Optimal Rating Design under Moral Hazard*

Maryam Saeedi[†] Ali Shourideh[‡]

January 15, 2026

Abstract

We study optimal rating design under moral hazard and strategic manipulation. An intermediary observes a noisy indicator of effort and commits to a rating policy that shapes market beliefs and pay. We characterize optimal ratings via concavification of a *gain function*. Optimal ratings depends on interaction of effort and risk: for activities that raise tail risk, optimal ratings exhibit *lower censorship*, pooling poor outcomes to insure and encourage risk-taking; for activities that reduce tail risk, *upper censorship* increases penalties for negligence. In multi-task environments with window dressing, *less informative* ratings deter manipulation. In redistributive test design, optimal tests exhibit *mid censorship*.

Keywords: Information Design, Moral Hazard, Window Dressing, Manipulation, Majorization, Censorship **JEL codes:** D82, D86, D83, L15, L86, D63

*This paper was previously circulated under the title "Optimal Rating Design". We thank Nageeb Ali, Ian Ball, James Best, Aislinn Bohren, Odilon Camara, Joyee Deb, Rahul Deb, Maryam Farboodi, Hugo Hopenhayn, Emir Kamenica, Navin Kartik, Alexey Kushnir, Stephen Morris, Paula Onuchic, Jacopo Perego, Ilya Segal, Vasiliki Skreta and Alex Wolitzky as well as numerous seminar and conference participants for their helpful comments.

[†]Carnegie Mellon University, Email: msaeedi@andrew.cmu.edu

[‡]Carnegie Mellon University, Email: ashourid@andrew.cmu.edu

1 Introduction

Many markets rely on information disclosure or ratings to facilitate trade and incentivize quality provision. ESG rating agencies aim to incentivize companies to improve their environmental and social impact. Online platforms such as Amazon, Airbnb, Upwork, and eBay design reputation systems to incentivize and signal providers' quality. Standardized tests communicate student ability to universities. In each case, an intermediary observes signals about agent behavior and must decide how to convey this information to a market that rewards agents based on perceived quality. A fundamental challenge arises: agents strategically respond to rating policies, and the information disclosed shapes the incentives. Additionally, these incentives can lead to *window-dressing* activities to *manipulate* the ratings.

Despite the ubiquity of these systems in markets suffering from moral hazard, several core theoretical questions are yet to be answered: What are the fundamental trade-offs in designing rating systems when participants can anticipate and react to them? How should an intermediary design information structures to account for window-dressing incentives? This paper answers these questions by developing a theoretical framework for optimal rating design under moral hazard, utilizing a variant of the career concerns model introduced by Holmström (1999).

Our model features an agent (e.g., a company seeking an ESG rating or a seller on eBay) who takes costly actions that create value for a competitive market. These actions also generate a noisy indicator observed by an intermediary (e.g., an ESG rating firm or an online platform) which must then decide how to convey this information via a rating. The market, in turn, pays the agent its expected value based on the signal and its belief about the agent's action.

Our primary objective is to find the optimal information structure to maximize a flexible welfare function for the intermediary. This function can target a particular action or a particular distribution of payoffs for the agent, capturing environments where market values do not fully internalize the social value of actions—such as the positive externalities of ESG activities—or where a platform has concerns for fairness or redistribution.[1]

Our paper consists of two parts: In the first part, we provide the theoretical foundations of our analysis, and in the second part, we provide general properties of optimal ratings and apply the method to a few practical applications. The main technical challenge in formulating the optimal rating problem is the interplay between information structures and the agent incentive constraints. We introduce the concept of *interim prices*—the agent's interim expectation of market's

[1] Since Holmström (1999), it is well known that the *implicit incentives* provided by career concerns do not necessarily lead to efficient effort levels because they fail to fully internalize the social benefit of the agent's action. This externality is also present in our model and the intermediary's objective can be thought of as addressing such externalities.

posteriors—as the sufficient statistic that determines incentives from the agent's perspective.[2] This object, which can be interpreted as the agent's second-order belief, allows us to transform the problem of choosing an information structure into a tractable mechanism design problem. In Proposition 1, we show that when these interim prices are comonotone with market values, then a price schedule can be implemented by some rating if and only if they are a mean-preserving contraction of the market values. This implies that we can cast the problem of rating design as a standard moral hazard problem with transfers subject to a majorization constraint.

With this result in hand, in Theorem 1, we show that optimal ratings can be characterized through concavification of a *gain function* in the quantile space. This gain function depends both on the distributional motives of the intermediary and the local effect of the agent's action on the quantile distribution of the indicator (holding fixed market beliefs). The concavification approach provides a sharp characterization: regions where the gain function coincides with its concave envelope correspond to full information disclosure, while regions where concavification requires linear interpolation correspond to pooling. Thus, the optimal information structure is a deterministic monotone partition of the indicator space. Intuitively, the intermediary either fully reveals the indicator on some regions or pools contiguous intervals—a sharp foundation for the prevalence of coarse, threshold-based rating systems.

The second part of the paper applies this framework to derive general properties of optimal ratings and how they depend on the agent's technology and the intermediary's objective. When the intermediary's objective is to maximize effort (absent distributional motives), the design problem reduces to finding the rating that achieves the highest level of effort. Under the canonical Monotone Likelihood Ratio Property (MLRP) assumption in the moral hazard literature, we show that the gain function in this case is concave and as a result full information disclosure is optimal.

However, many economically relevant activities violate MLRP in systematic ways. Innovative activities often increase both upside potential and downside risk, i.e., R&D effort can lead to breakthroughs or failures. Conversely, maintenance activities typically reduce variance through more consistent outcomes. To capture these patterns, we introduce two new distributional properties: the *expanding likelihood ratio property* (ELRP), where increased effort expands the distribution's tails, and the *compressing likelihood ratio property* (CLRP), where increased effort compresses outcomes toward the center. Under ELRP, optimal ratings take the form of lower censorship providing insurance against downside risk to encourage risk-taking. Under CLRP, optimal ratings are upper censorship, pooling high realizations while revealing low ones, which punishes poor outcomes and encourages variance-reducing effort.

We also characterize how distributional concerns interact with incentive provision. When intermediary places higher weights on lower realizations of the indicator, either because of fairness

[2] See also Doval and Smolin (2024).

concerns or redistributive objectives, the gain function may become non-concave even under MLRP at low quantiles. Thus leading to optimality of lower-censorship ratings. This creates a fundamental tension between maximizing effort and protecting agents from downside risk.

Finally, in Section 5, we use these insights to study a model of multi-tasking à la Holmström and Milgrom (1991) and redistributive test design. In the multi-task model, the agent allocates efforts across productive tasks and window-dressing ones (actions that boost the indicator more than market values) which differentially impact the observed indicator and market value. With normal additive noise and a convex cost, this model is reducible to a single action model and thus the results from Section 4 apply.[3]

Since the additive normal model satisfies MLRP, fully revealing ratings implement the highest level of effort. However, when window dressing makes those actions welfare-reducing, optimal policy involves withholding information to temper manipulation incentives. Moreover, similar to Holmström and Milgrom (1991), a decline in the cost of window dressing leads to further reductions in informativeness. We also study a nonreducible two-task example and show that when window dressing disproportionately drives extreme indicator realizations, upper censorship can be strictly welfare-improving relative to full revelation by disproportionately discouraging manipulative effort. Finally, we apply our framework to redistributive test design with heterogeneous students, showing that optimal tests may involve "mid censorship" to balance incentive provision across student types.

Beyond its technical contributions, our analysis offers practical guidance for regulators and rating system design. As data collection has intensified, several institutions have formed around using data to incentivize behavior which in turn has created incentives for manipulation and window dressing (see for example Mayzlin et al. (2014)). Our results provide guidance on how ratings should be designed in such environments supporting the observed heterogeneity in the rating system. For example, platforms have adopted a variety of ratings: some platforms (such as Shipt or Instacart) allow for low rating forgiveness and fresh start which could be interpreted as lower censorship; others such as Airbnb have too many high ratings (see for example Zervas et al. (2021)) that can be interpreted as upper censorship. Our results suggest these differences may reflect optimal responses to underlying differences in how effort affects outcome distributions. More broadly, the paper provides a toolkit for evaluating rating policies across domains—from ESG certification to educational testing to online marketplaces—by connecting observable features of agent technology to the optimal structure of information disclosure.

[3] Mathematically, this is equivalent to the set of equilibrium actions for arbitrary ratings having dimension one.

1.1 Related Literature

Our paper is related to a few strands of the literature in information economics and mechanism design. It is closely related to a recent literature that studies information design when strategic behavior affects the state by the choice of the information structure (e.g., Frankel and Kartik (2019), Ball (2025), and Perez-Richet and Skreta (2022)). In contrast with Ball (2025) and Frankel and Kartik (2019), our mathematical result on second-order expectations allows us to study a larger class of problems without any restrictions on information structures. Our analysis, thus, identifies both the precise shape of the optimal information structure and when it is optimal to use uncertain rating systems. In our model, presence of window–dressing incentives is similar to the falsification model in Perez-Richet and Skreta (2022). The main difference with our setting is the existence of noise in the ability of the agent to manipulate the signal observed by the intermediary.

A related paper to ours is Boleslavsky and Kim (2020). They study a model of Bayesian persuasion with moral hazard, similar to ours, in which an agent chooses an effort level that affects the distribution of the state, and a sender affects a receiver's action using an information structure. The papers differ in terms of focus and technique. We focus on a career concern model where the information structure only affects the agent's incentive. Additionally, we use majorization tools which allows us to work with larger state spaces.[4]

From a technical perspective, our results are related to the new literature in information economics that uses optimization under majorization constraints; Kleiner et al. (2021). Their solution method uses the characterization of extreme points of the set of monotone functions that majorizes a certain function. Similarly, Bergemann et al. (2022a) and Bergemann et al. (2022b) use the same strategy as our work to cast the problem in terms of quantiles and use concavification to derive optimal mechanisms. While their focus is on screening models with hidden information, ours is closer to classic moral hazard.

Our paper is also related to the literature concerned with the problem of certification and its interactions with moral hazard: Albano and Lizzeri (2001), Zubrickas (2015), Onuchic and Ray (2023), and Zapechelnyuk (2020). A notable contribution is that of Albano and Lizzeri (2001), where the key assumption that the intermediary can charge an arbitrary fee schedule leads to an indeterminacy between using transfers and ratings to implement desired outcomes. Zubrickas (2015), Zapechelnyuk (2020), and Onuchic and Ray (2023) also study related problems, but they focus on *deterministic* technologies where the agent's effort deterministically translates into values for the market. In contrast and in our model, the presence of noise allows us to disentangle

[4]Relatedly, a recent paper by Madsen et al. (2025) studies a moral hazard model with non–monetary incentives. Our paper is related to their work to the extent that our agent is incentivized using ratings; a non–monetary instrument.

the indicator from market values which in turn leads to an inefficient level of effort under full information and enables us to study window dressing and manipulation incentives. Relatedly and in the context of team production, Halac et al. (2021) show that uncertainty about a worker's compensation ranking in a team can remove low effort equilibria. Our result on how censorship for some technologies can improve incentives can be regarded as the single agent version of their result. [5]

Finally, our paper complements the empirical literature on certification and disclosure in markets with asymmetric information, such as online platforms (e.g., Hui et al. (2023) and Nosko and Tadelis (2015)), health insurance markets (Vatter (2025)), food labeling (Barahona et al. (2023)), and ESG investing (Berg et al. (2022)). We contribute to this literature by developing theoretical methods and general lessons for the optimal design of rating systems.

The remainder of the paper proceeds as follows. Section 2 presents the model and introduces interim prices as the key analytical object. Section 3 derives our main characterization result through concavification of the gain function. Section 4 establishes general properties of optimal ratings under alternative distributional assumptions, including MLRP, ELRP, CLRP, and redistributive motives. Section 5 applies our framework to multi-task moral hazard with window dressing and redistributive test design. Proofs are relegated to the Appendix.

2 A Model of Moral Hazard

In this section, we describe our basic model of rating design and provide some preliminary analysis of the restrictions implied by the fact that incentives are provided through ratings.

We are interested in settings in which an intermediary observes some information about an agent's chosen actions and decides how to convey this information to a competitive market, henceforth "the market," who then pays its posterior mean as a price to the agent.

More specifically, the agent exerts an effort vector $a \in A \subset \mathbb{R}^N$ at a cost $c(a)$. This action generates a random outcome $(v, y) \in \mathbb{R}^2$, where v represents the value of the output to the market and y is a noisy *indicator* observed by the intermediary. We denote the cumulative distribution function of the indicator y conditional on action a by $G(y|a)$.

The market consists of competitive buyers who value the agent's output at v. However, the market observes neither the true value v nor the agent's action a directly. Instead, it forms expectations based on information provided by the intermediary. If the market observes the indicator

[5] Ali et al. (2022) study a model with adverse selection (i.e., exogenous state), where optimal disclosure involves uncertainty, but it is a way of uniquely implementing an intermediary's desirable outcome.

y and holds a belief $\hat{a}$ regarding the agent's action, the expected value of the output is given by:

$$\overline{v}(y;\hat{a}) = \mathbb{E}[v \mid y, a = \hat{a}]$$

We refer to $\overline{v}(y;\hat{a})$ as *market values*, i.e., the most informative assessment of the valuation of the market. Throughout the paper we impose the following monotonicity assumption:

Assumption 1. *For all market beliefs $\hat{a}$, the market value $\overline{v}(y;\hat{a})$ is increasing in the indicator y.*

This assumption states that market values are ranked based on the values of the indicator. Without any other assumption on the distribution function $G(y|a)$, e.g., increasing in FOSD, MLRP, this assumption is innocuous as one can always relabel the values of the indicator according to the market values. While our main characterization results – Proposition 1 and Theorem 1 – hold without Assumption 1, we maintain this assumption for tractability and convenience.

The intermediary observes the indicator y (at no cost) and controls the information observed by the market. Specifically, the intermediary commits to an information structure $(S, \pi(\cdot|y))$, where S is a set of signal realizations and $\pi(\cdot|y) \in \Delta(S)$ is the distribution over signals conditional on realization of y. Having observed s, the market pays its expected payoff $\mathbb{E}[v|s]$ to the agent.[6] This expectation is calculated using the information available, s, and the common belief about equilibrium play.[7]

The timing of the model is as follows. First, the intermediary chooses and commits to an information structure $(S, \pi(\cdot|y))$. Subsequently observing the intermediary's choice, the agent chooses her action, a, which in turn generates a realization of indicator y for the intermediary. The intermediary then draws a rating s according to $\pi(\cdot|y)$ and sends it to the market. Finally, the market observes s, updates its beliefs and pays the agent $\mathbb{E}[v|s]$. Figure 1 depicts the structure of the model and actions.

[6] We assume that the buyers are on the long side of the market, thus willing to pay their expected value. Our analysis remains unchanged if the market keeps a constant fraction of their expected value.

[7] An information structure is a family of probability spaces $\{(S, \mathscr{S}, \pi(\cdot|y))\}_{y\in Y}$, where S is the space of signal realizations and $\mathscr{S}$ is a σ-algebra. Throughout the paper, we work with S as a compact subset of some Euclidean space, and $\mathscr{S}$ as the Borel σ-algebra associated with topology induced by the Euclidean norm and a compact space for S. Henceforth, we drop references to σ-algebra in our analysis. Additionally, when describing subsets, we refer to Borel subsets.

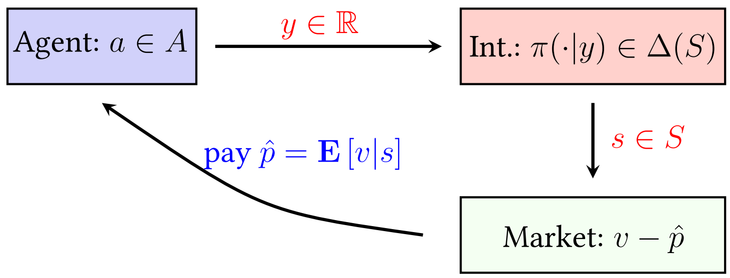


Figure 1: General structure of the model

Given an information structure $(S, \pi(\cdot|y))$ and action a, the agent's expected payoff is given by

$$\int_Y \int_S \mathbb{E}[v|s]\, d\pi(s|y)\, dG(y|a) - c(a). \tag{1}$$

In equilibrium, the agent chooses a to maximize (1).

The ex post market price $\mathbb{E}[v|s]$ depends on the information structure (S, π) and also on the market's prior about the distribution of (a, y), which depends on the agent's equilibrium strategy . More specifically, the market uses its beliefs about the equilibrium strategy of the agent a to form a prior $G(y|a)$ and uses Bayes' rule to form the posterior expectation $\mathbb{E}[v|s]$ satisfying

$$\int_Y \int_{S'} \mathbb{E}[v|s]\, d\pi(s|y)\, dG(y|a) = \int_Y \overline{v}(y;a)\, \pi(S'|y)\, dG(y|a), \forall S' \subset S. \tag{2}$$

The above defines a Bayesian Nash equilibrium given the information structure (S, π). More specifically, given an information structure (S, π), an equilibrium is an effort a together with market beliefs $\mathbb{E}[v|s]$ such that a maximizes expression (1), and given a, the market beliefs satisfy Bayesian updating as defined in equation (2).[8]

Examples

To clarify the scope and applicability of our analysis, we now describe several environments that fit the model above.

1. **Reputation Mechanisms in Online Platforms:** Online platforms face challenges in designing their reputation systems because of moral hazard. These platforms have access to performance data about providers (i.e., hosts on Airbnb, sellers on eBay, and freelancers on Upwork) not available to the market.[9] The platform's certification policy, such as Airbnb's

[8]We have focused on equilibria in which the agent plays a pure effort strategy. Our main characterization results, Proposition 2 and Theorem 1, hold when allowing for mixed effort strategy by the agent.

[9]As documented by Saeedi (2019), Hui et al. (2016), and Nosko and Tadelis (2015), there are many performance indicators available to eBay that are not conveyed to the market directly, such as total quantities sold, and previous claims and their outcomes.

Superhost, eBay's Top Rated Seller or Upwork's Talent Badge, is based on performance measures and they can be regarded as the information structure in our model. According to Hui et al. (2023) among others, the changes in such policies influence provider behavior. Our model examines the resulting issues and trade-offs for both platform and providers.

2. **Manipulation and Window Dressing:** Rating systems frequently incentivize agents to manipulate signals or engage in "window dressing"—costly actions that inflate observed indicators without enhancing fundamental values.[10] Online platforms are frequently plagued by data manipulation by providers.[11] For example, some third-party sellers on Amazon pay customers for positive reviews and higher ratings, He et al. (2022). In our model, this can be captured by letting the agent take costly actions to increase the observed indicator y without affecting market valuation v. This creates a trade-off in rating design: information provision incentivizes productive actions but simultaneously raises the incentives for window dressing. In Sections 5.1 and 5.2, we develop a multi-tasking model à la Holmström and Milgrom (1991) to describe how the presence of window-dressing motives affects the optimal design of ratings.

3. **Career Concerns and Externalities:** Since Holmström (1999)'s seminal model of career concerns, it has been known that in absence of long-term contracts and when agents (i.e., CEOs or government workers) care about their careers, they exert inefficient levels of effort.[12] In our framework, this occurs when market values $\overline{v}\,(y;a)$ change with a. Since the agent does not account for the effect of her effort on market values, equilibrium is inefficient. It is thus natural to ask whether ratings can be used to possibly reduce such inefficiencies. As we will show, our main characterization result can be used to shed light on this question. Specifically, we show that under the often used MLRP condition (Monotone Likelihood Ratio Property), perfect information implements the highest possible value of effort.[13] We also identify properties of the indicators distribution, $G\,(y|a)$, under which censoring parts of information is beneficial giving rise to non-trivial rating policies.

[10] In recent years, several lawsuits have involved rating manipulation in different industries, such as education (e.g., the case of Temple University, Temple Business School Dean Fraud, and the case of Columbia University in NYT on Columbia's ranking manipulation and Michael Thadeuss on ranking manipulation) and financial markets (e.g., the case of Greenwashing by Deutsche Bank). Along the same lines, Agarwal et al. (2018) show that greater transparency leads to fund managers' forgoing long term profits and short-termism.

[11] Feedback manipulation has long been a debated issue on e-commerce platforms (e.g., Hui et al. (2018)).

[12] See also Prat (2005) for highlighting situations in which information about actions can lead to conformism by the agent and as a result, inefficient outcome.

[13] Relatedly, Dewatripont et al. (1999) show that under MLRP, it is always optimal to use all the information available.

2.1 Interim Prices: Definition and Characterization

In this section, we introduce a mathematical object, *interim prices*, that allows us to simplify the problem of rating design in the environment described above.

The notion of interim price is simple. This mathematical object determines the agent's incentives in choice of effort and will be present in the incentive constraints for the agent. Specifically, we define *interim prices* as

$$p(y) = \int \mathbb{E}[v|s]\, d\pi(s|y). \tag{3}$$

In words, p is the expected payment to the agent conditional on the indicator y, integrating over possible signals s given the rating system. Additionally, it is an equilibrium object as it depends on $\mathbb{E}[v|s]$ which depends on the market's beliefs about the agent's action profile. It can also be interpreted as the agent's "second-order belief": their beliefs about the beliefs of the market on values.

Critically, it is a sufficient statistic for the information structure from the agent's perspective. Specifically, for any choice of a, the agent's payoff is given by

$$\int p(y)\, dG(y|a) - c(a).$$

Thus, the problem of designing an optimal rating system is isomorphic to the problem of choosing an interim price schedule p, subject to the constraint that p must be implementable via some information structure (S, π). Thus, we need to characterize the set of feasible interim prices.

Generally, there are no simple conditions to characterize the set of interim price profiles that result from a particular information structure and action profiles. However, as we will show next, under some restriction on information structures, a simple characterization exists.

To understand the notion of interim prices, recall that market values are given by $\overline{v}(y;a) = \mathbb{E}[v|y]$. These are the interim prices associated with a fully revealing information structure, i.e., the most informative information structure. Now, from the perspective of the market, $\mathbb{E}[v|s]$, is a garbling of $\overline{v}(y;a)$. Similarly, from the perspective of an observer that sees the realization of y, interim price p is a garbling of $\mathbb{E}[v|s]$ and thus itself a garbling of $\overline{v}$. In other words, if we view them as random variables, we must have $p \succeq_{\text{cv}} \overline{v}$ or equivalently p is mean preserving contraction of $\overline{v}$.[14]

Given the results in the literature – see for example Rothschild and Stiglitz (1970) or Gentzkow and Kamenica (2016) – it is tempting to suggest that the reverse of the above observation is also

[14] The relation $p \succeq_{\text{cv}} \overline{v}$ represents the concave order which implies that for all concave functions $\phi : \mathbb{R} \to \mathbb{R}$, $\mathbb{E}[\phi(p)] \geq \mathbb{E}[\phi(\overline{v})]$. Since Bayes plausibility implies $\mathbb{E}[p] = \mathbb{E}[\overline{v}]$, this definition is equivalent to majorization, second order stochastic dominance and increasing concave order – see for example Shaked and Shanthikumar (2007) section 4.A.

true: that mean preserving contraction is a sufficient condition for existence of ratings. Below we show that this is indeed true when interim prices and market values are comonotone.[15] Formally we say that p and $\overline{v}$ are comonotone if:

$$p(y) > p(y') \Rightarrow \overline{v}(y;a) > \overline{v}(y';a).$$

In words, higher prices are associated with higher market values, so the two random variables never move in opposite directions.

Proposition 1. *Suppose that p is a function that maps values of y into $\mathbb{R}$ such that*

1. *p is comonotone with $\overline{v}$, and*

2. *$p \succeq_{cv} \overline{v}$.*

Then, there exists an information structure (S, π) such that $p(y) = \int \mathbb{E}[v|s]\, d\pi(s|y)$.

The proof is a straightforward application of Kleiner et al. (2021)'s result on the extreme points of the set of monotone functions that satisfy a majorization constraint.

This proposition implies that for any arbitrary information structures with an action a and interim price function p, we can characterize the comonotone equilibria of the game as follows:

1. The action a is incentive compatible,

$$a \in \arg\max_{\hat{a} \in A} \int p(y)\, dG(y|\hat{a}) - c(\hat{a}) \tag{4}$$

2. Interim prices $p(y)$ dominate $\overline{v}(y;a) = \mathbb{E}[v|y]$ according to the concave order.

3. Interim prices and market valuations are comonotone.

This reduction allows us to transform the optimal rating design problem as a standard mechanism design problem with transfers, where the "transfers" are the interim prices constrained by the concave order.

3 Optimal Ratings: A General Characterization

In this section, we use Proposition 1 to provide our main theoretical characterization result for optimal ratings under moral hazard. In the rest of the paper, we discuss various applications and the implications of our characterization result.

[15] In the Appendix D, we provide an example that illustrates that without comonotonicity mean preserving contraction is no longer sufficient and additional conditions are needed. We also discuss its relationship with similar results in the literature.

3.1 The Intermediary's Problem

The intermediary chooses an information structure to maximize an objective that may differ from total surplus. We consider a class of objectives in the form

$$W(a) + \int p(y)\,\alpha(y)\,dG(y|a), \tag{5}$$

where $W(a)$ captures externalities or direct preferences over effort, and $\alpha(y) \geq 0$ represents distributional weights on agent payoffs. This class of objective functions fits several applications in which rating design interacts with moral hazard:

1. **Targeting an Action:** It is possible that market values do not necessarily reflect the social value of the agent's actions. This may occur for two reasons: first, direct externalities. In this case, $W(a)$ is different from the total surplus $V(a) = \mathbb{E}[v|a] - c(a)$. The difference of the two $W(a) - V(a)$ represents the external effects that are not captured by the market. Second, as discussed in Section 2, when market beliefs directly affect market values, a fully revealing equilibrium can be inefficient due to career concerns. In this case, the objective is simply total surplus or $V(a)$.

2. **Distributional Concerns:** The weights $\alpha(y)$ can be interpreted as distributional concerns. For example, in the context of platform design, platforms might aim to guarantee a minimum payoff level for sellers to maintain a minimum market size. In educational contexts, critics often argue that standardized tests create biases against lower-income students and minorities. Given such disparities in outcome distribution, a college or school with distributional concerns could reweight test outcomes for its admission policies. This reweighting can be achieved using an objective function similar to that in (5).

Given this class of objectives and the comonotonicity restriction, the problem of optimal rating design can be stated as maximizing the objective in (5) subject to incentive compatibility (4), comonotonicity and majorization.

3.2 Quantile Formulation

To simplify working with concave order constraints, we transform the interim prices and market values to their quantile formulation. This would allow us to characterize of optimal ratings via concavification of a gain function.[16]

Let $v_Q(i)$ denote the market value associated with the i-th quantile of the indicator distribution. Formally,

[16] See also Bergemann et al. (2022b) and Bergemann et al. (2022a) for a similar approach.

$$v_Q(i) = \overline{v}(G^{-1}(i|a); a). \tag{6}$$

Note that we have dropped a from the expression of quantile value for ease of exposition. Similarly, let $p_Q(i)$ be the quantile representation of the interim price. Given the comonotonicity assumption of p and $\overline{v}$, $p_Q(i)$ is the interim price associated with market value $v_Q(i)$.

Given this inversion and using integration by part, for any arbitrary integrable function $h(y)$, we can write

$$\int_Y h(y)\, p(y)\, dG(y|a) = \int_0^1 \int_{\left\{y:\overline{v}(y)>v_Q(i)\right\}} h(y)\, dG(y|a)\, dp_Q(i) = \int_0^1 H(i)\, dp_Q(i),$$

where

$$H(i) = \int_{\left\{y:\overline{v}(y)>v_Q(i)\right\}} h(y)\, dG(y|a) \tag{7}$$

collects the contribution of h over all realizations whose associated market value exceeds $v_Q(i)$. In the appendix, we use the above and the fact that $p \succcurlyeq_{\text{cv}} \overline{v}$ to prove the following lemma:

Lemma 1. *Let h be an integrable function and H be defined by (7).Let cav$H$ be the concave envelope of H, i.e., the lowest concave function dominating $H(i)$. Then*

$$\max_{\substack{p\,:\,p \succcurlyeq_{\text{cv}} \overline{v},\\ p,\overline{v}\,:\,\textit{comonotone}}} \int h(y)\, p(y)\, dG(y|a) = \int_0^1 \textit{cav}H(i)\, dv_Q(i) \tag{8}$$

Moreover, the optimal p satisfies:

1. $p(y) = \overline{v}(y;a)$ when $H(G(y|a)) = \text{cav}H(G(y|a))$.
2. If cav$H(i) > H(i)$ for all i in some maximal interval $I \subset [0,1]$, then, $p(y) = \mathbb{E}[\overline{v}|G(y|a) \in I]$.

The result of Lemma 1 is depicted in Figure 2. The function $H(i)$ is constructed by integrating $h(y)$ for values of y above a threshold, y'. This threshold is one for which $\overline{v}(y') = v_Q(i)$. When the concave envelope of $H(i)$ does not coincide with $H(i)$, optimal interim prices are constant and equal to average market value of the interval. Conversely when it coincides with $H(i)$, optimal interim prices coincide with market valuations $\overline{v}(y)$. In the proof of Lemma 2 we make use of the fact that if $p \succcurlyeq_{\text{cv}} \overline{v}$, the reverse is true for their quantiles (or c.d.f.'s). We then apply Blackwell's theorem to construct a concavification of H as the optimum.

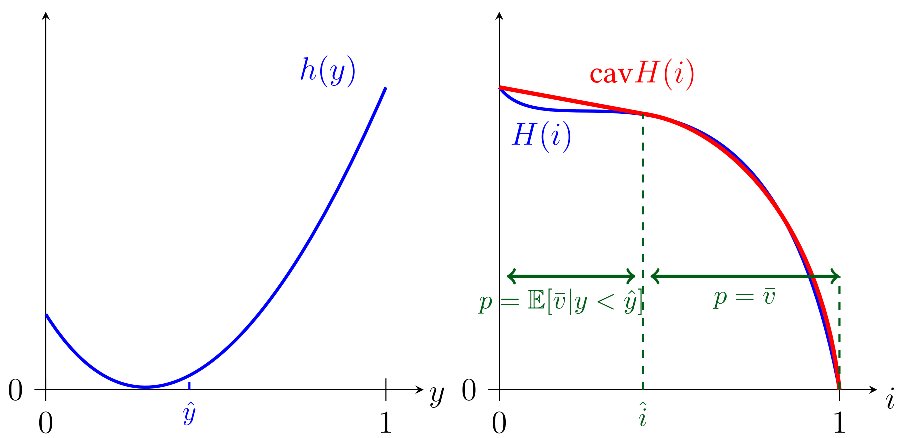


Figure 2: Concavification of H (right) and Construction of Optimal Ratings. The value $\hat{y}$ is associated with the quantile $\hat{\hat{i}}$.

3.3 Incorporating Incentives: the Main Characterization

In the above, we used a generic unconstrained objective function of the form $\int h(y)\,p(y)\,d\mu_y$. Our optimal rating design problem is a constrained optimization problem that has to respect incentive compatibility constraints (4). One can thus use standard Lagrangian arguments to transform our problem to an unconstrained optimization.

The incentive constraint can be written in quantile space by defining

$$F(i|\hat{a};a) = G\left(G^{-1}(i|a)\,|\hat{a}\right),$$

where F is the distribution over quantiles when the agent chooses $\hat{a}$ but quantiles are defined according to a. Then we can write any incentive compatibility as

$$-\int_0^1 F(i|a,a)\,dp_Q(i) - c(a) \geq -\int_0^1 F(i|\hat{a};a)\,dp_Q(i) - c(\hat{a}), \quad \forall \hat{a} \in A. \tag{9}$$

In order to sidestep many of the complications that typically arise in moral hazard problems, we will use the first order approach (FOA) throughout the paper. That is, we replace the incentive constraint (9) with its first order condition. In the Appendix C, we will use an approach similar to Chade and Swinkels (2020) to provide sufficient conditions on the distribution functions for the validity of the first order approach.

Using the IC in quantile space, FOA, and Lemma 1 we show the following result:

Theorem 1. *If w^* is the highest value of the objective (5) and under the validity of FOA, there exists* *a real vector* $\lambda \in \mathbb{R}^N$ *such that*

$$w^* = \max_a W(a) + \int \textit{cav}\Gamma(i;\lambda,a)\, dv_Q(i) \tag{D}$$
$$\textit{s.t. } i = G(\{y : \overline{v}(y;a) \le v_Q(i)\,|a\})$$

where

$$\Gamma(i;\lambda,a) = \int_{\left\{y:\overline{v}(y;a)>v_Q(i)\right\}} \alpha(y)\, dG - \sum_{n=1}^{N} \lambda_n \left[\frac{\partial}{\partial \hat{a}_n} F(i|\hat{a};a)\Big|_{\hat{a}=a} + \frac{\partial}{\partial a_n} c(a) \right]$$

We refer to Γ as the gain function. It summarizes the weight that the intermediary puts on a particular type and the associated IC conditions.

Theorem 1 implies that the rating design problem can be solved by solving a one-dimensional concavification problem of the gain function and then finding optimal values of effort and the multipliers associated with the incentive compatibility constraints (9). While its intuition is captured by the discussion above, its proof uses a notion of duality which is rather standard.

The first implication of Theorem 1 is that optimal rating systems are simple. In fact, since the function to be *concavified* is only a function of the quantile i, by Caratheodory theorem any convex combination of values of $\Gamma(i;\lambda,a)$ can be achieved by using at most two points. This logic establishes that the optimum in (D) is always achieved by a deterministic monotone partition. However, the optimum should also satisfy the incentive compatibility. To ensure that this is indeed possible, we make the following assumption on the distribution function $G(y|a)$:

Assumption 2. ***Independence.*** *For all* $a \in A$*:*

1. $G(y|a)$ *is full support over a convex subset of* $\mathbb{R}$.
2. *For any interval* $I \subset$ *Supp*$G(y|a)$*, then the function* $\alpha(y)\, g(y|a)$ *cannot be written as a non–zero linear combination of* $g(y|a)\,, \left\{\frac{\partial g(y|a)}{\partial a_n}\right\}_{n=1}^{N}$ *for all values of* $y \in I$.

The independence assumption ensures that there is enough variation in y conditional on effort a.[17] Given Assumption 2, Theorem 1, and second part of Lemma 1, we have the following proposition:

Proposition 2. *Suppose that Assumption 2 holds. Then the optimal interim price in (D) is always associated with a deterministic monotone partitional rating. Moreover, whenever cav*$\Gamma(i;\lambda,a) =$

[17] An example that violates Assumption 2 is one in which $y = \overline{y}(a)$, an increasing function of a. In this case, any change in the interim price function affects the incentives of the DM. In a previous version of this paper, we have established that if Assumption 2 is violated, optimal ratings can involve randomization.

$\Gamma(i;\lambda,a)$, *optimal rating reveals the value* $\overline{v} = v_Q(i)$ *to the market. When* cav$\Gamma(i;\lambda,a) > \Gamma(i;\lambda,a)$*, then there exists an interval* $i \in [i_1, i_2]$ *such that optimal rating reveals that* $\overline{v} \in [v_Q(i_1), v_Q(i_2)]$.

As we discuss above, the maximum value of the Lagrangian is always achieved by an interim price associated with a deterministic monotone partitional signal. The independence assumption guarantees that the optimum in (D) cannot be achieved by a (non–extreme) supporting point of the set $\{p : p, \overline{v} : \text{comonotone}, p \succcurlyeq_{\text{cv}} \overline{v}\}$ and only a unique extreme point of this set can achieve the unconstrained optimum in (D).[18]

Theorem 1 and Proposition 2 together provides a full characterization of optimal ratings. They tie the problem of optimal rating to concavification of a simple statistics of the outcome distribution: the response of the quantiles to local changes in actions along each dimension. In what follows, we describe how properties of the technology that generates the indicator and its correlation with market values determine the general properties of optimal ratings.

4 General Properties of Optimal Ratings

In this section, we provide general properties of optimal ratings and how they depend on the joint distribution of the indicator function and market values. For clarity, we focus on problems in which effort is one dimensional. In Section 5.1, we study a multi-tasking application where effort is allowed to be multi–dimensional.

4.1 Targeting An Action

We start our analysis by considering objectives that only target an action, i.e., $\alpha(y) = 0$ in (5). In this case, the problem of solving optimal rating design boils down to a characterization of the set of implementable efforts. As we have shown, an effort $a \in A$ is implementable when there exists an interim price function $p(y)$ such that $p(y)$ is a mean-preserving contraction of market values $\overline{v}(y;a)$ and a is incentive compatible given $p(y)$.

To make the model tractable, let us assume the following:

Assumption 3. *The action space* A *and the distribution function* $g(y|a)$ *satisfy the following*

1. *The action space is* $A = [0, \overline{a}] \subset \mathbb{R}$.
2. *For all* $a > 0$*, the support of* $g(y|a)$ *is a (potentially unbounded) interval* $I = [\underline{y}, \overline{y}]$ *and* $g(y|a)$ *is twice differentiable.*

[18] Formally, a supporting point of a convex set C is one that belongs to a supporting hyperplane of C.

3. *Cost function* $c(a)$ *is non-negative, strictly convex, increasing and twice differentiable for all* $a \in A$.

4. *For any effort,* $a \in A$, $\overline{v}(y;a)$ *is increasing in* y.

The first three parts of Assumption 3 are fairly common in the moral hazard literature. The last assumption notably implies that y is an indicator that is positively correlated with market values. This assumption on its own is innocuous since the indicator y itself is not payoff relevant.

By Theorem 1, under FOA, the optimal rating is found by a concavification of the function $\Gamma(i;\lambda,a) = -\lambda\left[\left.\frac{\partial F(i|\hat{a};a)}{\partial\hat{a}}\right|_{\hat{a}=a} + c'(a)\right]$ where λ is the Lagrange multiplier on local incentive-compatibility constraint, and $F(i|\hat{a};a)$ is the induced distribution of the quantiles of the indicator when the DM chooses effort $\hat{a}$ while the market believes it to be a. The following calculation ties the object to be concavified to properties of the distribution function $G(y|a)$:

$$-\lambda\frac{\partial^2}{\partial i^2}\left.\frac{\partial F(i|\hat{a})}{\partial\hat{a}}\right|_{\hat{a}=a} = -\lambda\frac{\partial^2}{\partial i^2}G_a\left(G^{-1}(i|a)|a\right) = -\lambda\frac{\partial}{\partial i}\frac{g_a\left(G^{-1}(i|a)|a\right)}{g\left(G^{-1}(i|a)|a\right)}$$
$$= \frac{-\lambda\left.\frac{\partial}{\partial y}\frac{g_a(y|a)}{g(y|a)}\right|_{y=G^{-1}(i|a)}}{g\left(G^{-1}(i|a)|a\right)} = \frac{-\lambda\left.\frac{\partial^2}{\partial y\partial a}\log g(y|a)\right|_{y=G^{-1}(i|a)}}{g(y|a)}$$

In other words, the concavity of $\Gamma(i;\lambda,a)$ at a particular quantile i is determined by the sign of the cross partial of the log-likelihood function $\log g(y|a)$. This implies that optimal ratings are directly tied to the supermodularity of the log-likelihood function $\log g(y|a)$. In what follows, we discuss a few cases and their economic interpretation and implication for optimal rating.

Let us start from the canonical assumption made in the moral hazard literature, the so-called *MLRP* assumption:

Definition 1. A distribution function $g(y|a)$ is said to satisfy Monotone Likelihood Ratio Property (**MLRP**) when $g(y|a)$ is log–supermodular. That is $\frac{\partial^2}{\partial a\partial y}\log g(y|a) = \frac{\partial}{\partial y}\frac{g_a(y|a)}{g(y|a)} \geq 0, \forall y \in I, a \in A$.

MLRP implies that an increase in effort leads to a rightward shift of the distribution of indicator realizations. Moreover, it also implies that the same is true for the conditional distribution of the indicator when restricted to an interval of values of y.[19]

Our first result establishes that in the presence of MLRP, highest implementable effort is indeed associated with full information:

[19]Formally, MLRP is equivalent to the statement that an increase in a increases the distributions over the indicator y according to the likelihood ratio order. See Shaked and Shanthikumar (2007), section 1C.

Proposition 3. *Suppose Assumptions 2 and 3 hold, FOA is valid, and $g(y|a)$ satisfies MLRP. Then the highest implementable effort is associated with interim price $p(y) = \overline{v}(y; a)$. That is, it is the highest effort level that satisfies*

$$a_{FI} \in \arg\max_{a \in A} \int \overline{v}(y; a_{FI})\, g(y|a)\, dy - c(a).$$

The above result states that the often assumed MLRP has strong implications for what can be achieved via ratings. Specifically, it states that using ratings, it is not possible to increase the level of effort beyond what the market can achieve by fully observing the indicators.

We should note that the definition of highest implementable effort a_{FI} involves calculation of a fixed point. This is because, market values $\overline{v}(y; a)$ should be calculated under the belief of the market that the action taken is a_{FI} while the DM is able to deviate from it. In Proposition 3 a_{FI} is defined as the highest such fixed point.

The proof of Proposition 3 follows straight from Theorem 1. Specifically, under MLRP, the function $\Gamma(i; \lambda, a)$ is either concave or convex for all values of i depending on the sign of λ. This means that when $\lambda > 0$, Γ is concave and coincides with its concavification. Thus given our construction of optimal ratings in Section 3, optimal rating becomes fully revealing. In turn, if $\lambda < 0$, Γ is convex in i and thus, its concavification is simply the 0 function – since $\Gamma(0; \lambda, a) = \Gamma(1; \lambda, a) = 0$. In other words, optimal rating involves providing no information which results in $a = 0$ which cannot be optimal, so λ cannot be negative.

4.1.1 Expanding and Compressing Likelihood Ratios

Many economic activities violate MLRP in systematic ways. For example, activities where greater effort affects not just the mean outcome but also its variance. Innovative activities often increase both upside potential and downside risk—greater R&D effort can lead to breakthroughs or failures. On the other hand, activities such as maintenance typically reduce variance—more careful attention produces more consistent outcomes. These patterns correspond to distributions where the cross-derivative of the log-likelihood changes sign.

In what follows, we define two classes of distributions and characterize the optimal ratings.

Definition 2. A distribution function $g(y|a)$ is said to satisfy:

1. Expanding likelihood ratio property **(ELRP)** if for any $a \in A$, there exists $\hat{y}$ such that $\frac{\partial^2}{\partial a \partial y} \log g(y|a) \geq 0$ when $y \geq \hat{y}$ and $\frac{\partial^2}{\partial a \partial y} \log g(y|a) \leq 0$ when $y \leq \hat{y}$,
2. Compressing likelihood ratio property **(CLRP)** if for any $a \in A$, there exists $\hat{y}$ such that $\frac{\partial^2}{\partial a \partial y} \log g(y|a) \leq 0$ when $y \geq \hat{y}$ and $\frac{\partial^2}{\partial a \partial y} \log g(y|a) \geq 0$ when $y \leq \hat{y}$.

The terminology reflects how effort affects the signal distribution's tails. Under ELRP, increased effort expands the tails, while under CLRP, increased effort compresses the distribution toward the center.

For example, consider $\log y \sim \mathcal{N}(\log a, a^\gamma)$, that is, $\log y$ has a normal distribution with mean $\log a$ and variance a^γ. In this case, we can use the definition of the density of the normal distribution to show that

$$\frac{\partial^2}{\partial a \partial y} \log g(y|a) = \frac{1 + \gamma \log y/a}{a^{1+\gamma} y}.$$

When $\gamma > 0$, g satisfies ELRP since the above is positive if and only if $y/a \geq e^{-\gamma}$. In contrast, when $\gamma < 0$, g satisfies CLRP since the above is negative if and only if $y/a \leq e^{-\gamma}$.

A version of these examples are depicted in Figure 3. As it can be seen, in case of ELRP, the tail densities increase as effort a increases while the densities for mid–realizations decline. In contrast, under CLRP, tail densities decline while the densities for mid–realizations increase. In both cases, the two densities intersect exactly twice which is in contrast with single crossing of MLRP.

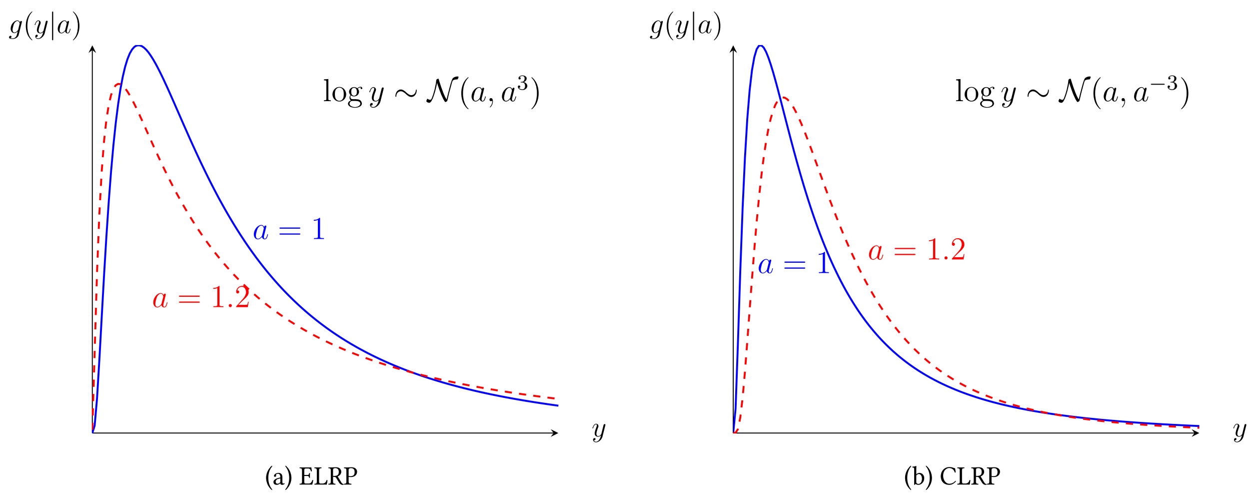


Figure 3: Distributions with ELRP (left) and CLRP (right)

Given these definitions, we can state our result on optimal ratings for these classes distributions:

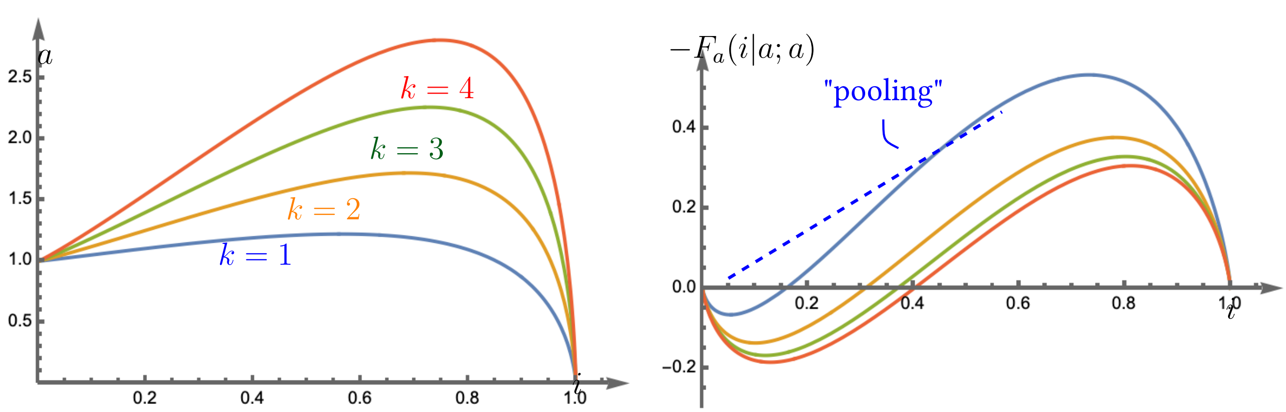
Proposition 4. *Suppose Assumptions 2 and 3 hold and FOA is valid.*

1. *If $g(y|a)$ satisfies ELRP, then the highest implementable effort a_{LC} is the highest value of effort that is incentive compatible for an interim price associated with lower-censorship ratings, i.e., ratings that pool values of y below a threshold and reveal higher values.*

2. *If $g(y|a)$ satisfies CLRP, then the highest implementable effort a_{UC} is the highest value of effort that is incentive compatible for an interim price associated with upper-censorship ratings, i.e., ratings that pool values of y above a threshold and reveal lower values.*

As the above proposition establishes, optimal ratings for ELRP and CLRP distributions are fairly simple. They involve either upper censorship (in case of ELRP) or lower censorship (in case of CLRP). In what follows we provide an example and discuss its implications for various tasks and technologies.

Suppose that $y \sim \mathcal{N}\left(a, (ka)^2\right)$, that y determines market values, i.e., $\overline{v}(y; a) = y$, and that cost is $c(a) = a^2/a$. Under a full information rating, $p(y) = y$, profit of the DM is $a - a^2/2$ which is maximized at $a_{FI} = 1$. It can be easily checked that in this case $G(y|a)$ satisfies ELRP. For any value of $i \in [0, 1]$, we can find the highest level of effort that is a best response to pooling of i lowest realizations of the indicator y. This is depicted in Figure 4 (left panel) for values of $k = 1, 2, 3, 4$. At the lowest value, $i = 0$, optimal effort is $a_{FI} = 1$. Optimal effort peaks at some threshold, 0.56, 0.69, 0.73, 0.75 respectively, and falls to zero as i tends to 1. It should be noted that as variance of y becomes steeper as a function of a, the highest possible value of effort increases. Figure 4 (right panel) depicts the marginal change in the quantile as a result of an increase in a, $-F_a(i|a; a)$, and its concavification (in the case of $k = 1$). The threshold for pooling on the right coincides with the peak of the left plot since optimal ratings take the form of lower censorship.



(a) Optimal effort for pooling the lowest i realizations of y (b) Concavification of the marginal change in quantiles

Figure 4: Optimal efforts for lower-censorship policies

We should note that the notion of ELRP and CLRP are tied to whether an increase in effort a leads to a higher or lower variance of the indicator y. Specifically, suppose that $y = f(\sigma(a)\varepsilon + m(a))$ with m and f increasing, and ε has density $e^{h(\varepsilon)}$ such that $h(\varepsilon)$ is concave and

$\varepsilon + h'(\varepsilon)/h''(\varepsilon)$ is increasing in ε. In this case, y exhibits ELRP (CLRP) only if $\sigma(a)$ is increasing (decreasing) in a. Several classes distributions satisfy these properties for h: Normal, Gumbel, Generalized Normal, Logistic, etc. For this class of distributions, a constant $\sigma(a)$ leads to MLRP.

The above findings point to a practical property of optimal ratings in targeting an action. Namely that how variance of indicator interacts with the desired action determines the best way to incentivize it. Specifically, an activity where more effort leads to a more precise outcome (CLRP) such as maintenance activities, full revelation at low values and pooling at higher values, encourages higher effort to avoid low value punishments. On the other hand, if higher effort leads to riskier outcome (ELRP) such as innovative activities, then pooling of low realization via lower-censorship ratings provides insurance against possible downsides and encourages risk taking (increasing effort). When effort doesn't affect variance (MLRP), optimal rating is full revelation and no pooling is needed.

We should end this section by emphasizing that while we have focused on the highest possible effort that is implementable, any lower value of effort can also be targeted by rating systems. The analysis in this section specifically is useful in identifying values of effort that are higher than those achieved by a fully revealing rating system. Especially in markets with positive externalities where fully informative ratings lead to inefficiently low levels of effort, one can use ratings (absent MLRP) to improve market efficiency.

4.2 Redistributive Motives

Here, we discuss optimal ratings in presence of redistributive motives. This could happen because of societal values – see for example Dessein et al. (2025), to guarantee a minimal level of ex-post payoff. In an earlier version of this paper Saeedi and Shourideh (2022), we provide examples in which the intermediary wishes to maximize fees from providing the rating to the market and showed that this also gives rise to redistributive motives.

To see the effect of redistributive motives, suppose that $\alpha(y)$ is positive, decreasing in y, and $\int_I \alpha(y)\,dG(y|a) < \infty$ for all $a \in A$. In this case and using Assumption 3, we can apply the result of Theorem 1 which implies that optimal ratings are determined by concavifying the following function

$$\int_{G^{-1}(i|a)}^{\infty} \alpha(y)\,dG(y|a) - \lambda G_a\left(G^{-1}(i|a)\right).$$

We have extensively discussed the properties of the second term that captures the incentive effect of ratings. The first term is decreasing and convex function of i since[20]

$$\frac{\partial^2}{\partial i^2}\int_{G^{-1}(i|a)}^{\infty}\alpha\left(y\right)dG\left(y|a\right)=-\frac{\partial}{\partial i}\frac{\alpha\left(G^{-1}\left(i|a\right)\right)g\left(G^{-1}\left(i|a\right)|a\right)}{g\left(G^{-1}\left(i|a\right)|a\right)}=-\frac{\alpha'\left(G^{-1}\left(i|a\right)\right)}{g\left(G^{-1}\left(i|a\right)|a\right)}\geq 0$$

Evidently, if $\lambda=0$, since the above function is convex, optimal ratings must be one that pools all values of y. Obviously, such a rating does not provide any incentive for exerting effort. The total gain function combines this convex redistributive term with the incentive effect. Even when the incentive effect is concave (as under MLRP), strong redistributive preferences can make the total gain function non-concave for low quantiles.

The following proposition illustrates how redistributive motives affect optimal ratings:

Proposition 5. *Suppose that*

$$\lim_{i\to 0}\alpha\left(G^{-1}\left(i|a\right)\right)+\lambda\frac{g_a\left(G^{-1}\left(i|a\right)|a\right)}{g\left(G^{-1}\left(i|a\right)|a\right)}>\int_0^1\alpha\left(G^{-1}\left(i|a\right)\right)di.$$

Then there exists an interval around $\underline{y}=G^{-1}\left(0|a\right)$ *where the optimal rating is pooling. As a special case, the same is true if* $\alpha\left(G^{-1}\left(i|a\right)\right)\to\infty$ *as* $i\to 0$ *and* $\lim_{i\to 0}\left|\frac{g_a\left(G^{-1}(i|a)|a\right)}{g\left(G^{-1}(i|a)|a\right)}\right|<\infty$.

Proposition 5 illustrates the key force of redistributive motives. The assumption implies that the function $\Gamma\left(i;\lambda,a\right)$ satisfies $\left.\frac{\partial\Gamma(i;\lambda,a)}{\partial i}\right|_{i=0}<-\Gamma\left(0;\lambda,a\right)$. Since $\Gamma\left(1;\lambda,a\right)=0$, the line connecting $\left(0,\Gamma\left(0;\lambda,a\right)\right)$ and $\left(1,\Gamma\left(1;\lambda,a\right)\right)$ is above $\Gamma\left(i;\lambda,a\right)$ for an interval of values of $i>0$ and thus the concave envelope of Γ lies strictly above Γ for an interval of values of i above 0. This means that optimal rating should be pooling for an interval of values above $i=0$. Intuitively, when redistribution motives are high, the redistribution effects of pooling at low values is higher than the incentive costs due to this pooling.

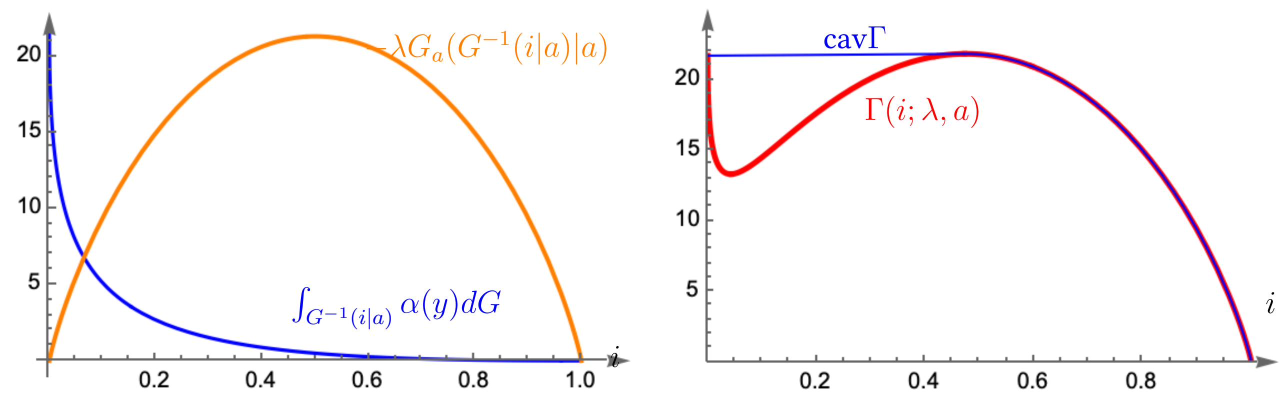
As an example, suppose that $y=a\varepsilon,\log\varepsilon\sim\mathcal{N}\left(-1/2,1\right)$, $c\left(a\right)=a^2/2$, and $\alpha\left(y\right)=y^{-\beta}$ for some $\beta>0$. In this case,

$$\begin{aligned}\frac{\partial}{\partial i}\Gamma\left(i;\lambda,a\right)&=-\alpha\left(y\right)-\lambda\left.\frac{g_a\left(y|a\right)}{g\left(y|a\right)}\right|_{y=G^{-1}(i|a)}\\&=-y^{-\beta}-\lambda\left.\frac{\frac{1}{2}+\log\frac{y}{a}}{a}\right|_{y=G^{-1}(i|a)}\end{aligned}$$

The above function is increasing in $i=G^{-1}\left(y|a\right)$ for values of y below a threshold and decreasing for values of y above it. This implies that H is convex below this threshold and concave above it. As a result, optimal rating must be lower censorship. Figure 5a shows the components of the

[20]While we are assuming α to be differentiable, this is really not needed for convexity of the first term.

function Γ when $\beta = 2$ at the optimum value of effort, a, and multiplier, λ. Since G satisfies MLRP, the incentive component is concave while adding the redistributive motives makes the sum convex for low realizations. The resulting sum and its concavification is depicted in Figure 5b. The optimal rating pools the lowest 47.28 percent of realizations of y.



(a) Components of $\Gamma(i;\lambda,a)$: Redistributive (blue) and incentives (orange)

(b) The function $\Gamma(i;\lambda,a)$ (red) and its concavification (blue)

Figure 5: Concavification of the marginal change in quantiles

It is useful to conclude this section with a summary of the results:

1. When targeting an effort, the set of implementable efforts is determined by the supermodularity of $\log g(y|a)$.

2. Optimal rating associated with highest implementable effort is lower (upper) censorship when $g(y|a)$ exhibits ELRP (CLRP).

3. Strong redistributive motives lead to pooling of low realizations of y.

In what follows, we use the insights in this section to shed light on two important examples: Multi-tasking and optimal redistributive test design with heterogeneity.

5 Applications

In this section, we illustrate the value of our characterization results above by applying them to a multi-task moral hazard model and optimal test design.

5.1 Rating Design, Multi-task Moral Hazard and Window Dressing

Since the seminal work of Holmström and Milgrom (1991), the multi-task principal-agent models have become the workhorse of analyzing incentives in setting where agents can use several actions to affect the observed outcomes[21] – see also Baker (1992) and Dewatripont et al. (1999).[22] In this section, we consider a variant of the model in Dewatripont et al. (1999) to understand how rating design can be used to mitigate multi-task incentive problems.

The DM chooses a vector of efforts $a = (a_1, \cdots, a_N) \in [0, \overline{a}]^N$, which determines market values and the indicator as follows:

$$v = b_v \cdot a + \varepsilon_v, y = b_y \cdot a + \varepsilon_y$$

where $b_v, b_y \in \mathbb{R}_+^N$ capture the effect of a on market values and indicator, respectively, and:

$$\begin{pmatrix} \varepsilon_y \\ \varepsilon_v \end{pmatrix} \sim \mathcal{N}\left(0, \begin{pmatrix} \sigma_v^2 & \sigma_{vy} \\ \sigma_{vy} & \sigma_y^2 \end{pmatrix}\right), \sigma_{vy} > 0$$

In words, choosing a vector of effort levels creates a value for the market while it affects the indicator observed by the intermediary differently. Since ε_v and ε_y are positively correlated, high values of the indicator signal a higher value for the market.

Using properties of the normal distribution, we can show that market values conditional on y and belief $\hat{a}$ are:

$$\overline{v}(y; \hat{a}) = \mathbb{E}[v|y; \hat{a}] = \frac{\sigma_{vy}}{\sigma_y^2}(y - b_y \cdot \hat{a}) + b_v \cdot \hat{a} = \beta y + (b_v - \beta b_y) \cdot \hat{a}$$

where $\beta = \sigma_{vy}/\sigma_y^2 > 0$. Suppose that the cost of effort is

$$c(a) = \frac{1}{2}\sum_{n=1}^{N} \kappa_n a_n^2$$

where $\kappa_n > 0$ is marginal cost of task n.

When $b_v \neq b_y$, the indicator is a distorted measure of market value. A special case is window

[21] Several empirical studies have looked at variants of the multi-task moral hazard model. A partial list includes Dumont et al. (2008) and Alexander (2020) for compensation of doctors, Acemoglu et al. (2020) for incentives in military and security forces, De Janvry et al. (2023) for incentives of government employees, Andrabi and Brown (2022) for incentives of teachers, and Mayzlin et al. (2014) and Hui et al. (2025) for incentives on online platforms.

[22] Since Dewatripont et al. (1999) analyze a variant of Holmström (1999)'s career concern model, their model is closer to ours where the agent's compensation is determined by market expectations as opposed to an endogenous contract chosen by the principal as in Holmström and Milgrom (1991) and Baker (1992).

dressing: an effort that boosts the indicator without affecting market values. More precisely, we say that task n *exhibits window–dressing* if $b_{y,n} > 0$ and $b_{v,n} = 0$. In this case, efficiency requires $a_i = 0$ since a_i does not add to market values and is costly.

A key feature of this model is that it is reducible to the single effort model of section 4. This is because the effort vector a affects the distribution of the indicator solely through $m_y = b_y \cdot a$. Additionally given m_y, there is a unique cost minimizing values for vector of efforts:

$$\min_{a \in [0,\bar{a}]^N} \frac{1}{2} \sum_{n=1}^{N} \kappa_n a_n^2$$
$$\text{subject to } b_y \cdot a = m_y$$

The solution to the above is given by

$$\tilde{a}_n (m_y) = \frac{b_{y,n}}{\kappa_n} \frac{m_y}{\sum_j b_{y,j}^2 / \kappa_j},$$

and the resulting indirect cost function is

$$C (m_y) = \frac{m_y^2}{2 \sum_j b_{y,j}^2 / \kappa_j}.$$

Since the DM can choose the vector a given any level of m_y, she will always choose $\tilde{a}_n (m_y)$ to minimize her cost. Thus the rating design problem is reducible to a choice of an interim price function $p(\cdot)$ and m_y.

Now, consider the problem of finding the optimal rating that maximizes total surplus. Total surplus in this environment is given by

$$b_v \cdot \tilde{a} (m_y) - C (m_y)$$

and thus the welfare maximizing m_y^* is given by

$$m_y^* = \sum_{n=1}^{N} \frac{b_{y,n} b_{v,n}}{\kappa_n}.$$

Given that $y | m_y \sim \mathcal{N} \left(m_y, \sigma_y^2 \right)$ and the normal distribution satisfies MLRP, our result in section 4 implies that the highest implementable level of effort is the one associated with full information. Under full information, the DM solves the following:

$$\max_{m > 0} \beta b_y \cdot \tilde{a} (m) - C (m)$$

whose solution is given by

$$m_{FI}^{*} = \beta \sum_{i} \frac{b_{y,i}^{2}}{\kappa_{i}}$$

We thus have the following proposition:

Proposition 6. *If $m_y^* \geq m_{FI}^*$, then the welfare maximizing rating is fully revealing and implements $\tilde{a}\left(m_{FI}^*\right)$. If $m_y^* < m_{FI}^*$, then the welfare maximizing rating implements $\tilde{a}\left(m_y^*\right)$. An optimal rating that implements $\tilde{a}\left(m_y^*\right)$ is a lower-censorship rating that pools the realizations of y below $\overline{y}$ given by*

$$\beta \overline{z} \phi\left(\overline{z}\right) + \beta \frac{\phi\left(\overline{z}\right)^{2}}{\Phi\left(\overline{z}\right)} + \beta\left[1 - \Phi\left(\overline{z}\right)\right] = \frac{\sum_{i=1}^{N} b_{y,i} b_{v,i}/\kappa_{i}}{\sum_{i=1}^{N} b_{y,i}^{2}/\kappa_{i}}, \overline{z} = \frac{\overline{y} - m_{y}^{*}}{\sigma_{y}} \tag{10}$$

where Φ is the c.d.f. of the standard normal distribution and $\phi = \Phi'$.

Given the above discussion and the results of section 4, Proposition 6 is immediate. We should note that the lower-censorship rating is one of possibly many optimal ratings that implement the efficient outcome. This is because m_y^* is an interior point of the set of implementable values of m_y given by $[0, m_{FI}^*]$.

Proposition 6 also identifies the sufficient statistic that determines the optimal lower-censorship rating. This is given by $\frac{\sigma_y^2}{\sigma_{vy}} \frac{\sum_{i=1}^{N} b_{y,i} b_{v,i}/\kappa_i}{\sum_{i=1}^{N} b_{y,i}^2/\kappa_i}$. An object of interest is the effect of changes in cost of manipulation or window–dressing on the optimal rating. So, suppose that some effort a_i, exhibits window dressing, i.e., $b_{v,i} = 0, b_{y,i} > 0$. An increase in κ_i reduces the denominator of the sufficient statistic and thus increases its value. Since the left hand side of (10) is decreasing in $\overline{z}$, an increase in the sufficient statistic leads to a reduction in $\overline{z}$ and hence a more informative rating. We thus have the following:

Proposition 7. *Suppose that a_i exhibits window dressing effort, i.e., $b_{v,i} = 0, b_{y,i} > 0$ and that $m_y^* < m_{FI}^*$. Then, a decrease in κ_i leads to a less informative optimal rating.*

This result is reminiscent of Holmström and Milgrom (1991)'s result on optimality of low–powered incentives. It highlights that when manipulation becomes easier, optimal ratings should become less informative in order to reduce window–dressing incentives.

For a general task i, whether a decline in cost κ_i leads to a less or more informative rating depends on the relationship between the values of $b_{v,i}$ and $b_{y,i}$. The following proposition illustrates this dependence:

Proposition 8. *A decline in cost of task i, κ_i leads to a less informative signal, i.e., higher value of $\overline{z}$, if and only if*

$$\frac{b_{y,i}}{\sum_{j=1}^{N} b_{y,j}^{2}/\kappa_{j}} - \frac{b_{v,i}}{\sum_{j=1}^{N} b_{y,j} b_{v,j}/\kappa_{j}} \geq 0$$

We should end this section by emphasizing that while an extensive literature has studied multi-tasking model and their empirical applications, the idea of using rating policy to mitigate issues like window–dressing remains unexplored. Our analysis here means to illustrate the benefits of using rating policies to reduce inefficiencies caused by such motives.

5.2 A Nonreducible Two–Task Model

The key benefit of the setup above was that it was reducible to the single effort setup characterized in section 4. Here we discuss an example that is not reducible to single effort and discuss its implication on optimal rating design.

Suppose there are two tasks a_1, a_2 and the market values and the indicator are given by

$$v = a_1 (\varepsilon_1 + 1)$$
$$y = ba_1 (\varepsilon_1 + 1) + a_2 (\varepsilon_2 + 1)$$

where ε_i's are standard normal distributions and independent and $b > 0$. Since v and y are positively correlated, we can calculate the expected market values using properties of the normal distribution

$$\overline{v} (y; a) = \frac{ba_1^2}{b^2 a_1^2 + a_2^2} (y - ba_1 - a_2) + a_1 = \beta (a) \times (y - ba_1 - a_2) + a_1$$

Using Theorem 1, when the objective of rating design is independent of the distribution of interim prices, the shape of the optimal rating is determined by a weighted value of the marginal change in the quantiles:

$$-\lambda_1 \left. \frac{\partial G \left(G^{-1} (i|\hat{a}) |a\right)}{\partial a_1} \right|_{\hat{a}=a} - \lambda_2 \left. \frac{\partial G \left(G^{-1} (i|\hat{a}) |a\right)}{\partial a_2} \right|_{\hat{a}=a}$$

Since $y \sim \mathcal{N} \left(ba_1 + a_2, (ba_1)^2 + a_2^2\right)$, we can use properties of the normal distribution to show that the above is either concave for low values of i and convex for high values or vice versa. As a result, optimal rating must be either upper or low. The following result refines this further:

Proposition 9. *Suppose that the cost of effort is $c (a) = \frac{\kappa}{2} (a_1^2 + a_2^2)$ and $b < \sqrt{\sqrt[3]{1 + \sqrt{2}} - \sqrt[3]{\sqrt{2} - 1}} \approx 0.772$. Then, the welfare maximizing rating in the non–reducible two–task model is upper censorship and delivers welfare that is strictly higher than fully informative rating.*

Note that in the two task model of this section both productive effort, a_1, and window–dressing effort a_2, increase the mean and variance of the indicator y. However, since the impact of productive effort on y is lower, censoring higher values of y has a bigger impact on window–dressing.

Since window–dressing actions only destroy surplus, some pooling of high observations has a high impact on window–dressing incentives while its impact on productive effort is mild.

Overall, our analysis of the multi-task model presented here highlights the importance of ratings when the indicator is a distorted measure of market values.

5.3 Redistributive Test Design

Recent public discourse in the education realm has highlighted the biases of standardized testing (such as the SAT) and testing of difficult subjects (such as math) against students with socioeconomic disadvantages.[23] Inspired by these observations, there has been a movement for more relaxation of requiring students to participate in such tests. This includes several universities' policies to make the SAT optional (see Dessein et al. (2025)) and attempts at making mathematics education more accessible and easier. Inspired by this debate, in this section, we provide an alternative answer to this question in the form of optimal test design.

To see this, consider a student that could be of $\theta \in \{R, P\}$with probabilities f_R, f_P . Suppose that both types can exert effort a_θ which leads to a distribution of an indicator y which is distributed according to $g(y|a_\theta)$ with support given by $I = [\underline{y}, \overline{y}]$ – with the possibility that $\underline{y} = -\infty$ and $\overline{y} = \infty$. The cost of effort for each type is $k_\theta c(a)$ where $c(a)$ is a convex and increasing function where $0 < k_R < k_P$. Finally, let $v = y$ so that the market values are simply the value of the indicators and let $p(y)$ be the interim price function that is increasing.

Consider a rating designer who wishes to maximize the following objective

$$\alpha_P f_P \left[\int_I p(y)\, dG(y|a_P) - k_P c(a_P)\right] + \alpha_R f_R \left[\int_I p(y)\, dG(y|a_R) - k_R c(a_R)\right] \tag{11}$$

where $\alpha_P f_P + a_R f_R = 1$ and $\alpha_P > 1 > \alpha_R$. Let us assume that an increase in a leads to an increase in $g(y|a)$ in the sense of first order stochastic dominance. This would imply that since the student of type R has a lower marginal cost, her associated distribution of the indicator is shifted to the right.

The problem of optimal rating design is then to find $p(y)$ and a_R, a_P to maximize the objective in (11) subject to incentive compatibility for both types and that $p(y)$ is mean preserving contraction of y which is distributed according to $f_P G(y|a_P) + f_R G(y|a_R) = \overline{G}(y|a_R, a_P)$. In this case, a similar proof to that of Theorem 1 implies that under the validity of FOA, the optimal

[23] In 2023, California Board of Education passed the controversial California Mathematics Framework which sets guidelines for mathematics education in California public schools. Citing the students' socioeconomic disadvantages, the framework calls for some relaxations in testing standards and education of mathematics. For more information, see the article in the New Yorker on the California Mathematics Framework.

rating can be found by concavification of the gain function

$$\Gamma(i) = -\alpha_P f_P G\left(\overline{G}^{-1}(i)\,|a_P\right) - \alpha_R f_R G\left(\overline{G}^{-1}(i)\,|a_R\right)$$
$$-\lambda_P G_a\left(\overline{G}^{-1}(i)\,|a_P\right) - \lambda_R G_a\left(\overline{G}^{-1}(i)\,|a_R\right)$$

where in the above i is the quantile of y according to $\overline{G}$ and λ_θ's are the multipliers on the associated incentive compatibility constraint.

The following proposition guarantees that for a class of distribution functions, the second derivative of the above object switches sign at most three times. This would imply that optimal rating is always switching between at most four regions of pooling and separation:

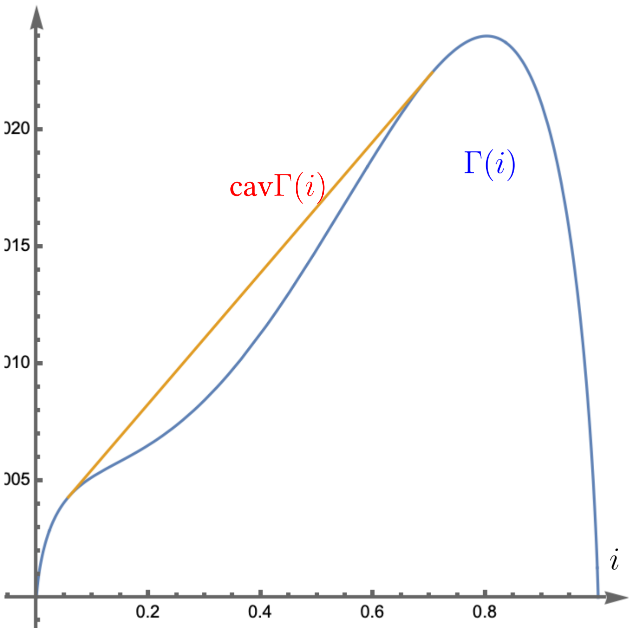
Proposition 10. *Suppose that* $\log g(y|a) = f(y) + r(y)\,m(a) - b(a)$ *where* $r(y), m(a), b(a)$ *are increasing functions and* $m(a) > 0$*. Then optimal rating that maximizes (11) always has at most four alternating intervals of pooling and revelation.*

Note that the class of distributions considered in Proposition 10 includes some of the fairly common ones that are used in applied work including 1. a normally distributed y where one of the mean or variance is controlled by the action a, 2. a log–normally distributed y such that a controls mean of $\log y$, 3. when y is distributed according to a extreme–valued distribution of type 1 and 2 (Gumbel and Frechet) where the scale parameter is controlled by a, among others. Moreover, the assumption implies that $\log g$ is supermodular in (y, a), i.e., it satisfies MLRP.

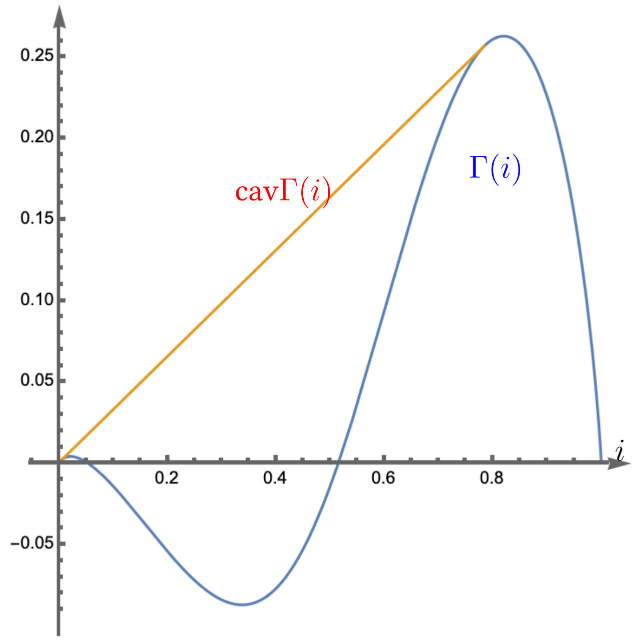
Interestingly, one might think that, in light of the results on redistributive optimal ratings in section 4, it should always be optimal to pool low realizations. However, the difference here is that there are two incentive constraints. Under certain conditions on the likelihood function g_a/g for low realizations of y – specifically as it becomes arbitrarily large as $y \to \underline{y}$, the incentive effect dominates the redistributive motives for low realizations and optimal ratings become mid censorship. To see this, let us consider the following example.

Example 1. Suppose that $y = a + \varepsilon$ where $\varepsilon \sim \mathcal{N}(0,1)$ and that $c(a) = a^2/2$. Let us also assume that $\alpha_P = 1/f_P, \alpha_R = 0$ so that objective is to maximize the payoff of the high cost type. As mentioned before, this example satisfies the requirement of Proposition 10 and hence optimal ratings switch at most three times between pooling and revelation. Our calculations illustrate that indeed optimal ratings are mid censorship: those that pool middle observations of y and separate the extreme realizations. We further assume that $k_R = 1/2$ and that $f_P = f_R = 1/2$. Figure 6 depicts the optimal rating and the concavification of the gain function described above for different values of costs for type P.[24] In the left panel, the cost of type P is closer to that

[24] In order for the difference between the concavification and the function to be more visible, we subtract $(\alpha_R f_R + \alpha_P f_P)(1-i)$ from the gain function Γ in the plot. Since this subtraction is linear, it does not affect the resulting concavification in terms of the pooling and separating intervals.



(a) Low cost for type P, $k_P = 3/4$



(b) High cost for type P, $k_P = 5/4$

Figure 6: Determining the optimal rating for Example 1

of type R. In this case, optimal rating pools observations approximately between the 6th and 71st quantile of the y distribution. When the cost for type−P increases to 5/4, the optimal rating pools observations below the 78th percentile – the lower threshold is at 1.3th percentile. As it can be seen, as the difference between the two cost types increases the rating policy becomes less informative in order to redistribute more across the types.

6 Conclusion

In this paper, we have developed a general framework for the design of rating systems in the presence of moral hazard and strategic manipulation. Our approach makes this problem tractable by identifying "interim prices" as the sufficient statistic for the agent's incentives. Under a natural monotonicity restriction implementability is equivalent to a majorization restriction: feasible interim prices are mean-preserving contractions of the full-information market value. This converts rating design into the classic moral-hazard problem with a majorization constraint.

Building on this characterization, we provide a general solution method. Under a first-order approach to incentive compatibility, optimal rating design reduces to concavifying a gain function in quantile space. This formulation yields two broad takeaways. First, it delivers a set of sufficient statistics for optimal transparency: the technology matters through how effort shifts the quantile distribution of the indicator and, in turn, the distribution of market values. Second, it implies structure on optimal information policies. Under mild conditions, optimal ratings are simple deterministic monotone partitions of the indicator space: the intermediary either fully reveals the indicator on some regions or pools contiguous regions into coarse categories.

The economic insights that emerge from our analysis connect the statistical properties of the task to the structure of optimal disclosure. In the canonical benchmark with monotone likelihood ratios, full information disclosure achieves the highest implementable effort. Departures from this benchmark generate systematic and testable patterns of censorship. For innovative activities where greater effort expands outcome variance (ELRP), *lower-censorship* ratings that pool poor realizations provide insurance against downside risk, encouraging risk-taking. For maintenance activities where effort compresses variance (CLRP), *upper-censorship* ratings that pool high realizations punish poor outcomes by deterring negligence and encouraging consistency. Strong redistributive motives create a fundamental tension between maximizing effort and protecting agents from downside risk, generally favoring policies that pool low realizations. More broadly, the framework clarifies why "more transparency" is not a universal remedy: whether additional information strengthens incentives depends on which parts of the outcome distribution effort affects.

Two applications illustrate how the theory speaks to current design problems. In multi-task environments with window dressing, more informative ratings can intensify incentives for manipulable activities that improve measured performance without an increase in underlying value. The optimal response is often to reduce informativeness to mitigate manipulation incentives. In particular, when manipulation becomes cheaper, welfare-optimal rating policies become more opaque, and in a nonreducible setting upper censorship can strictly dominate full disclosure by disproportionately discouraging extreme realizations driven by window dressing. In redistributive test design, we show how the same concavification logic rationalizes *mid-censorship* rules that pool intermediate outcomes while separating extremes. These results formalize a common policy intuition: the optimal granularity of evaluation depends jointly on incentive provision, manipulability, and distributional objectives.

References

Acemoglu, D., L. Fergusson, J. Robinson, D. Romero, and J. F. Vargas (2020): "The Perils of High-Powered Incentives: Evidence from Colombia's False Positives," *American Economic Journal: Economic Policy*, 12, 1–43. 24

Agarwal, V., R. Vashishtha, and M. Venkatachalam (2018): "Mutual fund transparency and corporate myopia," *The Review of Financial Studies*, 31, 1966–2003. 9

Albano, G. L. and A. Lizzeri (2001): "Strategic certification and provision of quality," *International economic review*, 42, 267–283. 5

Alexander, D. (2020): "How Do Doctors Respond to Incentives? Unintended Consequences of Paying Doctors to Reduce Costs," *Journal of Political Economy*, 128, 4046–4096. 24

Ali, S. N., N. Haghpanah, X. Lin, and R. Siegel (2022): "How to sell hard information," *The Quarterly Journal of Economics*, 137, 619–678. 6

Andrabi, T. and C. Brown (2022): "Subjective versus Objective Incentives and Teacher Productivity," *Berkeley Working Paper*. 24

Baker, G. P. (1992): "Incentive Contracts and Performance Measurement," *Journal of Political Economy*, 100, 598–614. 24

Ball, I. (2025): "Scoring Strategic Agents," *American Economic Journal: Microeconomics*, 17, 97–129. 5

Barahona, N., C. Otero, and S. Otero (2023): "Equilibrium Effects of Food Labeling Policies," *Econometrica*, 91, 839–868. 6

Berg, F., J. F. Koelbel, and R. Rigobon (2022): "Aggregate confusion: The divergence of ESG ratings," *Review of Finance*, 26, 1315–1344. 6

Bergemann, D., T. Heumann, and S. Morris (2022a): "Screening with persuasion," *arXiv preprint arXiv:2212.03360*. 5, 12

Bergemann, D., T. Heumann, S. Morris, C. Sorokin, and E. Winter (2022b): "Optimal Information Disclosure in Classic Auctions," *American Economic Review: Insights*, 4, 371–388. 5, 12

Blackwell, D. (1953): "Equivalent comparisons of experiments," *The Annals of Mathematical Statistics*, 265–272. 59

Boleslavsky, R. and K. Kim (2020): "Bayesian persuasion and moral hazard," *Working Paper, Emory University*. 5

Chade, H. and J. Swinkels (2020): "The No-Upward-Crossing Condition, Comparative Statics, and the Moral-Hazard Problem," *Theoretical Economics*, 15, 445–476. 14, 54

De Janvry, A., G. He, E. Sadoulet, S. Wang, and Q. Zhang (2023): "Subjective Performance Evaluation, Influence Activities, and Bureaucratic Work Behavior: Evidence from China," *American Economic Review*, 113, 766–799. 24

Dessein, W., A. Frankel, and N. Kartik (2025): "Test-Optional Admissions," *American Economic Review*, 115, 3130–3170. 21, 28

Dewatripont, M., I. Jewitt, and J. Tirole (1999): "The economics of career concerns, part II: Application to missions and accountability of government agencies," *The Review of Economic Studies*, 66, 199–217. 9, 24

Doval, L. and A. Smolin (2024): "Persuasion and Welfare," *Journal of Political Economy*, 132, 2451–2487. 3

Dumont, E., B. Fortin, N. Jacquemet, and B. Shearer (2008): "Physicians' Multitasking and Incentives: Empirical Evidence from a Natural Experiment," *Journal of Health Economics*, 27, 1436–1450. 24

Frankel, A. and N. Kartik (2019): "Muddled information," *Journal of Political Economy*, 127, 1739–1776. 5

Gentzkow, M. and E. Kamenica (2016): "A Rothschild-Stiglitz Approach to Bayesian Persuasion," *American Economic Review*, 106. 10, 36

Halac, M., E. Lipnowski, and D. Rappoport (2021): "Rank Uncertainty in Organizations," *American Economic Review*, 111, 757–786. 6

He, S., B. Hollenbeck, and D. Proserpio (2022): "The market for fake reviews," *Marketing Science*. 9

Holmström, B. (1999): "Managerial incentive problems: A dynamic perspective," *The Review of Economic Studies*, 66, 169. 2, 9, 24

Holmström, B. and P. Milgrom (1991): "Multitask principal–agent analyses: Incentive contracts, asset ownership, and job design," *The Journal of Law, Economics, and Organization*, 7, 24–52. 4, 9, 24, 26

Hui, X., G. Z. Jin, and M. Liu (2025): "Designing Quality Certificates: Insights from eBay," *Journal of Marketing Research*, 62, 40–60. 24

Hui, X., M. Saeedi, Z. Shen, and N. Sundaresan (2016): "Reputation and regulations: evidence from eBay," *Management Science*, 62, 3604–3616. 8

Hui, X., M. Saeedi, G. Spagnolo, and S. Tadelis (2023): "Raising the bar: Certification thresholds and market outcomes," *American Economic Journal: Microeconomics*, 15, 599–626. 6, 9

Hui, X., M. Saeedi, and N. Sundaresan (2018): "Adverse selection or moral hazard, an empirical study," *The Journal of Industrial Economics*, 66, 610–649. 9

Jewitt, I. (1988): "Justifying the First-Order Approach to Principal-Agent Problems," *Econometrica*, 1177–1190. 54

Kleiner, A., B. Moldovanu, and P. Strack (2021): "Extreme Points and Majorization: Economic Applications," *Econometrica*, 89, 1557–1593. 5, 11, 35, 40

Kolotilin, A. (2018): "Optimal information disclosure: A linear programming approach," *Theoretical Economics*, 13, 607–635. 36

Luenberger, D. G. (1997): *Optimization by Vector Space Methods*, John Wiley & Sons. 40

Madsen, E., B. Williams, and A. Skrzypacz (2025): "Performance Incentives with Multiple Instruments," *Available at SSRN 5262226*. 5

Mayzlin, D., Y. Dover, and J. Chevalier (2014): "Promotional Reviews: An Empirical Investigation of Online Review Manipulation," *American Economic Review*, 104, 2421–2455. 4, 24

Nosko, C. and S. Tadelis (2015): "The limits of reputation in platform markets: An empirical analysis and field experiment," Tech. rep., National Bureau of Economic Research. 6, 8

Onuchic, P. and D. Ray (2023): "Conveying Value via Categories," *Theoretical Economics*, 18, 1407–1439. 5

Perez-Richet, E. and V. Skreta (2022): "Test design under falsification," *Econometrica*, 90, 1109–1142. 5

Prat, A. (2005): "The Wrong Kind of Transparency," *The American Economic Review*, 95, 862–877. 9

Rockafellar, R. T. (1970): *Convex Analysis*, Princeton University Press. 40

Rothschild, M. and J. E. Stiglitz (1970): "Increasing risk: I. A definition," *Journal of Economic theory*, 2, 225–243. 10, 59

Saeedi, M. (2019): "Reputation and adverse selection: theory and evidence from eBay," *The RAND Journal of Economics*, 50, 822–853. 8

Saeedi, M. and A. Shourideh (2022): "Optimal Rating Design," *Working Paper, Carnegie Mellon University*. 21

SHAKED, M. AND J. G. SHANTHIKUMAR (2007): *Stochastic orders*, Springer Science & Business Media. 10, 17

STRASSEN, V. (1965): "The existence of probability measures with given marginals," *The Annals of Mathematical Statistics*, 36, 423–439. 59

VATTER, B. (2025): "Quality Disclosure and Regulation: Scoring Design in Medicare Advantage," *Econometrica*, 93, 959–1001. 6

ZAPECHELNYUK, A. (2020): "Optimal quality certification," *American Economic Review: Insights*, 2, 161–76. 5

ZERVAS, G., D. PROSERPIO, AND J. W. BYERS (2021): "A First Look at Online Reputation on Airbnb, Where Every Stay is Above Average," *Marketing Letters*, 32, 1–16. 4

ZUBRICKAS, R. (2015): "Optimal grading," *International Economic Review*, 56, 751–776. 5

A Proofs

A.1 Proof of Proposition 1

Proof. That if p is constructed from some information structure (π, S) then $p \succcurlyeq_{\text{c.v.}} \overline{v}$ is immediate from the text.

Now, suppose that p and $\overline{v}$ are comonotone and $p \succcurlyeq_{\text{c.v.}} \overline{v}$. Comonotonicity of p and $\overline{v}$ implies that there exists a monotone function $\hat{p}$ where $p(y) = \hat{p}(\overline{v}(y; a))$. By the main result of Kleiner et al. (2021), $\hat{p}$ has to be a linear combination of a set of monotone functions each of which partition the possible values of $\overline{v}$ into a collection of intervals $\mathcal{I} = \{I_\alpha\}_{\alpha \in A} \cup \{J_\beta\}_{\beta \in B}$ for which either $\hat{p}(v) = v, \forall v \in I_\alpha$ or $\hat{p}(v) = \mathbb{E}[\overline{v} | \overline{v} \in J_\beta], \forall v \in J_\beta$. We can represent each function with its associated partition $\mathcal{I}$. By the Krein–Milman theorem, $\hat{p}(v)$ must be a convex combination. For the sake of convenience, suppose that there are a finite[25] number of such function $\{\hat{p}_j(v)\}_{j=1}^J$ with partitions $\{\mathcal{I}_j\}_{j=1}^J = \left\{\{I_\alpha\}_{\alpha \in A_j} \cup \{J_\beta\}_{\beta \in B_j}\right\}_j$ and a probability distribution $\{\tau_j\}_{j=1}^J$ so that

$$\hat{p}(v) = \sum_{j=1}^{J} \hat{p}_j(v)\, \tau_j, \forall v$$

[25] It is fairly straightforward to see that this proof generalizes to arbitrary distributions. In order to avoid clutter we omit the general case.

We can define $S=\bigcup_j\{s_j\}\times\left(\bigcup_{\alpha\in A_j}I_\alpha\cup B_j\right)$ and

$$\pi\left(s_j,C|v\right)=\sum_{j=1}^{J}\tau_j\mathbf{1}\left[v\in C\text{ or }\exists\beta\in C,v\in J_\beta\right],\forall C\subset\bigcup_{\alpha\in A_j}I_\alpha\cup B_j$$

In words, π is associated with a signal that reveals which partition $\mathcal{I}_j$ is used with probability τ_j and then reveals v if the signal associated with $\mathcal{I}_j$ reveals v and otherwise the interval J_β that v belongs to. Under this signal, the market posterior $\mathbb{E}\left[v|s_j\right]$ is either $\overline{v}$ if $\overline{v}$ is fully revealed in $\mathcal{I}_j$ or it is $\mathbb{E}\left[v|v\in J_\beta\right],\beta\in B_j$ if s_j and interval J_β are revealed. Since these values are equal to $\hat{p}_j(v)$, it implies that $\hat{p}(v)=\mathbb{E}\left[\mathbb{E}\left[\overline{v}|s\right]\right]$. This concludes the proof. □

A.2 Proof of Lemma 1

Proof. In this proof, we assume that $-\infty<v_Q(0)<v_Q(1)<\infty$. The cases with $v_Q(1)=\infty$ or $v_Q(0)=-\infty$ can be proved using a limiting argument. Before proceeding, we prove the following lemma:

Lemma 2. *Let $p,\overline{v}$ be comonotone and $p_Q(i),\overline{v}_Q(i)$ be their associated quantile representation as defined in (6). If $F_p(v),F_v(v)$ are the cumulative distribution functions of $p,\overline{v}$ respectively, then $p\succcurlyeq_{\text{cv}}\overline{v}$ if and only if $F_v(v)\succcurlyeq_{\text{cv}}F_p(v)$ where v is uniformly distributed over $V=[v_Q(0),v_Q(1)]$. In other words,*

$$\int_0^1\phi\left(p_Q(i)\right)di\geq\int_0^1\phi\left(\overline{v}_Q(i)\right)di,\forall\phi:V\to\mathbb{R}:\textit{concave}\Leftrightarrow$$
$$\int_V\psi\left(F_v(v)\right)dv\geq\int_V\psi\left(F_p(v)\right)dv,\forall\psi:[0,1]\to\mathbb{R}:\textit{concave}$$

An example that illustrates Lemma 2 is depicted in Figure 7. On the left, we have an example of $p,\overline{v}$ (not their quantile version) where p is a mean–preserving contraction of $\overline{v}$ while this is reversed for their c.d.f.'s. The idea behind Lemma 2 is simple. Since c.d.f.'s are inverses of the quantile functions mean preserving contraction for one implies mean preserving spread for the other. We provide its proof in the online Appendix. Note that in the above, we can view $i=F_v(v)$ as having a distribution according to $\frac{v_Q(i)-v_Q(0)}{v_Q(1)-v_Q(0)}$ and $j=F_p(v)$ as having a distribution $\frac{p_Q(j)-v_Q(0)}{v_Q(1)-v_Q(0)}$ with probability $\frac{v_Q(1)-p_Q(1)}{v_Q(1)-v_Q(0)}$ on $j=1$.

Now, if $p\succcurlyeq_{\text{cv}}\overline{v}$, Lemma 2 implies that $F_v\succcurlyeq_{\text{cv}}F_p$. By Blackwell's theorem (see also Kolotilin (2018) and Gentzkow and Kamenica (2016)), there must exist a signal structure (or a garbling) where $F_v(v)=i$ is the conditional mean of $F_p=j$ upon realization of the signal with i having distribution $dv_Q(i)/\left(v_Q(i)-v_Q(0)\right)$ and similarly for j. Let $\mu(\cdot|i)\in\Delta[0,1]$ be the posterior

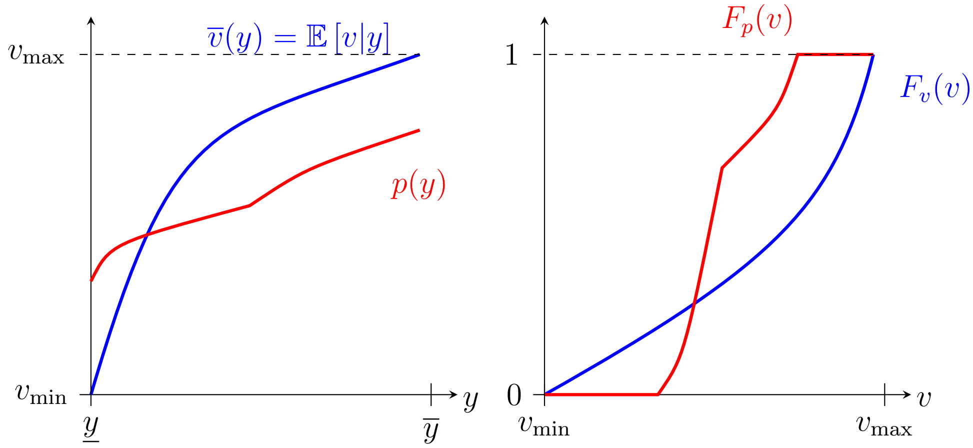


Figure 7: Example of $p \succcurlyeq_{\text{cv}} \overline{v}$ (left) and their associated CDF's (right) that satisfy $F_v \succcurlyeq_{\text{cv}} F_p$

associated with $F_v(v) = i$. Since the distribution of i is given by $dv_Q(i) / (v_Q(i) - v_Q(0))$ we can write Bayes plausibility as

$$\int \mu(A|i) \frac{dv_Q(i)}{|V|} = \frac{\int_A dp_Q(j)}{|V|}, A \subset [0,1] \tag{12}$$

such that

$$i = \int j d\mu(j|i), \forall i \in [0,1] \tag{13}$$

where in the above $|A|$ is the Lebesgue measure of A and $|V| = v_Q(1) - v_Q(0)$. Note that (13) simply states that $i = F_v(v)$ is the posterior mean of $j = F_p(v)$ according to the distribution $\mu(\cdot|i)$.

We can then write

$$\begin{aligned}
\int_Y p(y) h(y) dG &= -\int_0^1 p_Q(j) dH(j) \\
&= -\int_0^1 \mu([0,j]|i) dv_Q(i) dH(j) \text{ by setting } A\text{=}[0,j] \text{ in (12)} \\
&= -\int_0^1 \int_0^1 \mu([0,j]|i) dH(j) dv_Q(i) \text{ Fubini's Theorem} \\
&= -\int_0^1 \left[H(1) - \int_0^1 H(j) d\mu(j|i)\right] dv_Q(i) \text{ Int. by parts for inside integral} \\
&= \int_0^1 \int_0^1 H(j) d\mu(j|i) dv_Q(i) \text{ since by construction } H(1) = 0
\end{aligned}$$

Now, consider the expression $\int_0^1 H(j) d\mu(j|i)$. In this expression $\mu(\cdot|i)$ is a probability distribution over j whose mean value is i. In other words, this is a convex combination of values of $H(j)$ with average of i. Given the definition of the concave envelope, we have that $\int_0^1 H(j)\,d\mu(j|i) \le \text{cav}H(i)$. Moreover, since the space of measures over $[0,1]$ is compact according to weak–* topology (Banach-Alaoglu theorem) for each i there must exist $\mu(\cdot|i) \in \Delta[0,1]$ such that $\int_0^1 H(j)\,d\mu(j|i) = \text{cav}H(i)$. We can then use the above procedure to construct an interim price function that delivers $\int \text{cav}H(i)\,dv_Q(i)$. This proves the equality (8).

To prove the second part, note that by Caratheodory theorem, for any i either $\text{cav}H(i) = H(i)$ or that there exists $i_1 < i < i_2$ such that

$$\text{cav}H(i) = \frac{i_2 - i}{i_2 - i_1} H(i_1) + \frac{i - i_1}{i_2 - i_1} H(i_2)$$

In the first case, $\mu(\{i\}|i) = 1$ and in the second case $\mu(\{i_1\}|i) = \frac{i_2 - i}{i_2 - i_1} = 1 - \mu(\{i_2\}|i)$. In other words, concavification of H partitions the range of $[0,1]$ into subintervals $(i_{1,\alpha}, i_{2,\alpha})$, $[i_{1,\beta}, i_{2,\beta}]$ where for any α there exists β such that $i_{2,\alpha} = i_{1,\beta}$ and another β for which $i_{1,\alpha} = i_{2,\beta}$ where for all $i \in [i_{1,\beta}, i_{2,\beta}]$, $\mu(\{i\}|i) = 1$ and for all $i \in (i_{1,\alpha}, i_{2,\alpha})$, $\mu(\{i_{1,\alpha}\}|i) = \frac{i_{2,\alpha} - i}{i_{2,\alpha} - i_{1,\alpha}} = 1 - \mu(\{i_{2,\alpha}\}|i)$. From above we know that

$$p_Q(j) = \int_0^1 \mu([0,j]|i)\,dv_Q(i)$$

Given the construction of μ, we have

$$\mu([0,j]|i) = \begin{cases} \mathbf{1}[j \ge i] & \text{cav}H(i) = H(i) \\ \frac{i_{2,\alpha} - i}{i_{2,\alpha} - i_{1,\alpha}} \mathbf{1}[i_{2,\alpha} > j \ge i_{1,\alpha}] + \mathbf{1}[j \ge i_{2,\alpha}] & \text{otherwise} \end{cases}$$

Now, suppose that $j \in [i_{1,\beta}, i_{2,\beta}]$ for some β. This means that if $i > j$, $\mu([0,j]|i) = 0$ since all higher quantiles either put weights only on i or on values of the form $i_{1,\alpha}, i_{2,\alpha}$ which are higher than j. Then we can write

$$\begin{aligned} p_Q(j) &= \int_0^1 \mu([0,j]|i)\,dv_Q(i) \\ &= \int_0^j dv_Q(i) = v_Q(j) \end{aligned}$$

Moreover, if $j \in (i_{1,\alpha}, i_{2,\alpha})$ for some α, then it has to be that if $i > i_{2,\alpha}$, then $\mu([0,j]|i) = 0$. So we can write

$$\begin{aligned}
p_Q(j) &= \int_0^{i_{2,\alpha}} \mu([0,j]|i)\, dv_Q(i) \\
&= \int_0^{i_{1,\alpha}} dv_Q(i) + \int_{i_{1,\alpha}}^{i_{2,\alpha}} \mu([0,j]|i)\, dv_Q(i) \\
&= v_Q(i_{1,\alpha}) + \int_{i_{1,\alpha}}^{i_{2,\alpha}} \frac{i_{2,\alpha}-i}{i_{2,\alpha}-i_{1,\alpha}} dv_Q(i) \\
&= v_Q(i_{1,\alpha}) - v_Q(i_{1,\alpha}) + \int_{i_{1,\alpha}}^{i_{2,\alpha}} \frac{v_Q(i)}{i_{2,\alpha}-i_{1,\alpha}} di \\
&= \mathbb{E}\left[v_Q(i)\,|i \in (i_{1,\alpha}, i_{2,\alpha})\right]
\end{aligned}$$

This establishes the claim. □

A.3 Proof of Theorem 1

Proof. Consider the problem of finding the best interim price function in quantile form for a given action a:

$$\max_{p_Q} W(a) - \int p_Q(i)\, dH(i;a) = \max_{p_Q} W(a) + T_H p_Q$$

subject to $p_Q \succcurlyeq_{\text{cv}} \overline{v}_Q$, monotonicity of p_Q and the first order IC constraints. We will show that for any $a \in A$, Lagrange multipliers associated with the first order IC constraints exist so that the constrained optimization gives the same value as the unconstrained optimization of the Lagrangian over the space of p_Q's that are a mean preserving contraction of $\overline{v}_Q$ and are monotone. This combined with Lemma 1 implies the desired result.

We view p_Q as a member of any arbitrary $L_p([0,1])$ space for some $p \geq 1$. Let us refer to the first order IC constraints with respect to a_n as $T_n p_Q = 0$ where T_n is an affine transformation that maps $L_p([0,1])$ to $\mathbb{R}$ (same is true for T_H) and we can define $T p_Q = (T_1 p_Q, \cdots, T_N p_Q)$ which is an affine transformation from $L_p([0,1])$ to $\mathbb{R}^N$. Finally, let us refer to the set of p_Q's that satisfy $p_Q \succcurlyeq_{\text{cv}} \overline{v}_Q$ and monotonicity of p_Q as $\mathcal{P}$ and the subset of $\mathcal{P}$ that satisfies $T p_Q = 0$ as $\mathcal{Q}$.

For any subset $S \subset \{1, \cdots, N\}$, let $S^c = \{1, \cdots, N\} \backslash S$ and let us consider the following sets

$$\mathcal{P}(S) = \{p_Q \in L_p([0,1])\,|p_Q \succcurlyeq_{\text{c.v.}} v_Q, p_Q \text{ increasing}, T_n p_Q \geq 0, n \in S, T_{n'} p_Q \leq 0, n' \in S^c\}$$

Lemma 3. *There exists* S *such that* $\max_{p_Q \in \mathcal{P}(S)} T_H p_Q = \max_{p_Q \in \mathcal{Q}} T_H p_Q$.

Proof. Since all members of $\mathcal{Q}$ satisfy $T_n p_Q = 0$, it must be that $\mathcal{Q} \subset \mathcal{P}(S)$ for all S. This implies that $\max_{p_Q \in \mathcal{P}(S)} T_H p_Q \geq \max_{p_Q \in \mathcal{Q}} T_H p_Q$. Now, suppose to the contrary that for all S, the left hand side is strictly higher than the right hand side. This implies that for any $S \subset$

$\{2, \cdots, n\}$, there exists $p \in \mathcal{P}(S), p' \in \mathcal{P}(S \cup \{1\})$ such that $T_H p, T_H p' > \max_{p_Q \in \mathcal{Q}} T_H p_Q$. Since $T_1 p \leq 0 \leq T_1 p'$ there must exists λ such that $T_1(\lambda p + (1-\lambda) p') = 0$. Let $p^{(1),S} = \lambda p + (1-\lambda) p'$ and recall that $p^{(1),S} \in \mathcal{P}(S)$. Note that we must also have that $T_H p^{(1),S} > \max_{p_Q \in \mathcal{Q}} T_H p_Q$. Now, we know that $T_2 p^{(1),S} \leq 0 \leq T_2 p^{(1),S \cup \{2\}}$ for any $S \subset \{3, \cdots, n\}$. By using the same argument, we can find $p^{(1,2),S}$ such that $T_1 p^{(1,2),S} = T_2 p^{(1,2),S} = 0, p^{(1,2),S} \in \mathcal{P}(S)$ and $T_H p^{(1,2),S} > \max_{p_Q \in \mathcal{Q}} T_H p_Q$. By continuing this construction, we can find $p^{(1,2,\cdots,N)}$ such that $T_1 p^{(1,\cdots,N)} = \cdots = T_N p^{(1,\cdots,N)} = 0$ and that $T_H p^{(1,\cdots,N)} > \max_{p_Q \in \mathcal{Q}} T_H p_Q$ which is a contradiction since $p^{(1,\cdots,N)} \in \mathcal{Q}$. □

Now, suppose that $\hat{S} \subset \{1, \cdots, N\}$ satisfies the condition in Lemma 3. Let us define $\mathcal{P} = \{p_Q | p_Q \succcurlyeq_{\text{c.v.}} v_Q, p_Q \text{ increasing}\}$. Then, T maps members of $\mathcal{P}$ into $\mathbb{R}^N$. Moreover, T maps members of $\mathcal{P}\left(\hat{S}\right)$ into a convex cone. Since the image of $\mathcal{P}\left(\hat{S}\right)$ under T is convex in $\mathbb{R}^N$, it must have a non–empty relative interior.[26] This implies that we can apply standard results for existence of Lagrange multipliers (strong duality) – see for example, Theorem 8.3.1. in Luenberger (1997). Hence, it must be that $\lambda \neq 0 \in \mathbb{R}^N$ exists such that $\lambda_n \geq 0$ for all $n \in \hat{S}$ and $\lambda_n \leq 0$ for all $n \in \hat{S}^c$ such that

$$\max_{p_Q \in \mathcal{P}} W(a) - \int p_Q(i)\, dH(i;a) =$$

$$\max_{p_Q \in \mathcal{P}} W(a) + \int \left[H(i;a) - \sum_{n=1}^{N} \lambda_n \frac{\partial}{\partial \hat{a}_n} F(i|\hat{a};a)\bigg|_{\hat{a}=a} \right] dp_Q - \sum_{n=1}^{N} \lambda_n \frac{\partial c(a)}{\partial a_n}$$

The rest of the claim follows from Lemma 1. □

A.4 Proof of Proposition 2

Proof. Given the statement of Theorem 1, we know that the unconstrained objective in (D) can be achieved by a monotone partition. Note that by Kleiner et al. (2021), the extreme points of the convex set $\mathcal{P}$ – the set of p_Q's that are mean preserving contractions of v_Q and are monotone – are associated with the monotone partitions. In what follows, we show that under the Assumption 2, no two extreme points of $\mathcal{P}$ can deliver the same value of the Lagrangian

$$L(p_Q, \lambda; a) = W(a) - \int p_Q(i)\, d\left(H(i;a) - \sum_{n=1}^{N} \lambda_n \frac{\partial}{\partial \hat{a}_n} F(i|\hat{a};a)\bigg|_{\hat{a}=a} \right) - \sum_{n=1}^{N} \lambda_n \frac{\partial c(a)}{\partial a_n}$$

This would imply that $L(p_Q, \lambda; a)$ has a unique maximand for any λ which establishes the claim.

[26] See for example Theorem 6.2 in Rockafellar (1970).

Suppose that there are two interim price functions $p_1, p_2 \in \mathcal{P}$ that are associated with monotone partitions. If $p_1 \neq p_2$, then there must exist an interval $I \subset [0,1]$ so that all of its members satisfy $p_{1,Q}(i) = v_{1,Q}(i)$ and $p_{2,Q}(i)$ is constant for all $i \in I$. By Lemma 1, if $p_{1,Q}(i) = v_{1,Q}(i)$ is optimal for an interval I, then $\Gamma(i;a,\lambda) = H(i;a) - \sum_{n=1}^{N} \lambda_n \frac{\partial}{\partial \hat{a}_n} F(i|\hat{a};a)\Big|_{\hat{a}=a}$ should coincide with its concave envelope. Moreover, suppose that the maximal interval containing I for which $p_{2,Q}$ is constant is $\tilde{I} = (i_1, i_2)$. Suppose that the value of $p_{2,Q}$ is $\tilde{p}$ over this interval.

We show that this implies that $\Gamma(i;a,\lambda)$ is linear over the interval I. Suppose to the contrary that for some sub–interval $I' \subset I$, $\Gamma(i;a,\lambda)$ is strictly concave. Then,

$$-\int_{\tilde{I}} p_{2,Q}(i)\, d\Gamma(i;a,\lambda) = -\tilde{p}\left[\Gamma(i_2;a,\lambda) - \Gamma(i_1;a,\lambda)\right]$$

We also have that

$$v_Q(i_1) < \tilde{p} = \frac{\int_{\tilde{I}} v_Q(i)\, di}{i_2 - i_1} < v_Q(i_2)$$

Let us define

$$\hat{p}_Q(i) = \begin{cases} p_{Q,2}(i) & i \in [0,1] \setminus \tilde{I} \\ \underline{p} & i \in (i_1, j) \\ \overline{p} & i \in (j, i_2) \end{cases}$$

where

$$\underline{p} = \frac{\int_{i_1}^{j} v_Q(i)\, di}{j - i_1}, \overline{p} = \frac{\int_{j}^{i_2} v_Q(i)\, di}{i_2 - j}$$

We have

$$\begin{aligned} -\int \hat{p}_{2,Q}(i)\, d\Gamma(i;\lambda,a) + \int p_{2,Q}(i)\, d\Gamma(i;\lambda,a) = \\ -\frac{\int_{i_1}^{j} v_Q(i)\, di}{j - i_1}\left[\Gamma(j;a,\lambda) - \Gamma(i_1;a,\lambda)\right] - \frac{\int_{j}^{i_2} v_Q(i)\, di}{i_2 - j}\left[\Gamma(i_2;a,\lambda) - \Gamma(j;a,\lambda)\right] \\ + \frac{\int_{i_1}^{i_2} v_Q(i)\, di}{i_2 - i_1}\left[\Gamma(i_2;a,\lambda) - \Gamma(i_1;a,\lambda)\right] \end{aligned}$$

In the above, since $\Gamma(i;a,\lambda)$ is strictly concave over parts of I, we must have that

$$\frac{\Gamma(j;a,\lambda) - \Gamma(i_1;a,\lambda)}{j - i_1} > \frac{\Gamma(i_2;a,\lambda) - \Gamma(j;a,\lambda)}{i_2 - j}$$

Let us also define $\pi_1 = \frac{j-i_1}{i_2-i_1} = 1 - \pi_2$. Since $\underline{p} < \overline{p}$,we can write

$$\pi_1 \underline{p} \frac{\Gamma(j;a,\lambda) - \Gamma(i_1;a,\lambda)}{j-i_1} + \pi_2 \overline{p} \frac{\Gamma(i_2;a,\lambda) - \Gamma(j;a,\lambda)}{i_2-j} <$$

$$(\pi_1 \underline{p} + \pi_2 \overline{p}) \left(\pi_1 \frac{\Gamma(j;a,\lambda) - \Gamma(i_1;a,\lambda)}{j-i_1} + \pi_2 \frac{\Gamma(i_2;a,\lambda) - \Gamma(j;a,\lambda)}{i_2-j} \right) =$$

$$\frac{\int_{i_1}^{i_2} v_Q(i)\, di}{i_2 - i_1} \frac{\Gamma(i_2;a,\lambda) - \Gamma(i_1;a,\lambda)}{i_2-i_1}$$

which implies that

$$-\int \hat{p}_{2,Q}(i)\, d\Gamma(i;\lambda,a) + \int p_{2,Q}(i)\, d\Gamma(i;\lambda,a) > 0$$

and thus $p_{2,Q}$ cannot be optimal. Therefore,

$$\Gamma'(i;a,\lambda) = c, \forall i \in I$$

for some c. Using the definition of H and F, we have

$$\Gamma'(i;a,\lambda) = -\alpha\left(G^{-1}(i|a)\right) - \frac{1}{g(y|a)} \sum \lambda_n \frac{\partial g\left(G^{-1}(i|a)\,|a\right)}{\partial a_n} = c$$

This is indeed in contradiction with the independence assumption which establishes the claim. □

A.5 Proof of Proposition 3

Proof. In this case, the function $\Gamma(i;a,\lambda)$ satisfies

$$\Gamma'(i;a,\lambda) = -\lambda \frac{\partial g\left(G^{-1}(i|a)\,|a\right)}{\partial a} \frac{1}{g\left(G^{-1}(i|a)\,|a\right)}$$

Since g exhibits MLRP, if $\lambda > 0$ then, Γ' is decreasing in i and so Γ is concave. If $\lambda < 0$, then Γ' is increasing in i and so Γ is convex.

By Lemma 1, if Γ is concave, optimal rating should be fully informative which proves the desired result. If on the other hand, Γ is convex, then the optimal rating should be uninformative and as a result $\int \frac{\partial F}{\partial a} dp_Q = 0$ which means that no effort with a positive cost can be incentivized. □

A.6 Proof of Proposition 4

Proof. Recall that $\Gamma(i;\lambda,a) = -\lambda \left.\frac{\partial F(i;\hat{a},a)}{\partial \hat{a}}\right|_{\hat{a}=a}$ and as we have shown in section 4,

$$\Gamma''(i;a,\lambda) = -\lambda \frac{1}{g(y|a)} \left.\frac{\partial^2 \log g(y|a)}{\partial a \partial y}\right|_{y=G^{-1}(i|a)}$$

Given our definition of ELRP, when λ is negative, the above is concave–convex and when λ is positive, the above is convex–concave. We wish to show that under ELRP, λ is positive and thus optimal rating has to be lower censorship.

Suppose to the contrary that λ is negative. In this case, since $\Gamma(0;a,\lambda) = \Gamma(1;a,\lambda) = 0$, there are two possibilities: 1. $\Gamma'(0;a,\lambda) < 0$ in which case $\Gamma(i;a,\lambda)$ is non–positive for all values of i and its concave envelope is the zero function associated with no information; 2. $\Gamma'(0;a,\lambda) > 0$ in which case for an interval $[0,i_1]$ the concave envelope coincides with Γ and for higher values cavΓ is linear. This is associated with an upper-censorship optimal rating. This is depicted in Figure 8 on the left. In the first case, the marginal return to effort is zero and effort with positive marginal cost cannot be supported.

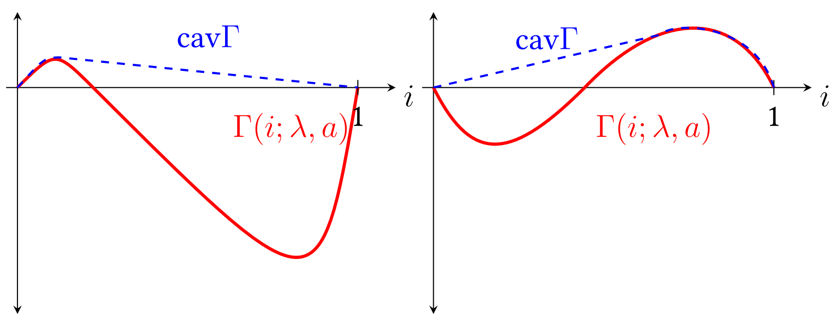


Figure 8: Concave envelope of the marginal change in distribution is concave–convex (left) and convex–concave (right).

In the second case, let y_1 be the value of indicator associated with i_1. In this case, the marginal benefit of effort is given by

$$\int p(y) \frac{\partial g(y|a)}{\partial a} dy =$$
$$\int_{\underline{y}}^{y_1} \overline{v}(y;a) \frac{\partial \log g(y|a)}{\partial a} dG + \overline{p} \int_{y_1}^{\overline{y}} \frac{\partial g(y|a)}{\partial a} dy$$

where in the above $\overline{p}$ is the average value of $\overline{v}$ when $y \geq y_1$. Since Γ is concave over $[0,i_1]$, $\frac{\partial \log g(y|a)}{\partial a}$ is decreasing in y over the interval $[\underline{y}, y_1]$ and thus, using the fact that $\overline{v}$ is increasing in y, we can use Chebyshev's sum inequality to write the above as

$$\int_{\underline{y}}^{y_1} \overline{v}(y;a) \frac{\partial \log g(y|a)}{\partial a} dG + \overline{p} \int_{y_1}^{\overline{y}} \frac{\partial g(y|a)}{\partial a} dy \leq$$
$$\frac{\int_{\underline{y}}^{y_1} \overline{v}(y;a)\, dG}{G(y_1|a)} \int_{\underline{y}}^{y_1} \frac{\partial \log g(y|a)}{\partial a} dG + \overline{p} \int_{y_1}^{\overline{y}} \frac{\partial g(y|a)}{\partial a} dy =$$
$$\frac{\int_{\underline{y}}^{y_1} \overline{v}(y;a)\, dG}{G(y_1|a)} \frac{\partial G(y_1|a)}{\partial a} - \overline{p} \frac{\partial G(y_1|a)}{\partial a}$$

where in the above we have used the fact that $\frac{\partial G(\overline{y}|a)}{\partial a} = 0$. Since $\Gamma(i_1) = -\lambda \frac{\partial G(y_1|a)}{\partial a} > 0$, the above expression satisfies

$$\left(\mathbb{E}[\overline{v}|y \leq y_1] - \mathbb{E}[\overline{v}|y \geq y_1]\right) \frac{\partial G(y_1|a)}{\partial a} \leq 0$$

which cannot be the case since the cost of effort is increasing. Hence, $\lambda \geq 0$ and thus optimal rating is lower censorship. When g exhibits CLRP, the argument is the mirror of the current argument. □

A.7 Proof of Proposition 5

Proof. To show the result, it is sufficient to show that cavΓ cannot coincide with Γ for an interval of values of i including 0. Suppose that to contrary that there exists an interval $[0, i_1]$ where cav$\Gamma = \Gamma$ and as a result Γ is concave in $[0, i_1]$. Consider the linear function $\tilde{\Gamma}(i) = (1-i) \int_0^1 \alpha\left(G^{-1}(i|a)\right) di$. This function coincides with $\Gamma(i;\lambda,a)$ at $i = 0, 1$ since $\frac{\partial G}{\partial a}\left(G^{-1}(0|a)\,|a\right) = \frac{\partial G}{\partial a}\left(G^{-1}(1|a)\,|a\right) = 0$. This implies that any concave function that is above Γ is also (weakly) higher than $\tilde{\Gamma}$ and thus for all values of i, cav$\Gamma(i;\lambda,a) \geq \tilde{\Gamma}(i)$. Since by our contrary assumption for all values of $i \in [0, i_1]$, Γ is concave and by the Assumption in the statement of the Proposition

$$\Gamma'(0;a,\lambda) = -\alpha\left(G^{-1}(0|a)\right) - \lambda \frac{\partial G\left(G^{-1}(0|a)\,|a\right)}{\partial a} < \tilde{\Gamma}'(i)$$

we must have that for all values of $i \in [0, i_1]$,

$$\Gamma(i;a,\lambda) < \tilde{\Gamma}(i)$$

As we argued, $\tilde{\Gamma}(i) \leq \text{cav}\Gamma(i;a,\lambda)$ which with the above gives a contradiction. □

A.8 Proof of Proposition 9

Proof. If we let $\hat{c}(a) = a^2/2$, then the optimal rating design in this case is given by

$$\max_{p,a_1,a_2} a_1 - \kappa\hat{c}(a_1) - \kappa\hat{c}(a_2)$$

subject to

$$\int_{-\infty}^{\infty} p(y) \frac{\partial g(y|a_1,a_2)}{\partial a_1} dy = \kappa\hat{c}'(a_1)$$
$$\int_{-\infty}^{\infty} p(y) \frac{\partial g(y|a_1,a_2)}{\partial a_2} dy = \kappa\hat{c}'(a_2)$$
$$p \succcurlyeq_{\text{c.v.}} \overline{v}(y;a), p\text{: non-decreasing}$$

The proof that in the above, p is determined by the concavification of

$$\Gamma(i;\lambda,a) = -\lambda_1 \frac{\partial G(y|a)}{\partial a_1} - \lambda_2 \frac{\partial G(y|a)}{\partial a_2}\Big|_{y=G^{-1}(i|a)}$$

is identical to that of Theorem 1. Note that

$$y|a \sim \mathcal{N}\left(\mu(a), \sigma(a)^2\right),$$
$$\mu(a) = ba_1 + a_2, \sigma(a) = \sqrt{(ba_1)^2 + a_2^2}$$

This implies that $G(y|a) = \Phi\left(\frac{y-\mu(a)}{\sigma(a)}\right)$ and therefore

$$\Gamma(G(y|a);\lambda,a) = \frac{\phi\left(\frac{y-\mu(a)}{\sigma(a)}\right)}{\sigma(a)}\left(\lambda_1 b + \lambda_2 + \frac{y-\mu(a)}{\sigma(a)}\frac{b^2a_1\lambda_1 + a_2\lambda_2}{\sigma(a)}\right)$$

where $\phi(x) = \frac{e^{-x^2/2}}{\sqrt{2\pi}}$ with $\phi'(x) = -x\phi(x)$. As a result,

$$\Gamma'(G(y|a);\lambda,a) = \frac{b^2a_1\lambda_1 + a_2\lambda_2}{\sigma(a)^2} + \frac{\phi'\left(\frac{y-\mu(a)}{\sigma(a)}\right)}{\sigma(a)\phi\left(\frac{y-\mu(a)}{\sigma(a)}\right)}\left(\lambda_1 b + \lambda_2 + \frac{y-\mu(a)}{\sigma(a)}\frac{b^2a_1\lambda_1 + a_2\lambda_2}{\sigma(a)}\right)$$
$$= \frac{b^2a_1\lambda_1 + a_2\lambda_2}{\sigma(a)^2} - \frac{y-\mu(a)}{\sigma(a)^2}\left(\lambda_1 b + \lambda_2 + \frac{y-\mu(a)}{\sigma(a)}\frac{b^2a_1\lambda_1 + a_2\lambda_2}{\sigma(a)}\right)$$
$$\Gamma''(G(y|a);\lambda,a)\,g(y|a) = -\frac{1}{\sigma(a)^2}\left(\lambda_1 b + \lambda_2 + 2\frac{y-\mu(a)}{\sigma(a)}\frac{b^2a_1\lambda_1 + a_2\lambda_2}{\sigma(a)}\right)$$

The right hand side of the last expression is linear in y and thus changes sign only once. This means that optimal rating is either lower or upper censorship.

Now, consider a lower-censorship rating that pools values of y below $\hat{y}$ and reveals those above it. In Online Appendix, we show that the welfare increases as $\hat{y}$ decreases. This implies that full information ratings deliver higher level of welfare than lower censorship. Additionally, we show that the welfare associated with an upper-censorship rating that pools values of y above $\hat{y}$, is decreasing at $\hat{y} = \infty$. Since $\hat{y} = \infty$ is full revelation, this proves the result. □

A.9 Proof of Proposition 10

Proof. It is immediate by using an argument similar to Theorem 1 that optimal ratings can be found by concavification of the function

$$\Gamma(i) = -\alpha_P f_P G\left(\overline{G}^{-1}(i)\,|a_P\right) - \alpha_R f_R G\left(\overline{G}^{-1}(i)\,|a_R\right) - \lambda_P G_a\left(\overline{G}^{-1}(i)\,|a_P\right) - \lambda_R G_a\left(\overline{G}^{-1}(i)\,|a_R\right)$$

for some Lagrange multipliers λ_P, λ_R. We have

$$\Gamma'(i) = -\alpha_P f_P \frac{g\left(\overline{G}^{-1}(i)\,|a_P\right)}{\overline{g}\left(\overline{G}^{-1}(i)\right)} - \alpha_R f_R \frac{g\left(\overline{G}^{-1}(i)\,|a_R\right)}{\overline{g}\left(\overline{G}^{-1}(i)\right)} - \lambda_P \frac{g_a\left(\overline{G}^{-1}(i)\,|a_P\right)}{\overline{g}\left(\overline{G}^{-1}(i)\right)} - \lambda_R \frac{g_a\left(\overline{G}^{-1}(i)\,|a_R\right)}{\overline{g}\left(\overline{G}^{-1}(i)\right)}$$

where in the above $\overline{g}(y) = f_P g(y|a_P) + f_R g(y|a_R)$. Given the functional form of g, we have

$$\begin{aligned} g(y|a) &= e^{f(y)+r(y)m(a)-b(a)} \\ \frac{g(y|a_P)}{g(y|a_R)} &= e^{r(y)(m(a_P)-m(a_R))+b(a_R)-b(a_P)} \\ \frac{g_a(y|a)}{g(y|a)} &= m'(a)\,r(y) - b'(a) \end{aligned}$$

Replacing in the formula for Γ' implies

$$\Gamma'\left(\overline{G}(y)\right) = -\alpha_P f_P \frac{1}{f_P + f_R \frac{g(y|a_R)}{g(y|a_P)}} - \alpha_R f_R \frac{\frac{g(y|a_R)}{g(y|a_P)}}{f_P + f_R \frac{g(y|a_R)}{g(y|a_P)}} - \lambda_P \frac{\frac{g_a(y|a_P)}{g(y|a_P)}}{f_P + \frac{g(y|a_R)}{g(y|a_P)}} - \lambda_R \frac{\frac{g(y|a_R)}{g(y|a_P)}\frac{g_a(y|a_R)}{g(y|a_R)}}{f_P + f_R \frac{g(y|a_R)}{g(y|a_P)}}$$

If we refer to $m(a_R) - m(a_P)$ as Δm and similarly for $b(a_R) - b(a_P)$, we can write the above as

$$\Gamma'\left(\overline{G}(y)\right) = -\frac{\alpha_P f_P + \alpha_R f_R e^{r\Delta m - \Delta b} + \lambda_P (m'_P r - b'_P) + \lambda_R (m'_R r - b'_R)\, e^{r\Delta m - \Delta b}}{f_P + f_R e^{r\Delta m - \Delta b}}$$

If we define $x = e^{r\Delta m}$, then the above has the form

$$-\frac{A_1 + A_2 x + A_3 \log x + A_4 x \log x}{B_1 + B_2 x}$$

Note that we can argue that $\lambda_R > 0$. This is because if we consider the problem by replacing the IC for the R type with its inequality version imposing that the marginal return to a_R be higher than its cost. In this problem, if this constraint remains slack, one can simply increase a_R and shifts all $p\left(y\right)$'s upwards by the same amount and improve the payoffs. Hence, $\lambda_R \geq 0$. Since $m_R' \geq 0$ by assumption, we must have that $A_4 > 0$.

We then have

$$\begin{aligned} \left(B_1 + B_2 x\right)^2 \frac{\Gamma''\left(\overline{G}\left(y\right)\right)\overline{g}\left(y\right)}{\frac{dx}{dy}} = & \left(B_2 A_3 - B_1 A_4\right)\log x - B_1 A_3/x - B_2 A_4 x \\ & + B_2\left(A_1 - A_3\right) - B_1\left(A_1 + A_4\right) \\ = & B_2 A_4\left(\alpha_1 \log x + \alpha_2/x - x + \alpha_3\right) \end{aligned}$$

Note that in the above $A_4 B_2 > 0$. The derivative of the above with respect to x is given by $\frac{\alpha_1}{x} - \alpha_2/x^2 - 1 = \frac{\alpha_1 x - \alpha_2 - x^2}{x}$. Suppose that $\alpha_2 > 0$. Since the numerator is a quadratic function, it has at most two roots and this means that Γ'' switches sign at most three times. This establishes the claim. $\square$

Online Appendix

B Proof of Optimality of Upper Censorship in Section 5.2

Proof. We have established that the optimal rating is either upper or lower censorship.

Consider a rating that pools that values of y below $\hat{y}$. Since $y \sim \mathcal{N}\left(\mu\left(a\right), \sigma\left(a\right)^2\right)$ and that $\frac{d\overline{v}}{dy} = \beta\left(a\right)$, we can decompose the marginal return to each action, a_1, a_2, into their effect on the mean and the variance of the distribution. Specifically, the expected interim price when the mean and variance chosen by the agent are m and s^2 and the market belief is μ, σ^2 is given by

$$\beta\left(a\right) \times \left(p_L \Phi\left(\frac{\hat{y} - m}{s}\right) + \int_{\hat{y}}^{\infty} \frac{y}{s}\phi\left(\frac{y-m}{s}\right) dy\right)$$

where $p_L = \frac{\int_{-\infty}^{\hat{y}} y d\Phi\left(\frac{y-\mu}{\sigma}\right)}{\Phi\left(\frac{\hat{y}-\mu}{\sigma}\right)}$. If we let $\hat{z} = \frac{\hat{y}-m}{s}$, then we can rewrite the above as

$$\beta\left(a\right) \times \left(p_L \Phi\left(\hat{z}\right) + \int_{\hat{z}}^{\infty} \left(sz + m\right)\phi\left(z\right) dz\right)$$

The derivatives of the term in the bracket with respect to m and s after imposing $m = \mu$ and $s = \sigma$ are given by

$$r_L\left(\hat{z}\right) = -p_L \frac{1}{\sigma}\phi\left(\hat{z}\right) + \left(\frac{\mu}{\sigma} + \hat{z}\right)\phi\left(\hat{z}\right) + \int_{\hat{z}}^{\infty} \phi\left(z\right) dz$$
$$t_L\left(\hat{z}\right) = -p_L \frac{\hat{z}}{\sigma}\phi\left(\hat{z}\right) + \left(\sigma\hat{z} + \mu\right)\frac{\hat{z}}{\sigma}\phi\left(\hat{z}\right) + \int_{\hat{z}}^{\infty} z\phi\left(z\right) dz$$

Additionally, we can use integration by parts and write

$$p_L = \frac{\int_{-\infty}^{\hat{y}} y d\Phi\left(\frac{y-\mu}{\sigma}\right)}{\Phi\left(\frac{\hat{y}-\mu}{\sigma}\right)} = \frac{\sigma\int_{-\infty}^{\hat{z}} z\phi\left(z\right) dz}{\Phi\left(\hat{z}\right)} + \mu$$
$$= -\sigma\frac{\int_{-\infty}^{\hat{z}} \Phi\left(z\right) dz}{\Phi\left(\hat{z}\right)} + \mu + \sigma\hat{z}$$

We can replace for p_L in $r_L\left(\hat{z}\right)$ and $t_L\left(\hat{z}\right)$ to arrive at

$$r_L\left(\hat{z}\right) = \frac{\int_{-\infty}^{\hat{z}} \Phi\left(z\right) dz}{\Phi\left(\hat{z}\right)}\phi\left(\hat{z}\right) + 1 - \Phi\left(\hat{z}\right)$$
$$t_L\left(\hat{z}\right) = \frac{\int_{-\infty}^{\hat{z}} \Phi\left(z\right) dz}{\Phi\left(\hat{z}\right)}\hat{z}\phi\left(\hat{z}\right) + \int_{\hat{z}}^{\infty} z\phi\left(z\right) dz = s_L\left(\hat{z}\right) r_L\left(\hat{z}\right)$$

Taking a derivative of the above gives us

$$r_L'(\hat{z}) = \int_{-\infty}^{\hat{z}} \Phi(z)\, dz \frac{d}{d\hat{z}} \frac{\phi(\hat{z})}{\Phi(\hat{z})} < 0$$

$$s_L'(\hat{z}) = \frac{d}{d\hat{z}} \frac{\frac{\int_{-\infty}^{\hat{z}} \Phi(z)dz}{\Phi(\hat{z})} \hat{z}\phi(\hat{z}) + \int_{\hat{z}}^{\infty} z\phi(z)\, dz}{\frac{\int_{-\infty}^{\hat{z}} \Phi(z)dz}{\Phi(\hat{z})} \phi(\hat{z}) + 1 - \Phi(\hat{z})} = \frac{d}{d\hat{z}} \frac{\frac{\int_{-\infty}^{\hat{z}} \Phi(z)dz}{\Phi(\hat{z})} \hat{z}\phi(\hat{z}) + \hat{z}(1 - \Phi(z)) + \int_{\hat{z}}^{\infty} (1 - \Phi(z))\, dz}{\frac{\int_{-\infty}^{\hat{z}} \Phi(z)dz}{\Phi(\hat{z})} \phi(\hat{z}) + 1 - \Phi(\hat{z})}$$

$$= \frac{d}{d\hat{z}} \hat{z} + \frac{\int_{\hat{z}}^{\infty} (1 - \Phi(z))\, dz}{\frac{\int_{-\infty}^{\hat{z}} \Phi(z)dz}{\Phi(\hat{z})} \phi(\hat{z}) + 1 - \Phi(\hat{z})} = 1 - \frac{1 - \Phi(\hat{z})}{r_L(\hat{z})} - \frac{\int_{\hat{z}}^{\infty} (1 - \Phi(z))\, dz}{r_L(\hat{z})} \frac{r_L'(\hat{z})}{r_L(\hat{z})} > 0$$

where the above holds because $\phi(z)/\Phi(z)$ is decreasing or $\Phi(z)$ is log–concave. The last inequality holds because $1 - \Phi(\hat{z}) < r_L(\hat{z})$ and that $r_L'(\hat{z}) < 0$. Note further that $r_L(-\infty) = 1, r_L(\infty) = 0$. Since r_L is strictly decreasing, we can thus define the function $\hat{s}_L(r) = s_L\left(r_L^{-1}(r)\right)$. By varying r between 0 and 1, the function $\hat{s}_L(r)$ decreases to 0 as r increases to 1.

Given $\hat{y}$, the best response of the agent should satisfy

$$\beta(a)\, rb + \beta(a) \frac{b^2 a_1}{\sigma(a)} r\hat{s}_L(r) = \kappa\hat{c}'(a_1)$$

$$\beta(a)\, r + \beta(a) \frac{a_2}{\sigma(a)} r\hat{s}_L(r) = \kappa\hat{c}'(a_2)$$

Since $\beta(a) = ba_1^2/\sigma(a)^2$, if we let $x = b\beta(a)$, then, $a_1 = \frac{\sigma\sqrt{x}}{b}, a_2 = \sigma\sqrt{1-x}$ and we can write the above as

$$rx\left(1 + \sqrt{x}\hat{s}_L(r)\right) = \kappa\frac{\sigma\sqrt{x}}{b}$$

$$r\frac{x}{b}\left(1 + \sqrt{1-x}\hat{s}_L(r)\right) = \kappa\sigma\sqrt{1-x}$$

which determine x, σ for a given value of r. Since x and σ determine a_1, a_2, we can refer to the values as $\hat{a}_1(r), \hat{a}_2(r)$ If we divide the top equation by the bottom one, we have

$$b^2 \frac{1 + \sqrt{x}\hat{s}_L(r)}{1 + \sqrt{1-x}\hat{s}_L(r)} = \sqrt{\frac{x}{1-x}}$$

and this implies that

$$\hat{s}_L(r) = \frac{1}{1 - b^2}\left(\frac{b^2}{\sqrt{x}} - \frac{1}{\sqrt{1-x}}\right) \tag{14}$$

Let us refer to the solution of this as $\hat{x}(r)$. The right hand side of the above is decreasing in x while the LHS is decreasing in r. This means that an increase in r increases $\hat{x}(r)$. Thus, the highest value of x is associated with $r = 1$ and $\hat{s}_L(1) = 0$ which is given by $\frac{b^4}{1+b^4}$. We also have that

$$\kappa\hat{a}_1(r) = r\hat{x}(r)(1+\hat{x}(r)\hat{s}_L(r)) = r\hat{x}(r)\left(1-\sqrt{\frac{\hat{x}(r)}{1-\hat{x}(r)}}\right)$$

The function $x\left(1-\sqrt{\frac{x}{1-x}}\right)$ is maximized at $\overline{x} = 1-\frac{1}{2}\sqrt[3]{\frac{2-\sqrt{2}}{4}}-\frac{1}{2}\sqrt[3]{\frac{2+\sqrt{2}}{4}} \approx 0.262$ and is increasing below this value. This implies that as long as $\frac{b^4}{1+b^4} \leq \overline{x} \to b \leq 0.772$, an increase in r leads to an increase in x and as a result a_1. Hence, the highest value of $\hat{a}_1$ is attained at $r=1$ and is given by $\frac{1}{\kappa}\frac{b^4}{1+b^4}$.

Note also that $a_2 = ba_1\sqrt{\frac{1-x}{x}}$ and thus total surplus is given by

$$\hat{W}(r) = \hat{a}_1(r) - \kappa\left(\frac{\hat{a}_1(r)^2}{2}+\frac{\hat{a}_2(r)^2}{2}\right) = \hat{a}_1(r) - \frac{\kappa}{2}\hat{a}_1(r)^2\left(1+b^2\frac{1-\hat{x}(r)}{\hat{x}(r)}\right)$$

The unconstrained optimal value of a_1 for a given x is $\frac{1}{\kappa}\frac{1}{1+b^2\frac{1-x}{x}}$. This value is decreasing in x and since $x \leq \frac{b^4}{1+b^4}$, it attains its lowest value at $\frac{1}{\kappa}\frac{1}{1+b^{-2}} = \frac{1}{\kappa}\frac{b^2}{1+b^2}$. Since the above function is hump–shaped in a_1, and is maximum value is always above $\hat{a}_1(1)$ and from above we know that $\hat{a}_1(r) \leq \hat{a}_1(1)$, it must be that total surplus satisfies

$$\begin{aligned}\hat{W}(r) &\leq \hat{a}_1(1) - \frac{\kappa}{2}\hat{a}_1(1)^2\left(1+b^2\frac{1-\hat{x}(r)}{\hat{x}(r)}\right)\\ &\leq \hat{a}_1(1) - \frac{\kappa}{2}\hat{a}_1(1)^2\left(1+b^2\frac{1-\hat{x}(1)}{\hat{x}(1)}\right) = \hat{W}(1)\end{aligned}$$

which implies that the best lower-censorship rating is full revelation, i.e., $r(\hat{z}) = 1 \to \hat{z} = -\infty$.

Now, consider an upper-censorship rating that pools values of y above $\hat{y}$. In this case, when the mean and variance chosen by the agent are m and s^2 and the market belief is μ, σ^2 is given by

$$\beta(a) \times \left(p_H\left(1-\Phi\left(\frac{\hat{y}-m}{s}\right)\right)+\int_{-\infty}^{\hat{y}}\frac{y}{s}\phi\left(\frac{y-m}{s}\right)dy\right)$$

where $p_L = \frac{\int_{\hat{y}}^{\infty} yd\Phi\left(\frac{y-\mu}{\sigma}\right)}{1-\Phi\left(\frac{\hat{y}-\mu}{\sigma}\right)}$. If we let $\hat{z} = \frac{\hat{y}-m}{s}$, then we can rewrite the above as

$$\beta(a) \times \left(p_H(1-\Phi(\hat{z}))+\int_{-\infty}^{\hat{z}}(sz+m)\phi(z)dz\right)$$

The derivatives of the term in the bracket with respect to m and s after imposing $m=\mu$ and

$s = \sigma$ are given by

$$r_U\left(\hat{z}\right) = p_H \frac{1}{\sigma}\phi\left(\hat{z}\right) - \left(\frac{\mu}{\sigma} + \hat{z}\right)\phi\left(\hat{z}\right) + \int_{-\infty}^{\hat{z}} \phi\left(z\right) dz$$

$$t_U\left(\hat{z}\right) = p_H \frac{\hat{z}}{\sigma}\phi\left(\hat{z}\right) - \left(\sigma\hat{z} + \mu\right)\frac{\hat{z}}{\sigma}\phi\left(\hat{z}\right) + \int_{-\infty}^{\hat{z}} z\phi\left(z\right) dz$$

We can use the same simplification as above and write

$$r_U\left(\hat{z}\right) = \frac{\int_{\hat{z}}^{\infty}\left(1 - \Phi\left(z\right)\right) dz}{1 - \Phi\left(\hat{z}\right)}\phi\left(\hat{z}\right) + \Phi\left(\hat{z}\right)$$

$$t_U\left(\hat{z}\right) = \frac{\int_{\hat{z}}^{\infty}\left(1 - \Phi\left(z\right)\right) dz}{1 - \Phi\left(\hat{z}\right)}\hat{z}\phi\left(\hat{z}\right) + \int_{-\infty}^{\hat{z}} z\phi\left(z\right) dz = s_U\left(\hat{z}\right) r_U\left(\hat{z}\right)$$

Similar to the above, we can show that r_U is increasing with values between 0 and 1 while s_U is increasing and negative with values between $-\infty$ and 0.

Similar to before, FOCs are given by

$$rx\left(1 + \sqrt{x}\hat{s}_U\left(r\right)\right) = \kappa\frac{\sigma\sqrt{x}}{b} = \kappa a_1$$

$$r\frac{x}{b}\left(1 + \sqrt{1 - x}\hat{s}_U\left(r\right)\right) = \kappa\sigma\sqrt{1 - x} = \kappa a_2$$

From before, at $r = 1$, $x = \frac{b^4}{1+b^4}$, and we have

$$\hat{s}_U\left(r\right) = \frac{1}{1 - b^2}\left(\frac{b^2}{\sqrt{x}} - \frac{1}{\sqrt{1 - x}}\right)$$

Taking a derivative of the above at $r = 1$, we have

$$\hat{s}_U'\left(r\right) = \frac{1}{1 - b^2}\left(-\frac{b^2}{2x\sqrt{x}} - \frac{1}{2\left(1 - x\right)\sqrt{1 - x}}\right)x'\left(r\right)$$

Since $\hat{s}_U'\left(r\right) \geq 0$, the above implies that $x'\left(r\right) \leq 0$. Note also that

$$s_U'\left(\hat{z}\right) r_U\left(\hat{z}\right) + s_U\left(\hat{z}\right) r_U'\left(\hat{z}\right) = t_U'\left(\hat{z}\right)$$

$$\int_{\hat{z}}^{\infty}\left(1 - \Phi\left(z\right)\right) dz \frac{d}{d\hat{z}}\left(\frac{\hat{z}\phi\left(\hat{z}\right)}{1 - \Phi\left(\hat{z}\right)}\right) = t_U'\left(\hat{z}\right)$$

$$\int_{\hat{z}}^{\infty}\left(1 - \Phi\left(z\right)\right) dz \frac{d}{d\hat{z}}\left(\frac{\phi\left(\hat{z}\right)}{1 - \Phi\left(\hat{z}\right)}\right) = r_U'\left(\hat{z}\right)$$

We know that $\hat{s}_U(r_U(\hat{z})) = s_U(\hat{z}) \to \hat{s}'_U(r_U(\hat{z})) = \frac{s'_U(\hat{z})}{r'_U(\hat{z})}$ and hence,

$$\hat{s}'_U(r_U(\hat{z})) = \frac{s'_U(\hat{z})}{r'_U(\hat{z})} = \frac{t'_U(\hat{z})}{r_U(\hat{z})\, r'_U(\hat{z})} - \frac{s_U(\hat{z})}{r_U(\hat{z})}$$

We have that

$$\begin{aligned}
\lim_{\hat{z}\to\infty} \frac{t'_U(\hat{z})}{r'_U(\hat{z})} &= \lim_{\hat{z}\to\infty} \frac{\frac{d}{d\hat{z}}\left(\frac{\hat{z}\phi(\hat{z})}{1-\Phi(\hat{z})}\right)}{\frac{d}{d\hat{z}}\left(\frac{\phi(\hat{z})}{1-\Phi(\hat{z})}\right)} \\
&= \lim_{\hat{z}\to\infty} \frac{\hat{z}\frac{d}{d\hat{z}}\left(\frac{\phi(\hat{z})}{1-\Phi(\hat{z})}\right) + \frac{\phi(\hat{z})}{1-\Phi(\hat{z})}}{\frac{d}{d\hat{z}}\left(\frac{\phi(\hat{z})}{1-\Phi(\hat{z})}\right)} \\
\lim_{\hat{z}\to\infty} \frac{d}{d\hat{z}}\left(\frac{\phi(\hat{z})}{1-\Phi(\hat{z})}\right) &= \lim_{\hat{z}\to\infty} \frac{-\hat{z}\phi(\hat{z})(1-\Phi(\hat{z})) + \phi(\hat{z})^2}{(1-\Phi(\hat{z}))^2} \\
&= \lim_{\hat{z}\to\infty} \frac{-\phi(\hat{z})(1-\Phi(\hat{z})) + \hat{z}^2\phi(\hat{z})(1-\Phi(\hat{z})) + \hat{z}\phi(\hat{z})^2 - 2\hat{z}\phi(\hat{z})^2}{-2\phi(\hat{z})(1-\Phi(\hat{z}))} \\
&= \lim_{\hat{z}\to\infty} \frac{-(1-\Phi(\hat{z})) + \hat{z}^2(1-\Phi(\hat{z})) - \hat{z}\phi(\hat{z})}{-2(1-\Phi(\hat{z}))} \\
&= \lim_{\hat{z}\to\infty} \frac{\phi(\hat{z}) + 2\hat{z}(1-\Phi(\hat{z})) - \hat{z}^2\phi(\hat{z}) - \phi(\hat{z}) + \hat{z}^2\phi(\hat{z})}{2\phi(\hat{z})} \\
&= \lim_{\hat{z}\to\infty} \frac{\hat{z}(1-\Phi(\hat{z}))}{\phi(\hat{z})} = \lim_{\hat{z}\to\infty} \frac{1-\Phi(\hat{z}) - \hat{z}\phi(\hat{z})}{-\hat{z}\phi(\hat{z})} = 1
\end{aligned}$$

Since $r_U(\hat{z}) \to 1$ as $\hat{z} \to \infty$ and $s_U(\hat{z}) \to 0$ as $\hat{z} \to \infty$, we must have that

$$\lim_{r\to 1} \hat{s}'_U(r) = \infty$$

This implies that $\hat{x}'(1) = -\infty$ and moreover since $\hat{x}(1) = \frac{b^4}{1+b^4}$:

$$\begin{aligned}
\lim_{r\to 1} \frac{\hat{s}'_U(r)}{\hat{x}'(r)} &= -\frac{1}{1-b^2}\left(\frac{b^2}{2\hat{x}(1)\sqrt{\hat{x}(1)}} + \frac{1}{2(1-\hat{x}(1))\sqrt{1-\hat{x}(1)}}\right) \\
&= -\frac{1}{2(1-b^2)}\left(\frac{b^2}{\frac{b^6}{(1+b^4)^{3/2}}} + \frac{1}{\frac{1}{(1+b^4)^{3/2}}}\right) = -\frac{(1+b^4)^{\frac{5}{2}}}{2(1-b^2)\, b^4}
\end{aligned}$$

Let us also calculate the value of $a_1'(1)$. We have

$$\begin{aligned} a_1'(r) =& r\hat{x}(r)\left(1+\sqrt{\hat{x}(r)}\hat{s}_U(r)\right) \\ =& a_1(r)\left(\frac{1}{r}+\frac{\hat{x}'(r)}{\hat{x}(r)}+\frac{\frac{\hat{x}'(r)\hat{s}_U(r)}{2\sqrt{\hat{x}(r)}}+\hat{s}_U'(r)\sqrt{\hat{x}(r)}}{1+\sqrt{\hat{x}(r)}\hat{s}_U(r)}\right) \end{aligned}$$

Hence,

$$\begin{aligned} \lim_{r\to 1}\frac{a_1'(r)}{\hat{x}'(r)} &= a_1(1)\left(\frac{1}{\hat{x}'(1)}+\left(\frac{1}{\hat{x}(1)}+\frac{\frac{\hat{s}_U(1)}{2\sqrt{\hat{x}(1)}}}{1+\sqrt{\hat{x}(1)}\hat{s}_U(1)}\right)+\lim_{r\to 1}\frac{\hat{s}_U'(r)}{\hat{x}'(r)}\frac{\sqrt{\hat{x}(1)}}{1+\sqrt{\hat{x}(1)}\hat{s}_U(1)}\right) \\ &= \frac{1}{\kappa}\frac{b^4}{1+b^4}\left(0+\frac{1+b^4}{b^4}-\frac{b^2}{\sqrt{1+b^4}}\frac{(1+b^4)^{\frac{5}{2}}}{2(1-b^2)b^4}\right) \\ &= \frac{1}{\kappa}\frac{b^4}{1+b^4}\left(\frac{1+b^4}{b^4}-\frac{(1+b^4)^2}{2(1-b^2)b^2}\right)=\frac{1}{\kappa}\left(1-\frac{b^2(1+b^4)}{2(1-b^2)}\right) \\ &= \frac{2-3b^2-b^6}{2\kappa(1-b^2)}\geq 0 \text{ if } b\leq 0.772 \end{aligned}$$

The above expression is positive for the same cutoff for b as in the lower-censorship case. Therefore, $a_1'(1)=-\infty$.

Finally, the derivative of welfare at $r=1$ satisfies

$$\begin{aligned} W'(r) &= a_1'(r)-\kappa a_1'(r)a_1(r)\left(1+b^2\frac{1-\hat{x}(r)}{\hat{x}(r)}\right)+\frac{\kappa}{2}a_1(r)^2 b^2\frac{\hat{x}'(r)}{\hat{x}(r)} \\ \lim_{r\to 1}\frac{W'(r)}{\hat{x}'(r)} &= \left(1-\kappa a_1(1)\left(1+b^2\frac{1-\hat{x}(1)}{\hat{x}(1)}\right)\right)\lim_{r\to 1}\frac{a_1'(r)}{\hat{x}'(r)}+\frac{\kappa}{2}b^2\frac{a_1(1)^2}{\hat{x}(1)} \\ &= \left(1-\frac{b^4}{1+b^4}\frac{1+b^2}{b^2}\right)\lim_{r\to 1}\frac{a_1'(r)}{\hat{x}'(r)}+\frac{b^2\times b^4}{2\kappa(1+b^4)} \\ &= \frac{1-b^2}{1+b^4}\lim_{r\to 1}\frac{a_1'(r)}{\hat{x}'(r)}+\frac{b^6}{2\kappa(1+b^4)}>0 \end{aligned}$$

Hence, $W'(1)=-\infty$ which implies that pooling some observations at the top improves welfare. This concludes the proof. □

C Validity of the First Order Approach for Upper and Lower-censorship Ratings

In this section, we describe conditions that make the first order approach valid. To do so, we use an approach similar to that of Jewitt (1988) and more recently Chade and Swinkels (2020). More specifically, it is sufficient to show that given our optimal ratings, (lower– or upper– censorship), the payoff of the agent is quasi–concave in her effort. To show quasi–concavity of the payoff, it is sufficient to show that $U'(a) = 0$ implies $U''(a) < 0$ if U is twice continuously differentiable. This would imply that U cannot have more than one local maximum, i.e., it is single peaked, and is thus quasi–concave.

To see this, suppose that there are two points $a_1 < a_2$ such that $U'(a_1) = U'(a_2) = 0$. Since $U''(a_1) < 0$ it must be that there is an interval of values above a_1 for which $U'(a) < 0$. Now, without lost of generality, let us assume that $a_2 = \inf_{a'>a_1, U'(a')=0} a'$. Since U is assumed to be twice continuously differentiable, U' is continuous and thus $U'(a_2) = 0$ and since $U'(a) < 0$ for an interval around a_1, $a_1 < a_2$. Moreover, we must have that for all $a \in (a_1, a_2)$, $U'(a) < 0$. This implies that $U'(a) < U'(a_2) = 0$ for values of a below a_2. Since U is assumed to be twice continuously differentiable, we must have that $U''(a_2) \geq 0$ which is a contradiction. This implies that U is single peaked.

This allows us to make the following claim:

Lemma 4. *Suppose that the family of distributions $\{G(y|a)\}_{a\in A}$ is twice continuously differentiable and A is a convex subset of $\mathbb{R}$. Suppose further that G satisfies the following properties*

$$0 > \int_{-\infty}^{\hat{y}} (v(y,\hat{a}) - \mathbb{E}[v|y \geq \hat{y}]) \frac{\partial}{\partial a} \frac{g_a(y|a)}{c'(a)} dy, \forall a, \hat{a} \in A, \hat{y} \in \mathbb{R},$$
$$0 > \int_{\hat{y}}^{\infty} (v(y,\hat{a}) - \mathbb{E}[v|y \leq \hat{y}]) \frac{\partial}{\partial a} \frac{g_a(y|a)}{c'(a)} dy, \forall a, \hat{a} \in A, \hat{y} \in \mathbb{R}.$$

Then the FOA is valid under a lower- and upper-censorship policy.

Proof. Consider an upper-censorship policy that pools realizations of y above $\hat{y}$. If market believes the agent chooses effort $\hat{a}$, then the payoff of the agent is given by

$$U(a) = \int_{-\infty}^{\hat{y}} v(y,\hat{a}) g(y|a) dy + \frac{\int_{\hat{y}}^{\infty} v(y,\hat{a}) g(y|\hat{a}) dy}{1 - G(\hat{y}|\hat{a})} (1 - G(\hat{y}|a)) - c(a)$$

If we let $p_H = \mathbb{E}\left[v\left(y,\hat{a}\right)|y \geq \hat{y}\right] > v\left(\hat{y},\hat{a}\right)$, we can write

$$\begin{aligned} U'\left(a\right) &= \int_{-\infty}^{\hat{y}} v\left(y,\hat{a}\right) g_a\left(y|a\right) dy - p_H G_a\left(\hat{y}|a\right) - c'\left(a\right) \\ &= \int_{-\infty}^{\hat{y}} \left(v\left(y,\hat{a}\right) - p_H\right) g_a\left(y|a\right) dy - c'\left(a\right) \end{aligned}$$

Now, suppose that $U'\left(a_1\right) = 0$ at some effort level a_1. Then,

$$\begin{aligned} U''\left(a_1\right) &= \int_{-\infty}^{\hat{y}} \left(v\left(y,\hat{a}\right) - p_H\right) g_{aa}\left(y|a_1\right) dy - c''\left(a_1\right) \\ &= \int_{-\infty}^{\hat{y}} \left(v\left(y,\hat{a}\right) - p_H\right) g_{aa}\left(y|a_1\right) dy \\ &\quad - \frac{c''\left(a_1\right)}{c'\left(a_1\right)} c'\left(a_1\right) \\ &= \int_{-\infty}^{\hat{y}} \left(v\left(y,\hat{a}\right) - p_H\right) g_{aa}\left(y|a_1\right) dy \\ &\quad - \frac{c''\left(a_1\right)}{c'\left(a_1\right)} \int_{-\infty}^{\hat{y}} \left(v\left(y,\hat{a}\right) - p_H\right) g_a\left(y|a_1\right) dy \\ &= \int_{-\infty}^{\hat{y}} \left(v\left(y,\hat{a}\right) - p_H\right) \left[g_{aa}\left(y|a_1\right) - \frac{c''\left(a_1\right)}{c'\left(a_1\right)} g_a\left(y|a_1\right)\right] dy \\ &= \int_{-\infty}^{\hat{y}} \left(v\left(y,\hat{a}\right) - p_H\right) c'\left(a_1\right) \frac{\partial}{\partial a} \frac{g_a\left(y|a_1\right)}{c'\left(a_1\right)} dy \end{aligned}$$

Since $c'\left(a_1\right) > 0$, the assumption on G in the statement of lemma guarantees that $U''\left(a_1\right) < 0$. Given our argument above, this implies that U is quasi–concave and thus FOA is valid. The argument for lower-censorship ratings is the mirror of this argument. □

The essence of the conditions Lemma 4 is that they put a restriction on how convex the cost is, captured by $c''\left(a\right)/c'\left(a\right)$ relative to that of expected interim price. Indeed, the conditions can be rewritten as

$$\frac{\int_{-\infty}^{\infty} p\left(y;\hat{y},\hat{a}\right) g_{aa}\left(y|a\right) dy}{\int_{-\infty}^{\infty} p\left(y;\hat{y},\hat{a}\right) g_a\left(y|a\right) dy} < \frac{c''\left(a\right)}{c'\left(a\right)}, \forall a \in A$$

where in the case of lower and upper censorship respectively, interim prices are

$$p(y;\hat{y},\hat{a}) = \begin{cases} v(y,\hat{a}) & y \geq \hat{y} \\ \mathbb{E}_{\hat{a}}[v|y \leq \hat{y}] & y \leq \hat{y} \end{cases}$$

$$p(y;\hat{y},\hat{a}) = \begin{cases} \mathbb{E}_{\hat{a}}[v|y \geq \hat{y}] & y \geq \hat{y} \\ v(y,\hat{a}) & y \leq \hat{y} \end{cases}$$

In the special case where $\frac{\partial}{\partial y} v(y,a) = \beta(a)$ and $y = a + \varepsilon$ with $\varepsilon \sim H(\varepsilon)$ and density $h(\varepsilon) = H'(\varepsilon)$, these become

$$\frac{\int_{-\infty}^{\hat{y}} (y - \mathbb{E}_{\hat{a}}[y|y \geq \hat{y}]) g_{aa}(y|a)\, dy}{\int_{-\infty}^{\hat{y}} (y - \mathbb{E}_{\hat{a}}[y|y \geq \hat{y}]) g_{a}(y|a)\, dy} < \frac{c''(a)}{c'(a)}$$

$$\frac{\int_{\hat{y}}^{\infty} (y - \mathbb{E}_{\hat{a}}[y|y \leq \hat{y}]) g_{aa}(y|a)\, dy}{\int_{\hat{y}}^{\infty} (y - \mathbb{E}_{\hat{a}}[y|y \leq \hat{y}]) g_{a}(y|a)\, dy} < \frac{c''(a)}{c'(a)}$$

and we have

$$\begin{aligned} \int_{-\infty}^{\hat{y}} (y - \mathbb{E}_{\hat{a}}[y|y \geq \hat{y}]) g_{a}(y|a)\, dy &= \\ \int_{-\infty}^{\hat{y}} (y - \hat{y} - \mathbb{E}_{\hat{a}}[y - \hat{y}|y \geq \hat{y}])\, dG_{a}(y|a) &= \\ -\int_{-\infty}^{\hat{y}} G_{a}(y|a)\, dy - \mathbb{E}_{\hat{a}}[y - \hat{y}|y \geq \hat{y}]\, G_{a}(\hat{y}|a) &= \\ -\int_{-\infty}^{\hat{y}} \frac{\partial}{\partial a} H(y-a)\, dy - \mathbb{E}_{\hat{a}}[y - \hat{y}|y \geq \hat{y}] \frac{\partial}{\partial a} H(\hat{y}-a) &= \\ H(\hat{y}-a) + \mathbb{E}_{\hat{a}}[y - \hat{y}|y \geq \hat{y}]\, h(\hat{y}-a) & \\ \int_{-\infty}^{\hat{y}} (y - \mathbb{E}_{\hat{a}}[y|y \geq \hat{y}]) g_{aa}(y|a)\, dy &= \\ -h(\hat{y}-a) - \mathbb{E}_{\hat{a}}[y - \hat{y}|y \geq \hat{y}]\, h'(\hat{y}-a) & \end{aligned}$$

Let us make the following assumption on H:

Assumption 4. *The cumulative distribution function $H(\varepsilon)$ satisfies:*

1. $\log(1 - H(\varepsilon))$ is concave in ε,

2. $\log H(\varepsilon)$ is concave in ε,

3. $\forall d > 0$, $\frac{h(x)}{1-H(x)} - \frac{h(x-d)}{1-H(x-d)} \leq \kappa_1(d)$, $\frac{h(x-d)}{H(x-d)} - \frac{h(x)}{H(x)} \leq \kappa_2(d)$

The first two conditions are standard log–concavity conditions while the last condition implies that the variations in the derivative of $\log(1-H(x))$ are uniformly bounded above. A sufficient condition for the latter is that the expressions $h'(x)/(1-H(x))+h(x)^2/(1-H(x))^2$ and $h'(x)/H(x)-h(x)^2/(H(x))^2$ are bounded above.

Under the above assumptions, $h(\varepsilon)/(1-H(\varepsilon))$ is increasing in ε and therefore,

$$\mathbb{E}_{\hat{a}}\left[y-\hat{y}|y\geq\hat{y}\right]=\frac{\int_{\hat{y}}^{\infty}(y-\hat{y})\,dH(y-\hat{a})}{1-H(\hat{y}-\hat{a})}$$
$$=\frac{\int_{\hat{y}-\hat{a}}^{\infty}\frac{1-H(z)}{h(z)}dH(z)}{1-H(\hat{y}-\hat{a})}\leq\frac{1-H(\hat{y}-\hat{a})}{h(\hat{y}-\hat{a})}$$

We then have that

$$\frac{\int_{-\infty}^{\hat{y}}(y-\mathbb{E}_{\hat{a}}[y|y\geq\hat{y}])\,g_{aa}(y|a)\,dy}{\int_{-\infty}^{\hat{y}}(y-\mathbb{E}_{\hat{a}}[y|y\geq\hat{y}])\,g_{a}(y|a)\,dy}=$$
$$-\frac{h(\hat{y}-a)+\mathbb{E}_{\hat{a}}[y-\hat{y}|y\geq\hat{y}]\,h'(\hat{y}-a)}{H(\hat{y}-a)+\mathbb{E}_{\hat{a}}[y-\hat{y}|y\geq\hat{y}]\,h(\hat{y}-a)}=$$
$$-\frac{\frac{1}{\mathbb{E}_{\hat{a}}[y-\hat{y}|y\geq\hat{y}]}+\frac{h'(\hat{y}-a)}{h(\hat{y}-a)}}{\frac{1}{\mathbb{E}_{\hat{a}}[y-\hat{y}|y\geq\hat{y}]}+\frac{h(\hat{y}-a)}{H(\hat{y}-a)}}\frac{h(\hat{y}-a)}{H(\hat{y}-a)}=\left(-1+\frac{\frac{h(\hat{y}-a)}{H(\hat{y}-a)}-\frac{h'(\hat{y}-a)}{h(\hat{y}-a)}}{\frac{1}{\mathbb{E}_{\hat{a}}[y-\hat{y}|y\geq\hat{y}]}+\frac{h(\hat{y}-a)}{H(\hat{y}-a)}}\right)\frac{h(\hat{y}-a)}{H(\hat{y}-a)}$$

By $\log$–concavity of H, $h/H-h'/h\geq 0$ and hence, we have the following inequality

$$-1+\frac{\frac{h(\hat{y}-a)}{H(\hat{y}-a)}-\frac{h'(\hat{y}-a)}{h(\hat{y}-a)}}{\frac{1}{\mathbb{E}_{\hat{a}}[y-\hat{y}|y\geq\hat{y}]}+\frac{h(\hat{y}-a)}{H(\hat{y}-a)}}\leq-1+\frac{\frac{h(\hat{y}-a)}{H(\hat{y}-a)}-\frac{h'(\hat{y}-a)}{h(\hat{y}-a)}}{\frac{h(\hat{y}-\hat{a})}{1-H(\hat{y}-\hat{a})}+\frac{h(\hat{y}-a)}{H(\hat{y}-a)}}=-\frac{\frac{h(\hat{y}-\hat{a})}{1-H(\hat{y}-\hat{a})}+\frac{h'(\hat{y}-a)}{h(\hat{y}-a)}}{\frac{h(\hat{y}-\hat{a})}{1-H(\hat{y}-\hat{a})}+\frac{h(\hat{y}-a)}{H(\hat{y}-a)}}$$

by the above property of $\mathbb{E}_{\hat{a}}[y-\hat{y}|y\geq\hat{y}]$. If we define $\hat{y}-a=x$ and $d=\hat{a}-a$, then

$$\frac{\int_{-\infty}^{\hat{y}}(y-\mathbb{E}_{\hat{a}}[y|y\geq\hat{y}])\,g_{aa}(y|a)\,dy}{\int_{-\infty}^{\hat{y}}(y-\mathbb{E}_{\hat{a}}[y|y\geq\hat{y}])\,g_{a}(y|a)\,dy}\leq-\frac{\frac{h(x-d)}{1-H(x-d)}+\frac{h'(x)}{h(x)}}{\frac{h(x-d)}{1-H(x-d)}+\frac{h(x)}{H(x)}}\frac{h(x)}{H(x)}$$
$$=-\frac{\frac{h(x-d)}{1-H(x-d)}+\frac{h'(x)}{h(x)}}{\frac{h(x-d)}{1-H(x-d)}\frac{H(x)}{h(x)}+1}$$

Since $1-H$ is concave, then $h'/h\geq-h/(1-H)$ and the right hand side of the above inequality satisfies

$$-\frac{\frac{h(x-d)}{1-H(x-d)}+\frac{h'(x)}{h(x)}}{\frac{h(x-d)}{1-H(x-d)}\frac{H(x)}{h(x)}+1}\leq\frac{\frac{h(x)}{1-H(x)}-\frac{h(x-d)}{1-H(x-d)}}{\frac{h(x-d)}{1-H(x-d)}\frac{H(x)}{h(x)}+1}$$

Note that in the above if $d<0$, since $h/(1-H)$ is increasing (log–concavity of $1-H$) the RHS of the above inequality is negative which is guaranteed to be less than $c''(a)/c'(a)$ since c is convex.

By Assumption 4, the RHS of the above is less than $\kappa_1(d)$. If $A = [0, \overline{a}]$, then the highest value of d is $2\overline{a}$ which means that it is sufficient to for c to satisfy

$$\max_{d \leq 2\overline{a}} \kappa_1(d) \leq \frac{c''(a)}{c'(a)}$$

In a similar fashion, we can show that for lower-censorship ratings to satisfy the requirement of Lemma 4, we must also have

$$\max_{d \leq 2\overline{a}} \kappa_2(d) \leq \frac{c''(a)}{c'(a)}$$

Thus validity of FOA is equivalent to c having a high enough curvature.

D Importance of Comonotonicity in Proposition 1

The following counterexample demonstrates that there exist price schedules satisfying the mean-preserving contraction property that cannot be generated by any information structure.

Example 2. Suppose that $A = \{0, 1/3, 1\} = \{a_1, a_2, a_3\}$, $v = a_i$. The indicator is deterministic: $G(\{a_i\} | a_i) = 1$, and prior is uniform $\mu(\{a_i\}) = 1/3$. In words, the market cares only about the action of the seller, and y coincides with it. Figure 9 depicts the feasible interim prices in the space of $(p(a_1), p(a_2))$; (the third coordinate is pinned down by Bayes rule since $\mathbb{E}[p] = \mathbb{E}[v] = \frac{4}{3}$). Area A shows the set of random variables that are mean preserving spread of $(0, 1/3, 1)$. They are depicted by their first two variables while the third is agin pins down by Bayes rule. [27] The set of interim prices is denoted by area B in Figure 9. We find this set by solving the optimization problem associated with the highest and lowest value of $p(a_2)$ as a function of $p(a_1)$.[28] Evidently, the set B does not coincide with A. This is mainly due to the restrictions put by the second order expectations. For example, it can be easily shown that the coefficient of $\overline{v}(a)$ in $p(a)$ is at least $1/3$ which means that $p(a_3)$ cannot become lower than $4/9$.

Finally, note that the points a, b, c, d are associated with deterministic ratings that either separate or pool the states and for which $p(a_1) \leq p(a_2) \leq p(a_3)$. Interestingly, if we consider the set of random variables whose realizations are less dispersed than $\overline{v}(a)$ and satisfy monotonicity, this coincides with the convex hull of the points a, b, c, and d. In Proposition 1 below, we show that

[27]The conditions are $0 \leq x_i \leq 1, 1/3 \leq x_i + x_j \leq 4/3$, for all i, j, and $x_1 + x_2 + x_3 = 4/3$.

[28]The upper and lower bound of $p(a_2)$ can be found via standard concavification method. The lower bound is given by $2(4 - p(a_1)) / \left(10 - 3p(a_1) + \sqrt{(8 - 3p(a_1))^2 - 12}\right)$ and the upper bound is $3 - (6 - 2p(a_1)) / (6 - 3p(a_1))$.

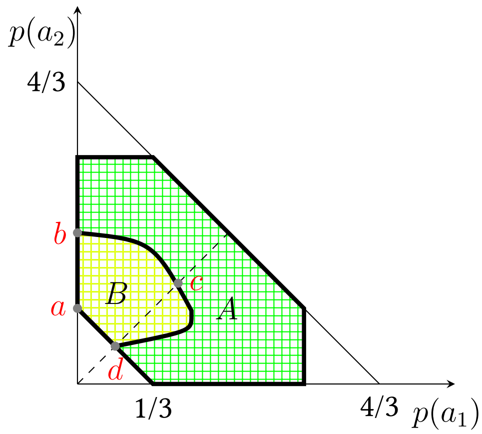


Figure 9: The set of interim prices and mean-preserving contractions of market valuations for Example 2. The green area, A, represents the three state random variables that are a mean-preserving contraction of a. The yellow area, B, is the set of feasible interim prices under some information structure.

this insight holds generally and allows us to significantly simplify the problem of rating design under a comonotonicity condition.

The result in Proposition 1 is reminiscent of the result of Blackwell (1953) and Rothschild and Stiglitz (1970), the general version of which can be found in Strassen (1965). That result states that for any two random variables x and y, there exists a random variable s such that $\mathbb{E}[x|s]$ has the same distribution as y if and only if y second-order stochastically dominates x.

While similar, our result is different in two aspects. First, it is stated for the second-order *conditional* expectation, and thus Blackwell's result cannot be applied. The key intricacy is that the same signal structure that generates the random variable $\mathbb{E}[v|s]$ must be used to generate $\mathbb{E}[\mathbb{E}[v|s]|y]$. Second, as illustrated by Example 2, the equivalent of Blackwell's result does not hold in general and can be shown only when v and p are comonotone.

We also note that the comonotonicity is effectively a form of uni-dimensionality for the indicator. Since the indicator y matters for the market and payoffs only through its effect on $\overline{v}(y;a)$, we can relabel the indicator to be $\overline{v}(y;a)$. Under this reformulation, comonotonicity implies that interim prices $p(y)$ are a well-defined function of $\overline{v}(y;a)$. In other words, comonotonicity means that the indicator can always be reduced to a one-dimensional signal.