# Optimal Rating Design under Moral Hazard

Maryam Saeedi  Ali Shourideh

Oxford, May 2026

# Introduction

- Many markets rely on ratings and information disclosure
  - ESG rating agencies, online platforms (Amazon, Airbnb, eBay, Upwork)
  - Standardized tests, security ratings, Google ranking

# Introduction

- Many markets rely on ratings and information disclosure
  - ESG rating agencies, online platforms (Amazon, Airbnb, eBay, Upwork)
  - Standardized tests, security ratings, Google ranking
- **Key Tensions:**
  - Ratings incentivize productive effort
  - But also induce **window dressing** and **manipulation**
    - ESG greenwashing, USNews ranking fraud, Amazon fake reviews

# Introduction

- Many markets rely on ratings and information disclosure
  - ESG rating agencies, online platforms (Amazon, Airbnb, eBay, Upwork)
  - Standardized tests, security ratings, Google ranking
- **Key Tensions:**
  - Ratings incentivize productive effort
  - But also induce **window dressing** and **manipulation**
    - ESG greenwashing, USNews ranking fraud, Amazon fake reviews
- Ratings are an information structure
- How should an intermediary design ratings when agents *strategically respond*?

## This Paper

- A career-concerns model (à la Holmstrom 1999) with rating design
  - Agent takes costly action;
  - Intermediary
    - observes an indicator of outcome
    - commits to rating mechanism
  - Market pays expected value based on the signal from intermediary

## This Paper

- A career-concerns model (à la Holmstrom 1999) with rating design
  - Agent takes costly action;
  - Intermediary
    - observes an indicator of outcome
    - commits to rating mechanism
  - Market pays expected value based on the signal from intermediary
- **Main Results:**
  - **Part I:** Cast as standard mechanism design with majorization
    - **Interim prices** as the sufficient statistic
    - Optimal ratings from ***concavification of a gain function***

## This Paper

- A career-concerns model (à la Holmstrom 1999) with rating design
  - Agent takes costly action;
  - Intermediary
    - observes an indicator of outcome
    - commits to rating mechanism
  - Market pays expected value based on the signal from intermediary
- **Main Results:**
  - **Part I:** Cast as standard mechanism design with majorization
    - **Interim prices** as the sufficient statistic
    - Optimal ratings from ***concavification of a gain function***
  - **Part II:** How effort–variance interactions shape ratings
    - Effort keeps risk constant ⇒ *full information*
    - Effort raises tail risk ⇒ *lower censorship* (insurance)
    - Effort reduces tail risk ⇒ *upper censorship* (punish negligence)

## This Paper

- A career-concerns model (à la Holmstrom 1999) with rating design
  - Agent takes costly action;
  - Intermediary
    - observes an indicator of outcome
    - commits to rating mechanism
  - Market pays expected value based on the signal from intermediary
- **Main Results:**
  - **Part I:** Cast as standard mechanism design with majorization
    - **Interim prices** as the sufficient statistic
    - Optimal ratings from ***concavification of a gain function***
  - **Part II:** How effort–variance interactions shape ratings
    - Effort keeps risk constant ⇒ *full information*
    - Effort raises tail risk ⇒ *lower censorship* (insurance)
    - Effort reduces tail risk ⇒ *upper censorship* (punish negligence)
  - **Applications:** Multi-tasking, window dressing, redistributive test design

## Roadmap

- The General Model and Interim Prices
- Concavification of the Gain Function
- General Properties of Optimal Ratings
  - MLRP, ELRP, CLRP
  - Redistributive Motives
- Applications:
  - Multi-tasking and Window Dressing
  - Redistributive Test Design

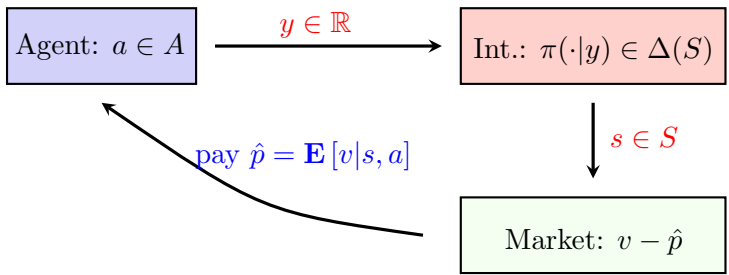
# The General Model

## The Model

- Agent chooses an action $a \in A \subset \mathbb{R}^N$ at cost $c(a)$
- Action generates $(y, v) \in \mathbb{R}^2$
  - $y$ : indicator observed by intermediary (e.g., performance data)
  - $v$ : value to the market
  - $y \sim G(y|a)$, $\overline{v}(y; \hat{a}) = \mathbb{E}[v|y, \hat{a}]$ market values
- Intermediary observes $y$, commits to signal $\pi(\cdot|y) \in \Delta(S)$

## The Model

- Agent chooses an action $a \in A \subset \mathbb{R}^N$ at cost $c(a)$
- Action generates $(y, v) \in \mathbb{R}^2$
  - $y$ : indicator observed by intermediary (e.g., performance data)
  - $v$ : value to the market
  - $y \sim G(y|a)$, $\overline{v}(y; \hat{a}) = \mathbb{E}[v|y, \hat{a}]$ market values
- Intermediary observes $y$, commits to signal $\pi(\cdot|y) \in \Delta(S)$



## Examples

- **Reputation Mechanisms in Online Platforms**
  - Airbnb's Superhost, eBay's Top Rated Seller, Upwork's Talent Badge
  - Platform has performance data; chooses what to disclose to market
- **Manipulation and Window Dressing**
  - ESG greenwashing (e.g., DWS), USNews ranking manipulation (Temple, Columbia)
  - Fake reviews on Amazon, eBay; mutual fund window dressing
- **Career Concerns and Externalities**
  - Career concerns produce inefficient effort (Holmstrom 1999)
  - Can ratings improve over full information? *Yes, when MLRP fails.*

## The Agent's Problem

- Given $(S, \pi)$ and market belief $\hat{a}$, agent's payoff:

$$\int_Y \int_S \mathbb{E}\left[v|s, \hat{a}\right] d\pi\left(s|y\right) dG\left(y|a\right) - c\left(a\right)$$

- **Information:**
  - $a$: private to the agent
  - $y$: observed only by intermediary
  - $s$: observed by market

## The Agent's Problem

- Given $(S, \pi)$ and market belief $\hat{a}$, agent's payoff:

$$\int_Y \int_S \mathbb{E}\left[v|s, \hat{a}\right] d\pi\left(s|y\right) dG\left(y|a\right) - c\left(a\right)$$

- **Information:**
  - $a$: private to the agent
  - $y$: observed only by intermediary
  - $s$: observed by market
- **Equilibrium**: Bayesian Nash
  - Given $\pi$ and market beliefs, $a$ maximizes agent's payoff
  - Market beliefs are consistent with $\pi$, $a$, and prior via Bayes' rule

## Interim Prices: A Sufficient Statistic

- Incentive compatibility for the agent:

$$a \in \arg\max_{a' \in A} \int_Y \underbrace{\int_S \mathbb{E}\left[v|s\right] d\pi\left(s|y\right)}_{p(y)} dG\left(y|a'\right) - c\left(a'\right)$$

- Define **interim prices**:

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

  - Agent's "second-order belief" – expectation of market posterior given $y$
  - Sufficient statistic for incentives

## Interim Prices: A Sufficient Statistic

- We say that $p$ and $\overline{v}$ are comonotone if:

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

- Higher prices are associated with higher market values.

## Interim Prices: A Sufficient Statistic

- We say that $p$ and $\overline{v}$ are comonotone if:

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

- Higher prices are associated with higher market values.

**Proposition.** If $p$ is *comonotone* with $\overline{v}$ and $p \preccurlyeq_{\text{maj}} \overline{v}$, then there exists an information structure $(S,\pi)$ such that $p(y) = \int \mathbb{E}[v|s]\, d\pi(s|y)$.

## Interim Prices: A Sufficient Statistic

- We say that $p$ and $\overline{v}$ are comonotone if:

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

- Higher prices are associated with higher market values.

**Proposition.** If $p$ is *comonotone* with $\overline{v}$ and $p \preccurlyeq_{\text{maj}} \overline{v}$, then there exists an information structure $(S, \pi)$ such that $p(y) = \int \mathbb{E}[v|s]\, d\pi(s|y)$.

- Comonotonicity assumption simplifies the problem dramatically
  - Textbook moral hazard *plus* a majorization constraint
    - Interim prices play the role of transfers

# Main Characterization: Concavification of the Gain Function

## The Intermediary's Objective

- Class of objectives:

$$W(a) + \int p(y)\,\alpha(y)\,dG(y|a)$$

with $\alpha(y) \geq 0$, is the weight on type $y$.

## The Intermediary's Objective

- Class of objectives:

$$W(a) + \int p(y)\,\alpha(y)\,dG(y|a)$$

  with $\alpha(y) \geq 0$, is the weight on type $y$.

- **Targeting an Action:** $\alpha(y) = 0$,
  - Direct externalities (e.g., ESG public good)
  - Career concerns (à la Holmstrom 1999): $\frac{\partial}{\partial a}\overline{v}(y;a) \neq 0$

## The Intermediary's Objective

- Class of objectives:

$$W(a) + \int p(y)\,\alpha(y)\,dG(y|a)$$

  with $\alpha(y) \geq 0$, is the weight on type $y$.

- **Targeting an Action:** $\alpha(y) = 0$,
  - Direct externalities (e.g., ESG public good)
  - Career concerns (à la Holmstrom 1999): $\frac{\partial}{\partial a}\overline{v}(y;a) \neq 0$
- **Distributional Concerns:** $\alpha(y)$ varies with $y$
  - Fairness: platform's interest in seller protection
  - Redistribution in test design across socio-economic groups

## Quantile Formulation

- $v_Q(i) = \overline{v}(G^{-1}(i|a); a)$
- $p_Q(i) =$ interim price at quantile $i$
- For any integrable $h(y)$:

$$\int h(y)p(y)dG(y|a) = \int_0^1 H(i)dp_Q(i)$$

$$\text{Where: } H(i) = \int_{\{y:\overline{v}(y)>v_Q(i)\}} h(y)dG(y|a)$$

- $H(i)$ collects the contribution of h from all outcomes ranked above the i-th quantile.

## Quantile Formulation

- The problem reduces to
- $$\max_{p_Q \preccurlyeq_{\text{maj}} v_Q} \int_0^1 H(i) dp_Q(i)$$
- One unknown function on [0,1], one objective, one constraint.

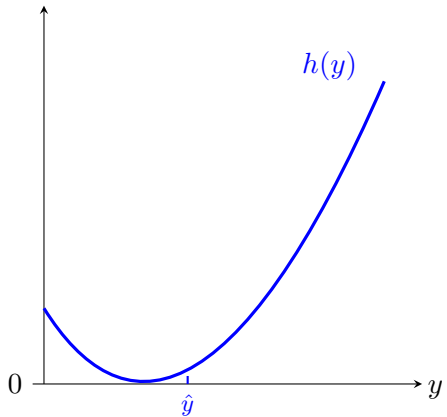
## Quantile Formulation

- The problem reduces to
- $$\max_{p_Q \preccurlyeq_{\text{maj}} v_Q} \int_0^1 H(i) dp_Q(i)$$
- One unknown function on [0,1], one objective, one constraint.

**Lemma.** The maximum of $\int h(y)p(y)dG$ s.t. $p \preccurlyeq_{\text{maj}} \overline{v}$ and comonotonicity equals $\int_0^1 \text{cav}H(i)dv_Q(i)$.
Moreover, the optimal $p$ satisfies:

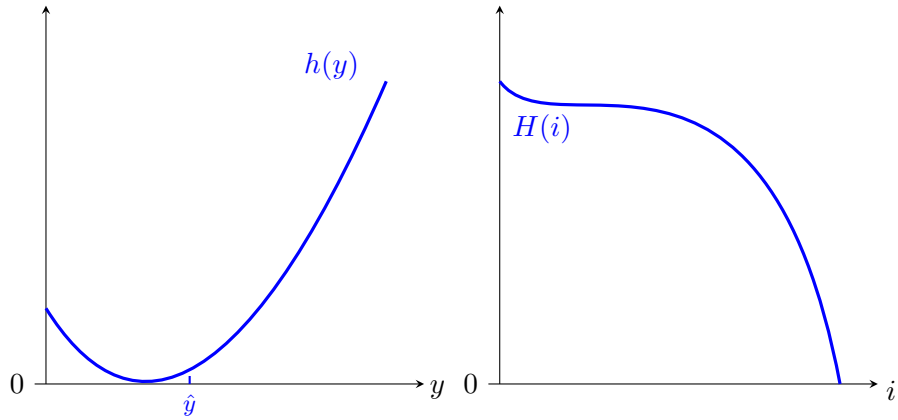
1. $p(y) = \overline{v}(y;a)$ when $H(G(y|a)) = \text{cav}H(G(y|a))$.
2. If $\text{cav}H(i) > H(i)$ for all $i$ in some maximal interval $I \subset [0,1]$, then, $p(y) = \mathbb{E}[\overline{v}|G(y|a) \in I]$.

## Concavification: A Picture



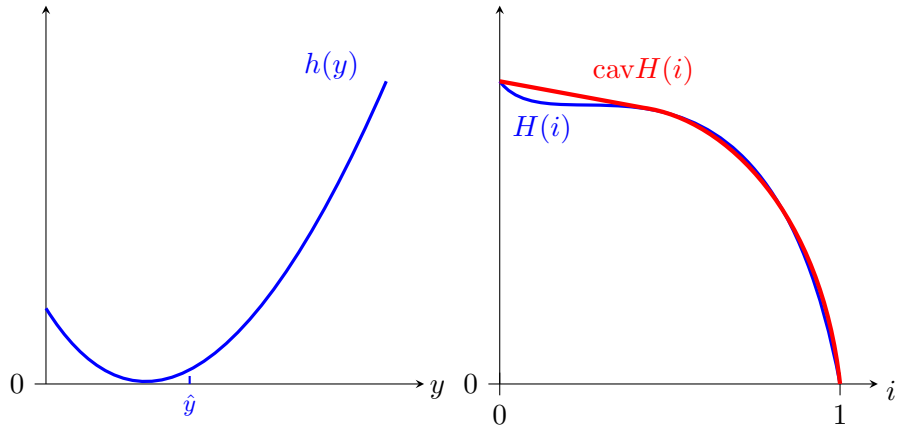

- **Pooling** where cav$H > H$; **revelation** where they coincide

## Concavification: A Picture



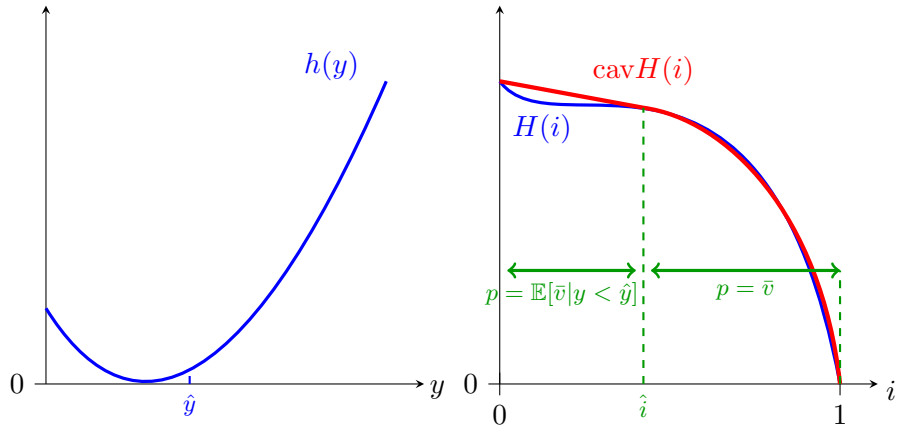

- **Pooling** where cav$H > H$; **revelation** where they coincide

## Concavification: A Picture




- **Pooling** where $\text{cav}H > H$; **revelation** where they coincide

## Concavification: A Picture




- **Pooling** where $\text{cav}H > H$; **revelation** where they coincide

## Theorem 1: The Main Characterization

**Theorem 1.** Under FOA, the optimal rating problem solves

$$\max_a W(a) + \int \text{cav}\Gamma(i;\lambda,a)\, dv_Q(i)$$

where the *gain function* is

$$\Gamma(i;\lambda,a) = \underbrace{\int_{\{y:\overline{v}(y;a)>v_Q(i)\}} \alpha(y) dG}_{\text{cumulative welfare weights}} - \sum_n \lambda_n \underbrace{\left[\frac{\partial}{\partial \hat{a}_n} F(i|\hat{a};a)\Big|_{\hat{a}=a} + \frac{\partial}{\partial a_n} c(a)\right]}_{\text{incentive component}}$$

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

and $\lambda$ are multipliers on incentive compatibility constraints.

## Implications of Theorem 1

- Reduces design problem to *one-dimensional concavification*
- **Simple ratings are optimal:** a deterministic monotone partition
  - Caratheodory: any convex combination uses at most two points

## Implications of Theorem 1

- Reduces design problem to *one-dimensional concavification*
- **Simple ratings are optimal:** a deterministic monotone partition
  - Caratheodory: any convex combination uses at most two points
- Two key drivers in the gain function:
  - Cumulative welfare weights (distributional motives)
  - Marginal change in quantile distribution (incentive effect)

## Implications of Theorem 1

- Reduces design problem to *one-dimensional concavification*
- **Simple ratings are optimal:** a deterministic monotone partition
  - Caratheodory: any convex combination uses at most two points
- Two key drivers in the gain function:
  - Cumulative welfare weights (distributional motives)
  - Marginal change in quantile distribution (incentive effect)
- No need for first-order approach a priori; FOA used for tractability
- Proof: Rockafellar–Fenchel duality
  - Also used in Dworczak-Kolotilin (2024), Corrao-Kolotilin-Wolitzky (2024)

## Simple Ratings Are Optimal

**Assumption (Independence).** $G(y|a)$ has full support on a convex subset of $\mathbb{R}$, and $\alpha(y)g(y|a)$ is not a non-zero linear combination of $g(y|a)$ and $\{\partial g/\partial a_n\}$ on any interval.

## Simple Ratings Are Optimal

**Assumption (Independence).** $G(y|a)$ has full support on a convex subset of $\mathbb{R}$, and $\alpha(y)g(y|a)$ is not a non-zero linear combination of $g(y|a)$ and $\{\partial g/\partial a_n\}$ on any interval.

**Proposition.** Under Independence, optimal rating is a monotone partition. Moreover:

- Where $\text{cav}\Gamma = \Gamma$: optimal rating fully reveals $\overline{v} = v_Q(i)$
- Where $\text{cav}\Gamma > \Gamma$ on $[i_1, i_2]$: optimal rating reveals $\overline{v} \in [v_Q(i_1), v_Q(i_2)]$ (pooling)

# General Properties:

# MLRP, ELRP, CLRP

## Targeting an Action ($\alpha(y) = 0$)

- Question: which efforts $a^*$ are implementable?
- Concavity of $\Gamma(i; \lambda, a)$ tied to log-supermodularity:

$$-\lambda \frac{\partial^2}{\partial i^2} F(i|\hat{a}; a)\Big|_{\hat{a}=a} = \frac{-\lambda \frac{\partial^2}{\partial y \partial a} \log g(y|a)\Big|_{y=G^{-1}(i|a)}}{g(y|a)}$$

## Targeting an Action ($\alpha(y) = 0$)

- Question: which efforts $a^*$ are implementable?
- Concavity of $\Gamma(i; \lambda, a)$ tied to log-supermodularity:

$$-\lambda \frac{\partial^2}{\partial i^2} F(i|\hat{a}; a)\Big|_{\hat{a}=a} = \frac{-\lambda \frac{\partial^2}{\partial y \partial a} \log g(y|a)\Big|_{y=G^{-1}(i|a)}}{g(y|a)}$$

- **Concavity of $\Gamma$ depends on sign of $\partial^2 \log g / \partial y \partial a$**

## MLRP: Full Information is Optimal

**Proposition.** Suppose $g(y|a)$ satisfies MLRP. Then the highest implementable effort is associated with $p(y) = \overline{v}(y; a_{FI})$ (*full information*):

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

- Under MLRP, $\Gamma$ is concave in $i$ (for $\lambda > 0$)
  - Concave envelope coincides with $\Gamma$ everywhere
  - Optimal rating is fully revealing
- Ratings cannot increase effort beyond what full information achieves
- See also Dewatripont, Jewitt and Tirole (1999)

## Beyond MLRP: Effort and Variance

- Many activities violate MLRP *systematically*:
  - **Innovation**: more effort $\Rightarrow$ *higher* variance (breakthroughs or failures)
  - **Maintenance**: more effort $\Rightarrow$ *lower* variance (consistent outcomes)

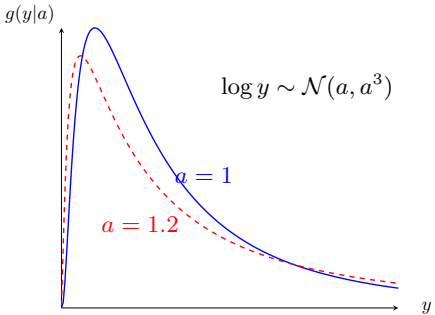
## Beyond MLRP: Effort and Variance

- Many activities violate MLRP *systematically*:
  - **Innovation**: more effort ⇒ *higher* variance (breakthroughs or failures)
  - **Maintenance**: more effort ⇒ *lower* variance (consistent outcomes)
- Two new distributional properties:

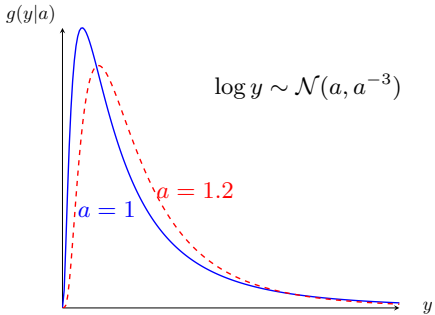
**Definition.** $g(y|a)$ satisfies:

- **ELRP (Expanding Likelihood Ratio Property)**: $\exists \hat{y}$ s.t. $\frac{\partial^2 \log g}{\partial a \partial y} \geq 0$ for $y \geq \hat{y}$ and $\leq 0$ for $y \leq \hat{y}$
- **CLRP (Compressing Likelihood Ratio Property)**: $\exists \hat{y}$ s.t. $\frac{\partial^2 \log g}{\partial a \partial y} \leq 0$ for $y \geq \hat{y}$ and $\geq 0$ for $y \leq \hat{y}$

## ELRP and CLRP Densities




(a) **ELRP**




(b) **CLRP**

- ELRP: tails expand with $a$ (innovation)
- CLRP: tails compress with $a$ (maintenance)

## ELRP/CLRP: Optimal Ratings

**Proposition.**

1. If $g(y|a)$ satisfies **ELRP**, highest implementable effort comes from **lower-censorship**: pool low $y$, reveal high $y$.
2. If $g(y|a)$ satisfies **CLRP**, highest implementable effort comes from **upper-censorship**: reveal low $y$, pool high $y$.

- **Intuition (ELRP):** effort raises tail risk $\Rightarrow$ pooling low realizations *insures* the agent and encourages risk-taking
- **Intuition (CLRP):** effort reduces tail risk $\Rightarrow$ pooling high realizations *punishes* poor outcomes and deters negligence

# Redistributive Motives

## Redistributive Motives

- Suppose $\alpha(y) > 0$ and *decreasing* in $y$
- Gain function combines redistributive and incentive components:

$$\Gamma(i;\lambda,a) = \underbrace{\int_{G^{-1}(i|a)}^{\infty} \alpha(y) dG(y|a)}_{\text{redistributive (convex)}} - \underbrace{\lambda G_a(G^{-1}(i|a))}_{\text{incentive (concave under MLRP)}}$$

## Redistributive Motives

- Suppose $\alpha(y) > 0$ and *decreasing* in $y$
- Gain function combines redistributive and incentive components:

$$\Gamma(i;\lambda,a) = \underbrace{\int_{G^{-1}(i|a)}^{\infty} \alpha(y) dG(y|a)}_{\text{redistributive (convex)}} - \underbrace{\lambda G_a(G^{-1}(i|a))}_{\text{incentive (concave under MLRP)}}$$

- Even under MLRP, strong redistributive motives can make $\Gamma$ **non-concave** at low quantiles
- $\Rightarrow$ optimal rating involves *lower-censorship pooling*

## Redistributive Motives

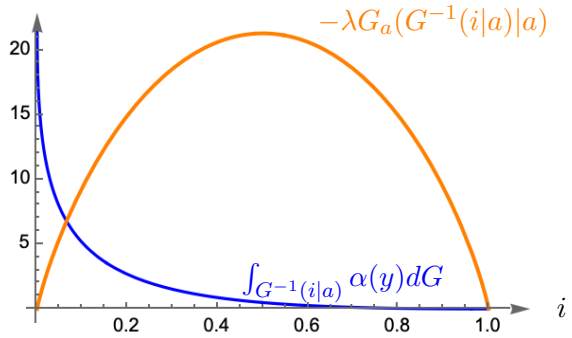
- Suppose $\alpha(y) > 0$ and *decreasing* in $y$
- Gain function combines redistributive and incentive components:

$$\Gamma(i;\lambda,a) = \underbrace{\int_{G^{-1}(i|a)}^{\infty} \alpha(y)dG(y|a)}_{\text{redistributive (convex)}} - \underbrace{\lambda G_a(G^{-1}(i|a))}_{\text{incentive (concave under MLRP)}}$$

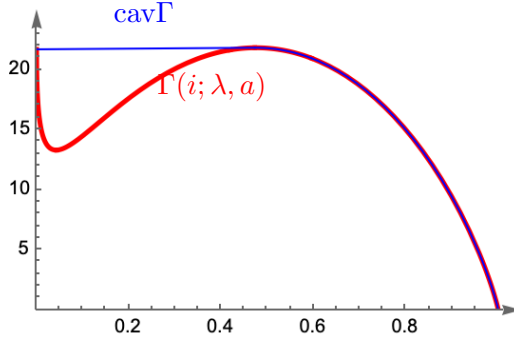
- Even under MLRP, strong redistributive motives can make $\Gamma$ **non-concave** at low quantiles
- $\Rightarrow$ optimal rating involves *lower-censorship pooling*

**Proposition.** If $\lim_{i\to 0} \alpha(G^{-1}(i|a)) + \lambda\frac{g_a}{g} > \int_0^1 \alpha(G^{-1}(i|a))di$, then optimal rating pools an interval near $\underline{y}$.

**Example:** $y = a\varepsilon$, $\log \varepsilon \sim \mathcal{N}(-1/2, 1)$, $\alpha(y) = y^{-\beta}$




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




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

Figure: Concavification of the marginal change in quantiles

# Application 1:

# Multi-Tasking and Window Dressing

## A Multi-Tasking Model

- Agent chooses $a = (a_1, \ldots, a_N) \in [0, \overline{a}]^N$
- Market value and indicator:

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

with

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

- Quadratic cost: $c(a) = \frac{1}{2} \sum \kappa_n a_n^2$

## A Multi-Tasking Model

- Agent chooses $a = (a_1, \ldots, a_N) \in [0, \overline{a}]^N$
- Market value and indicator:

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

with

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

- Quadratic cost: $c(a) = \frac{1}{2} \sum \kappa_n a_n^2$
- **Window dressing** task $n$: $b_{y,n} > 0$ but $b_{v,n} = 0$
  - Boosts indicator without affecting market value
  - Welfare requires $a_n = 0$

## Reducibility to a Single Effort

- Key insight: $a$ affects $y$ only through $m_y = b_y \cdot a$
- Cost-minimizing efforts given $m_y$:

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

- Surplus-maximizing $m_y^* = \sum_n b_{y,n} b_{v,n}/\kappa_n$
- Full-info maximum effort: $m_{FI}^* = \beta \sum_i b_{y,i}^2/\kappa_i$ where $\beta = \sigma_{vy}/\sigma_y^2$

## Reducibility to a Single Effort

- Key insight: $a$ affects $y$ only through $m_y = b_y \cdot a$
- Cost-minimizing efforts given $m_y$:

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

- Surplus-maximizing $m_y^* = \sum_n b_{y,n} b_{v,n}/\kappa_n$
- Full-info maximum effort: $m_{FI}^* = \beta \sum_i b_{y,i}^2/\kappa_i$ where $\beta = \sigma_{vy}/\sigma_y^2$
- Since $y|m_y \sim \mathcal{N}(m_y, \sigma_y^2)$: MLRP holds
  - Highest implementable $m_y$ is $m_{FI}^*$

## Optimal Rating in the Multi-Task Model

**Proposition.** If $m_y^* \geq m_{FI}^*$: full revelation. If $m_y^* < m_{FI}^*$: optimal rating is **lower-censorship**, pooling $y \leq \overline{y}$ with $\overline{z} = (\overline{y} - m_y^*)/\sigma_y$ solving

$$\beta \overline{z} \phi(\overline{z}) + \beta \frac{\phi(\overline{z})^2}{\Phi(\overline{z})} + \beta[1 - \Phi(\overline{z})] = \frac{\sum_i b_{y,i} b_{v,i}/\kappa_i}{\sum_i b_{y,i}^2/\kappa_i}$$

## Optimal Rating in the Multi-Task Model

**Proposition.** If $m_y^* \geq m_{FI}^*$: full revelation. If $m_y^* < m_{FI}^*$: optimal rating is **lower-censorship**, pooling $y \leq \overline{y}$ with $\overline{z} = (\overline{y} - m_y^*)/\sigma_y$ solving

$$\beta \overline{z} \phi(\overline{z}) + \beta \frac{\phi(\overline{z})^2}{\Phi(\overline{z})} + \beta[1 - \Phi(\overline{z})] = \frac{\sum_i b_{y,i} b_{v,i}/\kappa_i}{\sum_i b_{y,i}^2/\kappa_i}$$

**Proposition (Holmstrom-Milgrom revisited).** Suppose $a_i$ exhibits window dressing ($b_{v,i} = 0, b_{y,i} > 0$). Then a *decrease in* $\kappa_i$ (cheaper manipulation) leads to a **less informative** optimal rating.

# Application 2:

# Redistributive Test Design

## Redistributive Test Design

- Motivation: debate over SAT, California Mathematics Framework
  - Concern that standardized testing disadvantages low-SES students
  - Common "solution": make tests optional
  - Our question: what does *optimal test design* look like?

## Redistributive Test Design

- Motivation: debate over SAT, California Mathematics Framework
  - Concern that standardized testing disadvantages low-SES students
  - Common "solution": make tests optional
  - Our question: what does *optimal test design* look like?
- Two student types $\theta \in \{R, P\}$ with masses $f_R, f_P$
- Both choose effort $a_\theta$; indicator $y \sim g(y|a_\theta)$
- Cost: $k_\theta c(a)$ with $k_R < k_P$ (advantaged vs. disadvantaged)
- Market value $v = y$

## Redistributive Test Design

- Motivation: debate over SAT, California Mathematics Framework
  - Concern that standardized testing disadvantages low-SES students
  - Common "solution": make tests optional
  - Our question: what does *optimal test design* look like?
- Two student types $\theta \in \{R, P\}$ with masses $f_R, f_P$
- Both choose effort $a_\theta$; indicator $y \sim g(y|a_\theta)$
- Cost: $k_\theta c(a)$ with $k_R < k_P$ (advantaged vs. disadvantaged)
- Market value $v = y$
- Designer's objective (with $\alpha_P > 1 > \alpha_R$):

$$\alpha_P f_P \left[ \int p \, dG(\cdot|a_P) - k_P c(a_P) \right] + \alpha_R f_R \left[ \int p \, dG(\cdot|a_R) - k_R c(a_R) \right]$$

## Mid-Censorship Optimal

- Gain function combines two incentive components plus weights:

$$\begin{aligned}\Gamma(i) = &- \alpha_P f_P G(\overline{G}^{-1}(i)|a_P) - \alpha_R f_R G(\overline{G}^{-1}(i)|a_R) \\ &- \lambda_P G_a(\overline{G}^{-1}(i)|a_P) - \lambda_R G_a(\overline{G}^{-1}(i)|a_R)\end{aligned}$$

## Mid-Censorship Optimal

- Gain function combines two incentive components plus weights:

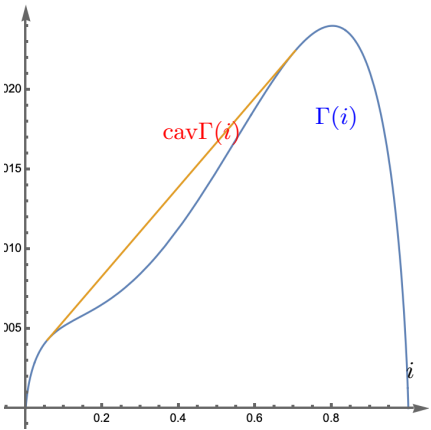
$$\begin{aligned}\Gamma(i) = &- \alpha_P f_P G(\overline{G}^{-1}(i)|a_P) - \alpha_R f_R G(\overline{G}^{-1}(i)|a_R) \\ &- \lambda_P G_a(\overline{G}^{-1}(i)|a_P) - \lambda_R G_a(\overline{G}^{-1}(i)|a_R)\end{aligned}$$

**Proposition.** If $\log g(y|a) = f(y) + r(y)m(a) - b(a)$ (with $r, m, b$ increasing and $m > 0$), then optimal rating has at most **four alternating intervals** of pooling and revelation.

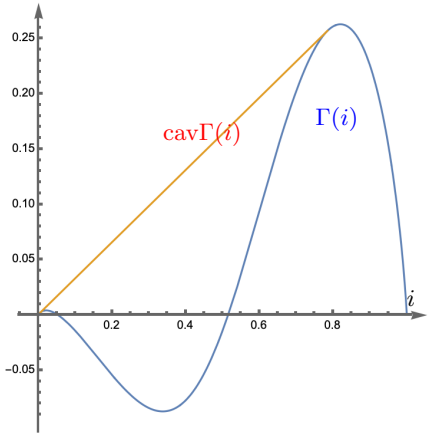
- Covers normals, log-normals, Gumbel, Fréchet – where $a$ controls either mean or scale
- Optimal rating can be **mid-censorship**: pool middle, reveal extremes

## Example: Maximize payoff of high-cost type $y = a + \varepsilon$, $\varepsilon \sim \mathcal{N}(0,1)$

- $\alpha_P = 1/f_P$, $\alpha_R = 0$, $k_R = 1/2$, $f_P = f_R = 1/2$




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




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

## Related Literature

- **Information design with strategic state:** Frankel and Kartik (2019), Ball (2025), Perez-Richet and Skreta (2022)
- **Bayesian persuasion with moral hazard:** Boleslavsky and Kim (2018), Madsen (2025)
- **Majorization in mechanism design:** Kleiner, Moldovanu and Strack (2021), Bergemann, Heumann and Morris (2022)
- **Certification with moral hazard:** Albano and Lizzeri (2001), Zubrickas (2015), Zapechelnyuk (2020), Onuchic and Ray (2023)
- **Multi-task moral hazard:** Holmstrom and Milgrom (1991), Dewatripont, Jewitt and Tirole (1999)
- **Empirical certification & disclosure:** Hui, Saeedi et al., Nosko and Tadelis (2015), Vatter (2021), Barahona, Otero and Otero (2023), Berg, Kölbel and Rigobon (2022)

## Conclusion

- A general framework for rating design under moral hazard
  - Interim prices as sufficient statistic
  - Optimal ratings from *concavifying a gain function*
  - Simple ratings (monotone partitions) are optimal

## Conclusion

- A general framework for rating design under moral hazard
  - Interim prices as sufficient statistic
  - Optimal ratings from *concavifying a gain function*
  - Simple ratings (monotone partitions) are optimal
- Effort–variance interaction shapes optimal ratings:
  - MLRP $\rightarrow$ full revelation
  - ELRP $\rightarrow$ lower censorship (insure innovation)
  - CLRP $\rightarrow$ upper censorship (deter negligence)
  - Redistribution $\rightarrow$ pool low realizations

## Conclusion

- A general framework for rating design under moral hazard
  - Interim prices as sufficient statistic
  - Optimal ratings from *concavifying a gain function*
  - Simple ratings (monotone partitions) are optimal
- Effort–variance interaction shapes optimal ratings:
  - MLRP $\rightarrow$ full revelation
  - ELRP $\rightarrow$ lower censorship (insure innovation)
  - CLRP $\rightarrow$ upper censorship (deter negligence)
  - Redistribution $\rightarrow$ pool low realizations
- Applications:
  - Window-dressing: cheaper manipulation $\rightarrow$ less informative ratings
  - Nonreducible multi-task: upper censorship can dominate full revelation
  - Redistributive test design: *mid-censorship* optimal

## Conclusion

- A general framework for rating design under moral hazard
  - Interim prices as sufficient statistic
  - Optimal ratings from *concavifying a gain function*
  - Simple ratings (monotone partitions) are optimal
- Effort–variance interaction shapes optimal ratings:
  - MLRP $\rightarrow$ full revelation
  - ELRP $\rightarrow$ lower censorship (insure innovation)
  - CLRP $\rightarrow$ upper censorship (deter negligence)
  - Redistribution $\rightarrow$ pool low realizations
- Applications:
  - Window-dressing: cheaper manipulation $\rightarrow$ less informative ratings
  - Nonreducible multi-task: upper censorship can dominate full revelation
  - Redistributive test design: *mid-censorship* optimal
- "*More transparency*" is not a universal remedy

## Majorization

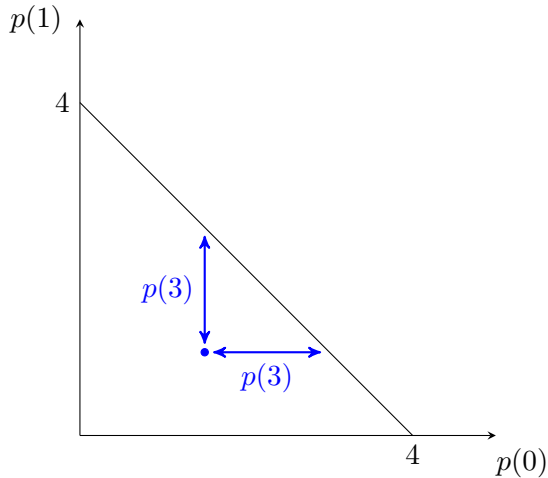
**Definition.** For a r.v. $y \sim H$, $f(y) \succcurlyeq_{\text{maj}} g(y)$ (equivalently, $g(y) \succcurlyeq_{\text{cv}} f(y)$ or $f(y) \succcurlyeq_{\text{cx}} g(y)$) iff

$$\int u(f(y))\, dH \geq \int u(g(y))\, dH, \ \forall u : \text{convex}$$

Back

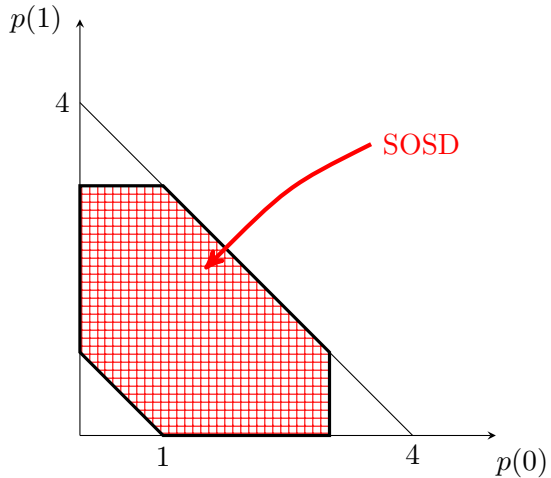
## Example

- $Y = A = \{0, 1, 3\}$, $v = y$, prior: $\mu(\{y\}) = 1/3$
- Interim prices: $\frac{p(0)+p(1)+p(3)}{3} = \frac{4}{3}$



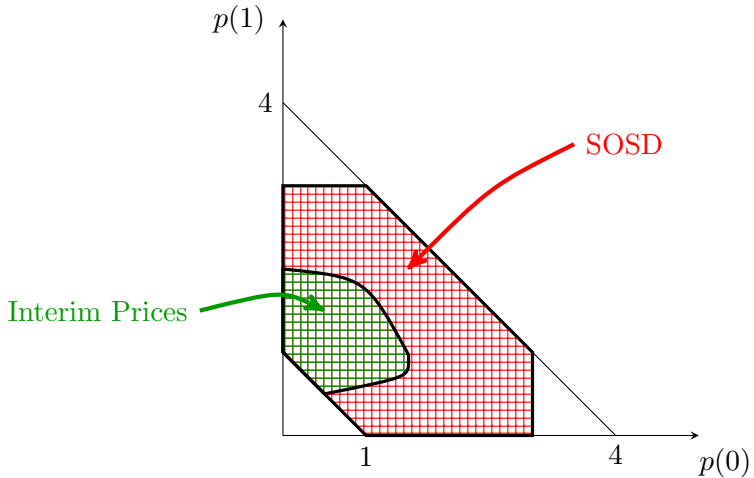
## Example

- $Y = A = \{0, 1, 3\}$, $v = y$, prior: $\mu(\{y\}) = 1/3$
- Interim prices: $\frac{p(0)+p(1)+p(3)}{3} = \frac{4}{3}$



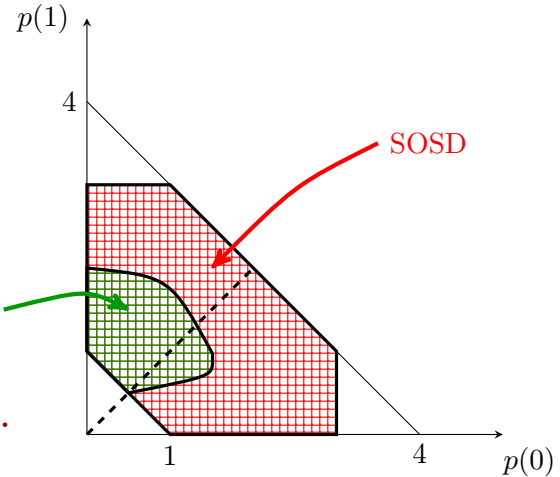
## Example

- $Y = A = \{0, 1, 3\}$, $v = y$, prior: $\mu(\{y\}) = 1/3$
- Interim prices: $\frac{p(0)+p(1)+p(3)}{3} = \frac{4}{3}$



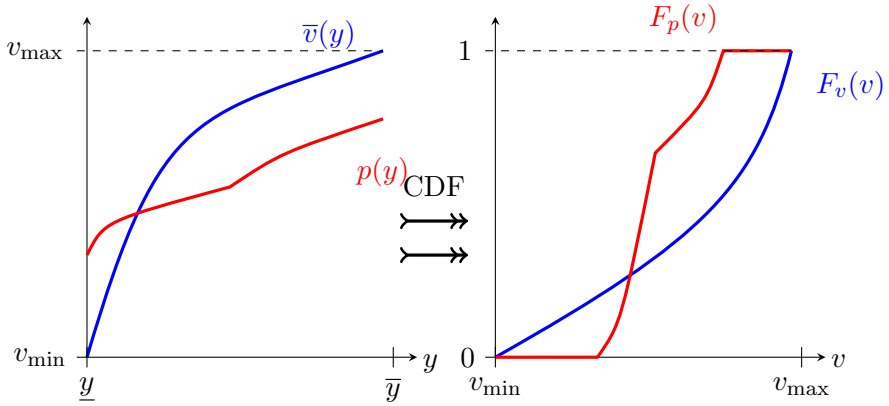
## Example

- $Y = A = \{0, 1, 3\}$, $v = y$, prior: $\mu(\{y\}) = 1/3$
- Interim prices: $\frac{p(0)+p(1)+p(3)}{3} = \frac{4}{3}$




**Weight of $y$ in $p(y)$ is at least $1/3$.**

# Reformulation: Quantiles and Concavification




$$\overline{v}(y) \succcurlyeq_{\text{maj}} p(y) \;\Leftrightarrow\; F_v(v) \succcurlyeq_{\text{cv}} F_p(v)$$

## Idea of Proof of Proposition T1

- Steps:
  - Assume support of $y$, $Y$, is finite
  - Use induction to construct $\pi$
  - Approximate compact $Y$'s
- Suppose $Y$ is finite, market values $\{\overline{v}_1 < \cdots < \overline{v}_n\}$
- Comonotonicity: $p_1 \leq \cdots \leq p_n$
- Class of signal structures: for given $i : 1 \leq i \leq n-1$

$$\pi(\{s\}|y) = \begin{cases} \lambda & s = y \\ (1-\lambda)\hat{\pi}(\{s\}|y) & s \in \hat{S} \end{cases}, \quad \hat{\pi}(\{s\}|y_i) = \hat{\pi}(\{s\}|y_{i+1})$$

- Reveals state with prob. $\lambda$; otherwise pools $i$ and $i+1$
- Use induction hypothesis