An Information–Theoretic Approach to Partially Identified Auction Models[*]

Sung Jae Jun[†] and Joris Pinkse[‡]

Department of Economics
The Pennsylvania State University

January 15, 2019[§]

We consider the question of how to determine the optimal auction mechanism based on data from English auctions with symmetric bidders, independent private values, and exogenous entry, in which bidders can only select bids from a prespecified finite set. A consequence of the format of the observed auctions is that the bidders' value distribution is only partially identified. Our assumptions about the observed data resemble those in Haile and Tamer (2003, HT) who obtain bounds on the value distributions which, under additional assumptions (including a strict pseudoconcavity assumption), can be used to form bounds on the optimal reserve price in second price auctions. Most of our paper is dedicated to the case in which the seller intends to use a second price auction, which means that the problem becomes an optimal reserve price r^* selection problem. Since the seller has to pick a single reserve price in a given auction, bounds on the optimal reserve price are of limited use when each point in the bounds of the optimal reserve price has different revenue implications. Further, it requires strong assumptions on the shape of the value distribution to get bounds on r^* from bounds on the value distribution.

We formulate the seller's decision problem as a maxmin problem subject to a lower bound $\mathcal{E}^*$ on the entropy of the value distribution. Any choice of $\mathcal{E}^*$ leads to a single (but different) value distribution in the identified set: for $\mathcal{E}^* = -\infty$ one obtains pure maxmin (Aryal and Kim, 2013, AK) whereas choosing $\mathcal{E}^*$ maximally results in the *maximum entropy (ME)* solution from information theory. The choice of $\mathcal{E}^*$ thus reflects a seller's trade–off between *ambiguity aversion* and a desire to rule out 'unusual' distributions as measured by the entropy.

Most of our paper focuses on the ME solution. We derive asymptotic properties of our estimators for the ME reserve price and expected revenue, from which we develop an inference procedure. Our methodology addresses the more general problem of an estimator that optimizes a deterministic objective function subject to both equality and inequality moment conditions that are estimated.

We further explore the possibility of using the ME value distribution to implement Myerson's optimal mechanism, something that would be difficult (if not impossible) to achieve using existing methods. However, we find that the gains from doing so are unlikely to be worthwhile in practice.

Keywords: English auctions, optimal mechanism, partial identification, maximum entropy, inequality constraints, and nonparametric inference.

JEL: C01; C10; C57.

---

[*] We thank Andres Aradillas–Lopez, Gaurab Aryal, Michael Gechter, Ed Green, Nima Haghpanah, Phil Haile, Marc Henry, Kei Hirano, Dong–Hyuk Kim, Vijay Krishna, Peter Newberry, Ran Shorrer, Guofu Tan, and participants of the SVEC conference in Vancouver, the conference in memory of Art Shneyerov in Paris, and departmental seminars at Penn State and the University of British Columbia for their helpful comments.

[†] 619 Kern Bldg, University Park, PA 16802, sjun@psu.edu

[‡] joris@psu.edu

[§] The first circulated version was on May, 2017.

1. Idea and context

We consider English auctions with exogenous entry and symmetric bidders under the independent private values (IPV) paradigm in which the number of potential bidders n and some bids are observed. Since this is an English auction, there is no one–to–one mapping between bids made by a given bidder and his valuation as there would be in a symmetric sealed bid IPV auction without a reserve price. Indeed, bidders can continue to submit bids after they have submitted one or may not submit a bid at all because the current bid level already exceeds their private value.

We assume that bidders can only select bids from a prespecified finite set. We focus on the simplest case in which bids belong to $\{\Delta, 2\Delta, \ldots\}$ with $\Delta > 0$, but this is not essential. Haile and Tamer (2003, HT) consider the possibility that there is a minimum positive bid increment, but HT's results apply in our environment, also. Like HT, we allow for jump bidding. From here on, we will discuss HT's methodology in the context of our environment. If $\Delta > 0$ then the bidders' value distribution is only partially identified.[1] Indeed, HT obtain bounds on the unknown distribution function $\mathcal{F}$ of private values which, both in HT and here, is assumed to be continuous. HT then add a strict pseudoconcavity assumption on the *pseudoprofit* function $\mathcal{Q}(r) = r\{1 - \mathcal{F}(r)\}$ to obtain bounds on the optimal reserve price.[2]

In HT, like here, the assumptions needed for the optimal reserve price in an English auction to define an optimal auction mechanism go beyond pseudoconcavity and they are stronger than those needed for identification of bounds on the value distribution function. Indeed, like HT, we think about the problem of setting an optimal reserve price in a hypothetical second price auction with $\Delta = 0$. The rationale for this is, as HT have shown, that an English auction with a reserve price and no minimum bid increment is revenue–equivalent to a second price sealed bid auction with the same reserve price under the behavioral assumptions made in HT, as long as the English auction can be represented by a "feasible auction mechanism."[3] However, the revenue equivalence theorem (Myerson, 1981) does *not* imply that a second price auction is optimal. We revisit these issues in sections 2 and 4.

Although there is precedence for HT's pseudoconcavity assumption in the literature, there are

[1] Athey and Haile (2002) contains identification results on English auctions.

[2] The pseudoprofit function is the seller's profit as a function of the reserve price if there is only one bidder. We set the seller's value to zero for presentational simplicity.

[3] A feasible auction mechanism requires a Nash equilibrium in the bidding game (Myerson, 1981, p62). The case $\Delta = 0$, both here and in HT, rules out the bulk of situations giving rise to jump bidding in equilibrium. For instance, Avery (1998) requires affiliation, whereas we follow HT in assuming independence. Further, Daniel and Hirshleifer (1998) is mostly about a scenario in which bidding is costly. Daniel and Hirshleifer (1998) do provide an example of a Nash equilibrium with zero costs, no minimum bid increment, and symmetric bidders in which there exists an equilibrium with jump bidding, but it is both fragile and does not satisfy HT's requirement that bidders will not let rivals win at a price that they are willing to beat. In fact, we are not aware of an example in which there are jump bids in equilibrium with symmetric bidders, exogenous entry, independent private values, no minimum bid increment, and in which HT's behavioral assumptions are satisfied.

many plausible value distributions that do not satisfy it and there is no clear path to usable bounds on the optimal reserve price in a second price auction from bounds on $\mathscr{F}$ without it: example 1 in section 2 and figures 6 and 7 in section 5 contain such distributions.

Further, bounds on the optimal reserve price are interesting, but they are of limited use to a seller faced with the problem of choosing a reserve price: she will still have to pick a single number. Indeed, as we show in section 5, if the number of bidders is small then expected revenue can vary substantially with the choice of reserve price, even if one were only to consider reserve prices that belong to the HT identified set. The question of interest in this paper, then, is how to pick a reserve price if the value distribution $\mathscr{F}$ is only partially identified.

This question already has an answer in Aryal and Kim (2013, AK) in the somewhat different case in which only the winning bid is ever observed. Their idea can be summarized as follows. AK assume that the winning bid is no greater than the highest value. So, the n–th power of $\mathscr{F}$ is bounded above by the distribution function of the winning bid. They further assume that the seller is ambiguity–averse (Gilboa and Schmeidler, 1989), but *not* risk–averse, which means that the seller's objective is to maximize minimum expected revenue over all value distributions.[4] AK's results show the following: for two candidate value distributions $\mathscr{F}_1$ and $\mathscr{F}_2$, if $\mathscr{F}_2$ is first order stochastically dominated by $\mathscr{F}_1$ then the expected revenue associated with $\mathscr{F}_1$ is no less than that with $\mathscr{F}_2$ for any reserve price. Therefore, the lower bound to expected revenue obtains when the value distribution is set to be equal to its upper bound. In other words, the value distribution $\mathscr{F}_{\text{AK}}$ that corresponds to the smallest expected revenue is such that $\mathscr{F}_{\text{AK}}^n$ equals the distribution of the winning bid. Therefore, the seller can choose a reserve price as if $\mathscr{F}_{\text{AK}}$ were the true value distribution.

The methodology we propose can be used in the environment analyzed in AK, but we choose to develop our methodology in an environment similar to HT's world which is methodologically more interesting for our approach: our methodology uses both the upper and the lower bounds of the identified set but in AK's environment the lower bound is always trivial.[5] AK's methodology can be extended to HT's case, also. Indeed, one can simply look at the worst possible value distribution that is consistent with HT's bounds, i.e. the HT upper bound of the value distribution. As far as we know, this has not been done, but it is not difficult. From here on, we will take that leap and discuss AK's approach in the context of HT's conditions only.

A more general way of addressing the reserve price selection problem is to express it in terms of a decision–theoretic framework by assuming that a seller chooses the reserve price to maximize minimum expected revenue where the minimum is taken over all value functions $\mathscr{F}$ that belong to the HT identified set and whose (relative) entropy is no less than some lower bound $\mathscr{E}^*$ specified by the seller. With this setup, the choice of $\mathscr{E}^*$ reflects the seller's attitude about ambiguity and

[4] She is not risk–averse, because she is still trying to maximize expected revenue instead of the expectation of a strictly concave function of revenue. In AK, the seller is ambiguity–averse because she does not like the ambiguity resulting from the fact that the value distribution is unknown.

[5] Further, the upper bound of the identified set in AK's environment is greater than that in HT's.

plausibility: she balances an aversion to ambiguity with a desire to rule out distributions she considers implausible as measured by the entropy. It allows her to provide a guess of the value distribution known as a *reference distribution* (which we take to be a uniform) and only permit deviations within a certain distance from the reference distribution, where the distance is measured by (minus) the (relative) entropy, i.e. the Kullback–Leibler divergence criterion. There is hence a *shrinkage* aspect to this approach in that (absent point identification in the original problem) the choice of distribution function is shrunk towards the reference distribution. The value of $\mathcal{E}^*$ determines the tradeoff between ambiguity aversion and plausibility: for $\mathcal{E}^* = -\infty$ one obtains AK's pure maxmin approach and with $\mathcal{E}^*$ chosen maximally one obtains *maximum entropy,* a method from the *information theory* literature. A discussion of this entropy–constrained decisions–theoretic approach can be found in section 6. Most of our paper, however, makes the assumption that the seller chooses $\mathcal{E}^*$ maximally, i.e. uses maximum entropy.

The main question with maximum entropy is how to pick the reference distribution. As noted, one possibility is to ask the seller for a guess of what the true value distribution looks like. This need can be compared to the need in Bayesian decision theory to specify a prior distribution, but the demand here is less onerous: Bayesian decision theory would require the seller to specify a prior over the set of value distributions, which is probably an unreasonable amount of information to elicit from a seller.

A second possibility is to start with a more restricted model, in which there is point identification, e.g. by adding behavioral assumptions. One can then estimate the restricted model which produces an estimated value distribution which can be used as a reference distribution in our procedure, albeit that the inference methods proposed in this paper would need to be extended to accommodate sampling errors in the reference distribution. This approach has a passing resemblance to the ideas of *robust control* advocated by Hansen and Sargent (2008) in a quite different context.

There are two reasons why we set the reference distribution to a uniform. First, it simplifies notation noting that the extension to other reference distributions is tedious but not difficult, albeit that the maximum entropy solution will then not be piecewise linear, as we will show it is here.[6] Second, it is a natural choice in the information theory framework since it means that we would be maximizing Shannon's entropy subject to the identified bounds of the value distribution. Note that the uniform distribution reflects our ex ante 'ignorance' about what the value distribution should look like, which is in line with the rationale of *Occam's razor*. Below we discuss the relationship between decision theory and the maximum entropy principle.

In addition to being a limit case of the entropy–constrained decision problem presented above, maximum entropy itself also has a decision–theoretic interpretation. Indeed, Topsøe (1979), Harremoës and Topsøe (2001), and Grünwald and Dawid (2004) inter alia provide a justification of the

[6] The maximum entropy distribution function will be piecewise linear; the corresponding density function hence piecewise constant.

maximum entropy principle from the usual minmax perspective of decision theory. Grünwald and Dawid (2004) show that maximizing entropy and minimizing worst–case expected loss are each other's dual in a statistical game in which a decision maker specifies *a* distribution and nature reveals values from an unknown distribution, albeit that their results do not cover the case of inequality restrictions and it is outside the scope of this paper to generalize their theory to the case of inequality restrictions.

With our procedure it is straightforward to add any additional information or insights the seller might have, or indeed information or insights that can be inferred from the seller's actions, such as the reserve price she chose to set in the observed auctions.[7]

Our maximum entropy method yields a unique solution $\mathcal{F}^*$ that satisfies the HT distribution function bounds and is piecewise linear.[8] The maximum entropy value distribution $\mathcal{F}^*$ then implies an optimal reserve price r^*, which is necessarily between the HT bounds whenever HT's strict pseudoconcavity assumption is satisfied. However, the maximum entropy distribution itself need *not* satisfy the pseudoconcavity assumption.

Neither AK's methodology nor our maximum entropy approach requires the pseudoconcavity assumption to select a reserve price. However, absent strict pseudoconcavity, Myerson (1981)'s regular case assumptions are violated and a second price auction and Myerson's optimal auction may no longer be revenue–equivalent.

Having the maximum entropy value distribution in hand, one can use it to implement Myerson's optimal auction mechanism instead of simply choosing a reserve price in a second price auction,[9] which is not possible with either HT or AK. Indeed, we discuss optimal auctions in section 4. However, as noted in section 4, the gain from doing this is typically small and Myerson's mechanism is likely to be too cumbersome to implement in a real world environment. Hence, we focus on second price auctions afterwards.

Aside from the philosophical differences between AK's approach and ours, it is evident that AK's approach leads to simpler estimation and inference procedures since one only needs to estimate a distribution of observables (the winning bids) nonparametrically. On the other hand, the reason that our approach leads to more complicated estimation and inference procedures is, in part, that we use more information. Indeed, HT use multiple distribution function bounds and pick the tightest ones.[10] With AK's approach (in our environment) one would, as noted, only use the HT upper bound on the

[7] For instance, if $\Delta = 0$ and $\mathcal{Q}$ is strictly pseudoconcave then r^* satisfying $\mathcal{Q}'(r^*) = 0$ would be the unique optimal reserve price. So, when the reserve price r^* the seller chose to set is observed, imposing $\mathcal{Q}'(r^*) = 0$ in the maximum entropy optimization problem could be a reasonable, albeit imperfect, choice. We thank Peter Newberry for this suggestion.

[8] If the seller had additional information that was added as restrictions then the ME solution might not be piecewise linear.

[9] Myerson assumes continuity of the value density function, but his results extend to our discontinuous maximum entropy value density.

[10] These are essentially 'intersection bounds.'

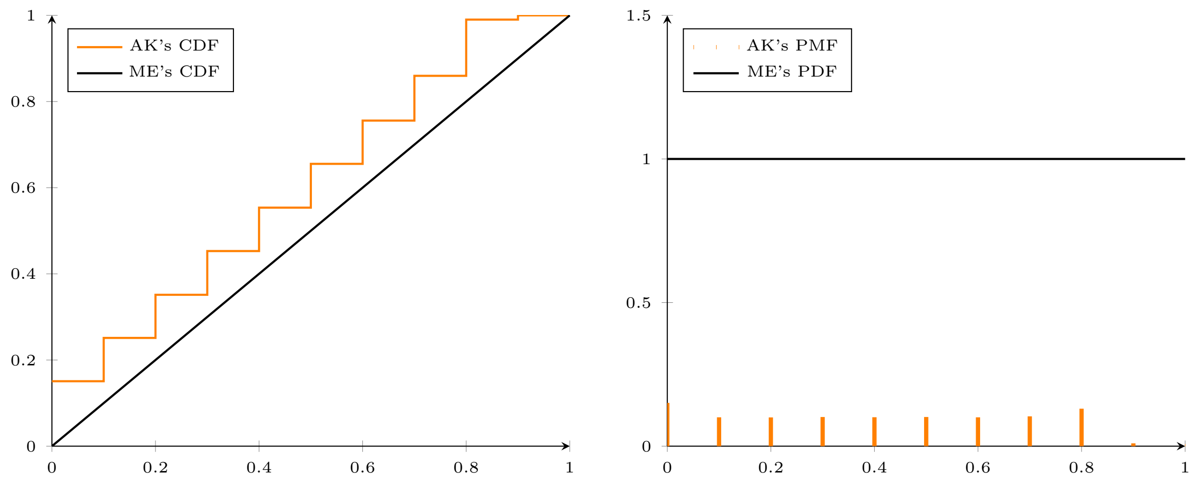


Figure 1: Comparison of AK's and ME's choice of the value distribution when $n = 2$, $\Delta = 0.1$, and the true value distribution is uniform. The left panel shows the distribution functions and the right panel shows the probability mass function of AK and the probability density function of ME.

distribution function. Maximum entropy uses the HT upper and lower bounds, plus an information criterion. Indeed, maximum entropy entails a constrained optimization problem, which leads to interesting inference problems as will become apparent below. Further, the entropy–maximizing value distribution is continuous, whereas the value distribution used for AK's reserve price is discrete, or perhaps more precisely: the AK value distribution results from the discrete limit of a sequence of continuous value distributions: the value distribution must be continuous for the bounds in HT to apply. Figure 1 shows an example, in which the true value distribution is uniform.

Another difference between AK's approach and ours arises when the reserve price in the observed auctions is binding: the AK approach *never* results in an optimal reserve price that is less than what is used in the data in contrast to our approach.[11] If there are many auctions in the data without positive bids then AK's optimal reserve price is the same as the reserve price observed in the data,[12] which we think is extreme. In contrast, the maximum entropy optimal reserve price can be as small as half the reserve price used in the data. We discuss this issue in more detail in section 2.

But neither method generally dominates the other or, put differently, either method is better than the other depending on which property is deemed preferable. If the seller is ambiguity–averse but not risk–averse then AK is the method of choice. In other casaes, the maximum entropy has a lot going for it. Either method can produce a higher expected revenue under different circumstances. Indeed, these are not the only two possibilities. In other work (Jun and Pinkse, 2016) and a different problem, we pursue the possibility of imposing a (Dirichlet process) prior on the class of distribution

[11] We thank Andres Aradillas–Lopez and Vijay Krishna for suggesting that we look into the effect of the reserve price used in the data on results.

[12] We are using a single reserve price here, but similar issues arise when the reserve price in the data varies by auction.

functions, but that approach comes with unpleasantries: it is painful to obtain a solution and it requires the arbitrary choice of input parameters. Since different reserve prices are optimal for different value distributions (consistent with the data), which method is better is a matter of philosophy, not econometrics. In section 5 we explore some differences between AK's approach and ours in a simulation study.

For our simulations we consider both 'regular' and 'irregular' cases, i.e. cases in which HT's strict pseudoconcavity assumption is satisfied and cases in which it is violated. In section 5 we simulate the 'population.' In the regular case our results are not surprising: HT's bounds for the optimal reserve price were well–defined but they are too wide to help the seller select a reserve price. Both AK's solution and ours are always contained in the HT bounds in the regular case. For the irregular case, we consider two different scenarios: one where the value distribution is regular and unimodal but the function $\mathcal{Q}$ has a flat area, and the other in which the value distribution is bimodal and $\mathcal{Q}$ has two distinct local maxima.

HT's results do not apply in the irregular case, but if one nevertheless follows their procedure here one obtains uninformative bounds or nonconvex identified sets for the optimal reserve price, depending on the shape of the $\mathcal{Q}$ function. Both AK's approach and ours produced a point decision, but neither dominated the other in terms of true expected revenue, which is in line with the theory. Again, the choice is between insurance against the worst case expected revenue (AK) and the most 'reasonable' value distribution as defined by the entropy criterion (JP).

The discussion has thus far focused on the case in which the distributions of observables are known, in which case $\mathcal{F}^*$ and r^* can be determined. In practice, however, $\mathcal{F}^*$, r^*, and the corresponding expected revenue would have to be estimated. Addressing issues resulting from estimation error can be relevant to the seller, also. She may, for instance, only wish to deviate from established practice if the ME reserve price is significantly different from the one she currently uses, or indeed if the corresponding expected revenue is significantly higher. For the purpose of inference, we assume that the HT bounds can be $\sqrt{N}$–consistently estimated, and we express the distribution of the objects of interest in terms of the distribution of the estimators of the bounds, where N is the number of auctions observed; all of HT's distribution function bounds are transformations of distribution functions of observables. We start with the estimation of the ME value *density* function f^*, which is piecewise constant since $\mathcal{F}^*$ is piecewise linear.

We obtain an estimator of f^* that is at worst $\sqrt{N}$–consistent: if no constraints are binding in the 'population' then convergence rate is arbitrarily fast, otherwise the rate is $\sqrt{N}$. However, even under $\sqrt{N}$–convergence asymptotic normality does not obtain even if the bound estimates were asymptotically normal since the limit distribution depends on which constraints are binding in the population.[13] We propose a constraint selection procedure which resembles the moment

[13] The bounds estimators are not generally asymptotically normal, either, because they are minima of several possibly asymptotically normal quantities.

selection procedure of Andrews and Soares (2010, AS) inter multa alia, but differs in three important respects: *(a)* in AS uniform inference is an important objective but our goal is to derive an expression for the limit distribution of the estimator of f^*; *(b)* AS is intended for the case of set identified parameters whereas f^* is unique; *(c)* in AS moment equalities and inequalities are all information available, whereas here we have an optimization problem with a known concave objective function with estimated inequality constraints. Like AS, our constraint selection procedure requires a sample–size–dependent input parameter. We pair the constraint selection procedure with a simulation method to obtain a procedure that is (pointwise) asymptotically similar for some value distributions $\mathcal{F}^*$ and conservative for others. There is nothing that makes our estimation problem unique to auction environments or indeed to the maximum entropy problem, so our new theoretical results should have wider application. For example, in a game–theoretic discrete choice model where moment inequalities are available to partially identify payoff parameters, our inference methods can be used to construct a confidence interval for the payoff for each player when the only information available is that the payoff parameters satisfy the moment inequalities. For a nonzero length identified set, this problem translates into two separate constrained optimization problems: one for the minimum and one for the maximum.

We further obtain results for the optimal reserve price and the corresponding expected revenue. Since the expected revenue function corresponding to the ME value distribution $\mathcal{F}^*$ need not be strictly pseudoconcave, the ME reserve price may not be unique. However, we show that for $\mathcal{F}^*$ the set of reserve price values for which expected revenue is maximized is at most finite: each maximizer can be consistently estimated.

Given that there can be multiple expected revenue maximizers, one still needs to pick one. Note that all elements in the solution set yield the same expected revenue for the ME value distribution in contrast to, say, the HT identified set. Therefore, by the entropy criterion, all elements in the solution set are equivalent. For most purposes, simply choosing *a* reserve price that maximizes the *estimated* expected revenue will be satisfactory. Doing so has the advantage that one does not have to estimate the entire set of maximizers, which entails the choice of an input parameter: this is our recommended approach. If one insists on estimating the entire set then one can introduce tie breakers. For example, one could compare the worst case loss for each maximizer. A more sophisticated alternative approach is to search a value density that maximizes entropy among the distributions that are orthogonal to the original ME density f^*. Finally, one can simply focus on the smallest element of the set of ME optimal reserve price values. This approach can be justified from a welfare perspective.[14]

If one intends to construct a confidence interval for the optimal reserve price or maximum attainable expected revenue for the ME distribution then our recommended approach will not suffice. Hence, for the purpose of inference on the optimal reserve price, we focus our efforts on the simplest

[14] The seller is indifferent and the buyer gains if the probability of a sale increases with the same expected revenue.

approach, i.e. choosing the smallest element in the set of ME optimal reserve price values. We show that our estimator of the smallest ME optimal reserve price has similar statistical properties to our estimator of $\mathscr{f}^*$: a convergence rate no worse than $\sqrt{N}$ and a limit distribution that is not normal, but can be simulated. Our estimator of expected revenue, however, is $\sqrt{N}$–consistent in all cases. For expected revenue a simulation–based method can be used to conduct inference, also.

Our paper is organized as follows. In section 2 we set up the environment. In section 3, we derive the formulation of the ME solution of $\mathscr{f}^*$. In section 5 we use simulations to compare HT's, AK's, and our approaches. Section 6 contains a description of the entropy–constrained maxmin problem. Then in sections 7 and 8 we develop statistical properties of our estimator of $\mathscr{f}^*$. Section 9 contains our results on the optimal reserve price and expected revenue.

2. Preliminaries

As mentioned in section 1, we consider an English auction with symmetric bidders and exogenous entry under the IPV paradigm, in which there is a minimum bid increment $\Delta > 0$. Unlike HT, we assume that bid increments are multiples of Δ, i.e. one of $\Delta, 2\Delta, \ldots,$ albeit that any other known discrete scheme works, also.[15] In what follows, we shall refer to HT as HT's approach with this additional assumption. Thus, bidders' values $v_1, \cdots, v_n$ are independent and identically distributed (i.i.d.) draws from an unknown continuous distribution function $\mathscr{F}$ with positive density function $\mathscr{f}$. We assume that the potential number of bidders n is known. Throughout the paper we assume that the support of v_i is the unit interval [0, 1]. The text is phrased as though we observe the highest bid of each bidder, but observing the winning bid (and the number of potential bidders) is sufficient if one sets the unobserved bids to zero, albeit that the distribution function bounds are then wider. Because this is an English auction, there is no one–to–one correspondence between the observed bid and a bidder's value. Further, since $\Delta > 0$, the bid distribution is discrete.

The objective in the empirical auctions literature is typically to uncover $\mathscr{F}$ from the bid distribution, which is then used to obtain policy–relevant objects such as the expected revenue and optimal reserve price. However, as HT point out, in our setup (point) identification does not obtain. One reason is that $\Delta > 0$. If Δ were equal to zero then $\mathscr{F}$ can be point identified under plausible assumptions such as the absence of jump bidding. If Δ is nonzero then the bid distributions provide bounds on $\mathscr{F}$. Below we briefly review HT's results and issues surrounding them.

It is well–known from the order statistics literature that if $u_1, \cdots, u_n$ are independent with standard uniform distributions then $u_{i:n}$, the i^{th} (smallest) order statistic, has a beta distribution with parameters i and $n-i+1$, i.e. $u_{i:n} \sim \mathscr{B}(i, n-i+1)$. Consequently, $\mathscr{F}(v_{i:n})$ has a $\mathscr{B}(i, n-i+1)$ distribution whose distribution function will be denoted by $\mathscr{H}_{i:n}$. Let $\mathscr{G}_{i:n}$ be the distribution function of the i^{th} lowest bid $b_{i:n}$ and let $r \in [0, 1]$ be the reserve price.

[15] For instance, small increments at low bid levels and large increments at high bid levels.

Theorem 1 (Haile and Tamer, 2003)**.** Suppose that bidders do not bid more than they are willing to pay and that they do not allow an opponent to win at a price that they are willing to beat. Then, for all $v \in [r, 1]$, we have $\mathscr{F}_L(v) \leq \mathscr{F}(v) \leq \mathscr{F}_U(v)$, where

$$\mathscr{F}_L(v) = \mathscr{H}_{n-1:n}^{-1}\{\mathscr{G}_{n:n}(v - \Delta)\} \quad \text{and} \quad \mathscr{F}_U(v) = \min_{i=1,2,\cdots,n} \mathscr{H}_{i:n}^{-1}\{\mathscr{G}_{i:n}(v)\}. \qquad \square$$

There are a few comments to make. First, to facilitate the discussion we assume that the number of participants is constant across auctions: an extension to different numbers of bidders is simple but messy. Second, the bounds in theorem 1 can be estimated. Indeed, given a sample of size N, $\sqrt{N}$–consistent estimators of $\mathscr{F}_L, \mathscr{F}_U$ can be constructed. Further, the bounds $\mathscr{F}_L, \mathscr{F}_U$ are step functions because the support of b_i is discrete since $\Delta > 0$. Finally, the bounds on $\mathscr{F}$ do not say much about the density function $f = \mathscr{F}'$, which is needed to analyze e.g. the optimal reserve price. In this section we will be mostly concerned with the density issue.

We now turn to the analysis of the optimal reserve price in a counterfactual environment in which $\Delta = 0$ and the assumptions for the revenue equivalence theorem in Myerson (1981) are satisfied. This is essentially the same exercise as in HT, albeit that we do *not* impose HT's pseudoconcavity assumption. As noted in section 1,[16] jump bidding is effectively ruled out in this counterfactual environment. Let $\tilde{\pi}(r; \mathscr{F})$ be the expected revenue function in this thought experiment, i.e.

$$\begin{aligned}\tilde{\pi}(r, \mathscr{F}) &= \mathbb{E}\{\max(v_{n-1:n}, r)\mathbb{1}(v_{n:n} > r)\} \\ &= 1 - r\mathscr{F}^n(r) + \int_r^1 \{(n-1)\mathscr{F}^n(v) - n\mathscr{F}^{n-1}(v)\}\, \mathrm{d}v. \qquad (1)\end{aligned}$$

So, like HT and AK, we use data on auctions with $\Delta > 0$ to study the optimal reserve price in regular second price auctions, i.e. second price auctions with $\Delta = 0$ in which the assumptions for the revenue equivalence theorem are satisfied. This is clearly not ideal but, without further assumptions about bidder behavior, it is the best that can be done. Indeed, with only the minimal behavioral assumptions made in HT, AK, and here, the conditions for Myerson (1981)'s revenue equivalence theorem are not met: for instance, the assumptions are not sufficient for the existence of Nash equilibrium bidding strategies in English auctions.[17] Conversely, if the assumptions necessary for the revenue equivalence theorem were satisfied for the auctions in the data then point identification could obtain.[18]

[16] See footnote 3.

[17] To establish lemma 4 in their paper, HT make additional assumptions, including the feasible auction mechanism requirement we mentioned in the introduction.

[18] The requirements for Myerson's revenue equivalence theorem are not met and point identification does not obtain for several reasons. For starters, one cannot identify a continuous value distribution from a discrete bid distribution. Further, if $\Delta > 0$ then the winning bidder can end up paying more or less than the value of the second highest bidder: the sequence in which bids are submitted matters. If the winning bidder wins with bid $\bar{b}$ then that only means that the second highest bidder's value is in the interval $[\bar{b} - \Delta, \bar{b} + \Delta)$. (This is not intended to be a complete list.)

Thus, the optimal reserve price r_0 satisfies

$$\frac{\partial_r \tilde{\pi}(r_0, \mathcal{F})}{n\mathcal{F}^{n-1}(r_0)} = 1 - \mathcal{F}(r_0) - r_0 \mathcal{f}(r_0) = 0, \tag{2}$$

provided that $r_0 \in (0, 1)$.

Solving the first order condition in (2) requires knowledge about the density function $\mathcal{f}$. Even then, the solution need not be unique. Therefore, the bounds in theorem 1 generally do not provide much, if any, information about r_0.

HT's approach in this situation is to note that the right–hand side of (2) is the derivative of $\mathcal{Q}(r) = r\{1 - \mathcal{F}(r)\}$ at r_0, and they restrict the function $\mathcal{Q}$ to be strictly pseudoconcave. This extra assumption on the shape of $\mathcal{Q}$ ensures that r_0 is uniquely defined and it allows them to derive bounds on r_0 from the bounds on $\mathcal{F}$. This solution is defensible because similar assumptions have been made in the auctions literature. But there are limitations: strict pseudoconcavity of $\mathcal{Q}$ rules out some plausible value distributions. For instance, the distribution described in example 1 has a unimodal density that appears to be perfectly 'regular' yet violates the strict pseudoconcavity assumption.

Example 1. Suppose that $\mathcal{f}(v) = 1/(4v^2)$ for $v \in (1/2, \eta)$ with $\eta < 1$ and that $\mathcal{f}$ is unrestricted on $[0, 1/2] \cup [\eta, 1]$ except that $\mathcal{F}(1/2) = 1/2$. Then, for any $v \in (1/2, \eta)$, we have $\mathcal{F}(v) = 1 - 1/4v$ so that $\mathcal{Q}'(v) = 0$. Therefore, $\mathcal{Q}$ is not strictly pseudoconcave. □

The density $\mathcal{f}$ in example 1 satisfies neither Myerson (1981, page 66)'s condition nor HT's. But Myerson shows that his condition is not needed for the characterization of an optimal auction mechanism: it is simply helpful for the computation of one. In section 5 we provide more examples using mixtures of two Beta distributions, where the corresponding function $\mathcal{Q}$ has two distinct local maxima and hence is not pseudoconcave.

Recall that instead of using HT's assumption that restricts the set of admissible value distributions, we apply the maximum entropy principle. Letting $\mathcal{F}^*$ be the entropy–maximizing value distribution, our approach is to maximize $\tilde{\pi}(r, \mathcal{F}^*)$, whereas AK's proposal is to maximize $\tilde{\pi}(r, \mathcal{F}_U)$. To highlight the difference between the two, suppose that the reserve price used in the data is $\bar{r} > 0$. Then, all bidders with values less than $\bar{r}$ (and some with values greater than r) will not bid, i.e. they will bid zero.[19] Therefore, $\mathcal{F}_U(r)$ is flat on $[0, \bar{r})$, and it will be positive when the price distribution has positive probability mass at zero. But then (1) shows that no point in $[0, \bar{r})$ can be the optimal reserve price suggested by AK, unless $\mathcal{F}_U(r) = 1$ for all r, since $\tilde{\pi}(r, \mathcal{F}_U)$ is increasing in $r \in [0, \bar{r})$. So AK's optimal reserve price cannot be less than $\bar{r}$. This is not the case with the entropy approach. As will become clear in section 3, $\mathcal{F}^*$ is linear on $[0, \bar{r})$. Therefore, (1) shows that if $\Omega = \mathcal{F}_U(0)$ is large enough then the maximum entropy optimal reserve price is $\bar{r} \ / \ 2\Omega$. To put that into perspective suppose that in 9,999 out of 10,000 observed auctions in the data with two bidders and a reserve

[19] Bidders with values greater than zero may not bid because others have already bid past their value.

price equal to 1 / 2 there are no positive bids. So, $\hat{\mathscr{G}}_{1:2}(r) \geq 0.9999$ and $\hat{\mathscr{G}}_{2:2}(r) = 0.9999$ for all $r \in [0, 1 / 2)$. Therefore, $\hat{\mathscr{F}}_U(r)$ equals either 0.99 or $\sqrt{0.9999}$ on $[0, 1 / 2)$. Then AK would suggest to keep the reserve price at 1 / 2 and our maximum entropy solution would suggest dropping the reserve price to approximately 1 / 4. Keeping the reserve price at 1 / 2 would be optimal if the value distribution indeed has a large mass point at zero (or something extremely close to it), but in most other cases dropping the reserve price would be better.

3. Maximum entropy density

The maximum entropy principle says that the probability distribution that best represents the current state of knowledge is the one that maximizes entropy subject to constraints provided by assumptions we are willing to make. All we know about the value distribution is that it satisfies the bounds in theorem 1. Thus, the maximum entropy density, i.e. the least informative density given the information provided by the identified bounds is given by

$$f^* = \operatorname*{argmin}_{f} \int_0^1 f(s) \log f(s) \, \mathrm{d}s \quad \text{subject to} \quad \begin{cases} \displaystyle\int_0^1 f(s) \, \mathrm{d}s = 1, \\ \forall v : \mathscr{F}_L(v) \leq \displaystyle\int_0^v f(s) \, \mathrm{d}s \leq \mathscr{F}_U(v), \end{cases} \tag{3}$$

where the bounds $\mathscr{F}_L, \mathscr{F}_U$ were given in theorem 1 and $\mathscr{F}(v) = \int_0^v f(s) \, \mathrm{d}s$ is continuous in v.

The optimization problem in (3) is simpler than a typical infinite–dimensional optimization problem because the bounds $\mathscr{F}_L$ and $\mathscr{F}_U$ are step functions, as noted earlier. Below we reformulate (3) as a finite–dimensional convex optimization problem.

Suppose that for some $J \geq 2$ the support of the bid distribution consists of the points $0 = \beta_0 < \beta_1 < \cdots < \beta_J < \beta_{J+1} = 1$. For the sake of presentational simplicity, we assume that the β_j's are equally spaced, i.e. $\beta_j = j\Delta$.

Lemma 3.1. The solution f^* to (3) is constant on each $I_j = [\beta_{j-1}, \beta_j)$ for $j = 1, 2, \cdots, J+1$. □

Therefore, solving (3) is a finite–dimensional problem in terms of its complexity. Specifically, (3) can be solved by finding

$$g_1^*, \ldots, g_{J+1}^* := \operatorname*{argmin}_{g_1, \ldots, g_{J+1} \geq 0} \sum_{j=1}^{J+1} g_j (\log g_j - \log \Delta)$$

$$\text{subject to} \quad \begin{cases} \sum_{j=1}^{J+1} g_j = 1, \\ \Lambda_{0j} \le \sum_{k=1}^{j} g_k \le \Upsilon_{0j} \quad \text{for } j = 1, 2, \ldots, J, \end{cases} \tag{4}$$

where $g_j = \Delta \ell(\beta_{j-1})$ and $0 \le \Lambda_{0j} := \mathcal{F}_L(\beta_j) \le \Upsilon_{0j} := \mathcal{F}_U(\beta_{j-1}) \le 1$.[20] Since $g_1^*, \cdots, g_{J+1}^*$ sum to one, we define $g^* = [g_1^*, \cdots, g_J^*]^\intercal \in \mathbb{R}^J$, omitting g_{J+1}^*. We show in appendix A that the sign constraints are never binding, i.e. $g_j^* > 0$ for all j.

The bounds $\Lambda_0 = [\Lambda_{01}, \cdots, \Lambda_{0J}]^\intercal$ and $\Upsilon_0 = [\Upsilon_{01}, \cdots, \Upsilon_{0J}]^\intercal$ are unknown but can be estimated at the parametric rate. Estimation and inference will be discussed later. The vector of all the bounds will be denoted by $D_0 = [\Lambda_0^\intercal, \Upsilon_0^\intercal]^\intercal$. When we wish to emphasize the dependence of g^* on D_0, we will write $g^*(D_0)$.

The optimization problem in (4) is a finite–dimensional convex programming problem, for which many well–known algorithms are available (e.g. Bertsekas, 2015). In fact, since the objective function in (4) is strictly convex (e.g. by the log sum inequality) and all constraints are linear, the solution to (4) is unique.

The maximum entropy density function ℓ^* can be obtained from g^*:

$$\ell^*(v) = \ell^*\{v, g^*(D_0)\} = \sum_{j=1}^{J+1} g_j^*(D_0) \mathbb{1}(v \in I_j) \,/\, \Delta, \tag{5}$$

where $\mathbb{1}$ is the indicator function. Consequently, the maximum entropy distribution function $\mathcal{F}^*$ is given by

$$\mathcal{F}^*(v) = \mathcal{F}^*\{v, g^*(D_0)\} = \int_0^v \ell^*\{s, g^*(D_0)\}\, \mathrm{d}s = \sum_{j=1}^{J+1} \mathbb{1}(v \in I_j) a_j^\intercal(v) g^*(D_0), \tag{6}$$

where $a_j(v) = \left[1, 1, \cdots, 1, v/\Delta - j + 1, 0, 0 \cdots, 0\right]^\intercal$.[21] Figure 2 shows an example using simulated bids.

For figure 2, the bounds are estimated with S =100,000 simulated auctions with $n \in \{2, 6\}$ potential bidders, $\Delta = 0.1$, and two different value distributions. Note that the bounds depend on the bidding strategies used. Here, bids were generated by randomly choosing among the currently losing bidders whose values were at least equal to the current bid level plus Δ and assign a bid equal to the current bid level plus Δ to the chosen bidder.

[20] The lower bound in (4) comes from the fact that $\mathcal{F}$ is known to be continuous such that $\lim_{v \uparrow \beta_j} \mathcal{F}(v) = \mathcal{F}(\beta_j) \ge \mathcal{F}_L(\beta_j)$. For details, see the proof of lemma 3.1 in appendix A.

[21] For $v \in I_j = [\beta_{j-1}, \beta_j) = \Delta[j-1, j)$, we have $\mathcal{F}^*(v) = a_j^\intercal(v) g^* = \sum_{k=1}^{j-1} g_k^* + (v/\Delta - j + 1) g_j^*$.

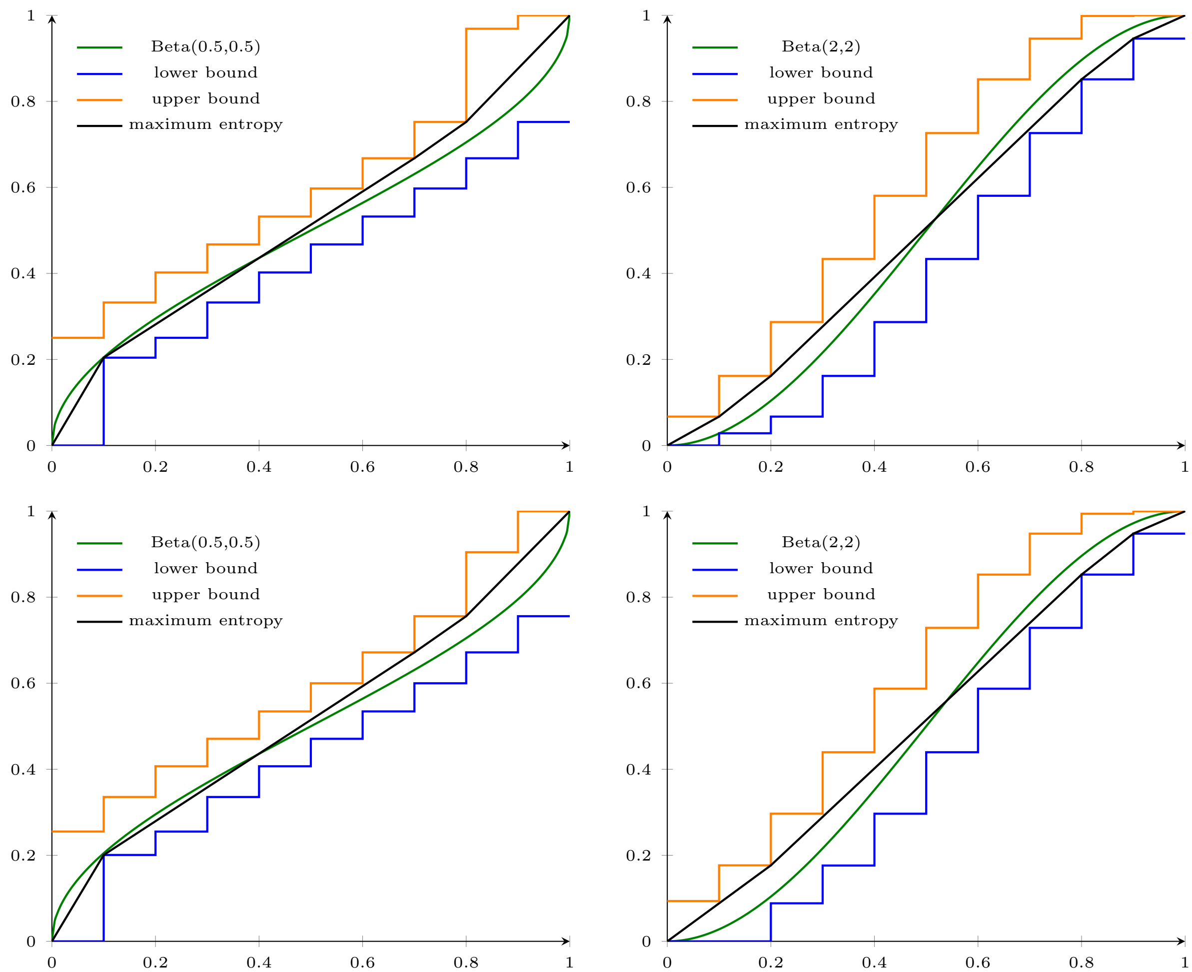


Figure 2: The HT bounds and the maximum entropy solution: the top figures are based on 100,000 simulated auctions with $n = 2$, $\Delta = 0.10$ using Beta$(0.5, 0.5)$ (left) and Beta$(2, 2)$ (right) as the true value distributions. The bottom figures use $n = 6$ with the other settings the same.

When n is larger, the lower bounds tend to be zero in a wide rage of small values of v, which is because the empirical distribution function of the largest bid tends to be zero for "small" values of v with finite S. This becomes an undesirable issue for the upper bound: the upper bound estimated by the usual empirical distribution function will be zero for "small" values of v. In order to avoid this problem, we used $\hat{\mathcal{G}}_{i:n}(v) = \left\{\sum_{s=1}^{S} \mathbb{1}(b_{i:n,s} \leq v) + 1\right\} / (S + 1)$, where $b_{i:n,s}$ is the i^{th} (smallest) bid in auction s. Please note that $\hat{\mathcal{G}}_{i:n}$ is asymptotically equivalent to the usual empirical distribution but is always positive with finite S.

As figure 2 shows, the maximum entropy density generally differs from the true value density. This is not surprising since the true value density is not point identified. We propose the maximum entropy density as a representative value distribution under partial identification: by the maximum entropy principle it is the least informative choice among the value distributions in the identified set.

Now consider the maximum entropy expected revenue function, $\pi(r, g^*) = \tilde{\pi}\{r, \mathcal{F}^*(\cdot, g^*)\}$. The maximizer and the maximum of $\pi(\cdot, g^*)$ are the maximum entropy optimal reserve price and the corresponding expected revenue, respectively. As we mentioned in section 1, the maximum entropy optimal reserve price is always contained in HT's interval for the optimal reserve price. However, the maximum entropy distribution itself need not satisfy HT's strict pseudoconcavity assumption.

4. Optimal auctions

Myerson (1981)'s strict monotonicity assumption on the so–called virtual valuation function implies the strict pseudoconcavity of the function $\mathcal{Q}$. Therefore, Myerson (1981)'s results show that absent strict pseudoconcavity of $\mathcal{Q}$ a second price auction is not necessarily optimal.

With the maximum entropy value distribution in hand, one can construct a Myerson optimal auction, which can yield higher expected revenue for the seller than a second price auction with a maximum entropy optimal reserve price. Doing so would be impossible if one merely had bounds on the value distribution function.

The optimal auction mechanism works as follows. Let

$$T^*(q) = -F^{-1}(q)(1-q), \quad q \in [0,1], \tag{7}$$

and let T be its *convex hull,* i.e. the maximum convex function below or equal to T^*. Define

$$C(v) = T'\{F(v)\}.^{22} \tag{8}$$

C is continuous and nondecreasing, which plays a critical role in describing the allocation rule of Myerson's optimal auction. If T^* were convex already, then C would be the same as C^* with $C^*(v) = v - \{1 - F(v)\} / f(v)$, which would be increasing.

To see how Myerson's allocation rule can be different from that of a second price auction, suppose that $n = 2$ and $v_1 > v_2$. We continue to assume that the seller values the object at zero. Let r_m^* denote the smallest value of v for which $C(v) = 0$. Further, define $\bar{b} = \max\left[\max\{v : C(v) = C(v_2)\}, r_m^*\right]$, and let $\underline{b}$ be the greater of r_m^* and $\min\{v : C(v) = C(v_2)\}$. If $v_1 > \bar{b}$ then player 1 wins the auction and pays $(\bar{b} + \underline{b}) / 2$. If $v_1 \leq \bar{b}$ then each player wins the auction with probability $1 / 2$ and the winner pays $\underline{b}$. If C were strictly convex then $\underline{b} = \bar{b} = \max(v_2, r_m^*)$ and we would be back in the standard second price auction case.

To illustrate, consider figure 3 which is based on a value distribution that is a mixture of Betas.[23] Here, as is apparent from the dotted lines, $r_m^* = 0.2$. If $v_1 < r_m^*$ then the object is not sold and if

[22] We are ignoring the easily addressed nuisance that T may not be differentiable at a countable set of points.

[23] 0.95 times a Beta(2,10) plus 0.05 times a Beta(20,2). The corresponding $\mathcal{Q}$ function has two distinct local maxima. See section 5.

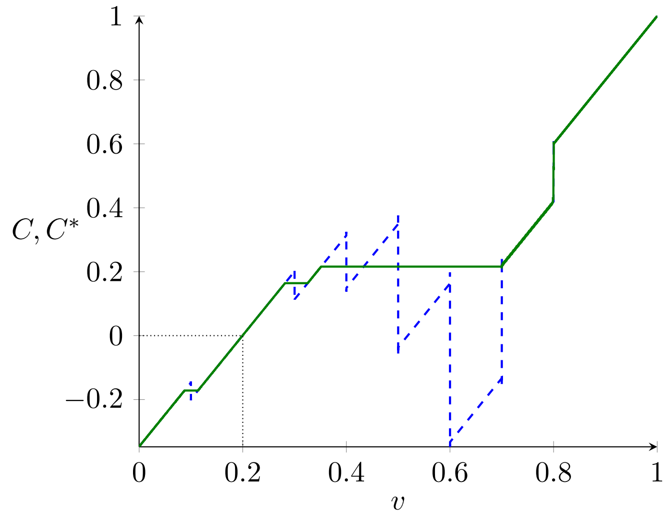


Figure 3: The functions C^* (blue dashed) and C (green solid) for a mixture of Betas.

$v_2 < r_m^* < v_1$ then player 1 wins the auction and pays r_m^*. So suppose that $v_2 \geq r_m^*$. If v_2 belongs to an area in which C is increasing (i.e. not flat) then player 1 wins and pays v_2. Finally, suppose that v_2 is in an area in which C is flat, say the large flat segment that extends from about 0.38 to about 0.70. If v_1 is also between 0.38 and 0.70 then each player wins with probability 0.50 and the winner pays 0.38. Otherwise, player 1 wins with certainty and pays 0.54.

As noted, the Myerson mechanism yields an expected profit for the seller that is equal to or exceeds that of a second price auction with an optimally chosen reserve price. However, the difference in expected profit is generally small. For instance, in the example of figure 3 the difference is about 0.001 despite the substantial convexity correction evidenced by the difference between dashed and solid lines in figure 3. This is not entirely surprising in view of Hartline (2016, corollary 5.3), which provides bounds on the gain from an optimal auction compared to a second price one and is a corollary to the Bulow–Klemperer theorem (Bulow and Klemperer, 1996), albeit that the bounds are fairly wide if the number of bidders is small. Further, as noted, it can be cumbersome to implement Myerson's auction in practice.

We therefore focus on the choice of an optimal reserve price in a second price auction in the remainder of this paper.

5. Comparison of methods

We now use simulations to compare the three approaches to the seller's problem, namely the bounding approach of Haile and Tamer, the maxmin approach of Aryal and Kim, and our entropy–based approach. We draw values independently from distributions chosen by us and simulate bids, for which we adopt the algorithm described in HT's example 2 in their appendix B. We then compute the identified bounds of the value distribution function using S =100,000 auctions to make estimation error negligible. The bounds on the value distribution function are then used to find the (bounds on the) reserve prices proposed by HT, AK, and the present paper.[24] In the discussion below, for a given value distribution function $\mathcal{F}$, we shall refer to the function $\tilde{\pi}(r, \mathcal{F})$ defined in (1) as the profit function and to $\mathcal{Q}(r) = r\{1 - \mathcal{F}(r)\}$ as the pseudoprofit function.

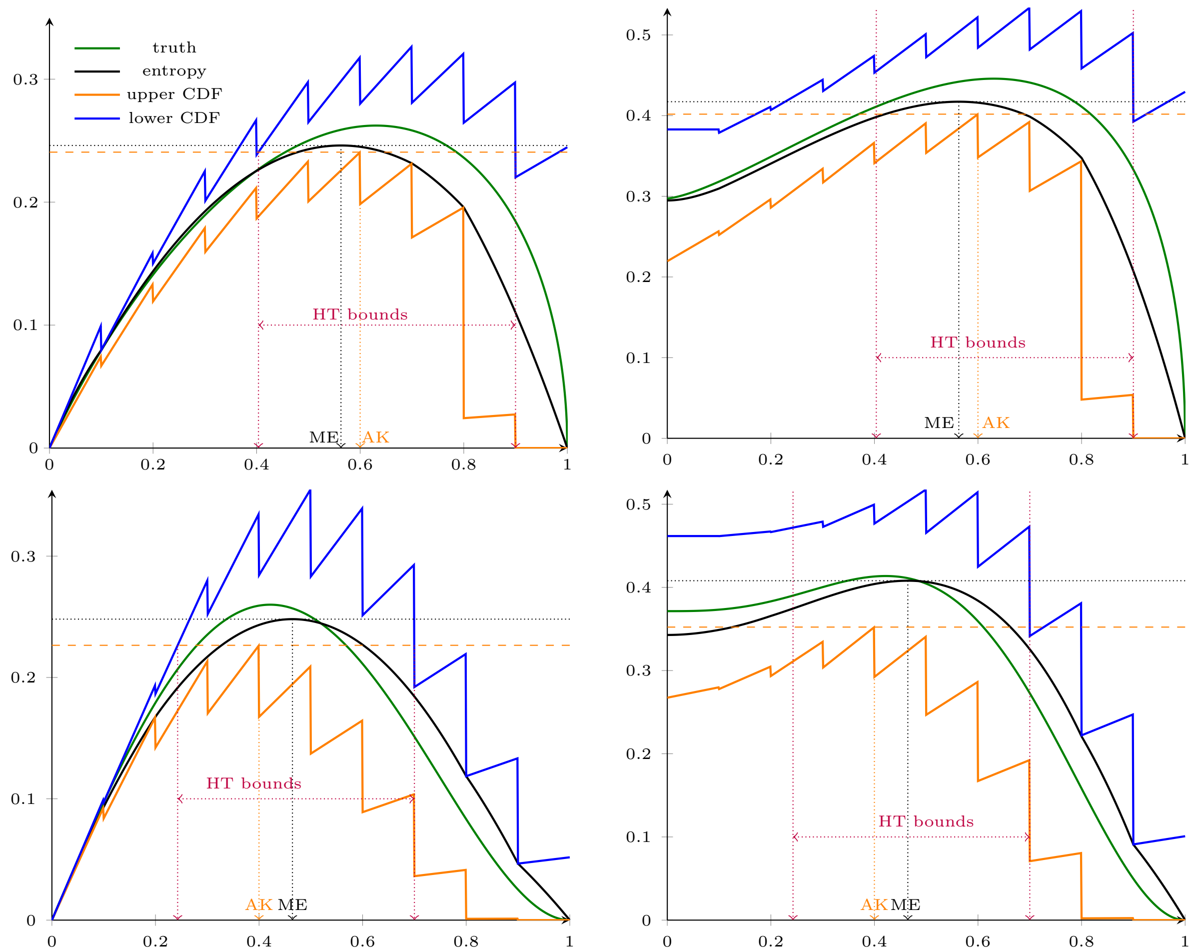


Figure 4: Pseudoprofit (left) and profit (right) for the Beta(0.5, 0.5) (top) and Beta(2, 2) (bottom) value distributions; $n = 2$, $\Delta = 0.1$. Green represents the design and black the maximum entropy solution.

In our experiments we use $n = 2$ since the reserve price has the greatest impact when the number of bidders is small: there is sufficient competition when n is large. Note that n here is the number of bidders in the hypothetical second price auction: the number of bidders in the data is immaterial for the purposes of the present exercise. The results described are based on $\Delta = 0.10$ unless otherwise indicated.

[24]So, what is marked as AK uses all bids, as we mentioned in section 1.

For figure 4 we used Beta$(0.5, 0.5)$ and Beta$(2, 2)$ as the value distributions, resulting in strictly pseudoconcave pseudoprofit functions, as HT requires. Figure 4 has four panels: for each of these two distributions there is one graph for the pseudoprofit function and one for the profit function. In each graph we draw the 'truth,' the (pseudo)profit function corresponding to the maximum entropy solution and its maximizer (marked ME), the HT bounds, and the maxmin optimal reserve price, marked AK. The punkish nature of the lower and upper bounds is due to the fact that the bounds on the distribution function are step functions: see figure 2.[25] The maximum entropy solution is a value at which the maximum entropy profit function is maximized which, in the case of strict pseudoconcavity, coincides with the point at which the maximum entropy pseudoprofit function is maximized. The maxmin solution is the point at which the lower bound to the profit function is maximal, which is always at one (or more) of the nodes. Finally, to obtain the HT bounds one takes the AK solution and extends it left and right to the point at which the upper bound to the pseudoprofit function attains the maximum of the lower bound.

The HT bounds for the optimal reserve price are well defined and informative in both designs depicted in figure 4, albeit that they are too wide to be of much use. In the Beta$(0.5, 0.5)$ design, maxmin yields a higher profit whereas in the Beta$(2, 2)$ design maximum entropy does better.[26] Also, note that the HT bounds in both designs contain the maximum entropy solution. This phenomenon arises because of the continuity of the maximum entropy distribution function. Since the pseudoprofit function corresponding to the maximum entropy distribution is continuous, its maximum value can never be smaller than the maximum of the worst case pseudoprofit function.

Figures 5 to 7 illustrate what can happen if HT's strict pseudoconcavity assumption is violated. First, in figure 5 the pseudoprofit function has a flat area; see example 1. In this example, the support of the value distribution is larger than $[0, 1]$; we only draw the unit interval because that is where the action is. We set $\not{f}(v) = -12v^2 + 8v$ on $[0, 0.5)$ and $\not{f}(v) = 1/(4v^2)$ on $[0.5, 1)$, with the remainder of the mass at or beyond 1. In this example, the HT identified set is still convex.

Second, figures 6 and 7 are based on a mixture of two Beta distributions, where the corresponding pseudoprofit functions have two distinct local maxima. Because HT's assumptions are violated, their results do not apply here. However, applying HT's machinery to the bounds of the pseudoprofit function produces a nonconvex set for the optimal reserve price. Neither the entropy approach nor the maxmin solution requires the pseudoconcavity assumption to yield a point decision for the reserve price.

Figure 7 shows that the maximum entropy reserve price and the AK solution can be substantially different. In this example, the profit generated by maximum entropy is considerably higher than that generated by maxmin. However, this is not always true, as can be seen in figure 6.

Figure 8 depicts results for a uniform value distribution with $\Delta = 0.10$ and $\Delta = 0.20$. In

[25] HT choose to smooth out the bounds, but there is no information on the bounds between nodes of the distribution.

[26] The green profit curve is higher at the value marked AK than at the value marked ME in the top graph and lower in the bottom one.

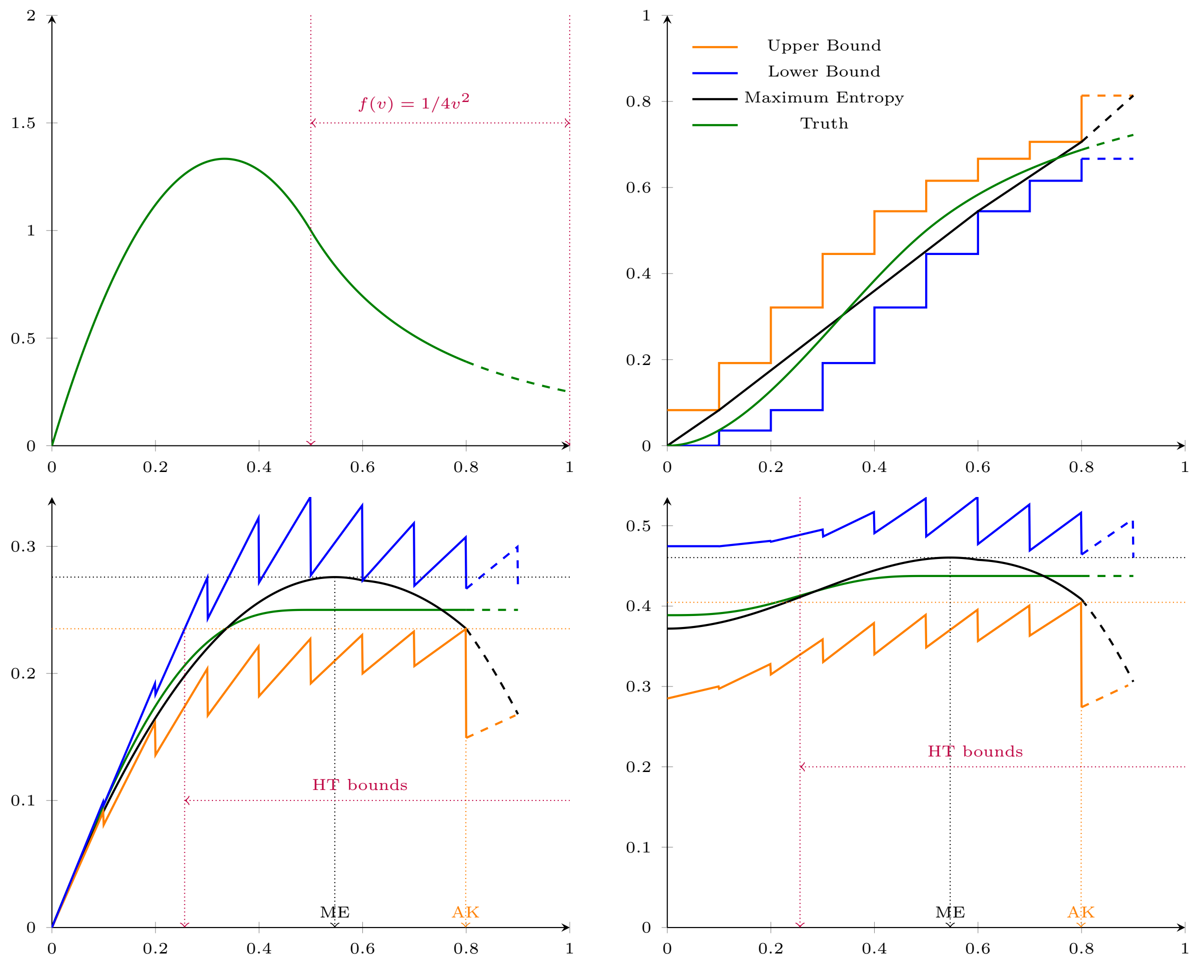


Figure 5: The top graphs show a density similar to the one discussed in example 1, along with the corresponding distribution function (green), its bounds (orange and blue), and the maximum entropy solution (black). The bottom figures show the pseudoprofit function (left) and the profit function (right).

this case the entropy–maximizing distribution is uniform, also, so the true pseudoprofit and profit functions coincide with their maximum entropy counterparts. There are a few points to note here. First, the HT bounds become less informative as Δ increases: for $\Delta = 0.20$, one end point of the HT bounds is uninformative in this example. Second, in contrast to maximum entropy, AK always take a value from the mass points of the bid distribution. For instance, if $\Delta = 0.20$ then the optimal reserve price 0.5 is not in the support of the bid distribution.

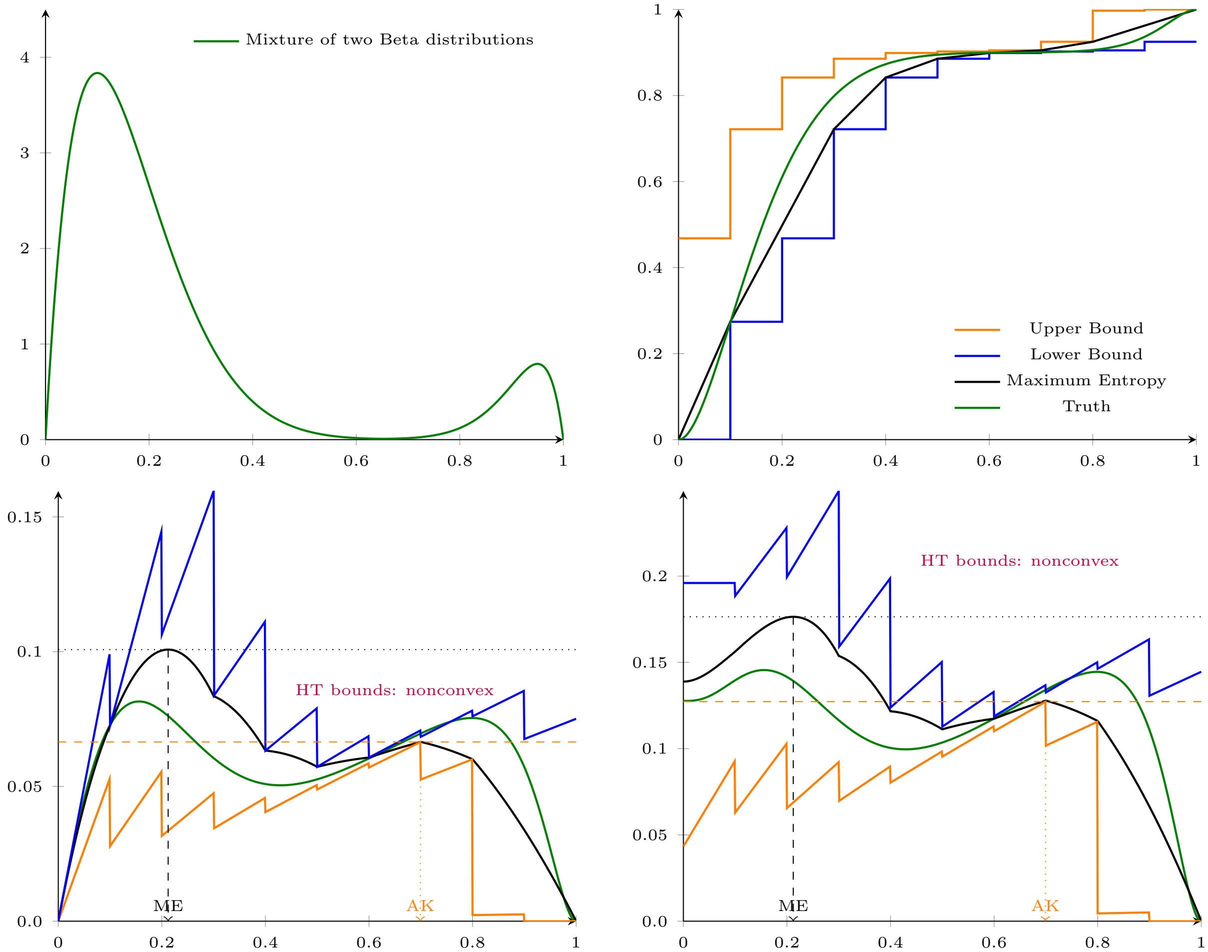


Figure 6: The top graphs shows the mixture density $0.9 \cdot \text{Beta}(2, 10) + 0.1 \cdot \text{Beta}(20, 2)$ and the corresponding distribution function (green), along with its bounds (orange and blue), and the maximum entropy solution (black). The bottom figures show the pseudoprofit function (left) and the profit function (right).

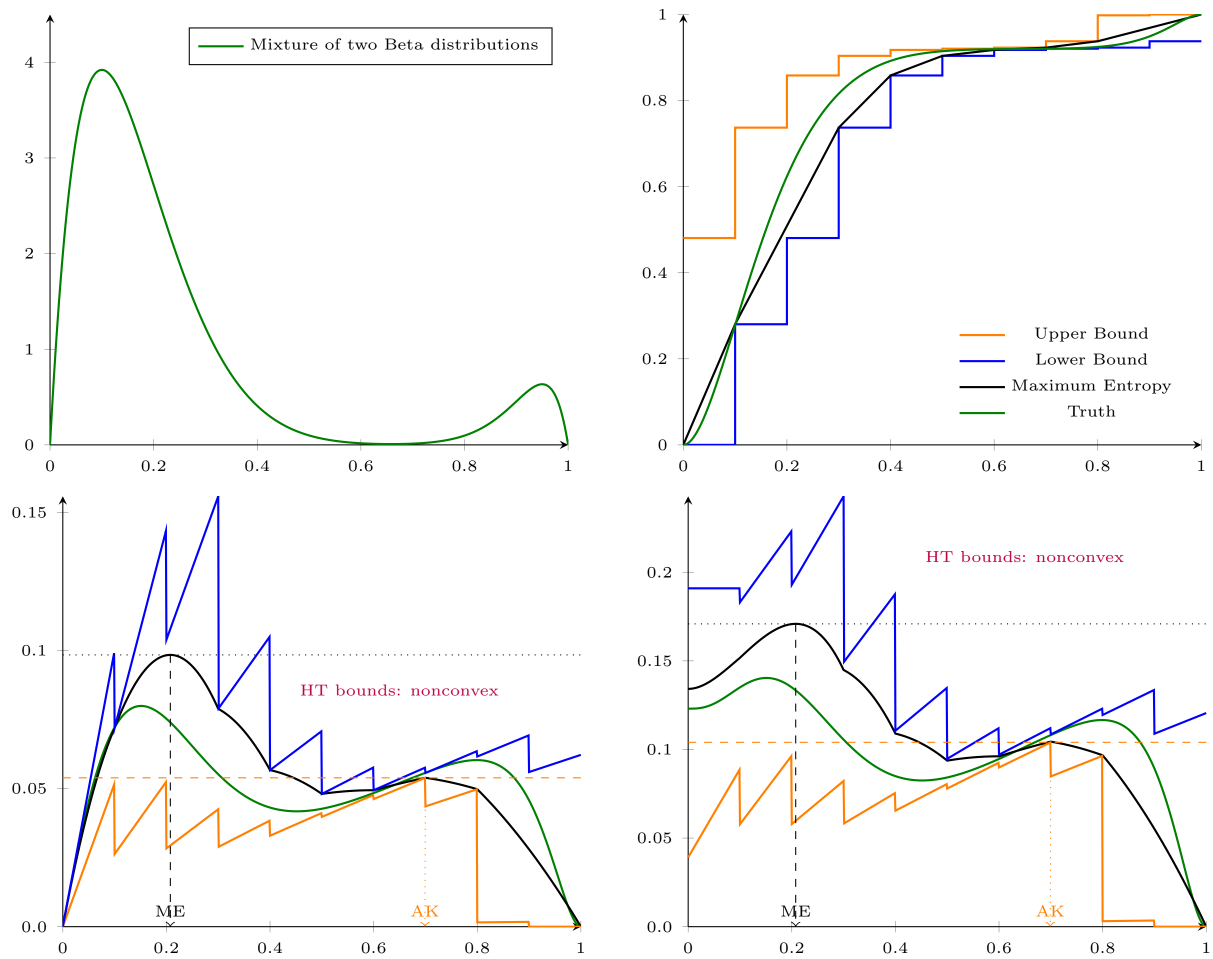


Figure 7: The top graphs shows the mixture density $0.92 \cdot \text{Beta}(2, 10) + 0.08 \cdot \text{Beta}(20, 2)$ and the corresponding distribution function (green), along with its bounds (orange and blue), and the maximum entropy solution (black). The bottom figures show the pseudoprofit function (left) and the profit function (right).

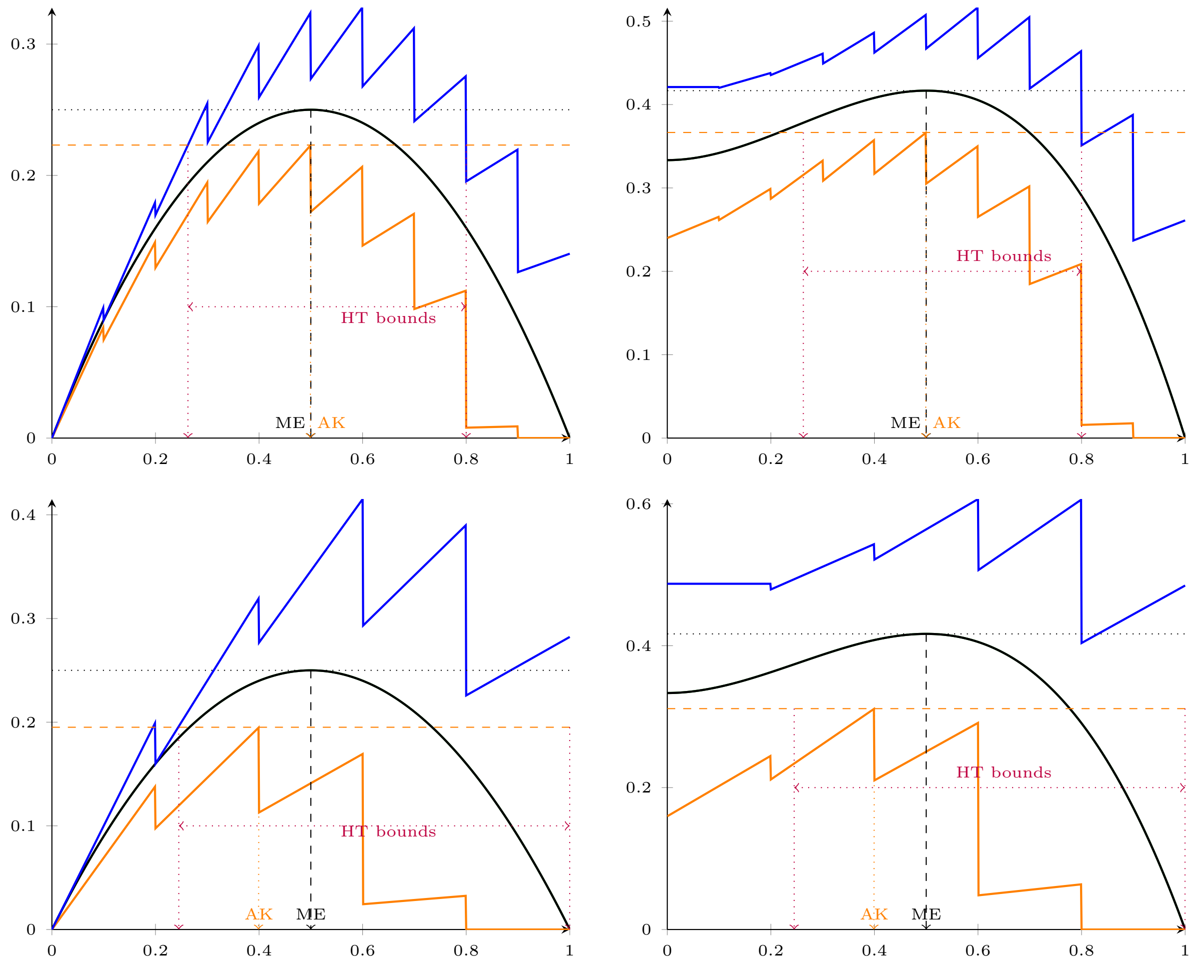


Figure 8: The true value distribution is uniform on the unit interval. The top graphs are the pseudoprofit (Left) and the profit function (Right) with $\Delta = 0.10$. The bottom graphs use $\Delta = 0.20$.

6. Entropy, maxmin, and decision theory

We now provide a decision–theoretic view of the ME solution, which is the focus of this paper. Specifically, we consider choosing the reserve price to maximize minimum expected revenue where the minimum is taken over all distributions which satisfy the bound constraints and are moreover such that their entropy is bounded below by a prespecified value $\mathcal{E}^*$. Like in AK, the seller is ambiguity–averse but she now also wants to rule out value distributions that are too crazy. Thus, $\mathcal{E}^*$ represents a tradeoff between ambiguity–aversion and plausibility. In this framework the AK and ME solutions can be understood as special cases: the AK solution corresponds to the case in which $\mathcal{E}^* = -\infty$, whereas the ME solution obtains if $\mathcal{E}^*$ is the maximum achievable entropy.

The worst case revenue for given $\mathcal{E}^*$ is

$$\min_{\ell \geq 0} \quad \tilde{\pi}(r, \mathcal{F}) \quad \text{subject to} \quad \begin{cases} -\int_0^1 \ell(s) \log \ell(s) \, \mathrm{d}s \geq \mathcal{E}^*, \\ \int_0^1 \ell(s) \, \mathrm{d}s = 1, \\ \forall v : \mathcal{F}_L(v) \leq \int_0^v \ell(s) \, \mathrm{d}s \leq \mathcal{F}_U(v), \end{cases} \tag{9}$$

where $\tilde{\pi}$ is given in (1), $\mathcal{E}^*$ is the minimum desired entropy level, and $\mathcal{F}_U, \mathcal{F}_L$ are still the identified bounds of the distribution function of valuation. Below we show how to solve (9). We focus on the simplest case, i.e. the case with two bidders: if the number of (independent) bidders is large then choosing the optimal reserve price becomes less important.

The problem in (9) involves infinite–dimensional objects and its solution is not generally closed–form. However, it can always be reformulated as a finite–dimensional parametric problem. To see this, note that $\mathcal{F}_U, \mathcal{F}_L$ are step functions, so the distribution bound constraints in (9) only contain finitely many restrictions. We can thus write the problem as

$$\min_{\ell \geq 0} \quad \left\{ 2r\bar{\mathcal{F}}(r) - r\bar{\mathcal{F}}^2(r) + \int_r^1 \bar{\mathcal{F}}^2(v) \, \mathrm{d}v \right\}$$
$$\text{subject to} \quad \begin{cases} -\int_0^1 \ell(s) \log \ell(s) \, \mathrm{d}s \geq \mathcal{E}^*, \\ \bar{\mathcal{F}}(1) = 0, \\ 1 - \Upsilon_{0j} \leq \bar{\mathcal{F}}(\beta_j) \leq 1 - \Lambda_{0j}, \quad j = 1, 2, \ldots, J, \end{cases} \tag{10}$$

where $\bar{\mathcal{F}}(v) = 1 - \mathcal{F}(v)$, $\Lambda_{0j} = \mathcal{F}_L(\beta_j)$, and $\Upsilon_{0j} = \mathcal{F}_U(\beta_{j-1})$ as explained right below (4). We then use the first order condition from variational calculus to obtain the following result. Let $k_r = \min\{j : \beta_j \geq r\}$, $\bar{Z} = \bar{\mathcal{F}}(r)$, and $M_j = \mathcal{F}(\beta_j)$, where we focus on $r > 0$.

Theorem 2. The solution $\mathcal{F}$ to (10) is such that *(i)* $\bar{\mathcal{F}}$ is continuous; *(ii)* for $v \leq r$, $\bar{\mathcal{F}}$ is piecewise linear; and *(iii)* for $v > r$,

$$\bar{\mathcal{F}}(v;\alpha,\nu,\omega,\tilde{\beta}) = \begin{cases} -\alpha\omega_j \tan\{(\nu_j - v)\omega_j\} & \text{for} \quad \max(\beta_{j-1}, r) \leq v < \min(\tilde{\beta}_j, \beta_j) \\ 1 - \Upsilon_{0j} & \text{for} \quad \min(\tilde{\beta}_j, \beta_j) \leq v < \beta_j, \end{cases} \tag{11}$$

for $\alpha < 0$ and $k_r \leq j \leq J + 1$. Further, if neither inequality condition holds with equality at β_j then $\omega_j = \omega_{j+1}$ and $\nu_j = \nu_{j+1}$. Finally, the parameter vectors ν, ω, and $\tilde{\beta}$ are known functions of $\bar{Z}, \alpha$, and the identities of the distribution bound inequalities that hold with equality. □

The proof of theorem 2 is provided in appendix C. The shape of $\bar{\mathcal{F}}$ for $v \leq r$ does not affect expected profit, so for $v \leq r$ the solution coincides with the ME solution. The parametric function $\mathcal{F}(\cdot;\alpha,\nu,\omega,\tilde{\beta})$ turns out to be the unique solution to a differential equation implied by the first order condition of (10). Here, α is the Lagrange multiplier for the entropy constraint, and ω, ν are parameters that show up in the solution to the differential equation. The $\tilde{\beta}$–parameters appear when the sign restrictions on f produce corner solutions. The intuition for this is as follows. As $\mathcal{E}^* \to -\infty$, the solution gets closer to AK's pure maxmin solution, i.e. the upper bound of the value distribution, which is discrete. Unless $\mathcal{E}^*$ is large negative, $\tilde{\beta}_j = \beta_j$.

Theorem 2 says that $\alpha < 0$ in the solution: the entropy constraint is binding for any finite $\mathcal{E}^*$. However, as we will show later in this section, the solution to (9) is continuous in $\mathcal{E}^*$, Λ_{j0}, Υ_{0j}. As $\mathcal{E}^* \to -\infty$, $\alpha \to 0$, eventually resulting in $\tilde{\beta}_j < \beta_j$.

It is convenient that the problem can be characterized as an optimization problem in two unknowns if one fixes r and the identities of the bound constraints that hold with equality. This is explained in the proof in appendix C.

By using theorem 2 we compute the solution to (10) in multiple steps. Fixing the set of constraints that hold with equality and the values of $\bar{\mathcal{F}}(r)$ and the Lagrange multiplier for the entropy constraint, we solve the parametric problem, which we then minimize over all candidate values of $\bar{\mathcal{F}}(r)$ and the entropy constraint Lagrange multiplier. For each set of constraints we thus have a solution. Finally, we minimize over all finitely many binding constraint sets.

We have implemented our method in three designs. The results of our efforts are depicted in figures 9 to 12: Figures 9, 11 and 12 depict the distribution function for which expected profit is minimized for the entropy–constrained maxmin solution for r. Figure 10 depicts the entropy–constrained minimum profit functions as a function of the reserve price.

In all three designs the nodes are at $0, 0.2, \ldots, 0.8, 1$. The main difference between the designs is in the bounds: in figures 9, 11 and 12 the lower bounds are yellow and the upper bounds are blue. In all cases, the optimal r was chosen from the grid $0.01, 0.03, \ldots, 0.99$.

For each design, four different values of $\mathcal{E}^*$ were used: -10.0, -1.0, -0.1, -0.01 for design 1 and -10.0, -1.0, -0.1, -0.05 for designs 2 and 3. The reason for the discrepancy is that for the bounds

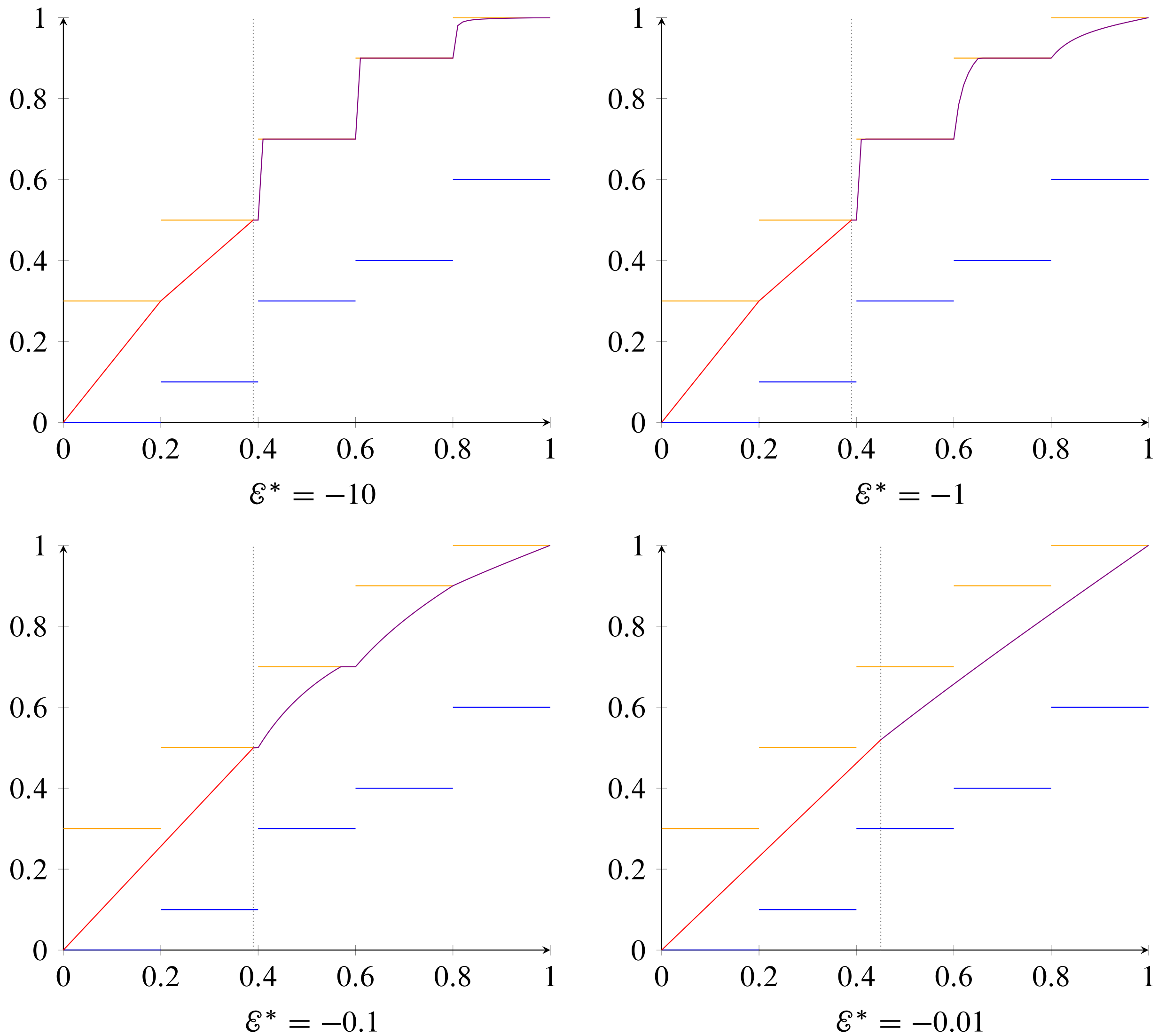


Figure 9: The graphs depict the distribution functions of the least favorable distribution for the optimal choice of r in design 1, where the least favorable distribution is determined by taking the minimum over all distributions whose entropy exceeds a threshold $\mathcal{E}^*$: each panel corresponds to a different choice of $\mathcal{E}^*$. In all cases the nodes are at $0, 0.2, \ldots, 1.0$. The graphs were created using a grid of r–values, namely $0.01, 0.03, \ldots, 0.99$. The maximal value of r is indicated by a vertical dotted line.

used in designs 2 and 3 there exist no distributions that yield an entropy value in excess of -0.01.

In design 1, the uniform distribution is the maximum entropy solution resulting in an optimal reserve price equal to 0.5. As can be seen in figures 9 and 10, for $\mathcal{E}^* = -0.01$ our methodology produces a profit function close to the profit function for a uniform value distribution and a distribution function at the optimal reserve price close to a uniform. As $\mathcal{E}^*$ decreases, however, the distribution function to the right of r places more weight on values slightly above node values. This is natural,

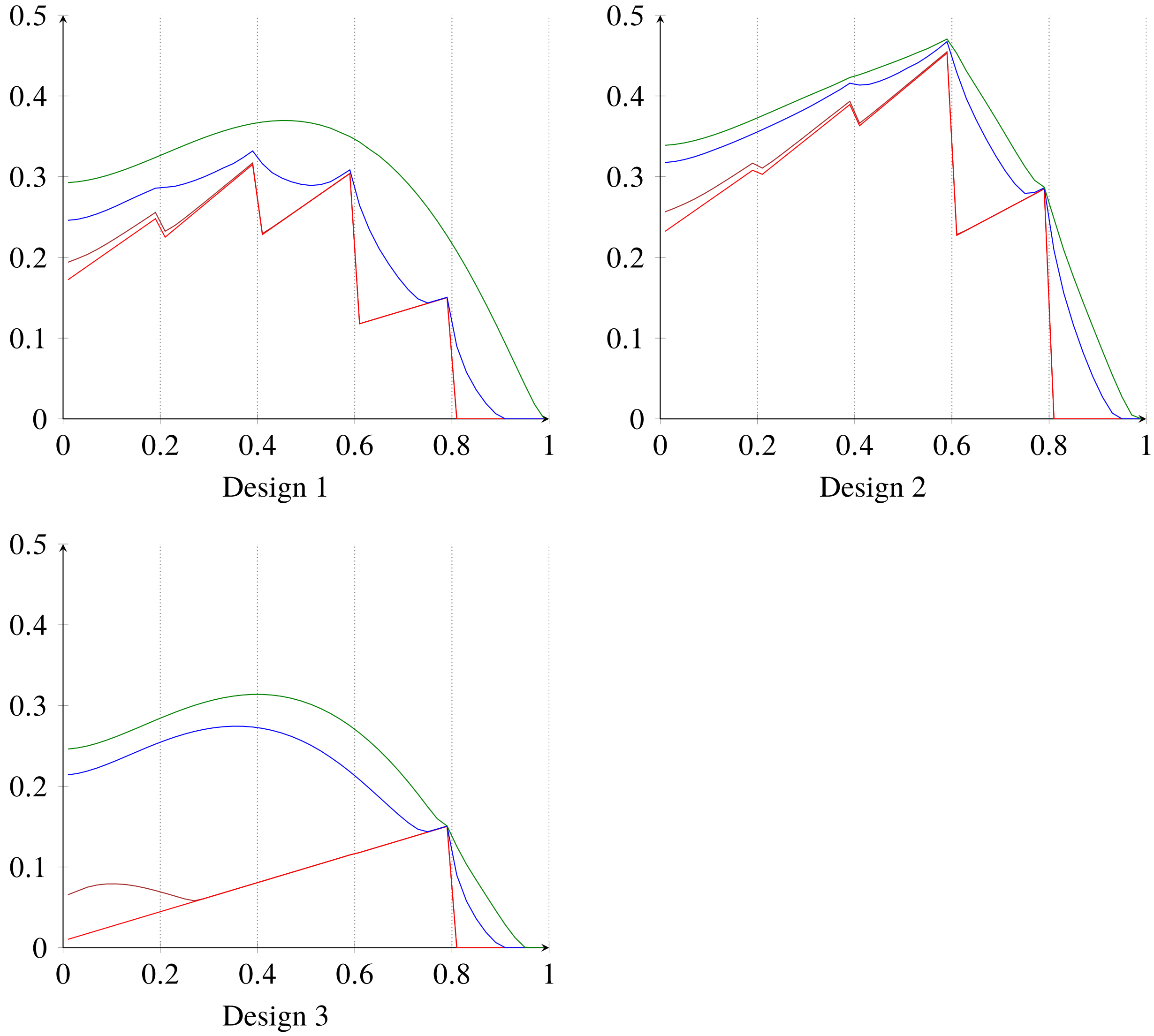


Figure 10: The graphs depict minimum expected profit as a function of r for three different choices of upper and lower bounds. The minimum at each value of r is taken over all distributions whose entropy exceeds a threshold $\mathcal{E}^*$: each curve in a given panel corresponds to a different choice of $\mathcal{E}^*$ with greater values of $\mathcal{E}^*$ resulting in higher expected profit. In all cases the nodes are at $0, 0.2, \ldots, 1.0$. The graphs were created using a grid of r–values, namely $0.01, 0.03, \ldots, 0.99$.

since putting a lot of weight there corresponds to having a pessimistic view of the world. Since the profit is unaffected by the shape of the distribution left of r, the distribution left of r is for all values of $\mathcal{E}^*$ fairly regular.

Now, minimum expected profit as a function of r for design 1 has an intuitive shape, also. For large negative values of $\mathcal{E}^*$, we get the familiar sawtooth pattern with drops at each of the nodes. The drop arises because if the seller sets the reserve price above the node value then she loses revenue from all bidders with values between the node value and r: for large negative values of $\mathcal{E}^*$ the

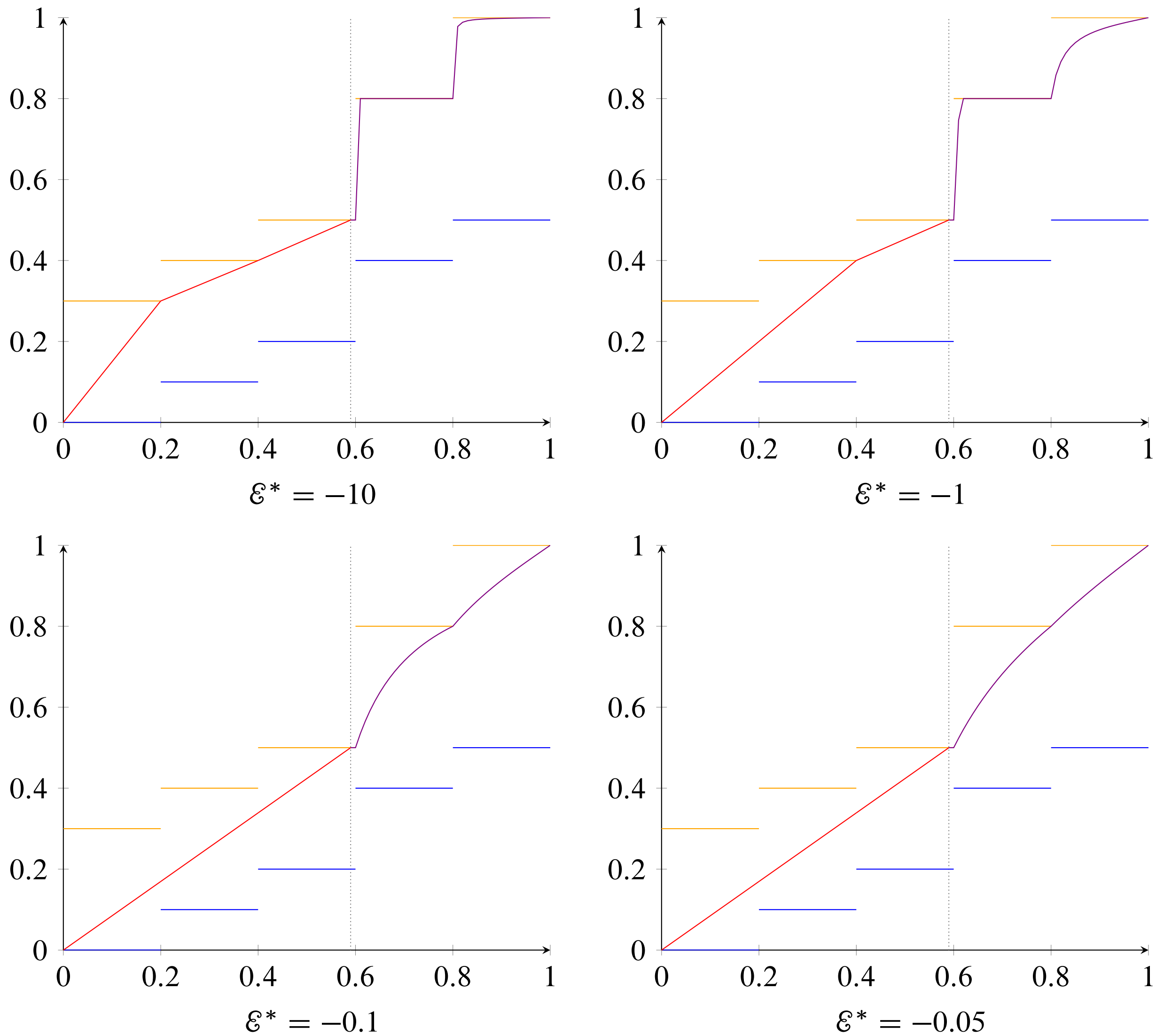


Figure 11: The graphs depict the distribution functions of the least favorable distribution for the optimal choice of r in design 2, where the least favorable distribution is determined by taking the minimum over all distributions whose entropy exceeds a threshold $\mathcal{E}^*$: each panel corresponds to a different choice of $\mathcal{E}^*$. In all cases the nodes are at $0, 0.2, \ldots, 1.0$. The graphs were created using a grid of r–values, namely $0.01, 0.03, \ldots, 0.99$. The maximal value of r is indicated by a vertical dotted line.

distribution of values is concentrated near the node values.

In design 2, the uniform distribution does not satisfy the bound constraints and hence the solution for values of $\mathcal{E}^*$ close to zero is not a uniform and $\mathcal{E}^*$ cannot be pushed to 0. Like with design 1, the profit function becomes more sawtoothed as $\mathcal{E}^*$ decreases because the least favorable distribution function at each r gets closer to the unconstrained least favorable distribution. In other ways, design 2 is similar in behavior to design 1.

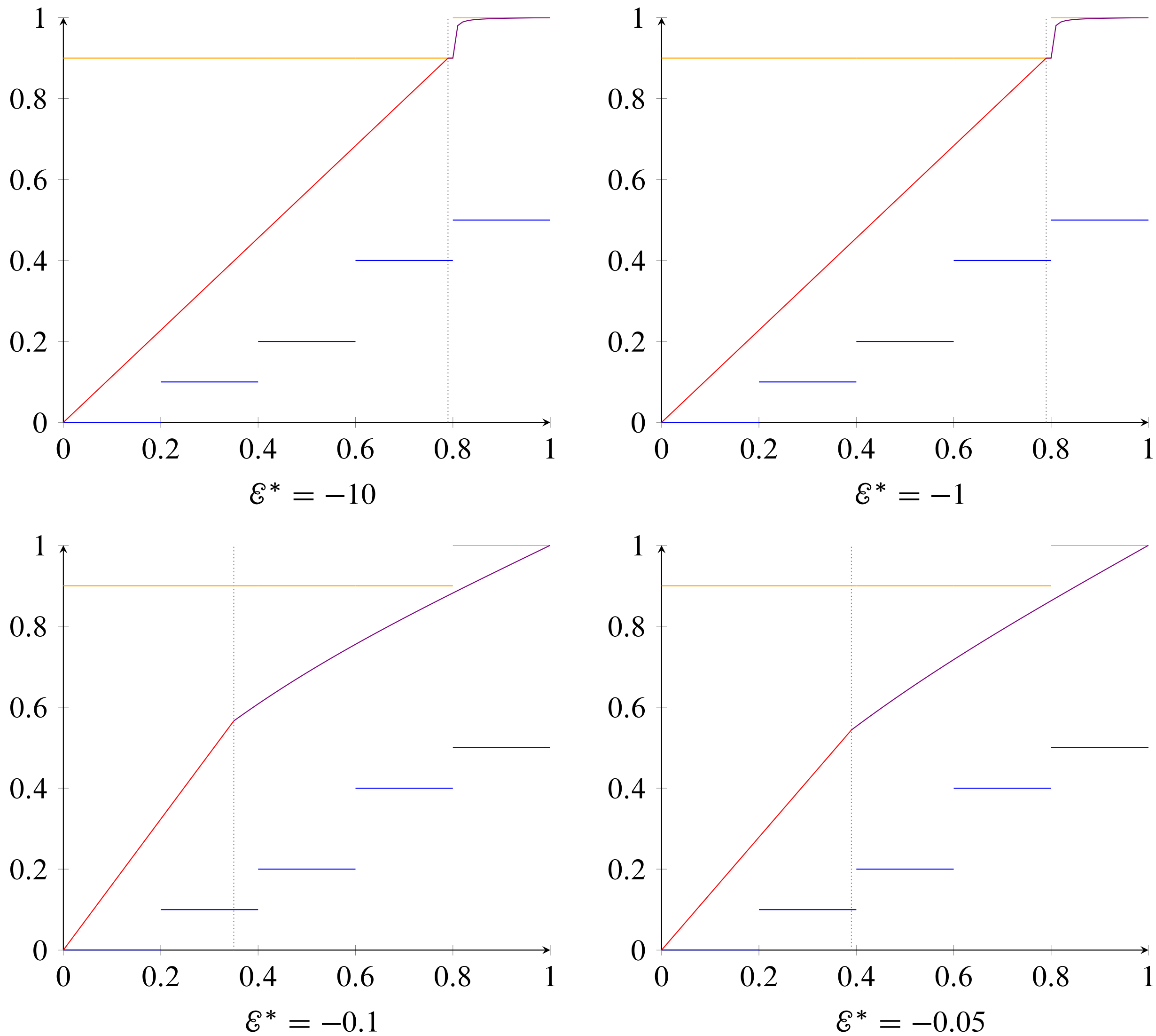


Figure 12: The graphs depict the distribution functions of the least favorable distribution for the optimal choice of r in design 3, where the least favorable distribution is determined by taking the minimum over all distributions whose entropy exceeds a threshold $\mathcal{E}^*$: each panel corresponds to a different choice of $\mathcal{E}^*$. In all cases the nodes are at $0, 0.2, \ldots, 1.0$. The graphs were created using a grid of r–values, namely $0.01, 0.03, \ldots, 0.99$. The maximal value of r is indicated by a vertical dotted line.

We deliberately chose a fairly extreme case for design 3 with an upper bound that is constant over a large range of values. That said, the solution would be much the same if the upper bound increased only slightly going from one node value to the next. In contrast to designs 1 and 2, the optimal reserve price changes dramatically depending on the choice of $\mathcal{E}^*$ and so does expected profit. For instance, if the seller sets the optimal reserve price according to $\mathcal{E}^* = -10$ then her profit is substantially less if the actual value distribution is closer to the maximum entropy distribution.

The converse is also true, but the loss is less.

As promised, we conclude by showing that the solution to (10) has some continuity properties. Let $\tilde{\mathcal{E}}^*(D_0)$ be the maximum entropy level that is attainable for given $D_0 = [\Lambda_0^\intercal, \Upsilon_0^\intercal]^\intercal$. Define $\bar{\mathcal{E}}^* = \inf_{D_0 \in \mathcal{D}} \tilde{\mathcal{E}}^*(D_0)$, where $\mathcal{D}$ is an arbitrary neighborhood of D_0. We denote the solution to (10) by $\pi^\bullet(r, \mathcal{E}^*, D_0)$, which is a function on $[0,1] \times [\underline{\mathcal{E}}^*, \bar{\mathcal{E}}^*] \times \mathcal{D}$ for $\underline{\mathcal{E}}^* > -\infty$.

Theorem 3. The function $\pi^\bullet$ is continuous on $[0,1] \times [\underline{\mathcal{E}}^*, \mathcal{E}^*] \times \mathcal{D}$. Further, $\tilde{r}(\mathcal{E}^*, D_0) = \operatorname{argmax}_r \pi^\bullet(r, \mathcal{E}^*, D_0)$ is upper–hemicontinuous on $[\underline{\mathcal{E}}^*, \bar{\mathcal{E}}^*] \times \mathcal{D}$. □

The proof of theorem 3 can be found in appendix D.

In the remainder of the paper we assume that the seller shares our preference for choosing $\mathcal{E}^*$ maximally, i.e. for maximum entropy.

7. Estimation

We now address the fact that D_0 is unknown and needs to be estimated in practice. We assume that we observe data on N homogeneous auctions which produces an estimator $\hat{D}$ of D_0.

7.1 An estimator of g^*: Once we estimate the bounds D_0, we can estimate g^*, so we consider $\hat{g}^* = g^*(\hat{D})$. Below we analyze the statistical properties of $\hat{g}^*$, assuming that $\hat{D}$ is $\sqrt{N}$–consistent.

Assumption A. *For some random vector Φ, $\hat{\Omega} = \sqrt{N}(\hat{D} - D_0) \xrightarrow{d} \Phi$ as $N \to \infty$.*

Since D_0 is a finite–dimensional vector of bounds (see theorem 1) which are probabilities, constructing a $\sqrt{N}$–consistent estimator $\hat{D}$ is a routine exercise. However, the limit distribution Φ is not usually normal because of the minimum function that appears in the upper bound; it is however easy to simulate from Φ.

Theorem 4 establishes the consistency of $\hat{g}^*$. Since g^* is the unique solution to (4), it can be shown by the maximum theorem that g^* is a continuous function of the bounds D_0. Therefore, consistency of $\hat{g}^*$ follows from the continuous mapping theorem. Formal proofs of all results are provided in an appendix.

Theorem 4. Suppose that assumption A holds. Then, $\hat{g}^* - g^* = o_p(1)$. □

Inference results for g^* follow in section 8.

7.2 An estimator of the optimal reserve price: We now briefly turn to the optimal reserve price for the maximum entropy distribution. There is no guarantee that there is a unique optimal reserve price: there can be multiple ones albeit that, as will become apparent in section 9, there are only finitely many ones. If one only desires to know *an* optimal reserve price then choosing an element from $\mathcal{R}(\hat{g}^*) = \operatorname{argmax}_r \pi(r, \hat{g})$ will do. If only the optimal profit is desired then $\mathcal{P}(\hat{g}^*) = \max \pi(r, \hat{g})$ is a consistent estimator. For a full discussion of all inference–related issues, we refer to section 9.

8. Asymptotic distribution and inference for $\hat{g}^*$

8.1 Asymptotic distribution: We now pursue the asymptotic distribution of $\hat{g}^*$ with the objective of doing inference on g^*. The theory developed in this section is not specific to the entropy problem but can be useful more generally for optimization problems with estimated inequality constraints.

For our purpose, we need to understand how perturbations to D_0 affect g^*. The function g^* is not generally differentiable at D_0. However, in lemma B.2 we show that it is directionally differentiable in every direction, which can be used to study the asymptotic distribution of $\sqrt{N}(\hat{g}^* - g^*)$. In order to describe the directional derivative, we need to discuss the behavior of the Lagrange multipliers of the optimization problem in (4).

We first eliminate g_{J+1} and one equality constraint by replacing g_{J+1} with $1 - \sum_{j=1}^J g_j$. The objective function becomes

$$Q(g) = \sum_{j=1}^J g_j(\log g_j - \log \Delta) + \left(1 - \sum_{j=1}^J g_j\right)\left\{\log\left(1 - \sum_{j=1}^J g_j\right) - \log \Delta\right\} \tag{12}$$

with constraints

$$G_j \le \Upsilon_{0j} \quad \text{and} \quad G_j \ge \Lambda_{0j} \quad \text{for } j = 1, 2, \cdots, J, \tag{13}$$

where $G_j = \sum_{k=1}^j g_k$. For the solution g^*, we define $G_j^* = \sum_{k=1}^j g_j^*$. Let $\lambda_{uj}^* \ge 0$ and $\lambda_{\ell j}^* \ge 0$ be the Lagrange multipliers corresponding to the Υ_{0j} and Λ_{0j} constraints, respectively. Further, let $\gamma_j^* = \lambda_{uj}^* - \lambda_{\ell j}^*$, where we note that $\lambda_{uj}^*\lambda_{\ell j}^* = 0$ for all j because $\Upsilon_{0j} \ge \Lambda_{0j}$.[27] The bounds $D_0 = [\Upsilon_0^\intercal, \Lambda_0^\intercal]^\intercal$ belong to the parameter space $[0,1]^{2J}$. Below, we will partition $[0,1]^{2J}$ into finitely many areas in such a way that the signs of all the γ_j^*'s are the same for two points in the same area and at least one multiplier has a different sign for two points in different areas; this partition is unique. Thus, each area corresponds to a set of nonzero Lagrange multipliers.

Let $\gamma^* = [\gamma_1^*, \cdots, \gamma_J^*]^\intercal$ and make its dependence on D_0 explicit by writing $\gamma^* = \gamma^*(D_0)$. Further, let $K = (K_u, K_\ell)$ be a pair of disjoint sets such that $K_u \cup K_\ell \subseteq \{1, 2, \ldots, J\}$ and define

$$S_K = \left\{D \in [0,1]^{2J} : K_+(D) = K_u,\ K_-(D) = K_\ell,\ \forall j : \Lambda_j \le G_j^*(D) \le \Upsilon_j\right\}, \tag{14}$$

where $K_+(D) = \{j : \gamma_j^*(D) > 0\}$ and $K_-(D) = \{j : \gamma_j^*(D) < 0\}$. So, K_u, K_ℓ represent

[27] It would be more precise to say that there exist solutions for which $\lambda_{uj}^*\lambda_{\ell j}^* = 0$ because the Lagrange multipliers are not unique when $\Upsilon_{0j} = \Lambda_{0j}$.

binding upper and lower bounds, respectively, where we define *binding constraints* to be constraints whose Lagrange multipliers are nonzero. Please note that Lagrange multipliers can equal zero even if constraints hold with equality.[28] This is an important distinction as will become apparent in section 8.2.

The S_K sets are distinct and form a partition of $[0,1]^{2J}$ by construction. Consider the following example.

Example 2. Suppose that $J = 2$ and there are only upper bound constraints: i.e.

$$\min_{g_1,g_2,g_3} \{g_1 \log g_1 + g_2 \log g_2 + g_3 \log g_3 - (g_1+g_2+g_3)\log \Delta\}$$
$$\text{subject to} \begin{cases} g_1+g_2+g_3 = 1, \\ g_1 \le \Upsilon_{10}, \\ g_1+g_2 \le \Upsilon_{20}. \end{cases}$$

We first eliminate g_3 by using the equality constraint and focus on g_1, g_2. Then, the Karush–Kuhn–Tucker (KKT) conditions are

$$\begin{cases} \log g_1^* - \log(1-g_1^*-g_2^*) + \gamma_1^* + \gamma_2^* = 0, & \gamma_1^*(g_1^* - \Upsilon_{10}) = 0, & g_1^* \le \Upsilon_{10}, \\ \log g_2^* - \log(1-g_1^*-g_2^*) + \gamma_2^* = 0, & \gamma_2^*(g_1^*+g_2^* - \Upsilon_{20}) = 0, & g_1^*+g_2^* \le \Upsilon_{20}, \\ \gamma_1^* \ge 0, & \gamma_2^* \ge 0. \end{cases}$$

So, there are four cases:[29] $(\gamma_1^*, \gamma_2^*) \in \{(0,0),(+,0),(0,+),(+,+)\}$. Each case represents a

(γ_1^*, γ_2^*)	g_1^*	g_2^*	$D_0 = (\Upsilon_{10}, \Upsilon_{20}) \in$
$(0,0)$	$1/3$	$1/3$	$S_{(\emptyset,\emptyset)} = \{3\Upsilon_1 \ge 1,\ 3\Upsilon_2 \ge 2\}$
$(+,0)$	Υ_{10}	$(1-\Upsilon_{10})/2$	$S_{(\{1\},\emptyset)} = \{3\Upsilon_1 < 1, \Upsilon_1 + 1 \le 2\Upsilon_2\}$
$(0,+)$	$\Upsilon_{20}/2$	$\Upsilon_{20}/2$	$S_{(\{2\},\emptyset)} = \{3\Upsilon_2 < 2, \Upsilon_2 \le 2\Upsilon_1\}$
$(+,+)$	Υ_{10}	$\Upsilon_{20} - \Upsilon_{10}$	$S_{(\{1,2\},\emptyset)} = \{2\Upsilon_1 < \Upsilon_2, 2\Upsilon_2 < \Upsilon_1 + 1\}$

Table 1: Solutions for the case $J = 2$ with no lower bounds

unique set of constraints with nonzero multipliers and corresponds to a polygon in $[0,1]^2$ as shown in figure 13: in table 1 each such polygon is denoted by $S_{(K_u,\emptyset)}$ for some K_u.[30] If two polygons share a boundary then the boundary belongs to the polygon with more multipliers equal to zero. □

[28] For instance, if one minimizes x^2 subject to $x \le 0$ then the Lagrange multiplier equals zero but the constraint holds with equality, i.e. $x = 0$.

[29] For general J and with both upper and lower bounds, there are $\sum_{r=0}^{J} 2^r \binom{J}{r}$ different cases.

[30] Since there are no lower bound constraints in this example, $K_\ell = \emptyset$.

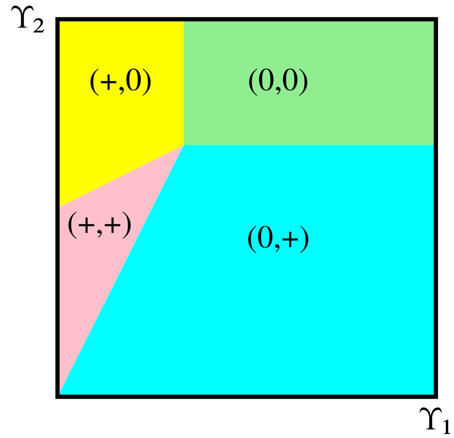


Figure 13: Graphical illustration of the S_K sets in table 1

In example 2, the partitioning sets S_K are polygons. As we show in lemma A.1 in appendix A, the S_K sets are always polyhedra, which is crucial for the directional differentiability of g^*. Further, it is also generally true that boundaries belong to the area with more multipliers equal to zero, a property that will prove useful for developing an inference procedure.

Recall that each of the S_K sets corresponds to a set of *binding constraints*. Therefore, if $D_0 \in S_K$ then the relevant constraints in (13) can be expressed as

$$R_K^\intercal g = D_{0K}, \tag{15}$$

where $R_K \in \mathbb{R}^{J \times \Xi}$ matrix and $D_{0K} = [\Upsilon_{0K_u}^\intercal, \Lambda_{0K_\ell}^\intercal]^\intercal \in \mathbb{R}^\Xi$ with Ξ the cardinality of $K_u \cup K_\ell$. Here, Υ_{0K_u} and Λ_{0K_ℓ} are the subvectors of Υ_0 and Λ_0 determined by the sets K_u, K_ℓ of indices, respectively: for a vector $D = [\Upsilon^\intercal, \Lambda^\intercal]^\intercal \in \mathbb{R}^{2J}$ and a set of indices $K = (K_u, K_\ell)$, the operation of finding the subvector $D_K = [\Upsilon_{K_u}^\intercal, \Lambda_{K_\ell}^\intercal]^\intercal$ will be denoted by the function ϕ (i.e. $D_K = \phi(D, K)$). Please note that R_K is a matrix of full column rank consisting of zeros and ones, unless $K_u = K_\ell = \emptyset$.[31]

Example 3. Again consider example 2, where there are four S_K sets. Since there are no lower bound constraints, we have $K_\ell = \emptyset$ in all four cases.

(a) $K = (\emptyset, \emptyset)$: R_K is void and no constraints are relevant.

(b) $K = (\{1\}, \emptyset)$: $R_K = [1, 0]$ and $D_{0K} = \Upsilon_{10}$.

(c) $K = (\{2\}, \emptyset)$: $R_K = [1, 1]$ and $D_{0K} = \Upsilon_{20}$.

(d) $K = (\{1, 2\}, \emptyset)$: $R_K = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$ and $D_{0K} = \begin{bmatrix} \Upsilon_{10} \\ \Upsilon_{20} \end{bmatrix}$. □

[31] If $K_u = K_\ell = \emptyset$ then R_K is void and all constraints evaporate.

We now derive an asymptotic expansion of $\hat{Z} = \sqrt{N}(\hat{g}^* - g^*)$, which will be the basis for our discussion in section 8.2. Let $K_0 = (K_{u0}, K_{\ell 0})$, where

$$K_{u0} = \{j : \gamma_j^*(D_0) > 0\} \quad \text{and} \quad K_{\ell 0} = \{j : \gamma_j^*(D_0) < 0\}. \tag{16}$$

So, $D_0 \in S_{K_0}$ by definition. Define $\hat{K}$ by replacing D_0 in the definition of K_0 with $\hat{D}$, so $\hat{D} \in S_{\hat{K}}$ by definition as well. Further, let $\Theta(d, K) = H^{-1} R_K \bigl(R_K^\intercal H^{-1} R_K\bigr)^{-1} \phi(d, K)$ when $K \neq (\emptyset, \emptyset)$ and $\Theta(d, K) = 0$ when $K = (\emptyset, \emptyset)$. We then have the following theorem.

Theorem 5. Suppose that assumption A holds. Then,

$$\hat{Z} = \begin{cases} \Theta(\Phi, K_0) + o_p(1), & \text{if } D_0 \in \operatorname{int}(S_{K_0}), \\ \Theta(\hat{\Omega}, \hat{K}) + o_p(1), & \text{if } D_0 \in \operatorname{bdr}(S_{K_0}), \end{cases}$$

where H is the Hessian of Q at g^*. □

Please note that theorem 5 does by itself not provide guidance for inference. For instance, we do not know whether or not $D_0 \in \operatorname{int}(S_{K_0})$. If D_0 is a boundary point of S_{K_0}, then $\hat{K}$ is random and depends on $\hat{\Omega}$ even in the limit.[32] We propose an inference procedure and establish its validity in section 8.2.

8.2 Inference: Our inference procedure detects automatically whether or not D_0 is an interior point. If D_0 is an interior point then our procedure is (asymptotically) similar, whereas if D_0 is a boundary point then our procedure is conservative. Our bound selection procedure is similar to what is used in the moment inequality literature (e.g. Andrews and Soares, 2010), but our inference procedure is different. Indeed, unlike Andrews and Soares (2010) we have an optimization problem with inequality constraints with estimated bounds that has a unique solution. Here, the optimization problem stems from our use of maximum entropy.

Recall that $K_0 = (K_{u0}, K_{\ell 0})$ are the sets of indices of the constraints that have nonzero multipliers at D_0: see (16). Recall from section 7.1 that we refer to K_0 as the set of *binding* constraints at D_0, but that there can be constraints that hold with equality that are not in K_0.

Indeed, define $K_0^* = (K_{u0}^*, K_{\ell 0}^*)$ with

$$K_{u0}^* = \bigl\{j : G_j^*(D_0) = \Upsilon_{0j}\bigr\} \quad \text{and} \quad K_{\ell 0}^* = \bigl\{j : G_j^*(D_0) = \Gamma_{0j}\bigr\}. \tag{17}$$

We then have the following lemma.

Lemma 8.1. $K_0 = K_0^*$ if and only if $D_0 \in \operatorname{int}(S_{K_0})$.[33] □

[32] For instance, suppose that in example 2, $D_0 = (\Upsilon_{10}, \Upsilon_{20}) = (1/3, 1/3) \in S_{(\emptyset,\emptyset)}$. Hence, $\gamma_1^*(D_0) = 0$. However, for any small $t > 0$, $\bigl\{d : \gamma_1^*(D_0 + td) = 0, \|d\| = 1\bigr\}$ and $\bigl\{d : \gamma_1^*(D_0 + td) > 0, \|d\| = 1\bigr\}$ are continua.

[33] By the KKT condition, we always have $K_0 \subseteq K_0^*$. So, $K_0 \subsetneq K_0^*$ if and only if $D_0 \in \operatorname{bdr}(S_{K_0})$.

Lemma 8.1 is a consequence of the continuity of the solution and the multipliers. The proof is provided in appendix E.

Below, we develop estimators $\tilde{K}$ and $\tilde{K}^*$ such that $\mathbb{P}(\tilde{K} \neq K_0) = o(1)$, $\mathbb{P}(\tilde{K}^* \neq K_0^*) = o(1)$, and $\mathbb{P}(\tilde{K} \subseteq \hat{K} \subseteq \tilde{K}^*) = 1$.[34] In view of theorem 5 and lemma 8.1, we propose simulating the distribution of $\hat{T} = T(\tilde{K}, \tilde{K}^*)$ for given values of $\tilde{K}$ and $\tilde{K}^*$, where

$$T(K^\circ, K^*) = \max_{K^\circ \subseteq K \subseteq K^*} \Theta(\Phi^*, K), \tag{18}$$

with Φ^* an independent copy of Φ. So, in each replication we are using a different draw Φ^* but the same estimates $\tilde{K}, \tilde{K}^*$ based on the original data. Let $T = T(K_0, K_0^*)$. Since $\tilde{K} = K_0$ and $\tilde{K}^* = K_0^*$ with probability approaching one, $\hat{T} = T$ with probability approaching one, also, so the distinction between $\hat{T}$ and T is moot for our asymptotic analysis. The quantiles of T provide an upper bound for the corresponding quantiles of $\hat{Z}$. Further, if D_0 is an interior point then $K_0 = K_0^*$ and hence $T = \Theta(\Phi^*, K_0)$ has the same distribution as $\Theta(\Phi, K_0) = \hat{Z} + o_p(1)$. Hence, the quantiles of T and $\hat{Z}$ coincide in the limit and inference is asymptotically similar.

Let $\tilde{K} = (\tilde{K}_u, \tilde{K}_\ell)$ and $\tilde{K}^* = (\tilde{K}_u^*, \tilde{K}_\ell^*)$, where

$$\begin{cases} \tilde{K}_u = \{j : \gamma_j^*(\hat{D}) \geq \kappa_N\}, & \tilde{K}_\ell = \{j : \gamma_j^*(\hat{D}) \leq -\kappa_N\}, \\ \tilde{K}_u^* = \{j : G_j^*(\hat{D}) \geq \hat{\Upsilon}_j - \kappa_N\}, & \tilde{K}_\ell^* = \{j : G_j^*(\hat{D}) \leq \hat{\Gamma}_j + \kappa_N\}, \end{cases}$$

with $0 < \kappa_N = o(1)$ and $1 = o(\kappa_N \sqrt{N})$. Input parameters like κ_N are much discussed in the moment inequality literature; the BIC choice $\kappa_N^2 = \log N \,/\, N$ is popular.

Lemma 8.2. Suppose that assumption A is satisfied. Then, *(a)* $\mathbb{P}(\tilde{K} \neq K_0) = o(1)$, *(b)* $\mathbb{P}(\tilde{K}^* \neq K_0^*) = o(1)$, *(c)* $\mathbb{P}(\tilde{K} \subseteq \hat{K} \subseteq \tilde{K}^*) = 1$. □

Lemma 8.2, together with theorem 5, provides the basis for using (18) for inference. The following theorem formalizes the idea.

Theorem 6. Suppose that assumption A is satisfied. Then $\mathbb{P}(\hat{T} \neq T) = o(1)$. Further, for any $x \in \mathbb{R}^J$, $\mathbb{P}(\hat{Z} \leq x) \geq \mathbb{P}(T \leq x) + o(1)$, where the inequality holds with equality whenever $D_0 \in \text{int}(S_{K_0})$. □

9. Inference on expected revenue and optimal reserve price

We now consider estimation of and inference for the maximum attainable revenue and optimal reserve price corresponding to the maximum entropy solution for the value distribution. We build on our discussion in section 8.2.

[34] For pairs of sets $K = (K_u, K_\ell)$ and $K^* = (K_u^*, K_\ell^*)$, we write $K \subseteq K^*$ when the inclusion holds elementwise.

Consider

$$\mathcal{P}(g^*) = \max_{r \in [0,1]} \pi(r, g^*) \quad \text{and} \quad \mathcal{R}(g^*) = \operatorname{argmax}_{r \in [0,1]} \pi(r, g^*), \tag{19}$$

where $\pi(r, g) = \tilde{\pi}\{r, \mathcal{F}^*(\cdot, g)\}$: $\tilde{\pi}(r, \mathcal{F})$ and $\mathcal{F}^*(v, g)$ were introduced in sections 2 and 3, respectively. So, $\pi(r, g^*)$ is the maximum entropy expected profit function. $\mathcal{P}(g^*)$ can be estimated by $\mathcal{P}(\hat{g}^*)$ but $\mathcal{R}(\hat{g}^*)$ need not be a consistent estimator of $\mathcal{R}(g^*)$, albeit that $\mathcal{R}(\hat{g}^*)$ is contained in $\mathcal{R}(g^*)$ with probability approaching one. So if the sole objective is to select an optimal reserve price, then selecting an element from $\mathcal{R}(\hat{g}^*)$ suffices.

Below we construct an estimator of $\mathcal{R}(g^*)$ and determine its properties. Once we establish the limiting distribution of the estimator $\hat{r}_1$ of the smallest element in $\mathcal{R}(g^*)$, $\hat{\mathcal{P}}^* = \pi(\hat{r}_1, \hat{g}^*)$ is an estimator of $\mathcal{P}(g^*)$ which is more convenient for inference purposes than $\mathcal{P}(\hat{g}^*)$.

Although $\mathcal{R}(g^*)$ need not be a singleton, it is at most finite. The reason is that, for any $g > 0$, the function $\theta(\cdot, g)$ with

$$\theta(r, g) = \theta^+(r, g) = \frac{\partial_r^+ \tilde{\pi}\{r, \mathcal{F}^*(r, g)\}}{n\{\mathcal{F}^*(r, g)\}^{n-1}}, \tag{20}$$

where ∂_r^+ denotes a right derivative, is piecewise linear in r with negative slope coefficients.[35] Indeed, for all g,

$$\theta(r, g) = 1 - M_{j-1}(g) + (j-1)g_j - 2g_j r \,/\, \Delta \quad \text{if } r \in I_j, \tag{21}$$

where $M_j(g) = G_j = \sum_{k=1}^{j} g_k$ and $M_0(g) = 0$. Note that θ is possibly discontinuous at the β_j's, i.e. on the support of the bid distribution, because $\theta(\beta_j, g)$ need not equal $\theta^-(\beta_j, g)$ for any or all j, which is defined as θ in (21) but with ∂_r^+ replaced with the left derivative ∂_r^-.

Lemma 9.1. For some integer $1 \le m < \infty$ and some $0 < r_1^* < r_2^* < \cdots < r_m^* < 1$, $\mathcal{R}(g^*) = \{r_1^*, r_2^*, \cdots, r_m^*\}$. Further, each $\bar{I}_j = [\beta_{j-1}, \beta_j]$ contains at most one element of $\mathcal{R}(g^*)$. □

We show that both the number and the identity of the r_j^*–values can be estimated consistently: the rate at which m is estimated is arbitrarily fast. The proof is simple. Let $\hat{\mathcal{R}}_N = \{r \in [0,1] : \pi(r, \hat{g}^*) \ge \mathcal{P}(\hat{g}^*) - \kappa_n\}$. Since $\pi(\cdot, \hat{g}^*)$ is continuous, $\hat{\mathcal{R}}_N$ is a compact subset of $[0,1]$ by construction. Consistency is obtained in a similar manner as in Chernozhukov, Hong, and Tamer (2007), albeit that here the identified set is known to be a collection of isolated points.

Theorem 7. $d_H\{\hat{\mathcal{R}}_N, \mathcal{R}(g^*)\} = o_p(1)$, where d_H denotes the Hausdorff distance. □

Note that $\hat{\mathcal{R}}_N$ is set–valued. We can use it to create point estimates of each of the r_j^*–values as follows. Let $\tilde{\kappa}_N = \sqrt[3]{\kappa_N}$, $\hat{\mathcal{R}}_{N,1} = \{r \in \hat{\mathcal{R}}_N : r - \min \hat{\mathcal{R}}_N \le \tilde{\kappa}_N\}$, $\hat{\mathcal{R}}_{N,2} = \{r \in \hat{\mathcal{R}}_N \setminus \hat{\mathcal{R}}_{N,1} :$

[35] Note that (2) and (20) are identical except that we now do not implicitly assume the existence of the partial derivative and only focus on distribution functions of the form $\mathcal{F}^*(\cdot, g)$.

$r - \min(\hat{\mathcal{R}}_N \setminus \hat{\mathcal{R}}_{N,1}) \le \tilde{\kappa}_N\}$, etcetera: so, $\hat{\mathcal{R}}_N = \cup_{k=1}^{\hat{m}} \hat{\mathcal{R}}_{N,k}$ for some estimator $\hat{m}$ of m. We define our estimator of r_k^* by

$$\hat{r}_k = \min \operatorname*{argmax}_{r \in \hat{\mathcal{R}}_{N,k}} \pi(r, \hat{g}^*).$$

Our definition of $\hat{r}_k$ allows for the possibility that the maximizer of $\pi(r, \hat{g}^*)$ in $\hat{\mathcal{R}}_{N,k}$ is not unique. Consistency of $\hat{r}_k$ for r_k^* is straightforward to establish.

Theorem 8. $\mathbb{P}(\hat{m} \neq m) = o(1)$ and for all $k = 1, \dots, m$, $\hat{r}_k - r_k^* = O_p(\sqrt{\kappa_N})$.[36] □

As we will show below, the convergence rate of $\hat{r}_k$ is in fact better than the rate obtained in theorem 8.

We now turn to the construction of confidence intervals and focus on r_1^*: the arguments are analogous for $r_2^*, \dots, r_m^*$ if $m > 1$. Note that r_1^* is the unique maximizer of $\pi(r, g^*)$ on some compact interval A_1: such an interval exists by lemma 9.1. Define $r_1(g) = \min \tilde{R}_1(g)$, where

$$\tilde{R}_1(g) = \operatorname*{argmax}_{r \in A_1} \pi(r, g). \tag{22}$$

So the difference between $r_1(\hat{g}^*)$ and $\hat{r}_1$ is that the maximization is conducted over a(n asymptotically) larger set in the former case: A_1 versus $\hat{\mathcal{R}}_{N,1}$. However, the difference is minute from a theoretical perspective.

Lemma 9.2. $\mathbb{P}\{r_1(\hat{g}^*) \neq \hat{r}_1\} = o(1)$. □

Typically, $\tilde{R}_1(g)$ is a singleton for all g in a small enough neighborhood of g^*, but there is one exceptional case in which $\tilde{R}_1(g)$ has two elements for some g near g^*.[37]

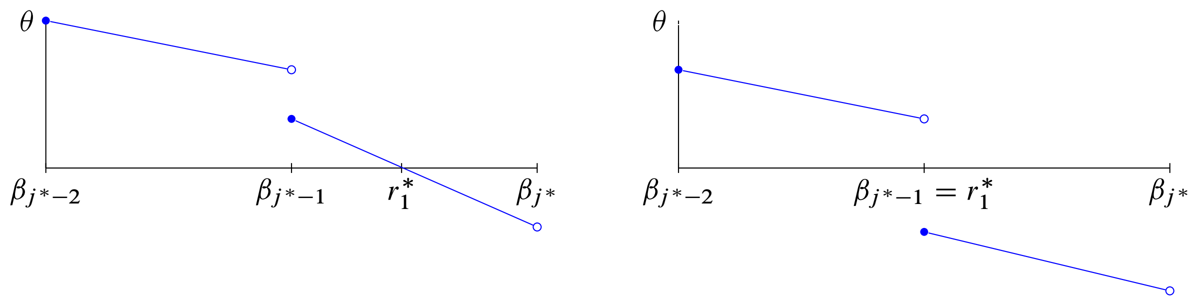
In view of lemma 9.2 we consider the asymptotic distribution of $\sqrt{N}\{r_1(\hat{g}^*) - r_1(g^*)\}$. If the function r_1 is (Hadamard–) differentiable at g^* then the delta method will apply. Unfortunately, as we commented earlier, r_1 may fail to be Hadamard differentiable. However, even in that case, small perturbations on g^* have only limited effects on the function r_1, which is sufficient for our purpose. Below we discuss the Hadamard derivatives of r_1, which requires us to analyze the first order condition of (22).

Recall that $\theta(\cdot, g^*)$ is piecewise linear but not necessarily continuous. So we distinguish between the case in which $r_1(g^*) = r_1^*$ is a continuity point of θ and the case in which it is not. Let j^* be such that $r_1^* \in I_{j^*}$.

Since θ is linear and downward sloping on I_{j^*}, there are two possibilities, which are illustrated in figure 14. First, if $r_1^* > \beta_{j^*-1}$ then $\theta\{r_1(g^*), g^*\} = 0$, in which case the derivative of r_1^* at g^*

[36] If $\hat{m} < m$ then $\hat{r}_{\hat{m}+1}, \dots, \hat{r}_m$ can be defined arbitrarily.

[37] The exception arises when $\tilde{r}_1(g^*) = \beta_{j^*-1}$ for some j^* *and* $\theta^+(\beta_{j^*-1}, g^*) = \theta^-(\beta_{j^*-1}, g^*) = 0$. Then Hadamard differentiability fails and there can be two maximizers. Absent this technicality, $r_1(g)$ is the unique maximizer of $\pi(r, g)$ on A_1 in a neighborhood of g^*.



(a) Interior solution

(b) Boundary solution

Figure 14: Graphical illustration of the first order condition for r_1^*

can be obtained via the implicit function theorem. However, if $r_1^* = \beta_{j^*-1}$, then r_1^* may not be differentiable at g^* and there are then four separate (sub)cases to consider as shown in table 2.

case	condition
I	$\beta_{j^*-1} < r_1^* < \beta_{j^*}$;
II	$r_1^* = \beta_{j^*-1}$ with $\theta(r_1^*, g^*) < 0 < \theta^-(r_1^*, g^*)$;
III	$r_1^* = \beta_{j^*-1}$ with $\theta(r_1^*, g^*) = 0 < \theta^-(r_1^*, g^*)$;
IV	$r_1^* = \beta_{j^*-1}$ with $\theta(r_1^*, g^*) < 0 = \theta^-(r_1^*, g^*)$;
V	$r_1^* = \beta_{j^*-1}$ with $\theta(r_1^*, g^*) = 0 = \theta^-(r_1^*, g^*)$.

Table 2: Separate cases for the sensitivity analysis

In appendix F.2 we show that r_1 is Hadamard–differentiable in cases I–IV. Case V is the exceptional case mentioned earlier, but even in case V valid inference for r_1^* is possible.

Let $\delta, \delta^- \in \mathbb{R}^J$ with

$$\delta_k = -\frac{\Delta}{2g_{j^*}^*} \times \begin{cases} 1, & k < j^*, \\ j^* - 1, & k = j^*, \\ 0, & k > j^*. \end{cases} \qquad \delta_k^- = -\frac{\Delta}{2g_{j^*-1}^*} \times \begin{cases} 1, & k < j^* - 1, \\ j^*, & k = j^* - 1, \\ 0, & k > j^* - 1. \end{cases}$$

Further, for $v \in \mathbb{R}^J$, define

$$\psi_{g^*}(v) = \begin{cases} v^\intercal \delta & \text{in case I,} \\ 0 & \text{in case II,} \\ \max(0, v^\intercal \delta) & \text{in case III,} \\ \min(0, v^\intercal \delta^-) & \text{in case IV,} \\ \max(|v^\intercal \delta|, |v^\intercal \delta^-|) & \text{in case V,} \end{cases} \tag{23}$$

which is the Hadamard derivative of r_1 at g^* in the direction v in cases I–IV and an upper bound of the effect of perturbing g^* on r_1 in case V.

The first case in (23) is the most common: it arises with an interior solution ($r_1^* > \beta_{j^*-1}$), in which case $\theta(r_1^*, g^*) = 0 = \theta^-(r_1^*, g^*)$. However, this case is dramatically different from $\theta(r_1^*, g^*) = 0 = \theta^-(r_1^*, g^*)$ with $r_1^* = \beta_{j^*-1}$, i.e. case V. Case II is pictured in the right panel of figure 14: small changes in g^* do not affect r_1^*. The remaining two cases in (23) arise if one of the two line segments in the right panel of figure 14 has an endpoint at $(\beta_{j^*-1}, 0)$.

We now characterize the asymptotic distribution of $\hat{\tau} = \sqrt{N}(\hat{r}_1 - r_1^*)$. Let Θ be the dominant right hand side term in theorem 5, i.e. depending on the location of D_0 either $\Theta(\Phi, K_0)$ or $\Theta(\hat{\Omega}, \hat{K})$.

Theorem 9. If assumption A is satisfied then in cases I–IV, we have $\hat{\tau} = \psi_{g^*}(\Theta) + o_p(1)$, and in case V, $|\hat{\tau}| \le \psi_{g^*}(\Theta) + o_p(1)$. □

Now consider revenue. Recall that $\mathcal{P}(g^*) = \pi(r_1^*, g^*)$ is estimated by $\hat{\mathcal{P}}^* = \pi(\hat{r}_1, \hat{g}^*)$.

Theorem 10. If assumption A is satisfied then $\sqrt{N}\{\hat{\mathcal{P}}^* - \mathcal{P}(g^*)\} = \partial_g \pi(r_1^*, g^*)^\intercal \Theta + o_p(1)$. □

Recall that the function ψ_{g^*} depends on the case. For instance, in case II, $\hat{r}_1$ is superconsistent. In fact, convergence in this case can be shown to be arbitrarily fast. Therefore, inference on r_1 requires that we know which of the five cases we are in. This is not hard to do because theorem 9 shows that the rate of $\hat{r}_1$ cannot be worse than $\sqrt{N}$ and hence we can find out which case is relevant with probability approaching one. Therefore, we can rely on theorem 6 to conduct inference on r_1^*.

For example, suppose that $\hat{r}_1 \in (\beta_{\hat{j}^*-1} + \kappa_N, \beta_{\hat{j}^*} - \kappa_N)$ for some $\hat{j}^*$. We can then conduct inference for r_1^* by using the distribution of $\delta^\intercal \Theta$ for which theorem 6 can be applied. Inference will be conservative in case V but case V is extreme.

Theorem 10 does not distinguish case V from the other four cases. In fact, $\hat{g}^*$ is the only relevant object for the limit distribution of $\pi(\hat{r}_1, \hat{g}^*)$. This phenomenon is an implication of the envelope theorem. Indeed, in cases I and V, both the right and left derivatives of $\pi(\cdot, g^*)$ at r_1^* are zero. In case II, neither of the directional derivatives is zero but $\sqrt{N}(\hat{r}_1 - r_1^*)$ is asymptotically negligible. Similar arguments apply to cases III and IV, also. Therefore, inference on maximum revenue is straightforward: we can simply use theorem 6 without knowing which case is relevant.

Appendices

A. Lemmas for the solution

Proof of lemma 3.1: Fix any $j^\circ \in \{1, 2, \cdots, J+1\}$ and let $\mathcal{A} = \big\{\ell : \forall s \notin I_{j^\circ} : \ell(s) = \ell^*(s)\big\}$. Then f^* is not only a solution to (3), but since $\mathcal{F}_L, \mathcal{F}_U$ are flat on I_{j° also to

$$\min_{f \in \mathcal{A}} \int_{\beta_{j^\circ-1}}^{\beta_{j^\circ}} \ell(s) \log \ell(s)\, \mathrm{d}s \quad \text{s.t.} \quad \begin{cases} \displaystyle\int_{\beta_{j^\circ-1}}^{\beta_{j^\circ}} \ell(s)\, \mathrm{d}s = c_{j^\circ}, \\ \forall v \in I_{j^\circ} : \mathcal{F}_L(\beta_{j^\circ-1}) \le \underline{\mathrm{S}} + \displaystyle\int_{\beta_{j^\circ-1}}^{v} \ell(s)\, \mathrm{d}s \le \mathcal{F}_U(\beta_{j^\circ}-), \\ \mathcal{F}_L(\beta_{j^\circ}) \le \underline{\mathrm{S}} + \displaystyle\int_{\beta_{j^\circ-1}}^{\beta_{j^\circ}} \ell(s)\, \mathrm{d}s \le \mathcal{F}_U(\beta_{j^\circ}-) \end{cases}$$

where $c_j = \int_{\beta_{j-1}}^{\beta_j} \ell^*(s)\, \mathrm{d}s$, $\underline{\mathrm{S}} = \sum_{j=1}^{j^*-1} c_j$, and $\mathcal{F}_U(\beta_{j^*}-) = \lim_{v \uparrow \beta_{j^*}} \mathcal{F}_U(v)$: the last inequality constraint is a condition coming from the fact that $\mathcal{F}(\cdot)$ is a continuous function.

Because ℓ^* satisfies the inequality constraints, we must have $\mathcal{F}_L(\beta_{j^\circ-1}) \le \underline{\mathrm{S}} \le \underline{\mathrm{S}} + c_{j^\circ} \le \mathcal{F}_U(\beta_{j^\circ}-)$ and $\mathcal{F}_L(\beta_{j^\circ}) \le \underline{\mathrm{S}} + c_{j^\circ}$, and therefore the inequality constraints are redundant. Hence f^* is constant on I_{j°. □

The Karush–Kuhn–Tucker (KKT) conditions for minimizing (12) subject to (13) are for $j = 1, \ldots, J$ given by

$$\begin{cases} \log g_j^* - \log(1 - G_J^*) + \sum_{k=j}^{J} (\lambda_{uk}^* - \lambda_{\ell k}^*) = \lambda_{sj}^*, & \text{(24a)} \\ \lambda_{uj}^* (G_j^* - \Upsilon_{0j}) = 0, \quad G_j^* \le \Upsilon_{0j}, \quad \lambda_{uj}^* \ge 0, & \text{(24b)} \\ \lambda_{\ell j}^* (G_j^* - \Lambda_{0j}) = 0, \quad G_j^* \ge \Lambda_{0j}, \quad \lambda_{\ell j}^* \ge 0, & \text{(24c)} \\ \lambda_{sj}^* g_j^* = 0, \quad g_j^* \ge 0, \quad \lambda_{sj}^* \ge 0. & \text{(24d)} \end{cases}$$

Recall that $\gamma_j^* = \lambda_{uj}^* - \lambda_{\ell j}^*$, where $\lambda_{uj}^* \lambda_{\ell j}^* = 0$. Here, if $g_j^* = 0$, then the conditions in (24a) and (24d) cannot be satisfied simultaneously. Thus $g_j^* > 0$ and $\lambda_{sj}^* = 0$ for all j.

Lemma A.1. The solution g^* depends on which constraints are binding but is otherwise an affine function of Υ_0, Λ_0. Therefore, each of the S_K sets defined in (14) is a polyhedron.

Proof. Note that $\lambda_{uj}^* \lambda_{\ell j}^* = 0$ for all j because $\Upsilon_{0j} \ge \Lambda_{0j}$.[38] So, the conditions in (24a) to (24c)

[38] It would be more precise to say that there exist solutions for which $\lambda_{uj}^* \lambda_{\ell j}^* = 0$ because the Lagrange multipliers are not unique when $\Upsilon_{0j} = \Lambda_{0j}$.

contain $2J$ unknowns: $g_1^*, \dots, g_J^*$ and $\gamma_1^*, \dots, \gamma_J^*$. We need to solve

$$\begin{cases} \log g_j^* - \log(1 - G_J^*) + \sum_{k=j}^{J} \gamma_k^* = 0, & \text{(25a)} \\ \gamma_j^*(G_j^* - B_{j0}) = 0, & \text{(25b)} \end{cases}$$

for $j = 1, \dots, J$, where B_{j0} is either Υ_{0j} or Λ_{0j}. The conditions in (25a) imply that

$$\gamma_j^* = \log g_{j+1}^* - \log g_j^*, \qquad j = 1, 2, \cdots, J, \tag{26}$$

where $g_{J+1}^* = 1 - G_J^*$. Suppose that there are r multipliers γ_j^* that equal zero and $J - r$ that are nonzero. For $\gamma_j^* \neq 0$, by (25b) we have $G_j^* = B_{j0}$. For $\gamma_j^* = 0$, (26) implies that $g_{j+1}^* = g_j^*$ with $1 - G_J^* = g_J^*$ as a special case. Therefore, $g^* = [g_1^*, \cdots, g_J^*]^\intercal$ is the solution to a linear equation system whose right hand side is linear in the B_{j0}'s. □

Lemma A.2. The solutions g^*, γ^* are continuous functions of Υ_0, Λ_0.

Proof. It suffices to show the continuity of g^*: the continuity of γ^* then follows from (26). The solution $g^* = [g_1^*, \cdots, g_J^*]^\intercal$ minimizes Q (defined in (12)) subject to $g \in \Xi(\Upsilon_0, \Lambda_0)$, where Ξ is the correspondence $\Xi(\Upsilon_0, \Lambda_0) = \{g : \Lambda_{0j} \leq \sum_{k=1}^{j} g_k \leq \Upsilon_{0j} \quad \text{for } j = 1, 2, \cdots, J\}$. Since Q is a continuous function and Ξ is a continuous correspondence, it follows from the maximum theorem that $g^* = g^*(\Upsilon_0, \Lambda_0)$ is upper hemicontinuous as a correspondence. Further, by the convexity of the problem, $g^*(\Upsilon_0, \Lambda_0)$ is a single element correspondence, i.e. a function, and therefore upper hemicontinuity is equivalent to continuity. □

Proof of theorem 4: Follows from lemma A.2 and the continuous mapping theorem. □

B. Sensitivity of the solution

In this section we consider the effect on the solution g^* of perturbations of Υ_0, Λ_0 in a given direction d. We will use these results for statistical inference.

Recall that $\mathcal{S} = \{S_K\}$ is a (finite) partition of $[0, 1]^{2J}$, where S_K is defined in (14). As we discussed in section 7.1, the solution $g^*(D)$ at $D = [\Upsilon^\intercal, \Lambda^\intercal]^\intercal \in S_K$ can be expressed as

$$g^*(D) = \underset{g}{\operatorname{argmin}} \; Q(g) \quad \text{subject to} \quad R_K^\intercal g = D_K, \tag{27}$$

where R_K is a matrix of ones and zeroes with full column rank and $D_K = [\Upsilon_{K_u}, \Lambda_{K_\ell}]$ is a subvector of D that is determined by $K = (K_u, K_\ell)$; if none of the constraints in S_K are binding (i.e. $K_u = K_\ell = \emptyset$) then the restrictions in (27) evaporate.

Now, suppose that we perturb a given $D_0 = (\Upsilon_0^\intercal, \Lambda_0^\intercal)^\intercal \in S_{K_0}$ in the direction d, where S_{K_0} is implicitly defined. So we consider $D_0 + td$, where d is given and $t > 0$ is small. The most important insight is that for all sufficiently small $t > 0$, $D_0 + td$ lies within the set S_{K_d} which only depends on d. The following lemma formalizes this idea.

Lemma B.1. There exist an $S_{K_d} \in \mathcal{S}$ and an $\epsilon > 0$ such that $D_0 + td \in S_{K_d}$ for all $0 < t < \epsilon$.

Proof. If D_0 is in the interior of S_{K_0}, then the assertion is true with $S_{K_d} = S_{K_0}$. Suppose that D_0 is on the boundary of S_{K_0}. By lemma A.1 all S_K sets in $\mathcal{S}$ are polyhedra, and therefore there are only four possibilities: for a sufficiently small $\epsilon > 0$, the (open) line segment $\{D_0 + td : 0 < t < \epsilon\}$ *(i)* is a subset of the boundary of S_{K_0}; *(ii)* is a subset of the boundary of some $S_K \neq S_{K_0}$; *(iii)* belongs to the interior of S_{K_0}; *(iv)* belongs to the interior of some $S_K \neq S_{K_0}$. □

By lemma B.1, for all sufficiently small $t > 0$, the solution at $D_0 + td$ is given by

$$g^*(D_0 + td) = \operatorname*{argmin}_g \; Q(g) \quad \text{subject to} \quad R_{K_d}^\intercal g = D_{0K_d} + td_{K_d}, \tag{28}$$

where D_{0K_d}, d_{K_d} are the subvectors of D_0, d corresponding to the indices in $K_d = (K_{du}, K_{d\ell})$, as described in the paragraph after (15). The formulation in (28) is convenient for obtaining directional derivatives of g^*.

Now, note that

$$\lim_{t \downarrow 0} g^*(D_0 + td) = \operatorname*{argmin}_g \; Q(g) \quad \text{subject to} \quad R_{K_d}^\intercal g = D_{0K_d}. \tag{29}$$

Since g^* is continuous by lemma A.2, (29) is equivalent to

$$g^*(D_0) = \operatorname*{argmin}_g \; Q(g) \quad \text{subject to} \quad R_{K_0}^\intercal g = D_{0K_0}, \tag{30}$$

where D_{0K_0} is a subvector of D_0 defined b $K_0 = (K_{0u}, K_{0\ell})$ as in (15).[39]

So the directional derivative $\nabla g^*(D_0, d)$ of g^* at D_0 in the direction d can be obtained by differentiating (28) with respect to t.[40] Recall H be the Hessian of the objective function Q at D_0.

Lemma B.2. *(i)* If $K_d = (\emptyset, \emptyset)$ then $\nabla g^*(D_0, d) = 0$. *(ii)* Otherwise, $\nabla g^*(D_0, d) = H^{-1} R_{K_d} \left(R_{K_d}^\intercal H^{-1} R_{K_d} \right)^{-1} d_{K_d}$.

[39] Although S_{K_0} and S_{K_d} need not be the same when D_0 is a boundary point, equivalence of (29) and (30) is intuitive. Indeed, suppose that D_0 is a boundary point of S_{K_0} and $D_0 + td \in S_{K_d}$ for all sufficiently small $t > 0$, where $K_0 \neq K_d$. Now, $K_0 \subset K_d$ because $\gamma_j^*(\cdot)$ is continuous in view of lemma A.2. Thus, any constraints that are binding at $D_0 + td$ become *redundant* in the limit.

[40] So, for $\bar{g}^*(t) = g^*(D_0 + td)$, $\nabla g^*(D_0, d) = \partial_t \bar{g}^*(0)$.

Proof. Since part *(i)* is trivial, we prove part *(ii)*. Note that H is positive definite: the typical element h_{ij} is given by $h_{ij} = \{\mathbb{1}(i = j) \,/\, g_i^*\} + (1 \,/\, g_{J+1}^*)$. The first order conditions of (28) are $\partial_g Q\{g^*(D_0 + td)\} = R_{K_d}\mu(t)$ and $R_{K_d}^\intercal g^*(D_0 + td) = D_{K_d} + td_{K_d}$, where $\mu(t)$ is the vector of the Lagrange multipliers. By differentiating with respect to t at 0, we obtain $H\nabla g^*(D_0, d) = R_{K_d}\partial_t\mu(0)$ and $R_{K_d}^\intercal \nabla g^*(D_0, d) = d_{K_d}$, which implies that $d_{K_d} = R_{K_d}^\intercal H^{-1} R_{K_d}\partial_t\mu(0)$. Therefore, the assertion follows. □

Lemma B.3. γ^* is directionally differentiable at D_0. If $K_d = (\emptyset, \emptyset)$ then $\nabla\gamma_j^*(D_0, d) = 0$. Otherwise, $\left|\nabla\gamma_j^*(D_0, d)\right| \leq \left\|\left(R_{K_d}^\intercal R_{K_d}\right)^{-1} H\nabla g^*(D_0, d)\right\|$.

Proof. By continuity of γ_j^*, there are three relevant cases: *(i)* $\gamma_j^*(D_0) = 0$ and $\gamma_j^*(D_0 + td) = 0$; *(ii)* $\gamma_j^*(D_0) = 0$ and $\gamma_j^*(D_0 + td) \neq 0$; *(iii)* $\gamma_j^*(D_0) \neq 0$ and $\gamma_j^*(D_0 + td) \neq 0$. In the first case, $\nabla\gamma_j^*(D_0, d) = 0$. In the other two cases, $\nabla\gamma_j^*(D_0, d)$ is an element of $\partial_t\mu(0)$ in the proof of lemma B.2. □

Proof of theorem 5: We first show that $\hat{Z} = \nabla g^*\big(D_0, \hat{d} \,/\, \|\hat{d}\|\big)\sqrt{N}\|\hat{d}\| + o_p(1)$, where $\hat{d} = \hat{D} - D_0$. Let for any distance $\rho > 0$ and direction a with $\|a\| = 1$,

$$\xi_N(a, \rho) = \sqrt{N}\big\{g^*\big(D_0 + \rho a \,/\, \sqrt{N}\big) - g^*(D_0)\big\} - \nabla g^*(D_0, a)\rho.$$

From lemma B.2, we know that $\xi_N(a, \rho) = o(1)$ for any a, ρ. Now, for any measure ζ and any $\epsilon > 0$, $\iint \mathbb{1}\big\{\|\xi_N(a, \rho)\| > \epsilon\big\}\, \mathrm{d}\zeta(a, \rho) = o(1)$, by the dominated convergence theorem. Let $\hat{\rho} = \sqrt{N}\|\hat{d}\|$ and $\hat{a} = \sqrt{N}\hat{d} \,/\, \hat{\rho}$. Take ζ to be the distribution of $(\hat{a}, \hat{\rho})$. Therefore, the boundary case is proved since $\hat{D} \in S_{\hat{K}}$ is always true by definition. For the interior case, it follows from the fact that $D_0 \in \mathrm{int}(S_{K_0})$ implies $\hat{D} \in S_{K_0}$ with probability approaching one. □

C. Hybrid problem — solution

Proof of theorem 2: The first order conditions to (10) are given by

$$2r\mathcal{F}(r)\mathbb{1}(v \leq r) + 2\int_{\max(v,r)}^{1} \bar{\mathcal{F}}(s)\, \mathrm{d}s + \alpha\big\{\log f(v) + 1\big\} - \sum_{j=1}^{J}(\tilde{\gamma}_j - \breve{\gamma}_j)\mathbb{1}(v \leq \beta_j) + \tilde{\lambda} = 0, \quad (31)$$

and

$$\begin{cases} -\displaystyle\int_0^1 f(v)\log f(v)\, \mathrm{d}v = \mathcal{E}^*, & \\ \tilde{\gamma}_j\Big(\displaystyle\int_0^{\beta_j} f(s)\, \mathrm{d}s - \Upsilon_{0j}\Big) = 0, & j = 1, \ldots, J, \\ \breve{\gamma}_j\Big(\displaystyle\int_0^{\beta_j} f(s)\, \mathrm{d}s - \Lambda_{0j}\Big) = 0, & \end{cases} \quad (32)$$

where $\tilde{\gamma}_j, \breve{\gamma}_j \geq 0$ and $\alpha < 0$: α cannot be zero since then $\mathcal{F}_U$ minimizes the profit which implies $\mathcal{E}^* = -\infty$. Therefore, for $k = 1, 2, \ldots, J$ and $\beta_{k-1} < v \leq \beta_k$, we have

$$2r\mathcal{F}(r)\mathbb{1}(v \leq r) + 2\int_{\max(v,r)}^{1} \bar{\mathcal{F}}(s)\,\mathrm{d}s + \alpha\{\log \ell(v) + 1\} - \sum_{j=k}^{J}(\tilde{\gamma}_j - \breve{\gamma}_j) + \tilde{\lambda} = 0. \tag{33}$$

Now, for $v \leq r$, the solution ℓ to (33) is piecewise constant. For $v > r$, differentiating (33) yields

$$2\bar{\mathcal{F}}(v) - \alpha\frac{\ell'(v)}{\ell(v)} = 0, \tag{34}$$

whenever $\ell(v) > 0$. Therefore, there is no loss of generality in merging $(\beta_{k-1}, \beta_k]$ and $(\beta_k, \beta_{k+1}]$ when the distribution bound constraints are not holding with equality at β_k. Below we focus on $(\beta_{k-1}, \beta_k]$, where we have $\tilde{\gamma}_{k-1} > 0$ and $\tilde{\gamma}_k > 0$, i.e. the (upper) bounds at the two end points are binding; binding lower bounds are dealt with analogously.

If $\ell(v) = 0$ at any $v \in (\beta_{k-1}, \beta_k]$ then there must be some $\tilde{\beta}_k \in (\beta_{k-1}, \beta_k]$ such that $\ell(v) = 0$ for all $\tilde{\beta}_k < v \leq \beta_k$: putting $\ell(v) = 0$ anywhere other than on the right side of $(\beta_{k-1}, \beta_k]$ would not minimize profit. So (34) must be satisfied for all $\beta_{k-1} < v \leq \tilde{\beta}_k$. The proposed parametric form of ℓ is a solution to the second order differential equation in (34).

To see that the final statement of the theorem is true consider the following. By the definition of k_r we start with $3(J + 2 - k_r)$ unknowns: one set $\{\omega_j, \nu_j, \tilde{\beta}_j\}$ for each interval. We have $\bar{\mathcal{F}}(r) = \bar{Z}$ and $\bar{\mathcal{F}}(1) = 0$ as two binding equalities. Suppose that there are no binding constraints at $\beta_{k_r}, \beta_{k_{r+1}}, \ldots, \beta_J$. Then $\tilde{\beta}_j = \beta_j$, $\omega_j = \omega_{j+1}$, and $\nu_j = \nu_{j+1}$ for all $j \geq k_r$. That gives us $2 + (J + 2 - k_r) + 2(J + 1 - k_r) = 3(J + 2 - k_r)$ equalities, as desired.[41] If a constraint is binding at some β_j then we have two intervals with binding endpoint constraints and the above argument can be repeated for each interval separately. For large negative values of $\mathcal{E}^*$ it can happen that there are no values of ω_j, ν_j to make both bounds at the endpoints of the interval binding in which case $\tilde{\beta}_j$ is set to the largest value for which both bounds can be made binding and ω_j, ν_j are then the corresponding solutions. □

D. Hybrid problem — continuity

We now turn to the proof of theorem 3. Let $\ell = 1, \ldots, \bar{L}$ denote each of the cases in which the identity of the inequality constraints that hold with equality is given. Define $\pi_\ell^\circ(\bar{Z}, \alpha, \mathcal{E}^*, D_0; r)$ to

[41] The equalities are nonlinear, but that is a minor nuisance. For example, consider the case where $0 < r < \beta_1$ so that $k_r = 1$. Then we must have $\omega_1 = \omega_2 = \cdots = \omega_{J+1}$ and $\nu_1 = \cdots = \nu_{J+1}$, which leads to the following two equations: $\alpha\omega_1 \tan\{(\nu_1 - r)\omega_1\} = -\bar{Z}$ and $\alpha\omega_1 \tan\{(\nu_1 - 1)\omega_1\} = 0$. Our computations have not revealed invertibility problems.

be the minimum revenue for a given r as a function of $\bar{Z} = \bar{\mathcal{F}}(r), \alpha, \mathcal{E}^*$, and D_0. So,

$$\pi^\bullet(\mathcal{E}^*, D_0; r) = \min_{\ell=1,2,\cdots,\bar{L}} \left\{ \min_{(\bar{Z},\alpha)\in\mathcal{C}_\ell(\mathcal{E}^*,D_0)} \pi_\ell^\circ(\bar{Z}, \alpha, \mathcal{E}^*, D_0; r) \right\}, \tag{35}$$

where $\mathcal{C}_\ell$ is a correspondence defined by

$$\mathcal{C}_\ell(\mathcal{E}^*, D_0) = \left\{(\bar{Z}, \alpha) : \text{ entr}_\ell(\bar{Z}, \alpha, D_0) \geq \mathcal{E}^*, \Lambda_{0j} \leq \mathcal{F}_\ell(\beta_j; \bar{Z}, \alpha) \leq \Upsilon_{0j}\right\}, \tag{36}$$

where $\mathcal{F}_\ell(\beta_j; \bar{Z}, \alpha)$ is the distribution function that has the parametric form described in theorem 2 after partialing out $\nu, \omega, \tilde{\beta}$ and entr_ℓ is the entropy level corresponding to $\mathcal{F}_\ell(\cdot; \bar{Z}, \alpha)$. So $\mathcal{F}_\ell(\beta_j; \bar{Z}, \alpha) = \mathcal{F}\left\{\beta_j; \alpha, \nu_\ell(\bar{Z}, \alpha), \omega_\ell(\bar{Z}, \alpha), \tilde{\beta}_\ell(\bar{Z}, \alpha)\right\}$, where ν_ℓ, ω_ℓ, and $\tilde{\beta}_\ell$ are defined by recursively solving for ν, ω, and $\tilde{\beta}$ for given values of $\bar{Z}, \alpha$, and the identity ℓ of the bound constraints that hold with equality as explained in appendix C. Note that $\mathcal{F}_\ell(\beta_j; \cdot, \cdot)$ is a continuous function for all j. The function $\text{entr}_\ell(\cdot, \cdot, \cdot)$ can similarly be shown to be continuous.

Lemma D.1. For any $\mathcal{E}^* \in [\underline{\mathcal{E}}^*, \bar{\mathcal{E}}^*]$ and D_0, $\mathcal{C}_\ell(\mathcal{E}^*, D_0)$ is compact.

Proof. Since $\bar{Z}$ is a probability and α is a Lagrange multiplier, $\mathcal{C}_\ell(\mathcal{E}^*, D_0)$ is a bounded set. The closedness of $\mathcal{C}_\ell(\mathcal{E}^*, D_0)$ follows from continuity of $\text{entr}_\ell(\cdot, \cdot, \cdot)$ and $\mathcal{F}_\ell(\beta_j; \cdot, \cdot)$. □

Lemma D.2. The correspondence $\mathcal{C}_\ell$ is continuous on $[\underline{\mathcal{E}}^*, \bar{\mathcal{E}}^*] \times \mathcal{D}$.

Proof. Since $\text{entr}_\ell(\cdot, \cdot, \cdot)$ and $\mathcal{F}_\ell(\beta_j; \cdot, \cdot)$ are continuous, by lemma D.1 the graph of $\mathcal{C}_\ell$ is closed and the image of $\mathcal{C}_\ell$ is compact. Hence, $\mathcal{C}_\ell$ is upper–hemicontinuous. We now show that $\mathcal{C}_\ell$ is also lower–hemicontinuous. Choose an arbitrary point $(\mathcal{E}^*, D_0, \bar{Z}, \alpha)$ from the graph of $\mathcal{C}_\ell$. Let $(\mathcal{E}_n^*, D_n)$ be an arbitrary sequence in $[\underline{\mathcal{E}}^*, \bar{\mathcal{E}}^*] \times \mathcal{D}$ that converges to $(\mathcal{E}^*, D_0)$ as $n \to \infty$. We need to show that there exists $\{(\bar{Z}_n, \alpha_n) : n = 1, 2, \cdots\}$ such that $(\bar{Z}_n, \alpha_n) \in \mathcal{C}_\ell(\mathcal{E}_n^*, D_n)$ for all sufficiently large n and for which $(\bar{Z}_n, \alpha_n) \to (\bar{Z}, \alpha)$. Let $\{\epsilon_n\}$ be a sequence such that $\epsilon_n \downarrow 0$ and $\epsilon_n \leq \mathcal{E}^* - \mathcal{E}_n^* \leq 2\epsilon_n$. Since $\text{entr}_\ell(\cdot, \cdot, \cdot)$ is continuous at $(\bar{Z}, \alpha, D_0)$, there exists $\delta_n \downarrow 0$ such that

$$(\bar{z}, a, d) \in \mathcal{N}_{\delta_n}(\bar{Z}, \alpha, D_0) \Longrightarrow \left|\text{entr}_\ell(\bar{z}, a; d) - \text{entr}_\ell(\bar{Z}, \alpha; D_0)\right| \leq \epsilon_n, \tag{37}$$

where $\mathcal{N}_{\delta_n}$ is a δ_n–neighborhood. Choose n large enugh to ensure that D_n is in a δ_n neighborhood of D_0. Then choose $(\bar{Z}_n, \alpha_n)$ such that $(\bar{Z}_n, \alpha_n, D_n) \in \mathcal{N}_{\delta_n}(\bar{Z}, \alpha, D_0)$. It follows that for all such n, $\mathcal{E}^* - \epsilon_n \leq \text{entr}_\ell(\bar{Z}, \alpha, D_0) - \epsilon_n \leq \text{entr}_\ell(\bar{Z}_n, \alpha_n; D_n)$, where the first inequality holds because $(\mathcal{E}^*, D_0, \bar{Z}, \alpha)$ belongs to the graph of $\mathcal{C}_\ell$ and the second inequality follows from (37). Now choose $\mathcal{E}_n^*$ such that $\mathcal{E}^* - 2\epsilon_n \leq \mathcal{E}_n^* \leq \mathcal{E}^* - \epsilon_n$. Then by definition we have $\mathcal{E}_n^* \leq \mathcal{E}^* - \epsilon_n \leq \text{entr}_\ell(\bar{Z}_n, \alpha_n; D_n)$ and $(\bar{Z}_n, \alpha_n) \to (\bar{Z}, \alpha)$. Repeat the argument for $\mathcal{F}_\ell$ in lieu of entr_ℓ. □

Proof of theorem 3: For continuity of $\pi^\bullet$ it suffices to show that

$$\pi_\ell^\bullet(\mathcal{E}^*, D_0; r) = \min_{(\bar{Z},\alpha)\in\mathcal{C}(\mathcal{E}^*,D_0)} \pi_\ell^\circ(\bar{Z}, \alpha, \mathcal{E}^*, D_0; r)$$

is a continuous function for each ℓ. Continuity in r is trivial. Continuity in $\mathcal{E}^*$ and D_0 follows from the theorem of the maximum because of lemmas D.1 and D.2. Finally, since $\pi^\bullet$ is continuous, upper hemi–continuity of $\tilde{r}$ follows from the theorem of the maximum as well. □

E. Constraint selection and inference

Proof of lemma 8.1: If $D_0 \in \text{int}(S_{K_0})$, then by the definition of S_{K_0} and the KKT conditions,

$$\begin{cases} \gamma_j^*(D_0) > 0, \ \sum_{k=1}^{j} g_k^*(D_0) = \Upsilon_{0j} & \text{for } j \in K_{u0}, \\ \gamma_j^*(D_0) = 0, \ \sum_{k=1}^{j} g_k^*(D_0) < \Upsilon_{0j} & \text{for } j \notin K_{u0}, \\ \gamma_j^*(D_0) < 0, \ \sum_{k=1}^{j} g_k^*(D_0) = \Lambda_{j0} & \text{for } j \in K_{\ell 0}, \\ \gamma_j^*(D_0) = 0, \ \sum_{k=1}^{j} g_k^*(D_0) > \Lambda_{j0} & \text{for } j \notin K_{\ell 0}. \end{cases} \tag{38}$$

Therefore, $K_{u0}^* = K_{u0}$ and $K_{\ell 0}^* = K_{\ell 0}$ by definition. Now, instead suppose that $D_0 \in \text{bdr}(S_{K_0})$. Then, there exist $j \notin K_{u0} \cup K_{\ell 0}$ and $\{D_t\}$ with $D_t \to D_0$ such that $\gamma_j^*(D_t) \neq 0$ for all t but $\gamma_j^*(D_t) \to \gamma_j^*(D_0) = 0$. Fix such j and $\{D_t\}$. By the KKT conditions, for all t we have either $\sum_{k=1}^{j} g_k^*(D_t) = \Upsilon_{jt}$ or $\sum_{k=1}^{j} g_k^*(D_t) = \Lambda_{jt}$. Therefore, it follows from lemma A.2 that we have either $\sum_{k=1}^{j} g_k^*(D_0) = \Upsilon_{0j}$ or $\sum_{k=1}^{j} g_k^*(D_0) = \Lambda_{j0}$, so $j \in K_{u0}^* \cup K_{\ell 0}^* = K_0^*$ and hence $K_0 \neq K_0^*$. □

Proof of lemma 8.2: Let $\hat{\gamma}_j^* = \gamma_j^*(\hat{D})$ and $\gamma_{j0}^* = \gamma_j^*(D_0)$. First parts *(a)* and *(b)*. By theorem 5 and assumption A, $\hat{Z} = O_p(1)$. By lemma B.3 and theorem 5, we also have $\sqrt{N}(\hat{\gamma}_j^* - \gamma_{j0}^*) = O_p(1)$. Since parts *(a)* and *(b)* are similar, we focus on part *(a)*, for which it suffices to show that $\mathbb{P}\big(\tilde{K}_u \neq K_{u0}\big) = o(1)$ and $\mathbb{P}\big(\tilde{K}_\ell \neq K_{\ell 0}\big)$. By symmetry, it suffices to establish the former. If $j \notin K_{u0}$ then $\sqrt{N}\hat{\gamma}_j^* = O_p(1)$. But then, $\mathbb{P}(j \in \tilde{K}_u) = \mathbb{P}\big(\sqrt{N}\hat{\gamma}_j^* > \sqrt{N}\kappa_N\big) = o(1)$. Conversely, if $j \in K_{u0}$ then $\gamma_{j0}^* > 0$ and $\sqrt{N}\big(\hat{\gamma}_j^* - \gamma_{j0}^*\big) = O_p(1)$, in which case $\mathbb{P}(j \notin \tilde{K}_u) = \mathbb{P}\big(\hat{\gamma}_j^* \leq \kappa_N\big) = \mathbb{P}\big\{\sqrt{N}\big(\hat{\gamma}_j^* - \gamma_{j0}^*\big) \leq \sqrt{N}\big(\kappa_N - \gamma_{j0}^*\big)\big\} = o(1)$.

Finally, part *(c)*. For the upper bounds, note that $\hat{\gamma}_j^* > \kappa_N \Rightarrow \hat{\gamma}_j^* > 0 \Rightarrow \sum_{k=1}^{j} \hat{g}_k^* = \hat{\Upsilon}_j \Rightarrow$

$\sum_{k=1}^{j} \hat{g}_k^* > \hat{\Upsilon}_j - \kappa_N$. The argument for the lower bounds is similar. □

Proof of theorem 6: By lemma 8.1, $\mathbb{P}(\hat{T} \neq T) = o(1)$.

Suppose first that D_0 is a boundary point, such that by lemma 8.1 $K_0 \subsetneq K_0^*$. By theorem 5 and lemma 8.2, and the continuity of G in its first argument,

$$\mathbb{P}(\hat{Z} \leq x) = \mathbb{P}\{\Theta(\hat{\Omega}, \hat{K}) > x\} + o(1) \geq \mathbb{P}\left\{\max_{\tilde{K} \subseteq K \subseteq \tilde{K}^*} \Theta(\hat{\Omega}, K) \leq x\right\} + o(1)$$
$$= \mathbb{P}\left\{\max_{K_0 \subseteq K \subseteq K_0^*} \Theta(\hat{\Omega}, K) \leq x\right\} + o(1) = \mathbb{P}(T \leq x) + o(1).$$

Now, suppose instead that D_0 is an interior point. Lemma 8.1 implies that then $K_0 = K_0^*$, such that $\mathbb{P}(\hat{Z} \leq x) = \mathbb{P}\{\Theta(\Phi, K_0) \leq x\} + o(1) = \mathbb{P}(T \leq x) + o(1)$. □

F. Optimal reserve price and maximum revenue

F.1 Consistency of $\hat{\mathcal{R}}_N$:

Proof of lemma 9.1: Follows from the fact that $\pi(\cdot, g^*)$ is continuous on $[0, 1]$ and (strictly) concave on each I_j. □

Lemma F.1. $\sup_{v\in[0,1]}\left|\hat{f}^*(v) - f^*(v)\right| + \sup_{v\in[0,1]}\left|\hat{\mathcal{F}}^*(v) - \mathcal{F}^*(v)\right| = O_p(1/\sqrt{N})$.

Proof. It follows from (5) and (6) and theorem 5. □

Lemma F.2. $\sup_{r\in[0,1]}\left|\pi(r, \hat{g}^*) - \pi(r, g^*)\right| = O_p(1/\sqrt{N})$.

Proof. It follows from lemma F.1 and (1). □

Lemma F.3. $\left|\mathcal{P}(\hat{g}^*) - \mathcal{P}(g^*)\right| = O_p(1/\sqrt{N})$

Proof. Follows from lemma F.2. □

Proof of theorem 7: Note that by continuity of $\pi(\cdot, g^*)$ and by lemma 9.1, $\mathcal{R}_\epsilon = \{r : \pi(r, g^*) \geq \mathcal{P}(g^*) - \epsilon\}$ consists for any sufficiently small $\epsilon > 0$ of a union of disjoint compact intervals $\mathcal{R}_{\epsilon 1}, \ldots, \mathcal{R}_{\epsilon m}$ each containing one r_j^*. Choose $\epsilon > 0$. By lemmas F.2 and F.3,

$$\begin{cases} \mathbb{P}(\hat{\mathcal{R}}_N \not\subseteq \mathcal{R}_\epsilon) \leq \mathbb{P}\left[\max_{r\in[0,1]}\{\pi(r, \hat{g}^*) - \pi(r, g^*)\} \geq \mathcal{P}(\hat{g}^*) - \mathcal{P}(g^*) + \epsilon - \kappa_n\right] = o(1), \\ \mathbb{P}(\mathcal{R}_0 \not\subseteq \hat{\mathcal{R}}_N) \leq \mathbb{P}\left[\max_{r\in[0,1]}\{\pi(r, g^*) - \pi(r, \hat{g}^*)\} \geq \mathcal{P}(g^*) - \mathcal{P}(\hat{g}^*) + \kappa_n\right] = o(1). \end{cases}$$

Let $\epsilon \to 0$. □

Proof of lemma 9.2: Note that $r_1(\hat{g}^*) \neq \hat{r}_1 \Rightarrow r_1(\hat{g}^*) \notin \hat{\mathcal{R}}_{N,1}$. Now, $r_1(\hat{g}^*) \in \hat{\mathcal{R}}_N$ with probability approaching one because otherwise $\mathcal{P}(\hat{g}^*) - \kappa_N \leq \pi(\hat{r}_1, \hat{g}^*) \leq \pi\{r_1(\hat{g}^*), \hat{g}^*\} < \mathcal{P}(\hat{g}^*) - \kappa_N$, which is a contradiction. But since $r_1(\hat{g}^*) \xrightarrow{p} r_1^*$ by construction, $r_1(\hat{g}^*) \in \hat{\mathcal{R}}_{N,1}$ because $\hat{\mathcal{R}}_{N,1}$ is the collection of elements in $\hat{\mathcal{R}}_N$ that converges to r_1^*. □

Lemma F.4. If $\hat{r}_* - r_k^* = o_p(1)$ for some $\hat{r}_* \in \hat{\mathcal{R}}_N$ and some $k = 1, \ldots, m$ then $\hat{r}_* - r_k^* = O_p(\sqrt{\kappa_N})$.

Proof. Let j^* be such that $r_k^* \in I_{j^*}$. If $r_k^* > \beta_{j^*-1}$ then $O_p(\kappa_N) = \pi(\hat{r}_*, g^*) - \pi(r_k^*, g^*) \simeq \partial_r^2 \pi(r_k^*, g^*)(\hat{r}_* - r_k^*)^2 \ / \ 2$.[42] Since $\partial_r^2 \pi(r_k^*, g^*) < 0$ it follows that $\hat{r}_* - r_k^* = O_p(\sqrt{\kappa_N})$. If $r_k^* = \beta_{j^*-1}$ then one should consider $\partial_r^+ \pi(r_k^*, g^*)$ and $\partial_r^- \pi(r_k^*, g^*)$ separately. If both are nonzero then it follows that $\hat{r}_* - r_k^* = O_p(\kappa_N)$. If either the right or left derivative equals zero then the convergence rate is no worse than $\sqrt{\kappa_N}$, as before. □

Proof of theorem 8: Suppose first that $m = 1$. Consistency of $\hat{r}_m$ follows trivially from theorem 7 and lemma F.4 establishes that $\hat{r}_m - r_m^* = O_p(\sqrt{\kappa_N})$. We now argue that $\hat{m} = m$ with probability approaching one. Suppose that $\hat{m} > m$. Then $\hat{r}_{m+1} \xrightarrow{p} r_m^*$ by theorem 7 and hence $\hat{r}_{m+1} - r_m^* = O_p(\sqrt{\kappa_N})$ by lemma F.4. But the construction of $\hat{\mathcal{R}}_{N,m+1}$ implies that

$$\sqrt[3]{\kappa_N} = \tilde{\kappa}_N \leq \hat{r}_{m+1} - \min \hat{\mathcal{R}}_{N,m} = \hat{r}_{m+1} - r_m^* + r_m^* - \min \hat{\mathcal{R}}_{N,m} \leq \hat{r}_{m+1} - r_m^* + O_p(\sqrt{\kappa_N}),$$

which is at odds with $\hat{r}_{m+1} - r_m^* = O_p(\sqrt{\kappa_N})$. So $\hat{m} = m$ with probability approaching one and the proof is complete for $m = 1$.

Now suppose that $m = 2$. Consistency and convergence rate of $\hat{r}_{m-1}$ follow as above. Further, $\hat{m} \geq m$ with probability approaching one since $\max \hat{\mathcal{R}}_N \xrightarrow{p} r_m^*$. As established above $\hat{r}_m$ does not converge to r_{m-1}^*, so $\hat{r}_m$ must converge to r_m^* and hence $\hat{r}_m - r_m^* = O_p(\sqrt{\kappa_N})$. Using the same argument as when $m = 1$, now $\hat{m} = m$ with probability approaching one. Iterate the argument made for $m = 2$ for $m > 2$, noting that m is finite. □

F.2 Sensitivity: Fixing a direction d, let

$$\Psi(d, t) = \frac{r_1\{g^*(D_0 + td)\} - r_1\{g^*(D_0)\}}{t}.$$

The existence of $\lim_{t \downarrow 0} \Psi(d, t)$ requires the Hadamard differentiability of r_1 at $g^* = g^*(D_0)$: the chain rule fails if r_1 is only directionally differentiable. So, we first consider the Hadamard directional derivative of r_1: for any $v_t = v + o(1)$,

$$\nabla_H r_1(g^*, v) = \lim_{t \downarrow 0} \frac{r_1(g^* + tv_t) - r_1(g^*)}{t}.$$

[42] We define $\simeq$ to mean that any remaining terms are asymptotically negligible.

Lemma F.5. r_1 is continuous at g^*.

Proof. By the maximum theorem, $\tilde{R}_1(g) = \operatorname{argmax}_{r \in A_1} \pi(r, g)$ is upper hemicontinuous. So, the conclusion follows from the fact that $\tilde{R}_1(g^*)$ is a singleton. □

Lemma F.6. For cases I–IV, $\nabla_H r_1(g^*, v) = \psi_{g^*}(v)$. First case I. By lemma F.5, we must have $\beta_{j^*-1} < r_1(g) < \beta_{j^*}$ in a neighborhood of g^*. Hence $\theta\{r_1(g), g\} = 0$ in a neighborhood of g^*. Apply the implicit function theorem.

For case II, note that $\theta(\cdot, g)$ is an invertible function near $r_1^* = \beta_{j^*-1}$, that both θ and its inverse are continuous in g, and that θ's inverse is flat in its first argument. For cases III and IV, combine the arguments for cases I and II. □

Lemma F.7. For case V and any $v_t = v + o(1)$, we have $\limsup_{t \downarrow 0} \left| r_1(g^* + t v_t) - r_1(g^*) \right| / t \leq \max\left(|v^\intercal \delta|, |v^\intercal \delta^-| \right)$.

Proof. Follows immediately by noting that

$$\limsup_{t \downarrow 0} \frac{\left| r_1(g^* + t v_t) - r_1(g^*) \right|}{t} \leq \max\left\{ \left| \frac{\partial_g \theta^+(r^*, g^*)}{\partial_r \theta^+(r^*, g^*)} \right|, \left| \frac{\partial_g \theta^-(r^*, g^*)}{\partial_r \theta^-(r^*, g^*)} \right| \right\}. \qquad \square$$

Lemma F.8. In cases I–IV we have $\nabla(r_1 \circ g^*)(D_0, d) = \nabla_H r_1\{g^*, \nabla g^*(D_0, d)\}$, where $\nabla_H r_1(g^*, v)$ is given in lemma F.6. For V, we have $\limsup_{t \downarrow 0} |\Psi(d, t)| \leq \max\{|\nabla g^*(D_0, d)^\intercal \delta|, |\nabla g^*(D_0, d)^\intercal \delta^-|\}$.

Proof. Follows from lemmas B.2 and F.6. □

Let $\tilde{d} = (\hat{D} - D_0)/\|\hat{D} - D_0\|$.

Lemma F.9. In cases I–IV, we have

$$\sqrt{N}\{r_1(\hat{g}^*) - r_1(g^*)\} = \nabla_H r_1\{g^*, \nabla g^*(D_0, \tilde{d})\} \sqrt{N} \|\hat{D} - D_0\| + o_p(1),$$

whereas in case V,

$$\sqrt{N} |r_1(\hat{g}^*) - r_1(g^*)| \leq \max\{|\nabla g^*(D_0, \tilde{d})^\intercal \delta|, |\nabla g^*(D_0, \tilde{d})^\intercal \delta^-|\} \sqrt{N} \|\hat{D} - D_0\| + o_p(1).$$

Proof. Note that $r_1(\hat{g}^*) = r_1\{g^* + \tilde{d} \|\hat{D} - D_0\|\}$. Using lemma F.8, apply the same logic as the proof of theorem 5. □

Proof of theorem 9: Follows from lemmas 9.2 and F.9. □

Proof of theorem 10: The function π is differentiable in g and directionally differentiable in r, so

$$\pi(r,g) - \pi(r_1^*, g^*) = \partial_r^+ \pi(r_1^*, g^*) \max(0, r - r_1^*) + \partial_r^- \pi(r_1^*, g^*) \min(0, r - r_1^*) \\ + \partial_{g^\top} \pi(r_1^*, g^*)(g - g^*) + o\big(|r - r_1^*| + \|g - g^*\|\big),$$

where ∂_r^+ and ∂_r^- denote the right and left derivative, respectively. Therefore, by theorems 5 and 9,

$$\sqrt{N}\big\{\pi(\hat{r}_1, \hat{g}^*) - \pi(r_1^*, g^*)\big\} = \partial_r^+ \pi(r_1^*, g^*) \max\big\{0, \sqrt{N}(\hat{r}_1 - r_1^*)\big\} \\ + \partial_r^- \pi(r_1^*, g^*) \min\big\{0, \sqrt{N}(\hat{r}_1 - r_1^*)\big\} + \partial_{g^\top} \pi(r_1^*, g^*) \sqrt{N}(\hat{g}^* - g^*) + o_p(1). \quad (39)$$

Now, note that by theorem 9,

1. $\partial_r^+ \pi(r_1^*, g^*) = 0 = \partial_r^- \pi(r_1^*, g^*)$ in cases I and V;
2. $\sqrt{N}(\hat{r}_1 - r_1^*) = o_p(1)$ in case II;
3. $\partial_r^+ \pi(r_1^*, g^*) = 0$ and $\sqrt{N}(\hat{r}_1 - r_1^*) - o_p(1) \geq 0$ in case III;
4. $\partial_r^- \pi(r_1^*, g^*) = 0$ and $\sqrt{N}(\hat{r}_1 - r_1^*) - o_p(1) \leq 0$ in case IV.

Therefore, the first right–hand side terms in (39) are asymptotically negligible in all cases. □

References

Andrews, Donald WK and Gustavo Soares (2010). "Inference for parameters defined by moment inequalities using generalized moment selection". In: *Econometrica* 78.1, pp. 119–157.

Aryal, Gaurab and Dong-Hyuk Kim (2013). "A point decision for partially identified auction models". In: *Journal of Business & Economic Statistics* 31.4, pp. 384–397.

Athey, Susan and Philip A Haile (2002). "Identification of standard auction models". In: *Econometrica* 70.6, pp. 2107–2140.

Avery, Christopher (1998). "Strategic jump bidding in English auctions". In: *The Review of Economic Studies* 65.2, pp. 185–210.

Bertsekas, Dimitri P (2015). *Convex optimization algorithms*. Athena Scientific Belmont.

Bulow, Jeremy and Paul Klemperer (1996). "Auctions versus negotiations". In: *The American Economic Review*, pp. 180–194.

Chernozhukov, Victor, Han Hong, and Elie Tamer (2007). "Estimation and confidence regions for parameter sets in econometric models". In: *Econometrica* 75.5, pp. 1243–1284.

Daniel, Kent D and David A Hirshleifer (1998). *A theory of costly sequential bidding*. Tech. rep.

Gilboa, Itzhak and David Schmeidler (1989). "Maxmin expected utility with non-unique prior". In: *Journal of mathematical economics* 18.2, pp. 141–153.

Grünwald, Peter D and A Philip Dawid (2004). "Game theory, maximum entropy, minimum discrepancy and robust Bayesian decision theory". In: *Annals of Statistics*, pp. 1367–1433.

Haile, Philip A and Elie Tamer (2003). "Inference with an incomplete model of English auctions". In: *Journal of Political Economy* 111.1, pp. 1–51.

Hansen, Lars Peter and Thomas J Sargent (2008). *Robustness*. Princeton university press.

Harremoës, Peter and Flemming Topsøe (2001). "Maximum entropy fundamentals". In: *Entropy* 3.3, pp. 191–226.

Hartline, Jason D (2016). *Mechanism design and approximation*. draft.

Jun, Sung Jae and Joris Pinkse (2016). *Point prediction under partial identification*. Tech. rep. Penn State.

Myerson, Roger B (1981). "Optimal auction design". In: *Mathematics of operations research* 6.1, pp. 58–73.

Topsøe, Flemming (1979). "Information-theoretical optimization techniques". In: *Kybernetika* 15.1, pp. 8–27.