Optimal Simple Ratings*

Hugo Hopenhayn†

UCLA

Maryam Saeedi‡

Tepper, CMU

February 19, 2024

Abstract

Simple and coarse ratings are ubiquitous on many online platforms and within rating agencies. In this paper, we address the practical challenge faced by these intermediaries: how should these ratings be designed to maximize payoffs while considering various market conditions? And how do these simple ratings perform? We first propose a method to determine nearly optimal thresholds with minimal data requirements. Then, we provide theoretical and numerical evidence demonstrating that the value loss from using these simple ratings is small and diminishes significantly as the number of ratings increases. An application to two markets (eBay sellers and Medicare Advantage providers) is provided to illustrate our findings.

Keywords: Certification, Simple Ratings, Consumer Surplus, Producer Surplus. **JEL codes:** D21, D47, D60, D82, L11

*We thank Marina Halac, Andy Skrzypacz, Ali Shourideh, and Ariel Zetlin-Jones for very valuable comments. We also thank Benjamin Vatter for generously sharing the data related to Medicare Advantage providers with us.

†hopen@econ.ucla.edu

‡msaeedi@andrews.cmu.edu

1 Introduction

Simple ratings, which categorize sellers into a limited number of tiers, are ubiquitous across trading platforms and rating agencies. For instance, California's Department of Public Health classifies restaurants into three categories (A, B, and C), Michelin employs its traditional 3-star rating system, eBay categorizes sellers into badged and non-badged sellers, AirBnB provides a superhost badge to top-quality hosts, and most certification agencies also categorize the respective populations into certified and non-certified groups.[1] The design of these ever more popular simple rating systems raises two fundamental questions: first, what criteria should be employed to optimally design these ratings , and second, to what extent are market designers foregoing potential value by opting for simple ratings instead of more sophisticated ones?

To address these questions, we investigate the optimal design and performance of rating systems when the market designer is constrained to use a small number of ratings. To address the first question, we show that the optimal mechanism is given by threshold partitionsgive the necessary conditions for the thresholds and provide an algorithm for to calculate them.In the special case, when there are only two tiers, our results show how stringent the standards for certification should be, and how they vary with market characteristics, such as the distribution of firm quality and the elasticity of firm supply.[2] To address the second question, we theoretically and numerically analyze the welfare loss from using a coarse rating system instead of the optimal mechanism without constraints.

Our baseline model considers a competitive market with a large set of buyers and sellers, though our most important results apply equally to the canonical model of Cournot competition with constant marginal cost. Sellers are endowed with different levels of quality, which is the only source of product differentiation in our model.[3] Our analysis focuses on two main sources

[1]For example, the website ecolabelindex.com currently lists 456 certifiers for food and consumer products across 199 countries and 25 industry sectors. To the best of our knowledge, they all use these simple mechanisms, mostly certifying only a subset of the firms in the sector that meet some minimum requirements. Accessed September, 2023.

[2]For example, Hui et al. (2023) study the effects of increasing the requirements to become a badged seller on eBay and find that this leads to a higher market share of high-quality sellers while decreasing the sales of sellers in the medium range of quality.

[3]Moral hazard is a critical consideration in some markets, while adverse selection might play a more critical role

of market heterogeneity: the distribution of seller qualities and the responsiveness of sellers' supply to prices. Intuitively, the heterogeneity and skewness of seller quality affect the spread of prices across ratings, while the responsiveness of supply determines the resulting reallocation of output across these categories.

A particular design challenge faced by certifiers is how strict and how selective should the standards for certification be, and how does their choice depend on the distribution of quality and market characteristics. In answering these questions, we derive a necessary condition for defining optimal thresholds that correspond to an intuitive criterion. Consider a marginal seller with quality at the threshold between two adjacent intervals. Pooling this seller with those in the upper interval distorts its output upwards, while pooling with those in the lower interval results in a downward distortion. The threshold is optimal when the cost of these two distortions is the same. In other words, the benefit of the increased output by pooling the seller with the upper group instead of the lower one, should equate the extra cost of production. This intuitive criterion provides good guidance in the design of ratings, as pooling is more costly in areas where there is more quality dispersion or where supply is more responsive to prices. Congruently, we find that an increase in right (resp., left) skewness reduces (resp., increases) the share of producers with high ratings and increases (resp., decreases) the share of those with lower ratings. Similar considerations apply to the degree of convexity of the supply function.

While the necessary condition provides an intuitive criteria, solving for the optimal thresholds requires specifying the structure of demand and supply in the market and repeated computations of equilibrium values. As a first approximation, we provide a very simple heuristic to determine thresholds, which involves only information regarding the distribution of qualities, and corresponds to the solution of a standard clustering problem.[4] These thresholds are exact for two canonical cases: competitive equilibrium with a linear supply and the Cournot competition

in others, as suggested by Hui et al. (2018) in an empirical study using eBay data. Optimal rating design with moral hazard and adverse selection is considered in Saeedi and Shourideh (2022) in a simplified market environment. Shi et al. (2020) and Vatter (2021) also study optimal information disclosure in a model that includes moral hazard.

[4]This clustering problem can be solved by the $k-$means algorithm as introduced by MacQueen et al. (1967) and used extensively in machine learning and statistics.

model with constant marginal cost. Overall, they serve as very good approximations in a more general conditions.

Regarding the performance of ratings, we show that a one-threshold partition closes at least half of the surplus gap between the no-information case and full-information case for quality distributions that follow some general conditions, such as log-concave density. In our numerical computations, we find that this partition closes from 46% to nearly 77% of the gap, depending on the underlying distribution of qualities. This loss due to coarse ratings diminishes rapidly as the number of thresholds increases, implying that a simple and cost-effective system with a few tiers can achieve a large part of the full-information disclosure value. Our numerical results also show that these simple heuristic thresholds provide remarkably good performance even when the supply function is not linear, with the exception of cases where supply is very elastic and the distribution of sellers' quality features long tails. In those cases it becomes optimal to certify a very small share of sellers and uncertified sellers' market share is driven to zero.To further demonstrate our methods, we calculate optimal certification thresholds based on data from two markets: eBay sellers and Medicare Advantage providers.

In summary, these findings indicate that limiting the market designer to employ a limited number of ratings does not result in significant welfare loss. Additionally, given the intricacies involved in designing and deciphering more complex information systems, our results justify the widespread use of simple rating systems. Concerning the practical aspect of rating design, the strong overall performance and modest data prerequisites of our linear thresholds offer a good approximation to the optimal thresholds or, at the very least, serve as an excellent starting point.

Related Literature This paper uses the canonical model we developed in Hopenhayn and Saeedi (2023), where we study optimal information disclosure in a theoretical framework. In this paper, we use this model to answer a verydifferent question, motivated by the wide use of simple ratings. While our previous paper provides broad qualitative results (e.g. the optimality of interval ratings), this one goes much further, giving a detailed characterization for the design of optimal ratings with a finite partition. We provide an algorithm that solves for the optimal thresholds in the case of linear supply or Cournot and a gives good approximation more generally and illustrate our results in two applied settings. Here, we want to find the value loss that arises as a result of coarse rating instead of using the optimal unconstrained ratings that is usually the aim of theoretical papers, such as our own previous work and many others. A lot of empirical work on ratings point to the use of coarse ratings in practice, Vatter (2021), Saeedi (2019), Elfenbein et al. (2015), Fan et al. (2013), and Jin and Leslie (2003).[5] Other empirical papers analyze the effects of changes in marketplace feedback mechanisms on price and quality (e.g., Hui et al. (2016), Filippas et al. (2018), and Nosko and Tadelis (2015)); Hui et al. (2020) study the benefit of adding a second tier to the reputation mechanism to mitigate the cold-start problem and to promote entry. Even though the coarse ratings are very commonplace in practice and have been studied very extensively in empirical studies, they are not the outcome of most canonical theory papers.

In our study, we show that the losses due to coarse ratings can be low and as a result, they might be justified because of the simplicity they provide over a more complex mechanism. Our results parallel the results of the coarse matching literature. Wilson (1989) shows that losses relative to the full-information case are of order $1/n^2$ for a partition with n classes. This finding is consistent with our computed bounds in Section 4. Our theoretical bound on the gains from a two-tier certification is also related to the bounds found by the coarse matching literature, such as in McAfee (2002), Hoppe et al. (2011), and Shao (2016).

Section 2 describes the model. In Section 3, we find the conditions for the design of optimal simple ratings, and in Section 4, we study the performance of these simple ratings relative to unconstrained optimal information disclosure. Section 5 concludes the paper. Proofs are relegated to the appendix unless otherwise specified.

[5] Dranove and Jin (2010) provide an excellent survey of the earlier papers on certification.

2 The Model

There is a unit mass of sellers with qualities z distributed according to a continuous cumulative distribution function (cdf) $F(z)$ on the real numbers. Production technology is the same for all sellers and is given by a differentiable, strictly increasing, and strictly convex cost function of quantity $c(q)$ and, correspondingly, a strictly increasing twice continuously differentiable supply function $S(p)$. On the demand side, there is mass M of consumers who face a discrete choice problem, with preferences

$$U(z, \theta, p) = z + \theta - p,$$

where z is the quality of the good purchased, θ is a taste parameter measuring the preference for goods offered in this market vis-à-vis an outside option, and p is the price of the good. The taste parameter $\theta \geq 0$ is distributed according to a continuous and strictly increasing cdf $\Psi(\theta)$, while the outside good's utility (no purchase) is normalized to zero. Goods are differentiated only by their quality level, which is equally valued by all consumers. Given the linearity of the utility function in z, we can replace a good of quality z with a good of expected quality z and the utility of the consumer stays the same. Throughout the paper, we use z interchangeably as the quality or expected quality of a good.[6]

We assume the following timing: (a) information about seller qualities is provided by the planner, (b) based on this information, consumers form posteriors about each seller's expected quality; (c) given these posteriors, perfectly competitive equilibrium prices are determined as a function of expected quality, considering the supply response of each seller to the corresponding price.[7] We interpret $F(z)$ as the distribution of sellers' quality which is observed by the planner.[8] We assume all market participants have the same posterior information about the expected qualities of sellers after receiving any set of signals from the planner, represented by the distribution

[6]We used this canonical model in a previous paper, Hopenhayn and Saeedi (2023), to answer another question, what is the optimal solution to the optimal rating without considering the practical problem of simple ratings.

[7]While throughout the paper we use the assumption of perfect competition among sellers, our main characterization of an optimal simple rating holds unchanged in the canonical case of Cournot equilibrium with constant marginal cost, as shown in Section 3.3.

[8]Without loss of generality this can a be a garbling of this distribution.

function $G(z)$.[9] In particular, in a finite rating system, G will have a discrete distribution with point masses at the conditional mean qualities associated with each rating. We say that a seller has expected quality z if conditional on all signals received, that is the quality expected by all consumers.[10]

Given expected quality z, equilibrium prices arbitrage away the differences in expected quality, taking the form $p(z) = p(0) + z$, where $p(0)$ corresponds to the demand price of a hypothetical good of quality zero. This expression for prices guarantees that consumers are indifferent between goods with different signal realizations with any positive sales, which is a necessary condition for an equilibrium. The baseline price $p(0)$ determines the extent of the market, where the marginal consumer's θ is found by setting $U(0, \theta, p(0)) = 0$, or simply $\theta(p(0)) = p(0)$. All consumers with $\theta \geq p(0)$ will make their unit purchase, so aggregate demand is $Q = M(1 - \Psi(p(0)))$. Inverting this function, we can define an *inverse baseline demand* function,

$$P(Q) = \Psi^{-1}(1 - Q/M) = p(0). \tag{1}$$

On the supply side, each seller with expected quality z chooses its output, $q = S(p(z))$, so aggregate supply $Q = \int S(p(z))\, dG(z)$.

Definition. An (interior) *equilibrium,* given the distribution of expected qualities $G(z)$, is given by prices $p(z) = P(Q) + z$, where total quantity

$$Q = \int S(P(Q) + z)\, dG(z)\,. \tag{2}$$

[9] This representation of the information structure is consistent with the approach followed in Ganuza and Penalva (2010) and Gentzkow and Kamenica (2016). Given a common prior $F(z_0)$ over seller qualities and a signal structure π, we can let $G(z)$ be the distribution of the expected posterior of seller quality. Any information structure can thus be represented as a garbling of $F(z)$,which corresponds to the finest level of information available to the planner.

[10] If the planner gives all the information to buyers, then expected quality z will be equal to the actual quality of the seller, but when there is some pooling, this expected quality will be a function of other sellers that are pooled with the target seller. For example, a simple certification rating divides sellers into two groups, those certified and those not certified. As a result, there will be two different levels of expected quality and equilibrium prices, one for each group, with possibly many different levels of heterogeneity in quality within each group that are unobserved by buyers.

It is immediate to see that this last equation gives a necessary and sufficient condition for an equilibrium. Moreover, if $P(Q)$ is strictly decreasing and continuous, for any distribution G there will be a unique equilibrium value Q^* and under some regularity conditions, it will be interior.

Using the following assumption, Hopenhayn and Saeedi (2023) prove that a unique interior equilibrium exists.

Assumption 1. *There exists* $\widetilde{\theta}$ *in the support of* Ψ *such that*

$$M > \int S\left(\widetilde{\theta} + z\right) dG(z)$$

for all distributions G *such that* F *is a mean-preserving spread of* G*. In addition,* $\int S(p(0) + z)\, dG(z) > 0$ *for the same class of distributions.*

The first assumption rules out the possibility that all consumers make purchases in this market; in other words, we assume that the consumers are on the long side of the market.[11] The second assumption rules out no output as an equilibrium. Although a corner equilibrium, if it exists, is also unique, we rule this out as a matter of convenience.

2.1 Optimal Ratings

In order to define optimal ratings, we first derive expressions for consumer and producer surplus. Consider a consumer of type θ who buys a good of expected quality z and therefore receives utility $\theta + z - p(z)$. Given the equilibrium price $p(z) = P(Q) + z$, the consumer's net utility is $\theta - P(Q)$. It follows that total consumer surplus is

$$M \int_{P(Q)} (\theta - P(Q))\, d\Psi(\theta) = \int_0^Q (P(x) - P(Q))\, dx,$$

[11] As explained below, the assumption spans the set of all possible information structures.

where the equality follows from the change of variables $x = M(1 - \Psi(\theta))$ and our definition of $P(Q)$: $P(Q) = \Psi^{-1}(1 - Q/M)$. This implies that consumer surplus will move in the same direction as market size, as given by total quantity Q. Producer surplus is defined more simply as total profits:

$$\Pi(Q) = \int \pi(P(Q) + z)\, dG(z).$$

A couple of observations are in order. First, notice that for fixed Q, producers benefit from better information. This follows from the convexity of the profit function and the fact that better information is defined by mean preserving spreads. But changes in the information structure might also affect the equilibrium value of total output Q, which is also market size as measured by the number of consumers served. This equilibrium effect has opposite impacts on consumer and producer surplus: whereas consumers prefer a larger market, producers prefer a smaller one. This countervailing effect of total quantity leads to a conflict of interest between consumers and producers in the design of an optimal rating system.

We take the planner's objective to be the maximization of total surplus, the unweighted sum of consumer and producer surplus. For any information structure as given by the posterior distribution of mean quality G, total surplus is then given by

$$TS(G) = \int \pi(P(Q) + z)\, dG(z) + \int^{Q} (P(x) - P(Q))\, dx,$$

where Q is the unique equilibrium output corresponding to G. This equation simplifies to

$$TS(G) = \int^{Q(G)} (P(x))\, dx + \int (zq(z) - c(q(z)))\, dG(z). \tag{3}$$

An *optimal simple rating* is given by the distribution G that maximizes this objective among a class of distributions defined in the following section.

3 Simple Ratings: Design

In this section, we focus on simple ratings that partition the set of sellers into N groups. We assume that consumers have no information other than that provided by the certifier.

Following our earlier discussion on information structures, the certifier's information can be summarized by a distribution of posterior mean qualities that, in order to avoid further notation, we denote by $F(z)$. This is the basis on which the certifier classifies sellers into rating bins. To simplify the exposition, we refer to the expected value z as the quality of the seller. We assume F is differentiable on its support with density $f(z)$.

A *threshold partition* totally orders sellers into N quality intervals. The following lemma proves that threshold partitions are the best among the set of finite partitions.

Lemma 1. *An optimal simple rating is given by a threshold partition.*

Given this lemma, the design of an optimal simple rating system reduces to finding the vector of optimal thresholds, $\boldsymbol{z} = (z_1....z_{N-1})$, that divide sellers into N partitions, $\{[z_0, z_1], [z_1, z_2], ..., [z_{N-1}, z_N]\}$, where z_0 and z_N are the lower and upper supports of the distribution of expected qualities given the planner's information ($-\infty$ or $+\infty$ if unbounded), respectively.[12]

3.1 A Necessary Condition

In this section, we derive a simple and intuitive necessary condition to characterize these optimal thresholds. Let $M_k = m(z_{k-1}, z_k), k = 1, ..., N$ denote the conditional means of sellers' quality within the interval $[z_{k-1}, z_k]$. Let $Q(\boldsymbol{z})$ denote the unique equilibrium total quantity at the optimal threshold vector $\boldsymbol{z} = (z_1....z_{N-1})$. The prices for sellers in partition $[z_{k-1}, z_k]$ are denoted by $p_k = P(Q(\boldsymbol{z})) + M_k$, and quantities, by $q_k = S(p_k)$. From equation (3) it follows that total surplus for partition $\boldsymbol{z}$ is given by

$$W(z) = \int_0^{Q(\boldsymbol{z})} P(x)\,dx + \sum_{k=1}^{N} [F(z_k) - F(z_{k-1})][M_k q_k - c(q_k)]. \tag{4}$$

[12] Bergemann and Pesendorfer (2007) find a similar result in the context of optimal information design in auctions.

Taking first-order conditions with respect to z_k proves the following necessary condition:

Lemma 2. *Let the thresholds $\boldsymbol{z} = (z_1, ..., z_{N-1})$ maximize (3). Then*

$$(P(Q(\boldsymbol{z})) + z_k)(q_{k+1} - q_k) \quad = \quad c(q_{k+1}) - c(q_k) \tag{5}$$

for all z_k.

Condition (5) has an intuitive interpretation. Consider a marginal seller with quality at the threshold between two adjacent intervals. For this threshold to be optimal, the planner should be indifferent between pooling this marginal seller with those in the lower or upper interval. For this threshold to be optimal, the planner should be indifferent between pooling this marginal seller with those in the lower or upper interval.; the seller underproduces when pooled to the lower group and overproduces when pooled to the upper one. The left hand side of (5) shows the marginal value obtained by increasing the quantity of the marginal seller with quality z_k, from q_k to q_{k+1}. The right hand side shows the difference in cost. This condition highlights the relevance of the supply behavior of sellers, in particular, the curvature of the supply function, as it impacts both the changes in total output and the production cost.

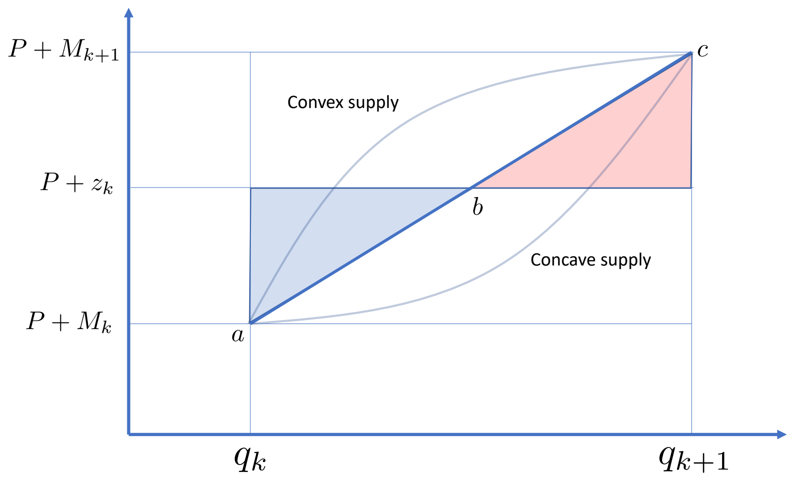
Figure 1 provides a graphical representation of this necessary condition and its connection to the supply function. Three cases are considered: a linear supply function, given by the solid diagonal line, an upper convex supply (concave marginal cost function), and a lower concave supply (convex marginal cost function).[13] The area below the marginal cost function, i.e., supply function, between q_k and q_{k+1} equals the right hand side of (5), whereas the area under the line $P + z_k$ equals the left hand side of the equation. The difference between these two areas is

$$\int_{q_k}^{q_{k+1}} (P + z_k - C'(q))\, dq,$$

which equals zero if and only if condition (5) holds, at the optimal threshold level for z_k. In the linear case, the integrand is positive up to point b (the triangle in blue) and negative thereafter.

[13] We follow the practice of putting price on the y-axis and quantity on the x-axis.

Figure 1: Optimal Pattern



Point $P + z_k$ is such that the regions from a to b (the triangle in blue) and from b to c (the triangle in red) have the same areas. It is immediate that in the linear case, the corresponding value of $z_k = (M_{k+1} + M_k)/2$. It also follows easily that for the convex supply case, $P + z_k$ must be higher so the two corresponding areas will have the same area, whereas the converse holds for the concave supply case.

Proposition 1. *Let the thresholds* $\boldsymbol{z} = (z_1, ..., z_{N-1})$ *maximize (3), and denote by* $M_k = E(z|z_{k-1} \leq z \leq z_k)$ *the corresponding conditional means. Then*

1. $z_k = (M_k + M_{k+1})/2$ *if the supply function* $S(p)$ *is linear;*

2. $z_k > (M_k + M_{k+1})/2$ *if the supply function is strictly convex in the interval* $[M_k, M_{k+1}]$*; and*

3. $z_k < (M_k + M_{k+1})/2$ *if the supply function is strictly concave in the interval* $[M_k, M_{k+1}]$*.*

A simple characterization for the solution in the linear supply case and conditions for uniqueness are provided in the following proposition.

Proposition 2. *If the supply function is linear, the optimal thresholds* $\boldsymbol{z} = (z_1, ..., z_N)$ *are the ones* *that minimize*

$$\sum_{k=1}^{N} \int_{z_{k-1}}^{z_k} (z - M_k)^2 \, dF(z) \, . \tag{6}$$

If in addition F has log-concave density, the solution to this minimization problem is unique.

According to this proposition, the optimal thresholds for the linear supply case are the ones that minimize the sum of the variance of qualities within partitions. This objective coincides with the popular $k-$means criteria for clustering as introduced by MacQueen et al. (1967), commonly used in the machine learning and statistics literature. Therefore, estimating the optimal thresholds is a trivial task, because many software programs incorporate algorithms to solve this problem. Additionally, the linear supply function helps us simplify the profit function for the sellers. For a seller with expected quality z, profits are equal to $p(z)^2/2 = (P(Q)+z)^2/2$. Therefore, for any distribution G of observed qualities, total profits are

$$\begin{aligned} \Pi &= \frac{1}{2} \int (P(Q)+z)^2 \, dG(z) \\ &= \frac{1}{2} P(Q)^2 + P(Q)\bar{z} + \frac{1}{2} \int z^2 dG(z) \, . \end{aligned} \tag{7}$$

In addition, when the supply function is linear, the equilibrium condition 2 reduces to

$$Q = \alpha \int P((Q)+z) \, dG(z) = \alpha P(Q) + \bar{z},$$

so the equilibrium quantity Q does not depend on G and thus on the partition. Consequently, maximizing total profits as given in equation (7) is equivalent to maximizing $\int z^2 dG(z)$.

Proposition 1 suggests that the thresholds for the linear case can be lower (resp., upper) bounds for the case of convex (resp., concave) supply. Although this proposition gives the criteria for local deviations for a single threshold, convex to the right and concave to the left, starting with those obtained for the linear case, it does not imply an ordering of the whole vector of thresholds. The following proposition gives the conditions for total ordering.

Proposition 3. *Suppose the quality distribution $F(z)$ has a log-concave density. Let $\left(z_1^L, \ldots, z_{N-1}^L\right)$*

be the optimal thresholds for the linear supply case. The optimal vector of thresholds $(z_1, ..., z_{N-1})$ *for a convex (resp., concave) supply function is pointwise higher (resp., lower) than* $\left(z_1^L, ..., z_{N-1}^L\right)$ *.*

The formula in equation (6) gives a simple characterization for the optimal thresholds in the linear supply case that depends only on the distribution of qualities, and, by the previous proposition, provides a lower (resp., upper) bound when the supply function is convex (resp., concave).

This proposition also suggests that the thresholds for the linear case can provide a good reference point in solving for the optimal thresholds for non-linear supply. In practice, online platforms, such as eBay and Airbnb, often experiment with different certification thresholds. The thresholds for the linear case are very easy to compute, without the need of knowing market characteristics aside from the distribution of qualities, and provide a good starting point. Moreover, as seen in Section 4.2, these simple linear thresholds perform remarkably well even when the supply function is not linear, thus providing a simple heuristic for the calculation of approximately optimal thresholds in general.

There is an additional reason why the solution to the linear case is of interest. As we show in the following section, this solution coincides with the optimal thresholds under Cournot competition with constant marginal costs, a workhorse model of imperfect competition. In consequence, the numerical results and applications we provide in Section 4 for the linear supply case also apply to Cournot competition.

3.2 Skewness and Optimal Thresholds

The optimal thresholds, as depicted in equation 6, will depend on the distribution of sellers' quality (i.e., F distribution). In this section, we study how skewness in the distribution of qualities impacts this optimal choice. In particular, we show that in the simple case of a two-tier certification, the optimal threshold is skewed in the same direction as the distribution of qualities. Then, we extend this result, providing general comparative statics for the vector of thresholds with respect to an appropriately defined skewness ordering.

Before proceeding to the analysis, we provide some intuition behind our results. Consider the case of one certification threshold, z^*. The following trade-off appears when deciding how strictly to draw the line separating the upper and lower segments. When putting z^* in the upper group, there is an upward distortion of the supply of the seller at z^*, which is a function of the distance $M_H - z^*$. This distance also measures the extent to which the seller at z^* gains from being pooled with higher-quality sellers. When putting z^* in the lower group, there is a downward distortion of the supply of the seller at z^*, which is a function of $M_L - z^*$. This distance also measures the extent to which the seller at z^* loses from being pooled with lower-quality sellers. Right skewness (resp., left skewness) of the distribution $F(z)$ will increase (resp., decrease) the upward distortion and decrease (resp., increase) the downward distortion, making it optimal to have more restrictive (resp., less restrictive) certification standards.

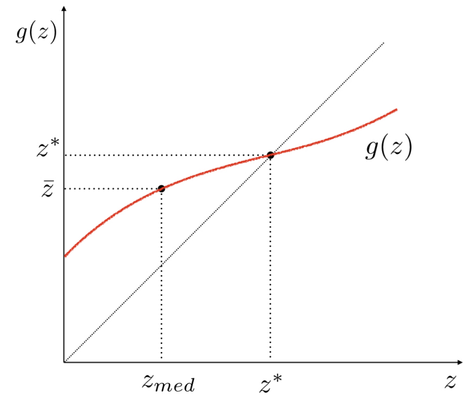
The condition given in Proposition 2 for linear supply functions implies

$$z^* = \frac{1}{2}\left(M_L(z^*) + M_H(z^*)\right), \tag{8}$$

which can be used to relate this threshold to properties of the distribution. Consider first the case of a symmetric distribution (i.e., where the median, z_{median}, equals the mean, $\bar{z}$). Given that any z^*, $F(z^*) M_L + (1 - F(z^*)) M_H = \bar{z}$, setting the threshold $z^* = \bar{z} = z_{median}$ would satisfy the above condition.

The same reasoning suggests that when F is skewed, the optimal threshold will also be skewed relative to the mean in the same direction. This can be easily proved, as follows. Consider the case of a right skewed distribution where $\bar{z} > z_{median}$. Let $M_L(.)$ and $M_H(.)$ denote functions equal to the conditional average of the quality of sellers below and above any value within the range of qualities, respectively. Furthermore, denote $g(z) = \frac{1}{2}(M_L(z) + M_H(z))$. Following Proposition 1, the optimal threshold is a fixed point of this function. When $z \to z_{max}$ (or as $z \to \infty$ in the case of unbounded support), $g(z) \to \frac{1}{2}\bar{z} + \frac{1}{2}z < z$, and when $z \to z_{min}$(or as $z \to -\infty$ in the case of unbounded support), $g(z) \to \frac{1}{2}z_{min} + \frac{1}{2}\bar{z} > z$. For $z = z_{median}$, $g(z) = \bar{z} > z$. Because the function $g(z)$ is increasing and continuous, the unique fixed point

Figure 2: z^* When the Mean Is Greater than the Median



z^* must be to the right of z_{median} and, as a consequence, $z^* > \bar{z}$, as illustrated in Figure 2. The result for the case of left skewness can be shown similarly. Consistently with these results, the empirical examples in Section 4.4 have both left-skewed distributions and the optimal thresholds we found are below the respective means.

Table 1 shows the optimal threshold for a series of distributions, as well as the corresponding fraction of certified sellers. All distributions in our example are skewed to the right except for one, so according to our argument, $z^* > \bar{z} > z_{median}$ and it is optimal to have a smaller share of sellers certified. This is shown in the fifth column of Table 1. As an example, for the Pareto distributions only a small fraction should get certified, 5% when the power parameter is 3, and 9% when the power parameter is 4.[14] For the exponential distribution, only 20% of sellers should be certified regardless of the hazard rate.

Now, we extend our findings to the case of multiple signals under a stronger skewness order. This skewness order, the convex (resp., concave) order, was originally proposed by Van Zwet (1964).

Definition. Distribution $\tilde{F}$ is more skewed to the right than F if $\tilde{F}^{-1}(F(x))$ is convex; equivalently, there exists an increasing convex function $g(x)$ such that $\tilde{F}(g(x)) = F(x)$.[15]

We can think of this ordering as stretching to the right the quality scale with the transforma-

[14]When $\alpha \leq 2$, the value of z^* is undefined, as total surplus is strictly increasing in z^* in all the support.
[15]Note that this definition implies that $F^{-1}(F(x)) = g^{-1}(x)$ is concave.

tion $g(x)$. As an example, if F is a uniform distribution in $[0,1]$ and $g(x) = x^2$, then $\tilde{F}(x^2) = x$ or, equivalently, $\tilde{F}(x) = x^{1/2}$.

Proposition 4. *Suppose the supply function is linear. Let F be a distribution with log-concave density and $\tilde{F}$ a distribution such that $\tilde{F}(g(z)) = F(z)$, where g is a strictly convex increasing function. Let $\{l_k\}$ be the optimal thresholds for F, and $\{g(z_k)\}$, the optimal thresholds for $\tilde{F}$. Then $z_k > l_k$ for all k.*

This proposition implies that for all k, $\tilde{F}(g(z_k)) = F(z_k) > F(l_k)$, so the percentiles defined by the two optimal thresholds are ordered. In particular, for a two-tier certification rating, the share of certified sellers should be lower for distribution $\tilde{F}$. An example is given in Table 1 for the case of power distributions $F(z) = z^\alpha$. It is easily shown that the distribution with $\alpha = 0.5$ is more skewed to the right than the one with $\alpha = 2$.[16] Consistently with the previous proposition, the share of certified sellers is lower when $\alpha = 0.5$.

3.3 Cournot Competition

Up to here we have assumed that sellers are price takers and are in perfect competition. In this section, we extend our analysis to the case of Cournot competition among sellers with constant marginal cost, c. Our main result is that the optimal thresholds in this setting coincide with those derived above for the perfect competition case with linear supply.

There are in total n sellers in the market. Assuming an interior solution, the first-order condition for a seller of quality level z is given by

$$MR(z) = P'(Q)\,q(z) + P(Q) + z = c. \tag{9}$$

[16] Take $g(x) = x^4$.

Summing over all sellers gives

$$P'(Q)\,Q + nP(Q) + n\bar{z} = nc, \tag{10}$$

where $\bar{z}$ is the average quality of sellers in the market, which is assumed to be exogenous.[17] Equation (10) determines Q independently of the distribution of z up to its mean, which is thus independent of the information structure.[18] Furthermore, from equation (9) it follows that

$$q(z) = \frac{P(Q) + z - c}{-P'(Q)},$$

so it is linear in z, given Q. Seller profits are given by

$$\begin{aligned} \pi(z) &= [(P(Q) + z) - c]\,q(z) \\ &= -P'(Q)\,q(z)^2 \\ &= \frac{[(P(Q) + z) - c]^2}{-P'(Q)} \\ &= \frac{(P(Q) - c)^2 + 2(P(Q) - c)\,z + z^2}{-P'(Q)}. \end{aligned}$$

Consumer surplus is independent of the information structure, because it depends only on Q. Thus, maximizing total surplus is equivalent to maximizing total profits. For any information structure $G(z)$, total profits are given by

$$\Pi = \frac{1}{-P'(Q)}\left[(P(Q) - c)^2 + 2(P(Q) - c)\,\bar{z} + \int z^2 dG(z)\right].$$

All the terms involving Q and $\bar{z}$ are independent of the information structure G. Thus, total

[17] The value Q that solves this equation is unique provided marginal revenue is decreasing for the average-sized seller.

[18] This is true for an interior equilibrium where no sellers are excluded from production. A sufficient condition is that $P(Q) + \underline{z} > c$, where Q is the solution to equation (10). .

surplus can be written as

$$S_0 + a \int z^2 dG(z) \tag{11}$$

for constants S_0 and a. Hence surplus maximization is equivalent to maximizing $\int z^2 dG(z)$, as in the case of perfect competition with linear supply (see equation 7).[19]

4 Performance of Simple Ratings

In this section, we seek to answer two questions: How well do simple thresholds perform and how good an approximation are the thresholds obtained for the linear case more generally. We first start with the linear supply case where, as shown above, the optimal thresholds are very easy to solve for. In section 4.1, we show that, under some regularity assumptions, a two-tier partition (i.e., certifying a subset of sellers) achieves at least half of the maximum possible welfare gain from an unrestricted information design. Section 4.2, provides numerical calculations for a broad set of distribution functions, and shows that, the gap between the maximum possible welfare gains from full information and those obtained with simple ratings rapidly shrink as the number of ratings increases. Section 4.3, shows that the optimal thresholds for the linear supply provide a good approximation to the optimal thresholds for a wide range of supply functions and distribution of sellers' quality. To illustrate our findings, Section 4.4 solves the optimal thresholds for two empirically relevant distributions of quality, corresponding to Medicare Advantage insurers and eBay sellers, respectively.

Total surplus is maximized with full information, as shown in Hopenhayn and Saeedi (2023). Our measure of performance of a simple rating system is the fraction of the gap between the full-information and no-information cases that is closed with this rating system. Given that in the case of linear supply, or Cournot equilibrium, total quantity and consumer surplus are invariant to the information structure, our performance measure coincides with the fraction of the gap in producer surplus that is covered by the simple rating, using equation 7. Under full information,

[19] Here we have considered quantity competition. For a model of price competition with partially informed consumers, see Moscarini and Ottaviani (2001).

total profits are $\frac{1}{2}P(Q)^2 + P(Q)\bar{z} + \frac{1}{2}\int z^2 dF(z)$. Therefore, the surplus gap with respect to the full-information case is

$$\Delta\Pi = \frac{1}{2}\left(\int z^2 dF(z) - \int z^2 dG(z)\right)$$

for any distribution of a rating system, G, that is a garbling of F.[20] In particular, the maximum surplus gap 12between the full-information and no-information cases is

$$\overline{\Delta\Pi} = \frac{1}{2}\left(\int z^2 dF(z) - \bar{z}^2\right). \tag{12}$$

For a threshold partition $(z_1, ..., z_{N-1})$, G has N mass points at the conditional means $M_1, ..., M_N$; therefore, we can write the surplus gap as

$$\begin{aligned} \Delta\Pi &= \frac{1}{2}\sum_{k=1}^{N}\int_{z_{k-1}}^{z_k}\left(z^2 - M_k^2\right) dF(z) \\ &= \frac{1}{2}\sum_{k=1}^{N}\int_{z_{k-1}}^{z_k}\left[(z - M_k)^2\right] dF(z). \end{aligned} \tag{13}$$

This equation corresponds to the loss function used in $k - means$ clustering, given that at the optimal thresholds, as defined earlier, the expected values M_k are precisely the centroids of the corresponding intervals $[z_{k-1}, z_k]$. We are interested in seeing how much of the possible total surplus gain from information, as expressed in equation 12, is captured by 13, or, simply, the following ratio:

$$\begin{aligned} \gamma \equiv \frac{\overline{\Delta\Pi} - \Delta\Pi}{\overline{\Delta\Pi}} &= 1 - \frac{\sum_{k=1}^{N}\int_{z_{k-1}}^{z_k}(z - M_k)^2 dF(z)}{\int z^2 dF(z) - \bar{z}^2} \\ &= \frac{\sum_{k=1}^{N}(F(z_k) - F(z_{k-1}))(M_k - \bar{z})^2}{\int (z - \bar{z})^2 dF(z)}, \end{aligned}$$

which is the ratio of the variance between the conditional mean qualities and total variance. Intuitively, this bound is a measure of the relative importance of the variance between the means

[20]The total quantity stays the same given the assumption of linear supply.

of the partitions, separated by their ratings, and the variance that remains in each pool. This connection to variance decomposition is used below to derive a theoretical bound on ratings' performance.

4.1 Theoretical Bounds

The simplest coarse rating scheme is a two-tier certification, widely used in many settings. The next proposition provides a useful bound for the gains from certification that builds on the variance decomposition described earlier. The corollary that follows gives sufficient conditions so that a two-tier rating achieves at least half of the surplus of the full-information case.

Proposition 5. *The relative performance of a two-tier setting satisfies*

$$\gamma \geq \frac{1}{1 + \max\left\{cv_1^2, cv_2^2\right\}},$$

where cv_1 is the coefficient of variation of $z - \bar{z}$ conditional on $z < \bar{z}$, and cv_2 is the coefficient of variation of $z - \bar{z}$ conditional on $z \geq \bar{z}$, where $\bar{z}$ is the mean of sellers' qualities.

Corollary 1. *Suppose that the distribution F has an increasing hazard rate and a decreasing reverse hazard rate. Then a two-tier rating achieves at least half of the surplus of the full-information case.*

Proof. From a well-known result from Stoyan and Daley (1983) (pp. 16–19), the conditions of this corollary imply that $cv_1 < 1$ and $cv_2 < 1$. Using the bound in Proposition 5 completes the proof. □

The conditions given in the corollary are satisfied by a large class of distributions that include all those with log-concave densities, such as uniform, normal, exponential and double exponential, logistic, extreme value, and many others with some restriction on parameters (e.g., power function $F(z) = z^c$ for $c \geq 1$.) Related bounds for two-sided matching problems can be found in McAfee (2002); Hoppe et al. (2011); Shao (2016). The results of Wilson (1989) imply that the losses from $N-$ratings are of order $1/N^2$.

This proposition in addition to the corollary has an important practical implication, for a wide range of distributions, certification can achieve at least half of total possible surplus. Proposition 5 suggests that even using just median as the certification threshold can be enough to reach this level. Next section we numerically show that these losses decrease rapidly as the number of signals increase. In Section 4.3 we show that even though our result here is for linear supply case, or counot with fixed marginal cost, using the same thresholds for non-linear supply achieves a signifanct share of surplus.

4.2 Numerical Results

We now examine the numerical results for a variety of distribution functions often used in the economics literature. Table 1 reports the share of the total surplus gap that is closed with partitions of different sizes n. As can be seen from the calculations, a one-threshold (certification) partition closes from near 50% to almost 80% of the total surplus gap, depending on the underlying distribution of qualities. The only case where one threshold cannot reach 50% of the benefits is the Pareto distribution, which does not satisfy the conditions stated in 1. The gains are diminishing as the number of thresholds increases. Even though total surplus increases with the number of tiers, our numerical results suggest that most gains are attained with a small number of ratings.

The numerical results and the previous theoretical bounds suggest that in practice we may not need very complicated mechanisms to attain most of the benefits from information provision. In this paper, we do not explicitly model any cost related to providing information; but in practice, providing information may involve various costs for both the market designer and consumers. First, we have assumed that the market designer has costless access to information on the quality of sellers; however, in practice, getting precise information may increase the market designer's cost of designing the rating system. Secondly, it might be costly for the market designer to convey finer information to consumers and for them to process this information. Therefore, these costs and the small benefits from more complex information mechanisms might justify the coarse

Table 1: Optimal Thresholds

Distribution	Case	Mean/Median	z^*	Certified	% Surplus Gap Closed			
					$n=2$	$n=3$	$n=5$	$n=10$
Pareto	$\alpha=3$	1.19	2.73	5	46	68	84	94
	$\alpha=4$	1.12	1.84	9	54	74	89	97
Exponential	all	1.45		20	65	82	93	98
$F(z)=z^{\alpha}$	$\alpha=0.5$	1.32	0.41	36	77	90	97	99
$z\in[0,1]$	$\alpha=2$	0.94	0.62	62	72	87	95	99
Log-normal	$\sigma=0.25$	1.03	1.09	36	63	81	92	98
$(\mu=0)$	$\sigma=1$	1.64	4.25	7	55	75	89	97

Note: The above calculations correspond to the linear supply case.

information mechanisms seen in practice.

4.3 Performance of Linear Thresholds for Nonlinear Supply

The thresholds derived above for the case of linear supply are easy to calculate, but how well do they perform more broadly? Table 2 compares the performance of the linear thresholds to the optimal ones for the distributions considered earlier and supply functions $s(p)=p^{\theta}$ for $\theta\in\left\{\frac{1}{4},\frac{1}{2},2,4\right\}$. The table provides values for the fraction of certified firms and the welfare gap closed, using optimal and linear thresholds.[21] Remarkably, the performance of the linear thresholds is almost indistinguishably close to the optimal ones, with the exception of three cases: Lognormal(0,1) when $\theta=2$ and $\theta=4$, and when the distribution is Exponential and $\theta=4$.

These three cases correspond to situations where supply is strongly convex and very elastic, and the distributions right skewed. As a result, an increase in thresholds leads to a large redistribution of output to certified firms and also an increase in total output. It is optimal then to certify a small fraction of firms that end up accounting for a large share of output. Note that in the lognormal case when $\theta=2$,it is optimal to certify only 0.6% of firms, while 7.4% are certified with the linear threshold. Moreover, in the linear case the solution is not interior and the market share of certified firms equals one. The same occurs when $\theta=4$, where in both the optimal and linear case only certified firms produce, and it is optimal to certify a negligible fraction of firms.

[21]The empty cells correspond to cases where firm supply is not integrable as a result of a very elastic supply and long tails of the respective distributions.

Table 2: Performance of Linear Supply Thresholds on Non-Linear Supply Functions

Distribution	$\theta = 1/4$				$\theta = 1/2$			
	% certified		% gap closed		% certified		% gap closed	
	optimal	linear	optimal	linear	optimal	linear	optimal	linear
Pareto $\alpha = 3$	15.3	4.9	57.3	55.9	8.2	4.9	55.6	54.2
Pareto $\alpha = 4$	12.5	8.7	60.9	59.6	11.3	8.7	58.8	58.3
Exponential	31.0	20.3	67.3	64.2	27.0	20.3	66.7	65.3
$F(z) = z^{1/2}$	41.1	36.0	77.6	76.6	38.9	36.0	77.6	77.3
$F(z) = z^2$	66.2	61.8	70.1	69.6	64.5	61.8	71.0	70.8
Log-normal $(0, 0.25)$	40.5	36.5	63.9	63.2	39.4	36.5	63.5	63.3
Log-normal $(0, 1)$	17.5	7.41	62.6	56.8	14.0	7.41	60.5	57.6
	$\theta = 2$				$\theta = 4$			
Pareto $\alpha = 3$	–	–	–	–	–	–	–	–
Pareto $\alpha = 4$	4.8	8.7	37.5	35.2	–	–	–	–
Exponential	9.8	20.3	62.6.9	55.9	0.4(*)	20.3(*)	70.7	23.6
$F(z) = z^{1/2}$	31.6	36.0	77.5	76.7	24.4	36.0	79.8	73.5
$F(z) = z^2$	57.3	61.8	74.0	73.4	49.7	61.8	77.5	73.5
Log-normal $(0, 0.25)$	30.8	36.5	62.2	61.6	20.0	36.5	60.5	55.5
Log-normal $(0, 1)$	0.6	7.4 (*)	44.9	27.9	$\simeq 0$(*)	7.4 (*)	45.5	0.2

(*) Market share of certified = 1

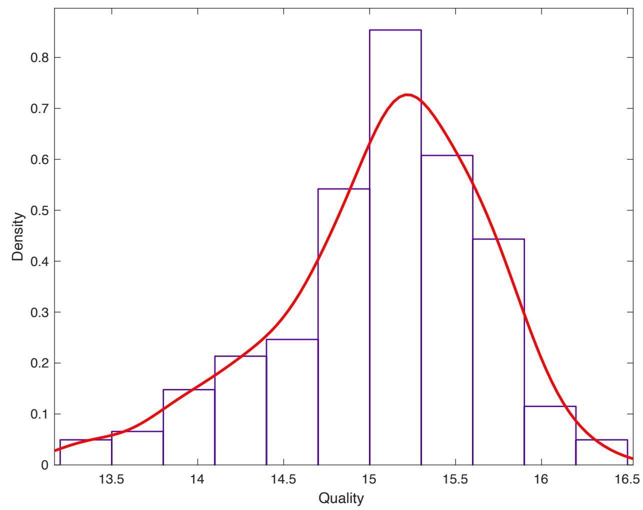
Similarly, for the exponential case with $\theta = 4$, it is optimal to certify only 0.2% of the firms, while 20.3% are certified with the linear threshold, and for both thresholds the market share of certified firms equals one.

Our results suggest that except in cases where supply is very elastic and the distribution of quality has a long right tail, linear thresholds perform very well relative to the optimal ones.

4.4 Empirical Application

Medicare Advantage Providers Here we apply our method of finding optimal thresholds to Medical Advantage providers. The data, which generously shared with us by Benjamin Vatter,

Figure 3: Medicare Advantage Quality Distribution



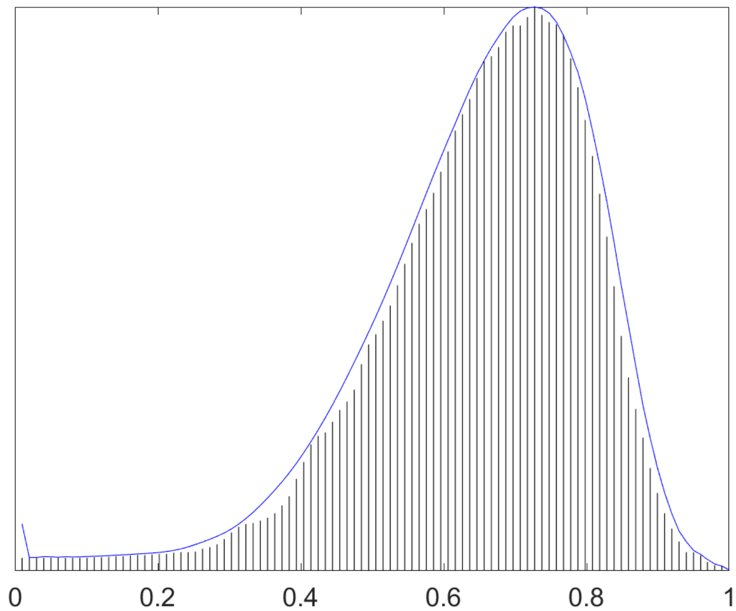
consists of estimates of the quality of products offered by insurance firms across the United States (details on the source of this data are given in Vatter (2021)). Our application is mainly meant to illustrate the workings of our method using an empirically relevant distribution, and is an overly-simplified version of how this market works. In particular, we pool all the national data and abstract from the local nature of many of these insurance markets, as well as abstracting from moral hazard considerations.[22]

Figure 3 gives the baseline discrete distribution of qualities along with a density kernel estimation, which is the one used in our calculations. The optimal two-tier rating is defined by a threshold quality equal to 14.9, which is slightly below the mean quality, implying that 66.5% of the firms are certified. This threshold achieves 64% of the gap between the full-information and no-information cases. Notice that by certifying a high fraction of firms, this rating serves mostly the purpose of screening out the lower tail of quality, which as seen in Section 3.2 is optimal when the distribution of qualities is left skewed, as in this case.

eBay Sellers' Ratings Nosko and Tadelis (2015) provides a quality measure given by the per-

[22] All these features and other important considerations relevant to the design of a rating system for this industry are carefully considered in Vatter's excellent paper.

Figure 4: Distribution of Percent Positive Responses for eBay Sellers



centage time a seller got positive feedback (as opposed to negative or none). The distribution of this statistic across sellers is given in Figure 4 together with a density kernel estimator.[23] If we interpret this statistic as an ex ante probability of a good (vs. a bad) experience and the expected utility from this purchase as $(1 - P\,(good))\,u\,(bad) + P(good)u(good)$, then expected utility is an affine transformation of the probability of a good experience. Based on this interpretation, we can use this distribution to calculate the optimal certification threshold, as we did for other distributions above. Table 1 reports the results for the kernel estimate of this distribution. According to our calculations, more than 65% of sellers should be certified, closing about 63% of the surplus gap. As in the Medicare application, the distribution of qualities is also left-skewed and certification serves the purpose of mainly screening out the lower tail.

[23]The data for the histogram comes directly from Table 4 in Nosko and Tadelis (2015).

5 Final Remarks

In this paper we considered the optimal design of quality ratings in markets with limited signals. Specifically, ratings reallocate demand across producers, which impacts not only the average quality of goods consumed but also the average cost. The optimal thresholds in a discrete rating system optimize this trade-off, and depend on the characteristics of the market, including the distribution of producers' quality and the elasticity of supply. For example, we find that the optimal thresholds in the case of a convex (resp., concave) supply function are pointwise higher (resp., lower) than those in the linear case. Intuitively, in the case of a simple certification rating with two groups, more elastic supply leads to a higher threshold and a lower share of certified sellers. Additionally, we also find that skewness in the distribution of seller qualities matters for optimal ratings, which move in the direction of the skew.

We characterized the optimal thresholds in the case of linear supply as the solution to standard clustering problems in order to provide a simple and easy-to-compute approach for the design of rating systems. This method is used to derive bounds on the performance of the rating system as a function of the number of categories. For instance, we first theoretically showed that a simple certification mechanism, or a two-tier rating, is enough to reach half of the benefits of the best rating mechanism in a large family of sellers' quality distributions such as log-concave densities. We then apply our method numerically. As an example, we found that for the exponential family of distributions, 65% of the total surplus gains from the full-information case can be achieved with only two categories. In addition to these theoretical and numerical results, we also applied our method to two empirical problems in order to further illustrate how it can be used to find the optimal thresholds and the welfare gains from them.

Our results suggest that in practice, we may not need very complicated mechanisms to attain most of the benefits from information provision. In this paper, we do not explicitly model any cost related to providing information; however, in practice, providing information may involve various costs for both the market designer and consumers. For example, getting precise information may increase the market designer's cost of designing the rating system, and it might be costly for the market designer to convey finer information to consumers and for them to process this information. These costs, as well as the potential small benefits from more complex information mechanisms, may justify the use of more simplified, or "coarse," information mechanisms seen in practice.

References

Bergemann, D. and M. Pesendorfer (2007): "Information structures in optimal auctions," *Journal of economic theory*, 137, 580–609. 10

Dranove, D. and G. Z. Jin (2010): "Quality disclosure and certification: Theory and practice," *Journal of Economic Literature*, 48, 935–963. 5

Elfenbein, D. W., R. Fisman, and B. McManus (2015): "Market structure, reputation, and the value of quality certification," *American Economic Journal: Microeconomics*, 7, 83–108. 5

Fan, Y., J. Ju, and M. Xiao (2013): "Losing to Win: Reputation Management of Online Sellers," . 5

Filippas, A., J. J. Horton, and J. Golden (2018): "Reputation inflation," in *Proceedings of the 2018 ACM Conference on Economics and Computation*, ACM, 483–484. 5

Ganuza, J.-J. and J. S. Penalva (2010): "Signal Orderings Based on Dispersion and the Supply of Private Information in Auctions," *Econometrica*, 78, 1007–1030. 7

Gentzkow, M. and E. Kamenica (2016): "A Rothschild-Stiglitz Approach to Bayesian Persuasion," *American Economic Review*, 106. 7

Hopenhayn, H. and M. Saeedi (2023): "Optimal information disclosure and market outcomes," *Theoretical Economics*, 18, 1317–1344. 4, 6, 8, 19, 31

Hoppe, H. C., B. Moldovanu, and E. Ozdenoren (2011): "Coarse matching with incomplete information," *Economic Theory*, 47, 75–104. 5, 21

Hui, X., Z. Liu, and W. Zhang (2020): "Mitigating the Cold-start Problem in Reputation Systems: Evidence from a Field Experiment," *Available at SSRN 3731169*. 5

Hui, X., M. Saeedi, Z. Shen, and N. Sundaresan (2016): "Reputation and regulations: evidence from eBay," *Management Science*, 62, 3604–3616. 5

Hui, X., M. Saeedi, G. Spagnolo, and S. Tadelis (2023): "Raising the Bar: Certification Thresholds and Market Outcomes," *American Economic Journal: Microeconomics*. 2

Hui, X., M. Saeedi, and N. Sundaresan (2018): "Adverse selection or moral hazard, an empirical study," *The Journal of Industrial Economics*, 66, 610–649. 3

Jin, G. Z. and P. Leslie (2003): "The effect of information on product quality: Evidence from restaurant hygiene grade cards," *The Quarterly Journal of Economics*, 118, 409–451. 5

MacQueen, J. et al. (1967): "Some methods for classification and analysis of multivariate observations," in *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, Oakland, CA, USA, vol. 1, 281–297. 3, 13

McAfee, R. P. (2002): "Coarse matching," *Econometrica*, 70, 2025–2034. 5, 21

Mease, D. and V. N. Nair (2006): "Unique optimal partitions of distributions and connections to hazard rates and stochastic ordering," *Statistica Sinica*, 1299–1312. 33

Moscarini, G. and M. Ottaviani (2001): "Price competition for an informed buyer," *Journal of Economic Theory*, 101, 457–493. 19

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. 5, 25, 26

Saeedi, M. (2019): "Reputation and adverse selection: theory and evidence from eBay," *The RAND Journal of Economics*, 50, 822–853. 5

Saeedi, M. and A. Shourideh (2022): "Optimal Rating Design," . 3

Shao, R. (2016): "Generalized coarse matching," *Games and Economic Behavior*, 100, 142–148. 5, 21

Shi, Z. J., K. Srinivasan, and K. Zhang (2020): "Design of Platform Reputation Systems: Optimal Information Disclosure," *Kannan and Zhang, Kaifu, Design of Platform Reputation Systems: Optimal Information Disclosure (March 14, 2020)*. 3

Stoyan, D. and D. J. Daley (1983): "Comparison Methods for Queues and Other Stochastic Models." *JOHN WILEY & SONS, INC., 605 THIRD AVE., NEW YORK, NY 10158, USA, 1983*. 21

Van Zwet, W. (1964): "Convex transformations of random variables, Math," *Centrum, Amsterdam*, 1964. 16

Vatter, B. (2021): "Quality disclosure and regulation: Scoring design in medicare advantage," Tech. rep., Working Paper. 3, 5, 25

Wilson, R. (1989): "Efficient and competitive rationing," *Econometrica: Journal of the Econometric Society*, 1–40. 5, 21

A Proofs

Proof of Lemma 1

Proof. Consider a partition of the set of sellers into sets $S_1, ..., S_N$. Suppose there are two sets S_k, S_{k+1} that are not totally ordered in quality with means $M_k \leq M_{k+1}$ and mass G_k and G_{k+1}. By reordering elements of these two sets, one can substitute S_k and S_{k+1} with two new disjoint sets S'_k and S'_{k+1} of equal measures to the original ones, where $S_k \cup S_{k+1} = S'_k \cup S'_{k+1}$ and $S'_k < S'_{k+1}$, element-wise. By construction, $M'_k \leq M_k \leq M_{k+1} \leq M'_{k+1}$ and $G'_k M'_k + G'_{k+1} M'_{k+1} = G_k M_k + G_{k+1} M_{k+1}$. This corresponds to a mean preserving spread of the original distribution of means. Using Proposition 1 in Hopenhayn and Saeedi (2023), this results in higher total surplus. □

Proof of Lemma 2

To totally differentiate equation (3) with respect to z_k, first note that by the envelope condition, we can ignore the effect on the output choices $q_1, ..., q_N$. In particular, this implies that $\partial Q / \partial z_k = f(z_k)(q_k - q_{k+1})$. Given that $M_k = \int_{z_{k-1}}^{z_k} z dF(z) / (F(z_k) - F(z_{k-1}))$, it follows that

$$\frac{\partial \left(F\left(z_k\right) - F\left(z_{k-1}\right)\right) M_k}{\partial z_k} = f\left(z_k\right) z_k, \frac{\partial \left(F\left(z_{k+1}\right) - F\left(z_k\right)\right) M_{k+1}}{\partial z_k} = -f\left(z_k\right) z_k.$$

The result now follows by totally differentiating (3) and setting it equal to zero.

Proof of Proposition 1

To prove this proposition, we need an intermediate step, which is proven using the following lemma.

Lemma 3. *The optimal thresholds satisfy the following condition:*

$$\frac{z_k - M_k}{M_{k+1} - M_k} S\left(p_k\right) + \frac{M_{k+1} - z_k}{M_{k+1} - M_k} S\left(p_{k+1}\right) = \frac{\int_{p_k}^{p_{k+1}} S\left(p\right) dp}{p_{k+1} - p_k}, \tag{14}$$

where M_k and M_{k+1} are the conditional mean qualities for the two groups, and p_k and p_{k+1} are the equilibrium prices.

Proof. First note that

$$\begin{aligned}
\left(P\left(Q\right)+z_k\right)\left(q_{k+1}-q_k\right) &= \left(P\left(Q\right)+M_{k+1}-M_{k+1}+z_k\right)q_{k+1} \\
&- \left(P\left(Q\right)+M_k-M_k+z_k\right)q_k \\
&= p_{k+1}q_{k+1}-p_kq_k-\left(M_{k+1}-z_k\right)q_{k+1}-\left(z_k-M_k\right)q_k.
\end{aligned}$$

Substituting in (5) and rearranging gives

$$\left(M_{k+1}-z_k\right)q_{k+1}+\left(z_k-M_k\right)q_k=\pi_{k+1}-\pi_k.$$

Equation (14) follows by substituting $\pi_{k+1}-\pi_k=\int_{p_k}^{p_{k+1}}S\left(p\right)dp$, using $q_{k+1}=S\left(p_{k+1}\right)$ and $q_k=S\left(p_k\right)$, and dividing the left hand side by $\left(M_{k+1}-M_k\right)$ and the right hand side by the equivalent value $p_{k+1}-p_k$. □

We use the expression found in Lemma 3. Equation (14) equates the expected value of $S\left(p\right)$ under two lotteries. The left hand side lottery has weights $\alpha=\left(z_k-M_k\right)/\left(M_{k+1}-M_k\right)$ on price p_k and $\left(1-\alpha\right)$ on price p_{k+1}. The second lottery is uniform between these two extreme prices. When S is linear, it must be the case that $\alpha=1/2$, and this implies that

$$z_k-M_k=M_{k+1}-z_k. \tag{15}$$

When S is convex, $\alpha>1/2$ so $z_k-M_k>M_{k+1}-z_k$, so the optimal threshold is above the one defined by equation (15), although the reverse occurs when s is concave. This concludes the proof.

Proof of Proposition 2

Without loss of generality, let $S(p) = p$, so the cost function $c(q) = \frac{1}{2}q^2$. Consider now the objective function (3) for this case:

$$W(\boldsymbol{z}) = \int_0^Q P(x)\,dx + \sum_{k=1}^{N} [F(z_k) - F(z_{k-1})] \left[M_k (P + M_k) - \frac{1}{2}\left((P + M_k)^2\right) \right] \quad (16)$$

$$= \int_0^Q P(x)\,dx + \sum_{k=1}^{N} [F(z_k) - F(z_{k-1})] \left[\frac{1}{2} M_k^2 - \frac{1}{2} P^2 \right]. \quad (17)$$

After suppressing the terms that are unaffected by the partition, maximizing this expression is equivalent to maximizing

$$\sum_{k=1}^{N} [F(z_k) - F(z_{k-1})] (M_k - \bar{z})^2, \quad (18)$$

where $\bar{z} = \sum_{k=1}^{N} [F(z_k) - F(z_{k-1})] M_k$ is the mean seller quality, which is independent of the partition. The above expression is the variance between partitions. Because total variance is fixed, maximizing (18) is equivalent to minimizing (6). Uniqueness of the thresholds is guaranteed when the distribution has log-concave density, as shown in Mease and Nair (2006).

Proof of Proposition 3

We use the following properties of distributions with log-concave densities (see Lemma 1 in Mease and Nair (2006)):

$$\mathbb{E}(z|s \leq z \leq s + d) - s \text{ is decreasing in } s \text{ for } d > 0 \text{ and} \quad (19)$$

$$s - \mathbb{E}(z|s - d \leq z \leq s) \text{ is increasing in } s \text{ for } d > 0, \quad (20)$$

and these properties are preserved when conditioning on intervals.

Lemma 4. *Suppose F is a distribution with log-concave density, and let $m(a, b) = E_F(z|a \leq z \leq b)$.*

Suppose the vector of thresholds $\{l_k\}_{k=1}^{N-1}$ *satisfies*

$$l_k - m\left(l_{k-1}, l_k\right) = m\left(l_k, l_{k+1}\right) - l_k, \tag{21}$$

and let $z_1, ..., z_{N-1}$ *be a vector such that*

$$z_k - m\left(z_{k-1}, z_k\right) > m\left(z_k, z_{k+1}\right) - z_k. \tag{22}$$

Then $z_k > l_k$ *for all* k.

To prove Lemma 4, we first use the following:

Claim. Under the assumptions of Lemma 4, suppose that for some k, $z_k < l_k$ and $z_{k+1} - z_k \geq l_{k+1} - l_k$. Then $z_{k-1} < l_{k-1}$ and $z_k - z_{k-1} \geq l_k - l_{k-1}$.

Proof. Note that

$$\begin{aligned} z_k - m\left(z_{k-1}, z_k\right) &> m\left(z_k, z_{k+1}\right) - z_k \\ &\geq m\left(z_k, z_k + l_{k+1} - l_k\right) - z_k \\ &\geq m\left(l_k, l_{k+1}\right) - l_k \\ &= l_k - m\left(l_{k-1}, l_k\right). \end{aligned} \tag{23}$$

The first inequality follows from (22), the second one, from the monotonicity of m, the third, from (19), and the last, from (21). Now consider $k-1$. We will show that $z_k - z_{k-1} \geq l_k - l_{k-1}$.

Suppose, by way of contradiction, that $z_k - z_{k-1} < l_k - l_{k-1}$. Then

$$\begin{aligned} z_k - m\left(z_{k-1}, z_k\right) &\leq l_k - m\left(l_k - \left(z_k - z_{k-1}\right), l_k\right) \\ &\leq l_k - m\left(l_{k-1}, l_k\right) \end{aligned}$$

where the first inequality follows from condition (20), and the second one, from the monotonicity of m. This inequality contradicts (23), proving that $z_k - z_{k-1} \geq l_k - l_{k-1}$. Given that $z_k < l_k$, this also guarantees that $z_{k-1} < l_{k-1}$. □

We now prove Lemma 4. Let h denote the highest k for which $z_k < l_k$. By the definition of h, $z_{h+1} - z_h > l_{h+1} - l_h$. Using inductively the previous claim, it follows that the same is true for all $k = 1, ..., h$. For $k = 1$, the claim would imply that $z_0 < l_0$, which cannot be true if the distribution had a lower bound, because in that case both z_0 and l_0 should equal this lower bound. For unbounded support, an argument similar to the one used in the claim can be used to generate a contradiction. This completes the proof.

Proof of Proposition 3.

Let $\{l_k\}$ denote the optimal thresholds for the linear supply function, and $\{z_k\}$, those for the convex supply function. Lemma 1 and equations (21) and (23) hold, so Lemma 4 proves the proposition.

Proof of Proposition 4

To prove this proposition, we first need to show the following lemma.

Lemma 5. *Let $g(z_1), ..., g(z_{N-1})$ be the optimal thresholds for $\tilde{F}$. Let $M_k = m(z_{k-1}, z_k) = E_F(z_{k-1} \leq z \leq z_k)$. Then $z_k - M_k > M_{k+1} - z_k$.*

Proof. Let $\tilde{M}_k = E_{\tilde{F}}(g(\tilde{z}_{k-1}) \leq z \leq g(\tilde{z}_k))$. Note that by strict convexity of g, $\tilde{M}_k > g(M_k)$. It follows that

$$
\begin{aligned}
z - M_k &> z_k - g^{-1}\left(\tilde{M}_k\right) \\
&= g^{-1}(g(z_k)) - g^{-1}\left(\tilde{M}_k\right) \\
&= g^{-1}\left(\tilde{M}_{k+1}\right) - g^{-1}(g(z_k)) \\
&> M_{k+1} - z_k.
\end{aligned}
$$

□

To prove the proposition, let the vector $\{l_k\}$ be the optimal thresholds for F, and $\{z_k\}$, the optimal thresholds for $\tilde{F}$. Equation (21) follows from the necessary condition for optimal thresholds, and (23) follows from the previous lemma.

Proof of Proposition 5.

Proof. Let $\bar{M}_1$ and $\bar{M}_2$ be the conditional mean of z below and above the mean $\bar{z}$, respectively. By the variance decomposition,

$$\begin{aligned}\int (z-\bar{z})^2 dF(z) &= \int^{\bar{z}} \left(z-\bar{M}_1\right)^2 dF(z) + \int_{\bar{z}} \left(z-\bar{M}_2\right)^2 dF(z) \\ &\quad + F(\bar{z})\left(\bar{M}_1-\bar{z}\right)^2 + (1-F(\bar{z}))\left(\bar{M}_2-\bar{z}\right)^2 \\ &= F(\bar{z})\left(cv_1^2+1\right)\left(\bar{M}_1-\bar{z}\right)^2 + (1-F(\bar{z}))\left(cv_2^2+1\right)\left(\bar{M}_2-\bar{z}\right)^2 \\ &\leq \left(\max\left\{cv_1^2, cv_2^2\right\}+1\right)\left(F(\bar{z})\left(\bar{M}_1-\bar{z}\right)^2 + (1-F(\bar{z}))\left(\bar{M}_2-\bar{z}\right)^2\right),\end{aligned}$$

where the second equality follows from

$$cv_1 = \frac{\int^{\bar{z}} \left((z-\bar{z}) - \left(\bar{M}_1-\bar{z}\right)\right)^2 dF(z)}{F(\bar{z})\left(\bar{M}_1-\bar{z}\right)^2}$$

,and similarly for cv_2. From the above inequality,

$$\frac{F(\bar{z})\left(\bar{M}_1-\bar{z}\right)^2 + (1-F(\bar{z}))\left(\bar{M}_2-\bar{z}\right)}{\int (z-\bar{z})^2 dF(z)} \geq \frac{1}{1+\max\left\{cv_1^2, cv_2^2\right\}}.$$

This gain corresponds to setting $z^* = \bar{z}$, so it is a lower bound to the gains under the optimal threshold. □