Distributions of Posterior Quantiles and Economic Applications*

Kai Hao Yang† Alexander K. Zentefis‡

April 30, 2022

Abstract

We characterize the distributions of posterior quantiles under a given prior. Unlike the distributions of posterior means, which are known to be mean-preserving contractions of the prior, the distributions of posterior quantiles reside in a first-order-stochastic-dominance interval bounded by an upper and a lower truncation of the prior. We apply this characterization to several environments, ranging from political economy, Bayesian persuasion, industrial organization, econometrics, finance, and accounting.

JEL classification: C31, D72, D82, D83, G28, L12, M41

Keywords: Posterior quantiles, stochastic dominance, gerrymandering, Bayesian persuasion, two-sided market, quantile regression.

*This paper subsumes a previously circulated paper with the title "Gerrymandering and the Limits of Representative Democracy." We are grateful to Yuehao Bai, Peter Caradonna, Kuan-ming Chen, Xiaohong Chen, Yu-ting Chiang, Joyee Deb, Tan Gan, Paul Goldsmith-Pinkham, Marina Halac, Ryota Iijima, Alexey Kushnir, Leslie Marx, Elliot Lipnowski, Aniko Öry, Doron Ravid, Anne-Katrin Roesler, Hamid Sabourian, Chris Shannon, Quitzé Valenzuela-Stookey, Vasiliki Skreta, Roland Strausz, Philipp Strack, Alexander Wolitzky, Wenji Xu, Nathan Yoder, and Jidong Zhou for their helpful comments. All errors are our own.

†Yale School of Management, Email: kaihao.yang@yale.edu

‡Yale School of Management, Email: alexander.zentefis@yale.edu

1 Introduction

A political body is redrawing the boundaries of electoral districts for partisan gain. Election results are dictated by the median voter theorem. To what extent can this gerrymandering affect the composition of the legislature?

A ride-sharing app is a platform between riders and drivers, and it can segment both sides of the market. An inelastic supply must be held fixed at 75% of the number of riders in each segment, so that the wait time remains approximately the same across riders. This means that, in each segment, the top 75th percentile of riders' willingness-to-pay determines the price. What is the optimal segmentation to maximize the platform's revenue?

An econometrician observes data on income and education from two different samples of the population. What can she infer about the relation between the top 1% of earners and their years of schooling?

Despite these scenarios roaming varied economic fields; reaching the areas of political economy, industrial organization, and econometrics; all are connected by their shared concern over the distribution of different posterior quantiles. This paper characterizes the distributions of posterior quantiles in a general setting, and it answers each of the scenario's questions and more.

In our environment, a one-dimensional variable $\omega \in \mathbb{R}$ follows a prior distribution F_0. Given any signal for ω (i.e., a joint distribution of ω and signal realizations with the marginal of ω being F_0), each signal realization induces a *posterior* distribution of ω via Bayes' rule. Therefore, any signal induces a distribution of posterior beliefs. Suppose that, for each posterior, one computes the mean. Strassen's theorem (Strassen 1965) implies that the distribution of posterior means must be a mean-preserving contraction of F_0. At the same time, any mean-preserving contraction of F_0 is a distribution of posterior means induced by some signal.

Instead of posterior means, one can derive many other statistics of a posterior. Suppose that, for each posterior, instead of the mean, one computes a τ-quantile. A natural question then follows: What is the distribution of posterior quantiles?

Theorem 1 provides that characterization. Two distributions are important in this regard:

$$\underline{F}_0^\tau(\omega) := \min\left\{\frac{1}{\tau}F_0(\omega), 1\right\}, \quad \overline{F}_0^\tau(\omega) := \max\left\{\frac{F_0(\omega)-\tau}{1-\tau}, 0\right\}.$$

The distribution $\underline{F}_0^\tau$ can be interpreted as the conditional distribution of F_0 in the event that ω is smaller than a τ-quantile of F_0. Similarly, $\overline{F}_0^\tau$ can be interpreted as the conditional distribution of F_0 in the event that ω is larger than the same τ-quantile. Theorem 1 states that any distribution of posterior τ-quantiles induced by a signal must be bounded by $\underline{F}_0^\tau$ and $\overline{F}_0^\tau$ in the sense of first-order stochastic dominance. In the meantime, any distribution bounded by $\underline{F}_0^\tau$ and $\overline{F}_0^\tau$ in the sense of first-order stochastic dominance must be a distribution of posterior τ-quantiles under some signal.

With the characterization of the distributions of posterior quantiles in hand, we then apply it to several economic setting. The first is to gerrymandering, or the manipulation of electoral district boundaries. In this setting, citizens identify with an ideal position on political issues along a spectrum. The variety of the citizenry's positions is represented as a distribution, which we can call a prior. An electoral map segments citizens into districts, which splits the prior distribution of into different parts. The distribution of ideal positions within each district can be interpreted as a posterior.

Assuming each district elects a representative holding the district's median position, the composition of the legislative body (i.e., the distribution of ideal positions of elected representatives) can then be interpreted as a distribution of posterior medians.[1] In this regard, Theorem 1 fully describes the scope of legislatures that unrestrained gerrymandering can achieve.[2] According to Theorem 1, gerrymandering can induce *any* legislature within the bounds of two extremes: an "all-left" body and an "all-right" body. In the former, every representative occupying the legislature has an ideal position that is left of the median voter's ideal, whereas in the latter, every representative is to the right. These two bounds imply that unrestricted gerrymandering, in the most extreme scenario, can lead half the population to lose all representation of their views.

With knowledge about the composition of the legislative body, we then study the effects of gerrymandering on enacted legislation. If we further assume that the enacted legislation must be a median of the legislative body, an immediate consequence of Theorem 1 is that any position between the interquartile range under the prior can be enacted by the legislature under some map. As a result, the set of legislation that can be enacted expands as the population becomes more polarized.

As every citizen has single-peaked preferences, if a referendum were held among the voters over the choice of legislative policy, a Condercet winner—the outcome that has majority support when compared to any other alternative—would always exist and coincide with the median policy. Another consequence of Theorem 1 is that we can derive the set of legislative outcomes that can defeat the population median by majority support in the congress under some map. In fact, under any polarized distribution of the citizenry's positions, *any*

[1] Political economy theory and empirical evidence suggest that the median voter property applies well to many election results (Downs 1957; Black 1958; Congleton 2004).

[2] In Section 4, we connect the splitting of the prior distribution of citizens' ideal political positions to the splitting of geographic areas on a two-dimensional surface, as gerrymandering is conducted in practice

outcome can potentially defeat the median by simple majority in the congress under some map, arguably reversing the median's Condercet property.

In addition to gerrymandering, we apply our characterization to Bayesian persuasion. Kamenica and Gentzkow (2011) provide a framework for studying communication under the commitment assumption. A practical challenge, however, is that the concavafication approach loses tractability as the number of states increases. An exception is when the state is one-dimensional and only posterior means are payoff-relevant for the sender. Theorem 1 complements this literature, as it expands its range of applications by bringing tractability to settings where only posterior *quantiles* are payoff-relevant for the sender. We demonstrate this by revisiting the two examples of Kamenica and Gentzkow (2011).

Bayesian persuasion has notably been applied to industrial organization settings involving market segmentation, in which a market is split into several segments to further price discrimination (e.g., Bergemann, Brooks, and Morris 2015; Ichihashi 2020). In our next application, we use the characterization to derive the optimal segmentation of a two-sided market. When markets are of this kind, the problem enlarges to finding not only the best way to split markets, but also the optimal way to match sides (see, e.g., Hagiu and Jullien 2011; Condorelli and Szentes 2022; Guinsburg and Saraiva 2022).

In our example, we consider a two-sided market (e.g., ride-sharing). The demand side is populated by riders with unit demand; whereas the supply side is populated by drivers who have inelastic supply, which is plausible during peak hours at a major airport or at the conclusion of a large sporting event. A third-party platform (the ride-sharing app) can segment the market to affect prices, and it retains a fixed share of sales revenue per segment. If the platform could arbitrarily segment both sides of the market, it would optimally segment each rider and driver, then match demand with supply in an assortative manner until the supply is exhausted.

But in practice, the platform might be obligated to keep the ratio of supply to demand of each segment the same (i.e., it faces a *market thickness* constraint), so that all riders would wait approximately the same time before matching with a driver. Theorem 1 provides a solution to an optimal segmentation problem with a fixed thickness constraint. Perhaps surprisingly, one can show that the platform's total revenue under optimal pricing and segmentation rules exactly matches that under first-degree price discrimination, thus rendering any thickness constraints irrelevant.

Because the distribution of posterior quantiles is simply a conditional distribution, another natural discipline ready for applications is econometrics. We apply the characterization to quantile regression, which models the quantiles of the conditional distribution of a response variable Y as a function of covariates X.

To facilitate the analysis and maintain tractability, econometricians often impose parametric assumptions regarding the quantile function, such as presuming linearity. Theorem 1 provides a model mis-specification test that relies only on the marginal of Y. The reliance on information from just the marginal allows one to bypass estimation of the joint distribution of (Y, X), which may be computationally demanding.

When the number of covariates equals to 1 and the marginals of both X and Y are known, an econometrician can go beyond evaluating model mis-specification to partially identifying the quantile function. Taking a concrete example, one might have Y representing income and X standing for education, and the two sets of data come from non-overlapping groups of people. If the quantile function is known to be increasing (such as income rising with years of schooling), we show how Theorem 1 provides a non-parametric partial identification of the quantile function.

Our final applications are to topics in finance and accounting. For the financial application, we take a setting in which a bank regulator considers requiring systemically important financial institutions to carry equity capital in accordance with their contribution to the Value-at-Risk of the financial system. Our characterization can pinpoint a robust regulatory policy that would guard against the greatest losses a financial system could be estimated to face. For accounting, we consider an auditor who worries about a manager misclassifying expenses to boost earnings. A certain dollar amount of misclassification constitutes material fraud. We provide a necessary condition for an auditor to engage in a closer inspection of the reported expenses.

Related Literature. This paper is related to several streams of literature. Belief-based characterizations of signals date back to seminal contributions of Blackwell (1953) and Harsanyi (1967-68). The characterization of distributions of posterior means can be derived from Strassen (1965), and their economic applications have been made clear by Rothschild and Stiglitz (1970). This paper can be regarded as a complement, which characterizes the distributions of posterior quantiles, instead of means. In addition, the characterization relies on identifying extreme points of a first-order stochastic dominance interval. Extreme points of orbits under the majorization order (and, hence, of second-order stochastic dominance intervals) have been studied since Hardy, Littlewood, and Pólya (1929), who examine finite-dimensional spaces. Ryff (1967) extends this result to an infinite dimensional space. Kleiner, Moldovanu, and Strack (2021) characterize the extreme points of a subset of orbits under an additional monotonicity assumption, which in turn leads to many economic applications. Extreme points of first-order stochastic dominance intervals exhibit a similar structure—albeit easier to characterize—in the sense that either the stochastic dominance constraints bind (on an interval) or there are at most finitely many mass points.

In terms of applications, our gerrymandering results are related to the literature on redistricting. Among the closest are Owen and Grofman (1988), Friedman and Holden (2008), Gul and Pesendorfer (2010), and Kolotilin and Wolitzky (2020), who adopt the same distribution-based approach and model a district map as a way to split the population distribution of voters. The main focus of these papers is finding the optimal gerrymandering for a political party who maximizes its expected number of seats. In contrast, our result characterizes the feasible compositions of a legislative body that a district map can induce.[3]

Our second application is related to the Bayesian persuasion literature. Based on the fundamental principles outlined by Kamenica and Gentzkow (2011), Gentzkow and Kamenica (2016) specialize preferences so that only the posterior means are payoff-relevant. Dworczak and Martini (2019) further generalize the results and provide a characterization of a sender's optimal signals for a general class of mean-based persuasion problems. Kleiner, Moldovanu, and Strack (2021) characterize the extreme points of the feasible set of this convex problem. We complement this literature by providing a foundation for solving persuasion problems where only the posterior quantiles are payoff-relevant.

In the meantime, the application of market segmentation to a two-sided market has features from both one-sided market segmentation and price discrimination (e.g., Bergemann, Brooks, and Morris 2015; Haghpanah and Siegel 2020, 2022; Yang 2022; Elliot, Galeotti, Koh, and Li 2022) and matching in a two-sided market (e.g., Zhao, Zhang, Khan, and Perrussel 2010; Hagiu and Jullien 2011; de Cornière 2016; Condorelli and Szentes 2022; Guinsburg and Saraiva 2022).

Finally, the econometric applications are related to problems of inferring the joint distribution from marginals, as studied by Horowitz and Manski (1995) and Cross and Manski (2002); the finance applications are related to the conditional Value-at-Risk measurement of systemic risk introduced by Adrian and Brunnermeier (2016); and the accounting application relates to classification shifting behavior identified in McVay (2006).

Outline. The remainder of the paper proceeds as follows. Section 2 establishes the general environment. Section 3 gives the paper's main result. Economic applications follow in Section 4 (gerrymandering), Section 5 (Bayesian persuasion and market segmentation), Section 6 (econometrics), and Section 7 (finance and accounting). Section 8 concludes.

[3] It is noteworthy that proposition 1 of Gomberg, Pancs, and Sharma (2021) shares the same flavor. However, the authors assume that each district elects a *mean* candidate as opposed to the median, and, hence, their characterization follows from Blackwell's theorem and properties of the majorization order, which do not apply to our setting.

2 Preliminaries

State and Signals Consider a one-dimensional variable $\omega \in \mathbb{R}$. Let $F_0 \in \mathcal{F}$ be the distribution of ω, where $\mathcal{F}$ denotes the collection of distribution functions on $\mathbb{R}$.[4] A particular distribution of interest is the uniform distribution on $[0,1]$, which is denoted by $U \in \mathcal{F}$. Namely, $U(\omega) = \omega$ for all $\omega \in [0,1]$. A *signal* for ω is defined as $\mu \in \Delta(\mathcal{F})$ such that

$$\int_{\mathcal{F}} F(\omega)\mu(\mathrm{d}F) = F_0(\omega), \tag{1}$$

for all $\omega \in \mathbb{R}$. Let $\mathcal{M}(F_0)$ denote the collection of all signals (under prior distribution F_0). For the ease of exposition, we sometimes write $\mathcal{M}$ instead of $\mathcal{M}(F_0)$ when there is no confusion. From Blackwell's theorem (Blackwell 1953), given any $\mu \in \mathcal{M}(F_0)$, each $F \in \text{supp}(\mu)$ can be interpreted as a *posterior* for ω obtained via Bayes' rule under a prior F_0, after observing the realization of a signal that is correlated with ω; and the marginal distribution of this signal is summarized by μ.

Quantiles and Quantile Selection Rules For any distribution $F \in \mathcal{F}$, let the quantile function F^{-1} be defined as

$$F^{-1}(\tau) := \inf\{\omega \in \mathbb{R} | F(\omega) \geq \tau\},$$

for all $\tau \in [0,1]$.[5] Denote the set of τ-quantiles of F by $\mathbb{Q}^\tau(F) := [F^{-1}(\tau), F^{-1}(\tau^+)]$. Furthermore, we say that a transition probability $r : \mathcal{F} \times [0,1] \to \Delta(\mathbb{R})$ is a *quantile selection rule* if $\text{supp}(r(\cdot|F,\tau)) \subseteq \mathbb{Q}^\tau(F)$ for all $F \in \mathcal{F}$ and for all $\tau \in [0,1]$. A quantile selection rule r selects (possibly through randomization) a τ-quantile for every CDF F and for every $\tau \in [0,1]$, whenever it is not unique. Let $\mathcal{R}$ be the collection of all selection rules.

Distributions of Posterior Quantiles For any $\tau \in [0,1]$, for any signal $\mu \in \mathcal{M}$, and for any selection rule $r \in \mathcal{R}$, let $H^\tau(\cdot|\mu,r)$ denote the distribution of the τ-quantile induced by μ and r. That is,

$$H^\tau(\omega|\mu,r) = \int_{\mathcal{F}} r((-\infty,\omega]|F,\tau)\mu(\mathrm{d}F), \tag{2}$$

for all $\omega \in \mathbb{R}$. Our main result characterizes the distributions of posterior quantiles induced by arbitrary signals and selection rules.

[4] $\mathcal{F}$ is endowed with the weak-* topology and the induced Borel σ-algebra.

[5] Note that F^{-1} is nondecreasing and left-continuous for all $F \in \mathcal{F}$. Moreover, for any $\tau \in [0,1]$ and for any $\omega \in \mathbb{R}$, $F^{-1}(\tau) \leq \omega$ if and only if $F(\omega) \geq \tau$.

Stochastic Dominance Interval Given any $F, F' \in \mathcal{F}$, recall that F *dominates* F' in the sense of first-order stochastic dominance, denoted by $F \succeq F'$ henceforth, if $F(\omega) \leq F'(\omega)$ for all $\omega \in \mathbb{R}$. For any $F, F' \in \mathcal{F}$ such that $F \succeq F'$, denote the set of CDFs that dominate F' and are dominated by F as $\mathcal{I}(F', F)$. That is,

$$\mathcal{I}(F', F) := \{H \in \mathcal{F} | F' \preceq H \preceq F\}.$$

3 Characterization of Distributions of Posterior Quantiles

Our main result characterizes the collection of all possible distributions of posterior quantiles. To state this result, for any $\tau \in [0, 1]$, let $\mathcal{H}_\tau$ denote the set of distributions that can be induced by some signal $\mu \in \mathcal{M}$ and selection $r \in \mathcal{R}$. Namely,

$$\mathcal{H}_\tau := \{H \in \mathcal{F} | H(\omega) = H^\tau(\omega|\mu, r), \forall \omega \in \mathbb{R}, \text{ for some } \mu \in \mathcal{M}, r \in \mathcal{R}\}.$$

In the meantime, define two distributions $\underline{F}_0^\tau$ and $\overline{F}_0^\tau$ as follows:[6]

$$\underline{F}_0^\tau(\omega) := \min\left\{\frac{1}{\tau}F_0(\omega), 1\right\}, \quad \overline{F}_0^\tau(\omega) := \max\left\{\frac{F_0(\omega) - \tau}{1 - \tau}, 0\right\}.$$

Note that $\overline{F}_0^\tau \succeq \underline{F}_0^\tau$ for all $\tau \in [0, 1]$. In essence, $\underline{F}_0^\tau$ is the conditional distribution of F_0 in the event that ω is smaller than a τ-quantile of F_0; whereas $\overline{F}_0^\tau$ is the conditional distribution of F_0 in the event that ω is larger than the same τ-quantile. This brings us to our main result.

Theorem 1 (Distributions of Posterior Quantiles)**.** *For any $\tau \in [0, 1]$,*

$$\mathcal{H}_\tau = \mathcal{I}(\underline{F}_0^\tau, \overline{F}_0^\tau).$$

Theorem 1 completely characterizes the distributions of posterior τ-quantiles by the stochastic dominance interval $\mathcal{I}(\underline{F}_0^\tau, \overline{F}_0^\tau)$. Figure I illustrates Theorem 1 for the case when $\tau = 1/2$. The distribution $\underline{F}_0^{1/2}$ is colored blue, whereas the distribution $\overline{F}_0^{1/2}$ is colored red. The green dotted curve represents the prior, F_0. According to Theorem 1, any distribution H bounded by $\underline{F}_0^{1/2}$ and $\overline{F}_0^{1/2}$ (for instance, the black curve in the figure) can be induced by a signal $\mu \in \mathcal{M}$ and a select rule $r \in \mathcal{R}$. Conversely, for any signal and for any selection rule, the induced graph of the distribution of posterior τ-quantiles must fall in the area bounded by the blue and the red curves.

In what follows, we explain the main steps for the proof of Theorem 1. Details of the

[6]For the case of $\tau \in \{0, 1\}$, define $\underline{F}_0^\tau$, $\overline{F}_0^\tau$ as the pointwise limit as $\tau \downarrow 0$ and $\tau \uparrow 1$, respectively.

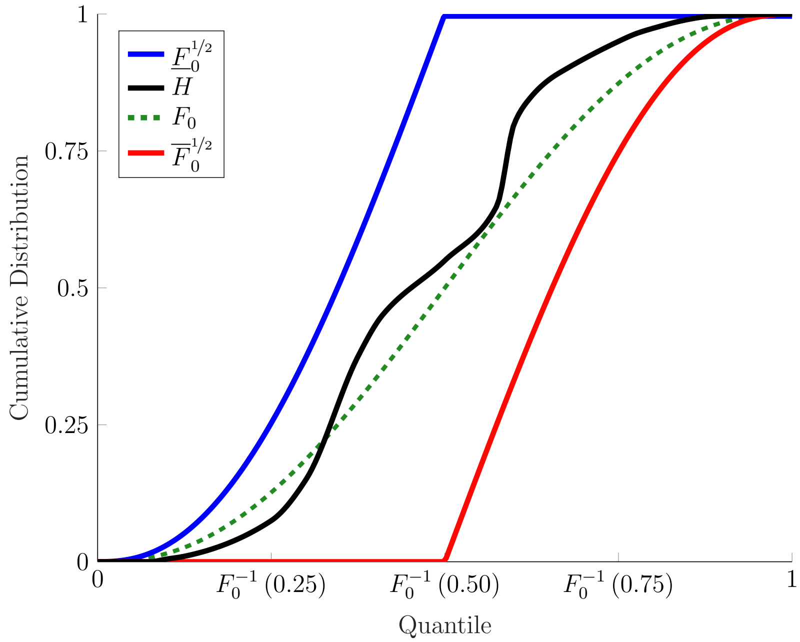


Figure I
Stochastic Dominance Interval

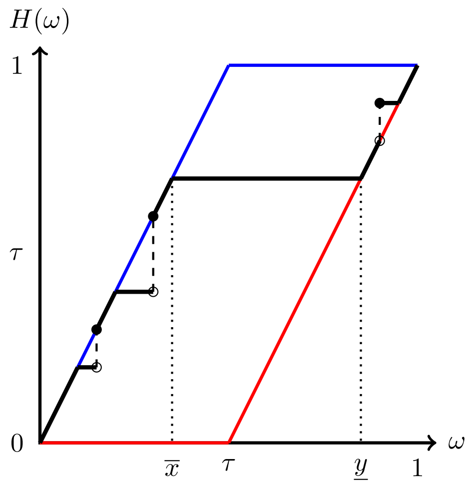
proof can be found in the appendix. To begin with, note that for any signal $\mu \in \mathcal{M}$ and for any $r \in \mathcal{R}$,

$$H^{\tau}(\omega|\mu, r) \leq \mu(\{F \in \mathcal{F}|F^{-1}(\tau) \leq \omega\}) = \mu(\{F \in \mathcal{F}|F(\omega) \geq \tau\}),$$

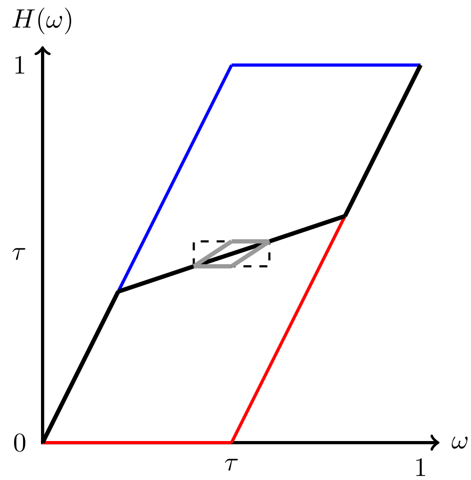
for all $\omega \in \mathbb{R}$, where the first inequality holds because the right-hand side corresponds to the distribution of posterior quantiles induced by μ when the lowest τ-quantile is selected with probability 1. Furthermore, for any $\omega \in \mathbb{R}$, if we regard $F(\omega) \in [0,1]$ as a random variable whose distribution is implied by μ, it then follows from (1) that its distribution must be a mean-preserving spread of $F_0(\omega)$. As a result, $\mu(\{F \in \mathcal{F}|F(\omega) \geq \tau\})$ can at most be $\min\{F_0(\omega)/\tau, 1\}$, since otherwise the mean of $F(\omega)$ can never be $F_0(\omega)$. This implies that $H^{\tau}(\cdot|\mu, r) \preceq \underline{F}_0^{\tau}$. A similar argument leads to the conclusion that $H^{\tau}(\cdot|\mu, r) \succeq \overline{F}_0^{\tau}$ as well. Thus, $\mathcal{H}_{\tau} \subseteq \mathcal{I}(\underline{F}_0^{\tau}, \overline{F}_0^{\tau})$.

The converse of the proof is relatively more involved. To begin with, since we are only interested in quantiles, it is without loss to assume that $F_0 = U$. Specifically, for any $\tau \in [0,1]$ and for any $\omega \in [0,1]$, let

$$\underline{U}^{\tau}(\omega) := \min\left\{\frac{\omega}{\tau}, 1\right\} \text{ and } \overline{U}^{\tau}(\omega) := \max\left\{\frac{\omega - \tau}{1 - \tau}, 0\right\},$$



(A) An Extreme Point



(B) Not an Extreme Point

Figure II
EXTREME POINTS OF $\mathcal{I}^*_\tau$

and let $\mathcal{I}^*_\tau := \mathcal{I}(\underline{U}^\tau, \overline{U}^\tau)$. In addition, let $\mathcal{H}^*_\tau$ be the collection of $H \in \mathcal{F}$ such that $H(\omega) = H^\tau(\omega|\tilde{\mu}, \tilde{r})$ for some $\tilde{\mu} \in \mathcal{M}(U)$ and $\tilde{r} \in \mathcal{R}$ for all $\omega \in \mathbb{R}$. Then, we have the following lemma:

Lemma 1. *Consider any $\tau \in [0,1]$. Then $\mathcal{H}^*_\tau \supseteq \mathcal{I}^*_\tau$ if and only if $\mathcal{H}_\tau \supseteq \mathcal{I}(\underline{F}^\tau_0, \overline{F}^\tau_0)$ for all $F_0 \in \mathcal{F}$.*

By Lemma 1, it suffices to show that $\mathcal{I}^*_\tau \subseteq \mathcal{H}^*_\tau$. To this end, we first characterize the extreme points of $\mathcal{I}^*_\tau$. This is stated by Lemma 2 below.

Lemma 2. *For any $\tau \in [0,1]$, H is an extreme point of $\mathcal{I}^*_\tau$ if and only if there exists $0 \leq \underline{x} \leq \overline{x} \leq \tau \leq \underline{y} \leq \overline{y}$; countable sets I, J; and sequences $\{\underline{x}_i, \overline{x}_i\}_{i\in I}, \{\underline{y}_j, \overline{y}_j\}_{j\in J} \subseteq \mathbb{R}$ such that $\underline{U}^\tau(\overline{x}) = \overline{U}^\tau(\underline{y})$; that $\underline{x} \leq \underline{x}_i \leq \overline{x}_i \leq \underline{x}_{i+1} \leq \overline{x} < \underline{y} \leq \underline{y}_j \leq \overline{y}_j \leq \underline{y}_{j+1} \leq \overline{y}$ for all $i \in I, j \in J$; and that*

$$H(\omega) = \begin{cases} \underline{U}^\tau(\underline{x}_i), & \text{if } \omega \in [\underline{x}_i, \overline{x}_i) \\ \underline{U}^\tau(\omega), & \text{if } \omega \in [\underline{x}, \overline{x}) \setminus \cup_{i\in I} [\underline{x}_i, \overline{x}_i) \\ \underline{U}^\tau(\overline{x}), & \text{if } \omega \in [\overline{x}, \underline{y}) \\ \overline{U}^\tau(\underline{y}_j), & \text{if } \omega \in [\underline{y}_j, \overline{y}_j) \\ \overline{U}^\tau(\omega), & \text{if } \omega \in [\underline{y}, \overline{y}) \setminus \cup_{j\in J} [\underline{y}_j, \overline{y}_j) \\ 1, & \text{if } \omega \geq \overline{y} \end{cases}, \tag{3}$$

for all $\omega \in \mathbb{R}$.

Figure IIA illustrates an extreme point of $\mathcal{I}^*_\tau$. According to Lemma 2, an extreme point H of $\mathcal{I}^*_\tau$ must have four cutoffs, $\underline{x}, \overline{x}$ and $\underline{y}, \overline{y}$, such that $\text{supp}(H) \subseteq [\underline{x}, \overline{x}] \cup [\underline{y}, \overline{y}]$. Moreover, on $[\underline{x}, \overline{x}]$, H must be distributed so that either H coincides with $\underline{U}^\tau$, or is constant over an interval. Similarly, on $[\underline{y}, \overline{y}]$, the left-limit of H must either coincide with $\overline{U}^\tau$, or is constant over an interval.

The main idea behind the proof is that, for any $H \in \mathcal{I}_\tau^*$ that does not exhibit this structure—as depicted in Figure IIB—there must exist a rectangle in the interior of the graph of $\mathcal{I}_\tau^*$ such that H is not a step function when restricted to that rectangle. Hence, in that rectangle, H can be split into two distinct nondecreasing functions, as depicted by the gray-scaled curves in Figure IIB. This, in turn, implies that H can be split into two distinct nondecreasing functions in $\mathcal{I}_\tau^*$, and hence, H is not an extreme point of $\mathcal{I}_\tau^*$.

Having characterized the extreme points of $\mathcal{I}_\tau^*$, we then show that for any extreme point H, there exists a signal and a selection rule such that the induced distribution of posterior τ-quantiles coincides with H. Details of the construction can be found in the appendix. The main intuition can be better understood by constructing a signal and a selection rule that attains the boundary $\overline{U}^\tau$. To this end, for any $\omega \in [\tau, 1]$, define a distribution U^ω:

$$U^\omega(x) := \begin{cases} 0, & \text{if } x < 0, \\ x, & \text{if } x \in [0, \tau) \\ \tau, & \text{if } x \in [\tau, \omega) \\ 1, & \text{if } x \geq \omega \end{cases},$$

for all $x \in \mathbb{R}$. Figure III illustrates U^ω for some $\omega \in [\tau, 1]$. Note that, by construction, $\omega = \max(\mathbb{Q}^\tau(U^\omega))$ for all $\omega \in [\tau, 1]$. Furthermore, let $\overline{\mu}$ be defined as

$$\overline{\mu}(\{U^\omega | \omega \leq y\}) := \frac{y - \tau}{1 - \tau},$$

for all $y \in [\tau, 1]$. It then follows that $\overline{\mu} \in \mathcal{M}(U)$. Together with the selection rule $\overline{r} \in \mathcal{R}$ that always selects the largest τ-quantile, it follows that $H^\tau(\omega|\overline{\mu}, \overline{r}) = \overline{U}^\tau(\omega)$ for all $\omega \in \mathbb{R}$. Finally, since any $\widetilde{H} \in \mathcal{I}_\tau^*$ can be represented as a mixture of the extreme points of $\mathcal{I}_\tau^*$, and since $(\mu, r) \mapsto H^\tau(\cdot|\mu, r)$ is affine, it then follows that $\mathcal{I}_\tau^* \subseteq \mathcal{H}_\tau^*$.

An immediate corollary of Theorem 1 characterizes the set of τ-quantiles of all distributions of posterior τ-quantiles. The corollary can be regarded as the analogue of the law of iterated expectations when means are replaced by quantiles.

Corollary 1 (Law of Iterated Quantiles)**.** *For any* $\tau \in [0, 1]$,

$$\bigcup_{H \in \mathcal{H}_\tau} \mathbb{Q}^\tau(H) = [F_0^{-1}(\tau^2), F_0^{-1}(\tau(2-\tau)^+)].$$

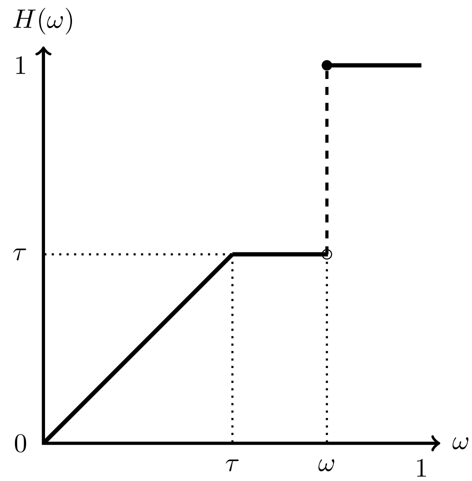


Figure III
AN EXAMPLE OF U^{ω}

Proof. Consider any $H \in \mathcal{H}_\tau$. By Theorem 1, $H \in \mathcal{I}(\underline{F}_0^\tau, \overline{F}_0^\tau)$. Thus,

$$F_0^{-1}(\tau^2) = (\underline{F}_0^\tau)^{-1}(\tau) \leq H^{-1}(\tau) \leq (\overline{F}_0^\tau)^{-1}(\tau^+) = F_0^{-1}(\tau(2-\tau)^+).$$

Conversely, consider any $\omega \in [F_0^{-1}(\tau^2), F_0^{-1}(\tau(2-\tau)^+)] = [(\underline{F}_0^\tau)^{-1}(\tau), (\overline{F}_0^\tau)^{-1}(\tau)]$. If $\omega \geq (\overline{F}_0^\tau)^{-1}(\tau)$, then $\omega \in \mathbb{Q}^\tau(\overline{F}_0^\tau)$. Otherwise, there must exist $\lambda \in [0,1]$ such that $\omega \in \mathbb{Q}^\tau(\lambda \underline{F}_0^\tau + (1-\lambda)\overline{F}_0^\tau)$. Since both $\overline{F}_0^\tau$ and $\lambda \underline{F}_0^\tau + (1-\lambda)\overline{F}_0^\tau$ are in $\mathcal{I}(\underline{F}_0^\tau, \overline{F}_0^\tau)$, Theorem 1 implies that $\omega \in \cup_{H \in \mathcal{H}_\tau} \mathbb{Q}^\tau(H)$, as desired. ■

4 Application I: Gerrymandering

4.1 The Limits of Gerrymandering

Our first application is to the consequences of political redistricting. The study of redistricting ranges across many fields: Legal scholars, political scientists, mathematicians, computer scientists, and economists have all contributed to this vast literature.[7] While existing economic theory on redistricting has largely focused on optimal redistricting or fair redistricting mechanisms (e.g., Owen and Grofman 1988; Friedman and Holden 2008; Gul and Pesendorfer 2010; Pegden, Procaccia, and Yu 2017; Ely 2019; Friedman and Holden 2020; Kolotilin and Wolitzky 2020), another fundamental question one may be curious about is the scope of redistricting's impact on a legislature. If *any* electoral map can be drawn, what kinds of

[7]See, for example, Sherstyuk (1998); Shotts (2001); Gilligan and Matsusaka (2006); Besley and Preston (2007); Coate and Knight (2007); McCarty, Poole, and Rosenthal (2009); Fryer Jr and Holden (2011); McGhee (2014); Stephanopoulos and McGhee (2015); Alexeev and Mixon (2018).

legislatures can be created? In other words, what are the "limits of gerrymandering"?

Theorem 1 describers the extent to which unrestrained gerrymandering can shape the composition of elected representatives. Specifically, consider an environment in which a continuum of citizens vote, and each citizen has single-peaked preferences over positions on political issues. Citizens have different ideal positions $\omega \in [0, 1]$, and these positions are distributed according to some $F_0 \in \mathcal{F}$.

In this setting, a signal $\mu \in \mathcal{M}$ can be thought of as an electoral *map*, which segments citizens into electoral *districts*, such that a district $F \in \mathrm{supp}(\mu)$ is defined as the conditional distribution of the ideal positions of citizens who belong to it. Each district elects a *representative*, and election results at the district-level follow the median voter property. That is, given any map $\mu \in \mathcal{M}$, the elected representative of each district F must have an ideal position that is a median of F. When there are multiple medians in a district, the representative's ideal position is determined by a selection rule $r \in \mathcal{R}$.[8]

Given any $\mu \in \mathcal{M}$ and any selection rule $r \in \mathcal{R}$, the induced distribution of posterior medians $H^{1/2}(\cdot|\mu, r)$ can be interpreted as a distribution of the ideal positions of the elected representatives. Meanwhile, the bounds $\underline{F}_0^{1/2}$ and $\overline{F}_0^{1/2}$ can be interpreted as distributions of representatives that only reflect one side of voters' political positions relative to the median of the population. Specifically, $\underline{F}_0^{1/2}$ describes an "all-left" legislature, in which each representative elected has an ideal position that is left of the median voter's ideal. Conversely, $\overline{F}_0^{1/2}$ represents an "all-right" legislature, in which all representatives are positioned to the right of the median voter.

An immediate implication of Theorem 1 is that *any* composition of the legislative body ranging from $\underline{F}_0^{1/2}$ to $\overline{F}_0^{1/2}$ can be procured by some map, as summarized by Proposition 1 below. Unrestrained gerrymandering can shape the legislative body all the way from the "all-left" legislature to the "all-right" legislature. This means that, in the most extreme scenario, unrestrained gerrymandering leads half the population to lose political representation of their views.

Proposition 1 (Limits of Gerrymandering)**.** *For any $H \in \mathcal{F}$, the following are equivalent:*

1. $H \in \mathcal{I}(\underline{F}_0^{1/2}, \overline{F}_0^{1/2})$.
2. *H is a distribution of the representatives' ideal positions under some map $\mu \in \mathcal{M}$ and some selection rule $r \in \mathcal{R}$.*

[8] Any district-level election system that meets the Condorcet criterion satisfies the median voter property (Downs 1957; Black 1958). An example is majority voting with two office-seeking candidates. Alternatively, any citizen-candidateship voting system with sufficiently high entry costs would elect a median representative in each district. An extensive empirical literature suggests that median voter preferences can more or less be mapped onto a single-issue space satisfying an ideal point (Congleton 2004). See also Poole and Daniels (1985), Congleton and Bennett (1995), and Gerber and Lewis (2004).

Having a complete characterization of possible compositions of the legislative body that can arise under gerrymandering, we may further explore the set of possible legislative outcomes. To this end, we assume that enacted legislation also satisfies the median voter property. Namely, the legislative outcome must be a median of the distribution of representatives' ideal positions.[9] Therefore, given a map $\mu \in \mathcal{M}$ and a selection rule $r \in \mathcal{R}$, we assume that the enacted *legislative outcomes* must be an element of $\mathbb{Q}^{1/2}(H^{1/2}(\cdot|\mu, r))$.

As an immediate implication of Corollary 1, the set of achievable legislative outcomes coincides with the interquartile range of political positions:

Proposition 2. *Under a legislative voting procedure satisfying the median voter property, a map can achieve a legislative outcome ω if and only if $\omega \in [F_0^{-1}(1/4), F_0^{-1}(3/4^+)]$.*

Proof. The set of achievable legislative outcomes is the union of medians of all distributions of posterior medians $\cup_{H \in \mathcal{H}_{1/2}} \mathbb{Q}^{1/2}(H)$, which, by Corollary 1, equals to $[F_0^{-1}(1/4), F_0^{-1}(3/4^+)]$. ■

Proposition 2 shows that any outcome within the interquartile range of citizens' ideal positions can possibly be enacted by a legislative body under some map, even if both the district-level elections and the legislative-level voting adhere to the median voter property. Furthermore, since the interquartile range of the population distribution is larger under more polarized distributions, Proposition 2 also implies that unrestrained gerrymandering could exacerbate polarization and translates citizens' polarized ideal positions into enacted legislation more easily.

A natural follow-up question from Proposition 2 is whether a voting procedure other than one satisfying the median voter property can reduce the range of possible legislation. To better formulate this question, we regard a legislative voting procedure as a mapping from the distribution of representatives' ideal positions to a legislative outcome. Of course, if any arbitrary voting procedure—regardless of complexity and practicality—is under consideration, then the answer to the question must be "yes." After all, always enacting the median citizen's ideal position regardless of how representatives vote (i.e., a constant mapping that maps every distribution to an element of $\mathbb{Q}^{1/2}(F_0)$) is feasible.

A more reasonable thought experiment would be to impose some minimal and reasonable requirements on the legislative voting procedure. One natural requirement would be that it must reflect the will of a majority whenever that will is unambiguous. In other words, for any distribution of representatives' ideal positions where more than 1/2 of the representatives have the same ideal position $\omega \in [0, 1]$, then a reasonable legislative voting system must yield outcome ω.

[9] See McCarty, Poole, and Rosenthal 2001; Bradbury and Crain 2005; and Krehbiel 2010 for evidence that the median legislator is decisive. See also Cho and Duggan (2009) for a micro foundation.

From this perspective, the aforementioned question can be reframed as the following: Does a mapping from the set of distributions of representatives' ideal positions to legislative outcomes exist such that its image is narrower than the interquartile range *and* reflects any unambiguous majority will at the same time? From Proposition 3 below, unfortunately, the answer is "no."

Proposition 3. *Consider any function $\mathbb{C} : \mathcal{H}_{1/2} \to [0,1]$. Suppose that $\mathbb{C}(H) = \omega$ for all H that assigns probability greater than $1/2$ to ω. Then, $\mathbb{C}(\mathcal{H}_{1/2}) \supseteq (F_0^{-1}(1/4), F_0^{-1}(3/4^+))$.*

Proof. Consider any $x \in (F^{-1}(1/4), F^{-1}(3/4^+))$. Let H_ω be a distribution that assigns probability $2 \cdot \min\{F_0(\omega), 1 - F_0(\omega)\}$ to ω, and probability $1 - 2 \cdot \min\{F_0(\omega), 1 - F_0(\omega)\}$ to $F_0^{-1}(1/2)$. Then $H_\omega \in \mathcal{I}(\underline{F}_0^{1/2}, \overline{F}_0^{1/2})$. Therefore, by Theorem 1, $H_\omega \in \mathcal{H}_{1/2}$, which in turn implies that $\mathbb{C}(H_\omega) = \omega$, as desired. ■

Proposition 2 and Proposition 3 suggest that under a wide variety of legislative voting procedures, the set of outcomes that can be enacted by a legislative body under some map is much larger than just the population medians, even if only the population medians are *Condorcet winners* in this setting.[10] A closely related question is: What is the set of legislative outcomes that can defeat the population medians by securing a majority of support among representatives elected under some map? Specifically, Proposition 4 below characterizes the set of legislative outcomes that are preferred by a fraction $\alpha \in [1/2, 1]$ of the representatives relative to any population medians under some map. To state this result, let $\underline{\omega}(\alpha) := \max\{2F_0^{-1}(\alpha/2) - F_0^{-1}(1/2), 0\}$ and $\overline{\omega}(\alpha) := \min\{2F_0^{-1}(1 - \alpha/2) - F_0^{-1}(1/2), 1\}$.

Proposition 4. *For any $\omega \in [0,1]$ and for any $\alpha \in [1/2, 1]$, the following are equivalent:*

1. $\omega \in [\underline{\omega}(\alpha), \overline{\omega}(\alpha)]$.
2. *There exists a map $\mu \in \mathcal{M}$ and a selection rule $r \in \mathcal{R}$ such that ω is preferred to any population median by at least α share of the representatives.*

Proof. Fix any $\alpha \in [1/2, 1]$, we first prove that 1 implies 2. Consider any $\omega \in [\underline{\omega}(\alpha), \overline{\omega}(\alpha)]$. If $\omega \in \mathbb{Q}^{1/2}(F_0)$, then 2 must hold since the map $\delta_{\{F_0\}} \in \mathcal{M}$ and the selection rule that selects ω with probability 1 induces a distribution of representatives that unanimously share an ideal position of ω. Now suppose that $\omega < F_0^{-1}(1/2)$. Note that if the distribution of representatives' ideal positions is $\underline{F}_0^{1/2}$, then the share of representatives whose ideal positions are closer to ω than to $F_0^{-1}(1/2)$ would be $2F((F^{-1}(1/2) + \omega)/2)$, which, in turn, is at least α,

[10] Recall that a Condorcet winner is defined as an outcome that has majority support when compared to any other alternative. As every citizen has single-peaked preferences over positions in $[0,1]$, a Condorcet winner always exists and the set of Condorcet winners coincides with the population medians $\mathbb{Q}^{1/2}(F_0)$.

as $\omega \geq \underline{\omega}(\alpha)$. Similarly, suppose that $\omega > F_0^{-1}(1/2^+)$. If the distribution of representatives' ideal positions is $\overline{F}_0^{1/2}$, then the share of representatives whose ideal position is closer to ω than to $F^{-1}(1/2^+)$ would be $2(1 - F((F^{-1}(1/2) + \omega)/2))$, which, in turn, is at least α, as $\omega \leq \overline{\omega}(\alpha)$. Therefore, by Theorem 1, 2 is satisfied for all $\omega \in [\underline{\omega}(\alpha), \overline{\omega}(\alpha)]$.

Conversely, to prove that 2 implies 1, fix any $\omega \in [0,1]$ and suppose that there exists a map $\mu \in \mathcal{M}$ and a selection rule $r \in \mathcal{R}$ such that under $H^{1/2}(\cdot|\mu, r)$, the share of representatives with ideal positions closer to ω than to either of $F_0^{-1}(1/2)$ or $F_0^{-1}(1/2^+)$ is at least α. That is, $H^{1/2}((F_0^{-1}(1/2) + \omega)/2|\mu, r) \geq \alpha$ if $\omega \leq F_0^{-1}(1/2)$ and $H^{1/2}((F_0^{-1}(1/2^+) + \omega)/2|\mu, r) \leq 1 - \alpha$ if $\omega \geq F_0^{-1}(1/2^+)$. By Theorem 1, it then follows that

$$2F_0\left(\frac{F_0^{-1}(1/2) + z}{2}\right) \geq H^{1/2}\left(\frac{F_0^{-1}(1/2) + z}{2}\middle|\mu, r\right) \geq \alpha$$

if $\omega \leq F_0^{-1}(1/2)$, and

$$2F_0\left(\frac{F_0^{-1}(1/2) + \omega}{2}\right) - 1 \leq H^{1/2}\left(\frac{F_0^{-1}(1/2) + \omega}{2}\middle|\mu, r\right) \leq 1 - \alpha$$

if $\omega \geq F^{-1}(1/2^+)$, which, in turn, implies $\underline{\omega}(\alpha) \leq \omega \leq \overline{\omega}(\alpha)$, as desired. ■

According to Proposition 4, even though the population medians are Condorcet winners, it is still possible for any outcome in $[\underline{\omega}(\alpha), \overline{\omega}(\alpha)]$ to secure an α-absolute majority of support among representatives under a map. A special case for this result is when F_0 has a symmetric and quasi-convex density. In this case, $\underline{\omega}(1/2) = 0$ and $\overline{\omega}(1/2) = 1$. That is, under a polarized population distribution (even only slightly), *any* outcome in $[0,1]$ can potentially defeat the population medians by a simple majority rule under some map, which is arguably a complete reversal of the population medians' Condorcet property. In the meantime, note that $\overline{\omega}$ is decreasing in α and $\underline{\omega}$ is increasing in α. Moreover, $\overline{\omega}(1) = F_0^{-1}(1/2^+)$ and $\underline{\omega}(1) = F_0^{-1}(1/2)$. This suggests that raising the voting threshold for an alternative to overturn the population median, such as requiring an absolute majority or even unanimous support, can mitigate the impact of gerrymandering in this regard.

Remark 1 (Districts on a Geographic Map). In practice, districts are drawn on a geographic map. Drawing districts in this manner can be regarded as partitioning a two-dimensional space that is spanned by latitude and longitude. More specifically, let a convex and compact set $\Theta \subseteq [0,1]^2$ denote a geographic map. Suppose that every citizen who resides at the same location $\theta \in \Theta$ shares the same ideal position $\boldsymbol{\omega}(\theta)$, where $\boldsymbol{\omega} : \Theta \to [0,1]$ is a measurable function. Furthermore, suppose that citizens are distributed on Θ according to a density function $\phi > 0$. Under this setting, theorem 1 of Yang (2020) ensures that for any $\mu \in \mathcal{M}$

with countable support, there exists a countable partition of Θ, such that the distributions of citizens' ideal positions within each element coincide with the distributions in the support of μ. If we further assume that $\boldsymbol{\omega}$ is non-degenerate, in the sense that each of its indifference curves $\{\theta \in \Theta | \boldsymbol{\omega}(\theta) = \omega\}_{\omega \in [0,1]}$ is isomorphic to the unit interval, then theorem 2 of Yang (2020) ensures that for any $\mu \in \mathcal{M}$, there exists a partition on Θ that generates the same distributions in each district. Therefore, the splitting of the distribution of citizens' ideal positions has an exact analogue to the splitting of geographic areas on a physical map.

4.2 Optimal Gerrymandering with Aggregate Uncertainty

In addition to characterizing the set of possible compositions of a legislative body that some map can induce, Theorem 1 also sheds lights on optimal gerrymandering problems in the presence of aggregate uncertainty. Consider the same unit mass of citizens whose ideal positions ω are distributed according to F_0. Suppose that there is a map drawer who designs the map of districts. The map drawer's objective depends on two political parties' seat shares in the legislative body. Specifically, consider the model with aggregate uncertainty but not individual uncertainty as in Kolotilin and Wolitzky (2020). Suppose that $X \sim G \in \mathcal{F}$ is an aggregate shock to citizens' views on the two parties. Given any realization $x \in [0,1]$, a citizen with ideal position ω votes for one party (Party 1) if and only if $\omega \geq x$; whereas, she votes for the other party (Party 0) if and only if $\omega < x$. Given a map $\mu \in \mathcal{M}$, a party wins a district if more than 50% of the citizens in that district vote for that party. Ties are broken by a tie-breaking rule $r \in \mathcal{R}$ selected by the map drawer.

The map drawer chooses a map $\mu \in \mathcal{M}$ and a tie-breaking rule $r \in \mathcal{R}$ to maximize her payoff $W(s)$, where $W : [0,1] \to \mathbb{R}$ and s denotes the share of districts that Party 1 wins (i.e., Party 1's seat share of the legislative body). Note that, for any district $F \in \mathcal{F}$ and for any realized aggregate shock $x \in \mathbb{R}$, Party 1 wins the district if $x \leq F^{-1}(1/2)$. Therefore, given any realized x, the map drawer's payoff under map μ and tie-breaking rule r is $1 - H^{1/2}(x^- | \mu, r)$, and hence the map drawer's problem can be written as

$$\sup_{\mu \in \mathcal{M}, r \in \mathcal{R}} \int_{\mathbb{R}} W\left(1 - H^{1/2}(x^- | \mu, r)\right) G(\mathrm{d}x).$$

But this problem, by Theorem 1, is equivalent to

$$\sup_{H \in \mathcal{I}(\underline{F}_0^{1/2}, \overline{F}_0^{1/2})} \int_{\mathbb{R}} W(1 - H(x^-)) G(\mathrm{d}x). \tag{4}$$

If W is increasing, then the solution of (4) is $\overline{F}_0^{1/2}$, which coincides with the solution stated in proposition 3 of Kolotilin and Wolitzky (2020). In general, Theorem 1 leads to an extension of proposition 3 of Kolotilin and Wolitzky (2020) and provides solutions to the map drawer's problem for any (measurable) objective function W. For instance, if the map drawer is a bipartisan commission who seeks to minimize the absolute difference between Party 1's seat share and $1/2$, then the objective function in (4) becomes $W(s) = -|s - 1/2|$. This, in turn, implies that the solution is $H^* \in \mathcal{I}(\underline{F}_0^{1/2}, \overline{F}_0^{1/2})$, where

$$H^*(\omega) := \begin{cases} \underline{F}_0^{1/2}(\omega), & \text{if } \omega < F_0^{-1}(1/4) \\ \frac{1}{2}, & \text{if } \omega \in [F_0^{-1}(1/4), F_0^{-1}(3/4^+)) \\ \overline{F}_0^{1/2}(\omega), & \text{if } \omega \geq F_0^{-1}(3/4^+) \end{cases}.$$

More generally, note that (4) can be written as

$$\sup_{H \in \mathcal{I}(\underline{F}_0^{1/2}, \overline{F}_0^{1/2})} \int_0^1 W(1-s) G \circ H^{-1}(\mathrm{d}s),$$

which can be further transformed into

$$\sup_{\widetilde{H} \in \mathcal{I}(G \circ (\overline{F}_0^{1/2})^{-1}, G \circ (\underline{F}_0^{1/2})^{-1})} \int_0^1 W(1-s) \widetilde{H}(\mathrm{d}s),$$

which becomes a linear programming problem. In particular, the extreme points of the feasible set can readily be characterized using similar arguments as in the proof of Lemma 2.

5 Application II: Bayesian Persuasion and Market Segmentation

5.1 Quantile-Based Bayesian Persuasion

Consider the canonical Bayesian persuasion problem of Kamenica and Gentzkow (2011). A state $\omega \in \mathbb{R}$ is distributed according to a common prior F_0. A sender chooses a signal $\mu \in \mathcal{M}$ to inform the receiver, who then picks an action $a \in A$ after seeing the signal's realization. The ex-post payoffs of the sender and receiver are $u_S(\omega, a)$ and $u_R(\omega, a)$, respectively. Kamenica and Gentzkow (2011) show that the sender's optimal signal and the value of persuasion can be characterized by the concave closure of the function $\hat{v} : \mathcal{F} \to \mathbb{R}$, where $\hat{v}(F) := \mathbb{E}_F[u_S(\omega, a^*(F))]$ and $a^*(F) \in A$ is the sender-preferred optimal action under posterior $F \in \mathcal{F}$.

When $|\mathrm{supp}(F_0)| \geq 2$, the "concavafication" method requires finding a concave closure of a multi-variate function, which is known to be computationally challenging, especially when $|\mathrm{supp}(F_0)| = \infty$. For tractability, many articles have restricted attention to preferences where the only payoff-relevant statistic for the sender is the posterior means (i.e., $\hat{v}(F)$ is measurable with respect with $\mathbb{E}_F[\omega]$). See, for example, Kolotilin, Li, Mylovanov, and Zapechelnyuk 2017, Gentzkow and Kamenica 2016, Kolotilin 2018, Dworczak and Martini 2019, and Ali, Haghpanah, Lin, and Siegel 2022. A natural analogue of this "mean-based" approach is to assume that the payoffs instead depend only on the posterior quantiles. Our main characterization provides a foundation for solving this class of problems.

Specifically, suppose that the sender's and receiver's payoffs are such that there exists $\tau \in [0,1]$ and a measurable function $v_S : \mathbb{R} \to \mathbb{R}$ in which

$$\hat{v}(F) = \sup_{\omega \in \mathbb{Q}^\tau(F)} v_S(\omega), \tag{5}$$

for all $F \in \mathcal{F}$. Under this assumption, Theorem 1 allows us to rewrite the sender's problem into

$$\sup_{H \in \mathcal{I}(\underline{F}_0^\tau, \overline{F}_0^\tau)} \int_{\mathbb{R}} v_S(\omega) H(\mathrm{d}\omega). \tag{6}$$

Namely, under the assumption of (5), Theorem 1 allows the sender to simply select a distribution belonging to the stochastic dominance interval $\mathcal{I}(\underline{F}_0^\tau, \overline{F}_0^\tau)$ that maximizes the expected value of $v_S(\omega)$, rather than concavafying the infinite-dimensional functional $\hat{v}$. A significant benefit of this simplification is that the sender only needs to solve a constrained maximization problem with an affine objective and a well-behaved feasible set.[11] In what follows, we demonstrate the aforementioned simplification by revisiting the two examples of Kamenica and Gentzkow (2011).

Lobbying under the Absolute Loss Function

A politician (receiver) chooses a one dimensional policy $a \in \mathbb{R}$ to match the state $\omega \in \mathbb{R}$, which is unknown to the politician and follows a common prior $F_0 \in \mathcal{F}$. The lobbyist (sender) can choose any signal for ω to affect the politician's choice of policy. Kamenica and Gentzkow (2011) assume that the lobbyist's payoff is given by $u_S(\omega, a) = -(a - \alpha\omega - (1-\alpha)\omega_0)^2$ for some fixed $\omega_0 \in \mathbb{R}$, $\alpha \in [0,1]$; and that the politician's payoff is $u_R(\omega, a) = -(a-\omega)^2$. The quadratic loss structure simplifies the lobbyist's problem to a mean-based persuasion problem that can easily be solved analytically.

Of course, one may argue that the quadratic loss structure is specific, and the general lobbying problem remains difficult to solve. Nonetheless, Theorem 1 allows us to solve another parameterization of this problem. Instead of quadratic loss, suppose now that the

[11] Indeed, notice that $\mathcal{I}(\underline{F}_0^\tau, \overline{F}_0^\tau)$ is compact under the weak-* topology, is a lattice under the partial order $\preceq$, and is a convex subset of a linear space whose extreme points can readily be derived from Lemma 1.

politician's payoff is given by the *absolute loss*: $u_R(\omega, a) = -|a - \omega|$. Also, suppose that the lobbyist's payoff is state-independent: $u_S(\omega, a) = v_S(a)$. Then, for any posterior $F \in \mathcal{F}$, the politician's optimal action is given by $\mathbb{Q}^{1/2}(F)$, and thus the sender's problem can be written as:

$$\sup_{H \in \mathcal{I}(\underline{F}_0^{1/2}, \overline{F}_0^{1/2})} \int_{\mathbb{R}} v_S(\omega) H(\mathrm{d}\omega),$$

which can now be solved analytically. For instance, suppose that v_S is increasing (resp., decreasing). Then the lobbyist's optimal signal is given by $\overline{F}_0^{1/2}$ (resp., $\underline{F}_0^{1/2}$). Alternatively, suppose that $v_S(a) = -|a - \omega_0|$ for some $\omega_0 \in \mathrm{supp}(F_0)$. Let $H^* \in \mathcal{I}(\underline{F}_0^{1/2}, \overline{F}_0^{1/2})$ be defined as:

$$H^*(\omega) := \begin{cases} \underline{F}_0^{1/2}(\omega), & \text{if } \omega < \omega_0 \\ \underline{F}_0^{1/2}(\omega_0), & \text{if } \omega \in [\omega_0, \omega_1) \\ \overline{F}_0^{1/2}(\omega), & \text{if } \omega \geq \omega_1 \end{cases},$$

where ω_1 is the unique solution of $\overline{F}_0^{1/2}(\omega_1^-) = \underline{F}_0^{1/2}(\omega_0)$. Then H^* is a solution to the lobbyist's problem. In general, for any (measurable) function $v_S : \mathbb{R} \to \mathbb{R}$, note that by the same arguments as the proof of Lemma 1, the lobbyist's problem can be characterized by

$$\sup_{H \in \mathcal{I}^*_{1/2}} \int_{\mathbb{R}} v_S(\omega) H(\mathrm{d}\omega).$$

Thus, by Lemma 1 and Lemma 2, it suffices to search for the optimal signal within the class of distributions satisfying (3).

Supplying Product Information to a Quantile-Maximizing Buyer

The second setting from Kamenica and Gentzkow (2011) has a seller choosing the price and information structure of its product when faced with a single potential buyer. Instead of expected utility preferences, suppose the buyer has quantile-maximizing preferences. This model of preferences was developed in Manski (1988), Chambers (2007), Rostek (2010), and de Castro and Galvao (2021). When selecting among lotteries, a quantile-maximizing individual chooses the one that gives the highest quantile of the utility distribution. For example, the person might maximize median utility instead of mean utility, as she would if she exhibited expected utility preferences.[12]

Suppose the buyer has unit demand and can be made aware of some information about

[12] When it comes to attitudes towards risk, Rostek (2010) argues that a quantile maximizer is concerned with downside risk, which means that she would rank the attractiveness of distributions of outcomes by comparing the likelihood of loss at different levels. The *maxmin* preferences of Roy (1952), Milnor (1954), and Rawls (1971)—in which a person selects the alternative offering the highest worst-case outcome—can be viewed as maximizing the lowest quantile, and such maxmin individuals would be the most risk averse.

the quality of the seller's product as a match for the buyer. Let the product's match quality be represented by $\theta \in \Theta$. Suppose that θ follows a common prior $\nu \in \Delta(\Theta)$. The buyer's indirect utility for the product given the item's match quality θ and price p is $u(\theta, p)$. The buyer chooses whether to purchase the product or pick an outside option worth $\overline{u}$ in utility with certainty. Meanwhile, the seller considers what information about the product's match quality to provide to the buyer. The seller commits to the signal for θ, but will not observe its realizations.

Kamenica and Gentzkow (2011) explore this problem under the assumptions that the buyer is an expected utility maximizer and that $u(\theta, p) = \theta - p$ and $\bar{u} = 0$. They conclude that an optimal signal for the seller, given price p, is to induce at most two signal realizations—one "low", one "high"—such that the buyer's posterior expected gains from trade under the "high" signal realization equals $\max\{p, \mathbb{E}_\nu[\theta]\}$. When the seller also optimizes with respect to price, it then follows that the seller's optimal price equals the expected gains from trade and the optimal signal reveals no information, as this allows the seller to fully extract the surplus.

Our main result enables us to explore this problem under the alternative assumption that the buyer is a quantile-maximizer, as defined by Rostek (2010). To see this, consider any fixed $p \geq 0$. Notice that it is without loss to represent a signal for θ by $\mu \in \mathcal{M}(F_p)$, where $F_p(\omega) := \nu(\{\theta \in \Theta | u(\theta, p) \leq \omega\})$ for all $\omega \in \mathbb{R}$. Quantile-maximizing preferences implies that the buyer will purchase the product under a posterior $F \in \mathcal{F}$ if $F^{-1}(\tau^+) > \bar{u}$ and will not purchase if $F^{-1}(\tau) < \bar{u}$, for some $\tau \in [0, 1]$. As a result, the seller's problem can be written as

$$\sup_{\mu \in \mathcal{M}, r \in \mathcal{R}} p(1 - H^\tau(\bar{u}^- | \mu, r)),$$

which is to maximize price times the probability of the buyer purchasing the product (i.e., expected revenue). The characterization in Theorem 1 implies that the seller's optimal revenue given price p is $p(1 - \overline{F}_p^\tau(\bar{u}^-))$, where $\overline{F}_p^\tau(\omega) := \max\{({F_p(\omega) - \tau})/({1 - \tau}), 0\}$. As this holds for all $p \geq 0$, it then follows that the seller's optimal price is the solution(s) of

$$\max_{p \geq 0} p(1 - \overline{F}_p^\tau(\bar{u}^- | p)).$$

Let p^* be the seller's optimal price. It is noteworthy that, unlike in the expected utility model, the seller does not fully extract all the gains from trade when facing a quantile-maximizing buyer. In the meantime, the seller would induce $\overline{F}_{p^*}^\tau$ as a distribution of the posterior τ-quantile, which is different from disclosing no information. Some information about the product's match quality is disclosed under the seller-optimal signal. In practice, inducing a distribution of the posterior quantile $\overline{F}^{\tau}_{p^*}$ means that the seller would share information for the buyer to distinguish products conditional on the products having high match quality, but the seller would retain information to prevent the buyer from completely discerning high match-quality products from low match-quality ones. For instance, the seller might disclose that a handbag were made in Italy, but withhold the means of production (machine versus handmade).

5.2 Optimal Market Segmentation with a Fixed Thickness Constraint

In addition to being a signal in Blackwell's sense, another common interpretation of $\mu \in \mathcal{M}$ is a *market segmentation* that splits a single market into several segments to facilitate price discrimination (see, for instance, Bergemann, Brooks, and Morris 2015; Haghpanah and Siegel 2020; Yang 2022; Haghpanah and Siegel 2022; Elliot, Galeotti, Koh, and Li 2022). From this perspective, the characterization of Theorem 1 further enables us to explore optimal market segmentations in a two-sided market. When the market is two-sided, market segmentation involves another dimension. Namely, in addition to segmenting both sides of the market, one needs to describe how segments on one side are matched with those on the other. Optimal market segmentation in a two-sided market contains features of both market segmentation of a single market and matching in a two-sided market (see, for instance, Zhao, Zhang, Khan, and Perrussel 2010; Hagiu and Jullien 2011; de Cornière 2016; Condorelli and Szentes 2022; Guinsburg and Saraiva 2022).

Consider a two-sided market (e.g., ride sharing) for an object (e.g., a car ride). The demand side is populated by a unit mass of agents (riders) who have unit demands for a ride. Their values ω for the object are distributed according to a distribution $F_0 \in \mathcal{F}$. The supply side is populated by a mass $\tau \in (0, 1)$ of agents (drivers) who have perfectly inelastic supply (e.g., during peak hours at a major airport or at the end of popular concert), and each driver provides one care ride. A third-party platform (a ride-sharing app) can segment both sides of the market to affect prices, which are in turn determined by the market-clearing condition in each segment. However, when segmenting the market, the platform is subject to a constraint that the *market thickness* of each segment must be the same.

Specifically, $\mu \in \mathcal{M}(F_0)$ can be regarded as a demand-side market segmentation. A segmentation μ induces market segments $\{F|F \in \text{supp}(\mu)\}$. Each segment is described by the distribution of agents' values within that segment. Suppose that the third-party platform retains a fixed share of sales revenue from each market segment, and it can arbitrarily segment both sides of the market and then match segments from one side of the market with the other. The platform's optimal segmentation without a thickness constraint is to completely segment both sides of the market, and then match the demand side with the supply side in an assortative manner until the supply side is exhausted. This corresponds to the standard first-degree price discrimination outcome when a monopolist has a capacity constraint.

However, for many practical reasons (e.g., regulation, fairness, corporate image, match efficiency, or customer satisfaction), platforms can rarely segment both sides of the market arbitrarily. Instead, they typically face several constraints when segmenting the market. One practical constraint is that the *market thickness* must be held fixed across all market segments. In this setting where the supply is perfectly inelastic, a market thickness constraint implies that the sizes of segments on the supply side must be completely determined by the sizes of segments on the demand side, so that the ratio of supply and demand remains τ in each segment.[13]

Without the ability to segment both sides of the market arbitrarily, it may seem difficult to solve the optimal segmentation problem with a fixed thickness constraint. Nonetheless, Theorem 1 provides a solution to this problem. To see this, note that for any segment $F \in \mathcal{F}$, given that the ratio of supply and demand is τ, the implied market-clearing price for this segment must be in $\mathbb{Q}^{(1-\tau)}(F)$. As a result, the platform's problem can be written as

$$\sup_{\mu \in \mathcal{M}, r \in \mathcal{R}} \tau \int_{\mathbb{R}} \omega H^{(1-\tau)}(\mathrm{d}\omega | \mu, r),$$

which, by Theorem 1, can be written as

$$\sup_{H \in \mathcal{I}(\underline{F}_0^{(1-\tau)}, \overline{F}_0^{(1-\tau)})} \tau \int_{\mathbb{R}} \omega H(\mathrm{d}\omega),$$

and thus the platform's optimal segmentation induces $\overline{F}_0^{(1-\tau)}$ as the distribution of prices across segments. Moreover, notice that total sales revenue is given by

$$\tau \int_{\mathbb{R}} \omega \overline{F}_0^{(1-\tau)}(\mathrm{d}\omega) = \mathbb{E}_{F_0}[\omega | \omega \geq F_0^{-1}((1-\tau)^+)],$$

which is exactly the same as that under first-degree price discrimination. Together, we can make the following observation:

Proposition 5. *Market thickness constraints are irrelevant to the platform's optimal revenue.*

[13]Practically speaking, a thickness constraint would imply that all riders would have to wait approximately the same time before being matched with a driver.

6 Application III: Econometrics

In this section, we apply our main result to subjects in econometrics. To this end, consider a random vector $(Y, X) \in \mathbb{R} \times \mathbb{R}^K$ on an underlying probability space with probability measure $\mathbb{P}$.[14] Our main interest is about the conditional distribution of Y given realizations $X = x \in \mathbb{R}^K$, which we denote by $F_{Y|X=x}$. Note that for each $x \in \mathbb{R}^K$, $F_{Y|X=x}$ can be regarded as a realized posterior induced by the joint distribution of (Y, X).

6.1 Model Misspecification and Partial Identification in Quantile Regression

Koenker and Bassett Jr (1978) introduced the econometric approach of quantile regression, which models the quantiles of the conditional distribution of a response variable as a function of observed covariates. Our characterization can evaluate whether a presumed model for the conditional quantiles is mis-specified.

In particular, consider a response variable Y_i and K-dimensional covariate vector X_i. The observations $\{Y_i, X_i\}_{i=1}^N$ are independently and identically drawn. The marginal distribution of Y_i is either known from the literature or can be correctly estimated to be $F_0 \in \mathcal{F}$. Consider now the τ-quantile function $g_\tau : \mathbb{R}^K \to \mathbb{R}$ such that

$$g_\tau(x) \in \mathbb{Q}^\tau(F_{Y_i|X_i=x}).$$

Quantile regression aims to estimate the function g_τ using the sample. To facilitate the analysis and maintain tractability, econometricians often impose some further assumptions on the functional form of g_τ. A commonly used model is the linear model $g_\tau(x) = (1, x')\beta$, for some $\beta \in \mathbb{R}^{K+1}$. However, these models may potentially be mis-specified. Theorem 1 provides a simple test for model mis-specification.

Consider the (potentially mis-specified) linear quantile regression model $g_\tau(x) = (1, x')\beta$. The estimand under this model solves the minimization problem

$$\min_{\beta \in \mathbb{R}^{K+1}} \mathbb{E}[\rho_\tau(Y_i - (1, X_i')\beta)], \tag{7}$$

where $\rho_\tau(u) := u(\tau - \mathbb{I}\{u < 0\})$ is the tilted absolute value function.

From Theorem 1, the linear model is correctly specified under $\mathbb{P}$ only if the distribution of $(1, X_i')\beta^*$ is in $\mathcal{I}(\underline{F}_0^\tau, \overline{F}_0^\tau)$, where β^* is the solution to (7). An econometrician could test for model mis-specification using only knowledge of the marginal of Y_i. If the empirical

[14] Throughout this section, we hold fix this probability space and assume that it is rich enough relative to the random variable Y in the sense of definition 2 of Yang (2020). That is, the probability space restricted to any pre-image of Y is isomorphic to a unit interval with the Lebesgue measure.

distribution of $(1, X_i')\beta^*$ fell outside the interval $\mathcal{I}(\underline{F}_0^\tau, \overline{F}_0^\tau)$, there would be strong evidence of mis-specification. A comparison of the empirical quantiles or a Kolmogorov–Smirnov test are two ways to implement the evaluation. The reliance on information from just the marginal of Y_i allows one to bypass estimation of the joint distribution of (Y_i, X_i), which may be computationally demanding. Meanwhile, as the test does not exploit any conditioning information from X_i, which would have restricted the set of eligible models, the test has the advantage of presenting low Type-I error, but also the disadvantage of having low power.

The bounds on the distribution of $g_\tau(X_i)$ can also institute parameter restrictions on β to mitigate bias resulting from potential model mis-specification. The restrictions would take the form of linear constraints imposed on the problem in the sample analogue of (7). That is, one may compute the estimator of β under the linear model by solving the following problem:

$$\begin{aligned} &\min_{\beta \in \mathbb{R}^{K+1}} \frac{1}{n} \sum_{i=1}^{n} \rho_\tau \left(Y_i - (1, X_i')\beta\right) \\ &\text{s.t.} \qquad \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}\{(1, X_i')\beta \leq y\} \leq \underline{F}_0^\tau(y), \forall y \in \mathbb{R}, \\ &\qquad\qquad \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}\{(1, X_i')\beta \leq y\} \ \geq \overline{F}_0^\tau(y), \forall y \in \mathbb{R}. \end{aligned} \tag{8}$$

The linear constraints in (8) confine the estimand β^* so that the distribution of the resultant conditional linear quantile function, $X_i'\beta^*$, respects the interval $\mathcal{I}(\underline{F}_0^\tau, \overline{F}_0^\tau)$ from Theorem 1.

When the number of covariates $K = 1$ and the marginals of both X_i and Y_i are known, an econometrician can go beyond evaluating model mis-specification to partially identifying the quantile function $g_\tau(X_i)$. Suppose now that $X_i \sim G$ and $Y_i \sim F_0$. Taking a concrete example, one might have Y_i representing income and X_i standing for years of schooling, but the two samples are reported in distinct datasets from potentially non-overlapping populations. Estimation of economic models that involve Y_i and X_i originating from different samples is part of the econometrics of *data combination* (Ridder and Moffitt 2007).

If g_τ is known to be increasing (such as wages increasing in years of schooling), then, for any $\omega \in \mathbb{R}$, the probability that the conditional quantile registers at or below ω, given by $\mathbb{P}(g_\tau(X_i) \leq \omega)$, is simply $G(g_\tau^{-1}(\omega))$. As a result, by Theorem 1, it must be that $G \circ g_\tau \in \mathcal{I}(\underline{F}_0^\tau, \overline{F}_0^\tau)$, and hence, for all $\omega \in \mathbb{R}$,

$$(\underline{F}_0^\tau)^{-1} \circ G(\omega) \leq g_\tau(\omega) \leq (\overline{F}_0^\tau)^{-1} \circ G(\omega), \tag{9}$$

for all $\omega \in \mathbb{R}$. Proposition 6 below formalizes this observation and provides a non-parametric partial identification of the function g_τ.

Proposition 6 (Identification Set)**.** *For any $\tau \in [0,1]$ and for any increasing function $g_\tau : \mathbb{R} \to \mathbb{R}$, the following are equivalent:*

1. *There exists a random variable X such that the marginal of X is G and $g_\tau(X) \in \mathbb{Q}^\tau(F_{Y_1|X})$ with probability 1.*

2. *$(\underline{F}_0^\tau)^{-1} \circ G(\omega) \leq g_\tau(\omega) \leq (\overline{F}_0^\tau)^{-1} \circ G(\omega)$, for all $\omega \in \mathbb{R}$.*

Proof. The proof for 1 implying 2 follows immediately from Theorem 1 and the fact that g_τ is increasing. To show that 2 implies 1, consider any increasing function g_τ satisfying (9). Let $H(\omega) := G(g_\tau^{-1}(\omega))$ for all $\omega \in \mathbb{R}$. Then, by Theorem 1, $H \in \mathcal{I}(\underline{F}_0^\tau, \overline{F}_0^\tau)$ implies that there exists a signal $\mu \in \mathcal{M}$ and a selection rule $r \in \mathcal{R}$ such that $H(\omega) = H^\tau(\omega|\mu, r)$ for all $\omega \in \mathbb{R}$. As $\omega \in \mathbb{R}$, theorem 2 of Yang (2020) ensures that there exists a random variable $\widetilde{X}$ such that $\mathbb{P}(F_{Y_1|\widetilde{X}} \in A) = \mu(A)$ for all measurable $A \subseteq \mathcal{F}$, and that

$$\mathbb{P}(\tilde{g}_\tau(\widetilde{X}) \leq \omega) = H(\omega),$$

for all $\omega \in \mathbb{R}$, with $\tilde{g}_\tau$ being an increasing function such that $\tilde{g}_\tau(x) \in \mathbb{Q}^\tau(F_{Y_1|\widetilde{X}=x})$ for all $x \in \mathbb{R}$. Now define a random variable $X := g_\tau^{-1}(\tilde{g}_\tau(\widetilde{X}))$. We claim that the marginal of X is G and that $g_\tau(X) \in \mathbb{Q}^\tau(F_{Y_1|X})$ with $\mathbb{P}$-probability 1. Indeed, since both $\tilde{g}_\tau$ and g_τ are increasing,

$$\mathbb{P}(X \leq x) = \mathbb{P}(g_\tau^{-1}(\tilde{g}_\tau(\widetilde{X})) \leq x) = \mathbb{P}(\tilde{g}_\tau(\widetilde{X}) \leq g_\tau(x)) = H(g_\tau(x)) = G(x).$$

In the meantime, since $g_\tau^{-1} \circ \tilde{g}_\tau$ is increasing, it must be that $\mathbb{Q}^\tau(F_{Y_1|\widetilde{X}=x}) = \mathbb{Q}^\tau(F_{Y_1|X=g_\tau^{-1}\circ\tilde{g}_\tau(x)})$ for $\mathbb{P} \circ \widetilde{X}^{-1}$-almost all $x \in \mathbb{R}$. Together with $g_\tau(X) = \tilde{g}_\tau(\widetilde{X})$, it then follows that $g_\tau(X) \in \mathbb{Q}^\tau(F_{Y_1|X})$ with $\mathbb{P}$-probability 1. This completes the proof. ■

Note that Proposition 6 provides a complete characterization of the identification set of g_τ. That is, under the given assumptions (i.e., only the marginals of Y_i, X_i are known and g_τ is only known to be increasing), the quantile function g_τ must satisfy (9). Conversely, for any function g_τ satysfying (9), there exists a model that meets the given assumptions and induces a quantile function g_τ. This identification result requires neither parametric assumptions, nor knowledge about the joint distribution, except for monotonicity of the quantile function. It allows an econometrician to make inferences on the conditional distribution of, say, income on schooling, just from knowing the marginal distributions of wages and school years, with the two potentially being measured from different population samples.

6.2 Inferences of Joint Distributions from Marginals

As hinted above, one common obstacle faced by econometricians is the lack of information about the joint distribution, even though information about the marginals is available. Specifically, given two random variables Y, X, with known marginals F_0 and G, respectively, what can one infer about their joint?

Horowitz and Manski (1995) provide a characterization in the case when F_0 has a positive density on its support and when X is binary, which might refer to the realization of an event that contaminates the dataset, or enrollment in the prescribed treatment of an experiment. If $X \in \{0,1\}$ and $\mathbb{P}(X = 1) = p \in (0, 1/2]$, the authors provide sharp bounds on the conditional distributions. In particular, for any $\tau \in [0,1]$, they show that $\mathbb{Q}^\tau(F_{Y|X=1}) \subseteq [F_0^{-1}(\tau p), F_0^{-1}(\tau p + 1 - p)]$. Moreover, there exists a joint distribution that attains the two bounds $F_0^{-1}(\tau p)$ and $F_0^{-1}(\tau p + 1 - p)$. Our Theorem 1 extends this result by demonstrating that *any* distribution within these bounds is attainable by some joint distribution of Y and X, with X being binary, as stated below.

Proposition 7. *For any $\tau \in [0,1]$, for any random variable Y with distribution F_0 and for any $\omega \in [F_0^{-1}(\tau p), F_0^{-1}(\tau p + 1 - p)]$, there exists a random variable $X \in \{0,1\}$ with $\mathbb{P}(X = 1) = p$ such that $\omega \in \mathbb{Q}^\tau(F_{Y|X=1})$.*

Proof. If $\omega \leq (\overline{F}_0^\tau)^{-1}(p)$, let

$$H_p(y) := \begin{cases} 0, & \text{if } y < \omega \\ p, & \text{if } y \in [\omega, (\overline{F}_0^\tau)^{-1}(p)) \\ 1, & \text{if } y \geq (\overline{F}_0^\tau)^{-1}(p) \end{cases},$$

for all $y \in \mathbb{R}$. Meanwhile, if $\omega > (\overline{F}_0^\tau)^{-1}(p)$, let

$$H_{1-p}(y) := \begin{cases} 0, & \text{if } y < (\underline{F}_0^\tau)^{-1}(1-p) \\ p, & \text{if } y \in [(\underline{F}_0^\tau)^{-1}(1-p), \omega) \\ 1, & \text{if } y \geq \omega \end{cases},$$

for all $y \in \mathbb{R}$. Note that $F_0^{-1}(\tau p) = (\underline{F}_0^\tau)^{-1}(p)$ and $F_0^{-1}(\tau p + 1 - p) = (\overline{F}_0^\tau)^{-1}(1-p)$. Therefore, if $\omega \leq (\overline{F}_0^\tau)^{-1}(p)$, then $H_p \in \mathcal{I}(\underline{F}_0^\tau, \overline{F}_0^\tau)$; while if $\omega > (\overline{F}_0^\tau)^{-1}(p)$, then $H_{1-p} \in \mathcal{I}(\underline{F}_0^\tau, \overline{F}_0^\tau)$. By Theorem 1, there always exists some $H \in \mathcal{H}_\tau$ with binary support that assigns probability p to ω, which, in turn, implies that there exists a random variable $X \in \{0,1\}$ with $\mathbb{P}(X = 1) = p$ such that $\omega \in \mathbb{Q}^\tau(F_{Y|X=1})$, as desired. ∎

In the meantime, Cross and Manski (2002) discuss identification of so-called *long* regressions when the *short* conditional distributions are known, but the *long* ones are not. The authors describe an environment in which each member of a population is associated with a tuple (Y, X, Z), such that $Y \in \mathbb{R}$, X takes values in a finite dimensional real space, and Z belongs to a K-element finite set. The issue at hand is identification of the long regression $\mathbb{E}[Y|X, Z]$ when the short conditional distributions $\mathbb{P}(Y|X)$ and $\mathbb{P}(Z|X)$ are known, but the long conditional distribution $\mathbb{P}(Y|X, Z)$ is unknown. The language "long" and "short" borrows from Goldberger (1991). In their article, the authors give bounds on $(\mathbb{E}[Y|X, Z = z_k])_{k=1}^K$, though the bounds might not be sharp. Only in the special case when both $K = 2$ and the response variable $Y \in \{0, 1\}$, do the authors completely identify the set of $(\mathbb{E}[Y|X, Z = z_k])_{k=1}^K$. Our Theorem 1 complements this result: The conditional quantiles, $(\mathbb{Q}^\tau(F_{Y|X,Z=z_k}))_{k=1}^K$, are completely identified for all $K < \infty$. Furthermore, Y need not be binary, as the theorem applies even if $|\text{supp}(Y)| = \infty$.

Proposition 8. *Let $P_0 := 0$ and let $P_k := \sum_{j=1}^k p_k$ for all $k \in \{1, \ldots, K\}$. For any random variable Y with distribution F_0, for any $\tau \in [0, 1]$, and for any vector $\boldsymbol{q} = (q_k)_{k=1}^K \in \mathbb{R}^K$ with $q_1 \leq \ldots \leq q_K$, the following are equivalent.*

1. *There exists a random variable Z with support $\{z_k\}_{k=1}^K$ and $\mathbb{P}(X = z_k) = p_k$ such that $q_k \in \mathbb{Q}^\tau(F_{Y|Z=z_k})$ for all $k \in \{1, \ldots, K\}$.*
2. *$q_k \in [(\underline{F}_0^\tau)^{-1}(P_k), (\overline{F}_0^\tau)^{-1}(P_{k-1}^+)]$.*

Proof. To show that 1 implies 2, consider any such random variable Z. Let H be the CDF of the induced probability distribution over $\{q_k\}_{k=1}^K$. Theorem 1 implies that $H \in \mathcal{I}(\underline{F}_0^\tau, \overline{F}_0^\tau)$, which in turn implies 2.

Conversely, to show that 2 implies 1, consider any $\boldsymbol{q} = (q_k)_{k=1}^K$ with $q_1 \leq \ldots \leq q_K$. Define a CDF H as follows:

$$H(\omega) := \begin{cases} 0, & \text{if } \omega < q_1 \\ P_k, & \text{if } \omega \in [q_{k-1}, q_k),\ k \in \{2, \ldots, K\} \\ 1, & \text{if } \omega \geq q_K \end{cases},$$

for all $\omega \in \mathbb{R}$. Then, 2 implies $H \in \mathcal{I}(\underline{F}_0^\tau, \overline{F}_0^\tau)$. By Theorem 1, $H \in \mathcal{H}_\tau$, and thus there exists a random variable Z with $\text{supp}(Z) = \{z_k\}_{k=1}^K$ and $\mathbb{P}(X = x_k) = p_k$ such that $q_k \in \mathbb{Q}^\tau(F_{Y|X=x_k})$ for all $k \in \{1, \ldots, K\}$, as desired. ■

It is noteworthy that although the statement of Proposition 8 does not include the control variable X for the short regression, there are no restrictions on the marginal of Y. Therefore, the conditional quantiles $(\mathbb{Q}^\tau(F_{Y|X,Z=z_k}))_{k=1}^K$ can be identified by applying Proposition 8 to the conditional distribution of Y given each realization of X. Moreover, the monotonicty restriction in Proposition 8 is merely a normalization. The identification set of $(\mathbb{Q}^\tau(F_{Y|X,Z=z_k}))_{k=1}^K$

can be obtain by applying permutations on the result of Proposition 8, as illustrated by Figure IV for the case of $K = 2$, $F_0 = U$, and $\tau = 1/2$.

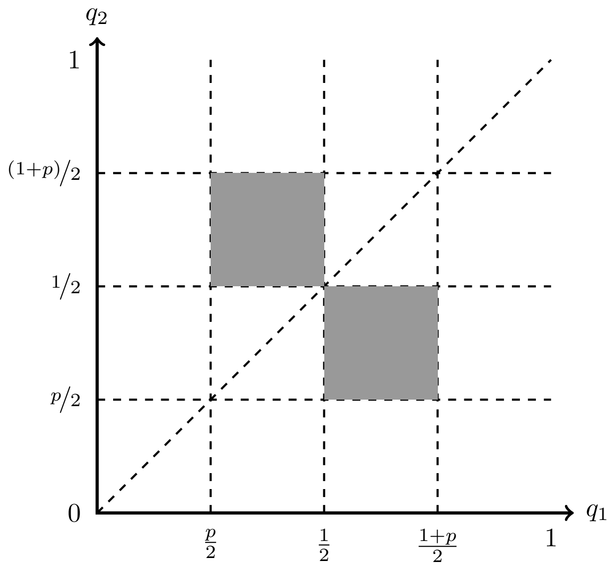


Figure IV
IDENTIFICATION SET OF $(\mathbb{Q}^{1/2}(F_{Y|Z=z_k}))_{k=1}^{2}$

7 Application IV: Finance and Accounting

7.1 Robust Macroprudential Policy

Macroprudential policy deals with regulatory practices that aim at ensuring, as best as possible, the stability of the financial system as a whole (Galati and Moessner 2013). A commonly suggested policy suitable to this aim is to link a large financial institution's required amount of equity capital to its contribution to systemic risk, which is the risk of the entire financial system collapsing (Brunnermeier, Gorton, and Krishnamurthy 2012; Orlov, Zryumov, and Skrzypacz 2020). Several measures of systemic risk abound, but a popular one is the Conditional Value at Risk from Adrian and Brunnermeier (2016), denoted $CoVaR$. The authors define this measure as the Value at Risk (VaR) of the financial system, conditional on a particular institution being under financial duress. In this context, the VaR would be the loss in the market value of the financial system that is exceeded with a certain (tail) probability (Duffie and Pan 1997).

But $CoVaR$ itself is inherently a quantile of a conditional probability distribution. Our characterization thus implies that we can derive bounds on $CoVaR$, irrespective of the financial state of the institution. If a macroprudential policy tool ties an institution's required equity capital to its $CoVaR$. our characterization would inform a policymaker of a "robust" policy that would set capital requirements uniformly across institutions to guard against the most calamitous losses a financial system could be estimated to face.

Specifically, let the unconditional VaR of the financial system with τ probability follow $V_\tau \sim F_0$. The $CoVaR$ of a financial institution would simply be V_τ, conditional on the event that the institution is under some amount of financial distress. Suppose that a financial system regulator is considering an equity capital requirement g that would be a monotone decreasing function of a financial institution's $CoVaR$. Orlov, Zryumov, and Skrzypacz (2020) provide a micro foundation for this kind of capital requirement. An immediate implication of Theorem 1 is that a robust macroprudential policy would set the capital requirement of each institution to be $g\left(\underline{F}_0^\tau\right)$, as the $CoVaR$ of any one institution can never fall below the stochastic bound $\underline{F}_0^\tau$. Capital requirements would be uniform across institutions and counter-cyclical when taking a dynamic perspective. As the risk of financial system collapse wanes, $\underline{F}_0^\tau$ would shift to the right, which would lower equity capital requirements across institutions. Conversely, as risk to the financial system rises, $\underline{F}_0^\tau$ would shift to the left, which would require banks to raise capital in anticipation of higher potential losses.

7.2 Classification Shifting

McVay (2006) describes a management tool to manipulate accounting earnings that involves deliberately misclassifying items within a firm's income statement. McVay refers to this practice as *classification shifting*, and she finds evidence consistent with managers moving expenses from the category of core expenses (e.g., cost of goods sold) to the category of special items (e.g., fines). Mangers are thought to engage in this conduct to overstate "core" earnings and meet Wall Street analyst earnings forecasts, as special items tend to be excluded from analysts' definitions of earnings. Fan, Barua, Cready, and Thomas (2010) document further evidence of classification shifting, and Dye (2002) presents a theoretical model of classification manipulation.

With our result, we can provide a necessary condition for an auditor to audit classification shifting. Consider the following scenario: A manager classifies an extensive set of expenses into categories (e.g., core versus special items) and is requested to report a τ-quantile of each category to an auditor. Each expense belongs to a rightful category consistent with Generally Accepted Accounting Principals (GAAP). A certain dollar threshold of misclassification is considered material and constitutes accounting fraud. The auditor can observe the distribution of all the expenses, but verifying each expense's classification is costly. The problem for the auditor is to determine whether a closer inspection of the manager's classification is warranted.

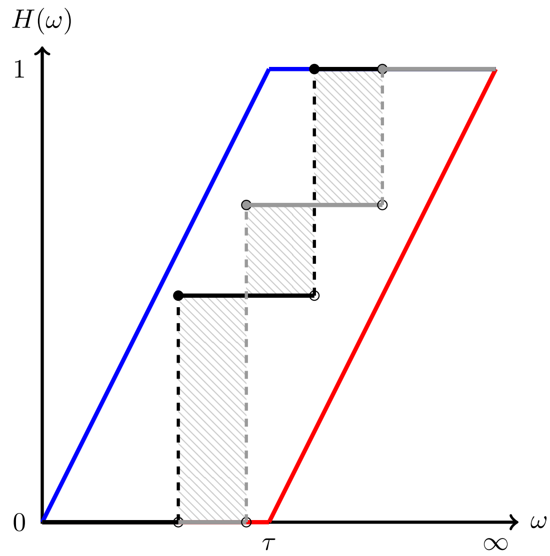


Figure V
A Measure of Misclassification

Specifically, suppose the cost of auditing is $K > 0$. Denote the distribution of expenses by F_0. Suppose that the auditor can select any $\tau \in [0, 1]$ and request the manager to report a τ-quantile of each category. Then, for any $\tau \in [0, 1]$, a report under any classification of spending induces a distribution of posterior τ-quantiles.

Let G_τ denote the distribution of τ-quantiles under the correct classification. If H is the distribution of τ-quantiles induced by the manager's classification, then the amount of misclassification, in the unit of dollars, can be measured as

$$\int_0^1 \left|G_\tau^{-1}(q) - H^{-1}(q)\right| \mathrm{d}q.$$

Figure V illustrates a distribution of quantiles induced by a classification with two categories. The manager's classification is represented in black, and the correct classification is in gray.

By Theorem 1, regardless of the classification, both distributions of τ-quantiles induced by the manager's classification and by the correct classification must reside within the stochastic dominance interval $\mathcal{I}(\underline{F}_0^\tau, \overline{F}_0^\tau)$. As a result, whenever

$$\sup_{\tau \in [0,1]} \int_0^\infty \left|\overline{F}_0^\tau(\omega) - \underline{F}_0^\tau(\omega)\right| \mathrm{d}\omega \le K, \tag{10}$$

then an audit is never warranted, since the left-hand side of (10) is the largest possible amount of misclassification.

8 Conclusion

We characterize the distributions of all possible posterior quantiles in a general environment. Unlike the distributions of posterior means, which are known to be mean-preserving contractions of the prior, the distributions of posterior quantiles reside between two first-order stochastic bounds that are simple functions of the prior. We apply this characterization to many economic scenarios, ranging from political economy, to Bayesian persuasion, industrial organization, econometrics, finance, and accounting.

Other applications involving quantiles undoubtedly exist. When consumers' values or firms' marginal costs follow distributions, different points on the inverse supply and demand curves are quantiles, which opens the door to further applications in consumer or firm theory. Inequality is often measured as an upper percentile of the wealth or income distribution, making it eligible for analysis. Likewise, settings in which threshold behavior is important, such as in theories of bank runs, protests, fads and fashions, or tipping points, are just some areas for future work.

References

Adrian, T. and M. K. Brunnermeier (2016): "CoVaR," *American Economic Review*, 1705–1741.

Alexeev, B. and D. G. Mixon (2018): "An impossibility theorem for gerrymandering," *The American Mathematical Monthly*, 125, 878–884.

Ali, S. N., N. Haghpanah, X. Lin, and R. Siegel (2022): "How to Sell Hard Information," *Quarterly Journal of Economics*, 137, 619–678.

Bergemann, D., B. Brooks, and S. Morris (2015): "The limits of price discrimination," *American Economic Review*, 105, 921–957.

Besley, T. and I. Preston (2007): "Electoral bias and policy choice: theory and evidence," *Quarterly Journal of Economics*, 122, 1473–1510.

Black, D. (1958): *The theory of committees and elections*, Springer.

Blackwell, D. (1953): "Equivalent Comparisons of Experiments," *Annals of Mathematical Statistics*, 24, 265–272.

Börgers, T. (2015): *An introduction to the theory of mechanism design*, Oxford University Press.

Bradbury, J. C. and W. M. Crain (2005): "Legislative district configurations and fiscal policy in American States," *Public Choice*, 125, 385–407.

Brunnermeier, M. K., G. Gorton, and A. Krishnamurthy (2012): "Risk topography," *Nber macroeconomics annual*, 26, 149–176.

Chambers, C. P. (2007): "Ordinal aggregation and quantiles," *Journal of Economic Theory*, 137, 416–431.

Cho, S.-j. and J. Duggan (2009): "Bargaining foundations of the median voter theorem," *Journal of Economic Theory*, 144, 851–868.

Coate, S. and B. Knight (2007): "Socially optimal districting: A theoretical and empirical exploration," *Quarterly Journal of Economics*, 122, 1409–1471.

Condorelli, D. and B. Szentes (2022): "Buyer-Optimal Matching in Two-Sided Platforms," Working Paper. University of Warwick, Coventry, UK.

Congleton, R. D. (2004): "The median voter model," in *The encyclopedia of public choice*, Springer, 707–712.

Congleton, R. D. and R. W. Bennett (1995): "On the political economy of state highway expenditures: Some evidence of the relative performance of alternative public choice models," *Public choice*, 84, 1–24.

Cross, P. J. and C. F. Manski (2002): "Regressions, short and long," *Econometrica*, 70, 357–368.

de Castro, L. and A. F. Galvao (2021): "Static and dynamic quantile preferences," *Economic Theory*, 1–33.

de Cornière, A. (2016): "Search advertising," *American Economic Journal: Microeconomics*, 8, 156–188.

Downs, A. (1957): *An economic theory of democracy*, Harper & Row New York.

Duffie, D. and J. Pan (1997): "An overview of value at risk," *Journal of derivatives*, 4, 7–49.

Dworczak, P. and G. Martini (2019): "The simple economics of optimal persuasion," *Journal of Political Economy*, 127, 1993–2048.

Dye, R. A. (2002): "Classifications manipulation and Nash accounting standards," *Journal of accounting research*, 40, 1125–1162.

Elliot, M., A. Galeotti, A. Koh, and W. Li (2022): "Market segmentation through information," Working Paper.

Ely, J. C. (2019): "A Cake-cutting solution to gerrymandering," Working Paper. Northwestern University, Evanston, IL.

Fan, Y., A. Barua, W. M. Cready, and W. B. Thomas (2010): "Managing earnings using classification shifting: Evidence from quarterly special items," *Accounting Review*, 85, 1303–1323.

Friedman, J. N. and R. Holden (2020): "Optimal Gerrymandering in a competitive environment," *Economic Theory Bulletin*, 8, 347–367.

Friedman, J. N. and R. T. Holden (2008): "Optimal gerrymandering: Sometimes pack, but never crack," *American Economic Review*, 98, 113–44.

Fryer Jr, R. G. and R. Holden (2011): "Measuring the compactness of political districting plans," *Journal of Law and Economics*, 54, 493–535.

Galati, G. and R. Moessner (2013): "Macroprudential policy–a literature review," *Journal of Economic Surveys*, 27, 846–878.

Gentzkow, M. and E. Kamenica (2016): "A Rothschild-Stiglitz approach to Bayesian persuasion," *American Economic Review: Papers and Proceedings.*

Gerber, E. R. and J. B. Lewis (2004): "Beyond the median: Voter preferences, district heterogeneity, and political representation," *Journal of Political Economy*, 112, 1364–1383.

Gilligan, T. W. and J. G. Matsusaka (2006): "Public choice principles of redistricting," *Public Choice*, 129, 381–398.

Goldberger, A. S. (1991): *A course in econometrics*, Harvard University Press.

Gomberg, A., R. Pancs, and T. Sharma (2021): "Electoral maldistricting," Working Paper. ITAM, Mexico.

Guinsburg, P. V. and G. Q. Saraiva (2022): "Price steering in two-sided platforms," Working Paper. São Paulo University, São Paulo, Brazil.

Gul, F. and W. Pesendorfer (2010): “Strategic redistricting,” *American Economic Review*, 100, 1616–41.

Haghpanah, N. and R. Siegel (2020): “Pareto Improving Segmentation of Multi-product Markets,” Working Paper.

——— (2022): “Pareto improving segmentation of multi-product markets,” Working Paper. Pennsylvania State University, Stage College, PA.

Hagiu, A. and B. Jullien (2011): “Why do intermediaries divert search,” *RAND Journal of Economics*, 42, 337–362.

Hardy, G., J. E. Littlewood, and G. Pólya (1929): “Some simple inequalities satisfied by convex functions,” *Messenger Math*, 58, 145–152.

Harsanyi, J. C. (1967-68): “Games of Incomplete Information Played by Bayesian Players, Part I, II, III,” *Management Science*, 14, 159–182, 320–334, 486–502.

Horowitz, J. L. and C. F. Manski (1995): “Identification and robustness with contaminated and corrupted data,” *Econometrica: Journal of the Econometric Society*, 281–302.

Ichihashi, S. (2020): “Online privacy and information disclosure by consumers,” *American Economic Review*, 110, 569–595.

Kamenica, E. and M. Gentzkow (2011): “Bayesian persuasion,” *American Economic Review*.

Kleiner, A., B. Moldovanu, and P. Strack (2021): “Extreme points and majorization: economic applications,” *Econometrica*, 89, 1557–1593.

Koenker, R. and G. Bassett Jr (1978): “Regression quantiles,” *Econometrica: journal of the Econometric Society*, 33–50.

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

Kolotilin, A., M. Li, T. Mylovanov, and A. Zapechelnyuk (2017): “Persuasion of a Privately Informed Receiver,” *Econometrica*, 85, 1949–1964.

Kolotilin, A. and A. Wolitzky (2020): “The economics of partisan gerrymandering,” Working Paper. Massachusetts Institute of Technology, Cambridge, MA.

Krehbiel, K. (2010): *Pivotal politics*, University of Chicago Press.

Manski, C. F. (1988): "Ordinal utility models of decision making under uncertainty," *Theory and Decision*, 25, 79–104.

McCarty, N., K. T. Poole, and H. Rosenthal (2001): "The hunt for party discipline in Congress," *American Political Science Review*, 95, 673–687.

——— (2009): "Does gerrymandering cause polarization?" *American Journal of Political Science*, 53, 666–680.

McGhee, E. (2014): "Measuring partisan bias in single-member district electoral systems," *Legislative Studies Quarterly*, 39, 55–85.

McVay, S. E. (2006): "Earnings management using classification shifting: An examination of core earnings and special items," *Accounting Review*, 81, 501–531.

Milnor, J. W. (1954): "Games Against Nature," in *Decision Processes*, ed. by R. M. Thrall, C. H. Coombs, and R. L. Davis, New York: Wiley, chap. 4, 49–68.

Orlov, D., P. Zryumov, and A. Skrzypacz (2020): "Design of macro-prudential stress tests," Working paper. Stanford Univeristy.

Owen, G. and B. Grofman (1988): "Optimal partisan gerrymandering," *Political Geography Quarterly*, 7, 5–22.

Pegden, W., A. D. Procaccia, and D. Yu (2017): "A partisan districting protocol with provably nonpartisan outcomes," Working Paper. Carnegie Mellon University, Pittsburgh, PA.

Poole, K. T. and R. S. Daniels (1985): "Ideology, party, and voting in the US Congress, 1959–1980," *American Political Science Review*, 79, 373–399.

Rawls, J. (1971): *A Theory of Justice*, Harvard University Press.

Ridder, G. and R. Moffitt (2007): "The econometrics of data combination," *Handbook of econometrics*, 6, 5469–5547.

Rostek, M. (2010): "Quantile maximization in decision theory," *The Review of Economic Studies*, 77, 339–371.

Rothschild, M. and J. Stiglitz (1970): "Increasing risk: I. A Definition," *Journal of Economic Theory*.

Roy, A. D. (1952): "Safety first and the holding of assets," *Econometrica: Journal of the econometric society*, 431–449.

Ryff, J. V. (1967): "Extreme points of some convex subsets of $L^(0,1)$," *Proceedings of the American Mathematical Society*, 18, 1026–1034.

Sherstyuk, K. (1998): "How to gerrymander: A formal analysis," *Public Choice*, 95, 27–49.

Shotts, K. W. (2001): "The effect of majority-minority mandates on partisan gerrymandering," *American Journal of Political Science*, 120–135.

Stephanopoulos, N. O. and E. M. McGhee (2015): "Partisan gerrymandering and the efficiency gap," *U. Chi. L. Rev.*, 82, 831.

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

Yang, K. H. (2020): "A note on generating arbitrary joint distributions using partitions," Working Paper. Yale University, New Haven, CT.

——— (2022): "Selling consumer data for profit: optimal market-segmentation design and its consequences," *American Economic Review*, 112, 1364–1393.

Zhao, D., D. Zhang, M. Khan, and L. Perrussel (2010): "Maximal matching for double auction," in *Australasian Joint Conference on Artificial Intelligence*, Springer, 516–525.

Appendix

Proof of Lemma 1

Since the CDF of the uniform distribution on $[0,1]$ is in $\mathcal{F}$, $\mathcal{H}_\tau = \mathcal{I}(\underline{F}_0^\tau, \overline{F}_0^\tau)$ for all $F_0 \in \mathcal{F}$ implies $\mathcal{H}_\tau^* \supseteq \mathcal{I}_\tau^*$.

Conversely, suppose that $\mathcal{H}_\tau^* \supseteq \mathcal{I}_\tau^*$. Let $\widetilde{H}(q) := H(F_0^{-1}(q))$ for all $q \in \mathbb{R}$. Then $\widetilde{H} \in \mathcal{I}_\tau^*$. Therefore, there exists $\tilde{\mu} \in \mathcal{M}(U)$ and $\tilde{r} \in \mathcal{R}$ such that

$$\widetilde{H}(q) = H^\tau(q|\tilde{\mu}, \tilde{r}) = \int_{\mathcal{F}} \tilde{r}((-\infty, q]|F)\tilde{\mu}(\mathrm{d}F),$$

for all $q \in [0,1]$. For any $F_0 \in \mathcal{F}$, define μ and r as

$$\mu(A) := \tilde{\mu}(\{F \in \mathcal{F} | F \circ F_0 \in A\}),$$

for all measurable $A \subseteq \mathcal{F}$, and

$$r((-\infty, \omega]|F, \tau) := \tilde{r}((-\infty, F_0(\omega)]|F \circ F_0^{-1}, \tau),$$

for all $\omega \in \mathbb{R}$, for all $\tau \in [0,1]$, and for all $F \in \mathcal{F}$. We claim that $\mu \in \mathcal{M}(U)$ and $r \in \mathcal{R}$. Indeed, for any measurable $A \subseteq \mathcal{F}$, $\mu(A) = \tilde{\mu}(\{F \in \mathcal{F}|F \circ F_0 \in A\}) \geq 0$. Meanwhile, $\mu(\mathcal{F}) = \tilde{\mu}(\{F \in \mathcal{F}|F \circ F_0 \in \mathcal{F}\}) = \tilde{\mu}(\mathcal{F}) = 1$. Furthermore, for any measurable set $A \subseteq \mathcal{F}$, let

$$F_0^{-1} \circ A := \{F_0^{-1} \circ F | F \in A\},$$

and note that $F \circ F_0 \in A$ if and only if $F \in F_0^{-1} \circ A$ for all $F \in \mathcal{F}$. Thus, for any disjoint collection of measurable sets $\{A_n\} \subseteq \mathcal{F}$,

$$\begin{aligned}
\mu\left(\bigcup_{n=1}^\infty A_n\right) =& \tilde{\mu}\left(\left\{F \in \mathcal{F} \middle| F \circ F_0 \in \bigcup_{n=1}^\infty A_n\right\}\right) \\
=& \tilde{\mu}\left(\left\{F \in \mathcal{F} \middle| F \in F_0^{-1} \circ \bigcup_{n=1}^\infty A_n\right\}\right) \\
=& \sum_{n=1}^\infty \tilde{\mu}(F_0^{-1} \circ A_n) \\
=& \sum_{n=1}^\infty \tilde{\mu}(\{F \in \mathcal{F}|F \circ F_0 \in A_n\}) \\
=& \sum_{n=1}^\infty \mu(A_n).
\end{aligned}$$

Together, μ is indeed a probability measure on $\mathcal{F}$. In the meantime, for any $F \in \mathcal{F}$,

$$r((-\infty, F^{-1}(\tau))|F, \tau) = \tilde{r}((-\infty, F_0(F^{-1}(\tau))|F \circ F_0^{-1}, \tau) = 0$$

and

$$r((-\infty, F^{-1}(\tau^+)]|F,\tau) = \tilde{r}((-\infty, F_0(F^{-1}(\tau^+))|F\circ F_0^{-1},\tau) = 1.$$

Thus, $\text{supp}(r(\cdot|F,\tau)) = \mathbb{Q}^\tau(F)$ for all $F \in \mathcal{F}$ and for all $\tau \in [0,1]$, and hence $r \in \mathcal{R}$.

In addition, note that for any $\omega \in \mathbb{R}$,

$$\int_{\mathcal{F}} F(\omega)\mu(\mathrm{d}F) = \int_{\mathcal{F}} F(F_0(\omega))\tilde{\mu}(\mathrm{d}F) = F_0(\omega),$$

which in turn implies that $\mu \in \mathcal{M}$.

As a result, for any $\omega \in \mathbb{R}$,

$$\begin{aligned}
H(\omega) = \widetilde{H}(F_0(\omega)) =& \int_{\mathcal{F}} \tilde{r}((-\infty, F_0(\omega)]|F,\tau)\tilde{\mu}(\mathrm{d}F) \\
=& \int_{\mathcal{F}} \tilde{r}((-\infty, F_0(\omega)]|F\circ F_0^{-1},\tau)\mu(\mathrm{d}F) \\
=& \int_{\mathcal{F}} r((-\infty,\omega]|F,\tau)\mu(\mathrm{d}F) \\
=& H^\tau(\omega|\mu,r).
\end{aligned}$$

Therefore, $H \in \mathcal{H}_\tau$. This completes the proof. ■

Proof of Lemma 2

Embed $\mathcal{I}_\tau^* \subseteq \mathcal{F}$ into the collection $L^1([0,1])$ of integrable functions on $[0,1]$. Note that $\mathcal{I}_\tau^*$ is a convex subset of a normed linear space $L^1([0,1])$. Consider any $H \in \mathcal{I}_\tau^*$ that takes the form of (3), and any $\widehat{H} \in L^1([0,1])$ such that $\widehat{H}(\tilde{\omega}) \neq 0$ for some $\tilde{\omega} \in [0,1]$. Suppose that $H(\tilde{\omega}) \in \{\underline{U}^\tau(\tilde{\omega}), \overline{U}^\tau(\tilde{\omega})\}$. Then clearly either $H(\tilde{\omega}) + \widehat{H}(\tilde{\omega}) > \underline{U}^\tau(\tilde{\omega})$ or $H(\tilde{\omega}) - \widehat{H}(\tilde{\omega}) < \overline{U}^\tau(\tilde{\omega})$ and hence, either $H + \widehat{H} \notin \mathcal{I}_\tau^*$ or $H - \widehat{H} \notin \mathcal{I}_\tau^*$. Meanwhile, suppose that $\tilde{\omega} \in [\underline{x}_i, \overline{x}_i)$ for some $i \in I$ or $\tilde{\omega} \in [\underline{y}_j, \overline{y}_j)$ for some $j \in J$. If either $H + \widehat{H} \notin \mathcal{F}$ or $H - \widehat{H} \notin \mathcal{F}$, then clearly either $H + \widehat{H} \notin \mathcal{I}_\tau^*$ or $H - \widehat{H} \notin \mathcal{I}_\tau^*$. If, on the other hand, both $H + \widehat{H}$ and $H - \widehat{H}$ are in $\mathcal{F}$, then it must be that either $H(\omega) + \widehat{H}(\omega) = \underline{U}^\tau(\underline{x}_i) + \widehat{H}(\tilde{\omega}) > \underline{U}^\tau(\underline{x}_i)$ for all $\omega \in [\underline{x}_i, \overline{x}_i)$, or $H(\omega) - \widehat{H}(\omega) = \overline{U}^\tau(\overline{y}_j) - \widehat{H}(\tilde{\omega}) < \overline{U}^\tau(\overline{y}_j)$, for all $\omega \in [\underline{y}_j, \overline{y}_j)$. Together, there must exist $\hat{\omega} \in \mathbb{R}$ such that either $H(\hat{\omega}) + \widehat{H}(\hat{\omega}) \notin \mathcal{I}_\tau^*$ or $H(\hat{\omega}) - \widehat{H}(\hat{\omega}) \notin \mathcal{I}_\tau^*$.

Conversely, suppose that $H \in \mathcal{I}_\tau^*$ does not take form of (3). Then there exists $\underline{\omega} < \overline{\omega}$ and $\underline{\eta} < \overline{\eta}$ such that $H(\underline{\omega}^-) \leq \underline{\eta} \leq H(\underline{\omega})$, $H(\overline{\omega}^-) \leq \overline{\eta} \leq H(\overline{\omega})$; that $\overline{U}^\tau(\overline{\omega}) \leq \underline{\eta} < \overline{\eta} \leq \underline{U}^\tau(\underline{\omega})$; and that $\underline{\eta} < H(\omega) < \overline{\eta}$ for some $\omega \in (\underline{\omega}, \overline{\omega})$. Then, since the set of extreme points of nondecreasing functions that map from $[\underline{\omega}, \overline{\omega}]$ to $[\underline{\eta}, \overline{\eta}]$ must only take values in $\{\underline{\eta}, \overline{\eta}\}$ (see, for instance, lemma 2.7 of Börgers 2015), there exists a non-zero, integrable function $\widetilde{H} : [\underline{\omega}, \overline{\omega}] \to [\underline{\eta}, \overline{\eta}]$ such that both $H + \widetilde{H}$ and $H - \widetilde{H}$ are nondecreasing, right-continuous functions from $[\underline{\omega}, \overline{\omega}]$ to $[\underline{\eta}, \overline{\eta}]$. As a result, for any $\omega \in [\underline{\omega}, \overline{\omega}]$, it must be that

$$\max\{H(\omega) + \widetilde{H}(\omega), H(\omega) - \widetilde{H}(\omega)\} \leq \overline{\eta} \leq \underline{U}^\tau(\underline{\omega}) \leq \underline{U}^\tau(\omega) \tag{A.11}$$

and that

$$\min\{H(\omega) + \widetilde{H}(\omega), H(\omega) - \widetilde{H}(\omega)\} \geq \underline{\eta} \geq \overline{U}^\tau(\overline{\omega}) \geq \overline{U}^\tau(\omega), \tag{A.12}$$

for all $\omega \in [\underline{\omega}, \overline{\omega}]$. Now let $\widehat{H} : [0,1] \to \mathbb{R}$ be defined as

$$\widehat{H}(\omega) := \begin{cases} \widetilde{H}(\omega), & \text{if } \omega \in [\underline{\omega}, \overline{\omega}] \\ 0, & \text{otherwise} \end{cases},$$

for all $\omega \in [0,1]$. Clearly, $\widehat{H} \in L^1([0,1])$. Moreover, for any $\omega \in [0,1]$, from (A.11) and (A.12), together with $H \in \mathcal{I}_\tau^*$, if follows that

$$\overline{U}^\tau(\omega) \leq \min\{H(\omega) + \widehat{H}(\omega), H(\omega) - \widehat{H}(\omega)\} \leq \max\{H(\omega) + \widehat{H}(\omega), H(\omega) - \widehat{H}(\omega)\} \leq \underline{U}^\tau(\omega),$$

for all $\omega \in [0,1]$. Meanwhile, since $\underline{\eta} \in [H(\underline{\omega}^-), H(\underline{\omega})]$ and $\overline{\eta} \in [H(\overline{\omega}^-), H(\overline{\omega})]$, it must be that

$$H(\omega) + \widehat{H}(\omega) = H(\omega) - \widehat{H}(\omega) = H(\omega) \leq H(\underline{\omega}^-) \leq \underline{\eta},$$

for all $\omega \leq \underline{\omega}$; while

$$H(\omega) + \widehat{H}(\omega) = H(\omega) - \widehat{H}(\omega) = H(\omega) \geq H(\overline{\omega}) \geq \overline{\eta},$$

for all $\omega \geq \overline{\omega}$. As a result, both $H + \widehat{H}$ and $H - \widehat{H}$ are nondecreasing and right-continuous. Together, it then follows that $H + \widehat{H} \in \mathcal{I}_\tau^*$ and $H - \widehat{H} \in \mathcal{I}_\tau^*$, and hence H is not an extreme point of $\mathcal{I}_\tau^*$. This completes the proof. ■

Proof of Theorem 1

To show that $\mathcal{H}_\tau \subseteq \mathcal{I}(\underline{F}_0^\tau, \overline{F}_0^\tau)$, consider any $H \in \mathcal{H}_\tau$. Let $\mu \in \mathcal{M}$ and any $r \in \mathcal{R}$ be a signal and a selection rule such that $H^\tau(\cdot|\mu, r) = H$. By the definition of $H^\tau(\cdot|\mu, r)$, it must be that, for all $\omega \in \mathbb{R}$,

$$H(\omega|\mu, r) \leq \mu(\{F \in \mathcal{F} | F^{-1}(\tau) \leq \omega\}) = \mu(\{F \in \mathcal{F} | F(x) \geq \tau\}).$$

Now consider any $\omega \in \mathbb{R}$. Clearly, $\mu(\{F \in \mathcal{F} | F(\omega) \geq \tau\}) \leq 1$ since μ is a probability measure. Moreover, let $M_\omega^+(q) := \mu(\{F \in \mathcal{F} | F(\omega) \geq q\})$ for all $q \in [0,1]$. From (1), it follows that the left-limit of $1 - M_x^+$ is a CDF and a mean-preserving spread of a Dirac measure at $F_0(\omega)$. Therefore, whenever $\tau \geq F_0(\omega)$, then $M_\omega^+(\tau)$ can at most be $F_0(\omega)/\tau$ to have a mean of $F_0(\omega)$.[15] Together, this implies that $\mu(\{F \in \mathcal{F} | F(x) \geq \tau\}) \leq \underline{F}_0^\tau(\omega)$ for all $\omega \in \mathbb{R}$.

At the same time, by the definition of $H^\tau(\cdot|\mu, r)$, it must be that, for all $\omega \in \mathbb{R}$,

$$H^\tau(\omega^-|\mu, r) \geq \mu(\{F \in \mathcal{F} | F^{-1}(\tau^+) < \omega\}) = \mu(\{F \in \mathcal{F} | F(x) > \tau\}).$$

Consider any $\omega \in \mathbb{R}$. Since μ is a probability measure, it must be that $\mu(\{F \in \mathcal{F} | F(\omega) > \tau\}) \geq 0$. Furthermore, let $M_\omega^-(q) := \mu(\{F \in \mathcal{F} | F(\omega) > q\})$ for all $q \in [0,1]$. From (1), it follows that $1 - M_x^-$ is a CDF and a mean-preserving spread of a Dirac measure at $F_0(\omega)$. Therefore, whenever $\tau \leq F_0(\omega)$, then $M_\omega^-(\tau)$ must be at least $(F_0(\omega) - \tau)/(1 - \tau)$ to have a mean of $F_0(\omega)$.[16] Together, this implies that $\mu(\{F \in$

[15] More specifically, to maximize the probability at τ, a mean-preserving spread of $F_0(\omega)$ must assign probability $F_0(\omega)/\tau$ at τ, and probability $1 - F_0(x)/\tau$ at 0.

[16] More specifically, to minimize the probability at τ, a mean-preserving spread of $F_0(x)$ must assign probability

$\mathcal{F}|F(\omega) > \tau\}) \geq \overline{F}_0^\tau$ for all $\omega \in \mathbb{R}$, which, in turn, implies that $\overline{F}_0^\tau(\omega) \leq H^\tau(\omega^-|\mu, r) \leq H^\tau(\omega|\mu, r) \leq \underline{F}_0^\tau(\omega)$ for all $\omega \in \mathbb{R}$, as desired.

To prove that $\mathcal{I}(\underline{F}_0^\tau, \overline{F}_0^\tau) \subseteq \mathcal{H}_\tau$, by Lemma 1, it suffices to show that $\mathcal{I}_\tau^* \subseteq \mathcal{H}_\tau^*$. To this end, we first show that for any extreme point H of $\mathcal{I}_\tau^*$, there exists a signal $\tilde{\mu} \in \mathcal{M}(U)$ and a selection rule $\tilde{r} \in \mathcal{R}$ such that $H(\omega) = H^\tau(\omega|\mu, r)$ for all $\omega \in \mathbb{R}$. Consider any extreme point H of $\mathcal{I}_\tau^*$. By Lemma 2, H must take the form of (3), for some $\underline{x}, \overline{x}, \underline{y}, \overline{y} \in [0, 1]$ and countable sequences $\{\underline{x}_i, \overline{x}_i\}_{i \in I}$ and $\{\underline{y}_j, \overline{y}_j\}_{j \in J}$ such that $\underline{x} \leq \underline{x}_i \leq \overline{x}_i \leq \underline{x}_{i+1} \leq \overline{x} < \underline{y} \leq \underline{y}_j \leq \overline{y}_j \leq \underline{y}_{j+1} \leq \overline{y}$ for all $i \in I, j \in J$. Now define two classes of distributions, $\{\underline{U}^\omega\}_{\omega \in [0, \overline{x}]}$ and $\{\overline{U}^\omega\}_{\omega \in [\underline{y}, 1]}$ as follows:

$$\underline{U}^\omega(x) := \begin{cases} 0, & \text{if } x < \omega \\ \frac{\overline{x}}{\underline{y} - \tau + \overline{x}}, & \text{if } x \in [\omega, \tau) \\ \frac{x - \tau + \overline{x}}{1 - \underline{y} + \overline{x}}, & \text{if } x \in [\tau, \underline{y}) \\ 1, & \text{if } x \geq \underline{y} \end{cases} ; \text{ and } \overline{U}^\omega(x) := \begin{cases} 0, & \text{if } x < \overline{x} \\ \frac{x - \overline{x}}{1 - \underline{y} + \tau - \overline{x}}, & \text{if } x \in [\overline{x}, \tau) \\ \frac{\tau - \overline{x}}{1 - \underline{y} + \tau - \overline{x}}, & \text{if } x \in [\overline{x}, \omega) \\ 1, & \text{if } x \geq \omega \end{cases} .$$

Note that since $\underline{U}^\tau(\overline{x}) = \overline{U}^\tau(\underline{y})$, it follows that $(1 - \tau)\overline{x} = \tau(\underline{y} - \tau)$, and hence, $\underline{U}^\omega(x) = \tau$ for all $x \in [\omega, \underline{y}]$ and $\overline{U}^\omega(x) = \tau$ for all $x \in [\overline{x}, \omega)$. As a result, $\mathbb{Q}^\tau(\underline{U}^\omega) = [\omega, \underline{y}]$ for all $\omega \in [0, \overline{x}]$ and $\mathbb{Q}^\tau(\overline{U}^\omega) = [\overline{x}, \omega]$ for all $\omega \in [\underline{y}, 1]$. Moreover, for any $i \in I$ and for any $j \in J$, let $\underline{U}^i$ and $\overline{U}^j$ be defined as

$$\underline{U}^i(x) := \frac{1}{\overline{x}_i - \underline{x}_i} \int_{\underline{x}_i}^{\overline{x}_i} \underline{U}^\omega(x) \, \mathrm{d}\omega; \text{ and } \overline{U}^j(x) := \frac{1}{\overline{y}_j - \underline{y}_j} \int_{\underline{y}_j}^{\overline{y}_j} \overline{U}^\omega(x) \, \mathrm{d}\omega,$$

for all $x \in \mathbb{R}$. Notice that, by construction, $\underline{U}^i, \overline{U}^j \in \mathcal{F}$ and $\overline{x}_i \in \mathbb{Q}^\tau(\underline{U}i)$, $\underline{y}_j \in \mathbb{Q}^\tau(\overline{U}^j)$, for all $i \in I$ and $j \in J$. For any $\omega \in \mathrm{supp}(H)$, let $F_\omega \in \mathcal{F}$ be defined as[17]

$$F_\omega(x) := \begin{cases} \underline{U}^\omega(x), & \text{if } \omega \in [0, \overline{x}] \backslash \cup_{i \in I} [\underline{x}_i, \overline{x}_i] \\ \underline{U}^i(x), & \text{if } \omega \in [\underline{x}_i, \overline{x}_i] \\ \overline{U}^\omega(x), & \text{if } \omega \in [\underline{y}, 1] \backslash \cup_{j \in J} [\underline{y}_j, \overline{y}_j] \\ \overline{U}^j(x), & \text{if } \omega \in [\underline{y}_j, \overline{y}_j] \end{cases} ,$$

for all $x \in \mathbb{R}$.

Now define $\tilde{\mu}$ as

$$\tilde{\mu}(\{F_\omega \in \mathcal{F} | \omega \leq x\}) := H(x),$$

for all $x \in \mathbb{R}$. By construction, $\mathrm{supp}(\tilde{\mu}) = \{\underline{U}^\omega\}_{\omega \in [0, \overline{x}] \backslash \cup_{i \in I} [\underline{x}_i, \overline{x}_i]} \cup \{\underline{U}^i\}_{i \in I} \cup \{\overline{U}^\omega\}_{\omega \in [\underline{y}, 1] \backslash \cup_{j \in J} [\underline{y}_j, \overline{y}_j]} \cup \{\overline{U}^j\}_{j \in J}$. Furthermore, for any $x \in [0, 1]$,

$$\int_{\mathcal{F}} F(x) \tilde{\mu}(\mathrm{d}F) = \int_0^1 F_\omega(x) H(\mathrm{d}\omega) = x,$$

$(F_0(\omega) - \tau)/(1 - \tau)$ at 1, and probability $1 - (F_0(\omega) - \tau)/(1 - \tau)$ at 0.

[17] As a convention, define $\underline{U}^i(x) := \underline{U}^\omega(x)$ for all x if $\underline{x}_i = \overline{x}_i = \omega$. Similarly, define $\overline{U}^j(x) := \overline{U}^\omega(x)$ for all x if $\underline{y}_j = \overline{y}_j = \omega$.

and hence $\tilde{\mu} \in \mathcal{M}(U)$. In the meantime, let $\tilde{r} : \mathcal{F} \to [0,1] \to \Delta(\mathbb{R})$ be defined as

$$\tilde{r}(F, \tau') := \begin{cases} \delta_{\{\max(\mathbb{Q}^{\tau'})\}}, & \text{if } F = F_\omega,\, \omega \in [\underline{y}, 1] \\ \delta_{\{\min(\mathbb{Q}^{\tau'})\}}, & \text{otherwise} \end{cases},$$

for all $F \in \mathcal{F}$ and for all $\tau' \in [0,1]$. Then, for all $x \in \mathbb{R}$,

$$H^\tau(x|\tilde{\mu}, \tilde{r}) = \begin{cases} 0, & \text{if } x < 0 \\ \tilde{\mu}(F_\omega | F_\omega^{-1}(\tau) \leq x), \text{if } x \in [0, \overline{x}) \\ \tilde{\mu}(F_\omega | F_\omega^{-1}(\tau^+) \leq x), & \text{if } x \in [\underline{y}, 1) \\ 1, & \text{if } x \geq 1 \end{cases}$$

for all $x \in \mathbb{R}$, and hence $H^\tau(\omega|\tilde{\mu}, \tilde{r}) = H(\omega)$ for all $\omega \in \mathbb{R}$, as desired.

Lastly, consider any $H \in \mathcal{I}_\tau^*$. Note that to show there exists $\tilde{\mu} \in \mathcal{M}(U)$ and $\tilde{r} \in \mathcal{R}$ such that $H(\omega) = H^\tau(\omega|\tilde{\mu}, \tilde{r})$ for all $\omega \in \mathbb{R}$, it suffices to show that there exists $\tilde{\mu} \in \mathcal{M}(U)$ such that

$$\tilde{\mu}(\{F \in \mathcal{F} | F^{-1}(\tau^+) < \omega\}) \leq H(\omega) \leq \tilde{\mu}(\{F \in \mathcal{F} | F^{-1}(\tau) \leq \omega\}), \tag{A.13}$$

for all $\omega \in \mathbb{R}$. To this end, first notice that by the Krein-Milman theorem, H belongs to the closed convex hull of extreme points of $\mathcal{I}_\tau^*$. Thus, there exists a sequence $\{H_n\} \subseteq \mathcal{I}_\tau^*$ such that for all $n \in \mathbb{N}$, $H_n = \sum_{i=1}^{I_n} \lambda_i H_i^n$ for some extreme points $\{H_i\}_{i=1}^{I_n}$ of $\mathcal{I}_\tau^*$ and $\{H_n\} \to H$ under the L^1 norm. By possibly taking a subsequence, it follows that $\{H_n\} \to H$ under the weak-* topology. For any $n \in \mathbb{N}$ and for any $i \in \{1, \dots, I_n\}$, since H_i^n is an extreme point of $\mathcal{I}_\tau^*$, there exists $\{\mu_i^n\}_{i=1}^{I_n} \subseteq \mathcal{M}(U)$ and $\{r_i^n\}_{i=1}^{I_n} \subseteq \mathcal{R}$ such that $H_n(\omega) = \sum_{i=1}^{I_n} H_i(\omega) = \sum_{i=1}^{I_n} H^\tau(\omega|\mu_i^n, r_i^n) = H^\tau(\omega|\mu_n, r_n)$, where $\mu_n := \sum_{i=1}^{I_n} \mu_i^n \in \mathcal{M}(U)$ and $r_n := \sum_{i=1}^{I_n} r_i \in \mathcal{R}$, and the last equality follows from the definition of H^τ. Meanwhile, by the definition of $H^\tau(\cdot|\mu, r)$, for any $n \in \mathbb{N}$, we must have

$$H_n(\omega) = H(\omega|\mu_n, r_n) \leq \mu_n(\{F \in \mathcal{F} | F^{-1}(\tau) \leq \omega\}) = \mu_n(\{F \in \mathcal{F} | F(\omega) \geq \tau\}),$$

for all $\omega \in \mathbb{R}$, and

$$H_n(\omega^-) = H(\omega^-|\mu_n, r_n) \geq \mu_n(\{F \in \mathcal{F} | F^{-1}(\tau) < \omega\}) = \mu_n(\{F \in \mathcal{F} | F(x) > \tau\}),$$

for all $\omega \in \mathbb{R}$.

Notice that, for any $\{F_n\} \subset \mathcal{F}$ and $\omega \in \mathbb{R}$ such that $F_n \to F$ under the weak-* topology and $F_n(\omega) \geq \tau$ for all $n \in \mathbb{N}$,

$$\tau \leq \lim_{n \to \infty} F_n(\omega) \leq \limsup_{n \to \infty} F_n(\omega) \leq F(\omega).$$

Hence, the set $\{F \in \mathcal{F} | F(\omega) \geq \tau\}$ is closed in $\mathcal{F}$. By similar arguments, the set $\{F \in \mathcal{F} | F(\omega) > \tau\}$ is open in $\mathcal{F}$. In addition, since $\mathcal{M}(U)$ is compact under the weak-* topology, by possibly taking a subsequence, there exists $\tilde{\mu} \in \mathcal{M}(U)$ such that $\{\mu_n\} \to \tilde{\mu}$. Together, for any $\omega \in \mathbb{R}$ at which H is continuous,

$$
\begin{aligned}
H(\omega) = \lim_{n\to\infty} H_n(\omega) = \lim_{n\to\infty} H^{\tau}(\omega|\mu_n, r_n) &\leq \limsup_{n\to\infty} \mu_n(\{F \in \mathcal{F}|F(\omega) \geq \tau\}) \\
&= \tilde{\mu}(\{F \in \mathcal{F}|F(\omega) \geq \tau\}) \\
&= \tilde{\mu}(\{F \in \mathcal{F}|F^{-1}(\tau) \leq \omega\}),
\end{aligned}
$$

and

$$
\begin{aligned}
H(\omega) = \lim_{n\to\infty} H_n(\omega^-) = \lim_{n\to\infty} H_n(\omega^-|\mu_n, r_n) &\geq \liminf_{n\to\infty} \mu_n(\{F \in \mathcal{F}|F(\omega) > \tau\}) \\
&\geq \tilde{\mu}(\{F \in \mathcal{F}|F(\omega) > \tau\}) \\
&= \tilde{\mu}(\{F \in \mathcal{F}|F^{-1}(\tau^+) < \omega\}).
\end{aligned}
$$

Moreover, since $\omega \mapsto \tilde{\mu}(\{F \in \mathcal{F}|F^{-1}(\tau) \leq \omega\})$ is right-continuous and $\omega \mapsto \tilde{\mu}(\{F \in \mathcal{F}|F^{-1}(\tau^+) < \omega\})$ is left-continuous, it must be that

$$\tilde{\mu}(\{F \in \mathcal{F}|F^{-1}(\tau^+) < \omega\}) \leq H(\omega^-) \leq H(\omega) \leq \mu(\{F \in \mathcal{F}|F^{-1}(\tau) \leq \omega\}),\ \forall \omega \in \mathbb{R},$$

which implies (A.13) as desired. This completes the proof. ■