Information Design for Many Parameters

Thành Nguyen* Rakesh Vohra†

January 2025

Abstract

A stochastic mapping from states to signals induces a distribution over possible posteriors. For various purposes, one is interested in parameters of these posteriors, possibly multi-dimensional, such as their quartiles or whether they first-order stochastically dominate some given distribution. Many parameters of interest can be encoded by requiring the posterior at each signal lie point-wise between two given functions. Given a collection of these bounding pairs, one for each signal, we characterize the set of posteriors that can be realized by some stochastic mapping of states to signals. We provide applications of this result to, among others, the design of securities, price and statistical discrimination.

Keywords: Bayesian Persuasion, Information Design, Farkas Lemma

1 Introduction

The problem of how an agent with superior information (called the Sender) can influence another (called the Receiver) by controlling what information is revealed is sometimes formulated in terms of the choice of a mapping, possibly stochastic, from states to signals (called a signal structure). The mapping is chosen *before* learning the underlying state. See Kamenica and Gentzkow (2011) and Bergemann and Morris (2016)) for example.

Once the state is realized, a signal is generated according to the chosen mapping and passed to the Receiver. Each signal induces a posterior that affects the Receiver's action choices and, by extension, the Sender's payoffs. Therefore, properties of the posterior distribution over states induced by a given signal structure will determine the Receiver's actions. For example, suppose the Receiver wishes to choose an action that matches a one-dimensional state, and the square of the distance between the action and the realized state measures their loss. Then, for any realized posterior, the Receiver's optimal choice is the mean of that posterior. Hence, in this case, choosing a signal structure is equivalent to choosing a distribution of posterior means.

*Daniels School of Business, Purdue University, 403 W. State Street, West Lafayette, Indiana, 47906, United States. E-mail: nguye161@purdue.edu.

†Department of Economics & Department of Electrical and Systems Engineering University of Pennsylvania. Email: rvohra@seas.upenn.edu.

For tractability, the literature focuses on cases where only a single parameter of the posterior, such as the mean, median, or quantile, matters. This is restrictive. For example, a policymaker evaluating economic growth forecasts might consider not only the median growth rate but also the inter-quartile range to assess the variability and associated risks.[1] Similarly, a financial analyst comparing investment options may prioritize a distribution that stochastically dominates a benchmark, ensuring that the returns of a new portfolio outperform those of a market index under all plausible scenarios.

Many parameters of interest can be represented by requiring that the realized posterior for each signal lies pointwise between a given pair of bounding functions. For example, to ensure a median of r, the cumulative probability of the posterior below r must be less than $\frac{1}{2}$, while at r, it must be at least $\frac{1}{2}$. For convenience, we call parameters that impose such upper and lower bounds on the cumulative distribution "nice."

We show that a distribution over nice parameters is implementable by some signal structure *if and only if*: (i) the prior's cumulative distribution lies between the posterior averages of the upper and lower bounding functions, and (ii) the total prior probability over any interval is at least the posterior average of the positive difference between these bounds.

Our characterization avoids the need to track the state–signal mapping, and explicitly links the prior and posterior distributions. This allows for straightforward applications to parameters such as the median, quartiles, and conditions under which the posterior stochastically dominates a given distribution. It also supports other applications, such as analyzing statistical discrimination or determining the range of prices and quantities that can arise under third-degree price discrimination.

Following Chambers and Echenique (2020) and Bharadwaj et al. (2024), one can model statistical discrimination by allowing the state space to include group identities representing different demographics or identity-related attributes. We show how our method can be used to rationalize observed disparities in outcomes by differences in the underlying prior distributions, the mappings from the state space to the signal space, or the decision rules used to determine outcomes.

In college admissions, for example, students possess both an academic score and an identity group, such as race. The stochastic mapping relates the prior distribution of student academic profiles to college admissions decisions; however, this mapping is not directly observed. Instead, the available data consist of the number of applicants and acceptance rates at each college for each identity group. Our characterization provides an *if and only if* condition that determines when such observed data on priors and aggregate admissions are compatible with an admission rule that ignores identity group. If the data are compatible, then disparities in outcomes could be attributed to differences in the underlying prior distributions or the mappings from the state space to the signal space. Conversely, if the data are incompatible, we conclude that the decision rules are correlated with the identity group, indicating taste-based discrimination. Our result, therefore, provides a novel test to detect such discrimination in college admission.

[1] This is especially relevant when the Receiver's action space is multidimensional, and their loss from mismatching the state is determined by an average of different 'pinball' loss functions. Such scenarios arise in forecasting competitions and conditional value-at-risk requirements (see Grushka-Cockayne et al. (2017)).

Another application is in the study of third-degree price discrimination, where unit-demand consumers are divided into segments, each offered a monopoly price tailored to that segment. Each segment corresponds to a price-quantity pair, and we characterize which distributions over such pairs can be realized by some segmentation. This contrasts with Bergemann et al. (2015), who characterize the *aggregate* revenue and consumer surplus distributions realizable from some segmentation.

Our result provides a *reduced-form* characterization linking the prior distribution of buyer types to the distribution of segment-specific prices and their associated sales volumes. This reduced form has two main uses. First, when only market segmentation data is observed but detailed demand information is unavailable, it allows inference on the underlying demand distribution. Second, given a known prior demand distribution, it characterizes segmentations that achieve particular goals, such as maximizing total surplus or consumer surplus.

A distinctive feature of our approach is the graphical interpretation of a segmentation called the *inverse consumption curve*. It is an inverse demand curve where each segment is treated as a single consumer with demand either zero or a fixed quantity, depending on the price. This interpretation simplifies both the design and analysis of market segmentations.

Besides the aggregate perspective emphasized by Bergemann et al. (2015), our approach captures the distribution of consumer surplus across segments, enabling identification of conditions under which efficient segmentations exist that limit surplus to given thresholds for all consumers. Furthermore, the framework naturally extends to settings with multidimensional but ordered consumer types.

Our work differs from the many papers inspired by Bergemann et al. (2015), in that we are not extending it to other settings such as auctions (e.g. Bergemann et al. (2019)), or competition (e.g. Elliott (2021), Farboodi et al. (2025)) or second degree price discrimination (e.g. Haghpanah and Siegel (2022)). Instead, within their setting, we seek to provide a more detailed picture of what can happen.

From a methodological perspective, this paper is most closely related to the work focused on characterizing properties of posterior distributions that can be induced by some signal structure. Examples of this are Benoît and Dubra (2011), Gentzkow and Kamenica (2016), Kolotilin and Wolitzky (2024), Vohra et al. (2023) and Yang and Zentefis (2022). However, as stated earlier, unlike this paper, their focus has been on one-dimensional parameters of the posterior, like mean, median, and quartile. This paper's main result widens the range of settings in which the information design framework can be deployed. We relegate more detailed discussion of relevant prior work to the appropriate sections.

2 Main Theorem

In this section, we state our main result. Let $\mathcal{N} = \{1, .., n\}$ be the set of states, $\pi_i \geq 0$ the prior probability of state $i \in \mathcal{N}$. Denote the finite set of signals by $\mathcal{S}$. Associated with each signal $s \in \mathcal{S}$ are two 'bounding' non negative functions, $F_s(\cdot)$ and $G_s(\cdot)$ over $\{1, .., n\}$.

A stochastic mapping from states to signals can be represented for each (i, s) pair by the variable $x_{is} \geq 0$, which is the joint probability of the state i and signal s. The probability that signal s is realized will be $\sum_{i\in\mathcal{N}} x_{is}$ and the probability that the state is k given signal s is $\frac{x_{ks}}{\sum_{i\in\mathcal{N}} x_{is}}$. We call a probability distribution $\{\rho_s\}_{s\in\mathcal{S}}$ over $\mathcal{S}$ *implementable* if there exists a stochastic mapping from states to signals such that:

1. Each state i is realized with probability π_i, that is, for all $i \in \mathcal{N}$: $\sum_{i\in\mathcal{S}} x_{is} = \pi_i$,
2. Each signal s is realized with probability ρ_s, that is, for all $s \in \mathcal{S}$: $\sum_{i\in\mathcal{N}} x_{is} = \rho_s$,
3. The posterior distribution conditional on signal s lies point-wise between the bounding functions $F_s(\cdot)$ and $G_s(\cdot)$. That is, for all $s \in \mathcal{S}, t \in \mathcal{N}$: $G_s(t) \cdot \rho_s \leq \sum_{i=1}^{t} x_{is} \leq F_s(t) \cdot \rho_s$.

THEOREM 1 *Given two non decreasing functions $F_s, G_s : \mathcal{N} \to \mathbb{R}$ with $0 \leq G_s(t) \leq F_s(t) \leq 1$ for all $s \in \mathcal{S}, t \in \mathcal{N}$ and $F_s(n) = 1$, the distribution $\{\rho_s\}_{s\in\mathcal{S}}$ over $\mathcal{S}$ is implementable if and only if :*

$$\sum_{s\in\mathcal{S}} G_s(t)\rho_s \leq \sum_{i=1}^{t} \pi_i \leq \sum_{s\in\mathcal{S}} F_s(t)\rho_s \ \forall t \in \mathcal{N} \tag{1}$$

$$\sum_{i=t+1}^{t'} \pi_i \geq \sum_{s\in\mathcal{S}} \max\{G_s(t') - F_s(t), 0\}\rho_s \ \forall t', t \in \mathcal{N} \quad s.t.\ t' > t. \tag{2}$$

Condition (1) requires that the prior cumulative probability from states 1 to t lies between the posterior averages of the lower- and upper-bound functions. Condition (2), in turn, requires that the prior probability of states $t+1$ through t' be at least the posterior average of the positive difference between the lower bound at t' and the upper bound at t.

A special case of Theorem 1, where the bounding functions are identical, i.e., $F_s(t) = G_s(t)$, implies the martingale condition, also known as the Splitting Lemma (see Blackwell (1953); Aumann et al. (1995)). In this case, condition (2) is trivially satisfied and thus, a posterior distribution $F_s(.)$ is implementable if and only if the prior distribution equals the average over the posterior distributions:

$$\sum_{i=1}^{t} \pi_i = \sum_{s\in\mathcal{S}} F_s(t)\rho_s.$$

There is a natural flow-based interpretation of Theorem 1. Think of each state $i \in \mathcal{N}$ as a supply node with supply π_i, and each signal $s \in \mathcal{S}$ as a demand node. However, unlike standard flow problems where demand nodes require a fixed quantity and are indifferent to the source, here each demand node s requires between $G_s(t)\rho_s$ and $F_s(t)\rho_s$ units from the subset of supply nodes $\{1, \ldots, t\}$, for each $t = 1, \ldots, n$.

Under this view, constraint (1) ensures that total supply is fully utilized: for each t, the cumulative supply from nodes $\{1, \ldots, t\}$ must lie within the total demand bounds—no less than the aggregate lower demand $\sum_s G_s(t)\rho_s$, and no more than the upper bound $\sum_s F_s(t)\rho_s$.

Constraint (2) is more subtle. Fix any interval of supply nodes $\{t+1,\dots,t'\}$ and a demand node s. The minimum amount that must be routed from these supply nodes to s is determined by the gap between what s must receive from all of $\{1,\dots,t'\}$ and what it could receive from just $\{1,\dots,t\}$. Specifically, s must receive at least $G_s(t')\rho_s$ units from $\{1,\dots,t'\}$, but can receive at most $F_s(t)\rho_s$ from $\{1,\dots,t\}$. Thus, if $(G_s(t') - F_s(t)) \cdot \rho_s > 0$, the shortfall must be filled by supply from $\{t+1,\dots,t'\}$.

The core idea of the proof of Theorem 1 is to apply Farkas' lemma to the system of upper- and lower-bound constraints. While this typically yields infinitely many conditions on the posterior distributions, in our setting tractability arises from exploiting the laminarity of constraints (4).[2]

Proof of Theorem 1. The "only if" direction is straightforward.

First,

$$\sum_{i=1}^{t} \pi_i = \sum_{i=1}^{t} \sum_{s\in\mathcal{S}} x_{is} = \sum_{s\in\mathcal{S}} \sum_{i=1}^{t} x_{is}.$$

Since $G_s(t)\rho_s \le \sum_{i=1}^{t} x_{is} \le F_s(t)\rho_s$ for all $s \in \mathcal{S}$, condition (1) follows.

Next,

$$\sum_{i=t+1}^{t'} \pi_i = \sum_{s\in\mathcal{S}} \sum_{i=t+1}^{t'} x_{is} = \sum_{s\in\mathcal{S}} \left(\sum_{i=1}^{t'} x_{is} - \sum_{i=1}^{t} x_{is} \right).$$

Using the same bounds, we get

$$\sum_{i=t+1}^{t'} x_{is} \ge \max\{0, G_s(t') - F_s(t)\}\rho_s \quad \text{for all } s \in \mathcal{S}, \qquad \text{which yields (2).}$$

To see the "if" direction, observe that to determine whether a distribution $\{\rho_s\}_{s\in\mathcal{S}}$ over $\mathcal{S}$ is implementable, it suffices to verify the feasibility of the following system:

$$\sum_{s\in\mathcal{S}} x_{is} \ge \pi_i \qquad \forall i \in \mathcal{N} \tag{3}$$

$$G_s(t)\rho_s \;\le\; \sum_{i=1}^{t} x_{is} \le F_s(t)\rho_s \qquad \forall s \in \mathcal{S},\ \forall t \in \{1,\dots,n\} \tag{4}$$

$$x_{is} \ge 0 \qquad \forall i \in \mathcal{N},\ \forall s \in \mathcal{S} \tag{5}$$

Note that we only require (3) to hold as an inequality rather than an equality, and we omit the constraints $\sum_{i=1}^{n} x_{is} = \rho_s$ for all $s \in \mathcal{S}$. This is because $\sum_{i\in\mathcal{N}} \pi_i = \sum_{s\in\mathcal{S}} \rho_s = 1$ and $F_s(n) = 1$

[2] An earlier version of this paper employed a generalization of (4) based on extended polymatroids (see Frank, 2011). In practice, however, (4) suffices for all applications considered here.

together imply these conditions implicitly. In particular, if a solution to (3–5) exists, then

$$1 = \sum_{i=1}^{n} \pi_i \leq \sum_{i=1}^{n} \sum_{s \in \mathcal{S}} x_{is} \leq \sum_{s \in \mathcal{S}} F_s(n)\rho_s = 1.$$

Thus, we must have $\sum_{s \in \mathcal{S}} x_{is} = \pi_i$ for all $i \in \mathcal{N}$, and $\sum_{i \in \mathcal{N}} x_{is} = \rho_s$ for all $s \in \mathcal{S}$.

Therefore, the feasibility of (3-5) is equivalent to the feasibility of the following system:

$$\begin{aligned}
&-\sum_{s \in \mathcal{S}} x_{is} \leq -\pi_i && \forall i \in \mathcal{N} \quad (u_i) \\
&\sum_{i=1}^{t} x_{is} \leq F_s(t)\rho_s && \forall s \in \mathcal{S},\ \forall t \in \{1, \ldots, n\} \quad (\lambda(t,s)) \\
&\sum_{i=t+1}^{n} x_{is} \leq (1 - G_s(t))\rho_s && \forall s \in \mathcal{S},\ \forall t \in \{1, \ldots, n\} \quad (\mu(t,s))
\end{aligned}$$

Notice that the constraint matrix of (3)-(5) corresponds to a laminar family and is therefore totally unimodular (see Schrijver (2003)). We will use this property, along with Farkas' Lemma, to derive a condition for the system's feasibility.

Assume for contradiction that $\{\rho_s\}_{s \in \mathcal{S}}$ is not implementable, that is, the system above is not feasible. By Farkas' Lemma, this system is infeasible if and only if its Farkas alternative is feasible. The variables shown in parentheses next to each constraint in the system correspond to the dual variables in the Farkas alternative, which is presented next:

$$-\sum_{i} u_i \pi_i + \sum_{s,t} \lambda(t,s)\rho_s F_s(t) + \sum_{s,t} \mu(t,s)\rho_s(1 - G_s(t)) < 0 \tag{6}$$

$$u_i \leq \sum_{t:t \geq i} \lambda(t,s) + \sum_{t:t<i} \mu(t,s)\ \forall i \in \mathcal{N}, s \in \mathcal{S} \tag{7}$$

$$u_i, \lambda(t,s), \mu(t,s) \geq 0\ \forall i \in \mathcal{N},\ s \in \mathcal{S}\ t \in \{1, \ldots, n\} \tag{8}$$

As the Farkas alternative is feasible, by scaling, we can focus on feasible solutions to (6-8) whose components have values between 0 and 1. This means the following linear program is feasible and has an objective function value strictly less than zero.

$$\begin{aligned}
\min \quad & -\sum_{i \in \mathcal{N}} u_i \pi_i + \sum_{s \in \mathcal{S}} \sum_{t=1}^{n} \lambda(t,s)\rho_s F_s(t) + \sum_{s \in \mathcal{S}} \sum_{t=1}^{n} \mu(t,s)\rho_s\big(1 - G_s(t)\big) && \\
\text{s.t.} \quad & u_i \leq \sum_{t \geq i} \lambda(t,s) + \sum_{t<i} \mu(t,s) && \forall i \in \mathcal{N},\ \forall s \in \mathcal{S} \\
& u_i,\ \lambda(t,s),\ \mu(t,s) \in [0,1] && \forall i, t \in \mathcal{N},\ \forall s \in \mathcal{S},
\end{aligned}$$

Denote by (u^*, λ^*, μ^*) an optimal extreme point solution to the linear program. Given the total unimodularity of the constraint matrix, all extreme point solutions take values in $\{0, 1\}$. Let

$A = \{i : u_i^* = 1\}$. Note that there might be multiple such solutions, and among them, we choose the solution where the size of A is largest.

Then, (λ^*, μ^*) must solve:

$$\begin{aligned} \phi(A) := \; & -\sum_{i \in A} \pi_i + \min_{\lambda,\, \mu} \left\{ \sum_{s,t} \lambda(t,s)\, \rho_s F_s(t) + \sum_{s,t} \mu(t,s)\, \rho_s \big(1 - G_s(t)\big) \right\} \\ \text{s.t.} \quad & \sum_{t \geq i} \lambda(t,s) + \sum_{t < i} \mu(t,s) \geq 1 \quad \forall i \in A,\ \forall s \in \mathcal{S} \\ & \lambda(t,s), \mu(t,s) \geq 0 \end{aligned} \tag{9}$$

First, we show that $A \neq \emptyset$. If not, there must exist $\lambda^*, \mu^* \geq 0$ such that $\sum_{s,t} \lambda^*(t,s)\, \rho_s F_s(t) + \sum_{s,t} \mu^*(t,s)\, \rho_s (1 - G_s(t)) < 0$. This is impossible since $G_s(t) \leq 1; F_s(t) \geq 0$ for all s, t.

Next, we show that it suffices to restrict attention to sets of the form $A = \{1, \ldots, l\}$, $\{l' + 1, \ldots, n\}$, or $\{1, \ldots, l\} \cup \{l' + 1, \ldots, n\}$ with $l < l'$; in other words, sets whose complement is an interval. If A is not of one of the specified forms, there exist $r' < r < r''$ such that $r \in A$, but $r', r'' \notin A$. Since $r \in A$ and (9) holds, either $\lambda^*(t,s) = 1$ for some $t \geq r$, or $\mu^*(t,s) = 1$ for some $t < r$. In the former case, consider $A' := A \cup \{r'\}$; in the latter, $A' := A \cup \{r''\}$. In both cases, (9) still holds and $\phi(A') \leq \phi(A)$, contradicting the optimality of A, which was chosen to have the largest size among optimal solutions.

If $A = \{1, \ldots, l\}$, then the dual takes the form $\lambda^*(l,s) = 1$ with all other components zero. In this case,

$$\phi(A) = -\sum_{i=1}^{l} \pi_i + \sum_{s \in \mathcal{S}} F_s(l) \rho_s \geq 0 \quad \text{by (1)}.$$

If $A = \{l' + 1, \ldots, n\}$, then the dual is $\mu^*(l', s) = 1$, yielding

$$\phi(A) = -\sum_{i=l'+1}^{n} \pi_i + \sum_{s \in \mathcal{S}} (1 - G_s(l')) \rho_s = \sum_{i=1}^{l'} \pi_i - \sum_{s \in \mathcal{S}} G_s(l') \rho_s \geq 0 \quad \text{by (1)}.$$

If $A = \{1, \ldots, l\} \cup \{l' + 1, \ldots, n\}$ with $l < l'$, then the optimal dual solution to the Farkas alternative must be one of the following, with all other components equal to zero: $\lambda^*(l,s) = \mu^*(l',s) = 1$, $\lambda^*(n,s) = 1$, or $\lambda^*(k,s) = \mu^*(k,s) = 1$ for some $k \in \{1, \ldots, n\}$. Thus, we have

$$\begin{aligned} \phi(A) &= -\sum_{i \in A} \pi_i + \sum_{s \in \mathcal{S}} \min\{F_s(l) + 1 - G_s(l'),\, 1,\, F_s(k) + 1 - G_s(k)\} \rho_s \\ &= -\sum_{i \in A} \pi_i + \sum_{s \in \mathcal{S}} \min\{F_s(l) + 1 - G_s(l'),\, 1\} \rho_s \\ &= \sum_{i=l+1}^{l'} \pi_i - \sum_{s \in \mathcal{S}} \max\{G_s(l') - F_s(l),\, 0\} \rho_s \\ &\geq 0 \quad \text{by (2)}. \end{aligned}$$

Here, the second equality holds because $F_s(k) \geq G_s(k)$. We conclude that $\phi(A) \geq 0$, contradicting (6). Therefore, the system (3)–(5) is feasible, and $\{\rho_s\}_{s\in\mathcal{S}}$ is implementable. ■

While Theorem 1 is stated—consistent with the existing literature—as characterizing the set of implementable signal distributions given a prior, it also admits a natural reverse interpretation: for any observed signal distribution, it characterizes the set of priors that rationalize it.

Theorem 1 does not, however, identify the specific mapping $\{x_{is}\}_{i\in\mathcal{N},s\in\mathcal{S}}$ from states to signals that implements a given distribution. Recovering such a mapping requires solving a different linear program, one subject to constraints (3)–(5).

3 Quantiles, Quartiles and Stochastic Dominance

As a warm up we begin by showing how Theorem 1 can be used to characterize which distribution over posterior quantiles can be implemented. This case was already covered in Benoît and Dubra (2011), Kolotilin and Wolitzky (2024) and Yang and Zentefis (2022). Subsequently, we cover the case of quartiles and stochastic dominance.

Quantiles

Each realized signal induces a posterior distribution over the states. We are interested in the posterior's q^{th} quantile. Different realized signals induce different posteriors, generating different q^{th} quantiles. Fix a distribution over q^{th} quantiles. Is there a mapping of states to signals that induces this distribution in q^{th} quantiles? More generally, can one characterize the set of distributions of posterior quantiles that can be generated by some mapping of states to signals? We call such distributions of q^{th} quantiles **implementable**.

Benoît and Dubra (2011) characterizes the implementable distributions for the median under a uniform prior and finite state and signal space. Kolotilin and Wolitzky (2024) and Yang and Zentefis (2022) characterize it in the continuum case, for *any* prior distribution and *any* quantile. The proofs in Benoît and Dubra (2011) and Yang and Zentefis (2022) are lengthy. Kolotilin and Wolitzky (2020) provides a direct proof by showing that all implementable distributions of the posterior q^{th} quantiles are implemented by a single mapping from states to signals. The characterization for the finite case is a straightforward consequence of Theorem 1.

If $H(\cdot)$ is a distribution over $\{1,\ldots,n\}$, then, the q-quantile of H is the smallest $r \in \{1,\ldots,n\}$ such that $H(r) \geq q$. For each $i \in \{1,\ldots,n\}$ and a fixed $q \in (0,1)$, we want to know if we can achieve the following: the probability that r is the q-quantile of the realized posterior is ρ_r (note $\rho_n = 0$). If so, then ρ is implementable.

A necessary condition for a distribution $H(\cdot)$ to have a q-quantile of r is the existence of an $\epsilon > 0$ such that $H(r) \geq q$ and $H(r-1) \leq q - \epsilon$. We use this observation to characterize which distributions $\{\rho_r\}_{r=1}^n$ over q-quantiles are implementable.

To apply Theorem 1 we choose $\mathcal{S}$, the set of signals, to be $\{1, \ldots, n\}$. Hence, each signal is associated with a particular state $r \in \mathcal{N}$. We wish to construct a state-to-signal mapping so that signal r is precisely the q^{th}-quantile of the posterior distribution at that signal.

The constraint for r to be the q^{th}-quantile is captured by suitable chosen bounding functions $(F_r(\cdot), G_r(\cdot))$ as follows:

$$G_r(t) = \begin{cases} 0 & \text{if } t < r, \\ q & \text{if } t \geq r, \end{cases} \qquad F_r(t) = \begin{cases} q - \epsilon & \text{if } t \leq r-1, \\ 1 & \text{if } t > r-1. \end{cases}$$

Inequality (1) of Theorem 1 requires that an implementable distribution $\{\rho_r\}_{r=1}^n$ satisfy

$$\sum_{i=1}^{t} \pi_i \geq \sum_{r} G_r(t)\rho_r = \sum_{i=1}^{t} q\rho_i$$

and

$$\sum_{i=1}^{t} \pi_i \leq \sum_{r=1}^{n} F_r(t)\rho_r = \sum_{i=1}^{t} \rho_i + (q-\epsilon) \sum_{i=t+1}^{n} \rho_i.$$

The same theorem requires (2) to hold as well, but we show that under a full support assumption on the prior, it is redundant. Observe that for $t' > t$: $\max\{G_r(t') - F_r(t), 0\} = \epsilon$ only if $t+1 \leq r \leq t'$. Hence, inequality (2) requires

$$\sum_{i=t+1}^{t'} \pi_i \geq \epsilon \sum_{r=t+1}^{t'} \rho_r.$$

This means $\pi_{t'} \geq \epsilon\rho_{t'}$. If we assume the distribution π is full support, that is $\pi_i > 0$ for all $i \in \mathcal{N}$, then for any distribution ρ, we can always choose ϵ small enough so that these constraints hold.

We obtain the following result, which is a discrete analog of Kolotilin and Wolitzky (2024).

COROLLARY 3.1 *Assume the prior π is full support, a distribution $\{\rho_r\}_{r=1}^n$ over q quantiles is implementable if and only if there exist $\epsilon > 0$ such that*

$$\max\{\frac{\sum_{i=1}^{k} \pi_i - q + \epsilon}{1 - q + \epsilon}, 0\} \leq \sum_{r=1}^{k} \rho_r \leq \min\{1, \frac{\sum_{i=1}^{k} \pi_i}{q}\}, \ \forall k \leq n.$$

The constraints that characterize implementable $\{\rho_r\}_{r=1}^n$ are triangular. It is well known that the problem of maximizing an additively separable function of the ρ_r's over such constraints can be solved using a greedy algorithm (see Vidal et al. (2019)).

Quartiles

We consider an extension of the problem of determining which distributions over quantiles are implementable. The Receiver upon realization of the signal chooses a *multi-dimensional* action $(a_1, a_2, \ldots, a_d)$ where $d \leq n$, each $a_r \in \{1, \ldots, n\}$ and $a_1 < a_2 < \ldots < a_d$. If $\phi(\cdot)$ is the realized

posterior, and y the underlying state, the Receiver's payoff from choosing $(a_1, a_2, \ldots, a_d)$ is

$$\sum_{j=1}^{d} \{s(\phi(a_j), i) - \phi(a_j)[i - a_j]^+ - (1 - \phi(a_j))[a_j - i]^+\}. \tag{10}$$

Here $s(\cdot, \cdot)$ is any real-valued function. When $d = 1$, the action that maximizes the expected payoff given $\phi(\cdot)$ will be the median of $\phi(\cdot)$. When $d > 1$, the action that maximizes the expected payoff given $\phi(\cdot)$ is a d-vector of quantiles (see Grushka-Cockayne et al. (2017)).

Candogan and Gurkan (2025) describes an inventory setting where such a payoff function arises. Grushka-Cockayne et al. (2017) provides similar examples as well as other examples. For instance, the Receiver may be a reinsurer offering a stop-loss contract or an option trader selling a European put option. We describe the second of these in greater detail.

Suppose that the put option is on a stock whose uncertain price at expiration, x, is governed by a probability distribution function $\phi(\cdot)$. If c is the option's cost to the buyer and the trader sells it for a strike price of K, the trader's expected profit is $\int_0^K (x - K) d\phi(x) + c$. Let $q = \phi(x)$. Then, the trader's expected payoff can be written as $\int_0^{\phi(K)} \phi^{-1}(q) dq - K\phi(K) + c$. Now, we can approximate $\int_0^{\phi(K)} \phi^{-1}(q) dq$ by subdividing $[0, K]$ into intervals of width $\frac{K}{m+1}$. For $i = 1, \ldots, m$ let $q_i - q_{i-1} = \frac{K}{m+1}$. Then, $\int_0^{\phi(K)} \phi^{-1}(q) dq$ is approximated by $\sum_{i=1}^{m} \phi^{-1}(q_i)[q_i - q_{i-1}]$. In words, the integral is approximated by a weighted sum of quantiles.

If the Receiver has a payoff function of the form given in (10), the Sender will be interested in characterizing the distribution over realized vectors of d quantiles. For an economy of exposition only, focus on the case of quartiles. Let $\mathcal{S}$ denote a set of possible quartiles; then, an element of $\mathcal{S}$ is a vector (r_1, r_2, r_3) with $r_1 < r_2 < r_3$. Here r_1 is the 25^{th} percentile, r_2 is the 50^{th} percentile, while r_3 is the 75^{th} percentile. Denote by $\rho(r_1, r_2, r_3)$ the 'desired' probability of realizing a posterior whose quartiles take values (r_1, r_2, r_3). Given $\mathcal{S}$ and a probability vector ρ over the elements of $\mathcal{S}$, is there a mapping of states to signals that will induce the distribution ρ over $\mathcal{S}$? Theorem 1 will provide us the answer once we have chosen appropriate pairs of bounding functions $(F_s(\cdot), G_s(\cdot))$ for $s = (r_1, r_2, r_3)$:

$$G_s(t) = \begin{cases} 0 & \text{if } t < r_1, \\ 0.25 & \text{if } r_1 \leq t < r_2 \\ 0.5 & \text{if } r_2 \leq t < r_3, \\ 0.75 & \text{if } r_3 \leq t \end{cases} \qquad F_s(t) = \begin{cases} 0.25 - \epsilon & \text{if } t < r_1, \\ 0.5 - \epsilon & \text{if } r_1 \leq t < r_2 \\ 0.75 - \epsilon & \text{if } r_2 \leq t < r_3, \\ 1 & \text{if } r_3 \leq t \end{cases}$$

Theorem 1 yields the following:

COROLLARY 3.2 *Assume the prior π is full support. Let $\mathcal{S}$ be a finite set of possible quartile vectors r and ρ a probability distribution over $\mathcal{S}$. There is a mapping of states to signals such that the probability of the realized posterior having quartiles $s \in \mathcal{S}$ is ρ_s if and only if there is an $\epsilon > 0$ such*

that for all $l < k \leq n$*:*

$$\sum_{i=1}^{k} \pi_i \geq 0.25 \sum_{s:\ r_1^j \leq k < r_2^j} \rho_s + 0.5 \sum_{s:\ r_2^j \leq k < r_3^j} \rho_s + 0.75 \sum_{s:\ r_3^j \leq k < n} \rho_s$$

$$\sum_{i=1}^{k} \pi_i \leq (0.25-\epsilon) \sum_{s:\ k < r_1^j} \rho_s + (0.5-\epsilon) \sum_{s:\ r_1^j \leq k < r_2^j} \rho_s + (0.75-\epsilon) \sum_{s:\ r_2^j \leq k < r_3^j} \rho_s + \sum_{s:\ r_3^j \leq k} \rho_s$$

$$\sum_{i=l}^{k} \pi_i \geq (0.5+\epsilon) \sum_{s:\ r_3 \leq k,\ r_1 > l} \rho_s + (0.25+\epsilon) \sum_{\substack{s:\ (r_3 \leq k,\ r_1 \leq l \leq r_2)\ or \\ (r_2 \leq k < r_3,\ l < r_1)}} \rho_s + \epsilon \sum_{\substack{s:\ (r_3 \leq k,\ r_2 \leq l \leq r_3)\ or \\ (r_2 \leq k < r_3,\ r_1 \leq l < r_2)\ or \\ (r_1 \leq k < r_2,\ l < r_1)}} \rho_s$$

Stochastic Dominance

We consider a different yet related question involving first-order stochastic dominance. Recall that a distribution $H(\cdot)$ over $\{1, \ldots, n\}$ is said to first-order stochastically dominate (FOSD) distribution $W(\cdot)$ if

$$H(t) \leq W(t)\ \forall t \leq n-1.$$

If the corresponding density functions are denoted $h(\cdot)$ and $w(\cdot)$ respectively, the FOSD condition can be written as:

$$\sum_{j=1}^{t} h(j) \leq \sum_{j=1}^{t} w(j), \quad \forall\, t \leq n-1,$$

$$\sum_{j=1}^{n} h(j) = 1.$$

If H FOSD W, then each q-quantile of H is no larger than the corresponding q-quantile of W.

To each signal $s \in \mathcal{S}$ we associate a probability distribution over $\{1, \ldots, n\}$ denoted $W_s(\cdot)$. Call a probability distribution $\{\rho_s : s \in \mathcal{S}\}$ implementable if there is a mapping of states ($\mathcal{N}$) to signals ($\mathcal{S}$) such that the realized posterior at signal s FOSD $W_s(\cdot)$.

One setting where this question may be relevant is in the design of derivatives for risk-averse investors. Stochastic dominance constraints are one approach to incorporating risk aversion into portfolio selection. The objective is to construct a portfolio that maximizes expected payoff subject to the portfolio's distribution of payoffs stochastically dominating the payoff distribution of a specified benchmark portfolio (see, for example, Porter and Gaumnitz (1972) and Dentcheva and Ruszczynski (2006)). This literature is largely silent on the choice of the benchmark. We propose an approach that accommodates *multiple* benchmarks.

Suppose one unit of a security that will realize value i with probability π_i tomorrow. Each $s \in \mathcal{S}$ corresponds to the payoff distribution, $W_s(\cdot)$, of some benchmark portfolio. The goal is to design a derivative based on this security whose payoffs will first-order stochastically dominate $W_s(\cdot)$ for each $s \in \mathcal{S}$. Specifically, for each outcome $i \in \mathcal{N}$, the holder of derivative $s \in \mathcal{S}$ will receive a fraction

$p(i,s)$ of the value of the security. If we set $x_{i,s} = p(i,s)\pi_i$, we obtain our stochastic mapping of states to signals. Then, $\rho_s = \sum_i x_{i,s}$ can be interpreted as the 'quantity' of derivative s that can be offered.

The following is immediate from Theorem 1.[3]

COROLLARY 3.3 *A distribution $\{\rho_s\}_{s\in\mathcal{S}}$ is implementable such that each posterior at $s \in \mathcal{S}$ first order stochastically dominates $W_s(.)$ if and only if*

$$\sum_{i=1}^{t} \pi_i \leq \sum_{s\in\mathcal{S}} \rho_s W_s(t) \ \forall t = 1, \ldots, n.$$

4 Statistical Discrimination

In this section, we show how to use Theorem 1 to test for discrimination within the information design framing introduced in Chambers and Echenique (2020) and used in Bharadwaj et al. (2024). It can be interpreted as a non-parametric generalization of the models in Phelps (1972) and Aigner and Cain (1977). The literature on discrimination is extensive, see Onuchic (2022) and Romei and Ruggieri (2014) for surveys but our focus is exclusively on that sliver that adopts the information design framing.

Agents of varying types belong to different identity groups. Each agent is assigned to a signal, which can be interpreted as some set of observable characteristics separate from group identity. Thus, agents with different types and group identities may be assigned to the same signal. Associated with each signal is an outcome, and that outcome is determined by a decision rule.

For example, in a labor market, an agent's type might be their productivity, which is unobserved by the employer. The signal represents a noisy proxy for their productivity, such as educational attainment and prior work experience, while group identity could refer to attributes like gender. An outcome is a wage. The decision rule determining wage could depend upon a combination of an agent's signal, group identity, prior knowledge about the distribution of productivities, and mapping from productivities to signals. In an educational setting, a student's type might be their SAT score, their group identity could refer to their race, and the signal corresponds to the college they apply to. A decision rule determines the admission outcome, which may depend on their score, the score distribution of applicants at that college, and the student's group identity.

If we observe a difference in outcomes between groups, we would like to know why. If it is because the decision rule conditions on the identity groups, we say the differences are a consequence of *taste-based discrimination*. If, on the other hand, we know that the decision rule does not condition on group membership, then outcome differences must be either the result of differences in the underlying distribution of types between groups or the mapping from types to signals differing between groups. In this case, the differences are ascribed to *statistical discrimination*.[4]

[3] Note that, in this case, we only have an upper bound function, so the condition is significantly simpler.

[4] If the state to signal mapping differs between groups, this begs the question of why? For instance, in the labor market setting, the signal could be an interview, and the difference arises because the interviewer engages in

We will consider two questions depending on the data available to the putative analyst. For the first (college admission), the analyst has access to the prior distribution over types conditional on group, the outcome distributions at each signal, but not the decision rule. Assuming the decision rule takes a specific form—a cutoff rule (with unobserved thresholds)- can the available data be rationalized by the hypothesis that the decision rule does not condition on group membership? In the second (labor market), the analyst has access to the decision rule and outcomes but not the distribution over types or the mapping of types to signals. Can the observed outcomes be rationalized by some distribution over types and type to signal mapping?

We now describe the generalization of Theorem 1 to answer this question, followed by two applications.

Let $\mathcal{G}$ be a finite set of identity groups, and let π_i^g denote the probability that a member of group $g \in \mathcal{G}$ has type $i \in \mathcal{N} = \{1, \ldots, n\}$. Hence, $\sum_{g\in\mathcal{G}} \sum_{i\in\mathcal{N}} \pi_i^g = 1$. As before, let $\mathcal{S}$ denote the set of signals. For each $g \in \mathcal{G}$ there is a group-specific mapping of $\mathcal{N}$ to $\mathcal{S}$ denoted $\{x_{is}^g\}_{i\in\mathcal{N},s\in\mathcal{S}}$. Each $x_{is}^g \geq 0$ and of course $\sum_{s\in\mathcal{S}} x_{is}^g = \pi_i^g$ for all $g \in \mathcal{G}$ and $i \in \mathcal{N}$. Each group-specific mapping will induce a prior distribution over $\mathcal{N}$ that depends on the realized signal *and* the relevant group.

As in Theorem 1 for each signal s, we require that the posterior conditional on s lie between a pair of bounding functions denoted $(G_s(\cdot), F_s(\cdot))$. We will also require that the posterior conditional on each signal s *and* group $g \in \mathcal{G}$ lie between a pair of bounding functions $(G_s^g(\cdot), F_s^g(\cdot))$. The following is an immediate result of Theorem 1.

COROLLARY 4.1 *$\{\rho^g\}_{g\in\mathcal{G}}$ is implementable if and only if for all $t = 1, \ldots, n$, $\forall g \in \mathcal{G}$*

$$\sum_{s\in\mathcal{S}} G_s^g(t)\rho_s \leq \sum_{i=1}^{t} \pi_i^g \leq \sum_{s\in\mathcal{S}} F_s^g(t)\rho_s$$

and

$$\sum_{i=t+1}^{t'} \pi_i^g \geq \sum_{s\in\mathcal{S}} \max\{G_s^g(t') - F_s^g(t), 0\}.$$

Although this result is a straightforward consequence of Theorem 1, we demonstrate next how it enables the detection of discrimination using limited aggregate information. The key insight lies in the fact that there is a common structure in the upper and lower bound functions, $G_s^g()$ and $F_s^g()$, across $g \in \mathcal{G}$ under the hypothesis that group membership does not influence the mapping from state to signal. This hypothesis imposes a constraint on the set of achievable posteriors, which serves as the basis for testing the hypothesis.

Discrimination in College Admissions

We propose a test to detect whether colleges use group membership when selecting based on cutoff rules. Each student, in addition to their group identity, is assumed to have a one-dimensional score

taste-based discrimination. This is not an issue we take up here.

that reflects their underlying quality or suitability for selection. The decision rule is modeled as threshold-based: an individual is selected if and only if their score exceeds a cutoff, which may vary across institutions or decision-makers.[5]

In our framework, each applicant has a score, and we identify the state space $\mathcal{N}$ with the set of possible scores. Let $\mathcal{S}$ represent the set of colleges. Without loss, we may assume that each student applies to exactly one college. A student applying to $k > 1$ colleges can be replaced by k clones of itself, with each clone applying to exactly one of the colleges the original student had applied to. We normalize the total number of students to 1. The number of students in group $g \in \mathcal{G}$ with score $i \in \mathcal{N}$ is π_i^g.

For each college $s \in \mathcal{S}$, we observe the following:

1. The overall acceptance rate, denoted α_s, as well as the group-specific acceptance rates $\{\alpha_s^g\}_{g\in\mathcal{G}, s\in\mathcal{S}}$.
2. The number of students who apply to college s, denoted by ρ_s, and the group-specific applicants, denoted by $\{\rho_s^g\}_{g\in\mathcal{G},\, s\in\mathcal{S}}$. Here, the total number satisfies $\rho_s = \sum_{g\in\mathcal{G}} \rho_s^g$.

We refer to the tuple of vectors (π, ρ, α) as a *college admission dataset*. Using this data, we want to verify whether each college uses a group-independent threshold. Specifically, for each college s, there is a threshold p_s, and each student, no matter their group, is accepted as long as their score exceeds p_s. We do not observe individual-level information linking applicants' scores to their admission decisions. Such detailed information may be unavailable to an analyst due to privacy concerns.

Definition 4.1 *We say that a college admission dataset (π, ρ, α) is* rationalizable *by group independent thresholds if there exists a mapping from students $\mathcal{N}$ to colleges (signals) $\mathcal{S}$, along with a set of thresholds $\{p_s\}_{s\in\mathcal{S}}$, such that:*

- *The number of students from each group $g \in \mathcal{G}$ applying to each college matches the application data specified by π and ρ;*
- *At each college, the acceptance decision is based solely on whether a student's score exceeds the group-independent threshold p_s, and the resulting acceptance rates match the data α.*

The next example illustrates Definition 4.1 and shows that even with identical aggregate data and colleges with different acceptance rates, one can still rationalize their admissions decisions using group-independent thresholds. This is because different students apply to different schools.

Example 4.1 *Assume two colleges, s_1 and s_2, and two groups, $\mathcal{G} = \{A, B\}$. Suppose the score distribution for each group $g \in \mathcal{G}$ is uniform over the interval $[1, 100]$, that is, $\pi_i^g = \frac{1}{|\mathcal{G}|n} = \frac{1}{200}$.*

[5] If admissions are based on a range of factors, one can interpret the score as a composite index or a latent institutional ranking common to all institutions. This would be true in countries where college admissions decisions are made by examination scores only. It would also be true in school choice districts where all schools must use the same criteria to assign priority scores to students. The key assumption is that admission decisions reduce to comparing individuals against a threshold, regardless of how the underlying score is constructed.

Assume that the number of students from each group applying to each college is the same, so $\rho_s^g = \frac{1}{|\mathcal{S}||\mathcal{G}|} = \frac{1}{4}$. *We assume that for group A, the acceptance rate at both colleges* $\alpha_1^A = \alpha_2^A = 10\%$, *while for group B,* $\alpha_1^B = \alpha_2^B = 20\%$.

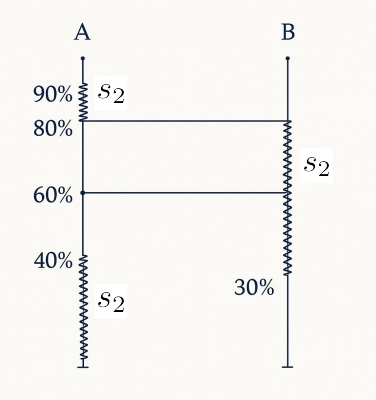


Figure 1: Mapping of students to colleges

To convey which students apply to which colleges, we use Figure 1. The vertical line on the left represents group A students sorted by their scores from bottom to top. The vertical line on the right corresponds to group B students. An interval demarcated by a zigzag indicates the students applying to college 2; the remaining students apply to college 1. This data can be rationalized by the colleges using group independent thresholds. To verify this, it is sufficient to identify thresholds, one for each college, consistent with this data. So, suppose college 1 admits students with scores above the 80th percentile, while college 2 admits those with scores above the 60th percentile. We show in Example 4.2, that increasing the acceptance rate for group B above 20% makes the data no longer rationalizable.

To capture the acceptance rate of college s, we observe that the cumulative posterior satisfies the property that the total probability of the realized score being less than or equal to p_s equals the rejection rate $1 - \alpha_s^g$. Thus, as before, for each $g \in \mathcal{G}$ and college s, we select a pair of bounding functions $(F_s^g(\cdot), G_s^g(\cdot))$ as follows:

$$G_s^g(t) = \begin{cases} 0 & \text{if } t < p_s, \\ 1 - \alpha_s^g & \text{if } t \geq p_s, \end{cases} \qquad F_s^g(t) = \begin{cases} 1 - \alpha_s^g & \text{if } t \leq p_s, \\ 1 & \text{if } t > p_s. \end{cases}$$

Theorem 1 yields the following:[6]

COROLLARY 4.2 *A college admission dataset* (π, α, ρ) *is rationalizable if and only if there exists a*

[6]In this case $F_s^g(t) \geq G_s^g(t')$ for all t, t', thus, inequality (2) in Theorem 1 is redundant.

set of thresholds $\{p_s\}_{s\in\mathcal{S}}$ *such that, for all* $t = 1, \ldots, n$ *and all* $g \in \mathcal{G}$*, the following conditions hold:*

$$\sum_{i=1}^{t} \pi_i^g \geq \sum_{s\in\mathcal{S}:p_s\leq t} (1-\alpha_s^g)\rho_s^g, \tag{11}$$

$$\sum_{i=t+1}^{n} \pi_i^g \geq \sum_{s\in\mathcal{S}:p_s\geq t} \alpha_s^g \rho_s^g. \tag{12}$$

In words, inequality (11) states that for each group g and score t, the number of *rejected* students at colleges with a threshold score at most t cannot exceed the number of students in the group who scored at most t. Inequality (12), on the other hand, states that for each group g and score t, the number of *accepted* students at colleges with a threshold score at least t cannot exceed the number of students in the group who scored above t.

Both (11) and (12) trivially hold if the data is rationalizable. Indeed, if a student is rejected by a college with a threshold at most t, then their score must be at most t; similarly, if a student is accepted by a college with a threshold at least t, then their score must be strictly greater than t. Corollary 4.2 states that these two conditions are necessary and sufficient to guarantee that the data is rationalizable.

For fixed thresholds $p_s s \in \mathcal{S}$, constraints (11–12) form a linear program. Thus, determining whether the data (π, ρ, α) is rationalizable reduces to finding thresholds $\{p_s\}_{s\in\mathcal{S}}$ that satisfy these constraints. The number of possible threshold configurations is $n^{|\mathcal{S}|}$, which can be large. However, as we explain below, if the relative order of the thresholds is known, then checking for feasibility—and finding a solution when one exists—becomes easy. In practice, colleges often have a pre-established ranking, which provides information about the relative ordering of their thresholds. Even when such a ranking is not known, we can exhaustively search over all possible orderings of colleges. The number of these orderings depends only on $|\mathcal{S}|$ and is independent of the maximum score n, offering a more tractable approach for testing rationalizability.

Suppose that the colleges are partitioned into vertically ordered tiers $T_1, \ldots, T_m$. Colleges in the same tier have the same admission threshold, and a higher-tier college will have a higher admission threshold than one in a lower tier. Let T_1 denote the lowest tier, T_2 denote the second lowest tier, and so on. We use (11) and (12) to identify a feasible interval for each possible admission threshold.

If there is a set of threshold, $\{p_k\}_{k=1}^m$ consistent with the tier structure, then, (11) is equivalent to the following statement: for every k the number of rejected students at school in tier k or lower is at most the total number of students with score p_k or less. i.e.,

$$\sum_{i=1}^{p_k} \pi_i^g \geq \sum_{s\in T_1\cup\ldots\cup T_k} (1-\alpha_s^g)\rho_s^g. \tag{13}$$

For each fixed p_k, the left-hand side of (13) is a constant determined by the given dataset. It increases in p_k. Hence, inequality (13) is satisfied if and only if p_k exceeds a certain lower bound that will depend on the dataset (π, ρ, α).

Similarly, (12) is equivalent to the requirement that the number of students accepted at colleges in tier k or higher is at most the number of students with score exceeding p_k, i.e.,

$$\sum_{i=p_k+1}^{n} \pi_i^g \geq \sum_{s \in T_k \cup \ldots \cup T_m} \alpha_s^g \rho_s^g. \tag{14}$$

As above, it is easy to see that (14) is equivalent to requiring that p_k be below a certain upper bound that depends on the dataset.

Thus, for each tier k and each group $g \in \mathcal{G}$, (13) and (14) are equivalent to a lower and upper bound on p_k. If the lower bound is larger than the upper bound, then the dataset is not rationalizable. In general, determining whether there exists a set of thresholds $\{p_k\}_{k=1}^m$ that satisfy these bounds and are consistent with the assumed order can be done efficiently using a greedy algorithm; In general, determining whether there exists a set of thresholds $\{p_k\}_{k=1}^m$ that satisfy these bounds and are consistent with the assumed order can be done efficiently using a greedy algorithm; see Appendix C for the details. Example 4.2 below illustrates how to compute these bounds and shows that, in this case, they imply that no such p_k can exist for all possible tier orders.

The next example, which is a slight modification of Example 4.1, demonstrates how to compute the bounds and shows that underlying the dataset can not be rationalized by group independent thresholds *regardless* of tier structure.

Example 4.2 *Assume two colleges, s_1 and s_2, and two groups, $\mathcal{G} = \{A, B\}$. Suppose the score distribution for each group $g \in \mathcal{G}$ is uniform over the interval $[1, n]$, that is, $\pi_i^g = \frac{1}{|\mathcal{G}|n} = \frac{1}{2n}$. The number of students from each group applying to each college is the same, so $\rho_s^g = \frac{1}{|\mathcal{S}||\mathcal{G}|} = \frac{1}{4}$. For group A, the acceptance rate at both colleges is 10%, while for group B, it is 21%.*

First, we assume the two colleges employ the same threshold and therefore belong to the same tier. Inequalities (13) and (14) apply to the tier consisting of both colleges. As the score distribution is the same across groups, these inequalities imply that if we treat s_1 and s_2 as a single hypothetical joint college, the acceptance rate must be the same across groups. However, in this example, group A faces a higher threshold than group B, violating that implication.

Next, assume college s_2 is in a higher tier than college s_1 (without loss by virtue of symmetry), that is, it has a higher threshold: $p_2 > p_1$. For each group $g \in \mathcal{G}$, applying inequality (13) for $p_k = p_2$, we deduce that the number of students in group g with score at most p_2 must be at least the number of students rejected by both colleges. That is,

$$\frac{p_2}{n|\mathcal{G}|} \geq (2 - \alpha_1^g - \alpha_2^g) \cdot \frac{1}{2|\mathcal{G}|} \quad \Leftrightarrow \quad \frac{p_2}{n} \geq 1 - \frac{\alpha_1^g + \alpha_2^g}{2}.$$

Applying inequality (14) with $p_k = p_2$, the number of students in group g with score greater than p_2 must be at least the number of students accepted at college s_2. That is,

$$\frac{n - p_2}{n|\mathcal{G}|} \geq \alpha_2^g \cdot \frac{1}{2|\mathcal{G}|} \quad \Leftrightarrow \quad \frac{p_2}{n} \leq 1 - \frac{\alpha_2^g}{2}.$$

Thus, for group A with $\alpha_1^A = \alpha_2^A = 10\%$*, we have* $0.9 = 1 - 0.1 \leq \frac{p_2}{n} \leq 1 - 0.05 = 0.95$*. And for group B with* $\alpha_1^B = \alpha_2^B = 21\%$*, we have* $0.79 = 1 - 0.21 \leq \frac{p_2}{n} \leq 1 - 0.105 = 0.895$*. These two intervals do not overlap, so no single value of* p_2 *can satisfy both conditions simultaneously. Therefore, we conclude that the admission data in this example above is not rationalizable by group independent thresholds independent of the ordering of the colleges.*

Wage Discrimination

Bharadwaj et al. (2024) uses the information design framing to propose a non-parametric test for when the observed wage distribution for two groups is consistent with a model of statistical discrimination. In their setting, workers belong to one of two groups, say male and female. The distribution of marginal productivities (the type) in the two groups may differ, but the mean productivities of the two groups coincide. Employers do not observe a worker's actual productivity but a noisy signal of it. Wage is assumed to be a monotone function of the *expected* productivity given the signal. The stochastic mapping of productivities to signals can differ between the two groups, which would be the cause of statistical discrimination.

Given the wage distribution for each group and knowing that the mean productivities are the same, are there two prior distributions and stochastic mappings of productivities to signals (one for each group) that rationalize the observed wage distributions? If none, then the observed wage distributions cannot be accounted for by statistical discrimination. Bharadwaj et al. (2024) shows that a pair of wage distributions is rationalizable under the assumption of equal mean productivities if neither wage distribution strictly first-order stochastically dominates the other.

What if the wage is conditioned on a different parameter of the conditional distribution of productivities, such as the median? If we replace the mean condition in Bharadwaj et al. (2024) with the median, then using Corollary 4.1, the condition for each group g is

$$\frac{1}{2}\sum_{i=1}^{t}\rho_i^g \leq \sum_{i=1}^{t}\pi_i^g \leq \left(\frac{1}{2}-\epsilon\right) + \left(\frac{1}{2}+\epsilon\right)\sum_{i=1}^{t}\rho_i^g \;\forall t \in \{1,\ldots,n\}.$$

This constraint is straightforward to satisfy, e.g. $\pi_i^g = \rho_i^g$ for all i. With the additional requirement that the median of the posterior is k (i.e. $\sum_{i=1}^{k}\pi_i^g = \frac{1}{2}$), a $\{\pi_i^g\}_{i=1}^n$ that satisfies these conditions exists if and only if $\sum_{i=1}^{k}\rho_i^g \geq \frac{\epsilon}{\epsilon+0.5}$.

Hence, for all practical purposes, the observed wage distributions can *always* be rationalized by two stochastic mappings of productivities to signals. Therefore, it is almost impossible to reject the hypothesis that statistical discrimination explains the observed wage distribution based on the median alone, making for a sharp contrast with Bharadwaj et al. (2024).

The signal, of course, can be based on other parameters, such as quartiles. The choice of signal defines the set of pairs of bounding functions, $F_s(\cdot)$ and $G_s(\cdot)$. For a prior distribution consistent with a state-to-signal mapping to exist, the cumulative probability of the prior, $\sum_{i=1}^{t}\pi_i$ must lie between the averages of $G_s(t)$ and $F_s(t)$ for $s \in \mathcal{S}$. This condition must hold for each group, with

additional constraints required to connect the two priors. As shown in Bharadwaj et al. (2024), one such constraint is that the means of both priors must be equal. These constraints give rise to a linear program, the feasibility of which can be characterized by Farkas' Lemma, providing a necessary and sufficient condition for existence.

5 Market Segmentation

This section demonstrates the application of Theorem 1 to market segmentation, extending the work of Bergemann et al. (2015). The setting consists of a population of buyers where the fraction of buyers with a reservation price of $i \in \{1, \ldots, n\}$ is π_i. The buyers are segmented into sets indexed by price p. Let $\mathcal{S}$ be the set of possible prices and $\{\rho_p\}_{p\in\mathcal{S}}$ a distribution over $\mathcal{S}$, that is $0 \leq \rho_p \leq 1$, and $\sum_{p\in\mathcal{S}} \rho_p = 1$. Associated with each segment p is the probability of a buyer in this segment purchasing the good, denoted q_p with $0 \leq q_p \leq 1$. Let $d_p = q_p\rho_p$, be the measure of buyers who purchase the good in that segment. We will call $\{(\rho_p, q_p)\}_{p\in\mathcal{S}}$ as well as $\{(\rho_p, d_p)\}_{p\in\mathcal{S}}$ a *probability–quantity profile*. A probability–quantity profile, $\{\rho_p, d_p\}_{p\in\mathcal{S}}$, is said to be *implementable* if the following conditions are met.[7]

1. The measure of buyers assigned to segment p is ρ_p.
2. $p \in \{1, \ldots, n\}$ is a revenue maximizing price (normalize production cost to 0) for the distribution of buyer reservation prices assigned to segment p.
3. The revenue maximizing mass of buyers in segment p purchasing the good is d_p.

Hence, every implementable probability–quantity profile corresponds to a segementation and therefore, we will sometimes refer to an implementable probability–quantity profile as a segmentation.

Theorem 1 provides a complete characterization of all implementable probability–quantity profiles that differs from that in Bergemann et al. (2015). They focus on the consumer-producer surplus pairs associated with a segmentation rather than the realized distribution of prices themselves. Nevertheless, as we demonstrate below, the characterization of achievable consumer-producer surplus pairs can be derived from our framework. Before summarizing these and other consequences, we highlight the mathematical differences between our approach and that of Bergemann et al. (2015).

In Bergemann et al. (2015) and our work, one can interpret the partition of buyers into segments as a stochastic mapping from states to signals. The state i represents the fraction of buyers with reservation price i, and the signal s corresponds to a segment. The approach in Bergemann et al. (2015) works directly with the variables $x_{i,s}$, which specify the fraction of buyers with reservation price i assigned to segment s. Beyond the usual mass balance constraints, these variables are subject to consistency conditions ensuring that the distribution of reservation prices within each segment p supports p as a revenue-maximizing price. In contrast, we employ a reduced-form representation

[7] Assuming that each price p is associated with at most one segment is without loss. If not, merge (p, q_1) and (p, q_2) into a single segment (p, q) with $\rho_{p,q} = \rho_{p,q_1} + \rho_{p,q_2}$, $q := \frac{q_1\rho_{p,q_1} + q_2\rho_{p,q_2}}{\rho_{p,q_1} + \rho_{p,q_2}}$.

centered on pairs $\{\rho_p, q_p\}$, where ρ_p is the probability of segment p and q_p the associated sales quantity. The relevant consistency constraints are encoded via carefully chosen bounding pairs, simplifying the problem structure.

Our reduced-form characterization explicitly links the distribution of segment prices and quantities, $\{\rho_p, q_p\}_{p\in\mathcal{S}}$, with the prior distribution of buyer types, $\{\pi_i\}_{i\in\mathcal{N}}$. This allows one to recover the underlying demand distribution from observed prices and quantities, enabling applications such as market analysis, demand forecasting, and the design of pricing or targeting strategies. For example, we characterize implementable probability–quantity profiles that maximize total or consumer surplus; determine all demand distributions π consistent with observed segment sizes and purchase quantities; impose fairness constraints by selecting, among surplus-maximizing distributions, those where consumer surplus within any segment does not exceed a threshold Δ; and extend these results to environments with multidimensional yet ordered consumer types.

For each price segment p, recall that $d_p = q_p\rho_p$, the measure of buyers who obtain the good in that segment. We characterize the set of probability-quantity profiles $\{(\rho_p, d_p)\}_{p=1}^n$ that are implementable. Below, we identify three inequalities and explain why any implementable probability–quantity profile satisfies them. Subsequently we show they are sufficient.

1. **Basic feasibility:** Since $0 \le q_p, \rho_p \le 1$, it follows that $d_p \le \rho_p$. Also, as the total measure of buyers is 1:
$$\sum_{p=1}^{n} \rho_p = 1 \quad \text{and} \quad 1 \ge \rho_p \ge d_p \ge 0 \quad \text{for all } p. \tag{15}$$

2. **Consumption constraint between types t and t':**
$$\sum_{p=t}^{t'} d_p \le \sum_{i=t}^{t'} \pi_i + \frac{1}{t'+1}\left(\sum_{p=t}^{t'} p \cdot d_p\right). \tag{16}$$

   The left-hand side is total consumption at prices between t and t'. Since buyers only purchase at prices weakly below their reservation price, these consumers must have a reservation price at least t, and the total mass of such buyers must be bounded by:

   - *Low reservation price:* Buyers with reservation price in $[t, t']$, whose total mass is at most $\sum_{i=t}^{t'} \pi_i$.
   - *High reservation price:* Buyers with reservation price $> t'$ who are assigned to segments priced in $[t, t']$. The number of such buyers cannot exceed the total revenue from these segments divided by $t'+1$ (the second term on the right-hand side of (16)). Otherwise, at least one segment could profitably deviate by raising its price to $t'+1$, contradicting optimality.

Note that when $t' = n$, there is no buyers of types $> n$. Thus, the inequality reduces to:

$$\sum_{p=t}^{n} d_p \leq \sum_{i=t}^{n} \pi_i. \tag{17}$$

This states that the total consumption of price t or higher is at most the total buyers of type t or higher.

3. **Revenue constraint for a group of segments A:** Because of the revenue maximizing condition, for any subset $A \subseteq \{1, \ldots, n\}$ of segments and any $t \in \{1, \ldots, n\}$, revenue maximization requires that the revenue generated from segments in A, scaled by $\frac{1}{t}$, must be at least the total mass of buyers of type t or higher assigned to A. This mass is at least the total number of buyers with type $\geq t$ minus the total mass of buyers *not* assigned to A. Formally,

$$\frac{1}{t}\left(\sum_{p \in A} p \cdot d_p\right) \geq \sum_{i=t}^{n} \pi_i - \sum_{p \notin A} \rho_p. \tag{18}$$

One such inequality exists for each (A, t) pair. Thus the number of inequalities is exponential in n. However, for given (ρ, γ) it suffices to check for each t, one constraint in (18) only to verify that all others are satisfied. Specifically, the set $A := \{p : \frac{p}{t} \cdot d_p \leq \rho_p\}$.

Our main result in this section (Theorem 2) establishes that inequalities (15), (16), (17), and (18) are not only necessary but also sufficient for the probability-quantity profile $\{\rho_p, d_p\}_{p=1}^n$ to be implementable.

THEOREM 2 *The probability-quantity profile $\{\rho_p, d_p\}_{p=1}^n$ is implementable if and only if the buyer mass ρ_p and the mass of buyers who obtain the good, d_p, at each price segment p, satisfy inequalities (15), (16), (17), and (18) for all $t = 1, \ldots, n$ and all subsets $A \subseteq \{1, \ldots, n\}$.*

The proof is given in the Appendix. The key idea is to construct upper and lower bounding functions for each segment based on its price–quantity pair (p, q_p) (recall $q_p = \frac{d_p}{\rho_p}$) and apply Theorem 1.

In particular, for a segment with price p and quantity q_p, it must be that a fraction $1 - q_p$ of the population has reservation values strictly below p—these are the buyers who do not purchase. This provides an *upper bound* on the cumulative distribution of types below p. That is,

$$F_p(t) = \begin{cases} 1 - q_p & \text{if } t \leq p - 1, \\ 1 & \text{if } t \geq p. \end{cases}$$

To encode the condition that p is the revenue-maximizing price for the segment, we impose the following: for any alternative price p', the revenue from selling to buyers with reservation price at least p' must be no more than pq_p. That is, the mass of buyers with reservation price at least p'

must be at most $\frac{pq_p}{p'}$. This leads to a *lower bound* on the cumulative distribution function, requiring that the fraction of buyers with reservation value below p' is at least.

$$G_p(p'-1) = \max\{1 - \frac{pq_p}{p'}, 0\}.$$

A special case of (18) with $A = \{1, \ldots, n\}$ implies that the total revenue of the segmentation satisfies

$$\sum_{p=1}^{n} p \cdot d_p \;\geq\; \Pi^* := \max_t \; t \cdot \sum_{i=t}^{n} \pi_i. \tag{19}$$

That is, the revenue from any segmentation is at least as large as the maximum revenue attainable under uniform pricing. This represents one edge of the 'surplus triangle' is introduced in Bergemann et al. (2015). Let C denote the total consumer surplus and R the total profit. It is obvious that $C + R \leq \sum_{i=1}^{n} i\pi_i$ and $R \geq \Pi^*$). Let

$$T = \{(R, C) : R \geq \Pi^*,\; C \geq 0,\; C + R \leq \sum_{i=1}^{n} i\pi_i\}.$$

The set T is the surplus triangle of Bergemann et al. (2015) and they show that it describes the set of (R, C) pairs that can be obtained by some segmentation. To prove this, it suffices to show that each corner of the triangle T corresponds to a segmentation. We show below how our result can be used to construct these "extremal" segmentations.

Efficient Segmentation

We characterize segmentations that maximize total surplus, which we call *efficient segmentations*. We then use this characterization to provide a new proof of the existence of an efficient segmentation that maximizes consumer surplus.

An efficient segmentation serves all consumers, meaning that $q_p = 1$ for every p such that $\rho_p > 0$. This implies $d_p = \rho_p$. Therefore, for efficient segmentations, we simplify the notation and refer to them simply as distributions $\{\rho_p\}_{p=1}^n$. With this simplification, we apply Theorem 2 to obtain the following result.

COROLLARY 5.1 *An implementable probability–quantity profile, $\{\rho_p\}_{p=1}^n$ with $\sum_p \rho_p = 1$ is efficient*

if and only if

$$\sum_{i=t}^{n} \pi_i \geq \sum_{p=t}^{n} \rho_p \forall t = 1, \ldots, n \tag{20}$$

$$\sum_{i=t}^{t'} \pi_i \geq \sum_{p=t}^{t'} (1 - \frac{p}{t'+1}) \rho_p \forall t < t' \in \{1, \ldots, n-1\} \tag{21}$$

$$\sum_{i=t}^{n} \pi_i \leq \sum_{p=1}^{t-1} \rho_p + \frac{1}{t} \sum_{p=t}^{n} p \cdot \rho_p \ \forall t = 1, \ldots, n \tag{22}$$

Next, we show that there exists an efficient segmentation that attains the revenue lower bound. This corresponds to the corner point $(R, C) = \left(\Pi^*, \sum_i i \cdot \pi_i - \Pi^*\right)$. This result is established in Bergemann et al. (2015); here, for completeness, we provide a proof based on this LP formulation (see Appendix).

THEOREM 3 (BERGEMANN ET AL. (2015)) *For any distribution π there exits an efficient segmentation $\{\rho_p\}_{p=1}^n$ with total revenue $\sum_p p\rho_p = \Pi^* = \max_p p \cdot \sum_{i=p}^n \pi_i$.*

Inverse consumption Curve

We introduce a visual interpretation of the probability-quantity profile $\{\rho_p, d_p\}_{p=1}^n$ called an inverse consumption curve, and use it to characterize all segmentations that maximize total surplus and provide a new proof of the main result in Bergemann et al. (2015).

Associate with each probability–quantity profile $\{\rho_p, d_p\}_{p=1}^n$ (not necessarily implementable) an inverse consumption curve. Informally, treat each putative segment p as a single consumer whose demand is $d_p = \rho_p q_p$ for any price at most p and zero otherwise.

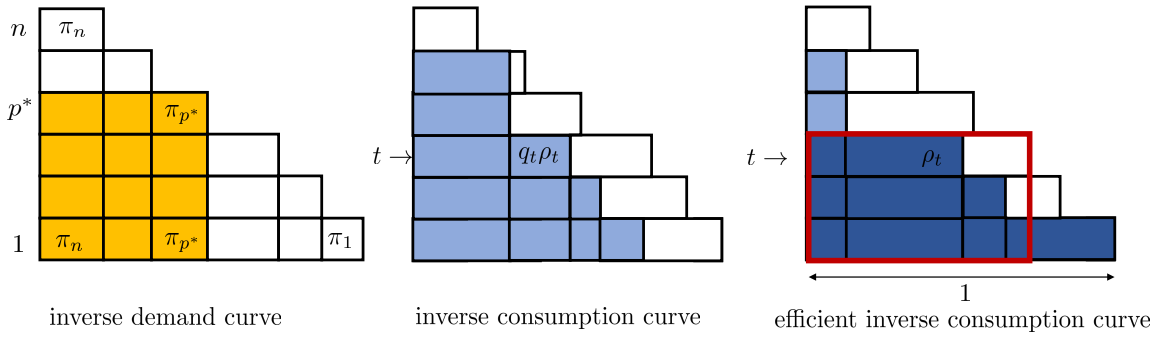


Figure 2: Inverse demand curve and inverse consumption curve

The inverse consumption curve can be constructed by stacking vertical columns, each representing a price p and quantity $d_p = \rho_p q_p$, arranged from left to right in decreasing order of price.

Figure 2 provides a visualization of an inverse demand curve and an inverse consumption curve. Prices appear on the vertical axes while the horizontal axes corresponds to quantities. The north-east

boundary in each panel depicts the inverse demand curve corresponding to the prior distribution, $\{\pi_i\}_{i=1}^n$. The area of region shaded in yellow in the leftmost panel corresponds to the maximum possible revenue under a uniform price.

The middle panel illustrates the inverse consumption curve as a geometric object. Imagine stacking n vertical columns from left to right, with heights decreasing from $t = n$ down to $t = 1$, where the column for price t has height t and width $d_t = q_t\rho_t$. Viewed horizontally, the figure can be sliced into *slabs* of unit height, ordered from longest to shortest, starting at the bottom. The slab at height i extends a distance $\sum_{p\geq i} d_p$, representing the total quantity consumed at prices at least i. Tracing the north–east boundary of the blue shaded region yields the inverse consumption curve. The total area under this curve corresponds to total revenue—the sum of price $\times$ quantity across all sales.

Inequality (17) states that for any implementable probability-quantity profile its corresponding inverse consumption curve must lie weakly below the inverse demand curve.

Corollary 5.1 characterizes efficient segmentation. Here, $\sum_p d_p = \sum_p \rho_p = 1$, and thus the bottom slab of the inverse consumption curve spans the entire unit interval. Condition (20) requires the inverse consumption curve to lie below the demand curve, while the remaining constraints determine its overall shape. For example, constraint (22) is equivalent to

$$t \cdot \sum_{i=t}^{n} \pi_i \;\leq\; \sum_{p=1}^{n} \min\{t, p\}\rho_p.$$

This means that the area of the inverse consumption curve below height t must be at least as large as the rectangle at height t under the inverse demand curve.

Full Inverse Consumption Curve. The first inverse consumption curve coincides with the inverse demand curve (see the left panel of Figure 3). Under this inverse consumption curve, the seller captures the entire surplus, leaving zero consumer surplus. This corresponds to a segmentation in which each type is assigned to its own segment. The associated corner of T is $(R, C) = (\sum_{i=1}^n i\pi_i, 0)$.

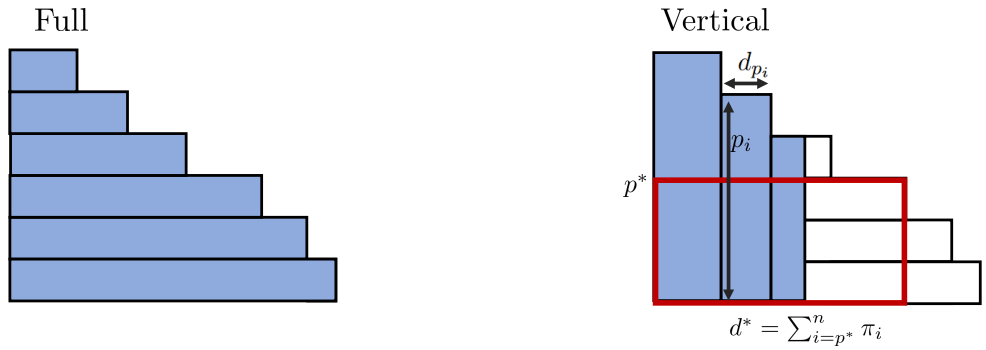


Figure 3: Two extreme consumption curves

Vertical Inverse Consumption Curve. To construct this inverse consumption curve, begin at the upper left, following the inverse demand curve up to a cutoff, beyond which the curve drops vertically to zero. The cutoff is chosen so that the total area under the curve equals Π^* (see the right panel of Figure 3).

If an implementable probability-quantity profile corresponds to this inverse consumption curve, the seller's revenue equals Π^*. Moreover, the number of consumers who do not receive the good equals 1 minus the width of the consumption curve—that is, the length of the bottom slab not covered by the vertical consumption curve.

Note that the area under the inverse demand curve beyond the cutoff (the white region in the right panel of Figure 3) represents the total reservation value of the excluded buyers. Since this area, together with the area under the inverse consumption curve, accounts for the entire area under the inverse demand curve (which corresponds to total surplus), it follows that the buyers who do purchase enjoy zero surplus. Therefore, this segmentation corresponds to the corner $(R, C) = (\Pi^*, 0)$.

Thus, it suffices to verify that the vertical consumption curve corresponds to a segmentation. As illustrated in Figure 3, this curve consists of k disjoint vertical rectangles, each with height p_i and width d_{p_i} for $i = 1, \ldots, k$. Define

$$\rho_{p_i} := \frac{p_i d_{p_i}}{\Pi^*} \quad \text{and thus} \quad q_{p_i} := \frac{d_{p_i}}{\rho_i} = \frac{\Pi^*}{p_i}.$$

Using Theorem 2, in the Appendix, we show that this probability-quantity profile is implementable.

COROLLARY 5.2 *$\{\rho_{p_i}, q_{p_i}\}_{i=1}^k$ as defined above is an implementable probability-quantity profile.*

Surplus Distribution

The many segmentations that maximize consumer surplus differ in how consumer surplus is distributed across segments. In particular, the high reservation price buyers benefit as they are assigned to segments with buyers whose reservation price is low. This would be a concern for an information intermediary choosing a segmentation to maximize consumer surplus whilst ensuring an equitable distribution between consumers. One way to formalize this concern is to limit the consumer surplus that any consumer enjoys, i.e., each consumer's surplus is at most Δ. A segmentation that yields zero surplus to all buyers trivially satisfies this constraint, but does not maximize consumer surplus. Any segmentation that maximizes consumer surplus must leave the seller with a profit Π^*. Depending on the magnitude of Δ, this may be impossible. We use our characterization of implementability to determine when this is possible.

Recall that efficiency requires each segment to serve all consumers, so $q = 1$ for every segment (p, q). We therefore associate each segment (or signal) directly with a price p and denote its probability by ρ_p. Let $x_{i,p}$ be the joint probability of state i and signal p. Under this notation, the

constraint that the seller achieves the minimum required revenue is

$$\sum_{p=1}^{n} p \cdot \rho_p = \Pi^*.$$

Next, we include an additional constraint: no consumer in a segment with price p enjoys a surplus more than Δ:

$$\sum_{i=1}^{p+\Delta} x_{i,p} \geq \rho_p.$$

To accommodate this additional constraint, we modify the definition of the lower bound function as follows:

$$G_p(t) := \begin{cases} 0, & \text{for } t < p-1, \\ 1 - \frac{p}{t+1}, & \text{for } p-1 \leq t < p+\Delta, \\ 1, & \text{for } t \geq p+\Delta. \end{cases}$$

The upper bound remains unchanged, but now with $q = 1$; that is,

$$F_p(t) = \begin{cases} 0, & \text{for } t < p, \\ 1, & \text{for } t \geq p. \end{cases}$$

Applying Theorem 1, we deduce the following characterization.

COROLLARY 5.3 *An efficient, consumer surplus maximizing segmentation in which the consumer surplus of each consumer is at most Δ exists if and only if there exists a distribution $\{\rho_p\}_{p=1}^n$ that satisfies:*

$$\sum_{p=1}^{n} p \cdot \rho_p = \Pi^*, \tag{23}$$

$$\sum_{p=t}^{n} \rho_p \leq \sum_{i=t}^{n} \pi_i \quad \forall t \in \{1, \ldots, n\}, \tag{24}$$

$$\sum_{p \geq t-\Delta} \min\{p, t\} \cdot \rho_p \geq t \cdot \sum_{i=t}^{n} \pi_i \quad \forall t \in \{1, \ldots, n\}. \tag{25}$$

$$\sum_{p: t < p \leq t' - \Delta} \rho_p + \sum_{p: \max\{t, t'-\Delta\} < p \leq t'+1} (1 - \frac{p}{t'+1})\rho_p \leq \sum_{i=t+1}^{t'} \pi_i. \tag{26}$$

If we interpret $\{\rho_p\}_{p=1}^n$ as assigning probability ρ_p to value p, equation (23) states that the expected value of ρ is Π^*. Inequality (24) reflects the fact that the "inverse consumption curve" stays below the inverse demand curve. It is equivalent to saying that ρ is first-order stochastically dominated by π.

Inequality (25) follows from

$$\sum_{i=1}^{t-1} \pi_i \geq \sum_p G_p(t)\rho_p = 1 - \frac{1}{t} \sum_{p:p\leq t\leq p+\Delta} p \cdot \rho_p - \frac{1}{t} \sum_{p:p>t} t \cdot \rho_p = 1 - \frac{1}{t} \sum_{p:t\leq p+\Delta} \min\{p, t\} \cdot \rho_p,$$

The left-hand side of (25) is the total truncated revenue from sales at prices $t - \Delta$ or higher. The right-hand side is the revenue generated by selling at a uniform price of t.

We can extend the approach to allow *different consumer surplus thresholds* Δ_p and *upper bounds* α_p on the fraction of consumers in each segment p who exceed this threshold. See Section B in the Appendix.

Multi-Unit Demand

Under certain conditions, our technique can be applied to multi-unit demand setting. Let D be a set of m possible quantities a consumer may consume. The type $t \in \{1, 2, .., n\}$ of a consumer corresponds to vector $(v_t^1, .., v_t^m)$, where v_t^d is the reservation price of type t for $d \in D$ units of the good. The cost to produce d units is $c_d \geq 0$. We assume that v_t^d is strictly increasing in t for all $d \in D$. In effect, each consumer is characterized by a demand curve, and the demand curves corresponding to distinct types do not cross. As before, the prior probability of type t is π_t.

We allow a seller to target a segment using both a price and a quantity.[8] Each segment is labeled by a quantity d, price p, and fraction q of the segment that is purchased. To satisfy individual rationality of the seller, $p \geq c_d$, and the seller's revenue is $R = (p - c_d) \cdot q \geq 0$.

Denote the corresponding segment by (d, p, q). For this segment, $p = v_t^d$ for one of the types t in the segment, it is profit-maximizing for the seller, and the seller cannot earn a higher profit by offering a different quantity to this segment. These conditions will again be translated into a pair of bounding functions for each segment (d, p, q).

As in the unit demand case, the following upper bound function guarantees that the fraction of consumers with valuation less than p is at most $1 - q$.

$$F_{d,p,q}(t) = \begin{cases} 1 - q & \text{if } v_t^d < p, \\ 1 & \text{if } v_t^d \geq p. \end{cases}$$

If the seller offers quantity $d' \in D$ at price p' and q' percentage of consumers consume, the revenue will be $R' = (p' - c_{d'}) \cdot q'$. To ensure that $R' \leq R$ we impose

$$q' \leq \frac{(p - c_d) \cdot q}{(p' - c_{d'})}.$$

Thus, we will need to constrain the percentage of consumers with valuation for quantity d' at or

[8] See Haghpanah and Siegel (2022) and Haghpanah and Siegel (2023), which allows the seller to offer a menu of quantities to each segment.

above p' to be at most $\frac{(p-c_d)\cdot q}{(p'-c_{d'})}$.

To capture this constraint using a lower bound on the cumulative posterior distribution, define

$$u_t^{\max} := \max\left\{\max_{d' \in D}\left(v_t^{d'} - c_{d'}\right), 0\right\}.$$

Type t, $u_t^{\max}$ represents the highest net (of cost) social value across the m possible quantities, or zero if all net values are negative. We consider the following lower bound function:

$$G_{p,q,d}(t) = \begin{cases} 0 & \text{if } u_{t+1}^{\max} < (p - c_d)\cdot q \\ 1 - \dfrac{(p-c_d)\cdot q}{u_{t+1}^{\max}} & \text{if } u_{t+1}^{\max} \geq (p-c_d)\cdot q. \end{cases}$$

If the cumulative posterior at segment $s = (d, p, q)$ satisfies $\sum_{k=1}^{t-1} \pi_{k|s} \geq G_{p,q,d}(t-1)$, then $\sum_{k=t}^{n} \pi_{k|s} \leq (p - c_d)q/u_t^{\max}$. Hence, for any quantity d', we have

$$(v_t^{d'} - c_{d'}) \cdot \sum_{k=t}^{n} \pi_{k|s} \leq (p - c_d)q.$$

This implies that offering any other quantity d' at price p' cannot yield higher revenue than $(p-c_d)q$ for segment $s = (d, p, q)$.

Applying Theorem 1, we obtain.

COROLLARY 5.4 *The distribution probability-quantity profile $\{\rho_{d,p,q}\}_{(d,p,q)\in\mathcal{S}}$ is implementable if and only if $\forall t \in \{1, \ldots, n\}$:*

$$\sum_{i=1}^{t-1} \pi_i \geq \sum_{(d,p,q)\in\mathcal{S}} \max\{(1 - \frac{(p-c_d)q}{v_t^{max}}), 0\}\rho_{d,p,q} = 1 - \frac{1}{u_t^{max}} \sum_{(d,p,q)\in\mathcal{S}} \min\{(p-c_d)q, u_t^{max}\}\rho_{d,p,q} \tag{27}$$

$$\sum_{i=1}^{t} \pi_i \leq \sum_{(d,p,q)\in\mathcal{S}: v_t^d < p} (1-q)\rho_{d,p,q} + \sum_{(d,p,q)\in\mathcal{S}: v_t^d \geq p} \rho_{d,p,q} = 1 - \sum_{(d,p,q)\in\mathcal{S}: v_t^d < p} q\rho_{d,p,q} \tag{28}$$

$$\sum_{i=t+1}^{t'} \pi_i \geq \sum_{(d,p,q)\in\mathcal{S}: v_t^d < p \leq u_{t'+1}^{max} + c_d} (q - \frac{(p-c_d)q}{u_{t+1}^{max}}) \tag{29}$$

Corollary 5.4 also has an economic interpretation. Rewrite inequality (27) as follows:

$$\sum_{(d,p,q)\in\mathcal{S}} \min\{(p-c_d)\,q, u_t^{max}\}\,\rho_{d,p,q} \geq u_t^{\max} \cdot \sum_{i=t}^{n} \pi_i.$$

The left-hand side is the total truncated revenue at level t generated by this segmentation. The right-hand side is the revenue attainable when the seller offers a single quantity at a uniform price—that

is, employs a single segment at reservation price level t.

Inequality (28) is equivalent to:

$$\sum_{(d,p,q)\in\mathcal{S}:v_t^d<p} q\rho_{d,p,q} \leq \sum_{i=t+1}^{n} \pi_i \ \forall t \in \{1,\ldots,n\}.$$

For any $t \in \{1,\ldots,n\}$, we refer to v_t^d as the level-t reserve price for the d-unit product. The left-hand side represents the total sales of products priced above this reserve level, while the right-hand side captures the volume of customers whose type exceeds t. This inequality is analogous to the condition in the unit-demand case where the inverse consumption curve lies below the inverse demand curve. It must hold under any segmentation, since a consumer purchasing at a price above the level-t reserve price must have a type exceeding t. Corollary 5.4 establishes that, together with the revenue constraint, this condition is not only necessary, but also sufficient for the distribution ρ and the associated quantities to be implementable.

6 Conclusion

This paper proposes a general framework for information design that accommodates multiple parameters in both the posterior and prior. It offers two key advantages. First, by extending beyond single-parameter settings, the framework captures a broader range of applications, enabling its use in more complex economic and strategic scenarios. Second, despite its generality, the framework features a mathematically clean formulation that provides simpler and more intuitive economic insights.

We believe this framework holds significant potential for advancing both theoretical research and empirical investigations. On the theoretical side, it opens avenues for extensions, such as analyzing market segmentation with multiple units or exploring second-degree price discrimination. On the empirical side, it could be applied to datasets, including those related to college admissions, house rentals, or bank loan decisions.

References

Aigner, D. J. and G. G. Cain (1977): "Statistical Theories of Discrimination in Labor Markets," *Industrial and Labor Relations Review*, 30, 175–187.

Aumann, R. J., M. Maschler, and R. E. Stearns (1995): *Repeated games with incomplete information*, MIT press.

Benoît, J.-P. and J. Dubra (2011): "Apparent Overconfidence," *Econometrica*, 79, 1591–1625.

Bergemann, D., B. Brooks, and S. Morris (2015): "The Limits of Price Discrimination," *American Economic Review*, 105, 921–57.

——— (2019): "Revenue Guarantee Equivalence," *American Economic Review*, 109, 1911–1929.

Bergemann, D. and S. Morris (2016): "Information Design, Bayesian Persuasion, and Bayes Correlated Equilibrium," *American Economic Review*, 106, 586–91.

Bharadwaj, P., R. Deb, and L. Renou (2024): "Statistical Discrimination and the Distribution of Wages," NBER Working Papers 32562, National Bureau of Economic Research, Inc.

Blackwell, D. (1953): "Equivalent comparisons of experiments," *The annals of mathematical statistics*, 265–272.

Candogan, O. and H. Gurkan (2025): "The Value of Information Design in Supply Chain Management," *Management Science*, 71, 6545–6558.

Chambers, C. P. and F. Echenique (2020): "A Characterisation of 'Phelpsian' Statistical Discrimination," *The Economic Journal*, 131, 2018–2032.

Dentcheva, D. and A. Ruszczynski (2006): "Portfolio optimization with stochastic dominance constraints," *Journal of Banking and Finance*, 30, 433–451.

Elliott, M. (2021): "Market Segmentation through Information," *Review of Economic Studies*, 88, 1–29.

Farboodi, M., N. Haghpanah, and A. Shourideh (2025): "Good Data and Bad Data: The Welfare Effects of Price Discrimination," *arXiv preprint arXiv:2502.03641*, abstract in ACM EC 2025.

Frank, A. (2011): *Connections in Combinatorial Optimization*, Oxford Lecture Series in Mathematics and Its Applications, OUP Oxford.

Gentzkow, M. and E. Kamenica (2016): "A Rothschild-Stiglitz Approach to Bayesian Persuasion," *American Economic Review*, 106, 597–601.

Grushka-Cockayne, Y., K. C. Lichtendahl, V. R. R. Jose, and R. L. Winkler (2017): "Quantile Evaluation, Sensitivity to Bracketing, and Sharing Business Payoffs," *Operations Research*, 65, 712–728.

Haghpanah, N. and R. Siegel (2022): "The Limits of Multiproduct Price Discrimination," *American Economic Review: Insights*, 4, 443–58.

——— (2023): "Pareto-Improving Segmentation of Multiproduct Markets," *Journal of Political Economy*, 131, 1546– 1575.

Kamenica, E. and M. Gentzkow (2011): "Bayesian Persuasion," *American Economic Review*, 101, 2590–2615.

Kolotilin, A. and A. Wolitzky (2020): "Assortative Information Disclosure," Discussion Papers 2020-08, School of Economics, The University of New South Wales.

——— (2024): "Distributions of posterior quantiles via matching," *Theoretical Economics*, 19.

Onuchic, P. (2022): "Recent Contributions to Theories of Discrimination," *arXiv preprint arXiv:2205.05994*, revised version published June 1, 2023.

Phelps, E. S. (1972): "The Statistical Theory of Racism and Sexism," *American Economic Review*, 62, 659–661.

Porter, R. B. and J. E. Gaumnitz (1972): "Stochastic Dominance vs. Mean-Variance Portfolio Analysis: An Empirical Evaluation," *The American Economic Review*, 62, 438–446.

Romei, A. and S. Ruggieri (2014): "A Multidisciplinary Survey on Discrimination Analysis," *The Knowledge Engineering Review*, 29, 582–638.

Schrijver, A. (2003): *Combinatorial Optimization - Polyhedra and Efficiency*, Springer.

Vidal, T., D. Gribel, and P. Jaillet (2019): "Separable Convex Optimization with Nested Lower and Upper Constraints," *INFORMS Journal on Optimization*, 1, 71–90.

Vohra, A., J. Toikka, and R. Vohra (2023): "Bayesian persuasion: Reduced form approach," *Journal of Mathematical Economics*, 107, 102863.

Yang, K. H. and A. Zentefis (2022): "Distributions of Posterior Quantiles and Economic Applications," *Working Paper*.

A Omitted Proofs

A.1 Proof of Theorem 2

Proof. Constraint (15) holds trivially since $d_p = q_p \cdot \rho_p$. We will show that, together with (16) and (18), these conditions imply the implementability of the distribution $\{\rho_p, q_p\}_{p=1}^n$.

Let $x_{i,p}$ denote the mass of buyers with reservation value i assigned to segment (p, q_p). A

distribution $\{\rho_p, q_p\}_{p=1}^n$ is implementable if and only if the following system is feasible:

$$\sum_{p=1}^{n} x_{i,p} = \pi_i \qquad \forall i \in \{1,\dots,n\}, \tag{30}$$

$$\sum_{i\in\mathcal{N}} x_{i,p} = \rho_p \qquad \forall p \in \{1,..,n\}, \tag{31}$$

$$\sum_{i=t}^{n} x_{i,p} \le \frac{pq_p}{t}\rho_p \qquad \forall t \in \{1,\dots,n\},\ \forall p \in \{1,..,n\}, \tag{32}$$

$$\sum_{i=1}^{p-1} x_{i,p} \le (1-q_p)\rho_p \qquad \forall p \in \{1,..,n\}, \tag{33}$$

$$x_{i,p} \ge 0 \qquad \forall i \in \{1,\dots,n\},\ \forall p \in \{1,..,n\}. \tag{34}$$

Constraint (30) ensures that each buyer type is assigned to exactly one segment. Constraint (31) enforces that the total mass assigned to each segment (p, q_p) is ρ_p. Constraints (32) and (33) ensure that within each segment (p, q_p), the assigned distribution respects p as the revenue-maximizing price and q_p as the corresponding quantity.

Apply Theorem 1 with the following bounding functions:

$$G_p(t) = \begin{cases} 0 & \text{if } t \le pq_p - 1, \\ 1 - \frac{pq_p}{t+1} & \text{if } t > pq_p - 1, \end{cases} \qquad F_p(t) = \begin{cases} 1-q_p & \text{if } t \le p-1, \\ 1 & \text{if } t \ge p. \end{cases}$$

Applying the lower bound using G, we obtain:

$$\begin{aligned} \sum_{i=1}^{t} \pi_i &\ge \sum_{p=1}^{n} G_p(t)\rho_p \\ &= \sum_{p:pq_p<t+1} \left(1 - \frac{pq_p}{t+1}\right)\rho_p \\ &= \left(1 - \sum_{p:pq_p\ge t+1} \rho_p\right) - \frac{1}{t+1}\sum_{p:pq_p<t+1} pq_p\rho_p, \end{aligned}$$

which is equivalent to:

$$\frac{1}{t+1}\sum_{p:pq_p<t+1} pq_p\rho_p + \sum_{p:pq_p\ge t+1} \rho_p \ge \sum_{i=t+1}^{n} \pi_i.$$

With $d_p = q_p \cdot \rho_p$, this corresponds to constraint (16).

Next, applying the upper bound F, we obtain:

$$\begin{aligned}\sum_{i=1}^{t} \pi_i &\leq \sum_{p} F_p(t)\rho_p \\ &= \sum_{p:p\leq t} \rho_p + \sum_{p:p>t} (1-q_p)\rho_p \\ &= 1 - \sum_{p:p>t} q_p\rho_p.\end{aligned}$$

This gives:

$$\sum_{i=t+1}^{n} \pi_i \geq \sum_{p:p\geq t+1} q_p\rho_p,$$

which corresponds to constraint (17).

Finally

$$\max\{G_p(t') - F_p(t),\, 0\} = \begin{cases} q_p - \dfrac{pq_p}{t'+1} & \text{if } t+1 \leq p < t'+1, \\ 0 & \text{otherwise.} \end{cases}$$

Thus,

$$\sum_{i=t+1}^{t'} \pi_i \geq \sum_{p,q} \max\{G_p(t') - F_p(t),\, 0\}\rho_p = \sum_{p,q:t+1\leq p\leq t'} q_p(\rho_p - \frac{pq_p\rho_p}{t'+1}).$$

This corresponds to (18). ∎

A.2 Proof of Theorem 3

Proof. Call the desired segmentation a *consumer-optimal*. We prove that every distribution π over reservation prices admits a consumer-optimal segmentation by induction on the size of the support of π.

First, observe that if two distributions π^1 and π^2 share the same revenue maximizing price p^*, and each admits a consumer-optimal segmentation, then any convex combination $\alpha\pi^1 + (1-\alpha)\pi^2$ also admits a consumer-optimal segmentation. This follows directly from Corollary 5.1, since if $\{\rho_p^1\}$ and $\{\rho_p^2\}$ are consumer-optimal segmentations for π^1 and π^2, respectively, then the convex combination $\{\alpha\rho_p^1 + (1-\alpha)\rho_p^2\}$ is a consumer-optimal segmentation for $\alpha\pi^1 + (1-\alpha)\pi^2$.

Second, let $t^* = \min_i\{\pi_i > 0\}$. It is straightforward to verify from Corollary 5.1 that a distribution π admits a distribution of segments of the form $\rho_{t^*} = 1$, $\rho_p = 0$ for $p > t^*$, if and only if

$$t \cdot \sum_{i\geq t} \pi_i \leq t^* \quad \text{for all } t.$$

Let $p^* \in \arg\max_p p\sum_{i\geq p}\pi_i$. If $p^* = t^*$, then the segmentation $\rho_{t^*} = 1$, $\rho_p = 0$ is consumer-

optimal. If $p^* > t^*$, we can decompose $\pi = \gamma^1 + \gamma^2$ as follows:

$$\gamma_i^1 = \begin{cases} (1-\alpha)\cdot\pi_i & \text{if } i > t^*, \\ 0 & \text{if } i \le t^*, \end{cases} \qquad \gamma_i^2 = \begin{cases} \alpha\cdot\pi_i & \text{if } i > t^*, \\ \pi_{t^*} & \text{if } i = t^*, \\ 0 & \text{if } i < t^*. \end{cases}$$

Here, α is chosen so that $\alpha p^* \cdot \sum_{i\ge p^*}\pi_i = t^*(\pi_{t^*} + \alpha\sum_{i>t^*}\pi_i)$.

It is easy to verify that the distributions $\pi^j = \frac{1}{\sum_i \gamma_i^j}\gamma^j$ for $j = 1, 2$ support the same uniform monopoly price p^*. Moreover, the distribution π^2 admits a consumer-optimal segmentation with $\rho_{t^*} = 1$, while π^1, having smaller support than π, admits a consumer-optimal segmentation by the induction hypothesis. It follows that π, which is a convex combination of π^1 and π^2 also admits a consumer-optimal segmentation. ∎

A.3 Proof of Corollary 5.2

Proof. Our goal is to show the following hold:

- The segmentation matches the given inverse consumption curve, i.e., $d_{p_i} = q_{p_i}\cdot\rho_{p_i}$ and $\sum_{i=1}^k \rho_{p_i} = 1$. This ensures that inequality (15) and (17) of Theorem 2 hold.
- The collection $\{(p_i, q_{p_i}, \rho_{p_i})\}_{i=1}^k$ satisfies inequality (18) and (16) of Theorem 2.

Notice that

$$\rho_{p_i} := \frac{p_i d_{p_i}}{\Pi^*} \quad \text{and} \quad q_{p_i} := \frac{\Pi^*}{p_i}.$$

Thus, by construction, the total area under the vertical consumption curve is exactly Π^*, so

$$\sum_{i=1}^k \rho_{p_i} = \frac{\sum_{i=1}^k p_i d_{p_i}}{\Pi^*} = 1.$$

To verify that $q_{p_i} \le 1$ for all i, let $p^* := \arg\max_p \left\{p\cdot\sum_{i=p}^n \pi_i\right\}$, so that $\Pi^* = p^*\cdot\sum_{i=p^*}^n\pi_i$. Since the area of the vertical consumption curve is exactly Π^*, $p_i \ge p^*$ for all i, it follows that

$$q_{p_i} = \frac{\Pi^*}{p_i} = \frac{p^*\cdot\sum_{i=p^*}^n}{p_i} \le \sum_{i=p^*}^n \pi_i \le 1.$$

Since $q_{p_i}\cdot\rho_{p_i} = d_{p_i}$, the resulting consumption curve constructed from $\{\rho_{p_i}, d_{p_i}\}_{i=1}^k$ exactly matches the vertical consumption curve and lies below the inverse demand curve. Hence, inequality (17) is satisfied.

To check inequality (18), we must verify that for all $t = 1, \ldots, n$,

$$\sum_{j=t}^n \pi_j - \sum_{(p_i,q_{p_i}):p_i q_{p_i} > t} \rho_{p_i} \le \sum_{(p_i,q_{p_i}):p_i q_{p_i}\le t} \frac{p_i q_{p_i}}{t}\cdot\rho_{p_i}.$$

Since $p_i \cdot q_{p_i} = \Pi^*$ for all i, we only need to consider two cases:

- If $p_i q_{p_i} > t$ for all i, then $\sum_{(p_i,q_{p_i}):p_i q_{p_i}>t} \rho_{p_i} = 1$, so the left-hand side becomes $\sum_{j=t}^{n} \pi_j - 1 \leq 0$, and the inequality trivially holds.
- If $p_i q_{p_i} \leq t$ for all i, then the inequality becomes $\sum_{j=t}^{n} \pi_j \leq \sum_{i=1}^{k} \frac{p_i q_{p_i}}{t} \cdot \rho_{p_i}$. This holds because

$$\sum_{i=1}^{k} p_i q_{p_i} = \Pi^* = \max_j \left\{ j \cdot \sum_{j=t}^{n} \pi_j \right\}.$$

Finally, condition (16) holds trivially because, along the vertical consumption curve, the total mass of consumers consuming at price t is at most π_t for each type p.

Therefore, all conditions of Theorem 2 are satisfied, and the probability-quantity profile is implementable. ∎

B Surplus Distribution

We can extend Corollary 5.3 to allow *different consumer surplus thresholds* Δ_p and *upper bounds* α_p on the fraction of consumers in each segment p who exceed this threshold. A segment is now a tuple (p, Δ_p, α_p), where Δ_p is the surplus threshold and α_p is the allowed fraction exceeding it. The earlier special case corresponds to $\alpha_p = 0$. Since price p must be revenue-maximizing for the seller, we require:

$$\alpha_p \leq \frac{p}{p + \Delta_p}.$$

With this, it is straightforward to see that the following bounding functions can be used to capture the constraint. In particular, for a segment $s = (p, \Delta_p, \alpha_p)$, the bounding functions $(F_s^*(\cdot), G_s^*(\cdot))$ are defined as follows:

$$G_s^*(t) = \begin{cases} 0 & \text{for } t < p, \\ 1 - \frac{p}{t+1}, & \text{for } t < p + \Delta_p, \\ 1 - \alpha_p & \text{for } t \geq p + \Delta_p. \end{cases} \qquad F_r^*(t) = \begin{cases} 0 & \text{if } t < p, \\ 1 - \alpha_p & \text{if } p \leq t \leq p + \Delta_p \\ 1 & \text{if } t > p + \Delta_p \end{cases}$$

C An algorithm for determining the thresholds

We show that determining whether there exist thresholds p_k that satisfy the upper- and lower-bound conditions from Section 4, while remaining consistent with the tier structure, can be accomplished using a simple greedy algorithm.

The idea is straightforward: for each tier T_k, calculate the tightest possible lower and upper bounds on the threshold p_k implied by the constraints. Then check, tier by tier, whether there is a threshold that lies within these bounds and also respects the requirement that thresholds increase

strictly across tiers. Starting from the lowest tier, assign the smallest feasible threshold, and then proceed upward, updating the bounds to account for the already chosen thresholds. If at any point the feasible range becomes empty, we conclude that no set of thresholds is consistent with both the constraints and the tier structure.

Formally, for each tier T_k:

- The *lower bound* for the threshold score p_k is

$$l_k := \min\, t \text{ s.t. } \sum_{i=1}^{t} \pi_i^g \;\geq \sum_{s \in T_1 \cup \cdots \cup T_k} (1 - \alpha_s^g)\, \rho_s^g \text{ for all } g \in \mathcal{G}. \tag{35}$$

- The *upper bound* for p_k is

$$u_k := \max\, t \text{ s.t. } \sum_{i=t+1}^{n} \pi_i^g \;\geq \sum_{s \in T_k \cup \cdots \cup T_m} \alpha_s^g\, \rho_s^g \text{ for all } g \in \mathcal{G}. \tag{36}$$

The algorithm implementing this procedure is given below.

Algorithm 1: Greedy Threshold Selection

Input: Tier structure $T_1, \ldots, T_m$; parameters $\pi_i^g, \alpha_s^g, \rho_s^g$

Output: Thresholds $p_1, \ldots, p_m$ or infeasibility conclusion

for $k \leftarrow 1$ ***to*** m **do**

  Compute l_k and u_k as in (35) and (36), respectively.

  if $k = 1$ **then**

    if $l_1 > u_1$ **then**

      return "No feasible thresholds – data not rationalizable"

    Set $p_1 \leftarrow l_1$

  else

    Set lower_candidate $\leftarrow \max\{p_{k-1} + 1,\ l_k\}$;

    if *lower_candidate* $\leq u_k$ **then**

      Set $p_k \leftarrow$ lower_candidate

    else

      return "No feasible thresholds – data not rationalizable"

return $p_1, p_2, \ldots, p_m$