Inference on Auctions with Weak Assumptions on Information*

Vasilis Syrgkanis[†]   Elie Tamer[‡]   Juba Ziani[§]

May 17, 2021

Abstract

Given a sample of bids from independent auctions, this paper examines the question of inference on auction objects (like valuation distributions, welfare measures, etc) under weak assumptions on information. We leverage the recent contributions of Bergemann and Morris [2013] in the robust mechanism design literature that exploit the link between Bayesian Correlated Equilibria and Bayesian Nash Equilibria in incomplete information games, to construct an econometrics framework that is computationally feasible and robust to assumptions about information. Checking whether a particular valuation distribution belongs to the identified set is as simple as determining whether a *linear* program (LP) is feasible. This is the key characteristic of our framework. A similar LP can be used to learn about various welfare measures and policy counterfactuals. For inference and to summarize statistical uncertainty, we propose novel finite sample methods using tail inequalities that are used to construct confidence sets on identified sets. Monte Carlo experiments show adequate finite sample properties. We illustrate our approach by applying our methods to a data set from search Ad auctions and to data from OCS auctions.

Keywords: inference, auctions, Bayes-correlated equilibrium, information robustness

1   Introduction

A recent literature in robust mechanism design studies the following question: given a game (e.g. an auction), what are the possible outcomes - such as welfare or revenue - that arise under different information structures? This literature is motivated by *robustness*, i.e., characterizing outcomes that can occur in a given game under weak assumptions on information (See Bergemann and Morris [2013]). For example, in auction models, in addition to specifying the details of the game in terms of bidder utility function, and bidding rules, one needs to specify the information structure (what do players' information set contain, i.e. what do they know about the state of the world and the information possessed by other players) to be able to derive the Nash equilibrium.

In particular, in the Independent Private Values (IPV) setting, players know their value for the item, which is assumed independent from other player values, and receive no further information about their opponents' values. The latter typically yields a unique equilibrium outcome. However, different assumptions on what signals players have about opponent values prior to bidding in the auction, lead to different equilibrium outcomes. Auction data rarely contain information on what bidders knew and what their information sets included and did not include, and given that this information leads to different outcomes, it would be interesting to analyze what can happen when we relax the independence assumption in such auctions by allowing bidders to know some information about their opponents' valuations. Bergemann, Brooks, and Morris [2017] (BBM) examine exactly this question in an auction game and provide achievable bounds on various outcomes, such as the revenue of the auction as a function of auction fundamentals like the distribution of the common value.

In this paper, we address the following econometrics question, which is the reverse of the one posed by BBM above: given an i.i.d. (or exchangeable) sample of auction data (independent copies of bids from a set of similar auctions), what can we learn (econometrically) in a computationally feasible way about auction funda-

mentals, such as the distribution of values, or policy counterfactuals when we make weak assumptions on the information structure? We maintain throughout that players play Bayesian Nash equilibrium (BNE) but allow these bidders to have different information structures in different auctions, i.e. receive different types of signals prior to bidding. In particular, we use these observations (bids and other observables) from a set of independent auctions to construct *sets* of valuation distributions that are consistent with both the observed distribution of bids (the data) and the known auction rules maintaining that players are Bayesian. We exploit the robust predictions in a given auction à la BBM to conduct econometrically robust predictions of auction fundamentals given a set of data; i.e., robust economic prediction leads to robust inference.

Key to our approach -and the main contribution of this paper- is the characterization of sharp sets of valuation distributions (and other functionals of interest) via computationally attractive procedures based on solving Linear Programs (LP) that work naturally with sample data. This linear in parameter structure that characterizes equilibrium is a result of the equivalence between a particular class of *Bayes Correlated Equilibria* (or BCE) and Bayes Nash Equilibria (or BNE) for a similar game with an arbitrary information structure. It is well known that BCEs can be computed efficiently since they are solutions to linear programs (as opposed to BNE which are hard to compute). Moreover, exploiting a result of Bergemann and Morris [2016] (see also Aumann [1987]), we show that there is an equivalence between the set of fundamentals that obey the BCE restrictions and the fundamentals that obey the BNE constraints under *some* information structure. This equivalence is the key to our econometrics approach in that it forms the linear program (LP) that is the main engine for our computations. In particular, this LP structure from theory is inherited by the econometrics problem in that the formal statistical program that ensues is one where the sharp set satisfies a set of linear equality and inequality constraints at the

true bid distribution (the distribution across these iid observations). If we knew this true distribution of bids, then we can solve the LP to obtain the sharp sets. We do not observe the true bid distribution, but this distribution can be estimated from the observed data. We are also able to characterize sampling uncertainty to obtain various notions of confidence regions covering the identified with a pre-specified probability.

Importantly, we address the problem of counterfactual estimation: for example, what would the revenue or surplus have been had we changed the auction rules? We show how bounds on counterfactuals can be constructed using a modified LP that also takes into account statistical uncertainty from sample data and we apply this to an Adsearch data set. We formulate notions of informationally robust counterfactual analysis and we show that such counterfactual questions can also be phrased as solutions to a single linear program, simultaneously capturing equilibrium constraints in the current auction as well as the new target auction. We also show that even without recovering the information structure from the data, an analyst can perform robust counterfactual analysis and answer for example the following question: under an arbitrary information structure in the current auction which produced the data at hand, what is the best and worst value of a given quality measure (e.g. welfare, revenue) in the new target auction under an arbitrary information structure? Our counterfactual analysis is also permissive in that the information structure of the bidders' is allowed to change between the auction that generated the data and the target counterfactual auction we are evaluating. We note though that our approach is also amenable to enforcing consistency constraints among the two settings as was shown in the very recent work of Bergemann, Brooks, and Morris [2019]. Enforcing consistency of the information structure boils down to adding a set of extra linear constraints in the counterfactual linear program that we formulate (see Section 2.3 for more details). We also show how signal constraints can be used to narrow the size of the identified set. In particular, we use two types of constraints: the first bounds the maximum deviation of the signal from the true state, and the second uses a functional form

assumption on the conditional distribution of signals given the true state.

Finally we apply our inference framework to two data sets. In the first, we analyze a large scale data set on auctions (millions of observations) from the BingAds sponsored search auction marketplace. These data concern sales of search queries to a set of bidders using the generalized second price auction format. We derive this optimization based framework for these auctions and showcase how our approach to inference can be used to learn about bidder valuations for a set of keywords (bidder/keyword specific valuations) and also conduct a counterfactual to examine revenue and welfare under a first price auction format. The second application uses the familiar to economists OCS auction data set and examine there inference on underlying valuation distributions.

Literature. The closest work to our paper is Bergemann, Brooks, and Morris 2017, where the authors provide worst-case bounds on the revenue of a common value first price auction as a function of the distribution of values. Their approach does not use the bid distribution as input, unlike our approach which obtains an estimate of the bid distribution from the data. The main approach in BBM is to show that the revenue cannot be too small since at BCE, no player wants to deviate to any other action and so players do not want to deviate to a specific type of a deviation which is the following: conditional on your bid, deviate uniformly at random above your bid (upwards deviation). Hence, the bound on the mean they provide uses a *subset* of the set of best response deviations that are allowed so as to bound the bid of a player as a function of his value. In drawing a connection between the equilibrium bid and the value, this bound is by definition loose (and can be very loose - bound twice as large as identified set - as we show in an example in Appendix A). Given data, we are able to learn the bid distribution and hence are not constrained to look at only these bid-distribution-oblivious upwards deviations. We can instead compute an optimal deviating bid for this given bid distribution and use the constraint that the player does not want to deviate to this distribution-tailored action. This allows us to bound

the unobserved value of the player as a function of the observed bid leading to a *sharp characterization* of auction fundamentals using the data. Also, the approach taken to inference in this paper is deliberately conservative in that we try to make weak or no assumptions on information while maintaining Nash behavior. This is in the same spirit as Haile and Tamer [2003] who study the question of inference in English auctions under minimal assumptions and derive estimable bounds on the distribution of bidder valuations.

Another paper that uses a similar insight to study the econometrics of games with weak information is the recent interesting work of Magnolfi and Roncoroni [2016] on inference in entry game models. They demonstrate identification of parameters of interaction when both players choose to enter a market. Also, a defining characteristic of our setup (which can handle a large class of games including discrete games) is the linear program that defines the BCE set. We find that this structure is attractive in reducing the computational burden (especially with large data sets) and allowing for more direct inference methods on primitives and also on policy counterfactuals, especially when the models are partially identified (as they typically are in our setup). The LP also is amenable directly to statistical inference procedures that also exploit the LP structure to build confidence regions on functionals of the identified sets. On the other hand, Magnolfi and Roncoroni [2016] provide a characterization of the BNEs of the game through the use of support functions that characterize convex sets and hence estimate the parameters of the game using methods from the random set literature. We view this work on auctions and that of Magnolfi and Roncoroni [2016] on entry models as complementary.

The paper is organized as follows. Section 2 introduces the problem and provides formal definitions of the objects of interest. We then state our identification results given an i.i.d. set of data on bids. This identification is constructed via a linear program where we show how various constraints (such as symmetry, parametric re-

6

strictions, etc) can be incorporated. We also show how computing sharp sets for the expected value of moments of the fundamentals amounts to solving two linear programs and how robust counterfactual analysis of some metric function, with respect to changes in the auction, can also be handled in a computationally efficient manner. We then provide two example applications of the general setup: one for common value auctions (Section 3) and another for private value auctions (Section 4). Section 5 provides our estimation approach for constructing confidence intervals on the estimated quantities from sampled datasets using sub-sampling methods and finite sample concentration inequality approaches. Section 6 examines the finite sample performance of the large linear program using a set of Monte Carlo simulations. These show adequate performance in IPV and Common Value (CV) setups. Section 7 illustrates our inference approach using auction data from AdSearch keywords auctions while Section 8 uses OCS wildcat oil auctions. In these sections, we show how the statistical algorithm can be used to derive bounds on valuation distributions and on counterfactual measures using both nonparametric and parametric assumptions on valuations. We also highlight the role signal constraints play. Finally, the Appendix contains results on the sharpness of the BBM bounds, and bounds on the mean of the valuation in common value auctions with different smoothness assumptions.

2 Bayes-Correlated Equilibria and Information Structure Uncertainty

We consider a game of incomplete information among n players. There is an unknown payoff-relevant state of the world $\theta \in \Theta$. This state of the world enters directly in each player's utility. Each player i chooses among a set of actions A_i and receives utility which is a function of the payoff-relevant state of the world θ and the action profile of the players $\mathbf{a} \in A \equiv A_1 \times \ldots \times A_n$: $u_i(\mathbf{a}; \theta)$. This along with a prior on θ (defined below) will represent the game structure that we denote by G, as separate from the information structure which we will define next.

Conditional on the state of the world each player receives some minimal signal $t_i \in T_i$. The state of the world $\theta \in \Theta$ and the vector of signals $\mathbf{t} \in T \equiv T_1 \times \ldots \times T_n$ are drawn from some joint measure[1] $\pi \in \Pi \subseteq \Delta(\Theta \times T)$. The signals t_is can be arbitrarily correlated with the state of the world, subject only that the joint density lying in the space Π.[2] We denote such signal structure with S. This defines the game (G, S).[3]

We consider a setting where prior to picking an action each player receives some additional information in the form of an extra signal $t'_i \in T'_i$. The signal vector $t' = (t'_1, \ldots, t'_n)$ can be arbitrarily correlated with the true state of the world and with the original signal vector $t = (t_1, \ldots, t_n)$. We denote such augmenting signal structure with S' and the set of all possible such augmenting signal structures with $\mathcal{S}'$. This will define a game (G, S'). Subsequent to observing the signals t_i and t'_i, the player picks an action a_i. A Bayes-Nash equilibrium or BNE in this game (G, S') is a mapping $\sigma_i : T_i \times T'_i \to \Delta(A_i)$ from the pair of signals t_i, t'_i to a distribution over actions, for each player i, such that each player maximizes his expected utility conditional on the signals s/he received.

A fundamental result in the literature on robust predictions (see Bergemann and Morris [2013, 2016]), is that the set of joint distributions of outcomes $a \in A$, unknown states θ and signals t, that can arise as a BNE of an incomplete information game under an arbitrary additional information structure in (G, S'), is equivalent to the set

---

[1]The setup is general in that the set T is unrestricted.

[2]For instance, in the case of private values, Π can encode the fact that players receive at least their private type as signal: let $\theta = (v_1, \ldots, v_n)$ and $t = (t_1, \ldots, t_n)$ and Π is defined by the constraints $\pi(\theta, t) = 0$ for all θ, t, such that $\theta_i \neq t_i$. Another special case of the set Π is when the conditional density $q(t \mid \theta)$ of t given θ is known. In this case Π is defined by the set of linear constraints: $\pi(\theta, t) = q(t \mid \theta)\pi(\theta)$ and $\sum_\theta \pi(\theta) = 1$. In this sense, restrictions on the information structure can be handled by placing constraints on the set Π.

[3]We will assume throughout that the set of n players is fixed and known. If in the data the set of participating players varies across auctions and players do not necessarily know the number or bidders, then we can consider the superset of all players and simply assign a special type to each non-participating player. Conditional on this type, players always choose a default action (e.g. bidding zero in an auction). Then dependent on whether players observe the number of participants before submitting an action can be encoded by whether the players receive as part of their default signal, whether some player's type is the special non-participating type. In our auction applications we will assume that players do not necessarily observe the entrants in the auction before bidding.

of Bayes-Correlated Equilibria, or BCE in (G, S). So, every BCE in (G, S) is a BNE in (G, S') for some augmenting information structure S'. We give the formal definition of BCE next.

Definition 1 (Bayes-Correlated Equilibrium). *A joint distribution $\psi \in \Delta(\Theta \times T \times A)$ is a Bayes-correlated equilibrium of (G, S) if for each player i, signal t_i, action a_i and deviating action a_i':*

$$\mathbb{E}_{(\theta, t, a) \sim \psi} [u_i(a; \theta) \mid a_i, t_i] \geq \mathbb{E}_{(\theta, t, a) \sim \psi} [u_i(a_i', a_{-i}; \theta) \mid a_i, t_i] \tag{1}$$

and such that the marginals with respect to signals and payoff states are preserved, i.e.:

$$\forall \theta \in \Theta, t \in T : \sum_{a \in A} \psi(\theta, t, a) = \pi(\theta, t) \tag{2}$$

An equivalent and simpler way of phrasing the Bayes-correlated equilibrium conditions is that:

$$\forall t_i, a_i, a_i' : \sum_{\theta, t_{-i}, a_{-i}} \psi(\theta, t, a) \cdot \partial u_i(a_i', a; \theta) \geq 0 \tag{3}$$

where we define:

$$\partial u_i(a_i', a; \theta) := u_i(a; \theta) - u_i(a_i', a_{-i}; \theta) \tag{4}$$

to be the utility loss for player i when deviating to action a_i', while the current action profile is a and the state of the world is θ. We state the main result in Bergemann and Morris [2016] next.

Theorem 1 (Bergemann and Morris [2016]). *A distribution $\psi \in \Delta(\Theta \times T \times A)$ can arise as the outcome of a Bayes-Nash equilibrium under some augmenting information structure $S' \in \mathcal{S}'$, if and only if it is a Bayes-correlated equilibrium in (G, S).*

The robustness property of this result is as follows. The set of BNE for (G, S') (think of an auction with unknown information) is the same as the set of BCE for (G, S)

where S' is an augmented information structure derived from S. So, we will not need to know what is in S', but rather we could compute the set of BCE for (G, S) and the Theorem shows that for each BCE, there exists a corresponding information structure S' in $\mathcal{S}'$ and a BNE of the game $(G, \mathcal{S}')$ that implements the same outcome.

2.1 The Econometric Inference Question

We consider the question of inference on auction fundamentals using data under weak assumptions on information. In particular, assume we are given a sample of observations of action profiles $a^1, \ldots, a^N$ from an incomplete information game G. Assume that we do not know the exact augmenting signal structure $S^1, \ldots, S^N$ that occurred in each of these samples where it is implicitly maintained that S^i can be different from S^j for $i \neq j$, i.e., the signal structure in the population is drawn from some unknown mixture. Also, maintaining that players play Nash, or that a^t was the outcome of some Bayes-Nash equilibrium or BNE under signal structure S^t, we study the question of inference on the distribution π of the fundamentals of the game and counterfactuals of interest.

The key question for our approach is to allow for observations on different auctions to use different (and unobserved to the econometrician) information structures, and, given the information structures, that different markets or observations on auctions to come from a different BNE. Given this equivalence between set of BNEs under some information structure and the set of BCE, and given that the set of BCE is convex, it is possible to allow for this kind of heterogeneity. Heuristically, given a distribution over action profiles $\phi \in \Delta(A)$ (which is constructed using the data), there exists a mixture of information structures and equilibria under which ϕ was the outcome. The process by which we arrived at ϕ is by first picking an information structure from this mixture and then selecting one of the Bayes-Nash equilibria for this information structure. This is possible if and only if there exists a distribution $\psi(\theta, t, a)$, that is a

Bayes-correlated equilibrium and such that $\sum_{\theta,t} \psi(\theta, t, a) = \phi(a)$ for all $a \in A$. Again, we start with the elementary information structure S and maintain that observations in the data are expansions of this information structure. So, for a given market i, any BNE using information structure S^i is a BCE under S. The data distribution of action is a mixture of such BNEs over various information structures and hence it would map into a mixture of BCEs under the same S. Since the set of BCEs under S is convex, any mixtures of elements in the set is also a BCE. So then the set of primitives that are consistent with the model and the data is the set of BCEs, $\psi(\theta, t, a)$ such that $\sum_{\theta,t} \psi(\theta, t, a) = \phi(a)$ for all $a \in A$.

To conclude, the convexity of the set of BCEs allows us to relate a distribution of bids from an iid sample to a mixture of BCEs. This is possible since the distribution of bids uses a mixture of signal structures, which essentially coincides with a mixture of BCEs, itself another BCE by convexity. We summarize this discussion with a formal result.

Lemma 2. *Consider a model which conditional on the unobservables, yields a convex set of possible predictions on the observables. Then the sharp identified set of the unobservables under the assumption that exactly one of these predictions is selected in our dataset, is identical to the sharp identified set under the assumption that our dataset is a mixture of selections from these predictions.*

Proof. Suppose that what we observe in the data is a convex combination of feasible predictions of our model. Then by convexity of the prediction set, this convex combination is yet another feasible prediction of our model. Hence, this is equivalent to the assumption that this single prediction is selected in our dataset. ∎

Next, we provide the main engine that allows for construction of the observationally

equivalent set of primitives that obey model assumptions and result in a distribution on the observables that match that with the data. We state this as a Result.

Result 3. *Let there be a distribution ϕ defined on the space of action profiles. Given the setup and results above, the set of feasible joint distributions of signals and types $\pi \in \Delta(\Theta \times T)$ that are consistent with ϕ is the set of distributions $\pi(.;.)$ for which the following linear program is feasible:*

$$\mathbf{LP}(\phi, \pi) \qquad \forall t_i, a_i, a_i' : \qquad \sum_{\theta, t_{-i}, a_{-i}} \phi(a) \cdot x(\theta, t|a) \cdot \partial u_i(a_i', a; \theta) \geq 0 \qquad (5)$$

$$\forall(\theta, t) : \qquad \pi(\theta, t) = \sum_{a \in A} \phi(a) \cdot x(\theta, t|a) \qquad (6)$$

$$x(\cdot, \cdot|\alpha) \in \Delta(\Theta \times T), \quad \pi \in \Pi \qquad (7)$$

where $\sum_{\theta, t} \psi(\theta, t, \alpha) = \phi(\alpha)$ for all $\alpha \in A$, and $\psi(\theta, t, \alpha) = \phi(\alpha)x(\theta, t|\alpha)$.

Equivalently, the sharp set for the distribution $\pi \in \Delta(\Theta \times T)$:

$$\Pi_I(\phi) = \{\pi \in \Pi : \mathbf{LP}(\phi, \pi) \text{ is feasible }\}$$

The above result is generic, in that it handles general games with generic states of the world θ. In particular, it nests both standard private and common value auction models and provides a mapping between the distribution of bids ϕ and the set of feasible distribtions over signals and θ. An iid assumption allows us to learn ϕ. This iid (or exchangeability) assumption is a maintained assumption that we require throughout and is what is required to get information on ϕ. Given ϕ, the above result tells us how to map ϕ to the set of BNEs that are consistent with the data and are robust to any information structure that is an expansion of a minimal information structure S. Suppose we assume that both t and θ take finitely many values (an assumption we maintain throughout), then a joint distribution on (t, θ), $\pi(t, \theta)$ is consistent with the model and the data if and only if it solves the above *linear program.* So, feasibility can be used as a basis for testability of the model. In addition, it is possible to

use parametric distributions for the π or specialize the LP to specific setups such as common or private values. In addition, we can also allow for observed heterogeneity by using covariate information (see an example such adaptation in the common value Section 3).

Note that generally, the player utilities may depend on a parameter β which can be an object of interest. For instance, in auctions with risk averse bidders, the utility function depends on a risk aversion parameter that one may want to estimate. In these cases, the LP above can still be used to learn this parameter. For instance, the set $\Pi_I(\phi; \beta)$ is indexed now by β. So, $\beta = b$ belongs to the identified for β if and only if $\Pi(\phi, b)$ is nonempty. The identified set for β then is the set of all b's for which $\Pi(\phi; b)$ is nonempty.

2.2 Identified Sets of Moments of $\Pi(.)$

In the case of non-parametric inference where we put no constraint on the distribution of fundamentals, i.e. $\Pi = \Delta(\Theta \times T)$, note that the sharp set is linear in the density function of the fundamentals $\pi(\cdot, \cdot)$. Therefore, maximizing or minimizing any linear function of this density can be performed via solving a single linear program. This implies that we can evaluate the upper and lower bounds of the expected value of any function $f(\theta, t)$ of these fundamentals, in expectation over the true underlying distribution. The latter holds, since the expectation of any function with respect to the underlying distribution is a linear function of the density. We state this as a corollary next.

Corollary 4. *Let Result (3) hold. Also, let $f(.,.)$ be any function of the state of the world and the profile of minimal signals (eg, $f(\theta, t) = \theta$ or $f(\theta, t) = \theta^2$, ...). The sharp identified set for the expected value of $f(.,.)$ w.r.t. the true distribution of*

fundamentals, i.e. $E_{(\theta,t)\sim\pi}[f(\theta,t)]$ is an interval $[L,U]$ such that:

$$L = \min_{\pi\in\Pi_I(\phi)} \sum_{\theta\in\Theta,t\in T} f(\theta,t)\cdot\pi(\theta,t) \qquad (8)$$

$$U = \max_{\pi\in\Pi_I(\phi)} \sum_{\theta\in\Theta,t\in T} f(\theta,t)\cdot\pi(\theta,t) \qquad (9)$$

When the space Π can be represented by a set of linear constraints, then these are two linear programming problems.

2.3 Robust Counterfactual Analysis

One key objective of our inference procedures is to understand the performance of some other auction when deployed in the same market, with respect to some objective or metric: $F : \Theta \times T \times A \to \mathbb{R}$ that is a function of the unknown fundamentals and the action vector. Examples of such metrics in single-item auctions could be social welfare: $F(\theta,t,a) = \sum_{i=1}^{n} \theta_i x_i(a)$ or revenue, i.e. $F(\theta,t,a) = \sum_{i=1}^{n} p_i(a)$, where θ_i is the value of player i for the item at sale, $x_i(a)$ is the probability of allocating to player i under action profile a and $p_i(a)$ is the expected payment of player i under action profile a. Counterfactual evaluations in our setup can be done as long as our LP structure can be maintained.

In particular, we are interested in computing an upper and lower bound on this analyst specified metric F under this new auction which has different utilities $\tilde{u}_i(a;\theta)$ and under any Bayes-correlated equilibrium which would map into a BNE with an augmenting information structures. The welfare bounds computed in this manner will inherit the robustness property in that they will be valid under all information structures.

This is straightforward in our setup since computing a sharp identified set for any such counter-factual can be done in a computationally efficient manner, in both the common value setting and in the correlated private value setting. The upper bound of the counter-factual can be obtained using the following linear program, that takes

as input the observed distribution of bids in our current auction, the metric F, and the primitive utility form $\tilde{u} = (\tilde{u}_1, \ldots, \tilde{u}_n)$ under the alternative auction. We state this result in the next Theorem.

Theorem 5. *Given a metric function $F : \Theta \times B \to \mathbb{R}$, a distribution over action profiles $\phi(.)$ and vector of alternative auction utilities $\tilde{u}$, the sharp upper and lower bounds on the expected metric under the new auction can be computed using the following LP:*

$$LP(\phi, F, \tilde{u}) \qquad \min_{\tilde{\psi}} / \max \sum_{\theta, t, a} F(\theta, t, a) \cdot \tilde{\psi}(\theta, t, a)$$

$$\forall t_i, a_i, a_i' : \sum_{\theta, t_{-i}, a_{-i}} \tilde{\psi}(\theta, t, a) \cdot (\tilde{u}_i(a; \theta) - \tilde{u}_i(a_i', a_{-i}; \theta)) \geq 0$$

$$\forall (\theta, t) \in \Theta \times T : \ \pi(\theta, t) = \sum_{a \in A} \tilde{\psi}(\theta, t, a)$$

$$\tilde{\psi} \in \Delta(\Theta \times T \times A) \ \text{and} \ \pi \in \Pi_I(\phi),$$

where $\Pi_I(\phi) = \{\pi \in \Pi : \mathbf{LP}(\phi, \pi) \text{ is feasible }\}$. In the non-parametric case, where $\Pi = \Delta(\Theta \times T)$, the latter is a linear program.

Note that getting sharp bounds on welfare measures for example using the above Theorem does not require one to infer in a prior step the distribution over the primitives. Rather, the above procedure provides sharp bounds on this welfare measure by appending to the above program the fact that $\pi \in \Pi_I(\phi)$, which, in the non-parametric case, is also a set of linear inequalities. Thus, in the non-parametric case, inference on counterfactuals boils down to solving a single linear program.

The linear inequalities in $\Pi_I(\phi)$, in the non-parametric case, could be augmented with further constraints on the density π, as long as they can be expressed as linear constraints. For instance, if π is a density in one dimension, then we can easily encode shape constraints, such as Lipschitz constraints (e.g. $|\pi(x) - \pi(x+1)| \leq L$), monotonicity constraints (e.g. $\pi(x) \leq \pi(x+1)$), and concavity/convexity constraints (e.g. $\pi(x) - \pi(x-1) \leq \pi(x+1) - \pi(x)$). The linear inequalities can also be augmented with minimal informativeness constraints, as outlined in Section 2.4 below.

Coupling Information Structures in Counterfactuals. Note that the above counterfactuals do not couple the signaling structure in the original game and the counterfactual game and allows for the information structure to change[4]. As was proven in the follow up paper of Bergemann, Brooks, and Morris [2019], one can also easily incorporate the extra constraint that the information structure between the two games does not change. This holds the information structure constant in the counterfactual. One would simply need to add the constraints that:

$$\tilde{\psi}(\theta, t, a') = \sum_{a \in A} m(\theta, t, a, a') \tag{10}$$

$$x(\theta, t|a)\phi(a) = \sum_{a' \in A} m(\theta, t, a, a') \tag{11}$$

$$\sum_{\theta, t, a, a'} m(\theta, t, a, a') = 1 \tag{12}$$

These constraints essentially enforce that the BCE in the first game and the BCE in the second game must be marginals of a joint distribution in a meta-game where the player decides on the actions in both games simultaneously.

2.4 Adding Informativeness Constraints on Player Signals

Our main setup above can incorporate constraints on a players' signals to be at least as informative as some baseline, using the auxiliary signal t_i, in our general formulation. Imposing such constraints can significantly shrink the size of the identified set. We explore two types of constraints: the first constraint bounds the *maximum deviation* that a signal can have from the true state; the second posits a parametric signal distribution. We examine also the empirical content of these restrictions in Monte Carlo Experiments and with real data.

---

[4]This means that in the counterfactual game where we are computing the bounds on these metrics, the information structure is not preserved. For instance, if the players observed a signal that is informative in the status quo game, then this signal is not maintained in the counterfactual.

Informativeness via support constraints. In particular, we can encode informativeness of a player's signal, by simply saying that $|t_i - \theta| < \epsilon$ (i.e. that player i receives a minimal signal t_i that is within some ϵ of the true state θ^5). More generally, we can assume that θ, t have a joint support such that $t \in S(\theta)$ for some set $S(\theta)$. Then we obtain the following modified LP, where we define variables $x(\theta, t \mid a)$ for any $\theta \in \Theta$ and $t \in S(\theta)$ and constraints:

$$LP(\phi, \pi) \quad \forall t_i, a_i, a_i' : \qquad \sum_{\theta, t_{-i}, a_{-i}: t_i \in S(\theta)} \phi(a) \cdot x(\theta, t|a) \cdot \partial u_i(a_i', a; \theta) \geq 0 \quad (13)$$

$$\forall \theta, \forall t \in S(\theta) : \quad \pi(\theta, t) = \sum_{a \in A} \phi(a) \cdot x(\theta, t|a) \quad (14)$$

$$x(\cdot, \cdot|\alpha) \in \Delta(\Theta \times T), \quad \pi \in \Pi \quad (15)$$

See also this type of constraint explicitly used in the context of a common value auction in Section 6.

Informativeness via known correlation structure. In another approach we can also assume that we know the correlation structure of the joint distribution of θ, t, i.e. we know the conditional density $q(t \mid \theta)$. In that case, we know that $\pi(\theta, t) = q(t \mid \theta)\pi(\theta)$. Then we can simply modify the linear constraints:

$$LP(\phi, \pi) \quad \forall t_i, a_i, a_i' : \qquad \sum_{\theta, t_{-i}, a_{-i}: t_i \in S(\theta)} \phi(a) \cdot x(\theta, t|a) \cdot \partial u_i(a_i', a; \theta) \geq 0 \quad (16)$$

$$\forall (\theta, t) : \quad q(t \mid \theta) \, \pi(\theta) = \sum_{a \in A} \phi(a) \cdot x(\theta, t|a) \quad (17)$$

$$x(\cdot, \cdot|\alpha) \in \Delta(\Theta \times T), \quad \pi \in \Pi \quad (18)$$

This incorporates the restriction that we know $q(t \mid \theta)$ while the above remains a linear program. For instance, we can incorporate informativeness of the signals, by assuming that players receive a minimal signal t_i that is normally distributed with mean θ and a known standard deviation σ, which will govern the amount of informativeness, i.e.,

---

[5]Note that this just says that the signal is within ϵ of θ and does not make any assumptions on joint distribution of θ and and the signal t.

$q(t|\theta) \sim \mathcal{N}(\theta, \sigma)$. We can examine how the identified sets shrink as a function of this informativeness parameter.

3   Common Value Auctions

We specialize the above results to important classes of auctions. We begin with the common value model in which the game of incomplete information G is a single item common value auction. In this case the unknown state of the world is the unknown value of the object v, which we assume to take values in some finite set V. Moreover, we initially assume that the minimal information structure is degenerate, i.e., players receive no minimal signal about this unknown common value[6]. The signal set T becomes a singleton and is irrelevant. Thus we will denote with $\pi \in \Delta(V)$ the distribution of the unknown common value, which is the parameter that we wish to identify. This is a particularly simple model to illustrate the structure of the LP approach and showcase the flexibility of our methods.

The sequence of actions in these auctions is as follows. Prior to bidding in the auction, the players receive some signal which is drawn from some distribution; this signal can be correlated with the unknown common value and with the signals of their opponents. We wish to be ignorant about which information structure realized in each auction sample and want to identify the sharp identified set for π. Moreover, we will assume that the players' bids take values in some discrete set B and players play a BNE with this unknown and unobserved general information structure. With the minimal information structure being the noninformative one, the Theorem below characterizes the identified set for π in this model.

Theorem 6. *Let the common value model above hold with n bidders. Given a dis-*

---

[6]Other constraints on the initial signals are allowed and here we take the degenerate signal for simplicity.

tribution of bid profiles ϕ supported on a set $S \subseteq B^n$, the set of distributions of bids $\pi \in \Delta(V)$ that are consistent with ϕ are ones where the following program is feasible:

$$LP(\phi, \pi) \qquad \forall b_i^*, b_i' \in B : \qquad \sum_{v \in V, \ \mathbf{b} \in S: b_i = b_i^*} \phi(\mathbf{b}) \cdot x(v|\mathbf{b}) \cdot (u_i(\mathbf{b}; v) - u_i(b_i', \mathbf{b}_{-i}; v)) \geq 0$$

$$\forall v \in V : \qquad \pi(v) = \sum_{\mathbf{b} \in S} \phi(\mathbf{b}) \cdot x(v|\mathbf{b})$$

$$\forall \mathbf{b} \in S : \qquad x(\cdot|\mathbf{b}) \in \Delta(V)$$

Observe, that in this setting, the latter linear program is also linear in π (and $x(.)$). Thus we get that the sharp identified set is a convex set and is defined as the set of solutions to the above linear program, where π is also a variable. Of course, feasibility of the linear program can also be the basis for testability of the model assumptions. If these constraints are not feasible, then the model is falsified. The above LP does not restrict the density of valuation. If one is willing to assume that this density is known up to a finite dimensional parameter β, then one can add that as a constraint where the LP becomes now $LP(\phi, \beta)$ appending the following to the above:

$$\forall v \in V : \pi(v, \beta) = \sum_{v \in V} \phi(\mathbf{b}) \cdot x(v|\mathbf{b})$$

The identified set for β is defined similarly in that it is the set of β's where $LP(\phi, \beta$ is feasible.

It is also possible to easily infer upper and lower bounds on any linear function of the unknown distribution π such as mean or other moments. This is stated next as a Corollary to the above Theorem.

Corollary 7. *Let the Assumptions in Theorem (6) hold. Also, let $f(.)$ be any function of valuations (such as $f(v) = v$ or $f(v) = v^2$, ...). Then, using the LP above, we can*

get upper and lower bounds on such moments as follows:

$$\max_{\pi \in \Pi_I(\phi)} \sum_{v \in V} f(v) \cdot \pi(v) \tag{19}$$

$$\min_{\pi \in \Pi_I(\phi)} \sum_{v \in V} f(v) \cdot \pi(v) \tag{20}$$

where

$$\Pi_I(\phi) = \{\pi \in \Pi : LP(\phi, \pi) \text{ is feasible }\}$$

Observe that the latter linear expressions are simply: $\mathbb{E}_{v \sim \pi}[f(v)]$ and so the above shows that we can compute in polynomial time upper and lower bounds of any moment of the unknown distribution of the common value.

Also, note that the Corollary above shows that to do set inference with respect to any moment of the unknown distribution, we do not need to discretize the space of probability distributions and enumerate over all probability vectors, checking whether they are inside the sharp identified set. Rather we can just solve the above LP.

Essentially this observation says that we can easily compute the support function $h(z; \Pi_I(\phi))$ of the identified set $\Pi_I(\phi)$ at any direction z, by simply solving a linear program.

Remark 1 (Winning bid). *The above procedure can be easily modified if indeed as it may be the case, only winning bids are observed. In particular, given that I observe the CDF $F(.)$ of the winning bid, I know that an equilibrium is consistent with said CDF if and only if for every possible bid in B,*

$$F(x) = \sum_{v \in V} \sum_{\mathbf{b}/b_i \leq x \ \forall i} \psi(\mathbf{v}, \mathbf{b})$$

This has up to B linear constraints of B^{n+1} variables, but if we assume the bid vector distribution has small support of size K, we only need $K \cdot B$ variables.

Remark 2 (**Covariate Heterogeneity**)**.** *Suppose we have covariates and want to allow for observed heterogeneity where we maintain the assumption that the vector of covariates x takes finitely many values. Then, one nonparametric approach is to repeat and solve the above LP for $\pi(v|x = x_0)$ for every value x_0 that x takes. In addition, in cases where we have $E[v|x] = x'\beta$, then we can solve directly for the identified set for β by solving the following LP, which we denote with $LP(\phi, \beta)$:*

$$\forall b_i^*, b_i' \in B, x_0 \in X : \quad \sum_{v \in V, \ \mathbf{b} \in S : b_i = b_i^*} \phi(\mathbf{b}|x_0) \cdot x(v|\mathbf{b}, x_0) \cdot (u_i(\mathbf{b}; v) - u_i(b_i', \mathbf{b}_{-i}; v)) \geq 0$$

$$\forall x_0 \in X : \qquad\qquad x_0'\beta = \sum_{\mathbf{b} \in S} \sum_v v\phi(\mathbf{b}|x_0) \cdot x(v|\mathbf{b}, x_0)$$

$$\forall \mathbf{b} \in S, x_0 \in X : \qquad\qquad x(\cdot|\mathbf{b}, x_0) \in \Delta(V)$$

Here, the $n \times k$ vector $x_0'\beta$ allows for different players to have different β's (and the x's in this case would be auction specific heterogeneity where the different β's would allow the mean valuation of different players to depend differently on auction characteristics).

The latter is attractive as it allows us to couple the identified set of the mean of the unknown common value across multiple covariate realizations, in a computationally tractable manner. Otherwise we would have to generate the identified sets of the mean for each covariate realization and then solve a second stage problem which would try to find the set of joint solutions in each of these identified sets (via some form of joint grid search) that are consistent with a model of how the conditional mean $E[v|x]$ varies as a function of x. However, joint grid search would grow exponentially with the number of realizations of the covariates. The latter remark, shows that when the model of $E[v|x]$ is linear, then we can save this exponential blow-up in the computation whilst leveraging the statistical power of coupling data from separate covariate realizations.

Remark 3 (Informativeness via support constraints on minimal signals)**.** *We specialize informativeness constraints on the player's signals to the common value case and assume that prior to the auction all players receive some common signal t, which is*

within some ϵ of the common value, i.e. $|t - v| \leq \epsilon$. In this case we can augment the common value LP with this minimal signal that player's receive and enforce the support constraint:

$$\forall t, b_i^*, b_i' : \sum_{\substack{v \in V: \ |v-t| \leq \epsilon \\ \mathbf{b} \in S: \ b_i = b_i^*}} \phi(\mathbf{b}) \cdot x(v, t | \mathbf{b}) \cdot (u_i(\mathbf{b}; v) - u_i(b_i', \mathbf{b}_{-i}; v)) \geq 0$$

$$\forall v, t : |v - t| \leq \epsilon : \ \pi(v, t) = \sum_{\mathbf{b} \in S} \phi(\mathbf{b}) \cdot x(v, t | \mathbf{b})$$

$$x(\cdot, \cdot | \mathbf{b}) \in \Delta(V \times T), \quad \pi \in \Pi$$

(21)

4   Private Value Auctions

We now consider the case of a private value single item auction. In this case the (unknown) state of the world is the a vector of private values $\mathbf{v} = (v_1, \ldots, v_n) \in V^n$. We assume that these private values come from some unknown joint distribution $\pi \in \Delta(V^n)$. Moreover, we initially assume that players know at least their own private value. Thus the (minimal) signal set T_i is equal to V and moreover, we have that conditional on a value vector v, $t_i = v_i$, deterministically. Since the signal is a deterministic function of the unknown state of the world, we will again denote with $\pi \in \Delta(V^n)$ the distribution of the unknown valuation vector, which is the parameter that we wish to identify. Here, each player first draws a valuation (as an element of the state of the world), and then each player's own valuation is revealed to the player through a signal. After that, a signal (unobserved to the econometrician) is further revealed before players play a BNE given this signal.

In this setting the sharp identified set is again slightly simplified. The result is stated in the next Theorem.

Theorem 8. *Let the above private values auction model hold. Given a distribution of bid profiles ϕ supported on a set $S \subseteq B^n$, the set of distributions of bids $\pi \in \Delta(V^n)$ that are consistent with ϕ are the ones where the following linear program, denoted*

$LP(\phi, b)$, *is feasible:*

$$\forall v_i^* \in V, b_i^*, b_i' \in B : \qquad \sum_{\mathbf{v}:v_i=v_i^*, \mathbf{b}\in S:b_i=b_i^*} \phi(\mathbf{b}) \cdot x(\mathbf{v}|\mathbf{b}) \cdot (u_i(\mathbf{b}; v_i) - u_i(b_i', \mathbf{b}_{-i}; v_i)) \geq 0$$

$$\forall \mathbf{v} \in V^n : \qquad \pi(\mathbf{v}) = \sum_{\mathbf{b}\in S} \phi(\mathbf{b}) \cdot x(\mathbf{v}|\mathbf{b})$$

$$\forall \mathbf{b} \in S : \qquad x(\cdot|\mathbf{b}) \in \Delta(V^n)$$

Observe, that in this setting, the latter linear program is also linear in π. Thus we get that the sharp identified set is a convex set and is defined as the set of solutions to the above linear program, where π is also a variable. Note also here that no assumption is made on the correlation between player valuation. This result allows for the recovery of valuation distribution with arbitrary correlation (and general signaling structures). As a special case of the above, we study next the IPV model of auctions.

Independent Private Values. The situation becomes more complex if we also want to impose an extra assumption that the distribution of private values is independent. In that case, we have the extra condition that π must be a product distribution which is a non-convex constraint. For instance, if we want to assume that the value of each player is independently drawn from the same distribution ρ, and hence ρ is what we wish to identify, then we also have the extra constraint that:

$$\forall \mathbf{v} \in V^n : \pi(\mathbf{v}) = \rho(v_1) \cdot \ldots \cdot \rho(v_n) \tag{22}$$

Adding this constraint into the above LP, makes the LP non-convex with respect to the variables $\rho(v)$ (even though checking whether a given ρ is in the identified set, is still an LP). Thus in this case we cannot compute in polynomial time upper and lower bounds on the moments of the distribution ρ using the above LP.

However, we make the following observation which simplifies the constraints of the LP: we note that conditional on a player's valuation and on a bid profile b, the effect of a deviation $u_i(\mathbf{b}; v_i) - u_i(b_i', \mathbf{b}_{-i}; v_i)$ is independent of the values of opponents, in a

private value setting. Thus we can re-write the best response constraint as:

$$\sum_{\mathbf{b}\in S:b_i=b_i^*} \phi(\mathbf{b}) \cdot x_i(v_i^*|\mathbf{b}) \cdot (u_i(\mathbf{b};v_i^*) - u_i(b_i', \mathbf{b}_{-i}; v_i^*)) \geq 0 \tag{23}$$

where $x_i(v_i|b) = \Pr[V_i = v_i|B = b]$, where V_i is the random variable representing player i's value and B is the random variable representing the bid profile at a BCE.

Then we can formulate the consistency constraints, by simply imposing a constraint per player, i.e. if ρ_i is the distribution of player i's value, then it must be that:

$$\rho_i(v_i) = \sum_{\mathbf{b}\in S} \phi(\mathbf{b}) \cdot x_i(v_i|\mathbf{b}) \tag{24}$$

These are constraints that are still linear in $\rho_i(v_i)$. We state the LP as a corollary next.

Corollary 9. *Assume that the above IPV model hold. The following LP, denoted $LP(\phi,\rho)$, characterizes the sharp identification under the independent private values model:*

$$\forall i \in [n], v_i \in V, b_i^*, b_i' \in B: \quad \sum_{\mathbf{b}\in S:b_i=b_i^*} \phi(\mathbf{b}) \cdot x_i(v_i|\mathbf{b}) \cdot (u_i(\mathbf{b};v_i) - u_i(b_i', \mathbf{b}_{-i}; v_i)) \geq 0$$

$$\forall i \in [n], v_i \in V: \quad \rho_i(v_i) = \sum_{\mathbf{b}\in S} \phi(\mathbf{b}) \cdot x_i(v_i|\mathbf{b})$$

$$\forall \mathbf{b} \in S: \quad x_i(\cdot|\mathbf{b}) \in \Delta(V)$$

In particular if we assume that player's are symmetric, i.e. $\rho_i(v) = \rho(v)$, then we can compute upper and lower bounds on any moment $E[f(v)]$ of the common value distribution:

$$\max_{x_i(\cdot|\mathbf{b})} \sum_{v\in V} f(v) \cdot \rho(v)$$

$$\forall i \in [n], v_i \in V, b_i^*, b_i' \in B: \quad \sum_{\mathbf{b}\in S:b_i=b_i^*} \phi(\mathbf{b}) \cdot x_i(v_i|\mathbf{b}) \cdot (u_i(\mathbf{b};v_i) - u_i(b_i', \mathbf{b}_{-i}; v_i)) \geq 0$$

$$\forall i \in [n], v \in V: \quad \rho(v) = \sum_{\mathbf{b}\in S} \phi(\mathbf{b}) \cdot x_i(v|\mathbf{b})$$

$$\forall i \in [n], \mathbf{b} \in S: \quad x_i(\cdot|\mathbf{b}) \in \Delta(V)$$

24

A by-product of this analysis is that the linear program allows us to test for symmetry in the independent private values model. In particular it is not clear that when we assume that all marginals are the same, then the LP is feasible. Thus by checking feasibility of the LP we can refute the assumption of symmetric independent private values. It is important to note here that given that we allow the augmenting signals to be arbitrary correlated, we are not able to infer any information about the *joint* distribution of valuation, given a bid profile.

5 Inference on Sharp Sets in CV Auctions

In general, we do no have access to the distribution $\phi(\mathbf{b})$. Instead, we have access to N i.i.d. observations $\omega_1, ..., \omega_N$ from said distribution ; let $\omega_t(\mathbf{b}) = 1\{\omega_t = \mathbf{b}\}$. The sampled distribution is then given by

$$\phi_N(\mathbf{b}) = \frac{1}{N} \sum_{t=1}^{N} \omega_t(\mathbf{b}) \tag{25}$$

In this section, we develop techniques for estimating the sharp identified set using samples from $\phi(\mathbf{b})$. We showcase our techniques for the CV setup for simplicity. Also, we focus on inference on the identified set for π which is the object of inference here. Often times, we parametrize the distribution of the common values, and so inference will be on the identified set for the vector of parameters[7]. We start with finite sample approaches to constructing confidence regions for sets using concentration inequalities. These seem to be the first application of such results on using such inequalities to settings with partial identification[8]. We also show how existing set inference methods can also be used to construct confidence regions.

---

[7]It is possible to construct the CI for the unknown parameters -rather than the identified set by inverting test statistics.

[8]Finite sample inference results are particularly attractive in models with partial identification since standard asymptotic approximations are not typically uniformly valid especially in models that are close to/or are point identified.

5.1 Finite Sample Inference via Concentration Inequalities

We explore first the question of set inference using finite sample concentration inequalities. This allows us to obtain a confidence set for the identified set where the coverage property holds for every sample size. In addition, we highlight tools from the concentration of measure literature applied to inference on sets in partially identified models.

As a reminder, given a probability distribution over bids $\phi(\mathbf{b})$, the sharp set of compatible equilibria $\phi(v, \mathbf{b}) = \phi(\mathbf{b})x(v|\mathbf{b})$ is the set Θ_I of joint probability distributions satisfying:

$$\forall b_i^*, b_i' \in B : \qquad \sum_{v \in V, \ \mathbf{b} \in S: b_i = b_i^*} \phi(\mathbf{b}) \cdot x(v|\mathbf{b}) \cdot (u_i(\mathbf{b}; v) - u_i(b_i', \mathbf{b}_{-i}; v)) \geq 0 \qquad (26)$$

for the proper $x(.|.)$. See the statement of Theorem (6) above. Let $\pi(.)$ be the vector characterizing the distribution of the common value. Let $\{F_j(x; \pi, \phi) : j \in M\}$, denote the negative of the best-response and density constraints, associated with the BCE LP in Theorem (6). Then in the population π is feasible iff:

$$\min_x \max_{j \in M} F_j(x; \pi, \phi) \leq 0 \qquad (27)$$

Then we have that the identified set is defined as:

$$\Pi_I = \{\pi : \min_x \max_{j \in M} F_j(x; \pi, \phi) \leq 0\} \qquad (28)$$

Now we consider a finite sample analogue. Observe that $F_j(x; \pi, \phi) = \mathbb{E}[f_j(x; \pi, \omega)]$, where expectation is over the random vector ω and the function $f_j(\cdot; \pi, \omega)$ takes the form:

$$f_j(x; \pi, \omega) = \sum_{v \in V, \ \mathbf{b} \in S: b_i = b_i^*} \omega(\mathbf{b}) \cdot x(v|\mathbf{b}) \cdot (u_i(b_i', \mathbf{b}_{-i}; v) - u_i(\mathbf{b}; v)) \qquad (29)$$

for some triplet (i, b_i^*, b_i') for the case of best-response constraints and similarly for density consistency constraints. Then we consider the sample analogue:

$$F_j^N(x; \pi) = \frac{1}{N} \sum_{t=1}^{N} f_j(x; \pi, \omega_t) \qquad (30)$$

26

Observe that due to the linearity of F_j with respect to ϕ, we can re-write: $F_j^N(x; \pi) = F_j(x; \pi, \phi_N)$. One can then define the estimated identified set by analogy, i.e., replacing $\phi(b)$ with $\phi_N(b)$:

$$\Pi_I^N = \{\pi : \min_x \max_{j \in M} F_j(x; \pi, \phi_N) \leq \sigma^N\} \tag{31}$$

for some decaying tolerance constant σ^N (which can be set to zero).

Because the pdf of a non-parametric distribution is a very high-dimensional object, inference on it will require many samples. Hence, we will instead focus on two more structured inference problems. In the first one we are interested in inferring the identified set for *a moment* of the distribution and in the second we make assume that the said pdf is known up to a finite dimensional parameter and infer the identified set of these lower dimensional parameters. One can in principle recover non-parametric inference by simply making the parameters be the values of the pdf at the discrete support points, albeit at a cost in the sample complexity.

5.1.1 Inference on Identified Set for Moments

We begin with the non-parametric setting and show how to construct inference on the identified set of any moment function $m : V \to [-H, H]$ of the common value distribution that is valid in finite samples. Observe that the identified set for any moment $m : V \to [-H, H]$ is an interval $[L, U]$ defined by:

$$
\begin{aligned}
L &= \min_{x: \ \max_{j \in M} F_j(x;\phi) \leq 0} \sum_{v \in V} m(v) \cdot \sum_{b \in S} \phi(\mathbf{b}) \cdot x(v|\mathbf{b}) \\
U &= \max_{x: \ \max_{j \in M} F_j(x;\phi) \leq 0} \sum_{v \in V} m(v) \cdot \sum_{b \in S} \phi(\mathbf{b}) \cdot x(v|\mathbf{b})
\end{aligned}
\tag{32}
$$

where x in both optimization problems is ranging over the convex set of conditional distributions, i.e. $x(\cdot|\mathbf{b}) \in \Delta(V)$, which we omit for simplicity of notation. We can then define their finite sample analogues as:

$$
\begin{aligned}
L^N(\sigma^N) &= \min_{x: \ \max_{j \in M} F_j^N(x) \leq \sigma^N} \sum_{v \in V} m(v) \cdot \sum_{b \in S} \phi_N(\mathbf{b}) \cdot x(v|\mathbf{b}) \\
U^N(\sigma^N) &= \max_{x: \ \max_{j \in M} F_j^N(x) \leq \sigma^N} \sum_{v \in V} m(v) \cdot \sum_{b \in S} \phi_N(\mathbf{b}) \cdot x(v|\mathbf{b})
\end{aligned}
\tag{33}
$$

Where $F_j^N(x)$ and ϕ_N are given in Equations (30) and (25) respectively.

The following result gives finite sample high probability bounds on the coverage of the interval $[L^N(\sigma^N) - \epsilon^N, U^N(\sigma^N) + \epsilon^N]$. As a reminder, we use n to designate the number of bidders, N is sample size, and $|B|$ is the number of support points for bids.

Theorem 10. *Suppose that $u_i(\mathbf{b}; v) \in [-H, H]$ for all $i \in [n]$, $\mathbf{b} \in B^n$ and $v \in V$ and let $\sigma^N = 2H\sqrt{\frac{\log(4n|B|^2/\delta)}{N}}$ and $\varepsilon^N = 2H\sqrt{\frac{\log(4/\delta)}{N}}$. Then:*

$$\Pr\left[[L, U] \subseteq \left[L^N(\sigma^N) - \epsilon^N, U^N(\sigma^N) + \epsilon^N\right]\right] \geq 1 - \delta \tag{34}$$

where L, U are defined by Equation (32) and $L^N(\sigma^N)$, $U^N(\sigma^N)$ are defined by Equation (33).

5.1.2 Inference on Identified Set for Parametric Distributions.

For the parametric case, we assume that the distribution $\pi(v)$ is parametric of the form $\pi(v, \theta)$ for some finite parameter set Θ. In the parametric setting we need to augment the constraint set M, apart from containing the best-response constraints, to contain the parametric form consistency constraints:

$$\forall v \in V : \pi(v, \theta) = \sum_{v \in V} \phi(\mathbf{b}) \cdot x(v|\mathbf{b}) \tag{35}$$

These can also be written in the form $F_j(x; \phi) \leq 0$ for some function $F_j(x; \phi) = \mathbb{E}[f_j(x; \omega)]$.[9] We overload notation and let M be this augmented set of constraints. Then the parameter of interest is θ and the identified set for θ takes the form:

$$\Theta_I = \{\theta \in \Theta : \min_x \max_{j \in M} F_j(x; \theta, \phi) \leq 0\} \tag{36}$$

and its sample equivalent

$$\Theta_I^N\left(\sigma^N\right) = \{\theta \in \Theta : \min_x \max_{j \in M} F_j(x; \theta, \phi_N) \leq \sigma^N\} \tag{37}$$

for some decaying tolerance constant σ^N (which can be set to zero).

The next Theorem provides finite sample high probability coverage bounds for Θ_I.

---

[9]Simply add one constraint of the form $\pi(v, \theta) - \sum_{v \in V} \phi(\mathbf{b}) \cdot x(v|\mathbf{b}) \leq 0$ and one of the form $\sum_{v \in V} \phi(\mathbf{b}) \cdot x(v|\mathbf{b}) - \pi(v, \theta) \leq 0$.

Theorem 11. *Consider Θ_I and $\Theta_I^N\left(\sigma^N\right)$ as defined in Equations (36) and (45). Suppose that $u_i(\mathbf{b};v) \in [-H, H]$ for all $i \in [n]$, $\mathbf{b} \in B^n$ and $v \in V$. If $\sigma^N = 2H\sqrt{\frac{\log(|\Theta|\cdot(n|B|^2+|V|)/\delta)}{N}}$, then:*

$$\Pr\left[\Theta_I \subseteq \Theta_I^N\left(\sigma^N\right)\right] \geq 1 - \delta \tag{38}$$

Sample Variance Based Bounds. We now show how to improve upon the prior analysis and getting tolerance variables σ^N whose leading $1/\sqrt{N}$ term does not depend on H, but rather on the sample variance of the constraints. We will modify the finite sample identified set $\Theta_I^N(\sigma^N)$ defined in Equation (45), as follows: instead of using a uniform tolerance σ^N for all the constraints, we will define tolerance for each constraint differently based on its sample variance:

$$\hat{\Theta}_I^N(\lambda, \sigma^N) = \left\{\theta : \min_x \max_{j \in M}\left(F_j^N(x;\theta) - \lambda\sqrt{\frac{\mathrm{Var}_N(f_j(x;\theta,\omega))}{N}}\right) \leq \sigma^N\right\} \tag{39}$$

Where $\mathrm{Var}_N(X)$ for a random variable X, denotes the sample variance:

$$\mathrm{Var}_N(X) = \frac{1}{N(N-1)} \sum_{1 \leq t < t' \leq n} (X_t - X_t')^2.$$

The extra variance modification, which is reminiscent of *sample variance penalization* in empirical risk minimization [Maurer and Pontil, 2009] and *optimism* in bandit algorithms [Audibert, Munos, and Szepesvári, 2009] , will allow us to set σ^N to be of order $O(H/N)$ rather than $O(H/\sqrt{N})$.

Theorem 12. *Consider Θ_I and $\hat{\Theta}_I^N\left(\lambda, \sigma^N\right)$ as defined in Equations (44) and (39). Suppose that $u_i(\mathbf{b};v) \in [-H, H]$ for all $i \in [n]$, $\mathbf{b} \in B^n$ and $v \in V$. If $\lambda = \sqrt{2\log(2|\Theta||M|/\delta)}$ and $\sigma^N = \frac{14H\log(2|\Theta||M|/\delta)}{3(N-1)}$ (with $|M| = n|B|^2 + |V|$), then:*

$$\Pr\left[\Theta_I \subseteq \hat{\Theta}_I^N\left(\lambda, \sigma^N\right)\right] \geq 1 - \delta \tag{40}$$

The latter theorem is asymptotically an improvement over Theorem 11, when the variances of the constraints are not as large as their worst case bound of $4H^2$, which is essentially what is assumed in Theorem 11. However, one drawback of this theorem

is that the new estimated set $\hat{\Theta}_I^N(\lambda, \sigma^N)$ requires solving more than linear programs. In particular for every parameter $\theta \in \Theta$ we need to solve an optimization problem of the form:

$$\min_x \max_j \left(\langle \alpha_j, x \rangle - \kappa_j \sqrt{\sum_{b,b' \in S} \gamma_{b,b'} \left(\sum_v (x_{vb}\beta_{jvb} - x_{vb'}\beta_{jvb'}) \right)^2} \right) \qquad (41)$$

The negative part is a concave function of x, as it can be thought of as a norm of a vector that is a linear function of x. Thus this is a non-convex minimization problem. Thus even though it offers a statistical improvement, it is at the expense of computational efficiency.

5.2 Alternative Set Inference Approaches

Here, we use existing set inference approaches and adapt them to the problem at hand. The first approach exploits the mapping between the reduced form parameter (the distribution of bids) and the identified set via the linear program, and the second uses subsampling approaches to set inference.

5.2.1 Inference via the Bayesian Bootstrap

In this section, we describe approaches to inference on the identified set via a Bayesian Bootstrap for both parametric and nonparametric models. We start with the nonparametric case. As a reminder, the identified set in this nonparametric case is

$$\Pi_I = \{\pi : \min_x \max_{j \in M} F_j(x; \pi, \phi) \leq 0\} \qquad (42)$$

This is a case of a separable problem where knowing $\phi^* = \{\phi^*(b_1), \ldots, \phi^*(b_B)\}$ we can solve for the identified set via the mapping:

$$\Pi(\phi) = \{\pi : \min_x \max_{i \in M} F_i(x; \pi, \phi^*) \leq \sigma^N\} \qquad (43)$$

for some decaying tolerance constant σ^N (which can be set to zero). Inference on parameter sets in separable models is analyzed in (Kline and Tamer [2016]) where

(posterior) probability statements related to the identified set[10] can be computed using the mapping between the reduced form parameters, $\phi(\mathbf{b})$ in this case, and the structural parameters π. Intuitively, for every draw s from the posterior for ϕ (this is a multinomial distribution and so obtaining draws from the multinomial is simple using the Bayesian bootstrap[11]) we can solve via (43) for a "copy" of the implied identified set Π_N^s by solving the LP above. Using this procedure, we can get a sequence $\{\Pi_N^s\}_{s=1}^S$ that we can use to answer probability statements about the identified set Π_I. The computational constraint in this nonparametric problem is that the parameter of interest π is a vector of probabilities with $|V|$ support points. So, the bigger $|V|$ is the larger the number of parameters the more difficult it is to solve for the identified set via (31) above. Hence, with the finite support condition on the bids standard (Bayesian) inference on $\phi(\mathbf{b})$ can be easily mapped into inference on the set Π_I. The exact same bootstrap procedure can be used to provide (posterior) probability statement on identified intervals $[L, U]$ that are defined through the mapping in (33) above. This is relevant if one is interested in linear functional of the distribution. Inference on such objects reduces the computational burden substantively.

In the parametric case, we assume that the distribution function of v belongs to a parametric class that is known up to a finite dimensional parameter θ and so the LP in Theorem 6 is modified by setting $\pi(v) = \pi(v, \theta)$. Then the parameter of interest is θ and the identified set for θ takes the form:

$$\Theta_I = \{\theta : \min_x \max_{j \in M} F_j(x; \theta, \phi) \leq 0\} \tag{44}$$

---

[10]Examples of such statements are: the probability that the identified set belongs to a particular set (e.g, when Θ_I is an interval, the probability that this interval lies in $[a, b]$) or the probability that a particular vector belongs to the identified set, etc.

[11]For example, under a limiting uninformative Dirichlet prior for $\phi(\mathbf{b})$, the posterior for this $\phi(\mathbf{b})$ approaches the Dirichlet posterior $Dir(n_1, \ldots, n_B)$ where $n_j = \sum_{i=1}^N 1[b_i = b_j]$. Therefore, using results that connect the Gamma and Dirichlet distributions, a draw from given posterior can be approximated by a weighted Bootstrap with Gamma weights. See also Chamberlain and Imbens [2003].

and so the separable mapping between ϕ^* and θ takes the form again of

$$\Theta_I\left(\phi^*; \sigma^N\right) = \{\theta : \min_x \max_{j \in M} F_j(x; \theta, \phi^*) \le \sigma^N\} \tag{45}$$

for some decaying tolerance constant σ^N (which can be set to zero).

Here again, we can use the approach in (Kline and Tamer [2016]) to answer probability statements on Θ_I. The computational step here is much easier than the nonparametric case as now solving for the identified set for every draw from the posterior for ϕ is a lower dimensional problem. The Bayesian bootstrap is used to draw vectors of bid probabilities from the multinomial distribution for bids.

In addition to this Bayesian bootstrap approach, we also provide methods based on subsampling next.

5.2.2 Parametric Case via Subsampling

Subsampling methods can be used to conduct inference on the identified set in a large sample framework. In particular, let θ be the parameter of interest. Again, our problem is isomorphic to one where the objective function $Q(\theta)$ is as follows:

$$Q(\theta) = \min_x \max_{j \in M} F_j(x; \theta) \tag{46}$$

with a the corresponding sample analogue

$$Q^N(\theta) = \min_x \max_{j \in M} F_j^N(x; \theta) \tag{47}$$

We first re-state the above random variable as a minimax problem over convex and compact sets. Let $p \in \Delta_M$ lie on the simplex on M constraints. Then $Q^N(\theta)$ is equivalently defined as:

$$Q^N(\theta) = \min_x \sup_p \sum_{j \in M} p_j F_j^N(x; \theta) \tag{48}$$

and

$$Q^*(\theta) = \min_x \sup_p \sum_{j \in M} p_j F_j(x; \theta) \tag{49}$$

32

where Q^* is zero for all feasible $\theta's$. In addition, following the approach in Chernozhukov, Hong, and Tamer [2007], we get a confidence region for the identified set $\Theta_I = \{\theta : Q^*(\theta) \leq 0\}$ by studying the asymptotic distribution[12] of

$$\mathcal{C}_N = \sup_{\theta \in \Theta_I} \min_{x} \sup_{p} \sum_{j \in M} p_j F_j^N(x; \theta)$$

Let $\tau_N(1 - \alpha)$ be the $(1 - \alpha)-$quantile of $\mathcal{C}$ where $\mathcal{C}$ is the nondegenerate limit of $\sqrt{N}\mathcal{C}_N$. Then, define the $C_N(1 - \alpha)$ as follows

$$C_N(1 - \alpha) = \{\theta : Q^N(\theta) \leq \tau_N^+(1 - \alpha)\}$$

where $\tau_N^+(1 - \alpha) = \max(\tau_N(1 - \alpha), 0)$. Notice here that the event $\Theta_I \subseteq C_N(1 - \alpha)$ is equivalent to the event $\mathcal{C}_N \leq \tau_N^+(1 - \alpha)$. The next Theorem states the result.

Theorem 13. *Assume that N increases to infinity, and $\sqrt{N}(\mathcal{C}_N)$ converges in distribution to $\mathcal{C}$, a nondegenerate random variable. Then the set $C_N(1 - \alpha)$ defined above has the following coverage property:*

$$\lim_{N \to \infty} \Pr[\Theta_I \subseteq C_N(1 - \alpha)] = 1 - \alpha.$$

The asympotic distribution $\mathcal{C}$ above can be characterized using for example results from Shapiro [2009] and sufficient conditions for nondegeneracy are given there which requires second moments to hold (these hold in our case trivially since we maintain the assumptions that bids take finitely many values).

Finally, to be able to feasibly implement the above approach, one needs to get a value for the cutoff $\tau_N(1 - \alpha)$. It is clear here that the standard nonparametric bootstrap may not work. Even if the asymptotic distribution is normal[13], we may not be able to estimate its variance because it depends on the number of binding

---

[12]Alternatively, we can define the objective function $\tilde{Q}^N(\theta) = [Q^N(\theta)]_+$ which is the positive part of $Q^N(\theta)$. The identified set now is the minimizer of $\tilde{Q}^N(\theta)$ and then similar approaches to Chernozhukov et al. [2007] can be used to construct a CI based on subsampling.

[13]The distribution of $\mathcal{C}$ is likely to be the supremum of a Gaussian process since the optimal value of the LP is generally not unique.

constraints. Given the nondegeneracy of the limit, one approach is to use subsampling to compute the cutoff. This can be accomplished by getting m subsamples of the data such that $m/N \to 0$ as $N \to \infty$. For every subsample, we compute $\mathcal{C}_N^m$ using a preliminary estimate $\widehat{\Theta}_I$ and using the sequence $\{\mathcal{C}_N^m\}_{m=1}^M$ to get its upper $(1 - \alpha)$ quantile. This will result in an estimate of $C_N(1 - \alpha)$. We can then use that as our new $\widehat{\Theta}_I$ and iterate one or two times. This approach was implemented in the Monte Carlo section below and in the empirical application and it provided adequate results.

6 Monte-Carlo Analysis: Common Value Auctions

In this section, we examine the techniques of Section 5 to obtain an estimated set on simulated data. We assume in all the simulations that the values and bids are taken from discrete sets, respectively V and B. In the whole section, we fix the following parameters: i) The maximum value/bid H is given by $H = 20$. ii) The set of possible bids is given by $B = \{0, \ldots, H\}$. The set of possible common values is $V = B = \{0, \ldots, H\}$. We generate equilibrium observations as follows: First, we generate a density $f(v)$ of valuations with support V. We consider the four following densities:

Normal density. $f_n(.)$ is the density of a normal random variable with mean parameters $\mu = 4$ and standard deviation parameter $\sigma = 1$, discretized and truncated to have support V, i.e.

$$f_n(v) = \frac{\exp(-(v - \mu)^2/\sigma^2)}{\sum\limits_{w \in V} \exp(-(w - \mu)^2/\sigma^2)} \tag{50}$$

Poisson density. $f_p(.)$ is the density of a Poisson distribution with parameters $\lambda = 4$, truncated inside V, i.e.:

$$f_p(v) = \frac{\lambda^v \exp(-\lambda)/v!}{\sum\limits_{w \in V} \lambda^w \exp(-\lambda)/w!} \tag{51}$$

Binomial density. $f_b(.)$ is the density of a binomial random variable with probability $p = 0.2$ and number of draws $n = H = 20$, i.e.:

$$f_b(v) = \binom{N}{v} p^v (1-p)^{H-v} \tag{52}$$

Geometric density. $f_g(.)$ is the density of a geometric random variable with probability $p = 0.2$ truncated to have support V, i.e.:

$$f_g(v) = \frac{p(1-p)^v}{\sum\limits_{w \in V} p(1-p)^w} \tag{53}$$

For each of those densities, we then generate one distribution of equilibrium bids ϕ, through solving the BCE linear program for the given distribution of values and with variables $\phi(\mathbf{b})$ rather than $\pi(v)$. We then generate N samples of bid vectors from ϕ, to generate the observed empirical bid distribution ϕ_N.

In the non-parametric setting, since we cannot directly formulate the estimated set of the variance (as it cannot be written as $E[m(v)]$ for some m), we obtain a superset for the identified set as follows: First obtain upper and lower bounds $E_{min}, E_{max}, E_{min}^{2nd}, E_{max}^{2nd}$ for the first and second moments respectively. Then set the bounds on the standard deviation by the conservative ones: $E_{min}^{2nd} - E_{max}^2 \leq Var \leq E_{max}^{2nd} - E_{min}^2$. A computationally more tedious procedure of estimating the identified set for the variance is to first get the identified set for the (discrete) distribution of V and then using that we can "solve" for a bound on the variance[14].

In the parametric case, bounds on the variance can be obtained directly from recovering bounds on the possible parameters for the distribution we consider, and tight bounds on the parameters imply tight bounds on the second and higher order moments. All simulations in the parametric case are therefore presented in terms of identified and estimated sets on the parameters of the distribution of values.

---

[14]For example, for every "draw" from the identified set for the distribution of V, we can obtain a variance. We can repeat the process to build the identified set for the variances.

6.1 Parametric vs. Non-Parametric Identified Sets

In this section, we compare the identified sets when we know (parametric) the functional form of the distribution of v (up to a finite dimensional parameter) and when we don't (non-parametric). Figure 1 uses the true distribution ϕ and shows:

- The set of (mean, standard deviation) pairs that belong to the identified set when solving the linear program with parametric constraints, in brown.

- The set of (mean, standard deviation) pairs that belong to the identified set when solving the original LP without parametric constraints, in green.

We do so for the four different distributions of the common value (Gaussian, Poisson, binomial and geometric) mentioned above. We remark that the non-parametric linear programs seems to recover the mean of the common value accurately in all cases. However, the bounds obtained on the second moment/standard deviation of the distribution of common values are far from being tight.

Figure 2 compares the identified set for the true distribution to the estimated set using Hoeffding with $\delta = 0.10$, as described in Section 5.1 for the Gaussian density function, for a number of samples $N \in \{10^3, 10^4, 10^5, 10^6\}$. We see that as N grows larger, the estimated set grows smaller and smaller and closer to the true identified set.

6.2 Parametric Identified Sets

In this section, we characterize the identified and estimated sets in the parametric case for the four distributions described above. For the Gaussian distribution, we provide figures of the identified parameters in the (μ, σ) space. For the other, 1-dimensional parameter distributions, we provide intervals for the parameter of the chosen distribution. We consider two different techniques to determine the estimated set:

- Using tolerances determined by Hoeffding's inequality, given in Section 5.1.2

(a) Gaussian density

(b) Poisson density

(c) binomial density

(d) geometric density
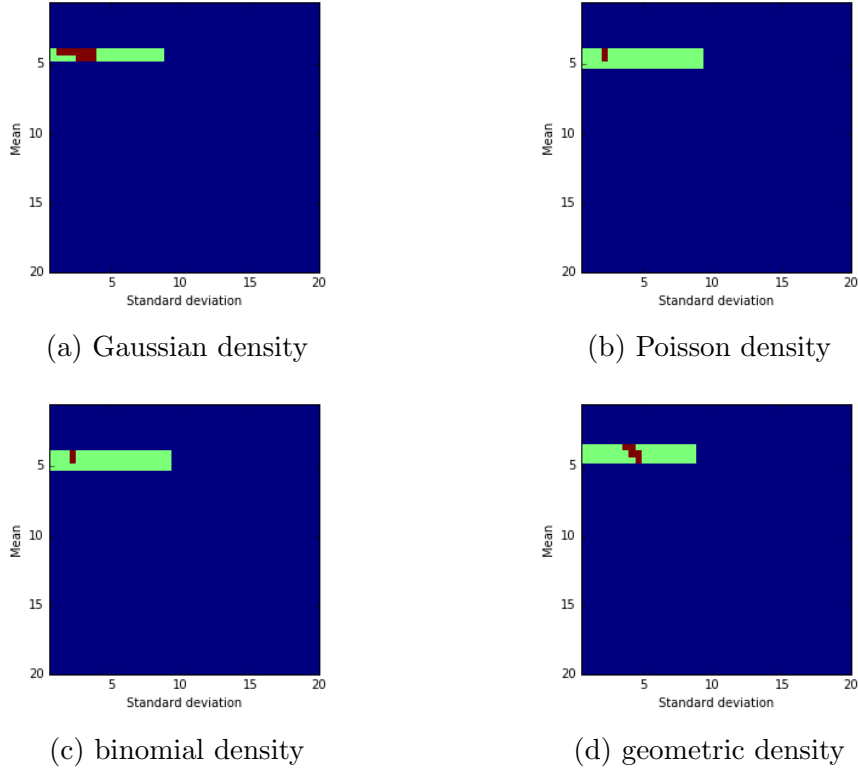
Figure 1: Plots of the non-parametric and parametric identified sets in the (mean, standard deviation) space

- Using quantiles of the tolerance via subsampling, as discussed in Section 5.2.2

In all figures, the brown region is the true identified set while the union of the brown and the green region is the estimated set. Figure 3 plots the identified and estimated set when using a Gaussian distribution for the common value and tolerances determined through Hoeffding with 90 percent confidence ($\delta = 0.10$), for the number of samples $N \in \{10^3, 10^4, 10^5, 10^6\}$. We remark that the number of samples needs be large (of the order of at least 10^5) for Hoeffding to perform well.

Figure 4 plots the identified and estimated set when using a Gaussian distribution for the common value and tolerances determined through subsampling and quantile estimation for the 90, 95 and 99 percent quantiles, as seen in Section 5.2.2; we only use $N = 100$ samples for the bid distribution in the three figures and $k = 50$ subsamples of size $s = N/4 = 25$. We note that the quantile estimation technique covers the true

(a) N = 1000

(b) N = 10000

(c) N = 100000

(d) N = 1000000
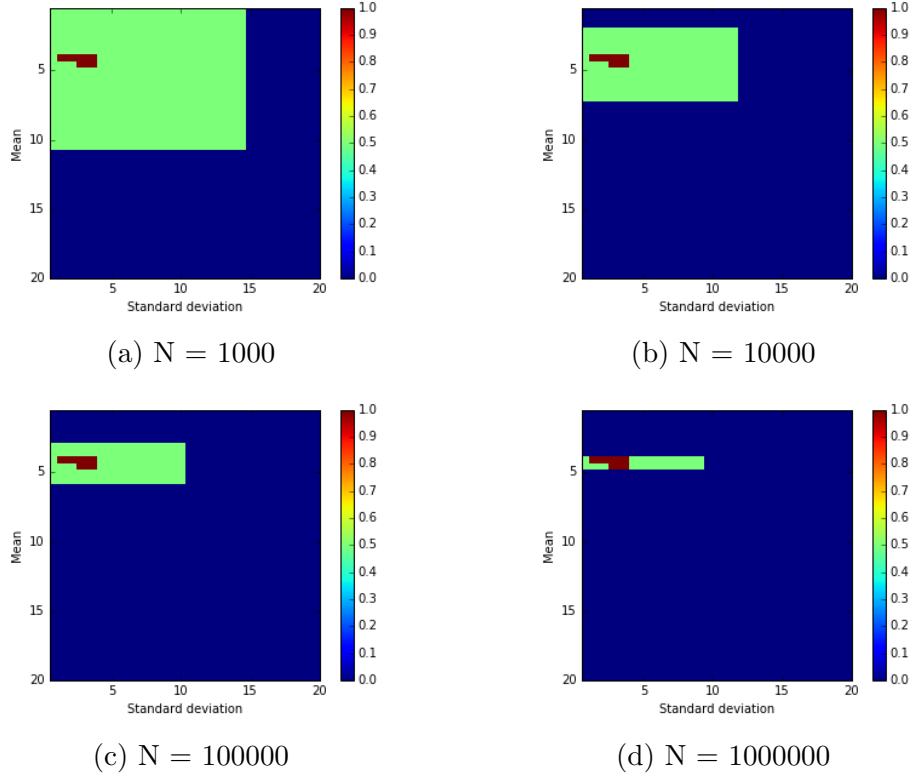
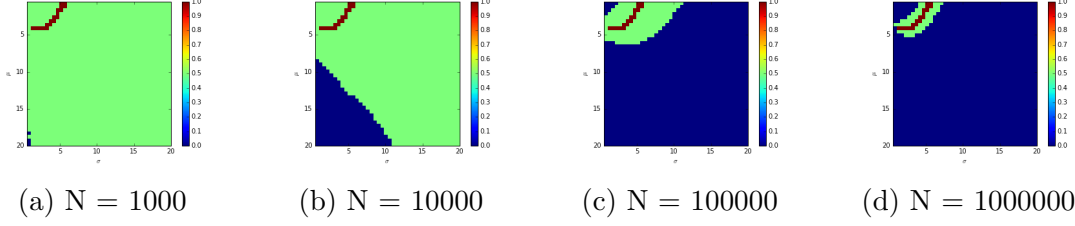Figure 2: Plots of the parametric identified sets and the non-parametric estimated set through Hoeffding with $\delta = 0.10$ in the (mean, standard deviation) space

identified set fairly sharply even though N is only equal to 100 and hence should be preferred to Hoeffding when small amounts of data are available.

Figure 5 gives the true identified interval for the parameters of the Poisson, binomial and geometric distributions and the estimated interval using subsampling for quantile estimation, for the 90, 95 and 99 percent quantiles – see 5.2.2 – for a bid distribution sampled with $N = 100$. We see that for the binomial distribution, the estimated set is very close to the true identified set. While the recovered sets for the Poisson distribution are not as sharp as the recovered sets for the binomial distribution, they still restrict the space of possible parameters in a reasonable way: the recovered interval is $[1.5, 6.5]$ while the space of possible parameters is $[0, 20]$. However, the recovered sets for the geometric distribution contains more than half of the possible parameters, which is unsatisfying. A reason for this comes from the fact that

(a) N = 1000   (b) N = 10000   (c) N = 100000   (d) N = 1000000

Figure 3: Plots of the parametric identified sets and the parametric estimated set through Hoeffding with $\delta = 0.10$ in the (μ, σ) space

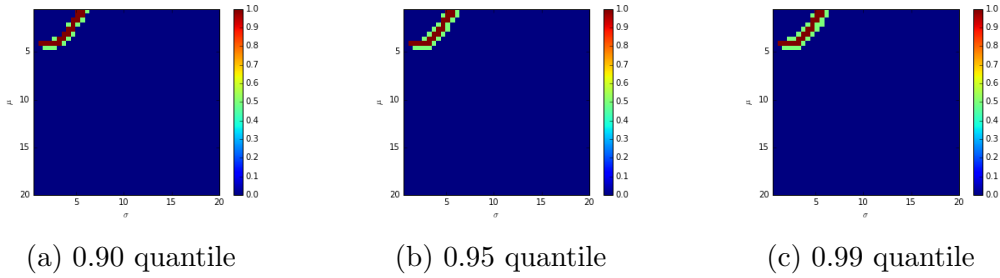(a) 0.90 quantile   (b) 0.95 quantile   (c) 0.99 quantile
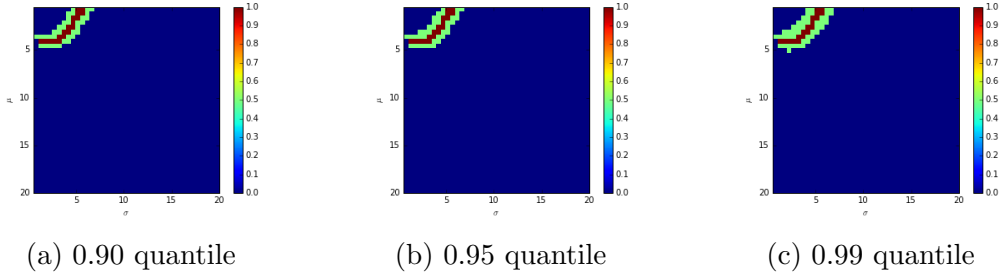
Figure 4: Plots of the parametric identified sets and the parametric estimated set through the quantile method in the (μ, σ) space for $N = 100$

the geometric distribution has a much higher variance than a Poisson or binomial with comparable means, hence there is a lot of variability across different subsamples that can lead to large tolerances.

Using $N = 500$ samples and $k = 50$ subsamples of size $s = N/4 = 125$, we obtain Figure 6. We can see that the estimated sets for the binomial and geometric distributions are now significantly tighter.

Finally, Figure 7 plots the identified and estimated set when using a Gaussian distribution for the common value and tolerances determined through subsampling and quantile estimation for the 90, 95 and 99 percent quantiles, as Figure 7. However,

Distribution	True identified set	0.90 quantile set	0.95 quantile set	0.99 quantile set
Binomial	[0.19,0.22]	[0.17,0.25]	[0.16,0.25]	[0.15,0.27]
Poisson	[4.0,4.5]	[1.5,6.5]	[1.5,6.5]	[1.5,6.5]
Geometric	[0.17,0.22]	[0.06,0.56]	[0.04,0.66]	[0.02,0.70]

Figure 5: Identified and estimated sets of the parameters of the distributions (N=100).

Distribution	True identified set	0.90 quantile set	0.95 quantile set	0.99 quantile set
Binomial	[0.19,0.22]	[0.18,0.24]	[0.17,0.25]	[0.17,0.25]
Poisson	[4.0,4.5]	[3.5,5.0]	[3.0,5.5]	[2.5,5.5]
Geometric	[0.17,0.22]	[0.16,0.25]	[0.15,0.28]	[0.14,0.29]

Figure 6: Identified and estimated sets of the parameters of the distributions (N=500).

we now use $N = 500$ samples for the bid distribution in the three figures and $k = 50$ subsamples of size $s = N/4 = 125$. We note that the estimated set now almost coincides with the true identified set.

(a) 0.90 quantile   (b) 0.95 quantile   (c) 0.99 quantile

Figure 7: Plots of the parametric identified sets and the parametric estimated set through the quantile method in the (μ, σ) space for $N = 500$

6.3 Parametric Identified Sets with Signal Constraints

In this section, we plot the identified sets in the parametric case for a Gaussian distribution, under additional assumptions on the signal structure.

More precisely, for $H = 20$, we use a Gaussian distribution for the common value with parameters $\mu = 4$ and $\sigma = 1$, as per Section 6.2. Under this distribution, we generate a Bayes correlated equilibrium of the game played between two bidders, that receive a common signal t on the common value v, such that $|v - t| \leq \epsilon^*$, as described in Remark 3 above. We pick $\epsilon^* = 2$.

We then compute the feasible sets of values of (μ, σ) for which the LP in Remark 3 is feasible. Here and for illustration, we consider the restriction that $|t - v| \leq \epsilon$, for ϵ equal to 2, 3, 5, and 20. We plot[15] these feasible sets in Figure 8; in Figure 9, we

---

[15]Due to numerical errors in the generation of the BCE, we allow ourselves some tolerance on the

translate these feasible sets into the (mean, standard deviation) space.

(a) $\epsilon = 2$

(b) $\epsilon = 3$
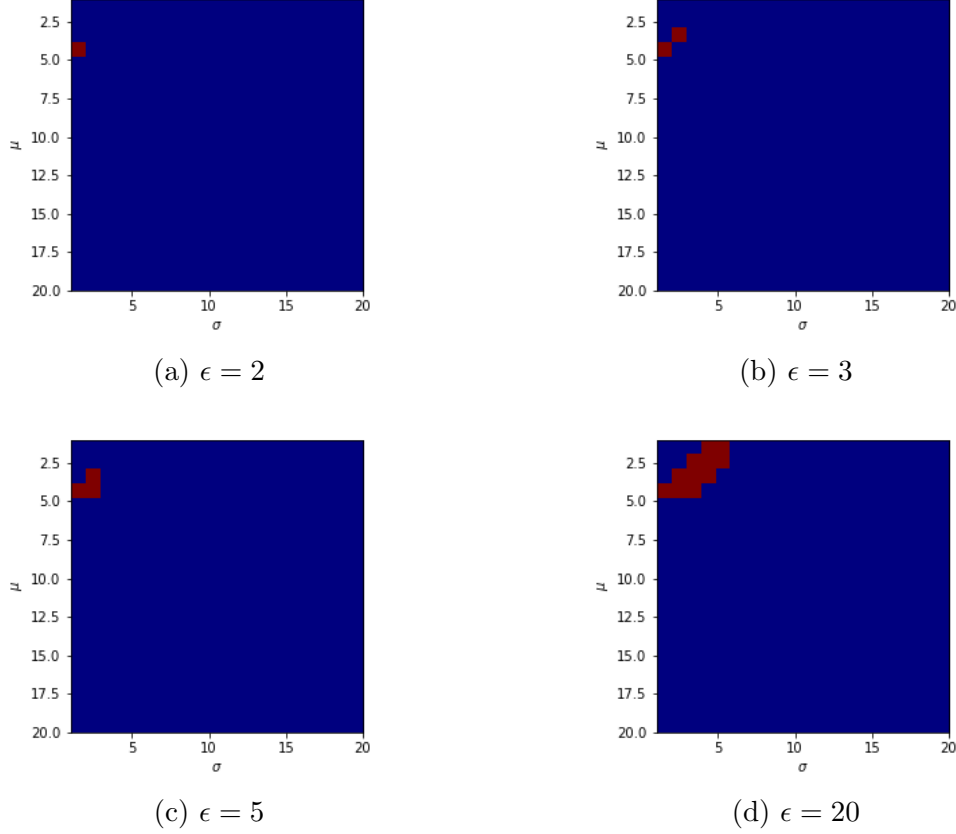
(c) $\epsilon = 5$

(d) $\epsilon = 20$

Figure 8: Plots of the parametric identified sets under signal constraints

Note that $\epsilon = 2$ corresponds to the strongest assumption we can make on the signal, knowing that the true data was generated according to $\epsilon^* = 2$. On the other hand, noting that $H = 20$, $\epsilon = 20$ corresponds to the assumption that the signal t is completely uninformative (equivalently, no assumption is made on the signal). We note on both figures that sharp knowledge of the support of the signal t leads to a significant improvement in the quality of recovery and identification, compared to when no assumption is made on the signal.

---

best response constraints in Program 21. We pick this tolerance to be 10^{-8}, as a result of noting that the minimum tolerance recovered under the assumption that $\epsilon = 2$ (when $\mu = 4$ and $\sigma = 1$) is of the order of $5 \cdot 10^{-9}$, while the second lowest tolerance is several orders of magnitude higher, of the order of 10^{-6}.

(a) $\epsilon = 2$ (b) $\epsilon = 3$

(c) $\epsilon = 5$ (d) $\epsilon = 20$
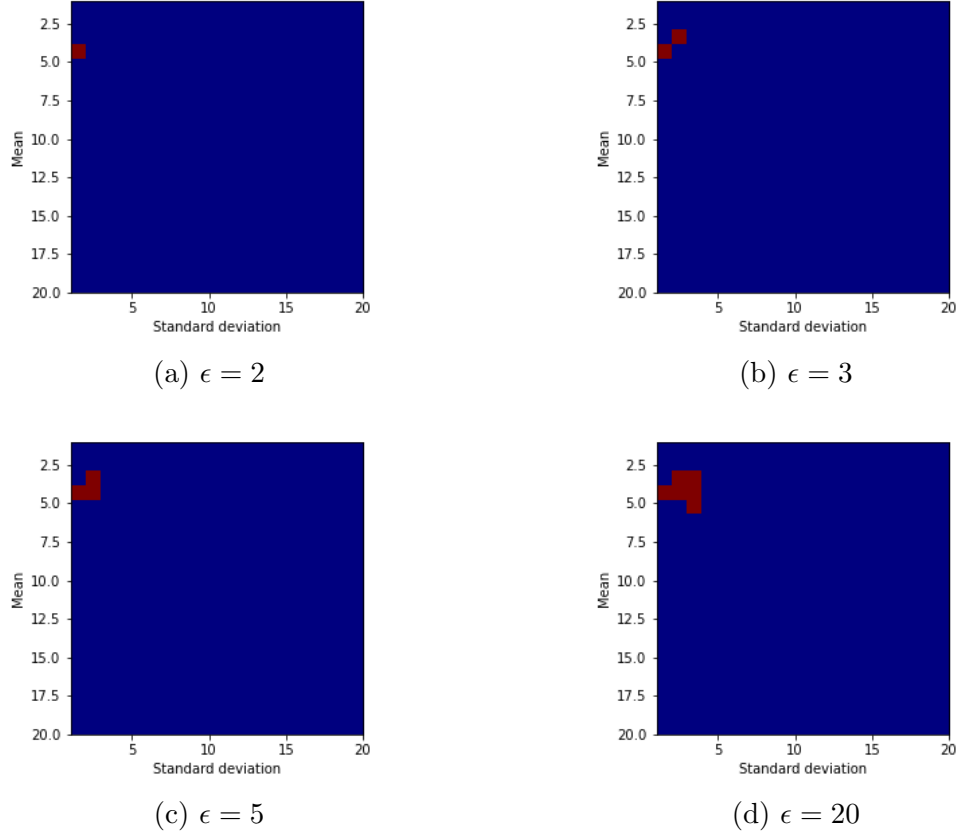
Figure 9: Plots of the parametric identified sets under signal constraints

7   Bidding and Counterfactuals in Ad Auctions

In a large scale application, we analyze auction data from the BingAds sponsored search auction market place. The dataset[16] consists of bidding data for 13 high-volume keywords[17] collected in a period of two weeks. The observations are ad auctions where the auctioneer, Bing Ads Platforms, is selling multiple slots for a particular search query (typically a keyword), among n bidders. These bidders participate in multiple

---

[16]See Noti and Syrgkanis [2020] for recent work using these data.

[17]Due to data confidentiality agreement, we are not able to reveal what these keywords are.

auctions. We aggregate the data by hour for each bidder. The hourly aggregated bid of a player is the average of the bids the player placed during that hour. The dataset further filters the observations and keeps bidders who participated in auctions in at least 100 hours and won the top position at least once during the whole two weeks of data. In addition, we require that the players place only positive bids and have non-zero variance. The full dataset includes multiple thousands of bidders, with an average of 202.4 active hours per bidder. This leads to an aggregate total number of auctions on the order of multiple millions.

The auction format used is a variant of the familiar Generalized Second Price auction (GSP). For analysis of GSP auctions, see Athey and Nekipelov [2010] and Edelman, Ostrovsky, and Schwarz [2007]. In these auctions, each bidder i has a private value v_i for getting a click, and each value is drawn independently across bidders and auctions, from a distribution with density $\rho_i(\cdot)$. Each slot has a different click through rate a_i. We consider an arbitrary black-box auction mechanism, in which each bidder submits a bid $b_i \in \mathbb{R}_+$. Based on the bid profile, the mechanism decides the allocation of slots to bidders and how much each bidder is going to pay. We let $X_i(b)$ denote the probability of click of bidder i for the slot that they are allocated and $C_i(b)$ the cost-per-click or the expected payment. We assume bidders have quasi-linear utility, i.e.:

$$u_i(b; v_i) = X_i(b)\ (v_i - C_i(b)) \tag{54}$$

and let $P_i(b) = X_i(b)\, C_i(b)$.

Our framework allows bidders to use different signals before submitting a bid. For example, in practice in these ad auctions, bidders may use different learning algorithms to decide on their bidding strategies (see e.g. Wordstream [2018]), and each learning algorithm may have access to different information based on history of play. In this case, the set of all possible signals is given by the history of play, and different users may have access to different parts of it. Moreover, depending on the sophistication of the bidder, they might be using some of the feedback that the auctioneer is providing

(e.g. the past average cost-per-click curves and allocation curves [Search Engine Land, 2014, Search Marketing Standard, 2014, Google, 2018a,b,c, Microsoft, 2018]).

7.1   An optimization framework for GSP auctions

In the rest of the section, we make the following conditional independence Assumption 14 on the Bayes correlated equilibrium selected by the players:

Assumption 14. *For every bidder i, and conditional on bid b_i, the bidder's value v_i is independent of the vector of other bidders' bids b_{-i}.*

Even though our framework does not require this assumption, given the large set of bidders in our auctions, maintaining it allows us to considerably simplify our procedures, by rewriting our estimation problem as a simple, separate problem for each bidder. To illustrate the computational problem with and without independence, first note the LP that recovers the density of each bidder's value in a non-parametric manner in the case of private value auctions is:

$$\forall i \in [n], v_i \in V, b_i^*, b_i' \in B : \quad \sum_{\mathbf{b} \in S : b_i = b_i^*} \phi(\mathbf{b}) \cdot x_i(v_i | \mathbf{b}) \cdot (u_i(\mathbf{b}; v_i) - u_i(b_i', \mathbf{b}_{-i}; v_i)) \geq 0$$

$$\forall i \in [n], v_i \in V : \quad \rho_i(v_i) = \sum_{\mathbf{b} \in S} \phi(\mathbf{b}) \cdot x_i(v_i | \mathbf{b})$$

$$\forall \mathbf{b} \in S : \quad x_i(\cdot | \mathbf{b}) \in \Delta(V)$$

Under our conditional independence assumption, we can simplify this program to get:

$$\forall i \in [n], v_i \in V, b_i^*, b_i' \in B : \quad \sum_{\mathbf{b} \in S : b_i = b_i^*} \phi(\mathbf{b}) \cdot x_i(v_i | b_i) \cdot (u_i(\mathbf{b}; v_i) - u_i(b_i', \mathbf{b}_{-i}; v_i)) \geq 0$$

$$\forall i \in [n], v_i \in V : \quad \rho_i(v_i) = \sum_{\mathbf{b} \in S} \phi(\mathbf{b}) \cdot x_i(v_i | b_i)$$

$$\forall b_i \in B : \quad x_i(\cdot | b_i) \in \Delta(V),$$

where now we only need to identify/estimate a lower dimensional object $x_i(\cdot | b_i)$, in-

stead of $x_i(\cdot|\mathbf{b})$. We are then keeping track of a much smaller number of variables[18].
Furthermore, note that the optimization program is immediately separable, and can be divided into n optimization programs, one for each bidder. This allows us to further reduce the running time of our algorithm, by focusing the optimization on only the subsets of bidders whose valuation distributions we want to infer.

Using the utility function of each bidder as in (54), we can re-write the best response constraints and obtain:

$$x_i(v_i \mid b_i^*) \cdot \left(v_i \sum_{\mathbf{b} \in S: b_i = b_i^*} \phi(\mathbf{b}) \cdot (X_i(\mathbf{b}) - X_i(b_i', \mathbf{b}_{-i})) \right.$$

$$\left. - \sum_{\mathbf{b} \in S: b_i = b_i^*} \phi(\mathbf{b}) \cdot (P_i(b) - P_i(b_i', \mathbf{b}_{-i})) \right) \geq 0$$

Note that when dividing the best response constraints by $\Pr[b_i^*]$, we obtain:

$$x_i(v_i \mid b_i^*) \left(v_i \cdot \mathbb{E}[X_i(\mathbf{b}) - X_i(b_i', \mathbf{b}_{-i}) \mid b_i^*] - \mathbb{E}[P_i(\mathbf{b}) - P_i(b_i', \mathbf{b}_{-i}) \mid b_i^*] \right) \geq 0. \quad (55)$$

Let us define

$$F_i(t \mid b_i) \triangleq \mathbb{E}[X_i(t, \mathbf{b}_{-i}) \mid b_i] \qquad \text{(conditional allocation curve)}$$

$$G_i(t \mid b_i) \triangleq \mathbb{E}[P_i(t, \mathbf{b}_{-i}) \mid b_i] \qquad \text{(conditional payment curve)}$$

Then, the best response condition in (55) becomes:

$$\Delta(i, v_i, b_i^*, b_i') := x_i(v_i \mid b_i^*) \left(v_i \left(F_i(b_i^* \mid b_i^*) - F_i(b_i' \mid b_i^*) \right) - \left(G_i(b_i^* \mid b_i^*) - G_i(b_i' \mid b_i^*) \right) \right) \geq 0$$

In turn, we obtain the following feasible set:

$$\forall i \in [n], v_i \in V, b_i^*, b_i' \in B: \qquad \Delta(i, v_i, b_i^*, b_i') \geq 0$$

$$\forall i \in [n], v_i \in V: \qquad \rho_i(v_i) = \sum_{b_i \in S} \phi_i(b_i) \cdot x_i(v_i|b_i) \qquad (56)$$

$$\forall b_i \in B: \qquad x_i(\cdot|b_i) \in \Delta(V)$$

---

[18]To illustrate, in the first LP above, and for each bidder i, there was one variable $x_i(v_i|\mathbf{b})$ for each bid vector $\mathbf{b}$. In turn, the number of such variables was growing exponentially in the number of bidders. However, in our simplified program, for each bidder i, there is only a variable for each (value,bid) pair $v_i, b_i \in V \times B$, i.e. $|V| \cdot |B|$ such variables, irrespective of the number of bidders, making the optimization problem tractable, even with large numbers of bidders.

where $\phi_i(b_i)$ is the marginal empirical distribution of the bid of player i. We use these sets of linear inequalities as the basis for estimating the distribution of valuations in these auctions.

7.2   GSP Counterfactuals

In addition to inference on the auction fundamentals, our framework allows us to conduct counterfactual analysis. For instance, what would the revenue or welfare be if we switch to a first price auction instead of the current mechanism, which is GSP like. For instance, we can write the *forward counterfactual LP* presented in Theorem 5 as follows:

$$LP(\phi, F, \tilde{u}) \quad \min_{\tilde{\psi}} / \max \sum_{\mathbf{v}, \mathbf{b}} F(\mathbf{v}, \mathbf{b}) \cdot \tilde{\psi}(\mathbf{v}, \mathbf{b})$$

$$\forall v_i, b_i, b_i' : \sum_{\mathbf{v}_{-i}, \mathbf{b}_{-i}} \tilde{\psi}(\mathbf{v}, \mathbf{b}) \cdot (\tilde{u}_i(\mathbf{b}; v_i) - \tilde{u}_i(b_i', \mathbf{b}_{-i}; v_i)) \geq 0$$

$$\forall v_i \in V : \ \rho(v_i) = \sum_{\mathbf{v}_{-i} \in V^{n-1}, \mathbf{b} \in B^n} \tilde{\psi}(\mathbf{v}, \mathbf{b})$$

$$\tilde{\psi} \in \Delta(V^n \times B^n) \text{ and } \rho \in \Pi_I(\phi)$$

where $F(\mathbf{v}, \mathbf{b})$ is an analyst specified objective such as revenue or welfare. See Section 2.3 for more on setting up the counterfactual LP. Note here that the above program for simulating the counterfactual uses the restrictions obtained from the status quo GSP auctions in that it requires that $\rho \in \Pi_I(\phi)$. Note also that for counterfactuals, we still need to solve an LP where the number of variables grows exponentially with the number of players. Hence, we have decided to focus on the largest two bidders in terms of average bids across all auctions and perform such counterfactuals on these bidders.

An interesting counterfactual that we estimate is the revenue that would ensue if we were to run a *first price auction* instead and just for the top slot. Given a click-through rate (i.e., the probability that someones who sees the ad clicks on it) of α_1

for the top slot, bidder i's utility is given by

$$u_i(\mathbf{b}; v_i) = \alpha_1 (v_i - b_i) 1\{b_i \geq \max_{j \neq i} b_j\}$$

and the counterfactual objective is the revenue, i.e. $F(v, b) = \alpha_1 \cdot \max_{i \in [n]} b_j$. We report the results from this counterfactual in Section 7.4 below.

7.3   Data Pipeline and LP Setup Details

For each keyword in our data, for every hour h within a two week period, and for every bidder i that participated in the auctions for that keyword, we are able to provide averaged hourly curves as follows:

- the average bid of each player $b_i^{(h)}$

- the average hourly (estimated) allocation curve $\hat{X}_i^{(h)}(t, \mathbf{b}_{-i}^{(h)})$

- the average hourly cost-per-click curve $C_i^{(h)}(t, \mathbf{b}_{-i}^{(h)})$. Using this, we can also construct the average hour payment curve as:

$$P_i^{(h)}(t, \mathbf{b}_{-i}^{(h)}) = \hat{X}_i^{(h)}(t, \mathbf{b}_{-i}^{(h)}) \cdot C_i^{(h)}(t, \mathbf{b}_{-i}^{(h)}).$$

In recent work, Noti and Syrgkanis [2020] examined the question of how best to estimate these allocation curves. The allocation and payment curves at the auction/hour level is generated by an auction engine provided by the platform [Bayir, Xu, Zhu, and Shi, 2019], that simulates outcomes for different bids of an advertiser. Subsequently, we fit parametric allocation and payment curves at the hourly level, based on minimizing a mean squared error criterion, across a variety of different parametrizations. Noti and Syrgkanis [2020] found that the following parametric form achieves the smallest mean squared error and so we use this specification at the hour/bidder level:

$$X_i^{(h)}(t, \mathbf{b}_{-i}^{(h)}) = \frac{a_i^{(h)}}{t + \gamma_i^{(h)}}$$

and the cost-per-click curve takes the form:

$$C_i^{(h)}(t, \mathbf{b}_{-i}^{(h)}) = c_i^{(h)} \cdot t$$

Thus the hourly payment curve takes the form:

$$P_i^{(h)}(t, \mathbf{b}_{-i}^{(h)}) = \frac{a_i^{(h)}}{t + \gamma_i^{(h)}} \cdot c_i^{(h)} \cdot t$$

Note that we are estimating a bidder and hour specific curve. In turn, the data are represented as a table where each row contains the id of the bidder i, the hour period h, the average bidder's bid $b_i^{(h)}$, and the three parameters $a_i^{(h)}, \gamma_i^{(h)}, c_i^{(h)}$. Note that these parameters are sufficient for obtaining empirical estimates of the hourly allocation and payment curves, and in turn to compute empirical estimates $\hat{F}_i(t \mid b_i)$ and $\hat{G}_i(t \mid b_i)$ of curves $F_i(t \mid b_i)$ and $G_i(t \mid b_i)$. More concretely, we treat each hour as a separate sample, select the set of samples in which bidder i submitted bid b_i^*, and we calculate the empirical allocation and payment curves $\hat{F}_i(t \mid b_i^*)$ and $\hat{G}_i(t \mid b_i^*)$ by using solely this subset of the samples. For $S(b_i) = \{h : b_i^{(h)} = b_i\}$, we have the following empirical analogues of the conditional allocation curve and the conditional payment curve:

$$\hat{F}_i(t \mid b_i) \triangleq \frac{1}{|S(b_i)|} \sum_{h \in S(b_i)} X_i^{(h)}(t, \mathbf{b}_{-i}) = \frac{1}{|S(b_i)|} \sum_{h \in S(b_i)} \frac{a_i^{(h)}}{t + \gamma_i^{(h)}} \tag{57}$$

$$\hat{G}_i(t \mid b_i) \triangleq \frac{1}{|S(b_i)|} \sum_{h \in S(b_i)} P_i^{(h)}(t, \mathbf{b}_{-i}) = \frac{1}{|S(b_i)|} \sum_{h \in S(b_i)} \frac{a_i^{(h)}}{t + \gamma_i^{(h)}} \cdot c_i^{(h)} \cdot t \tag{58}$$

The above estimated conditional allocation and payment curves will be the input to the optimization LP in (56) above. We can then conduct inference on the identified set of the distribution of each player's type either parametrically (e.g. by parameterizing the density) or non-parametrically, e.g. by estimating the identified sets of moments of the distribution.

7.4   Empirical Results

We first assume the distribution of each bidder's valuation has bounded and finite support $V = \{0, \ldots, H\}$. We renormalize the bids to be in $[0, \lceil \frac{H}{2} \rceil]$ – we pick $H/2$

because bidders may bid lower than their valuations, which can often be the case in non-truthful auctions such as GSP auctions, and discretize the set of bids to be $\{0, \ldots, \lceil \frac{H}{2} \rceil\}$. We do so by rounding each renormalized bid in each auction to the closest integer. In the simulations of this section, we pick $H = 20$.

For the parametric models, we assume that the valuation distribution of each bidder is a truncated Gaussian with support V. I.e., for each bidder i, there exists $\mu_i, \sigma_i \in \mathbb{R}$ such that the probability density function of the bidder's valuation is given by

$$f_i(v_i; \mu_i, \sigma_i) = \frac{\exp\left(-(v_i - \mu_i)/2\sigma_i^2\right)}{\sum_{v=0}^{V} \exp\left(-(v - \mu_i)/2\sigma_i^2\right)}.$$

In our inference approach, the distribution of the observed bids, $\phi(.)$, was estimated from the data. In here, in addition, we also estimate the $X(.)$ and $P(.)$ curves and so our confidence regions need to reflect this statistical uncertainty. We use the sub-sampling procedure described in our synthetic experiments in Section 6(see also Section 5.2.2) to construct confidence regions for the sharp identified sets, where sub-samples of the data are used to estimate the tolerance level that should use in the LP constraints. To incorporate the uncertainty in the estimates of $X(.)$ and $P(.)$, we also re-estimate these functions on each subsample. Finally, for each keyword, we select the 3 bidders that have the 3 highest average bids on that keyword across all the ad auctions. We then plot the confidence regions for the sharp sets of mean and variance of the value distribution for those top 3 bidders. For the construction of the confidence sets, for each bidder i we consider, we use 100 subsamples, each of size $n_i/4$—where n_i is the number of times i appeared in an auction for that keyword—, and we pick the tolerance level to be the 95% quantile of the minimum tolerances of the sub-samples. We run the experiments for 13 different keywords. Figure 10 shows the results for the 3 bidders for 4 keywords. Figures 19 and 20 in the Appendix provide bounds for the rest of the keywords.

We note that in many of our results in the parametric case, the confidence regions are relatively tight. We also note that even though bidders are ordered based on their

bid level, it is not necessary that their valuation distributions should also respect that order. The reason is that sponsored search auctions typically multiply the bid of an advertiser by a quality score, which is a proxy of the quality of the ad for the specific keyword. Thus, advertisers can potentially lose an auction even if they bid higher, because they might be less relevant. In turn, different advertisers are essentially facing a different price landscape and hence might need to bid higher than others to win, even when they have lower values.

To see the identifying power of the truncated normal parametric distributional assumptions, Tables 1 and 2 provide confidence intervals on the mean valuation for the 3 bidders and all 13 keywords maintaining the truncated normal assumption (Table 1) and when the valuation distribution is nonparametric (Table 2). Notice for instance that for Keyword0, the CI for the mean valuation for bidder 1 under the parametric model is $[4.06, 12.09]$ while under the nonparametric model it is $[2.14, 12.87]$, a modest difference. This is in contrast to the mean valuation for this keyword for Bidder 3 where the parametric CI is $[2.01, 2.01]$ while the nonparametric is $[1.11, 4.58]$, a large increase.

	bidder1	bidder2	bidder3
keyword0	[4.06, 12.09]	[3.69, 5.90]	[2.01, 2.01]
keyword1	[6.00, 6.00]	[5.00, 5.00]	[4.00, 4.00]
keyword2	[7.00, 19.48]	[0.52, 6.50]	[8.67, 13.58]
keyword3	[2.01, 2.01]	[3.00, 3.00]	[3.00, 3.00]
keyword4	[9.00, 9.00]	[6.00, 7.00]	[6.00, 6.00]
keyword5	[3.00, 3.00]	[0.52, 0.52]	[0.52, 0.52]
keyword6	[4.06, 6.11]	[5.02, 5.02]	[3.00, 3.18]
keyword7	[8.02, 14.78]	[7.00, 7.00]	[7.00, 7.00]
keyword8	[11.84, 14.71]	[4.06, 4.06]	[4.06, 4.41]
keyword9	[9.00, 9.00]	[6.00, 6.00]	[4.06, 5.23]
keyword10	[3.00, 3.00]	[2.01, 2.01]	[2.01, 2.89]
keyword11	[0.52, 0.52]	[0.52, 0.52]	[3.00, 3.00]
keyword12	[9.00, 9.00]	[7.28, 11.07]	[7.00, 8.00]

Table 1: Parametric 95% Confidence Intervals for the mean of the private value distribution for each keyword and each of the top three bidders, based on a truncated normal parametric distribution. The minimum and maximum value allowed for the value distribution is $[0, 20]$.
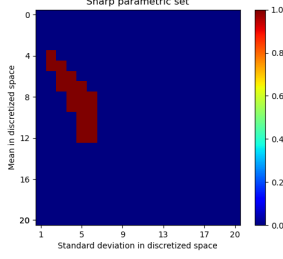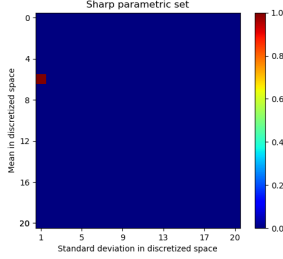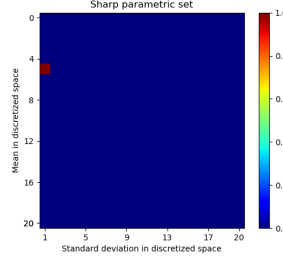
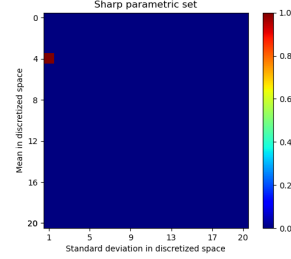(a) 1st bidder, keyword 0
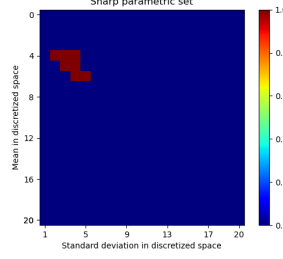
(b) 2nd bidder, keyword 0
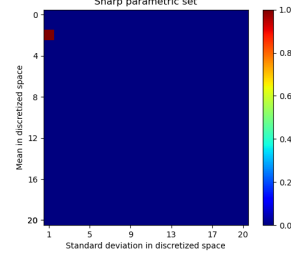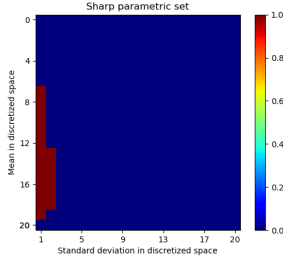
(c) 3rd bidder, keyword 0
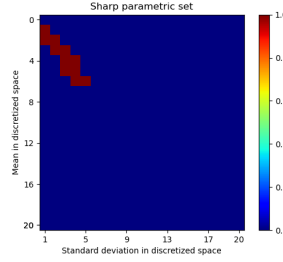
(d) 1st bidder, keyword 1
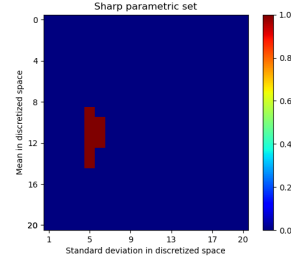
(e) 2nd bidder, keyword 1
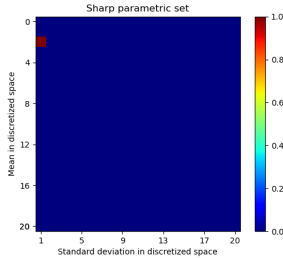
(f) 3rd bidder, keyword 1
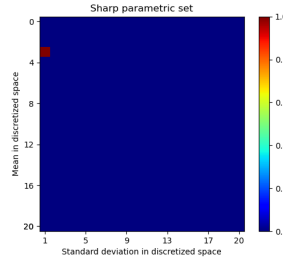
(g) 1st bidder, keyword 2

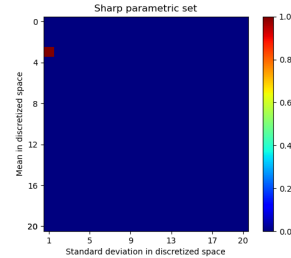(h) 2nd bidder, keyword 2

(i) 3rd bidder, keyword 2

(j) 1st bidder, keyword 3

(k) 2nd bidder, keyword 3

(l) 3rd bidder, keyword 3

Figure 10: 95% confidence sets on identified sets of the mean and standard deviation of the distribution of values for each of the top three bidders of each keyword, for a sample of three of the 13 keywords in our analysis, based on a truncated normal parametric distribution assumption.

51

	bidder1	bidder2	bidder3
keyword0	[2.14, 12.87]	[1.06, 8.14]	[1.11, 4.58]
keyword1	[4.88, 7.35]	[3.98, 6.56]	[3.59, 4.89]
keyword2	[1.00, 20]	[0.21, 8.25]	[7.06, 14.60]
keyword3	[1.61, 4.00]	[1.75, 4.89]	[1.68, 4.74]
keyword4	[7.51, 11.02]	[5.00, 8.67]	[4.81, 7.99]
keyword5	[1.74, 5.52]	[0.00, 4.40]	[0.00, 4.66]
keyword6	[3.45, 7.93]	[3.33, 6.69]	[1.68, 4.95]
keyword7	[6.23, 19.12]	[6.06, 9.05]	[6.15, 8.86]
keyword8	[9.97, 16.61]	[2.29, 7.00]	[2.26, 6.93]
keyword9	[7.86, 11.24]	[4.51, 8.00]	[2.74, 7.50]
keyword10	[1.72, 5.20]	[1.52, 4.40]	[0.81, 5.95]
keyword11	[0.00, 4.81]	[0.00, 4.88]	[1.60, 5.79]
keyword12	[8.00, 11.37]	[5.45, 13.65]	[5.75, 9.66]

Table 2: Non-parametric 95% Confidence Intervals for the mean of the private value distribution for each keyword and each of the top three bidders. The minimum and maximum value allowed for the value distribution is $[0, 20]$

Counterfactuals. In Table 3, we provide intervals for simulated counterfactual revenues for running a first price auction with a click through rate of 1 among the top two bidders in the original GSP auctions. These counterfactuals were constructed non-parametrically, without any assumptions on the marginal distribution of valuations. As for the results, note here that for certain keywords, these counterfactuals lead to tight predictions (such as Keywords 1, 3 and 10) while for some others (Keywords 0 and 2) the intervals are wider. Of course a parametric restriction on the valuation distribution would tighten up these counterfactuals. Similarly, shape constraints on the non-parametric density, as described in Section 2.3, could also tighten these counterfactuals.

8   Empirical Illustration 2: OCS Auctions

In this section, we use the Outer Continental Shelf (OCS) Auction Dataset that was used in the seminal work of Hendricks and Porter (see Hendricks and Porter [1988]). The dataset contains bidding information on 3036 tract auctions in Louisiana and Texas. In particular, for each auction, our dataset contains the acreage of the tract and the total bid of each participant in the auction. We assume that the bidders

	lower bound	upper bound
keyword0	0.48	9.83
keyword1	3.84	5.19
keyword2	0.01	10.00
keyword3	0.05	2.05
keyword4	4.92	8.41
keyword5	0.07	6.68
keyword6	2.09	5.30
keyword7	5.72	9.72
keyword8	2.22	9.99
keyword9	4.59	8.49
keyword10	0.12	2.03
keyword11	0.01	3.52
keyword12	4.92	9.42

Table 3: Non-parametric identified sets with 95% confidence, for counterfactual revenue in a first price auction for a single slot with a click-through-rate of 1 and among the top two bidders of the GSP auction.

participate in a first price common value auction, where the value is defined per acre; our goal is i) to show that indeed, the bidders' behavior in the data can be explained by a common value auction (via the testable restriction of whether the estimated identified set is empty), and ii) to recover the first and second order moments of the distribution of said value. This is under weak assumptions on information in that the framework allows bidders in different auctions to know more information about the environment.

Pre-processing of data: The dataset contains 3036 auction with varying number of players. We consider 2-player common value auctions, hence we only keep the entries in the dataset that contain exactly 2 bidders; there are 584 such auctions. We model the two bidders as being the same over the 584 auctions, and assign bidders' identities to be 1 or 2 uniformly at random in each auction. Many of the bids we have are zero and Figure 11 plots the distribution of bids; it has mean $991.48 and standard deviation $1825.43.
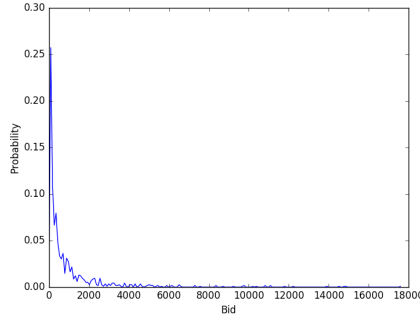
Figure 11: Distribution of bids

We assume the distribution of the common value per acre has bounded finite support $V = \{0, \ldots, H\}$. We renormalize the bids per acre to be in $[0, \lceil \frac{H}{2} \rceil]$ – we pick $H/2$ because the common value could have a distribution whose support goes beyond the observed bids –, and discretize the set of bids to be $\{0, \ldots, \lceil \frac{H}{2} \rceil\}$; we do so by rounding each renormalized bid in each auction to the closest integer. We remark that the dataset contains a few outliers whose bid per acre is significantly higher than in all other auctions; we therefore delete the auctions that contain bids over threshold $t = \$20000$. We further assume that the distribution of the common value is given by a truncated normal distribution that takes discrete values in $\{0, \ldots, H\}$, exactly as described in the simulations of Section 6, and parametrize all optimization problems we solve accordingly.

8.1 Results Absent Signal Constraints

In all figures, the value of (μ, σ) are given as the values in dollars instead of the corresponding discretized and renormalized value, for the sake of comparison with the $\$20,000$ threshold and the corresponding maximum value of $\$40,000$. These parameter pairs are placed on a grid. Figure 12 plots a heat map of the estimated set as a function of the chosen tolerance in the (μ, σ) space for two different values of H. Each color on the heat map corresponds to a tolerance level, and the mapping from tolerance levels to colors is given by the colorbar on the right of each figure. The color that is assigned

to any given (μ, σ) pair corresponds to the minimum level of tolerance that the analyst needs to add to the equilibrium constraints for (μ, σ) to belong to the estimated set; therefore, the heat map shows how the estimated set grows as the tolerance picked by the analyst increases. Note also that there seems to be a region of the parameter where the minimum tolerance is close to zero indicating that the model is not rejected by the data[19].
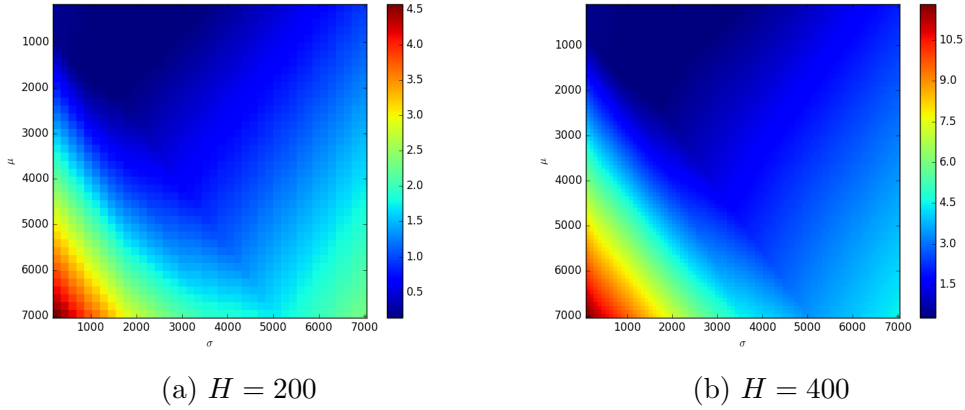

(a) $H = 200$ (b) $H = 400$

Figure 12: Heat map of minimum tolerance needed for (μ, σ) pairs to belong to the estimated set

Figure 13 plots the estimated set when using the tolerance determined by the subsampling approach of Section 5.2.2, in green using 95% level, and compares it to the estimated set using the minimum tolerance for which the estimated set is non-empty, in brown, for $H = 400$. No matter what method is used for picking the tolerance and determining the corresponding confidence intervals, the region that is obtained cannot be possibly smaller than the minimum tolerance set unless it is empty. We remark that the estimated set for (μ, σ) remains small relatively to the upper bound of $\$20,000$ on the bids and of $\$40,000$ on the maximum common value. We can see from the confidence regions that the mean of the common values varies from zero 0 to around $\$4000$ while the standard deviation varies from close to zero to 6000. It is

---

[19]This can formalized into a statistical test (of whether behavior conforms to BNE with augmenting information structure). We do not deal with the testability question formally in this paper and will leave that for future work.

also possible to estimate these means as functions of covariates and hence to allow for observed heterogeneity.
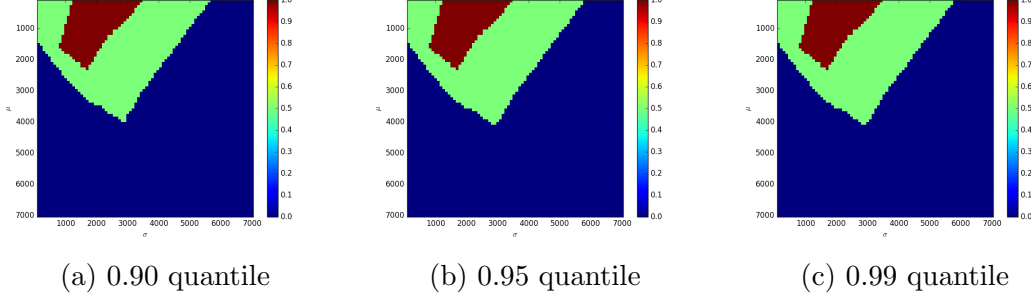

(a) 0.90 quantile   (b) 0.95 quantile   (c) 0.99 quantile

Figure 13: Estimated sets using the 90, 95 and 99 percent quantiles, for $H = 400$.

The above plots the mean and variance of a normal density that is then truncated. Since the truncated normal has a different mean and variance than the underlying normal, we plot in Figure 14 the minimum tolerance and the estimated sets as a function of this overall mean and standard deviation of the distribution of common values that we identify, instead of the parameters μ, σ of the truncated normal distribution. We obtain these plots by computing a mapping from (μ, σ) pairs to (mean, standard deviation) pairs, and plot the image of Figure 13 by said mapping in the (mean, standard deviation) space. We note that the estimated set is fairly small, indicating that our approach identifies the first two moments of the true, underlying distribution of the common value in an accurate fashion, despite only having access to a limited number of samples.

(a) 0.90 quantile   (b) 0.95 quantile   (c) 0.99 quantile
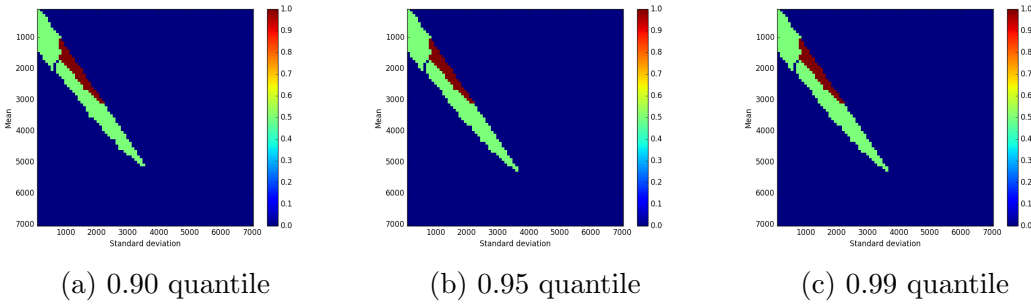
Figure 14: Estimated sets using the 90, 95 and 99 percent quantiles, for $H = 400$.

8.2 Results With Signal Constraints

Here, we use a particular form of a signal constraint to try and shrink the size of the identified set. In particular, we assume that the analyst believes all bidders observe a common signal t such that $|t-v| \leq \epsilon$, where v is the common value of the auction and ϵ is a (small) integer. This analysis based on constraints on signals follows the treatment in Section 2.4 above. Note again that these type of constraints allow the signals to be *arbitrarily* correlated with the underlying valuation within the specified bound. No assumption is made on the joint distribution of the signal and the valuation.

Since we are solving a larger optimization problem with more constraints, we consider smaller values of H compared to the previous experiments, to optimize the running time. The Figures below plot the estimated confidence sets when using the tolerance determined by the subsampling approach of Section 5.2.2. These are colored in green at the 95% level. The sets using the minimum tolerance are in brown and here we use $H = 20$ and repeat the exercise for different values of ϵ. The way to read these graphs is to compare ones where ϵ is large (no constraint on signals) to plots where ϵ is small. Any reduction of the estimated sets as ϵ decreases is due to constraints on the signal space (and hence if there is no shrinkage in the size of these sets, then one can deduce that signals do not play a role). Figure 15 plots the mean/variance CIs of the truncated normal while Figure 16 maps those results to the (mean,standard deviation) space. We note that sharper assumptions (smaller values of ϵ) on the informativeness of signal t lead to a significant reduction in the size of the estimated sets and hence using prior information on ϵ can further reduce the size of the identified sets. This reduction is replicated in Figures 17 and 18. These Figures are easier to see since we consider here a more fine-grained discretization of the value space, with $H = 60$. We note that even going from $\epsilon = 0$ to $\epsilon = 5$ (which is already a significantly informative signal in the context of $H = 60$), leads to a non-negligible increase in the sizes of the estimated identified sets. These constraints on signals seem to be an attractive approach to allow for arbitrary signals while controlling their

signal "strength." Plotting these sets for multiple values of ϵ (with larger ϵ indicating wider signals) is an informative way to visualize the importance of arbitrary signals in affecting the size of the identified set.
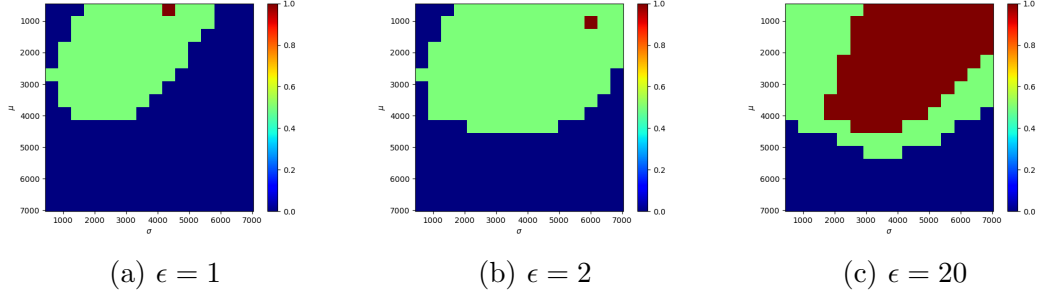

(a) $\epsilon = 1$ (b) $\epsilon = 2$ (c) $\epsilon = 20$

Figure 15: Estimated sets using $\epsilon = 1$, 2, and 20, for $H = 20$.

(a) $\epsilon = 1$ (b) $\epsilon = 2$ (c) $\epsilon = 20$
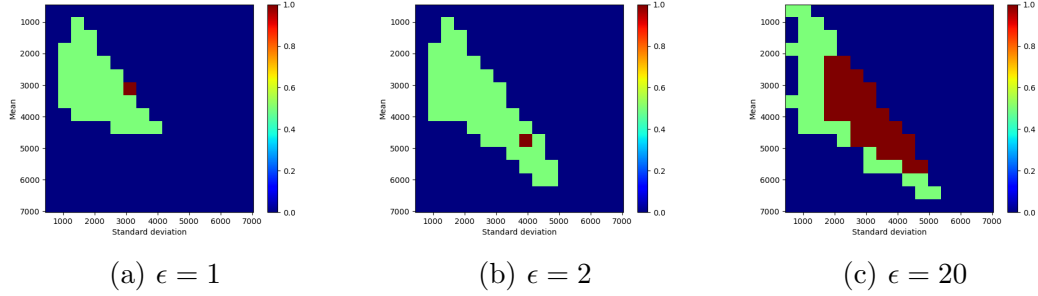
Figure 16: Confidence Sets on the mean and variance from a truncated normal using $\epsilon = 1$, 2, and 20, for $H = 20$.

9 Conclusion

We provide a framework for inference on auction fundamentals and counterfactuals without making strong restrictions on information. Using data, we use the recent results in theory to characterize the identified set for these objects by exploiting the linear programming structure of the set of Bayesian correlated equilibria. Our results can also be used by mechanism designers in that the data allows us to restrict the domain of signal/state of the world distribution, which would lead to sharper mechanisms. We also provide approaches to inference, and propose finite sample approaches

(a) $\epsilon = 0$ (b) $\epsilon = 2$ (c) $\epsilon = 5$
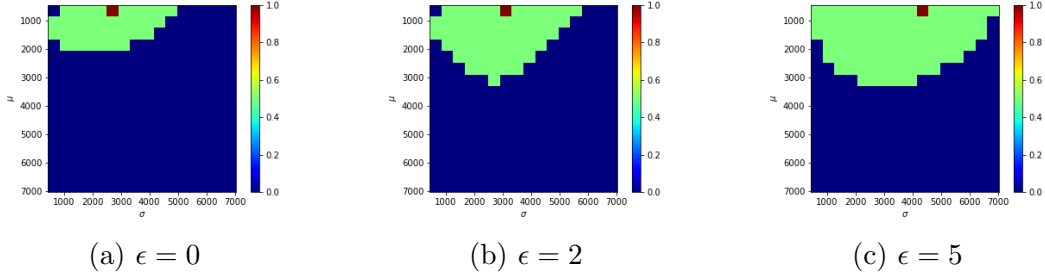
Figure 17: Confidence Set for mean and variance in the overall mean/st. deviation space using $\epsilon = 0$, 2, and 5, for $H = 60$.

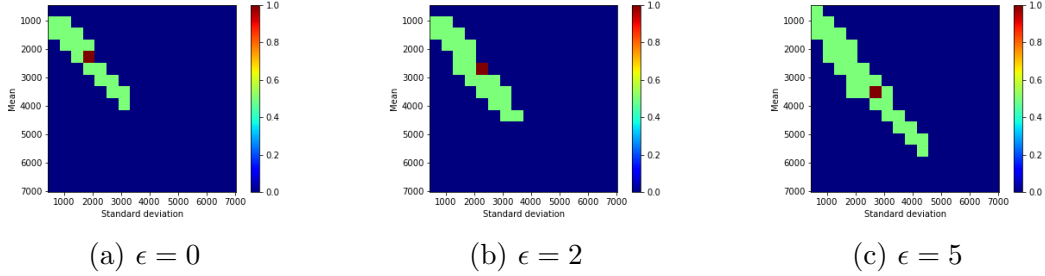(a) $\epsilon = 0$ (b) $\epsilon = 2$ (c) $\epsilon = 5$

Figure 18: Estimated sets using $\epsilon = 0$, 2, and 5, for $H = 60$.

to building confidence regions for sets in partially identified models. Finally, we apply our results to two data sets where we construct confidence regions on the mean and variance of the underlying valuation distributions and also on counterfactual revenues.

Though our focus is on the primitives and counterfactuals, our LP can in principle be used to learn about the joint distribution of signals. Since this is a possibly high dimensional object, a parametric approach to inference on such distributions is likely to be more informative. We leave this question for future work as estimating information structures and the distribution of signals is an important ingredient in some design problems.

References

S. Athey and D. Nekipelov. A structural model of sponsored search advertising auctions. In *Sixth ad auctions workshop*, 2010.

J.-Y. Audibert, R. Munos, and C. Szepesvári. Exploration–exploitation tradeoff using variance estimates in multi-armed bandits. *Theoretical Computer Science*, 410(19):1876 – 1902, 2009. ISSN 0304-3975. doi: https://doi.org/10.1016/j. tcs.2009.01.016. URL `http://www.sciencedirect.com/science/article/pii/ S030439750900067X`. Algorithmic Learning Theory.

R. J. Aumann. Correlated equilibrium as an expression of bayesian rationality. *Econometrica: Journal of the Econometric Society*, pages 1–18, 1987.

M. A. Bayir, M. Xu, Y. Zhu, and Y. Shi. Genie: An open box counterfactual policy estimator for optimizing sponsored search marketplace. In *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining*, WSDM'19, pages 465–473, New York, NY, USA, 2019. Association for Computing Machinery. ISBN 9781450359405. doi: 10.1145/3289600.3290969. URL `https://doi.org/10.1145/ 3289600.3290969`.

D. Bergemann and S. Morris. Robust predictions in games with incomplete information. *Econometrica*, 81(4):1251–1308, 2013. ISSN 1468-0262. doi: 10.3982/ ECTA11105. URL `http://dx.doi.org/10.3982/ECTA11105`.

D. Bergemann and S. Morris. Bayes correlated equilibrium and the comparison of information structures in games. *Theoretical Economics*, 11(2):487–522, 2016. URL `https://econtheory.org/ojs/index.php/te/article/view/20160487/0`.

D. Bergemann, B. Brooks, and S. Morris. First-price auctions with general information structures: Implications for bidding and revenue. *Econometrica*, 85(1):107–143, 2017. ISSN 1468-0262. doi: 10.3982/ECTA13958. URL `http://dx.doi.org/10. 3982/ECTA13958`.

D. Bergemann, B. A. Brooks, and S. Morris. Counterfactuals with latent information. 2019.

G. Chamberlain and G. W. Imbens. Nonparametric applications of bayesian inference. *Journal of Business & Economic Statistics*, 21(1):12–18, 2003.

V. Chernozhukov, H. Hong, and E. Tamer. Estimation and confidence regions for parameter sets in econometric models. *Econometrica*, 75(5):1243–1284, 2007.

B. Edelman, M. Ostrovsky, and M. Schwarz. Internet advertising and the generalized second-price auction: Selling billions of dollars worth of keywords. *American economic review*, 97(1):242–259, 2007.

Google. AdWords Bid Simulator. `https://support.google.com/adwords/answer/2470105?hl=en&ref_topic=3122864`, 2018a. [Online; accessed 15-February-2018].

Google. Bid Landscapes. `https://developers.google.com/adwords/api/docs/guides/bid-landscapes`, 2018b. [Online; accessed 15-February-2018].

Google. Bid Lanscapes. `https://developers.google.com/adwords/api/docs/reference/v201710/DataService.BidLandscape`, 2018c. [Online; accessed 15-February-2018].

P. A. Haile and E. Tamer. Inference with an incomplete model of english auctions. *Journal of Political Economy*, 111(1):1–51, 2003.

K. Hendricks and R. H. Porter. An empirical study of an auction with asymmetric information. *The American Economic Review*, pages 865–883, 1988.

D. Hoy, D. Nekipelov, and V. Syrgkanis. Robust data-driven efficiency guarantees in auctions. *CoRR*, abs/1505.00437, 2015. URL `http://arxiv.org/abs/1505.00437`.

B. Kline and E. Tamer. Bayesian inference in a class of partially identified models. *Quantitative Economics*, 7(2):329–366, 2016.

L. Magnolfi and C. Roncoroni. Estimation of discrete games with weak assumptions on information. Technical report, mimeo, 2016.

A. Maurer and M. Pontil. Empirical Bernstein Bounds and Sample Variance Penalization. *ArXiv e-prints*, July 2009.

Microsoft. BingAds, Bid Landscapes. `https://advertise.bingads.microsoft.com/en-us/resources/training/bidding-and-traffic-estimation`, 2018. [Online; accessed 15-February-2018].

G. Noti and V. Syrgkanis. Bid prediction in repeated auctions with learning. *arXiv preprint arXiv:2007.13193*, 2020.

T. Peel, S. Anthoine, and L. Ralaivola. Empirical bernstein inequalities for u-statistics. In *NIPS*, 2010.

Search Engine Land. Bing ads launches "bid landscape", a keyword level bid simulator tool. `https://searchengineland.com/bing-ads-launches-bid-landscape-keyword-level-bid-simulator-tool-187219`, 2014. [Online; accessed 15-February-2018].

Search Marketing Standard. Google adwords improves the bid simulator tool feature. `http://www.searchmarketingstandard.com/google-adwords-improves-the-bid-simulator-tool-feature`, 2014. [Online; accessed 15-February-2018].

A. Shapiro. *5. Statistical Inference*, pages 155–252. 2009. doi: 10.1137/1.9780898718751.ch5. URL `http://epubs.siam.org/doi/abs/10.1137/1.9780898718751.ch5`.

V. Syrgkanis and E. Tardos. Composable and efficient mechanisms. In *Proceedings of the Forty-fifth Annual ACM Symposium on Theory of Computing*, STOC '13, pages 211–220, New York, NY, USA, 2013. ACM. ISBN 978-1-4503-2029-0. doi: 10.1145/2488608.2488635. URL `http://doi.acm.org/10.1145/2488608.2488635`.

Wordstream. Bid management tools. `https://www.wordstream.com/bid-management-tools`, 2018. [Online; accessed 15-February-2018].

A Non-sharpness of the BBM bound

We examine the upper and lower bounds on the mean from BMM. In a common value auction an easy worst-case bound on the mean of the common value distribution comes from the following fact: at any BCE players are getting non-zero expected utility. Thus an obvious lower bound on mean value is the observed revenue. We can try to show that mean value is also not much more than the revenue or some function of the observed bids that is close to the revenue. For instance, the BBM bound shows that:

$$v(q) - \frac{1}{q^{(n-1)/n}} \int_0^{v(q)} (F(x))^{(n-1)/n} dx \leq B(q) \tag{59}$$

where $q \in [0, 1]$ is a quantile, $v(q) = F^{-1}(q)$, $F(x)$ is the CDF of the value distribution and $B(q)$ is the quantile function of the maximum bid distribution. This is equivalent to:

$$\frac{1}{q^\alpha} \int_0^q \alpha y^{\alpha-1} v(y) dy \leq B(q) \tag{60}$$

From this we can try to upper bound the mean of F (i.e. $\int_0^1 v(q) dq$) as a function of the revenue, which is simply the mean of the maximum value distribution (i.e. $\int_0^1 B(q) dq$). If the CDF is convex, $F(0) = 0$ and $F(H) = 1$, then $F(x) \leq x/H$ hence the BBM bound gives:

$$v(q) \left(1 - \frac{n}{2n-1} \frac{1}{H^{(n-1)/n}} \left(\frac{v(q)}{q} \right)^{(n-1)/n} \right) \leq B(q) \tag{61}$$

By integrating and if we let μ be the mean of the common value and R be the observed revenue of the auction:

$$\mu \leq R + \frac{n}{2n-1} \frac{1}{H^{(n-1)/n}} \int_0^1 \frac{v(q)^{(2n-1)/n}}{q^{(n-1)/n}} dq$$

This shows that there is a reasonable upper bound. For instance for the uniform distribution, i.e. $v(q) = q \cdot H$, this gives $\mu \leq R + \frac{n}{2n-1} H \int_0^1 q dq = R + H \frac{n}{2n-1} \frac{1}{2} \leq R + H/3$. Observe that if the uniform distribution is allowed in our set of possible value distributions then we cannot hope to prove any better upper bound.

For the case of two bidders we can now show that the BBM bound implies that the mean of the auction must be at most:

$$\mu \leq R \cdot \left(2\sqrt{\frac{H}{R}} - 1 \right) \tag{62}$$

where R is the observed revenue of the auction and H is the assumed upper bound on the distribution. Thus if the revenue is at least H/α for some α, then we can conclude that the mean must leave in the range: $[R, R(2\sqrt{\alpha} - 1)]$ which is a non-trivial sharp identified set. In other words the ratio of the upper to lower bound of the sharp identified set is at most $2\sqrt{\alpha} - 1$.

For n bidders we show (see Section A.1 below) that the BBM bound implies that the mean of the distribution cannot be more than: $\sqrt{\frac{2n}{n-1} \cdot R \cdot H}$ or alternatively $\sqrt{\frac{2n}{n-1} \cdot R \cdot \left(\frac{1}{n}R + L \right)}$ if the density of values is bounded from below by $1/L$. Thus if the revenue is at least H/α, then we have that μ lies in the range: $[R, R \cdot \sqrt{\frac{2n}{n-1}\alpha}]$. In fact the above bound is tight if one only uses the deviation of BBM, i.e. if all we know is the inequality implied by these deviations, then there exists a distribution that satisfies that inequality and which achieves the above bound.

Non-sharpness of BBM: The upper bound of BBM can be very far from the identified set. Before we give a concrete example, we first argue why the bound from Bergemann et al. [2017] cannot possibly be sharp: The main point of the work of BBM was getting worst-case bounds on the revenue of a common value first price auction as a function of the distribution of values. For that reason in the analysis, the bid distribution is not part of the input. Their main approach is to claim that the revenue cannot be too small. The idea behind the analysis is as follows: since something is a BCE, no player wants to deviate to any other action. Hence, they do not want to deviate to a specific style of a deviation which is: conditional on your bid deviate uniformly at random above your bid (upwards deviation). Such deviation arguments were also used in Hoy, Nekipelov, and Syrgkanis [2015], Syrgkanis and Tardos [2013] to give bounds on the welfare of non-truthful auctions such as the first price auction.

However, this means that the final bound given in (59) is the product of a subset of the best-response deviation constraints. If we actually knew what the bid distribution is in a BCE (which is the case in the econometrics task), then we wouldn't look at only these bid-oblivious upwards deviations. We would instead compute an optimal best-response bid for this given bid distribution and right the constraint that the player does not want to deviate to this tailored best-response.

We now give a concrete "extreme" example where the upper bound provided by (59) can be as large as twice the sharp bound. In fact, in this example the mean can be point identified, but the bound (59) gives a large interval.

Example. Consider the case of two bidders, where the observed maximum bid distribution is a singleton i.e. $\{b^*\}$. First we argue that in this case, we can conclude that the mean μ of the common value is equal to b^*. First, we know that the total expected utility of the bidders must be non-negative, hence the mean of the common value is at least the expected revenue of the auction, which is equal to b^*, i.e. $\mu \geq b^*$.

Now suppose that the mean was strictly larger than b^*, i.e. $b^* + \epsilon$. We will show that this yields a contradiction.

First consider the case where in the support of the BCE, there exist bid vectors of the form $(b, 5)$ or $(5, b)$ for $b < 5$. Then a player when seeing a bid of $b < 5$, he wants to deviate to bidding $b^* + \zeta$ for some $\zeta \in (0, \epsilon)$. The reason is that with a bid of b he knows he is not winning and hence he is getting zero utility, while with a bid of $b^* + \zeta$ he knows he is deterministically winning, getting a value of μ and paying $b^* + \zeta < \mu$, yielding strictly positive utility. Thus it must be that the BCE contains only one bid vector, i.e. (b^*, b^*). But under this BCE, it is obvious that the expected value conditional on winning is μ and it is also clear that one of the two players is winning with probability less than 1, i.e $1 - \delta$. Thus if this player deviates to $b^* + \zeta$, he wins deterministically and pays $b^* + \zeta$. Thus the net effect of this deviation is $\mu - b^* - \zeta - (1 - \delta)(\mu - b^*) = \delta \cdot (\mu - C) - \zeta = \delta\epsilon - \zeta$. Taking $\zeta \to 0$, yields a deviation

with positive net effect.

Thus we get a contradiction, and we conclude that the mean is point identified and $\mu = b^*$.

Now consider the bound that is derived from Equation (59) for this case of a bid distribution. First we note that for a singleton bid distribution we have that $B(q) = b^*$ for all $q \in [0, 1]$. Thus the constraint on the distribution of values implied by Equation (60) is simplified to:

$$\forall q \in [0, 1] : \frac{1}{\sqrt{q}} \int_0^q \frac{v(y)}{2\sqrt{y}} dy \leq b^* \tag{63}$$

We now give a distribution of values, i.e. a function $v(\cdot)$, that satisfies the above constraint and whose mean μ^* is at least $2 \cdot b^* - (b^*)^2/H$, where H is an externally assumed upper bound on the distribution of values. If H is large, then this gives an interval such that the ratio of the upper to the lower bound is of size 2, which is far from point identification.

Consider a value function $v(\cdot)$, as follows (two point-mass distribution):

$$v(q) = \begin{cases} H & \text{if } q \geq \theta \\ 0 & \text{o.w.} \end{cases} \tag{64}$$

Then the constraint simplifies to:

$$b^* \geq \frac{1}{\sqrt{q}} \int_{\min\{\theta, q\}}^q \frac{H}{2\sqrt{y}} dy$$
$$= \frac{1}{\sqrt{q}} H [\sqrt{y}]_{\min\{\theta, q\}}^q = H \left(1 - \sqrt{\frac{\min\{\theta, q\}}{q}} \right)$$

If $\theta \leq q$, then this constraint is definitely satisfied, since $b^* \geq 0$. Thus we need to only check the constraints for $q > \theta$. The tightest of these constraints with respect to θ, is when $q = 1$, leading to:

$$\theta \geq \left(\frac{H - b^*}{H} \right)^2 \tag{65}$$

By setting θ to be equal to the above lower bound, we get a feasible value function and this value function has mean:

$$\mu^* = H(1 - \theta) = H - \frac{(H - b^*)^2}{H} = \frac{H^2 - H^2 - (b^*)^2 + 2b^* H}{H} = 2b^* - \frac{(b^*)^2}{H}$$

67

which concludes the point of the example.   ∎

A.1   Upper bound on mean of common value

Theorem 15. *Let $R = \mathbb{E}_{b \sim D}[\max_i b_i]$ is the expected revenue of a first price auction when bids are drawn from some BCE and let $\mu = \mathbb{E}[v]$ be the expected common value. If we assume that $v \in [0, H]$ and that the inverse of the CDF of the distribution of values is continuously differentiable in $[0, H]$, then it holds that:*

$$\mu \leq \sqrt{\frac{2n}{n-1} H \cdot R} \tag{66}$$

If we also assume that the inverse of the CDF of values is L-Lipschitz (equivalently the density of values is bounded below by $1/L$), then we don't need to assume an upper bound on the distribution and we alternatively get:

$$\mu \leq \sqrt{\frac{2n}{n-1}\left(\frac{1}{n}R + L\right) \cdot R} = R\sqrt{\frac{2}{n-1}} + \sqrt{\frac{2n}{n-1}L \cdot R} \tag{67}$$

Proof. By the main theorem of Bergemann et al. [2017] we have that if F is the CDF of common values and H is the CDF of the maximum bid distribution at a BCE, while $v(\cdot) = F^{-1}(\cdot)$ and $B(\cdot) = H^{-1}(\cdot)$ are the quantile functions of these distributions, then for any $q \in [0, 1]$:

$$B(q) \geq v(q) - \frac{1}{q^\alpha} \int_0^{v(q)} (F(x))^\alpha dx \tag{68}$$

with $\alpha = \frac{n-1}{n}$. Let $I(q)$ denote the right hand side.

We first do a change of variables in the integral: let $x = v(y)$. Then $dx = v'(y)dy$ and since the integral ranges from $x \in [0, v(q)]$, in the new variable the integral ranges from $y \in [0, q]$. Hence:

$$I(q) = v(q) - \frac{1}{q^\alpha} \int_0^q (F(v(y)))^\alpha v'(y)dy = v(q) - \frac{1}{z^\alpha} \int_0^q y^\alpha v'(y)dy \tag{69}$$

68

Applying integration by parts on the integral we have:

$$I(q) = v(q) - \frac{1}{q^\alpha}\left([y^\alpha v(y)]_0^q - \int_0^q \alpha y^{\alpha-1} v(y) dy\right)$$

$$= v(q) - v(q) + \frac{1}{q^\alpha}\int_0^q \alpha y^{\alpha-1} v(y) dy$$

$$= \frac{1}{q^\alpha}\int_0^q \alpha y^{\alpha-1} v(y) dy$$

Now observe that, $\mu = \int_0^1 v(q) dq$. This can be easily verified pictorially, but also algebraically as follows:

$$\mu = \int_0^H 1 - F(x) dx = \int_0^1 (1 - F(v(q))) v'(q) dq = [(1-q)v(q)]_0^1 + \int_0^1 v(q) dq = \int_0^1 v(q) dq \tag{70}$$

Similarly, $R = \int_0^1 B(q) dq$.

Thus integrating the inequality $B(q) \geq I(q)$ over q we get:

$$R \geq \int_0^1 \frac{1}{q^\alpha}\int_0^q \alpha y^{\alpha-1} v(y) dy dq \tag{71}$$

Exchanging the integration order, we get:

$$R \geq \int_0^1 \int_y^1 \frac{1}{q^\alpha}\alpha y^{\alpha-1} v(y) dq dy = \int_0^1 \alpha y^{\alpha-1} v(y) \int_y^1 \frac{1}{q^\alpha} dq dy \tag{72}$$

$$= \int_0^1 \alpha y^{\alpha-1} v(y) \left[\frac{q^{1-\alpha}}{1-\alpha}\right]_y^1 dy \tag{73}$$

$$= \int_0^1 \frac{\alpha}{1-\alpha} y^{\alpha-1} v(y) \left(1 - y^{1-\alpha}\right) dy \tag{74}$$

$$= \int_0^1 \frac{\alpha}{1-\alpha} v(y) \left(y^{\alpha-1} - 1\right) dy \tag{75}$$

Thus by re-arranging we have that:

$$\int_0^1 v(y) \left(y^{\alpha-1} - 1\right) dy \leq \frac{1-\alpha}{\alpha} R \tag{76}$$

Therefore if we are given as fixed the revenue of the auction R, then the mean of the common value distribution can be at most the solution to the following optimization

69

program over the set of all possible quantile functions $v(\cdot)$:

$$\max_{v(\cdot)\in[0,H]} \int_0^1 v(y)dy \tag{77}$$

$$s.t. \int_0^1 v(y)\left(y^{\alpha-1}-1\right)dy \leq \frac{1-\alpha}{\alpha}R \tag{78}$$

Now consider the function $w(y) = y^{\alpha-1} - 1$. This is a monotone decreasing and non-negative function of y, starting from ∞ at $y = 0$ and ending at 0 at $y = 1$.

The goal of the above linear program is to push as much possible value in $v(y)$ as possible, while satisfying the constraint. Observe that a quantity $v(y)$ is multiplied by a smaller value in the above constraint than any $v(y')$ for $y' \leq y$. Thus it is easy to see that the optimal solution to the above linear program puts as much value on the high y's and sets the remainder y's to zero (another way of arguing this is considering arbitrarily small discretizations of the y space and then arguing that the solution to the linear program with respect to $v(y)$, has the above form). Thus the optimal $v(\cdot)$ takes the form:

$$v(y) = \begin{cases} 0 & \text{if } y \leq \theta \\ H & \text{o.w.} \end{cases} \tag{79}$$

for some threshold $\theta \in [0,1]$.

Thus the optimal value of the above program simplifies to:

$$\max_{\theta\in[0,1]} H(1-\theta) \tag{80}$$

$$s.t. \ H \int_\theta^1 \left(y^{\alpha-1}-1\right)dy \leq \frac{1-\alpha}{\alpha}R \tag{81}$$

Thus we want to minimize θ, such that:

$$\frac{1-\alpha}{\alpha}\frac{R}{H} \geq \int_\theta^1 \left(y^{\alpha-1}-1\right)dy = \frac{1}{\alpha} - \frac{\theta^\alpha}{\alpha} - 1 + \theta \tag{82}$$

Case of $n = 2$ For $n = 2$, i.e. $\alpha = 1/2$, we can exactly solve the latter, since it corresponds to a quadratic inequality, i.e.:

$$\frac{R}{H} \geq 1 - 2\sqrt{\theta} + \theta = (1 - \sqrt{\theta})^2 \tag{83}$$

which yields that: $\sqrt{\theta} \geq 1 - \sqrt{\frac{R}{H}}$ or equivalently, $\theta \geq 1 + \frac{R}{H} - 2\sqrt{\frac{R}{H}}$. Therefore, the highest possible mean is at most: $H(1 - \theta) = H\left(2\sqrt{\frac{R}{H}} - \frac{R}{H}\right) = 2\sqrt{R \cdot H} - R$.

General n. For arbitrary $\alpha \in [0, 1]$, we can lower bound the right hand side with a second order Taylor expansion around $\theta = 1$: i.e. for any $\theta \in [0, 1]$, for some $t \in [\theta, 1]$:

$$\theta^\alpha = 1 + (\theta - 1)\alpha + \frac{\alpha(\alpha - 1)}{2}t^{\alpha-2}(\theta - 1)^2 \leq 1 + (\theta - 1)\alpha + \frac{\alpha(\alpha - 1)}{2}(\theta - 1)^2 \quad (84)$$

(since for $\alpha \in [0, 1]$, $t \in [0, 1]$: $t^{\alpha-2} \geq 1$ and $\alpha \leq 1$) Replacing the above approximation in the constraint, gives:

$$\frac{1 - \alpha}{\alpha}\frac{R}{H} \geq \frac{1}{\alpha} - \frac{1}{\alpha} - \theta + 1 - \frac{(\alpha - 1)}{2}(\theta - 1)^2 - 1 + \theta = \frac{1 - \alpha}{2}(\theta - 1)^2 \quad (85)$$

Simplifying further, we want to minimize θ such that: $(\theta - 1)^2 \leq \frac{2}{\alpha}\frac{R}{H}$. The optimal value sets:

$$\theta = 1 - \sqrt{\frac{2}{\alpha}\frac{R}{H}} \quad (86)$$

Leading to a mean of:

$$\mu^* = H(1 - \theta) \leq \sqrt{\frac{2}{\alpha} \cdot R \cdot H} \quad (87)$$

For the second part of the theorem, we simply observe that if the function $v(\cdot)$ is L-Lipschitz, then it must be that $v(1) \leq (1 - \alpha)R + L$, thereby, we can do the exact same analysis as above, but with $H = (1 - \alpha)R + L$. The upper bound on $v(1)$, comes from the fact, that if $v(\cdot)$ is lipschitz, then $v(y) \geq v(1) - L(1 - y) \geq v(1) - L$. Thus we have:

$$\frac{1 - \alpha}{\alpha}R \geq \int_0^1 v(y)\left(y^{\alpha-1} - 1\right) dy \geq (v(1) - L)\int_0^1 \left(y^{\alpha-1} - 1\right) dy = \frac{v(1) - L}{\alpha} \quad (88)$$

Re-arranging gives the property and concludes the proof of the theorem. ∎

B Omitted Proofs from Section 5

B.1 Proof of Theorem 10

Proof. We will show that with probability at least $1 - \delta/2$: $L^N(\sigma^N) - \epsilon^N \leq L$. The theorem follows by also showing that $U^N(\sigma^N) + \epsilon^N \geq U$ with probability at least

$1 - \delta/2$ and then using a union bound of the bad events. The second inequality, follows along identical lines by simply replacing $m(v)$ with $-m(v)$. So it suffices to show the first inequality.

We remind the reader the definitions of the two quantities:

$$L = \min_{x:\ \max_{j \in M} F_j(x;\phi) \leq 0} \sum_{v \in V} m(v) \cdot \sum_{b \in S} \phi(\mathbf{b}) \cdot x(v|\mathbf{b}) \tag{89}$$

$$L^N(\sigma^N) = \min_{x:\ \max_{j \in M} F_j^N(x) \leq \sigma^N} \sum_{v \in V} m(v) \cdot \sum_{b \in S} \phi_N(\mathbf{b}) \cdot x(v|\mathbf{b}) \tag{90}$$

Let x^* be the optimal solution to the population LP of Equation (89). We will argue that w.h.p. for the choice of σ^N it remains a solution of the finite sample LP of Equation (90) and the value of the finite sample LP under x^* is at least the value of the population LP plus ϵ^N.

Observe that $F_j^N(x^*)$ is the sum of N i.i.d. random variables bounded in $[-H, H]$ with mean $F_j(x^*; \phi)$. By Hoeffding's inequality with probability $1 - \kappa$:

$$F_j^N(x^*) \leq F_j(x^*; \phi) + 2H\sqrt{\frac{\log(1/\kappa)}{N}} \leq 2H\sqrt{\frac{\log(1/\kappa)}{N}} \tag{91}$$

where the second inequality follows by feasibility of x^* for the LP of Equation (89). By a union bound over all $j \in M$ and since $|M| = n|B|^2$ we get that with probability at least $1 - \kappa n|B|^2$:

$$\max_{j \in M} F_j(x^*; \phi) \leq 2H\sqrt{\frac{\log(1/\kappa)}{N}} \tag{92}$$

For $\kappa = \frac{\delta}{4n|B|^2}$ and $\sigma^N = 2H\sqrt{\frac{\log(4n|B|^2/\delta)}{N}}$ we get that with probability $1 - \frac{\delta}{4}$: $\max_{j \in M} F_j^N(x^*) \leq \sigma^N$ and thereby x^* is feasible for the finite sample LP.

Finally, the value of the finite sample solution can be written as

$$Q^N(x) = \frac{1}{N} \sum_{t=1}^{N} q(x; \omega_t) \tag{93}$$

with:

$$q(x; \omega_t) = \sum_{v \in V} m(v) \sum_{\mathbf{b} \in S} \omega_t(\mathbf{b}) \cdot x(v|\mathbf{b}) \tag{94}$$

Thus it is also the sum of N i.i.d. random variables bounded in $[-H, H]$,[20] and with mean $Q(x; \phi) = \sum_{v \in V} m(v) \cdot \sum_{b \in S} \phi(\mathbf{b}) \cdot x(v|\mathbf{b})$. Hence, by Hoeffding's inequality, with probability at least $1 - \delta/4$:

$$Q_j^N(x^*) \leq Q_j(x^*; \phi) + 2H\sqrt{\frac{\log(4/\delta)}{N}} = L + \epsilon^N \tag{95}$$

By a union bound we get that with probability at least $1 - \delta/2$, x^* is feasible for the finite sample LP and achieves value $Q_j^N(x^*) \leq L + \epsilon^N$. In that case, by the definition of $L^N(\sigma^N)$, we get: $L^N(\sigma^N) \leq Q_j^N(x^*) \leq L + \epsilon^N$ and the theorem follows. [21] ■

B.2   Proof of Theorem 11

Proof. To show the statement we need to show that with probability $1 - \delta$, for all $\theta \in \Theta_I$ it must be that $\theta \in \Theta_I^N(\sigma^N)$. Equivalently, if $\min_x \max_{j \in M} F_j(x; \theta, \phi) \leq 0$ then $\min_x \max_{j \in M} F_j^N(x; \theta) \leq \sigma^N$. The latter follows along similar lines as in the proof of Theorem 10. Let $\theta \in \Theta_I$ and let x^* be the solution to the population problem. Then by Hoeffding's inequality and a union bound over M,[22] for the chosen σ^N we have that with probability at least $1 - \frac{\delta}{|\Theta|}$: $\max_{j \in M} F_j^N(x^*; \theta) \leq \sigma^N$. Hence, the latter inequality also holds for the minimum over x and thereby $\theta \in \Theta_I^N(\sigma^N)$. By a union bound the latter holds uniformly over θ with probability at least $1 - \delta$. ■

B.3   Proof of Theorem 12

Proof. To show the statement we need to show that with probability $1 - \delta$, for all $\theta \in \Theta_I$ it must be that $\theta \in \hat{\Theta}_I^N(\lambda, \sigma^N)$. Let $\theta \in \Theta_I$ and let x^* be the solution to the population problem, i.e. $x^* \in \arg\min_x \max_{i \in M} F_i(x; \theta, \phi)$.

---

[20]Since $m(v) \in [-H, H]$ and $x(\cdot|\mathbf{b}) \in \Delta(V)$.

[21]We could have used an empirical Bernstein inequality instead of Hoeffding's inequality and replaced the H in the quantities σ^N by $\sqrt{\max_j \text{Var}_N(F_j^N(x^*))}$, where Var_N is the empirical variance, at the expense of adding a lower order term of $O\left(\frac{H}{N}\right)$ (see e.g. Maurer and Pontil [2009], Peel, Anthoine, and Ralaivola [2010]). Similarly, for ϵ^N. However, the latter requires knowledge of x^* and taking a supremum over x^* in the latter seems to be as conservative as a Hoeffding bound.

[22]Now M also contains the density consistency constraints and thereby $|M| = n|B|^2 + |V|$.

By the empirical Bernstein bound (Theorem 4 of Maurer and Pontil [2009]), we have that with probability at least $1 - \kappa$:

$$F_j^N(x^*; \theta) \leq F_j(x^*; \theta, \phi) + \sqrt{\frac{2\mathrm{Var}_N(f_j(x^*; \theta, \omega))\log(2/\kappa)}{N}} + \frac{14H\log(2/\kappa)}{3(N-1)} \quad (96)$$

Since x^* is feasible $F_j(x^*; \theta, \phi) \leq 0$ and thereby with probability $1 - \kappa$:

$$F_j^N(x^*; \theta) - \sqrt{\frac{2\mathrm{Var}_N(f_j(x^*; \theta, \omega))\log(2/\kappa)}{N}} \leq \frac{14H\log(2/\kappa)}{3(N-1)} \quad (97)$$

Hence, also:

$$F_j^N(x^*; \theta) - \sqrt{\frac{2\mathrm{Var}_N(f_j(x^*; \theta, \omega))\log(2/\kappa)}{N}} \leq \frac{14H\log(2/\kappa)}{3(N-1)} \quad (98)$$

By the union bound the latter happens for all $j \in M$ with probability $1 - \kappa|M|$. Thus setting $\kappa = \frac{\delta}{|\Theta||M|}$, we have that with probability $1 - \frac{\delta}{|\Theta|}$:

$$\max_{j \in M} F_j^N(x^*; \theta) - \sqrt{\frac{2\mathrm{Var}_N(f_j(x^*; \theta, \omega))\log(2|\Theta||M|/\delta)}{N}} \leq \frac{14H\log(2|\Theta||M|/\delta)}{3(N-1)} \quad (99)$$

The latter then also holds for the minimum over x of the LHS:

$$\min_x \max_{j \in M} F_j^N(x; \theta) - \sqrt{\frac{2\mathrm{Var}_N(f_j(x; \theta, \omega))\log(2|\Theta||M|/\delta)}{N}} \leq \frac{14H\log(2|\Theta||M|/\delta)}{3(N-1)} \quad (100)$$

Hence, $\theta \in \hat{\Theta}_I^N(\lambda, \sigma^N)$ for the λ and σ^N stated in the Theorem. The Theorem then follows by a union bound over Θ. ∎

C   Further Empirical Results

C.1   Identified Sets of Remaining Keywords in Ad Auctions
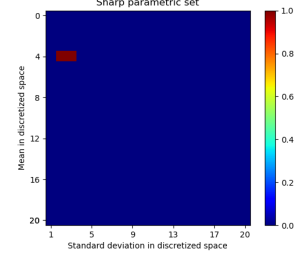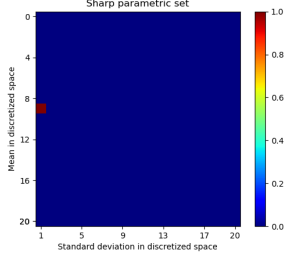

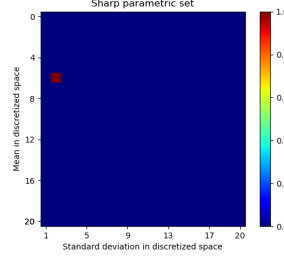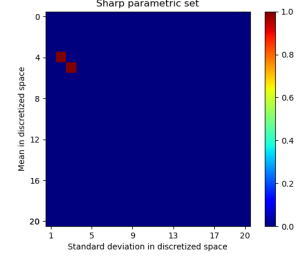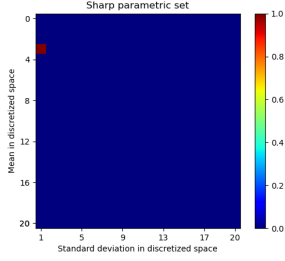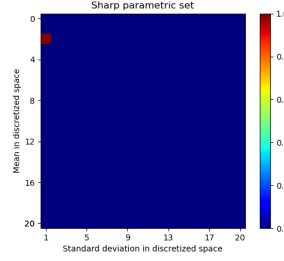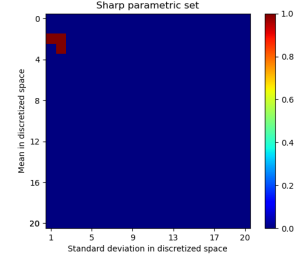|		
(a) 1st bidder, keyword 4	(b) 2nd bidder, keyword 4	(c) 3rd bidder, keyword 4
(d) 1st bidder, keyword 5	(e) 2nd bidder, keyword 5	(f) 3rd bidder, keyword 5
(g) 1st bidder, keyword 6	(h) 2nd bidder, keyword 6	(i) 3rd bidder, keyword 6
(j) 1st bidder, keyword 7	(k) 2nd bidder, keyword 7	(l) 3rd bidder, keyword 7
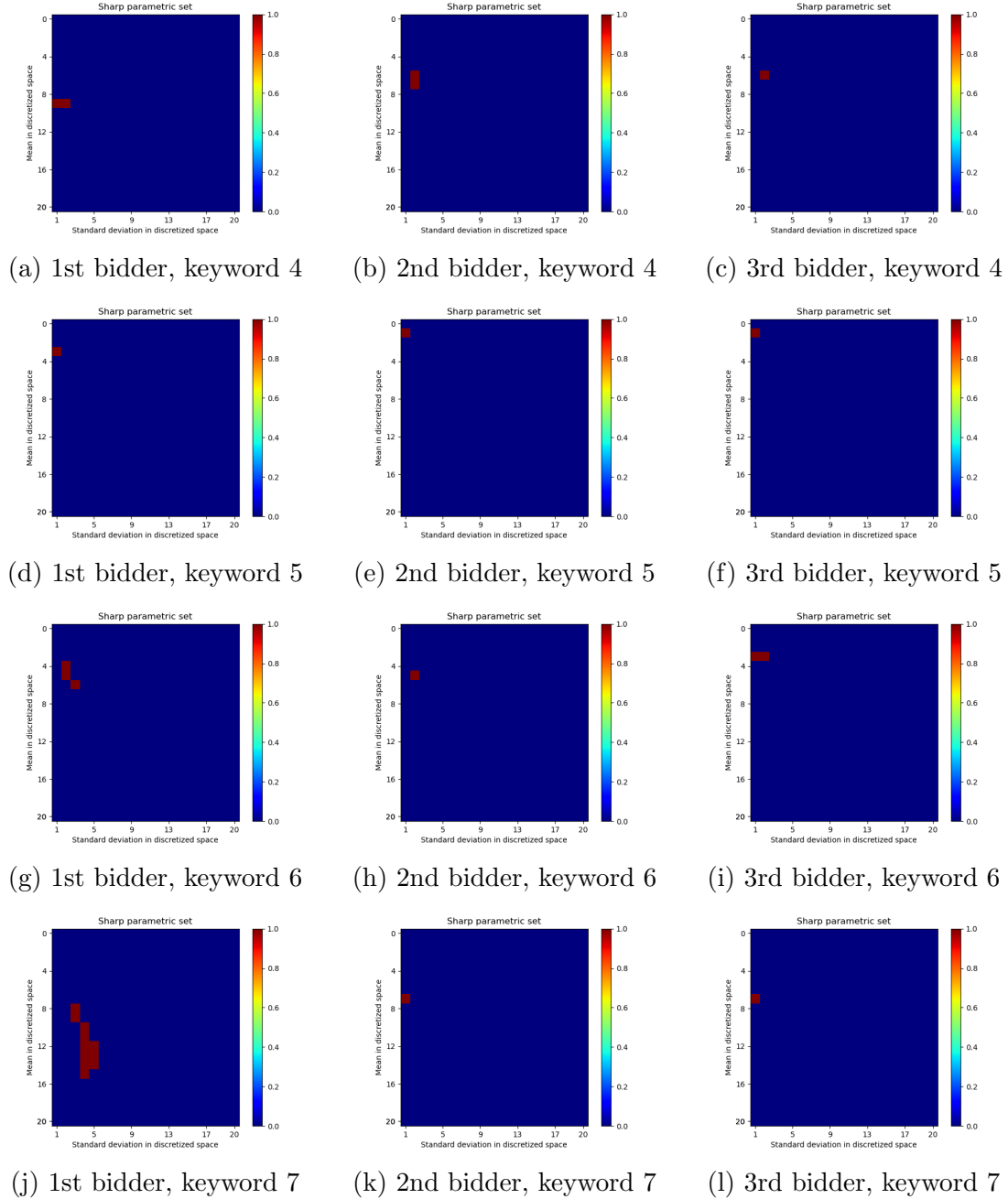
Figure 19: Identified sets with 95% confidence, of the mean and standard deviation of the distribution of values for each of the top three bidders of each keyword, for a sample of three of the 13 keywords in our analysis, based on a truncated normal parametric distribution assumption.
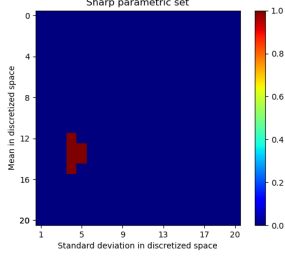
(a) 1st bidder, keyword 8   (b) 2nd bidder, keyword 8   (c) 3rd bidder, keyword 8
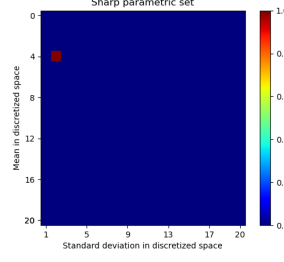
(d) 1st bidder, keyword 9   (e) 2nd bidder, keyword 9   (f) 3rd bidder, keyword 9
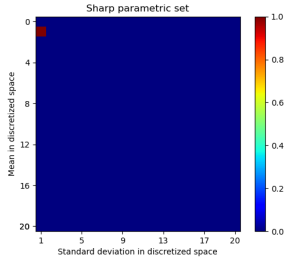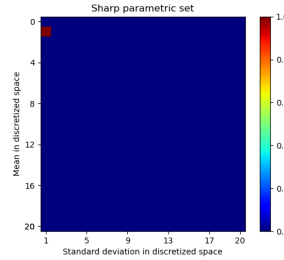
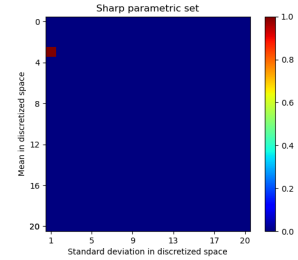(g) 1st bidder, keyword 10   (h) 2nd bidder, keyword 10   (i) 3rd bidder, keyword 10
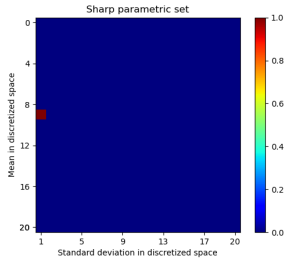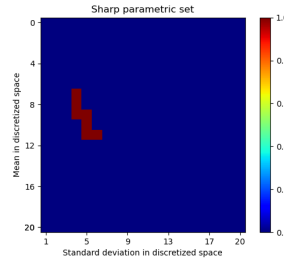
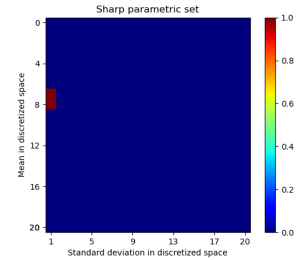(j) 1st bidder, keyword 11   (k) 2nd bidder, keyword 11   (l) 3rd bidder, keyword 11

(m) 1st bidder, keyword 12   (n) 2nd bidder, keyword 12   (o) 3rd bidder, keyword 12

Figure 20: Identified sets with 95% confidence, of the mean and standard deviation of the distribution of values for each of the top three bidders of each keyword, for a sample of three of the 13 keywords in our analysis, based on a truncated normal parametric distribution assumption.