Buyer-Optimal Information with Nonlinear Technology*

Kai Hao Yang[†]

July 09, 2019

Abstract

This paper explores the buyer-optimal information structures in a monopolistic screening context with nonlinear production technology. It shows that the buyer's optimal surplus may increase even when the production cost becomes more uncertain or when the efficient surplus decreases. Under a binary prior, this paper further shows that the buyer-optimal information structures must lie in a family described by truncated Pareto distributions. Such characterization effectively reduces the surplus maximization problem to a monopsony's pricing problem, which further implies that the buyer-optimal surplus is quasi-convex in technologies that are ranked by the rotational order.

KEYWORDS: Buyer surplus, information structure, mechanism design, nonlinear pricing, comparative statics, virtual valuation.

JEL CLASSIFICATION: C72, D42, D44, D82, D83

*This paper was previously circulated under the title "Strategic Information Acquisition in Auctions" and "Buyer-Optimal Information under Uncertainty". This version supersedes the previous ones.

[†]Department of Economics, University of Chicago, khyang@uchicago.edu. I am indebted to Ben Brooks, Elliot Lipnowski, and Doron Ravid for comments and suggestions on strategies for the proofs. I am grateful for discussions with and comments from Simon Board, Steven Davis, Alex Frankel, Emir Kamenica, Eric Maskin, Stephen Morris, Roger Myerson, Phil Reny, and Nancy Stokey.

1 Introduction

Information that consumers possess when buying a product from a monopolist greatly affects their purchasing decisions and the monopolist's selling strategy, which in turn affects the market outcomes. Moreover, in the information age, there have been increasingly more channels through which consumers can be informed about the products. As a result, a natural question arises: what is the best way for the consumers to be informed about a product so that they can retain as much surplus as possible? Essentially, the answer to this question pertains to balancing the trade-off between the effect of consumers' information on the monopolist's selling strategy and how such information benefits the consumers' purchasing decision.

In the recent literature, Roesler and Szentes (2017) provide a clear answer to the above question in a linear environment. In their paper, the monopolist is assumed to have a (commonly known) constant marginal cost of production.[1] The linearity assumption greatly simplifies the monopolist's selling strategy under a given information structure. After all, with a constant marginal cost and a unit demand, a posted price mechanism is always optimal for the monopolist. With this simplification, Roesler and Szentes (2017) completely characterize the (least informative) buyer-optimal information structure, which features a Pareto-shaped distribution of interim expected value.

However, in many real-world contexts, a monopolist's production technology often features some forms of nonlinearity. More specifically, the monopolist's (constant) marginal cost might be uncertain. For example, when a third party[2] designs information structures to maximize consumer surplus, it often does not have complete information about the monopolist's cost. On the other hand, as summarized in Wilson (1993), many industries actually feature nonlinear pricing instead of a simple posted price (e.g., railroad tariffs, electricity tariffs, airline fairs etc.). That is, there are many situations where the monopolist would prefer certain form of second degree price discrimination to a simple posted price mechanism. These environments naturally entail the assumption of nonlinear production technologies, especially the technology with increasing marginal cost (see, for instance, Mussa and Rosen (1978)).

[1] In the main text, Roesler and Szentes (2017) restrict their attention to a zero marginal cost. In the online appendix, they extend their results to allow positive but constant marginal costs.

[2] For instance, a developer who designs a mobile app that discloses information about restaurants in order to maximize consumer surplus, or a regulator who designs policies that regulate a pharmaceutical firm to disclose information about its products to benefit the consumers.

More importantly, in addition to characterizing the buyer-optimal information structure, it is economically relevant to understand how the production technology affects the optimal consumer surplus. Specifically, given that the consumers can always be informed according to the buyer-optimal information structure, are the consumers always better-off when gains from trade increase? Is cost uncertainty always detrimental to consumer surplus? To answer these questions, it is necessary to examine the buyer-optimal information structures under nonlinear environments.

In nonlinear environments, the monopolist's selling strategy given an information structure becomes much more convoluted than a posted price mechanism. Nonlinearity of the production technology often leads to infinite-dimensional optimal selling mechanisms. As a result, the techniques developed in Roesler and Szentes (2017) are no longer applicable. In fact, finding the buyer-optimal information structures becomes an intricate problem.

This paper attempts to provide a better understanding of the buyer-optimal information structures under nonlinear environments. In this paper, I first underline an observation that the aforementioned two forms of nonlinearity in technology—cost uncertainty and increasing marginal cost—are in fact isomorphic. I then examine how production technology affects consumer surplus. It turns out that after introducing the possibility of nonlinear technologies, many counter-intuitive comparative statics results appear. Specifically, consumer surplus does not necessarily decrease when the production cost becomes more uncertain. Similarly, a decrease in gains from trade does not necessarily lead to a decrease in consumer surplus.

Motivated by these comparative statics, which call for a more comprehensive understanding of the buyer-optimal information structure in nonlinear environments, I then provide a complete characterization of the buyer-optimal information structures under a simplifying assumption that the prior has binary support.[3] By using a local perturbation argument, I show that it is without loss to restrict attention to a family of information structures that induce a Pareto-shaped distribution of interim expected values and can be described by finitely many parameters. Using this

[3]The binary prior assumption greatly simplifies the characterization. Section 7 summarizes the extension to arbitrary prior. However, with arbitrary prior, the set of optimal information structure might involve (countably) infinite partitions on the interval $[0,1]$, and the economic implications are less clear. Nevertheless, from the characterization in Blackwell (1953)), any prior with support on a subset of $[0,1]$ can be thought of as a garbling of a prior with binary support on $\{0,1\}$. Hence, results in this paper can also be regarded as an upper bound of what can be achieved in the richest possible set of information structures that can be considered.

result, maximizing consumer surplus is then effectively equivalent to maximizing a monopsony's surplus where the production technology plays the same rule as a supply function. Together with the insights of Johnson and Myatt (2006), a more comprehensive comparative static result can then be obtained. That is, the buyer-optimal surplus is quasi-convex along a family of technologies ranked by the rotational order. Furthermore, this characterization justifies the outcome selection imposed by Roesler and Szentes (2017).[4] More specifically, the buyer-optimal surplus in a linear environment can be approximated by a sequence of buyer-optimal surplus in nonlinear environments where all the optimal selling mechanisms induce the same outcome. Finally, I conclude this paper by applying the characterization to the case where there are multiple buyers acquiring information simultaneously before an auctioneer designs an optimal auction and solve for the unique symmetric Nash equilibrium.

2 Related Literature

This paper is closely related to the literature on monopolistic pricing and information structures. Lewis and Sappington (1991) and Johnson and Myatt (2006) solve for the optimal information structure within a restricted family for a seller who has a nonnegative production cost, showing that either full or no information is optimal. Bergemann et al. (2015) examine welfare implications of the seller's information about the buyer's value, finding that all the possible surplus divisions—as long as the seller's profit is no less than the uniform pricing profit, the buyer's surplus is nonnegative and the sum of these two is bounded by the efficient surplus—are attainable by some information structure. Roesler and Szentes (2017) characterize the buyer-optimal information structures and provide a similar set-valued prediction in a monopolistic pricing context when the seller has no production cost. Libgober and Mu (2019) consider the max-min mechanism and the min-max information structure for the seller in a dynamic context. Yang (2018) studies an environment in which a seller has private information about the production cost and an intermediary provides information to the buyer while extracting surplus from the seller. A distinct feature of this paper, in contrast, is that no further screening technology is available when designing information and the objective is to maximize the buyer's surplus rather than revenue.

[4] Recall that the buyer-optimal information structure in Roesler and Szentes (2017) requires the seller to charge lowest optimal price, even though there might be many optimal prices.

This paper also relates to the literature on mechanism design and information design. Bergemann and Välimäki (2011) examine an efficiency implementation problem in which agents can acquire information rationally under a set of restricted information structures. Shi (2012) assesses an optimal auction problem where buyers can acquire information rationally under the assumptions that feasible information structures are restricted to one dimension and that buyers acquire information after an auction is chosen.[5] Bergemann and Pesendorfer (2007) examine an optimal disclosure problem where buyers are uninformed *a priori* and the seller can disclose any information to each buyer independently and can design any mechanism. Esö and Szentes (2007) and Li and Shi (2017) explore an optimal disclosure problem of a seller who can disclose information to multiple, partially-informed buyers. In addition, a growing literature on robust mechanism design (e.g., Neeman (2003), Bergemann and Schlag (2011), Carrasco et al. (2018), Bergemann et al. (2017), Du (2018) and Brooks and Du (2019)) studies a setting where the mechanism designer designs a mechanism against Nature, who chooses the worst-case information structure accordingly.

This paper is closest to Roesler and Szentes (2017), who solve for the buyer-optimal information structure when the seller has a constant marginal cost that is common knowledge. I consider the same maximization problem, but allows the marginal cost to be either uncertain or increasing in quantity. In addition, the main characterization of this paper provides a foundation of the outcome selection used in Roesler and Szentes (2017). Methodologically, this paper relies on a generalization of virtual valuation induced by arbitrary distributions, which can be found in Monteiro and Svaiter (2010).

3 Model

3.1 Primitives

A seller (she) sells one unit of a divisible good to a buyer (he) who has outside option zero and a quasi-linear preference with value v. If the quantity consumed by the buyer is q and the amount of payment is p, his *ex-post* payoff would be $qv-p$. Assume that v follows a common prior, which is denoted by a CDF F with $\text{supp}(F)=[0,1]$. Let

[5] In contrast to Shi (2012), the auction application in section 6 considers a model where buyers acquire information *before* the seller designs an auction, without any restriction on the set of feasible information structures.

μ denote the expected value under the prior F. That is,

$$\mu := \int_0^1 vF(\mathrm{d}v).$$

3.2 Information Structure

The buyer does not know the value *a priori*. Instead, he must learn about the value through *information structures*.[6] Due to quasi-linearity, the only payoff-relevant statistic of an information structure for the buyer is the interim expected value. Thus, the information structures can be represented by the collection of all possible marginal distributions of the buyer's interim expected value. Using the characterization of Blackwell (1953), such information structures can be represented by the family of CDFs that are mean-preserving contractions of F.[7] That is, the set of feasible information structures is

$$\mathcal{I}_F := \left\{ G : \mathbb{R}_+ \to [0,1] \middle| \int_0^x (1 - G(z))\,\mathrm{d}z \geq \int_0^x (1 - F(z))\,\mathrm{d}z,\ \forall x \in [0,1], \right.$$
$$\left. \int_0^1 (1 - G(z))\,\mathrm{d}z = \mu,\ G \text{ is a CDF} \right\}.$$

Given an information structure, assume that only the buyer observes the signal realization and hence the interim expected value is his private information. On the other hand, the seller sees the information structure (but not signal realizations) and designs optimal selling mechanism accordingly.

3.3 Production Technology

Below, I introduce two forms of production technology that will be discussed in this paper—uncertain cost and increasing marginal cost.

Uncertain cost— The seller has constant marginal cost of production c drawn from a CDF γ with $\text{supp}(\gamma) \subseteq [0,1]$. Assume that only the seller observes the realization of c. By quasi-linearity, given any information structure $G \in \mathcal{I}_F$ and any realized

[6] An information structure is a Blackwell experiment that specifies a conditional distribution of signal given the true value.

[7] The same observation has been made in recent literature, including Gentzkow and Kamenica (2016), Kolotilin et al. (2017), Roesler and Szentes (2017), Brooks and Du (2019), Dworczak and Martini (2019) and Yang (2018).

marginal cost c, it is without loss to consider posted price mechanisms.[8] Thus, for any $c \in [0,1]$ and any information structure $G \in \mathcal{I}_F$, the seller simply solves

$$\max_{p\in[0,1]} (p-c)(1-G(p^-)),$$

where $G(p^-) := \lim_{\delta\downarrow 0} G(p-\delta)$ is the left-limit of G at p.[9] For any optimal price $p(G,c)$ that the seller sets, the buyer's surplus is

$$\int_{p(G,c)}^1 (x - p(G,c))G(\mathrm{d}x) = \int_{p(G,c)}^1 (1-G(x))\,\mathrm{d}x.$$

Consequently, the buyer's expected surplus under information structure $G \in \mathcal{I}_\mu$ and any optimal price $\{p(G,c)\}_{c\in[0,1]}$ is

$$\int_0^1 \left(\int_{p(G,c)}^1 (1-G(x))\,\mathrm{d}x \right) \gamma(\mathrm{d}c).$$

If a tie is always broken in favor of the buyer (i.e., the seller always sets the smallest optimal posted price when indifferent), the buyer's expected surplus can be written as[10]

$$\int_0^1 \gamma(\psi(x|G))(1-G(x))\,\mathrm{d}x, \tag{1}$$

where $\psi(\cdot|G)$ is the *virtual valuation* induced by distribution G. More formally, as it is well-known that for any regular distribution G with a density $g > 0$, the virtual valuation induced by G, $\psi(x|G)$, is $x - (1-G(x))/g(x)$. For general distributions, Monteiro and Svaiter (2010) offer a formal definition (included in the Supplemental Material). Heuristically, the virtual valuation of G at x, $\psi(x|G)$, is the cost that incentivizes the seller to optimally set a posted price at x (see Lemma 1 for formal discussions). Thus, (1) can be derived by interchanging the order of integration,

[8]To be more precise, notice that the seller has private information about c when designing selling mechanisms. This means that the seller is facing an informed principal problem. Nevertheless, posted price mechanisms are still optimal. To see this, notice that the environment features independent private values and quasi-linear payoffs, and both the seller's and the buyers' payoffs are monotone in their types. By Proposition 8 in Mylovanov and Tröger (2014), it is as if c is commonly known when the seller designs selling mechanisms. Therefore, by Myerson (1981) and Maskin and Zeckhauser (1983), it is without loss to restrict attention to posted price mechanisms.

[9]For subgame perfect equilibrium in this take-it-or-leave-it game to exist, the buyer must choose to buy when indifferent.

[10]When γ is atomless, whichever tie-breaking rule is selected will not affect the expected buyer surplus, and hence this formula always applies regardless of the selected optimal price.

since when $p(G, c)$ is the smallest optimal posted price for a seller with cost c under information structure G, $p(G, c) \leq x$ if and only if $c \leq \psi(x|G)$.[11]

Increasing marginal cost— The seller has a continuous, nondecreasing and convex cost function $C : [0, 1] \to \mathbb{R}_+$ so that it costs $C(q)$ for the seller to produce $q \in [0, 1]$ units. Given any information structure $G \in \mathcal{I}_F$, the Myersonian approach suggests that the optimal mechanism for the seller can be identified by solving[12]

$$\max_{q:[0,1]\to[0,1]} \int_0^1 \Big(q(x)\psi(x|G) - C(q(x)) \Big) G(\mathrm{d}x).$$

Let $V(z) := \max_{q\in[0,1]}[qz - C(q)]$ and let $\gamma(z)$ be a selection of $\mathrm{argmax}_{q\in[0,1]}[qz - C(q)]$ for all $z \in [0, 1]$. Also, let $\gamma(z) := 0$ for any $z \in (-\infty, 0)$. Notice that by the properties of C, γ is well-defined and nondecreasing on $(-\infty, 1]$. With this definition, under any information structure $G \in \mathcal{I}_F$, the seller's optimal menu entails $q(x) = \gamma(\psi(x|G))$ for all reported interim expected values $x \in [0, 1]$, which yields the buyer's surplus

$$\int_0^1 \gamma(\psi(x|G))(1 - G(x))\,\mathrm{d}x. \tag{2}$$

An isomorphism— Notice that, from (1) and (2), the formulae that represent the buyer's surplus under both classes of technology greatly resemble each other. The reason is more than just a coincidental notation choice. In fact, the aforementioned two classes of production technology are isomorphic to each other. To see this, notice that given any convex cost function C, the induced optimal quantity γ can be regarded as a CDF with $\mathrm{supp}(\gamma) \subseteq [0, 1]$ (after a normalization so that $\gamma(1) = 1$).

[11] More precisely, for any $G \in \mathcal{I}_\mu$ and the smallest selection optimal prices $p(c, G) := \min \mathrm{argmax}_{p\in[0,1]}(p - c)(1 - G(p^-))$ for all $c \in [0, 1]$, $p(c, G) \leq x$ if and only if $c \leq \psi(x|G)$. Therefore,

$$\int_0^1 \left(\int_{p(c,G)}^1 (1 - G(x))\,\mathrm{d}x \right) \gamma(\mathrm{d}c) = \int_{p(0,G)}^1 \int_0^{\psi(x|G)} 1\gamma(\mathrm{d}c)(1 - G(x))\,\mathrm{d}x$$
$$= \int_{p(0,G)}^1 \gamma(\psi(x|G))(1 - G(x))\,\mathrm{d}x = \int_0^1 \gamma(\psi(x|G))(1 - G(x))\,\mathrm{d}x,$$

where the last equality follows from the fact that $\gamma(c) = 0$ for all $x < 0$ and that $\psi(x|G) < 0$ for all $x < p(0, G)$.

[12] Again, $\psi(\cdot|G)$ here is the generalized (ironed) virtual valuation of the buyer when the information structure is $G \in \mathcal{I}_\mu$, as formally defined in Monteiro and Svaiter (2010). By Toikka (2011), this is still valid even if the seller has an increasing marginal cost.

Conversely, given any CDF γ with $\mathrm{supp}(\gamma) \subseteq [0,1]$, let

$$C(q) := \int_0^q \gamma^{-1}(z)\,\mathrm{d}z,\ \forall q \in [0,1],$$

where $\gamma^{-1}(q) := \inf\{c \in [0,1] | \gamma(c) \geq q\}$ for all $q \in [0,1]$. Then C is a continuous, nondecreasing and convex cost function. With this observation, the buyer's surplus maximization problem under both classes of production technologies can be summarized by a nondecreasing function γ on $\mathbb{R}$ with $\gamma \equiv 0$ on $(-\infty, 0)$, $0 \leq \gamma \leq 1$ on $[0,1]$ and is a constant on $[1, \infty)$. Let Γ denote the collection of such functions. Then, for any $\gamma \in \Gamma$, finding the buyer-optimal information structure is equivalent to solving

$$\sigma^*(\gamma) := \sup_{G \in \mathcal{I}_F} \int_0^1 \gamma(\psi(x|G))(1 - G(x))\,\mathrm{d}x. \tag{3}$$

It is noteworthy that the objective function of (3) is neither convex nor concave. As a result, no existing methodology can be directly applied to characterize the solution of (3).[13] Thus, solving (3) requires a new set of methods and/or some more further assumptions about the environment. More details about the complete solution can be found in section 5.

4 Buyer-Optimal Surplus and Production Technology

In this section, I examine how $\sigma^*(\gamma)$ differs under different production technology γ. To begin with, I first summarize the characterization in Roesler and Szentes (2017) for linear environments. In this case, the technology is represented by the function γ^c, for some $c \in [0,1]$, where

$$\gamma^c(z) := \begin{cases} 1, & \text{if } z \geq c \\ 0, & \text{if } z < c \end{cases}.$$

As shown in the online appendix of Roesler and Szentes (2017), the (least informative) buyer-optimal information structure under technology γ^c must lie in the following family

$$\mathcal{I}_F^{\mathrm{RS}}(c) := \{G_{\pi,c}^{\eta,\lambda,b} \in \mathcal{I}_F\},$$

[13] In particular, the methods developed in Dworczak and Martini (2019) cannot be applied.

where for any $\pi, \eta, \lambda, b \in [0, 1]$ such that $b \geq \pi/(1-\eta) + c \geq \lambda$,

$$G_{\pi,c}^{\eta,\lambda,b}(x) := \begin{cases} 0, & \text{if } x \in [0, \lambda) \\ \eta, & \text{if } x \in \left[\lambda, \frac{\pi}{1-\eta} + c\right) \\ 1 - \frac{\pi}{x-c}, & \text{if } x \in \left[\frac{\pi}{1-\eta} + c, b\right) \\ 1, & \text{if } x \in [b, 1] \end{cases}.$$

Thus, the (least informative) buyer-optimal information structure and the buyer-optimal surplus $\sigma^*(\gamma^c)$ can be identified by choosing (π, η, b, λ) to maximize the buyer's surplus subject to the constraint that $G_{\pi,c}^{\eta,\lambda,b} \in \mathcal{I}_F$.

One of the most significant implications of the result in Roesler and Szentes (2017) is that if $c = 0$,[14] whenever the buyer-optimal surplus $\sigma^*(\gamma^0)$ is attained, the outcome must be efficient. That is, the buyer's expected surplus and the seller's profit must sum to the efficient surplus μ, which means that there is no distortion at all even though the buyer has some private information under the buyer-optimal information structure. However, as shown in Proposition 1, this surprising result is not generic and does not hold when there is nonlinearity.

Proposition 1. *For any F that has full support, any $\gamma \in \Gamma$ that is strictly increasing on $[0, 1]$, and any information structure $G \in \mathcal{I}_F$, the sum of the buyer's surplus and the seller's profit must be strictly below the efficient surplus level.*

In addition to efficiency, with the introduction of nonlinear production technology, many comparative statics analyses can be conducted. More specifically, a natural and economically relevant question regarding the effects of production technology on the buyer's optimal surplus is that whether or not having more precise knowledge about the seller's (constant) marginal cost benefits the buyer. Alternatively, it is also crucial to understand whether or not the buyer is always worse-off when total surplus decreases. In what follows, I provide two results indicating that the answers to the questions above are both negative. That is, it is possible that some form of cost uncertainty and a reduction in total surplus can actually benefit the buyer.

Proposition 2. *There exists $c \in (0, 1)$ and $\gamma \in \Gamma$ such that γ is a mean-preserving spread of γ^c and*

$$\sigma^*(\gamma) > \sigma^*(\gamma^c).$$

[14] In their online appendix, Roesler and Szentes (2017) also demonstrates that the same property is true for $c > 0$ that is small enough, under a binary prior assumption.

Although it seems that the buyer would be worse-off when more cost uncertainty is introduced—after all, with more information about the production cost, the information structure that maximizes the buyer's surplus can be better tailored—Proposition 2 shows that more cost uncertainty can actually sometimes benefit the buyer. The reason is that when cost uncertainty increases in terms of a mean-preserving spread, some realized marginal cost would be below the expected marginal cost. As the buyer's expected surplus is neither convex nor concave in the chosen information structure, it is possible that the benefit of lower realized marginal cost outweighs the higher realized marginal cost and increases the buyer's surplus.

While Proposition 2 provides a counter-intuitive comparative statics analysis regarding cost uncertainty, Proposition 3 below provides another counter-intuitive comparative statics that involves technologies with increasing marginal cost. To state the result, for any convex and continuous cost function C, let $S(C)$ denote the efficient surplus induced by the cost function C. That is,

$$S(C) := \int_0^1 \max_{q\in[0,1]} [qz - C(q)] F(\mathrm{d}z).$$

When C is affine (i.e., when $C'(q) = c$ for all q), I slightly abuse the notation and let $S(c)$ denote the efficient surplus.

Proposition 3. *There exists $c \in (0, 1)$, $\gamma \in \Gamma$, and a continuous, nondecreasing and convex cost function $C : [0, 1] \to \mathbb{R}_+$ such that $S(C) \leq S(c)$, $\gamma(z) \in \mathrm{argmax}_{q\in[0,1]}[qz - C(q)]$ for all $z \in [0, 1]$ and that*

$$\sigma^*(\gamma) > \sigma^*(\gamma^c).$$

To some extent, Proposition 3 is even more counter-intuitive than Proposition 2. According to Proposition 3, there exists a cost function that induces smaller total gains from trade than what is induced by a constant marginal cost c, yet gives the buyer larger optimal surplus.[15] As shown in the proof of Proposition 3, the reason still relates to the non-concavity of the buyer's surplus as a function of information structures. With this property, by constructing a similar "spread" (not necessarily

[15]It is noteworthy that such a feature cannot be found when restricting attention to linear environments. More specifically, according to the characterization in Roesler and Szentes (2017), the buyer's optimal surplus is always decreasing in the (commonly known) constant marginal cost. As such, the result that smaller total gains from trade may lead to higher buyer surplus can only be discovered after introducing nonlinear technologies.

mean-preserving) of γ^c and then use this spread to back out the associated cost function, Proposition 3 then follows.

With the analyses above, it appears that the comparative statics for the effects of production technology on buyer-optimal surplus is nontrivial. In order to gain a more comprehensive understanding of the relationship between production technology and buyer surplus, a more complete characterization for the buyer-optimal information in nonlinear environments would be necessary. This is the goal for the next section.

5 Characterizing the Buyer-Optimal Information Structures

In this section, I completely characterize the buyer-optimal information structures in any (regular) nonlinear environments under an additional assumption that the prior has binary support. Using this characterization, I then revisit the comparative statics introduced in section 4 and provide a more comprehensive analysis. Finally, I show that the selection imposed when characterizing the buyer-optimal surplus in linear environments can be justified as a limit of a sequence of buyer-optimal surplus in nonlinear environments where the optimal selling mechanisms always induce a unique outcome.

To begin with, consider any prior with binary support $\{0,1\}$. Recall that μ denotes the expected value under the prior. Henceforth, I slightly abuse the notation and let $\mathcal{I}_\mu$ denote the set of all feasible information structures. That is,

$$\mathcal{I}_\mu := \left\{ G : \mathbb{R}_+ \to [0,1] \,\middle|\, \int_0^1 (1 - G(x))\,\mathrm{d}x = \mu,\ \ G \text{ is a CDF, } \mathrm{supp}(G) = [0,1] \right\}.$$

Also, define the class of *regular* technologies as the following.

Definition 1. A production technology $\gamma \in \Gamma$ is said to be *regular* if γ is right-continuous and strictly increasing on $[0,1]$, and is Lipschitz continuous on $(0,1)$.

The main characterization shows that even though $\mathcal{I}_\mu$ is an infinite dimensional space, the surplus maximization problem (3) can be effectively reduced to a finite dimensional constraint maximization problem. To state the result, define a crucial family of information structures $\mathcal{I}^*_\mu$ as the following: For any $\pi, \eta, k \in [0,1]$ such that $\pi/(1-\eta) + k \leq 1$, define $G^\eta_{\pi,k}$ as

$$G^\eta_{\pi,k}(x) := \begin{cases} \eta, & \text{if } x \in \left[0, \frac{\pi}{1-\eta} + k\right) \\ 1 - \frac{\pi}{x-k}, & \text{if } x \in \left[\frac{\pi}{1-\eta} + k, 1\right) \\ 1, & \text{if } x = 1 \end{cases} . \tag{4}$$

Let $\mathcal{I}^*_\mu$ be the collection of CDFs of form (4) that satisfy the mean constraint

$$\mathcal{I}^*_\mu := \left\{ G^\eta_{\pi,k} : [0,1] \to [0,1] \middle| \int_0^1 (1 - G^\eta_{\pi,k}(x))\,\mathrm{d}x = \mu \right\}.$$

In words, an element $G^\eta_{\pi,k} \in \mathcal{I}^*_\mu$ is a *truncated Pareto distribution* that has two mass points at 0 and 1, with mass η and $\pi/(1-k)$, respectively. It is noteworthy that in any linear environment γ^c, as shown in Roesler and Szentes (2017), the family where the (least informative) buyer-optimal information structure lie, $\mathcal{I}^{\mathrm{RS}}_F(c)$, can be reduced to $\mathcal{I}^{\mathrm{RS}}_\mu(c) := \{G^\eta_{\pi,c} \in \mathcal{I}_\mu\}$. Thus, the collection $\mathcal{I}^*_\mu$ is exactly the union of the parameterized families that contains the buyer-optimal information structures for all constant marginal costs c. That is, $\mathcal{I}^*_\mu = \cup_{c\in[0,1]}\mathcal{I}^{\mathrm{RS}}_\mu(c)$. With the definition of $\mathcal{I}^*_\mu$, Theorem 1 below summarizes the main characterization.

Theorem 1. *For any regular $\gamma \in \Gamma$, the set of optimal information structures under γ is nonempty and is a subset of $\mathcal{I}^*_\mu$. That is,*

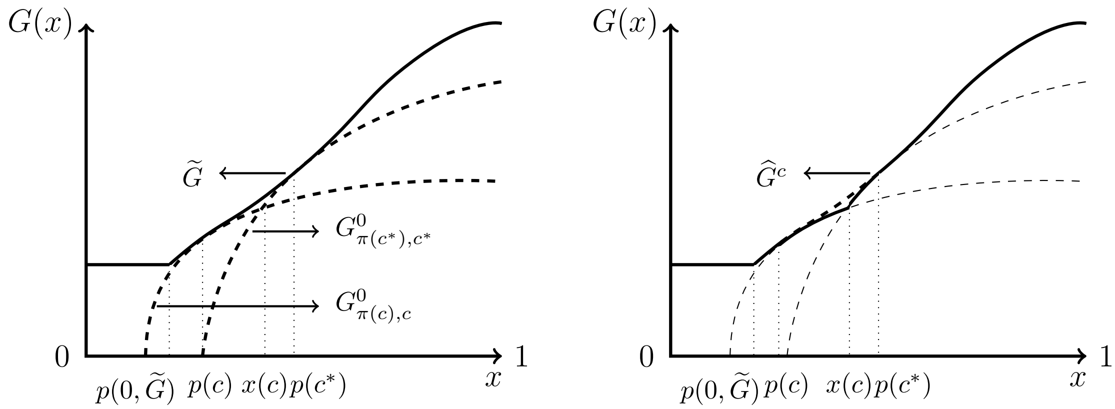
$$\sigma^*(\gamma) = \max_{G^\eta_{\pi,k}\in\mathcal{I}^*_\mu} \int_0^1 \gamma(\psi(x|G^\eta_{\pi,k}))(1 - G^\eta_{\pi,k}(x))\,\mathrm{d}x.$$

The proof of Theorem 1 can be found in the appendix. In brief, Theorem 1 is proved by first arguing that the buyer's surplus as a function of information structure is upper-semicontinuous under any regular technology $\gamma \in \Gamma$. Then, for any information structure $G \in \mathcal{I}_\mu \backslash \mathcal{I}^*_\mu$, I construct a local perturbation of G so that the perturbed information structure is feasible and strictly improves the buyer's surplus.

While the formal proof of Theorem 1 is relatively complex as it involves local perturbations of any arbitrary CDF, the reason why an information structure $\widetilde{G} \notin \mathcal{I}^*_\mu$ cannot be optimal can be well illustrated when $\widetilde{G}$ is regular (i.e., when $\psi(x|\widetilde{G}) = x - (1 - \widetilde{G}(x))/\widetilde{g}(x)$ is continuous and strictly increasing). Heuristically, suppose that for the maximization problem (3), the solution exists and strong duality holds.[16] Then solving (3) is equivalent to solving the following Lagrangian

$$\max_G \int_0^1 \gamma(\psi(x|G))(1 - G(x))\,\mathrm{d}x + \lambda^* \left(\mu - \int_0^1 (1 - G(x))\,\mathrm{d}x \right),$$

[16] Recall that the objective of (3) depends on the choice variable in a complicated way. As a result, no existing sufficient conditions for strong duality can be applied for (3). The discussion here is merely for illustrations. The formal proof in the appendix does not (directly) rely on any arguments about the duality theorem. On the other hand, existence of the solution is in fact not an assumption but a result of upper-semicontinuity of the buyer's surplus as a function of information structure, see the Lemma 3 in the appendix.



(a) $\widetilde{G}$ (b) $\widehat{G}$

Figure 1: Constructing $\widehat{G}^c$ from $\widetilde{G}$

where $\lambda^* > 0$ is a Lagrange multiplier and the maximum is taken over all CDF G with $\text{supp}(G) \subseteq [0, 1]$.[17] As a result, proving Theorem 1 is equivalent to showing that any $\widetilde{G} \notin \mathcal{I}^*_\mu$ cannot be a solution of the following maximization problem

$$\max_G \int_0^1 (\gamma(\psi(x|G)) - \lambda^*)(1 - G(x))\,\mathrm{d}x. \tag{5}$$

To see this, first notice that for any CDF $\widetilde{G}$, $\widetilde{G}$ is optimal for (5) only if $\widetilde{G}(x) = \widetilde{G}(p(0, \widetilde{G}))$ for all $x \in [0, p(0, \widetilde{G})]$, where $p(0, \widetilde{G})$ is smallest optimal price for a seller with marginal cost 0 under $\widetilde{G}$. Indeed, recall that $0 = \gamma(z) < \lambda^*$ for all $z < 0$. Thus, if $\widetilde{G}(x) < G(p(0, \widetilde{G}))$ for some $x \in [0, p(0, \widetilde{G}))$, the objective of (5) can be strictly improved by moving all the probability weights on $(0, p(0, \widetilde{G})]$ to 0.[18]

Now suppose that $\lambda^* > \gamma(1)$. Then clearly any $\widetilde{G}$ except for $\widetilde{G} \equiv 1$ on $(0, 1]$ cannot be optimal. On the other hand, suppose that $\lambda^* \in (0, \gamma(1))$. Consider any

[17]More specifically, as it can be shown that (3) is in fact equivalent to

$$\begin{aligned} &\sup_G \int_0^1 \gamma(\psi(x|G))(1 - G(x))\,\mathrm{d}x \\ &\text{s.t.} \int_0^1 (1 - G(x))\,\mathrm{d}x \le \mu, \end{aligned}$$

where the supremum is taken over all the CDF G with $\text{supp}(G) \subseteq [0, 1]$. When strong duality holds, the associated Lagrange multiplier must be positive.

[18]This also explains why $G^\eta_{\pi,k} \in \mathcal{I}^*_\mu$ can have a mass point at zero.

$\widetilde{G} \notin \mathcal{I}^*_\mu$ with $\widetilde{G}(x) = \widetilde{G}(p(0,\widetilde{G}))$ for all $x \in [0, p(0,\widetilde{G})]$ (the solid curve in Figure 1a) and suppose that $\psi(\cdot|\widetilde{G})$ is continuous and strictly increasing on $[p(0,\widetilde{G}), 1]$. Since γ is regular, $c^* := \gamma^{-1}(\lambda^*)$ is well-defined and positive. Moreover, since $\psi(\cdot|\widetilde{G})$ is strictly increasing, $p(c^*) := \psi^{-1}(c^*|\widetilde{G})$ is well-defined and is (the unique) optimal price for a seller with marginal cost c^* (see Lemma 1). Therefore, if

$$\pi(c^*) := \max_{p\in[0,1]} (p - c^*)(1 - \widetilde{G}(p^-)) = (p(c^*) - c^*)(1 - \widetilde{G}(p(c^*))),$$

then $1 - \pi(c^*)/(x - c^*) \leq \widetilde{G}(x)$ for all $x \in [\pi(c^*) + c^*, 1]$, as illustrated by the dotted curve $G^0_{\pi(c^*),c^*}$ in Figure 1a. Similarly, for any $c \in [0, c^*)$, $p(c) := \psi^{-1}(c|\widetilde{G})$ is also well-defined and is (the unique) optimal price for a seller with marginal cost c. Thus, for

$$\pi(c) := \max_{p\in[0,1]} (p - c)(1 - \widetilde{G}(p^-)) = (p(c) - c)(1 - \widetilde{G}(p(c))),$$

$1 - \pi(c)/(x - c) \leq \widetilde{G}(x)$ for all $x \in [\pi(c) + c, 1]$ as well, as shown by the dotted curve $G^0_{\pi(c),c}$ in Figure 1a. With these observations, for each $c \in [0, c^*]$, define $\widehat{G}^c$ as the following

$$\widehat{G}^c(x) := \begin{cases} \widetilde{G}(x), & \text{if } x \in [0, p(c)) \\ 1 - \frac{\pi(c)}{x-c}, & \text{if } x \in [p(c), x(c)) \\ 1 - \frac{\pi(c^*)}{x-c}, & \text{if } x \in [x(c), p(c^*)) \\ \widetilde{G}(x), & \text{if } x \in [p(c^*), 1] \end{cases},$$

as illustrated by the solid curve in Figure 1b, where $x(c)$ is the unique intersection between the Pareto curves induced by c and c^*.[19] Then since $\widetilde{G} \notin \mathcal{I}^*_\mu$, $\widetilde{G} \neq \widehat{G}^c$ for all $c \in [0, c^*)$. Moreover, the gain in the objective of (5) from moving $\widetilde{G}$ to $\widehat{G}^c$ is given by

$$\Psi(c) := \int_{p(c)}^{x(c)} (\gamma(c) - \lambda^*)(1 - \widehat{G}^c(x))\, \mathrm{d}x - \int_{p(c)}^{p(c^*)} (\gamma(\psi(x|\widetilde{G})) - \lambda^*)(1 - \widetilde{G}(x))\, \mathrm{d}x.$$

It can then be shown that $\Psi(c^*) = \Psi'(c^*) = 0$ and $\Psi''(c^*) > 0$. Therefore, there exists $c < c^*$ such that $\Psi(c) > 0$, which means that $\widehat{G}^c$ improves the objective of (5) relative to $\widetilde{G}$ and hence $\widetilde{G}$ is not a solution.

[19] More precisely,

$$x(c) := \frac{\pi(c)c^* - \pi(c^*)c}{\pi(c) - \pi(c^*)}$$

and thus

$$\frac{\pi(c)}{x(c) - c} = \frac{\pi(c^*)}{x(c) - c^*}.$$

As a remark, notice that the construction of $\widehat{G}^c$ is only possible when $\widetilde{G} \notin \mathcal{I}_\mu^*$, otherwise $x(c)$ would not be well-defined. This property also plays a crucial role in the formal proof of Theorem 1.

In essence, Theorem 1 suggests that even when facing a nonlinear production technology, the buyer-optimal information structure must still have a Pareto shape and can be described by three parameters. Furthermore, as a feature of Pareto distributions, for any $G_{\pi,k}^\eta \in \mathcal{I}_\mu^*$, the induced virtual valuation takes a simple form. That is, $\psi(x|G_{\pi,k}^\eta) = k$ for all $x \in \mathrm{supp}(G_{\pi,k}^\eta)\backslash\{0,1\}$. Therefore, (3) is effectively reduced to a finite-dimensional constraint optimization problem, as the buyer's surplus under any $G_{\pi,k}^\eta \in \mathcal{I}_\mu^*$ is simply

$$\int_0^1 \gamma(\psi(x|G_{\pi,k}^\eta))(1 - G_{\pi,k}^\eta(x))dx = \gamma(k)(\mu - (\pi + (1-\eta)k))$$

and the constraint is simply

$$\int_0^1 (1 - G_{\pi,k}^\eta(x))\,\mathrm{d}x = \pi + (1-\eta)k + \pi \log\left(\frac{(1-\eta)(1-k)}{\pi}\right) = \mu.$$

In fact, this finite-dimensional problem can be further simplified by first solving

$$\begin{aligned} &\min_{\pi,\eta} \left[\pi + (1-\eta)k\right] \\ &\text{s.t. } \pi \log\left(\frac{(1-\eta)(1-k)}{\pi}\right) + \pi + (1-\eta)k = \mu, \\ &\qquad \eta \in [0,1],\ \pi \in [0, (1-k)(1-\eta)]. \end{aligned} \tag{6}$$

for each $k \in [0,1]$. Let $\omega(k)$ the value of (6). The surplus maximization problem (3) then reduces to a one-dimensional problem

$$\max_{k\in[0,1]} \gamma(k)(\mu - \omega(k)). \tag{7}$$

Furthermore, it is straightforward to show that ω is continuous on $[0,1]$, differentiable on $(0,1)$ and strictly increasing on $[0,1]$ with $\omega(1) = \mu$. As a result, an equivalent way to write (7) is

$$\max_{p\in[\omega(0),\mu]} \gamma \circ \omega^{-1}(p)(\mu - p). \tag{8}$$

Thus, (8) suggests that maximizing buyer's surplus is in fact equivalent to maximizing a monopsonist's surplus whose value of the good is μ and is facing a supply function $\gamma \circ \omega^{-1}$. This observation, together with the insight of Johnson and Myatt (2006), yields an unambiguous comparative static analysis. More specifically, consider a family of technologies $\{\gamma_\alpha | \alpha \in [0,1]\} \subseteq \Gamma$ such that $\partial \gamma_\alpha(c)/\partial \alpha$ exists for all $\alpha \in [0,1]$ and for all $c \in [0,1]$. Define an ordering on $\{\gamma_\alpha | \alpha \in [0,1]\}$ as the following.

Definition 2. A family $\{\gamma_\alpha | \alpha \in [0,1]\} \subseteq \Gamma$ is said to be ranked by the *rotational order* if there exists $\{c_\alpha | \alpha \in [0,1]\} \subseteq [0,1]$ such that c_α is nondecreasing in α and that

$$\frac{\partial \gamma_\alpha(c)}{\partial \alpha} \leq 0 \text{ if } c < c_\alpha, \quad \text{and} \quad \frac{\partial \gamma_\alpha(c)}{\partial \alpha} \geq 0 \text{ if } c > c_\alpha, \, \forall \alpha \in [0,1].$$

Motivated by the analyses in Johnson and Myatt (2006), as a corollary of Theorem 1, the following comparative statics can then be derived.

Theorem 2. *For any family of technologies $\{\gamma_\alpha | \alpha \in [0,1]\} \subseteq \Gamma$ ranked by the rotational order, the buyer-optimal surplus is quasi-convex in α. That is, for any $\alpha_1, \alpha_2 \in [0,1]$ and any $\lambda \in [0,1]$,*

$$\sigma^*(\gamma_{\lambda\alpha_1 + (1-\lambda)\alpha_2}) \leq \max\{\sigma^*(\gamma_{\alpha_1}), \sigma^*(\gamma_{\alpha_2})\}.$$

Proof. By Theorem 1 and (7), for any $\alpha \in [0,1]$,

$$\sigma^*(\gamma_\alpha) = \max_{k \in [0,1]} \gamma_\alpha(k)(\mu - \omega(k)),$$

Furthermore, since $\{\gamma_\alpha | \alpha \in [0,1]\}$ is ranked by the rotational order, for any $k \in [0,1]$, the function $\alpha \mapsto \gamma_\alpha(k)(\mu - \omega(k))$ is quasi-convex. Therefore, as a function of α, $\sigma^*(\gamma_\alpha)$ is quasi-convex in α as it is a pointwise supremum of a family of quasi-convex functions. This completes the proof. ∎

Theorem 2 means that for any family of technologies $\{\gamma_\alpha | \alpha \in [0,1]\} \subseteq \Gamma$ ranked by the roational order, the buyer's optimal surplus as a function of the parameter α is either increasing, decreasing, or U-shaped. Therefore, as the technology is "rotated" clockwisely, the buyer's surplus either increases, or decreases, or decreases first and then increases. In terms of probability distributions, higher α corresponds more dispersion; while in terms of cost functions, higher α means that the cost function becomes "more convex". As a result, Theorem 2 implies that for any family of technologies $\{\gamma_\alpha | \alpha \in [0,1]\}$ ranked by the rotational order, the buyer would always prefer one of the extremes (γ_0 or γ_1) than any other technologies in the family. In particular, for a family of distributions of cost ranked by the rotational order, the buyer would prefer either the most dispersed or the least dispersed distribution.

Similarly, for a family of cost functions such that the optimal quantity can be ranked by the rotational order, the buyer would always prefer the "most convex" or the "least convex" technology.

Returning to the comparative statics analyses in section 4, notice that in both of the proofs of Proposition 2 and Proposition 3, the constructed γ is a rotation of γ^c. This means that when specialized to the case where F has support $\{0, 1\}$, Proposition 2 and Proposition 3 are examples of the general result given by Theorem 2. Combining with the monopsony interpretation given by (8), Theorem 2 gives a further reason for the counter-intuitive results in the previous section. More generally, in addition to Theorem 2, Theorem 1 implies that when the prior has support $\{0, 1\}$, the comparative statics for how the changes in production technology affect the buyer-optimal surplus is equivalent to the comparative statics for how the changes of a supply curve affect a monopsonist's surplus.

In addition to comparative statics, Theorem 1 is useful for providing a foundation for the selection of outcome imposed by Roesler and Szentes (2017) when studying the linear environments. More specifically, in linear environments, the buyer-optimal information structure characterized by Roesler and Szentes (2017) has a feature that the seller is *indifferent* in charging any price in the support of the distribution of interim expected value. Only by imposing a selection that prescribes the seller to charge the lowest price can the buyer-optimal surplus be attained. Such ability to select the most preferred optimal price is essential to the analyses in the linear environment, especially because there is also a price that is optimal for the seller but gives zero surplus to the buyer under the same (buyer-optimal) information structure.

In contrast, under regular nonlinear environments, there is generically a unique outcome that can be induced by the seller's optimal selling mechanisms under any information structures. As a result, the buyer-optimal surplus in linear environments can be approximated by a sequence of buyer-optimal surplus in nonlinear environments, without imposing any selections of the outcomes. In other words, in terms of buyer surplus, only the selection that is the best for the buyer is robust to small perturbation of the production technology. This provides a foundation that justifies the selection imposed by Roesler and Szentes (2017). The formal result is stated in Theorem 3 below.

Theorem 3. *For any $c \in [0, 1)$, there exists a sequence of regular technologies $\{\gamma_n\} \subseteq \Gamma$ such that $\{\gamma_n\} \to \gamma^c$ pointwisely; that for any $n \in \mathbb{N}$ and any $G \in \mathcal{I}_\mu$, the seller's* *optimal selling mechanisms under γ_n induce a unique outcome; and that*

$$\lim_{n\to\infty} \sigma^*(\gamma_n) = \sigma^*(\gamma^c).$$

6 Strategic Information Acquisition in Auctions

Besides facilitating the comparative statics and providing a foundation for the selection imposed in linear environments, another application of Theorem 1 can be found in the context of strategic information acquisition when there are multiple buyers. In many real-world scenarios, when a seller tries to sell a single object to multiple buyers, the buyers might not have complete information about the object *a priori*. Instead, the buyers must learn about the object and acquire information about their own valuation. For example, in online auctions, buyers typically learn about objects by viewing advertisements, examining product details, and obtaining information elsewhere. In fine art or house auctions, buyers examine objects in person and gather information. In these contexts, the seller is often able to observe how buyers examine objects,[20] and therefore buyers' information acquisition behaviors affect how the seller sells the object. As a result, buyers behave strategically when acquiring information because their behaviors affect both their posterior beliefs about the object's value and the information rents they retain.

Formally, consider a model in which a seller sells an indivisible good. There are $N \geq 2$ potential buyers with quasi-linear preferences and outside option zero. Each buyer i has a private value $v_i \in \{0, 1\}$ for the good so that, if the probability of having the good is q_i and the amount of payment is p_i, the *ex-post* payoff would be $q_i v_i - p_i$. The values are independently and identically distributed according to the common prior so that $\mathbb{P}(v_i = 1) = \mu$ for all i. Buyer i does not know the realization of v_i *a priori* and must choose an information structure $G \in \mathcal{I}_\mu$ to learn about v_i.[21] Timing of the events is:

1. Nature draws $\{v_i\}_{i=1}^N$ independently from the prior.

2. Each buyer, without seeing v_i, chooses an information structure in $\mathcal{I}_\mu$ simultaneously.

[20] In online auctions, the seller is able to do so with accesses to buyers' cookies. In fine art or house auctions, buyers' physical presence enables the seller to observe.

[21] As buyers acquire information independently, their signals conditional on v_i are assumed to be independent. Thus, interim expected values are also independent, and each buyer's information structure can still be represented by a CDF $G \in \mathcal{I}_\mu$.

3. The seller sees each buyer's chosen information structure and designs a mechanism.

4. The buyers see their own signal realizations, make participation decisions, and take actions in the mechanism if participating.

At the mechanism design stage, the seller knows the distributions of each buyer's interim expected value, and the interim expected values are independent as buyers acquire information independently and simultaneously. Thus, although buyers cannot observe each other's chosen information structure when playing in the mechanism, the seller's optimal mechanism is the same as if buyers can, because the optimal Bayesian mechanism is dominant-strategy implementable in this independent private value environment (Gershkov et al., 2013). Given a profile of chosen information structures $\{G_i\}_{i=1}^N \subseteq \mathcal{I}_\mu$, the seller's optimal mechanism is the Myersonian optimal auction (Myerson, 1981).[22] That is, the allocation rule is to sell the product to the buyer with the highest nonnegative virtual valuation[23] according to the reported interim expected value.[24] Transfers are then pinned down by incentive compatibility and individual rationality constraints. Therefore, given a profile of chosen information structures $\{G_i\}_{i=1}^N$, buyer i's expected surplus is

$$U(G_i, G_{-i}) := \int_0^1 \gamma_{G_{-i}}(\psi(x|G_i))(1 - G_i(x))dx, \tag{9}$$

where $\gamma_{G_{-i}}$ is the distribution of the highest virtual valuation among all buyers $j \neq i$, with adjustments made to incorporate tie-breaking rules.[25]

[22]Although the optimal auction characterized in Myerson (1981) requires buyers' value distributions to admit continuous densities, Monteiro and Svaiter (2010) provide a formal proof that extends the result to arbitrary distributions.

[23]Implicitly, it is assumed that the seller always breaks ties in favor of the buyers.

[24]Since the interim expected value is the only payoff-relevant statistic to each buyer, it is without loss to restrict attention to mechanisms that ask buyers to report interim expected values when solving for optimal mechanisms.

[25]Formally, given any $G \in \mathcal{I}_\mu$, define $\varphi_G(q) := \int_0^1 \mathbf{1}\{\psi(x|G) \leq q\} G(dx)$ for all $q \in (-\infty, 1]$. For any $i \in \{1, \dots, N\}$ and any $G_{-i} \in \mathcal{I}_\mu^{N-1}$, let $\mathbb{W}_i := \{W \subseteq \{1, \dots, N\} | i \in W\}$ and let

$$\gamma_{G_{-i}}(q) := \mathbf{1}\{q \geq 0\} \sum_{W \in \mathbb{W}_i} \beta_i(W) \prod_{j \in W \setminus \{i\}} (\varphi_{G_j}(q) - \varphi_{G_j}(q^-)) \prod_{j \notin W} \varphi_{G_j}(q^-), \, \forall q \in (-\infty, 1].$$

Here, $\beta_i(W)$ is induced by a given tie-breaking rule that gives the good to buyer i with probability $\beta_i(W)$ when W is the set of winners. For example, the uniform tie-breaking rule is the one where $\beta_i(W) = 1/|W|$ for all $W \in \mathbb{W}_i$ and for all $i \in \{1, \dots, N\}$.

Taking this into account, at the information-acquisition stage, buyers effectively play a symmetric (mixed extension) strategic form game with strategy space $\Delta(\mathcal{I}_\mu)$[26] and payoff function $\bar{U} : \Delta(\mathcal{I}_\mu)^N \to \mathbb{R}$, where $\bar{U}$ is the expected payoff function induced by U.[27] From (9), all the strategically relevant components in this game are encoded in functions $\{\gamma_{G_{-i}}\}_{i=1}^N$. Thus, each buyers' payoff function $\bar{U}$ is

$$\bar{U}(\nu_i, \nu_{-i}) = \int_{\mathcal{I}_\mu} \left(\int_0^1 \gamma_{\nu_{-i}}(\psi(x|G_i))(1 - G_i(x))dx \right) \nu_i(dG_i),$$

where $\gamma_{\nu_{-i}}(z) := \int_{\mathcal{I}_\mu^{N-1}} \gamma_{G_{-i}}(z)\nu_{-i}(dG_{-i})$ and $\nu_{-i} := \prod_{j \neq i} \nu_j$. Therefore, the information-acquisition stage can be studied by solving for Nash equilibria of this strategic form game. Since buyers are *ex-ante* symmetric, I restrict attention to symmetric equilibria.

Theorem 1 is useful for characterizing symmetric equilibria of the information-acquisition game. According to Theorem 1, if strategy profile $(\nu_1, \dots, \nu_N)$ is such that $\gamma_{\nu_{-i}} \in \Gamma$ and is regular, then buyer i's best response must lie in $\mathcal{I}_\mu^*$. Intuitively, as Theorem 1 studies a single buyer's surplus maximization problem with uncertainty of the seller's production cost, the multiple-buyer information-acquisition game endogenizes this uncertainty from the perspective of a single buyer. After all, the seller has an *opportunity cost* to sell to a single buyer because when doing so, she forfeits the opportunity to extract surplus from others. As buyers do not observe each other's signal realization, it is as if each single buyer does not know the realization of the seller's production costs.

Formally, Theorem 1 yields a characterization of the symmetric equilibria of the information-acquisition game. To see this, suppose that all buyers randomize according to a strategy ν, with $\text{supp}(\nu) \subseteq \mathcal{I}_\mu^*$. To simplify the notations, use γ_ν to denote $\gamma_{(\nu,\dots,\nu)}$. For $(\nu, \dots, \nu)$ to be an equilibrium, it must be that for each buyer, his best response against ν is a subset of $\text{supp}(\nu) \subseteq \mathcal{I}_\mu^*$. Moreover, for each buyer, when all opponents use strategy ν, choosing an information structure $G_{\pi,k}^\eta \in \mathcal{I}_\mu^*$ gives payoff

$$\gamma_\nu(k)(\mu - (\pi + (1 - \eta)k)).$$

By (6), this buyer's best response against ν must further be restricted to the set $\{G_{\pi(k),k}^{\eta(k)} | k \in [0, 1]\}$, where $(\pi(k), \eta(k))$ is the (unique) solution of (6).

[26]Endowed with the Borel σ-algebra generated by the weak-*topology on $\Delta(\mathcal{I}_\mu)$.

[27]That is, $\bar{U}(\nu_1, \dots, \nu_N) := \int_{\mathcal{I}_\mu^N} U(G_1, \dots, G_N)\nu_1(dG_1) \cdots \nu_N(dG_N)$ for any $(\nu_1, \dots, \nu_N) \in \Delta(\mathcal{I}_\mu)^N$.

As a result, for $(\nu, \dots, \nu)$ to be an equilibrium, ν must be represented by the push-forward measure

$$\Lambda(k) := \nu\left(\left\{G^{\eta(z)}_{\pi(z),z}\middle| z \leq k\right\}\right), \forall k \in [0,1].$$

Therefore, for such a strategy profile to be an equilibrium, it must be that:

$$\gamma_\nu(k)(\mu - \omega(k)) \geq \gamma_\nu(k')(\mu - \omega(k')), \tag{10}$$

for all $k \in \text{supp}(\Lambda)$ and all $k' \in [0,1]$. By the property of ω and the fact that γ_ν is nondecreasing, it can be shown that Λ is atomless and $\text{supp}(\Lambda) = [0, \bar{k}]$ for some $\bar{k} \in (0,1)$. Together, the indifference condition implied by (10) can be written as

$$\gamma_\nu(k) = \frac{\bar{\nu}(\mu - \omega(\bar{k}))}{\mu - \omega(k)}, \forall k \in [0, \bar{k}],$$

where $\bar{\nu} := \gamma_\nu(\bar{k})$. Moreover, since under this strategy profile, γ_ν is the distribution of the highest virtual valuation among $N-1$ buyers, it must be that[28]

$$\frac{\bar{\nu}(\mu - \omega(\bar{k}))}{\mu - \omega(k)} = \gamma_\nu(k) = \left(\int_0^k \left(1 - \frac{\pi(z)}{1-z}\right)\Lambda(dz) + \int_k^{\bar{k}} \eta(z)\Lambda(dz)\right)^{N-1}, \forall k \in [0, \bar{k}]. \tag{11}$$

(11) is a functional equation that involves CDF Λ and constants $\bar{k}, \bar{\nu}$. Evaluating (10) at $k = \bar{k}$, it further implies that

$$\bar{\nu}^{\frac{1}{N-1}} = \int_0^{\bar{k}} \left(1 - \frac{\pi(z)}{1-z}\right)\Lambda(dz).$$

By the differentiability of ω, any Λ that solves (11) must be absolutely continuous and therefore admits a density λ that can be pinned down by differentiating both sides of (11). Therefore, let

$$\lambda(k) := \frac{1}{N-1}\left(\frac{(\mu - \omega(k))}{\bar{\nu}(\mu - \omega(\bar{k}))}\right)^{\frac{N-2}{N-1}} \left(\frac{\bar{\nu}(\mu - \omega(\bar{k}))\omega'(k)}{(\mu - \omega(k))^2\left(1 - \frac{\pi(k)}{1-k} - \eta(k)\right)}\right).$$

[28]Notice that the term in the parenthesis on the right hand side is the probability of one buyer's virtual valuation being less than or equal to k when randomizing on $\{G^{\eta(k)}_{\pi(k),k} | k \in [0,1]\}$ according to Λ. Therefore, the right hand side is the induced distribution of the highest virtual valuation among $N-1$ buyers when they use an (atomless) strategy Λ.

and let $\bar{k}$, $\bar{\nu}$ be the unique solution of the system[29]

$$\begin{aligned}&\int_0^{\bar{k}} \left(1 - \frac{\pi(z)}{1-z}\right) \lambda(z) dz = \bar{\nu}^{\frac{1}{N-1}} \\ &\int_0^{\bar{k}} \lambda(z) dz = 1.\end{aligned} \tag{12}$$

Furthermore, let

$$\lambda^*(k) := \begin{cases} \lambda(k), & \text{if } k \in [0, \bar{k}] \\ 0, & \text{otherwise} \end{cases}, \tag{13}$$

and

$$\Lambda^*(k) := \int_0^k \lambda^*(z) dz. \tag{14}$$

It then follows that Λ^* is the unique solution of (11). Therefore, for $\nu^* \in \Delta(\mathcal{I}_\mu)$ defined by

$$\nu^* \left(\left\{ G_{\pi(z),z}^{\eta(z)} \middle| z \leq k \right\} \right) := \Lambda^*(k), \, \forall k \in [0, 1], \tag{15}$$

$(\nu^*, \dots, \nu^*)$ is the only possible candidate for a symmetric equilibrium that has support in $\mathcal{I}_\mu^*$. As Proposition 4 below suggests, since γ_{ν^*} satisfies the conditions in Theorem 1, it follows that ν^* indeed constitutes an equilibrium. Combined with a local perturbation analogous to the proof of Theorem 1, such equilibrium is in fact the unique symmetric equilibrium of the information-acquisition game.

Proposition 4. *The information-acquisition game possesses a unique symmetric equilibrium $(\nu^*, \dots, \nu^*)$, where $\nu^* \in \Delta(\mathcal{I}_\mu^*)$ is defined by (12), (13), (14) and (15).*

To understand the intuitions behind Proposition 4, first, by a similar local perturbation argument as used in the proof of Theorem 1, it can be shown that for any (symmetric) equilibrium, the support of the equilibrium strategy must lie in $\mathcal{I}_\mu^*$. Since any information structure in $\mathcal{I}_\mu^*$ induces virtual valuation that is constant with positive probability, the equilibrium strategy must entail randomization so that the induced γ_ν does not allow positive probability of a tie. Otherwise, any buyer who has positive probability of encountering a tie can perturb his strategy locally and win the object in this event at an ignorable cost. Together, these imply that any symmetric equilibrium must be a randomization on $\mathcal{I}_\mu^*$. The indifference condition

[29]Existence and uniqueness are ensured by the intermediate value theorem and the properties that $\lim_{k \uparrow \bar{k}} \omega(k) = \mu$ and that ω is strictly increasing.

further pins down the exact shape of such randomization. On a higher level, while Theorem 1 effectively reduces the surplus maximization problem for a single buyer to a monopsonist's problem, when there are multiple buyers competing in information acquisition, Proposition 4 reduces the game to a Bertrand competition—just as when more buyers are introduced to a monopsony environment.

Using the characterization in Proposition 4, several economic implications can be discovered. First, comparing to the single buyer case in Roesler and Szentes (2017), when there are multiple buyers, competition in information acquisition decreases buyers' expected surplus and increases the seller's revenue. For example, when $N = 1$, $\mu = 3/4$, the buyer's surplus is approximately 0.38 and the seller's profit is approximately 0.37. In contrast, when $N = 2$, $\mu = 1/2$,[30] in the unique symmetric equilibrium, buyers' total surplus is approximately 0.18 while the seller's revenue is about 0.52. In addition, a limiting result (Proposition 5) suggests that although each individual buyer wishes to refrain from acquiring full information to retain information rents, as the number of buyers approaches infinity, the equilibrium information acquisition still converges to full information. Meanwhile, buyers' surplus converges to zero and the seller's revenue converges to 1. Specifically, for each $N \in \mathbb{N}$, let $\Lambda_N \in \Delta([0,1])$ be the push-forward measure associated with the unique symmetric equilibrium ν_N^* when there are N buyers. Let $\bar{\sigma}_N$ be the buyers' total expected surplus and $\bar{\pi}_N$ the seller's expected profit in this equilibrium. Also, let $\bar{\tau}_N := \bar{\sigma}_N + \bar{\pi}_N$ be the total surplus. Proposition 5 below summarizes the limiting behaviors of these outcomes.

Proposition 5. *As $N \to \infty$,*

1. *$\{\Lambda_N\}$ converges to the degenerate distribution $\delta_{\{1\}}$ under the weak-* topology.*

2. *$\lim_{N\to\infty} \bar{\sigma}_N = 0$.*

3. *$\lim_{N\to\infty} \bar{\pi}_N = \lim_{N\to\infty} \bar{\tau}_N = 1$.*

7 Conclusion

In this paper, I explore the effects of the production technology on the buyer-optimal information structure and the associated buyer-optimal surplus. For any prior of the

[30] When $N = 2$, there are two independent draws of values and thus the efficient surplus mechanically increases comparing to the case when $N = 1$. As such, the prior mean is normalized to keep the efficient surplus at the same level.

buyer's value, I show that it is possible that increased cost uncertainty can benefit the buyer. In addition, I show that there exists a cost function that induces less efficient surplus than a linear cost function but increases the buyer's optimal surplus. Then I completely characterize the buyer-optimal information structures under an assumption that the prior has binary support. With this complete characterization, I then show that maximizing buyer's surplus is equivalent to maximizing a monopsonist's surplus. Finally, I apply this characterization to study an environment where there are multiple buyers for a single object and each buyer can acquire information independently. I characterize the unique symmetric equilibrium and show that this equilibrium must involve randomization. Limiting behaviors of this equilibrium are further examined. As the number of buyers approaches to infinity, buyers' acquired information converges to full information and the seller achieves full surplus extraction.

Several extensions and related economic problems can be further investigated. For instance, the assumption that the prior of the buyer's value has binary support, which greatly simplifies the representation of information structures and hence simplifies local perturbation arguments, can be relaxed. For a general prior, using similar local perturbation arguments, the buyer-optimal information structure must exhibit a property that resembles the *monotone partition*, as discussed by Kolotilin et al. (2017) and Dworczak and Martini (2019). More specifically, there exists an increasing sequence $\{x_n\}$ that forms a partition of $[0,1]$, such that in the interior of each element of this partition, (x_n, x_{n+1}), the buyer's interim expected value must follow a Pareto distribution, and the second-order stochastic dominance constraint binds at x_n and x_{n+1}. However, for an arbitrarily distributed value, the representation of information structures involves pointwise integral constraints and is thus much less tractable. In particular, the sequence $\{x_n\}$ can be (countably) infinite, which means that the buyer's surplus maximization problem can still be infinite-dimensional even when restricting attention to this family of Pareto distributions. The exact characterization of the solution depends on further assumptions about the prior. This can be a topic for future research.

Another related economic problem is to explore the set of all possible surplus division that can be induced by some information structure for the buyer in nonlinear environments. In a companion paper (Yang, 2019), I extend Theorem 1 and provide such a characterization. Using this characterization, I further show that for any technology that converges to the linear technology, the set of possible surplus division also converges, generalizing Proposition 4. On the other hand, I also show that there exists some technology such that the set of possible surplus divisions collapses to a one-dimensional in some regions.

An additional possible application considers a seller's optimal selling mechanism that is robust to mis-specification of a buyer's information structure when the seller has an increasing marginal production cost. That is, the seller designs an optimal selling mechanism to solve the max-min problem against Nature, who chooses the worst-case information structure for any mechanism selected. When the production cost is a constant and is common knowledge, the buyer-optimal information structure corresponds to a min-max solution of Nature, as the buyer-optimal information structure in Roesler and Szentes (2017) achieves efficient surplus. With an increasing marginal production cost, buyer-optimal information structures may not achieve efficient surplus in general, as shown in Proposition 1. However, the result in Yang (2019) indicates that the local perturbation arguments can also be applied to solve the associated min-max problem, which appears to be useful for identifying the max-min solution. These can be topics for future studies.

References

Bergemann, D., B. Brooks, and S. Morris (2015): "The Limits of Price Discrimination," *American Economic Review*, 105, 921–957.

——— (2017): "Informationally Robust Optimal Auction Design," Working Paper.

Bergemann, D. and M. Pesendorfer (2007): "Information Structures in Optimal Auctions," *Journal of Economic Theory*, 137, 580–609.

Bergemann, D. and K. Schlag (2011): "Robust Monopoly Pricing," *Journal of Economic Theory*, 146, 2527–2543.

Bergemann, D. and J. Välimäki (2011): "Information Acquisition and Efficient Mechanism Design," *Econometrica*, 70, 2510–2526.

Blackwell, D. (1953): "Equivalent Comparisons of Experiments," *Annals of Mathematical Statistics*, 24, 265–272.

Brooks, B. and S. Du (2019): "Optimal Auction Design with Common Values: An Informationally-Robust Approach," Working Paper.

Carrasco, V., V. F. Luz, N. Kos, M. Messner, P. Monteiro, and H. Moreira (2018): "Optimal Selling Mechanisms under Moment Conditions," *Journal of Economic Theory*, 177, 2527–2543.

Du, S. (2018): "Robust Mechanism under Common Valuation," *Econometrica*, 86, 1569–1588.

Dworczak, P. and G. Martini (2019): "The Simple Economics of Optimal Persuasion," *Journal of Political Economy*, forthcoming.

Esö, P. and B. Szentes (2007): "Optimal Information Disclosure in Auctions and the Handicap Auction," *Review of Economic Studies*, 74, 705–731.

Gentzkow, M. and E. Kamenica (2016): "A Rothschild-Stiglitz Approach to Bayesian Persuasion," *American Economic Review: Papers and & Proceedings*, 106, 597–601.

Gershkov, A., J. K. Goeree, A. Koshnir, B. Moldovanu, and X. Shi (2013): "On the Equivalence of Bayesian and Dominant Strategy Implementation," *Econometrica*, 81, 197–220.

Johnson, J. P. and D. P. Myatt (2006): "On the Simple Economics of Advertising, Marketing, and Product Design," *American Economic Review*, 96, 756–784.

Kolotilin, A., M. Li, T. Mylovanov, and A. Zapechelnyuk (2017): "Persuasion of a Privately Informed Receiver," *Econometrica*, 85, 1949–1964.

Lewis, T. R. and D. E. M. Sappington (1991): "Supplying Information to Facilitate Price Discrimination," *International Economic Review*, 35, 309–327.

Li, H. and X. Shi (2017): "Discriminatory Information Disclosure," *American Economic Review*, 107, 3363–3385.

Libgober, J. and X. Mu (2019): "Information Robustness in Intertemporal Pricing," Working Paper.

Maskin, E. and R. Zeckhauser (1983): "Optimal Selling Strategies: When to Haggle, When to Hold Firm," *MQuarterly Journal of Economics*, 98, 267–289.

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

Monteiro, P. K. (2015): "A Note on the Continuity of the Optimal Auction," *Economics Letters*, 137, 127–130.

Monteiro, P. K. and B. F. Svaiter (2010): "Optimal Auction with a General Distribution: Virtual Valuation without Densities," *Journal of Mathematical Economics*, 46, 21–31.

Mussa, M. and S. Rosen (1978): "Monopoly and Product Quality," *Journal of Economic Theory*, 18, 301–307.

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

Mylovanov, T. and T. Tröger (2014): "Mechanism Design by an Informed Principal: The Quasi-Linear Private-Values Case," *Review of Economic Studies*, 81, 1668–1707.

Neeman, Z. (2003): "The Effectiveness of English Auctions," *Games and Economic Behavior*, 43, 214–218.

Roesler, A.-K. and B. Szentes (2017): "Buyer-Optimal Learning and Monopoly Pricing," *American Economic Review*, 107, 2072–2080.

Shi, X. (2012): "Optimal Auctions with Information Acquisition," *Games and Economic Behavior*, 74, 666–686.

Toikka, J. (2011): "Ironing without Control," *Journal of Economic Theory*, 146, 2510–2526.

Wilson, R. B. (1993): *Nonlinear Pricing*, Oxford: Oxford University Press.

Yang, K. H. (2018): "Informational Monopolist: An Equivalence Result," Working Paper.

——— (2019): "Informationally Robust Welfare Predictions under Second-Degree Price Discrimination," Working Paper.

Appendix

This appendix contains all the proofs for the result presented in the main text. Section A provides the proof for the main characterization result, Theorem 1. Section B contains the proofs for all the other results. For some of the proofs, several claims are used. The proof of these claims can be found in the Supplemental Material. Also, the proof of Lemma 7, which is used in section B.4 is completely analogous to the proof of Lemma 2 and thus is also relegated to the Supplemental Material.

Throughout this appendix, for any CDF G with $\text{supp}(G) \subseteq [0,1]$, let $a := \inf\{x \in [0,1] | G(x) > 0\}$ and define

$$H(x|G) := \begin{cases} 0, & \text{if } x \in [0,a) \\ a - x(1-G(x)), & \text{if } x \in [a,1] \end{cases}.$$

Let $\phi(\cdot|G)$ be the *generalized convex hull* of H under measure G (see Milgrom and Segal (2002) and the Supplemental Material for formal definitions). By definition, for any $x \in [0,1]$ the virtual valuation at x, $\psi(x|G)$, is a G-sub-gradient of $\phi(\cdot|G)$. The following Lemma will be used in the proofs of the rest the of results.

Lemma 1. *Fix any CDF R with $\text{supp}(R) \subseteq [0,1]$. Suppose that $x_0 \in (0,1)$ is such that $\psi(x_0^+|R) > \psi(x|R)$ for all $x \in [0, x_0)$. Then for any $k \in [\psi(x_0^-|R), \psi(x_0^+|R)]$,*

$$(x-k)(1-R(x)) \leq (x-k)(1-R(x^-)) \leq (x_0-k)(1-R(x_0^-)), \forall x \in [0,1]$$

and the inequality is strict if $\phi(x|R) < H(x|R)$ or $\psi(x|R) \neq k$. In particular, let $r := \sup\{x \in [0,1] | \psi(x|R) < 0\}$. Then

$$x(1-R(x)) \leq x(1-R(x^-)) \leq r(1-R(r^-)), \forall x \in [0,1].$$

Proof. Suppose that there exists some $x_0 \in (0,1)$ such that $\psi(x_0^+|R) > \psi(x|R)$ for all $x \in [0, x_0)$. For any $k \in [\psi(x_0^-|R), \psi(x_0^+|R)]$. Consider a profit maximization problem of a monopolist with a constant marginal cost k. Standard envelope arguments give that a mechanism (p,y) is incentive compatible if and only if $y(x) = y(0) + p(x)x - \int_0^x p(t)dt$ for all $x \in [0,1]$ and p is nondecreasing. Let

$$p^*(x) := \begin{cases} 1, & \text{if } \psi(x|R) \geq k \\ 0, & \text{if } \psi(x|R) < k \end{cases}.$$

for all $x \in [0,1]$. Then, for any nondecreasing $q : [0,1] \to [0,1]$,

$$\begin{aligned}
&\int_0^1 q(x)(x-k)R(\mathrm{d}x) - \int_0^1 q(x)(1-R(x))\,\mathrm{d}x \\
\leq &\int_0^1 q(x)(\psi(x|R)-k)R(\mathrm{d}x) \\
\leq &\int_0^1 p^*(x)(\psi(x|R)-k)R(\mathrm{d}x) \\
= &\int_0^1 p^*(x)(x-k)R(\mathrm{d}x) - \int_0^1 p^*(x)(1-R(x))\,\mathrm{d}x,
\end{aligned}$$

where the first and the last equality follow from Theorem 2 and Theorem 3 in Monteiro and Svaiter (2010) since q is nondecreasing and p^* is measurable with respect to the σ-algebra generated by $\psi(\cdot|R)$, which follows from the fact that $\psi(x_0^+|R) > \psi(x|R)$ for all $x \in [0, x_0)$. The second inequality follows from the definition of p^*. Since p^* corresponds to a posted price mechanism with a price x_0, the maximal profit is given by $(x_0-k)(1-R(x_0^-))$. Thus, for any other posted price $x \in [0,1]$,

$$(x-k)(1-R(x)) \leq (x-k)(1-R(x^-)) \leq (x_0-k)(1-R(x_0^-)),$$

and the inequality is strict when $\phi(x|R) < H(x|R)$ or $\psi(x|R) \neq k$, as desired. In particular, since by definition $\psi(r^+|R) > 0 > \psi(x|R)$ for all $x \in [0,r)$, we may take $r = x_0$. Finally, since $0 \in [\psi(r^-|R), \psi(r^+|R)]$, k can always be replaced by 0. ■

A Proof of Theorem 1

The proof of Theorem 1 is divided into two subsidiary lemmas. Lemma 2 shows that for any distribution R that is not in $\mathcal{I}_\mu^*$, one can always find a local perturbation $\widehat{R}$ that strictly improves the buyer's surplus. Lemma 3 then shows that the buyer's surplus, as a function of information structure, is upper-semicontinuous under the weak-* topology. Together, these two lemmas prove Theorem 1.

Lemma 2. *Consider any $\gamma \in \Gamma$ that is regular. For any $R \in \mathcal{I}_\mu \backslash \mathcal{I}_\mu^*$, there exists some $\widehat{R} \in \mathcal{I}_\mu$ such that*

$$\int_0^1 \gamma(\psi(x|R))(1-R(x))\,\mathrm{d}x < \int_0^1 \gamma(\psi(x|\widehat{R}))(1-\widehat{R}(x))\,\mathrm{d}x.$$

Lemma 3. *For any $\gamma \in \Gamma$ that is regular, the function*

$$R \mapsto \int_0^1 \gamma(\psi(x|R))(1-R(x))dx$$

is upper-semicontinuous under the weak- topology on $\mathcal{I}_\mu$.*

A.1 Proof of Lemma 2

Proof of Lemma 2. Consider any $R \in \mathcal{I}_\mu \backslash \mathcal{I}_\mu^*$. Let $a := \inf\{x \in [0,1] | R(x) > 0\}$, $b := \sup\{x \in [0,1] | R(x) < 1\}$ be the lower bound and upper bound of $\text{supp}(R)$, respectively, and $r := \inf\{x \in [0,1] | \psi(x|R) > 0\}$. Since $R \in \mathcal{I}_\mu$,

$$1 - \mu - \int_0^1 R(x)\,\mathrm{d}x = 1 - \mu - \int_r^1 R(x)\,\mathrm{d}x - \int_a^r R(x)\,\mathrm{d}x = 0. \tag{A.1}$$

Since R is nondecreasing, it must be that

$$1 - \mu - \int_r^1 R(x)\,\mathrm{d}x - rR(r) \le 1 - \mu - \int_r^1 R(x)\,\mathrm{d}x - (r-a)R(r) \le 0.$$

Consider three cases separately:

Case 1: $R(r^-) > 0$ and

$$1 - \mu - \int_r^1 R(x)\,\mathrm{d}x - rR(r) < 0. \tag{A.2}$$

Take any $k > 0$ and let $\rho := \inf\{x \in [0,1] | \psi(x|G) \ge k\}$. Clearly, $\rho \ge r$. From Lemma 1, $(x-k)(1-R(x)) \le (\rho - k)(1 - R(\rho^-))$ for all $x \in [0,1]$. Therefore,

$$R(x) \ge \max\left\{0, 1 - \frac{\zeta}{x-k}\right\}, \tag{A.3}$$

for all $x \in [0,r]$, where $\zeta := (\rho - k)(1 - R(\rho^-))$. Moreover, since $\psi(x|R) < 0$ for all $x \in [0,r)$ and $R(r^-) > 0$, for k close enough to r, there exists $z \in [0,r)$ such that $R(x) > 1 - \zeta/(x-k) > 0$ for all $x \in (z,r)$. Therefore, for k close enough to 0,

$$1 - \mu - \int_\rho^1 R(x)\,\mathrm{d}x - \int_0^\rho \left(1 - \frac{\zeta}{(x-k)^+}\right)^+ \mathrm{d}x > 1 - \mu - \int_0^1 R(x)\,\mathrm{d}x = 0.$$

Together with (A.2), by the intermediate value theorem, there exists $\underline{x} \in (0,r)$ such that

$$1 - \mu - \int_\rho^1 R(x)\,\mathrm{d}x - \int_{\underline{x}}^\rho \left(1 - \frac{\zeta}{x-k}\right)\mathrm{d}x - \underline{x}\left(1 - \frac{\zeta}{\underline{x}-k}\right) = 0 \tag{A.4}$$

with $1 - \zeta/(\underline{x} - k) \ge 0$. Now define

$$\widehat{R}(x) := \begin{cases} R(x), & \text{if } x \in (\rho, 1] \\ 1 - \frac{\zeta}{x-k}, & \text{if } x \in (\underline{x}, \rho] \\ 1 - \frac{\zeta}{\underline{x}-k}, & \text{if } x \in [0, \underline{x}] \end{cases}.$$

By (A.4), $\widehat{R} \in \mathcal{I}_\mu$. Furthermore, by construction,

$$\psi(x|\widehat{R}) = \begin{cases} \psi(x|R), & \text{if } x \in (\rho, 1] \\ k, & \text{if } x \in (\underline{x}, \rho] \\ \underline{\psi}, & \text{if } x \in [0, \underline{x}]. \end{cases},$$

for some $\underline{\psi} < 0$. Therefore,

$$\begin{aligned} &\int_0^1 \gamma(\psi(x|\widehat{R}))(1 - \widehat{R}(x))\,\mathrm{d}x \\ =& \int_{\underline{x}}^{\rho} \gamma(k)\left(1 - \frac{\zeta}{x-k}\right)\mathrm{d}x + \int_\rho^1 \gamma(\psi(x|R))(1-R(x))\,\mathrm{d}x \\ >& \int_r^1 \gamma(\psi(x|R))(1-R(x))\,\mathrm{d}x \\ =& \int_0^1 \gamma(\psi(x|R))(1-R(x))\,\mathrm{d}x, \end{aligned}$$

where the strict inequality follows from $k > 0$ and γ being strictly increasing. Thus, there is some $\widehat{R} \in \mathcal{I}_\mu$ such that

$$\int_0^1 \gamma(\psi(x|R))(1-R(x))\,\mathrm{d}x < \int_0^1 \gamma(\psi(x|\widehat{R}))(1-\widehat{R}(x))\,\mathrm{d}x,$$

as desired.

Case 2:

$$1 - \mu - \int_r^1 R(x)\,\mathrm{d}x - rR(r) = 0. \tag{A.5}$$

Then by (A.1) and (A.5),

$$0 = 1 - \mu - \int_r^1 R(x)\,\mathrm{d}x - \int_a^r R(x) \geq 1 - \mu - \int_r^1 R(x)\,\mathrm{d}x - (r-a)R(r) = aR(r) \geq 0$$

and hence $R(r) = R(r^-)$, $\int_a^r (R(r) - R(x))\,\mathrm{d}x = 0$, $aR(r) = 0$ and therefore $R(r) = R(a)$.

First notice that if $\psi(x|R) = \bar{\psi}$ for all $x \in (r, b)$, let $\zeta := (r - \bar{\psi})(1 - R(r^-))$, then by Lemma 1

$$R(x) \geq 1 - \frac{\zeta}{x - \bar{\psi}}$$

for all $x \in [0, 1]$. Furthermore, since $R \notin \mathcal{I}^*_\mu$, $\phi(x|R) < H(x|R)$ for some $x \in (r, b)$ and thus, by Lemma 1 again, the inequality is strict for some $x \in (r, b)$ such that $1 - \zeta/(x - \bar{\psi}) > 0$. Together with (A.5), there exists $\bar{b} \in [\mu, b)$ such that for

$$\widehat{R}(x) := \begin{cases} 0, & \text{if } x \in \left[0, \frac{\zeta}{1-R(r)} + \bar{\psi}\right) \\ 1 - \frac{\zeta}{x - \bar{\psi}}, & \text{if } x \in \left[\frac{\zeta}{1-R(r)} + \bar{\psi}, \bar{b}\right) \\ 1, & \text{if } x \in [\bar{b}, 1]. \end{cases},$$

$\widehat{R} \in \mathcal{I}_\mu$. Then,

$$\int_0^1 \gamma(\psi(x|R))(1-R(x))\,\mathrm{d}x = \gamma(\bar{\psi})\int_r^1 (1-R(x))\,\mathrm{d}x = \gamma(\bar{\psi})\,(\mu - r(1-R(r)))$$

$$=\gamma(\bar{\psi})(\mu-(r-\bar{\psi})(1-R(r))-\bar{\psi}(1-R(r))) = \gamma(\bar{\psi})(\mu-\zeta-\bar{\psi}(1-R(r))) = \int_0^1 \gamma(\psi(x|\widehat{R}))(1-\widehat{R}(x))\,\mathrm{d}x.$$

Furthermore, since $\bar{b} < b < 1$, there exists $\pi \in [0, 1-\bar{\psi}]$, $\eta \in [0, 1-\mu]$ such that

$$\int_0^1 \gamma\left(\psi\left(x\middle|G^\eta_{\pi,\bar{\psi}}\right)\right)\left(1-G^\eta_{\pi,\bar{\psi}}(x)\right)\mathrm{d}x = \gamma(\bar{\psi})(\mu-\pi-(1-\eta)\bar{\psi}) > \gamma(\bar{\psi})(\mu-\zeta-\bar{\psi}(1-R(r))).$$

Together, there exists $G^\eta_{\pi,\bar{\psi}} \in \mathcal{I}_\mu$ such that

$$\int_0^1 \gamma\left(\psi\left(x\middle|G^\eta_{\pi,\bar{\psi}}\right)\right)\left(1-G^\eta_{\pi,\bar{\psi}}(x)\right)\mathrm{d}x > \int_0^1 \gamma(\psi(x|R))(1-R(x))\,\mathrm{d}x,$$

as desired.

Now consider the case when there exists $x_1, x_2 \in (r, b)$ such that $\psi(x_2|R) > \psi(x_1|R) \geq 0$. Suppose first that there exists some $x_0 \in (r, 1)$ and a sequence $\{x_n\}$ such that $x_n < x_{n+1} < x_0$ and $\psi(x_n|R) < \psi(x_{n+1}|R) < \psi(x_0^-|R)$ for all $n \in \mathbb{N}$ and that $\{x_n\} \uparrow x_0$. Since $\psi(x_{n+1}|R) > \psi(x_n|R)$ for all $n \in \mathbb{N}$ and $\psi(\cdot|R)$ is nondecreasing, we may take such $\{x_n\}$ so that $\psi(x_n^+|R) > \psi(y|R)$ for all $y \in [r, x_n)$ for all $n \in \mathbb{N}$. Let $k := \psi(x_0^-|R)$ and $\zeta_0 := (x_0 - k)(1-R(x_0^-))$. Also, for each $n \in \mathbb{N}$, let $\zeta_n := (x_n - \psi(x_n^+|R))(1-R(x_n^-))$. Then by Lemma 1, since $\psi(x_n^+|R) > \psi(y|R)$ for all $y \in [r, x_n)$ and since $\psi(x_{n+1}|R) > \psi(x_n|R)$, we must have

$$R(x) \geq \max\left\{1-\frac{\zeta_0}{x-k},\, 1-\frac{\zeta_n}{x-\psi(x_n|R)},\, 0\right\}, \tag{A.6}$$

for all $x \in [x_n, x_0]$, for all $n \in \mathbb{N}$, with strict inequality holding at some $x \in (x_n, x_0)$. Therefore, there exists $\bar{n} \in \mathbb{N}$ such that whenever $n > \bar{n}$, there exists $\underline{x}_n \in (r, x_n)$ and $\hat{x}_n \in (x_n, x_0)$ such that

$$\int_0^{x_0} R(x)\,\mathrm{d}x = \underline{x}_n R(\underline{x}_n) + \int_{\underline{x}_n}^{x_n} R(x)\,\mathrm{d}x + \int_{x_n}^{x_0} \max\left\{1-\frac{\zeta_0}{x-k},\, 1-\frac{\zeta_n}{x-\psi(x_n|R)}\right\}\mathrm{d}x \tag{A.7}$$

$$1-\frac{\zeta_0}{\hat{x}_n-k} = 1-\frac{\zeta_n}{\hat{x}_n-\psi(x_n|R)}. \tag{A.8}$$

As such, for any $n > \bar{n}$, define

$$\widehat{R}^{x_n}(x) := \begin{cases} R(x), & \text{if } x \in [x_0, 1] \\ 1-\frac{\zeta_0}{x-k}, & \text{if } x \in [\hat{x}_n, x_0) \\ 1-\frac{\zeta_n}{x-\psi(x_n|R)}, & \text{if } x \in [x_n, \hat{x}_n) \\ R(x), & \text{if } x \in [\underline{x}_n, x_n) \\ R(\underline{x}_n), & \text{if } x \in [0, \underline{x}_n) \end{cases},$$

where $\hat{x}_n$ and $\underline{x}_n$ are uniquely defined by (A.7) and (A.8). Notice that by (A.6), $\hat{x}_n < \hat{x}_{n+1}$ for all $n \in \mathbb{N}$, $\{\hat{x}_n\} \uparrow x_0$ and $\underline{x}_n > \underline{x}_{n+1}$ for all $n \in \mathbb{N}$, $\{\underline{x}_n\} \downarrow r$.

By construction, for all $x \in [0,1]$, for any $n > \bar{n}$,

$$\psi(x|\widehat{R}^{x_n}) = \begin{cases} \psi(x|R), & \text{if } x \in [x_0, 1] \\ k, & \text{if } x \in [\hat{x}_n, x_0) \\ \psi(x_n|R), & \text{if } x \in [x_n, \hat{x}_n) \\ \psi(x|R), & \text{if } x \in [\underline{x}_n, x_n) \\ \underline{\psi}_n, & \text{if } x \in [0, \underline{x}_n) \end{cases},$$

for some $\underline{\psi}_n < 0$. As such, the difference in expected surplus between choosing $\widehat{R}^{x_n}$ and R is

$$\begin{aligned}
&\int_0^1 \gamma(\psi(x|\widehat{R}^{x_n}))(1-\widehat{R}^{x_n}(x))\,\mathrm{d}x - \int_0^1 \gamma(\psi(x|R))(1-R(x))\,\mathrm{d}x \\
=&\gamma(k)(r(1-R(r)) - \underline{x}_n(1-R(\underline{x}_n))) + \int_r^{\underline{x}_n} (\gamma(k)-\gamma(\psi(x|R)))(1-R(x))\,\mathrm{d}x \\
&- \int_{x_n}^{\hat{x}_n} \big((\gamma(k)-\gamma(\psi(x_n|R)))(1-\widehat{R}^{x_n}(x))\big)\,\mathrm{d}x \\
&+ \int_{x_n}^{x_0} \big(\gamma(k)-\gamma(\psi(x|R)))(1-R(x))\big)\,\mathrm{d}x.
\end{aligned}$$

Notice that by Lemma 1 and $R(r) = R(r^-)$, $r(1-R(r)) = r(1-R(r^-)) \geq \underline{x}_n(1-R(\underline{x}_n))$ for all $n > \bar{n}$.

On the other hand, for any $\underline{x} \in (r, x_0)$, let

$$\begin{aligned}
\Psi(\underline{x}) :=& \int_r^{\underline{x}} \big(\gamma(k)-\gamma(\psi(x|R))\big)(1-R(x))\,\mathrm{d}x \\
&- \int_{\bar{x}(\underline{x})}^{\hat{x}(\underline{x})} \big(\gamma(k)-\gamma(\psi(\bar{x}(\underline{x})|R)\big)\big(1-\widehat{R}^{\underline{x}}(x)\big)\,\mathrm{d}x \qquad \text{(A.9)}\\
&+ \int_{\bar{x}(\underline{x})}^{x_0} \big(\gamma(k)-\gamma(\psi(x|R))\big)(1-R(x))\,\mathrm{d}x,
\end{aligned}$$

where $\hat{x}(\underline{x})$ and $\bar{x}(\underline{x})$ are uniquely defined by

$$\int_0^{x_0} R(x)\,\mathrm{d}x = \underline{x}R(\underline{x}) + \int_{\underline{x}}^{\bar{x}(\underline{x})} R(x)\,\mathrm{d}x + \int_{\bar{x}(\underline{x})}^{x_0} \max\left\{1-\frac{\zeta_0}{x-k}, 1-\frac{\zeta_{\bar{x}(\underline{x})}}{x-\psi(\bar{x}(\underline{x})|R)}\right\}\mathrm{d}x \tag{A.10}$$

$$1-\frac{\zeta_0}{\hat{x}(\underline{x})-k} = 1-\frac{\zeta_{\bar{x}(\underline{x})}}{\hat{x}(\underline{x})-\psi(\bar{x}(\underline{x})|R)}, \tag{A.11}$$

and $\widehat{R}^{\underline{x}}$ is defined by

$$\widehat{R}^{\underline{x}}(x) := \begin{cases} R(x), & \text{if } x \in [x_0, 1] \\ 1 - \frac{\zeta_0}{x-k}, & \text{if } x \in [\hat{x}(\underline{x}), x_0) \\ 1 - \frac{\zeta_{\bar{x}(\underline{x})}}{x-\psi(\bar{x}(\underline{x})|R)}, & \text{if } x \in [\bar{x}(\underline{x}), \hat{x}(\underline{x})) \\ R(x), & \text{if } x \in [\underline{x}, \bar{x}(\underline{x})) \\ R(\underline{x}), & \text{if } x \in [0, \underline{x}) \end{cases},$$

where $\zeta_x := (x - \psi(x|R))(1 - R(x^-))$ for any $x \in [r, 1]$.

Notice that by (A.7), (A.8), (A.10) and (A.11), for any $n > \bar{n}$, $x_n = \bar{x}(\underline{x}_n)$ and $\hat{x}_n = \hat{x}(\underline{x}_n)$. Also, $\hat{x}$ and $\bar{x}$ are decreasing in $\underline{x}$ and $\lim_{\underline{x}\to r} \bar{x}(\underline{x}) = \lim_{\underline{x}\to r} \hat{x}(\underline{x}) = x_0$, $\lim_{\underline{x}\to r} \psi(\bar{x}(\underline{x})|R) = \lim_{n\to\infty} \psi(x_n|R) = \psi(x_0^-|R)$.

Since $\psi(\cdot|R)$ and γ are nondecreasing, by (A.10) and (A.11), $\bar{x}$ and $\hat{x}$ are differentiable Lebesgue-almost everywhere and therefore Ψ is differentiable Lebesgue-almost everywhere. Thus, for Lebesgue almost all $\underline{x}$,

$$\begin{aligned} \Psi'(\underline{x}) &= (\gamma(k) - \gamma(\psi(\underline{x}|R)))(1 - R(\underline{x})) - \hat{x}'(\underline{x})(\gamma(k) - \gamma(\psi(\bar{x}(\underline{x})|R)))(1 - \widehat{R}^{\underline{x}}(\hat{x}(\underline{x}))) \\ &\quad - \int_{\bar{x}(\underline{x})}^{\hat{x}(\underline{x})} \frac{\partial}{\partial \underline{x}}(\gamma(k) - \gamma(\psi(\bar{x}(\underline{x})|R)))(1 - \widehat{R}^{\underline{x}}(x))\,\mathrm{d}x. \end{aligned}$$

Since $\lim_{\underline{x}\to r} \hat{x}(\underline{x}) = \lim_{\underline{x}\to r} \bar{x}(\underline{x}) = x_0$, $\lim_{\underline{x}\to r} \psi(\bar{x}(\underline{x})|R) = k$ and since $\hat{x}$ is decreasing, there exists $\delta > 0$ such that for $\underline{x}$ sufficiently close to r, whenever Ψ is differentiable at $\underline{x}$,

$$\Psi'(\underline{x}) > (\gamma(k) - \gamma(\psi(\underline{x}|R)))(1 - R(\underline{x})) - \delta > 0, \tag{A.12}$$

where the second inequality follows from the hypothesis that γ is strictly increasing. Therefore, since Ψ is continuous in $\underline{x}$ and $\Psi'(\underline{x}) > 0$ for $\underline{x}$ sufficiently close to r, there exists $\hat{r}$ such that $\Psi(\underline{x}) > 0$ for all $\underline{x} \in (r, \hat{r})$.

As such, for n sufficiently large so that $\underline{x}_n \in (r, \hat{r})$,

$$\begin{aligned} 0 < \Psi(\underline{x}_n) &= \int_r^{\underline{x}_n} (\gamma(k) - \gamma(\psi(x|R)))(1 - R(x))\,\mathrm{d}x \\ &\quad - \int_{x_n}^{\hat{x}_n} \big((\gamma(k) - \gamma(\psi(x_n|R)))(1 - \widehat{R}^{x_n}(x))\big)\,\mathrm{d}x \\ &\quad + \int_{x_n}^{x_0} \big(\gamma(k) - \gamma(\psi(x|R)))(1 - R(x))\big)\,\mathrm{d}x. \end{aligned}$$

Together, for n sufficiently large,

$$\int_0^1 \gamma(\psi(x|\widehat{R}^{x_n}))(1 - \widehat{R}^{x_n}(x))\,\mathrm{d}x > \int_0^1 \gamma(\psi(x|R))(1 - R(x))\,\mathrm{d}x,$$

as desired.

If, on the other hand, for any $x \in (r,1)$ and for any sequence $\{x_n\}$ such that $\{x_n\} \uparrow x$, there exists $\bar{n} \in \mathbb{N}$ such that $\psi(x_n|R) = \psi(x^-|R)$ for all $n > \bar{n}$, then for any $x \in (r,1)$, there exists $\delta > 0$ such that $\psi(y|R) = \psi(x^-|R)$ for all $y \in (x-\delta, x)$. Let $\delta_x := \sup\{\delta > 0 | \psi(y|R) = \psi(x^-|R), \forall y \in (x-\delta, x)\}$. Then $\delta_x > 0$ for all $x \in (r,1)$. Moreover, for any $x, y \in (r,1)$, if $\psi(x^-|R) \neq \psi(y^-|R)$, then it must be that $(x-\delta_x, x) \cap (y-\delta_y, y) = \emptyset$. Therefore, $\{\psi(x^-|R)\}_{x\in(r,1)}$ is at most countable. Since $\psi(\cdot|R)^+$ is nondecreasing, it must be a step function. Consider first the case when for any $\delta > 0$, there exists $x, y \in (b-\delta, b)$ with $x < y$ such that $\psi(x|R) < \psi(y|R) < \psi(b^-|R)$. Since $\psi(\cdot|R)^+$ is a step function, $\psi(\cdot|R)^+$ has countably infinitely many jumps and therefore we may represent $\psi(\cdot|R)^+$ as

$$\psi(x|R) = \sum_{n=1}^{\infty} \psi_n \mathbf{1}\{x \in (\alpha_n, \beta_n)\}, \, \forall x \in [r,1] \backslash [\{\alpha_n\}_{n=1}^{\infty} \cup \{\beta_n\}_{n=1}^{\infty}].$$

for some $\{\alpha_n\}$, $\{\beta_n\}$ such that $\alpha_n < \beta_n$ for all $n \in \mathbb{N}$ and some $\{\psi_n\}_{n=1}^{\infty}$ such that $\psi_n > \psi(r^+|R)$ for all $n \in \mathbb{N}$, $n \geq 2$. Since for any $\delta > 0$, there exists $x, y \in (b-\delta, b)$, $x < y$, such that $\psi(x|R) < \psi(y|R) < \psi(b^-|R)$, there exists a sequence $\{x_j\}$ such that $x_j \in \{\alpha_n\}_{n=1}^{\infty} \cup \{\beta_n\}_{n=1}^{\infty}$, $x_j < x_{j+1}$, $\psi_j := \psi(x_j^+|R) = \psi(x_{j+1}^-|R) < \psi(x_{j+1}^+|R) =: \psi_{j+1}$ for all $j \in \mathbb{N}$, and that $\{x_j\} \uparrow b$ and $\{\psi_j\} \uparrow \psi(b^-|R)$.

Since $\psi(\cdot|R)^+$ is a step function and $\psi(x|R) = b$ for all $x \in [b,1]$, we must have $\psi(b^-|R) < \psi(b|R) = b$ and $R(b^-) < 1$. As such, let $\zeta_b := (b - \psi(b^-|R))(1 - R(b^-))$. Then $\zeta_b > 0$. Also, for each $j \in \mathbb{N}$, let $\zeta_j := (x_j - \psi_j)(1 - R(x_j^-))$, by Lemma 1, since $\psi_{j-1} < \psi_j < \psi_{j+1}$, we have

$$R(x) > \max\left\{1 - \frac{\zeta_b}{x - \psi(b^-|R)}, \, 1 - \frac{\zeta_j}{x - \psi_j}\right\},$$

for all $x \in (x_j, b)$.

As such, there exists a sequence $\{r_j\}$ such that $\{r_j\} \downarrow r$ and $\widehat{R}^j \in \mathcal{I}_\mu$ for j large enough, where

$$\widehat{R}^j(x) := \begin{cases} R(x), & \text{if } x \in [b,1] \\ 1 - \frac{\zeta_b}{x - \psi(b^-|R)}, & \text{if } x \in [\hat{x}_j, b) \\ 1 - \frac{\zeta_j}{x - \psi_j}, & \text{if } x \in [\bar{x}_j, \hat{x}_j) \\ R(x), & \text{if } x \in [r_j, \bar{x}_j) \\ R(r_j), & \text{if } x \in [0, r_j) \end{cases}.$$

and $x_j < \hat{x}_j < \bar{x}_j < b$ are uniquely defined by

$$\int_0^b \widehat{R}^j(x) \, \mathrm{d}x = \int_0^b R(x) \, \mathrm{d}x$$

$$1 - \frac{\zeta_b}{\hat{x}_j - \psi(b^-|R)} = 1 - \frac{\zeta_j}{\hat{x}_j - \psi_j},$$

Similar to the previous case, for each $j \in \mathbb{N}$ such that $\widehat{R}^j \in \mathcal{I}_\mu$, the deviation gain from R to $\widehat{R}^j$ is

$$\gamma(\psi(b^-|R))(r(1-R(r)) - r_j(1-R(r_j))) + \int_r^{r_j} (\gamma(\psi(b^-|R)) - \gamma(\psi(x|R)))(1-R(x))\,\mathrm{d}x$$
$$- \int_{\bar{x}_j}^{\hat{x}_j} \big(\gamma(\psi(b^-|R)) - \gamma(\psi_j)\big)(1-\widehat{R}^j(x))\,\mathrm{d}x$$
$$+ \int_{\bar{x}_j}^{x_j} \big(\gamma(\psi(b^-|R)) - \gamma(\psi(x|R))\big)(1-R(x))\,\mathrm{d}x.$$

As noted above, from Lemma 1, $r(1-R(r)) = r(1-R(r^-)) \geq r_j(1-R(r_j))$ for all $j \in \mathbb{N}$. Also, as shown in the previous case, from (A.9) and (A.12), with $x_0 = b$ and $k = \psi(b^-|R)$, there exists $\delta > 0$ such that for $\underline{x}$ sufficiently close to r, $\Psi'(\underline{x}) > \gamma(\psi(b^-|R)) - \gamma(\psi(r^+|R)) - \delta > 0$ since $\psi_j > \psi(r^+|R)$ for all $j \in \mathbb{N}$ and γ is strictly increasing. Thus, since by (A.10) and (A.11), $\hat{x}_j = \hat{x}(r_j)$ and $\bar{x}_j = \bar{x}(r_j)$, and since $\{r_j\} \downarrow r$, for j sufficiently large,

$$\begin{aligned} 0 < \Psi(r_j) = &\int_r^{r_j} (\gamma(\psi(b^-|R)) - \gamma(\psi(x|R)))(1-R(x))\,\mathrm{d}x \\ &- \int_{\bar{x}_j}^{\hat{x}_j} \big((\gamma(\psi(b^-|R)) - \gamma(\psi_j)\big)(1-\widehat{R}^j(x))\,\mathrm{d}x \\ &+ \int_{\bar{x}_j}^{x_j} \big(\gamma(\psi(b^-|R)) - \gamma(\psi(x|R)))\big)(1-R(x))\,\mathrm{d}x. \end{aligned}$$

Together, for j large enough, $\widehat{R}^j \in \mathcal{I}_\mu$ and

$$\int_0^1 \gamma(\psi(x|\widehat{R}^j))(1-\widehat{R}^j(x))\,\mathrm{d}x > \int_0^1 \gamma(\psi(x|R))(1-R(x))\,\mathrm{d}x,$$

as desired.

Finally, if there exists $\delta > 0$ such that for any $x, y \in (b-\delta, b)$, $\psi(x|R) = \psi(y|R) = \psi(b^-|R)$. Let $\underline{b} := \inf\{b_0 \in [r,1) | \psi(x|R) = \psi(y|R), \forall x, y \in (b_0, b)\}$ and let $\underline{\psi} := \psi(b^-|R)$. Then $\underline{b} < b$ and $\underline{\psi} > \psi(r^+|R)$ since $\psi(\cdot|R)$ is not a constant on $[r, b)$. We claim that it is without loss to suppose that $\phi(x|R) = H(x|R)$ for all $x \in (\underline{b}, b)$. Indeed, if there exists $x \in (\underline{b}, b)$ such that $\phi(x|R) < H(x|R)$, let $\zeta := (\underline{b} - \underline{\psi})(1 - R(\underline{b}^-))$. Lemma 1 then ensures that

$$R(x) \geq 1 - \frac{\zeta}{x - \underline{\psi}},$$

for all $x \in (\underline{b}, b)$ with strict inequality for some $x \in (\underline{b}, b)$ and therefore, there exists $\hat{b} \in (\underline{b}, b)$ such that

$$\int_{\underline{b}}^{b} (1-R(x))\,\mathrm{d}x = \int_{\underline{b}}^{\hat{b}} \frac{\zeta}{x - \underline{\psi}}\,\mathrm{d}x.$$

Therefore, for

$$\widehat{R}(x) := \begin{cases} R(x), & \text{if } x \in [0, \underline{b}) \\ 1 - \frac{\zeta}{x - \underline{\psi}}, & \text{if } x \in [\underline{b}, \hat{b}) \\ 1, & \text{if } x \in [\hat{b}, 1] \end{cases},$$

$\widehat{R} \in \mathcal{I}_\mu$ and

$$\int_0^1 \gamma(\psi(x|R))(1 - R(x))\,\mathrm{d}x = \int_0^1 \gamma(\psi(x|\widehat{R}))(1 - \widehat{R}(x))\,\mathrm{d}x$$

and $H(x|\widehat{R}) = \phi(x|\widehat{R})$ for all $x \in (\underline{b}, b)$.

Therefore, as $H(x|R) = \phi(x|R)$ for all $x \in (\underline{b}, b)$ and $\psi(x|R) = \underline{\psi}$ for all $x \in (\underline{b}, b)$, we must have

$$R(x) = 1 - \frac{\zeta}{x - \underline{\psi}}, \; \forall x \in [\underline{b}, b),$$

for some $\zeta > 0$. Now take and fix any $\bar{x} \in (\underline{b}, b)$ and notice that for any $h \in (\psi(\underline{b}^-|R), \underline{\psi})$ and any $k \in (\underline{\psi}, b)$, let $\zeta(k) := (\bar{x} - k)(1 - R(\bar{x}^-))$ and $\zeta(h) := (\underline{b} - h)(1 - R(\underline{b}^-))$, we must have

$$R(x) > \max\left\{1 - \frac{\zeta(k)}{x - k}, 1 - \frac{\zeta(h)}{x - h}\right\},$$

for all $x \in (\underline{b}, \bar{x})$. Thus, for any $\underline{x} > r$ that is close enough to r, there exists $h(\underline{x}) \in (\psi(\underline{b}^-|R), \underline{\psi})$,$k(\underline{x}) \in (\underline{\psi}, b)$ and $\hat{x}(\underline{x}) \in (\underline{b}, b)$ such that $\lim_{\underline{x} \downarrow r} h(\underline{x}) = \lim_{\underline{x} \downarrow r} k(\underline{x}) = \underline{\psi}$,

$$\frac{\zeta(k(\underline{x}))}{\hat{x}(\underline{x}) - k(\underline{x})} = \frac{\zeta(h(\underline{x}))}{\hat{x}(\underline{x}) - h(\underline{x})}$$

and

$$\int_0^{\bar{x}} R(x)\,\mathrm{d}x = \int_0^{\bar{x}} \widehat{R}^{\underline{x}}(x)\,\mathrm{d}x,$$

where

$$\widehat{R}^{\underline{x}}(x) := \begin{cases} R(x), & \text{if } x \in [\bar{x}, 1] \\ 1 - \frac{\zeta(k(\underline{x}))}{x - k(\underline{x})}, & \text{if } x \in [\hat{x}(\underline{x}), \bar{x}) \\ 1 - \frac{\zeta(h(\underline{x}))}{x - h(\underline{x})}, & \text{if } x \in [\underline{b}, \hat{x}(\underline{x})) \\ R(x), & \text{if } x \in [\underline{x}, \underline{b}) \\ R(\underline{x}), & \text{if } x \in [0, \underline{x}) \end{cases}$$

and thus $\widehat{R}^{\underline{x}} \in \mathcal{I}_\mu$. Moreover, such $h(\cdot)$ and $k(\cdot)$ can be selected so that $\hat{x}(\underline{x})$ is decreasing in $\underline{x}$ and $\lim_{\underline{x} \downarrow r} \hat{x}(\underline{x}) = \bar{x}$.

Notice that for any such $\widehat{R}^{\underline{x}}$,

$$\psi(x|\widehat{R}^{\underline{x}}) = \begin{cases} \underline{\psi}, & \text{if } x \in [0, \underline{x}) \\ \psi(x|R), & \text{if } x \in [\underline{x}, \underline{b}) \\ h(\underline{x}), & \text{if } x \in [\underline{b}, \hat{x}(\underline{x})) \\ \tilde{k}(\underline{x}), & \text{if } x \in [\hat{x}(\underline{x}), b) \\ b, & \text{if } x \in [b, 1]. \end{cases},$$

for some $\underline{\psi} < 0$ and some $\tilde{k}(\underline{x}) \in (\underline{\psi}, k(\underline{x}))$. As such, for any $\underline{x}$ such that $\widehat{R}^{\underline{x}} \in \mathcal{I}_\mu$, the deviation gain from R to $\widehat{R}^{\underline{x}}$ is

$$
\begin{aligned}
&\int_0^1 \gamma(\psi(x|\widehat{R}^{\underline{x}}))(1-\widehat{R}^{\underline{x}}(x))\,\mathrm{d}x - \int_0^1 \gamma(\psi(x|R))(1-R(x))\,\mathrm{d}x \\
&>\gamma(\tilde{k}(\underline{x}))(r(1-R(r)) - \underline{x}(1-R(\underline{x}))) + \int_r^{\underline{x}} (\gamma(\tilde{k}(\underline{x})) - \gamma(\psi(x|R)))(1-R(x))\,\mathrm{d}x \\
&-\int_{\underline{b}}^{\hat{x}(\underline{x})} (\gamma(\tilde{k}(\underline{x})) - \gamma(h(\underline{x})))(1-\widehat{R}^{\underline{x}}(x))\,\mathrm{d}x,
\end{aligned}
$$

where the stritc inequality follows from that $\tilde{k}(\underline{x}) > \underline{\psi}$ and that γ is strictly increasing.

Again, by Lemma 1, $r(1-R(r)) \geq \underline{x}(1-R(\underline{x}))$ for all $\underline{x} \geq r$. Also, let

$$
\Psi(\underline{x}) := \int_r^{\underline{x}} (\gamma(\tilde{k}(\underline{x})) - \gamma(\psi(x|R)))(1-R(x))\,\mathrm{d}x - \int_{\underline{b}}^{\hat{x}(\underline{x})} (\gamma(\tilde{k}(\underline{x})) - \gamma(h(\underline{x})))(1-\widehat{R}^{\underline{x}}(x))\,\mathrm{d}x.
$$

Then Ψ is differentiable Lebesgue-almost everywhere and the derivative converges to

$$
(\gamma(\psi(\underline{b}^-|R)) - \gamma(\psi(r^+|R)))(1-R(r)) - \lim_{k,h\to\underline{\psi},\, k>h} (\gamma(k)-\gamma(h)) \cdot \lim_{\underline{x}\downarrow r} \hat{x}'(\underline{x}) \cdot (1-R(\bar{x})) > 0
$$

as $\underline{x}$ approaches r, since $\hat{x}(\underline{x})$ is decreasing in $\underline{x}$, $\psi(b^-|R) > \psi(r^+|R)$ and γ is strictly increasing. Thus, $\Psi(\underline{x}) > 0$ for $\underline{x}$ sufficiently close to r.

Together, for $\underline{x}$ close enough to r, there exists $\widehat{R}^{\underline{x}} \in \mathcal{I}_\mu$ such that

$$
\int_0^1 \gamma(\psi(x|\widehat{R}^{\underline{x}}))(1-\widehat{R}^{\underline{x}}(x))\,\mathrm{d}x > \int_0^1 \gamma(\psi(x|R))(1-R(x))\,\mathrm{d}x,
$$

as desired.

Case 3: $R(r^-) = 0$

Notice that the arguments in **case 2** rely only on the observations that $R(r) = R(r^-) = R(a)$ and $aR(r) = 0$. As such, if $R(r) = R(r^-) = 0$, this is exactly **case 2**. On the other hand, if $R(r) > R(r^-) = 0$, we may define

$$
G(x) := \begin{cases} R(r), & \text{if } x \in [0, r) \\ R(x), & \text{if } x \in [r, 1]. \end{cases} .
$$

Then since the the arguments in **case 2** do not depend on particular value of $\mu = \int_0^1 (1 - R(x))\,\mathrm{d}x$, by the same arguments, there exists $\widehat{G} \in \mathcal{I}_{\bar{\mu}}$ such that

$$
\int_0^1 \gamma(\psi(x|\widehat{G}))(1-\widehat{G}(x))\,\mathrm{d}x > \int_0^1 \gamma(\psi(x|G))(1-G(x))\,\mathrm{d}x,
$$

and that $G(\underline{x}) = \widehat{G}(\underline{x}) = \widehat{G}(0)$ for some $\underline{x} \in (r, 1)$, where $\bar{\mu} := \int_0^1 (1 - G(x))\,\mathrm{d}x$. Since $\underline{x} > r$ and $G(\underline{x}) \geq G(r) = R(r)$, there exists $\epsilon > 0$ such that $\underline{x}(\widehat{G}(\underline{x}) - \epsilon) = rR(r)$ and therefore $\widehat{R} \in \mathcal{I}_\mu$, where

$$\widehat{R}(x) := \begin{cases} \widehat{G}(\underline{x}) - \epsilon, & \text{if } x \in [0, \underline{x}) \\ \widehat{G}(x), & \text{if } x \in [\underline{x}, 1] \end{cases}.$$

Furthermore, by construction, $\psi(x|\widehat{G}) = \psi(x|\widehat{R})$ for all $x \in (\underline{x}, 1]$, $\psi(x|G) = \psi(x|R)$ for all $x \in (r, 1]$, $\psi(x|\widehat{G}) < 0$ if and only if $\psi(x|\widehat{R}) < 0$ and $\psi(x|G) < 0$ if and only if $\psi(x|R) < 0$. Together, we have

$$\begin{aligned} \int_0^1 \gamma(\psi(x|R))(1 - R(x))\,\mathrm{d}x &= \int_0^1 \gamma(\psi(x|G))(1 - G(x))\,\mathrm{d}x \\ &< \int_0^1 \gamma(\psi(x|\widehat{G}))(1 - \widehat{G}(x))\,\mathrm{d}x \\ &= \int_0^1 \gamma(\psi(x|\widehat{R}))(1 - \widehat{R}(x))\,\mathrm{d}x \end{aligned}$$

for some $\widehat{R} \in \mathcal{I}_\mu$, as desired. ■

A.2 Proof of Lemma 3.

Proof of Lemma 3. Consider any sequence $\{G_n\} \subset \mathcal{I}_\mu$ such that $\{G_n\} \to G$ under the weak-* topology. Suppose that G is continuous and suppose that, by way of contradiction,

$$\limsup_{n\to\infty} \int_0^1 \gamma(\psi(x|G_n))(1 - G_n(x))\,\mathrm{d}x > \int_0^1 \gamma(\psi(x|G))(1 - G(x))\,\mathrm{d}x.$$

Take a subsequence $\{G_{n_k}\} \subset \{G_n\}$ such that

$$\lim_{k\to\infty} \int_0^1 \gamma(\psi(x|G_{n_k}))(1 - G_{n_k}(x)) > \int_0^1 \gamma(\psi(x|G))(1 - G(x))\,\mathrm{d}x.$$

For any $R \in \mathcal{I}_\mu$, let

$$\hat{\psi}(x|R) := \begin{cases} \psi(x|R), & \text{if } \psi(x|R) \geq 0 \\ -1, & \text{if } \psi(x|R) < 0. \end{cases}.$$

Then $\{\hat{\psi}(\cdot|G_{n_k})\}$ is a sequence of uniformly bounded and nondecreasing functions. Helley's selection theorem ensures that there exists a further subsequence $\{\psi_{n_{kl}}\} \subset \{\hat{\psi}(\cdot|G_{n_k})\}$ such that $\{\psi_{n_{kl}}\} \to \bar{\psi}$ pointwisely for some $\bar{\psi} : [0, 1] \to [-1, 1]$.

Claim 1. *There exists a countable set $\mathbb{C}$ such that $\bar{\psi}(x) = \hat{\psi}(x|G)$ for all $x \in \mathrm{supp}(G) \backslash \mathbb{C}$.*

Let $\mu_G \in \Delta([0, 1])$ be the probability measure associated with G, since G is continuous, $\mu_G(\mathbb{C}) = 0$ as it is atomless. Furthermore, by Claim 1, since γ is continuous on $(0, 1)$,

$\{\gamma(\psi_{n_{kl}}(x))(x-\psi_{n_{kl}}(x))\} \to \gamma(\hat{\psi}(x|G))(x-\hat{\psi}(x|G)) = \gamma(\psi(x|G))(x-\psi(x|G))$ pointwisely μ_G-almost everywhere on $\operatorname{supp}(G)$. Thus, for any $\varepsilon > 0$, by Egroff's theorem, there exists a set $A \subset \operatorname{supp}(G)$ such that $\mu_G([0,1]\backslash A) < \varepsilon$ and $\{\gamma(\psi_{n_{kl}}(x))(x-\psi_{n_{kl}}(x))\} \to \gamma(\psi(x|G))(x-\psi(x|G))$ uniformly on A. Also, since G is continuous, $\bar{\psi}$ and G do not share common jumps, by the Portmanteau lemma,

$$\lim_{l\to\infty} \int_0^1 \gamma(\bar{\psi}(x))(x-\bar{\psi}(x))G_{n_{kl}}(\mathrm{d}x) = \int_0^1 \gamma(\bar{\psi}(x))(x-\bar{\psi}(x))G(\mathrm{d}x) = \int_0^1 \gamma(\psi(x|G))(x-\psi(x|G))G(\mathrm{d}x).$$

Together, for any $\varepsilon > 0$,

$$\begin{aligned}
&\lim_{l\to\infty} \int_0^1 \gamma(\psi_{n_{kl}}(x))(1-G_{n_{kl}}(x))\,\mathrm{d}x \\
=&\lim_{l\to\infty} \int_0^1 \gamma(\psi_{n_{kl}}(x))(x-\psi_{n_{kl}}(x))G_{n_{kl}}(\mathrm{d}x) \\
=&\lim_{l\to\infty} \left[\int_0^1 \gamma(\bar{\psi}(x))(x-\bar{\psi}(x))G_{n_{kl}}(\mathrm{d}x) + \int_0^1 [\gamma(\psi_{n_{kl}}(x))(x-\psi_{n_{kl}}(x)) - \gamma(\bar{\psi}(x))(x-\bar{\psi}(x))]G_{n_{kl}}(\mathrm{d}x)\right] \\
\leq&\int_0^1 \gamma(\psi(x|G))(x-\psi(x|G))G(\mathrm{d}x) + \lim_{l\to\infty} \mu_{G_{n_{kl}}}([0,1]\backslash A) \\
&+ \lim_{l\to\infty} \sup_{x\in A} |\gamma(\psi_{n_{kl}}(x))(x-\psi_{n_{kl}}(x)) - \gamma(\bar{\psi}(x))(x-\bar{\psi}(x))| \\
<&\int_0^1 \gamma(\psi(x|G))(1-G(x))\,\mathrm{d}x + \varepsilon.
\end{aligned}$$

However, since $\{\int_0^1 \gamma(\psi_{n_{kl}}(x)))(1-G_{n_{kl}}(x))\,\mathrm{d}x\}$ is a subsequence of $\{\int_0^1 \gamma(\psi_{n_k}(x)))(1-G_{n_k}(x))\,\mathrm{d}x\}$,

$$\lim_{l\to\infty} \int_0^1 \gamma(\psi_{n_{kl}}(x)))(1-G_{n_{kl}}(x))\,\mathrm{d}x \leq \int_0^1 \gamma(\psi(x|G))(1-G(x))\,\mathrm{d}x,$$

a contradiction. Thus, the functional is upper-semicontinuous at continuous distributions.

Now suppose that $\{G_n\} \to G$ under the weak-* topology and G has finitely many jumps.

Claim 2. *There exists a sequence $\{G_m\}$ such that $\{G_m\} \to G$ under the weak-* topology, G_m is continuous for all $m \in \mathbb{N}$ and*

$$\limsup_{m\to\infty} \int_0^1 \gamma(\psi(x|G_m))(1-G_m(x))\,\mathrm{d}x \leq \int_0^1 \gamma(\psi(x|G))(1-G(x))\,\mathrm{d}x.$$

For each $m \in \mathbb{N}$, since G_m is continuous and thus the functional is upper-semicontinuous at G_m, for any $\varepsilon > 0$, since the weak-* topology on $\Delta([0,1])$ is metrizable, there exists

$\delta_m > 0$ such that for all $R \in N_{\delta_m}(G_m)$,[31]

$$\int_0^1 \gamma(\psi(x|R))(1 - R(x))\,\mathrm{d}x < \int_0^1 \gamma(\psi(x|G_m))(1 - G_m(x))\,\mathrm{d}x + \frac{\varepsilon}{2}.$$

On the other hand, by Claim 2, for the same $\varepsilon > 0$, there exists $M \in \mathbb{N}$ such that for any $m > M$,

$$\int_0^1 \gamma(\psi(x|G_m))(1 - G_m(x))\,\mathrm{d}x < \int_0^1 \gamma(\psi(x|G))(1 - G(x))\,\mathrm{d}x + \frac{\varepsilon}{2}.$$

Since $\{G_m\} \to G$ under the weak-* topology and since $\Delta([0,1])$ is a compact metrizable topological space,

$$G \in \overline{\bigcup_{m=M+1}^{\infty} N_{\frac{\delta_m}{4}}(G_m)} = \bigcup_{m=M+1}^{\infty} \overline{N_{\frac{\delta_m}{4}}(G_m)}.$$

Thus, $G \in N_{\delta_m/2}(G_m)$ for some $m > M$. Take and fix any such m. Since $\{G_n\} \to G$ under the weak-* topology as well, there exists $N \in \mathbb{N}$ such that $G_n \in N_{\delta_m/2}(G)$ for all $n > N$ and thus $G_n \in N_{\delta_m}(G_m)$ for all $n > N$. Together, for any $n > N$,

$$\int_0^1 \gamma(\psi(x|G_n))(1-G_n(x))\,\mathrm{d}x < \int_0^1 \gamma(\psi(x|G_m))(1-G_m(x))\,\mathrm{d}x+\frac{\varepsilon}{2} < \int_0^1 \gamma(\psi(x|G))(1-G(x))\,\mathrm{d}x+\varepsilon.$$

Thus,

$$\limsup_{n\to\infty} \int_0^1 \gamma(\psi(x|G_n))(1 - G_n(x))\,\mathrm{d}x \leq \int_0^1 \gamma(\psi(x|G))(1 - G(x))\,\mathrm{d}x.$$

Therefore, the functional is upper-semicontinuous at the distributions with finitely many jumps as well.

Finally, suppose that $\{G_n\} \to G$ under the weak-* topology and that G has countably infinitely many jumps.

Claim 3. *There exists a sequence $\{G_m\}$ such that $\{G_m\} \to G$ under the weak-* topology, G_m has finitely many jumps for all $m \in \mathbb{N}$ and*

$$\limsup_{m\to\infty} \int_0^1 \gamma(\psi(x|G_m))(1 - G_m(x))\,\mathrm{d}x \leq \int_0^1 \gamma(\psi(x|G))(1 - G(x))\,\mathrm{d}x.$$

By Claim 3, by the conclusion above that the functional is upper-semicontinuous at the distributions that have finitely many jumps and by repeating the same argument as above, we have

$$\limsup_{n\to\infty} \int_0^1 \gamma(\psi(x|G_n))(1 - G_n(x))\,\mathrm{d}x \leq \int_0^1 \gamma(\psi(x|G))(1 - G(x))\,\mathrm{d}x.$$

[31] $N_\delta(G) := \{R \in \mathcal{I}_\mu | \rho(R, G) < \delta\}$ denotes the δ-ball around G, where ρ is the Lévy-Prokhorov metric.

This completes the proof. ■

A.3 Proof of Theorem 1

Proof of Theorem 1. Notice that $\mathcal{I}_\mu$ is compact under the weak-* topology by the Riesz representation theorem, Alaoglu's theorem and the dominated convergence theorem. Together with Lemma 3, optimal information structure always exists. That is,

$$\operatorname*{argmax}_{R\in\mathcal{I}_\mu} \int_0^1 \gamma(\psi(x|R))(1-R(x))\,\mathrm{d}x \neq \emptyset.$$

On the other hand, by Lemma 2, such optimal information structure must be a subset of $\mathcal{I}_\mu^*$. That is,

$$\operatorname*{argmax}_{R\in\mathcal{I}_\mu} \int_0^1 \gamma(\psi(x|R))(1-R(x))\,\mathrm{d}x \subseteq \mathcal{I}_\mu^*.$$

Together, the proof is complete. ■

B Other Omitted Proofs

B.1 Proof of Proposition 1

The proof of Proposition 1 involves a unified representation of the seller's profit under two different classes of technologies. To this end, consider first the uncertain cost technology. Let $\gamma \in \Gamma$ be the CDF of the seller's marginal cost. For any $c \in [0,1]$ and for any $G \in \mathcal{I}_F$, let

$$\pi(c|G) := \max_{p\in[0,1]} (p-c)(1-G(p^-))$$

be the seller's profit under information structure G when the realized marginal cost is c.

On the other hand, consider the increasing marginal cost technology. Given any continuous and convex cost function C, it is well-known (see Mussa and Rosen (1978) and Myerson (1981)) that the seller's profit under an information structure $G \in \mathcal{I}_\mu$ is

$$\int_0^1 [\gamma(\psi(x|G))\psi(x|G) - C(\gamma(\psi(x|G)))]G(\mathrm{d}x),$$

where $\gamma(z) \in \operatorname{argmax}_{q\in[0,1]}[qz - C(q)]$ for all $z \in [0,1]$. The following lemma, whose proof can be found in Yang (2019), shows that the seller's profit under an uncertain cost environment is as if it is the profit under a convex cost environment defined by the cost function associated with the CDF γ and vice versa.

Lemma 4. *For any $\gamma \in \Gamma$ and for any $G \in \mathcal{I}_F$, let*

$$C(q) := \int_0^q \gamma^{-1}(z)\,\mathrm{d}z,\ \forall q \in [0,1].$$

Then
$$\int_0^1 \pi(c|G)\gamma(\mathrm{d}c) = \int_0^1 [\gamma(\psi(x|G))\psi(x|G) - C(\gamma(\psi(x|G)))]G(\mathrm{d}x).$$
With Lemma 4, the proof of Proposition 1 then follows.

Proof of Proposition 1. For any technology $\gamma \in \Gamma$ such that γ is strictly increasing on $[0,1]$, let
$$C(q) := \int_0^q \gamma^{-1}(z)\,\mathrm{d}z.$$
Then, by Lemma 4, for any $G \in \mathcal{I}_\mu$, the sum of the buyer's surplus and the seller's profit is
$$\int_0^1 \gamma(\psi(x|G))(x - \psi(x|G))G(dx) + \int_0^1 \big[\gamma(\psi(x|G))\psi(x|G) - C(\gamma(\psi(x|G)))\big]G(dx).$$
Moreover, by Theorem 3 in Monteiro and Svaiter (2010),
$$\int_0^1 \gamma(\psi(x|G))(x - \psi(x|G))G(dx) + \int_0^1 \big[\gamma(\psi(x|G))\psi(x|G) - C(\gamma(\psi(x|G)))\big]G(dx)$$
$$= \int_0^1 \big[\gamma(\psi(x|G))x - C(\gamma(\psi(x|G)))\big]G(dx).$$
On the other hand, since γ is strictly increasing and thus C is strictly convex,
$$\operatorname*{argmax}_{q\in[0,1]}[qx - C(q)]$$
is unique for all $x \in [0,1]$. Therefore,
$$\gamma(\psi(x|G))x - C(\psi(x|G)) \leq V(x) := \max_{q\in[0,1]}[qx - C(q)],$$
for all $x \in [0,1]$, with strict inequality for all $x \in [0,1]$ such that $\psi(x|G) \neq x$. Furthermore, since $V'(z) = \gamma(z)$ for almost all $z \in [0,1]$ by the envelope theorem (Milgrom and Segal, 2002) and since $\gamma \geq 0$ and is strictly increasing, V is increasing and strictly convex. Together, since any $G \in \mathcal{I}_\mu$ is a mean-preserving contraction of the distribution F, we have
$$\int_0^1 V(x)G(\mathrm{d}x) \leq \int_0^1 V(x)F(\mathrm{d}x),$$
for any $G \in \mathcal{I}_\mu$, with strict inequality for $G \neq F$.

Together, it must be that for any $G \in \mathcal{I}_\mu$,
$$\int_0^1 \big[\gamma(\psi(x|G))x - C(\gamma(\psi(x|G)))\big]G(\mathrm{d}x) \leq \int_0^1 V(x)G(\mathrm{d}x) \leq \int_0^1 V(x)F(\mathrm{d}x),$$
with at least one inequality being strict since F has full support and hence $\psi(x|F) \neq x$ for all $x \in [0,1)$. Finally, notice that given any technology γ, the efficient surplus is exactly
$$\int_0^1 V(x)F(\mathrm{d}x).$$
This completes the proof. ∎

B.2 Proofs of Proposition 2 and Proposition 3

The proofs of Proposition 2 and Proposition 3 essentially rely on the follows two lemmas. Before stating the lemmas, define a crucial function $\omega : [0,1] \to [0,1]$ as the following. For any $c \in [0,1]$, let

$$\begin{aligned} \omega(c) := \min_{\pi,\eta,\lambda,b\in[0,1]^4} & [\pi + (1-\eta)c + \eta\lambda] \\ \text{s.t. } & \eta\lambda + \pi + (1-\eta)c + \pi \log\left(\frac{(b-c)(1-\eta)}{\pi}\right) = \mu \\ & \pi \log\left(\frac{b-c}{x-c}\right) \leq \int_x^1 (1-F(z))\,\mathrm{d}z,\ \forall x \in \left[\frac{\pi}{1-\eta}+c, b\right]. \\ & \eta(x-\lambda) \leq \int_0^x F(z)\,\mathrm{d}z,\ \forall x \in \left[\lambda, \frac{\pi}{1-\eta}+c\right] \\ & \lambda \leq \frac{\pi}{1-\eta} + c \leq b. \end{aligned} \tag{B.1}$$

By the theorem of maximum and the envelope theorem, it can be verified that ω is nondecreasing and continuous on $[0,1]$, differentiable on $(0,1)$, and $\omega(1) = \mu$.

Lemma 5. *For any $c \in [0,1]$,*

$$\sigma^*(\gamma^c) = \mu - \omega(c).$$

Proof. This immediately follows from rewriting the simplified buyer-surplus maximization problem characterized in Roesler and Szentes (2017),

$$\max_{G\in\mathcal{I}_F^{\mathrm{RS}}(c)} \int_0^1 \gamma^c(\psi(x|G))(1-G(x))\,\mathrm{d}x,$$

as choosing parameters π, η, λ, b to maximize the buyer's surplus subject to the constraint $G_{\pi,c}^{\eta,\lambda,b} \in \mathcal{I}_F$. ■

Lemma 6. *For any function $f : [0,1]\times[0,1]$ that is differentiable on $(0,1)\times(0,1)$. Suppose that $f(c,c) = 1$ for all $c \in [0,1]$ and that the indefinite integral of $\phi(c) := f_c(c,c)$, $\Phi(c)$, exists. Then there exists $c \in (0,1)$ such that*

$$\max_{c'\in[0,c]} f(c,c')(\mu - \omega(c')) > (\mu - \omega(c)).$$

Proof. Suppose that, by way of contradiction, for any $c \in [0,1]$,

$$\max_{c'\in[0,c]} f(c,c')(\mu - \omega(c')) \leq (\mu - \omega(c)),$$

Then, since $f(c,c) = 1$,

$$W(c) := \max_{c'\in[0,c]} f(c,c')(\mu - \omega(c')) = (\mu - \omega(c)),\ \forall c \in (0,1). \tag{B.2}$$

Since ω is differentiable, W is differentiable on $(0,1)$. Thus, by the envelope theorem (see Milgrom and Segal (2002), Lemma 1), for any $c \in (0,1)$

$$-\omega'(c) = W'(c) = f_c(c,c)(\mu - \omega(c)). \tag{B.3}$$

Let $\bar{\omega}(c) := \mu - \omega(c)$ for all $c \in [0,1]$, (B.3) can be written as

$$\bar{\omega}'(c) = f_c(c,c)\bar{\omega}(c).$$

Together with continuity of $\bar{\omega}$ and the fact that the indefinite integral of $\phi(c) := f_c(c,c)$ exists,

$$\bar{\omega}(c) = \kappa e^{\Phi(c)},\ \forall c \in [0,1]$$

for some $\kappa \in \mathbb{R}$. Furthermore, since $\omega(1) = \mu$, $\bar{\omega}(1) = 0$ and hence $\kappa = 0$, which then implies that $\omega(c) = \mu$ for all $c \in [0,1]$, a contradiction. ■

With Lemma 5 and Lemma 6, Proposition 2 and Proposition 3 can now be proved.

Proof of Proposition 2. By Lemma 6, when $f(c,c') = (1-c)/(1-c')$, there exists $0 \leq \tilde{c} < c < 1$ such that

$$\frac{1-c}{1-\tilde{c}}(\mu - \omega(\tilde{c})) > \mu - \omega(c). \tag{B.4}$$

Take and fix such $c, \tilde{c}$, let $\tilde{\gamma}$ be defined as

$$\tilde{\gamma}(z) := \begin{cases} 0, & \text{if } z \in [0,\tilde{c}) \\ \beta, & \text{if } z \in [\tilde{c},1) \\ 1, & \text{if } z = 1 \end{cases},$$

where $\beta := (1-c)/(1-\tilde{c})$. Then, by construction, $\tilde{\gamma} \in \Gamma$ and is a mean-preserving spread of γ^c. Furthermore, under the technology $\tilde{\gamma}$, the information structure $G_{\tilde{\pi},\tilde{c}}^{\tilde{\eta},\tilde{\lambda},\tilde{b}} \in \mathcal{I}_F$ yields surplus

$$\int_0^1 \tilde{\gamma}(\psi(x|G_{\tilde{\pi},\tilde{c}}^{\tilde{\eta},\tilde{\lambda},\tilde{b}}))(1 - G_{\tilde{\pi},\tilde{c}}^{\tilde{\eta},\tilde{\lambda},\tilde{b}}(x))\,\mathrm{d}x = \tilde{\gamma}(\tilde{c})(\mu - \omega(\tilde{c})),$$

where $(\tilde{\pi},\tilde{\eta},\tilde{\lambda},\tilde{b})$ is a solution of (B.1). Finally, by the definition of $\tilde{\gamma}$ and by (B.4) and Lemma 5,

$$\sigma^*(\tilde{\gamma}) \geq \tilde{\gamma}(\tilde{c})(\mu - \omega(\tilde{c})) = \frac{1-c}{1-\tilde{c}}(\mu - \omega(\tilde{c})) > (\mu - \omega(c)) = \sigma^*(\gamma^c),$$

as desired. ■

Proof of Proposition 3. By Lemma 6, when $f(c,c') = \int_c^1 (1-F(z))\,\mathrm{d}z / \int_{c'}^1 (1-F(z))\,\mathrm{d}z$, there exists $0 \le \tilde{c} < c < 1$ such that

$$\frac{\int_c^1 (1-F(z))\,\mathrm{d}z}{\int_{\tilde{c}}^1 (1-F(z))\,\mathrm{d}z}(\mu - \omega(\tilde{c})) > \mu - \omega(c). \tag{B.5}$$

Take and fix such $c, \tilde{c}$ and let $\tilde{C}$ be defined as

$$\tilde{C}(q) := \begin{cases} \tilde{c}q, & \text{if } q \in [0, q^*) \\ q - q^* + \tilde{c}q^*, & \text{if } q \in [q^*, 1] \end{cases},$$

where

$$q^* := \frac{\int_c^1 (1-F(z))\,\mathrm{d}z}{\int_{\tilde{c}}^1 (1-F(z))\,\mathrm{d}z}.$$

Then, let

$$\tilde{\gamma}(z) := \max \operatorname*{argmax}_{q \in [0,1]} [qz - \tilde{C}(q)],$$

we then have

$$\tilde{\gamma}(z) = \begin{cases} 0, & \text{if } z \in [0, \tilde{c}) \\ q^*, & \text{if } z \in [\tilde{c}, 1) \\ 1, & \text{if } z = 1 \end{cases}.$$

Furthermore, notice that for any $z \in [0,1]$

$$\max_{q \in [0,1]} [qz - \tilde{C}(q)] = \begin{cases} 0, & \text{if } z \in [0, \tilde{c}) \\ q^*(z - \tilde{c}), & \text{if } z \in [\tilde{c}, 1] \end{cases}.$$

Therefore,

$$\begin{aligned} S(\tilde{C}) :=& \int_0^1 \max_{q \in [0,1]} [qz - \tilde{C}(q)] F(\mathrm{d}z) \\ =& q^* \int_{\tilde{c}}^1 (z - \tilde{c}) F(\mathrm{d}z) \\ =& q^* \int_{\tilde{c}}^1 (1 - F(z))\,\mathrm{d}z \\ =& \int_c^1 (1 - F(z))\,\mathrm{d}z \\ =& \int_c^1 (z - c) F(\mathrm{d}z) \\ =& S(c). \end{aligned}$$

On the other hand, under the technology $\tilde{C}$ and the information structure $G^{\tilde{\eta},\tilde{\lambda},\tilde{b}}_{\tilde{\pi},\tilde{c}} \in \mathcal{I}_F$, where $(\tilde{\pi}, \tilde{\eta}, \tilde{\lambda}, \tilde{b})$ is a solution of (B.1),

$$\begin{aligned}\sigma^*(\tilde{\gamma}) \geq & \int_0^1 \tilde{\gamma}(\psi(x|G^{\tilde{\eta},\tilde{\lambda},\tilde{b}}_{\tilde{\pi},\tilde{c}}))(1 - G^{\tilde{\eta},\tilde{\lambda},\tilde{\lambda},\tilde{b}}_{\tilde{\pi},\tilde{c}}(x))\,\mathrm{d}x \\ =& q^*(\mu - \omega(\tilde{c})) \\ =& \frac{\int_c^1 (1-F(z))\,\mathrm{d}z}{\int_{\tilde{x}}^1 (1-F(z))\,\mathrm{d}z}(\mu - \omega(c)) \\ >& (\mu - \omega(c)) \qquad \text{(from (B.5))} \\ =& \sigma^*(\gamma^c),\end{aligned}$$

as desired ■

B.3 Proof of Theorem 3

Proof of Theorem 3. Consider any $c \in [0,1)$. Take any sequence of regular $\{\gamma_n\} \subseteq \Gamma$ such that $\gamma_n(0)$ for all $n \in \mathbb{N}$ and that $\{\gamma_n\} \to \gamma^c$ pointwisely. For each $n \in \mathbb{N}$, notice that if γ_n represents an increasing marginal cost technology, then the associated cost function, C_n, is strictly convex. Thus, γ_n is the unique element of

$$\operatorname*{argmax}_{q \in [0,1]}[qz - C_n(q)]$$

for all $z \in [0,1]$. Therefore the buyer's surplus is

$$\int_0^1 \gamma_n(\psi(x|G))(1-G(x))\,\mathrm{d}x.$$

On the other hand, if γ_n represents an uncertain cost technology, then since $\gamma_n(0) = 0$, γ_n is atomless. Thus, for any information structure $G \in \mathcal{I}_\mu$ and for any selection of optimal prices $p(\cdot, G)$, the induced buyer's surplus is unique and is

$$\begin{aligned}\int_0^1 \left(\int_{p(c,G)}^1 (x - p(c,G))G(\mathrm{d}x)\right)\gamma_n(\mathrm{d}c) =& \int_0^1 \left(\int_{p(c,G)}^1 (1-G(x))\,\mathrm{d}x\right)\gamma_n(\mathrm{d}c) \\ =& \int_0^1 \gamma_n(\psi(x|G))(1-G(x))\gamma_n(\mathrm{d}x).\end{aligned}$$

As a result, by Theorem 1 and (7), for any $n \in \mathbb{N}$, there is a unique outcome induced by the buyer-optimal information structure and the buyer's surplus is

$$\gamma_n(k_n)(\mu - \omega(k_n)),$$

where

$$k_n \in \operatorname*{argmax}_{k\in[0,1]} \gamma_n(k)(\mu - \omega(k)).$$

Since $\{\gamma_n\} \to \gamma^c$ pointwisely,

$$\liminf_{n\to\infty} \gamma_n(k_n)(\mu - \omega(k_n)) \geq \lim_{n\to\infty} \gamma_n(c)(\mu - \omega(c)) = (\mu - \omega(c)).$$

This then implies that, after possibly taking a subsequence, $\{k_n\} \to c$. Thus,

$$\limsup_{n\to\infty} \gamma_n(k_n)(\mu - \omega(k_n)) \leq \lim_{n\to\infty} (\mu - \omega(k_n)) = (\mu - \omega(c)).$$

Together, by Theorem 1,

$$\lim_{n\to\infty} \sigma^*(\gamma_n) = \lim_{n\to\infty} \gamma_n(k_n)(\mu - \omega(k_n)) = (\mu - \omega(c)) = \sigma^*(\gamma^c),$$

as desired. ■

B.4 Proof of Proposition 4

Proof of Proposition 4. To show that $(\nu^*, \dots, \nu^*)$ is an equilibrium, first notice Theorem 1 and (6) ensure that each buyer's best response against ν^* must be in the subset $\{G^{\eta(k)}_{\pi(k),k} | k \in [0,1]\}$. As such, it is sufficient to check if for all $k \in [0,1]$ such that $G^{\eta(k)}_{\pi(k),k} \in \text{supp}(\nu^*)$,

$$\gamma_{\nu^*}(k)(\mu - \omega(k)) \geq \gamma_{\nu^*}(k')(\mu - \omega(k')),\ \forall k' \in [0,1]$$

Indeed, by construction of ν^*, for any $k \in [0, \bar{k}]$,

$$\gamma_{\nu^*}(k)(\mu - \omega(k)) = \gamma_{\nu^*}(\bar{k})(\mu - \omega(\bar{k})) \geq \gamma_{\nu^*}(k')(\mu - \omega(k')),\ \forall k' \in [\bar{k}, 1],$$

where the inequality follows from the facts that $\gamma_{\nu^*}(k') = \gamma_{\nu^*}(\bar{k})$ for all $k' \in [\bar{k}, 1)$, that $\lim_{k\uparrow 1} \omega(k) = \mu$ and that ω is increasing. As such, since by construction, $\text{supp}(\nu^*) = \{G^{\eta(k)}_{\pi(k),k} | k \in [0, \bar{k}]\}$, $(\nu^*, \dots, \nu^*)$ is indeed a symmetric equilibrium. Finally, uniqueness is implied by Lemma 7 below, (6) and (10).[32] ■

Lemma 7. *Suppose that $(\nu, \dots, \nu)$ is a symmetric equilibrium. Then* $\text{supp}(\nu) \subseteq \mathcal{I}^*_\mu$.

Proof. See Supplemental Material. ■

[32]The proof of Lemma 7 is entirely analogous to the local perturbation argument used in the proof of Lemma 2 and hence is relegated to the Supplemental Material.

B.5 Proof of Proposition 5

Proof of Proposition 5. From Proposition 4, we know that the unique equilibrium strategy ν_N can be represented by a push-forward measure Λ_N, with a density given by

$$\lambda_N(k) := \mathbf{1}\{k \leq \bar{k}_N\} \frac{1}{N-1} (\bar{\nu}_N(\mu - \omega(\bar{k}_N)))^{\frac{1}{N-1}} \left(\frac{1}{\mu - \omega(k)} \right)^{\frac{N}{N-1}} \left(\frac{\omega'(k)}{1 - \frac{\pi(k)}{1-k} - \eta(k)} \right),$$

for all $k \in [0, 1]$, where $\omega(k)$ is the value of the minimization problem (6) and $(\pi(k), \eta(k))$ is the solution, $\bar{k}_N \in [0, 1)$ is the unique solution such that

$$\int_0^{\bar{k}_N} \frac{1}{N-1} (\mu - \omega(\bar{k}_N))^{\frac{1}{N-1}} \left(1 - \frac{\pi(k)}{1-k} \right) \left(\frac{1}{\mu - \omega(k)} \right)^{\frac{N}{N-1}} \left(\frac{\omega'(k)}{1 - \frac{\pi(k)}{1-k} - \eta(k)} \right) = 1. \quad \text{(B.6)}$$

and $\bar{\nu}_N$ is given by

$$\bar{\nu}_N = \left(\int_0^{\bar{k}_N} \frac{1}{N-1} (\mu - \omega(\bar{k}_N))^{\frac{1}{N-1}} \left(\frac{1}{\mu - \omega(k)} \right)^{\frac{N}{N-1}} \left(\frac{\omega'(k)}{1 - \frac{\pi(k)}{1-k} - \eta(k)} \right) \right)^{-(N-1)}.$$

Notice that since $\{\bar{\nu}_N\}$ is bounded, $\lim_{N \to \infty} \lambda_N(k) = 0$ for all $k \in [0, 1)$. Thus, by the dominated convergent theorem, for any $k \in [0, 1)$,

$$\lim_{N \to \infty} \Lambda_N(k) = \lim_{N \to \infty} \int_0^k \lambda_N(z) \, \mathrm{d}z = 0,$$

which establishes 1. Now observe that since $\{\lambda_N(k)\} \to 0$ for all $k \in [0, 1)$, from (B.6), it must be that $\{\bar{k}_N\} \to 1$ and therefore, for each buyer $i \in \{1, \ldots, N\}$, as $N \to \infty$, expected surplus is

$$\lim_{N \to \infty} \bar{\nu}_N(\mu - \omega(\bar{k}_N)) = 0,$$

since ω is continuous and $\omega(1) = \mu$. Furthermore, let

$$\pi_N := \int_0^1 \left(\int_{[\frac{\pi(k)}{1-\eta(k)} + k, 1)} \gamma_{\nu_N}(k) \psi \left(x \middle| G^{\eta(k)}_{\pi(k),k} \right) G^{\eta(k)}_{\pi(k),k}(\mathrm{d}x) + \gamma_{\nu_N}(1) \frac{\pi(k)}{1-k} \right) \Lambda_N(dk)$$

and let

$$\tau_N := \left(\int_{[\frac{\pi(k)}{1-\eta(k)} + k, 1)} \gamma_{\nu_N}(k) x G^{\eta(k)}_{\pi(k),k}(\mathrm{d}x) + \gamma_{\nu_N}(1) \frac{\pi(k)}{1-k} \right) \Lambda_N(dk)$$

be the seller's expected revenue from a single buyer and the surplus generated by a single buyer, respectively. Since the equilibrium is symmetric, $\bar{\pi}_N = N\pi_N$, $\bar{\tau}_N = N\tau_N$. Together,

$$1 = \lim_{N \to \infty} \frac{\bar{\sigma}_N + \bar{\pi}_N}{\bar{\tau}_N} = \lim_{N \to \infty} \frac{\bar{\nu}_N(\mu - \omega(\bar{k}_N)) + \pi_N}{\tau_N} = \lim_{N \to \infty} \frac{\pi_N}{\tau_N} = \lim_{N \to \infty} \frac{\bar{\pi}_N}{\bar{\tau}_N},$$

and hence $\lim_{N\to\infty}\bar{\sigma}_N = 0$. This establishes 2. Now it suffices to show that $\lim_{N\to\infty}\bar{\tau}_N = 1$. As shown in Monteiro (2015), the seller's revenue is lower semi-continuous in distributions of buyers' valuation and therefore is lower-semicontinuous under the weak-* topology on $\Delta(\mathcal{I}_\mu)^N$ in mixture of distributions by the Portmanteau lemma, for each $N \in \mathbb{N}$. Thus, as $\{\Lambda_N\} \to \delta_{\{1\}}$ under the weak-* topology and $\lim_{N\to\infty}\bar{\sigma}_N = 0$,

$$1 \geq \liminf_{N\to\infty} \bar{\tau}_N \geq 1.$$

As such, $\lim_{N\to\infty}\bar{\tau}_N = 1$, as desired. ■