Informative Certification: Screening vs. Acquisition*

Gorkem Celik† and Roland Strausz‡

July 22, 2026

Abstract

We study monopolistic certification in markets where sellers possess partial private information about product quality. A certifier can provide information through two channels: screening sellers' private information (soft information) and acquiring new quality data (hard information). We prove that any certification menu achieving less than maximal screening is Pareto dominated by one with full screening. Among Pareto-efficient menus, the certifier's profit-maximising menu provides maximal soft information while restricting hard information provision. The two channels diverge because screening creates value the certifier can fully capture, whereas hard information amplifies costly information rents. Using power value functions, we derive comparative statics showing that information restrictions target low-quality sellers when information value is moderate, but high-quality sellers receive perfect quality revelation when information value is high.

Keywords: certification, disclosure, screening, information acquisition, monopolistic distortions

JEL classification: D82

*We are most indebted to Roberto Corrao for pointing out the broader implications of our analysis of the general informativeness of certification. As a result, this paper supersedes our earlier paper "Selling Certification of Private and Market Information". In addition, we thank Gary Biglaiser, Li Hao, In-Uck Park, Curt Taylor, Nikhil Vellodi, Ran Weksler, Boaz Zik, and seminar participants at University of Bristol, Paris Dauphine, Santa Clara University, UNC Chapel Hill, University of British Columbia, University of Colorado Boulder, Berlin School of Economics, Africa Business School at UM6P Rabat, Nazarbayev University, Virtual MD Seminar, Nordic Economic Theory Seminar, 24th Journées Louis-André Gérard-Varet, 24th Annual SAET Conference, IIOC 2025, and EEA Congress in Bordeaux for helpful discussions. Gorkem Celik acknowledges support from research program FSM CY Initiative. Roland Strausz acknowledges financial support from the European Union through the ERC-grant PRIVDIMA (project number 101096682) and the Deutsche Forschungsgemeinschaft through CRC-TRR 190 (project number 280092119).

†ESSEC Business School & THEMA Research Center. Email: celik@essec.edu.

‡Humboldt Universität zu Berlin & Berlin School of Economics. Email: strauszr@hu-berlin.de.

1 Introduction

Certification intermediaries have become central to the functioning of modern markets, serving as information gatekeepers between privately informed sellers and uninformed buyers. In financial markets, rating agencies evaluate complex securities whose risks are opaque to most investors, with their assessments directly affecting global capital allocation. Accounting firms audit corporate financial statements, transforming unverifiable management claims into verified public information. Digital platforms employ user-generated ratings to aggregate dispersed private experiences into public quality signals. These diverse institutions share a common economic function: they transform private or unobservable information into public, verifiable signals that facilitate market transactions.

The economic consequences of certification design are first-order. The role of credit rating agencies in the 2007-2008 financial crisis has been extensively documented, with evidence suggesting that the structure of their ratings was associated with systematic risk mispricing (Bolton, Freixas, and Shapiro, 2012; Bar-Isaac and Shapiro, 2013). The 2001 Enron-Arthur Andersen collapse illustrates a different certification failure: when auditors fail their gatekeeping function, the informational foundation of capital markets erodes (Healy and Palepu, 2003). These episodes highlight that how certifiers reveal and acquire information matters for market outcomes, yet fundamental questions about how certification provides information remain unresolved.

Despite a substantial body of theoretical literature on certification, a fundamental distinction remains unexplored: whether certifiers primarily serve to reveal existing private information or to generate new information previously unknown to any party. This distinction matters because these two roles operate through different economic mechanisms. When certification reveals private information, it functions as a screening device, using incentive compatibility to extract soft information from informed parties. When certification generates new information, it functions as an acquisition device, producing hard, verifiable data through direct investigation or testing.

The existing literature has largely treated these roles separately. One strand, beginning with Viscusi (1978) and Lizzeri (1999), models certifiers as revealing sellers' private information about product quality. Another strand, building on information design (Bizzotto, Rüdiger, and Vigier, 2020; Ali et al., 2022; Asseyer and Weksler, 2024), models certifiers as generating entirely new information. Yet in practice, certification typically performs both functions simultaneously. For instance, in the certification of pre-owned cars, even the seller learns new hard information, such as precise tyre-wear measurements or emissions readings, while the seller's decision whether to certify at all provides clear soft information about her private knowledge of the car's quality. Understanding how these dual roles interact—and whether a profit-maximising certifier has different incentives for screening versus acquisition—is essential for explaining observed certification market outcomes and designing optimal regulatory interventions.

This paper provides the first analysis of how these dual certification roles interact. Our main insight is that a monopolistic certifier exploits the screening role to the fullest, completely revealing all private information, while deliberately restricting its information acquisition role, leaving market uncertainties unresolved. This overturns the conventional intuition that information rents necessitate limited screening, and that a cost-free technology to acquire information leads to maximal acquisition. The economic mechanism driving this puzzle is that the screening and acquisition channels create value for the certifier in different ways. For screening, full separation creates value that the certifier can completely capture: by revealing each seller's type, the certifier enables perfect price discrimination while the soft information improves buyer-seller matching. Crucially, moving from any partial pooling to full separation increases the certifier's profit while leaving the buyer's and every seller type's payoff unchanged. This turns maximal screening into a tool for free value creation.

By contrast, the economic effects of hard information are entirely different: providing hard information requires the certifier to pay information rents that increase with the informativeness of certificates. When low-quality sellers receive more informative certificates, high-quality sellers, who benefit even more from informative certificates due to their higher probability of triggering favorable ones, must receive correspondingly larger rents to prevent them from mimicking low types. The complementarity between channels means that providing hard information to low types amplifies the screening costs for high types. Thus, hard information amplifies information rents while soft information does not. This reveals the main intuition for our result: soft information creates free value whereas hard information can only create value at a cost. Consequently, while the certifier always maximises screening to capture its free value, it restricts hard information provision to minimize costly rents. This latter effect can be so powerful that the certifier may provide no hard information at all, achieving perfect screening through menu design without paying any information rents.

Our findings have important policy implications. While monopolistic certification exhibits no distortions in screening—completely eliminating information asymmetries between sellers and buyers—it systematically underprovides hard information relative to the social optimum. Market inefficiencies persist not because of private information, but because the certifier deliberately preserves market uncertainty. This suggests regulatory interventions should focus on encouraging information acquisition rather than improving screening mechanisms.

We derive these results in a tractable buyer-seller model where quality is uncertain and the seller possesses partial private information. A monopolistic certifier can both screen the seller's type through menu design and provide hard information about true quality through information design. We characterize the optimal certification menu, showing it involves maximal screening—fully separating all seller types—while limiting information acquisition. The key insight is that information rents are minimized when hard information provision is restricted.

While the certification literature is vast (see Dranove and Jin, 2010), we are unaware of existing work that explicitly distinguishes between certification's screening and acquisition roles.

Models with either full private information (Viscusi, 1978; Lizzeri, 1999) or no private information (Bizzotto, Rüdiger, and Vigier, 2020; Ali et al., 2022) cannot capture this distinction. Existing models with partial private information that feature the dual roles of soft and hard information provision (e.g., Rosar, 2017; Bergemann, Bonatti, and Smolin, 2018; Kartik, Lee, and Suen, 2021; Ichihashi and Smolin, 2023; Weksler and Zik, 2023, 2024) do not explicitly analyze how these dual roles interact.

The remainder of the paper proceeds as follows. Section 2 presents the model setup. Section 3 establishes benchmarks and states our main theorems. Section 4 characterizes optimal maximal-screening menus. Section 5 proves that non-maximal screening is Pareto dominated. Section 6 analyzes the comparative statics of optimal certification. Section 7 discusses the robustness of our results. Section 8 concludes.

2 Setup

Allocations and Payoffs. We consider a seller (she) who sells a quantity x of a perfectly divisible good to a single, representative buyer (he). The seller's quality ω is either high (h) or low (l), i.e., $\omega \in \{h, l\}$. Production costs are independent of quality, preventing the seller from inferring quality from her costs. The seller faces a competitive fringe producing only low-quality goods at zero cost, establishing a competitive benchmark price of zero for low quality.[1]

The buyer's utility from consuming quantity x of quality level ω at price p is $U = u_\omega(x) - p$. We assume standard properties: $u_\omega(x)$ is twice differentiable with $u''_\omega(x) < 0$ (strict concavity), and satisfies single-crossing with $u'_h(x) > u'_l(x)$. Initial consumption has positive marginal value ($u'_\omega(0) > 0$), while $\lim_{x\to\infty} u'_\omega(x) < 0$. These properties ensure a unique efficient quantity x^*_ω satisfying $u'_\omega(x^*_\omega) = 0$, with $x^*_h > x^*_l > 0$ by single-crossing. Measuring utility relative to the buyer's outside option of purchasing low quality from the competitive fringe, we set $u^*_l = u_l(x^*_l) = 0$, and further normalize $u^*_h = u_h(x^*_h) = 1$. To reflect that h represents a higher quality than l, we also assume that $u_h(x) > u_l(x)$ for all $x \in [x^*_l, x^*_h]$.

Information structure. Let $\bar{\theta} \in (0, 1)$ denote the ex ante probability of high quality. While the buyer knows only this prior, the seller receives a private signal that updates her belief to some $\theta \in \Theta \equiv [0, 1]$, where θ represents both her type and her posterior probability of high quality.[2] Types are distributed according to CDF $F(\theta)$ with full-support density $f(\theta) > 0$. This partial private information structure, where quality is uncertain even to the seller, allows

[1] We discuss the role of this and other modeling assumptions in Section 7.

[2] It is without loss to assume that the signal coincides with the seller's updated belief as we can represent any signal this way. Formally, this means that the signal θ has the cumulative distribution $F^\omega(\theta)$ in state ω and $F(\theta) \equiv \bar{\theta}F^h(\theta) + (1-\bar{\theta})F^l(\theta)$ has support Θ and is such that its expectation equals the prior $\bar{\theta}$. Moreover, after receiving the signal θ, the seller's belief updates to θ. That is, the pdfs $f^l(\theta)$ and $f^h(\theta)$ are such that $\theta/(1-\theta) = \bar{\theta}f^h(\theta)/(1-\bar{\theta})f^l(\theta)$ for all $\theta \in \Theta$. The full-support assumption is for technical convenience and not crucial.

us to study certification's dual roles without the complications of multidimensional screening, because quality is binary.

Given that a seller type θ has high quality with probability θ and low quality with probability $1-\theta$, and that with high quality the efficient quantity x_h^* yields a surplus of $u_h^* = 1$, while with low quality the efficient quantity x_l^* yields a surplus of $u_l^* = 0$, the efficient expected surplus that a seller type θ can generate is

$$v^*(\theta) \equiv \theta\, u_h(x_h^*) + (1-\theta)\, u_l(x_l^*) = \theta.$$

Certification. A monopolistic certifier publicly offers a menu of certification contracts. Both the seller's choice from this menu and the resulting certificate outcome are publicly observed by all parties; the seller has no disclosure decision.[3] This framework generates two distinct information channels corresponding to different stages of the certification game.

First, following monopolistic screening, the certifier offers a *certification menu* $\Gamma = \{\gamma_i\}$ of *certification contracts*. Each contract $\gamma_i = (\sigma_i, t_i)$ commits the certifier to providing a *certification structure* σ_i at a flat price t_i that does not condition on the realised certificate.[4] The seller's choice from this menu reveals *soft* information.

Second, following information design, each certification structure $\sigma_i = (C_i, \pi_i^l, \pi_i^h)$ specifies a finite set of certificates C_i and the conditional probability distributions π_i^ω over these certificates in different states $\omega \in \{l, h\}$. After the seller selects contract γ_i, the certificate is drawn from σ_i. This generates *hard* information.

The certifier's objective. The certifier chooses a certification menu Γ and an equilibrium of the induced certification game so as to maximise its expected profit, where the expectation is taken across seller types according to the prior F. Throughout the paper, *the certifier's profit-maximising menu* refers to a solution of this problem.

The certification game. A certification menu $\Gamma = \{(\sigma_i, t_i)\}$ induces the following *certification game* between seller and buyer:

$t = 1$ Seller publicly selects some menu item i from the menu Γ, or declines to get certified;

$t = 2$ Certificate c is publicly drawn according to the selected certification structure σ_i (if any);

$t = 3$ Seller offers a quantity x at a price p to the buyer;

$t = 4$ Buyer accepts or rejects;

$t = 5$ Payoffs are realised.

[3] Section 7 provides motivation for this mandatory disclosure assumption and discusses robustness to the alternative where the seller controls disclosure.

[4] In Section 7 we explain in detail that the assumption of a flat price t_i that does not condition on the realised certificate, as in related literature, represents a crucial friction.

Our equilibrium concept is perfect Bayesian Equilibrium, describing both players' strategies and beliefs in this game. The strategy of the seller describes, for each type θ, her selection, $i(\theta)$, from the certification menu as well as her offer (x, p) conditional on the chosen menu item and the realized certificate. The strategy of the buyer is an accept-or-reject decision conditional on the seller's choice from the menu, the realized certificate, and the seller's offer (x, p). The buyer's equilibrium-path beliefs are consistent with the seller's strategy of certification-contract selection and the probability distributions governing the realization of different certificates.[5]

Illustrative Example. To ground our analysis, we use the following fully parameterized uniform-quadratic example throughout the paper.

Example: The seller's type θ is distributed uniformly over $[0, 1]$, implying $\bar{\theta} = 1/2$. For a low-quality good, the buyer has the quadratic utility function $u_l(x) = -(1 - x)^2$ so that the efficient quantity is $x_l^* = 1$, generating the normalized aggregate surplus of $u_l^* = u_l(1) = 0$. For a high-quality good, the buyer has the quadratic utility function $u_h(x) = 1 - (2 - x)^2$ so that the efficient quantity is $x_h^* = 2$, generating the normalized aggregate surplus of $u_h^* = u_h(2) = 1$. Hence, the efficient solution generates an ex ante expected surplus of $\mathbb{E}\{v^*(\theta)\} = \int_0^1 \theta d\theta = \bar{\theta} u_h(x_h^*) + (1 - \bar{\theta}) u_l(x_l^*) = 1/2$. $\square$

We use this example not only to illustrate the insights of our more general model but also to point out the extent to which this natural example is special. For instance, while the optimal certification menu for this example will feature maximal revelation of soft information and no revelation of any hard information, the former feature is general, whereas the latter is not.

With this framework in place, we now turn to a series of benchmarks to build intuition before presenting our main theorems.

3 Benchmarks and Theorems

In this section, we present four benchmarks that clarify how certification's dual roles, screening (soft information) and acquisition (hard information), generate different economic incentives. These benchmarks motivate our two main theorems.

1. No Certification. Without a certifier, the seller cannot credibly convey her private information to the buyer. Since the seller's costs and outside options are type-independent, no adverse selection arises à la Akerlof (1970). The buyer maintains belief $\bar{\theta}$ regardless of the seller's offer. Consequently, he expects a quantity x to yield a gross utility

$$\bar{u}(x) \equiv \mathbb{E}_\omega\{u_\omega(x)\} = \bar{\theta} u_h(x) + (1 - \bar{\theta}) u_l(x).$$

[5]Regarding beliefs, we assume that the buyer's beliefs are *passive* vis-à-vis the seller's offer (x, p) *off the equilibrium path.* This is justified by the observation that the seller's payoff function is independent of the quality of the good and her private information about it.

Hence, it is sequentially rational for the buyer to accept an offer (x, p) if and only if $p \leq \bar{u}(x)$. Thus, for some quantity x, the monopolistic seller optimally charges $p = \bar{u}(x)$.

It follows that her profit-maximising quantity, $\bar{x}$, maximises $\bar{u}(x)$, yielding her the revenue $\bar{\Pi} \equiv \bar{u}(\bar{x})$. The expressions $\bar{x}$ and $\bar{\Pi}$ depend on the ex ante probability, $\bar{\theta}$, that quality is high. To express this dependence explicitly, note that the buyer's optimal quantity at a belief θ is

$$\hat{x}(\theta) \equiv \arg\max_{x} \theta u_h(x) + (1-\theta)u_l(x) \tag{1}$$

so that the seller can extract the revenue

$$\hat{v}(\theta) \equiv \bar{u}(\hat{x}(\theta)) = \theta u_h(\hat{x}(\theta)) + (1-\theta)u_l(\hat{x}(\theta)). \tag{2}$$

The following lemma plays a crucial role in our analysis.

Lemma 1 *The revenue function $\hat{v}(\theta)$ is strictly increasing and strictly convex in θ. Moreover, $\hat{v}(\theta) \in [0, v^*(\theta))$ for all $\theta \in (0,1)$.*

Intuitively, the positive slope of $\hat{v}$ reflects that a high θ represents "good news" about the good's quality. Its strict convexity expresses that quality information is efficiency-relevant so that revealing additional information about the quality has strictly positive social value. The revenue function $\hat{v}(\theta)$ lies below the efficient surplus $v^*(\theta)$ because, due to the buyer's concave utility function, $\hat{x}(\theta) \in (x_l^*, x_h^*)$ for all $\theta \in (0,1)$.

Thus, without certification the seller offers quantity $\bar{x} = \hat{x}(\bar{\theta})$ and fully extracts the buyer's payoff, yielding her the revenue $\bar{\Pi} = \hat{v}(\bar{\theta})$.

Example: Applying this to our running example, the buyer's expected utility from a quantity x given a belief θ equals

$$\bar{u}(x) = \theta u_h(x) + (1-\theta)u_l(x) = (x-1)(1+2\theta-x).$$

It follows that the utility-maximising quantity equals $\hat{x}(\theta) = 1+\theta$, yielding the buyer the expected utility θ^2, which the seller can fully extract. Hence, in our example with quadratic utilities, we have

$$\hat{v}(\theta) = \theta^2,$$

confirming that $\hat{v}$ is increasing, convex, and lies below v^*. It follows that $\bar{x} = 1+\bar{\theta}$ and $\bar{\Pi} = \bar{\theta}^2$. $\square$

2. Certification Without Soft or Hard Information. Next consider a monopolistic certifier who can provide neither soft nor hard information to market participants. Effectively, the certifier can only use the uninformative certification structure $\sigma^u \equiv (\{c_\emptyset\}, 1, 1)$, which, irrespective of actual quality ω, always generates the uninformative certificate $c_\emptyset$.

We first verify that, in line with Lizzeri (1999), the certifier can nevertheless extract the entire surplus that seller types $\theta > 0$ generate. To see this, let the certifier charge a certification price $t = \bar{\Pi}$ for providing the uninformative certificate σ^u. That is, the certifier offers the certification menu $\Gamma^u \equiv \{(\sigma^u, \bar{\Pi})\}$. The menu is degenerate because it contains only a single certification contract.

The degenerate menu Γ^u induces a certification game with the following equilibrium outcome. All seller types $\theta > 0$ apply for certification, while seller type $\theta = 0$ does not. Upon seeing no certification, the buyer correctly anticipates $\theta = 0$; upon seeing a certificate, the buyer correctly anticipates $\mathbb{P}\{\omega = h\} = \bar{\theta}$.[6] Given these Bayes-consistent beliefs, it is then indeed optimal for a seller type $\theta = 0$ not to apply for certification and sell quantity x_l^* at price $p = 0$, while it is optimal for seller types $\theta > 0$ to apply for certification and, subsequently, sell a quantity $\hat{x}(\bar{\theta})$ of the certified good at the price $p = \bar{\Pi}$. In this outcome, the certifier fully extracts the generated surplus, yielding it the payoff $V^u = \bar{\Pi}$, while the buyer and all seller types obtain zero payoffs. The resulting aggregate surplus coincides with Benchmark 1, demonstrating that uninformative certification is purely rent-extracting without creating any value.

3. Certification with only Soft Information. Next consider a monopolistic certifier that can disclose the seller's private information through screening, but cannot provide any hard information about the product's quality ω. That is, the certifier only uses the uninformative certification structure σ^u, as in Benchmark 2.

We first argue that screening enables the certifier to *fully* disclose the seller's private information by offering the seller a menu of uninformative tests at different prices t. To show this, we consider a certification menu Γ that only contains the uninformative certification structure σ^u, but offers it at different prices. In particular, this menu allows the seller to buy the uninformative σ^u for any price between 0 and 1, i.e., $\Gamma^s \equiv \{(\sigma^u, \hat{v}(\theta))\}_{\theta \in [0,1]}$.

We claim that the menu Γ^s is incentive compatible—it is optimal for any seller type θ to reveal her type honestly by picking menu item θ from the menu Γ^s. To substantiate this claim, note that if the buyer anticipates that each seller type reveals her type, then a buyer who observes the seller picking menu item θ from the menu Γ^s updates his beliefs to θ. Facing a buyer with belief θ, it is optimal for any seller type to offer the quantity $\hat{x}(\theta)$ at a price $\hat{v}(\theta)$. This offer yields the seller the (gross) revenue $\hat{v}(\theta)$, which exactly matches the price of menu item θ. Hence, any pick from the menu Γ^s yields any seller type a profit of zero. As a result, any seller type θ is indifferent among all menu items in Γ^s, implying that it is indeed (weakly) optimal for a seller type θ to reveal herself by picking menu item θ from Γ^s.[7] Because, in this

[6] Hence, treating seller type $\theta = 0$ as a type who does not certify clarifies that the equilibrium does not depend on some contrived out-of-equilibrium belief when the buyer does not see any certification.

[7] As we explicitly show in Section 7, the menu Γ^s is the limit of a sequence of certification menus $\Gamma^s_\varepsilon \equiv \{(\sigma_\varepsilon(\theta), t_\varepsilon(\theta))\}_{\theta \in [0,1]}$ that provide strict incentives for truth-telling for any $\varepsilon > 0$. Hence, truth-telling in menu Γ^s is natural though it is only weakly optimal.

equilibrium of Γ^s, the certifier obtains the transfer $\hat{v}(\theta)$ from a seller type $\theta > 0$, the certification menu Γ^s yields the certifier the revenue

$$V^s \equiv \mathbb{E}_\theta\{\hat{v}(\theta)\} > \hat{v}(\mathbb{E}_\theta\{\theta\}) = \hat{v}(\bar{\theta}) = V^u, \tag{3}$$

where the inequality follows from the strict convexity of the revenue function $\hat{v}(\theta)$ as established in Lemma 1.

4. Certification with only Hard Information. Finally, consider the polar opposite of the previous benchmark: a certifier who only provides hard information and no soft information, foregoing any screening. In particular, suppose the certifier uses a certification scheme with a maximal degree of hard information, fully disclosing the good's quality ω. Hence, we consider a certifier who offers to the seller the fully informative certification structure $\sigma^f \equiv (\{c_l, c_h\}, (1,0), (0,1))$, where certificate c_l obtains for sure when quality is low ($\pi_l^l = 1$), while for high quality only certificate c_h obtains ($\pi_h^h = 1$). Because the certifier does not screen, the certification menus in this benchmark therefore take the form $\Gamma^h = \{(\sigma^f, t^f)\}$, leaving only the price t^f to be chosen.

From an aggregate surplus perspective, the certification structure σ^f is highly attractive as it results in the efficient quantity x_l^* when quality is low, and the efficient quantity x_h^* when quality is high.

For the certifier, however, providing maximal hard information to *all* seller types is unattractive. This is because a seller type θ obtains at most $(1-\theta)\hat{v}(0) + \theta\hat{v}(1) = \theta$ from buying σ^f. Consequently, if the certifier wants to induce all seller types $\theta > 0$ to participate in Γ^h, t^f must equal zero, forgoing any profits. Hence, we conclude that while feasible, it is not optimal for the certifier to maximise hard information and induce all seller types to accept it.

Two Theorems. We next compare the four benchmarks to develop an initial intuition about hard and soft information. The comparison also motivates the two theorems that we prove in the remainder of this paper.

Comparing Benchmarks 1 and 2 reveals that completely uninformative certification allows the monopolistic certifier to extract all generated rents in the market. It does so without raising overall surplus. This result is fully in line with Lizzeri (1999), who identified this parasitic nature of monopolistic certification in a unit-good framework where quality information is not efficiency-relevant.[8] In this case, it is optimal for the certifier to provide no information to the market.

Inequality (3) in Benchmark 3, however, qualifies Lizzeri's extreme result. It shows that when quality information is efficiency-relevant, the certifier obtains larger profits when pro-

[8]More precisely, the uninformative certification structure σ^u allows the certifier to extract all (non-maximal) rents in excess of what the worst seller type $\theta = 0$ obtains from the buyer. See also Kartik, Lee, and Suen (2021).

viding soft information by fully screening the seller's private information. Moreover, a direct comparison to Benchmark 2 reveals that this increased profit is not at the expense of the seller types or the buyer—their payoffs in Benchmarks 2 and 3 are identical. This means that the outcome of Benchmark 3 Pareto dominates the outcome of Benchmark 2.

As we prove in the remainder, this observation is a special case of a more general principle:

Theorem 1 *For every certification menu Γ, there exists a maximal-screening certification menu Γ' that yields expected payoffs for the certifier, the buyer, and every seller type that are at least as high as those under Γ. It is therefore without loss for Pareto efficiency to restrict attention to maximal-screening menus.*

Benchmark 4, in turn, shows that while the certifier could, in principle, provide certification that reveals all hard information perfectly, it would be suboptimal for the certifier to do so for all seller types. This identifies a misalignment between the certifier's incentives to provide hard information and social welfare because, from a pure welfare perspective, revealing all hard information is optimal: it maximises aggregate surplus. Note that the benchmarks suggest that this misalignment is specific to hard information. Indeed, it follows from Benchmark 4 that, when using certification to reveal hard information perfectly, the certifier cannot screen different seller types by asking them to pay different prices. By contrast, Benchmark 3 shows that the certifier can ask different seller types to pay different prices when using certification to reveal soft information perfectly.

In other words, Benchmark 4 implies that a certifier faces a trade-off between profit maximization and providing hard information, while Benchmark 3 indicates that there is no such trade-off between profit maximization and providing soft information. This suggests our second theorem:

Theorem 2 *There exists a profit-maximising certification menu for the certifier, Γ^*, that is a maximal-screening menu and that induces each seller type to reveal herself fully. Moreover, in Γ^* there exists a positive-measure subset $\hat{\Theta} \subseteq \Theta$ of types that receive a certification structure that is not fully revealing of the true quality ω.*[9]

Consequently, the certifier's profit-maximising certification menu exhibits no inefficiencies in the provision of soft information, whereas it exhibits the inefficiency of providing too little hard information.

In the remainder of this paper, we prove these two theorems. In doing so, we identify further properties of the certifier's optimal menu. In particular, we show the crucial role that the curvature of the convex revenue function, $\hat{v}(\theta)$, plays—namely, whether its average slope is larger or smaller than the geometric mean of its slopes.

[9] We do not claim uniqueness of Γ^*. On any subset of types receiving the fully revealing certification structure, pooling preserves payoffs, so profit-maximising menus exist that are not maximal-screening on that subset. The example in Section 6 exhibits such a non-empty subset $\Theta \setminus \hat{\Theta}$.

We prove the theorems in two steps. First, we focus, in Section 4, on maximal-screening menus which provide a maximal degree of soft information. We derive the profit-maximising one among these menus and establish its properties. In particular, we show that, in general, optimal maximal-screening menus may also provide some degree of hard information, but, in line with Benchmark 4, it is never strictly optimal for them to provide hard information to the fullest extent. Second, we show, in Section 5, that any certification menu that only partially screens the seller's private information is Pareto inferior to a certification menu that fully screens the seller's private information. Taken together, the two sections prove the two theorems above.

Example: We close this section by analyzing the different benchmarks for our running example. We already established that, in the first benchmark, where buyers do not learn more about quality than its ex ante probability $\bar{\theta}$, our example generates aggregate surplus $\hat{v}(\bar{\theta}) = 1/4$. In Benchmark 2, the certifier extracts this surplus completely; the certifier's maximum profit is 1/4. Moving from Benchmark 2 to Benchmark 3, where the certifier fully discloses the seller's private information to the buyer but acquires no hard information, the certifier's payoff and ex ante expected aggregate surplus increase from 1/4 to $\mathbb{E}_\theta\{\hat{v}(\theta)\} = \int_0^1 \theta^2 d\theta = 1/3$.[10]

Finally, in the last benchmark of providing maximal hard information, a surplus of $\hat{v}(0) = 0$ obtains for low quality and the surplus $\hat{v}(1) = 1$ obtains for high quality. From an ex ante perspective, the expected aggregate surplus is $\bar{\theta}\hat{v}(0) + (1 - \bar{\theta})\hat{v}(1) = 1/2$, while a seller type θ expects the (gross) payoff θ and the certifier's payoff is zero. Note that the certifier appropriates the entire surplus in Benchmarks 2 and 3, but does not earn any rents in Benchmark 4, as seller types arbitrarily close to 0 have no willingness to pay for the certification. □

4 Optimal Maximal-Screening Menus

We now study maximal-screening menus—those that fully disclose the seller's private information to the buyer, thus providing soft information perfectly. We denote such menus by Γ^S. As we prove in Section 5, these menus are optimal for the monopolistic certifier because any non-maximal-screening menu is Pareto dominated.

[10] We may consider the certifier's profits in an alternative "minimum soft information" arrangement, in which the certifier offers a single certification contract (σ, t) rather than a full menu, so that soft information arises only on the extensive margin of the seller's accept/reject decision. For our running example, the certifier-optimal one among such contracts uses an uninformative σ^u with the certifier-selected threshold $\theta^* = 1/4$ and price $t = 3/8$, yielding profit 9/32. The optimal hard information is therefore zero, the same conclusion as in our full menu solution. The difference between this benchmark and the full solution lies entirely in the soft-information dimension, which rises from the extensive margin to full revelation between the two.

4.1 Incentive-compatible direct mechanisms

Without loss of generality, we represent maximal-screening menus as incentive-compatible direct mechanisms. A direct menu is denoted by

$$\Gamma^D \equiv \{(C(\theta^r), \pi^l(\theta^r), \pi^h(\theta^r), t(\theta^r)\}_{\theta^r \in \Theta},$$

where $\mathcal{M}^D$ is the set of all such direct menus.

Because a maximal-screening menu fully discloses the seller's private information to the buyer, it resolves private information completely. After the seller has picked her certification contract from the menu, the buyer and seller have identical beliefs $\theta^B = \theta^S = \theta$.

After the seller reports θ and the certificate c is revealed, both buyer and seller share the posterior belief given by Bayes' rule:

$$\theta_c(\theta) \equiv \frac{\theta \pi_c^h(\theta)}{\theta \pi_c^h(\theta) + (1-\theta)\pi_c^l(\theta)} = \mathbb{P}\{\omega = h | c, \theta\}. \tag{4}$$

At the selling stage $t = 3$, for a given (θ, c), both parties expect a quantity x to yield the buyer the utility

$$\mathbb{E}_\omega\{u_\omega(x)|\theta, c\} = \theta_c(\theta)u_h(x) + (1 - \theta_c(\theta))u_l(x).$$

The buyer accepts an offer (x, p) if and only if $p \leq \mathbb{E}_\omega\{u_\omega(x)|\theta, c\}$. Thus, the seller optimally charges a price $p = \mathbb{E}_\omega\{u_\omega(x)|\theta, c\}$ and offers quantity $\hat{x}(\theta_c(\theta))$, extracting revenue $\hat{v}(\theta_c(\theta))$ as defined by equations (1)-(2).

We next formalize incentive compatibility for the direct menu Γ^D. When a seller type θ reports θ^r, she obtains revenue $\hat{v}(\theta_c(\theta^r))$ if certificate c realizes. Her expected payoff from report θ^r is therefore

$$\tilde{V}(\theta^r|\theta) \equiv \sum_c \left[\theta \pi_c^h(\theta^r) + (1-\theta)\pi_c^l(\theta^r)\right] \hat{v}(\theta_c(\theta^r)) - t(\theta^r).$$

This can be rewritten as $\tilde{V}(\theta^r|\theta) = \tilde{V}(\theta^r|\theta^r) + (\theta - \theta^r)I(\theta^r)$ because the difference in expected revenues between types θ and θ^r is proportional to their type difference so that

$$I(\theta) \equiv \sum_c [\pi_c^h(\theta) - \pi_c^l(\theta)]\hat{v}(\theta_c(\theta)) \tag{5}$$

represents the marginal information rent of type θ.

Denoting by $V(\theta) \equiv \tilde{V}(\theta|\theta)$ the seller type θ's payoff from revealing her type truthfully, Γ^D is incentive compatible (with truth-telling) if and only if, for all $\theta, \theta^r \in [0, 1]$,

$$V(\theta) \geq V(\theta^r) + (\theta - \theta^r)I(\theta^r). \tag{IC_{θ,θ^r}}$$

Moreover, a direct menu Γ^D is individually rational if for all $\theta \in \Theta$

$$V(\theta) \geq 0. \tag{IR_θ}$$

Hence, a maximal-screening menu can be represented as a direct menu Γ^D satisfying (IC_{θ,θ^r}) and (IR_θ).

Because the seller extracts full surplus from the buyer at the trading stage, the certifier's profit from type θ equals the total surplus generated minus the seller's information rent. The type-specific surplus is therefore

$$S(\theta) \equiv \sum_c [\theta \pi_c^h(\theta) + (1-\theta)\pi_c^l(\theta)]\hat{v}(\theta_c(\theta)) \tag{6}$$

yielding the certifier's expected profit

$$\Pi^c(\Gamma) = \int_0^1 [S(\theta) - V(\theta)]\ dF(\theta). \tag{7}$$

The certifier's optimal maximal-screening menu is therefore a direct one that solves the following maximization program:

$$\mathcal{P}^S : \max_{\Gamma \in \mathcal{M}^D} \Pi^c(\Gamma) \text{ s.t. } (IC_{\theta,\theta^r}) \text{ and } (IR_\theta) \text{ for all } \theta, \theta^r \in \Theta. \tag{8}$$

We denote a solution to the program by $\hat{\Gamma}^S$ and its associated value by $\hat{V}^S = \Pi^c(\hat{\Gamma}^S)$.

To solve this problem, we first show that standard monotonicity and payoff-equivalence conditions fully characterize incentive compatibility and individual rationality. To do so, we say that a direct menu Γ^D satisfies *monotonicity* if

$$I(\theta) \text{ is increasing;} \tag{MON}$$

and satisfies *payoff-equivalence* if

$$V(\theta) = \int_0^\theta I(\tau)d\tau + V(0). \tag{PE}$$

The following lemma characterizes a menu's incentive compatibility and its individual rationality in terms of monotonicity and payoff equivalence.

Lemma 2 *A direct menu Γ^D is incentive compatible if and only if it satisfies both (MON) and (PE). An incentive-compatible direct menu Γ^D exhibits an increasing V. An incentive-compatible direct menu Γ^D is individually rational if and only if $V(0) \geq 0$. If an incentive compatible direct menu Γ^D exhibits a strictly increasing $I(\theta)$, then all incentive compatibility constraints are strictly satisfied.*

The lemma confirms that $I(\theta)$ represents the marginal information rent. Using integration by parts on equation (7) and applying (PE), we obtain the virtual surplus formulation:

$$\hat{\mathcal{P}}^S : \max_{\Gamma \in \mathcal{M}^D} \hat{\Pi}^c(\Gamma) \equiv \int_0^1 \left[S(\theta) - \frac{1 - F(\theta)}{f(\theta)} I(\theta) \right] dF(\theta) \text{ s.t. } (MON). \tag{9}$$

The simplified problem, however, still maximises over a large set of direct menus. In particular, the optimal number of certificates remains undetermined.

In an indivisible, unit-good model in which there are only the two allocations—to buy or not to buy—feasibility considerations alone already imply that it suffices to consider only two certificates, as this leads to at most two posteriors. By contrast, in our setup with continuous quantities, feasibility considerations alone imply no limit on the optimal number of certificates.[11]

To address the question of the optimal number of certificates, we first derive the following optimality result that is specific to our framework.

Proposition 1 *There exists an optimal maximal-screening menu in which every contract uses at most three certificates.*

The proposition provides the crucial technical result that makes the optimization problem tractable. We prove it by showing that we can always replicate the value of any optimal maximal-screening menu that uses more than three certificates by one that uses at most three. The main idea is to reduce any solution with more than three posteriors to one with at most three posteriors. More specifically, the reduction follows because we can express it as the outcome of a linear optimization problem with 3 constraints. The proposition then follows from the general insight that there is a solution to a linear problem that has at most as many non-zero entries as there are constraints.

The proof reveals that one of the three constraints ensures monotonicity. When monotonicity is not binding, only 2 certificates suffice. This insight motivates a two-step procedure for finding optimal certification menus, analogous to standard mechanism design approaches:

Step 1: Solve program $\hat{\mathcal{P}}^S$ ignoring the monotonicity constraint (MON) and restricting to 2-certificate menus. This requires only point-wise maximization, making it analytically tractable. If the solution satisfies (MON), then by Lemma 3 (stated below), it is optimal among all maximal-screening menus.

Step 2: If the Step 1 solution violates (MON), solve $\hat{\mathcal{P}}^S$ with 3-certificate menus, explicitly imposing the monotonicity constraint.

[11] More specifically, the sufficiency of two posteriors in a unit good model reflects the revelation principle's insight that an optimal direct mechanism can be interpreted as providing incentive compatible recommendations over possible allocations—whether the good is bought or not. With continuous quantities, there is a continuum of possible allocations rather than only two and hence the revelation principle provides no limits on the number of possible recommendations.

4.2 Two-certificate menus

Following this two-step procedure, we next study 2-certificate menus in detail. Since 2-certificate menus have $C(\theta) = \{c_0, c_1\}$ for all θ, we can characterize such a menu by triples

$$\{(t(\theta), \pi^l(\theta), \pi^h(\theta))\}_{\theta \in \Theta},$$

where $t(\theta)$ is the transfer and $\pi^\omega(\theta)$ is the probability that certificate c_1 is revealed when quality is $\omega \in \{l, h\}$.

Based on a relabeling argument, it is without loss to assume that $\pi^l(\theta) \leq \pi^h(\theta)$ so that certificate c_0 is more indicative of low quality and certificate c_1 is more indicative of high quality. Indeed, if the seller reports θ, thus selecting the pair $(\pi^l(\theta), \pi^h(\theta))$ from a 2-certificate menu, then this induces the belief

$$\hat{\theta}_0(\theta) = \frac{\theta(1 - \pi^h(\theta))}{\theta(1 - \pi^h(\theta)) + (1 - \theta)(1 - \pi^l(\theta))} \text{ and } \hat{\theta}_1(\theta) = \frac{\theta \pi^h(\theta)}{\theta \pi^h(\theta) + (1 - \theta)\pi^l(\theta)} \tag{10}$$

after observing, respectively, certificate c_0 and c_1, whenever this certificate occurs with positive probability. This confirms that $\pi^l(\theta) \leq \pi^h(\theta)$ implies $\hat{\theta}_0(\theta) \leq \theta \leq \hat{\theta}_1(\theta)$ for all θ.

Inverting the expressions in (10), we obtain the probabilities $\pi^h(\theta)$ and $\pi^l(\theta)$ that, for a given prior θ, generate the posteriors $q = \hat{\theta}_0$ after seeing certificate $c = c_0$ and $p = \hat{\theta}_1$ after seeing certificate $c = c_1$:

$$\pi^h(\theta) = \frac{(\theta - q)p}{(p - q)\theta} \text{ and } \pi^l(\theta) = \frac{(\theta - q)(1 - p)}{(p - q)(1 - \theta)}. \tag{11}$$

For 2-certificate contracts, it is easier to work directly with the posteriors (p, q) rather than their inducing probabilities (π^l, π^h). Hence, we express an incentive compatible 2-certificate menu Γ_2^D as a triple consisting of the transfer plus a pair of posterior beliefs (p, q) with $0 \leq q \leq \theta \leq p \leq 1$:

$$\Gamma_2^D = \{(t(\theta), q(\theta), p(\theta))\}_{\theta \in \Theta}.$$

We denote the set of 2-certificate menus with maximal-screening by $\mathcal{M}_2^D$.

We next define

$$I_2(p, q|\theta) \equiv \frac{[p - \theta][\theta - q]}{(1 - \theta)\theta} \frac{[\hat{v}(p) - \hat{v}(q)]}{(p - q)}. \tag{12}$$

The first fraction in this expression captures the informational spread, while the second fraction represents the average value gain from the belief change. We can further exploit the simple structure of 2-certificate menus to simplify the monotonicity requirement; the summation expression in (MON) reduces to $I_2(p(\theta), q(\theta)|\theta)$. It follows from Lemma 2 that a 2-certificate menu Γ_2^D is incentive compatible if and only if

$$I_2(p(\theta), q(\theta)|\theta) \text{ is increasing in } \theta. \tag{MON_2}$$

Rewriting the objective of program $\hat{\mathcal{P}}^S$ for the specific case of 2-certificate menus in terms of the two posteriors, we obtain

$$V_2 \equiv \int_0^1 \left[S_2(p(\theta), q(\theta)|\theta) - \frac{1-F(\theta)}{f(\theta)} I_2(p(\theta), q(\theta)|\theta) \right] dF(\theta) \tag{13}$$

where

$$S_2(p, q|\theta) \equiv \frac{\theta - q}{p - q} \hat{v}(p) + \frac{p - \theta}{p - q} \hat{v}(q). \tag{14}$$

$S_2(p, q|\theta)$ expresses the surplus generated by a type θ who induces posteriors q and p from a 2-certificate menu. The function $I_2(p, q|\theta)$ captures the associated marginal information rent, scaled by the inverse hazard rate.

Accordingly, we can express the first step of the procedure as follows. Find a pair of posterior functions $(p(\theta), q(\theta))$ with $p(\theta) \geq \theta \geq q(\theta)$ that maximises the virtual surplus expression V_2 point-wise. That is, we solve for each $\theta \in \Theta$ the problem

$$\mathcal{R}_\theta^S : \max_{(p,q): q \leq \theta \leq p} S_2(p, q|\theta) - \frac{1-F(\theta)}{f(\theta)} I_2(p, q|\theta).$$

The next lemma confirms that if the resulting pair of functions $(\hat{p}(\theta), \hat{q}(\theta))$ satisfies the monotonicity constraint (MON_2), then it solves $\hat{\mathcal{P}}^S$.

Lemma 3 *Suppose $(\hat{p}(\theta), \hat{q}(\theta))$ with value $\hat{V}_2$ solves the unconstrained optimization problem $\mathcal{R}_\theta^S$ for each $\theta \in \Theta$. If $(\hat{p}(\theta), \hat{q}(\theta))$ satisfies (MON_2), then $\hat{V}_2$ is the value of program $\hat{\mathcal{P}}^S$, i.e. $\hat{V}^S = \hat{V}_2$.*

The proof of Lemma 3 mirrors the proof of Proposition 1, showing that, with the help of a linear maximization problem that now only consists of two constraints, one can replicate any solution with more than 2 posteriors by a maximal-screening menu with at most 2 posteriors.

To illustrate the procedure, we now apply the first step of our procedure to our running example; it alone already yields the optimal maximal-screening menu.

Example: Applying the first step, we need to find, for each $\theta \in \Theta$, a pair $(\hat{p}(\theta), \hat{q}(\theta))$ that solves

$$\max_{(p,q): q \leq \theta \leq p} S_2(p, q|\theta) - \frac{1-F(\theta)}{f(\theta)} I_2(p, q|\theta) = 2\theta(p+q) - 3qp - p^2 - q^2 + \frac{pq}{\theta}(p+q).$$

The first-order conditions of this quadratic optimization problem yield the optimum $\hat{p}(\theta) = \theta$ and $\hat{q}(\theta) = 0$ so that by (12), we have

$$I_2(\hat{p}(\theta), \hat{q}(\theta)|\theta) = I_2(\theta, 0|\theta) = 0,$$

which satisfies the monotonicity condition (MON_2) trivially. With $\hat{p}(\theta) = \theta$ and $\hat{q}(\theta) = 0$, the certification structure effectively collapses to a single certificate that maintains the prior belief

θ, providing no hard information while still achieving maximal screening through the menu of prices. $\square$

4.3 When two certificates suffice: a curvature condition

Lemma 3 implies that problems for which the monotonicity condition is non-binding possess a high degree of tractability in two respects. First, they require only the consideration of 2-certificate menus. Second, they only require a point-wise maximization rather than a full-fledged dynamic optimization. Consequently, it would be helpful to characterize sufficient conditions on the model's primitives guaranteeing that (MON) is non-binding. While in classical monopolistic screening problems, such sufficient conditions often exist in the form of assumptions only on the distribution of types, a sufficient condition in our certification context also requires assumptions involving the revenue function $\hat{v}(\theta)$.[12] The fact that our uniform-quadratic example has non-binding (MON) despite not satisfying these sufficient conditions illustrates that these conditions are sufficient but not necessary.

For this reason, we next develop a more practical condition for 2-certificate optimality based on the curvature of $\hat{v}$. The key insight, illustrated in Figure 1, involves comparing the average slope between two posteriors to the geometric mean of the slopes at those points. To formalize this, consider a pair $p, q \in [0,1]$ with $p > q$ and let

$$w(p,q) \equiv \frac{\hat{v}(p) - \hat{v}(q)}{p-q} \tag{15}$$

express the average slope of $\hat{v}$ between the two points p and q, and

$$g(p,q) \equiv \sqrt{\hat{v}'(p)\hat{v}'(q)}$$

express the geometric mean of the slopes at p and q.

The convexity of $\hat{v}$ ensures that $\hat{v}'(q) < w(p,q) < \hat{v}'(p)$. To determine which inequality dominates, we examine

$$A(p,q) \equiv g(p,q) - w(p,q). \tag{16}$$

When $A(p,q) < 0$, the average slope exceeds the geometric mean, indicating that the revenue function's growth decelerates between q and p. Conversely, when $A(p,q) > 0$, the geometric mean exceeds the average slope, suggesting accelerating growth. This distinction proves crucial for characterizing optimal posteriors.

In particular, we obtain the following insight:

[12] A sufficient condition for general 2-certificate menus is a decreasing inverse hazard rate $[1-F(\theta)]/f(\theta)$ together with a non-negative cross-partial derivative, $B_{I\theta}(I,\theta)$, of the function

$$B(I,\theta) \equiv \max_{\{p,q: q \le \theta \le p \wedge I_2(p,q) = I\}} S_2(p,q).$$

Figure 1: Curvature measure comparing the average slope $w(p,q)$ to the geometric mean $g(p,q)$ of the marginal slopes $\hat{v}'(q)$ and $\hat{v}'(p)$

Lemma 4 *There exists an optimal maximal-screening menu, $\hat{\Gamma}^S = \{\sigma(\theta), t(\theta)\}_{\theta\in\Theta}$, such that for any $\theta \in \Theta$, and for any two distinct posteriors p and q with $p > q$ induced by $\sigma(\theta)$, we have*

$$\begin{aligned} A(p,q) &\leq 0 \quad \text{if } q = 0 \text{ and } p < 1; \\ A(p,q) &= 0 \quad \text{if } q > 0 \text{ and } p < 1; \\ A(p,q) &\geq 0 \quad \text{if } q > 0 \text{ and } p = 1. \end{aligned} \tag{17}$$

To understand the lemma, consider a seller type θ whose contract induces posteriors $q < \theta < p$. The certifier can perturb the contract by raising or lowering both posteriors while keeping the marginal information rent $I(\theta)$ fixed. The conditions in the lemma are exactly the first-order conditions for optimality of $S(\theta)$ along this constant-rent path.

We next explain why a comparison of the average slope, $w(p,q)$, and the geometric mean, $g(p,q)$, determines the optimal posteriors p and q that maximize $S(\theta)$ subject to a constant information rent $I(\theta)$. For this, note first that the type-specific surplus $S(\theta)$ is the weighted average of the values that $\hat{v}(\cdot)$ takes at all the induced posteriors, weighted by the probabilities of these posteriors. Next note that changing two posteriors, say p and q, while keeping the information rent $I(\theta)$ constant, requires moving these posteriors in the same direction. Now consider a simultaneous increase of these posteriors p and q, with $p > q$. On the one hand, their common increase raises the values of $\hat{v}(\cdot)$ as measured by the derivatives $\hat{v}'(p)$ and $\hat{v}'(q)$. This raises the type-specific surplus $S(\theta)$. On the other hand, the increase in both p and q implies a reallocation of weight from the high posterior p to the low posterior q because Bayes consistency requires that the mean posterior remains at θ. This reallocation lowers $S(\theta)$ and does so at a rate scaled by the average slope $w(p,q)$. Thus, the requirement of keeping the information rent $I(\theta)$ constant yields a trade-off between a positive effect that depends on $\hat{v}'(p)$ and $\hat{v}'(q)$, and a negative effect that depends on $w(p,q)$. The proof of Lemma 4 makes this precise, showing that the overall net change carries exactly the sign of $\hat{v}'(p)\hat{v}'(q) - w(p,q)^2$. This sign is equivalent to whether the geometric mean $g(p,q)$ is larger or smaller than the average slope $w(p,q)$.

We note that the conditions in the lemma are independent of the seller type θ that chooses the contract. They reference only the geometric mean and the average slope of $\hat{v}$ between the two posteriors.

Exploiting this structural property of our certification game, Lemma 4 directly implies an alternative, more practical sufficient condition for the optimality of 2-certificate menus than the one based on non-binding monotonicity constraints, as discussed above. More specifically, if the sign of $A(p,q)$ is independent of (p,q), there exists an optimal menu relying on only two certificates. This follows because Lemma 4 shows that with constant sign of $A(p,q)$, at least one posterior must be at a boundary (0 or 1, depending on the sign of A). Hence it is impossible to induce three distinct posteriors, since two of them would then have to share the same boundary value.

Corollary 1 *Suppose for all $p > q$, the average slope $w(p,q)$ exceeds the geometric mean $g(p,q)$, i.e., $A(p,q) \leq 0$. Then there exists an optimal maximal-screening menu that is a 2-certificate menu with $q^*(\theta) = 0$. Suppose for all $p > q$, the geometric mean $g(p,q)$ exceeds the average slope $w(p,q)$, i.e., $A(p,q) \geq 0$. Then there exists an optimal maximal-screening menu that is a 2-certificate menu with $p^*(\theta) = 1$. Moreover, contracts within the menu degenerate to uninformative certification when one of the optimal posteriors coincides with the seller's prior θ in either case.*

The corollary follows directly from the lemma. Suppose $A(p,q) < 0$ holds for every $p > q$. If a contract induced two distinct posteriors $p > q$ with $q > 0$, then the lemma would require $A(p,q) \geq 0$, contradicting the constant-sign assumption. Hence q must equal 0, ruling out three distinct posteriors. The case $A(p,q) > 0$ is symmetric and pins p at 1.[13]

The two cases expressed in the corollary do not necessarily imply that certification that provides hard information is optimal. Indeed, this would only be so if the fully revealing certificate is sent with a strictly positive probability, which the corollary does not claim. On the contrary, the corollary allows this probability to be zero, implying that certificate c_1 is actually sent with probability 1 so that the seller's and the buyer's beliefs about the state remain constant—the certification does not reveal any hard information. Our example demonstrates this subtle but important point.

Example: For our example, we established that $\hat{v}(\theta) = \theta^2$. As a result, $A(p,q) = -\left[\sqrt{p} - \sqrt{q}\right]^2 < 0$ for all $p > q$. From Corollary 1 it then follows that 2-certificate contracts are optimal and one of the two certificates must exhibit a degenerate posterior $q^*(\theta) = 0$. A naive interpretation of this result would suggest that fully revealing the good's low quality ($\omega = l$) is optimal. However, as established in our previous calculation, the optimal certification contract also exhibits $p^*(\theta) = \theta$. The latter implies that an optimal certification contract sends the low-quality certificate with probability 0. Although the structure is formally a 2-certificate one, only certificate c_1 is ever observed, effectively degenerating to uninformative certification. Thus, the certifier's optimal menu involves no acquisition of hard information but only discloses the seller's type θ to the buyer through screening.

It is important to point out that the optimality of providing no hard information is a special feature of the uniform distribution. For instance, when the type distribution has the concave

[13] When $A(p,q) \equiv 0$ on $\{(p,q) : p > q\}$, every 2-certificate maximal-screening menu that generates the same marginal information rent yields the same surplus at every type. For this case, starting with any (p,q) pair, as we move these posteriors either up or down while keeping the marginal information rent constant, the surplus function does not change. The convex value function $\hat{v}(\theta) = \theta/(2-\theta)$ is the leading example.

form $F(\theta) = \sqrt{\theta}$, both certificates are sent with strictly positive probabilities for each seller type in the interior of the support $[0, 1]$. In this case, optimal certification menus do provide a strictly positive degree of hard information.[14] In addition to the uniform distribution, the quadratic utility in the example is also crucial for the optimality of no hard information. Indeed, the corollary itself already indicates that the specific functional form of $\hat{v}$ matters. $\square$

Because the sign of $A(p, q)$ depends on two variables, the corollary's condition is cumbersome to check for general value functions $\hat{v}(\theta)$. For this reason, we provide a more practical sufficient condition that depends only on one variable. In particular, consider the following *scaled curvature measure*, which measures how the slope, $\hat{v}'(\theta)$, changes relative to its curvature, $\hat{v}''(\theta)$:

$$\hat{A}(\theta) \equiv \frac{\hat{v}'(\theta)\sqrt{\hat{v}'(\theta)}}{\hat{v}''(\theta)}.$$

This measure captures how rapidly the marginal value $\hat{v}'(\theta)$ grows relative to the rate of change in that growth, $\hat{v}''(\theta)$. When $\hat{A}(\theta)$ is increasing, the revenue function exhibits decelerating growth, whereas the growth is accelerating when $\hat{A}(\theta)$ is decreasing.

Lemma 5 *Suppose the scaled curvature measure $\hat{A}(\theta)$ is strictly increasing for $\theta > 0$. Then $A(p, q) < 0$ for all $p > q$. Suppose the scaled curvature measure $\hat{A}(\theta)$ is strictly decreasing for $\theta > 0$. Then $A(p, q) > 0$ for all $p > q$.*

The lemma will be helpful in extending the analysis of our example to more general value functions than the quadratic one. Indeed, we will do so in Section 6 when discussing the comparative statics of the optimal certification menu and further economic implications.

5 Maximal-Screening and Pareto Efficiency

This section proves Theorem 1: for every certification menu, there exists a maximal-screening menu that yields the certifier, the buyer, and every seller type expected payoffs at least as high as those under the original menu. The constructed maximal-screening alternative preserves the buyer's utility and all seller types' rents and weakly raises the certifier's expected profit, with strict improvement whenever the original menu pools types on a certification structure that is not fully revealing.

Section 3 illustrated this principle through Benchmarks 2 and 3. Both the non-screening menu $\Gamma^u = \{(\sigma^u, \bar{\Pi})\}$ and the maximal-screening menu $\Gamma^S \equiv \{(\sigma^u, \hat{v}(\theta))\}_{\theta\in[0,1]}$ leave all seller types with zero payoff. Yet Γ^S generates strictly higher expected surplus, demonstrating that the certifier gains from moving to maximal screening at no cost to the other parties.

We generalize this insight by leveraging results from Section 4 in three steps. First, we extend Lemma 2's characterization to non-maximal-screening menus. Second, we use the

[14] One may show that, optimally, $q^*(\theta) = 0$ and $p^*(\theta) = \frac{1}{4}\sqrt{\theta} + \frac{3}{4}\theta > \theta$.

certificate-reduction argument from Proposition 1 to show that menus with at most three certificates yield the certifier, the buyer, and every seller type expected payoffs at least as high as menus with more certificates. Third, we prove the same comparison between maximal-screening menus and non-maximal-screening menus that use at most three certificates.

Notation. To capture non-maximal-screening menus in general, we extend the notation from Section 4. Intuitively, a non-maximal-screening menu induces different seller types to pool by choosing the same certification contract $\{C, \pi^l, \pi^h, t\}$ from the menu.

To formalize this pooling, fix a (possibly non-direct) menu of contracts

$$\Gamma = \left\{C(m), \pi^l(m), \pi^h(m), t(m)\right\}_{m \in M},$$

where M is some message set.

The menu Γ leads to a certification game in which the seller first picks a certification contract from Γ by sending some message m. Let $\hat{m} : \Theta \to M$ represent the seller's *messaging strategy*, meaning that seller type θ picks message $\hat{m}(\theta)$.[15] It is without loss to assume that M contains only messages that at least some type θ sends—any contract that is not used can be pruned from Γ without affecting the equilibrium outcome of the induced certification game. Consequently, the subset of types

$$\hat{\Theta}(m) \equiv \{\theta | \hat{m}(\theta) = m\} \subseteq \Theta$$

that pick message m is non-empty for each m, and the collection of sets $\{\hat{\Theta}(m)\}_{m \in M}$ forms a partition $\mathcal{M}$ of Θ.

If $\hat{\Theta}(m)$ is a singleton for each $m \in M$, the menu achieves maximal screening as analyzed in Section 4. By contrast, the menu is non-maximal-screening if for some message $m \in M$, the set $\hat{\Theta}(m)$ contains multiple types. The coarseness of the partition $\mathcal{M}$ thus measures the degree of pooling.

To capture the information that a seller reveals when picking a message m, let $\bar{\theta}^M : M \to \Theta$ denote the average type associated with a message m:

$$\bar{\theta}^M(m) \equiv \frac{\int_{\theta \in \hat{\Theta}(m)} \theta dF(\theta)}{\int_{\theta \in \hat{\Theta}(m)} dF(\theta)}.$$

The mapping $\bar{\theta}^M(m)$ captures the soft information of certification. It reflects the buyer's updating of his belief from $\bar{\theta}$ to $\bar{\theta}^M(m)$ when observing the seller pick the certification contract m from the menu Γ. In particular, a maximal-screening menu exhibits $\bar{\theta}^M(\hat{m}(\theta)) = \theta$ for all $\theta \in \Theta$. Hence, $\bar{\theta}^M$ extends the updating step of soft information in the previous section, where the seller's pick from the menu Γ fully reveals her type.

To capture the menu's acquisition of hard information, let $\hat{\theta}_c^M : M \to \Theta$ denote the mapping

[15] While our formal analysis focuses on sellers sending a deterministic message, the argument can be extended to sellers picking menu items randomly, but this requires additional notation to track the respective mixing.

such that $\hat{\theta}_c^M(m)$ represents the Bayes-consistent expectation of θ given the seller's message $m \in M$ and certificate $c \in C(m)$. For any $m \in M$ and $c \in C(m)$, we have

$$\hat{\theta}_c^M(m) \equiv \frac{\bar{\theta}^M(m)\pi_c^h(m)}{\bar{\theta}^M(m)\pi_c^h(m) + (1-\bar{\theta}^M(m))\pi_c^l(m)}.$$

Thus, for a non-maximal-screening menu, the change in beliefs from $\bar{\theta}$ to $\bar{\theta}^M(m)$ represents the disclosure via soft information, whereas the change in beliefs from $\bar{\theta}^M(m)$ to $\hat{\theta}_c^M(m)$ expresses the extent to which the certification menu provides hard information.

Anticipating the buyer's belief $\hat{\theta}_c^M(m)$, all seller types choosing message m and seeing certificate $c \in C(m)$ optimally offer the same contract with quantity $x = \hat{x}(\hat{\theta}_c^M(m))$ and a price equal to the buyer's expected utility at that quantity. We now characterize the incentive-compatibility constraints induced by this messaging structure.

Step 1: Characterizing IC. Equipped with this notation, we can extend the notion of incentive compatibility to non-maximal menus. In particular, we say that a menu Γ together with the seller's messaging strategy $\hat{m}$ is incentive compatible if for any type θ, the message $\hat{m}(\theta)$ yields type θ the highest payoff among all messages $m \in M$. That is, for each type $\theta \in \Theta$, the following incentive compatibility constraint is satisfied

$$\begin{aligned} \sum_c \left[\theta\pi_c^h(\hat{m}(\theta)) + (1-\theta)\pi_c^l(\hat{m}(\theta))\right] \hat{v}(\hat{\theta}_c^M(\hat{m}(\theta))) - t(\hat{m}(\theta)) \geq \\ \sum_c \left[\theta\pi_c^h(m) + (1-\theta)\pi_c^l(m)\right] \hat{v}(\hat{\theta}_c^M(m)) - t(m), \quad \forall m \in M. \end{aligned} \tag{18}$$

Consequently, we say that a (general) certification menu is incentive compatible with respect to messaging strategy $\hat{m} : \Theta \to M$ if for all $\theta, \theta^r \in \Theta$ we have

$$\hat{V}(\theta) \geq \hat{V}(\theta^r) + (\theta - \theta^r)I^{\hat{m}}(\theta^r), \tag{$\widehat{IC}_{\theta,\theta^r}$}$$

where

$$\hat{V}(\theta) \equiv \sum_c \left[\theta\pi_c^h(\hat{m}(\theta)) + (1-\theta)\pi_c^l(\hat{m}(\theta))\right] \hat{v}(\hat{\theta}_c^M(\hat{m}(\theta))) - t(\hat{m}(\theta)) \tag{19}$$

expresses the utility of seller type θ sending her message $\hat{m}(\theta)$, and

$$I^{\hat{m}}(\theta) \equiv \sum_c \left[\pi_c^h(\hat{m}(\theta)) - \pi_c^l(\hat{m}(\theta))\right] \hat{v}(\hat{\theta}_c^M(\hat{m}(\theta))) \tag{20}$$

expresses the marginal information rent of type θ. Note that ($\widehat{IC}_{\theta,\theta^r}$) differs from ($IC_{\theta,\theta^r}$) in that for a non-maximal menu, we have to keep track of the message $\hat{m}(\theta)$ that seller type θ sends.

Lemma 6 *A (general) certification menu Γ is incentive compatible with respect to messaging* *strategy* $\hat{m} : \Theta \to M$ *if and only if it satisfies both:*

(i) Monotonicity:

$$I^{\hat{m}}(\theta) \text{ is increasing in } \theta. \tag{$\widehat{MON}$}$$

(ii) Payoff-equivalence:

$$\hat{V}(\theta) = \int_0^\theta I^{\hat{m}}(\tau)d\tau + \hat{V}(0). \tag{$\widehat{PE}$}$$

The lemma extends Lemma 2 because, for a direct maximal-screening menu, we implicitly have the identity $\hat{m}(\theta) = \theta$ as the messaging strategy, in which case ($\widehat{MON}$) and ($\widehat{PE}$) reduce to (MON) and (PE), respectively.

Step 2: Optimizing the Number of Certificates. Generalizing the certifier's problem in Section 4, we can view the certifier as solving a two-step maximization procedure. First, the certifier chooses a partition $\mathcal{M}$ of the type space, determining which types pool together. In this step, maximal-screening amounts to choosing the finest partition. Second, for each element of the partition, the certifier designs the optimal certification contract, explicitly choosing the number of certificates each contract contains.

The proof of Proposition 1 shows that one can reduce the number of certificates of any contract to at most three without affecting the seller's marginal information rent or the buyer's payoff, while (potentially) raising the certifier's payoff. Hence, we obtain the following result.

Lemma 7 *For every certification menu* Γ *that contains contracts with more than three certificates, there exists a certification menu* $\hat{\Gamma}$ *consisting only of contracts with at most three certificates that yields the certifier, the buyer, and every seller type expected payoffs at least as high as under* Γ.

This allows us to focus on menus with at most three certificates without loss of generality.

Step 3: Maximal-screening dominates non-maximal-screening. By Lemma 7, it remains to show that for every non-maximal-screening menu that uses contracts with at most three certificates, there exists a maximal-screening menu that yields the certifier, the buyer, and every seller type expected payoffs at least as high.

Step 3a: 2-certificate menus. To establish the result for 2-certificate contracts, consider a general (possibly non-maximal) 2-certificate menu Γ_2 with $C(\theta) = \{c_p, c_q\}$ for all $\theta \in \Theta$. We characterize it by a partition $\mathcal{M}$ of Θ and posterior mappings $(p(\theta), q(\theta))$ with $q(\theta) \leq p(\theta)$. Crucially, types within the same partition cell must induce identical posteriors: $(p(\theta_1), q(\theta_1)) = (p(\theta_2), q(\theta_2))$ if θ_1 and θ_2 belong to the same element of $\mathcal{M}$.

In terms of posterior pairs $(p(\theta), q(\theta))$ and using (15), a seller type θ generates surplus

$$\hat{S}(\theta) \equiv \hat{v}(q(\theta)) + [\bar{\theta}(\theta) - q(\theta)] \left[\frac{\theta p(\theta)}{\bar{\theta}(\theta)} + \frac{(1-\theta)(1-p(\theta))}{(1-\bar{\theta}(\theta))} \right] w(p(\theta), q(\theta)), \tag{21}$$

where $\bar{\theta}(\theta) = \bar{\theta}^M(\hat{m}(\theta))$ is the average type in the pool containing θ.

Based on (20), we can also use (15) to rewrite type θ's marginal information rent as

$$\hat{I}(\theta) = \frac{p(\theta) - \bar{\theta}(\theta)}{1 - \bar{\theta}(\theta)} \frac{\bar{\theta}(\theta) - q(\theta)}{\bar{\theta}(\theta)} w(p(\theta), q(\theta)). \tag{22}$$

The information rent reflects the spread between posteriors $(p(\theta), q(\theta))$ relative to the pool's average belief $\bar{\theta}(\theta)$, scaled by the average value gain $w(p(\theta), q(\theta))$. By Lemma 6, the pair $(p(\theta), q(\theta))$ is implementable if and only if the marginal information rent $\hat{I}(\theta)$ is weakly increasing in θ.

Integrating over seller types, we obtain the total surplus

$$\hat{S} = \int_0^1 \hat{S}(\theta) dF(\theta). \tag{23}$$

The total information rent, after integration by parts using the payoff-equivalence condition, equals

$$\hat{I} = \int_0^1 \frac{1 - F(\theta)}{f(\theta)} \hat{I}(\theta) dF(\theta), \tag{24}$$

where the inverse hazard rate weights reflect the informational externality each type imposes on higher types. Thus, an implementable 2-certificate menu Γ_2 yields the certifier the expected profit $\hat{V}^c = \hat{S} - \hat{I}$.

Lemma 8 *For every non-maximal-screening 2-certificate menu Γ_2, there exists a maximal-screening menu $\hat{\Gamma}$ that yields the certifier, the buyer, and every seller type expected payoffs at least as high as Γ_2. The inequality for the certifier is strict whenever the bunching contract is not the fully revealing structure $(\bar{p}, \bar{q}) = (1, 0)$.*

The intuition rests on the observation that, whenever a 2-certificate menu is non-maximal, some subset $\hat{\Theta} \subseteq \Theta$ is bunched in the sense that there exist $\bar{q} \leq \bar{p}$ such that $q(\theta) = \bar{q}$ and $p(\theta) = \bar{p}$ for all $\theta \in \hat{\Theta}$. We show in the proof that we can fully separate the types in $\hat{\Theta}$ while keeping each seller type's marginal information rent $I(\theta)$ constant, weakly raising the certifier's surplus without affecting the buyer's or any seller type's payoff. The key step in the proof is to construct type-specific posteriors $q(\theta)$ and $p(\theta)$ for all θ so that we obtain a surplus function $\check{S}(\theta)$ that is convex.

It is easiest to see the construction of the alternative separating certification menu when the types in subset $\hat{\Theta}$ are bunched by a certification structure that does not provide any hard information, i.e., when either $\bar{q}$ or $\bar{p}$ equals the average type $\bar{\theta}(\hat{\Theta})$ in $\hat{\Theta}$. Such a certification contract extracts the full surplus $\hat{S} = \hat{v}(\bar{\theta}(\hat{\Theta}))$ without leaving any information rent ($I(\theta) = 0$) for $\theta \in \hat{\Theta}$. The corresponding separating menu fully discloses the types in $\hat{\Theta}$ while still leaving zero information rent to each of them. The surplus function resulting from this separating menu $\check{S}(\theta) = \hat{v}(\theta)$ is indeed convex, implying higher revenue for the certifier than under the original bunching contract because $\mathbb{E}\{\hat{v}(\theta)\} > \hat{v}(\mathbb{E}\{\theta\})$.

When a subset of types is bunched by a contract revealing some market information to the buyer ($\bar{q} < \bar{\theta}(\hat{\Theta}) < \bar{p}$), there exist different separating menus ($q(\theta)$ and $p(\theta)$) resulting in the same marginal information rent $I(\theta)$ as in the original bunching contract. The proof of the lemma identifies one such menu that yields a convex surplus function $\check{S}(\theta)$. Convexity of $\check{S}(\theta)$ then yields strictly higher revenue for the certifier than the original bunching contract, except in the degenerate case where bunching occurs on the fully revealing structure with $\bar{q} = 0$ and $\bar{p} = 1$.

Step 3b: 3-certificate menus. We finally extend the insight of the lemma to 3-certificate contracts. We do so by showing that, as in the case of 2-certificate contracts, we can also construct a convex surplus function $\check{S}(\theta)$ for 3-certificate contracts that pool a subset of seller types. The construction, however, is more involved because the convex function $\check{S}(\theta)$ will typically consist of two parts: an interval of types for which three certificates are used and an interval of types for which the surplus function $\check{S}(\theta)$ uses only two certificates. Combined with Lemma 7, this yields the following.

Proposition 2 *For every non-maximal-screening certification menu* Γ*, there exists a maximal-screening certification menu* $\hat{\Gamma}$ *that yields the certifier, the buyer, and every seller type expected payoffs at least as high as* Γ*.*

The proposition implies Theorem 1 directly. Moreover, it implies that the certifier's optimal certification menu is maximal-screening, thereby completing the proof of Theorem 2.

Theorem 1 shows that maximal-screening is without loss for Pareto efficiency, and Theorem 2 establishes that there exists a certifier-optimal menu that is maximal-screening. Together they yield the paper's main point. The certifier's profit-maximising menu fully separates the seller's private information, so soft information is provided without distortion; yet on a positive-measure subset of types the certification structure is not fully revealing of the underlying quality, so hard information is provided below the welfare-maximising level. The intuition for this contrast is the rent-amplification mechanism discussed in the introduction. We use the remainder of the paper to develop comparative statics and to assess the robustness of the analysis.

6 Comparative Statics of Optimal Certification

Having established Theorems 1 and 2 for general convex value functions $\hat{v}$, we now examine how optimal certification menus vary with key economic parameters. These comparative statics generate additional testable implications for monopolistic certification markets. Specifically, we focus on (i) the seller's type θ, (ii) the value of information captured by the convexity of $\hat{v}$, and (iii) the degree of private information, as captured by changes in $F(\theta)$.

To obtain sharp results, we focus on the tractable class of power value functions

$$\hat{v}(\theta|\alpha) = \theta^{\alpha},$$

where $\alpha > 1$ measures the degree of convexity. Higher values of α imply greater curvature, which increases the value of resolving quality uncertainty. Hence, as α increases, the (social) value of information increases because resolving quality uncertainty becomes more important. Note, however, that α does not affect the setup's maximum aggregate surplus, which obtains when the buyer learns the quality perfectly and therefore equals $\bar{\theta}\hat{v}(1) + (1-\bar{\theta})\hat{v}(0) = \bar{\theta}$, which is independent of α. By contrast, when no information is provided, the buyer's belief remains at the ex ante prior $\bar{\theta}$, implying that the ex ante expected social value is $\hat{v}(\bar{\theta}) = \bar{\theta}^{\alpha}$. Thus, in the class of power value functions, the value of information, quantified as the gap between informed and uninformed outcomes, equals $\bar{\theta} - \bar{\theta}^{\alpha}$, approaching its maximum of $\bar{\theta}$ in the limit $\alpha \to \infty$.

Corollary 1 and Lemma 5 imply that power functions exhibit a tractable characterization of optimal certification. To see this, first note that they satisfy the regularity condition that $\hat{A}(\theta)$ is strictly increasing because

$$\hat{A}(\theta) = \frac{\theta\sqrt{\alpha\theta^{\alpha-1}}}{\alpha - 1} \text{ and } \hat{A}'(\theta) = \frac{(1+\alpha)\sqrt{\alpha\theta^{\alpha-1}}}{2(\alpha-1)} > 0.$$

The optimal certification menu is therefore a 2-certificate menu that induces a two-point posterior with $q(\theta) = 0$ and $p(\theta) \geq \theta$ for each seller type θ.

By applying (12) and (14), it follows that

$$I(p, 0|\theta) = \frac{p-\theta}{1-\theta}p^{\alpha-1} \text{ and } S(p, 0|\theta) = \theta p^{\alpha-1}. \tag{25}$$

The certification menu therefore yields the certifier the payoff

$$\int_0^1 \left[S(p(\theta), 0|\theta) - \frac{1-F(\theta)}{f(\theta)}I(p(\theta), 0|\theta)\right] dF(\theta) = \int_0^1 \left[\theta - \frac{1-F(\theta)}{f(\theta)}\frac{p(\theta)-\theta}{1-\theta}\right] p(\theta)^{\alpha-1} dF(\theta). \tag{26}$$

The optimal posterior $p^*(\theta) \in [\theta, 1]$ maximises (26) subject to the monotonicity condition (MON_2) that $I(p(\theta), 0|\theta)$ is non-decreasing.

Comparative Statics in Type. To assess the comparative statics in seller type θ, we focus on the uniform case $F(\theta) = \theta$ so that the comparative statics are not confounded by variations in the type-specific density, $f(\theta)$. For the uniform distribution, the inverse hazard rate simplifies to $(1-\theta)$ so that the (unconstrained) point-wise first-order condition yields $p(\theta) = 2\theta(\alpha-1)/\alpha$.

Because $p(\theta)$ must satisfy $p(\theta) \in [\theta, 1]$, we obtain, for the case $\alpha \leq 2$, the solution $p^*(\theta) = \theta$

for all θ, while, for the case $\alpha > 2$, we obtain[16]

$$p^*(\theta) = \begin{cases} \frac{2(\alpha-1)}{\alpha}\theta & \text{if } \theta \leq \hat{\theta}(\alpha) \equiv \frac{\alpha}{2(\alpha-1)} \\ 1 & \text{otherwise.} \end{cases} \tag{27}$$

The solution shows that the suboptimality of providing hard information as established for the quadratic example $\alpha = 2$ extends to any $\alpha \in (1, 2]$. By contrast, for the more convex power functions $\alpha > 2$, the optimal maximal-screening menu provides hard information to all types, because $p^*(\theta) > \theta$ for all $\theta \in (0, 1)$.

More specifically, (27) reveals that types exceeding the threshold $\hat{\theta}(\alpha)$ receive perfect quality revelation ($p^* = 1, q^* = 0$), while types below this threshold receive partial hard information with $p^*(\theta) = 2(\alpha - 1)\theta/\alpha > \theta$. These results translate our general theoretical findings into specific, testable predictions about how certification varies with seller characteristics.

Moreover, these results refine Theorem 2's finding that the certifier underprovides hard information. The underprovision specifically targets low types, revealing the following economic logic: providing hard information to low types would increase the information rents that must be paid to high types. To see why, recall that the marginal information rent $I(\theta)$ increases in the informativeness of certificates. When low types receive more informative certificates, they obtain higher rents. But incentive compatibility requires that high types, who benefit even more from informative certificates due to their higher probability of triggering favourable ones, must receive correspondingly larger rents to prevent them from mimicking low types.

This targeting reveals that although the certifier distorts the overall information provision, more hard information is provided where it generates the most value. Hence, the targeting obtains from two reinforcing forces: (i) providing hard information to low types makes screening expensive (as it amplifies the rents to high types), and (ii) providing hard information to low types generates less profit (as low types have lower willingness to pay for it). As a result, low types receive minimal hard information while high types receive perfect quality revelation ($q^* = 0$, $p^* = 1$).

Comparative Statics in the Value of Information. We next investigate the model's comparative statics in α, which measures the convexity of the value function and thereby the social value of quality information.

Equality (27) reveals two effects as α increases: First, the threshold $\hat{\theta}(\alpha)$ decreases, approaching 1/2 as $\alpha \to \infty$. This means more types receive perfect quality revelation. Second, for types below the threshold, $p^*(\theta)$ increases, providing more informative certification. These effects reflect the following economic logic: as quality information becomes more socially valuable (higher α), the efficiency gains from hard information increasingly outweigh the informa-

[16] Noting that for $p^*(\theta) = \theta$ it follows $I(\theta, 0|\theta) = 0$, (MON_2) is trivially satisfied for the case $\alpha \in (1, 2]$. To see that the monotonicity constraint (MON_2) is also satisfied for $\alpha > 2$, note that for $p^*(\theta) = 2\theta(\alpha - 1)/\alpha$ we obtain $I(p^*(\theta), 0|\theta) = 2^{\alpha-1}(\alpha - 2)\alpha^{-\alpha}(\alpha - 1)^{\alpha-1}\theta^\alpha(1 - \theta)^{-1}$ increasing in θ, while for $p^*(\theta) = 1$ it follows $I(1, 0|\theta) = 1$.

Figure 2: Comparative statics with respect to the convexity parameter α of the power function $\hat{v}(\theta|\alpha)$, for ex ante expected aggregate surplus S^*, ex ante expected certifier profit V_C^*, and ex ante expected seller information rent V_S^* under a uniform distribution of seller types. The thin black dashed line represents the certifier's profit when not providing any hard information.

tional rent costs. The certifier responds by expanding hard information provision, both at the extensive margin (for more types information is fully revealing) and at the intensive margin (partial revelation becomes more informative).

We next address how α affects players' payoffs and the aggregate surplus. For this, we first recall that an increase in α has a negative first-order effect when information is not fully provided—an increasing α lowers $\hat{v}(\theta)$ for any non-degenerate belief $\theta \in (0,1)$. For the range $\alpha < 2$, the comparative statics on payoffs and surplus are therefore straightforward. In this case, optimal certification is independent of α and only reveals the seller's private information, without leaving her any information rents, enabling the certifier to extract the entire generated aggregate surplus. That is, for $\alpha < 2$, we have

$$V_C^*(\alpha) = S^*(\alpha) = \int_0^1 \hat{v}(\theta|\alpha)d\theta = \frac{1}{1+\alpha},$$

which is decreasing in α.

Comparative statics are richer when α increases beyond the quadratic case $\alpha = 2$, because optimal certification then starts to reveal hard information. Building on the intuition provided by earlier results, at least some seller types must benefit from increasing α beyond 2 because revealing hard information implies a strictly positive information rent. Using (25) and (27), we can make this intuition more precise, because for $\alpha > 2$ and $\theta < \hat{\theta}(\alpha)$ these equalities imply that

$$I^*(\theta|\alpha) = I(p^*(\theta), 0|\alpha) = \frac{2^{\alpha-1}(\alpha-2)}{(\alpha-1)(1-\theta)}\left(\frac{(\alpha-1)\theta}{\alpha}\right)^{\alpha}.$$

It follows

$$\frac{\partial I^*}{\partial \alpha} = \frac{(2\alpha-2)^{\alpha-1}}{1-\theta}\left(\frac{\theta}{\alpha}\right)^{\alpha}\left[1+(\alpha-2)\ln\left(\frac{2(\alpha-1)\theta}{\alpha}\right)\right].$$

Hence, whether the marginal information rents of a type are increasing or decreasing in α depends on the sign of the term in the square brackets.

In the limit $\alpha \to 2$, the term equals 1, implying that marginal information rents of all seller types are increasing in α at $\alpha = 2$. This confirms our earlier intuition that the seller must gain when α increases beyond 2 because by the payoff-equivalence condition (PE), a seller type's information rent is the integral of the marginal information rents of lower types.

However, because for any $\alpha > 2$, the term is negative for the lower seller types, marginal information rents are strictly decreasing for these types. The payoff-equivalence condition then implies that, for all lower seller types, the information rents themselves are non-monotonic in

α. Indeed, it follows for all seller types $\theta < 1/2$ that their type-specific information rents are zero for $\alpha \leq 2$ and strictly increase over a non-empty interval beyond $\alpha > 2$, after which they strictly decrease, vanishing to zero for $\alpha \to \infty$.[17]

Turning next to an ex ante perspective, we use (25) and (27) to obtain the following expression for ex ante expected aggregate surplus

$$S^*(\alpha) = \int_0^1 S\left(p^*(\theta)|\theta\right) d\theta \;=\; \int_0^{\frac{\alpha}{2(\alpha-1)}} \left(\frac{2\left(\alpha-1\right)}{\alpha}\right)^{\alpha-1} \theta^\alpha d\theta + \int_{\frac{\alpha}{2(\alpha-1)}}^1 \theta d\theta = \frac{3\alpha^2-4}{8\alpha^2-8},$$

which is increasing in α for $\alpha > 2$, converging to $3/8$. Hence, aggregate surplus is decreasing for $\alpha < 2$ and increasing for $\alpha > 2$, implying that aggregate surplus is non-monotonic in α.

Likewise, we obtain for the seller's ex ante expected information rents

$$V_S^*(\alpha) = \int_0^1 \int_0^\theta I(p^*(t)|\alpha) dt d\theta = \frac{\alpha^2-4}{8\alpha^2-8},$$

which is increasing in α, converging to $1/8$. The certifier's ex ante expected profit is therefore

$$V_C^*(\alpha) = S^*(\alpha) - V_S^*(\alpha) = \frac{\alpha^2}{4\alpha^2-4},$$

which is decreasing in α, converging to $1/4$. Hence, the certifier loses when α increases, reflecting that the certifier's response of providing more informative certification when α increases over the range $\alpha > 2$ does not fully mitigate the negative first-order effect of increasing α.

Figure 2 illustrates these comparative statics. The difference between the black thin dashed line and the green solid line, V_C^*, represents the certifier's gain in profit from providing hard information optimally, while the difference between the black dashed line and the red dashed line, S^*, represents the aggregate social value of the certifier providing hard information optimally.

Comparative Statics in Private Information. Finally, we consider the role of the distribution function $F(\theta)$, which captures the degree of private information. The first insight that our analysis provides is that, as in standard models of monopolistic screening, the inverse hazard rate $[1 - F(\theta)]/f(\theta)$ shapes the trade-off between surplus extraction and information rent minimization. Indeed, if the inverse hazard rate term in expression (26) were zero, then the certifier's payoff would coincide with the surplus $S(p, 0|\theta)$, inducing the certifier to provide an undistorted, full revelation of hard information, $p(\theta) = 1$, for all θ. This confirms that it is the presence of private information, its intensity measured by the positive inverse hazard rate, that induces the certifier to distort the degree of hard information provision downward because it requires conceding information rents to elicit soft information.

To make this concrete, note that, for general type distributions $F(\theta)$, a point-wise maxi-

[17] To see that the non-monotonic comparative statics holds for all types $\theta < 1/2$, note that the bracketed term is increasing in θ and zero when θ equals $\hat{\theta}_I(\alpha) \equiv \frac{\alpha}{2(\alpha-1)} e^{\frac{-1}{\alpha-2}}$, which approaches $1/2$ for $\alpha \to \infty$.

mization of (26) for each θ yields, respecting the domain restriction $p(\theta) \in [\theta, 1]$, that

$$p(\theta) = \max\left\{\theta, \min\left\{1, \frac{\alpha - 1}{\alpha}\theta\left[1 + \frac{f(\theta)(1-\theta)}{1 - F(\theta)}\right]\right\}\right\}.$$

Hence, for any bounded density $f(\theta)$, we obtain $p(0) = 0 < 1$, implying that the acquisition of hard information is distorted downward ($p(\theta) < 1$) for θ small enough whenever $f(\theta)$ is also continuous. This shows that downward distortions in hard information are a typical feature of optimal certification when the density $f(\theta)$ is bounded and continuous. However, if the term in the square brackets exceeds the fraction $\alpha/(\alpha - 1)$ for an interior $\theta \in (0, 1)$, we have $p(\theta) > \theta$, implying at least some positive degree of hard information.

7 Robustness

In this section, we discuss the robustness of our results. In particular, we address the extent to which they depend on equilibrium selection and our specific modeling assumptions.

Equilibrium Selection. In our analysis of the certifier's optimal certification menu, we follow standard mechanism design by assuming that the certifier has full commitment power to offer any menu and, when multiple equilibria exist, can select its preferred equilibrium. This approach could be criticized for downplaying equilibrium multiplicity. Indeed, under Perfect Bayesian Equilibrium, off-equilibrium beliefs are unrestricted, potentially creating multiple equilibria. This section demonstrates, however, that our main results are not driven by contrived out-of-equilibrium beliefs.

Because the multiplicity problem arises in different guises, we distinguish four cases: (i) binding incentive constraints, (ii) informed-principal problems, (iii) off-equilibrium belief manipulation, and (iv) signaling-what-you-don't-know paradoxes. We discuss these cases in turn.

First, following standard mechanism design, we work with weak incentive compatibility, leading to binding constraints at the optimum. For example, with power functions and $\alpha \in (1, 2]$, optimal certification provides no hard information so that, as introduced in Benchmark 3, the maximal-screening menu Γ^s is optimal, leaving all seller types indifferent among all menu items.[18] It would however be wrong to interpret this result as implying that the optimality depends on a specific form of equilibrium selection.

In particular, we make precise the claim of Footnote 7 that we can ensure strict incentive compatibility "virtually". To see this, construct the menu $\Gamma^s_\varepsilon \equiv \{(\gamma(\theta))\}_{\theta\in[0,1]}$ for any $\varepsilon > 0$ as follows. Each certification contract $\gamma_\varepsilon(\theta) = (C(\theta), \pi^l_\varepsilon(\theta), \pi^h_\varepsilon(\theta), t_\varepsilon(\theta))$ has $C(\theta) = \{c_\emptyset, c_l, c_h\}$, $\pi^l_\varepsilon = (1 - \varepsilon\theta, \varepsilon\theta, 0)$, and $\pi^h_\varepsilon = (1 - \varepsilon\theta, 0, \varepsilon\theta)$. Hence, the certification contract $\gamma_\varepsilon(\theta)$ consists of

[18] The optimality of no hard information is not specific to power functions with $\alpha \in (1, 2]$, exhibiting $A(p, q) < 0$. In Appendix B we analyze a class of value functions exhibiting $A(p, q) > 0$ and show that, when these value functions are not too convex, optimal certification also provides no hard information.

three certificates, a fully uninformative certificate $c_\emptyset$ that is triggered with probability $1-\varepsilon\theta$ and two fully revealing ones that are triggered in the corresponding state ω with probability $\varepsilon\theta$.

For the certification menu Γ^s_ε, we calculate the marginal information rent $I_\varepsilon(\theta)$ using its definition

$$I_\varepsilon(\theta) \quad = \quad \sum_c \left(\pi^h_c(\theta) - \pi^l_c(\theta)\right)\hat{v}\left(\hat{\theta}_c(\theta)\right) = \varepsilon\theta\left(\hat{v}(1) - \hat{v}(0)\right) = \varepsilon\theta,$$

which is strictly increasing for any $\varepsilon > 0$. Lemma 2 then implies that the transfer schedule of the certification menu Γ^s_ε is such that it leaves seller type θ a rent of

$$V_\varepsilon(\theta) = \int_0^\theta I_\varepsilon(\tau)d\tau = \int_0^\theta \varepsilon\tau d\tau = \varepsilon\theta^2/2,$$

and induces a certification game in which the incentive compatibility constraints hold *strictly.* To the certifier, Γ^s_ε generates the payoff

$$V_\varepsilon = \int_0^1 [S_\varepsilon(\theta) - V_\varepsilon(\theta)]d\theta = \int_0^1 [(1-\varepsilon\theta)\hat{v}(\theta) + \varepsilon\theta[\theta\hat{v}(1) + (1-\theta)\hat{v}(0)] - \varepsilon\theta^2/2]d\theta$$

which converges to V^* as ε tends to zero.

Second, our model features an informed-principal problem: the privately informed seller makes the offer, potentially signaling her type. We assumed passive beliefs so that the buyer does not update from the seller's offer. This rules out the buyer interpreting the offer as a signal, so our results do not rely on such signalling. Alternatively, we could eliminate this issue entirely by using an auction mechanism: with two competing buyers who observe only the certificate (not who selected it), Bertrand competition drives price to $\hat{v}(\hat{\theta})$ based solely on the certificate-updated belief $\hat{\theta}$, eliminating any signaling through the mechanism choice.

Third, and related to the second issue, the fact that perfect Bayesian equilibrium does not restrict out-of-equilibrium beliefs allows it to sustain outcomes that fully block the certifier's ability to screen. This arises when the buyer's out-of-equilibrium belief interprets the seller's pick from the certifier's menu in contradiction to the Bayesian incentive constraints. To see this, suppose that the certifier offers a direct menu Γ^s_ε for some $\varepsilon > 0$ as introduced above. That is, the menu exhibits minimal hard information and satisfies all incentive constraints strictly. Despite the strict incentive constraints, the certification game induced by Γ^s_ε sustains the full-pooling outcome of all seller types selecting the menu item $\bar{\theta}$ as a perfect Bayesian equilibrium, supported by the buyer's out-of-equilibrium belief that any menu item other than $\bar{\theta}$ is chosen by type $\theta = 0$. These out-of-equilibrium beliefs make it optimal for any seller type to pick menu item $\bar{\theta}$, destroying the certifier's screening. Section 5, however, shows that the equilibrium outcomes that block screening are Pareto dominated by screening ones. Hence, applying the Pareto dominance criterion as a refinement, any equilibrium outcome must feature maximal screening, fully revealing the seller's private information. Having established that screening prevails, our analysis further identifies that the certifier optimally chooses a menu that limits hard information revelation.

Finally, we address the "signaling-what-you-don't-know" paradox (Fudenberg and Tirole, 1991). Without restrictions, Perfect Bayesian Equilibrium in the overall game—where the certifier chooses the certification menu first—could sustain any certification menu Γ as an equilibrium if the buyer interprets deviations as signaling $\theta = 0$. Such equilibrium outcomes, however, depend on the paradox that, despite the certifier not knowing θ, the buyer nevertheless interprets the certifier's action as informative and updates his beliefs accordingly. Our analysis avoids this paradox by having buyers enter the certification game with prior $\bar{\theta}$ regardless of the menu offered, preventing the certifier's action from being interpreted as informative about seller types.

Modeling Qualifications. We next address the extent to which our striking analytical result—full soft information but restricted hard information—depends on our specific modeling framework. These qualifications identify three key modeling assumptions that generate the economic forces that underlie our result and separate them from four assumptions that merely help to isolate these key forces cleanly.

First, we assume that the seller's private information is fundamental. The certifier can learn it only via screening, which in general requires information rents. If, instead, the certifier could design a certification signal that correlates with the seller's private information, then the seller's private information becomes verifiable. In this case, we obtain a model in which the certifier can use hard information to reveal both new information and the seller's private information. The economic forces in models with verifiable private information are fundamentally different.

Second, we do not allow certificate-contingent pricing, where the certifier charges based on the realized certificate. This restriction is crucial: with pure contingent pricing, the certifier could reveal quality perfectly and charge only for favourable certificates, extracting full surplus without information rents. Following the literature (e.g., Bergemann et al., 2018), we exclude such pricing, noting that regulation often prohibits certificate-contingent pricing due to manipulation concerns.

Third, our results depend on the value function's convexity, which captures that information has economic value. This convexity drives the trade-off between soft and hard information in any certification context where information improves allocation. However, convexity is not universal—Bayesian persuasion models à la Kamenica and Gentzkow (2011) feature non-convex value functions; full separation would be optimal were these functions convex. This explains why models analyzing pure persuasion effects obtain different results (Chopra and Ely, 2025; Mäkimattila et al., 2025).

Fourth, we model two quality levels to maintain one-dimensional private information. With multiple quality levels, screening becomes multidimensional and substantially more complex.

We do not attempt that extension here.

Fifth, we assume mandatory public disclosure: both the seller's menu choice and the certificate outcome are automatically observed by all parties, with no seller disclosure decision. This assumption serves two purposes. First, it isolates our focus on the certifier's optimal information provision, separating this from strategic disclosure considerations. Second, it reflects institutional features of many certification markets where disclosure is mandatory (e.g., food safety grades, financial audit results, environmental compliance reports) or where withholding certification results is contractually prohibited or legally sanctioned. An alternative modeling approach would allow the seller to choose whether to disclose her menu choice and the certificate outcome. In our setup, however, this alternative yields equivalent equilibrium outcomes. The intuition follows from standard unraveling arguments for verifiable information: because the menu choice and certificate are verifiable and costless to disclose, and buyers interpret non-disclosure as negative information, the seller always discloses both in equilibrium, replicating our mandatory disclosure baseline.

Sixth, we assume type-independent production costs. With type-dependent costs, even pure screening would require information rents, as higher types need compensation for higher outside options. This would reduce the stark contrast between the two channels, but our core mechanism persists: hard information amplifies whatever rents are necessary for screening, creating incentives to restrict it relative to soft information.

Finally, we highlight and qualify the role of the buyer's outside option. The buyer can always secure his outside utility U^{out} by, for instance, purchasing a low-quality good from the competitive fringe. Our baseline sets U^{out} equal to the surplus u_l^* from efficient consumption of low quality, which we normalise to $u_l^* = 0$, so that $U^{out} = 0$. The competitive fringe is a natural microfoundation: in many certification markets a generic or uncertified version of the good is available at a commodity price, and the fringe represents perfectly competitive supply of low quality at zero cost, which fixes that price and hence the buyer's outside option. The fringe enters the analysis only through this outside option. In particular, the revenue function $\hat{v}(\theta)$, the information rents, and the certification game are determined entirely by the buyer's utility functions u_h and u_l, the seller's type distribution F, and the outside option level U^{out}; they do not depend on which economic institution delivers that level. Any alternative microfoundation that fixes U^{out} at the same value — a regulated price floor, a non-strategic outside supplier of low quality, or simply an exogenous reservation utility — leaves the entire analysis unchanged.

The substantive content of the assumption is the inequality $\Delta U \equiv u_l^* - U^{out} \geq 0$; the fringe delivers the knife-edge $\Delta U = 0$, but our results require only $\Delta U \geq 0$. To see this, note that a lower outside option ($\Delta U > 0$) raises the participation rent floor from $V(0) = 0$ to $V(0) = \Delta U$, shifting every type's rent $V(\theta)$ up by ΔU and the type-specific surplus $S(\theta)$ up by the same constant. Since the certifier's profit integrates $S(\theta) - V(\theta)$, the constant cancels and the optimal menu is unchanged. The certifier simply concedes a uniform rent ΔU to every type on top of the information rents, and all qualitative results persist for any $\Delta U \geq 0$. In particular, Theorems 1 and 2 continue to hold. The assumption does however bind for $\Delta U < 0$. When the buyer's outside option exceeds u_l^*, the efficient and equilibrium allocations acquire an exclusion region: at beliefs low enough that the surplus from the seller's good falls below U^{out}, both the planner and the buyer prefer the outside option to any trade. Characterising optimal certification then requires a participation margin alongside the screening and acquisition margins, a different problem that we leave aside.

8 Conclusion

This paper establishes a fundamental result about certification markets: monopolistic certifiers have strong incentives to restrict the provision of hard but not soft information. By developing a model that distinguishes between certification's screening role (extracting soft information) and its acquisition role (generating hard information), we uncover a contrast in how the two channels create value for the certifier, a finding that challenges conventional wisdom. Although hard information is technologically costless to produce, the certifier deliberately limits its provision to minimize information rents, while screening achieves maximal separation.

The economic mechanism driving this result reveals a subtle complementarity between information channels. Information rents arise only when hard information is provided, and these rents increase disproportionately for high-quality types when low-quality types receive informative certificates. This forces the certifier to balance the value of information against the cost of screening, leading to the systematic underprovision of hard information, particularly for low-quality sellers. For natural model specifications, the underprovision effect is so strong that the certifier does not provide any hard information at all. Meanwhile, screening creates free value through price discrimination that the certifier can fully capture.

The welfare implications directly inform regulatory design. Current regulatory approaches often focus on conflicts of interest and fee structures, implicitly targeting the screening dimension of certification. One such concern is rating inflation. Our results suggest this emphasis may be misplaced. Instead, our analysis rationalizes rating inflation as resulting from a non-maximal provision of hard information, deliberately garbling high-state information with low-state information—a symptom of inadequate information production, not strategic manipulation. Mandating minimum standards for certification tests, measurements, and investigations could address the underprovision of hard information directly. Such mandates may, however, discourage low-type sellers from seeking certification, as higher information provision requires larger information rents to maintain incentive compatibility. This creates a non-trivial trade-off between increasing hard information and potentially reducing soft information through lower participation. While this trade-off represents an important direction for policy design, a proper treatment requires a separate analysis, which we leave to future research.

This paper contributes to our broader understanding of information markets by demonstrating that the method of information production matters as much as the information itself.

The distinction between revealing existing information and generating new information fundamentally shapes market outcomes. As certification intermediaries become more important in market design and regulation, understanding these distinctions becomes essential for both positive analysis of market behavior and normative design of regulatory interventions.

Appendix A: Proofs

This appendix collects the proofs of the propositions and lemmas in the main text.

Proof of Lemma 1: We first prove that $\hat{x}(\theta)$ is strictly increasing. To see this, note that $\hat{x}$ is defined by the first-order condition

$$\theta u_h'(\hat{x}) + (1-\theta)u_l'(\hat{x}) = 0.$$

Applying the implicit function theorem yields

$$u_h'(\hat{x}) + \theta u_h''(\hat{x})\frac{\partial \hat{x}}{\partial \theta} - u_l'(\hat{x}) + (1-\theta)u_l''(\hat{x})\frac{\partial \hat{x}}{\partial \theta} = 0 \Rightarrow \frac{\partial \hat{x}}{\partial \theta} = -\frac{u_h'(\hat{x}) - u_l'(\hat{x})}{\theta u_h''(\hat{x}) + (1-\theta)u_l''(\hat{x})} > 0,$$

where the strict inequality follows from the single-crossing condition $u_h'(\hat{x}) > u_l'(\hat{x})$, and the strict concavity of u.

To see that $\hat{v}$ is strictly increasing, note that by the envelope theorem:

$$\hat{v}'(\theta) = \frac{\partial \hat{v}}{\partial \theta} + \frac{\partial \hat{v}}{\partial x}\frac{\partial \hat{x}}{\partial \theta} = \frac{\partial \hat{v}}{\partial \theta} = u_h(\hat{x}(\theta)) - u_l(\hat{x}(\theta)) > 0,$$

where the strict inequality follows from single-crossing.

To see that $\hat{v}$ is strictly convex, consider a pair of beliefs $\theta_1 < \theta_2$, and a convex combination $\tilde{\theta} = \pi\theta_1 + (1-\pi)\theta_2$ with $\pi \in (0,1)$ so that $\theta_1 < \tilde{\theta} < \theta_2$. Because, as established above, $\hat{x}(\theta)$ is strictly increasing, we have $\hat{x}(\theta_1) < \hat{x}(\tilde{\theta}) < \hat{x}(\theta_2)$. Hence, it follows

$$\begin{aligned}
\hat{v}\,(\theta_1) = \theta_1 u_h(\hat{x}(\theta_1)) + (1-\theta_1)u_l(\hat{x}(\theta_1)) &> \theta_1 u_h(\hat{x}(\tilde{\theta})) + (1-\theta_1)u_l(\hat{x}(\tilde{\theta}));\\
\hat{v}\,(\theta_2) = \theta_2 u_h(\hat{x}(\theta_2)) + (1-\theta_2)u_l(\hat{x}(\theta_2)) &> \theta_2 u_h(\hat{x}(\tilde{\theta})) + (1-\theta_2)u_l(\hat{x}(\tilde{\theta})).
\end{aligned}$$

Taking the average of these two strict inequalities with the weights π and $(1-\pi)$ yields

$$\pi\hat{v}(\theta_1) + (1-\pi)\hat{v}(\theta_2) > \tilde{\theta}u_h(\tilde{x}) + (1-\tilde{\theta})u_l(\tilde{x}) = \hat{v}(\tilde{\theta}).$$

This shows the strict convexity of $\hat{v}$. □

Proof of Lemma 2: To see that incentive compatibility implies (MON), consider (IC_{θ,θ^r}) and $(IC_{\theta^r,\theta})$:

$$V(\theta) \geq V(\theta^r) + (\theta - \theta^r)I(\theta^r);$$

$$V(\theta^r) \geq V(\theta) + (\theta^r - \theta)I(\theta).$$

Together they imply

$$(\theta - \theta^r)[I(\theta) - I(\theta^r)] \geq 0.$$

Hence, incentive compatibility implies (MON).

To see that incentive compatibility implies (PE), rewrite, for $\delta > 0$, the incentive constraint

$(IC_{\theta+\delta,\theta})$ as

$$\frac{V(\theta+\delta)-V(\theta)}{\delta} \geq I(\theta)$$

Likewise, for $\delta > 0$, rewrite $(IC_{\theta-\delta,\theta})$ as

$$\frac{V(\theta)-V(\theta-\delta)}{\delta} \leq I(\theta)$$

At a point of differentiability of $V(\theta)$, we have

$$I(\theta) \leq \lim_{\delta\downarrow 0} \frac{V(\theta+\delta)-V(\theta)}{\delta} = V'(\theta) = \lim_{\delta\downarrow 0} \frac{V(\theta)-V(\theta-\delta)}{\delta} \leq I(\theta),$$

implying $V'(\theta) = I(\theta)$ so that (PE) then follows from the fundamental theorem of calculus.

We now show sufficiency of (MON) and (PE) for incentive compatibility. When $\theta > \theta^r$, applying (PE), condition (IC_{θ,θ^r}) can be rewritten as

$$V(\theta) - V(\theta^r) = \int_{\theta^r}^{\theta} I(\tau)d\tau \geq (\theta - \theta^r)I(\theta^r), \tag{28}$$

which follows from (MON), since $I(\tau) \geq I(\theta^r)$ for all $\tau \geq \theta^r$. Similarly, when $\theta < \theta^r$, condition (IC_{θ,θ^r}) reduces to

$$V(\theta) - V(\theta^r) = -\int_{\theta}^{\theta^r} I(\tau)d\tau \geq -(\theta^r - \theta)I(\theta^r), \tag{29}$$

which also holds, since $I(\tau) \leq I(\theta^r)$ for all $\tau \leq \theta^r$.

To see the second statement, note that by (IC_{θ,θ^r}) the result follows directly if for any θ we have

$$I(\theta) = \sum_c \left[\pi_c^h(\theta) - \pi_c^l(\theta)\right] \hat{v}(\hat{\theta}_c(\theta)) \geq 0 \tag{30}$$

To see that this inequality holds, fix θ and, without loss, order the certificates in $C(\theta)$ such that the likelihood ratio $\pi_c^h(\theta)/\pi_c^l(\theta)$ is increasing in c. Hence, $\pi^h(\theta)$ dominates $\pi^l(\theta)$ in the sense of FOSD. Note that because the likelihood ratio is increasing, also $\hat{v}(\theta_c(\theta))$ is increasing in c. As FOSD implies that the expectation of any increasing function is larger under $\pi^h(\theta)$ than under the FOSD-ed $\pi^l(\theta)$, it holds in particular for the increasing sequence $\hat{v}(\theta_c(\theta))$. Hence, $\sum_c \pi_c^h(\theta)\hat{v}(\hat{\theta}_c(\theta)) \geq \sum_c \pi_c^l(\theta)\hat{v}(\theta_c(\theta))$, which implies (30). The third statement follows because condition (IR_0), $V(0) \geq 0$, is sufficient for (IR_θ) for all θ as $V(\theta)$ is increasing. The final statement holds because inequalities in (28) and (29) are strict when $I(\theta)$ is strictly increasing. $\square$

Proof of Proposition 1: Fix some solution $\hat{\Gamma} = \{(\hat{C}(\theta), \hat{\pi}^l(\theta), \hat{\pi}^h(\theta), \hat{t}(\theta)\}_{\theta\in[0,1]}$ to the problem $\hat{\mathcal{P}}^S$ with associated value $\hat{\Pi}^c(\hat{\Gamma})$. Let $\Theta^+ \subseteq [0,1]$ denote the subset of types θ for which the solution $\hat{\Gamma}$ uses more than three certificates. We next argue that if $\Theta^+ \neq \emptyset$, then we can construct an alternative solution $\check{\Gamma}$ that has at most three certificates for each θ that is also a solution to $\hat{\mathcal{P}}^S$ with associated value $\hat{\Pi}^c(\hat{\Gamma})$. To see this, let

$$\hat{S}(\theta) \equiv \sum_c \left[\theta \hat{\pi}_c^h(\theta) + (1-\theta)\hat{\pi}_c^l(\theta)\right]\hat{v}(\hat{\theta}_c(\theta))$$

represent the surplus that a type θ generates under solution $\hat{\Gamma}$. Likewise, let

$$\hat{I}(\theta) = \sum_c [\hat{\pi}_c^h(\theta) - \hat{\pi}_c^l(\theta)]\hat{v}(\hat{\theta}_c(\theta))$$

express the marginal information rent of seller type θ under solution $\hat{\Gamma}$. We have

$$\hat{\Pi}^c(\hat{\Gamma}) = \int_0^1 \hat{\Pi}(\theta) dF(\theta) \text{ where } \hat{\Pi}(\theta) \equiv \hat{S}(\theta) - \frac{1-F(\theta)}{f(\theta)}\hat{I}(\theta).$$

Next for any $\theta \in \Theta^+$, consider the linear problem indexed by θ of finding a solution $\alpha(\theta) = (\alpha_1(\theta), \ldots, \alpha_N(\theta))$ to:

$$\hat{\mathcal{P}}^\alpha(\theta): \max_{(\alpha_1,\ldots,\alpha_N)\geq 0} \sum_c \alpha_c \left\{\theta\hat{\pi}_c^h(\theta) + (1-\theta)\hat{\pi}_c^l(\theta) - \frac{1-F(\theta)}{f(\theta)}\left[\hat{\pi}_c^h(\theta) - \hat{\pi}_c^l(\theta)\right]\right\}\hat{v}(\hat{\theta}_c(\theta))$$
$$\text{s.t. } \sum_c \alpha_c \hat{\pi}_c^l(\theta) = 1; \sum_c \alpha_c \hat{\pi}_c^h(\theta) = 1; \sum_c \alpha_c[\hat{\pi}_c^h(\theta) - \hat{\pi}_c^l(\theta)]\hat{v}(\hat{\theta}_c(\theta)) = \hat{I}(\theta).$$

As $\alpha_1(\theta) = \ldots = \alpha_N(\theta) = 1$ is feasible in program $\hat{\mathcal{P}}^\alpha(\theta)$, the program's value, $\hat{\Pi}^\alpha(\theta)$, is at least $\hat{\Pi}(\theta)$. Moreover, note that $\hat{\mathcal{P}}^\alpha(\theta)$ is a linear program with only 3 constraints. Consequently, it has a solution $\hat{\alpha}(\theta)$ with at most 3 entries being non-zero. For each $\theta \in \Theta^+$, let these three entries be indexed by the triple $(k(\theta), l(\theta), m(\theta))$. Given the solution $\hat{\alpha}(\theta)$, we construct a 3-certificate menu $\check{\Gamma}$ as follows. For each $\theta \in \Theta^+$ and $\omega \in \{l, h\}$, let $\check{C}(\theta) \equiv \{c_{k(\theta)}, c_{l(\theta)}, c_{m(\theta)}\}$, $\check{\pi}^\omega_{k(\theta)}(\theta) \equiv \hat{\alpha}_{k(\theta)}(\theta)\pi^\omega_{k(\theta)}(\theta)$, $\check{\pi}^\omega_{l(\theta)}(\theta) \equiv \hat{\alpha}_{l(\theta)}(\theta)\pi^\omega_{l(\theta)}(\theta)$, $\check{\pi}^\omega_{m(\theta)}(\theta) \equiv \hat{\alpha}_{m(\theta)}(\theta)\pi^\omega_{m(\theta)}(\theta)$, and

$$\check{t}(\theta) \equiv \check{S}(\theta) - \left[V(0) + \int_0^\theta I(\tau)d\tau\right].$$

Now construct $\check{\Gamma} = \{\check{\Gamma}(\theta)\}_{\theta\in[0,1]}$ by setting $\check{\Gamma}(\theta) = (\check{C}(\theta), \check{\pi}^l(\theta), \check{\pi}^h(\theta), \check{t}(\theta))$ for $\theta \in \Theta^+$ and $\check{\Gamma}(\theta) = \hat{\Gamma}(\theta)$ for $\theta \notin \Theta^+$. By construction, the 3-certificate menu $\check{\Gamma}$ satisfies monotonicity, because the marginal information rents are unchanged. It is therefore incentive compatible and yields $\hat{\Pi}^c(\check{\Gamma}) = \int_{\theta\in\Theta^+} \hat{\Pi}^\alpha(\theta)dF + \int_{\theta\notin\Theta^+} \hat{\Pi}(\theta)dF \geq \int_0^1 \hat{\Pi}(\theta)dF = \hat{\Pi}^c(\hat{\Gamma})$. But since $\hat{\Gamma}$ itself is optimal, we must also have $\hat{\Pi}^c(\hat{\Gamma}) \geq \hat{\Pi}^c(\check{\Gamma})$, implying that also the 3-certificate menu $\check{\Gamma}$ must be optimal. This proves the claim that 3-certificate menus are optimal. □

Proof of Lemma 3: Let $(\hat{p}(\theta), \hat{q}(\theta))$ with value $\hat{V}_2$ (as defined by (13)) be such that it satisfies (MON_2) and solves $\mathcal{R}^S_\theta$ for each $\theta \in \Theta$. Let $\hat{\Gamma}_2 = \{(\{c_0, c_1\}, \hat{\pi}^l(\theta), \hat{\pi}^h(\theta), \hat{t}(\theta)\}_{\theta\in[0,1]}$ represent the corresponding 2-certificate menu in terms of the probabilities $(\hat{\pi}^l, \hat{\pi}^h)$, satisfying

(11) with $p = \hat{p}(\theta)$ and $q = \hat{q}(\theta)$, and the transfer $\hat{t}(\theta)$, yielding seller type θ the information rent $\hat{I}(\theta)$ induced by $(\hat{p}(\theta), \hat{q}(\theta))$. Recalling (7), let $\Pi^c(\hat{\Gamma}_2)$ represent the certifier's profit of the 2-certificate menu $\hat{\Gamma}$. As it is a solution to $\mathcal{R}^S_\theta$, it corresponds to the value, $V^R_2(\theta)$, of program $\mathcal{R}^S_\theta$, i.e., $\hat{V}_2 = \Pi^c(\hat{\Gamma}_2) = \int_0^1 V^R_2(\theta)dF$.

Next note that $\hat{\Gamma}_2$ is feasible in program $\hat{\mathcal{P}}^S$. Hence, the value $\hat{V}^S$ of program $\hat{\mathcal{P}}^S$ is at least $\hat{V}_2$, i.e. $\hat{V}^S \geq \hat{V}_2$.

Next consider a solution of a relaxed version of $\hat{\mathcal{P}}^S$ that disregards the constraint (MON). Let $\hat{V}^R$ denote the value of this program and note that $\hat{V}^R \geq \hat{V}^S$.

Let $\hat{\Gamma} = \{(\hat{C}(\theta), \hat{\pi}^l(\theta), \hat{\pi}^h(\theta), \hat{t}(\theta)\}_{\theta\in[0,1]}$ denote a solution to the relaxed program. Note that for each θ, the collection $\{(\hat{\pi}^l_c(\theta), \hat{\pi}^h_c(\theta))\}_{c\in C(\theta)}$ solves

$$\hat{V}(\theta) = \max_{\{\pi^l_c, \pi^h_c\}_{c\in C(\theta)}} \sum_c \left[\left\{ \theta\pi^h_c + (1-\theta)\pi^l_c - \frac{1-F(\theta)}{f(\theta)} \left[\pi^h_c - \pi^l_c\right] \right\} \hat{v}(\theta_c(\theta)) \right]. \tag{31}$$

Hence, $V^R = \int_0^1 \hat{V}(\theta)dF$

Next, consider the linear program of finding a solution $\alpha(\theta) = (\alpha_1, \ldots, \alpha_N)$ to:

$$\begin{aligned} \mathcal{P}^\alpha(\theta): \max_{(\alpha_1,\ldots,\alpha_N)\geq 0} & \sum_c \left[\alpha_c \left\{ \theta\hat{\pi}^h_c(\theta) + (1-\theta)\hat{\pi}^l_c(\theta) - \frac{1-F(\theta)}{f(\theta)} \left[\hat{\pi}^h_c(\theta) - \hat{\pi}^l_c(\theta)\right] \right\} \hat{v}(\theta_c(\theta)) \right] \\ \text{s.t. } & \sum_c \alpha_c \pi^l_c(\theta) = 1; \sum_c \alpha_c \pi^h_c(\theta) = 1. \end{aligned}$$

As $\alpha_1 = \ldots = \alpha_N = 1$ is feasible in program $\mathcal{P}^\alpha(\theta)$, its value, $\hat{V}^\alpha(\theta)$, is at least $\hat{V}(\theta)$, i.e. $\hat{V}^\alpha(\theta) \geq \hat{V}(\theta)$. Moreover, note that this is a linear optimization problem under two linear constraints. Hence, a solution $\hat{\alpha}(\theta) = (\hat{\alpha}_1(\theta), \ldots, \hat{\alpha}_N(\theta))$ exists with at most two entries in $\hat{\alpha}(\theta)$ as non-zero, say $k(\theta)$ and $l(\theta)$.

Given solutions $\hat{\alpha}(\theta)$, define the 2-certificate menu $\check{\Gamma} = \{(\check{C}(\theta), \check{\pi}^l(\theta), \check{\pi}^h(\theta), \check{t}(\theta)\}_{\theta\in[0,1]}$ by $\check{C}(\theta) \equiv \{c_{k(\theta)}, c_{l(\theta)}\}$, $\check{\pi}^\omega_{c_{k(\theta)}}(\theta) \equiv \hat{\alpha}_{k(\theta)}(\theta)\hat{\pi}^\omega_{c_{k(\theta)}}(\theta)$, $\check{\pi}^\omega_{c_{l(\theta)}}(\theta) \equiv \hat{\alpha}_{l(\theta)}(\theta)\hat{\pi}^\omega_{c_{l(\theta)}}(\theta)$, $\check{t}(\theta) \equiv \check{S}(\theta) - \check{V}(\theta)$, where

$$\check{S}(\theta) \equiv \sum_{j\in\{k(\theta),l(\theta)\}} \left[\theta\check{\pi}^h_{c_j}(\theta) + (1-\theta)\check{\pi}^l_{c_j}(\theta)\right] \hat{v}\big(\check{\theta}_{c_j}(\theta)\big),$$

$$\check{I}(\theta) \equiv \sum_{j\in\{k(\theta),l(\theta)\}} \left[\check{\pi}^h_{c_j}(\theta) - \check{\pi}^l_{c_j}(\theta)\right] \hat{v}\big(\check{\theta}_{c_j}(\theta)\big) \text{ and } \check{V}(\theta) \equiv \int_0^\theta \check{I}(\tau)\, d\tau, .$$

By construction, this 2-certificate contract has value $V^{\hat{\alpha}} \equiv \int_0^1 \hat{V}^\alpha(\theta)dF \geq \int_0^1 \hat{V}(\theta)dF = \hat{V}^R$

But given that the menu $\check{\Gamma}$ is a 2-certificate menu, it cannot generate a higher value than the optimal 2-certificate menu $\hat{\Gamma}_2$, i.e. $\hat{V}_2 \geq V^{\hat{\alpha}}$. We therefore obtain the string of weak inequalities $\hat{V}_2 \geq V^{\hat{\alpha}} \geq \hat{V}^R \geq \hat{V}^S \geq \hat{V}_2$, so that, in fact, they must hold with equality so that $\hat{V}^S = \hat{V}_2$. □

Proof of Lemma 4: Following Proposition 1, we prove the lemma by considering an optimal maximal-screening menu $\hat{\Gamma}^S = \{\hat{\sigma}(\theta), \hat{t}(\theta)\}_{\theta\in\Theta}$ that has at most 3 certificates. That is, for any $\theta \in \Theta$, the certification structure $\hat{\sigma}(\theta) = (\hat{C}(\theta), \hat{\pi}^l(\theta), \hat{\pi}^h(\theta))$ induces at most three distinct posteriors. If $\hat{\sigma}(\theta)$ exhibits only one certificate, then $\hat{\sigma}(\theta)$ does not induce two distinct posteriors and the lemma holds trivially. Thus, to prove the lemma we distinguish between the case that $\hat{\sigma}(\theta)$ induces, with a strictly positive probability, exactly 2 distinct posteriors, and the case in which $\hat{\sigma}(\theta)$ induces, with a strictly positive probability, exactly 3 distinct posteriors. We prove that for either case, there is an optimal Γ^S satisfying (17).

Before doing so, we define

$$\tilde{A}(p,q) \equiv \hat{v}'(p)\hat{v}'(q) - w(p,q)^2$$

so that the sign of $A(p,q)$ coincides with the sign of $\tilde{A}(p,q)$.

Case 1: $\hat{\sigma}(\theta) = (\{\hat{c}_p, \hat{c}_q\}, (1-\hat{\pi}^l_p(\theta), \hat{\pi}^l_p(\theta)), (1-\hat{\pi}^h_p(\theta), \hat{\pi}^h_p(\theta)))$ *with* $\hat{\pi}^h_p(\theta) > 0$, $\hat{\pi}^l_p(\theta) < 1$, *and* $\hat{\pi}^h_p(\theta) > \hat{\pi}^l_p(\theta)$.[19] By Bayes' consistency and (4), it holds $q < \theta < p$ and

$$q = \frac{\theta(1-\hat{\pi}^h_p)}{\theta(1-\hat{\pi}^h_p) + (1-\theta)(1-\hat{\pi}^l_p)}; \quad p = \frac{\theta\hat{\pi}^h_p}{\theta\hat{\pi}^h_p + (1-\theta)\hat{\pi}^l_p}, \tag{32}$$

where we dropped the arguments θ in $\hat{\pi}^h_p(\theta)$ and $\hat{\pi}^l_p(\theta)$.

Solving for $\hat{\pi}^h_p$ and $\hat{\pi}^l_p$ in (32) yields

$$\hat{\pi}^l_p(p,q) = \frac{(1-p)(\theta-q)}{(p-q)(1-\theta)}; \; \hat{\pi}^h_p(p,q) = \frac{p(\theta-q)}{(p-q)\theta}. \tag{33}$$

Using (5), it then follows that with 2 certificates, type θ's marginal information rent $I(p,q|\theta)$ equals

$$I(p,q|\theta) = [\pi^h_p(p,q) - \pi^l_p(p,q)][\hat{v}(p) - \hat{v}(q)] = \frac{(p-\theta)(\theta-q)}{(1-\theta)\theta} w(p,q). \tag{34}$$

Hence, by the implicit function theorem, the information rent does not change if we change p and q such that

$$\frac{\partial p}{\partial q} = -\frac{\partial I/\partial q}{\partial I/\partial p} = \frac{(p-\theta)[(p-\theta)w(p,q) + (\theta-q)\hat{v}'(q)]}{(\theta-q)[(p-\theta)\hat{v}'(p) + (\theta-q)w(p,q)]} \geq 0. \tag{35}$$

Hence, for $q > 0$, we can decrease q and p marginally without affecting type θ's information rent. Similarly, we can increase q and p marginally without affecting type θ's information rent if $p < 1$.

To assess the effect of such changes of q and p on the certifier's profits, it suffices to determine how it affects the surplus, because the certifier's profit is the difference between surplus and information rents, and the increase in q and p is such that the latter is kept constant.

Using (33), we can express the surplus (6) with two certificates in two equivalent formula-

[19] The conditions $\hat{\pi}^h_p(\theta) > 0$ and $\hat{\pi}^l_p(\theta) < 1$ ensure that the two certificates $\hat{c}_p$ and $\hat{c}_q$ obtain with a strictly positive probability, while $\hat{\pi}^h_p(\theta) > \hat{\pi}^l_p(\theta)$ ensures $p > q$, as stated in the lemma.

tions

$$\begin{aligned} S(p,q|\theta) &= \hat{v}(q) + [\theta(1-q) - (1-\theta)q]w(p,q) \\ &= \hat{v}(p) + [\theta(1-p) - (1-\theta)p]w(p,q). \end{aligned}$$

Consequently,

$$\frac{\partial S}{\partial p} = [\theta(1-q) - (1-\theta)q]\frac{\partial w}{\partial p} \text{ and } \frac{\partial S}{\partial q} = [\theta(1-p) - (1-\theta)p]\frac{\partial w}{\partial q}.$$

It therefore follows that an increase in p and q that leaves type θ's information rent constant changes the surplus by

$$\frac{dS}{dq} = \frac{\partial S}{\partial p}\frac{\partial p}{\partial q} + \frac{\partial S}{\partial q} = \frac{p-\theta}{(p-\theta)\hat{v}'(p) + (\theta-q)w(p,q)}\tilde{A}(p,q),$$

where the fraction is strictly positive. Hence, the sign of $\tilde{A}(p,q)$ determines the effect on the certifier's profits. As the sign of $\tilde{A}(p,q)$ coincides with the sign of $A(p,q)$, we obtain (17). Because, if $\tilde{A}(p,q) > 0$ and $p < 1$, we can increase the certifier's profits by increasing q and p. And, if, by contrast, $\tilde{A}(p,q) < 0$ and $q > 0$, we can increase the certifier's profits by decreasing q and p.

Case 2: $\hat{\sigma}(\theta) = (\{\hat{c}_1, \hat{c}_2, \hat{c}_3\}, (\hat{\pi}_1^l(\theta), \hat{\pi}_2^l(\theta), \hat{\pi}_3^l(\theta)), (\hat{\pi}_1^h(\theta), \hat{\pi}_2^h(\theta), \hat{\pi}_3^h(\theta)))$ *with* $\hat{\pi}_1^h(\theta) + \hat{\pi}_1^l(\theta) > 0$, $\hat{\pi}_2^h(\theta) + \hat{\pi}_2^l(\theta) > 0$, $\hat{\pi}_3^h(\theta) + \hat{\pi}_3^l(\theta) > 0$.[20] Dropping the argument θ, the certification structure $\hat{\sigma}$ induces posteriors

$$\hat{p}_1 = \frac{\theta\hat{\pi}_1^h}{\theta\hat{\pi}_1^h + (1-\theta)\hat{\pi}_1^l}; \hat{p}_2 = \frac{\theta\hat{\pi}_2^h}{\theta\hat{\pi}_2^h + (1-\theta)\hat{\pi}_2^l}; \hat{p}_3 = \frac{\theta\hat{\pi}_3^h}{\theta\hat{\pi}_3^h + (1-\theta)\hat{\pi}_3^l},$$

where we label the posteriors such that $\hat{p}_1 > \hat{p}_2 > \hat{p}_3$ and, hence, Bayes' consistency implies $\hat{p}_1 > \theta > \hat{p}_3$. These posteriors yield seller type θ the marginal information rent

$$\hat{I} \equiv [\hat{\pi}_1^h - \hat{\pi}_1^l]\hat{v}(\hat{p}_1) + [\hat{\pi}_2^h - \hat{\pi}_2^l]\hat{v}(\hat{p}_2) + [\hat{\pi}_3^h - \hat{\pi}_3^l]\hat{v}(\hat{p}_3).$$

We show that, if $\hat{\sigma}(\theta)$ is part of an optimal menu, then there is no loss in assuming that the posteriors $(\hat{p}_1, \hat{p}_2, \hat{p}_3)$ satisfy the conditions stated in the lemma (where p and q in the lemma correspond to a pair of posteriors, i.e. $(p,q) \in \{(\hat{p}_1, \hat{p}_2), (\hat{p}_1, \hat{p}_3), (\hat{p}_2, \hat{p}_3)\}$).

To see this, first note that if $\hat{\sigma}$ is optimal, then it necessarily maximises the certifier's surplus

$$S = [\theta\pi_1^h + (1-\theta)\pi_1^l]\hat{v}(p_1) + [\theta\pi_2^h + (1-\theta)\pi_2^l]\hat{v}(p_2) + [\theta\pi_3^h + (1-\theta)\pi_3^l]\hat{v}(p_3) \qquad (36)$$

subject to the following six constraints:

[20] The conditions $\hat{\pi}_1^h(\theta) + \hat{\pi}_1^l(\theta) > 0$, $\hat{\pi}_2^h(\theta) + \hat{\pi}_2^l(\theta) > 0$, and $\hat{\pi}_3^h(\theta) + \hat{\pi}_3^l(\theta) > 0$ ensure that each of the three certificates $\hat{c}_1, \hat{c}_2, \hat{c}_3$ obtains with strictly positive probability.

1. The information rent being constant at $\hat{I}$,

$$[\pi_1^h - \pi_1^l]\hat{v}(p_1) + [\pi_2^h - \pi_2^l]\hat{v}(p_2) + [\pi_3^h - \pi_3^l]\hat{v}(p_3) = \hat{I}; \tag{37}$$

2. the feasibility conditions

$$\pi_1^l + \pi_2^l + \pi_3^l = 1; \tag{38}$$
$$\pi_1^h + \pi_2^h + \pi_3^h = 1; \tag{39}$$

3. the three posterior conditions

$$[\theta\pi_1^h + (1-\theta)\pi_1^l]p_1 = \theta\pi_1^h; \tag{40}$$
$$[\theta\pi_2^h + (1-\theta)\pi_2^l]p_2 = \theta\pi_2^h; \tag{41}$$
$$[\theta\pi_3^h + (1-\theta)\pi_3^l]p_3 = \theta\pi_3^h. \tag{42}$$

The six constraints (37)-(42) form a linear system of the six variables $(\pi_1^h, \pi_1^l, \pi_2^h, \pi_2^l, \pi_3^h, \pi_3^l)$. For a combination (θ, p_1, p_2, p_3) with $1 \geq p_1 > p_2 > p_3 \geq 0$ and $p_1 > \theta > p_3$, define

$$D(p_1, p_2, p_3, \theta) \equiv (p_2 - p_3)(p_1 - \theta)(\hat{v}(p_1) - \hat{v}(p_3)) + (p_1 - p_3)(\theta - p_2)(\hat{v}(p_2) - \hat{v}(p_3))$$

so that the solution of this linear system with respect to (p_1, p_2, p_3, θ) exhibits

$$\begin{aligned}
\pi_1^h &= \frac{p_1}{\theta}\pi_1;\ \pi_1^l = \frac{1-p_1}{1-\theta}\pi_1 \text{ with } \pi_1 \equiv \frac{(\theta - p_2)(\theta - p_3)[\hat{v}(p_2) - \hat{v}(p_3)] + (1-\theta)\theta(p_2 - p_3)\hat{I}}{D(p_1, p_2, p_3, \theta)};\\
\pi_2^h &= \frac{p_2}{\theta}\pi_2;\ \pi_2^l = \frac{1-p_2}{1-\theta}\pi_2 \text{ with } \pi_2 \equiv \frac{(p_1 - \theta)(\theta - p_3)[\hat{v}(p_1) - \hat{v}(p_3)] - (1-\theta)\theta(p_1 - p_3)\hat{I}}{D(p_1, p_2, p_3, \theta)};\\
\pi_3^h &= \frac{p_3}{\theta}\pi_3;\ \pi_3^l = \frac{1-p_3}{1-\theta}\pi_3 \text{ with } \pi_3 \equiv \frac{(p_1 - \theta)(p_2 - \theta)[\hat{v}(p_1) - \hat{v}(p_2)] + (1-\theta)\theta(p_1 - p_2)\hat{I}}{D(p_1, p_2, p_3, \theta)}.
\end{aligned} \tag{43}$$

Substituting these solutions for $(\pi_1^h, \pi_1^l, \pi_2^h, \pi_2^l, \pi_3^h, \pi_3^l)$ into (36), we obtain after a rearrangement of terms that $(\hat{p}_1, \hat{p}_2, \hat{p}_3)$ maximises the expression

$$S(p_1,p_2,p_3,\theta)=\frac{1}{D(p_1,p_2,p_3,\theta)}\times\Bigg[[\theta-p_2][\theta-p_3][\hat{v}(p_2)-\hat{v}(p_3)]\hat{v}(p_1) \\ +[p_1-\theta][\theta-p_3][\hat{v}(p_1)-\hat{v}(p_3)]\hat{v}(p_2) \\ +[p_1-\theta][p_2-\theta][\hat{v}(p_1)-\hat{v}(p_2)]\hat{v}(p_3) \\ +\{[p_2-p_3]\hat{v}(p_1)+[p_3-p_1]\hat{v}(p_2)+[p_1-p_2]\hat{v}(p_3)\}\theta(1-\theta)\hat{I}\Bigg]. \tag{44}$$

Taking derivatives and using $w(.,.)$ as defined in (15), we can rewrite them as follows

$$\begin{aligned}
\frac{\partial S(p_1,p_2,p_3,\theta)}{\partial p_1} &= \frac{(p_1-p_3)\pi_1}{(p_1-\theta)w(p_1,p_2)+(\theta-p_3)w(p_2,p_3)}[w(p_2,p_3)\hat{v}'(p_1)-w(p_1,p_3)w(p_1,p_2)];\\
\frac{\partial S(p_1,p_2,p_3,\theta)}{\partial p_2} &= \frac{(p_1-p_3)\pi_2}{(p_1-\theta)w(p_1,p_2)+(\theta-p_3)w(p_2,p_3)}[w(p_1,p_3)\hat{v}'(p_2)-w(p_2,p_3)w(p_1,p_2)];\\
\frac{\partial S(p_1,p_2,p_3,\theta)}{\partial p_3} &= \frac{(p_1-p_3)\pi_3}{(p_1-\theta)w(p_1,p_2)+(\theta-p_3)w(p_2,p_3)}[w(p_1,p_2)\hat{v}'(p_3)-w(p_1,p_3)w(p_2,p_3)],
\end{aligned}$$

where all the fractions are strictly positive.

Now suppose posteriors $(\hat{p}_1, \hat{p}_2, \hat{p}_3)$ are optimal.

If any pair $(p, q) \in \{(\hat{p}_1, \hat{p}_2), (\hat{p}_1, \hat{p}_3), (\hat{p}_2, \hat{p}_3)\}$ of these posteriors are in the interior of $[0, 1]$, then their optimality implies that the two derivatives corresponding to the posteriors evaluated at $(\hat{p}_1, \hat{p}_2, \hat{p}_3)$ equals zero. For the case $(p, q) = (\hat{p}_1, \hat{p}_2)$, it follows

$$w(\hat{p}_2,\hat{p}_3)\hat{v}'(\hat{p}_1)=w(\hat{p}_1,\hat{p}_3)w(\hat{p}_1,\hat{p}_2) \text{ and } w(\hat{p}_1,\hat{p}_3)\hat{v}'(\hat{p}_2)=w(\hat{p}_2,\hat{p}_3)w(\hat{p}_1,\hat{p}_2).$$

From which it follows $\tilde{A}(p, q) = 0$. Likewise, the cases $(p, q) = (\hat{p}_1, \hat{p}_3)$ and $(p, q) = (\hat{p}_2, \hat{p}_3)$ also imply $\tilde{A}(p, q) = 0$.

Similarly, if $q = 0$ and $p < 1$, then their optimality implies that, evaluated at $(\hat{p}_1, \hat{p}_2, \hat{p}_3)$, we have $\partial S/\partial q \leq 0$ and $\partial S/\partial p = 0$. For the case $(p, q) = (\hat{p}_1, \hat{p}_2)$, it follows

$$w(\hat{p}_2,\hat{p}_3)\hat{v}'(\hat{p}_1)=w(\hat{p}_1,\hat{p}_3)w(\hat{p}_1,\hat{p}_2) \text{ and } w(\hat{p}_1,\hat{p}_3)\hat{v}'(\hat{p}_2)\leq w(\hat{p}_2,\hat{p}_3)w(\hat{p}_1,\hat{p}_2)$$

From which it follows $\tilde{A}(p, q) \leq 0$. Likewise, the cases $(p, q) = (\hat{p}_1, \hat{p}_3)$ and $(p, q) = (\hat{p}_2, \hat{p}_3)$ also imply $\tilde{A}(p, q) \leq 0$.

Similarly, if $q > 0$ and $p = 1$, then their optimality implies that, evaluated at $(\hat{p}_1, \hat{p}_2, \hat{p}_3)$, we have $\partial S/\partial q = 0$ and $\partial S/\partial p \geq 0$. That is, it holds

$$w(\hat{p}_2,\hat{p}_3)\hat{v}'(\hat{p}_1)\geq w(\hat{p}_1,\hat{p}_3)w(\hat{p}_1,\hat{p}_2) \text{ and } w(\hat{p}_1,\hat{p}_3)\hat{v}'(\hat{p}_2)=w(\hat{p}_2,\hat{p}_3)w(\hat{p}_1,\hat{p}_2)$$

From which it follows $\tilde{A}(p,q) \geq 0$. Likewise, the cases $(p,q) = (\hat{p}_1, \hat{p}_3)$ and $(p,q) = (\hat{p}_2, \hat{p}_3)$ also imply $\tilde{A}(p,q) \geq 0$. $\square$

Proof of Lemma 5: Define for $q, p \in [0,1]$ with $p > q$,

$$g(p,q) \equiv \hat{v}'(p)\hat{v}'(q) \text{ and } h(p,q) \equiv w(p,q)^2.$$

We extend the domain of $g(p,q)$ and $h(p,q)$ to include $p = q$ by setting $g(p,p) = \hat{g}(p)$ and $h(p,p) = \hat{h}(p)$, where $\hat{g}(p) \equiv \hat{v}'(p)^2$ and $\hat{h}(p) \equiv \lim_{q \uparrow p} h(q,p) = \hat{v}'(p)^2$.

Note that $g(p,q) \in [\hat{v}'(0)^2, \hat{v}'(1)^2]$ and $h(p,q) \in [\hat{v}'(0)^2, \hat{v}'(1)^2]$. Moreover,

$$A(p,q) = \sqrt{g(p,q)} - \sqrt{h(p,q)}$$

so that $A(p,q) = 0$ if and only if $g(p,q) = h(p,q)$. As $g(p,p) = h(p,p) = \hat{v}'(p)^2$, we have, in particular, $A(p,p) = \lim_{q \uparrow p} A(p,q) = 0$.

Next, for any $k \in [\hat{v}'(0)^2, \hat{v}'(1)^2]$, define functions $q_g(p|k)$ and $q_h(p|k)$ implicitly by

$$g(p, q_g(p|k)) = k \text{ and } h(p, q_h(p|k)) = k. \tag{45}$$

By the implicit function theorem, we obtain for $p > q$

$$q_g'(p|k) = -\frac{\hat{v}''(p)\hat{v}'(q)}{\hat{v}''(q)\hat{v}'(p)} \leq 0 \text{ and } q_h'(p|k) = -\frac{\hat{v}'(p) - w(p,q)}{w(p,q) - \hat{v}'(q)} \leq 0. \tag{46}$$

In addition, note that we have $\hat{g}(p) = \hat{h}(p) = \hat{v}'(p)^2$ so that the convexity of $\hat{v}(\theta)$ implies that $\hat{h}(p)$ and $\hat{g}(p)$ are continuous and increasing with $\hat{h}(0) = \hat{g}(0) = \hat{v}'(0)^2$ and $\hat{h}(1) = \hat{g}(1) = \hat{v}'(1)^2$. Hence, for any $k \in [\hat{v}'(0)^2, \hat{v}'(1)^2]$, we can find a $p(k) \in [0,1]$ such that

$$q_g(p(k)|k) = q_h(p(k)|k) = p(k).$$

Thus, we have established that for any $k \in [\hat{v}'(0)^2, \hat{v}'(1)^2]$, we can find a $p(k) \in [0,1]$ such that $A(p(k), p(k)) = 0$. Hence, the curves $q_g(p|k)$ and $q_h(p|k)$ intersect at the point $(p(k), p(k))$.

We next argue that, while the curves have the same slope of $q_g'(p(k)|k) = q_h'(p(k)|k) = -1$ at the intersection point $(p,q) = (p(k), p(k))$, $\hat{A}'(\theta) > 0$ implies $q_g''(p(k)|k) > q_h''(p(k)|k)$ so that $q_g(p(k) + \varepsilon|k) > q_h(p(k) + \varepsilon|k)$ for all $\varepsilon > 0$ small. Indeed, evaluating $q_g'(p|k)$ at $p = q = p(k)$ by taking limits, we obtain

$$q_g'(p(k)|k) = \lim_{\varepsilon \to 0} -\frac{\hat{v}''(p)\hat{v}'(p-\varepsilon)}{\hat{v}''(p-\varepsilon)\hat{v}'(p)} = -1$$

Likewise, evaluating $q_h'(p|k)$ at $p = q = p(k)$, we obtain after applying Hopital's rule twice

$$q_h'(p(k)|k) = \lim_{\varepsilon \to 0} -\frac{\hat{v}(p) - \hat{v}(p-\varepsilon) - \varepsilon\hat{v}'(p)}{\hat{v}(p) + \hat{v}(p-\varepsilon) + \varepsilon\hat{v}'(p-\varepsilon)} = \lim_{\varepsilon \to 0} -\frac{\hat{v}''(p-\varepsilon)}{\hat{v}''(p-\varepsilon) + \varepsilon\hat{v}'''(p-\varepsilon)} = -1$$

Implicitly differentiating (45) twice, we evaluate the second-order derivative of $q_g(p|k)$ at $(p,q) = (p(k), p(k))$ to obtain

$$q_g''(p(k)|k) = \frac{2\hat{v}''(p(k))}{\hat{v}'(p(k))} - \frac{2\hat{v}'''(p(k))}{\hat{v}''(p(k))}.$$

Evaluating moreover the second-order derivative of $q_h(p|k)$ at $(p,q) = (p(k), p(k))$ by taking limits, we obtain after using l'Hopital's rule multiple times that

$$q_h''(p(k)|k) = -\frac{2\hat{v}'''(p(k))}{3\hat{v}''(p(k))}.$$

Thus, we obtain

$$q_g''(p(k)|k) > q_h''(p(k)|k) \Leftrightarrow \frac{2[3\hat{v}''(p)^2 - 2\hat{v}'(p)\hat{v}'''(p)]}{3\hat{v}'(p)\hat{v}''(p)} \geq 0 \Leftrightarrow 3\hat{v}''(p)^2 - 2\hat{v}'(p)\hat{v}'''(p)) \geq 0 \Leftrightarrow \hat{A}'(\theta) > 0. \tag{47}$$

We next argue that if $\hat{A}(\theta)$ is strictly increasing then the previous result implies that the two curves do not intersect for any $p > p(k)$. We do so by showing that a strictly increasing $\hat{A}$ implies $q_g'(p|k) > q_h'(p|k)$ at any point (p,q) with $p > q$ where the curves intersect (i.e., where $q_g(p|k) = q_h(p|k)$) and, hence, at any intersection point the curve $q_g(p|k)$ cuts $q_h(p|k)$ from above, implying that the two curves have an intersection point only at $(p,q) = (p(k), p(k))$. To see $q_g'(p|k) > q_h'(p|k)$ for any intersection point, note first that at an intersection point (p,q), it holds

$$\hat{v}'(p)\hat{v}'(q) = w(p,q)^2. \tag{48}$$

Using (48), it holds for $q_h'(p|k)$ at an intersection point (p,q) that

$$q_h'(p|k) = -\frac{\hat{v}'(p) - w(p,q)}{w(p,q) - \hat{v}'(q)} = -\frac{\hat{v}'(p)w(p,q) - w(p,q)^2}{w(p,q)^2 - w(p,q)\hat{v}'(q)} = -\frac{\hat{v}'(p)w(p,q) - \hat{v}'(p)\hat{v}'(q)}{\hat{v}'(p)\hat{v}'(q) - w(p,q)\hat{v}'(q)} = \frac{\hat{v}'(p)}{\hat{v}'(q)}\frac{1}{q_h'(p|k)}.$$

Hence, it holds

$$q_h'(p|k)^2 = \frac{\hat{v}'(p)}{\hat{v}'(q)},$$

which with $q_h'(p|k) \leq 0$ implies

$$q_h'(p|k) = -\frac{\sqrt{\hat{v}'(p)}}{\sqrt{\hat{v}'(q)}}.$$

We therefore have

$$q_g'(p|k) > q_h'(p|k) \Leftrightarrow \frac{\hat{v}''(p)\hat{v}'(q)}{\hat{v}''(q)\hat{v}'(p)} < \frac{\sqrt{\hat{v}'(p)}}{\sqrt{\hat{v}'(q)}} \Leftrightarrow \frac{\hat{v}'(q)\sqrt{\hat{v}'(q)}}{\hat{v}''(q)} < \frac{\hat{v}'(p)\sqrt{\hat{v}'(p)}}{\hat{v}''(p)} \Leftrightarrow \hat{A}(q) < \hat{A}(p), \tag{49}$$

which, given that $\hat{A}(\theta)$ is strictly increasing for $\theta > 0$, holds true for any $p > q \geq 0$. This yields the first statement of the lemma.

By contrast, if $\hat{A}(\theta)$ is strictly decreasing all inequalities in (47) and (49) hold in reverse, and we obtain the second statement of the lemma. □

Proof of Lemma 8: Consider a 2-certificate menu $\hat{\Gamma}_2$ that is non-maximal. That is, the partition $\mathcal{M}$ contains a non-singleton subset $\hat{\Theta} \subseteq \Theta$ so that it holds $p(\theta) = \bar{p}$ and $q(\theta) = \bar{q}$ for all $\theta \in \hat{\Theta}$. Let $\bar{\theta}$ denote the average over $\hat{\Theta}$. Then all types $\theta \in \hat{\Theta}$ receive the same marginal information rent

$$\bar{I} \equiv \frac{\bar{p} - \bar{\theta}}{1 - \bar{\theta}} \frac{\bar{\theta} - \bar{q}}{\bar{\theta}} w(\bar{p}, \bar{q}).$$

Now consider a pair of mappings $(\check{p}(\theta), \check{q}(\theta))$ with $(\check{p}(\bar{\theta}), \check{q}(\bar{\theta})) = (\bar{p}, \bar{q})$ and the function

$$\check{S}(\theta) \equiv \hat{v}(\check{q}(\theta)) + [\theta - \check{q}(\theta)] w(\check{p}(\theta), \check{q}(\theta)). \tag{50}$$

If $\check{S}$ is convex, it follows

$$\begin{aligned}
\hat{S} &= \int_{\theta \in \hat{\Theta}} \left[\hat{v}(\bar{q}) + (\bar{\theta} - \bar{q}) \left[\frac{\theta \bar{p}}{\bar{\theta}} + \frac{(1-\theta)(1-\bar{p})}{(1-\bar{\theta})} \right] w(\bar{p}, \bar{q}) \right] dF(\theta) + \int_{\theta \notin \hat{\Theta}} \hat{S}(\theta) dF(\theta) \\
&= \int_{\theta \in \hat{\Theta}} \left[\hat{v}(\bar{q}) + (\bar{\theta} - \bar{q}) w(\bar{p}, \bar{q}) \right] dF(\theta) + \int_{\theta \notin \hat{\Theta}} \hat{S}(\theta) dF(\theta) \\
&= \int_{\theta \in \hat{\Theta}} \check{S}(\bar{\theta}) dF(\theta) + \int_{\theta \notin \hat{\Theta}} \hat{S}(\theta) dF(\theta) \\
&\leq \int_{\theta \in \hat{\Theta}} \check{S}(\theta) dF(\theta) + \int_{\theta \notin \hat{\Theta}} \hat{S}(\theta) dF(\theta).
\end{aligned}$$

To obtain the result, it therefore suffices to construct a pair of mappings $(\check{p}(\theta), \check{q}(\theta))$ with $(\check{p}(\bar{\theta}), \check{q}(\bar{\theta})) = (\bar{p}, \bar{q})$ so that $\check{S}$ is convex in θ and for all $\theta \in \hat{\Theta}$, it holds

$$\frac{\check{p}(\theta) - \theta}{1 - \theta} \frac{\theta - \check{q}(\theta)}{\theta} w(\check{p}(\theta), \check{q}(\theta)) = \bar{I}. \tag{51}$$

Given $(\check{p}(\bar{\theta}), \check{q}(\bar{\theta})) = (\bar{p}, \bar{q})$, Equality (51) holds for all $\theta \in \hat{\Theta}$ if $(\check{p}(\theta), \check{q}(\theta))$ satisfies the differential equation

$$(\theta - \check{q})[w(\check{p}, \check{q}) + (\check{p} - \theta) w_p(\check{p}, \check{q})]\check{p}' - (\check{p} - \theta)[w(\check{p}, \check{q}) - (\theta - \check{q}) w_q(\check{p}, \check{q})]\check{q}' = \psi(\theta, \check{p}, \check{q}) w(\check{p}, \check{q}), \tag{52}$$

where we dropped the argument θ and have

$$\psi(\theta, p, q) \equiv (\theta - q) \frac{1-p}{1-\theta} - (p - \theta) \frac{q}{\theta}.$$

We construct $(\check{p}(\theta), \check{q}(\theta))$ as follows. We fix $\check{p}(\bar{\theta}) = \bar{p}$ and $\check{q}(\bar{\theta}) = \bar{q}$. Then, depending on whether $\psi(\theta, \check{p}, \check{q})$ is positive or negative, marginally change only $\check{p}$ or $\check{q}$ so that type θ's marginal information rent remains at $\bar{I}$. In particular,

$$\check{p}'(\theta) = \begin{cases} \frac{\psi(\theta,\check{p},\check{q})}{\theta - \check{q}} \frac{w}{w + (\check{p}-\theta) w_p} & \text{if } \psi(\theta, \check{p}, \check{q}) \geq 0 \\ 0 & \text{otherwise;} \end{cases} \quad \text{and } \check{q}'(\theta) = \begin{cases} \frac{-\psi(\theta,\check{p},\check{q})}{\check{p} - \theta} \frac{w}{w - (\theta - \check{q}) w_q} & \text{if } \psi(\theta, \check{p}, \check{q}) \leq 0 \\ 0 & \text{otherwise.} \end{cases} \tag{53}$$

This construction yields two differential equations that define the mappings $\check{p}(\theta)$ and $\check{q}(\theta)$. They are both continuous and (weakly) increasing in θ. They are also continuously differentiable, as $\check{p}'(\theta^0) = \check{q}'(\theta^0) = 0$ with θ^0 such that $\psi(\theta^0, \check{p}(\theta^0), \check{q}(\theta^0)) = 0$. When $\bar{p} = 1$ and $\bar{q} = 0$, function $\psi(\theta, \check{p}, \check{q})$ equals zero for all values of θ, implying that $\check{p}(\theta) = \bar{p}$ and $\check{q}(\theta) = \bar{q}$ for all θ. It is easy to see that the resulting $\check{S}$ is linear in θ. This is the only $(\bar{p}, \bar{q})$ pair for which the constructed $\check{S}$ is linear; otherwise the constructed $\check{S}$ is strictly convex on the relevant non-degenerate intervals. For the continuation of the proof, suppose that $(\bar{p}, \bar{q}) \neq (1, 0)$. For θ such that $\psi \geq 0$, our construction implies that ψ is strictly increasing in θ, as

$$\begin{aligned}\left.\frac{d\psi}{d\theta}\right|_{\psi \geq 0} &= \frac{\partial \psi}{\partial \theta} + \frac{\partial \psi}{\partial \check{p}} \check{p}'(\theta) = \frac{(1-\check{p})(1-\check{q})}{(1-\theta)^2} + \frac{\check{p}\check{q}}{\theta^2} - \left(\frac{\theta - \check{q}}{1-\theta} + \frac{\check{q}}{\theta}\right) \check{p}'(\theta) \\ &> \frac{(1-\check{p})(1-\check{q})}{(1-\theta)^2} + \frac{\check{p}\check{q}}{\theta^2} - \left(\frac{\theta - \check{q}}{1-\theta} + \frac{\check{q}}{\theta}\right) \frac{(1-\check{p})}{(1-\theta)} = \frac{1-\check{p}}{1-\theta} + \frac{\check{q}}{\theta} \frac{\check{p} - \theta}{\theta(1-\theta)} > 0.\end{aligned}$$

This indicates that ψ changes sign at most once, and only from negative to positive. Hence, there is at most one value θ^0 such that $\psi(\theta^0, \check{p}(\theta^0), \check{q}(\theta^0)) = 0$. Moreover, $\psi(\theta, \check{p}(\theta), \check{q}(\theta)) > 0$ for $\theta > \theta^0$ and $\psi(\theta, \check{p}(\theta), \check{q}(\theta)) < 0$ for $\theta < \theta^0$.

Note that $\check{S}'(\theta)$ is continuous at θ^0.[21] As a result, it suffices to show convexity of $\check{S}$ by showing convexity of $\check{S}$ separately for $\psi < 0$ and $\psi > 0$. Hence, we distinguish two cases:

Case 1: Consider an interval over which $\psi \geq 0$. In this case $\check{q}'(\theta) = 0$ and

$$\check{p}'(\theta) = \frac{\psi(\theta, \check{p}, \check{q})}{\theta - \check{q}} \frac{w}{w + (\check{p} - \theta) w_p} = \underbrace{\left(\frac{1-\check{p}}{1-\theta} - \frac{(\check{p} - \theta)\check{q}}{(\theta - \check{q})\theta}\right)}_{\in[0,1]} \underbrace{\frac{w}{w + (\check{p} - \theta)\, w_p}}_{\in[0,1]} \in [0,1].$$

Using the middle expression in (50) and $\check{q}'(\theta) = 0$, we obtain

$$\check{S}'(\theta) = w + (\theta - \check{q}) w_p \check{p}'(\theta); \tag{54}$$

and differentiating once more yields

$$\begin{aligned}\check{S}''(\theta) &= 2w_p \check{p}'(\theta) + (\theta - \check{q})[w_{pp} \check{p}'(\theta)^2 + w_p \check{p}''(\theta)] &(55)\\ &= 2w_p \check{p}'(\theta) - 2\frac{\theta - \check{q}}{\check{p} - \check{q}} w_p \check{p}'(\theta)^2 + \frac{\theta - \check{q}}{\check{p} - \check{q}} \hat{v}''(\check{p}) \check{p}'(\theta)^2 + (\theta - \check{q}) w_p \check{p}''(\theta) &(56)\\ &= 2w_p \check{p}'(\theta) \left(1 - \frac{\theta - \check{q}}{\check{p} - \check{q}} \check{p}'(\theta)\right) + \frac{\theta - \check{q}}{\check{p} - \check{q}} \hat{v}''(\check{p}) \check{p}'(\theta)^2 + (\theta - \check{q}) w_p \check{p}''(\theta) &(57)\\ &\geq \frac{\theta - \check{q}}{\check{p} - \check{q}} \left(\hat{v}''(\check{p}) \check{p}'(\theta)^2 + (\check{p} - \check{q}) w_p \check{p}''(\theta)\right), &(58)\end{aligned}$$

[21] To see that $\check{S}'(\theta)$ is continuous at θ^0 (given its existence), note $\lim_{\theta \uparrow \theta^0} \check{S}'(\theta) = w(\check{p}(\theta^0), \check{q}(\theta^0)) - [\check{p}(\theta^0) - \theta^0] w_q(\check{p}(\theta^0), \check{q}(\theta^0)) \check{q}'(\theta^0) = w(\check{p}(\theta^0), \check{q}(\theta^0))$ which equals $\lim_{\theta \downarrow \theta^0} \check{S}'(\theta) = w(\check{p}(\theta^0), \check{q}(\theta^0)) - [\theta^0 - \check{q}(\theta^0)] w_p(\check{p}(\theta^0), \check{q}(\theta^0)) \check{p}'(\theta^0) = w(\check{p}(\theta^0), \check{q}(\theta^0))$.

where (56) follows from substituting out w_{pp}, where the definition of $w(p,q)$ implies

$$w_p = \frac{\hat{v}'(p) - w}{p-q} \text{ and } w_{pp} = \frac{\hat{v}''(p) - 2w_p}{p-q}, \tag{59}$$

and (58) follows because the first term in (57) is positive as $w_p \geq 0$, $\check{p}'(\theta) \in [0,1]$, and $0 < \theta - \check{q} < \check{p} - \check{q}$ so that the term in parenthesis is also positive.

We next show that also the remaining term (58) is positive, implying that $\check{S}''(\theta) \geq 0$ and, hence, $\check{S}$ is convex over the range $\psi \geq 0$. To sign (58), we first derive $\check{p}''(\theta)$ for $\psi \geq 0$. By the product rule, it follows from (53) that

$$\begin{aligned}
\check{p}''(\theta) &= \left(\frac{1-\check{p}}{1-\theta} - \frac{\check{p}-\theta}{\theta-\check{q}}\frac{\check{q}}{\theta}\right)\left(\frac{w}{w+(\check{p}-\theta)w_p}\right)' + \left(\frac{1-\check{p}}{1-\theta} - \frac{\check{p}-\theta}{\theta-\check{q}}\frac{\check{q}}{\theta}\right)'\left(\frac{w}{w+(\check{p}-\theta)w_p}\right) && (60)\\
&\geq \frac{2w_p\check{p}'(\theta) - 2\frac{\theta-\check{q}}{\check{p}-\check{q}}w_p\check{p}'(\theta)^2 - \frac{\check{p}-\theta}{\check{p}-\check{q}}\hat{v}''(\check{p})\check{p}'(\theta)^2}{w+(\check{p}-\theta)w_p} \geq -\frac{(\check{p}-\theta)\hat{v}''(\check{p})\check{p}'(\theta)^2}{(\check{p}-\check{q})(w+(\check{p}-\theta)w_p)}, && (61)
\end{aligned}$$

where the last inequality in (61) obtains because, taken together, the first two terms in the numerator of the first fraction are positive as $\theta - \check{q} < \check{p} - \check{q}$ and $\check{p}'(\theta) \in [0,1]$. To see the first inequality in (61) note that collecting terms after substituting out w_{pp} using (59), we obtain

$$\begin{aligned}
\left(\frac{w}{w+(\check{p}-\theta)\,w_p}\right)' &= \frac{[w+(\check{p}-\theta)w_p]w_p\check{p}'(\theta) - [w_p\check{p}'(\theta) + (\check{p}'(\theta)-1)w_p + (\check{p}-\theta)w_{pp}\check{p}'(\theta)]w}{(w+(\check{p}-\theta)\,w_p)^2}\\
&= \frac{ww_p - 2\frac{\theta-\check{q}}{\check{p}-\check{q}}ww_p\check{p}'(\theta) + w_p\check{p}'(\theta)(w+(\check{p}-\theta)w_p) - \frac{\check{p}-\theta}{\check{p}-\check{q}}w\hat{v}''(\check{p})\check{p}'(\theta)}{(w+(\check{p}-\theta)w_p)^2}.
\end{aligned}$$

Moreover, it follows

$$\begin{aligned}
\left(\frac{1-p}{1-\theta} - \frac{p-\theta}{\theta-q}\frac{q}{\theta}\right)' &= \frac{1-p}{(1-\theta)^2} + \frac{(p-q)}{(\theta-q)^2}\frac{q}{\theta} + \frac{p-\theta}{\theta-q}\frac{q}{\theta^2} - \frac{1}{1-\theta}p'(\theta) - \frac{1}{\theta-q}\frac{q}{\theta}p'(\theta)\\
&\geq \frac{1-p}{(1-\theta)^2} + \frac{1}{\theta-q}\frac{1-p}{1-\theta}\frac{q}{\theta} - \frac{1}{1-\theta}p'(\theta) - \frac{1}{\theta-q}\frac{q}{\theta}p'(\theta)\\
&\geq \left(\frac{1-p}{1-\theta} - \frac{p-\theta}{\theta-q}\frac{q}{\theta} - p'(\theta)\right)\left(\frac{1}{1-\theta} + \frac{1}{\theta-q}\frac{q}{\theta}\right)\\
&= \frac{(p-\theta)\,w_p}{w}\left(\frac{1}{1-\theta} + \frac{1}{\theta-q}\frac{q}{\theta}\right)p'(\theta),
\end{aligned}$$

where the first inequality holds since $\frac{p-q}{\theta-q} \geq 1 \geq \frac{1-p}{1-\theta}$ and $\frac{p-\theta}{\theta-q}\frac{q}{\theta^2}$ is non-negative, the second one holds since $\frac{p-\theta}{\theta-q}\frac{q}{\theta}$ is non-negative. Substitution of these terms in (60) yields the first inequality in (61) and, after collecting terms, the first fraction.

Using (61), we continue from (58) to obtain

$$\check{S}''(\theta) \geq \frac{\theta-\check{q}}{\check{p}-\check{q}}\left(1 - \frac{w_p(\check{p}-\theta)}{w+(\check{p}-\theta)w_p}\right)\hat{v}''(\check{p})\check{p}'(\theta)^2 = \frac{\theta-\check{q}}{\check{p}-\check{q}}\frac{w}{w+(\check{p}-\theta)w_p}\hat{v}''(\check{p})\check{p}'(\theta)^2 \geq 0.$$

This establishes convexity of $\check{S}$ over an interval for which $\psi \geq 0$.

Case 2: Consider an interval over which $\psi < 0$. In this case $\check{p}'(\theta) = 0$ and[22]

$$\check{q}'(\theta) = -\frac{\psi(\theta, \check{p}, \check{q})}{\check{p} - \theta} \frac{w}{w - (\theta - \check{q})w_q} = \underbrace{\left(\frac{\check{q}}{\theta} - \frac{\theta - \check{q}}{\check{p} - \theta} \frac{1 - \check{p}}{1 - \theta}\right)}_{\in[0,1]} \underbrace{\frac{w}{w - (\theta - \check{q})w_q}}_{\in\left[0, \frac{\check{p} - \check{q}}{\check{p} - \theta}\right]} \in \left[0, \frac{\check{p} - \check{q}}{\check{p} - \theta}\right].$$

We first establish that the definition of $w(p, q)$ implies

$$w_q = \frac{w - \hat{v}'(q)}{p - q} \text{ and } w_{qq} = \frac{2w_q - \hat{v}''(q)}{p - q}. \tag{62}$$

By the product rule, we have

$$\begin{aligned}
\check{q}''(\theta) &= \left(\frac{\check{q}}{\theta} - \frac{\theta - \check{q}}{\check{p} - \theta} \frac{1 - \check{p}}{1 - \theta}\right) \left(\frac{w}{w - (\theta - \check{q})\, w_q}\right)' + \left(\frac{\check{q}}{\theta} - \frac{\theta - \check{q}}{\check{p} - \theta} \frac{1 - \check{p}}{1 - \theta}\right)' \frac{w}{w - (\theta - \check{q})\, w_q}, &(63)\\
&\leq \frac{2(\check{p} - \check{q}) w_q \check{q}'(\theta) - 2(\check{p} - \theta) w_q \check{q}'(\theta)^2 - (\theta - \check{q}) \hat{v}''(\check{q}) \check{q}'(\theta)^2}{[\check{p} - \check{q}][w - (\theta - \check{q}) w_q]} &(64)
\end{aligned}$$

where the inequality follows from using (62) to obtain

$$\begin{aligned}
\left(\frac{w}{w - (\theta - \check{q})w_q}\right)' &= \frac{(w - (\theta - \check{q})w_q) w_q \check{q}'(\theta) - [w_q \check{q}'(\theta) - (1 - \check{q}'(\theta)) w_q - (\theta - \check{q}) w_{qq} \check{q}'(\theta)] w}{(w - (\theta - \check{q}) w_q)^2} \\
&= \frac{w w_q - 2\frac{\check{p} - \theta}{\check{p} - \check{q}} w w_q \check{q}'(\theta) + (w - (\theta - \check{q}) w_q) w_q \check{q}'(\theta) - \frac{\theta - \check{q}}{\check{p} - \check{q}} w \hat{v}''(\check{q}) \check{q}'(\theta)}{(w - (\theta - \check{q}) w_q)^2},
\end{aligned}$$

and

$$\begin{aligned}
\left(\frac{\check{q}}{\theta} - \frac{\theta - \check{q}}{\check{p} - \theta} \frac{1 - \check{p}}{1 - \theta}\right)' &= \left(\frac{1}{\theta} + \frac{1}{\check{p} - \theta} \frac{1 - \check{p}}{1 - \theta}\right) q'(\theta) - \frac{\check{q}}{\theta^2} - \frac{\check{p} - \check{q}}{\check{p} - \theta} \frac{1}{\check{p} - \theta} \frac{1 - \check{p}}{1 - \theta} - \frac{\theta - \check{q}}{\check{p} - \theta} \frac{1 - \check{p}}{(1 - \theta)^2} \\
&\leq \left(\frac{1}{\theta} + \frac{1}{\check{p} - \theta} \frac{1 - \check{p}}{1 - \theta}\right) \check{q}'(\theta) - \frac{\check{q}}{\theta^2} - \frac{\check{q}}{\theta} \frac{1}{\check{p} - \theta} \frac{1 - \check{p}}{1 - \theta} \\
&\leq \left(\frac{1}{\theta} + \frac{1}{\check{p} - \theta} \frac{1 - \check{p}}{1 - \theta}\right) \check{q}'(\theta) - \left(\frac{1}{\theta} + \frac{1}{\check{p} - \theta} \frac{1 - \check{p}}{1 - \theta}\right) \left(\frac{\check{q}}{\theta} - \frac{\theta - \check{q}}{\check{p} - \theta} \frac{1 - \check{p}}{1 - \theta}\right) \\
&= \frac{(\theta - \check{q})\, w_q}{w} \left(\frac{1}{\theta} + \frac{1}{\check{p} - \theta} \frac{1 - \check{p}}{1 - \theta}\right) \check{q}'(\theta),
\end{aligned}$$

using $\frac{\check{p} - \check{q}}{\check{p} - \theta} \geq 1 \geq \frac{\check{q}}{\theta}$ and $\frac{\theta - \check{q}}{\check{p} - \theta} \frac{1 - \check{p}}{(1 - \theta)^2}$ is non-negative to obtain the first inequality, and using $\frac{\theta - \check{q}}{\check{p} - \theta} \frac{1 - \check{p}}{1 - \theta} \geq 0$ to obtain the second inequality.

Using the right-hand-side expression in (50) and $\check{p}'(\theta) = 0$, we obtain

$$\check{S}'(\theta) = w - (\check{p} - \theta) w_q \check{q}'(\theta); \tag{65}$$

[22] To see the bounds for the second factor, note that $w_q = [w - \hat{v}'(\check{q})]/(\check{p} - \check{q})$ so that it follows $w/[w - (\theta - \check{q}) w_q] = (\check{p} - \check{q}) w/[(\check{p} - \theta) w + (\theta - \check{q}) \hat{v}'(\check{q})] \leq [\check{p} - \check{q}]/[\check{p} - \theta]$, since $\theta \geq \check{q}$ and $\hat{v}'(\check{q}) > 0$.

and differentiating once more yields

$$\check{S}''(\theta) = 2w_q\check{q}'(\theta) - (\check{p}-\theta)[w_{qq}\check{q}'(\theta)^2 + w_q\check{q}''(\theta)] \tag{66}$$
$$= 2w_q\check{q}'(\theta)\left(1 - \frac{\check{p}-\theta}{\check{p}-\check{q}}\check{q}'(\theta)\right) + \frac{\check{p}-\theta}{\check{p}-\check{q}}\hat{v}''(\check{q})\check{q}'(\theta)^2 - (\check{p}-\theta)w_q\check{q}''(\theta) \tag{67}$$
$$\geq 2w_q\check{q}'(\theta)\left(1 - \frac{\check{p}-\theta}{\check{p}-\check{q}}\check{q}'(\theta)\right)\left[1 - \frac{(\check{p}-\theta)w_q}{w-(\theta-\check{q})w_q}\right] \geq 0. \tag{68}$$

where the second equality follows from substituting out w_{qq} using (62) and the inequality from (64). The final inequality follows because $w_q \geq 0$, $\check{q}'(\theta) \geq 0$,

$$1 - \frac{\check{p}-\theta}{\check{p}-\check{q}}\check{q}'(\theta) \geq 0$$

by the bound on $\check{q}'$, and

$$1 - \frac{(\check{p}-\theta)w_q}{w-(\theta-\check{q})w_q} = \frac{w-(\check{p}-\check{q})w_q}{w-(\theta-\check{q})w_q} = \frac{\hat{v}'(\check{q})}{w-(\theta-\check{q})w_q} \geq 0.$$

This establishes convexity of $\check{S}$ over an interval for which $\psi < 0$. □

Proof of Proposition 2: It follows from Lemma 7 that Pareto optimal certification menus can be composed of contracts exhibiting at most three certificates. If a menu pools some subset of types $\hat{\Theta} \in \Theta$ at a two-certificate contract, then it follows from the proof of Lemma 8 that the certifier's revenue can be improved with an alternative menu fully separating these types, without changing the buyer's utility or the information rents of the seller types. It remains to show that certification menus pooling types at a three-certificate contract that induces three distinct posteriors are Pareto dominated as well.

To show this, we consider an incentive compatible menu that pools a subset of types $\hat{\Theta} \subset \Theta$ at a certification structure $\hat{\sigma} = (\{\hat{c}_1, \hat{c}_2, \hat{c}_3\}, (\bar{\pi}_1^l, \bar{\pi}_2^l, \bar{\pi}_3^l), (\bar{\pi}_1^h, \bar{\pi}_2^h, \bar{\pi}_3^h))$ such that each of the three certificates occurs with strictly positive probability: $\bar{\pi}_1^h + \bar{\pi}_1^l > 0$, $\bar{\pi}_2^h + \bar{\pi}_2^l > 0$, $\bar{\pi}_3^h + \bar{\pi}_3^l > 0$. We denote by $\bar{\theta} = \mathbb{E}\{\hat{\Theta}\}$ the expected type over the subset $\hat{\Theta}$ (with possibly $\bar{\theta} \notin \hat{\Theta}$). This certification structure induces posteriors

$$\bar{p}_1 = \frac{\bar{\theta}\bar{\pi}_1^h}{\bar{\theta}\bar{\pi}_1^h + (1-\bar{\theta})\bar{\pi}_1^l};\ \bar{p}_2 = \frac{\bar{\theta}\bar{\pi}_2^h}{\bar{\theta}\bar{\pi}_2^h + (1-\bar{\theta})\bar{\pi}_2^l};\ \bar{p}_3 = \frac{\bar{\theta}\bar{\pi}_3^h}{\bar{\theta}\bar{\pi}_3^h + (1-\bar{\theta})\bar{\pi}_3^l}, \tag{69}$$

such that $\bar{p}_1 > \bar{p}_2 > \bar{p}_3$. Incentive compatibility implies that each type $\hat{\theta} \in \hat{\Theta}$ obtains the marginal information rent

$$\bar{I} = [\bar{\pi}_1^h - \bar{\pi}_1^l]\hat{v}(\bar{p}_1) + [\bar{\pi}_2^h - \bar{\pi}_2^l]\hat{v}(\bar{p}_2) + [\bar{\pi}_3^h - \bar{\pi}_3^l]\hat{v}(\bar{p}_3). \tag{70}$$

A type $\hat{\theta} \in \hat{\Theta}$ generates the surplus

$$\hat{S}(\hat{\theta}) = [\hat{\theta}\bar{\pi}_1^h + (1-\hat{\theta})\bar{\pi}_1^l]\hat{v}(\bar{p}_1) + [\hat{\theta}\bar{\pi}_2^h + (1-\hat{\theta})\bar{\pi}_2^l]\hat{v}(\bar{p}_2) + [\hat{\theta}\bar{\pi}_3^h + (1-\hat{\theta})\bar{\pi}_3^l]\hat{v}(\bar{p}_3)$$

with average $\hat{S}(\bar{\theta})$.

We next construct an alternative incentive compatible menu that coincides with the old one for all $\theta \notin \hat{\Theta}$, but separates all types in $\hat{\Theta}$ such that type $\hat{\theta} \in \hat{\Theta}$ obtains certificate $\hat{c}_i$ in state $\omega \in \{l, h\}$ with probability $\pi_i^\omega(\hat{\theta})$ for $i = 1, 2, 3$, while keeping the marginal information rent constant at $\bar{I}$. The constant information rent ensures that the alternative menu is incentive compatible and it yields the same information rents for different seller types. We construct the type-dependent probability system $\{\pi_i^\omega(\hat{\theta})\}$ in such a way that the respective posteriors also remain constant at $(\bar{p}_1, \bar{p}_2, \bar{p}_3)$. It follows from the proof of Lemma 4 (Case 2) that $(\bar{p}_1, \bar{p}_2, \bar{p}_3, \hat{\theta})$ identify $\{\pi_i^\omega(\hat{\theta})\}$ as a solution to a system of six linear equations (see (43)). As the solution $\{\pi_i^\omega(\theta)\}$ is continuous in θ and since $\pi_i^\omega(\bar{\theta}) = \bar{\pi}_i^\omega$, we have $\pi_c^\omega(\theta) \in [0, 1]$ for θ close to $\bar{\theta}$. Indeed, as the numerator in $\pi_i^\omega(\theta)$ is quadratic in θ, we find for each certificate $\hat{c}_i$ two thresholds $\theta_i^l < \bar{\theta} < \theta_i^h$ such that $\pi_i^\omega(\theta) \in [0, 1]$ if and only if $\theta \in [\theta_i^l, \theta_i^h]$. Note that these thresholds are independent of the state ω.

Denoting by $\theta^l \equiv \max_i\{\theta_i^l\}$ the maximal lower threshold and by $\theta^h \equiv \min_i\{\theta_i^h\}$ the minimal upper threshold, it follows that the solution $\{\pi_i^\omega(\theta)\}$ satisfies the individual boundary conditions for any $\theta \in [\theta^l, \theta^h]$. Let $\check{S}(\theta)$ for $\theta \in [\theta^l, \theta^h]$ be the surplus generated by type θ given the solution $\{\pi_i^\omega(\theta)\}$.

$$\check{S}(\theta) = S(\bar{p}_1, \bar{p}_2, \bar{p}_3, \theta),$$

where function S is defined in (44).

Notice that the function $\check{S}$ gives the same value as function $\hat{S}$ when evaluated at the average type in subset $\hat{\Theta}$, that is $\check{S}(\bar{\theta}) = \hat{S}(\bar{\theta})$. Hence, if $\hat{\Theta} \subset [\theta^l, \theta^h]$, we obtain our result by showing that $\check{S}(\theta)$ is convex in θ. Convexity of $\check{S}$ holds if $\check{S}''(\theta)$ is positive for any $\theta \in [\theta^l, \theta^h]$.

By differentiating (44) with respect to θ twice, we find this second derivative as

$$\check{S}''(\theta) = \frac{K(\bar{p}_1, \bar{p}_2, \bar{p}_3)}{D(\bar{p}_1, \bar{p}_2, \bar{p}_3, \theta)^3}$$

with

$$\begin{aligned} K(\bar{p}_1, \bar{p}_2, \bar{p}_3) =& 2\{(\bar{p}_2 - \bar{p}_3)[\hat{v}(\bar{p}_1) - \hat{v}(\bar{p}_3)] - (\bar{p}_1 - \bar{p}_3)[\hat{v}(\bar{p}_2) - \hat{v}(\bar{p}_3)]\}\times \\ & \{(\bar{p}_1 - \bar{p}_3)(\bar{p}_2 - \bar{p}_3)(\bar{p}_1 - \bar{p}_2)[\hat{v}(\bar{p}_1) - \hat{v}(\bar{p}_3)][\hat{v}(\bar{p}_2) - \hat{v}(\bar{p}_3)][\hat{v}(\bar{p}_1) - \hat{v}(\bar{p}_2)] \\ & - [\bar{p}_1(\bar{p}_2 - \bar{p}_3)(\hat{v}(\bar{p}_1) - \hat{v}(\bar{p}_2)) - \bar{p}_3(\bar{p}_1 - \bar{p}_2)(\hat{v}(\bar{p}_2) - \hat{v}(\bar{p}_3))]\times \\ & [(1 - \bar{p}_3)(\bar{p}_1 - \bar{p}_2)(\hat{v}(\bar{p}_2) - \hat{v}(\bar{p}_3)) - (1 - \bar{p}_1)(\bar{p}_2 - \bar{p}_3)(\hat{v}(\bar{p}_1) - \hat{v}(\bar{p}_2))]\bar{I}\} \end{aligned} \quad (71)$$

and

$$D(\bar{p}_1, \bar{p}_2, \bar{p}_3, \theta) = (\bar{p}_2 - \bar{p}_3)(\bar{p}_1 - \theta)(\hat{v}(\bar{p}_1) - \hat{v}(\bar{p}_3)) + (\bar{p}_1 - \bar{p}_3)(\theta - \bar{p}_2)(\hat{v}(\bar{p}_2) - \hat{v}(\bar{p}_3))$$

To see that $D(\bar{p}_1, \bar{p}_2, \bar{p}_3, \theta)$ is positive, notice that convexity of $\hat{v}$ implies $(\bar{p}_2 - \bar{p}_3)[\hat{v}(\bar{p}_1) - \hat{v}(\bar{p}_3)] > (\bar{p}_1 - \bar{p}_3)[\hat{v}(\bar{p}_2) - \hat{v}(\bar{p}_3)]$. It follows from this inequality that

$$\begin{aligned} D(\bar{p}_1, \bar{p}_2, \bar{p}_3, \theta) &> (\bar{p}_1 - \bar{p}_3)(\bar{p}_1 - \theta)[\hat{v}(\bar{p}_2) - \hat{v}(\bar{p}_3)] + (\bar{p}_1 - \bar{p}_3)(\theta - \bar{p}_2)(\hat{v}(\bar{p}_2) - \hat{v}(\bar{p}_3)) \\ &= (\bar{p}_1 - \bar{p}_3)(\bar{p}_1 - \theta + \theta - \bar{p}_2)(\hat{v}(\bar{p}_2) - \hat{v}(\bar{p}_3)) \\ &= (\bar{p}_1 - \bar{p}_3)(\bar{p}_1 - \bar{p}_2)(\hat{v}(\bar{p}_2) - \hat{v}(\bar{p}_3)) \\ &> 0 \end{aligned}$$

Hence, the sign of $\check{S}''$ is positive if (71) is positive. As the sign of (71) coincides with the sign of the term in the second curly brackets of (71), $\check{S}''$ is positive if the following inequality holds

$$\begin{aligned} &(\bar{p}_1 - \bar{p}_3)(\bar{p}_2 - \bar{p}_3)(\bar{p}_1 - \bar{p}_2)[\hat{v}(\bar{p}_1) - \hat{v}(\bar{p}_3)][\hat{v}(\bar{p}_2) - \hat{v}(\bar{p}_3)][\hat{v}(\bar{p}_1) - \hat{v}(\bar{p}_2)] \geq \\ &\quad [\bar{p}_1(\bar{p}_2 - \bar{p}_3)(\hat{v}(\bar{p}_1) - \hat{v}(\bar{p}_2)) - \bar{p}_3(\bar{p}_1 - \bar{p}_2)(\hat{v}(\bar{p}_2) - \hat{v}(\bar{p}_3))] \times \\ &\quad\quad [(1 - \bar{p}_3)(\bar{p}_1 - \bar{p}_2)(\hat{v}(\bar{p}_2) - \hat{v}(\bar{p}_3)) - (1 - \bar{p}_1)(\bar{p}_2 - \bar{p}_3)(\hat{v}(\bar{p}_1) - \hat{v}(\bar{p}_2))]\bar{I} \end{aligned} \tag{72}$$

For any $\bar{I} \geq 0$, the right hand side of (72) is smaller than

$$\begin{aligned} &[\bar{p}_1(\bar{p}_2 - \bar{p}_3)(\hat{v}(\bar{p}_1) - \hat{v}(\bar{p}_2))][(1 - \bar{p}_3)(\bar{p}_1 - \bar{p}_2)(\hat{v}(\bar{p}_2) - \hat{v}(\bar{p}_3))] \\ &\quad + [\bar{p}_3(\bar{p}_1 - \bar{p}_2)(\hat{v}(\bar{p}_2) - \hat{v}(\bar{p}_3))][(1 - \bar{p}_1)(\bar{p}_2 - \bar{p}_3)(\hat{v}(\bar{p}_1) - \hat{v}(\bar{p}_2))]\bar{I}, \end{aligned}$$

which simplifies to

$$[(\bar{p}_2 - \bar{p}_3)(\bar{p}_1 - \bar{p}_2)(\bar{p}_1 + \bar{p}_3 - 2\bar{p}_1\bar{p}_3)(\hat{v}(\bar{p}_1) - \hat{v}(\bar{p}_2))(\hat{v}(\bar{p}_2) - \hat{v}(\bar{p}_3))]\bar{I}. \tag{73}$$

Hence, it suffices to show that the left hand side of (72) is larger than (73), or, equivalently that

$$(\bar{p}_1 - \bar{p}_3)[\hat{v}(\bar{p}_1) - \hat{v}(\bar{p}_3)] \geq (\bar{p}_1 + \bar{p}_3 - 2\bar{p}_1\bar{p}_3)\bar{I}. \tag{74}$$

Because $\theta \in [\theta^l, \theta^h]$ implies that $\pi_2 = [\theta\bar{\pi}_2^h + (1 - \theta)\bar{\pi}_2^l] \geq 0$ for $\theta \in [\theta^l, \theta^h]$, the numerator in the fraction for π_2 in (43) must be positive. That is, it holds

$$\bar{I} \leq \frac{(\bar{p}_1 - \theta)(\theta - \bar{p}_3)}{(1 - \theta)\theta} \frac{\hat{v}(\bar{p}_1) - \hat{v}(\bar{p}_3)}{\bar{p}_1 - \bar{p}_3}.$$

Hence, inequality (74) holds if

$$(\bar{p}_1 - \bar{p}_3)^2 \geq (\bar{p}_1 + \bar{p}_3 - 2\bar{p}_1\bar{p}_3)\frac{(\bar{p}_1 - \theta)(\theta - \bar{p}_3)}{(1-\theta)\theta}. \tag{75}$$

Because

$$\max_{\theta} \frac{(\bar{p}_1 - \theta)(\theta - \bar{p}_3)}{(1-\theta)\theta} = \bar{p}_1 + \bar{p}_3 - 2\bar{p}_1\bar{p}_3 - 2\sqrt{\bar{p}_1\bar{p}_3(1-\bar{p}_1)(1-\bar{p}_3)},$$

it follows that inequality (75) holds if

$$(\bar{p}_1 - \bar{p}_3)^2 \geq (\bar{p}_1 + \bar{p}_3 - 2\bar{p}_1\bar{p}_3)(\bar{p}_1 + \bar{p}_3 - 2\bar{p}_1\bar{p}_3 - 2\sqrt{\bar{p}_1\bar{p}_3(1-\bar{p}_1)(1-\bar{p}_3)}). \tag{76}$$

Using $\alpha = \bar{p}_1(1-\bar{p}_3)$ and $\beta = \bar{p}_3(1-\bar{p}_1)$, (76) rewrites as

$$(\alpha - \beta)^2 \geq (\alpha + \beta)(\alpha + \beta - 2\sqrt{\alpha\beta}), \tag{77}$$

where it follows

$$(\alpha - \beta)^2 \geq (\alpha + \beta)(\alpha + \beta - 2\sqrt{\alpha\beta}) \Leftrightarrow (\alpha + \beta)^2 \geq 4\alpha\beta \Leftrightarrow (\alpha - \beta)^2 \geq 0. \tag{78}$$

As the latter inequality is true, we conclude that also inequality (72) holds so that $\check{S}$ is convex.

Hence, if $\hat{\Theta} \subset [\theta^l, \theta^h]$, the 3-certificate menu is not optimal for the certifier, because the convexity of $\check{S}$ implies that we can raise the certifier's profit by separating all types in $\hat{\Theta}$, while keeping the three posteriors for each of these types at $\bar{p}_1$, $\bar{p}_2$, $\bar{p}_3$.

We next argue that, if $\hat{\Theta} \not\subset [\theta^l, \theta^h]$, the 3-certificate menu also cannot be optimal. To show this, we derive a convex function $\check{\check{S}}(\theta)$ that is defined for the entire interval Θ by combining the construction of $\check{S}(\theta)$ in this proof with the construction of $\check{S}(\theta)$ as in the proof of Lemma 8.

To develop this argument, first note that $\check{S}$ is defined only on the domain $[\theta^l, \theta^h]$, because for $\theta \notin [\theta^l, \theta^h]$ our construction implies that $\pi_i^\omega(\theta) \notin [0, 1]$ for some state $\omega \in \{l, h\}$ and certificate $c \in \{\hat{c}_1, \hat{c}_2, \hat{c}_3\}$.

In particular, we have at θ^l that $\pi_i^\omega(\theta^l) = 0$ for some state $\omega \in \{l, h\}$. This implies that at θ^l, we effectively have a two-certificate menu that induces two of the three posteriors $\bar{p}_1$, $\bar{p}_2$, $\bar{p}_3$. Hence, following the proof of Lemma 8, we can construct a pair of mappings $(\check{p}_l(\theta), \check{q}_l(\theta))$ satisfying the differential equation (53) with $(\check{p}_l(\theta^l), \check{q}_l(\theta^l))$ corresponding to the only two posteriors among $\bar{p}_1$, $\bar{p}_2$, $\bar{p}_3$ that occur with positive probability at θ^l. This then gives rise to a convex function $\check{S}_l(\theta)$ as defined in (50).

Likewise, we can construct $(\check{p}_h(\theta), \check{q}_h(\theta))$ with $(\check{p}_h(\theta^h), \check{q}_h(\theta^h))$ corresponding to the two posteriors among $\bar{p}_1$, $\bar{p}_2$, $\bar{p}_3$ that still occur with positive probability at θ^h. This yields a function $\check{S}_h(\theta)$ that is defined for all θ in Θ.

Thus, we obtain three convex functions $\check{S}(\theta)$, $\check{S}_l(\theta)$, and $\check{S}_h(\theta)$. The function $\check{S}(\theta)$ is supported by the three (fixed) posteriors $\bar{p}_1, \bar{p}_2, \bar{p}_3$ and has the limited domain $[\theta^l, \theta^h]$. The convex function $\check{S}_l(\theta)$ is defined for all Θ, satisfies $\check{S}(\theta^l) = \check{S}_l(\theta^l)$ and is supported by the two posteriors $\check{p}_l(\theta)$, $\check{q}_l(\theta)$. Likewise, the convex function $\check{S}_h(\theta)$ is also defined for all Θ, satisfies $\check{S}(\theta^h) = \check{S}_h(\theta^h)$ and is supported by the two posteriors $\check{p}_h(\theta)$, $\check{q}_h(\theta)$.

Define

$$\check{\check{S}}(\theta) \equiv \begin{cases} \max\{\check{S}(\theta), \check{S}_l(\theta), \check{S}_h(\theta)\} & \text{if } \theta \in [\theta^l, \theta^h] \\ \max\{\check{S}_l(\theta), \check{S}_h(\theta)\} & \text{otherwise.} \end{cases}$$

Being the maximum over convex functions, $\check{\check{S}}$ is convex. Hence, if $\hat{\Theta} \not\subset [\theta^l, \theta^h]$, the 3-certificate menu is Pareto dominated, because the convexity of $\check{\check{S}}$ implies that we can raise the certifier's profit by separating all types in $\hat{\Theta}$ using either two or three certificates, without changing the buyer's utility and the information rents of the seller types. $\square$

Appendix B: No hard information with $A(p,q) > 0$

Our analysis of the uniform-quadratic example and its extension to power functions with an exponent of $\alpha \leq 2$ yielded the extreme result that if the social value of information is not too high, the monopolistic certifier does not disclose any hard information. To show that this is not specific to the case in which $\hat{A}(\theta)$ is increasing, we next study a class of models for which $\hat{A}(\theta)$ is decreasing. An analytically convenient class is the class of convex logarithmic revenue functions of the form

$$\hat{v}(\theta|\alpha) = \frac{\ln(1-\alpha\theta)}{\ln(1-\alpha)},$$

defined for $\theta \in [0,1]$ and parameterized by $\alpha \in (0,1)$. This class provides a tractable parameterization of value functions for which the optimal maximal-screening menu is a 2-certificate menu with $p^*(\theta) = 1$ and $q^*(\theta) \leq \theta$.

To see this, note that

$$\hat{A}(\theta) = \sqrt{\frac{1-\alpha\theta}{-\alpha\ln(1-\alpha)}} \Rightarrow \hat{A}'(\theta) = -\frac{1}{2}\sqrt{\frac{\alpha}{-(1-\alpha\theta)\ln(1-\alpha)}} < 0,$$

so that Lemma 5 implies $A(p,q) > 0$ for all $0 \leq q \leq p \leq 1$. Corollary 1 then confirms that the optimal maximal-screening menu is indeed a 2-certificate menu with $p^*(\theta) = 1$ and $q^*(\theta) \leq \theta$.

We next show that, also for this case, the optimal certification menu for such value functions does not provide any hard information if the value function is not too convex. In particular, we show this for $\alpha \leq 2/3$ under a uniform distribution of types, $F(\theta) = \theta$.

Indeed, using (12) and (14), it follows

$$I(1,q|\theta) = \frac{(\theta-q)[\ln(1-\alpha)-\ln(1-\alpha q)]}{(1-q)\theta\ln(1-\alpha)} \text{ and } S(1,q) = \frac{\theta-q}{1-q} + \frac{(1-\theta)\ln(1-\alpha q)}{(1-q)\ln(1-\alpha)}.$$

With a uniform distribution, a 2-certificate maximal-screening menu that induces the two posteriors $p(\theta) = 1$ and $q(\theta) \leq \theta$ for a seller type θ, therefore yields the certifier:

$$\int_0^1 \underbrace{\frac{(\theta-q)(2\theta-1)}{(1-q)\theta} + \frac{(2\theta-q)(1-\theta)\ln(1-\alpha q)}{(1-q)\theta\ln(1-\alpha)}}_{V(q|\theta)} d\theta.$$

Hence, for each θ, the optimal $q^*(\theta)$ maximises $V(q|\theta)$ over the range $q \in [0,\theta]$.

We next show that for $\alpha \leq 2/3$, the solution is $q(\theta) = \theta$ for all $\theta \in [0,1]$, implying that the optimal certification menu does not provide any hard information.

To demonstrate this, we take the derivative with respect to q to obtain:

$$V'(q|\theta) = \underbrace{\frac{(1-\theta)}{(1-q)^2(1-\alpha q)\theta\ln(1-\alpha)}}_{-}[\underbrace{(1-\alpha q)(1-2\theta)(\ln(1-\alpha)-\ln(1-\alpha q)) - \alpha(1-q)(2\theta-q)}_{B}],$$

where the fraction is negative so that $V(q|\theta)$ is increasing in q if the expression in the squared bracket, B, is negative.

We claim that $\alpha \leq 2/3$ ensures that B is negative. To see this, note first that B is linear in θ. Evaluating B at $\theta = 0$ yields

$$B_0 = B|_{\theta=0} = (1-\alpha q)(\ln(1-\alpha) - \ln(1-\alpha q)) + \alpha q(1-q),$$

which is increasing in q as

$$\frac{\partial B_0}{\partial q} = \alpha[2(1-q) + \ln(1-\alpha q) - \ln(1-\alpha)] \geq 0.$$

Hence

$$B_0 \leq B_0|_{q=1} = 0.$$

It remains to show that B evaluated at $\theta = 1$,

$$B_1 = B|_{\theta=1} = -\alpha(1-q)(2-q) - (1-\alpha q)(\ln(1-\alpha) - \ln(1-\alpha q)),$$

is also non-positive for $\alpha \leq 2/3$. To see this, note that

$$\frac{\partial^2 B_1}{\partial q^2} = -2\alpha + \frac{\alpha^2}{1-\alpha q} \leq -2\alpha + \frac{\alpha^2}{1-\alpha} = \alpha\frac{3\alpha - 2}{1-\alpha}.$$

As the final expression is less than or equal to 0 for $\alpha \leq 2/3$, the expression B_1 is concave in q for $\alpha \leq 2/3$. Noting that for $q = 1$, it follows that the first-order condition

$$\left.\frac{\partial B_1}{\partial q}\right|_{q=1} = \alpha(2 - 2q + \ln(1-\alpha) - \ln(1-\alpha q))|_{q=1} = 0$$

is satisfied, concavity of B_1 in q implies that B_1 attains a maximum at $q = 1$ if

$$B_1|_{q=1} = 0.$$

Hence, for $\alpha \leq 2/3$, the overall expression B is non-positive, implying that $V(q|\theta)$ is increasing for the range $[0, \theta]$ so that $V(q|\theta)$ is maximised at the corner solution $q = \theta$.

References

AKERLOF, G. A. (1970), "The Market for Lemons: Quality Uncertainty and the Market Mechanism", *Quarterly Journal of Economics*, **84** (3), 488–500.

ALI, N., HAGHPANAH, N., LIN, X. and SIEGEL, R. (2022), "How to Sell Hard Information", *Quarterly Journal of Economics*, **137** (1), 619–678.

ASSEYER, A. and WEKSLER, R. (2024), "Certification Design with Common Values", *Econometrica*, **92** (3), 651–686.

BAR-ISAAC, H. and SHAPIRO, J. (2013), "Ratings quality over the business cycle", *Journal of Financial Economics*, **108**, 62–78.

BERGEMANN, D., BONATTI, A. and SMOLIN, A. (2018), "The Design and Price of Information", *American Economic Review*, **108**, 1–48.

BIZZOTTO, J., RÜDIGER, J. and VIGIER, A. (2020), "Testing, disclosure and approval", *Journal of Economic Theory*, **187**, 105002.

BOLTON, P., FREIXAS, X. and SHAPIRO, J. (2012), "The Credit Ratings Game", *The Journal of Finance*, **67**, 85–111.

CHOPRA, H. and ELY, J. (2025), "Incentive-Compatible Information Design" (mimeo, Northwestern University).

CORRAO, R. (2024), "Mediation Markets: The Case of Soft Information" (mimeo, MIT).

DRANOVE, D. and JIN, G. Z. (2010), "Quality Disclosure and Certification: Theory and Practice", *Journal of Economic Literature*, **48**, 935–963.

EVANS, R. and PARK, I.-U. (2024), "Selling Information for Bilateral Trade" (mimeo, University of Bristol).

FUDENBERG, D. and TIROLE, J. (1991), "Perfect Bayesian and Sequential Equilibrium", *Journal of Economic Theory*, **53**, 236–260.

HANCART, N. (2024), "The Optimal Menu of Tests" (mimeo, University of Oslo).

HEALY, P. and PALEPU, K. (2003), "The Fall of Enron", *Journal of Economic Perspectives*, **17**, 3–26.

ICHIHASHI, S. and SMOLIN, A. (2023), "Data Provision to an Informed Seller" (mimeo, Queen's University).

KAMENICA, E. and GENTZKOW, M. (2011), "Bayesian Persuasion", *American Economic Review*, **101** (6), 2590–2615.

KARTIK, N., LEE, F. and SUEN, W. (2021), "Information validates the prior: A theorem on Bayesian updating and applications", *American Economic Review: Insights*, **3** (2), 165–182.

KOESSLER, F. and RENAULT, R. (2012), "When does a firm disclose product information?", *RAND Journal of Economics*, **43** (4), 630–649.

LIZZERI, A. (1999), "Information Revelation and Certification Intermediaries", *The RAND Journal of Economics*, **30** (2), 214–231.

MÄKIMATTILA, M., SHANG, Y. and SHIRAKAW, R. (2025), "The Design and Price of Certification" (mimeo, MIT).

OKUNO-FUJIWARA, A., POSTLEWAITE, M. and SUZUMURA, K. (1990), "Strategic Information Revelation", *Review of Economic Studies*, **57**, 25–47.

ROSAR, F. (2017), "Test design under voluntary participation", *Games and Economic Behavior*, **104**, 632–655.

SEIDMANN, D. J. and WINTER, E. (1997), "Strategic Information Transmission with Verifiable Messages", *Econometrica*, **65**, 163–169.

VISCUSI, W. (1978), "A Note on 'Lemons' Markets with Quality Certification", *The Bell Journal of Economics*, **9**, 277–279.

WEKSLER, R. and ZIK, B. (2023), "Disclosure in Markets for Ratings", *American Economic Journal: Microeconomics*, **15** (3), 501–526.

WEKSLER, R. and ZIK, B. (2024), "Monopolistic Screening in Certification Markets" (mimeo, Hebrew University of Jerusalem).

.