Learning by Consuming: Optimal Pricing for a Divisible Good*

Huiyi Guo† Wei He‡ Bin Liu§

October 18, 2023

Abstract

We study the revenue-maximizing mechanism when a buyer's value evolves because of learning-by-consuming. The buyer chooses the initial consumption based on his rough valuation. Consuming more induces a finer valuation estimate, after which he determines the final consumption. The seller faces the tradeoff that selling more initially makes selling the rest more profitable but on a smaller base. The optimum is a try-and-decide contract. In equilibrium, a higher first-stage valuation buyer chooses more initial consumption and enjoys a lower second-stage per-unit price. Methodologically, we address the difficulty that without the single-crossing condition, monotonicity plus envelope condition is insufficient for incentive compatibility. Our results help to understand contracts with learning features, *e.g.*, course packages with included sessions and leasing agreements for experience goods.

JEL Classification Numbers: D44, D82, D86

Keywords: Dynamic mechanism design, Learning by consuming, Endogenous type distribution, Quantity effect, Experience good.

---

*We are grateful for helpful discussions with V. Bhaskar, Tilman Börgers, Ilwoo Hwang, Silvana Krasteva, R. Vijay Krishna, Rohit Lamba, Jingfeng Lu, Sergei Severinov, Vasiliki Skreta, Roland Strausz, Siyang Xiong, Jidong Zhou, as well as audiences at Texas A&M University, Univerity of Texas at Austin, University of California Berkeley, University of California Riverside, Southern Methodist University, University of British Columbia, Wuhan University, Sun Yat-sen University, 2022 Decentralization Conference, 2022 ES North American Summer Meeting, 2022 Conference on Mechanism and Institution Design, 21st SAET Conference, and 2022 Southern Economic Association Annual Meeting. He and Liu acknowledge financial support from the Research Grants Council of Hong Kong (14503318) and the National Natural Science Foundation of China (72192805 and 72073115), respectively.

†Huiyi Guo: Department of Economics, Texas A&M University, 4228 TAMU, College Station, TX 77843, United States of America. Email: huiyiguo@tamu.edu.

‡Wei He: Department of Economics, The Chinese University of Hong Kong, Hong Kong, China. Email: hewei@cuhk.edu.hk.

§Bin Liu: School of Management and Economics and Shenzhen Finance Institute, The Chinese University of Hong Kong, Shenzhen (CUHK-Shenzhen), China 518172. Email: binliu@cuhk.edu.cn.

1 Introduction

Situations are abundant in which a consumer is uncertain about how well a good's characteristics fit him at the outset, but by consuming (a portion of) the good, he could learn additional information to refine the value estimation. With such a more precise value estimate, the consumer then decides how many further units of the good to consume. Such kind of *learning by consuming* is widely observed in practice.

For example, when purchasing certain courses with fixed terms — a one-month package of fitness classes from a gym, a two-month playgroup for pre-school toddlers, or a summer sports course for children — the consumer often prefers to experience a few sessions first. After paying a fee and attending a few included sessions, which can be viewed as trial sessions, the consumer refines his valuation and decides whether to register for additional (or the remaining) sessions or not. In practice, the seller often sets a price for the included sessions and another price for additional sessions, and typically both prices depend on the number/length of included sessions. It is commonly observed that the seller offers different pricing packages to the consumer, who then decides which package to choose. For example, Orangetheory Fitness, a popular fitness chain with more than a million members in the U.S., offers three different monthly membership packages for consumers: basic, elite, and premier. These options differ with each other mainly in terms of the number of included sessions and fee structures. For instance, in Los Angeles, the basic membership charges 99 dollars and offers 4 classes, with discounted add-on classes; the elite membership charges 149 dollars and offers 8 classes, with discounted add-on classes; and the premier membership charges 209 dollars and includes all sessions.[1] Car leasing, as another example, also exhibits similar features.[2]

There are two important features in the examples above. First, the seller can choose to first sell a portion of the good to the consumer, through which the consumer better understands how well the good fits him and then decides the subsequent consumption. Second, since the consumer's learning is achieved through consuming, there is a tension between information acquisition and future consumption. Naturally, consuming more in the beginning would lead to a more precise value estimate, so that he could make a

[1]https://www.orangetheory.com/en-us/locations/california/southpark-la/1120-south-grand-street-suite-3 (accessed August 18, 2023).

[2]Car dealers in the U.S. usually provide a menu of contracts to potential consumers. Some consumers may choose to enter leasing contracts, which give them both the right to drive the car during the lease term and a lease-end option to buy out the car. The cost of the leasing contract and the buyout price depend on the length of the lease term, which usually ranges from two to four years.

better decision in the future. However, at the same time, it also means that the size of the remaining portion of the good decreases. For instance, experiencing more included sessions helps the consumer better learn the valuation but decreases the number of sessions that can be sold beyond the trial; while entering a contract with a long lease term induces sufficient learning, it may make the buyout option unattractive as the car will be getting old.

From the seller's perspective, she faces a tradeoff between *information* and *quantity*: The more she tries to sell initially, the more additional information the buyer acquires, and thus the more surplus she can extract from each unit of second-stage consumption; but by selling more initially, she has a smaller base to extract the second-stage surplus. How should the seller incorporate such kind of learning by consuming into her selling mechanism? To address this question, we study a two-stage model, in which a risk-neutral seller sells one unit of a *divisible* good to a risk-neutral buyer. The buyer's valuation depends on how well the good fits him, which is uncertain to him at stage one. Yet, at stage one, he has a prior – rough private valuation of the good. Relying on this rough valuation, he decides how many units to experience. Experiencing the good provides him with additional private information regarding the good's characteristics. Consuming more leads to more precise additional information.[3] With the updated private valuation, the buyer further determines his second-stage consumption level.

In our problem, the first-stage allocation (*i.e.*, consumption)[4] plays two roles. First, it is the device for information acquisition, since it provides the buyer with additional information, which will be more precise with a higher first-stage allocation. Meanwhile, it also defines an *intertemporal* problem: It "secures" some consumption in the early period, regardless of whether the newly acquired information is good or bad, and also determines the maximum amount of consumption in the later stage. The seller then faces the above-mentioned tradeoff between information and quantity. Clearly, a revenue-maximizing seller should incorporate both roles of the first-stage allocation into her pricing strategy.

We fully characterize the revenue-maximizing mechanism and find that the optimum can be implemented by a menu of *try-and-decide* contracts, consisting of a first-stage price-quantity pair and a second-stage per-unit price for the remaining quantity. When

[3] The precision of the additional information through consumption is defined in the sense of rotation order; see Section 2 for details.

[4] We use "allocation" and "consumption" interchangeably.

the buyer selects some try-and-decide contract, he needs to pay the corresponding first-stage price specified by the contract chosen. By doing this, the buyer not only gets to experience the corresponding first-stage portion (*i.e.*, quantity) of the good, but also obtains the option to buy the remaining portion at the prescribed second-stage per-unit price. In the optimal contract, a larger first-stage consumption level (quantity) is paired with a higher first-stage price but a lower per-unit second-stage price for the remaining portion of the good. Moreover, if the buyer ends up buying the entire good across two stages, a higher first-stage consumption leads to a lower total payment. In equilibrium, a higher first-stage valuation buyer pays more to consume more in the first stage, in exchange for a lower per-unit price for the remaining portion.

The intuition results from the seller's tradeoff between *information* and *quantity*. A key feature of our model is that the higher the initial consumption level is, the more additional private information the buyer acquires about the true value of the good. This yields a tradeoff between information and quantity on the seller's revenue when designing the first-stage allocation. The more she tries to sell initially, the more profitable each unit of second-stage consumption is, which is the information effect. But by selling more initially, the remaining quantity shrinks, and thus she has a smaller base to extract the second-stage surplus, while at the same time the surplus extracted from the first-stage consumption changes, which is the quantity effect. At the optimum, the information effect is positive to balance the negative quantity effect. Comparative statics with respect to the first-stage type implies that: For a slightly higher first-stage type, the information effect dominates the quantity effect at the margin, and thus at the optimum, the corresponding first-stage allocation must be higher as well.

From the buyer's perspective, when choosing the contract, a high first-stage valuation buyer is more confident that his updated valuation of the good is sufficiently high, so that he will likely buy the entire portion of the good. Thus, he is willing to pay to consume more in the first stage to enjoy a lower price for the additional consumption in the second stage, and also a lower total price for the entire portion of the good. However, this is quite risky for a low first-stage valuation buyer. If he does so, despite the second-stage per-unit price being lower, he has to pay to consume more in the first stage to enjoy this second-stage benefit. Yet, since his first-stage valuation is low, he really wants to experience the good just a bit to improve his decision in the second stage, rather than "blindly" having a high first-stage consumption level, which can lead to a low expected payoff given the low first-stage valuation.

The format of our optimal try-and-decide contracts resembles practical contracts. For instance, in the course registration example, the included sessions and the remaining sessions can be viewed as the consumption at the two stages, respectively. In the car-leasing example, the leasing price can be viewed as the first-stage price, while the lease term and the buyout price can be regarded as the first-stage quantity and second-stage price. In this sense, we provide a possible rationale for the common phenomena of sequential consumption with learning in reality.

On the technical side, we would like to point out that when solving the optimal mechanism, establishing the global incentive compatibility (IC) condition is non-standard and quite involved in our setting. In canonical sequential screening problems, *e.g.*, Courty and Li (2000) and Eső and Szentes (2007a), building on local IC, the monotonicity of an allocation rule leads to global IC, even if it may not be the optimal allocation rule. By imposing certain regularity conditions, one can check that the optimal allocation rule in the relaxed problem, which only uses local IC, is indeed monotone. As a result, such an allocation rule also satisfies global IC, so that it is indeed the optimum. However, we cannot merely verify the monotonicity of the first-stage allocation rule and expect the standard approach to go through in our setting. In fact, we provide an example where the first-stage allocation rule is monotone, but this allocation rule cannot be part of a global IC mechanism. Instead, in our setting, when establishing global IC, one also needs to use the optimality of the allocations that satisfy the local IC condition. Global IC follows from local IC and the optimality of the allocation, rather than local IC alone.

This difference can be better explained by focusing on the first-stage problem, where the sequential screening problem can be understood as a corresponding static screening problem (see Krähmer and Strausz (2017) for further discussions). In the corresponding static problem, a standard condition in the literature, often called the single-crossing condition (alternatively, the constant sign condition, or the Spence–Mirrlees condition), is missing. The lack of such a condition prevents us from establishing global IC only from the monotonicity of the first-stage allocation rule.

The rest of the paper is organized as follows. Section 2 sets up the model. We analyze the solution of a relaxed problem in Section 3 and the optimal mechanism in Section 4. Section 5 discusses the results, and Section 6 examines the robustness of our findings. Section 7 concludes, and the appendix collects omitted proofs.

Literature

Our paper joins the growing literature on dynamic mechanism design.[5] The canonical literature typically assumes that the agent has two (or more) stages of private information, where the distribution of the second-stage private information is exogenously determined by the first-stage private information. In particular, it is often assumed that a higher first-stage type corresponds to a better distribution of the second-stage type in the sense of FOSD. See, for example, Courty and Li (2000), Eső and Szentes (2007a), Krähmer and Strausz (2015, 2017), and more recently Li and Shi (2022), as well as Battaglini (2005) and Garrett and Pavan (2012) for infinite stages and Boleslavsky and Said (2013) for arbitrarily long time horizons. However, in the current work, the distribution of the second-stage valuation (type) depends on the first-stage consumption, which is endogenously chosen by the buyer. Due to rotation ordering,[6] such endogenously generated second-stage information is no longer ranked in terms of FOSD, which is a feature that does not exist in many canonical papers.[7]

Within the dynamic mechanism design literature, there is a strand that involves information acquisition and provision. Among these papers, the channel of information provision is often independent of the product sold by the principal. One approach to model information provision assumes that the principal releases information directly or through the sales of a second good. For example, in Eső and Szentes (2007a,b), Li and Shi (2017), and Guo et al. (2022), the principal directly controls how much information to release to the agent. In Hoffmann and Inderst (2011), the principal produces two goods: a product itself and an additional information provision service; the amount of service determines the precision of the agent's learning. Another approach models information acquisition as a moral hazard problem or an entry problem. For example, in Krähmer and Strausz (2011), the agent can take a hidden action to gather information; in Lu et al. (2021), the agent can incur an entry cost to fully observe the ex post value. Some studies, for example, Lu and Wang (2021), model the information acquisition as searches. An exception is Bonatti (2011), where learning by consuming also takes place. In that paper, an agent's private valuation does not evolve, but the symmetric information about product quality is revealed gradually through total consumption in the market. In the

[5]This is now an extensive literature; see Bergemann and Välimäki (2019) for an excellent survey.

[6]Studies that also use this information order include, for example, Johnson and Myatt (2006), Hoffmann and Inderst (2011), and Shi (2012).

[7]In Liu and Lu (2018, 2022), the second-stage type's distribution is also endogenous, but it is ranked by FOSD.

current paper, the buyer's first-stage consumption plays a dual role: The buyer not only enjoys a payoff but also acquires additional private information from the first-stage consumption. Private learning from the allocation (*i.e.*, consumption) itself differentiates the current work from the above-mentioned papers.

This paper also features an *intertemporal* problem: The first-stage allocation not only endogenously shapes the distribution of the second-stage valuation but also changes the feasibility constraint of the second-stage allocation. Pavan et al. (2014) accommodate this feature before the current work, but they focus on providing a general approach to tackle dynamic mechanism design problems. On the other hand, we explicitly characterize the optimum in a consumer-learning environment.

2 The Model

A risk-neutral monopolist sells one unit of a *divisible* good to a risk-neutral buyer in two stages. The buyer's true valuation V of the good depends on how well the good fits him. At stage one, the buyer is uncertain about V, but he observes a "rough" valuation of the good, v_1. Relying on this rough valuation, the buyer purchases $q_1 \in [0,1]$ units of the good. The true valuation V is jointly determined by v_1 and the additional information $\tilde{v}_2$, which is independent of v_1.[8] We assume that $E[\tilde{v}_2] = 0$ and $V = v_1 + \tilde{v}_2$.[9] Consuming/experiencing q_1 units of the good provides the buyer with additional information to learn about $\tilde{v}_2$. With a more precise assessment of the good, at stage two, the buyer decides how many further units $q_2 \in [0, 1 - q_1]$ to buy. The buyer's outside option is normalized to be 0.

From the seller's perspective, v_1 is a random draw from a cumulative distribution function (CDF) $G(\cdot)$, which admits a strictly positive continuous density function g over the support $[0,1]$.[10] The buyer learns additional information about $\tilde{v}_2$ through consumption: After buying q_1 units of the good, the buyer forms a posterior estimate

[8] The assumption that v_1 and $\tilde{v}_2$ are independent means that the ex ante information asymmetry does not depend on the precision of the second-stage information, which helps us provide a clean characterization of the optimal mechanism. On the other hand, when considering the more general setting that v_1 and $\tilde{v}_2$ are correlated, one has to impose additional restrictions on the information structure; see, for example, Courty and Li (2000) and Eső and Szentes (2007a). We focus on the current setting, as working with the more general structure will make the analysis much more complicated and the new insights less transparent.

[9] Such an additive form is not an assumption, because one can define the difference between the true valuation and the rough valuation as the additional information.

[10] It is without loss to assume that the support is $[0, 1]$.

v_2 of $\tilde{v}_2$. From an ex ante perspective, v_2 follows the CDF $F(\cdot|q_1)$. The realization of v_2 is again the buyer's private information. The seller's goal is to design a contract that maximizes her revenue.

Intuitively, consuming more at stage one helps the buyer acquire more precise information at stage two. The precision of the additional information v_2 through consumption is captured by the rotation order (cf. Johnson and Myatt, 2006). Specifically, for any $q_1 \in (0, 1]$ and $v_2 \in (-\infty, +\infty)$, $F(v_2|q_1)$ is continuously differentiable in q_1 such that

$$\frac{\partial F(v_2|q_1)}{\partial q_1} \begin{cases} > 0, & \text{when } v_2 < 0; \\ = 0, & \text{when } v_2 = 0; \\ < 0, & \text{when } v_2 > 0. \end{cases}$$

When $q_1 = 0$, $F(\cdot|q_1)$ degenerates to a mass at $E[\tilde{v}_2] = 0$, capturing no additional information gained if there is no consumption. For convenience, suppose that when $q_1 > 0$, $F(v_2|q_1)$ is twice continuously differentiable in v_2 and the corresponding density function $f(v_2|q_1) > 0$. We also assume that for any $v_2 \neq 0$, $\lim_{q_1 \to 0+} F(v_2|q_1) = F(v_2|0)$.

To illustrate the setting, consider the following "truth-or-noise" example, which has been widely used in the literature; see, for example, Lewis and Sappington (1994) and Johnson and Myatt (2006).

Example 1 (Truth-or-noise)**.** *After consuming q_1 units, the buyer observes a signal s of $\tilde{v}_2$. The signal is true (i.e., $s = \tilde{v}_2$) with probability q_1, and is completely noisy (i.e., s is an independent random draw from the same CDF as $\tilde{v}_2$) with probability $1 - q_1$. Denote the CDF of $\tilde{v}_2$ by $H(\cdot)$, which is twice continuously differentiable over the support $(-\infty, +\infty)$. Then, when observing the signal s, the buyer's posterior estimate of $\tilde{v}_2$ is*

$$v_2 = E[\tilde{v}_2|s, q_1] = q_1 s + (1 - q_1)E[\tilde{v}_2] = q_1 s.$$

Thus, ex ante, v_2 follows the CDF $F(v_2|q_1) = H(\frac{v_2}{q_1})$, which satisfies all the assumptions mentioned above.

After consuming q_1 units, imagine that the buyer adopts a simple threshold plan v_2. That is, he will buy the rest $1 - q_1$ units at stage two only when the additional information acquired is sufficiently good — *i.e.*, when the additional information is higher than the threshold v_2. Thus, the buyer's expected consumption in the second stage is

$(1-q_1)(1-F(v_2|q_1))$, and the expected total consumption across two stages is

$$C(v_2, q_1) = q_1 + (1-q_1)(1-F(v_2|q_1)) = 1-(1-q_1)F(v_2|q_1). \tag{1}$$

When $q_1 > 0$, it is clear that a higher threshold v_2 — *i.e.*, requiring better information — leads to a drop in C, as the partial derivative with respect to v_2,

$$\frac{\partial C(v_2, q_1)}{\partial v_2} = -(1-q_1)f(v_2|q_1) < 0.$$

However, the effect of a higher first-stage consumption q_1 on C is ambiguous. The "marginal rate of substitution" is

$$M(v_2, q_1) = \frac{\frac{\partial C(v_2,q_1)}{\partial q_1}}{\frac{\partial C(v_2,q_1)}{\partial v_2}} = -\frac{1}{1-q_1}\frac{F(v_2|q_1)}{f(v_2|q_1)} + \frac{\frac{\partial F(v_2|q_1)}{\partial q_1}}{f(v_2|q_1)}$$

for any $(v_2, q_1) \in \mathbb{R} \times (0,1)$.

We make the following assumption regarding $M(v_2, q_1)$.

Assumption 1. *For any fixed $q_1 \in (0,1)$ and $v_2 < v_2'$,*

$$M(v_2, q_1) \leq 0 \quad \Longrightarrow \quad M(v_2', q_1) < 0.$$

Remark 1. *Assumption 1 is a natural assumption regarding the substitution between the first-stage consumption and the additional information. It says that: If at a particular level of first-stage consumption q_1 and a certain requirement of additional information v_2, the buyer is willing to sacrifice his first-stage consumption in exchange for a lower requirement of information (i.e., a lower v_2), then he will still be willing to do so when the requirement of information is more stringent than v_2 (i.e., higher than v_2).*

Note that Assumption 1 holds when $F(v_2|q_1)/f(v_2|q_1)$ is increasing in v_2 and $\frac{\partial F(v_2|q_1)}{\partial q_1}/f(v_2|q_1)$ is decreasing in v_2. The former condition is a standard hazard rate assumption. The latter one is the same as Assumption 3 in Shi (2012), which can be interpreted as supermodularity. In the truth-or-noise example above, the latter assumption is automatically satisfied, while the former assumption is satisfied when $H(x)/h(x)$ is increasing in x.

Finally, we make the following standard hazard rate assumption about G.

Assumption 2. *We assume that $\frac{1-G(v_1)}{g(v_1)}$ is strictly decreasing in v_1.*

We shall focus on truthful direct mechanisms $\{q_1(r_1), q_2(r_1, r_2), t(r_1, r_2)\}_{r_1 \in [0,1], r_2 \in \mathbb{R}}$, which is without loss of generality; see Myerson (1986). In the first stage, when the buyer reports r_1, the seller allocates $q_1(r_1)$ units of the good to him. In the second stage, v_2 is realized according to $F(\cdot|q_1(r_1))$. Given the buyer's second-stage report r_2, the seller allocates $q_2(r_1, r_2)$ units of the good to the buyer and demands a payment $t(r_1, r_2)$.

2.1 The buyer's problem

Suppose that the buyer truthfully reported v_1 at stage one, but he reports r_2 despite that the true second-stage type is v_2. Let $\tilde{\pi}(v_1, r_2, v_2)$ be his expected payoff at stage two: $\tilde{\pi}(v_1, r_2, v_2) = (v_1 + v_2)q_2(v_1, r_2) - t(v_1, r_2)$. Envelope theorem yields

$$\frac{\mathrm{d}\tilde{\pi}(v_1, v_2, v_2)}{\mathrm{d}v_2} = \frac{\partial \tilde{\pi}(v_1, r_2, v_2)}{\partial v_2}|_{r_2 = v_2} = q_2(v_1, v_2).$$

Denote $\psi(v_1) = v_1 - \frac{1-G(v_1)}{g(v_1)}$, which is strictly increasing by Assumption 2. The following result is standard (see, *e.g.*, Eső and Szentes, 2007a), and its proof is omitted.

Lemma 1. *(i) Suppose that the buyer reports the first-stage type v_1 truthfully. The second-stage IC constraint is satisfied if and only if the following two conditions hold:*

(a) For any v_1 and v_2,[11]

$$\tilde{\pi}(v_1, v_2, v_2) = \tilde{\pi}(v_1, -\psi(v_1), -\psi(v_1)) + \int_{-\psi(v_1)}^{v_2} q_2(v_1, s)\,\mathrm{d}s. \tag{2}$$

(b) The second-stage allocation $q_2(v_1, v_2)$ is increasing in v_2 for any v_1.

(ii) On the other hand, suppose that the buyer's first-stage type is v_1 but he reported r_1 in the first stage. Then, when he observes v_2 in the second stage, he will report $r_2 = r_2(v_1, r_1, v_2)$ such that $r_1 + r_2(v_1, r_1, v_2) = v_1 + v_2$.

Based on Lemma 1, the expected payoff of the buyer with first-stage type v_1 and report r_1 can be expressed as

$$U(v_1, r_1) \quad = \quad q_1(r_1) \int_{-\infty}^{+\infty} (v_1 + v_2) F(\mathrm{d}v_2 | q_1(r_1))$$

[11] When $v_2 < -\psi(v_1)$, $\int_{-\psi(v_1)}^{v_2} q_2(v_1, s)\,\mathrm{d}s = -\int_{v_2}^{-\psi(v_1)} q_2(v_1, s)\,\mathrm{d}s$.

$$+\int_{-\infty}^{+\infty}\left[\begin{array}{c}(v_1+v_2)q_2(r_1,r_2(v_1,r_1,v_2))\\ -t(r_1,r_2(v_1,r_1,v_2))\end{array}\right]F(\mathrm{d}v_2|q_1(r_1)).$$

The first-stage IC constraint requires that for any v_1 and r_1, $U(v_1,v_1)\geq U(v_1,r_1)$. The following result provides a necessary condition for the first-stage IC constraint (all the proofs are relegated to the Appendix).

Lemma 2. *The first-stage IC constraint implies that for any* v_1

$$U(v_1,v_1)=U(0,0)+\int_0^{v_1}\left[q_1(s)+\int_{-\infty}^{+\infty}q_2(s,v_2)F(\mathrm{d}v_2|q_1(s))\right]\mathrm{d}s.$$

2.2 The seller's problem

The seller's expected revenue is the difference between the social welfare and the buyer's ex ante expected payoff. By Lemma 2, it can be written as

$$\begin{aligned}R &= \int_0^1\left\{v_1q_1(v_1)+\int_{-\infty}^{+\infty}(v_1+v_2)q_2(v_1,v_2)F(\mathrm{d}v_2|q_1(v_1))\right\}g(v_1)\,\mathrm{d}v_1\\ &\quad -\int_0^1 U(v_1,v_1)g(v_1)\,\mathrm{d}v_1\\ &= \int_0^1\left[\begin{array}{c}\psi(v_1)q_1(v_1)\\ +\int_{-\infty}^{+\infty}[\psi(v_1)+v_2]q_2(v_1,v_2)F(\mathrm{d}v_2|q_1(v_1))\end{array}\right]g(v_1)\,\mathrm{d}v_1-U(0,0).\end{aligned}$$

Now we are ready to state the seller's problem as follows.

$$\text{Problem (O)}:\qquad \max_{(q_1,q_2,t)} R$$

subject to

$$\text{constraint (2) and } q_2(v_1,v_2) \text{ is increasing in } v_2 \text{ for any } v_1; \tag{3}$$

$$U(v_1,v_1)\geq U(v_1,r_1),\ \text{for any } v_1,r_1; \tag{4}$$

$$0\leq q_1(v_1)\leq 1 \text{ and } 0\leq q_2(v_1,v_2)\leq 1-q_1(v_1),\ \text{for any } v_1,v_2; \tag{5}$$

$$U(v_1,v_1)\geq 0,\ \text{for any } v_1. \tag{6}$$

Here, (3) is the equivalent condition for the second-stage IC constraint, (4) is the first-stage IC constraint, (5) is the feasibility constraint imposed on allocations, and (6) is the individual rationality constraint.

Clearly, at the optimum, $U(0,0) = 0$, and thus the seller's ex ante revenue becomes

$$R = \int_0^1 \left[\begin{array}{c} \psi(v_1) q_1(v_1) \\ + \int_{-\infty}^{+\infty} [\psi(v_1) + v_2] q_2(v_1, v_2) F(\mathrm{d}v_2 | q_1(v_1)) \end{array} \right] g(v_1) \, \mathrm{d}v_1.$$

Notice that Lemma 2 implies that for an IC mechanism, (6) holds if and only if $U(0,0) \geq 0$. As such, with $U(0,0) = 0$, we can drop constraint (6). We further drop constraints (3) to (4) and thus can omit the choice variable t to form a relaxed problem, Problem (O-R):

$$\max_{(q_1, q_2)} R$$

subject to constraint (5).

If the solution to Problem (O-R) also satisfies (3) and (4), then such a solution also solves Problem (O). As such, we will first solve Problem (O-R), and then verify that its solution satisfies all the constraints in the original problem.

3 The solution to Problem (O-R)

In Problem (O-R), for each fixed first-stage allocation rule q_1, the following second-stage allocation rule q_2 obviously maximizes the objective function R:

$$q_2(v_1, v_2) = \begin{cases} 1 - q_1(v_1), & \text{if } \psi(v_1) + v_2 \geq 0; \\ 0, & \text{otherwise.} \end{cases}$$

Thus, we have[12]

$$\begin{aligned} R &= \int_0^1 \left[\psi(v_1) q_1(v_1) + \int_{-\psi(v_1)}^{+\infty} [\psi(v_1) + v_2](1 - q_1(v_1)) F(\mathrm{d}v_2 | q_1(v_1)) \right] g(v_1) \, \mathrm{d}v_1 && (7) \\ &= \int_0^1 \left[\psi(v_1) + (1 - q_1(v_1)) \int_{-\infty}^{-\psi(v_1)} F(v_2 | q_1(v_1)) \, \mathrm{d}v_2 \right] g(v_1) \, \mathrm{d}v_1. && (8) \end{aligned}$$

To facilitate the presentation, define the seller's revenue from a type-v_1 buyer with

[12] The detailed derivation can be found in the Appendix.

the first-stage consumption q_1 as

$$\Pi(q_1, v_1) = \psi(v_1) + (1 - q_1) \int_{-\infty}^{-\psi(v_1)} F(v_2|q_1)\, \mathrm{d}v_2. \tag{9}$$

For each v_1, let $q_1^*(v_1)$ be the maximizer of $\Pi(q_1, v_1)$ in $q_1 \in [0, 1]$. Finally, define

$$v_1^* = \psi^{-1}(0) \text{ and } \tilde{v}_1 = \inf\{v_1 \in [0, 1] : q_1^*(v_1) > 0\}.$$

We have the following observation.

Lemma 3. *The following allocation rule pair* $\{q_1^*(v_1), q_2^*(v_1, v_2)\}_{v_1 \in [0,1], v_2 \in \mathbb{R}}$ *solves Problem (O-R):*

(i) The first-stage allocation $q_1^*(v_1)$ is the maximizer of

$$\Pi(q_1, v_1) = \psi(v_1) + (1 - q_1) \int_{-\infty}^{-\psi(v_1)} F(v_2|q_1)\, \mathrm{d}v_2$$

in $q_1 \in [0, 1]$ for each v_1. In particular, there exists a cutoff $\tilde{v}_1 < v_1^*$ such that $q_1^*(v_1) = 0$ for all $v_1 < \tilde{v}_1$ if any. For all $v_1 \in [\tilde{v}_1, 1]$, $q_1^*(v_1) \in [0, 1)$ and can be characterized by the first-order condition:

$$\int_{-\infty}^{-\psi(v_1)} \left[-F(v_2|q_1^*(v_1)) + (1 - q_1^*(v_1)) \frac{\partial F(v_2|q_1^*(v_1))}{\partial q_1} \right] \mathrm{d}v_2 = 0. \tag{10}$$

Moreover, for all $v_1 \in (\tilde{v}_1, 1]$, $q_1^*(v_1) \in (0, 1)$.

(ii) The second-stage allocation rule q_2^* is

$$q_2^*(v_1, v_2) = \begin{cases} 1 - q_1^*(v_1), & \text{if } \psi(v_1) + v_2 \geq 0; \\ 0, & \text{otherwise.} \end{cases}$$

In particular, when $v_1 < \tilde{v}_1$, $F(\cdot|q_1^*(v_1))$ degenerates to a mass at 0 and $q_2^*(v_1, v_2) = 0$ for all v_2.

The above lemma implies that there is a first-stage cutoff type $\tilde{v}_1$, below which both the first-stage allocation and the second-stage allocation are zero. Hence, there is no consumption at the bottom of the distribution G. Moreover, this cutoff is strictly below v_1^*. This is because a buyer with a not-too-low first-stage type would still be willing to experience the good a bit, as he knows that his ex post value (*i.e.*, $v_1 + v_2$) is still likely to be high enough. The seller then should take advantage of this by setting a positive allocation for such first-stage types. However, when the first-stage type is so low (below $\tilde{v}_1$) that the ex post value is quite unlikely to be sufficiently high, the buyer does not find it worthwhile to experience the good. The seller, anticipating this, should set a zero first-stage allocation in this case.

The following result characterizes the monotonicity of the optimal first-stage allocation rule.

Lemma 4. *The first-stage allocation $q_1^*(v_1)$ is strictly increasing in $v_1 \in [\tilde{v}_1, 1]$.*

The intuition results from the seller's tradeoff between *information* and *quantity*. Recall that in (7), the seller's revenue from a type-v_1 buyer can be expressed as:

$$\underbrace{q_1}_{\text{1st-stage quantity}} \cdot \underbrace{\psi(v_1)}_{\text{virtual value}} + \underbrace{(1-q_1)}_{\text{remaining quantity}} \cdot \underbrace{\int_{-\psi(v_1)}^{+\infty} [\psi(v_1) + v_2] F(\mathrm{d}v_2 | q_1)}_{\text{surplus from per-unit 2nd-stage consumption}} .$$

The first term $q_1\psi(v_1)$ is the revenue gained from the first-stage consumption, while the second term is the revenue generated from the second-stage consumption. After rewriting the integral, as in (8), the above equation can be expressed as

$$\underbrace{q_1\psi(v_1)}_{\text{1st-stage surplus}} + \underbrace{(1-q_1)}_{\text{remaining quantity}} \left[\underbrace{\psi(v_1)}_{\text{per-unit surplus w/o learning}} + \underbrace{\int_{-\infty}^{-\psi(v_1)} F(v_2|q_1)\,\mathrm{d}v_2}_{\text{additional per-unit surplus w/ learning}} \right].$$

The square bracket decomposes the surplus from each unit of second-stage allocation into two parts: (i) the surplus gained had there been no learning (the "base surplus") and (ii) the *additional* surplus gained from offering the buyer the option of learning.

A key feature of our model is that the higher the initial consumption level q_1 is, the more additional private information (in the sense of rotation order) the buyer acquires about the true value of the good. This yields a tradeoff between *information* and *quantity* on the seller's revenue when designing the first-stage allocation. The more she tries to sell initially, the more profitable each unit of second-stage consumption is (the square-bracket term),[13] which is the information effect. But by selling more initially, the remaining

[13]The square-bracket term is increasing in q_1 due to the rotation order.

quantity shrinks so she has a smaller base to extract the second-stage surplus (the term $1 - q_1$ decreases in q_1), while at the same time the surplus extracted from the first-stage consumption changes,[14] which is the quantity effect.

Mathematically, the information effect is $(1 - q_1) \int_{-\infty}^{-\psi(v_1)} \frac{\partial F(v_2|q_1)}{\partial q_1} \, \mathrm{d}v$; the quantity effect is

$$\underbrace{\psi(v_1) - \psi(v_1)}_{\text{cancel out}} - \int_{-\infty}^{-\psi(v_1)} F(v_2|q_1) \, \mathrm{d}v_2 = - \int_{-\infty}^{-\psi(v_1)} F(v_2|q_1) \, \mathrm{d}v_2, \tag{11}$$

where the first two terms cancel out because the two stages' quantities sum up to one whenever there is consumption at both stages. The information effect is an integral, because an increase in the first-stage consumption level impacts the seller's revenue for all realizations of the additional private signal that induces selling of the remainder of the good.

Clearly, for each fixed first-stage type v_1, at the optimum, the information effect is positive to balance the negative quantity effect, which gives rise to the first-order condition (10). Now, consider comparative statics with respect to the first-stage type. It turns out that at the optimum, the higher the initial estimate v_1 is, the more the tradeoff between information and quantity will be resolved towards information, which then leads to Lemma 4. To see this, notice that for a slightly higher first-stage type $v_1 + \varepsilon$, the negative quantity effect is increased at the margin, so it gets weaker; while the positive information effect is decreased (if $\psi(v_i) \geq 0$) or increased (if $\psi(v_i) < 0$) at the margin. Assumption 1 ensures that the overall effect is increased at the margin. As such, for a slightly higher first-stage type, the information effect dominates the quantity effect at the margin, so at the optimum the corresponding first-stage allocation must be higher as well.

Remark 2. *To ease the notations and presentation, we do not take into account the possibility that for some v_1, the maximizer $q_1^*(v_1)$ may not be unique. This multiplicity issue can be addressed by imposing the following assumption analogous to Assumption 1:*

- *For any fixed $v_2 \leq 0$ and any $q_1, q_1' \in (0, 1)$ with $q_1 < q_1'$,*

$$M(v_2, q_1) \leq 0 \qquad \Longrightarrow \qquad M(v_2, q_1') < 0.$$

This assumption can be interpreted as a natural substitution condition, which says

[14] The term $q_1 \psi(v_1)$ increases or decreases in q_1 depending on the sign of $\psi(v_1)$.

that: If at a particular level of first-stage consumption q_1 and a certain requirement of additional information v_2, the buyer is willing to sacrifice his first-stage consumption in exchange for a lower requirement of information, then he will still be willing to do so when his first-stage consumption is higher than q_1. Under this condition, one can show that any selection of maximizers must be strictly increasing in v_1 when $v_1 \geq \tilde{v}_1$. The proof of this claim is given in the Appendix.

4 The solution to Problem (O)

4.1 Optimal direct mechanism

Having characterized the solution to Problem (O-R) as in Lemma 3, we can use the envelope conditions in Lemmas 1 and 2 to construct a payment rule t^*, the expression of which is provided in the following result.

Lemma 5. *The payment rule t^* is specified as follows:*

$$t^*(v_1, v_2) = \begin{cases} (1 - q_1^*(v_1))p_2^*(v_1) + p_1^*(v_1), & \text{if } \psi(v_1) + v_2 \geq 0, \\ p_1^*(v_1), & \text{otherwise,} \end{cases}$$

where

$$\begin{aligned} p_1^*(v_1) = (1 - q_1^*(v_1)) &\left[\int_{-\infty}^{-\psi(v_1)} F(v_2 | q_1^*(v_1))\, \mathrm{d}v_2 - \frac{1 - G(v_1)}{g(v_1)} \right] \\ &+ \int_0^{v_1} (1 - q_1^*(x)) F(-\psi(x) | q_1^*(x))\, \mathrm{d}x \end{aligned}$$

and

$$p_2^*(v_1) = \frac{1 - G(v_1)}{g(v_1)}.$$

Thus, we obtain a candidate mechanism $\{q_1^*(v_1), q_2^*(v_1, v_2), t^*(v_1, v_2)\}_{v_1 \in [0,1], v_2 \in \mathbb{R}}$ for Problem (O). If we can show that this candidate mechanism satisfies all constraints in Problem (O), then it must solve Problem (O). Clearly, one only needs to verify constraints (3) and (4) — *i.e.*, the first- and second-stage IC constraints.

To this end, we begin by considering a menu of *try-and-decide* option contracts $\{p_1^*(v_1), q_1^*(v_1); p_2^*(v_1)\}_{v_1 \in [0,1]}$, where functions p_1^*, q_1^*, and p_2^* are defined in Lemmas 3 and 5. The buyer needs to select a contract from the menu. If for some $r_1 \in [0, 1]$, option contract $\{p_1^*(r_1), q_1^*(r_1); p_2^*(r_1)\}$ is selected, then $p_1^*(r_1)$ is the advance payment for a buyer to enter this contract. By paying this advance payment, the buyer not only consumes $q_1^*(r_1)$ units of the good, but also reserves the right to buy the remaining $1 - q_1^*(r_1)$ units at the per-unit strike price $p_2^*(r_1)$.

In the Appendix, we show that this menu of contracts implements the above-mentioned direct mechanism. Hence, our direct mechanism satisfies constraints (3) and (4), and thus is a solution to Problem (O), which is the following result.

Proposition 1. *The direct mechanism* $\{q_1^*(v_1), q_2^*(v_1, v_2), t^*(v_1, v_2)\}_{v_1 \in [0,1], v_2 \in \mathbb{R}}$ *can be implemented by a menu of try-and-decide option contracts* $\{p_1^*(v_1), q_1^*(v_1); p_2^*(v_1)\}_{v_1 \in [0,1]}$. *Hence, this direct mechanism solves Problem (O).*

What is crucial in the proof of the above proposition is to establish global IC. The argument to establish IC in our problem is non-standard and quite involved. In many canonical sequential screening problems in the literature, *e.g.*, Courty and Li (2000) and Eső and Szentes (2007a), as long as the allocation rule satisfies certain monotonicity condition in private types, regardless of whether it is the solution of the relaxed problem or not, the allocation rule can be used to construct a direct mechanism satisfying global IC. However, this approach does not work for our problem. In fact, we provide an example in Section 4.3 where the first-stage allocation rule is monotone, but this allocation rule cannot be part of a global IC mechanism.

Most papers in the dynamic mechanism design literature does not involve intertemporal allocation, and global IC is established for *all* allocations that satisfy the local IC and certain monotonicity conditions, which is especially the case when the first-stage types are ranked by first-order stochastic dominance (FOSD). However, in our setting, there is also the quantity effect (intertemporal problem), so one needs to use the optimality of the allocations that satisfy the local IC. Global IC follows from local IC and the optimality of the allocation, instead of local IC alone. We will discuss this in detail in Section 5.

4.2 Implementation

The following lemma summarizes several useful properties of the payment rule, and we will discuss them after Proposition 2.

Lemma 6. *(i) For* $v_1 < \tilde{v}_1$ *and* v_2, $p_1^*(v_1) = 0$ *and* $t^*(v_1, v_2) = 0$.

(ii) The first-stage payment $p_1^*(v_1)$ is strictly increasing in $v_1 \in [\tilde{v}_1, 1]$ and equal to $q_1^*(\tilde{v}_1)\frac{1-G(\tilde{v}_1)}{g(\tilde{v}_1)} \geq 0$ for $v_1 = \tilde{v}_1$.

(iii) The second-stage per-unit payment $p_2^*(v_1)$ is strictly decreasing in $v_1 \in [0, 1]$ and equal to zero when $v_1 = 1$.

(iv) The function $p_1^*(v_1) + (1 - q_1^*(v_1))p_2^*(v_1)$ is strictly decreasing in $v_1 \in [0, 1]$.

(v) The function $p_1^*(v_1) + (1 - q_1^*(v_1))p_2^*(v_1)(1 - F(-\psi(v_1)|q_1^*(v_1)))$ is strictly increasing in $v_1 \in [\tilde{v}_1, 1]$ and equal to zero elsewhere.

Remark 3 (No participation at the bottom)**.** *According to Lemmas 3 and 6 (i), when $v_1 < \tilde{v}_1$, there is zero consumption in both stages and the total payment is zero. As such, the buyer with the first-stage type lower than $\tilde{v}_1$ is completely shut down — he does not participate.*

According to Remark 3, it is immediate that the "reduced" menu of try-and-decide option contracts $\{p_1^*(v_1), q_1^*(v_1); p_2^*(v_1)\}_{v_1 \in [\tilde{v}_1, 1]}$ implements the solution of Problem (O), since a buyer with first-stage type $v_1 < \tilde{v}_1$ simply does not participate. Hence, we have the following result.

Proposition 2. *The solution of Problem (O) can be implemented by a menu of try-and-decide option contracts $\{p_1^*(v_1), q_1^*(v_1); p_2^*(v_1)\}_{v_1 \in [\tilde{v}_1, 1]}$.*

According to Lemmas 4 and 6, the first-stage payment $p_1^*(v_1)$ and consumption $q_1^*(v_1)$ are strictly increasing when $v_1 \geq \tilde{v}_1$. However, the per-unit strike price $p_2^*(v_1)$ is strictly decreasing, and the total payment conditional on buying the entire portion of the good — *i.e.*, $p_1^*(v_1) + (1 - q_1^*(v_1))p_2^*(v_1)$ — is strictly decreasing in v_1.[15] This implies that in equilibrium, a buyer with a higher first-stage type will choose a contract with a higher advance payment and higher first-stage consumption, in exchange for a lower per-unit strike price for additional consumption in the second stage and a lower cost for purchasing the entire unit of the good.

The intuition is clear. For a high v_1-type buyer, he is more confident that his ex post valuation of the good, $v_1 + v_2$, is sufficiently high so that he will likely end up buying the entire good. The buyer is thus incentivized to choose a contract that "secures" a large first-stage consumption and first-stage payment so that he can enjoy a lower per-unit

[15]Note that this is the total payment *conditional on* buying the entire good. It is not the ex ante payment from a type-v_1 buyer, which is increasing in v_1 according to Lemma 6 (v).

strike price in stage two and a lower cost for purchasing the entire unit. However, this will be quite risky for a low v_1-type. If he does so, in spite of a lower per-unit second-stage price, he has to pay to consume more in the first stage. Yet, since his first-stage type is low, he really wants to experience the good a bit to make a better decision in the second stage, rather than "blindly" having a high first-stage consumption level, which can lead to a rather low expected payoff given his low first-stage type.

This intuition echoes phenomena seen in practice. For example, consumers who are optimistic about their matching quality with a course tend to register for the entire course outright, because doing so is usually cheaper than paying for the first few sessions with the intention to buy additional ones later. However, those who are not as optimistic may choose to attend the first few sessions before making a more substantial commitment. This practice can be expensive, but gives consumers an opportunity to experience the course before committing to it for a longer duration.

4.3 An illustrative example

In this section, we provide an illustrative example, which demonstrates the optimal mechanism that is identified in Proposition 1. In addition, we shall also illustrate that the monotonicity of the first-stage allocation rule does not imply global IC. To be precise, we explicitly construct an increasing first-stage allocation rule $\hat{q}_1$, and show that it cannot be part of an incentive-compatible mechanism.

Recall the truth-or-noise model in Example 1. Let G be the uniform distribution on $[0, 1]$ with the constant density $g \equiv 1$, and H be the normal distribution $N(0, 1)$ with mean 0 and variance 1 (*i.e.*, the density is $h(v_2) = \frac{1}{\sqrt{2\pi}} e^{-\frac{v_2^2}{2}}$). Then $\psi(v_1) = 2v_1 - 1$,

$$F(v_2|q_1) = H(\frac{v_2}{q_1}) = \int_{-\infty}^{\frac{v_2}{q_1}} \frac{1}{\sqrt{2\pi}} e^{-\frac{s^2}{2}} \, \mathrm{d}s, \qquad \text{and} \qquad f(v_2|q_1) = \frac{1}{q_1\sqrt{2\pi}} e^{-\frac{v_2^2}{2q_1^2}}.$$

In the Appendix, we verify that F satisfies the rotation order, and Assumptions 1 and 2 hold.

As shown in Proposition 1, the optimal first-stage allocation rule q_1^* must maximize

$$(1 - q_1) \int_{-\infty}^{-\psi(v_1)} F(v_2|q_1) \, \mathrm{d}v_2.$$

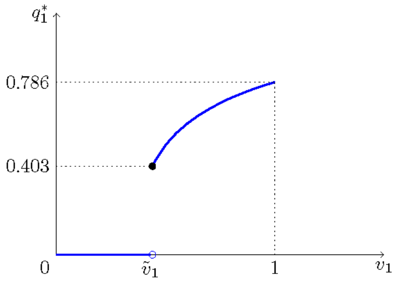

Figure 1: q_1^*

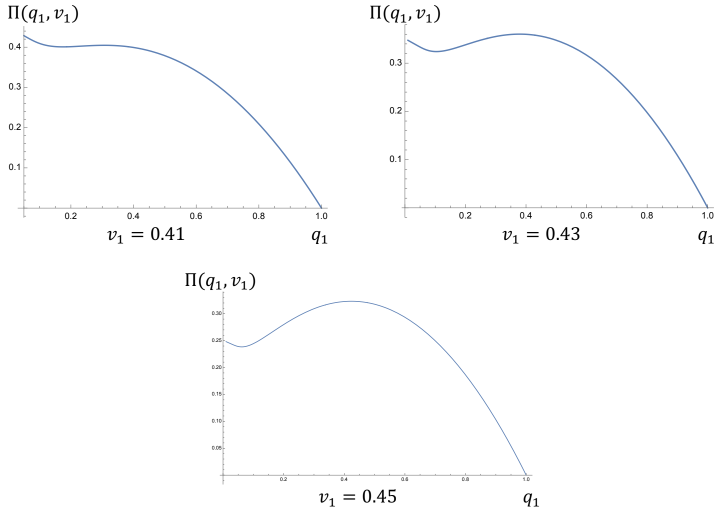

Figure 2: $\Pi(q_1, v_1)$

Figure 1 numerically illustrates the optimal allocation rule.[16] In this example, $\tilde{v}_1 \approx 0.43 < 0.5 = v_1^*$. In Figure 2, we plot $\Pi(q_1, v_1)$ when $v_1 = 0.41$, 0.43, and 0.45. As can be seen, $q_1^*(v_1)$ is unique and higher than $q_1^*(\tilde{v}_1) \approx 0.403$ when $v_1 > \tilde{v}_1 \approx 0.43$, is either 0 or almost 0.403 when $v_1 = \tilde{v}_1$, and is 0 when $v_1 < \tilde{v}_1$. This pattern explains the jump of q_1^* at $\tilde{v}_1$ in Figure 1.

Below, we construct another allocation rule $(\hat{q}_1, \hat{q}_2)$:

$$\hat{q}_1(v_1) = \begin{cases} \psi^2(v_1), & v_1 \geq \frac{1}{2}, \\ 0, & \text{otherwise}; \end{cases} \qquad \hat{q}_2(v_1, v_2) = \begin{cases} 1 - \hat{q}_1(v_1), & \psi(v_1) + v_2 \geq 0, \\ 0, & \text{otherwise}. \end{cases}$$

It is clear that $\hat{q}_1$ is increasing. In the Appendix, we show that $(\hat{q}_1, \hat{q}_2)$ cannot be the allocation rule in an incentive-compatible mechanism.

5 Discussions

As mentioned above, establishing global IC is non-standard and quite involved in our setting. This section discusses how and why our problem is different from the standard treatment and the related implications. To ease our discussion, we start with a brief overview of how global IC is established in our problem.

[16] In this example, we plot the figures by simulation.

5.1 Establishing global IC

Suppose that a type-v_1 buyer is considering the contract $\{p_1^*(r_1), q_1^*(r_1); p_2^*(r_1)\}$. Clearly, under this contract, the optimal second-stage strategy for a type-v_1 buyer who learns v_2 is to buy the remaining portion if and only if $v_2 \geq p_2^*(r_1) - v_1$, so his interim payoff is $U(v_1, r_1) = w(q_1^*(r_1), p_2^*(r_1), v_1) - p_1^*(r_1)$, where $w(q_1, p_2, v_1) \equiv q_1 v_1 + \int_{p_2 - v_1}^{+\infty}(v_1 + v_2 - p_2)(1 - q_1)F(\mathrm{d}v_2|q_1)$ for all $q_1 \in [0,1]$, $p_2 \geq 0$, and $v_1 \in [0,1]$.

It is shown in the proof of Proposition 1 that the difference of interim payoffs between selecting $\{p_1^*(v_1), q_1^*(v_1); p_2^*(v_1)\}$ and $\{p_1^*(r_1), q_1^*(r_1); p_2^*(r_1)\}$, $\Delta(v_1, r_1) = U(v_1, v_1) - U(v_1, r_1)$, is[17]

$$\int_{r_1}^{v_1} \int_{x}^{v_1} \Big[\, w_{31}(q_1^*(x), p_2^*(x), s) q_1^{*\prime}(x) + w_{32}(q_1^*(x), p_2^*(x), s) p_2^{*\prime}(x) \,\Big] \,\mathrm{d}s \,\mathrm{d}x. \tag{12}$$

Clearly, IC holds if and only if $\Delta(v_1, r_1) \geq 0$. To this end, it can be easily shown that $w_{32}(q_1, p_2, v_1) \leq 0$ for all $q_1 \in [0,1]$, $p_2 \geq 0$, and $v_1 \in [0,1]$. We have established in Lemma 5 that $p_2^*(v_1) = \frac{1-G(v_1)}{g(v_1)}$ and thus $p_2^{*\prime}(\cdot) < 0$ by Assumption 2. Hence, the second term in the integrand of the double integral (12), $w_{32}(q_1^*(x), p_2^*(x), s) p_2^{*\prime}(x) \geq 0$. Therefore, it suffices to show that the first term in the integrand is nonnegative.

By Lemmas 3 and 4, $q_1^{*\prime}(\cdot) \geq 0$. However, the sign of

$$w_{31}(q_1, p_2, v_1) = F(p_2 - v_1|q_1) - (1 - q_1)\frac{\partial F(p_2 - v_1|q_1)}{\partial q_1} \tag{13}$$

is ambiguous. Due to rotation order, when $p_2 > v_1$, $w_{31} > 0$; but when $p_2 < v_1$, w_{31} can be positive or negative. The ambiguity of the sign of w_{31} makes establishing IC complicated.[18]

In the proof of Proposition 1, we show that when q_1^* solves Problem (O-R), the double integral $\int_{r_1}^{v_1} \int_{x}^{v_1} w_{31}(q_1^*(x), p_2^*(x), s) q_1^{*\prime}(x) \,\mathrm{d}s \,\mathrm{d}x$ is positive, although its integrand has an ambiguous sign. Hence, we use the optimality of the solution to Problem (O-R) together with local IC to establish global IC, instead of local IC alone.

---

[17] In this paper, w_3 means partial derivative with respect to the third variable; likewise, w_{31} means the second-order partial derivative with respect to the first and the third variables. Other partial derivatives' notations are analogous.

[18] It is clear that if types were ranked by FOSD — *i.e.*, $\frac{\partial F(v_2|q_1)}{\partial q_1} < 0$ for any $q_1 > 0$ and v_2 — then the sign of w_{31} is no longer ambiguous (in fact, $w_{31} \geq 0$). It then follows that with FOSD, the first-stage allocation's monotonicity implies the first-stage IC, so establishing global IC would become standard.

5.2 Monotonicity does not imply global IC

In many canonical sequential screening problems in the literature (cf. Courty and Li, 2000; Eső and Szentes, 2007a), as long as the allocation rule satisfies certain monotonicity condition, regardless of whether it is the solution of the relaxed problem, the allocation rule can be used to construct a mechanism satisfying global IC. As such, the standard treatment in the literature focuses on identifying sufficient conditions under which the solution of the relaxed problem is monotone.

By nicely linking a canonical sequential screening problem to a static screening problem, Krähmer and Strausz (2017) show that the first-order stochastic dominance (FOSD) ranking of first-stage types in a canonical sequential screening problem as in Courty and Li (2000) corresponds to the single-crossing condition in the corresponding static screening problem; conversely, a sequential screening problem without FOSD corresponds to a static screening problem without the single-crossing condition.[19] With FOSD, the above-mentioned standard treatment in canonical sequential screening problems — *i.e.*, finding sufficient conditions under which the optimal allocation rule in a relaxed problem is monotone — works for establishing global IC. This is similar to the well-known result that in static screening problems, the single-crossing condition ensures that local IC plus monotonicity of the allocation rule implies global IC.

In our problem, types are not ranked by FOSD and the counterpart of the single-crossing condition does not hold in our corresponding static screening problem. More specifically, in the corresponding static screening problem, the condition requires that w_{31} has a constant sign in the respective integration region, but as we have discussed in Section 5.1, w_{31} does not satisfy this property.[20] In fact, the expression of w_{31} is complicated: The cutoff point at which w_{31} changes sign depends on v_1, p_2, and q_1 simultaneously. This is due to the *dual* role played by the first-stage allocation. First, the first-stage allocation is a device of *information acquisition*. Second, it defines an *intertemporal* problem: It affects the feasibility constraint of the second-stage allocation, as the second-stage allocation cannot exceed the remaining portion of the good. In problems where the first-stage allocation only affects information acquisition

[19]The single-crossing condition is also refereed to as the Spence-Mirrlees condition or the constant sign condition in the literature. Note that this condition should not be confused with the one required by our Assumption 1. To minimize confusion, we use the term single-crossing condition to refer to the standard condition imposed by canonical screening problems exclusively.

[20]The lack of single-crossing condition issue has been discussed by Araujo and Moreira (2010) and Schottmüller (2015) in their respective static screening environments.

(cf. Hoffmann and Inderst, 2011), the counterpart of w_{31} has a more clear-cut structure — the point at which their w_{31} changes sign depends on v_1 and p_2 only.

Due to the lack of the single-crossing condition, the monotonicity of allocation rule does not imply global IC, as seen in the example in Section 4.3. As is mentioned in Sections 4.3 and 5.1, the optimality of the first-stage allocation rule, *i.e.*, the fact that the first-stage allocation rule solves the relaxed problem, is explicitly used in establishing global IC, and this contrasts with many canonical sequential screening problems.

Finally, it is worth noting that Courty and Li (2000) also consider the case when the first-stage types are ranked by rotation order. The corresponding single-crossing condition also fails globally in their setting, because types are not ranked by FOSD. However, Courty and Li (2000) focus on subdomains where the single-crossing condition holds (though this condition fails over the whole domain) when solving the relaxed problem, so the issue in our paper does not arise in Courty and Li (2000); see also the discussion in Remark 5 of Krähmer and Strausz (2017).

6 Extension: Generalized feasibility

In our setting, the seller faces the key tradeoff between the information effect and the quantity effect. We showed, after Lemma 4, that with a slight increase in the first-stage type, the information effect dominates the quantity effect at the margin, so the optimal first-stage consumption increases with the first-stage type. Moreover, when establishing global IC, we need to use the optimality of the first-stage allocation together with the local IC conditions, instead of the latter alone.

In particular, the quantity effect takes a special form, because the consumption levels of the two stages sum up to one unit whenever there is consumption at both stages, *i.e.*, $q_1+q_2 = 1$, which is what divisibility of the good means. This leads to the special property that with respect to a slight change in the first-stage quantity, the marginal change in the first-stage surplus and the marginal change in the second-stage "base surplus" cancel out, so that this part is independent of the first-stage type, *i.e.*, the first two terms cancel out in equation (11). However, there can be other problems of learning by consuming which exhibit other forms of feasibility, so that the quantity effect no longer has this special property. How would this affect the monotonicity of the first-stage consumption? Would the method used when establishing global IC in the main model still apply?

This section aims to generalize the divisibility in our main model to a generalized feasibility setting and explore the tradeoff between information and quantity and the robustness of the approach used to establishing global IC. To this end, we replace the original feasibility constraint (5) with the following one:

$$0 \leq q_1(v_1) \leq 1 \text{ and } 0 \leq q_2(v_1, v_2) \leq s(q_1(v_1)), \text{ for any } v_1, v_2, \tag{14}$$

where $s : [0, 1] \to \mathbb{R}_+$ is a nonnegative, twice continuously differentiable function describing the maximum amount of the good that can be sold in the second stage. Under the generalized feasibility condition, the sum of the maximum first- and second-stage consumption levels, $q_1 + s(q_1)$, may not be constant and can depend on the first-stage allocation. In particular, we focus on the following two cases:

Case 1: $q_1 + s(q_1)$ is weakly increasing in $q_1 \in [0, 1]$.

Case 2: $q_1 + s(q_1)$ is strictly decreasing in $q_1 \in [0, 1]$.

Let us illustrate the relevance of these two cases by three examples. In our first example, suppose $s(q_1) = 1$ (or any other nonnegative constant) for all $q_1 \in [0, 1]$, *i.e.*, the seller can always produce up to one unit in the second stage, regardless of the quantity sold in the first stage. Note that for this example, as the second stage's maximum quantity is always one, it does not involve the intertemporal allocation tradeoff in the main model. As a second example, suppose $s(q_1) = \beta(1 - q_1)$ with $\beta \in [0, 1]$ for all $q_1 \in [0, 1]$, *i.e.*, the unsold portion from the first stage depreciates at a constant rate β. Note that the second example covers the benchmark case imposed by (5). The above two examples fit into our Case 1. In our third example where $s(q_1) = \beta(1 - q_1)$ with $\beta > 1$, *i.e.*, the unsold portion appreciates at a constant rate, the corresponding feasibility condition fits into Case 2.

Given function s, similar to (1), let $C^s(v_2, q_1) = q_1 + s(q_1)(1 - F(v_2|q_1))$ be the corresponding expected total consumption across two stages. Furthermore, we make the following assumption analogous to Assumption 1.

Assumption 3. *For any fixed $q_1 \in (0, 1]$ and $v_2 < v_2'$,*

$$\frac{\partial C^s(v_2, q_1)}{\partial q_1} \geq 0 \quad \Longrightarrow \quad \frac{\partial C^s(v_2', q_1)}{\partial q_1} > 0,$$

$$\int_{q_1}^{1} \frac{\partial C^s(v_2, \tilde{q}_1)}{\partial \tilde{q}_1} \, \mathrm{d}\tilde{q}_1 \geq 0 \quad \Longrightarrow \quad \int_{q_1}^{1} \frac{\partial C^s(v_2', \tilde{q}_1)}{\partial \tilde{q}_1} \, \mathrm{d}\tilde{q}_1 > 0.$$

As will be clear later, the first part of the assumption generalizes Assumption 1 and leads to the monotonicity or piecewise monotonicity property of the optimal first-stage allocation when the optimum is interior. The second part guarantees that the monotonicity or quasi-concavity property holds globally despite the potential corner solution.

Following Sections 2.2 and 3, it is easy to show that the seller's revenue from a type-v_1 buyer with the first-stage consumption q_1 under the generalized feasibility condition is

$$\Pi^s(q_1, v_1) = \underbrace{q_1}_{\text{1st-stage quantity}} \cdot \underbrace{\psi(v_1)}_{\text{virtual value}} + \underbrace{s(q_1)}_{\text{remaining quantity}} \cdot \underbrace{\int_{-\psi(v_1)}^{+\infty} [\psi(v_1) + v_2] F(\mathrm{d}v_2 | q_1)}_{\text{surplus from per-unit 2nd-stage consumption}} .$$

Using a similar argument as in (8), the revenue can be expressed as

$$\underbrace{q_1 \psi(v_1)}_{\text{1st-stage surplus}} + \underbrace{s(q_1)}_{\text{remaining quantity}} \left[\underbrace{\psi(v_1)}_{\text{per-unit surplus w/o learning}} + \underbrace{\int_{-\infty}^{-\psi(v_1)} F(v_2 | q_1) \, \mathrm{d}v_2}_{\text{additional per-unit surplus w/ learning}} \right]. \tag{15}$$

For each $v_1 \in [0,1]$, denote the maximizer of $\Pi^s(q_1, v_1)$ in $q_1 \in [0,1]$ by $q_1^s(v_1)$. Define

$$p_1^s(v_1) = s(q_1^s(v_1)) \left[\int_{-\infty}^{-\psi(v_1)} F(v_2 | q_1^s(v_1)) \, \mathrm{d}v_2 - \frac{1 - G(v_1)}{g(v_1)} \right] + \int_0^{v_1} s(q_1^s(x)) F(-\psi(x) | q_1^s(x)) \, \mathrm{d}x$$

and recall that $p_2^*(v_1)$ is defined in Lemma 5. The following result characterizes the optimal contract under the generalized feasibility condition.

Proposition 3. *Under the generalized feasibility condition, Assumption 2, and Assumption 3, the optimal mechanism can be implemented by a menu of try-and-decide option contracts* $\{p_1^s(v_1), q_1^s(v_1); p_2^*(v_1)\}_{v_1 \in [0,1]}$. *In particular, when* $q_1 + s(q_1)$ *is weakly increasing in* $q_1 \in [0,1]$, $q_1^s(v_1)$ *is weakly increasing in* $v_1 \in [0,1]$; *when* $q_1 + s(q_1)$ *is strictly decreasing in* $q_1 \in [0,1]$, $q_1^s(v_1)$ *is quasi-concave in* $v_1 \in [0,1]$.

This result generalizes Proposition 1. It implies that when $q_1 + s(q_1)$ is weakly increasing in q_1, the optimal first-stage allocation is still weakly increasing in the first-stage type; however, when $q_1 + s(q_1)$ is decreasing, the optimal first-stage allocation can be increasing, of a hump-shape, or even decreasing in the first-stage type.

This is intuitive. Recall the discussion about the key tradeoff between quantity and information after Lemma 4, which leads to the monotonicity of the first-stage allocation in our main setting. Now, with generalized feasibility, we do not have the special property of the quantity effect mentioned in the beginning of this section: In (11), the first two terms cancel out. The current quantity effect is more involved, as the total quantity can depend on the first-stage consumption level. More specifically, the information effect is $s(q_1) \int_{-\infty}^{-\psi(v_1)} \frac{\partial F(v_2|q_1)}{\partial q_1} \, \mathrm{d}v$; the quantity effect is

$$\psi(v_1) + s'(q_1)\psi(v_1) + s'(q_1) \int_{-\infty}^{-\psi(v_1)} F(v_2|q_1) \, \mathrm{d}v_2,$$

where the first two terms do not cancel out and their sum depends on v_1. This further implies that the shape of $q_1^s(v_1)$ relies on $q_1 + s(q_1)$.

With generalized feasibility, for a fixed first-stage type v_1, at the optimum the positive information effect balances the negative quantity effect[21] for an interior optimum, leading to the first-order condition of the seller's revenue with respect to the quantity q_1. Then, the shape of $q_1^s(v_1)$ in v_1 follows from the comparative statics with respect to these two effects. Specifically, when $q_1 + s(q_1)$ is weakly increasing, at the optimum, the higher v_1 is, the more the tradeoff between information and quantity will be resolved towards information, so $q_1^s(v_1)$ is weakly increasing in v_1.[22] However, when $q_1 + s(q_1)$ is decreasing in q_1, depending on how steep $q_1 + s(q_1)$ is, at the optimum, the higher v_1 is, the more the tradeoff between information and quantity will be resolved towards information or quantity, so the first-stage consumption $q_1^s(v_1)$ can be increasing, of a hump-shape, or even decreasing in v_1.[23]

[21] If the quantity effect is nonnegative—*e.g.*, when $s'(q_1) > 0$ and $\psi(v_1) > 0$—then the tradeoff disappears as the information effect is positive, so the optimum is a corner solution $q_1 = 1$. Thus, for the discussion on intuitions, we only focus on the case when the quantity effect is negative.

[22] To see this, notice that for a slightly higher first-stage type $v_1 + \varepsilon$, the negative quantity effect is increased at the margin (as the quantity effect is increasing in v_1), so it gets weaker; while the positive information effect may be increased or decreased at the margin. Assumption 3 ensures that the overall effect is increased at the margin. Thus, for $v_1 + \varepsilon$, the information effect dominates the quantity effect.

[23] Different from the previous case, for a slightly higher first-stage type $v_1 + \varepsilon$, the negative quantity effect can be increased or decreased at the margin, depending on the magnitude of $(1 + s'(q_1))\psi'(v_1)$. Thus, the quantity effect can dominate (resp. be dominated by) the information effect, which then implies that $q_1^s(v_1)$ is decreasing (resp. increasing) in v_1 locally. Assumption 3 ensures that globally such a switch of dominance between these two effects occurs at most once; if it happens, it must be a switch from the pattern that information effect dominates the quantity effect to the opposite pattern. Therefore, globally, $q_1^s(v_1)$ can be increasing, decreasing, or of a hump-shape.

Finally, like in our main setting, in the proof of Proposition 3, establishing global IC still needs to use the optimality of the first-stage allocation together with the local IC condition, rather than just the latter one. Therefore, our approach of establishing IC in the main setting extends to this general setting.

7 Concluding Remarks

In this paper, we study the two-stage revenue-maximizing mechanism when the buyer acquires additional information by first-stage consumption. The buyer's decision of first-stage consumption depends on his private, prior valuation of the good. A higher first-stage consumption level leads to a more precise value estimate of the good but reduces the available amount of consumption left for the second stage. The key feature of our model is that the first-stage consumption plays a dual rule: The buyer not only enjoys a payoff but also acquires additional information from the first-stage consumption. The higher the initial consumption level is, the more additional private information the buyer acquires about the true value of the good. This yields a tradeoff between information and quantity on the seller's revenue when designing the first-stage allocation.

We fully characterize the optimum and find that it can be implemented by a menu of try-and-decide option contracts, consisting of a first-stage price-quantity pair and a second-stage per-unit price for the remaining quantity. A larger first-stage quantity is paired with a higher first-stage price but a lower second-stage per-unit price. In equilibrium, a higher first-stage valuation buyer pays more to have higher first-stage consumption and enjoys a lower second-stage price.

Since the second-stage type's distribution is not ranked by first-order stochastic dominance, we face the difficulty of the failure of the single-crossing condition when establishing global IC. The monotonicity in the first-stage consumption plus local IC is not sufficient for global IC. As such, we cannot apply the usual approach as in many dynamic mechanism design papers, which assumes FOSD, to establish global IC.

In our current analysis, we assume that the second-stage type's distribution only depends on the first-stage consumption. A more general environment is when it depends on both the first-stage consumption level and the first-stage type. This is a highly meaningful but challenging direction to explore. We leave it for future work.

References

Araujo, A. and H. Moreira (2010). Adverse selection problems without the Spence–Mirrlees condition. *Journal of Economic Theory 145*(3), 1113–1141.

Battaglini, M. (2005). Long-term contracting with Markovian consumers. *American Economic Review 95*(3), 637–658.

Bergemann, D. and J. Välimäki (2019). Dynamic mechanism design: An introduction. *Journal of Economic Literature 57*(2), 235–74.

Boleslavsky, R. and M. Said (2013). Progressive screening: Long-term contracting with a privately known stochastic process. *Review of Economic Studies 80*(1), 1–34.

Bonatti, A. (2011). Menu pricing and learning. *American Economic Journal: Microeconomics 3*(3), 124–63.

Courty, P. and H. Li (2000). Sequential screening. *The Review of Economic Studies 67*(4), 697–717.

Eső, P. and B. Szentes (2007a). Optimal information disclosure in auctions and the handicap auction. *The Review of Economic Studies 74*(3), 705–731.

Eső, P. and B. Szentes (2007b). The price of advice. *The RAND Journal of Economics 38*(4), 863–880.

Garrett, D. F. and A. Pavan (2012). Managerial turnover in a changing world. *Journal of Political Economy 120*(5), 879–925.

Guo, Y., H. Li, and X. Shi (2022). Optimal discriminatory disclosure. Working paper.

Hoffmann, F. and R. Inderst (2011). Pre-sale information. *Journal of Economic Theory 146*(6), 2333–2355.

Johnson, J. P. and D. P. Myatt (2006). On the simple economics of advertising, marketing, and product design. *American Economic Review 96*(3), 756–784.

Krähmer, D. and R. Strausz (2011). Optimal procurement contracts with pre-project planning. *The Review of Economic Studies 78*(3), 1015–1041.

Krähmer, D. and R. Strausz (2015). Ex post information rents in sequential screening. *Games and Economic Behavior 90*, 257–273.

Krähmer, D. and R. Strausz (2017). Sequential versus static screening: An equivalence result. *Games and Economic Behavior 106*, 317–328.

Lewis, T. R. and D. E. Sappington (1994). Supplying information to facilitate price discrimination. *International Economic Review*, 309–327.

Li, H. and X. Shi (2017). Discriminatory information disclosure. *American Economic Review 107*(11), 3363–85.

Li, H. and X. Shi (2022). Stochastic sequential screening. Working paper.

Liu, B. and J. Lu (2018). Pairing provision price and default remedy: optimal two-stage procurement with private R&D efficiency. *The RAND Journal of Economics 49*(3), 619–655.

Liu, B. and J. Lu (2022). Sequential screening with type-enhancing investment. *American Economic Journal: Microeconomics 14*(4), 25–57.

Lu, J. and Z. Wang (2021). Optimal selling mechanisms with buyer price search. *Journal of Economic Theory 196*, 105307.

Lu, J., L. Ye, and X. Feng (2021). Orchestrating information acquisition. *American Economic Journal: Microeconomics 13*(4), 420–65.

Milgrom, P. and I. Segal (2002). Envelope theorems for arbitrary choice sets. *Econometrica 70*(2), 583–601.

Myerson, R. B. (1986). Multistage games with communication. *Econometrica*, 323–358.

Pavan, A., I. Segal, and J. Toikka (2014). Dynamic mechanism design: A Myersonian approach. *Econometrica 82*(2), 601–653.

Schottmüller, C. (2015). Adverse selection without single crossing: Monotone solutions. *Journal of Economic Theory 158*, 127–164.

Shi, X. (2012). Optimal auctions with information acquisition. *Games and Economic Behavior 74*(2), 666–686.

A Appendix

A.1 Proof of Lemma 2

By Lemma 1, the expected payoff of the buyer with first-stage type v_1 and report r_1 can be expressed as

$$\begin{aligned}
U(v_1, r_1) =& q_1(r_1) \int_{-\infty}^{+\infty} (v_1 + v_2) F(\mathrm{d}v_2 | q_1(r_1)) \\
&+ \int_{-\infty}^{+\infty} \left[\begin{array}{c} (v_1 + v_2) q_2(r_1, r_2(v_1, r_1, v_2)) \\ -t(r_1, r_2(v_1, r_1, v_2)) \end{array} \right] F(\mathrm{d}v_2 | q_1(r_1)). \\
=& q_1(r_1) v_1 + \int_{-\infty}^{+\infty} \left[\begin{array}{c} (v_1 + v_2) q_2(r_1, r_2(v_1, r_1, v_2)) \\ -t(r_1, r_2(v_1, r_1, v_2)) \end{array} \right] F(\mathrm{d}v_2 | q_1(r_1)),
\end{aligned}$$

where the second equality uses the fact that $E[v_2] = 0$. Taking the partial derivative with respect to v_1 leads to[24]

$$\frac{\partial U(v_1, r_1)}{\partial v_1} = q_1(r_1) + \int_{-\infty}^{+\infty} \left\{ \begin{array}{c} q_2(r_1, r_2(v_1, r_1, v_2)) \\ + \frac{\partial r_2(v_1,r_1,v_2)}{\partial v_1} \left[\begin{array}{c} (v_1 + v_2)\frac{\partial q_2(r_1,r_2(v_1,r_1,v_2))}{\partial r_2} \\ - \frac{\partial t(r_1,r_2(v_1,r_1,v_2))}{\partial r_2} \end{array} \right] \end{array} \right\} F(\mathrm{d}v_2|q_1(r_1)).$$

Since $r_2(v_1, r_1, v_2)$ is the optimal report following a lie in the first stage — *i.e.*, it maximizes the second-stage expected payoff after a lie — it must satisfy the first-order condition, so

$$(v_1 + v_2)\frac{\partial q_2(r_1, r_2(v_1, r_1, v_2))}{\partial r_2} - \frac{\partial t(r_1, r_2(v_1, r_1, v_2))}{\partial r_2} = 0.$$

Therefore,

$$\frac{\partial U(v_1, r_1)}{\partial v_1} = q_1(r_1) + \int_{-\infty}^{+\infty} q_2(r_1, r_2(v_1, r_1, v_2))F(\mathrm{d}v_2|q_1(r_1)).$$

Note that $r_2(v_1, r_1, v_2) = v_2$ when $r_1 = v_1$ (*i.e.*, a truthful report in the first stage). The envelope theorem (cf. Milgrom and Segal, 2002) implies that

$$U(v_1, v_1) = U(0, 0) + \int_0^{v_1} \left[q_1(s) + \int_{-\infty}^{+\infty} q_2(s, v_2)F(\mathrm{d}v_2|q_1(s)) \right] \mathrm{d}s.$$

This completes the proof. □

A.2 Derivation of Equation (8)

Before proving Equation (8), we first prove the following preparatory lemma.

Lemma 7. *For any $a \in (-\infty, +\infty)$ and $q_1 > 0$,*

$$\int_{-\infty}^{a} v_2 F(\mathrm{d}v_2|q_1) = aF(a|q_1) - \int_{-\infty}^{a} F(v_2|q_1)\,\mathrm{d}v_2.$$

Proof. We first claim that as $b \to -\infty$, $bF(b|q_1) \to 0$. Suppose that the claim is not true. Then there exists some $\epsilon > 0$ and a negative sequence $\{b_k\}$ that converges to $-\infty$ such that $|b_k|F(b_k|q_1) > \epsilon$ for any k. Since the integral $\int_{-\infty}^{+\infty} v_2 F(\mathrm{d}v_2|q_1)$ is well defined, there exists some sufficiently large K such that for any $k \geq K$, $\int_{-\infty}^{b_k} |v_2|F(\mathrm{d}v_2|q_1) < \epsilon$. It implies that

$$|b_k|F(b_k|q_1) = \int_{-\infty}^{b_k} |b_k|F(\mathrm{d}v_2|q_1) \leq \int_{-\infty}^{b_k} |v_2|F(\mathrm{d}v_2|q_1) < \epsilon,$$

[24]The (almost everywhere) differentiability of q_2 and t in r_2 follows from the second-stage IC constraint.

which is a contradiction.

For any $b < \min\{a, 0\}$, due to integral by parts,

$$\int_b^a v_2 F(\mathrm{d}v_2|q_1) = aF(a|q_1) - bF(b|q_1) - \int_b^a F(v_2|q_1)\,\mathrm{d}v_2.$$

It implies that

$$\int_b^a F(v_2|q_1)\,\mathrm{d}v_2 = aF(a|q_1) - bF(b|q_1) - \int_b^a v_2 F(\mathrm{d}v_2|q_1).$$

It is obvious that $\int_b^a F(v_2|q_1)\,\mathrm{d}v_2$ is decreasing in b. In addition, it is bounded as $bF(b|q_1) \to 0$ and $\int_b^a v_2 F(\mathrm{d}v_2|q_1) \to \int_{-\infty}^a v_2 F(\mathrm{d}v_2|q_1)$ when $b \to -\infty$. Thus, the limit $\lim_{b\to-\infty} \int_b^a F(v_2|q_1)\,\mathrm{d}v_2$ exists, which is $\int_{-\infty}^a F(v_2|q_1)\,\mathrm{d}v_2$. This completes the proof. □

Now we derive Equation (8).

$$\begin{aligned}
&q_1(v_1)\psi(v_1) + (1-q_1(v_1)) \int_{-\psi(v_1)}^{+\infty} [\psi(v_1) + v_2] F(\mathrm{d}v_2|q_1(v_1)) \\
=&q_1(v_1)\psi(v_1) + (1-q_1(v_1)) \left[\psi(v_1)[1 - F(-\psi(v_1)|q_1(v_1))] + \int_{-\psi(v_1)}^{+\infty} v_2 F(\mathrm{d}v_2|q_1(v_1))\right] \\
=&\psi(v_1) - (1-q_1(v_1)) \left[\psi(v_1)F(-\psi(v_1)|q_1(v_1)) + \int_{-\infty}^{-\psi(v_1)} v_2 F(\mathrm{d}v_2|q_1(v_1))\right] \\
=&\psi(v_1) - (1-q_1(v_1)) \left[\begin{array}{c} \psi(v_1)F(-\psi(v_1)|q_1(v_1)) \\ +\left(-\psi(v_1)F(-\psi(v_1)|q_1(v_1)) - \int_{-\infty}^{-\psi(v_1)} F(v_2|q_1(v_1))\,\mathrm{d}v_2\right) \end{array}\right] \\
=&\psi(v_1) + (1-q_1(v_1)) \int_{-\infty}^{-\psi(v_1)} F(v_2|q_1(v_1))\,\mathrm{d}v_2,
\end{aligned}$$

where the second equality uses the fact that

$$\int_{-\psi(v_1)}^{+\infty} v_2 F(\mathrm{d}v_2|q_1(v_1)) + \int_{-\infty}^{-\psi(v_1)} v_2 F(\mathrm{d}v_2|q_1(v_1)) = 0,$$

and the third equality holds because of Lemma 7. □

A.3 Proof of Lemma 3

We first establish (i). Since the optimal $q_1^*(\cdot)$ maximizes (9), the solution either satisfies the first-order condition or is the corner solution.

When $q_1 > 0$,

$$\frac{\partial \Pi(q_1, v_1)}{\partial q_1} = -\int_{-\infty}^{-\psi(v_1)} F(v_2|q_1)\,\mathrm{d}v_2 + (1-q_1)\int_{-\infty}^{-\psi(v_1)} \frac{\partial F(v_2|q_1)}{\partial q_1}\,\mathrm{d}v_2.$$

We first show that $q_1 = 1$ cannot be optimal. In fact,

$$\frac{\partial \Pi(q_1, v_1)}{\partial q_1}|_{q_1=1} = -\int_{-\infty}^{-\psi(v_1)} F(v_2|1)\,\mathrm{d}v_2 < 0,$$

which makes $q_1 = 1$ suboptimal.

On the other hand, if $q_1 = 0$ is optimal, it must be the case that $\psi(v_1) < 0$ — *i.e.*, $v_1 < v_1^*$. In fact, when $\psi(v_1) \geq 0$ and $q_1 = 0$, the seller's revenue is

$$\Pi(0, v_1) = \psi(v_1) + \int_{-\infty}^{-\psi(v_1)} F(v_2|0)\,\mathrm{d}v_2 = \psi(v_1),$$

which is strictly dominated by, for example, choosing $q_1 = \frac{1}{2}$:

$$\Pi(\frac{1}{2}, v_1) = \psi(v_1) + \frac{1}{2}\int_{-\infty}^{-\psi(v_1)} F(v_2|\frac{1}{2})\,\mathrm{d}v_2 > \psi(v_1).$$

This means that the value $\tilde{v}_1 \equiv \inf\{v_1 \in [0,1] : q_1^*(v_1) > 0\} < v_1^*$.

Finally, we show that if $q_1^*(v_1) = 0$ for some v_1, then $q_1^*(v_1') = 0$ for any $v_1' < v_1$. We have shown that if $q_1^*(v_1) = 0$, then $v_1 < v_1^*$, and thus

$$\Pi(0, v_1) = \psi(v_1) + \int_{-\infty}^{-\psi(v_1)} F(v_2|0)\,\mathrm{d}v_2 = \psi(v_1) - \psi(v_1) = 0.$$

In addition, $q_1^*(v_1) = 0$ implies that $\Pi(q_1, v_1) \leq \Pi(0, v_1) = 0$ for all $q_1 \in (0,1)$. When $v_1' < v_1$, $\psi(v_1') < \psi(v_1) < 0$, and thus for any $q_1 \in (0,1)$, we have

$$\begin{aligned}
\Pi(q_1, v_1') &= \psi(v_1') + (1-q_1)\int_{-\infty}^{-\psi(v_1')} F(v_2|q_1)\,\mathrm{d}v_2 \\
&= \psi(v_1') + (1-q_1)\left[\int_{-\infty}^{-\psi(v_1)} F(v_2|q_1)\,\mathrm{d}v_2 + \int_{-\psi(v_1)}^{-\psi(v_1')} \underbrace{F(v_2|q_1)}_{<1}\,\mathrm{d}v_2\right] \\
&< \psi(v_1') + (1-q_1)\left[\int_{-\infty}^{-\psi(v_1)} F(v_2|q_1)\,\mathrm{d}v_2 + \psi(v_1) - \psi(v_1')\right] \\
&< \psi(v_1') + (1-q_1)\int_{-\infty}^{-\psi(v_1)} F(v_2|q_1)\,\mathrm{d}v_2 + \psi(v_1) - \psi(v_1') \\
&= \Pi(q_1, v_1) \leq \Pi(0, v_1) = 0 = \Pi(0, v_1'),
\end{aligned}$$

implying that $q_1^*(v_1') = 0$.

As a result, for $v_1 < \tilde{v}_1$, $q_1^*(v_1) = 0$; for $v_1 \geq \tilde{v}_1$, $q_1^*(v_1) \in [0,1)$ and satisfies the first-order condition stated in the lemma; for $v_1 > \tilde{v}_1$, $q_1^*(v_1) \in (0,1)$.

To establish (ii), notice that for $v_1 < \tilde{v}_1$, $F(\cdot|q_1^*(v_1))$ reduces to a mass at $v_2 = 0$. In this case, $v_2 = 0 \geq -\psi(v_1)$ is impossible, as $-\psi(v_1) > -\psi(\tilde{v}_1) > -\psi(v_1^*) = 0$. Therefore, $q_2^*(v_1, 0) = 0$. The remainder of (ii) has been established in the text. $\square$

A.4 Proof of Lemma 4

Fix any $v_1 > \tilde{v}_1$. The problem is to choose $q_1 \in [0,1]$ to maximize

$$\Pi(q_1, v_1) = \psi(v_1) + (1 - q_1) \int_{-\infty}^{-\psi(v_1)} F(v_2|q_1)\,\mathrm{d}v_2. \tag{16}$$

Note that restricting the range of q_1 to $(0,1)$ is without loss of generality, because (i) by the definition of $\tilde{v}_1$, $q_1 = 0$ cannot be optimal; and (ii) by Lemma 3, $q_1 = 1$ is not optimal either.

Define a function $\xi : \mathbb{R} \times (0,1) \to \mathbb{R}$ as

$$\begin{aligned} \xi(v_2, q_1) &= M(v_2, q_1) \cdot (1 - q_1) \cdot f(v_2|q_1) \\ &= -F(v_2|q_1) + (1 - q_1) \frac{\partial F(v_2|q_1)}{\partial q_1}. \end{aligned} \tag{17}$$

Since the maximizer $q_1^*(v_1) \in (0,1)$, it satisfies the first-order condition of (16) with respect to q_1:

$$\frac{\partial \Pi(q_1, v_1)}{\partial q_1}|_{q_1 = q_1^*(v_1)} = \int_{-\infty}^{-\psi(v_1)} \xi(v_2, q_1^*(v_1))\,\mathrm{d}v_2 = 0. \tag{18}$$

By the second-order condition, $\frac{\partial^2 \Pi(q_1, v_1)}{\partial q_1^2}|_{q_1 = q_1^*(v_1)} \leq 0$.

Note that

$$\xi(-\psi(v_1), q_1^*(v_1)) < 0, \tag{19}$$

as otherwise $\xi(v_2, q_1^*(v_1)) > 0$ for any $v_2 < -\psi(v_1)$ by Assumption 1, which violates (18). It implies that

$$\frac{\partial^2 \Pi(q_1, v_1)}{\partial q_1 \partial v_1}|_{q_1 = q_1^*(v_1)} = -\psi'(v_1)\xi(-\psi(v_1), q_1^*(v_1)) > 0. \tag{20}$$

Now differentiating the first-order condition (18) with respect to v_1 on both sides of the equation leads to

$$\frac{\mathrm{d}q_1^*(v_1)}{\mathrm{d}v_1} \cdot \underbrace{\frac{\partial^2 \Pi(q_1, v_1)}{\partial q_1^2}|_{q_1 = q_1^*(v_1)}}_{\leq 0} + \underbrace{\frac{\partial^2 \Pi(q_1, v_1)}{\partial q_1 \partial v_1}|_{q_1 = q_1^*(v_1)}}_{>0} = 0,$$

which further implies that $\frac{\mathrm{d}q_1^*(v_1)}{\mathrm{d}v_1} > 0$. $\square$

A.5 Proof of the claim in Remark 2

Pick any $v_1, v_1' \in [\tilde{v}_1, 1]$ with $v_1 < v_1'$. Let q_1^* and $q_1^{*\prime}$ be a maximizer of $\Pi(q_1, v_1)$ and $\Pi(q_1, v_1')$, respectively. By Lemma 3, $q_1^* \in [0,1)$ (since v_1 may be $\tilde{v}_1$) and $q_1^{*\prime} \in (0,1)$. Our goal is to show that $q_1^* < q_1^{*\prime}$. Hence, it is without loss to focus on the case that $q_1^*, q_1^{*\prime} \in (0,1)$.

Note that (20) still applies because it only uses Assumption 1. Hence,

$$\xi(-\psi(v_1), q_1^*) < 0 \quad \text{and} \quad \xi(-\psi(v_1'), q_1^{*\prime}) < 0.$$

Suppose to the contrary that $q_1^* \geq q_1^{*\prime}$. We claim that

$$\xi(-\psi(s), q_1) < 0 \text{ for any } s \in [v_1, v_1'] \text{ and any } q_1 \in [q_1^{*\prime}, q_1^*]. \tag{21}$$

To see this, recall that $F(\cdot|\cdot)$ satisfies the rotation order. If $-\psi(s) > 0$, then $\xi(-\psi(s), q_1) < 0$ for any $q_1 \in [q_1^{*\prime}, q_1^*]$.

- If $-\psi(v_1') > 0$, then $-\psi(s) > 0$ for any $s \in [v_1, v_1']$, so (21) holds.
- Suppose that $-\psi(v_1') \leq 0$. Since $\xi(-\psi(v_1'), q_1^{*\prime}) < 0$, the additional condition mentioned in Remark 2 implies that $\xi(-\psi(v_1'), q_1) < 0$ for any $q_1 \in [q_1^{*\prime}, q_1^*]$. Then due to Assumption 1 and the monotonicity of ψ, $\xi(-\psi(s), q_1) < 0$ for any $q_1 \in [q_1^{*\prime}, q_1^*]$ and $s \in [v_1, v_1']$. This establishes (21).

Due to the definition of q_1^*, $\Pi(q_1^*, v_1) \geq \Pi(q_1^{*\prime}, v_1)$. If $q_1^* > q_1^{*\prime}$, then

$$\begin{aligned}
0 \;\leq\; \Pi(q_1^*, v_1) - \Pi(q_1^{*\prime}, v_1) &= \int_{q_1^{*\prime}}^{q_1^*} \frac{\partial \Pi(q_1, v_1)}{\partial q_1} \, \mathrm{d}q_1 \\
&= \int_{q_1^{*\prime}}^{q_1^*} \left(\int_{-\infty}^{-\psi(v_1)} \xi(v_2, q_1) \, \mathrm{d}v_2 \right) \mathrm{d}q_1 \\
&= \int_{q_1^{*\prime}}^{q_1^*} \left(\int_{-\infty}^{-\psi(v_1')} \xi(v_2, q_1) \, \mathrm{d}v_2 + \int_{-\psi(v_1')}^{-\psi(v_1)} \underbrace{\xi(v_2, q_1)}_{<0 \text{ by } (21)} \mathrm{d}v_2 \right) \mathrm{d}q_1 \\
&< \int_{q_1^{*\prime}}^{q_1^*} \left(\int_{-\infty}^{-\psi(v_1')} \xi(v_2, q_1) \, \mathrm{d}v_2 \right) \mathrm{d}q_1 \\
&= \int_{q_1^{*\prime}}^{q_1^*} \frac{\partial \Pi(q_1, v_1')}{\partial q_1} \, \mathrm{d}q_1 = \Pi(q_1^*, v_1') - \Pi(q_1^{*\prime}, v_1').
\end{aligned}$$

However, $\Pi(q_1^*, v_1') - \Pi(q_1^{*\prime}, v_1') > 0$ contradicts the optimality of $q_1^{*\prime}$.

Finally, what is left to show is that $q_1^* = q_1^{*\prime}$ also leads to a contradiction. In fact, if $q_1^* = q_1^{*\prime}$, the first-order condition (18) implies that

$$\int_{-\infty}^{-\psi(v_1)} \xi(v_2, q_1^*) \, \mathrm{d}v_2 = 0 = \int_{-\infty}^{-\psi(v_1')} \xi(v_2, q_1^*) \, \mathrm{d}v_2.$$

However, (21) implies that $\xi(v_2, q_1^*) < 0$ for any $v_2 \in [-\psi(v_1'), -\psi(v_1)]$. Then,

$$0 \;=\; \int_{-\infty}^{-\psi(v_1)} \xi(v_2, q_1^*) \, \mathrm{d}v_2 = \underbrace{\int_{-\infty}^{-\psi(v_1')} \xi(v_2, q_1^*) \, \mathrm{d}v_2}_{=0} + \int_{-\psi(v_1')}^{-\psi(v_1)} \xi(v_2, q_1^*) \, \mathrm{d}v_2$$

$$= \int_{-\psi(v_1')}^{-\psi(v_1)} \xi(v_2, q_1^*)\,\mathrm{d}v_2 < 0,$$

which is a contradiction. This completes the proof of the claim in Remark 2. $\square$

A.6 Proof of Lemma 5

We first construct the payment rule t^*. By Lemmas 2 and 3 and $U(0,0) = 0$, for each $v_1 \in [0,1]$, we have

$$\begin{aligned}
U(v_1, v_1) &= U(0,0) + \int_0^{v_1} \left[q_1^*(x) + \int_{-\infty}^{+\infty} q_2^*(x, v_2) F(\mathrm{d}v_2 | q_1^*(x)) \right] \mathrm{d}x \\
&= \int_0^{v_1} \left[q_1^*(x) + \int_{-\psi(x)}^{+\infty} (1 - q_1^*(x)) F(\mathrm{d}v_2 | q_1^*(x)) \right] \mathrm{d}x \\
&= \int_0^{v_1} \left[q_1^*(x) + (1 - q_1^*(x)) \left(1 - F(-\psi(x) | q_1^*(x))\right) \right] \mathrm{d}x \\
&= \int_0^{v_1} \left[1 - (1 - q_1^*(x)) F(-\psi(x) | q_1^*(x)) \right] \mathrm{d}x \\
&= v_1 - \int_0^{v_1} \big(1 - q_1^*(x)\big) F(-\psi(x) | q_1^*(x))\,\mathrm{d}x. \qquad (22)
\end{aligned}$$

On the other hand, from the envelope condition (2) in the second stage,

$$\begin{aligned}
(v_1 + v_2) q_2^*(v_1, v_2) - t^*(v_1, v_2) &= \tilde{\pi}(v_1, v_2, v_2) \\
&= \tilde{\pi}(v_1, -\psi(v_1), -\psi(v_1)) + \int_{-\psi(v_1)}^{v_2} q_2^*(v_1, s)\,\mathrm{d}s.
\end{aligned}$$

Thus,

$$t^*(v_1, v_2) = (v_1 + v_2) q_2^*(v_1, v_2) - \int_{-\psi(v_1)}^{v_2} q_2^*(v_1, s)\,\mathrm{d}s - \tilde{\pi}(v_1, -\psi(v_1), -\psi(v_1)).$$

- When $v_2 < -\psi(v_1)$, $q_2^*(v_1, v_2) = 0$ and

$$t^*(v_1, v_2) = -\tilde{\pi}(v_1, -\psi(v_1), -\psi(v_1)).$$

- When $v_2 \geq -\psi(v_1)$ and $v_1 \geq \tilde{v}_1$, $q_2^*(v_1, v_2) = 1 - q_1^*(v_1)$ and

$$\begin{aligned}
t^*(v_1, v_2) &= (v_1 + v_2)(1 - q_1^*(v_1)) - \int_{-\psi(v_1)}^{v_2} (1 - q_1^*(v_1))\,\mathrm{d}s - \tilde{\pi}(v_1, -\psi(v_1), -\psi(v_1)) \\
&= \frac{1 - G(v_1)}{g(v_1)}(1 - q_1^*(v_1)) - \tilde{\pi}(v_1, -\psi(v_1), -\psi(v_1)).
\end{aligned}$$

- When $v_1 < \tilde{v}_1$, $-\psi(v_1) > 0$, $q_1^*(v_1) = 0$, and $F(\cdot | q_1^*(v_1))$ reduces to a mass at 0, implying that $v_2 \geq -\psi(v_1) > 0$ is impossible. As a result, with probability one

$v_2 < -\psi(v_1)$, which implies that

$$q_2^*(v_1, v_2) = 0 \qquad \text{and} \qquad t^*(v_1, v_2) = -\tilde{\pi}(v_1, -\psi(v_1), -\psi(v_1)).$$

To construct the payment rule t^*, it remains to pin down $\tilde{\pi}(v_1, -\psi(v_1), -\psi(v_1))$. To this end, notice that by the definition of the first-stage expected payoff,

$$\begin{aligned}
U(v_1, v_1) &= q_1^*(v_1) \int_{-\infty}^{+\infty} (v_1 + v_2) F(\mathrm{d}v_2 | q_1^*(v_1)) + \int_{-\infty}^{+\infty} \tilde{\pi}(v_1, v_2, v_2) F(\mathrm{d}v_2 | q_1^*(v_1)) \\
&= q_1^*(v_1) v_1 + \int_{-\infty}^{+\infty} \Big[(v_1 + v_2) q_2^*(v_1, v_2) - t^*(v_1, v_2) \Big] F(\mathrm{d}v_2 | q_1^*(v_1)) \\
&= q_1^*(v_1) v_1 + \int_{-\psi(v_1)}^{+\infty} (1 - q_1^*(v_1)) \left[v_1 + v_2 - \frac{1 - G(v_1)}{g(v_1)} \right] F(\mathrm{d}v_2 | q_1^*(v_1)) \\
&\quad + \tilde{\pi}(v_1, -\psi(v_1), -\psi(v_1)) \\
&= q_1^*(v_1) v_1 + \int_{-\psi(v_1)}^{+\infty} (1 - q_1^*(v_1)) \left(\psi(v_1) + v_2 \right) F(\mathrm{d}v_2 | q_1^*(v_1)) \\
&\quad + \tilde{\pi}(v_1, -\psi(v_1), -\psi(v_1)).
\end{aligned}$$

Comparing with (22), we obtain

$$\begin{aligned}
& - \tilde{\pi}(v_1, -\psi(v_1), -\psi(v_1)) \\
=& q_1^*(v_1) v_1 + \int_{-\psi(v_1)}^{+\infty} (1 - q_1^*(v_1)) [\psi(v_1) + v_2] F(\mathrm{d}v_2 | q_1^*(v_1)) \\
& - \left[v_1 - \int_0^{v_1} (1 - q_1^*(x)) F(-\psi(x) | q_1^*(x)) \, \mathrm{d}x \right] \\
=& \int_{-\psi(v_1)}^{+\infty} (1 - q_1^*(v_1)) [\psi(v_1) + v_2] F(\mathrm{d}v_2 | q_1^*(v_1)) - v_1 (1 - q_1^*(v_1)) \\
& + \int_0^{v_1} (1 - q_1^*(x)) F(-\psi(x) | q_1^*(x)) \, \mathrm{d}x \\
=& (1 - q_1^*(v_1)) \left\{ \begin{array}{c} [1 - F(-\psi(v_1) | q_1^*(v_1))] \psi(v_1) - v_1 \\ + \int_{-\psi(v_1)}^{+\infty} v_2 F(\mathrm{d}v_2 | q_1^*(v_1)) \end{array} \right\} \\
& + \int_0^{v_1} (1 - q_1^*(x)) F(-\psi(x) | q_1^*(x)) \, \mathrm{d}x \\
\overset{E[v_2]=0}{=}& (1 - q_1^*(v_1)) \left\{ \begin{array}{c} [1 - F(-\psi(v_1) | q_1^*(v_1))] \psi(v_1) - v_1 \\ - \int_{-\infty}^{-\psi(v_1)} v_2 F(\mathrm{d}v_2 | q_1^*(v_1)) \end{array} \right\} \\
& + \int_0^{v_1} (1 - q_1^*(x)) F(-\psi(x) | q_1^*(x)) \, \mathrm{d}x \\
\overset{\text{Lemma } 7}{=}& (1 - q_1^*(v_1)) \left\{ \begin{array}{c} [1 - F(-\psi(v_1) | q_1^*(v_1))] \psi(v_1) - v_1 \\ + \psi(v_1) F(-\psi(v_1) | q_1^*(v_1)) + \int_{-\infty}^{-\psi(v_1)} F(v_2 | q_1^*(v_1)) \, \mathrm{d}v_2 \end{array} \right\} \\
& + \int_0^{v_1} (1 - q_1^*(x)) F(-\psi(x) | q_1^*(x)) \, \mathrm{d}x
\end{aligned}$$

$$=(1-q_1^*(v_1))\Big[\int_{-\infty}^{-\psi(v_1)} F(v_2|q_1^*(v_1))\,\mathrm{d}v_2 - \frac{1-G(v_1)}{g(v_1)}\Big] + \int_0^{v_1} (1-q_1^*(x))F(-\psi(x)|q_1^*(x))\,\mathrm{d}x.$$

Thus,

$$t^*(v_1,v_2) = \begin{cases} (1-q_1^*(v_1))p_2^*(v_1) + p_1^*(v_1), & \text{if } \psi(v_1)+v_2 \geq 0,\\ p_1^*(v_1), & \text{otherwise,} \end{cases}$$

where

$$p_1^*(v_1) = -\tilde{\pi}(v_1,-\psi(v_1),-\psi(v_1)) = (1-q_1^*(v_1))[\int_{-\infty}^{-\psi(v_1)} F(v_2|q_1^*(v_1))\,\mathrm{d}v_2 - \frac{1-G(v_1)}{g(v_1)}] + \int_0^{v_1} (1-q_1^*(x))F(-\psi(x)|q_1^*(x))\,\mathrm{d}x$$

and $p_2^*(v_1) = \frac{1-G(v_1)}{g(v_1)}$. □

A.7 Proof of Proposition 1

Let us first outline the roadmap of the proof. Fix any first-state type v_1. We first characterize the difference between his payoff under truth-telling and his payoff when he instead chooses the contract $\{p_1^*(r_1), q_1^*(r_1); p_2^*(r_1)\}$. We further express this difference in terms of q_1, p_1, and p_2 with a double-integral form. If we can show that this difference is always nonnegative, then the proof completes. However, as we discussed after Proposition 1 and Section 5, the standard approach in the dynamic mechanism design literature—which uses the allocation's monotonicity plus local IC to establish global IC—does not apply here, because types are not ranked by FOSD in our setting so the integrand of the double integral does not have a constant sign. We overcome this difficulty by explicitly using the optimality of $q_1^*(v_1)$ to show that type-v_1 buyer has no incentive to deviate.

Now let us start the proof. We need to show that the buyer has the incentive to follow the "recommendation" that for each $v_1 \in [0,1]$, (i) if type-v_1 buyer chooses the option contract $\{p_1^*(v_1), q_1^*(v_1); p_2^*(v_1)\}$, he should buy the remaining $1-q_1^*(v_1)$ portion in the second stage if and only if $v_1+v_2 \geq p_2^*(v_1)$; (ii) type-v_1 buyer should find it optimal to choose the option contract $\{p_1^*(v_1), q_1^*(v_1); p_2^*(v_1)\}$. The verification of (i) is trivial. Thus, we only need to establish (ii) in this proof.

Define a three-variable function

$$w(q_1,p_2,v_1) = q_1 v_1 + \int_{p_2-v_1}^{+\infty} (v_1+v_2-p_2)(1-q_1)F(\mathrm{d}v_2|q_1).$$

If type-v_1 buyer chooses the contract $\{p_1^*(r_1), q_1^*(r_1); p_2^*(r_1)\}$ for some r_1, he will buy the remaining $1-q_1^*(r_1)$ portion in the second stage if and only if $v_1+v_2 \geq p_2^*(r_1)$; that is, $v_2 \geq -v_1 + p_2^*(r_1)$. Hence, his expected utility when selecting $\{p_1^*(r_1), q_1^*(r_1); p_2^*(r_1)\}$ and following the optimal second-stage strategy is given by

$$U(v_1, r_1) = w(q_1^*(r_1), p_2^*(r_1), v_1) - p_1^*(r_1).$$

Our goal is to show that $\Delta(v_1, r_1) \equiv U(v_1, v_1) - U(v_1, r_1) \geq 0$, for any $v_1, r_1 \in [0, 1]$.

To this end, notice that the partial derivative of w with respect to the third variable,

$$w_3(q_1, p_2, v_1) = 1 - (1 - q_1)F(p_2 - v_1|q_1) \geq 0. \tag{23}$$

By the construction of option contracts, it is easy to verify that when the type-v_1 buyer chooses the contract $\{p_1^*(v_1), q_1^*(v_1); p_2^*(v_1)\}$, his expected utility $U(v_1, v_1)$ can be expressed as the form in (22). Therefore, we have

$$\frac{\mathrm{d}U(v_1, v_1)}{\mathrm{d}v_1} = 1 - (1 - q_1^*(v_1))F(-\psi(v_1)|q_1^*(v_1)) \overset{(23)}{=} w_3(q_1^*(v_1), p_2^*(v_1), v_1). \tag{24}$$

For any v_1 and $r_1 \in [0, 1]$, $\Delta(v_1, r_1) \equiv U(v_1, v_1) - U(v_1, r_1)$ is further equal to

$$\begin{aligned}
&U(v_1, v_1) - U(r_1, r_1) + U(r_1, r_1) - U(v_1, r_1) \\
=&U(v_1, v_1) - U(r_1, r_1) + [w(q_1^*(r_1), p_2^*(r_1), r_1) - p_1^*(r_1)] - [w(q_1^*(r_1), p_2^*(r_1), v_1) - p_1^*(r_1)] \\
\overset{(24)}{=}& \int_{r_1}^{v_1} w_3(q_1^*(s), p_2^*(s), s)\,\mathrm{d}s + w(q_1^*(r_1), p_2^*(r_1), r_1) - w(q_1^*(r_1), p_2^*(r_1), v_1) \\
=& \int_{r_1}^{v_1} [w_3(q_1^*(s), p_2^*(s), s) - w_3(q_1^*(r_1), p_2^*(r_1), s)]\,\mathrm{d}s \\
=& \int_{r_1}^{v_1} \int_{r_1}^{s} \left[w_{31}(q_1^*(x), p_2^*(x), s)q_1^{*\prime}(x) + w_{32}(q_1^*(x), p_2^*(x), s)p_2^{*\prime}(x)\right] \mathrm{d}x\,\mathrm{d}s \\
=& \int_{r_1}^{v_1} \int_{x}^{v_1} \left[w_{31}(q_1^*(x), p_2^*(x), s)q_1^{*\prime}(x) + w_{32}(q_1^*(x), p_2^*(x), s)p_2^{*\prime}(x)\right] \mathrm{d}s\,\mathrm{d}x,
\end{aligned}$$

where the third and the fourth equalities follow from the fundamental theorem of calculus and the last equality interchanges the order of integration.

Since $p_2^{*\prime}(x) \leq 0$ and $w_{32}(q_1^*(x), p_2^*(x), s) \leq 0$ for all x, $s \in [0, 1]$, to establish $\Delta(v_1, r_1) \geq 0$, it suffices to show that

$$\int_{r_1}^{v_1} \int_{x}^{v_1} w_{31}(q_1^*(x), p_2^*(x), s)q_1^{*\prime}(x)\,\mathrm{d}s\,\mathrm{d}x \geq 0.$$

Notice that

$$\begin{aligned}
\int_{x}^{v_1} w_{31}(q_1^*(x), p_2^*(x), s)\,\mathrm{d}s &\overset{(23)}{=} \int_{x}^{v_1} \frac{\partial[1 - (1 - q_1^*(x))F(p_2^*(x) - s|q_1^*(x))]}{\partial q_1^*(x)}\,\mathrm{d}s \\
&\overset{(17)}{=} - \int_{x}^{v_1} \xi(p_2^*(x) - s, q_1^*(x))\,\mathrm{d}s \\
&= \int_{-\psi(x)}^{x - v_1 - \psi(x)} \xi(y, q_1^*(x))\,\mathrm{d}y,
\end{aligned}$$

where the last equality follows the definition of $p_2^*(x)$ and the change of variable $y = \frac{1-G(x)}{g(x)} - s = x - s - \psi(x)$. Hence, to show $\Delta(v_1, r_1) \geq 0$, it suffices to show

$$\int_{r_1}^{v_1} {q_1^*}'(x) \int_{-\psi(x)}^{x-v_1-\psi(x)} \xi(y, q_1^*(x))\,\mathrm{d}y\,\mathrm{d}x \geq 0. \tag{25}$$

Now we discuss two cases and show that in both cases expression (25) holds; therefore $\Delta(v_1, r_1) \geq 0$.

Case 1: for $r_1 \geq v_1$ and $x \in [v_1, r_1]$, we must have $-\psi(x) \leq x - v_1 - \psi(x)$.

If x is such that $q_1^*(x) = 0$, it has been established that ${q_1^*}'(x) = 0$. Thus,

$${q_1^*}'(x) \int_{-\psi(x)}^{x-v_1-\psi(x)} \xi(y, q_1^*(x))\,\mathrm{d}y = 0.$$

If x is such that $q_1^*(x) > 0$, by (20), $\xi(-\psi(x), q_1^*(x)) < 0$. Assumption 1 implies that $\xi(y, q_1^*(x)) < 0$ for any $y \geq -\psi(x)$. Also, notice that ${q_1^*}'(x) \geq 0$. Thus,

$${q_1^*}'(x) \int_{-\psi(x)}^{x-v_1-\psi(x)} \xi(y, q_1^*(x))\,\mathrm{d}y \leq 0.$$

Since $r_1 \geq v_1$, expression (25) holds.

Case 2: for $r_1 \leq v_1$ and $x \in [r_1, v_1]$, we have $-\psi(x) \geq x - v_1 - \psi(x)$.

If x is such that $q_1^*(x) = 0$, again,

$${q_1^*}'(x) \int_{-\psi(x)}^{x-v_1-\psi(x)} \xi(y, q_1^*(x))\,\mathrm{d}y = 0.$$

If x is such that $q_1^*(x) > 0$ and $\xi(x - v_1 - \psi(x), q_1^*(x)) \leq 0$, then Assumption 1 implies that $\xi(y, q_1^*(x)) \leq 0$ for any $y \geq x - v_1 - \psi(x)$, which further implies that

$${q_1^*}'(x) \int_{-\psi(x)}^{x-v_1-\psi(x)} \xi(y, q_1^*(x))\,\mathrm{d}y \geq 0.$$

If x is such that $q_1^*(x) > 0$ and $\xi(x - v_1 - \psi(x), q_1^*(x)) > 0$, then Assumption 1 implies that $\xi(y, q_1^*(x)) \geq 0$ for any $y \leq x - v_1 - \psi(x)$. Since $q_1^*(x) > 0$, from (18), the optimality of $q_1^*(x)$ requires $\int_{-\infty}^{-\psi(x)} \xi(y, q_1^*(x))\,\mathrm{d}y = 0$. It then follows that

$$0 = \int_{-\infty}^{-\psi(x)} \xi(y, q_1^*(x))\,\mathrm{d}y = \underbrace{\int_{-\infty}^{x-v_1-\psi(x)} \xi(y, q_1^*(x))\,\mathrm{d}y}_{\geq 0} + \int_{x-v_1-\psi(x)}^{-\psi(x)} \xi(y, q_1^*(x))\,\mathrm{d}y,$$

which implies that $\int_{x-v_1-\psi(x)}^{-\psi(x)} \xi(y, q_1^*(x))\,\mathrm{d}y \le 0$, *i.e.*,

$$q_1^{*\prime}(x) \int_{-\psi(x)}^{x-v_1-\psi(x)} \xi(y, q_1^*(x))\,\mathrm{d}y \ge 0.$$

Expression (25) holds again, since $r_1 \le v_1$.

In both Cases 1 and 2, we conclude that $\Delta(v_1, r_1) \ge 0$. This completes the proof. $\square$

A.8 Proof of Lemma 6

For (i), notice that for $v_1 < \tilde{v}_1$, $q_1^*(v_1) = 0$. In this case, $F(\cdot|0)$ degenerates to a mass at 0 and with probability one $\psi(v_1) + v_2 = \psi(v_1) < 0$, where the inequality follows from Lemma 3. As a result,

$$\begin{aligned} p_1^*(v_1) &= \int_{-\infty}^{-\psi(v_1)} F(v_2|0)\,\mathrm{d}v_2 - \frac{1-G(v_1)}{g(v_1)} + \int_0^{v_1} F(-\psi(x)|0)\,\mathrm{d}x \\ &= -\psi(v_1) - \frac{1-G(v_1)}{g(v_1)} + v_1 \\ &= 0. \end{aligned} \tag{26}$$

By Lemma 3, there is no consumption in both stages when $v_1 \in [0, \tilde{v}_1)$. Then by Lemma 5, when $v_1 < \tilde{v}_1$, for all v_2,

$$t^*(v_1, v_2) = p_1^*(v_1) = 0.$$

For (ii), when $v_1 \in [\tilde{v}_1, 1]$,

$$\begin{aligned} p_1^{*\prime}(v_1) =& -\psi'(v_1)(1-q_1^*(v_1))F(-\psi(v_1)|q_1^*(v_1)) \\ &+ q_1^{*\prime}(v_1)\underbrace{\int_{-\infty}^{-\psi(v_1)} \frac{\partial(1-q_1^*(v_1))F(v_2|q_1^*(v_1))}{\partial q_1^*(v_1)}\,\mathrm{d}v_2}_{=0 \text{ by } (18)} \\ &- (1-q_1^*(v_1))(\frac{1-G(v_1)}{g(v_1)})' + q_1^{*\prime}(v_1)\frac{1-G(v_1)}{g(v_1)} \\ &+ (1-q_1^*(v_1))F(-\psi(v_1)|q_1^*(v_1)) \\ =& q_1^{*\prime}(v_1)\frac{1-G(v_1)}{g(v_1)} - (1-q_1^*(v_1))(1-F(-\psi(v_1)|q_1^*(v_1)))(\frac{1-G(v_1)}{g(v_1)})'. \end{aligned}$$

Recall that for $v_1 \in [\tilde{v}_1, 1]$, $q_1^{*\prime}(v_1) > 0$, $\frac{1-G(v_1)}{g(v_1)} \ge 0$ with strict inequality when $v_1 \in [\tilde{v}_1, 1)$, and $(\frac{1-G(v_1)}{g(v_1)})' < 0$ (Assumption 2). It can be seen that $p_1^{*\prime}(v_1) \ge 0$ with strict inequality when $v_1 \in [\tilde{v}_1, 1)$. Hence, p_1^* is strictly increasing on $[\tilde{v}_1, 1]$.

It is easy to see that $\Pi(0, \tilde{v}_1) = \Pi(q_1^*(\tilde{v}_1), \tilde{v}_1)$. As a result,

$$\int_{-\infty}^{-\psi(\tilde{v}_1)} F(x|0)\,\mathrm{d}x = (1-q_1^*(\tilde{v}_1)) \int_{-\infty}^{-\psi(\tilde{v}_1)} F(x|q_1^*(\tilde{v}_1))\,\mathrm{d}x,$$

which, jointly with Lemma 5, implies that

$$
\begin{aligned}
p_1^*(\tilde{v}_1) &= \int_{-\infty}^{-\psi(\tilde{v}_1)} F(x|0)\,\mathrm{d}x - (1-q_1^*(\tilde{v}_1))\frac{1-G(\tilde{v}_1)}{g(\tilde{v}_1)} + \int_0^{\tilde{v}_1}(1-q_1^*(x))F(-\psi(x)|q_1^*(x))\,\mathrm{d}x \\
&= -\psi(\tilde{v}_1) - (1-q_1^*(\tilde{v}_1))\frac{1-G(\tilde{v}_1)}{g(\tilde{v}_1)} + \tilde{v}_1 = q_1^*(\tilde{v}_1)\frac{1-G(\tilde{v}_1)}{g(\tilde{v}_1)}.
\end{aligned}
$$

For (iii), the result follows directly from Assumption 2.

For (iv), when $v_1 < \tilde{v}_1$, $q_1^*(v_1) = 0$,

$$
\underbrace{p_1^*(v_1)}_{=0 \text{ by } (26)} + p_2^*(v_1)(1-q_1^*(v_1)) = p_2^*(v_1) = \frac{1-G(v_1)}{g(v_1)},
$$

which is strictly decreasing in v_1.

Plugging in the expressions of p_1^* and p_2^*, we have that

$$
\begin{aligned}
&p_1^*(v_1) + p_2^*(v_1)(1-q_1^*(v_1)) \\
=&(1-q_1^*(v_1))\int_{-\infty}^{-\psi(v_1)} F(v_2|q_1^*(v_1))\,\mathrm{d}v_2 + \int_0^{v_1}(1-q_1^*(x))F(-\psi(x)|q_1^*(x))\,\mathrm{d}x.
\end{aligned}
$$

For $v_1 \geq \tilde{v}_1$, since (10) applies, the derivative of the above expression with respect to v_1 is

$$
(1-q_1^*(v_1))F(-\psi(v_1)|q_1^*(v_1)) \cdot (\frac{1-G(v_1)}{g(v_1)})' < 0.
$$

For (v), the expected payment of any type v_1 is given by

$$
\begin{aligned}
&p_1^*(v_1) + p_2^*(v_1)(1-q_1^*(v_1))(1-F(-\psi(v_1)|q_1^*(v_1))) \\
=&p_1^*(v_1) + p_2^*(v_1)(1-q_1^*(v_1)) - p_2^*(v_1)(1-q_1^*(v_1))F(-\psi(v_1)|q_1^*(v_1)).
\end{aligned}
$$

For $v_1 \in [0, \tilde{v}_1)$, $p_1^*(v_1) = 0$ and $F(-\psi(v_1)|q_1^*(v_1)) = F(-\psi(v_1)|0) = 1$ (since $\tilde{v}_1 < v_1^*$). Thus, the expected payment of $v_1 \in [0, \tilde{v}_1)$ is equal to zero. For $v_1 \in [\tilde{v}_1, 1]$, the derivative of the above expression with respect to v_1 is

$$
\begin{aligned}
&(1-q_1^*(v_1))F(-\psi(v_1)|q_1^*(v_1)) \cdot (\frac{1-G(v_1)}{g(v_1)})' - \Big(\frac{1-G(v_1)}{g(v_1)}(1-q_1^*(v_1))F(-\psi(v_1)|q_1^*(v_1))\Big)' \\
=&-\frac{1-G(v_1)}{g(v_1)}\big((1-q_1^*(v_1))F(-\psi(v_1)|q_1^*(v_1))\big)' \\
=&-\frac{1-G(v_1)}{g(v_1)}\big(\underbrace{-\psi'(v_1)}_{<0}(1-q_1^*(v_1))f(-\psi(v_1)|q_1^*(v_1)) + \underbrace{\xi(-\psi(v_1), q_1^*(v_1))}_{<0 \text{ by } (19)} q_1^{*\prime}(v_1)\big) > 0.
\end{aligned}
$$

The proof completes. $\square$

A.9 Proof of the claim in Section 4.3

Rotation order. Note that

$$\frac{\partial F(v_2|q_1)}{\partial q_1} = \frac{\partial \int_{-\infty}^{\frac{v_2}{q_1}} \frac{1}{\sqrt{2\pi}} e^{-\frac{s^2}{2}} \,\mathrm{d}s}{\partial q_1} = -\frac{v_2}{q_1^2} \frac{1}{\sqrt{2\pi}} e^{-\frac{v_2^2}{2q_1^2}}.$$

It is clear that F satisfies the rotation order, as

$$\frac{\partial F(v_2|q_1)}{\partial q_1} \begin{cases} > 0, & \text{when } v_2 < 0; \\ = 0, & \text{when } v_2 = 0; \\ < 0, & \text{when } v_2 > 0. \end{cases}$$

Assumption 1. By Remark 1, it suffices to show that $F(v_2|q_1)/f(v_2|q_1)$ is increasing in v_2 and $\frac{\partial F(v_2|q_1)}{\partial q_1}/f(v_2|q_1)$ is decreasing in v_2. The latter is straightforward as

$$\frac{\partial F(v_2|q_1)}{\partial q_1}/f(v_2|q_1) = \left(-\frac{v_2}{q_1^2} \frac{1}{\sqrt{2\pi}} e^{-\frac{v_2^2}{2q_1^2}} \right) / \left(\frac{1}{q_1\sqrt{2\pi}} e^{-\frac{v_2^2}{2q_1^2}} \right) = -\frac{v_2}{q_1}.$$

To show that $F(v_2|q_1)/f(v_2|q_1)$ is increasing in v_2, note that

$$F(v_2|q_1)/f(v_2|q_1) = \left(\int_{-\infty}^{\frac{v_2}{q_1}} \frac{1}{\sqrt{2\pi}} e^{-\frac{s^2}{2}} \,\mathrm{d}s \right) / \left(\frac{1}{q_1\sqrt{2\pi}} e^{-\frac{v_2^2}{2q_1^2}} \right).$$

By changing variables as $x = \frac{v_2}{q_1}$, one needs to show that $\varphi(x) = \left(\int_{-\infty}^{x} e^{-\frac{s^2}{2}} \,\mathrm{d}s \right) e^{\frac{x^2}{2}}$ is increasing in x. We have that

$$\varphi'(x) = 1 + \left(\int_{-\infty}^{x} e^{-\frac{s^2}{2}} \,\mathrm{d}s \right) e^{\frac{x^2}{2}} x = e^{\frac{x^2}{2}} \left(e^{-\frac{x^2}{2}} + x \int_{-\infty}^{x} e^{-\frac{s^2}{2}} \,\mathrm{d}s \right).$$

When $x \to -\infty$, $e^{-\frac{x^2}{2}} \to 0$, and $x \int_{-\infty}^{x} e^{-\frac{s^2}{2}} \,\mathrm{d}s \to 0$ by L'Hôspital's rule. In addition,

$$\left(e^{-\frac{x^2}{2}} + x \int_{-\infty}^{x} e^{-\frac{s^2}{2}} \,\mathrm{d}s \right)' = \int_{-\infty}^{x} e^{-\frac{s^2}{2}} \,\mathrm{d}s > 0.$$

Thus, $\varphi'(x) \geq 0$ and $\varphi(x)$ is increasing, which implies that Assumption 1 holds.

Assumption 2. It is clear that $\frac{1-G(v_1)}{g(v_1)} = 1 - v_1$ is decreasing, implying that Assumption 2 holds.

Monotonicity $\not\Rightarrow$ global IC. Finally, we show that $(\hat{q}_1, \hat{q}_2)$ cannot be the allocation rule in an incentive-compatible mechanism. Suppose that the claim does not hold. Then there exists some $\hat{t}$ such that $(\hat{q}_1, \hat{q}_2, \hat{t})$ is IC in both stages. We abuse the notation a bit by still using $U(v_1, r_1)$ to denote the buyer's utility with the first-stage type v_1 and report r_1.

By Lemma 2, for $v_1 \geq \frac{1}{2}$,

$$
\begin{aligned}
U(v_1, v_1) &= \int_0^{v_1} \left[\hat{q}_1(s) + \int_{-\infty}^{+\infty} \hat{q}_2(s, v_2) F(\mathrm{d}v_2 | \hat{q}_1(s))\right] \mathrm{d}s \\
&= \int_0^{v_1} \left[\hat{q}_1(s) + \int_{-\psi(s)}^{+\infty} (1 - \hat{q}_1(s)) F(\mathrm{d}v_2 | \hat{q}_1(s))\right] \mathrm{d}s \\
&= \int_0^{v_1} \left[\hat{q}_1(s) + (1 - \hat{q}_1(s))(1 - F(-\psi(s) | \hat{q}_1(s))\right] \mathrm{d}s \\
&= \int_0^{v_1} \left[1 - H\left(-\frac{1}{\psi(s)}\right)(1 - \psi^2(s))\right] \mathrm{d}s,
\end{aligned}
$$

where the last equality holds since $F(v_1|q_1) = H(\frac{v_1}{q_1})$. On the other hand,

$$
U(v_1, v_1) = \hat{q}_1(v_1) v_1 + \int_{-\psi(v_1)}^{+\infty} (1 - \hat{q}_1(v_1))(v_1 + v_2) F(\mathrm{d}v_2 | \hat{q}_1(v_1)) - \hat{T}(v_1),
$$

where

$$
\hat{T}(v_1) = \int_{-\infty}^{+\infty} \hat{t}(v_1, v_2) F(\mathrm{d}v_2 | \hat{q}_1(v_1)).
$$

Then we have

$$
\begin{aligned}
U(v_1, r_1) &= \hat{q}_1(r_1) v_1 + \int_{-\psi(r_1)}^{+\infty} (1 - \hat{q}_1(r_1))(v_1 + v_2) F(\mathrm{d}v_2 | \hat{q}_1(r_1)) - \hat{T}(r_1) \\
&= \hat{q}_1(r_1) v_1 + \int_{-\psi(r_1)}^{+\infty} (1 - \hat{q}_1(r_1))(v_1 + v_2) F(\mathrm{d}v_2 | \hat{q}_1(r_1)) \\
&\quad + U(r_1, r_1) - \hat{q}_1(r_1) r_1 - \int_{-\psi(r_1)}^{+\infty} (1 - \hat{q}_1(r_1))(r_1 + v_2) F(\mathrm{d}v_2 | \hat{q}_1(r_1)) \\
&= U(r_1, r_1) + \hat{q}_1(r_1)(v_1 - r_1) + (1 - \hat{q}_1(r_1))(v_1 - r_1)\left(1 - H\left(-\frac{1}{\psi(r_1)}\right)\right) \\
&= U(r_1, r_1) + (v_1 - r_1) - (1 - \psi^2(r_1))(v_1 - r_1) H\left(-\frac{1}{\psi(r_1)}\right).
\end{aligned}
$$

By simple algebra,

$$
U(v_1, v_1) \geq U(v_1, r_1) \qquad \Longleftrightarrow
$$

$$
\int_{r_1}^{v_1} H\left(-\frac{1}{\psi(s)}\right)(1 - \psi^2(s))\, \mathrm{d}s \leq (v_1 - r_1) H\left(-\frac{1}{\psi(r_1)}\right)(1 - \psi^2(r_1)),
$$

which may not be always true. We observe that $H(-\frac{1}{x})(1 - x^2) \geq 0$ for $x \in [0, 1]$, converges to 0 when either $x \to 0$ or $x \to 1$. Thus, there must be an open set $(a_1, a_2) \subseteq [0, 1]$ such that $H(-\frac{1}{x})(1 - x^2)$ is strictly increasing on (a_1, a_2). Pick r_1 and v_1 such that $a_1 < \psi(r_1) < \psi(v_1) < a_2$. Then for any $s \in (r_1, v_1]$,

$$
H\left(-\frac{1}{\psi(s)}\right)(1 - \psi^2(s)) > H\left(-\frac{1}{\psi(r_1)}\right)(1 - \psi^2(r_1)),
$$

which implies that $U(v_1, v_1) < U(v_1, r_1)$. This is a contraction. $\square$

A.10 Proof of Proposition 3

The seller's revenue from a type-v_1 buyer with the first-stage consumption q_1 is

$$\begin{aligned}\Pi^s(q_1, v_1) &= (q_1 + s(q_1))\psi(v_1) + s(q_1)\int_{-\infty}^{-\psi(v_1)} F(v_2|q_1)\,\mathrm{d}v_2 \\ &= (q_1 + s(q_1))\psi(v_1) + \int_{-\infty}^{-\psi(v_1)} [q_1 + s(q_1) - C^s(v_2, q_1)]\,\mathrm{d}v_2 \\ &= (q_1 + s(q_1))(\psi(v_1) + \int_{-\infty}^{-\psi(v_1)} \mathrm{d}v_2) - \int_{-\infty}^{-\psi(v_1)} C^s(v_2, q_1)\,\mathrm{d}v_2.\end{aligned}$$

Thus, we have

$$\frac{\partial \Pi^s(q_1, v_1)}{\partial q_1} = (1 + s'(q_1))(\psi(v_1) + \int_{-\infty}^{-\psi(v_1)} \mathrm{d}v_2) - \int_{-\infty}^{-\psi(v_1)} \frac{\partial C^s(v_2, q_1)}{\partial q_1}\,\mathrm{d}v_2. \tag{27}$$

Clearly,

$$\frac{\partial^2 \Pi^s(q_1, v_1)}{\partial q_1 \partial v_1} = \psi'(v_1)\frac{\partial C^s(-\psi_1(v_1), q_1)}{\partial q_1}. \tag{28}$$

For an interior optimum $q_1^s(v_1)$, it satisfies the first-order condition

$$\frac{\partial \Pi^s(q_1, v_1)}{\partial q_1}|_{q_1 = q_1^s(v_1)} = (1+s'(q_1^s(v_1)))(\psi(v_1)+\int_{-\infty}^{-\psi(v_1)} \mathrm{d}v_2)-\int_{-\infty}^{-\psi(v_1)} \frac{\partial C^s(v_2, q_1^s(v_1))}{\partial q_1^s(v_1)}\,\mathrm{d}v_2 = 0.$$

Moreover, by the second-order condition, $\frac{\partial^2 \Pi^s(q_1, v_1)}{\partial q_1^2}|_{q_1 = q_1^s(v_1)} \leq 0$. Further differentiating the above first-order condition with respect to v_1 on both sides of the equation leads to

$$\frac{\mathrm{d}q_1^s(v_1)}{\mathrm{d}v_1} \cdot \underbrace{\frac{\partial^2 \Pi^s(q_1, v_1)}{\partial q_1^2}|_{q_1 = q_1^s(v_1)}}_{\leq 0} + \frac{\partial^2 \Pi(q_1, v_1)}{\partial q_1 \partial v_1}|_{q_1 = q_1^s(v_1)} = 0. \tag{29}$$

<u>Step 1</u>. Identify the monotonicity or quasi-concavity property of $q_1^s(\cdot)$.

Case 1: $q_1 + s(q_1)$ is weakly increasing in $q_1 \in [0, 1]$.

If at v_1, the optimal first-stage allocation $q_1^s(v_1)$ is interior and thus satisfies $\frac{\partial \Pi^s(q_1^s(v_1), v_1)}{\partial q_1^s(v_1)} = 0$, then

$$\int_{-\infty}^{-\psi(v_1)} \frac{\partial C^s(v_2, q_1^s(v_1))}{\partial q_1^s(v_1)}\,\mathrm{d}v_2 = (1 + s'(q_1^s(v_1)))(\psi(v_1) + \int_{-\infty}^{-\psi(v_1)} \mathrm{d}v_2) \geq 0. \tag{30}$$

By Assumption 3, we have $\frac{\partial C^s(-\psi_1(v_1), q_1^s(v_1))}{\partial q_1^s(v_1)} > 0$.[25] Therefore, (28) implies that

[25] Strictly speaking, $\psi(v_1) + \int_{-\infty}^{-\psi(v_1)} \mathrm{d}v_2 = +\infty$. But this does not affect the analysis: When $1 +$

$\frac{\partial^2 \Pi(q_1,v_1)}{\partial q_1 \partial v_1}|_{q_1=q_1^s(v_1)} > 0$, which, combined with (29), further implies $q_1^{s\prime}(v_1) > 0$.

Following Lemma 3, it is easy to show that there exists a cutoff $\tilde{v}_1^s < v_1^*$ such that $q_1^s(v_1) = 0$ for $v_1 < \tilde{v}_1^s$ if any, and $q_1^s(v_1) \in (0,1]$ for $v_1 > \tilde{v}_1^s$.

We now show that if $q_1^s(v_1) = 1$, then $q_1^s(v_1') = 1$ for all $v_1' \geq v_1$. For any $q_1 \in (0,1]$, $\Pi^s(1,v_1) \geq \Pi^s(q_1,v_1)$ is equivalent to

$$\int_{-\infty}^{-\psi(v_1)} \int_{q_1}^{1} \frac{\partial C^s(v_2,\tilde{q}_1)}{\partial \tilde{q}_1} \,\mathrm{d}\tilde{q}_1 \,\mathrm{d}v_2 \leq (1+s(1)-q_1-s(q_1))(\psi(v_1)+\int_{-\infty}^{-\psi(v_1)} \mathrm{d}v_2).$$

Notice that the right-hand side expression is nonnegative and independent of v_1. Hence, by Assumption 3, for all $v_1' > v_1$,

$$\int_{-\infty}^{-\psi(v_1')} \int_{q_1}^{1} \frac{\partial C^s(v_2,\tilde{q}_1)}{\partial \tilde{q}_1} \,\mathrm{d}\tilde{q}_1 \,\mathrm{d}v_2 \leq (1+s(1)-q_1-s(q_1))(\psi(v_1')+\int_{-\infty}^{-\psi(v_1')} \mathrm{d}v_2).$$

This implies that $\Pi^s(1,v_1') \geq \Pi^s(q_1,v_1')$. Hence, $q_1^s(v_1') = 1$.

Overall, $q_1^s(\cdot)$ is weakly increasing.

Case 2: $q_1 + s(q_1)$ is strictly decreasing in $q_1 \in [0,1]$.

Notice that $\psi(v_1) + \int_{-\infty}^{-\psi(v_1)} \mathrm{d}v_2$ is independent of v_1. By Assumption 3, for each $q_1 \in [0,1]$, there are at most two $v_1 \in [0,1]$ such that $(27) = 0$: the v_1 such that $\frac{\partial C^s(-\psi(v_1),q_1)}{\partial q_1} \geq 0$ (resp. < 0) satisfies $q_1^{s\prime}(v_1) \geq 0$ (resp. < 0)[26] and is denoted by $\varphi_1(q_1)$ (resp. $\varphi_2(q_1)$). Moreover, whenever there are exactly two solutions, by Assumption 3 and the fact that $(27) = 0$: it must be the case that $\varphi_1(q_1) < \varphi_2(q_1)$ and

$$\int_{-\psi(\varphi_2(q_1))}^{-\psi(\varphi_1(q_1))} \frac{\partial C^s(v_2,q_1)}{\partial q_1} \,\mathrm{d}v_2 = 0. \tag{31}$$

It is easy to show that for any $v_1' < v_1 < v_1^*$, if $q_1^s(v_1) = 0$, then $q_1^s(v_1') = 0$, and for any $v_1' > v_1 > v_1^*$, if $q_1^s(v_1) = 0$, then $q_1^s(v_1') = 0$. Under Assumption 3, it is easy to show that if there are $v_1 < v_1''$ such that $q_1^s(v_1) = q_1^s(v_1'') = 1$, then $q_1^s(v_1') = 1$ for all $v_1' \in [v_1, v_1'']$.

Hence, $q_1^s(\cdot)$ is quasi-concave.

Step 2. Construct the transfer rule.

Once the optimal stage-one allocation $q_1^s(\cdot)$ in the relaxed problem is identified, one can follow Lemma 5 to show that the advance payment and the per-unit strike price are given by $p_1^s(\cdot)$ and $p_2^*(\cdot)$. We thus omit the details.

Step 3. Establish the global IC condition.

$s'(q_1^s(v_1)) = 0$, it goes back to the case in our main model, and thus $\frac{\partial C^s(-\psi_1(v_1),q_1^s(v_1))}{\partial q_1^s(v_1)} > 0$; when $1 + s'(q_1^s(v_1)) > 0$, (30) implies that when $N > 0$ is large enough, $\int_{-N}^{-\psi(v_1)} \frac{\partial C^s(v_2,q_1^s(v_1))}{\partial q_1^s(v_1)} \mathrm{d}v_2 \geq 0$, which implies $\frac{\partial C^s(-\psi_1(v_1),q_1^s(v_1))}{\partial q_1^s(v_1)} > 0$ by Assumption 3. Therefore, hereafter, to ease our notations and proof, we will also use similar notations and argument as in (30).

[26]The monotonicity of q_1^s can be seen from the argument in Case 1 and expression (29).

To do so, we generalize the w function defined in the proof of Proposition 1 to w^s as follows:

$$w^s(q_1, p_2, v_1) = q_1 v_1 + s(q_1) \int_{p_2 - v_1}^{+\infty} (v_1 + v_2 - p_2) F(\mathrm{d}v_2 | q_1),$$

which satisfies

$$w^s_{31}(q_1, p_2, v_1) = \frac{\partial C^s(p_2 - v_1, q_1)}{\partial q_1}, \quad w^s_{32}(q_1, p_2, v_1) \leq 0.$$

Similar to the proof of Proposition 1, for any v_1 and $r_1 \in [0, 1]$, to show that type v_1 has the incentive to truthfully report rather than misreporting r_1, it suffices to show that

$$\begin{aligned}\int_{r_1}^{v_1} \int_{x}^{v_1} w^s_{31}(q^s_1(x), p^*_2(x), z) q^{s\prime}_1(x) \, \mathrm{d}z \, \mathrm{d}x &= \int_{r_1}^{v_1} \int_{x}^{v_1} \frac{\partial C^s(p^*_2(x) - z, q^s_1(x))}{\partial q^s_1(x)} q^{s\prime}_1(x) \, \mathrm{d}z \, \mathrm{d}x \\ = \int_{r_1}^{v_1} \int_{x - v_1 - \psi(x)}^{-\psi(x)} \frac{\partial C^s(y, q^s_1(x))}{\partial q^s_1(x)} q^{s\prime}_1(x) \, \mathrm{d}y \, \mathrm{d}x \geq 0. & \qquad (32)\end{aligned}$$

When $q^s_1(\cdot)$ is monotone, one can follow Proposition 1 to show that the required inequality holds. The proof is very similar, which utilizes Assumption 3 and the optimality of $q^s_1(\cdot)$.

When $q^s_1(\cdot)$ is non-monotone (and thus is first increasing and then decreasing), the verification is complicated yet tedious. As an illustration, we only demonstrate the case where $v_1 < r_1$ with $0 < q^s_1(v_1) \leq q^s_1(r_1) < 1$, $q^{s\prime}_1(v_1) > 0$, and $q^{s\prime}_1(r_1) < 0$. All other cases can be verified analogously. In the current case, fix any $\hat{v}_1 \in \mathrm{argmax}_{v_1 \in [0,1]} q^s_1(v_1)$ and thus $v_1 < \hat{v}_1 < r_1$. Let $r'_1 \in [v_1, \hat{v}_1]$ be the "symmetric" type of r_1 in the sense that $r'_1 = \varphi_1(q^s_1(r_1))$.

We rewrite the double integral in (32) as follows

$$\int_{r'_1}^{v_1} \int_{x - v_1 - \psi(x)}^{-\psi(x)} \frac{\partial C^s(y, q^s_1(x))}{\partial q^s_1(x)} q^{s\prime}_1(x) \, \mathrm{d}y \, \mathrm{d}x \qquad (33)$$

$$+ \int_{\hat{v}_1}^{r'_1} \int_{x - v_1 - \psi(x)}^{-\psi(x)} \frac{\partial C^s(y, q^s_1(x))}{\partial q^s_1(x)} q^{s\prime}_1(x) \, \mathrm{d}y \, \mathrm{d}x + \int_{r_1}^{\hat{v}_1} \int_{x - v_1 - \psi(x)}^{-\psi(x)} \frac{\partial C^s(y, q^s_1(x))}{\partial q^s_1(x)} q^{s\prime}_1(x) \, \mathrm{d}y \, \mathrm{d}x. \qquad (34)$$

Again, $(33) \geq 0$ can be established by the same analysis as in the proof of Proposition 1. By performing a change of variable, (34) can be rewritten as

$$\begin{aligned}&\int_{q^s_1(\hat{v}_1)}^{q^s_1(r'_1)} \int_{\varphi_1(q_1) - v_1 - \psi(\varphi_1(q_1))}^{-\psi(\varphi_1(q_1))} \frac{\partial C^s(y, q_1)}{\partial q_1} \, \mathrm{d}y \, \mathrm{d}q_1 + \int_{q^s_1(r_1)}^{q^s_1(\hat{v}_1)} \int_{\varphi_2(q_1) - v_1 - \psi(\varphi_2(q_1))}^{-\psi(\varphi_2(q_1))} \frac{\partial C^s(y, q_1)}{\partial q_1} \, \mathrm{d}y \, \mathrm{d}q_1 \\ &= \int_{q^s_1(r'_1) = q^s_1(r_1)}^{q^s_1(\hat{v}_1)} \left[\int_{-\psi(\varphi_1(q_1))}^{\varphi_1(q_1) - v_1 - \psi(\varphi_1(q_1))} \frac{\partial C^s(y, q_1)}{\partial q_1} \, \mathrm{d}y + \int_{\varphi_2(q_1) - v_1 - \psi(\varphi_2(q_1))}^{-\psi(\varphi_2(q_1))} \frac{\partial C^s(y, q_1)}{\partial q_1} \, \mathrm{d}y \right] \mathrm{d}q_1 \\ &\overset{(31)}{=} \int_{q^s_1(r_1)}^{q^s_1(\hat{v}_1)} \left[\int_{\varphi_2(q_1) - v_1 - \psi(\varphi_2(q_1))}^{\varphi_1(q_1) - v_1 - \psi(\varphi_1(q_1))} \frac{\partial C^s(y, q_1)}{\partial q_1} \, \mathrm{d}y \right] \mathrm{d}q_1. \qquad (35)\end{aligned}$$

Fix any $q_1 \in [q_1^s(r_1), q_1^s(\hat{v}_1)]$. By Assumption 2, $\psi'(\cdot) \geq 1$ and thus $\frac{\psi(\varphi_2(q_1))-\psi(\varphi_1(q_1))}{\varphi_2(q_1)-\varphi_1(q_1)} \geq 1$. This further implies that $\varphi_1(q_1) - v_1 - \psi(\varphi_1(q_1)) \geq \varphi_2(q_1) - v_1 - \psi(\varphi_2(q_1))$.

- If for $y = \varphi_2(q_1) - v_1 - \psi(\varphi_2(q_1))$, $\frac{\partial C^s(y,q_1)}{\partial q_1} \geq 0$, then it is obvious that

$$\int_{\varphi_2(q_1)-v_1-\psi(\varphi_2(q_1))}^{\varphi_1(q_1)-v_1-\psi(\varphi_1(q_1))} \frac{\partial C^s(y,q_1)}{\partial q_1} \, \mathrm{d}y \geq 0$$

  by Assumption 3.

- If for $y = \varphi_2(q_1) - v_1 - \psi(\varphi_2(q_1))$, $\frac{\partial C^s(y,q_1)}{\partial q_1} < 0$, then by Assumption 3, there exists $\hat{y} \in [\varphi_2(q_1) - v_1 - \psi(\varphi_2(q_1)), \varphi_1(q_1) - v_1 - \psi(\varphi_1(q_1))]$ such that $\frac{\partial C^s(\hat{y},q_1)}{\partial q_1} = 0$.[27] Moreover,

$$\begin{aligned}
&\int_{\varphi_2(q_1)-v_1-\psi(\varphi_2(q_1))}^{\varphi_1(q_1)-v_1-\psi(\varphi_1(q_1))} \frac{\partial C^s(y,q_1)}{\partial q_1} \, \mathrm{d}y \\
&= \int_{\varphi_2(q_1)-v_1-\psi(\varphi_2(q_1))}^{\hat{y}} \underbrace{\frac{\partial C^s(y,q_1)}{\partial q_1}}_{\leq 0} \mathrm{d}y + \int_{\hat{y}}^{\varphi_1(q_1)-v_1-\psi(\varphi_1(q_1))} \underbrace{\frac{\partial C^s(y,q_1)}{\partial q_1}}_{\geq 0} \mathrm{d}y \\
&\overset{\substack{\varphi_2(q_1)-v_1\geq 0,\\ \varphi_1(q_1)-v_1\geq 0}}{\geq} \int_{-\psi(\varphi_2(q_1))}^{\hat{y}} \frac{\partial C^s(y,q_1)}{\partial q_1} \, \mathrm{d}y + \int_{\hat{y}}^{-\psi(\varphi_1(q_1))} \frac{\partial C^s(y,q_1)}{\partial q_1} \, \mathrm{d}y \overset{(31)}{=} 0.
\end{aligned}$$

Hence, we have (35) ≥ 0 and eventually (32) ≥ 0. This completes the proof of the global IC condition. □

[27] To see this, notice that (31) implies that $\frac{\partial C^s(-\psi(\varphi_1(q_1)),q_1)}{\partial q_1} > 0$. On the other hand, $\varphi_1(q_1) - v_1 - \psi(\varphi_1(q_1)) > -\psi(\varphi_1(q_1))$ for any $q_1 \in [q_1^s(r_1), q_1^s(\hat{v}_1)]$. Assumption 3 implies that: for $z = \varphi_1(q_1) - v_1 - \psi(\varphi_1(q_1))$, $\frac{\partial C^s(z,q_1)}{\partial q_1} > 0$. The existence of $\hat{y}$ then follows from the continuity of $\frac{\partial C^s(z,q_1)}{\partial q_1}$ in z.