase Opportunity, and Promote Equality*, Cambridge, MA: Harvard University Press.

Spence, M. (1977): "Nonlinear Prices and Welfare," *Journal of Public Economics*, 8, 1–18.

Stewart, E. (2022): "Why Must We Pay to Have a Slightly Less Miserable Time at the Airport?" *Vox*, https://www.vox.com/the-goods/23101906/tsa-precheck-clear-cost-airport-security-travel.

Vickrey, W. S. (1969): "Congestion Theory and Transport Investment," *American Economic Review (Papers and Proceedings)*, 59, 251–260.

Weitzman, M. L. (1977): "Is the Price System or Rationing More Effective in Getting a Commodity to Those Who Need It Most?" *Bell Journal of Economics*, 8, 517–524.

A Omitted Proofs

A.1 Proofs for Section 4.1

We first argue that Axiom 3 holds in this application. Consider a perturbation of the laissez-faire allocation. For $\varepsilon \in (0, 1/N)$, define

$$\mathbf{m}_\varepsilon := \left(\frac{1}{N} - \varepsilon, \frac{1}{N}, \ldots, \frac{1}{N}, \frac{1}{N} + \varepsilon\right),$$

so that a mass ε of agents is shifted from lane 1 to lane N, while all other lanes retain their laissez-faire masses. Writing F^ε for the induced distribution of qualities, the resulting loss in average quality is

$$Q_0 - \mathbf{E}_{F^\varepsilon}[Q] = 2\phi\left(\frac{1}{N}\right) - \phi\left(\frac{1}{N} - \varepsilon\right) - \phi\left(\frac{1}{N} + \varepsilon\right) = o(\varepsilon).$$

The final equality follows from the differentiability of ϕ: the two first-order effects cancel. By contrast, the increase in dispersion—measured by the cumulative improvement above Q_0 received by agents in lane 1—is

$$\mathbf{E}_{F^\varepsilon}[(Q - Q_0)_+] = \left(\frac{1}{N} - \varepsilon\right)\left[v\left(1 - w\left(\frac{1}{N} - \varepsilon\right)\right) - v\left(1 - w\left(\frac{1}{N}\right)\right)\right] = \frac{\varepsilon}{N} v'\left(1 - w\left(\frac{1}{N}\right)\right) w'\left(\frac{1}{N}\right) + o(\varepsilon).$$

Because $v', w' > 0$, this expression is positive for all sufficiently small ε and is of order ε. Hence, Axiom 3 holds:

$$\lim_{\varepsilon \downarrow 0} \frac{Q_0 - \mathbf{E}_{F^\varepsilon}[Q]}{\mathbf{E}_{F^\varepsilon}[(Q - Q_0)_+]} = 0.$$

We now prove Proposition 1. Suppose first that $N = 2$. Any deterministic lane assignment induces a quality distribution with at most two tiers. A one-tier distribution cannot be a Pareto improvement: incentive compatibility requires all payments to be equal, budget balance requires them to be zero, and feasibility implies that the common quality is no greater than Q_0. A two-tier distribution cannot be a Pareto improvement by Theorem 1, since Axiom 2 holds in this setting. Hence, no mechanism using only deterministic lane assignments can be a Pareto improvement when $N = 2$.

Now suppose that $N \geq 3$. For $\varepsilon \in (0, 1/N)$, consider the deterministic lane assignment

$$\mathbf{m}_\varepsilon \coloneqq \left(\frac{1}{N} - \varepsilon, \frac{1}{N}, \ldots, \frac{1}{N}, \frac{1}{N} + \varepsilon\right).$$

The induced distribution F^ε has three tiers. Writing

$$Q_L \coloneqq v\left(1 - w\left(\frac{1}{N} + \varepsilon\right)\right) \quad \text{and} \quad Q_H \coloneqq v\left(1 - w\left(\frac{1}{N} - \varepsilon\right)\right),$$

these tiers are $Q_L < Q_0 < Q_H$, with respective masses $1/N + \varepsilon$, $(N-2)/N$, and $1/N - \varepsilon$.

Consider the mechanism implementing F^ε. The threshold types between the low and middle tiers and between the middle and high tiers are

$$r_L = G^{-1}\left(\frac{1}{N} + \varepsilon\right) \quad \text{and} \quad r_H = G^{-1}\left(1 - \frac{1}{N} + \varepsilon\right).$$

Let p_L, p_0, and p_H denote the corresponding payments. Incentive compatibility requires

$$p_0 - p_L = r_L\,(Q_0 - Q_L) \quad \text{and} \quad p_H - p_0 = r_H\,(Q_H - Q_0).$$

Combining these equalities with budget balance gives

$$-p_0 = \left(\frac{1}{N} - \varepsilon\right) r_H\,(Q_H - Q_0) - \left(\frac{1}{N} + \varepsilon\right) r_L\,(Q_0 - Q_L).$$

As $\varepsilon \downarrow 0$,

$$\begin{cases} Q_H - Q_0 &= v'\left(1 - w\left(\frac{1}{N}\right)\right) w'\left(\frac{1}{N}\right)\varepsilon + o(\varepsilon), \\ Q_0 - Q_L &= v'\left(1 - w\left(\frac{1}{N}\right)\right) w'\left(\frac{1}{N}\right)\varepsilon + o(\varepsilon). \end{cases}$$

Therefore,

$$-p_0 = \frac{\varepsilon}{N} v'\left(1 - w\left(\frac{1}{N}\right)\right) w'\left(\frac{1}{N}\right)\left[G^{-1}\left(1 - \frac{1}{N}\right) - G^{-1}\left(\frac{1}{N}\right)\right] + o(\varepsilon).$$

Because $N \geq 3$, we have $1 - 1/N > 1/N$. Since G^{-1} is strictly increasing and $v', w' > 0$, it follows that

$-p_0 > 0$ for all sufficiently small $\varepsilon > 0$.

Finally, the utility gain relative to the laissez-faire outcome is

$$\begin{cases} -p_0 + (r - r_H)(Q_H - Q_0), & \text{if } r > r_H, \\ -p_0, & \text{if } r_L < r \le r_H, \\ -p_0 + (r_L - r)(Q_0 - Q_L), & \text{if } r \le r_L. \end{cases}$$

Every term is strictly positive because $-p_0 > 0$. Hence, for all sufficiently small $\varepsilon > 0$, the mechanism based on the deterministic lane assignment $\mathbf{m}_\varepsilon$ strictly Pareto-improves on the laissez-faire outcome.

A.2 Proofs for Section 4.2

We first prove that Axioms 2 and 3 hold. As in our lane-pricing application, since randomization does not affect average effective quality, it suffices to show that every deterministic line design has average effective quality at most Q_0 and that every design attaining Q_0 induces the distribution δ_{Q_0}. To this end, define $\phi(y) \coloneqq yv(1-w(\rho y))$; since w is increasing and convex on $(0,1)$, it follows that $\phi : (0, 1/\rho) \to \mathbf{R}$ is strictly concave. For any deterministic line design $(K, \mathbf{m}, \mathbf{s})$, average effective quality can be written as

$$\sum_{i=1}^{K} s_i \phi\left(\frac{m_i}{s_i}\right) \le \phi\left(\sum_{i=1}^{K} s_i \frac{m_i}{s_i}\right) = \phi(1) = v(1 - w(\rho)) = Q_0.$$

Since ϕ is strictly concave, equality holds if and only if m_i/s_i is constant across lines. Because $\sum_{i=1}^{K} m_i = \sum_{i=1}^{K} s_i = 1$, this constant must be 1. Hence, equality holds if and only if $m_i = s_i$ for every i, in which case every line has load ρ and the induced distribution is δ_{Q_0}. Thus, Axiom 2 holds.

Axiom 3 is also satisfied in this setting. To see this, fix $\eta \in (0, 1/2)$. For any $\varepsilon > 0$ sufficiently small that $\rho(1+\varepsilon) < 1$, consider a deterministic three-line design with capacity and routing shares

$$(s^F, s^0, s^S) = (\eta, 1 - 2\eta, \eta) \quad \text{and} \quad (m^F, m^0, m^S) = (\eta(1-\varepsilon), 1 - 2\eta, \eta(1+\varepsilon)).$$

The corresponding loads are $\rho(1-\varepsilon)$, ρ, and $\rho(1+\varepsilon)$, so the induced distribution F_ε has three support

points

$$v(1 - w(\rho\,(1-\varepsilon))) > Q_0 = v(1 - w(\rho)) > v(1 - w(\rho\,(1+\varepsilon)))\,.$$

These have corresponding masses $\eta\,(1-\varepsilon)$, $1-2\eta$, and $\eta\,(1+\varepsilon)$. By construction, the loss in average effective quality relative to the equity benchmark is

$$Q_0 - \mathbf{E}_{F_\varepsilon}[Q] = \eta\,[2\phi(1) - \phi(1-\varepsilon) - \phi(1+\varepsilon)] = o(\varepsilon).$$

By contrast, the cumulative improvement above Q_0 received by agents in the fast line is

$$\mathbf{E}_{F_\varepsilon}[(Q - Q_0)_+] = \eta\,(1-\varepsilon)\,[v(1 - w(\rho\,(1-\varepsilon))) - v(1 - w(\rho))] = \eta\rho v'(1 - w(\rho))w'(\rho)\varepsilon + o(\varepsilon).$$

Because $v', w' > 0$, this expression is positive for all sufficiently small ε and is of order ε. Consequently, Axiom 3 holds:

$$\lim_{\varepsilon \downarrow 0} \frac{Q_0 - \mathbf{E}_{F_\varepsilon}[Q]}{\mathbf{E}_{F_\varepsilon}[(Q - Q_0)_+]} = 0.$$

We now sketch the proof of Proposition 2, which closely parallels the proof of Proposition 1.

Suppose first that $K \leq 2$. Any deterministic line design with K lines induces a distribution with at most two support points. If this distribution is degenerate, it cannot be a Pareto improvement, by the argument given in Section 3 for one-tier distributions. If it has exactly two tiers, it cannot be a Pareto improvement by Theorem 1, since Axiom 2 holds in this setting.

Now suppose $K = 3$. Fix $\eta \in (0, 1/2)$ and, for $\varepsilon > 0$ sufficiently small that $\rho(1+\varepsilon) < 1$, consider the deterministic three-line design

$$(s^F, s^0, s^S) = (\eta, 1-2\eta, \eta) \quad \text{and} \quad (m^F, m^0, m^S) = (\eta(1-\varepsilon), 1-2\eta, \eta(1+\varepsilon))\,,$$

already used above to verify Axiom 3. As shown there, the induced distribution F_ε assigns mass $\pi_H := \eta(1-\varepsilon)$ to $Q_H := v(1 - w(\rho(1-\varepsilon)))$, mass $1-2\eta$ to Q_0, and mass $\pi_L := \eta(1+\varepsilon)$ to $Q_L := v(1 - w(\rho(1+\varepsilon)))$, with $Q_L < Q_0 < Q_H$.

Consider the mechanism implementing F_ε. The threshold types between the slow and middle lines

and between the middle and fast lines are

$$r_L \coloneqq G^{-1}(\eta(1+\varepsilon)) \quad \text{and} \quad r_H \coloneqq G^{-1}(1-\eta(1-\varepsilon)) .$$

Let p_L, p_0, and p_H denote the corresponding payments. Incentive compatibility requires

$$p_0 - p_L = r_L (Q_0 - Q_L) \quad \text{and} \quad p_H - p_0 = r_H (Q_H - Q_0) .$$

Combining these equalities with budget balance gives

$$-p_0 = \pi_H r_H (Q_H - Q_0) - \pi_L r_L (Q_0 - Q_L) .$$

By the same argument as used above,

$$Q_H - Q_0 = \rho\, v'(1 - w(\rho))\, w'(\rho)\, \varepsilon + o(\varepsilon) \quad \text{and} \quad Q_0 - Q_L = \rho\, v'(1 - w(\rho))\, w'(\rho)\, \varepsilon + o(\varepsilon).$$

As $\varepsilon \downarrow 0$, $r_L \to G^{-1}(\eta)$ and $r_H \to G^{-1}(1-\eta)$, so

$$-p_0 = \eta\rho\, v'(1 - w(\rho))\, w'(\rho) \left[G^{-1}(1-\eta) - G^{-1}(\eta)\right] \varepsilon + o(\varepsilon).$$

Because $\eta \in (0, 1/2)$, we have $1 - \eta > \eta$; since G^{-1} is strictly increasing and $v', w' > 0$, it follows that $-p_0 > 0$ for all sufficiently small $\varepsilon > 0$.

Finally, the utility gain relative to the laissez-faire outcome is

$$\begin{cases} -p_0 + (r - r_H)(Q_H - Q_0), & \text{if } r > r_H, \\ -p_0, & \text{if } r_L < r \le r_H, \\ -p_0 + (r_L - r)(Q_0 - Q_L), & \text{if } r \le r_L, \end{cases}$$

and every term is strictly positive because $-p_0 > 0$. Hence, for all sufficiently small $\varepsilon > 0$, the mechanism based on the deterministic three-line design $(s^F, s^0, s^S;\, m^F, m^0, m^S)$ strictly Pareto-improves

on the laissez-faire outcome.

A.3 Proofs for Section 4.3

We show that Axiom 3 holds. Fix $m \in (0, 1/2)$ and, for $\varepsilon > 0$ sufficiently small, consider a three-tier allocation that assigns mass m to physical quality $q_0 + \varepsilon$, mass m to physical quality $q_0 - \varepsilon$, and the remaining mass to physical quality q_0. This perturbation preserves average physical quality and is therefore feasible. Writing the induced distribution of effective qualities as

$$F_\varepsilon = m\delta_{v(q_0-\varepsilon)} + (1-2m)\,\delta_{v(q_0)} + m\delta_{v(q_0+\varepsilon)},$$

the corresponding loss in average effective quality is

$$Q_0 - \mathbf{E}_{F_\varepsilon}[Q] = m\,[2v(q_0) - v(q_0-\varepsilon) - v(q_0+\varepsilon)] = o(\varepsilon).$$

However, the cumulative improvement above Q_0 received by the mass m of agents assigned physical quality $q_0 + \varepsilon$ is

$$\mathbf{E}_{F_\varepsilon}[(Q-Q_0)_+] = m\,[v(q_0+\varepsilon) - v(q_0)] = mv'(q_0)\varepsilon + o(\varepsilon).$$

Consequently, Axiom 3 holds:

$$\lim_{\varepsilon \downarrow 0} \frac{Q_0 - \mathbf{E}_{F_\varepsilon}[Q]}{\mathbf{E}_{F_\varepsilon}[(Q-Q_0)_+]} = 0.$$

Proposition 3 is a direct corollary of Theorems 1 and 2.

A.4 Proofs for Section 4.4

We first show that the setting of Section 4.4 does not satisfy Axiom 3. For any feasible F such that $\mathbf{E}_F[(Q-Q_0)_+] > 0$, feasibility implies that both $\Pr_F[Q < Q_0]$ and $\Pr_F[Q > Q_0]$ are positive.

Chebyshev's integral inequality then implies that

$$\begin{cases} \displaystyle\int_{1-\Pr_F[Q>Q_0]}^{1} C(u)\left[F^{-1}(u)-Q_0\right]\,\mathrm{d}u & \geq \dfrac{\mathbf{E}_F[(Q-Q_0)_+]}{\Pr_F[Q>Q_0]}\displaystyle\int_{1-\Pr_F[Q>Q_0]}^{1} C(u)\,\mathrm{d}u, \\ \displaystyle\int_{0}^{\Pr_F[Q<Q_0]} C(u)\left[Q_0-F^{-1}(u)\right]\,\mathrm{d}u & \leq \dfrac{\mathbf{E}_F[(Q_0-Q)_+]}{\Pr_F[Q<Q_0]}\displaystyle\int_{0}^{\Pr_F[Q<Q_0]} C(u)\,\mathrm{d}u. \end{cases}$$

Feasibility requires

$$\begin{aligned} &\int_{1-\Pr_F[Q>Q_0]}^{1} C(u)\left[F^{-1}(u)-Q_0\right]\,\mathrm{d}u \leq \int_{0}^{\Pr_F[Q<Q_0]} C(u)\left[Q_0-F^{-1}(u)\right]\,\mathrm{d}u \\ \implies &\frac{\mathbf{E}_F[(Q_0-Q)_+]}{\Pr_F[Q<Q_0]}\int_{0}^{\Pr_F[Q<Q_0]} C(u)\,\mathrm{d}u \geq \frac{\mathbf{E}_F[(Q-Q_0)_+]}{\Pr_F[Q>Q_0]}\int_{1-\Pr_F[Q>Q_0]}^{1} C(u)\,\mathrm{d}u. \end{aligned}$$

Consequently,

$$\frac{Q_0-\mathbf{E}_F[Q]}{\mathbf{E}_F[(Q-Q_0)_+]} = \frac{\mathbf{E}_F[(Q_0-Q)_+]}{\mathbf{E}_F[(Q-Q_0)_+]} - 1 \geq \frac{\Pr_F[Q<Q_0]\int_{1-\Pr_F[Q>Q_0]}^{1} C(u)\,\mathrm{d}u}{\Pr_F[Q>Q_0]\int_{0}^{\Pr_F[Q<Q_0]} C(u)\,\mathrm{d}u} - 1.$$

The expression on the right is the ratio of the average of C over the uppermost $\Pr_F[Q>Q_0]$ quantiles to its average over the lowermost $\Pr_F[Q<Q_0]$ quantiles, minus one. Moreover, this lower bound is uniform. Viewed as a function of the two probabilities, the expression on the right extends continuously to the compact set of pairs in $[0,1]^2$ whose sum is at most one, where the lower and upper averages at zero equal $C(0)$ and $C(1)$, respectively. It is strictly positive throughout this set because C is strictly increasing, and so it has a strictly positive minimum. Thus, Axiom 3 does not hold in this setting.

We now prove Proposition 4.

Because Axiom 2 holds, Theorem 1 immediately implies that there does not exist a two-tier Pareto improvement over uniform coverage.

We next establish the positive result. Under condition (B),

$$\lim_{m\downarrow 0}\frac{G^{-1}(1-m)}{\frac{1}{m}\int_{1-m}^{1} C(u)\,\mathrm{d}u} = \frac{\overline{r}}{c(\overline{r})} > \frac{\underline{r}}{c(\underline{r})} = \lim_{m\downarrow 0}\frac{G^{-1}(m)}{\frac{1}{m}\int_{0}^{m} C(u)\,\mathrm{d}u}.$$

Hence, by continuity, there exists $m \in (0, 1/2)$ such that

$$\frac{G^{-1}(1-m)}{\frac{1}{m}\int_{1-m}^{1} C(u)\,\mathrm{d}u} > \frac{G^{-1}(m)}{\frac{1}{m}\int_{0}^{m} C(u)\,\mathrm{d}u}.$$

Define

$$Q_L \coloneqq Q_0 - \frac{\varepsilon}{m}\int_{1-m}^{1} C(u)\,\mathrm{d}u \quad \text{and} \quad Q_H \coloneqq Q_0 + \frac{\varepsilon}{m}\int_{0}^{m} C(u)\,\mathrm{d}u.$$

For $\varepsilon > 0$ sufficiently small, consider the three-tier distribution

$$F^{\varepsilon} \coloneqq m\delta_{Q_L} + (1-2m)\,\delta_{Q_0} + m\delta_{Q_H}.$$

Because $Q_0 \in (0, 1)$, all three support points lie in $[0, 1]$ and are distinct when ε is sufficiently small.

This distribution is feasible. Indeed, relative to uniform coverage, its aggregate expected cost is

$$\begin{aligned}\int_0^1 C(u)\left[(F^{\varepsilon})^{-1}(u) - Q_0\right]\mathrm{d}u &= -\frac{\varepsilon}{m}\left[\int_0^m C(u)\,\mathrm{d}u\right]\left[\int_{1-m}^1 C(u)\,\mathrm{d}u\right] \\ &\quad + \frac{\varepsilon}{m}\left[\int_{1-m}^1 C(u)\,\mathrm{d}u\right]\left[\int_0^m C(u)\,\mathrm{d}u\right] = 0.\end{aligned}$$

Let $r_L = G^{-1}(m)$ and $r_H = G^{-1}(1-m)$ denote the two threshold types. If p_L, p_0, and p_H are the corresponding premium adjustments, incentive compatibility requires

$$p_0 - p_L = r_L\,(Q_0 - Q_L) \quad \text{and} \quad p_H - p_0 = r_H\,(Q_H - Q_0)\,.$$

Combining these equalities with budget balance gives

$$\begin{aligned}-p_0 &= m r_H\,(Q_H - Q_0) - m r_L\,(Q_0 - Q_L) \\ &= \varepsilon\left[r_H\int_0^m C(u)\,\mathrm{d}u - r_L\int_{1-m}^1 C(u)\,\mathrm{d}u\right] > 0,\end{aligned}$$

where the inequality follows from the choice of m.

The utility gain relative to uniform coverage is therefore

$$\begin{cases} -p_0 + (r - r_H)\,(Q_H - Q_0)\,, & \text{if } r > r_H, \\ -p_0, & \text{if } r_L < r \le r_H, \\ -p_0 + (r_L - r)\,(Q_0 - Q_L)\,, & \text{if } r \le r_L. \end{cases}$$

Because $-p_0 > 0$, every type is strictly better off. Thus, under condition (B), F^{ε} yields a three-tier strict Pareto improvement over uniform coverage.

B Additional Results and Generalizations

B.1 Proof of Theorem 2*

By condition (ii) and continuity,

$$\lim_{m \downarrow 0} \frac{G^{-1}(1-m)}{\frac{1}{m}\int_{1-m}^{1} \kappa(u)\,\mathrm{d}u} = \frac{G^{-1}(1)}{\kappa(1)} > \frac{G^{-1}(0)}{\kappa(0)} = \lim_{m \downarrow 0} \frac{G^{-1}(m)}{\frac{1}{m}\int_{0}^{m} \kappa(u)\,\mathrm{d}u}.$$

Hence, there exists $m \in (0, 1/2)$ such that

$$\frac{G^{-1}(m)}{G^{-1}(1-m)} < \frac{\int_0^m \kappa(u)\,\mathrm{d}u}{\int_{1-m}^{1} \kappa(u)\,\mathrm{d}u}.$$

Fix such an m and define the bounded nondecreasing function

$$h(u) := \begin{cases} \dfrac{1}{2}\left[\dfrac{G^{-1}(m)}{G^{-1}(1-m)} + \dfrac{\int_0^m \kappa(z)\,\mathrm{d}z}{\int_{1-m}^{1} \kappa(z)\,\mathrm{d}z}\right], & 1-m \le u \le 1, \\ 0, & m \le u < 1-m, \\ -1, & 0 \le u < m. \end{cases}$$

By construction,

$$\int_0^1 \kappa(u)h(u)\,\mathrm{d}u = -\int_0^m \kappa(u)\,\mathrm{d}u + h(1)\int_{1-m}^1 \kappa(u)\,\mathrm{d}u < 0.$$

By condition (i), we obtain a sequence $\{(\varepsilon_n, F_n)\}_{n=1}^{\infty}$ such that $\varepsilon_n \downarrow 0$, $F_n \in \mathcal{F}$, and

$$\int_0^1 \left| \frac{F_n^{-1}(u) - Q_0}{\varepsilon_n} - h(u) \right| \mathrm{d}u \to 0. \tag{6}$$

Next, for each n, we construct a three-tier mean-preserving contraction $\hat{F}_n^{(3)}$ of F_n by pooling qualities within three quantile intervals. To this end, define

$$Q_{L,n} := \frac{1}{m}\int_0^m F_n^{-1}(u)\,\mathrm{d}u, \qquad Q_{M,n} := \frac{1}{1-2m}\int_m^{1-m} F_n^{-1}(u)\,\mathrm{d}u, \qquad Q_{H,n} := \frac{1}{m}\int_{1-m}^1 F_n^{-1}(u)\,\mathrm{d}u.$$

Consider

$$\hat{F}_n^{(3)} := m\delta_{Q_{L,n}} + (1-2m)\,\delta_{Q_{M,n}} + m\delta_{Q_{H,n}}.$$

Since $\hat{F}_n^{(3)}$ preserves the conditional means within each quantile interval, $\hat{F}_n^{(3)} \in \mathrm{MPC}(F_n)$ and is feasible by Axiom 1. Moreover, equation (6) implies that

$$Q_{L,n} = Q_0 - \varepsilon_n + o(\varepsilon_n), \qquad Q_{M,n} = Q_0 + o(\varepsilon_n), \qquad Q_{H,n} = Q_0 + h(1)\varepsilon_n + o(\varepsilon_n).$$

Thus, for all sufficiently large n,

$$Q_{L,n} < Q_0 < Q_{H,n} \quad \text{and} \quad Q_{L,n} < Q_{M,n} < Q_{H,n}.$$

We now modify $\hat{F}_n^{(3)}$ to obtain a three-tier distribution $F_n^{(3)}$ whose middle tier is exactly Q_0. If $Q_{M,n} = Q_0$, no further pooling is required. We thus consider two cases:

1. If $Q_{M,n} < Q_0$, pool the entire mass at $Q_{M,n}$ with enough of the mass at $Q_{H,n}$ for the pooled mass to have mean Q_0.

2. If $Q_{M,n} > Q_0$, pool the entire mass at $Q_{M,n}$ with enough of the mass at $Q_{L,n}$ for the pooled mass to have mean Q_0.

The masses taken from the respective tails are

$$\frac{(1-2m)\,|Q_{M,n}-Q_0|}{Q_{H,n}-Q_0} \quad \text{and} \quad \frac{(1-2m)\,|Q_{M,n}-Q_0|}{Q_0-Q_{L,n}}.$$

Since $Q_{M,n} = Q_0 + o(\varepsilon_n)$, these masses converge to zero for all sufficiently large n. Consequently, this pooling is possible for all sufficiently large n. We write the resulting distribution as

$$F_n^{(3)} = \pi_{L,n}\delta_{Q_{L,n}} + \pi_{0,n}\delta_{Q_0} + \pi_{H,n}\delta_{Q_{H,n}}, \qquad \text{where } \pi_{L,n} + \pi_{0,n} + \pi_{H,n} = 1.$$

Because $F_n^{(3)}$ is a mean-preserving contraction of $\hat{F}_n^{(3)}$, it is feasible by Axiom 1. Moreover, for all sufficiently large n, it has exactly three tiers with strictly positive masses.

Now, consider the mechanism implementing $F_n^{(3)}$. Let $r_{L,n} \coloneqq G^{-1}(\pi_{L,n})$ and $r_{H,n} \coloneqq G^{-1}(1-\pi_{H,n})$ denote the two threshold types. By construction,

$$r_{L,n} \to G^{-1}(m) \quad \text{and} \quad r_{H,n} \to G^{-1}(1-m).$$

Let $p_{L,n}$, $p_{0,n}$, and $p_{H,n}$ denote the corresponding payments. Incentive compatibility requires

$$p_{0,n} - p_{L,n} = r_{L,n}\left(Q_0 - Q_{L,n}\right) \quad \text{and} \quad p_{H,n} - p_{0,n} = r_{H,n}\left(Q_{H,n} - Q_0\right).$$

Combining these equalities with budget balance gives

$$-p_{0,n} = \pi_{H,n} r_{H,n}\left(Q_{H,n} - Q_0\right) - \pi_{L,n} r_{L,n}\left(Q_0 - Q_{L,n}\right).$$

It follows that

$$\frac{-p_{0,n}}{\varepsilon_n} \to m\left[G^{-1}(1-m)h(1) - G^{-1}(m)\right] > 0.$$

Hence, $-p_{0,n} > 0$ for all sufficiently large n.

Finally, the utility gain relative to the equity benchmark is

$$\begin{cases} -p_{0,n} + \left(r - r_{H,n}\right)\left(Q_{H,n} - Q_0\right), & \text{if } r > r_{H,n}, \\ -p_{0,n}, & \text{if } r_{L,n} < r \le r_{H,n}, \\ -p_{0,n} + \left(r_{L,n} - r\right)\left(Q_0 - Q_{L,n}\right), & \text{if } r \le r_{L,n}. \end{cases}$$

Every expression is strictly positive because $-p_{0,n} > 0$. Thus, for every sufficiently large n, the feasible distribution $F_n^{(3)}$ yields a three-tier strict Pareto improvement.

B.2 Proof of Lemma 1

In this appendix, we prove that Axiom 3* is weaker than Axiom 3.

Lemma 1. *Suppose Axioms 1 to 3 hold. Then Axiom 3* is satisfied by choosing* $\kappa \equiv 1$.

Suppose that Axioms 1 to 3 hold. Set $\kappa \equiv 1$; this function is continuous and strictly positive. Under this choice of κ, condition (ii) holds:

$$\frac{G^{-1}(1)}{\kappa(1)} = \overline{r} > \underline{r} = \frac{G^{-1}(0)}{\kappa(0)}.$$

To verify condition (i), fix any bounded nondecreasing function $h : [0,1] \to \mathbf{R}$ such that

$$\int_0^1 h(u)\,\mathrm{d}u < 0.$$

By Axiom 3, for every n, there exists $\hat{F}_n \in \mathcal{F}$ such that

$$\mathbf{E}_{\hat{F}_n}[(Q - Q_0)_+] > 0 \quad \text{and} \quad \frac{Q_0 - \mathbf{E}_{\hat{F}_n}[Q]}{\mathbf{E}_{\hat{F}_n}[(Q - Q_0)_+]} \le \frac{1}{n}.$$

Clearly, $\hat{F}_n \neq \delta_{Q_0}$; thus, Axiom 2 implies that $Q_0 - \mathbf{E}_{\hat{F}_n}[Q] > 0$. Since qualities lie in $[0,1]$,

$$0 < Q_0 - \mathbf{E}_{\hat{F}_n}[Q] \le \frac{1}{n}\mathbf{E}_{\hat{F}_n}[(Q - Q_0)_+] \le \frac{1}{n}.$$

Define

$$\varepsilon_n \coloneqq \frac{Q_0 - \mathbf{E}_{\hat{F}_n}[Q]}{-\int_0^1 h(u)\,\mathrm{d}u} > 0. \tag{7}$$

Passing to a subsequence and relabeling if necessary, we may take $\varepsilon_n \downarrow 0$.

Next, observe that $Q_0 \in (0, 1)$. Indeed, if $Q_0 = 1$, then $\mathbf{E}_F[(Q - Q_0)_+] = 0$ for every $F \in \mathcal{F}$, which contradicts Axiom 3. In addition, if $Q_0 = 0$, any distribution satisfying $\mathbf{E}_F[(Q - Q_0)_+] > 0$ would have strictly positive mean—and hence higher average quality than the equity benchmark, which contradicts the definition of Q_0.

Because h is bounded and nondecreasing, for all sufficiently large n there exists $F_n \in \Delta([0, 1])$ whose quantile function satisfies $F_n^{-1}(u) = Q_0 + \varepsilon_n h(u)$, almost everywhere. By construction, F_n shares the same mean with $\hat{F}_n$:

$$\mathbf{E}_{F_n}[Q] = Q_0 + \varepsilon_n \int_0^1 h(u)\,\mathrm{d}u = \mathbf{E}_{\hat{F}_n}[Q].$$

We now show that F_n is a mean-preserving contraction of $\hat{F}_n$ for all sufficiently large n. To see this, fix any point t between the lowest and highest support points of F_n. Then

$$\mathbf{E}_{F_n}[(Q - t)_+] \le 2\varepsilon_n \|h\|_\infty \quad \text{and} \quad \mathbf{E}_{\hat{F}_n}[(Q - t)_+] \ge \mathbf{E}_{\hat{F}_n}[(Q - Q_0)_+] - \varepsilon_n \|h\|_\infty.$$

Moreover, equation (7) implies that

$$\frac{\varepsilon_n}{\mathbf{E}_{\hat{F}_n}[(Q - Q_0)_+]} = \frac{Q_0 - \mathbf{E}_{\hat{F}_n}[Q]}{-\int_0^1 h(u)\,\mathrm{d}u \cdot \mathbf{E}_{\hat{F}_n}[(Q - Q_0)_+]} \le \frac{1}{-n\int_0^1 h(u)\,\mathrm{d}u} \to 0.$$

Hence, for all sufficiently large n and for all t lying between the lowest and highest support points of F_n,

$$\mathbf{E}_{F_n}[(Q - t)_+] \le \mathbf{E}_{\hat{F}_n}[(Q - t)_+].$$

If t lies below the support of F_n, then equality of means gives

$$\mathbf{E}_{F_n}[(Q - t)_+] = \mathbf{E}_{F_n}[Q] - t = \mathbf{E}_{\hat{F}_n}[Q] - t \le \mathbf{E}_{\hat{F}_n}[(Q - t)_+].$$

If t lies above the support of F_n, then the same inequality holds trivially:

$$\mathbf{E}_{F_n}[(Q-t)_+] = 0 \le \mathbf{E}_{\hat{F}_n}[(Q-t)_+].$$

Thus, this inequality holds for every $t \in \mathbf{R}$; hence, $F_n \in \mathrm{MPC}(\hat{F}_n)$. Since $\hat{F}_n$ is feasible, Axiom 1 implies that F_n is feasible as well.

Finally, observe that for all sufficiently large n, our definition of F_n satisfies

$$\int_0^1 \left| \frac{F_n^{-1}(u) - Q_0}{\varepsilon_n} - h(u) \right| \mathrm{d}u = 0.$$

Passing to a subsequence and relabeling if necessary, we obtain the sequence $\{(\varepsilon_n, F_n)\}_{n=1}^{\infty}$ required by condition (i). Hence, Axiom 3^* holds with $\kappa \equiv 1$.