

# Optimal Auction Design with Contingent Payments and Costly Verification\*

Ian Ball      Teemu Pekkariinen

15 April 2026

## Abstract

We study the design of an auction for an income-generating asset such as an intellectual property license. Each bidder has a signal about his future income from acquiring the asset. After the asset is allocated, the winner’s income from the asset is realized privately. The principal can audit the winner, at a cost, and then charge a payment contingent on the winner’s realized income. We solve for an auction that maximizes the principal’s revenue, net of auditing costs. The winning bidder is charged linear royalties up to a cap, beyond which there is no auditing. A higher bidder pays more in cash upfront and faces a lower royalty cap. (JEL D44, D82, D86)

---

\*Ball: Department of Economics, MIT (ianball@mit.edu); Pekkariinen: Department of Economics, University of Vaasa (teemu.pekkariinen@uwasa.fi). This paper was previously circulated under the titles “Optimal Auction Design with Flexible Royalty Payments” and “Optimal Auction Design with Contingent Payments.” For helpful comments, we are grateful to Dirk Bergemann, Daniel Clark, Laura Doval, Marina Halac, Daniel Hauser, Topi Hokkanen, Johannes Hörner, Deniz Kattwinkel, Jan Knoepfle, Bart Lipman, Laurenz Marsteller, Stephen Morris, Petteri Palonen, Juuso Toikka, Hannu Vartiainen, Juuso Välimäki, Stephan Waizmann, Alex Winter, Alex Wolitzky, and members of the microeconomic theory groups at the Helsinki GSE, MIT, Yale, TSE, and Bonn. Part of this paper was written while Ian Ball was visiting Northwestern University and Teemu Pekkariinen was visiting MIT.

## 1. INTRODUCTION

In 2017, Cardiva Medical Inc. entered into an asset purchase agreement with Interventional Therapies, LLC, a developer of medical devices.<sup>1</sup> Cardiva agreed to pay Interventional Therapies \$100,000 at closing and a 5% royalty on net sales of a specified list of products, subject to a royalty cap of \$7.5 million. To determine the royalties owed, the agreement requires the buyer to submit quarterly sales reports. To verify these reports, the seller may audit the buyer’s records using an independent certified public accountant. Any underpayment revealed by an audit must be paid, and any overpayment refunded.

Royalty agreements are standard in various forms of intellectual property licensing such as patents, copyrights, and franchises.<sup>2</sup> Under a royalty agreement, the amount that the buyer ultimately pays is contingent on subsequent sales. More generally, contingent payments are common whenever an asset generates verifiable cash flows. When there are multiple potential buyers, contingent-payment auctions are often used, as in government auctions for casino licenses, oil leases, and procurement contracts (Skrzypacz, 2013).

The theoretical literature on contingent-payment auctions, following DeMarzo et al. (2005), assumes that the income generated by the asset is publicly observed. But in many applications, such as in the royalty agreement described above, cash flows are privately observed and can only be verified at a cost. Indeed, an important consideration in patent licensing is that “a lump-sum royalty removes the administrative burden and costs of monitoring the actual use of the licensed technology because the royalty payment is independent of the licensee’s *actual* sales” (Sidak, 2016, p. 903).

In this paper, we study the design of an auction for an asset that generates income, which can be verified at a cost. For example, the asset could specify the right to use some intellectual property (such as a patent, trademark, copyright, or franchise). Or the asset could specify the right to operate in a particular industry (such as a government license for a casino or a marijuana dispensary).

---

<sup>1</sup>The full asset purchase agreement is available in the SEC archives at <https://www.sec.gov/Archives/edgar/data/1520325/000119312521001047/d833938dex21.htm>.

<sup>2</sup>See Contreras (2022) for a textbook treatment of intellectual-property licensing.

Procurement contracts can also be covered by our analysis after an appropriate sign change; costs, rather than income, are subsequently realized, so contingent payments are interpreted as cost-sharing rather than royalties.

We consider the following dynamic mechanism design setting. Each agent observes an independent, private signal about his future income from acquiring the asset. As in a standard auction, the principal solicits bids from the agents. Based on these bids, the principal charges upfront transfers and allocates the asset to at most one agent. The next phase is specific to our setting. If the asset is allocated, the winner's income from the asset is realized. The winner privately observes this income and makes a report to the principal. Based on this report, the principal charges the winner a payment. Finally, the principal can conduct a costly audit that reveals the winner's realized income. If the principal conducts the audit, she can then charge the winner a payment that depends on the winner's realized income. The slope of this payment, as a function of income, is bounded, as specified in the model.

We solve for a dynamic mechanism that maximizes the principal's revenue, net of auditing costs. To build intuition for the solution, we begin with two benchmarks.

If the principal cannot charge payments that are contingent on realized income, then the model reduces to the standard cash-payment auction setting of [Myerson \(1981\)](#). At the other extreme, suppose that the income generated by the asset is public and that the principal can charge *unrestricted* payments contingent on this realized income. In this case, the principal can fully extract the surplus using the following modified first-price auction. The allocation and upfront transfers are the same as in a standard first-price auction, but the winner is charged a penalty equal to the difference between his realized income from the asset and his bid. In this auction, each agent receives zero utility, no matter how he bids. In particular, it is an equilibrium for each agent to bid his expected income from the asset. In this equilibrium, the asset is allocated efficiently, so the principal extracts the full surplus.

Our main model interpolates between these benchmarks. In [Theorem 1](#), we identify an optimal mechanism. To solve for this mechanism, we find an

expression for the virtual value in our setting. This virtual value is weakly higher than Myerson’s virtual value (for cash auctions) because contingent payments reduce information rents. In our optimal mechanism, the asset is allocated to the agent with the highest virtual value, provided that this virtual value is positive. If the asset is allocated, the principal charges the winner a royalty based on his reported income. The royalty is linear up to a royalty cap (which depends on the winner’s bid). The winner is audited if and only if he claims that the royalty he owes is strictly less than the cap. If auditing reveals that the agent under-reported his income, then he is charged the underpaid royalties as a penalty. The increasing segment of the royalty schedule reduces information rents. Over this segment, costly auditing is necessary to motivate the winner to report his income truthfully. The flat segment beyond the royalty cap saves on auditing costs. Thus, our analysis helps to rationalize the use of royalty caps in practice, as in the agreement between Cardiva and Interventional Therapies.

With a single agent, the optimal auction in Theorem 1 can be implemented by posting a menu of contracts, each with a different royalty cap and upfront price. The agent chooses from this menu based upon his private signal about future income.

Theorem 2 analyzes comparative statics. Our optimal auction interpolates between the Myerson optimal cash auction and full-surplus extraction. If auditing is cheaper or if contingent payments are allowed to be steeper, then the virtual value increases, and royalty caps increase. For example, if technological advances reduce auditing costs, our model predicts that royalty caps will increase.

## **RELATED LITERATURE**

We depart from previous work on contingent-payment auctions by modeling the winner’s realized income as private information that can only be verified

at a cost.<sup>3</sup> Earlier work assumes that cash flows are public.<sup>4</sup> Hansen (1985) and Riley (1988) first observe that contingent payments can increase seller revenue relative to cash (non-contingent) payments. DeMarzo et al. (2005) present a general auction model in which the designer specifies an ordered set of securities and agents submit security bids to the auctioneer.<sup>5</sup> Crucially, the asset requires an upfront investment by the winning bidder. Their main finding is that “steeper” securities generate more revenue.<sup>6</sup> In a symmetric setting, they show that a first-price auction with call options is the revenue-maximizing auction within a class of symmetric, efficient security-bid auctions.

In our mechanism design approach, we do not impose symmetry or efficiency constraints. Figueroa and Inostroza (2025) also adopt a mechanism-design approach with contingent payments, but in a setting with a single agent and public cash flows. In their optimal mechanism, as in DeMarzo et al. (2005), the payment received by the principal takes the form of a call option, which is flat and then linearly increasing in the income generated by the asset.<sup>7</sup> By contrast, in our optimal mechanism with costly auditing, the payment to the principal is linearly increasing up to the royalty cap and then flat. Section 5.4 provides a more detailed comparison of our results with those in DeMarzo et al. (2005) and Figueroa and Inostroza (2025).

Several papers analyze optimal contingent-payment auctions under parametric

---

<sup>3</sup>There is a classical literature on optimal patent licensing (e.g., Kamien and Tauman (1986) and Katz and Shapiro (1986)). Those papers focus on the effect of oligopolistic competition in the market for the patented product. Our focus is on the adverse selection problem for the patent holder.

<sup>4</sup>Skrzypacz (2013, p. 666) comments on the potential cost of verification: “In other situations while values/costs are objective, they cannot be easily verified. For example, in auctions selecting contractors to repair a highway, the costs of completing the project are hard to verify. On the other hand, in many commercial settings, the value of an asset/contract is at least partially observed.”

<sup>5</sup>They also consider “informal” auctions in which the auctioneer cannot commit to a rule for selecting the winning bid.

<sup>6</sup>Sogo et al. (2016) show that with entry costs, steeper securities can further increase revenue by attracting more bidders. By contrast, Che and Kim (2010) observe that steeper securities can reduce revenues if each bidder’s opportunity cost from acquiring the asset is increasing in the cash flow that he can generate from the asset.

<sup>7</sup>Figueroa and Inostroza (2025) analyze the security received by the agent. This security is a debt contract. The payment received by the principal is therefore a call option.

restrictions on the form of the payments. In [Bernhardt et al. \(2020\)](#), payments are required to be in the form of cash plus a linear royalty at a fixed (exogenous) rate. In the setting of a takeover auction, [Liu \(2016\)](#) considers equity auctions, in which each bidder's payment is an (endogenous) equity share. [Liu and Bernhardt \(2021\)](#) allow for cash plus equity, and show that full extraction is possible. In all of these papers on security-bid auctions, the uninformed party designs the securities. By contrast, the finance literature on security design, beginning with [Nachman and Noe \(1994\)](#), typically assumes that the issuer has private information.<sup>8</sup>

We model auditing as costly state verification, following [Townsend \(1979\)](#). In models of lending to a single agent, [Townsend \(1979\)](#), [Diamond \(1984\)](#), and [Gale and Hellwig \(1985\)](#) show that it is optimal for the principal to offer a debt contract in which verification is performed only in the case of default. In those models, the agent has no private information prior to contracting. Our paper shows that the optimality of debt contracts extends to a costly verification setting with sequential screening. Moreover, we establish this optimality using the Myersonian approach.

There are also models of costly state verification with static screening, but they focus on different issues. [Border and Sobel \(1987\)](#) characterize optimal tax enforcement with costly auditing. In their optimal mechanism, taxes are monotonically increasing and auditing is monotonically decreasing as a function of reported wealth. After auditing, the agent is offered a rebate if his report is revealed to be truthful. [Ben-Porath et al. \(2014\)](#), [Mylovanov and Zapechelnyuk \(2017\)](#), and [Li \(2020\)](#) study optimal mechanisms for allocating a good among agents, using costly verification but no transfers. After verification, the principal can impose a limited punishment. This punishment, unlike a transfer, does not benefit the principal.

In our model, private information arrives sequentially. At the time of contracting, each agent has an imperfect signal of his valuation, as in the single-agent models of [Baron and Besanko \(1984\)](#) and [Courty and Li \(2000\)](#), and the

---

<sup>8</sup>A recent exception is [Gershkov et al. \(2025\)](#), who study optimal security design with informed, risk-averse investors.

multi-agent model of [Eső and Szentes \(2007\)](#).<sup>9</sup> In those models, each agent learns his true valuation for an object (like a plane ticket) after contracting but *before* allocation. Thus, the principal can use the allocation together with transfers to elicit the agent’s new information. In our model, the new information (income) is received by the winner of the asset *after* the asset is allocated. Eliciting this information requires contingent payments.<sup>10</sup>

## 2. MODEL

**Setting** There is a principal and there are  $N$  agents, labeled  $i = 1, \dots, N$ . The principal has an asset to allocate. Each agent observes a private signal about his future income from acquiring the asset. Agent  $i$ ’s signal realization, denoted  $\theta_i$ , is called his type. Each agent  $i$ ’s type  $\theta_i$  is independently drawn from a distribution  $F_i$  with a continuous, strictly positive density  $f_i$  on its support  $\Theta_i = [\underline{\theta}_i, \bar{\theta}_i]$ . The space of type profiles is  $\Theta = \prod_{i=1}^N \Theta_i$ .

If agent  $i$ , with type  $\theta_i$ , receives the asset, then his income  $\pi_i$  is drawn from a distribution  $G_i(\cdot|\theta_i)$  with a continuous, strictly positive density  $g_i(\cdot|\theta_i)$  over the interval  $\Pi_i(\theta_i) := [\underline{\pi}_i(\theta_i), \bar{\pi}_i(\theta_i)]$ , where  $0 \leq \underline{\pi}_i(\theta_i) < \bar{\pi}_i(\theta_i) < \infty$ . In particular, the support of the conditional income distribution can shift with the agent’s type.<sup>11</sup> We assume that the functions  $\underline{\pi}_i$  and  $\bar{\pi}_i$  are continuous and weakly increasing. As in the main setting of [DeMarzo et al. \(2005\)](#), agent  $i$ ’s income distribution depends only on his own type, not on the types of other agents. This is the analog, in our setting, of the private values assumption.<sup>12</sup>

---

<sup>9</sup>These settings are all nested by the general multi-period setting of [Pavan et al. \(2014\)](#). That paper provides conditions under which an envelope formula holds.

<sup>10</sup>In [Mezzetti \(2004, 2007\)](#) *every* agent receives private information (his realized interdependent payoff) after the allocation. The principal can elicit this information, without verification, provided that each agent’s report affects the other agents’ payments, not his own.

<sup>11</sup>With shifting supports, our regularity assumptions (imposed below) become less restrictive. Shifting supports present a few technical issues, but these can be resolved, as first observed in [Liu et al. \(2020\)](#).

<sup>12</sup>Only the winner’s income is realized, so there is no need to specify the *joint* distribution of  $\pi_1, \dots, \pi_N$ . Thus, our private-values model implicitly allows for some dependence between different agents’ realized incomes. For example, incomes could be given by  $\pi_i = \theta_i + \varepsilon$ , for some aggregate shock  $\varepsilon$  that is independent of  $(\theta_1, \dots, \theta_N)$ .

Let  $\Pi_i = \cup_{\theta_i \in \Theta_i} \Pi_i(\theta_i)$ . For each  $\pi_i \in \Pi_i$ , let

$$\Theta_i(\pi_i) = \{\theta_i \in \Theta_i : \underline{\pi}_i(\theta_i) < \pi_i < \bar{\pi}_i(\theta_i)\}.$$

For each agent  $i$  and each  $\pi_i \in \Pi_i$ , we assume that  $G_i(\pi_i|\cdot)$  is continuously differentiable over  $\Theta_i(\pi_i)$  and that

$$G_{i,2}(\pi_i|\theta_i) < 0,$$

for each  $\theta_i$  in  $\Theta_i(\pi_i)$ . Thus, the conditional income distribution  $G_i(\cdot|\theta_i)$  is strictly increasing (in the sense of first-order stochastic dominance) in the type  $\theta_i$ . We normalize types so that for each agent  $i$ ,

$$(1) \quad \mathbb{E}[\pi_i|\theta_i] = \theta_i.$$

Thus, each agent's type equals his expected income from acquiring the asset. Finally, we assume that there exists a constant  $L > 0$  such that for each agent  $i$ , we have  $|G_i(\pi_i|\theta_i) - G_i(\pi_i|\theta'_i)| \leq L|\theta_i - \theta'_i|$  for all  $\theta_i, \theta'_i \in \Theta_i$  and all  $\pi_i \in \Pi_i$ . This Lipschitz condition, together with our differentiability assumptions, allows us to differentiate under integrals.

The principal can audit the winner of the asset. This audit perfectly reveals the winner's income.<sup>13</sup> The principal's auditing cost,  $c_i$ , can depend on the identity  $i$  of the audited agent.<sup>14</sup>

The principal and the agents are risk-neutral. There is no discounting.<sup>15</sup> The principal maximizes the expectation of payments minus auditing costs. Each agent maximizes the expectation of income minus payments.

---

<sup>13</sup>In Remark 1, we show that this assumption can be relaxed.

<sup>14</sup>We assume that the principal pays the auditing cost. In some models (e.g., [Diamond \(1984\)](#)), the agent pays the auditing cost. As long as  $c_i$  is commonly known, our results are unaffected by who formally pays this cost since the cost can be re-allocated through transfers.

<sup>15</sup>This simplifies notation. It is straightforward to incorporate a common discount rate for the principal and the agents.

**Discussion** We now discuss the independent private values assumption. In design problems, it is standard to assume that types are independently distributed; otherwise, under general conditions, there exist (unrealistic) mechanisms that fully extract the surplus (Crémer and McLean, 1988). Our independence assumption should be interpreted as *conditional independence*, given any information that is public (or that is held by multiple agents and hence can be elicited freely). For example, in Liu’s (2016) model of a takeover auction, the standalone value of the target firm is public, but each bidding firm privately knows its additional synergy value from acquiring the target. These synergy values are independent across bidders. Our setting admits a similar interpretation: the baseline value of the intellectual property is public, but each bidder has a private signal of his additional synergy value above this baseline. Of course, with a single potential buyer, the independent private values assumption holds vacuously.

**Protocol** The principal commits to her strategy in the following multi-stage protocol.

1. Each agent’s type is realized.
2. Each agent submits a message to the principal.
3. The principal allocates the asset to at most one agent (called the winner) and charges each agent a *transfer*.  
[If the asset is not allocated, there is no winner and the protocol ends.]
4. The winner’s income is realized.
5. The winner submits a message to the principal.
6. The principal charges the winner a *royalty*.
7. With some probability, the principal audits the winner.
8. If the winner is audited, the principal observes the winner’s realized income and then charges the winner a *penalty*.

Stages 1–3 are standard. If the asset is allocated, then the procedure continues. The royalty charged to the winner in stage 6 can depend on the winner’s message in stage 5 (as well as on the winner’s identity and the messages in stage 2). The penalty charged to the winner in stage 8, after auditing, can be contingent on the winner’s realized income (as well as on the winner’s identity and the messages in stages 2 and 5).

The terms *royalty* and *penalty* are suggestive of our motivating applications. Formally, the three kinds of payments—transfers, royalties, and penalties—are distinguished only by the information that they can depend on. This protocol is motivated by the timing used in many applications, but all payments could equivalently be delayed until the end of the game.

**Mechanisms** By the revelation principle, there is no loss in restricting attention to direct mechanisms in which the principal elicits a type report from each agent in stage 2 and an income report from the winner in stage 5. To represent the principal’s stochastic allocation decision, let

$$\mathcal{Q} = \{(q_1, \dots, q_N) \in \mathbf{R}_+^N : q_1 + \dots + q_N \leq 1\}.$$

In particular, the principal does not have to allocate the asset to any of the agents. If the profile  $\theta' = (\theta'_1, \dots, \theta'_N)$  is reported and the asset is allocated to agent  $i$ , then it is sufficient for the principal to solicit from agent  $i$  an income report in the set  $\Pi_i(\theta'_i)$ . Denote the set of report histories (when agent  $i$  is the winner) by

$$\mathcal{H}_i = \{(\theta'_1, \dots, \theta'_N, \pi'_i) \in \Theta \times \Pi_i : \pi'_i \in \Pi_i(\theta'_i)\}.$$

A direct mechanism for the principal specifies the following:

- allocation rule  $q = (q_1, \dots, q_N): \Theta \rightarrow \mathcal{Q}$ ;
- transfer rule  $t = (t_1, \dots, t_N): \Theta \rightarrow \mathbf{R}^N$ ;
- royalty rule  $r_i: \mathcal{H}_i \rightarrow \mathbf{R}$ , for each agent  $i$ ;
- auditing rule  $a_i: \mathcal{H}_i \rightarrow [0, 1]$ , for each agent  $i$ ;

- penalty rule  $p_i: \mathcal{H}_i \times \Pi_i \rightarrow \mathbf{R}$ , for each agent  $i$ .

A mechanism is denoted by  $(q, t, r, a, p)$ , where  $r = (r_1, \dots, r_N)$ ,  $a = (a_1, \dots, a_N)$ , and  $p = (p_1, \dots, p_N)$ . The allocation and transfer rules are standard. We describe the royalty, auditing, and penalty rules. Suppose that the type profile  $\theta' = (\theta'_1, \dots, \theta'_N)$  is reported and the principal allocates the asset to agent  $i$ .<sup>16</sup> Since agent  $i$  is the winner, the mechanism applies the rules  $r_i$ ,  $a_i$ , and  $p_i$ . If agent  $i$  reports income  $\pi'_i \in \Pi_i(\theta'_i)$ , then the principal charges agent  $i$  the royalty payment  $r_i(\theta', \pi'_i)$ . Then the principal audits agent  $i$  with probability  $a_i(\theta', \pi'_i)$ . If the audit is conducted, then agent  $i$ 's realized income  $\pi_i \in \Pi_i$  is revealed to the principal, and the principal charges agent  $i$  the penalty  $p_i(\theta', \pi'_i, \pi_i)$ .<sup>17</sup>

Crucially, we impose a constraint on how the winner's payments depend on his realized income from the asset. Fix parameters  $\phi_1, \dots, \phi_N \in [0, 1]$ .

**Condition A** (Generalized double monotonicity). For each agent  $i$  and each report history  $(\theta', \pi'_i)$  in  $\mathcal{H}_i$ , the function  $p_i(\theta', \pi'_i, \cdot)$  satisfies, for all  $\pi_i, \hat{\pi}_i \in \Pi_i$ ,

$$\hat{\pi}_i > \pi_i \implies 0 \leq p_i(\theta', \pi'_i, \hat{\pi}_i) - p_i(\theta', \pi'_i, \pi_i) \leq (\hat{\pi}_i - \pi_i)\phi_i.$$

Condition A requires that for each report history, agent  $i$ 's post-audit penalty is weakly increasing in his realized income, and the rate of increase is at most  $\phi_i$ .<sup>18</sup> Since the penalty is the only payment that depends directly on realized income, Condition A equivalently restricts the sensitivity of agent  $i$ 's total payment (summing transfers, royalties, and penalties) to his realized income. In words, if agent  $i$  wins the asset, then for each additional dollar

---

<sup>16</sup>If  $q_i(\theta') = 0$ , then this history  $(\theta', \pi'_i)$  cannot be reached. Under our definition, a mechanism specifies decisions at more histories than is necessary. This convention simplifies notation.

<sup>17</sup>If agent  $i$  wins the asset after reporting  $\theta'_i$ , then the principal accepts income *reports* only from  $\Pi_i(\theta'_i)$ . But the mechanism must still specify the outcome if the agent's true income is subsequently revealed to be outside  $\Pi_i(\theta'_i)$ ; such an income realization proves that the agent misreported his type (and his income).

<sup>18</sup>The upper bound on the slope is critical; the lower bound is not. If we reduced the lower bound to  $-(\hat{\pi}_i - \pi_i)C$  for any fixed constant  $C > 0$ , the mechanism in Theorem 1 would remain optimal. Luo and Yang (2023) consider securities that are monotone and Lipschitz, but their focus is on coordination frictions.

generated by the asset, agent  $i$  must retain at least a fraction  $1 - \phi_i$  of the dollar (and at most the full dollar).<sup>19</sup>

If all the parameters  $\phi_i$  equal 1, then Condition A reduces to *double monotonicity*: both maps  $\pi_i \mapsto p_i(\theta', \pi'_i, \pi_i)$  and  $\pi_i \mapsto \pi_i - p_i(\theta', \pi'_i, \pi_i)$  are weakly increasing. Double monotonicity is a standard assumption in the security design literature; see, e.g., DeMarzo et al. (2005), Figueroa and Inostroza (2025), and Nachman and Noe (1994). The double-monotonicity assumption is motivated by an informal moral hazard argument. If the payment by the asset-holder ever decreases in the asset's income, then the asset-holder could generate a risk-free return by injecting cash into the project to raise measured income, thus decreasing the payment that he must make to the principal. Conversely, if the asset-holder's net profit ever decreases in the asset's income, then he could strictly benefit by burning the asset's cash flow. In many applications, the asset-holder can do better than burning money—he can divert cash for his own benefit. If the asset-holder can use fraction  $1 - \phi_i$  of each dollar that he diverts, then our generalization of double monotonicity is the appropriate condition.<sup>20</sup> More broadly, our condition captures the moral hazard concern, noted by (DeMarzo et al., 2005, p. 950), that the asset-holder's incentives are dampened if he retains too small a share of the asset's income.

In summary, the model primitives specify, for each agent  $i$ , the type distribution  $F_i$ , conditional income distribution  $G_i$ , auditing cost  $c_i$ , and maximum penalty sensitivity  $\phi_i$ .

---

<sup>19</sup>Condition A has antecedents in the literature on tax evasion. In their classic work, Allingham and Sandmo (1972) assume an exogenous linear penalty of  $(\pi_i - \pi'_i)\phi_i$ ; see also Kleven et al. (2011) and Palonen and Pekkarinen (2026). We restrict the maximum sensitivity of the penalty, but we allow nonlinear penalties.

<sup>20</sup>Condition A can alternatively be microfounded as follows. Suppose that an audit reveals agent  $i$ 's income with probability  $\phi_i$  and reveals nothing otherwise. In this case, requiring the penalty to be double monotone in the winner's income (when it is revealed) is equivalent to requiring the *expected* penalty to satisfy Condition A.

### 3. PRINCIPAL'S PROGRAM

We now formulate the principal's optimization problem. To state the incentive constraints, we introduce additional notation. For each allocation rule  $q_i$  and transfer rule  $t_i$ , define the associated interim rules by

$$Q_i(\theta'_i) = \mathbb{E}_{\theta_{-i}} [q_i(\theta'_i, \theta_{-i})], \quad T_i(\theta'_i) = \mathbb{E}_{\theta_{-i}} [t_i(\theta'_i, \theta_{-i})].$$

Consider a mechanism  $(q, t, r, a, p)$ . Fix an agent  $i$ . For any type report  $\theta'_i \in \Theta_i$ , income report  $\pi'_i \in \Pi_i(\theta'_i)$ , and true income  $\pi_i \in \Pi_i$ , define the utility

$$(2) \quad u_i(\theta'_i, \pi'_i | \pi_i) = \mathbb{E}_{\theta_{-i}} \left[ q_i(\theta'_i, \theta_{-i}) (\pi_i - r_i(\theta'_i, \theta_{-i}, \pi'_i) - a_i(\theta'_i, \theta_{-i}, \pi'_i)) p_i(\theta'_i, \theta_{-i}, \pi'_i, \pi_i) \right].$$

If  $Q_i(\theta'_i) = 0$ , then  $u_i(\theta'_i, \pi'_i | \pi_i) = 0$ . If  $Q_i(\theta'_i) > 0$ , then  $u_i(\theta'_i, \pi'_i | \pi_i) / Q_i(\theta'_i)$  is agent  $i$ 's expected gross utility (excluding upfront transfers) conditional upon reporting  $\theta'_i$ , winning the asset, privately observing income  $\pi_i$ , and then reporting income  $\pi'_i$ . This utility term  $u_i(\theta'_i, \pi'_i | \pi_i)$  allows us to conveniently express agent  $i$ 's income-reporting incentive constraint without conditioning on zero-probability events.

There is one subtlety introduced by shifting supports. If type  $\theta_i$  reports  $\theta'_i$  and  $\Pi_i(\theta'_i) \not\supseteq \Pi_i(\theta_i)$ , then the realized income  $\pi_i$  may be outside  $\Pi_i(\theta'_i)$ . In this case, it is not feasible for agent  $i$  to report his income truthfully. We introduce notation for reporting "as truthfully as possible." For any type report  $\theta'_i \in \Theta_i$  and any income  $\pi_i \in \Pi_i$ , let  $\text{proj}_{\Pi_i(\theta'_i)} \pi_i$  denote the element in  $\Pi_i(\theta'_i)$  that is closest to  $\pi_i$ . In particular,  $\text{proj}_{\Pi_i(\theta'_i)} \pi_i = \pi_i$  if  $\pi_i$  is in  $\Pi_i(\theta'_i)$ . For any types  $\theta_i, \theta'_i \in \Theta_i$ , let

$$(3) \quad U_i(\theta'_i | \theta_i) = \mathbb{E}_{\pi_i | \theta_i} \left[ u_i(\theta'_i, \text{proj}_{\Pi_i(\theta'_i)} \pi_i | \pi_i) \right].$$

This expectation is taken with respect to the true income distribution  $G_i(\cdot | \theta_i)$ . Thus,  $U_i(\theta'_i | \theta_i)$  is the expected gross utility (excluding upfront transfers) for type  $\theta_i$  if he reports type  $\theta'_i$  and then reports his income as truthfully as possible

whenever he wins the asset.

We consider the principal's problem of choosing a direct mechanism  $(q, t, r, a, p)$  satisfying Condition A to solve:

$$\max \sum_{i=1}^N \mathbb{E}_{\theta} [t_i(\theta) + q_i(\theta) \mathbb{E}_{\pi_i|\theta_i} [r_i(\theta, \pi_i) + a_i(\theta, \pi_i)(p_i(\theta, \pi_i, \pi_i) - c_i)]]$$

subject to the following constraints, for each agent  $i$ :

$$\begin{aligned} (\text{IC}_{\pi}) \quad & u_i(\theta_i, \pi_i|\pi_i) \geq u_i(\theta_i, \pi'_i|\pi_i), \quad \theta_i \in \Theta_i, \quad \pi_i, \pi'_i \in \Pi_i(\theta_i) \\ (\text{IC}_{\theta}) \quad & U_i(\theta_i|\theta_i) - T_i(\theta_i) \geq U_i(\theta'_i|\theta_i) - T_i(\theta'_i), \quad \theta_i, \theta'_i \in \Theta_i \\ (\text{IR}) \quad & U_i(\theta_i|\theta_i) - T_i(\theta_i) \geq 0, \quad \theta_i \in \Theta_i. \end{aligned}$$

In the principal's objective, the  $i$ -th term in the summation is the expected payment made by agent  $i$  (including transfers, royalties, and penalties) net of the principal's expected cost of auditing agent  $i$ . Technically, some of the constraints are relaxations of the full incentive constraints, as we describe below. We will solve this relaxed problem and then confirm that our solution satisfies the full incentive constraints.

The constraint  $(\text{IC}_{\pi})$  captures incentive compatibility for the winner of the asset at the income-reporting stage.<sup>21</sup> Suppose that type  $\theta_i$  reports truthfully. If  $Q_i(\theta_i) > 0$ , then there is a positive probability that agent  $i$  wins the asset. In this case, the constraint  $(\text{IC}_{\pi})$  ensures that if agent  $i$  wins the asset, it is optimal for him to truthfully report his income realization, whatever its value. If  $Q_i(\theta_i) = 0$ , then  $(\text{IC}_{\pi})$  automatically holds—both sides equal zero by the definition of  $u_i$ .<sup>22</sup>

---

<sup>21</sup>In particular,  $(\text{IC}_{\pi})$  takes into account agent  $i$ 's inference (about the other agents' type reports) from winning the asset. The constraint  $(\text{IC}_{\pi})$  does not take into account agent  $i$ 's further inference from the upfront transfer that he is charged. Thus,  $(\text{IC}_{\pi})$  is technically a relaxation of the income-reporting incentive constraint. In our solution of the relaxed problem, the reporting, auditing, and penalty rules for the winner are all independent of the other agents' type reports. Thus, the winner does not benefit from learning the others' type reports.

<sup>22</sup>A similar approach is followed in the definition of Bayes correlated equilibrium (Bergemann and Morris, 2016, Definition 1, p. 493).

The constraint  $(\text{IC}_\theta)$  says that every type  $\theta_i$  weakly prefers (a) truthfully reporting his type and then truthfully reporting his subsequent income if he wins the asset, to (b) misreporting his type and then reporting his income “as truthfully as possible” if he wins the asset. This is a relaxation of the full incentive-compatibility constraint, which requires that all double deviations are unprofitable.<sup>23</sup>

Finally,  $(\text{IR})$  ensures that every type weakly prefers participating in the mechanism to his outside option of zero utility.

#### 4. BENCHMARKS

In this section, we describe the optimal mechanism in two extreme cases: non-contingent payments and fully contingent payments.

**Non-contingent payments** Suppose that the principal cannot charge payments that are contingent on the winner’s realized income, i.e.,  $\phi_i = 0$  for all  $i$ .<sup>24</sup> After the asset is allocated, the principal can still ask the winner to report his income. But without the threat of contingent payments, the winner will make whichever income report minimizes his expected payment. Thus, incentive compatibility implies that the winner’s expected post-allocation payment cannot depend on his reported income. Therefore, expected payments can depend only on the agents’ initial type reports, and the model reduces to the classical cash auction setting of Myerson (1981). Agent  $i$ ’s valuation for the asset is  $\mathbb{E}[\pi_i|\theta_i]$ , which equals  $\theta_i$  by our normalization. The optimal cash auction allocates the asset according to each agent  $i$ ’s Myerson virtual value

$$\psi_i^M(\theta_i) = \theta_i - \frac{1 - F_i(\theta_i)}{f_i(\theta_i)}.$$

---

<sup>23</sup>This relaxation can be strict only with shifting supports. With common support,  $(\text{IC}_\pi)$  implies that it is optimal for the winner to report his income truthfully, even after misreporting his type.

<sup>24</sup>This benchmark also corresponds to the limit as auditing becomes prohibitively costly.

Transfers are pinned down by the envelope theorem. In the main model with verification and penalties, this cash auction is always feasible, but it is generally suboptimal.

**Fully contingent payments** Suppose that the winner’s income from the asset is publicly revealed and that the principal can charge the winner a payment that depends arbitrarily on the realized income. In this case, the principal can fully extract the surplus using the following modified first-price auction. The allocation and upfront transfers are the same as in a standard first-price auction. After allocating the asset, the principal charges the winner a penalty equal to the difference between his realized income and his earlier bid.

In this auction, no matter how the agents bid, the winner’s total payment equals his realized income from the asset, and the losers pay nothing. Each agent’s expected utility is zero, and hence every strategy profile is a dominant-strategy equilibrium. In the truthful equilibrium, the asset is allocated to the agent with the highest type, so the principal extracts the full surplus  $\mathbb{E}[\max\{0, \theta_1, \dots, \theta_N\}]$ . This full-extraction mechanism is feasible in the main model if auditing is costless ( $c_i = 0$  for all  $i$ ) and  $\phi_i = 1$  for each  $i$ .<sup>25</sup>

If penalties can depend arbitrarily on realized income, then even if auditing is costly, the principal can approximate full extraction by adjusting the modified first-price auction as follows. For any  $\varepsilon > 0$ , the principal can audit the winner with probability  $\varepsilon$  and then scale the penalty rule by  $1/\varepsilon$ . Expected payments are unchanged, and the auditing cost vanishes in the limit as  $\varepsilon$  tends to 0. In our model, Condition A rules out such mechanisms.

## 5. SOLVING FOR AN OPTIMAL MECHANISM

In this section, we use the Myersonian approach to solve for an optimal mechanism. Then we analyze comparative statics.

---

<sup>25</sup>Similarly, Cr mer (1987) observes that in the setting of Hansen (1985), full extraction is possible if the limited liability constraint is dropped.

## 5.1 MYERSONIAN APPROACH

We use the envelope theorem to obtain an upper bound on the principal's objective. This bound depends only on the allocation and auditing rules, not on any of the payment rules. First, we introduce additional notation. For each  $\theta_i \in \Theta_i$  and each  $\pi_i \in (\underline{\pi}_i(\theta_i), \bar{\pi}_i(\theta_i))$ , let

$$\mu_i(\theta_i, \pi_i) = -\frac{G_{i,2}(\pi_i|\theta_i)}{g_i(\pi_i|\theta_i)} \cdot \frac{1 - F_i(\theta_i)}{f_i(\theta_i)}.$$

Since the partial derivative  $G_{i,2}$  is strictly negative, we have  $\mu_i(\theta_i, \pi_i) > 0$  if  $\theta_i < \bar{\theta}_i$ , and we have  $\mu_i(\bar{\theta}_i, \pi_i) = 0$ .

**Lemma 1** (Payoff bound)

Let  $(q, t, r, a, p)$  be a mechanism that satisfies Condition A and the constraints (IC $_{\pi}$ ), (IC $_{\theta}$ ), and (IR) for each agent  $i$ . The principal's expected payoff from  $(q, t, r, a, p)$  is at most

$$\mathbb{E}_{\theta} \left[ \sum_{i=1}^N q_i(\theta) \Psi_i(\theta) \right],$$

where for each agent  $i$ , the function  $\Psi_i: \Theta \rightarrow \mathbf{R}$  is given by

$$\Psi_i(\theta) = \theta_i - \frac{1 - F_i(\theta_i)}{f_i(\theta_i)} + \mathbb{E}_{\pi_i|\theta_i} [a_i(\theta, \pi_i)(\mu_i(\theta_i, \pi_i)\phi_i - c_i)].$$

To prove Lemma 1, we apply the envelope theorem to the income-reporting and type-reporting incentive constraints. The difficulty is that the penalty can depend on the winner's *realized* income, so the allocation rule does not pin down the agents' information rents. Using Condition A, we derive a lower bound on each agent's information rent, which yields an upper bound on the principal's payoff. Below, we will find a mechanism that achieves this upper bound and hence is optimal.

In Lemma 1, the coefficient  $\Psi_i(\theta)$  is an *endogenous* virtual value that depends on the full type profile  $\theta$  and on the principal's auditing rule. Note that  $\Psi_i(\theta)$  is the sum of the Myerson virtual value  $\psi_i^M(\theta_i)$  and an additional expectation term, which we call the *auditing term*. This term takes a simple

form in two extreme cases. If the principal never audits agent  $i$ , then the auditing term vanishes and we get  $\Psi_i(\theta) = \psi_i^M(\theta_i)$ . If the principal always audits agent  $i$ , then it can be checked that

$$\Psi_i(\theta) = \theta_i - (1 - \phi_i) \frac{1 - F_i(\theta_i)}{f_i(\theta_i)} - c_i.$$

The sensitivity parameter  $\phi_i$  controls the share of agent  $i$ 's information rent that the principal can recover through penalties, upon auditing the agent.

In the auditing term, the expression inside the expectation reflects the direct cost  $c_i$  of auditing agent  $i$  and the indirect benefit from reducing agent  $i$ 's information rent. Here is a heuristic derivation. Auditing agent  $i$  at history  $(\theta, \pi_i)$  allows the principal to increase the slope of the royalty payment at that history by at most  $\phi_i$ . This raises the royalty paid by type  $\theta_i$  in the event that his realized income is at least  $\pi_i$ . The probability of this event is  $1 - G_i(\pi_i|\theta_i)$ . The sensitivity of this probability to agent  $i$ 's true type is  $-G_{i,2}(\pi_i|\theta_i)$ . Therefore, by the envelope theorem, auditing at history  $(\theta, \pi_i)$  reduces by at most  $-G_{i,2}(\pi_i|\theta_i)\phi_i q_i(\theta)$  the information rent of every type above  $\theta_i$ . The mass of such types is  $1 - F_i(\theta_i)$ , so the expected information rent is reduced by at most

$$-G_{i,2}(\pi_i|\theta_i)(1 - F_i(\theta_i))\phi_i q_i(\theta).$$

In the auditing term, this coefficient on  $q_i(\theta)$  is divided by the density  $g_i(\pi_i|\theta_i)f_i(\theta_i)$  to offset the scaling from the expectation over  $(\theta_i, \pi_i)$ .

## 5.2 OPTIMAL MECHANISM

To solve for an optimal mechanism, we find allocation and auditing rules that pointwise maximize the expression inside the expectation in the payoff bound (Lemma 1). Under suitable regularity conditions, we show that this mechanism satisfies global incentive compatibility.

The endogenous virtual value,  $\Psi_i(\theta)$ , depends on the principal's auditing rule. To maximize  $\Psi_i(\theta)$ , the principal audits agent  $i$  if and only if the

resulting reduction in information rent outweighs the direct cost of the audit:  $\mu_i(\theta_i, \pi_i)\phi_i \geq c_i$ . With this optimal auditing rule, the endogenous virtual value  $\Psi_i(\theta)$  becomes  $\psi_i(\theta_i)$ , where

$$\psi_i(\theta_i) = \theta_i - \frac{1 - F_i(\theta_i)}{f_i(\theta_i)} + \mathbb{E}_{\pi_i|\theta_i} [\max\{\mu_i(\theta_i, \pi_i)\phi_i - c_i, 0\}].$$

Hereafter, we call  $\psi_i(\theta_i)$  the *virtual value* of type  $\theta_i$ . Unlike  $\Psi_i(\theta)$ , the virtual value  $\psi_i(\theta_i)$  depends only on agent  $i$ 's type (not on the principal's auditing rule or on the types of the other agents).<sup>26</sup>

In order to solve for an optimal mechanism, we introduce suitable regularity conditions. Consider a real-valued function  $h$  whose domain,  $\text{dom } h$ , is a convex subset of  $\mathbf{R}$ . We say that  $h$  is *single-crossing from above* if for all  $x, x' \in \text{dom } h$  with  $x < x'$ , we have:  $h(x) \leq (<) 0 \implies h(x') \leq (<) 0$ .

**Assumption 1** (Regularity). For each agent  $i$ , the following hold:

- (a)  $\mu_i(\theta_i, \pi_i)\phi_i - c_i$  is single-crossing from above in  $\theta_i$  and in  $\pi_i$ ;
- (b)  $\psi_i(\theta_i)$  is strictly increasing in  $\theta_i$ .

Assumption 1.a ensures that the pointwise optimal rule for auditing agent  $i$  is weakly decreasing in agent  $i$ 's type and income. Assumption 1.b is the analog of Myerson regularity. It ensures that the virtual surplus-maximizing allocation rule for agent  $i$  is weakly increasing in agent  $i$ 's type, for each fixed type profile of the other agents. Assumption 1 is strictly weaker than the regularity assumptions in [Esó and Szentes \(2007\)](#), as we show in [Appendix A.2](#).

To describe our optimal mechanism, we introduce additional notation. Using Assumption 1.a, we define for each agent  $i$  a weakly decreasing function

---

<sup>26</sup>[Liu \(2016\)](#) derives an expression for the virtual value in equity auctions. The exact formula is different from ours, but both expressions reflect the fact that contingent payments reduce information rents. [Ball and Kattwinkel \(2025\)](#) also derive an expression for the virtual value in their model of probabilistic verification. Before the allocation, the principal can generate a binary, noisy signal of whether the agent is misreporting, and the principal can exclude the agent if a misreport is detected.

$\pi_i^*: \Theta_i \rightarrow [0, \infty]$  with the property that for all types  $\theta_i \in \Theta_i$  and all  $\pi_i \in (\underline{\pi}_i(\theta_i), \bar{\pi}_i(\theta_i))$ , we have  $\pi_i \leq \pi_i^*(\theta_i) \iff \mu_i(\theta_i, \pi_i)\phi_i \geq c_i$ .<sup>27</sup> Let

$$(4) \quad \Phi_i(\theta_i) = -\phi_i \int_0^{\pi_i^*(\theta_i)} G_{i,2}(\pi_i|\theta_i) d\pi_i.$$

The normalization (1) implies that  $-G_{i,2}(\cdot|\theta_i)$  integrates to 1. Since  $-G_{i,2}(\cdot|\theta_i)$  is nonnegative, we have  $0 \leq \Phi_i(\theta_i) \leq \phi_i$ . To describe our optimal mechanism, let  $[\cdot]$  denote the indicator function for the predicate it encloses.

**Theorem 1** (Optimal mechanism)

*Under Assumption 1, the following mechanism  $(q^*, t^*, r^*, a^*, p^*)$  is optimal. For each agent  $i$ , the allocation and auditing rules are given by<sup>28</sup>*

$$q_i^*(\theta) = [\psi_i(\theta_i) \geq 0 \text{ and } \psi_i(\theta_i) > \max_{j \neq i} \psi_j(\theta_j)], \quad a_i^*(\theta, \pi'_i) = [\pi'_i < \pi_i^*(\theta_i)].$$

*The royalties and penalties are given by*

$$r_i^*(\theta, \pi'_i) = \min\{\pi'_i, \pi_i^*(\theta_i)\}\phi_i, \quad p_i^*(\theta, \pi'_i, \pi_i) = \min\{\pi_i, \pi_i^*(\theta_i)\}\phi_i - \min\{\pi'_i, \pi_i^*(\theta_i)\}\phi_i.$$

*The upfront transfers are given by*

$$t_i^*(\theta) = q_i^*(\theta) \mathbb{E}_{\pi_i|\theta_i} [\pi_i - r_i^*(\theta, \pi_i)] - \int_{\underline{\theta}_i}^{\theta_i} q_i^*(z_i, \theta_{-i})(1 - \Phi_i(z_i)) dz_i.$$

*Moreover,  $(q^*, t^*, r^*, a^*, p^*)$  is dominant-strategy incentive compatible and dominant-strategy individually rational.*

We discuss the components of the mechanism  $(q^*, t^*, r^*, a^*, p^*)$  in turn. The asset is allocated to the agent with the highest virtual value, provided that

<sup>27</sup>For each  $\theta_i \in \Theta_i$ , let  $\Pi_i^*(\theta_i) = \{\pi_i \in (\underline{\pi}_i(\theta_i), \bar{\pi}_i(\theta_i)) : \mu_i(\theta_i, \pi_i)\phi_i \geq c_i\}$ . Define  $\pi_i^*$  by

$$\pi_i^*(\theta_i) = \begin{cases} \infty & \text{if } \Pi_i^*(\theta_i) = (\underline{\pi}_i(\theta_i), \bar{\pi}_i(\theta_i)), \\ \sup \Pi_i^*(\theta_i) & \text{if } \emptyset \subsetneq \Pi_i^*(\theta_i) \subsetneq (\underline{\pi}_i(\theta_i), \bar{\pi}_i(\theta_i)), \\ 0 & \text{if } \Pi_i^*(\theta_i) = \emptyset. \end{cases}$$

<sup>28</sup>To cover the case  $N = 1$ , we adopt the convention that the maximum of the empty set is  $-\infty$ .

this virtual value is nonnegative.<sup>29</sup> Otherwise, the asset is not allocated. The allocation probability  $q_i^*$  is weakly increasing in agent  $i$ 's type  $\theta_i$  because  $\psi_i$  is strictly increasing (by Assumption 1.b). The virtual value  $\psi_i(\theta_i)$  is weakly greater than the Myerson virtual value  $\psi_i^M(\theta_i)$ . Relative to Myerson's optimal cash auction, the asset is allocated at more type profiles (in the sense of weak set inclusion).

Next, we describe the auditing, royalty, and penalty rules. Agent  $i$ 's type report  $\theta_i$  determines his *royalty cap*  $\pi_i^*(\theta_i)\phi_i$ . Suppose that agent  $i$  wins the asset after reporting type  $\theta_i$ . If agent  $i$  subsequently reports income of at least  $\pi_i^*(\theta_i)$ , then he pays a royalty equal to the royalty cap  $\pi_i^*(\theta_i)\phi_i$  and he is not audited. If he instead reports income below  $\pi_i^*(\theta_i)$ , then he pays a royalty equal to the fraction  $\phi_i$  of his reported income, and he is subsequently audited. If the audit reveals that his true income is different from his reported income, then he is charged a penalty equal to the underpaid royalties (or he is given a refund for the overpaid royalties).<sup>30</sup> This penalty format is consistent with the medical device royalty agreement described in the introduction.

The royalty cap  $\pi_i^*(\theta_i)\phi_i$  is weakly decreasing in agent  $i$ 's type report  $\theta_i$ ; we show this in the proof of Theorem 1. If agent  $i$  reports a higher type, then upon winning the asset, he faces a lower royalty cap and hence is less likely to be audited. From a statistical standpoint, this auditing rule may be counterintuitive. When an agent's true type is higher, a low income report is *more suspicious*. But in equilibrium, all reports are truthful. The optimal auditing rule prescribes the minimal level of auditing that preserves the incentive compatibility of the associated royalty rule. In turn, the royalty rule optimally screens the agents at the type-reporting stage according to their beliefs about future income. Higher type reports result in lower royalty caps, which reward high income realizations. These high realizations are more likely

---

<sup>29</sup>Since  $\psi_i$  is strictly increasing, ties between agents occur with probability zero. To keep notation simple, we assume that the asset is not allocated in the case of a tie.

<sup>30</sup>Refunding the agent for overpaid royalties is not necessary. It would also be optimal to use the penalty function  $\bar{p}_i(\theta, \pi'_i, \pi_i) = \max\{p_i^*(\theta, \pi'_i, \pi_i), 0\}$ . More generally,  $p_i^*$  can be replaced with any feasible penalty function that agrees with  $p_i^*$  on path and is weakly larger than  $p_i^*$  off path.

for higher types. In particular, the highest type,  $\bar{\theta}_i$ , is never audited (provided that  $c_i > 0$ ), and the virtual value  $\psi_i(\bar{\theta}_i)$  equals the true value  $\bar{\theta}_i$ . Allocating to type  $\bar{\theta}_i$  does not increase information rent for higher types, so there is no motive to downward distort or audit.

Finally, the interim transfers are pinned down by incentive compatibility. The information rent for type  $\theta_i$  is given by

$$\int_{\theta_i}^{\theta_i} Q_i^*(z_i) (1 - \Phi_i(z_i)) dz_i.$$

This expression is similar to that in a cash auction, but here the allocation probability in the integral is scaled by the factor  $1 - \Phi_i$ . From (4), we see that for each report  $\theta_i$ , the term  $\Phi_i(\theta_i)$  is between 0 (when no income is subject to royalties) and  $\phi_i$  (when all income is subject to royalties at rate  $\phi_i$ ).

Bayesian incentive compatibility depends only on interim payments. The particular payment rules in Theorem 1 are chosen for their special properties. Transfers and royalty payments are always nonnegative. Equilibrium penalties are zero. Agent  $i$  pays an upfront transfer only if he wins the asset. Moreover, the mechanism  $(q^*, t^*, r^*, a^*, p^*)$  is dominant-strategy incentive compatible and dominant-strategy individually rational. Whatever agent  $i$  believes about his opponents' reports, it is optimal for him to participate in the mechanism and then report truthfully at every stage. Since  $q_i^*(\theta_i, \theta_{-i})$  is weakly increasing in  $\theta_i$  and  $\pi_i^*(\theta_i)$  is weakly decreasing in  $\theta_i$ , it follows from dominant-strategy incentive compatibility that  $t_i^*(\theta_i, \theta_{-i})$  is weakly increasing in  $\theta_i$ .

**Remark 1** (Noisy auditing). We have assumed that an audit perfectly reveals the winner's income. If the auditing technology instead generates a noisy, unbiased signal of the winner's income, then we show that the principal's optimal payoff is unchanged. Formally, suppose that an audit of the winning agent  $i$  generates a random signal  $\zeta_i$ , where  $\mathbb{E}[\zeta_i | \pi_i] = \pi_i$ . Assume that the conditional signal distribution is weakly increasing in  $\pi_i$  (with respect to first-order stochastic dominance). We impose the analog of Condition A, with  $\zeta_i$  in place of  $\pi_i$ . We show that the principal's optimal payoff is the same as in the main model. This payoff is achieved by using essentially the same mechanism

as in Theorem 1, except that the penalty rule is  $\tilde{p}_i(\theta, \pi'_i, \zeta_i) = (\zeta_i - \pi'_i)\phi_i$ . Thus,  $\mathbb{E}_{\zeta_i|\pi_i}[\tilde{p}_i(\theta, \pi'_i, \zeta_i)] = (\pi_i - \pi'_i)\phi_i$ . For the proof, see Appendix A.5.

We next introduce a general class of distributions, termed *transformed additive noise*, that satisfy Assumption 1.<sup>31</sup> It is convenient to first introduce these distributions in terms of an unnormalized signal,  $\gamma_i$ , for agent  $i$ . Then we normalize agent  $i$ 's signal by setting  $\theta_i = \mathbb{E}[\pi_i|\gamma_i]$ . Suppose that for each agent  $i$ , income is given by

$$\pi_i = \alpha_i(\gamma_i + \varepsilon_i),$$

where

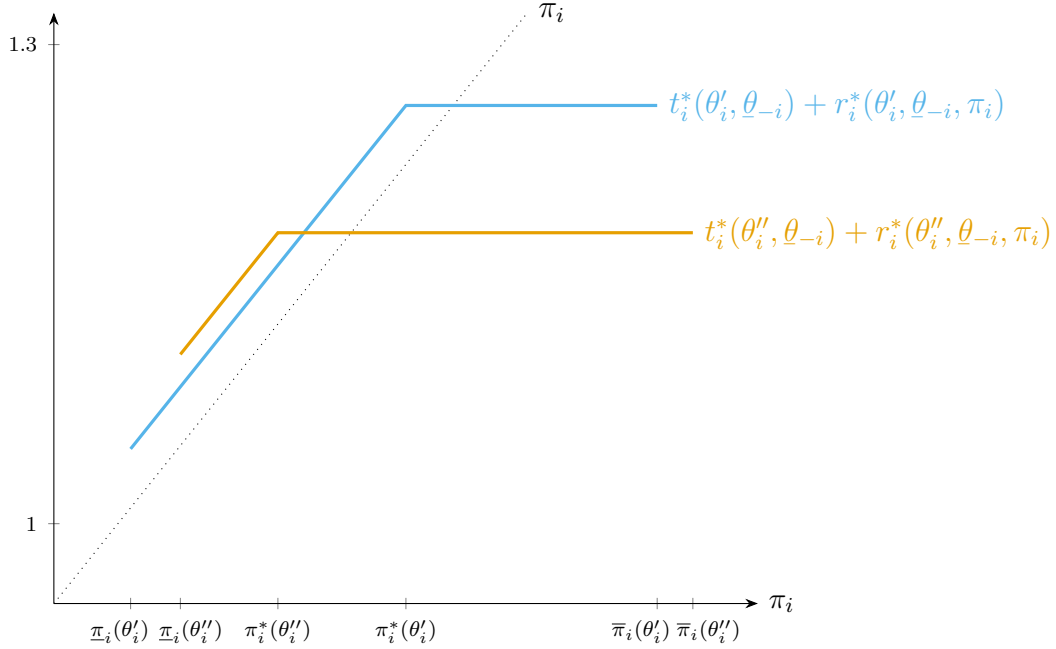
- $\gamma_i$  is agent  $i$ 's (unnormalized) signal, which has a continuous, positive density and a weakly increasing hazard rate over its support  $[\underline{\gamma}_i, \bar{\gamma}_i]$ ;
- $\varepsilon_i$  is a random variable, independent of  $\gamma = (\gamma_1, \dots, \gamma_N)$ , with continuous, positive density over its support  $[\underline{\varepsilon}_i, \bar{\varepsilon}_i]$ ;
- $\alpha_i: [\underline{\gamma}_i + \underline{\varepsilon}_i, \bar{\gamma}_i + \bar{\varepsilon}_i] \rightarrow [0, \infty)$  is a twice continuously differentiable function satisfying  $\alpha'_i(x) > 0$  and  $\alpha''_i(x) \leq 0$  for all  $x \in [\underline{\gamma}_i + \underline{\varepsilon}_i, \bar{\gamma}_i + \bar{\varepsilon}_i]$ .

With the renormalization  $\theta_i = \mathbb{E}[\pi_i|\gamma_i]$  for each agent  $i$ , Assumption 1 is satisfied; see Appendix A.3 for a proof. In the special case that  $\alpha_i$  is the identity function and  $\varepsilon_i$  has mean zero, no normalization is required, and we get  $\pi_i = \theta_i + \varepsilon_i$ . We call this special case the *additive noise* specification. Under the additive noise specification, it can be verified that  $\mu_i(\theta_i, \pi_i)$  depends only on  $\theta_i$ , not on  $\pi_i$ .

Figure 1 illustrates the mechanism from Theorem 1 in an example with transformed additive noise. Suppose that the  $N$  agents are symmetric. For each agent  $i$ , let  $\pi_i = \sqrt{\gamma_i + \varepsilon_i}$ , where agent  $i$ 's unnormalized signal  $\gamma_i$  is uniformly distributed over  $[1, 2]$ , and  $\varepsilon_i$  is independently and uniformly distributed

---

<sup>31</sup>This class generalizes many of the examples in Eső and Szentes (2007). Note that Example 3 in Eső and Szentes (2007, p. 712) contains an algebraic error: the correct mixed partial  $u_{i12}$  is nonnegative. This example must be modified to satisfy their regularity conditions.



**Figure 1:** Winner's total equilibrium payment as a function of income for two type reports.

over  $[0, 1]$ . Agent  $i$ 's normalized type is given by  $\theta_i(\gamma_i) = \mathbb{E}[\pi_i|\gamma_i]$ . In this case, we have  $\mu_i(\theta_i(\gamma_i), \pi_i) = (2 - \gamma_i)/(2\pi_i)$ . Let  $\phi_i = 1$  and  $c_i = 0.4$ . Given two types  $\theta'_i$  and  $\theta''_i$  with  $\theta'_i < \theta''_i$ , we consider the case in which agent  $i$ 's opponents all report the lowest possible type.<sup>32</sup> The plot is similar for other opponent reports, provided that agent  $i$  still wins the asset.<sup>33</sup> For each of these two types, Figure 1 plots agent  $i$ 's total equilibrium payment as a function of his income realization. Equilibrium penalties are zero, so they are omitted from the formulas.

In Figure 1, the type- $\theta''_i$  payment curve crosses the type- $\theta'_i$  payment curve

<sup>32</sup>This vector of opponent reports is denoted by  $\underline{\theta}_{-i}$ . Here,  $\Theta_i \approx [1.22, 1.58]$  for all  $i$ . The two types in the plot are  $\theta'_i \approx 1.23$  and  $\theta''_i \approx 1.26$  (corresponding to the unnormalized signals  $\gamma'_i = 1.02$  and  $\gamma''_i = 1.10$ ). The associated income supports are  $\Pi_i(\theta'_i) \approx [1.01, 1.42]$  and  $\Pi_i(\theta''_i) \approx [1.05, 1.45]$ , and the auditing thresholds are  $\pi_i^*(\theta'_i) \approx 1.23$  and  $\pi_i^*(\theta''_i) \approx 1.13$ . In this example, types above approximately 1.28 (equivalently, unnormalized signals above approximately 1.14) are never audited.

<sup>33</sup>More precisely, with a higher vector of opponent reports, each equilibrium payment curve is translated upward.

from above. That is, the higher type makes higher payments at low income realizations and lower payments at high income realizations. This crossing property is quite general; it follows from dominant-strategy incentive compatibility, as we show in the proof of Theorem 1. Agent  $i$ 's realized utility is the difference between  $\pi_i$  (shown as the 45-degree line) and the payment curve. For both types  $\theta'_i$  and  $\theta''_i$ , the realized utility is negative for some low income realizations and positive for some high income realizations. Individual rationality ensures that for each type, the expected difference between  $\pi_i$  and the payment curve is nonnegative, where the expectation is with respect to the income distribution for that type.

Figure 1 highlights that each equilibrium payment resembles a debt contract. The slope before the cap is  $\phi_i$  because generalized double monotonicity (Condition A) binds. In a debt contract, the corresponding slope is 1 because the standard double-monotonicity constraint binds. The flat section beyond the cap, like that in a debt contract, saves on auditing costs. Our model shows that the intuition from the static, single-agent model of [Townsend \(1979\)](#) extends to a richer setting with sequential screening. By contrast, in the optimal security-bid auction in [DeMarzo et al. \(2005\)](#), the equilibrium payments are call options, which are flat up to a threshold, and then increasing with slope 1; for a more detailed comparison with their results, see Section 5.4.

### 5.3 SPECIAL CASES

We consider two special cases in which the optimal mechanism in Theorem 1 takes a simple form.

First, suppose that auditing is free (i.e.,  $c_i = 0$  for all  $i$ ). In this case, it is optimal for the principal to always audit the winner. The virtual value  $\psi_i(\theta_i)$  becomes

$$\psi_i^0(\theta_i) = \theta_i - (1 - \phi_i) \frac{1 - F_i(\theta_i)}{f_i(\theta_i)}.$$

Compared with Myerson's virtual value, the information rent term is scaled by  $1 - \phi_i$ , because the fraction  $\phi_i$  of the information rent is extracted through royalty payments. In particular, if  $\phi_i = 1$  for all  $i$ , then the virtual value

coincides with the true value, and the principal achieves full extraction, consistent with the benchmark in Section 4.

With free auditing, Assumption 1 is satisfied as long as  $\psi_i^0$  is strictly increasing. This holds, in particular, if the type distribution  $F_i$  has a weakly increasing hazard rate. Recall that  $[\cdot]$  denotes the indicator function for the predicate it encloses. The following corollary is immediate as a special case of Theorem 1.

**Corollary 1** (Free auditing)

*Suppose that  $c_i = 0$  for all  $i$ . If  $\psi_i^0$  is strictly increasing for each agent  $i$ , then the mechanism  $(q^*, t^*, r^*, a^*, p^*)$  from Theorem 1 is optimal, and takes the following form:*

$$q_i^*(\theta) = [\psi_i^0(\theta_i) \geq 0 \text{ and } \psi_i^0(\theta_i) > \max_{j \neq i} \psi_j^0(\theta_j)],$$

$$t_i^*(\theta) = (1 - \phi_i) \left( q_i^*(\theta) \theta_i - \int_{\underline{\theta}_i}^{\theta_i} q_i^*(z_i, \theta_{-i}) dz_i \right),$$

and

$$a_i^*(\theta, \pi'_i) = 1, \quad r_i^*(\theta, \pi'_i) = \pi'_i \phi_i, \quad p_i^*(\theta, \pi'_i, \pi_i) = (\pi_i - \pi'_i) \phi_i.$$

Here, the winner pays a linear royalty at rate  $\phi_i$ . [Bernhardt et al. \(2020\)](#) derive the same virtual value  $\psi_i^0(\theta_i)$  in their analysis of optimal auctions in which payments must be in the form of cash plus a linear royalty at a fixed (exogenous) rate.<sup>34</sup> Their model features entry costs, so they must take into account which types of each bidder will participate.

For the second special case, suppose that there is only one potential buyer. For this case, we drop agent subscripts. With one buyer, the solution from Theorem 1 can be implemented by offering a menu of contracts. In this optimal menu, each contract specifies linear royalties at rate  $\phi$  up to a royalty cap, with

---

<sup>34</sup>Also, in [Liu's \(2016\)](#) optimal equity-auction model, in the limit that synergies are much smaller than standalone values, the virtual value converges to  $\psi_i^0(\theta_i)$ , with  $\phi_i = V_i/(V_i + V_T)$ , where  $V_i$  is firm  $i$ 's standalone value and  $V_T$  is the standalone value of the target ([Liu, 2016](#), p. 110, footnote 15).

auditing if and only if the agent claims to owe less than the royalty cap. Each contract in the menu is distinguished by its upfront price  $t^*(\theta)$  and its royalty cap  $\pi^*(\theta)\phi$ . A lower royalty cap requires a higher upfront price. In the additive noise specification, this menu takes an even simpler form.

**Corollary 2** (Binary menu)

*Under Assumption 1, suppose that  $N = 1$  and that the income distribution follows the additive noise specification. The optimal mechanism  $(q^*, t^*, r^*, a^*, p^*)$  from Theorem 1 can be implemented by offering the agent a menu with at most two contracts: (i) lump-sum with no royalties or auditing; and possibly (ii) linear royalties at rate  $\phi$ , certain auditing, and penalties  $(\pi - \pi')\phi$ .*

The proof gives an explicit formula for the upfront price of each contract. We also provide a condition that characterizes whether the menu has both items. If it does, then high types choose the lump-sum contract, middle types choose the linear royalty, and low types may not buy at all. In the proof, we give a formula for the cutoff types separating these intervals.

**5.4 COMPARISON WITH OTHER CONTINGENT-PAYMENT AUCTIONS**

In our optimal mechanism, the winner’s equilibrium payment, as a function of his realized income, is linearly increasing and then flat; see Figure 1. By contrast, in the optimal mechanisms in DeMarzo et al. (2005) and Figueroa and Inostroza (2025), the payment received by the principal, as a function of the realized income, is flat and then linearly increasing. To understand the reason for this difference, it is helpful to consider the obstacle to full extraction in each model. Roughly, the main obstacle in our model is the cost of verification. In each of their models, the obstacle is a form of limited liability, which binds following low income realizations.

Consider the special case of our model in which  $\phi_i = 1$  for all  $i$ . In this case, there exists a mechanism that allocates the asset efficiently and leaves no information rents to the agents, but this mechanism requires the principal to always audit the winner. To reduce auditing costs, the optimal mechanism

features royalty caps. If the winner claims to owe the full royalty cap, the principal does not audit.

In [DeMarzo et al. \(2005\)](#), the project requires an upfront investment by the winner. If this required investment were zero, then full extraction would be possible. With a positive required investment, however, the net cash flow of the project is negative following low income realizations. Thus, full extraction would violate the seller’s limited liability constraint.

In [Figueroa and Inostroza \(2025\)](#), there is a single agent who is interpreted as a liquidity supplier for the principal.<sup>35</sup> The principal is less patient than the agent. Therefore, the principal would like to receive a larger upfront transfer in exchange for paying out more to the agent in the second period. But the limited liability constraint requires these second-period payments to be smaller than the realized cash flow.

Our analysis focuses on costly verification rather than liquidity constraints. In our motivating examples, the asset is often transferred because of technological constraints, not liquidity constraints. For example, the government is not liquidity constrained, but it cannot efficiently drill for oil or manage a gambling operation.<sup>36</sup> In our solution, the winner of the asset may face an unlucky income realization that does not cover the upfront transfer that he pays for the asset.

## 5.5 COMPARATIVE STATICS

In this section, we show how the optimal auction in [Theorem 1](#) depends on the model primitives. Given distribution functions  $H_1$  and  $H_2$  with respective densities  $h_1$  and  $h_2$ , we say that  $H_1$  is weakly larger than  $H_2$  with respect to

---

<sup>35</sup>[Figueroa and Inostroza \(2025\)](#) and [DeMarzo et al. \(2005\)](#) use opposite sign conventions. In [DeMarzo et al. \(2005\)](#), the security specifies what is paid by the bidder to the auctioneer. With this convention, the optimal security is a call option. In [Figueroa and Inostroza \(2025\)](#), the principal is selling a security to the agent. Thus, the security specifies what is paid by the principal to the agent. With this convention, the optimal security is a debt contract.

<sup>36</sup>[Contreras \(2022, Chapter 1\)](#) lists many reasons other than liquidity constraints for licensing intellectual property.

the hazard rate order if

$$\frac{h_1(x)}{1 - H_1(x)} \leq \frac{h_2(x)}{1 - H_2(x)},$$

for all  $x \in \mathbf{R}$ .

**Theorem 2** (Comparative statics)

For each type  $\theta_i$ , the virtual value  $\psi_i(\theta_i)$  and the auditing threshold  $\pi_i^*(\theta_i)$  satisfy the following.

- (a)  $\psi_i(\theta_i)$  is weakly decreasing in  $F_i$  with respect to the hazard-rate order.
- (b)  $\pi_i^*(\theta_i)$  is weakly increasing in  $F_i$  with respect to the hazard-rate order.
- (c)  $\psi_i(\theta_i)$  is weakly decreasing in  $c_i$  and weakly increasing in  $\phi_i$ .
- (d)  $\pi_i^*(\theta_i)$  is weakly decreasing in  $c_i/\phi_i$ , provided that  $\phi_i > 0$ .

Suppose that agent  $i$ 's type distribution  $F_i$  increases in the hazard-rate order (and the conditional income distribution for each type remains fixed). Then each fixed type of agent  $i$  wins the asset less often (i.e., for a smaller set of type reports of his opponents). And conditional on winning the asset, each type faces a higher royalty cap. The higher royalty cap reduces the information rents of all higher types, who become relatively more common when the type distribution increases with respect to the hazard-rate order.

If the cost  $c_i$  of auditing agent  $i$  decreases, or the maximal penalty sensitivity  $\phi_i$  increases, the asset is allocated to agent  $i$  for a larger set of type reports. Moreover, when agent  $i$  wins the asset, he faces a higher royalty cap and hence is audited more. In particular, if  $\phi_i$  and  $c_i$  are scaled up by the same factor (so  $c_i/\phi_i$  remains constant), the principal's optimal payoff increases: agent  $i$  wins more often, but, conditional on winning, faces the same auditing rule.

Recall the optimal cash auction from Myerson (1981). If the bidders have different valuation distributions, then the winner of the asset is not necessarily the bidder with the highest valuation. The auction favors bidders whose valuation distributions are smaller (in the hazard-rate order). In our model,

there are additional reasons why the asset may be allocated to an agent whose expected income from the asset is not the highest. By Theorem 2, the allocation rule favors agents who are cheaper to audit or who can be charged steeper penalties (perhaps because their income is more difficult to manipulate).

Theorem 2 can also be used to compare the optimal royalty structures in different environments. Suppose that the asset is a patent license. We expect that the cost of auditing bidders will be higher if the bidders are in a different country than the seller. In this case, our model predicts that the principal will impose lower royalty caps in order to save on auditing costs.

## 6. CONCLUSION

We study the optimal design of an auction for an income-generating asset. Our model offers an explanation for the common practice of charging royalties up to a cap. If the winner of the asset claims to owe the full royalty cap, the principal does not audit him. For these income realizations, the cost of auditing outweighs the gains from reducing information rent.

In our setting, moral hazard is an important consideration, but it is modeled in reduced form, following much of the literature on security design. Incorporating a richer model of moral hazard into the contingent-payment auction setting is an interesting challenge for future work.

## A. PROOFS

### A.1 PROOF OF LEMMA 1

Fix a direct mechanism  $(q, t, r, a, p)$  that satisfies Condition A and the constraints  $(\mathbf{IC}_\pi)$ ,  $(\mathbf{IC}_\theta)$ , and  $(\mathbf{IR})$  for each agent  $i$ . We can modify the penalty function profile  $p$  in such a way that (a) Condition A and the constraints  $(\mathbf{IC}_\pi)$ ,  $(\mathbf{IC}_\theta)$ , and  $(\mathbf{IR})$  still hold, and (b) the principal's payoff is unchanged. For each agent  $i$  and each history  $(\theta, \pi'_i)$  in  $\mathcal{H}_i$ , let

$$\hat{p}_i(\theta, \pi'_i, \pi_i) = p_i(\theta, \pi'_i, \pi'_i) + (\pi_i - \pi'_i)_+ \phi_i.$$

Here and below, we use the notation  $x_+ = \max\{x, 0\}$  for any real  $x$ . By construction,  $\hat{p}$  satisfies Condition A. For each agent  $i$ , the modified penalty function  $\hat{p}_i$  agrees with  $p_i$  whenever  $\pi_i = \pi'_i$  and is weakly larger than  $p_i$  otherwise. Thus, (a) and (b) hold. Therefore, it suffices to prove Lemma 1 with  $\hat{p}$  in place of  $p$ . For the rest of the proof, we assume that  $p$  has already been replaced with  $\hat{p}$ . This replacement ensures that  $p$  satisfies additional regularity properties. Let  $p_{i,3+}$  denote the partial right-derivative of  $p_i$  with respect to realized income. For each  $(\theta, \pi_i)$  in  $\mathcal{H}_i$ , we have  $p_{i,3+}(\theta, \pi_i, \pi_i) = \phi_i$ . We will use this fact in the proof below.

**Income-reporting envelope theorem** For each agent  $i$ , define the support

$$S_i = \{(\theta_i, \pi_i) \in \Theta_i \times \Pi_i : \pi_i \in \Pi_i(\theta_i)\}.$$

In the definition of  $u_i$  in (2), the expression inside the expectation is weakly increasing and 1-Lipschitz in the true income  $\pi_i$ , by Condition A. We can right-differentiate under the expectation to get

$$(5) \quad u_{i,3+}(\theta_i, \pi_i | \pi_i) = \mathbb{E}_{\theta_{-i}} [q_i(\theta_i, \theta_{-i})(1 - a_i(\theta_i, \theta_{-i}, \pi_i)\phi_i)],$$

for each  $(\theta_i, \pi_i)$  in  $S_i$ .

By (IC $_{\pi}$ ), for each  $(\theta_i, \pi_i)$  in  $S_i$ , we have

$$u_i(\theta_i, \pi_i | \pi_i) = \max_{\pi'_i \in \Pi_i(\theta_i)} u_i(\theta_i, \pi'_i | \pi_i).$$

By Milgrom and Segal (2002, Theorems 1), it can be shown that for each type  $\theta_i$ , there exists a bounded measurable function  $m_i(\theta_i, \cdot)$  on  $\Pi_i(\theta_i)$  satisfying  $m_i(\theta_i, \pi_i) \geq u_{i,3+}(\theta_i, \pi_i | \pi_i)$  for each  $\pi_i$  in  $\Pi_i(\theta_i)$ , with equality whenever the derivative  $u_{i,3}(\theta_i, \pi_i | \pi_i)$  exists, such that

$$(6) \quad u_i(\theta_i, \pi_i | \pi_i) = u_i(\theta_i, \underline{\pi}_i(\theta_i) | \underline{\pi}_i(\theta_i)) + \int_{\underline{\pi}_i(\theta_i)}^{\pi_i} m_i(\theta_i, z_i) dz_i,$$

for each  $\pi_i$  in  $\Pi_i(\theta_i)$ .<sup>37</sup> Extend  $m_i$  to  $\Theta_i \times \Pi_i$  by setting  $m_i(\theta_i, \pi_i) = Q_i(\theta_i)$  if  $\pi_i < \underline{\pi}_i(\theta_i)$  and  $m_i(\theta_i, \pi_i) = u_{i,3+}(\theta_i, \bar{\pi}_i(\theta_i) | \pi_i)$  if  $\pi_i > \bar{\pi}_i(\theta_i)$ . Let  $\hat{\pi}_i = \inf \Pi_i$ . By the definition of  $p_i$ , it can be checked by cases that

$$(7) \quad u_i(\theta_i, \text{proj}_{\Pi_i(\theta_i)} \pi_i | \pi_i) = u_i(\theta_i, \underline{\pi}_i(\theta_i) | \hat{\pi}_i) + \int_{\hat{\pi}_i}^{\pi_i} m_i(\theta_i, z_i) dz_i,$$

for each  $(\theta_i, \pi_i)$  in  $\Theta_i \times \Pi_i$ .

**Type-reporting envelope theorem** Recall the definition of  $U_i$  from (3).

For any types  $\theta_i, \theta'_i \in \Theta_i$ , applying (7) gives

$$\begin{aligned} U_i(\theta'_i | \theta_i) &= \int_{\hat{\pi}_i}^{\infty} u_i(\theta'_i, \text{proj}_{\Pi_i(\theta'_i)} \pi_i | \pi_i) g_i(\pi_i | \theta_i) d\pi_i \\ &= u_i(\theta'_i, \underline{\pi}_i(\theta'_i) | \hat{\pi}_i) + \int_{\hat{\pi}_i}^{\infty} \left( \int_{\hat{\pi}_i}^{\pi_i} m_i(\theta'_i, z_i) dz_i \right) g_i(\pi_i | \theta_i) d\pi_i. \end{aligned}$$

---

<sup>37</sup>For each fixed type  $\theta_i$ , the map  $\pi_i \mapsto u_i(\theta_i, \pi_i | \pi_i)$  is 1-Lipschitz, so its derivative exists almost everywhere. Extend this almost-everywhere derivative to a measurable  $[0, 1]$ -valued function  $\tilde{m}_i(\theta_i, \cdot)$  on  $\Pi_i(\theta_i)$ . At each differentiability point  $\pi_i$ , Milgrom and Segal (2002, Theorem 1) implies that  $\tilde{m}_i(\theta_i, \pi_i) \geq u_{i,3+}(\theta_i, \pi_i | \pi_i)$ , with equality if  $u_{i,3}(\theta_i, \pi_i | \pi_i)$  exists. Finally, for each  $\pi_i \in \Pi_i(\theta_i)$ , let  $m_i(\theta_i, \pi_i) = \max\{\tilde{m}_i(\theta_i, \pi_i), u_{i,3+}(\theta_i, \pi_i | \pi_i)\}$ .

Change the order of integration (and switch the variable labels) to get<sup>38</sup>

$$U_i(\theta'_i|\theta_i) = u_i(\theta'_i, \pi_i(\theta'_i)|\hat{\pi}_i) + \int_{\hat{\pi}_i}^{\infty} (1 - G_i(\pi_i|\theta_i))m_i(\theta'_i, \pi_i) d\pi_i.$$

Differentiate under the integral with respect to  $\theta_i$  to get<sup>39</sup>

$$U_{i,2}(\theta_i|\theta_i) = \int_{\hat{\pi}_i}^{\infty} -G_{i,2}(\pi_i|\theta_i)m_i(\theta_i, \pi_i) d\pi_i.$$

By (IC $_{\theta}$ ), for each  $\theta_i$  in  $\Theta_i$ , we have

$$U_i(\theta_i|\theta_i) - T_i(\theta_i) = \max_{\theta'_i \in \Theta_i} \{U_i(\theta'_i|\theta_i) - T_i(\theta'_i)\}.$$

By our Lipschitz condition on  $G_i(\pi_i|\cdot)$ , we can apply the envelope theorem (Milgrom and Segal, 2002). For all  $\theta_i$  in  $\Theta_i$ , we have

$$\begin{aligned} & U_i(\theta_i|\theta_i) - T_i(\theta_i) \\ (8) \quad & = U_i(\theta_i|\theta_i) - T_i(\theta_i) + \int_{\theta_i}^{\theta_i} \left( \int_{\hat{\pi}_i}^{\infty} -G_{i,2}(\pi_i|y_i)m_i(y_i, \pi_i) d\pi_i \right) dy_i \\ & \geq \int_{\theta_i}^{\theta_i} \left( \int_{\pi_i(y_i)}^{\bar{\pi}_i(y_i)} -G_{i,2}(\pi_i|y_i)u_{i,3+}(y_i, \pi_i|\pi_i) d\pi_i \right) dy_i, \end{aligned}$$

where we get the final inequality by applying (IR) and noting that  $G_{i,2}$  vanishes outside  $S_i$  and that  $m_i(y_i, \pi_i) \geq u_{i,3+}(y_i, \pi_i|\pi_i)$  for  $(y_i, \pi_i)$  in  $S_i$ .<sup>40</sup>

Now take an expectation over  $\theta_i$ , change the order of integration, and multiply and divide by  $f_i(\theta_i)g_i(\pi_i|\theta_i)$  to conclude that

$$\begin{aligned} (9) \quad \mathbb{E}[U_i(\theta_i|\theta_i) - T_i(\theta_i)] & \geq \mathbb{E}[\mu_i(\theta_i, \pi_i)u_{i,3+}(\theta_i, \pi_i|\pi_i)] \\ & = \mathbb{E}[q_i(\theta_i, \theta_{-i})\mu_i(\theta_i, \pi_i)(1 - a_i(\theta_i, \theta_{-i}, \pi_i)\phi_i)], \end{aligned}$$

<sup>38</sup>This integral is well-defined since  $m_i$  is bounded.

<sup>39</sup>We can differentiate under the integral because  $m_i$  is bounded and the functions  $G_i(\pi_i|\cdot)$  are uniformly Lipschitz, by assumption.

<sup>40</sup>Note that the function  $m_i$  is guaranteed to be measurable in its second argument only, but this is sufficient for the proof.

where we have used the independence of  $(\theta_i, \pi_i)$  and  $\theta_{-i}$ .

**Principal's payoff** We can express the principal's expected payoff in terms of the functions  $U_i$  and  $T_i$  as

$$\mathbb{E}_\theta \left[ \sum_{i=1}^N (q_i(\theta) \mathbb{E}_{\pi_i|\theta_i} [\pi_i - c_i a_i(\theta, \pi_i)] - U_i(\theta_i|\theta_i) + T_i(\theta_i)) \right].$$

Plug in (9) and simplify to get the upper bound

$$(10) \quad \mathbb{E}_\theta \left[ \sum_{i=1}^N q_i(\theta) \mathbb{E}_{\pi_i|\theta_i} [\pi_i - \mu_i(\theta_i, \pi_i) + a_i(\theta, \pi_i)(\mu_i(\theta_i, \pi_i)\phi_i - c_i)] \right].$$

To get the desired expression, it remains to check that

$$(11) \quad \mathbb{E}_{\pi_i|\theta_i} [\pi_i - \mu_i(\theta_i, \pi_i)] = \theta_i - \frac{1 - F_i(\theta_i)}{f_i(\theta_i)}.$$

By our normalization and a standard probability identity,

$$\theta_i = \mathbb{E}[\pi_i|\theta_i] = \int_0^\infty (1 - G_i(\pi_i|\theta_i)) d\pi_i.$$

Differentiating under the integral sign gives

$$(12) \quad 1 = - \int_0^\infty G_{i,2}(\pi_i|\theta_i) d\pi_i = \mathbb{E}_{\pi_i|\theta_i} \left[ \frac{-G_{i,2}(\pi_i|\theta_i)}{g_i(\pi_i|\theta_i)} \right],$$

as needed.

## A.2 REGULARITY ASSUMPTIONS

In our notation, the regularity assumptions in [Eső and Szentes \(2007, p. 709\)](#) state that for each agent  $i$ ,

1.  $f_i(\theta_i)/(1 - F_i(\theta_i))$  is weakly increasing in  $\theta_i$ ;
2.  $G_{i,2}(\pi_i|\theta_i)/g_i(\pi_i|\theta_i)$  is weakly increasing in  $\theta_i$  and  $\pi_i$ .

We prove that these regularity assumptions imply Assumption 1.<sup>41</sup> Recall that

$$\mu_i(\theta_i, \pi_i) = -\frac{G_{i,2}(\pi_i|\theta_i)}{g_i(\pi_i|\theta_i)} \cdot \frac{1 - F_i(\theta_i)}{f_i(\theta_i)}.$$

Thus,  $\mu_i(\theta_i, \pi_i)$  is weakly decreasing in  $\theta_i$  and  $\pi_i$ . Assumption 1.a follows immediately. For Assumption 1.b, substitute in (11) to see that

$$\psi_i(\theta_i) = \theta_i - \mathbb{E}_{\pi_i|\theta_i} [\mu_i(\theta_i, \pi_i) - (\mu_i(\theta_i, \pi_i)\phi_i - c_i)_+].$$

The expression inside the expectation is decreasing in  $\theta_i$  and  $\pi_i$ , and we have  $G_{i,2}(\pi_i|\theta_i) < 0$ . Thus, the expectation is decreasing in  $\theta_i$ , and hence,  $\psi_i(\theta_i)$  is strictly increasing in  $\theta_i$ .

Next, we provide an example of distributions  $G_i$  and  $F_i$  for which Assumption 1 holds, but the ratio

$$\frac{G_{i,2}(\pi_i | \theta_i)}{g_i(\pi_i | \theta_i)}$$

is strictly *decreasing* in  $\theta_i$ , violating the regularity assumptions in [Eső and Szentes \(2007\)](#). We consider a single agent, and we drop agent indices. Assume that  $\theta$  is uniformly distributed on  $[1, 2]$ . Given  $\theta$ , the conditional distribution of  $\pi$  is uniform over  $[\theta - e^{-\theta/2}, \theta + e^{-\theta/2}]$ . For each  $\theta \in [1, 2]$  and  $\pi \in (\theta - e^{-\theta/2}, \theta + e^{-\theta/2})$ , it can be verified that

$$\frac{G_2(\pi|\theta)}{g(\pi|\theta)} = \frac{\pi - \theta}{2} - 1,$$

which is strictly increasing in  $\pi$  and strictly *decreasing* in  $\theta$ . However, we have

$$\mu(\theta, \pi) = (2 - \theta) \left( 1 - \frac{\pi - \theta}{2} \right),$$

which is weakly decreasing in both arguments. This implies that the virtual value  $\psi$  is strictly increasing. Hence, Assumption 1 is satisfied.

---

<sup>41</sup>In fact, the regularity assumptions in [Eső and Szentes \(2007\)](#), together with the type normalization (1), imply additive full-support noise; see [Ball and Pekkariinen \(2024\)](#).

### A.3 TRANSFORMED ADDITIVE NOISE

Let  $H_i$  and  $h_i$  denote the cumulative distribution function and probability density function of  $\varepsilon_i$ , respectively. In place of  $F_i, f_i, G_i, g_i, \mu_i$ , we write  $\hat{F}_i, \hat{f}_i, \hat{G}_i, \hat{g}_i, \hat{\mu}_i$  for the corresponding functions defined on unnormalized signals. We have

$$\hat{G}_i(\pi_i|\gamma_i) = \mathbf{P}_{\varepsilon_i}(\alpha_i(\gamma_i + \varepsilon_i) \leq \pi_i) = H_i(\alpha_i^{-1}(\pi_i) - \gamma_i).$$

Therefore,

$$\begin{aligned}\hat{G}_{i,2}(\pi_i|\gamma_i) &= -h_i(\alpha_i^{-1}(\pi_i) - \gamma_i) \\ \hat{g}_i(\pi_i|\gamma_i) &= \frac{h_i(\alpha_i^{-1}(\pi_i) - \gamma_i)}{\alpha_i'(\alpha_i^{-1}(\pi_i))}.\end{aligned}$$

First, we prove Assumption 1.a. We have

$$\hat{\mu}_i(\gamma_i, \pi_i) = -\frac{\hat{G}_{i,2}(\pi_i|\gamma_i)}{\hat{g}_i(\pi_i|\gamma_i)} \cdot \frac{1 - \hat{F}_i(\gamma_i)}{\hat{f}_i(\gamma_i)} = \alpha_i'(\alpha_i^{-1}(\pi_i)) \cdot \frac{1 - \hat{F}_i(\gamma_i)}{\hat{f}_i(\gamma_i)}.$$

Notice that the right side is multiplicatively separable in  $\pi_i$  and  $\gamma_i$ . The function  $\hat{\mu}_i$  is weakly decreasing in  $\pi_i$  since  $\alpha_i' > 0$  and  $\alpha_i'' \leq 0$ . And  $\hat{\mu}_i$  is weakly decreasing in  $\gamma_i$  because the distribution of  $\gamma_i$  has weakly increasing hazard rate. It is easily verified that  $\hat{\mu}_i(\gamma_i, \pi_i) = \mu_i(\mathbb{E}[\pi_i|\gamma_i], \pi_i)$ , so Assumption 1.a follows.

Next, we prove Assumption 1.b. Since we have not normalized the signal, we see from (10) that the associated virtual value (with respect to the unnormalized signal) is given by

$$\begin{aligned}\hat{\psi}_i(\gamma_i) &= \mathbb{E}_{\pi_i|\gamma_i}[\pi_i - \hat{\mu}_i(\gamma_i, \pi_i) + (\hat{\mu}_i(\gamma_i, \pi_i)\phi_i - c_i)_+] \\ &= \mathbb{E}_{\pi_i|\gamma_i}[\max\{\pi_i - (1 - \phi_i)\hat{\mu}_i(\gamma_i, \pi_i) - c_i, \pi_i - \hat{\mu}_i(\gamma_i, \pi_i)\}].\end{aligned}$$

We claim that  $\hat{\psi}_i$  is strictly increasing in  $\gamma_i$ . Inside the expectation, each function inside the maximum is strictly increasing in  $\pi_i$  and weakly increasing in  $\gamma_i$  (since the distribution of  $\gamma_i$  has weakly increasing hazard rate). Moreover, the conditional distribution of  $\pi_i$  given  $\gamma_i$  is strictly increasing in  $\gamma_i$ , with respect to first-order stochastic dominance. Thus,  $\hat{\psi}_i$  is strictly increasing. It

is easily verified that  $\hat{\psi}_i(\gamma_i) = \psi_i(\mathbb{E}[\pi_i|\gamma_i])$ , so Assumption 1.b follows.

#### A.4 PROOF OF THEOREM 1

Before the main proof, we first confirm that  $\pi_i^*$  is weakly decreasing. Recall that  $\pi_i^*$  is defined as follows. Here, we use the notation  $\Delta_i(\theta_i, \pi_i) = \mu_i(\theta_i, \pi_i)\phi_i - c_i$ . For each  $\theta_i \in \Theta_i$ , let

$$\Pi_i^*(\theta_i) = \{\pi_i \in (\underline{\pi}_i(\theta_i), \bar{\pi}_i(\theta_i)) : \Delta_i(\theta_i, \pi_i) \geq 0\}.$$

Define  $\pi_i^* : \Theta_i \rightarrow [0, \infty]$  by

$$\pi_i^*(\theta_i) = \begin{cases} \infty & \text{if } \Pi_i^*(\theta_i) = (\underline{\pi}_i(\theta_i), \bar{\pi}_i(\theta_i)), \\ \sup \Pi_i^*(\theta_i) & \text{if } \emptyset \subsetneq \Pi_i^*(\theta_i) \subsetneq (\underline{\pi}_i(\theta_i), \bar{\pi}_i(\theta_i)), \\ 0 & \text{if } \Pi_i^*(\theta_i) = \emptyset. \end{cases}$$

To prove that  $\pi_i^*$  is weakly decreasing, it suffices to show the following local monotonicity property: for any types  $\theta_i$  and  $\theta'_i$  with  $\theta_i < \theta'_i$  and  $\bar{\pi}_i(\theta_i) > \underline{\pi}_i(\theta'_i)$ , we have  $\pi_i^*(\theta_i) \geq \pi_i^*(\theta'_i)$ . It then follows from a compactness argument that  $\pi_i^*$  is globally weakly decreasing.<sup>42</sup> Fix types  $\theta_i$  and  $\theta'_i$  with  $\theta_i < \theta'_i$  and  $\bar{\pi}_i(\theta_i) > \underline{\pi}_i(\theta'_i)$ . Suppose for a contradiction that  $\pi_i^*(\theta_i) < \pi_i^*(\theta'_i)$ . In particular,  $\pi_i^*(\theta_i) < \infty$  and  $\pi_i^*(\theta'_i) > 0$ , so the intervals  $(\pi_i^*(\theta_i), \bar{\pi}_i(\theta_i))$  and  $(\underline{\pi}_i(\theta'_i), \pi_i^*(\theta'_i))$  are nonempty. Moreover, these intervals have nonempty intersection because  $\bar{\pi}_i(\theta_i) > \underline{\pi}_i(\theta'_i)$  and  $\pi_i^*(\theta_i) < \pi_i^*(\theta'_i)$ . Choose  $\pi_i \in (\pi_i^*(\theta_i), \bar{\pi}_i(\theta_i)) \cap (\underline{\pi}_i(\theta'_i), \pi_i^*(\theta'_i))$ . By the definition of  $\pi_i^*(\theta'_i)$ , there exists  $\pi'_i$  in  $(\pi_i, \pi_i^*(\theta'_i) \wedge \bar{\pi}_i(\theta'_i))$  with  $\Delta_i(\theta'_i, \pi'_i) \geq 0$ . Since  $\Delta_i$  is single-crossing from above in income, we have  $\Delta_i(\theta'_i, \pi_i) \geq 0$ . But we also have  $\Delta_i(\theta_i, \pi_i) < 0$  because  $\pi_i$  is in  $(\pi_i^*(\theta_i), \bar{\pi}_i(\theta_i))$ . This contradicts the assumption that  $\Delta_i$  is single-crossing from above in the type.

<sup>42</sup>Since  $\underline{\pi}_i$  and  $\bar{\pi}_i$  are continuous, for each type  $\theta_i$  there is an open interval  $U_{\theta_i}$  containing  $\theta_i$  such that for all types  $\theta'_i, \theta''_i \in U_{\theta_i}$  with  $\theta'_i < \theta''_i$  we have  $\bar{\pi}_i(\theta'_i) > \underline{\pi}_i(\theta''_i)$ . The local monotonicity property implies that  $\pi_i^*$  is weakly decreasing over  $U_{\theta_i}$ . The collection  $\{U_{\theta_i}\}_{\theta_i \in \Theta_i}$  is an open cover of  $\Theta_i$ . Since  $\Theta_i$  is compact,  $\{U_{\theta_i}\}_{\theta_i \in \Theta_i}$  has a finite subcover. Global monotonicity follows because any two types can be connected by a finite path where each pair of consecutive types lies in some interval in the finite cover.

Now we turn to the main proof. For any mechanism  $(q, t, r, a, p)$  satisfying Condition A and the constraints (IC $_{\pi}$ ), (IC $_{\theta}$ ), and (IR) for each agent  $i$ , the principal's expected payoff, denoted  $V(q, t, r, a, p)$ , satisfies

$$(13) \quad V(q, t, r, a, p) \leq \mathbb{E} \left[ \sum_{i=1}^N q_i(\theta) \Psi_i(\theta) \right]$$

$$(14) \quad \leq \mathbb{E} \left[ \max_i \Psi_i(\theta)_+ \right]$$

$$(15) \quad \leq \mathbb{E} \left[ \max_i \psi_i(\theta_i)_+ \right],$$

where (13) follows from Lemma 1 and (15) follows from the definitions of  $\Psi_i$  and  $\psi_i$ .<sup>43</sup> Note that  $\Psi_i$  depends on the auditing rule  $a_i$ .

For the mechanism  $(q, t, r, a, p) = (q^*, t^*, r^*, a^*, p^*)$ , we claim that the inequalities (13)–(15) hold with equality. For (13), follow the proof of Lemma 1 with  $(q, t, r, a, p) = (q^*, t^*, r^*, a^*, p^*)$ .<sup>44</sup> The inequality in (8) holds with equality because (a) (IR) holds with equality for the lowest type of each agent, and (b) for each  $(\theta_i, \pi_i) \in S_i$  with  $\pi_i \neq \pi_i^*(\theta_i)$ , the partial derivative  $u_{i,3}(\theta_i, \pi_i | \pi_i)$  exists and agrees with the expression in (5). For (14), equality is immediate from the definition of  $q^*$ . For (15), since  $\mu_i(\theta_i, \pi_i)\phi_i - c_i$  is single-crossing from above in  $\pi_i$  (by Assumption 1.a), note that

$$[\mu_i(\theta_i, \pi_i)\phi_i - c_i > 0] \leq a_i^*(\theta, \pi_i) \leq [\mu_i(\theta_i, \pi_i)\phi_i - c_i \geq 0].$$

To complete the proof, we show that  $(q^*, t^*, r^*, a^*, p^*)$  is dominant-strategy incentive compatible and dominant-strategy individually rational. For the rest of the proof, we focus on a fixed agent  $i$ .

First, we prove dominant-strategy individual rationality. Given a report vector  $\theta_{-i}$  in  $\Theta_{-i}$  from agent  $i$ 's opponents, the expected utility for type  $\theta_i$  from truthfully reporting his type and subsequent income is

$$\int_{\theta_i}^{\theta_i} q_i^*(z_i, \theta_{-i})(1 - \Phi_i(z_i)) dz_i.$$

<sup>43</sup>For any function  $h$ , we denote  $(h(x))_+$  by  $h(x)_+$ .

<sup>44</sup>Except in this case, do not replace  $p^*$  with the modified penalty rule  $\hat{p}$ .

This expression is nonnegative because  $\Phi_i(z_i) \leq 1$  for all  $z_i$ .

Next, we prove dominant-strategy incentive compatibility. If agent  $i$  wins the asset, the royalty and the penalty do not depend on the other agents' reports. We check that at the income-reporting stage, it is optimal for agent  $i$  to report his income as truthfully as possible after any type report  $\theta'_i$ . To see this, suppose agent  $i$ 's true income is  $\pi_i$ . Consider agent  $i$ 's additional (post-transfer) payment as a function of his report  $\pi'_i$ . If  $\pi'_i < \pi_i^*(\theta'_i)$ , then he pays  $\min\{\pi_i, \pi_i^*(\theta'_i)\}\phi_i$ . If  $\pi'_i \geq \pi_i^*(\theta'_i)$ , then he pays  $\pi_i^*(\theta'_i)\phi_i$ . Therefore, reporting  $\pi'_i = \text{proj}_{\Pi_i(\theta'_i)} \pi_i$  is optimal because  $\text{proj}_{\Pi_i(\theta'_i)} \pi_i \geq \pi_i^*(\theta'_i)$  if and only if  $\pi_i \geq \pi_i^*(\theta'_i)$ .

Now we consider the type-reporting stage. Fix a report vector  $\theta_{-i}$  in  $\Theta_{-i}$  from agent  $i$ 's opponents. Given the report  $\theta_{-i}$ , the difference in expected utility for type  $\theta_i$  between reporting type  $\theta_i$  (and then reporting income truthfully) and reporting type  $\theta'_i$  (and then reporting income as truthfully as possible) can be expressed as

$$(16) \quad \int_{\theta'_i}^{\theta_i} q_i^*(z_i, \theta_{-i}) [1 - \Phi_i(z_i)] dz_i \\ + q_i^*(\theta'_i, \theta_{-i}) \int_0^\infty (\pi_i - \phi_i \min\{\pi_i, \pi_i^*(\theta'_i)\}) [g_i(\pi_i|\theta'_i) - g_i(\pi_i|\theta_i)] d\pi_i.$$

Note that the sign of the first integral depends on the ordering of  $\theta_i$  and  $\theta'_i$ . For any  $z_i$  in  $\Theta_i$  and  $\pi_i$  in  $\Pi_i$ , let

$$m_i^*(z_i, \pi_i; \theta_{-i}) = q_i^*(z_i, \theta_{-i}) (1 - \phi_i [\pi_i \leq \pi_i^*(z_i)]).$$

Since  $q_i^*(z_i, \theta_{-i})$  is weakly increasing in  $z_i$ , and  $\pi_i^*(z_i)$  is weakly decreasing in  $z_i$ , it follows that  $m_i^*$  is weakly increasing in  $z_i$ . We will express each line of (16) in terms of  $m_i^*$ .

From the definition of  $\Phi_i$  in (4), and the implication (12) of our normalization, we have

$$1 - \Phi_i(z_i) = \int_0^\infty -G_{i,2}(\pi_i|z_i) (1 - \phi_i [\pi_i \leq \pi_i^*(z_i)]) d\pi_i,$$

so the first line in (16) can be expressed as

$$(17) \quad \int_{\theta'_i}^{\theta_i} \int_0^\infty -G_{i,2}(\pi_i|z_i) m_i^*(z_i, \pi_i; \theta_{-i}) d\pi_i dz_i.$$

For the integral in the second line of (16), integrating by parts<sup>45</sup>

$$\int_0^\infty (1 - \phi_i[\pi_i \leq \pi_i^*(\theta'_i)]) (G_i(\pi_i|\theta_i) - G_i(\pi_i|\theta'_i)) d\pi_i,$$

so the second line of (16) becomes

$$(18) \quad \int_{\theta'_i}^{\theta_i} \int_0^\infty G_{i,2}(\pi_i|z_i) m_i^*(\theta'_i, \pi_i; \theta_{-i}) d\pi_i dz_i.$$

Putting together (17) and (18), the expression in (16) becomes

$$\int_{\theta'_i}^{\theta_i} \int_0^\infty -G_{i,2}(\pi_i|z_i) (m_i^*(z_i, \pi_i; \theta_{-i}) - m_i^*(\theta'_i, \pi_i; \theta_{-i})) d\pi_i dz_i.$$

This expression is nonnegative since  $-G_{i,2}$  is nonnegative and  $m_i^*$  is weakly increasing in its first argument.

**Crossing payments** Here, we prove the crossing property referenced after Figure 1. First, we introduce some notation. For all  $\theta \in \Theta$  with  $q_i^*(\theta) = 1$ , let

$$s_i^*(\theta, \pi_i) = t_i^*(\theta) + r_i^*(\theta, \text{proj}_{\Pi_i(\theta_i)} \pi_i) + a_i^*(\theta, \text{proj}_{\Pi_i(\theta_i)} \pi_i) p_i^*(\theta, \text{proj}_{\Pi_i(\theta_i)} \pi_i, \pi_i).$$

Thus,  $s_i^*(\theta, \pi_i)$  is agent  $i$ 's total ex-post payment when the type vector  $\theta$  is reported, agent  $i$ 's true income is  $\pi_i$ , and agent  $i$  reports income  $\text{proj}_{\Pi_i(\theta_i)} \pi_i$ . Observe that  $\text{proj}_{\Pi_i(\theta_i)} \pi_i \geq \pi_i^*(\theta_i)$  if and only if  $\pi_i \geq \pi_i^*(\theta_i)$ . Therefore, for each  $\theta \in \Theta$  with  $q_i^*(\theta) = 1$  and each  $\pi_i \in \Pi_i$ , we have

$$s_i^*(\theta, \pi_i) = t_i^*(\theta) + \min\{\pi_i, \pi_i^*(\theta_i)\} \phi_i.$$

---

<sup>45</sup>Integration by parts still holds when the “antiderivative” is an absolutely continuous function with almost-sure derivative.

Note that Figure 1 plots the function  $s_i^*(\theta'_i, \theta_{-i}, \cdot)$  over the domain  $\Pi_i(\theta'_i)$  and the function  $s_i^*(\theta''_i, \theta_{-i}, \cdot)$  over the domain  $\Pi_i(\theta''_i)$ .

Consider agent  $i$ . Fix  $\theta_{-i} \in \Theta_{-i}$  and  $\theta'_i, \theta''_i \in \Theta_i$  with  $\theta'_i < \theta''_i$ . Suppose that  $q_i^*(\theta'_i, \theta_{-i}) = 1$  and  $q_i^*(\theta''_i, \theta_{-i}) = 1$ . Since  $\pi_i^*$  is weakly decreasing, we have  $\pi_i^*(\theta'_i) \geq \pi_i^*(\theta''_i)$ . We compare the functions  $s_i^*(\theta'_i, \theta_{-i}, \cdot)$  and  $s_i^*(\theta''_i, \theta_{-i}, \cdot)$  over  $\Pi_i$ . We claim that if  $\pi_i^*(\theta'_i)$  and  $\pi_i^*(\theta''_i)$  are both in  $(0, \infty)$ , then either (i)  $\pi_i^*(\theta'_i) = \pi_i^*(\theta''_i)$ , in which case the functions agree everywhere, or (ii)  $\pi_i^*(\theta'_i) > \pi_i^*(\theta''_i)$ , in which case the functions intersect at exactly one point, and this point is in  $(\pi_i^*(\theta''_i), \pi_i^*(\theta'_i))$ .

Now we verify the claim. Since  $q_i^*(\theta'_i, \theta_{-i}) = q_i^*(\theta''_i, \theta_{-i})$ , dominant-strategy incentive compatibility implies that

$$(19) \quad \begin{aligned} \mathbb{E}_{\pi_i|\theta'_i}[s_i^*(\theta'_i, \theta_{-i}, \pi_i)] &\leq \mathbb{E}_{\pi_i|\theta'_i}[s_i^*(\theta''_i, \theta_{-i}, \pi_i)], \\ \mathbb{E}_{\pi_i|\theta''_i}[s_i^*(\theta''_i, \theta_{-i}, \pi_i)] &\leq \mathbb{E}_{\pi_i|\theta''_i}[s_i^*(\theta'_i, \theta_{-i}, \pi_i)]. \end{aligned}$$

Suppose that  $\pi_i^*(\theta'_i)$  and  $\pi_i^*(\theta''_i)$  are both in  $(0, \infty)$ . If  $\pi_i^*(\theta'_i) = \pi_i^*(\theta''_i)$ , then the curves must agree everywhere; otherwise, one curve is pointwise strictly larger, contrary to (19). If  $\pi_i^*(\theta'_i) > \pi_i^*(\theta''_i)$ , then by comparing slopes, we see that over the interval  $[\pi_i^*(\theta''_i), \pi_i^*(\theta'_i)]$ , the curves intersect at exactly one point. We show that this point cannot be either endpoint. If the intersection is at  $\pi_i^*(\theta''_i)$ , then  $s_i^*(\theta'_i, \theta_{-i}, \cdot)$  is pointwise weakly larger than  $s_i^*(\theta''_i, \theta_{-i}, \cdot)$ . By (19), we must have  $\Pi_i(\theta'_i) \subseteq (-\infty, \pi_i^*(\theta''_i)]$ . In particular,  $\bar{\pi}_i(\theta'_i) \leq \pi_i^*(\theta''_i) < \pi_i^*(\theta'_i)$ , so  $\pi_i^*(\theta'_i) = \infty$ , which is a contradiction. If the intersection is at  $\pi_i^*(\theta'_i)$ , then  $s_i^*(\theta''_i, \theta_{-i}, \cdot)$  is pointwise weakly larger than  $s_i^*(\theta'_i, \theta_{-i}, \cdot)$ . By (19), we must have  $\Pi_i(\theta''_i) \subseteq [\pi_i^*(\theta'_i), \infty)$ . In particular,  $\pi_i^*(\theta''_i) < \pi_i^*(\theta'_i) \leq \bar{\pi}_i(\theta''_i)$ , so  $\pi_i^*(\theta''_i) = 0$ , which is a contradiction.

## A.5 NOISY AUDITING

Let  $Z_i$  denote the support of the signal  $\zeta_i$ . With noisy auditing, the principal chooses a penalty rule  $\tilde{p}_i: \mathcal{H}_i \times Z_i \rightarrow \mathbf{R}$ , for each agent  $i$ . We require that these penalty rules satisfy the analog of Condition A, with  $\zeta_i$  in place of  $\pi_i$ .

Given a penalty rule  $\tilde{p}_i$ , define the associated *effective penalty rule*  $p_i: \mathcal{H}_i \times$

$\Pi_i \rightarrow \mathbf{R}$  by

$$p_i(\theta, \pi'_i, \pi_i) = \mathbb{E}_{\zeta_i|\pi_i}[\tilde{p}_i(\theta, \pi'_i, \zeta_i)].$$

Using penalty rule  $\tilde{p}_i$  with noisy auditing is equivalent (in terms of expected payments and utilities) to using the associated effective penalty rule  $p_i$  in the original model with perfect auditing. In the main text, we present a mechanism in the noisy auditing model that achieves the principal's maximum payoff in the perfect auditing model. To prove that this mechanism is optimal, it suffices to check that noisy auditing (weakly) shrinks the principal's feasible set.

Formally, we show that if  $\tilde{p}_i$  satisfies Condition A, then the associated effective penalty rule  $p_i$  satisfies Condition A as well. Let  $(\theta', \pi'_i)$  be a report history in  $\mathcal{H}_i$ . Fix  $\pi_i, \hat{\pi}_i \in \Pi_i$  with  $\hat{\pi}_i > \pi_i$ . The associated conditional distributions of  $\zeta_i$  are ordered by first-order stochastic dominance, so we can construct a random vector  $(\zeta_i, \hat{\zeta}_i)$  satisfying  $\zeta_i \leq \hat{\zeta}_i$  such that the marginal distribution of  $\zeta_i$  (respectively,  $\hat{\zeta}_i$ ) agrees with the conditional signal distribution, given  $\pi_i$  (respectively  $\hat{\pi}_i$ ). By Condition A, we have

$$0 \leq \tilde{p}_i(\theta', \pi'_i, \hat{\zeta}_i) - \tilde{p}_i(\theta', \pi'_i, \zeta_i) \leq (\hat{\zeta}_i - \zeta_i)\phi_i.$$

Taking expectations gives

$$0 \leq p_i(\theta', \pi'_i, \hat{\pi}_i) - p_i(\theta', \pi'_i, \pi_i) \leq (\hat{\pi}_i - \pi_i)\phi_i,$$

as desired.

## A.6 PROOF OF COROLLARY 2

We drop the agent subscripts. Under the additive noise specification,

$$\mu(\theta, \pi) = \frac{1 - F(\theta)}{f(\theta)}.$$

Define the cutoff types by

$$\begin{aligned}\theta^* &= \sup \left\{ \theta \in \Theta : \frac{1 - F(\theta)}{f(\theta)} \phi - c \geq 0 \right\}, \\ \theta_0 &= \inf \left\{ \theta \in \Theta : \theta - \frac{1 - F(\theta)}{f(\theta)} + \left( \frac{1 - F(\theta)}{f(\theta)} \phi - c \right)_+ > 0 \right\},\end{aligned}$$

where we follow the convention that  $\sup \emptyset = -\infty$ . There are two cases.

If  $\theta_0 \geq \theta^*$ , then the mechanism  $(q^*, t^*, r^*, a^*, p^*)$  from Theorem 1 can be implemented by offering a singleton menu with a lump-sum contract at price  $\theta_0$ .

If  $\theta_0 < \theta^*$ , then the mechanism  $(q^*, t^*, r^*, a^*, p^*)$  from Theorem 1 can be implemented by offering a binary menu with a lump-sum contract at price  $(1 - \phi)\theta_0 + \phi\theta^*$  and a linear royalty contract at price  $(1 - \phi)\theta_0$ .

## A.7 PROOF OF THEOREM 2

Using the identity (12), we can express agent  $i$ 's virtual value as

$$\begin{aligned}\psi_i(\theta_i) &= \theta_i - \mathbb{E}_{\pi_i|\theta_i}[\mu_i(\theta_i, \pi_i)] + \mathbb{E}_{\pi_i|\theta_i}[(\mu_i(\theta_i, \pi_i)\phi_i - c_i)_+] \\ &= \mathbb{E}_{\pi_i|\theta_i}[\max\{\pi_i - (1 - \phi_i)\mu_i(\theta_i, \pi_i) - c_i, \pi_i - \mu_i(\theta_i, \pi_i)\}].\end{aligned}$$

Therefore,  $\psi_i(\theta_i)$  is weakly decreasing in  $c_i$  and weakly increasing in  $\phi_i$  (since  $\mu_i$  is nonnegative). Moreover,  $\psi_i(\theta_i)$  is weakly decreasing in  $\mu_i(\theta_i, \pi_i)$ , which is increasing in the hazard-rate order. Next, the comparative statics for  $\pi_i^*(\theta_i)$  are immediate upon noting that whenever  $\phi_i > 0$ , the auditing condition can be expressed as

$$(20) \quad \mu_i(\theta_i, \pi_i) \geq c_i/\phi_i.$$

If  $F_i$  increases in the hazard-rate order, then  $\mu_i(\theta_i, \pi_i)$  weakly increases, so (20) is relaxed and  $\pi_i^*(\theta_i)$  weakly increases. Conversely, if  $c_i/\phi_i$  increases, then (20) becomes more restrictive, so  $\pi_i^*(\theta_i)$  weakly decreases.

## REFERENCES

- Allingham, Michael G. and Agnar Sandmo**, “Income Tax Evasion: A Theoretical Analysis,” *Journal of Public Economics*, 1972, 1 (3–4), 323–338.
- Ball, Ian and Deniz Kattwinkel**, “Probabilistic Verification in Mechanism Design,” *Theoretical Economics*, 2025, 20 (4), 1247–1284.
- and **Teemu Pekkari**, “On Regularity and Normalization in Sequential Screening,” *Available at arXiv: 2410.17962*, 2024.
- Baron, David P. and David Besanko**, “Regulation and Information in a Continuing Relationship,” *Information Economics and Policy*, 1984, 1 (3), 267–302.
- Ben-Porath, Elchanan, Eddie Dekel, and Barton L. Lipman**, “Optimal Allocation with Costly Verification,” *American Economic Review*, 2014, 104 (12), 3779–3813.
- Bergemann, Dirk and Stephen Morris**, “Bayes Correlated Equilibrium and the Comparison of Information Structures in Games,” *Theoretical Economics*, 2016, 11 (2), 487–522.
- Bernhardt, Dan, Tingjun Liu, and Takeharu Sogo**, “Costly Auction Entry, Royalty Payments, and the Optimality of Asymmetric Designs,” *Journal of Economic Theory*, 2020, 188, 105041.
- Border, Kim C. and Joel Sobel**, “Samurai Accountant: A Theory of Auditing and Plunder,” *Review of Economic Studies*, 1987, 54 (4), 525–540.
- Che, Yeon-Koo and Jinwoo Kim**, “Bidding with Securities: Comment,” *American Economic Review*, September 2010, 100 (4), 1929–35.
- Contreras, Jorge L.**, *Intellectual Property Licensing and Transactions: Theory and Practice*, Cambridge University Press, 2022.
- Courty, Pascal and Hao Li**, “Sequential Screening,” *Review of Economic Studies*, 2000, 67 (4), 697–717.
- Crémer, Jacques**, “Auctions with Contingent Payments: Comment,” *American Economic Review*, 1987, 77 (4), 746.
- Crémer, Jacques and Richard P. McLean**, “Full Extraction of the Surplus in Bayesian and Dominant Strategy Auctions,” *Econometrica: Journal of the Econometric Society*, 1988, pp. 1247–1257.
- DeMarzo, Peter M., Ilan Kremer, and Andrzej Skrzypacz**, “Bidding with

- Securities: Auctions and Security Design,” *American Economic Review*, 2005, *95* (4), 936–959.
- Diamond, Douglas W.**, “Financial Intermediation and Delegated Monitoring,” *Review of Economic Studies*, 1984, *51* (3), 393–414.
- Eső, Péter and Balázs Szentes**, “Optimal Information Disclosure in Auctions and the Handicap Auction,” *Review of Economic Studies*, 2007, *74* (3), 705–731.
- Figuroa, Nicolas and Nicolas Inostroza**, “Optimal Screening with Securities,” *Journal of Economic Theory*, 2025, p. 106125.
- Gale, Douglas and Martin Hellwig**, “Incentive-Compatible Debt Contracts: The One-Period Problem,” *Review of Economic Studies*, 1985, *52* (4), 647–663.
- Gershkov, Alex, Benny Moldovanu, Philipp Strack, and Mengxi Zhang**, “Optimal Security Design for Risk-averse Investors,” *American Economic Review*, 2025, *115* (6), 2050–2092.
- Hansen, Robert G.**, “Auctions with Contingent Payments,” *American Economic Review*, 1985, *75* (4), 862–865.
- Kamien, Morton I. and Yair Tauman**, “Fees Versus Royalties and the Private Value of a Patent,” *Quarterly Journal of Economics*, 1986, *101* (3), 471–491.
- Katz, Michael L. and Carl Shapiro**, “How to License Intangible Property,” *Quarterly Journal of Economics*, 1986, *101* (3), 567–589.
- Kleven, Henrik Jacobsen, Martin B. Knudsen, Claus Thustrup Kreiner, Søren Pedersen, and Emmanuel Saez**, “Unwilling or Unable to Cheat? Evidence from a Tax Audit Experiment in Denmark,” *Econometrica*, 2011, *79* (3), 651–692.
- Li, Yunan**, “Mechanism Design with Costly Verification and Limited Punishments,” *Journal of Economic Theory*, 2020, *186*, 105000.
- Liu, Bin, Dongri Liu, and Jingfeng Lu**, “Shifting Supports in Eső and Szentes (2007),” *Economics Letters*, 2020, *193*, 109251.
- Liu, Tingjun**, “Optimal Equity Auctions with Heterogeneous Bidders,” *Journal of Economic Theory*, 2016, *166*, 94–123.
- **and Dan Bernhardt**, “Rent Extraction with Securities plus Cash,” *The Journal of Finance*, 2021, *76* (4), 1869–1912.
- Luo, Dan and Ming Yang**, “The Optimal Structure of Securities under Coordination Frictions,” *Available at SSRN: 4484914*, 2023.

- Mezzetti, Claudio**, “Mechanism Design with Interdependent Valuations: Efficiency,” *Econometrica*, 2004, *72* (5), 1617–1626.
- , “Mechanism Design with Interdependent Valuations: Surplus Extraction,” *Economic Theory*, 2007, *31* (3), 473–488.
- Milgrom, Paul and Ilya Segal**, “Envelope Theorems for Arbitrary Choice Sets,” *Econometrica*, 2002, *70* (2), 583–601.
- Myerson, Roger B.**, “Optimal Auction Design,” *Mathematics of Operations Research*, 1981, *6* (1), 58–73.
- Mylovanov, Tymofiy and Andriy Zapechelnyuk**, “Optimal Allocation with Ex Post Verification and Limited Penalties,” *American Economic Review*, 2017, *107* (9), 2666–94.
- Nachman, David C. and Thomas H. Noe**, “Optimal Design of Securities under Asymmetric Information,” *Review of Financial Studies*, 1994, *7* (1), 1–44.
- Palonen, Petteri and Teemu Pekkarinen**, “Avoidance, Auditing, and Optimal Contracts,” *Available at SSRN: 3729347*, 2026.
- Pavan, Alessandro, Ilya Segal, and Juuso Toikka**, “Dynamic Mechanism Design: A Myersonian Approach,” *Econometrica*, 2014, *82* (2), 601–653.
- Riley, John G.**, “Ex Post Information in Auctions,” *Review of Economic Studies*, 1988, *55* (3), 409–429.
- Sidak, J. Gregory**, “Converting Royalty Payment Structures for Patent Licenses,” *Criterion Journal of Innovation*, 2016, *1*, 901–915.
- Skrzypacz, Andrzej**, “Auctions with Contingent Payments—An Overview,” *International Journal of Industrial Organization*, 2013, *31* (5), 666–675.
- Sogo, Takeharu, Dan Bernhardt, and Tingjun Liu**, “Endogenous Entry to Security-Bid Auctions,” *American Economic Review*, 2016, *106* (11), 3577–3589.
- Townsend, Robert M.**, “Optimal Contracts and Competitive Markets with Costly State Verification,” *Journal of Economic Theory*, 1979, *21* (2), 265–293.