Estimation of Standard Auction Models

Yeshwanth Cherapanamjeri
University of California Berkeley
yeshwanth@berkeley.edu

Constantinos Daskalakis
Massachusetts Institute of Technology
costis@csail.mit.edu

Andrew Ilyas
Massachusetts Institute of Technology
ailyas@mit.edu

Manolis Zampetakis
University of California Berkeley
mzampet@berkeley.edu

Abstract

We provide efficient estimation methods for first- and second-price auctions under independent (asymmetric) private values and partial observability. Given a finite set of observations, each comprising the identity of the winner and the price they paid in a sequence of identical auctions, we provide algorithms for non-parametrically estimating the bid distribution of each bidder, as well as their value distributions under equilibrium assumptions. We provide finite-sample estimation bounds which are uniform in that their error rates do not depend on the bid/value distributions being estimated. Our estimation guarantees advance a body of work in Econometrics wherein only identification results have been obtained (e.g. Athey and Haile [2002, 2007]), unless the setting is symmetric (e.g. Morganti [2011], Menzel and Morganti [2013]), parametric (e.g. Athey et al. [2011]), or all bids are observable (e.g. Guerre et al. [2000]). Our guarantees also provide computationally and statistically effective alternatives to classical techniques from reliability theory [Meilijson, 1981]. Finally, our results are immediately applicable to Dutch and English auctions.

1 Introduction

Estimating value and/or bid distributions from an observed sequence of auctions is a fundamental challenge in Econometrics with direct practical applications. For example, these fundamentals allow one to analyze the performance of an auction and make counterfactual predictions about alternatives. The difficulty of this problem depends on the format of the auctions and the structure of the observed information from each one, as well as how the fundamentals of bidders are interrelated and vary across the sequence of observations.

In this paper, we study a basic version of the afore-described estimation challenge, wherein the auction format and the bidder distributions stay fixed across observations, and the bidders have independent private values (which are independently resampled across different observations). The auction formats that we consider are first- and second-price auctions, as well as Dutch and English auctions. What will make our problem challenging is that (i) our bidders are ex ante asymmetric, drawing their independent private values from different distributions; (ii) we will make no parametric assumptions about these distributions; and (iii) we will only be observing the identity of the winner and the price they paid but not the losing bids. Under this observational model and our independent private values assumption above, we can focus our attention on first- and second-price auctions, and our results automatically extend to Dutch and English auctions.

In the above settings, we give computationally and sample efficient methods for estimating all agents' bid distributions and (under equilibrium assumptions) value distributions:

- In the case of first-price auctions, we provide finite-sample estimation guarantees under Lévy, Kolmogorov and Total Variation distance with minimal assumptions. Under (a condition weaker than) a lower bound on the density of the bid distributions (although we actually do not need existence of densities), Theorem 2.2 shows that the bid distributions can be estimated to within ε in Lévy distance, using $1/\varepsilon^{O(k)}$ samples, where k is the number of bidders. Theorem 2.6 shows that the exponential dependence on k is necessary, and Theorem 2.7 shows that Lévy distance cannot be strengthened to Kolmogorov distance. Sidestepping the exponential sample dependence on k, strengthening the estimation distance, and removing the density lower bound assumption, Theorem 2.3 shows that, assuming only continuity of their cumulative functions, the bid distributions can be estimated to within ε in Kolmogorov distance on their effective supports, i.e. the part of the support that is likely to be observed, defined in Blum et al. [2015]. Finally, under Lipschitzness assumptions on the densities of the bid distributions, Theorem 2.4 improves the latter to ε-error in Total Variation distance. Our sample requirements for estimation over the effective supports of the bid distributions under either Kolmogorov or TV distance are dramatically improved to logarithmic in the number of bidders and benign in $1/\varepsilon$. Finally, assuming that bidders use Bayesian Nash equilibrium strategies, Theorems 2.20 and 2.15 show that bidders' value distributions can be estimated over their full and, respectively, effective supports with similar sample sizes as those needed for the estimation of bid distributions.

  All of our estimation algorithms run in polynomial time in their sample sizes, and all our estimation error bounds are uniform in that they do not depend on the bid/value distributions being estimated, unlike the instance-dependent rates that commonly arise from the use of kernel density estimation methods. It is also important to note that we estimate the value distributions in Lévy distance (in fact, in the stronger notion of Wasserstein distance) and this is sufficient for the purposes of performing counter-factual predictions about the revenue that would result from running alternative auctions [Brustle et al., 2020].

- In the case of second-price auctions, Theorem 3.3 establishes that bid distributions can be estimated to within ε in Kolmogorov distance over their entire supports assuming upper and lower bounds on their density functions. Again the sample complexity scales as $1/\varepsilon^{O(k)}$. This result poses major technical challenges, requiring a computationally and statistically effective, fixed point computation alternative to Meilijson [1981]'s method. We again sidestep the exponential dependence of the required sample size on k, by considering estimation over the effective support of the distributions in a setting, similar to that proposed by Blum et al. [2015] for first-price auctions, where we can insert bids to the auction or, equivalently, set a reserve price (see discussion in Section 2.3). In this setting, Theorem 3.12 shows that bid distributions can be estimated to within ε in Kolmogorov distance over their effective supports, using a sample size that is polynomial in both $1/\varepsilon$ and k. Similar to Theorem 2.4 estimation in Kolmogorov distance can be turned to estimation in Total Variation distance under Lipschitzness of the densities. Of course, assuming that the bidders bid according to the truthful bidding equilibrium, our estimation results for bid distributions automatically translate to estimation results for value distributions.

To the best of our knowledge, our results are the first finite-sample estimation guarantees for the general problem we consider. In particular:

- There is an extensive line of work on identification of bid and value distributions from complete or partial observations of bids; see Athey and Haile [2007] for a survey. In our setting of independent private values, Athey and Haile [2002] show that with infinite samples, bid distributions are identifiable from the distribution of the identity of the winner and the price they paid.

  Identification results for bid and value distributions have been established in the presence of correlated values, alternative auction formats, and unobserved heterogeneity, or unknown numbers of bidders [Paarsch, 1992, Laffont and Vuong, 1993, Laffont et al., 1995, Donald and Paarsch, 1996, Baldwin et al., 1997, Donald et al., 2003, Bajari and Ye, 2003, Haile, 2001, Luo and Xiao, 2020, Haile et al., 2003, Mbakop, 2017, Hu et al., 2013]. In contrast, here we focus on the IPV framework (albeit, in its more challenging asymmetric case) and standard auction mechanisms (first-price, second-price, Dutch, and English auctions), and provide finite-sample estimation results in these settings.

- On the estimation front, Morganti [2011] and Menzel and Morganti [2013] both provide estimators of bid distributions in first- and second-price auctions (in fact they provide estimation using any order statistics of the bid distributions) under the restrictive assumption that the bidders are symmetric. They obtain rates for estimation of the bid distributions over the full support, which degrade exponentially with the number of bidders. In comparison, we get similar rates for the significantly more challenging asymmetric setting, and also provide drastically better estimation rates (with logarithmic dependence on the number of bidders) on the effective supports of the distributions.

- In terms of non-parametric estimation of value distributions from bid distributions in first-price auctions, Guerre et al. [2000] provide estimation algorithms which operate under the restrictive assumption that the bidders are symmetric. Their estimation makes use of the explicit formula for the Bayesian Nash equilibrium in the symmetric case. Later work [Campo et al., 2003, Bajari and Ye, 2003, Krasnokutskaya, 2011, Haile et al., 2003] extends these results to the more general asymmetric setting, where the Bayesian Nash equilibrium has no closed-form expression; as such, these analyses are typically limited to the setting with only two unique bidder types. Our algorithms operate in the latter (significantly more challenging) setting, with the additional challenges of (a) allowing each bidder to have their own unique value distribution (b) only observing the winning bid, rather than all agents' bids; (c) not using higher-order differentiability assumptions used in prior work while also providing uniform convergence bounds (i.e., bounds not depending on the distributions being estimated).

- In the computer science literature, there has been work on non-parametric estimation of bid distributions in first-price auctions, under the stronger assumption that the econometrician can insert bids which do not influence the bidding behavior of the bidders Blum et al. [2015]. We compare to that setting and work in Section 2.3, explaining that our work obtains substantial improvements on their rates.

Besides the non-parametric identification work on auctions, discussed above, there has been more extensive work on estimation and identification under parametric or semi-parametric assumptions: see [Donald and Paarsch, 1996] and [Athey and Haile, 2006] for an overview. For example, Athey et al. [2011] fit the parameters of Weibull distributions to observed maximum bids in to estimate bid distributions in USFS timber auctions.

1.1 Preliminaries

The (asymmetric) independent private values model. In this work, we consider the asymmetric independent private values (IPV) model, with the additional stipulation that not all bids are observed. In this model, we observe a series of identical auctions between k agents (known $k \geqslant 2$): in each auction, every agent i submits a bid X_i, sampled independently from a (fixed) distribution with cumulative distribution function F_i. The sampled bids, together with the auction type, determine the winner Z of each auction (typically $Z = \arg\max_i X_i$) and a transaction price Y, i.e., what the winner pays for the auctioned item. In this work, we will only observe Z and (sometimes) Y, and rather than all bids X_i.

Two key differences between our setting and the typical IPV setup are (a) the aforementioned partial observability; and (b) asymmetry—in particular, a typical assumption is that all agents bid according to the same fixed distribution F, which simplifies both bid and value estimation significantly (e.g., we could estimate a first-price auction by learning the CDF of the largest bid, and then estimate the individual bid distributions as the k-th root).

Statistical distances. Throughout our work, we provide finite-sample convergence bounds in terms of the Wasserstein, Lévy, and Kolmogorov distances, depending on the setting. The Wasserstein distance $\mathcal{W}$ between two distributions P, Q supported on $[0,1]$ is

$$\mathcal{W}(P,Q) \triangleq \inf_{R} \mathbb{E}_{(x,y)\sim R}\left[|x-y|\right],$$

where the infimum is over all joint distributions R with support $[0,1]^2$ such that the marginal of x is equal to P and the marginal of y is equal to Q when $(x,y) \sim R$. The Kolmogorov distance $\mathrm{d_K}$ between two distributions P and Q over an interval I is defined as

$$\mathrm{d_K}(P,Q) \triangleq \sup_{x\in I} |F_P(x) - F_Q(x)|,$$

where F_P and F_Q are the cumulative distribution functions of P and Q, respectively. Finally, the Lévy distance D_L between P and Q is given by

$$D_L(P,Q) = \min\left\{\epsilon :\; F_P(x-\epsilon) - \epsilon \leqslant F_Q(x) \leqslant F_P(x+\epsilon) + \epsilon\right\}.$$

Note that Lévy distance is a strictly weaker notion than Wasserstein distance and Kolmogorov distance, in the sense that both $D_L(P,Q) \leqslant \mathrm{d_K}(P,Q)$ and $D_L(P,Q) \leqslant \sqrt{\mathcal{W}(P,Q)}$. Thus, all of our results in Wasserstein and Kolmogorov distance also effectively bound Lévy distance.

Finite-sample rates. We present our convergence results using order notation: in particular, the constants omitted from all order notation in this paper are absolute constants that do not depend on the distributions being estimated or any other parameters of the setting being considered. In other words, our bounds are uniform. For example, whenever a bound in a theorem statement reads as $O(f(k, 1/\varepsilon, L))$, where f is some function and k, ε, L are parameters of the setting, this means that there is an absolute constant C such that for any setting conforming to the setting examined in this theorem we can replace $O(f(k, 1/\varepsilon, L))$ in the theorem statement by $C \cdot f(k, 1/\varepsilon, L)$. Recall that $\tilde{O}(f(\cdot))$ means that for some absolute constants C and $k \geqslant 1$, the bound can be replaced by $Cf(\cdot) \log^k f(\cdot)$. Finally, in this work we present most of our bounds in terms of the number of samples necessary to attain a specific learning error ε—these can be straightforwardly converted to bounds on ε in terms of the number of samples n, providing estimation rates as $n \to \infty$.

2 Estimation from First-price Auction Data

In this section we show how to estimate the bid distributions from a finite number of first-price auction observations. We consider two regimes:

- In the *full-support* regime, our goal is to provide an estimation of the bid distributions in their whole support $[0, 1]$. As shown in Theorem 2.2, in this regime we estimate the probability distributions within ε in Wasserstein distance. The sample complexity here is $\approxeq (1/\varepsilon)^k$ and has exponential dependence on the number of the agents k. As we explain in Subsection 2.2, this dependence on k is necessary for the full-support regime (due to the exponentially low probability of observing winning bids near zero).

- In the *effective-support* regime, our goal is to provide an estimation of the bid distributions only at the bid values that have probability at least λ to be observed as an outcome of the first-price auction. As we show in Theorem 2.3, in this regime we avoid the exponential dependence on k and we are able to get an algorithm that depends only polynomially in ε and γ and only logarithmically in k. This is a doubly exponential improvement over the full-support regime and an exponential improvement on the best known effective-support result from Blum et al. [2015]. This result also provides the first algorithm with sublinear sample complexity for this problem.

Our first step is to formally define the procedure from which the first-price auction data are generated. The observation access that we assume is minimal in the sense that we only observe the outcome of the auction; who wins and how much they pay.

Definition 2.1 (First-Price Auction Data)**.** Let $\{F_i\}_{i=1}^k$ be k cumulative distribution functions with support $[0, 1]$, i.e. $F_i(x) = 0\ \forall x < 0$ and $F_i(1) = 1$. A sample (Y, Z) from a first-price auction with bid distributions $\{F_i\}_{i=1}^k$ is generated as follows:

1. first generate $X_i \sim F_i$ independently for all $i \in [k]$,
2. observe the tuple $(Y, Z) \triangleq (\max_{i \in [k]} X_i, \arg\max_{i \in [k]} X_i)$.

A different access model explored in Blum et al. [2015] gives the econometrician control over an additional agent that has the ability to bid arbitrarily, but only allows them to observe the identity of the winner of each auction (not the transaction price). Using our result we can also improve the result of Blum et al. [2015] under this model (see Subsection 2.3).

2.1 Estimation of Bid Distributions

We are now ready to state our main results for the estimation of the bid distributions given sample access to first-price auction data as defined in Definition 2.1. We start with our result for the full-support regime.

Theorem 2.2 (First-Price Auctions – Full Support)**.** *Let $\{(Y_i, Z_i)\}_{i=1}^n$ be n i.i.d. samples from the first-price auction as per Definition 2.1. Assume that the cumulative distribution functions F_i are continuous and satisfy $|F_i(x) - F_i(y)| \geqslant \lambda |x - y|$ for all $x, y \in [0, 1]$. Then, there is a polynomial-time algorithm that computes functions $\hat{F}_i$ for $i \in [k]$ such that*

$$\mathbb{P}\left(\mathcal{W}(\hat{F}_i, F_i) \leqslant \varepsilon\right) \geqslant 1 - \delta$$

for all $i \in [k]$ assuming that $n = \tilde{\Theta}\left(\left(\frac{2}{\lambda \cdot \varepsilon}\right)^{4k} \frac{\log(1/\delta)}{\varepsilon^2}\right)$, where $\mathcal{W}$ is the Wasserstein distance.

As we show in Subsection 2.2, the sample complexity of Theorem 2.2 is almost optimal. Nevertheless, as we already explained, the exponential dependence on the number of agents k can be reduced to only logarithmic dependence if we only focus on the part of the support that is likely to be observed. In this case, our estimation guarantee is also simpler: we estimate the cumulative distribution functions with additive error ε.

Theorem 2.3 (First-Price Auctions – Effective Support)**.** *Let $\{(Y_i, Z_i)\}_{i=1}^n$ be n i.i.d. samples from the same first-price auction as per Definition 2.1 and assume that the cumulative distribution functions F_i are continuous. Then, there exists a polynomial-time estimation algorithm, that computes the cumulative distribution functions $\hat{F}_i$ for $i \in [k]$, such that for every $p, \gamma \in \{p, \gamma \geqslant 0 : \mathbb{P}_{(Y,Z)\sim\mathcal{P}_1}(Y \leqslant p) \geqslant \gamma\}$, and every $\varepsilon \in (0, \gamma/2]$,*

$$\mathbb{P}\left(\max_{x\in[p,1]} \left|\hat{F}_i(x) - F_i(x)\right| \leqslant \varepsilon\right) \geqslant 1 - \delta$$

for all $i \in [k]$ assuming that $n = \tilde{\Theta}\left(\log(k/\delta)/(\gamma^4 \varepsilon^2)\right)$.

Finally, we establish estimation of the corresponding probability density functions $\{f_i\}$:

Theorem 2.4. *Let $\{(Y_i, Z_i)\}_{i=1}^n$ be n i.i.d. samples from the same first-price auction as per Definition 2.1 and assume the densities f_i of F_i are well-defined and Lipschitz continuous, i.e.,*

$$|f_i(x) - f_i(y)| \leqslant L|x - y| \text{ for all } x, y \in [0, 1].$$

Then, there exists a polynomial-time estimation algorithm, that computes functions $\hat{f}_i$ for $i \in [k]$, with the following guarantee; for every $p, \gamma \geqslant 0$ such that $\mathbb{P}(Y \leqslant p) \geqslant \gamma$, and for every $\varepsilon \in (0, \gamma/2]$ it holds that

$$\mathbb{P}\left(\int_p^1 \left|\hat{f}_i(x) - f_i(x)\right| dx \leqslant \varepsilon\right) \geqslant 1 - \delta$$

for all $i \in [k]$ assuming that $n = \tilde{\Theta}\left(L^2 \cdot \log(k/\delta)/(\gamma^4 \cdot \varepsilon^4)\right)$.

Before explaining the formal proofs of the above theorems we give some intuition behind our estimation algorithm. This intuition is given in a simplified setting where: (1) we assume the population model where we have access to infinitely many samples from the first-price auction data defined in Definition 2.1, and (2) the distributions are smoothed enough so that all the probability density functions that are involved are well defined. In this simplified setting we have access to the following distributions:

▷ H_i is the cumulative distribution function of Y conditioned on $Z = i$, for $i \in [k]$ and

▷ H is the cumulative distribution function of Y with no conditioning on Z.

If we assume that all the H_i's have well-defined densities h_i, then

$$h_i(x) = f_i(x) \cdot \prod_{j \neq i} F_j(x), \quad H(x) = \prod_{j \in [k]} F_j(x), \quad \text{and} \quad H(x) = \sum_{j \in [k]} H_j(x).$$

Based on the above relations we can solve for the distribution F_1 as follows

$$\frac{d}{dx} \log(F_i(x)) = \frac{f_i(x)}{F_i(x)} = \frac{h_i(x)}{H(x)} \implies F_i(x) = \exp\left(-\int_x^1 \frac{h_i(z)}{H(z)} dz\right).$$

This simple idea summarizes our approach in the population setting where infinite samples are available. Moving to the finite sample case an important observation is that the aforementioned expression of F_i can be also written as

$$F_i(x) = \exp\left(-\mathbb{E}_{(y,z)\sim \mathcal{P}_1}\left[\frac{\mathbf{1}\{z = i\}}{H(y)} \,\middle|\, y \geqslant x\right]\right).$$

The above expression allows the expectation involved to be estimated with an empirical expectation instead of an integral, assuming that a good estimation of $H(z)$ is computed. Towards designing our actual estimation algorithm and proving its exact sample complexity we face the following additional technical difficulties:

1. in the above outline we assume that all the distributions are smooth enough so that all the densities are well defined—in our main theorem this assumption is not necessary,
2. the usual estimation of H has an additive error, whereas in the above expression a multiplicative error guarantee is needed,
3. the term $1/H(z)$ that is crucial in our estimation is not numerically stable for z close to 0 where $H(z)$ can also be very close to 0 as well.

2.1.1 Proof of Theorem 2.3

We start by considering the effective-support setting. Our first result will be an information theoretic result enabling identification of F_i with access to the function H and the measure H_i (without requiring a density function).

Lemma 2.5. *For all $i \in [k]$ and all $x \in (0,1)$ such that $F_i(x) > 0$ and $H(x) > 0$,*

$$F_i(x) = \exp\left(-\int_x^1 \frac{1}{H(y)}\, dH_i\right).$$

Proof. Using Lemma 3.1 from Norvaiša [2002] we have that

$$\log(F_i(1)) - \log(F_i(x)) = \int_x^1 \frac{1}{F_i(y)}\, dF_i = \int_x^1 \frac{\prod_{j \neq i} F_j(y)}{\prod_{j \in [k]} F_j(y)}\, dF_i = \int_x^1 \frac{1}{H(y)} dH_i,$$

where the last equality follows from the continuity of F_i's and the properties of Riemann-Stieltjes integration. The lemma follows by observing that $F_i(1) = 1$. □

We now focus our attention on obtaining good estimates of the quantity within the exponential on the right hand side in Lemma 2.5. We introduce the following notation:

$$\hat{H}(x) \triangleq \frac{1}{n}\sum_{j=1}^{n} \mathbf{1}\left\{Y_j \leqslant x\right\} \qquad \hat{G}_i(x) \triangleq \frac{1}{n}\sum_{j=1}^{n} \frac{1}{\hat{H}(Y_j)}\mathbf{1}\left\{Y_j \geqslant x \text{ and } Z_j = i\right\}$$

Based on the above definitions we can define our estimate for F_i as $\hat{F}_i(x) = \exp\left(-\hat{G}_i(x)\right)$. Our next goal is to prove that $\hat{F}_i$ is close to F_i for every value $y \in [0,1]$ such that $H(y) \geqslant \gamma$.

Now we establish concentration of $\hat{H}$. By the DKW inequality Dvoretzky et al. [1956]:

$$\max_{x\in[0,1]} \left|\hat{H}(x) - H(x)\right| \leqslant \frac{1}{20} \cdot \gamma^2 \varepsilon$$

with probability at least $1 - \delta/2$ for our setting of n. Conditioning on the above event and observing that $H(x) \geqslant \gamma$ for all $x \geqslant p$,

$$\max_{x\in[p,1]} \left|1/H(x) - 1/\hat{H}(x)\right| \leqslant \frac{\varepsilon}{10}. \tag{1}$$

To establish concentration of $\hat{G}_i(x)$, we introduce another quantity $\widetilde{G}_i$, defined as follows:

$$\widetilde{G}_i(x) \triangleq \frac{1}{n} \cdot \sum_{j=1}^{n} \frac{1}{H(Y_i)}\mathbf{1}\left\{Y_j \geqslant x \text{ and } Z_j = i\right\}.$$

We have from (1) that $|\hat{G}_i(x) - \widetilde{G}_i(x)| \leqslant \frac{\varepsilon}{10}$ for all $x \in [p,1]$. Thus, it suffices to establish concentration of $\widetilde{G}$ around G. We first prove concentration on a discrete set of points and interpolate to the rest of the interval. Define U_i and V_i as:

$$U_i = \left\{\gamma + i \cdot \frac{\varepsilon}{10} : i \in [N] \cup \{0\} \text{ and } \gamma + i \cdot \frac{\varepsilon}{10} \leqslant 1\right\} \cup \{1\} \text{ and } V_i = F_i^{-1}(U_i).$$

and let $V = \cup_{i\in[k]} V_i$. We have for all $x \in V, i \in [k]$, by Hoeffding's inequality that:

$$|\widetilde{G}_i(x) - G_i(x)| \leqslant \frac{\varepsilon}{10} \quad \text{with probability at least } 1 - \delta/2. \tag{2}$$

We now condition on the above event as well. By combining Eqs. (1) and (2), we get:

$$\forall x \in V, i \in [k] : |\hat{G}_i(x) - G_i(x)| \leqslant \varepsilon/5, \text{and so for all } x \in V, i \in [k], \text{ we have:}$$
$$\exp\left\{-\varepsilon/5\right\} \cdot F_i(x) \leqslant \hat{F}_i(x) \leqslant \exp\left\{\varepsilon/5\right\} \cdot F_i(x).$$

We now extend from V to the rest of $[p,1]$. Note that $\hat{G}_i(x)$ is a decreasing function of x. Hence, $\hat{F}_i$ is an increasing function of x. Now, let $x \in (p,1) \setminus V$ and $i \in [k]$. We must have $x_l, x_h \in V$ with $x_l < x \leqslant x_h$ satisfying $F_i(x_h) - F_i(x_l) \leqslant \varepsilon/10$. We now get:

$$\hat{F}_i(x) \leqslant \hat{F}_i(x_h) \leqslant \exp\left\{\varepsilon/5\right\} \cdot F_i(x_h) \leqslant \exp\left\{\varepsilon/5\right\} F_i(x_l) + \varepsilon/8 \leqslant \exp\left\{\varepsilon/5\right\} F_i(x) + \varepsilon/8,$$
$$\hat{F}_i(x) \geqslant \exp\left\{-\varepsilon/5\right\} F_i(x_l) \geqslant \exp\left\{-\varepsilon/5\right\} F_i(x_h) + \varepsilon/10 \geqslant \exp\left\{-\varepsilon/5\right\} F_i(x) + \varepsilon/10.$$

The above two inequalities and our condition on ε conclude the proof. □

2.1.2 Proof of Theorem 2.2

We now leverage our effective-support recovery result to recover bid distributions on their full support (in Wasserstein distance). Under the "lower bound on density" assumption,

$$H(\eta) = \prod_{j\in[k]} F_j(\eta) \geqslant (\lambda \cdot \eta)^k. \tag{3}$$

Now, setting $\gamma = (\lambda \cdot \eta)^k$ and using Theorem 2.3 we have that $\tilde{\Theta}\left(\frac{\log(k/\delta)}{\lambda^k \cdot \eta^k \cdot \eta^2}\right)$ samples suffice to find estimates $\hat{F}_i$ such that the additive error between $\hat{F}_i$ and F_i is at most η in the interval $[\eta, 1]$. For every i, the maximum possible mass in the interval $[0, \eta]$ with respect to the measure F_i is 1. Therefore, any two measures with support $[0, \eta]$ mass at most 1 have a Wasserstein distance of at most η. Also, in the subset $[\eta, 1]$ of the support we have that since the longest distance in the support is at most 1 and $\max_{x\in[\eta,1]} \left|\hat{F}_i(x) - F_i(x)\right| \leqslant \eta$ we have that the Wasserstein distance of the measures $\hat{F}_i$ and F_i conditioned on the support $[\eta, 1]$ is at most $\varepsilon \cdot 1$. Thus,

$$\mathcal{W}(\hat{F}_i, F_i) \leqslant 2 \cdot \eta.$$

Setting $\eta = \varepsilon/2$ the theorem follows. □

2.1.3 Proof of Theorem 2.4

We are going to use the estimation $\hat{F}_i$ from Theorem 2.3 together with the Lipschitzness of f_i to prove this theorem. Let $h > 0$ and $\epsilon_0 > 0$ be parameters that we will determine later. We define, for every $x \in [p, 1]$, an density estimate

$$\hat{f}_i(x) \triangleq \frac{1}{h}(\hat{F}_i(x+h) - \hat{F}_i(x)),$$

where due to Theorem 2.3 we have $|\hat{F}_i(x+h) - F_i(x+h)| \leqslant \varepsilon_0$ and $|\hat{F}_i(x) - F_i(x)| \leqslant \varepsilon_0$ for $n = \tilde{\Theta}\left(\frac{\log(k/\delta)}{\gamma^4 \varepsilon^2}\right)$ samples. Then,

$$\begin{aligned}
\int_p^1 \left|\hat{f}_i(x) - f_i(x)\right| dx &= \int_p^1 \left|\frac{1}{h}(\hat{F}_i(x+h) - \hat{F}_i(x)) - f_i(x)\right| dx \\
&\leqslant \int_p^1 \left|\frac{1}{h}(F_i(x+h) - F_i(x)) - f_i(x)\right| dx + 2\varepsilon \\
&= \int_p^1 \left|\frac{1}{h}\left(\int_x^{x+h} f_i(z)\, dz\right) - f_i(x)\right| dx + \frac{2\varepsilon}{h} \\
&\leqslant \int_p^1 \frac{1}{h}\left(\int_x^{x+h} |f_i(z) - f_i(x)|\, dz\right) dx + \frac{2\varepsilon}{h}
\end{aligned}$$

now due to the Lipschitzness of f_i we have that

$$\begin{aligned}
\int_p^1 \left|\hat{f}_i(x) - f_i(x)\right| dx &\leqslant \int_p^1 \frac{1}{h}\left(\int_x^{x+h} L \cdot |z - x|\, dz\right) dx + \frac{2\varepsilon}{h} \\
&= \int_p^1 \frac{L}{h}\left(\frac{(x+h)^2}{2} - \frac{x^2}{2} - h \cdot x\right) dx + \frac{2\varepsilon}{h} \\
&= \int_p^1 \frac{L}{h} \cdot h^2\, dx + 2\varepsilon \leqslant L \cdot h + \frac{2\varepsilon}{h}.
\end{aligned}$$

Therefore, if we choose $h = \sqrt{\varepsilon_0/L}$ and we also set $\varepsilon = \varepsilon_0^2/(9L)$ the theorem follows.

2.2 Lower Bound for Full-Support Estimation

Here, we establish lower bounds proving the optimality of Theorem 2.2. We prove:

1. the exponential dependence on k incurred in Theorem 2.2 is necessary and
2. the distributions cannot be recovered in Kolmogorov distance in their whole support.

In both these cases, we will construct a pair of distributions $\{f_i\}_{i=1}^k$ and $\{f_i'\}_{i=1}^k$ satisfying the bounded density condition of Theorem 2.2 such that:

1. f_1 and f_1' have $\mathcal{W}(f_1, f_1') \geqslant \Omega(\varepsilon)$ and $\mathrm{d_K}(f_1, f_1') \geqslant 1/2$ and
2. Fewer than $\Omega((\lambda\varepsilon)^{-(k-1)})$ fail to distinguish them with large probability.

The main intuition behind our construction is that learning the behavior of any of the densities below ε requires observing $Y \leqslant \varepsilon$ and this only happens with probability ε^{-k}.

Theorem 2.6. *Let $k \in \mathbb{N}$, and let $\varepsilon, \lambda \in (0, 1/2)$. Then, there exist two tuples of distributions $\mathcal{D} = \{f_i\}_{i=1}^k$ and $\mathcal{D}' = \{f_i'\}_{i=1}^k$ with the first price auction model (Definition 2.1) on $\mathcal{D}, \mathcal{D}'$ satisfies the density bound condition from Theorem 2.2 such that for any estimator $\hat{\mu}$, we have:*

$$\max\left(\mathbb{P}\left\{\mathcal{W}\left(\hat{\mu}(\{(Y_i, Z_i)\}_{i=1}^n), \mathcal{D}\right) \geqslant \frac{\varepsilon}{8}\right\}, \mathbb{P}\left\{\mathcal{W}\left(\hat{\mu}(\{(Y_i', Z_i')\}_{i=1}^n), \mathcal{D}'\right) \geqslant \frac{\varepsilon}{8}\right\}\right) \geqslant \frac{1}{3}$$

where (Y_i, Z_i), (Y_i', Z_i') are drawn i.i.d from $\mathcal{D}$ and $\mathcal{D}'$ respectively if $n \leqslant \frac{1}{10} \cdot (\lambda\varepsilon)^{-(k-1)}$.

Proof. Let $\mathcal{D}_1 = \{f_1, \dots, f_k\}$ and $\mathcal{D}_2 = \{f_1', \dots, f_k'\}$ denote the two sets of distributions characterizing our first price auction model (Definition 2.1). We will have $f_i = f_i'$ for all $i > 1$.

$$f_i' = f_i = \lambda \cdot \mathrm{Unif}([0,1]) + (1-\lambda) \cdot \mathrm{Unif}([3/4, 1]) \quad \text{for all } i > 1.$$

However, f_1 and f_1' will have large Wasserstein and Kolmogorov distance:

$$\begin{aligned} f_1 &= \lambda \cdot \mathrm{Unif}([0,1]) + (1-\lambda) \cdot \mathrm{Unif}([0, \varepsilon/4]) \\ f_1' &= \lambda \cdot \mathrm{Unif}([0,1]) + (1-\lambda) \cdot \mathrm{Unif}([3\varepsilon/4, \varepsilon]). \end{aligned}$$

Let (Y, Z) and (Y', Z') be distributed according to the first price auction model with respect to $\mathcal{D}_1$ and $\mathcal{D}_2$. We now define the events E and E' on (Y, Z) and (Y', Z') as follows:

$$E = \{Y \in (\varepsilon, 1]\} \text{ and } E' = \{Y' \in (\varepsilon, 1]\} \implies \mathsf{P}\{E\} = \mathsf{P}\{E'\} = 1 - (\lambda\varepsilon)^{k-1} \tag{4}$$

By construction, (Y, Z) and (Y', Z') have the same distribution conditioned on E and E'.

Now, let $\boldsymbol{W} = \{(Y_i, Z_i)\}_{i \in [n]}$ and $\boldsymbol{W}' = \{(Y_i, Z_i)\}_{i=1}^n$ be collections of n i.i.d samples from $\mathcal{D}$ and $\mathcal{D}'$ respectively, and let $\hat{\mu}$ denote any estimator of the first price auction model. We show that $\hat{\mu}$ has large error on at least one of $\mathcal{D}$ or $\mathcal{D}'$. Letting F (respectively, F') denote the event that E (respectively, E') holds for all of the (Y_i, Z_i) (respectively, (Y_i', Z_i')), we have:

$$\mathbb{P}\left(\mathcal{W}(\hat{\mu}(\boldsymbol{W}), \mathcal{D}) \leqslant \varepsilon/8\right) = \mathbb{P}(F) \cdot \mathbb{P}\left(\mathcal{W}(\hat{\mu}(\boldsymbol{W}), \mathcal{D}) \leqslant \varepsilon/8 | F\right) + \mathbb{P}(\bar{F}) \cdot \mathbb{P}\left(\mathcal{W}(\hat{\mu}(\boldsymbol{W}), \mathcal{D}) \leqslant \varepsilon/8 | \bar{F}\right).$$

Now, if $n \leqslant \frac{1}{10}(\lambda\varepsilon)^{-(k-1)}$, we have from Eq. (4) and a union bound that $\mathbb{P}(F) \geqslant 9/10$. Furthermore, note that conditioned on F and F', $\boldsymbol{W}$ and $\boldsymbol{W}'$ have the same distribution and $\mathcal{W}(f_1, f_1') \geqslant \varepsilon/4$.

Assuming the probability in the above equation is greater than $2/3$, we may re-arrange the above equation as follows:

$$\frac{2}{3} \leqslant \mathbb{P}(F)\mathbb{P}\left\{\mathcal{W}(\hat{\mu}(\boldsymbol{W}), \mathcal{D}) \leqslant \frac{\varepsilon}{8}\middle| F\right\} + \frac{1}{10} \leqslant \mathbb{P}(F')\mathbb{P}\left\{\mathcal{W}(\hat{\mu}(\boldsymbol{W}'), \mathcal{D}') \geqslant \frac{\varepsilon}{8}\middle| F'\right\} + \frac{1}{10}.$$

By re-arranging the above equation, we have that either:

$$\mathbb{P}\left\{\mathcal{W}(\hat{\mu}(\boldsymbol{W}), \mathcal{D}) \leqslant \frac{\varepsilon}{8}\right\} \leqslant \frac{2}{3}, \text{ or } \mathbb{P}\left\{\mathcal{W}(\hat{\mu}(\boldsymbol{W}'), \mathcal{D}') \leqslant \frac{\varepsilon}{8}\right\} \leqslant \frac{2}{3}$$

concluding the proof of the theorem. □

Note that the probabilities $1/3$ chosen in the above theorem is not a substantial restriction as any algorithm successfully distinguishing between $\mathcal{D}$ and $\mathcal{D}'$ with probability bounded away from $1/2$ can be boosted to arbitrarily high probability by simple repetition. As a simple consequence of this construction, we can rule out estimation in Kolmogorov distance:

Theorem 2.7. *Let $n \in \mathbb{N}$ and $\hat{\mu}$ be an estimator for the First-Price-Auction model. Then, for all $\delta > 0$, there exists a First-Price-Auction model characterized by $\mathcal{D} = \{f_i\}_{i=1}^k$ satisfying the bounded density condition of Theorem 2.2 satisfying:*

$$\mathbb{P}\left(\mathrm{d_K}(\hat{\mu}(\boldsymbol{W}), \mathcal{D}) \leqslant \frac{1}{4}\right) \leqslant \frac{1}{2} + \delta.$$

where $\boldsymbol{W} = \{(Y_i, Z_i)\}_{i=1}^n$ are drawn i.i.d from the first price auction model on $\mathcal{D}$.

Proof. We will prove the lemma via contradiction. Let $n, \hat{\mu}$ be such that the there exists $\delta > 0$ such that for all First-Price-Auction models, $\mathcal{D}$, satisfying the bounded density condition:

$$\mathbb{P}\left(\mathrm{d_K}(\hat{\mu}(\boldsymbol{W}), \mathcal{D}) \leqslant \frac{1}{4}\right) \geqslant \frac{1}{2} + \delta.$$

Note that by repeating the experiment $\Omega(1/\delta^2)$ times, we may boost the success probability to $9/10$ by taking the pointwise median of the resulting estimates. However, from our construction in the proof of Theorem 2.6, we have by picking ε small enough in the construction that there exists a distribution, $\mathcal{D}$ such that:

$$\mathbb{P}\left\{\mathrm{d_K}\left(\hat{\mu}(\boldsymbol{W}), \mathcal{D}\right) \geqslant \frac{1}{4}\right\} \geqslant \frac{1}{3}$$

as all the distributions we construct have Kolmogorov distance greater than $1/2$ between them. This yields the contradiction, proving the theorem. □

2.3 Estimation from Partial Observations

In this section we show how our results in the previous sections can be translated to the partial observation model introduced by Blum et al. [2015] defined below.

Definition 2.8 (Partial Observation Data)**.** Let $\{F_i\}_{i=1}^k$ be k cumulative distribution functions with support $[0, 1]$, i.e. $F_i(x) = 0\ \forall x < 0$ and $F_i(1) = 1$. A sample (r, Y, Z) from a first-price auction with bid distributions $\{F_i\}_{i=1}^k$ is generated as follows:

1. we, the observer, pick a price $r \in [0,1]$, and let $X_{k+1} = r$
2. generate $X_i \sim F_i$ independently for all $i \in [k]$,
3. observe a winner $Z = \arg\max_{i \in [k+1]} X_i$.

At first glance, it seems like the access to partial observation data is more restrictive than the access to the first-price auction data that we defined in Definition 2.1. Nevertheless, we show that partial observations suffice to run the same estimation used in Subsection 2.1.

Theorem 2.9 (First-Price Auctions – Partial Observations)**.** *Let $\{Z_i\}_{i=1}^{n}$ be n i.i.d. partially observed samples from the same first-price auction as per Definition 2.8 and assume that the cumulative distribution functions F_i are continuous and admit Lipschitz-continuous densities f_i with constant L. Then, given $p, \gamma \in [0,1]$ such that $\mathbb{P}(X_{i \in [k]} \leqslant p) \geqslant \gamma$, there exists a polynomial-time estimation algorithm, that computes the cumulative distribution functions $\hat{F}_i$ for $i \in [k]$, so that for every $\varepsilon \in (0, \gamma/2]$ it holds that*

$$\mathbb{P}\left(\max_{x \in [p,1]} \left|\hat{F}_i(x) - F_i(x)\right| \leqslant \varepsilon\right) \geqslant 1 - \delta$$

for all $i \in [k]$ assuming that $n = \Theta\left(\frac{k}{\gamma^6 \epsilon^5} \log\left(\frac{k}{\gamma^2 \epsilon \alpha}\right) \log\left(\frac{L}{\gamma^2 \epsilon}\right)\right)$

Proof. Note that under the partial observation model, we can estimate $\mathbb{P}(Y \geqslant x)$ for any fixed x by setting the reserve price to x (i.e., bidding x) and counting the number of times that the planted bid wins the auction. More precisely, we can define

$$\hat{H}(x) = 1 - \frac{1}{n_1} \sum_{i=1}^{n_1} \mathbf{1}\{Z_i = k+1\}.$$

We can similarly define, for any given agent i, an estimator for the probability that the agent wins with price less than or equal to x:

$$\hat{H}_i(x) \triangleq \frac{1}{n_2} \sum_{j=1}^{n_2} \mathbf{1}\{Z_j = i \text{ at reserve price } 0\} - \frac{1}{n_2} \sum_{j=1}^{n_2} \mathbf{1}\{Z_j = i \text{ at reserve price } x\}.$$

By construction $\hat{H}_i \to H_i$ and $\hat{H} \to H$ where H_i and H are as defined earlier. Similarly to our strategy in the proof of Theorem 2.3, let

$$U = \{\gamma + i \cdot \delta : i \in \mathbb{N} \cup \{0\} \text{ and } \gamma + i \cdot \delta \leqslant 1\} \cup \{1\} \text{ and } V = H^{-1}(U) \text{ and } W = H_i^{-1}(U)$$

For convenience, define $N = |U| \leqslant \delta^{-1}$, and recall that k is the number of agents in the auction. Our first goal is to obtain a set of estimates $\hat{v}_j \approx v_j$ for the quantiles of H and another set $\hat{w}_j \approx w_j$ for the quantiles of each H_i. To accomplish this, we will run, for each $u_j \in U$, T iterations of binary search between 0 and 1. In particular, we initialize $\hat{v}_j^{(0)} = 1$ then, for each successive iteration t, a Hoeffding bound shows that

$$\mathbb{P}\left(\left|\hat{H}(\hat{v}_j^{(t)}) - H(\hat{v}_j^{(t)})\right| \geqslant \epsilon_1/2\right) < 2\exp\left\{-2\left(\epsilon_1/2\right)^2 n\right\}. \tag{5}$$

$$\mathbb{P}\left(\left|\hat{H}_i(\hat{w}_j^{(t)}) - H_i(\hat{w}_j^{(t)})\right| \geqslant \epsilon_1/2\right) < 2\exp\left\{-\left(\epsilon_1/2\right)^2 n/2\right\}. \tag{6}$$

We condition on the above events by taking a union bound over all agents, all search steps, and all points $u_i \in U$. Now, for each iteration t of the binary search:

1. If $|\hat{H}(\hat{v}_i^{(t)}) - u_t| \leqslant \epsilon_1/2$, then $|H(\hat{v}_i^{(t)}) - u_t| \leqslant \epsilon_1$, so we terminate and set $\hat{v}_i = \hat{v}_i^{(t)}$,
2. otherwise if $\hat{H}(\hat{v}_i^{(t)}) - u_t > \epsilon_1/2$ then $H(\hat{v}_i^{(t)}) > u_t$ and we search the upper interval,
3. otherwise $\hat{H}(\hat{v}_i^{(t)}) - u_t < -\epsilon_1/2$ and so $H(\hat{v}_i^{(t)}) < u_t$ and we search the lower interval.

We perform an analogous process to find the $\hat{w}_j$. This ensures the correctness of the binary search, and setting $T = \log(2L/\epsilon_1)$, where L is a Lipschitz constant of H and all H_i, guarantees that after performing this search for each u_i, we will find $\hat{V}$ and $\hat{W}$ such that

$$
\begin{aligned}
|H(\hat{v}_j) - u_j| \leqslant \epsilon_1 \qquad & \text{for all} \qquad j \in [|U|], \text{ and} \qquad (7)\\
|H_i(\hat{w}_j) - u_j| \leqslant \epsilon_1 \qquad & \text{for all} \qquad j \in [|U|] \text{ and } i \in [k] \qquad (8)\\
\text{w.p.} \qquad & 1 - 2TN \exp\left\{-2(\epsilon_1/2)^2\right\} - 2kTN \exp\left\{-(\epsilon_1/2)^2/2\right\}
\end{aligned}
$$

In order to define our approximation of G_i, we will consider the list of indices $X = V \cup W_i$, i.e., the union of the estimated quantiles of H and H_i. Using Hoeffding's inequality,

$$
\begin{aligned}
|H(x_j) - \hat{H}(x_j)| \leqslant \beta \qquad & \text{for all} \qquad j \in [|X|], \text{ and}\\
|H_i(x_j) - \hat{H}_i(x_j)| \leqslant \beta \qquad & \text{for all} \qquad j \in [|X|], \text{ and}\\
\text{w.p.} \qquad & 1 - 4kN \exp\left\{2n\beta^2\right\} \geqslant 1 - \frac{4k}{\delta} \exp\left\{2n\beta^2\right\}
\end{aligned}
$$

We further condition on the above and define an estimate of $G_i(x_j) = \int_{x_j}^1 \frac{1}{H(z)}\, dH_i(z)$,

$$
\hat{G}_i(x_j) = \sum_{s=j}^{|X|-1} \left(\hat{H}_i(x_{s+1}) - \hat{H}_i(x_s)\right) / \hat{H}(x_s).
$$

Using the mean value theorem (see the Appendix A.1 for more detail),

$$
\begin{aligned}
|G_i(\hat{v}_t) - \hat{G}_i(\hat{v}_t)| &\leqslant \frac{2}{\gamma} \cdot \sum_{s=1}^{|X|} |H_i(x_s) - \hat{H}_i(x_s)| + \max_{s \in [|X|]} \left| \frac{1}{H(x_{s+1})} - \frac{1}{\hat{H}(x_s)} \right|\\
&\leqslant \frac{2|X|\beta}{\gamma} + \frac{\delta + 2 \cdot \epsilon_1 + \beta}{\gamma^2}
\end{aligned}
$$

We now extend our approximation from the set of points $\{x_i\}$ to the entire interval $[\rho, 1]$. Note that for any x in this interval, there exists an $x_h, x_{h+1} \in X$ such that $x \in (x_h, x_{h+1}]$. Furthermore, both G_i and $\hat{G}_i$ are monotonic in x by construction, and

$$
|G_i(x_{h+1}) - G_i(x_h)| = \left| \int_{x_h}^{x_{h+1}} \frac{1}{H(z)}\, dH_i \right| \leqslant \frac{1}{\gamma} \cdot (\delta + 2\epsilon_1),
$$

since the x_i are at least as close as the quantiles of H_i, while $\frac{1}{H(z)} \leqslant \frac{1}{\gamma}$ by assumption. Using this inequality and the monotonicity of G_i yields:

$$
\begin{aligned}
G_i(x) \geqslant G_i(x_h) &\geqslant \hat{G}_i(x_{h+1}) - \frac{1}{\gamma}(\delta + 2\epsilon_1) - \frac{2|X|\beta\gamma + \delta + 2\epsilon_1 + \beta}{\gamma^2}\\
&\geqslant \hat{G}_i(x_{h+1}) - \frac{(2|X|\gamma + 1)\beta + (\gamma + 1)(\delta + 2\epsilon_1)}{\gamma^2}
\end{aligned}
$$

$$\geqslant \hat{G}_i(x) - \frac{4(k+1)\gamma\beta/\delta + 2\delta + 4\epsilon_1}{\gamma^2}$$

$$\text{Analogously, } G_i(x) \leqslant \hat{G}_i(x) + \frac{4(k+1)\gamma\beta/\delta + 2\delta + 4\epsilon_1}{\gamma^2}$$

Now, set: $\delta = \frac{\gamma^2\epsilon}{6}, \epsilon_1 = \frac{\gamma^2\epsilon}{24}$, and $\beta = \frac{\gamma\epsilon^2}{24(k+1)}$, so that $|\hat{G}_i(x) - G_i(x)| \leqslant \frac{\epsilon}{2}$ with probability

$$1 - \frac{12}{\gamma^2\epsilon}\log\left(\frac{48L}{\gamma^2\epsilon}\right)\exp\left\{-\frac{\gamma^4\epsilon^2}{1152}n\right\} - \frac{12k}{\gamma^2\epsilon}\log\left(\frac{48L}{\gamma^2\epsilon}\right)\exp\left\{-\frac{\gamma^4\epsilon^2}{4608}n\right\} - \frac{24k}{\gamma^2\epsilon}\exp\left\{-\frac{\gamma^2\epsilon^4}{288(k+1)^2}n\right\}.$$

Thus, setting

$$n = \frac{4608(k+1)^2}{\gamma^4\epsilon^4}\log\left(\frac{3}{\alpha}\frac{24k}{\gamma^2\epsilon}\log\left(\frac{48L}{\gamma^2\epsilon}\right)\right)$$

makes this probability $1-\alpha$. Now, the total number of samples required for this approach is $O(N \cdot k \cdot T \cdot n)$, concluding the proof. □

Remark 2.10 (Inserting bids may change equilibria)**.** As we highlighted above, using access to the partial observation data we can estimate the distributions H_i to within ϵ error and thus apply a similar algorithm to the ordinary first-price setting. Observe, however, that by inserting arbitrary bids to get good estimates of the functions F_i, the econometrician can affect the bidding strategy of the agents and thus interfere with the equilibrium point of the first-price auction. This is not true for our model in Definition 2.1, where the econometrician is a passive observer (in particular, observations do not interfere with the equilibrium of the agents) and hence the bid distributions can lead to an estimation of the value distributions as well (as we show in Subsection 2.4).

2.4 Estimation of Value Distributions

Theorems 2.2, 2.3, and 2.4 establish recovery results for the bid distribution of each agent. In a first-price auction at (Bayes-Nash) equilibrium, however, these bid distributions do not correspond to agents' value distributions. Instead, at equilibrium, each agent draws a value $v_i \sim G_i(\cdot)$ and bids the best responses to other agents, i.e.,

$$\beta_i(v_i) = \arg\max_b u_i(b; v_i) := \arg\max_b\,(v_i - b)\prod_{j\neq i} F_i(b). \tag{9}$$

As discussed in the introduction, in our asymmetric IPV setting, we (a) cannot write an explicit form for the optimal bid for a given value; and (b) cannot derive smoothness results for the bid distribution from smoothness assumptions on the value distribution. In fact, a unique Bayes-Nash equilibrium is not even guaranteed to exist.

Approach. Lebrun [2006] provides the following characterization of Bayes-Nash equilibria for asymmetric first-price auctions, and shows that such equilibria exist and are unique under under some (relatively mild) assumptions:

Lemma 2.11 (Lebrun [2006])**.** *Suppose the agents' values are distributed according to the right-continuous cumulative distribution functions $G_i(\cdot)$ with support $[0,1]$ and whose derivatives (i.e., the value density functions) $g_i(\cdot)$ are locally bounded away from zero. Then, a set of strategies (bid functions) $\alpha_i(\cdot) : [0,1] \rightarrow$*

$[0, \infty)$ *is a Bayesian equilibrium if and only if there exists an* $\eta \in [0,1]$ *such that the inverses* $\alpha_i(\cdot) = \beta_i^{-1}(\cdot)$ *exist, are strictly increasing, and form a solution over* $[0, \eta]$ *of the following system of differential equations:*

$$\frac{d}{db} \log G_i(\alpha_i(b)) = \frac{1}{n-1} \left(\frac{-(n-2)}{\alpha_i(b) - b} + \sum_{j \neq i} \frac{1}{\alpha_j(b) - b} \right), \ \alpha_i(0) = 0, \ \alpha_i(\eta) = 1. \tag{10}$$

If bidders are not permitted to bid above their values, and if one of the following two conditions are met, then the set of strategies $\beta_i(\cdot)$ *represents a unique equilibrium:*

(i) The value distributions have an atom at zero, i.e., $G_i(0) > 0$*,*

(ii) There exists $\delta > 0$ *such that the cumulative density function of the i-th agent's value is strictly log-concave over* $(0, \delta)$ *for all i.*

Corollary 2.12. *Rearranging Equation (10) from Lemma 2.11 yields*

$$\sum_{j \neq i} \frac{d}{db} \log G_j(\alpha_j(b)) = \frac{1}{\alpha_i(b) - b}, \quad \text{and in turn} \quad \sum_{j \neq i} \frac{f_j(b)}{F_j(b)} = \frac{1}{\alpha_i(b) - b}. \tag{11}$$

Since Lemma 2.11 guarantees that the inverse bidding strategies are strictly increasing at equilibrium, the inverse mapping theorem dictates that $G_i(v) = F_i(\beta_i(v))$. If the equilibrium strategies $\beta_i(\cdot)$ were known, we could apply our results for the bid distributions to estimate $G_i(v)$ directly. In our setting, however, we do not have access to the strategies $\beta_i(\cdot)$—in fact, a general closed form does not exist for asymmetric auctions.

Instead, we will use the characterization given by Equation (9) of the equilibrium bid as the best response to other bidders. In particular, since we have accurate estimates for each $F_i(\cdot)$, we can define the following empirical versions of each quantity introduced so far, including $\widehat{G}_i$, an estimate for the cumulative distribution functions of each agent's value:

$$\widehat{u}_i(b; v_i) = (v_i - b) \prod_{j \neq i} \widehat{F}_j(b), \quad \widehat{\beta}_i(v) = \arg\max_b \widehat{u}_i(b; v_i), \quad \widehat{G}_i(v) = \widehat{F}_i(\widehat{\beta}_i(v)). \tag{12}$$

Turning this into a formal argument requires tackling the following technical challenges:

1. Approximating the utility function via $\widehat{u}_i(\cdot; v_i)$ and efficiently maximizing the estimated utility function to find the optimal bid.

2. Showing that the maximizer of the $\widehat{u}_i(\cdot, v_i)$ is close to that of the true utility.

3. Bounding the combined error incurred from our empirical approximations.

We will start by tackling the above challenges in the effective-support regime. The following characterizes the setup as well as the additional assumptions used to estimate the value distribution:

Assumption 2.13 (Value Estimation)**.** We assume the preconditions of Lemma 2.11, as well as an upper bound on the density of the value distributions, i.e., $g_i(b) \leqslant \zeta$ for all $i \in [k]$.

Definition 2.14 (Effective Support)**.** In the effective-support setting, we are given $(p, \gamma) \in [0,1]$ such that $\prod_{i \in [k]} F_i(p) \geqslant \gamma$. This is identical to the effective-support setting for bid estimation, with the addition that p is pre-defined, since it is part of the estimation algorithm.

We are now prepared to tackle recovery of the valuation distribution in the effective-support regime. Our main result is captured in Theorem 2.15 below. After proving the result, we will show how it straightforwardly extends to full-support estimation, in a similar manner to our bid estimation results.

Theorem 2.15 (Estimation of Value Distributions – Effective Support)**.** *Let $\{(Y_i, Z_i)\}_{i=1}^n$ be n i.i.d. samples from the same first-price auction as per Definition 2.1. Under the setup of Definition 2.14, there exists a polynomial-time estimation algorithm that computes cumulative distribution functions $\hat{G}_i(\cdot)$ for $i \in [k]$ with the following guarantee:*

$$\sup_{v\in[p,1]} \left|\hat{G}_i(v) - G_i(v)\right| \leqslant \epsilon \qquad \text{if } n = \tilde{\Theta}\left(\frac{k^2\zeta^2 L^6 \log(1/\delta)}{\gamma^{10}\epsilon^6}\right) \text{ and } F_i \text{ is L-Lipschitz}$$

$$D_L\left(\hat{G}_i \cdot \mathbf{1}_{[p,1]}, G_i \cdot \mathbf{1}_{[p,1]}\right) \leqslant \epsilon \qquad \text{if } n = \tilde{\Theta}\left(\frac{k^2\zeta^2 \log(1/\delta)}{\gamma^{16}\epsilon^{12}}\right)$$

for all $i \in [k]$, where $D_L(\cdot,\cdot)$ is distance in the Lévy metric.

Remark 2.16 (Testing Lipschitzness)**.** The first guarantee given by Theorem 2.15 depends on the (global) Lipschitzness of the bid CDFs F_i. While efficiently testing global Lipschitzness of F_i from samples is impossible, if we have a specific $\epsilon_0 > 0$ in mind, we can instead test

$$\widehat{L} = \max_{|x-y|=\epsilon_0} \frac{1}{\epsilon_0}\left(\left|\widehat{F}_i(x) - \widehat{F}_i(y)\right| + 2\epsilon\right).$$

This maximization can be done efficiently in n steps, since $\widehat{F}_i$ is piecewise constant. Since we condition on accurate bid CDF estimation, we have that for any x and y,

$$\left|\widehat{F}_i(x) - \widehat{F}_i(y)\right| \geqslant |F_i(x) - F_i(y)| - 2\epsilon,$$

and so $\widehat{L} \geqslant \max_{|x-y|=\epsilon_0} \frac{1}{\epsilon_0}|F_i(x) - F_i(y)|$, and so we can use $\widehat{L}$ in place of L in the theorem.

Proof of Theorem 2.15. In this effective-support setup, Theorem 2.3 guarantees that with probability $1-\delta$, we can learn the bid CDFs on the interval $[\beta_i(\rho), 1]$ up to additive error ϵ_0, for any $\epsilon_0 > 0$, in $n = \tilde{\Theta}\left(\frac{\log(k/\delta)}{\gamma^4\epsilon_0^2}\right)$ queries. We thus assume that we have CDF estimates $\widehat{F}_i(\cdot)$ that are within ϵ_0 of the corresponding true CDFs in the effective support.

Our point of start is to show that we can estimate the approximate utility function efficiently. For each agent i, we can relabel each observed data point (Y, Z) as $(Y, \mathbf{1}_{Z=i})$ and run our estimation procedure on the corresponding two-agent auction to get piecewise-constant ϵ_0-approximations of $\prod_{j\neq i} F_j(b)$ for all $i \in [k]$. Since $v_i, b \in [0,1]$, we can condition on the event that $\widehat{u}_i(\cdot, v_i)$ is an ϵ_0-approximate estimate of the true utility function.

Now, the form of our estimate for $\prod_{j\neq i} F_j(b)$ is piecewise constant (with n pieces, where n is the sample complexity of Theorem 2.3) and monotonically increasing in b. Meanwhile, $(v_i - b)$ is strictly decreasing in b along any interval. We can thus exactly maximize $\widehat{u}_i$ by evaluating it at n locations (i.e., the beginning of each piecewise-constant interval).

Next, define b^* and $\widehat{b}$ to be the maximizer of $u_i(\cdot; v_i)$ and $\widehat{u}_i(\cdot; v_i)$ respectively over the interval $[p, 1]$. Since we have an ϵ_0-approximation of utility within this interval,

$$\left|u_i(b^*; v_i) - u_i(\widehat{b}; v_i)\right| \leqslant 2\epsilon_0. \tag{13}$$

Our next goal is to translate this proximity in utility-space to proximity in parameter-space, i.e., to show that $b^* \approx \widehat{b}$. To do so, we use the derivative of the utility function with respect to the bid, which is given by the following result:

Lemma 2.17 (Derivative of utility function)**.** *Fix any $v_i \in [0,1]$, and let $b^* = \beta_i(v_i)$ be the equilibrium bid for the i-th agent corresponding to value v_i. Let $u_i(\cdot; v_i)$ denote the utility function as defined in (9). Then,*

$$\frac{d}{db} u_i(b; v) = (\alpha_i(b^*) - \alpha_i(b)) \sum_{j \neq i} \frac{f_j(b)}{F_j(b)} \prod_{j \neq i} F_j(b).$$

Proof. Recalling the definition of the utility function from (9),

$$\begin{aligned} u_i(b; v) &= (v - b) \cdot \prod_{j \neq i} F_j(b) \\ \frac{d}{db} u_i(b; v) &= -\prod_{j \neq i} F_j(b) + (v - b) \sum_{j \neq i} f_j(b) \prod_{k \neq j,i} F_k(b) \\ &= \left((v - \alpha_i(b)) \sum_{j \neq i} \frac{f_j(b)}{F_j(b)} + (\alpha_i(b) - b) \sum_{j \neq i} \frac{f_j(b)}{F_j(b)} - 1 \right) \prod_{j \neq i} F_j(b). \end{aligned}$$

Observing that $v_i = \alpha_i(b^*)$ and using Corollary 2.12 concludes the proof. ☐

Now, returning to (13),

$$2\epsilon_0 \geqslant \left| \int_{\widehat{b}}^{b^*} (\alpha_i(b^*) - \alpha_i(x)) \cdot \frac{1}{\alpha_i(x) - x} \cdot \prod_{j \neq i} F_j(x)\, dx \right|. \qquad \text{(Corollary 2.12)} \qquad (14)$$

In order to bound $|b^* - \widehat{b}|$, we need the following lower bound on the derivative $\alpha_i'(x)$:

Lemma 2.18. *Under the conditions of Assumption 2.13, for all $b \in [\rho, 1]$,*

$$\frac{d}{db} \log\, G_i(\alpha_i(b)) > L(b) \qquad \textit{where } L(b) := \frac{\alpha_i(b) - b}{(k-1)^2 \cdot \zeta}.$$

We defer the proof of Lemma 2.18 to the Online Appendix (the proof is nearly identical to that of Lemma A-1 in [Lebrun, 2006], with the exception that we keep better track of constants to get a non-zero lower bound), and state the corollary:

Corollary 2.19. *Under the conditions of Lemma 2.18,*

$$\alpha_i'(b) = \frac{\gamma}{\zeta} \cdot \left(\frac{d}{db} \log\, G_i(\alpha_i(b)) \right) \geqslant \frac{\gamma}{(k-1)^2 \cdot \zeta^2} \cdot (\alpha_i(b) - b).$$

Now, let $\bar{b} = (\widehat{b} + b^*)/2$. By positivity of the integrand in (14),

$$2\epsilon_0 \geqslant \left| \int_{\bar{b}}^{\widehat{b}} (\alpha_i(x) - \alpha_i(\bar{b})) \cdot \frac{1}{\alpha_i(x) - x} \cdot \prod_{j \neq i} F_j(x)\, dx \right|.$$

Using the intermediate value theorem, there exists $z \in [\min(x, \bar{b}), \max(x, \bar{b})]$ so that

$$2\epsilon_0 \geqslant \left| \int_{\bar{b}}^{\widehat{b}} \frac{\left(x - \bar{b}\right) \alpha_i'(z)}{\alpha_i(x) - x} \cdot \prod_{j \neq i} F_j(x)\, dx \right| \geqslant \int_{\bar{b}}^{\widehat{b}} \frac{\gamma |x - \bar{b}|}{(k-1)^2 \cdot \zeta^2} \cdot \frac{\alpha_i(z) - z}{\alpha_i(x) - x} \cdot \prod_{j \neq i} F_j(x)\, dx.$$

Using our effective support definition and bid optimality,

$$(\alpha_i(z) - z) \cdot \prod_{j \neq i} F_j(z) \geqslant (\alpha_i(z) - p) \cdot \gamma \geqslant (z - p) \cdot \gamma \geqslant (\Delta/2) \cdot \gamma,$$

where $\Delta = |\widehat{b} - b^*|$. Thus, since $\alpha_i(x) - x \leqslant 1$,

$$\frac{\alpha_i(z) - z}{\alpha_i(x) - x} \cdot \prod_{j \neq i} F_j(x) \geqslant (\Delta/2) \cdot \gamma \cdot \left(\prod_{j \neq i} F_j(x) / \prod_{j \neq i} F_j(z) \right) \geqslant (\Delta/2) \cdot \gamma^2$$

Returning to the integral,

$$2\epsilon_0 \geqslant \frac{\Delta \gamma^3}{2(k-1)^2 \cdot \zeta^2} \cdot \int_{\bar{b}}^{\widehat{b}} |x - \bar{b}|\, dx \geqslant \frac{\Delta \gamma^3}{2(k-1)^2 \cdot \zeta^2} \cdot \left(\frac{\Delta^2}{8} \right). \tag{15}$$

Thus, for any $\epsilon_1 > 0$, setting $\epsilon_0 = \frac{\epsilon_1^3 \Gamma^3}{32 k^2 \zeta^2}$ implies that $\Delta < \epsilon_1$.

We are now ready to bound the error in our estimate of the valuation distribution. We consider two cases: first, when the bid CDFs $F_i(\cdot)$ satisfy a Lipschitz-like constraint; second, a more general setting where we only require a lower bound on the valuation densities $g_i(\cdot)$. In the first case, we learn the valuation distributions in Kolmogorov distance over the interval $[\rho, 1]$, whereas in the second case we learn in 1-Wasserstein distance.

In both cases, our estimate will be given by:

$$\widehat{G}_i(v) := \mathbf{1}\{v \in [p, 1]\} \cdot \widehat{F}_i(\widehat{b}(v))$$

Case 1: Lipschitz bid CDF In the first case, we assume that the cumulative distribution function of each bid distribution is L-Lipschitz continuous (note that any bid distribution with density bounded by L satisfies this). Then, for any $v \in [p, 1]$,

$$\left| G_i(v) - \widehat{F}_i(\widehat{b}(v)) \right| \leqslant |G_i(v) - F_i(b(v))| + \left| F_i(\widehat{b}(v)) - F_i(b(v)) \right| + \epsilon_0 \leqslant L \cdot \epsilon_1 + \epsilon_0 < 2L\epsilon_1,$$

where we recall that $\widehat{G}_i(v) := \widehat{F}_i(\widehat{b}(v))$. Thus, we define $\epsilon_1 = \epsilon/(2L)$ so that

$$\sup_{v \in [p,1]} \left| G_i(v) - \widehat{G}_i(v) \right| \leqslant \epsilon \text{ using } n = \tilde{\Theta}\left(\frac{k^2 \zeta^2 \log(1/\delta)}{\gamma^{10} \epsilon_1^6} \right) = \tilde{\Theta}\left(\frac{k^2 \zeta^2 L^6 \log(1/\delta)}{\gamma^{10} \epsilon^6} \right)$$

samples, which concludes the proof.

Case 2: General case. We can also obtain a convergence guarantee that is independent of the Lipschitz continuity of F_i by considering a slightly smaller interval $[p + d, 1]$ for some $d > 0$. In this setting, the "best response" property of bids implies that, for any $b \in [p + d, 1]$,

$$(\alpha_i(b) - b) \cdot \prod_{j \neq i} F_j(b) \geqslant (\alpha_i(b) - p) \cdot \gamma \geqslant (b - p) \cdot \gamma \geqslant d \cdot \gamma.$$

Thus from Corollary 2.12,

$$\sum_{j\neq i} \frac{f_j(b)}{F_j(b)} = \frac{1}{\alpha_i(b)-b} \leqslant \frac{1}{d\cdot\gamma} \implies f_j(b) \leqslant \frac{1}{d\cdot\gamma} \text{ for all } j\in[k].$$

Using the same method as Case I (with $L = 1/(d\cdot\gamma)$), we can guarantee that for $d > 0$,

$$\sup_{v\in[p+d,1]} \left|\widehat{G}_i(v) - G_i(v)\right| < \epsilon, \text{ as long as } n \in \tilde{\Theta}\left(\frac{k^2\zeta^2\log(1/\delta)}{\gamma^{16}d^6\epsilon^6}\right). \tag{16}$$

Setting $d = \epsilon$ concludes the proof of the Theorem. □

As a consequence of Theorem 2.15, we can accurately estimate valuation distributions in Wasserstein distance when the bid densities are lower bounded:

Theorem 2.20 (Estimation of Value Distributions – Full Support)**.** *Let $\{(Y_i, Z_i)\}_{i=1}^n$ be n i.i.d. samples from the same first-price auction as per Definition 2.1. Under the setup of Definition 2.14 and assuming $|F_i(x) - F_i(y)| \geqslant \lambda|x-y|$ for all $x, y \in [0,1]$, there exists a polynomial-time estimation algorithm that computes cumulative distribution functions $\hat{G}_i(\cdot)$ for $i\in[k]$ with the following guarantee for all $i \in [k]$:*

$$\mathcal{W}\left(\hat{G}_i(v), G_i(v)\right) \leqslant \epsilon \qquad \text{if } n = \tilde{\Theta}\left(\left(\frac{1024}{\lambda^{10}\epsilon^{10}}\right)^k \zeta^2 L^6 \log(1/\delta)\right) \text{ and } F_i \text{ is } L\text{-Lipschitz},$$

$$\mathcal{W}\left(\hat{G}_i(v), G_i(v)\right) \leqslant \epsilon \qquad \text{if } n = \tilde{\Theta}\left(\left(\frac{2048^2\cdot(11/8)^{16}\cdot(11/3)^6}{\lambda^{16}\epsilon^{22}}\right)^k \zeta^2 \log(1/\delta)\right) \text{ otherwise.}$$

Proof. We proceed identically to the proof of Theorem 2.2. In particular, for any $\eta > 0$,

$$H(\eta) = \prod_{j\in[k]} F_j(\eta) \geqslant (\lambda\cdot\eta)^k.$$

Set $\gamma = (\lambda\cdot\eta)^k$ and we use the first case of theorem Theorem 2.15, such that with

$$\tilde{\Theta}\left(\frac{k^2\zeta^2L^6\log(1/\delta)}{\lambda^{10k}\cdot\eta^{10k}\cdot\eta^6}\right)$$

samples we can find estimates $\hat{G}_i$ for all i such that the additive error between $\hat{G}_i$ and G_i is at most η in the interval $[\eta, 1]$. From here an identical argument to that of the proof of Theorem 2.2 (i.e., Subsection 2.1.2) shows that $\mathcal{W}(\hat{G}_i, G_i) \leqslant 2\cdot\eta$, after which setting $\eta = \varepsilon/2$ the first case in the theorem follows. For the second case, we use (16) with $p = 8\eta/11$, $d = 3\eta/11$, $\gamma = (8\lambda\eta/11)^k$, and $\eta = \epsilon/2$. □

3 Estimation from Second-price Auction Data

In this section, we will state and prove our main result for the estimation of bid distributions from second-price auction observations. Unlike the first-price-auction setting, our main result in this setting involves estimating the bid distributions under the *full-support* regime where we aim to obtain distributions approximating F_i up to small error in Kolmogorov distance. As in the first-price-setting, this incurs an exponential dependence on k. We leave the problem of estimation in the *effective-support* regime as an open problem for future work. (We do show in Subsection 3.4 that if the econometrician can insert bids of their own, then even just the identity of the winner and an indicator of whether the reserve price was paid suffice to estimate bid distributions over the effective support [Theorem 3.12], using a very simple algorithm.) The identification of the cumulative density functions, F_i, given access to the distribution of (Y, W) was previously established in Athey and Haile [2002] building on techniques from reliability theory Meilijson [1981]. However, this is to our knowledge, the first result establishing non-parametric finite-sample recovery from observations of second-price-auction data. We now formally introduce the observation model generating our data:

Definition 3.1. Let $\{f_i\}_{i=1}^k$ be k probability density functions on $[0, 1]$. An observation from the second-price selection model on $\{f_i\}_{i=1}^k$ is defined as follows:

1. First generate $X_i \sim f_i$ independently for $i \in [k]$
2. Define $W := \arg\max_{i \in [k]} X_i$
3. Observe the tuple (Y, W) where $Y := \max_{i \in [k] \setminus \{W\}} X_i$.

We now state the assumptions on F_i required for our guarantees to hold:

Assumption 3.2. The bid distributions F_i each admit densities $f_i(\cdot)$ satisfying $\alpha \leqslant f_i \leqslant \eta$ for some constants $\alpha, \eta > 0$.

Note that in comparison to Theorem 2.2, we require an upper bound on the densities, f_i, in addition to the lower bound property used previously. Our main result in this setting is the following theorem where we establish efficient, finite sample recovery guarantees from second-price-auction data satisfying Assumption 3.2:

Theorem 3.3. *Let $\varepsilon \in (0, 1)$ and $\boldsymbol{X} = \{(Y_i, W_i)\}_{i=1}^n$ denote n i.i.d observations from a Second-Price-Selection model (Definition 3.1) satisfying Assumption 3.2. Then, it is possible to learn in polynomial time cumulative distribution functions $\hat{F}_i$ satisfying:*

$$\sup_{x \in [0,1]} |\hat{F}_i(x) - F_i(x)| \leqslant \varepsilon \text{ with probability at least } 1 - \rho$$

as long as $\varepsilon \leqslant e^{-C^1_{\eta,\alpha} k}$ and $n \geqslant \left(\frac{1}{\varepsilon}\right)^{C^2_{\eta,\alpha} k} \log 1/\rho$ for some absolute constants $C^1_{\eta,\alpha}, C^2_{\eta,\alpha}$.

In the rest of the section, we prove Theorem 3.3. In Subsection 3.1, we present a high level overview of our proof strategy where analogously to the first-price case, we derive a differential equation relating the densities, f_i to the distribution of (Y, W) (Definition 3.1) leading to a fixed point equation satisfied by the cumulative functions $\{F_i\}$. Subsequently, in Subsection 3.2, we present a discretized version of the fixed point iteration that we analyze to prove our theorem. We carry out the formal analysis of our in Subsection 3.3.

3.1 Approach

We now provide a high-level overview of our proof of Theorem 3.3. For clarity, we will first outline the proof in the idealized population (infinite-sample) setting. In the next section, we formalize this outline and establish finite-sample guarantees.

We start by deriving a fixed point equation which plays a central part in our analysis. Note that in the population setting, we have access to the functions

$$G_i(x) = \mathbb{P}\left(W = i, Y \leqslant x\right).$$

For each $G_i(x)$, we can use the independence of the bid distributions F_i and some simple calculations to define a valid corresponding density:

$$g_i(x) = (1 - F_i(x)) \sum_{j \neq i} f_j(x) \prod_{l \neq i,j} F_l(x).$$

Rearranging and taking advantage of the product rule allows us to simplify this as

$$\prod_{j \neq i} F_j(x) = \int_0^x \frac{1}{1 - F_i(z)} g_i(z)\, dz.$$

Re-parameterizing the above by letting $U_i^* = \prod_{j \neq i} F_i$, we obtain the fixed point equation:

$$U_i^*(x) = \int_0^x \frac{1}{1 - H_i(U^*(z))} g_i(z)\, dz \text{ where } H_i(v) = \left(\prod_{j \neq i} v_j^{1/(k-1)}\right) / v_i^{(k-2)/(k-1)}$$

We divide the domain into smaller intervals and approximate $\widetilde{F}_i$ by a piecewise-constant function on each interval. The Banach fixed-point theorem is crucial to our analysis:

Theorem 3.4 (Banach Fixed-Point Theorem)**.** *Let (X, d) be a complete metric space with a contraction mapping $T : X \to X$, i.e., suppose there exists a metric d on X and a constant $\theta > 0$ such that $d(T(x), T(y)) \leqslant (1 - \theta) \cdot d(x, y)$ for all $x, y \in X$. Then T admits a unique fixed-point $x^* \in X$. Furthermore, x^* can be found: start with an arbitrary element $x_0 \in X$ and define a sequence $\{x_n\}$ by $x_n = T(x_{k-1})$ for $n \geqslant 1$. Then $x_n \to x^*$.*

Unfortunately, it turns out that the fixed point iteration just described is not contractive with respect to its input—instead, we proceed iteratively, starting with the origin and estimating $\widetilde{F}_i$ for each successive interval in turn. After solving for a set of intervals, we treat them as fixed, and construct a new fixed-point iteration for the next interval. We make this argument precise in the following sections. In doing so, the key technical difficulty faced by our approach is in the *amplification* of errors incurred at earlier stages of the algorithm into later stages. As we shall see, errors due to approximation, sampling or computation are compounded exponentially over the running of the algorithm (Lemma 3.8). Therefore, we use a careful data-based approach where samples are used to decide the widths of successive intervals, ensuring both that the fixed point equation remains contractive and crucially, that there are not too many stages where successive intervals are constructed (Lemma 3.9).

3.2 Fixed Point Definition

Here, we formally define the version of the fixed point iteration used in our algorithm. Recall that the CDFs of the bid generating distributions, F_i, satisfy:

$$\forall i \in [k] : U_i^*(x) := \prod_{j \neq i} F_j(x) = \int_0^x \frac{1}{1 - F_i(z)} \cdot g_i(z) dz.$$

Recasting the above equation in terms of the functions, U_i^*, we obtain:

$$\forall i \in [k] : U_i^*(x) := \prod_{j \neq i} F_j(x) = \int_0^x \frac{g_i(z)}{1 - H_i(z)} dz \text{ with } H_i(z) = \frac{\prod_{j \neq i} (U_j^*(z))^{1/(k-1)}}{(U_i^*(z))^{(k-2)/(k-1)}}.$$

In our algorithm, we approximate the functions, U_i^*, by piecewise constant functions on intervals of width δ (PAR) and approximate solutions to the above fixed point. We will subsequently prove that the approximation errors as well as errors due to computational and statistical constraints remain small despite these choices.

Our algorithm operates in stages: we divide the interval $[\nu, 1-\theta]$ (PAR) into a finite number of "macro-intervals", each of which contains a number of micro-intervals of width δ, defined by the points $\nu := x_0 < x_1 \cdots < x_T \leqslant 1 - \theta/2$ where T and the width of each macro-interval are chosen dynamically based on observed data to ensure the fixed point iteration remains suitably contractive. However, we must ensure that the total number of intervals, T, does not grow too rapidly as estimation errors incurred in earlier stages of the algorithm are exponentially amplified in later stages.

We construct these macro-intervals in a recursive fashion where the end point of the next interval, x_τ, is chosen based on the previous one $x_{\tau-1}$. Note the first point, x_0, is chosen to be ν. Conditioned on $x_{\tau-1} < 1 - \theta$, the end point of the subsequent macro-interval, x_τ, is defined in the following display where $\hat{U}_i$ is a function coarsely approximating U_i^* (see Lemma 3.5 for a formal definition) and $\hat{G}_i$ are empirical approximations of G_i:

$$\gamma_\ell^{(\tau)} := \max_{i \in [k]} 16 \cdot \left(\frac{\eta}{\alpha}\right)^6 \cdot \frac{1}{\hat{U}_i(x_{\tau-1})} \sum_{m=1}^{\ell} \frac{x_{\tau,m}}{(1 - x_{\tau,m})^2} \cdot \Delta_{i,m}^{(\tau)}$$

$$\Delta_{i,m}^{(\tau)} := \hat{G}_i(x_{\tau,m}) - \hat{G}_i(x_{\tau,m-1}),\ \hat{G}_i(x) := \frac{1}{n} \cdot \sum_{j=1}^{n} \mathbf{1}\left\{W_j = i, Y_j \leqslant x\right\},\ x_{\tau,l} := x_{\tau-1} + l\delta$$

$$\ell^{(\tau)} := \max\left\{\ell \in \mathbb{N} : x_{\tau,\ell} \leqslant 2x_{\tau-1}, x_{\tau,\ell} \leqslant 1 - \theta/2 \text{ and } \gamma_\ell^{(\tau)} \leqslant 1/4\right\}. \tag{MACRO}$$

We terminate once $x_\tau > 1 - \theta$. Note $x_{\tau,0} = x_{\tau-1}$ and $x_{\tau,\ell^{(\tau)}} = x_\tau$. The fixed point iteration for estimating $U_i^*(x)$ for $x \in [x_{\tau-1}, x_\tau]$ is now defined below with the $(i,l)^{th}$ entry of the variable $U^{(\tau)} \in \mathbb{R}^{k \times \ell^{(\tau)}}$, $U_{i,l}^{(\tau)}$ meant to approximate $U_i^*(x_{\tau,l})$:

$$\phi_{i,l}^{(\tau)}(U^{(\tau)}) = \mathrm{clip}\left(\sum_{m=0}^{l} \frac{1}{1 - H_{i,m}^{(\tau)}(U^{(\tau)})} \cdot \Delta_{i,m}^{(\tau)} + V_i^{(\tau)}, \frac{2}{\alpha} \cdot \hat{U}_i(x_{\tau,l}), \frac{1}{2\eta} \cdot \hat{U}_i(x_{\tau,l})\right)$$

$$H_{i,m}^{(\tau)}(U^{(\tau)}) = \max\left(\min\left(\frac{\prod_{j \neq i} (U_{j,m}^{(\tau)})^{\frac{1}{(k-1)}}}{(U_{i,m}^{(\tau)})^{\frac{(k-2)}{(k-1)}}}, 1 - \alpha(1 - x_{\tau,m}), \eta x_{\tau,m}\right), \alpha x_{\tau,m}\right) \tag{FP}$$

where $V_i^{(\tau)}$ is recursively chosen as the estimate of $U_i^*(x_{\tau-1})$ by running the fixed point iteration, $\phi^{(\tau-1)}$, L times (PAR). For initialization, we simply set $V_i^{(0)} := \hat{G}_i(\nu)$. The precise choices of our parameters are provided below:

$$\theta := \frac{\varepsilon}{16\eta}, \delta := \left(\frac{\alpha}{8\eta}\nu\right)^{32k}, \varepsilon_g := \delta \cdot \left(\frac{\alpha\nu}{2\eta}\right)^{24k}, L := \log(4/\varepsilon_g)$$

$$\nu := \min\left\{\left(\frac{\alpha}{2\eta}\right)^{256}, \exp\left(-2^{32}k\left(\frac{\eta}{\alpha}\right)\log\left(\frac{2\eta}{\alpha}\right)\right), \left(\frac{\theta}{2}\right)^{\left(\frac{4\eta}{\alpha}\right)^{16}}, \left(\frac{\alpha\varepsilon}{32\eta}\right)^{24}\right\} \tag{PAR}$$

3.3 Proof of Theorem 3.3

In this subsection, we prove Theorem 3.3. The sole probabilistic condition we require is the empirical concentration of $\hat{G}_i$ (Lemma B.4) for ε_g in PAR; i.e, we assume:

$$\forall i \in [k] : \|\hat{G}_i - G_i\|_\infty \leqslant \varepsilon_g. \tag{PROB-COND}$$

The remainder of the proof is structured as follows. In Subsection 3.3.1, we show the functions $\hat{U}_i$ in the definition of $\phi^{(\tau)}$ (FP) may be efficiently estimated from data. Subsequently, in Subsection 3.3.2, we analyze the contractivity properties of $\phi^{(\tau)}$ allowing application of Theorem 3.4. Then, in Subsection 3.3.3, we show how errors incurred in early stages of the procedure are exponentially compounded for each new *macro-interval* requiring careful control over the number of such intervals, T. Finally, we bound T and prove Theorem 3.3 in Subsection 3.3.4.

3.3.1 Approximate Estimation

Here we describe the construction of $\hat{U}_i$ used in FP, ensuring the truncation range contains the true parameter values; i.e. we establish the following lemma:

Lemma 3.5. *Let $\hat{U}_i : [0, 1-\theta/4] \to \mathbb{R}$ be monotonic functions defined as follows:*

$$\hat{U}_i(x) := \frac{1}{n}\sum_{j=1}^{n}\frac{1}{1-Y_j}\cdot \mathbf{1}\left\{W_j = i, Y_j \leqslant x\right\}.$$

Then, for all $x, y \in [0, 1-(\theta/4)]$ such that $U_i^*(x) - U_i^*(y) \geqslant \left(\frac{\alpha\nu}{2\eta}\right)^{16k}$, we have:

$$\frac{\alpha}{2}(U_i^*(x) - U_i^*(y)) \leqslant \hat{U}_i(x) - \hat{U}_i(y) \leqslant 2\eta(U_i^*(x) - U_i^*(y)).$$

Proof. Fix x, y satisfying the required constraints and consider the random variable:

$$\widetilde{U}^i = \frac{1}{(1-Y)}\cdot\mathbf{1}\left\{W = i, y < Y \leqslant x\right\}.$$

We have:

$$\mathbb{E}[\widetilde{U}^i] = \int_x^y \frac{1}{(1-z)}\cdot(1-F_i(z))\sum_{j\neq i} f_j(z)\prod_{k\neq j,i} F_k(z)dz$$

and hence, we get:

$$\alpha(U_i^*(x) - U_i^*(y)) \leqslant \mathbb{E}[\widetilde{U}^i] \leqslant \eta(U_i^*(x) - U_i^*(y)).$$

Now, we will show that the estimate U^i can be uniformly estimated for all i, x, y. For empirical analysis, we have by the integration by parts formula:

$$\hat{U}_i(x) - \hat{U}_i(y) := \frac{1}{n}\cdot\sum_{j=1}^{n}\frac{1}{1-Y_j}\cdot\mathbf{1}\left\{W_j = i, y < Y_j \leqslant x\right\}$$

$$= \left(\frac{\hat{G}_i(y)}{(1-y)} - \frac{\hat{G}_i(x)}{(1-x)} \right) - \int_y^x \frac{1}{(1-z)^2} \hat{G}_i(z) dz.$$

Similarly, we have for the population counterparts:

$$\mathbb{E}[\widetilde{U}_i] = \left(\frac{G_i(y)}{(1-y)} - \frac{G_i(x)}{(1-x)} \right) - \int_y^x \frac{1}{(1-z)^2} \cdot G_i(z) dz.$$

From the previous two displays, we get:

$$|\hat{U}_i(x) - \hat{U}_i(y) - \mathbb{E}[\widetilde{U}^i]| \leqslant \frac{16\|G_i - \hat{G}_i\|_\infty}{\theta^2}.$$

This establishes the lemma by PROB-COND and PAR. □

3.3.2 Contractivity Analysis

We first state a few simple lemmas on the behavior of the mappings, $\phi^{(\tau)}$. All the following results follow from direct calculation, provided in full in Appendix B.

Lemma 3.6. *The mapping $\phi^{(\tau)}$ maps the set $S^{(\tau)}$ defined as follows onto itself:*

$$S^{(\tau)} := \left\{ U^{(\tau)} \in \mathbb{R}^{k \times \ell^{(\tau)}} : \frac{1}{2\eta} \cdot \hat{U}_i(x_{\tau,l}) \leqslant U_{i,l}^{(\tau)} \leqslant \frac{2}{\alpha} \cdot \hat{U}_i(x_{\tau,l}) \text{ and } U_{i,l}^{(\tau)} \leqslant U_{i,l+1}^{(\tau)} \right\}.$$

Proof. The first constraint follows from FP while the second follows from the fact that ϕ^τ is a clipping of a monotonic function onto a monotonically growing range (Lemma 3.5). □

In our proof, we establish contractivity of $\phi^{(\tau)}$ in the infinity-norm; i.e, for some $\rho < 1$:

$$\|\phi^{(\tau)}(U) - \phi^{(\tau)}(U')\|_\infty \leqslant \rho \|U - U'\|_\infty \text{ where } \|M\|_\infty = \max_{i,j} |M_{i,j}|.$$

Denoting the Jacobian of $\phi^{(\tau)}$ by $J_{\phi^{(\tau)}}(\cdot)$, we bound its 1-norm defined below:

$$\|J_{\phi^{(\tau)}}(U^{(\tau)})\|_1 := \max_{i,l} \sum_{\substack{j \in [k] \\ m \in [\ell^{(\tau)}]}} \left(J_{\phi^{(\tau)}}(U^{(\tau)}) \right)_{(i,l),(j,m)} = \max_{i,l} \sum_{\substack{j \in [k] \\ m \in [\ell^{(\tau)}]}} \left| \frac{\partial \phi_{i,l}^{(\tau)}(U^{(\tau)})}{\partial U_{j,m}^{(\tau)}} \right|$$

Lemma 3.7 (Appendix B.1)**.** *We have, for all $U^{(\tau)} \in S^{(\tau)}$,*

$$\|J_{\phi^{(\tau)}}(U^{(\tau)})\|_1 \leqslant \max_{i \in [k]} 16 \cdot \left(\frac{\eta}{\alpha} \right)^6 \cdot \frac{1}{\hat{U}_i(x_{\tau-1})} \sum_{m=1}^{\ell^{(\tau)}} \frac{x_{\tau,m}}{(1 - x_{\tau,m})^2} \cdot \Delta_{i,m}^{(\tau)}.$$

3.3.3 Error Propagation

Before proceeding, we introduce some notation. Let $\widetilde{U}^{(\tau)}$ to be estimates produced when the fixed point iteration $\phi^{(\tau)}$ (FP) is run L times and $\bar{U}^{(\tau)} \in \mathbb{R}^{k \times \ell^{(\tau)}}$ be defined as:

$$\forall \tau \in [T], i \in [k], l \in \ell^{(\tau)} : \bar{U}_{i,l}^{(\tau)} := U_i^*(x_{\tau,l})$$

and is used to measure the accuracy of our estimates. Recall that we use $\{\widetilde{U}^{(\tau-1)}_{i,\ell^{(\tau-1)}}\}_{i\in[k]}$ as initializations, $V^{(\tau)}$, for the τ^{th} iteration, $\phi^{(\tau)}$ (FP). Finally, let $\widetilde{U} := [\widetilde{U}^{(1)} \cdots \widetilde{U}^{(T)}]$ and $\bar{U} := [\bar{U}^{(1)} \cdots \bar{U}^{(T)}]$ be the estimates and the true values aggregated into a single matrix. The main lemma of this subsection establishes that the error grows at most exponentially in the number of *macro-intervals*, T. Hence, it is crucial to obtain a bound on T that does not grow too rapidly with respect to ν which is carried out in Lemma 3.9.

Lemma 3.8 (Appendix B.2)**.** *We have:* $\|\widetilde{U} - \bar{U}\|_\infty \leqslant 2^T(2\eta\nu)^k$.

3.3.4 Bounding the Number of Macro Intervals

Here, we bound T (Lemma 3.9) as the most technical component of the proof and conclude the proof of Theorem 3.3. To prove Lemma 3.9, we employ a potential function argument where we track the growth of the functions U_i^* the end points of the macro-intervals $\{x_\tau\}_{\tau\in[T]}$. The key observation is that when $x_{\tau-1}$ is smaller than an appropriate constant $c_{\eta,\alpha}$, $U_i^*(x_\tau) \geqslant (U_i^*(x_{\tau-1}))^{1-1/(2(k-1))}$. Hence, the rate of growth is *doubly exponential* before $c_{\eta,\theta}$ allowing the bound on $T \approx \log_2(1/\nu) + k\log\log(1/\nu)$ while the initial error scales is at most ν^k. With Lemma 3.8, a simple post-processing step proves Theorem 3.3.

Lemma 3.9 (Appendix B.3)**.** *We have:*

$$T \leqslant 2(k-1)\log\log(1/(\alpha\nu)) + \log_2(2/\nu) + 2^{20}\left(\frac{\eta}{\alpha}\right)^{14}\left(k^2\log(2\eta/\alpha) + k\log_2(2/\theta)\right).$$

To complete the proof of Theorem 3.3, we have from Lemmas 3.8 and 3.9 and our setting of the parameter, ν, that $\|\bar{U} - \widetilde{U}\|_\infty \leqslant 2^T(2\eta\nu)^k \leqslant (2\eta\nu)^{k/8} \leqslant (\alpha\cdot\varepsilon/32\eta)^{2k}$.

We now recover estimates of F_i from estimates of U_i^*. Note, when $x \leqslant \theta$, $F_i(x) \leqslant \varepsilon/16$ and hence, 0 is suitable in this range. Likewise, when $x \geqslant 1-\theta$, $F_i(x) \geqslant 1-\varepsilon/16$ and 1 is correspondingly accurate. For the final case, assume $\theta \leqslant x \leqslant 1-\theta$. We will first estimate F_i on the grid points, $x_{\tau,l}$. Suppose now that $x = x_{\tau,l}$ for some τ, l. We have:

$$U_i^*(x) \geqslant \int_0^x \sum_{j\neq i} f_j(z)\prod_{m\neq i,j} F_m(z)dz \geqslant (k-1)\alpha^{k-1}\int_0^x z^{k-2}dz \geqslant \left(\frac{\alpha\varepsilon}{16}\right)^{k-1}.$$

And, as a consequence, we get: $\left(1-\frac{\varepsilon}{16}\right) \leqslant \frac{U_i^*(x)}{\widetilde{U}_{i,l}^{(\tau)}} \leqslant \left(1+\frac{\varepsilon}{16}\right)$. Defining our estimate:

$$\hat{F}_i(x) := \textstyle\prod_{j\neq i}(\widetilde{U}_{j,l}^{(\tau)})^{1/(k-1)} / \left((U_{i,l}^{(\tau)})^{(k-2)/(k-1)}\right)$$

we get by noting that $F_i(x) = \frac{\prod_{j\neq i}(U_j^*(x))^{1/(k-1)}}{(U_i^*(x))^{(k-2)/(k-1)}}$:

$$\hat{F}_i(x) \leqslant F_i(x)\cdot\left(1+\frac{\varepsilon}{16}\right)\cdot\left(1-\frac{\varepsilon}{16}\right)^{-1} \leqslant \left(1+\frac{\varepsilon}{4}\right)\cdot F_i(x)$$
$$\hat{F}_i(x) \geqslant F_i(x)\cdot\left(1-\frac{\varepsilon}{16}\right)\cdot\left(1+\frac{\varepsilon}{16}\right)^{-1} \geqslant \left(1-\frac{\varepsilon}{4}\right)\cdot F_i(x).$$

Finally, for any $\theta \leqslant x \leqslant 1-\theta$, there exists $x_{\tau,l}$ such that $|x - x_{\tau,l}| \leqslant \delta$. And we have:

$$\frac{|F_i(x) - \hat{F}_i(x_{\tau,l})|}{F_i(x)} \leqslant \frac{|F_i(x) - F_i(x_{\tau,l})| + |F_i(x_{\tau,l}) - \hat{F}_i(x_{\tau,l})|}{F_i(x_{\tau,l}) - |F_i(x_{\tau,l}) - F_i(x)|} \leqslant \frac{\delta\eta + (\varepsilon/4)F_i(x_{\tau,l})}{F_i(x_{\tau,l}) - \delta\eta} \leqslant \frac{\varepsilon}{2}$$

from our setting of δ and θ. This concludes the proof of the theorem. □

3.4 Estimation from Partial Observations

Finally, we explore a second-price analogue of the "partial observability" setting (introduced by Blum et al. [2015]) that we studied in the context of first-price auctions in Section 2.3. In particular, in this setting we observe the winner of each auction and a binary indicator of whether the reserve price was triggered, but not the price that the winner pays for the auctioned good. On the other hand, as in [Blum et al., 2015], in this setting the econometrician is given the ability to set the reserve price (or equivalently, insert bids into the auction). We formally define the setting below:

Definition 3.10 (Partial Observation Data – Second-price)**.** Let $\{F_i\}_{i=1}^k$ be k cumulative distribution functions with support $[0,1]$, i.e. $F_i(x) = 0\ \forall x < 0$ and $F_i(1) = 1$. A sample (r, Y, Z) from a first-price auction with bid distributions $\{F_i\}_{i=1}^k$ is generated as follows:

1. we, the observer, pick a price $r \in [0,1]$, and let $X_{k+1} = r$
2. generate $X_i \sim F_i$ independently for all $i \in [k]$,
3. observe a winner $Z = \arg\max_{i \in [k+1]} X_i$ and an indicator Q indicating whether the reserve price r was triggered.

We again operate in the effective-support setting (cf. Theorem 2.3), and—that is, we a tuple $p, \gamma \in [0,1]$ such that for all $j \in [k]$, $\prod_{l \neq j} F_l(p) \geqslant \gamma$. In other words, the transaction price of the auction will be less than p with probability at least γ.

It turns out that in this seemingly limited observation model, a very simple algorithm suffices for recovering agents' value distributions. We begin with the Lemma demonstrating pointwise recovery of the bid distributions for any $x \in [p, 1]$:

Lemma 3.11. *Fix any $x \in [p,1]$ and any $\epsilon > 0$. Using n samples from the we can obtain as estimate $\widehat{F}_{j \in [k]}(x)$ satisfying, for all $j \in [k]$,*

$$\left| \widehat{F}_j(x) - F_j(x) \right| \leqslant \epsilon \quad \textit{with probability at least } 1 - \delta,$$

as long as $n \geqslant \frac{48}{\gamma \epsilon^2} \log(2k/\delta)$.

Proof. First, suppose we set the reserve price of the auction to x, and define the random variable Z_j as the indicator of whether either (a) agent j won the auction and the reserve price was triggered; or (b) no one won the auction and the reserve price was triggered. By construction (and since (a) and (b) are disjoint),

$$\mathbb{P}(Z_j = 1) = \mathbb{P}(X_j \geqslant x, X_{-j} \leqslant x) + \mathbb{P}(X_{[k]} \leqslant x) = (1 - F_j(x)) \prod_{l \neq j} F_l(x) + \prod_{l \in [k]} F_l(x) = \prod_{l \neq j} F_l(x).$$

Thus, applying a (multiplicative) Chernoff bound combined with the lower bound $\prod_{l \neq j} F_l(x) \geqslant \gamma$ given by our effective support assumption,

$$\mathbb{P}\left(\left| \sum_{i=1}^n Z_j^{(i)} - \prod_{l \neq j} F_l(x) \right| \geqslant \epsilon \cdot \prod_{l \neq j} F_l(x) \right) \leqslant 2 \exp\left\{ -\epsilon^2 \gamma n / 3 \right\}$$

Now, by our effective support assumption, as long as $k \geqslant 2$,

$$\widehat{F}_j(x) := \frac{\prod_{l\in[k]}\left(\sum_{i=1}^n Z_j^{(i)}\right)^{\frac{1}{k-1}}}{\left(\sum_{i=1}^n Z_j^{(i)}\right)} \leqslant \frac{(1+\epsilon)^{k/(k-1)}\prod_{l\in[k]} F_l(x)}{(1-\epsilon)\prod_{l\neq j} F_l(x)} \leqslant (1+4\epsilon)F_j(x)$$

and an identical argument for the lower bound shows that $|\widehat{F}_j(x) - F_j(x)| \leqslant 4\epsilon$. Applying a union bound over all agents completes the proof. □

We can use this result to construct piecewise-constant approximations of $F_j(x)$ that is ϵ-close to the true bid distributions:

Theorem 3.12. *Assume the partially observed second-price setting, and suppose the cumulative density functions $F_{[k]}$ are all Lipschitz-continuous with Lipschitz constant L. For any pair $p, \gamma \in [0,1]$ that define an effective support, we can find piecewise-constant functions $\widehat{F}_j(\cdot)$ satisfying*

$$\sup_{x\in[p,1]}\left|\widehat{F}_j(x) - F_j(x)\right| \leqslant \epsilon \quad \textit{with probability at least } 1-\delta,$$

using $n = \Theta\left(\frac{k\log(k/\epsilon)\log(L/\epsilon)^2}{\epsilon^3\gamma}\right)$ samples from the partially observed second-price model.

Given Lemma 3.11, we can use the exact binary search and estimation procedure from Subsection 2.3 to prove Theorem 3.12—we give the full proof in Appendix B.4.

4 Conclusion

In this work, we presented efficient methods for estimating first- and second-price auctions under independent (asymmetric) private values and partial observability. Our methods come with convergence guarantees that are uniform in that their error rates do not depend on the bid/value distributions being estimated. These methods and the corresponding finite-sample guarantees build on a long line of work in Econometrics that establishes either identification results, or estimation results under restrictive assumptions such as symmetry or full bid observability.

5 Acknowledgements

This work is supported by NSF Awards CCF-1901292, DMS-2022448 and DMS2134108, a Simons Investigator Award, the Simons Collaboration on the Theory of Algorithmic Fairness, a DSTA grant, the DOE PhILMs project (DE-AC05-76RL01830), and an Open Philanthropy AI Fellowship.

References

Susan Athey and Philip Haile. Empirical models of auctions. 2006. 4

Susan Athey and Philip A Haile. Identification of standard auction models. *Econometrica*, 70(6): 2107–2140, 2002. 1, 3, 20

Susan Athey and Philip A Haile. Nonparametric approaches to auctions. *Handbook of econometrics*, 6:3847–3965, 2007. 1, 3

Susan Athey, Jonathan Levin, and Enrique Seira. Comparing open and sealed bid auctions: Evidence from timber auctions. *The Quarterly Journal of Economics*, 126(1):207–257, 2011. 1, 4

Patrick Bajari and Lixin Ye. Deciding between competition and collusion. *Review of Economics and statistics*, 85(4):971–989, 2003. 3

Laura H Baldwin, Robert C Marshall, and Jean-Francois Richard. Bidder collusion at forest service timber sales. *Journal of Political Economy*, 105(4):657–699, 1997. 3

Avrim Blum, Yishay Mansour, and Jamie Morgenstern. Learning valuation distributions from partial observation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 29, 2015. 2, 4, 5, 11, 26

Johannes Brustle, Yang Cai, and Constantinos Daskalakis. Multi-item mechanisms without item-independence: Learnability via robustness. In *Proceedings of the 21st ACM Conference on Economics and Computation*, pages 715–761, 2020. 2

Sandra Campo, Isabelle Perrigne, and Quang Vuong. Asymmetry in first-price auctions with affiliated private values. *Journal of Applied Econometrics*, 18(2):179–207, 2003. ISSN 08837252, 10991255. 3

S Donald, H Paarsch, and Jacques Robert. An empirical model of multi-unit sequential, ascending-price auctions. *Document de Travail, University of Iowa*, 2003. 3

Stephen G Donald and Harry J Paarsch. Identification, estimation, and testing in parametric empirical models of auctions within the independent private values paradigm. *Econometric Theory*, 1996. 3, 4

A. Dvoretzky, J. Kiefer, and J. Wolfowitz. Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *Ann. Math. Statist.*, 27:642–669, 1956. ISSN 0003-4851. doi: 10.1214/aoms/1177728174. URL https://doi.org/10.1214/aoms/1177728174. 8, 39

Emmanuel Guerre, Isabelle Perrigne, and Quang Vuong. Optimal nonparametric estimation of first-price auctions. *Econometrica*, 68(3):525–574, 2000. 1, 3

Philip Haile, Han Hong, and Matthew Shum. Nonparametric tests for common values at first-price sealed-bid auctions, 2003. 3

Philip A Haile. Auctions with resale markets: An application to us forest service timber sales. *American Economic Review*, 91(3):399–427, 2001. 3

Yingyao Hu, David McAdams, and Matthew Shum. Identification of first-price auctions with non-separable unobserved heterogeneity. *Journal of Econometrics*, 174(2):186–193, 2013. 3

Elena A. Krasnokutskaya. Identification and estimation of auction models with unobserved heterogeneity. *The Review of Economic Studies*, 78:293–327, 2011. 3

Jean Jacques Laffont and Quang Vuong. Structural econometric analysis of descending auctions. *European Economic Review*, 37(2-3):329–341, 1993. 3

Jean-Jacques Laffont, Herve Ossard, and Quang Vuong. Econometrics of first-price auctions. *Econometrica: Journal of the Econometric Society*, pages 953–980, 1995. 3

Bernard Lebrun. Uniqueness of the equilibrium in first-price auctions. 2006. 14, 17, 30

Yao Luo and Ruli Xiao. Identification of auction models using order statistics. *Available at SSRN 3599045*, 2020. 3

Eric Mbakop. Identification of auctions with incomplete bid data in the presence of unobserved heterogeneity. Technical report, Working paper, Northwestern University, 2017. 3

Isaac Meilijson. Estimation of the lifetime distribution of the parts from the autopsy statistics of the machine. *Journal of Applied Probability*, pages 829–838, 1981. 1, 2, 20

Konrad Menzel and Paolo Morganti. Large sample properties for estimators based on the order statistics approach in auctions. *Quantitative Economics*, 4(2):329–375, 2013. 1, 3

Paolo Riccardo Morganti. Estimating auction models from transaction prices with extreme value theory. 2011. 1, 3

R. Norvaiša. Chain rules and p-variation. *Studia Math*, 149:197–238, 2002. 7

Harry J Paarsch. Deciding between the common and private value paradigms in empirical models of auctions. *Journal of econometrics*, 51(1-2):191–215, 1992. 3

A Omitted Proofs for First-Price Auctions

A.1 Omitted Calculations from Proof of Theorem 2.9

We further condition on the above and define an estimate of $G_i(x_j) = \int_{x_j}^{1} \frac{1}{H(z)}\, dH_i(z)$,

$$\hat{G}_i(x_j) = \sum_{s=j}^{|X|-1} \left(\hat{H}_i(x_{s+1}) - \hat{H}_i(x_s)\right) / \hat{H}(x_s).$$

Using the mean value theorem, there exists a set of points $\{\zeta_j\}$ with $\zeta_j \in (x_j, x_{j+1}]$ and

$$G_i(x_j) = \int_{x_j}^{1} \frac{1}{H(z)}\, dH_i(z) = \sum_{s=j}^{|X|-1} \frac{H_i(x_{s+1}) - H_i(x_s)}{H(\zeta_s)}.$$

We first bound the difference between our piecewise estimate and the true G_i on the set X:

$$\begin{aligned}
|G_i(\hat{v}_t) - \hat{G}_i(\hat{v}_t)| &= \left| \sum_{s=j}^{|X|-1} \frac{H_i(x_{s+1}) - H_i(x_s)}{H(\zeta_s)} - \sum_{s=j}^{|X|-1} \frac{\hat{H}_i(x_{s+1}) - \hat{H}_i(x_s)}{\hat{H}(x_s)} \right| \\
&\leqslant \left| \sum_{s=j}^{|X|-1} \frac{H_i(x_{s+1}) - H_i(x_s)}{H(\zeta_s)} - \sum_{s=j}^{|X|-1} \frac{\hat{H}_i(x_{s+1}) - \hat{H}_i(x_s)}{H(\zeta_s)} \right| \\
&\quad + \left| \sum_{s=j}^{|X|-1} \frac{\hat{H}_i(x_{s+1}) - \hat{H}_i(x_s)}{H(\zeta_s)} - \sum_{s=j}^{|X|-1} \frac{\hat{H}_i(x_{s+1}) - \hat{H}_i(x_s)}{\hat{H}(x_s)} \right| \\
&\leqslant \frac{2}{\gamma} \cdot \sum_{s=1}^{|X|} |H_i(x_s) - \hat{H}_i(x_s)| + \max_{s \in [|X|]} \left| \frac{1}{H(x_{s+1})} - \frac{1}{\hat{H}(x_s)} \right| \\
&\leqslant \frac{2|X|\beta}{\gamma} + \frac{\delta + 2 \cdot \epsilon_1 + \beta}{\gamma^2} \\
&\leqslant \frac{2|X|\beta}{\gamma} + \frac{1}{\gamma^2} \max_{s \in [|X|]} \left| H(\zeta_s) - \hat{H}(x_s) \right|
\end{aligned}$$

A.2 Proof of Lemma 2.18

We will proceed similarly to the proof of [Lebrun, 2006], who use a similar technique to prove strict monotonicity (i.e., a lower bound of zero). In particular, proving a quantitative lower bound requires carefully controlling additional terms that cancel in the original proof.

For $1 \leqslant i \leqslant n$, we define

$$b_i' = \inf\left\{ b' \in [0,1] : \frac{d}{db} \log\left(G_i(v_i(b))\right) > L(b) \ \text{ for all } b \in (b', 1] \right\},$$

and let i be such that $b_i' = \max_{1 \leqslant k \leqslant n} b_k'$. Our goal is to prove that $b_i' < \rho$, since (by construction) this would imply that our desired property is true on the entire range.

By continuity of $(d/db) \log G_i(v_i(b))$ and of $L(b)$, at the point b_i' we must have that

$$\frac{d}{db} \log G_i(\alpha_i(b)) = L(b).$$

Suppose that $b_i' \geqslant \rho$—by our definition of effective support, $G_i(\alpha_i(b_i')) \geqslant \gamma$. Re-arranging the characterization of the Bayes-Nash equilibrium (10) (cf. Lemma 2.11),

$$(v_i(b) - b) \cdot \frac{d}{db} \log\left(G_i(v_i(b))\right) = \frac{1}{n-1}\left(-(n-2) + \sum_{j \neq i} \frac{v_i(b) - b}{v_j(b) - b}\right).$$

Taking the derivative with respect to b yields

$$D(b) = \sum_{j \neq i} \frac{v_i'(b)}{v_j(b) - b} - \sum_{j \neq i} \frac{(v_i(b) - b)v_j'(b)}{(v_j(b) - b)^2} + \sum_{j \neq i} \frac{v_i(b) - v_j(b)}{(v_j(b) - b)^2}. \tag{17}$$

Our next goal is to upper-bound the value of (17) at b_i'. First, note that for all $j \neq i$, our construction of b_i' implies that $b_i' \in [b_j', 1]$ (since $i = \arg\max_k b_k'$), and so

$$\frac{1}{v_i(b_i') - b_i'} - \frac{1}{v_j(b_i') - b_i'} = \frac{d}{db} \log\left(G_j(v_j(b_i'))\right) - \frac{d}{db} \log\left(G_i(v_i(b_i'))\right) \geqslant 0,$$

meaning that

$$\alpha_i(b_i') \leqslant \alpha_j(b_i') \implies \frac{v_i(b) - v_j(b)}{(v_j(b) - b)^2} \leqslant 0. \tag{18}$$

Thus, we can safely ignore the (negative) final term when upper bounding (17). Turning our attention to the second term, (11) implies the existence of at least one $j \neq i$ such that

$$\begin{aligned} &\frac{d}{db} \log\, G_j(\alpha_j(b)) \geqslant \frac{1}{n-1} \frac{1}{v_i(b) - b}, \text{ and rearranging,} \\ &(v_i(b) - b)v_j'(b_i') \geqslant \frac{G_j(v_j(b_i'))}{(n-1) \cdot g_j(v_j(b_i'))} \geqslant \frac{\gamma}{(n-1) \cdot \eta}. \end{aligned} \tag{19}$$

Since $v_j(b), b \in [0,1]$ and $v_j(b_i') \geqslant v_i(b_i')$, we have $0 \leqslant v_j(b_i') - b_i' \leqslant 1$, and in turn

$$\frac{(v_i(b) - b)v_j'(b_i')}{(v_j(b_i') - b_i')^2} \geqslant \frac{\gamma}{(n-1) \cdot \eta} \tag{20}$$

for at least one $j \neq i$, and thus

$$-\sum_{j \neq i} \frac{(v_i(b) - b)v_j'(b_i')}{(v_j(b_i') - b_i')^2} \leqslant -\frac{\gamma}{(n-1) \cdot \eta}. \tag{21}$$

Finally, recall that

$$L(b_i') = \frac{v_i'(b_i') \cdot g_i(\alpha_i(b_i'))}{G_i(\alpha_i(b_i'))} \implies v_i'(b_i') = \frac{L(b_i') \cdot G_i(\alpha_i(b_i'))}{g_i(\alpha_i(b_i'))}.$$

Combining this with (18) and (19), and using that $\alpha_j(b_i') \geqslant \alpha_i(b_i')$ yields

$$D(b_i') \leqslant \frac{-\gamma}{(n-1) \cdot \eta} + \sum_{j \neq i} \frac{L(b_i')}{v_j(b_i') - b_i'} \leqslant \frac{-\gamma}{(n-1) \cdot \eta} + \frac{(n-1) \cdot L(b_i')}{\alpha_i(b_i') - b_i'} = 0,$$

where the last equality is by definition of L. Meanwhile, by definition

$$
\begin{aligned}
D(b) &= \frac{d}{db}\left\{(v_i(b)-b)\cdot\frac{d}{db}\log\left(G_i(v_i(b))\right)\right\} \\
&= \left(v_i'(b)-1\right)\cdot\frac{d}{db}\log\, G_i(\alpha_i(b)) + (v_i(b)-b)\cdot\frac{d^2}{db^2}\log\, G_i(v_i(b)) \\
D(b_i') &\geqslant (0-1)\cdot L(b_i') + (v_i(b_i')-b_i')\cdot\frac{d^2}{db^2}\log\, G_i(v_i(b)). \qquad (22)
\end{aligned}
$$

Combining the above results yields $(v_i(b_i')-b_i')\cdot\frac{d^2}{db^2}\log\, G_i(v_i(b)) \leqslant -L(b_i') < 0$, and since $v_i(b_i')-b_i' > 0$, this must mean $\frac{d}{db}\log\, G_i(\alpha_i(b)) < 0$. However, this in turn implies that there exists an $\epsilon > 0$ such that $\log\, G_i(\alpha_i(b_i')) < x$, which is a contradiction of our definition of $\beta_i(')$. Thus, our initial assumption ($b_i' > \delta$) is impossible, which proves the desired result.

B Omitted Proofs for Second-Price Auctions

B.1 Proof of Lemma 3.7

Considering a term in the Jacobian of $\phi^{(\tau)}$, we get from Lemmas 3.5, 3.6 and B.3:

$$\left|\frac{\partial \phi_{i,l}^{(\tau)}(U^{(\tau)})}{\partial U_{j,m}^{(\tau)}}\right| = \left|\frac{\partial W_{i,l}^{(\tau)}(U^{(\tau)})}{\partial U_{j,m}^{(\tau)}} \cdot \mathbf{1}\left\{\frac{1}{2\eta}\cdot \hat{U}_i(x_{\tau,l}) \leqslant W_{i,l}^{(\tau)}(U^{(\tau)}) \leqslant \frac{2}{\alpha}\cdot \hat{U}_i(x_{\tau,l})\right\}\right|$$

$$\leqslant \left|\frac{1}{k-1} - \mathbf{1}\{i=j\}\right| \cdot \left|\frac{1}{(1-H_{i,m}^{(\tau)}(U^{(\tau)}))^2}\cdot\left(\frac{1}{U_{j,m}^{(\tau)}}\cdot H_{i,m}^{(\tau)}(U^{(\tau)})\right)\cdot \Delta_{i,m}^{(\tau)}\right|$$

$$\leqslant \frac{\eta}{\alpha^2}\cdot \left|\frac{1}{k-1} - \mathbf{1}\{i=j\}\right| \cdot \left|\frac{x_{\tau,m}}{(1-x_{\tau,m})^2}\cdot \frac{1}{U_{j,m}^{(\tau)}}\cdot \Delta_{i,m}^{(\tau)}\right|$$

$$\leqslant \frac{\eta}{\alpha^2}\cdot 4 \cdot \frac{\eta}{\alpha}\cdot \left|\frac{1}{k-1} - \mathbf{1}\{i=j\}\right| \cdot \left|\frac{x_{\tau,m}}{(1-x_{\tau,m})^2}\cdot \frac{1}{U_j^*(x_{\tau,m})}\cdot \Delta_{i,m}^{(\tau)}\right|$$

$$\leqslant 4\cdot \frac{\eta}{\alpha^2}\cdot \left(\frac{\eta}{\alpha}\right)^4 \cdot \left|\frac{1}{k-1} - \mathbf{1}\{i=j\}\right| \cdot \left|\frac{x_{\tau,m}}{(1-x_{\tau,m})^2}\cdot \frac{1}{U_i^*(x_{\tau,m})}\cdot \Delta_{i,m}^{(\tau)}\right|$$

$$\leqslant 4\cdot \frac{1}{\alpha}\cdot \left(\frac{\eta}{\alpha}\right)^5 \cdot 2\eta \cdot \left|\frac{1}{k-1} - \mathbf{1}\{i=j\}\right| \cdot \left|\frac{x_{\tau,m}}{(1-x_{\tau,m})^2}\cdot \frac{1}{\hat{U}_i(x_{\tau,0})}\cdot \Delta_{i,m}^{(\tau)}\right|$$

$$\leqslant 8\cdot \left(\frac{\eta}{\alpha}\right)^6 \cdot \left|\frac{1}{k-1} - \mathbf{1}\{i=j\}\right| \cdot \left|\frac{x_{\tau,m}}{(1-x_{\tau,m})^2}\cdot \frac{1}{\hat{U}_i(x_{\tau-1})}\cdot \Delta_{i,m}^{(\tau)}\right|.$$

Summing the previous equation over all j, m yields:

$$\left\|J_{\phi^{(\tau)}}\right\|_1 \leqslant \max_{i\in[k]} 16\cdot \left(\frac{\eta}{\alpha}\right)^6 \cdot \frac{1}{\hat{U}_i(x_{\tau-1})}\sum_{m=1}^{\ell^{(\tau)}} \frac{x_{\tau,m}}{(1-x_{\tau,m})^2}\cdot \Delta_{i,m}^{(\tau)}.$$

□

B.2 Proof of Lemma 3.8

Proof. We will prove the lemma by induction on the number of macro intervals. Concretely, we will prove the following claim via induction on τ:

$$\|\widetilde{U}^{(\tau)} - \bar{U}^{(\tau)}\| \leqslant 2^{\tau}(2\eta\nu)^k. \tag{IND}$$

For the base case, we have:

$$|\hat{G}_i(\nu) - U_i^*(\nu)| \leqslant \varepsilon_g + |G_i(\nu) - U_i^*(\nu)| = \varepsilon_g + \int_0^{\nu} (1-(1-F_i(z)))\cdot \sum_{j\neq i} f_j(z) \prod_{m\neq i,j} F_m(z)dz$$

$$\leqslant \varepsilon_g + \eta\nu \int_0^{\nu} \sum_{j\neq i} f_j(z) \prod_{m\neq i,j} F_m(z)dz \leqslant \varepsilon_g + \eta\nu U_i^*(\nu) \leqslant (2\eta\nu)^k$$

For the induction step, suppose IND is true for all intervals up to $\tau - 1$. From Lemma 3.7, $\phi^{(\tau)}$ is $1/4$ contractive. Letting $U_{\text{fixed}}^{(\tau)}$ denote the fixed point of $\phi^{(\tau)}$ (Theorem 3.4), we have:

$$\|\widetilde{U}^{(\tau)} - U_{\text{fixed}}^{(\tau)}\|_\infty \leqslant \frac{1}{4^L}. \tag{23}$$

Hence, it suffices to bound the error between $U_{\text{fixed}}^{(\tau)}$ and $\bar{U}^{(\tau)}$:

$$\begin{aligned}\|\bar{U}^{(\tau)} - U_{\text{fixed}}^{(\tau)}\|_\infty &\leqslant \|\bar{U}^{(\tau)} - \phi^{(\tau)}(\bar{U}^{(\tau)})\|_\infty + \|\phi^{(\tau)}(\bar{U}^{(\tau)}) - U_{\text{fixed}}^{(\tau)}\|_\infty \\ &= \|\bar{U}^{(\tau)} - \phi^{(\tau)}(\bar{U}^{(\tau)})\|_\infty + \|\phi^{(\tau)}(\bar{U}^{(\tau)}) - \phi^{(\tau)}(U_{\text{fixed}}^{(\tau)})\|_\infty \\ &\leqslant \|\bar{U}^{(\tau)} - \phi^{(\tau)}(\bar{U}^{(\tau)})\|_\infty + \|\bar{U}^{(\tau)} - U_{\text{fixed}}^{(\tau)}\|_\infty/4. \end{aligned} \tag{24}$$

For the RHS, we have for fixed $i \in [k], \ell \in [\ell^{(\tau)}]$:

$$\begin{aligned}&\left|\bar{U}_{i,\ell}^{(\tau)} - (\phi^{(\tau)}(\bar{U}^{(\tau)}))_{i,\ell}\right| = \left|\bar{U}_{i,\ell} - \sum_{l=1}^{\ell} \frac{1}{1 - H_{i,l}^{(\tau)}(\bar{U}^{(\tau)})}\Delta_{i,l}^{(\tau)} - V_i^{(\tau)}\right| \\ &= \left|\int_{x_{\tau-1}}^{x_{\tau,\ell}} \frac{1}{1-F_i(z)} \cdot g_i(z)dz + U_i^*(x_{\tau-1}) - \sum_{l=1}^{\ell} \frac{1}{1-F_i(x_{\tau,l})}\Delta_{i,l}^{(\tau)} - V_i^{(\tau)}\right| \\ &\leqslant \left|\int_{x_{\tau-1}}^{x_{\tau,\ell}} \frac{1}{1-F_i(z)} \cdot g_i(z)dz - \sum_{l=1}^{\ell} \frac{1}{1-F_i(x_{\tau,l})}\Delta_{i,l}^{(\tau)}\right| + \left|U_i^*(x_{\tau-1}) - V_i^{(\tau)}\right| \\ &\leqslant \left|\int_{x_{\tau-1}}^{x_{\tau,\ell}} \frac{1}{1-F_i(z)} \cdot g_i(z)dz - \sum_{l=1}^{\ell} \frac{1}{1-F_i(x_{\tau,l})}\Delta_{i,l}^{(\tau)}\right| + \left\|\widetilde{U}^{(\tau-1)} - \bar{U}^{(\tau-1)}\right\|_\infty. \end{aligned} \tag{25}$$

For the first term in the above expression, we have:

$$\begin{aligned}&\left|\int_{x_{\tau-1}}^{x_{\tau,\ell}} \frac{1}{1-F_i(z)} \cdot g_i(z)dz - \sum_{l=1}^{\ell} \frac{1}{1-F_i(x_{\tau,l})}\Delta_{i,l}^{(\tau)}\right| \\ &= \left|\sum_{l=1}^{\ell} \int_{x_{\tau,l-1}}^{x_{\tau,l}} \frac{1}{1-F_i(z)} \cdot g_i(z)dz - \frac{1}{1-F_i(x_{\tau,l})}(\hat{G}_i(x_{\tau,l}) - \hat{G}_i(x_{\tau,l-1}))\right| \\ &\leqslant \left|\sum_{l=1}^{\ell} \int_{x_{\tau,l-1}}^{x_{\tau,l}} \left(\frac{1}{1-F_i(z)} - \frac{1}{1-F_i(x_{\tau,l})}\right) \cdot g_i(z)dz\right| + \\ &\qquad \left|\sum_{l=1}^{\ell} \frac{1}{1-F_i(x_{\tau,l})}(\hat{G}_i(x_{\tau,l}) - \hat{G}_i(x_{\tau,l-1}) - (G_i(x_{\tau,l}) - G_i(x_{\tau,l-1})))\right| \\ &\leqslant \sum_{l=1}^{\ell} \int_{x_{\tau,l-1}}^{x_{\tau,l}} \frac{(F_i(x_{\tau,l}) - F_i(z))}{(1-F_i(z))(1-F_i(x_{\tau,l}))} \cdot g_i(z)dz + \frac{8\varepsilon_g \ell^{(\tau)}}{\alpha\theta} \\ &\leqslant \frac{4\eta\delta}{(\alpha\theta)^2} \int_{x_{\tau,0}}^{x_{\tau,\ell}} g_i(z)dz + \frac{8\varepsilon_g \ell^{(\tau)}}{\alpha\theta} \leqslant \frac{4\eta\delta}{(\alpha\theta)^2} + \frac{8\varepsilon_g \ell^{(\tau)}}{\alpha\theta}. \end{aligned} \tag{26}$$

Eqs. (24) to (26) and (23) conclude the induction step with PAR and IND.

□

B.3 Proof of Lemma 3.9

We start by stating and proving some useful Lemmata:

Lemma B.1. *We have, for all* $\forall i \in [k], x \in [0,1] : g_i(x) \leqslant \frac{\eta^2}{\alpha}$.

Proof. Fix $x \in [0,1]$ and let $i^* = \arg\max_{i\in[k]} F_i(x)$. For any $i \in [k]$, we get:

$$\begin{aligned}
g_i(x) &= (1 - F_i(x)) \sum_{j\neq i} f_j(x) \prod_{\ell\neq i,j} F_\ell(x) \leqslant \eta(1 - F_i(x)) \sum_{j\neq i} \prod_{\ell\neq i,j} F_\ell(x) \\
&\leqslant \left(\frac{\eta^2}{\alpha}\right) \cdot (1 - F_{i^*}(x)) \sum_{j\neq i} \prod_{\ell\neq i,j} F_\ell(x) \leqslant \left(\frac{\eta^2}{\alpha}\right) \cdot (1 - F_{i^*}(x)) \sum_{j\neq i} \prod_{\ell\neq i,j} F_{i^*}(x) \\
&= (k-1)\left(\frac{\eta^2}{\alpha}\right)(1 - F_{i^*}(x))(F_{i^*}(x))^{k-2} \leqslant \left(\frac{\eta^2}{\alpha}\right) \cdot \left(1 - \frac{1}{k-1}\right)^{k-2} \leqslant \frac{\eta^2}{\alpha}
\end{aligned}$$

where the first two inequalities follow from Assumption 3.2. □

Lemma B.2. *We have*

$$\begin{gathered}
\text{(a) } \forall \tau \in [T] : \ell^{(\tau)} > 0, \qquad \text{(b) } x_T \geqslant 1 - \theta \\
\text{(c) } \forall \tau \in [T] \text{ s.t } x_{\tau-1,\ell^{(\tau)}+1} \leqslant \min(2x_{\tau-1}, 1 - \theta/2) : \gamma^{(\tau)}_{\ell^{(\tau)}} \geqslant 1/8 \\
\text{(d) } \forall \tau \in [T], i \in [k] : U_i^*(x_\tau) - U_i^*(x_{\tau-1}) \geqslant \frac{1}{2048} \cdot \left(\frac{\alpha}{\eta}\right)^{10} \cdot \theta \cdot \left(\frac{\alpha\nu}{2}\right)^{k-1}.
\end{gathered}$$

Proof. We start with the first claim. From Eq. (FP), we have $x_{\tau-1} < 1 - \theta$. Furthermore, we have $x_{\tau-1,1} < 2x_{\tau-1}$ from our definition of δ and ν PAR. For the second condition, from Lemma B.1 and our bounds on $\varepsilon_g, \theta, \delta$ (PAR) and $\hat{U}_i(x_{\tau-1})$ (Lemma 3.5):

$$\begin{aligned}
\gamma_1^{(\tau)} &:= \max_{i\in[k]} 8 \cdot \left(\frac{\eta}{\alpha}\right)^6 \cdot \frac{1}{\hat{U}_i(x_{\tau-1})} \cdot \frac{x_{\tau-1} + \delta}{(1 - x_{\tau-1} - \delta)^2} \cdot \Delta^{(\tau)}_{i,m} \\
&\leqslant \max_{i\in[k]} 8 \cdot \left(\frac{\eta}{\alpha}\right)^6 \cdot \frac{1}{\hat{U}_i(x_{\tau-1})} \cdot \frac{x_{\tau-1} + \delta}{(1 - x_{\tau-1} - \delta)^2} \cdot (\hat{G}_i(x_{\tau-1} + \delta) - \hat{G}_i(x_{\tau-1})) \\
&\leqslant \max_{i\in[k]} 8 \cdot \left(\frac{\eta}{\alpha}\right)^6 \cdot \frac{1}{\hat{U}_i(x_{\tau-1})} \cdot \frac{x_{\tau-1} + \delta}{(1 - x_{\tau-1} - \delta)^2} \cdot \left(\frac{\eta^2\delta}{\alpha} + 2\varepsilon_g\right) < \frac{1}{4}
\end{aligned}$$

establishing the first claim. Note that the previous argument also establishes the second claim as if $x_T < 1 - \theta$, a new macro-interval exists (FP).

For the third claim, we have:

$$\begin{aligned}
\gamma^{(\tau)}_{\ell^{(\tau)}} &= \max_{i\in[k]} 16 \cdot \left(\frac{\eta}{\alpha}\right)^6 \cdot \frac{1}{\hat{U}_i(x_{\tau-1})} \sum_{m=1}^{\ell^{(\tau)}} \frac{x_{\tau,m}}{(1 - x_{\tau,m})^2} \cdot \Delta^{(\tau)}_{i,m} \\
&= \max_{i\in[k]} 16 \cdot \left(\frac{\eta}{\alpha}\right)^6 \cdot \frac{1}{\hat{U}_i(x_{\tau-1})} \left(\sum_{m=1}^{\ell^{(\tau)}+1} \frac{x_{\tau,m}}{(1 - x_{\tau,m})^2} \Delta^{(\tau)}_{i,m} - \frac{x_{\tau,\ell^{(\tau)}+1}}{(1 - x_{\tau,\ell^{(\tau)}+1})^2} \Delta^{(\tau)}_{i,\ell^{(\tau)}+1}\right) \\
&\geqslant \gamma^{(\tau)}_{\ell^{(\tau)}+1} - 16 \cdot \left(\frac{\eta}{\alpha}\right)^6 \cdot \frac{1}{\hat{U}_i(x_{\tau-1})} \cdot \frac{1 - \theta/4}{(\theta/4)^2} \cdot \left(\frac{\eta^2\delta}{\alpha} + 2\varepsilon_g\right) \geqslant \frac{1}{8}
\end{aligned}$$

where the final inequality follows from Lemma B.1, the fact that $\gamma^{(\tau)}_{\ell^{(\tau)}+1} \geqslant 1/4$, $\delta < \theta/4$ and our bounds on $\varepsilon_g, \delta, \theta$ (PAR) and $\hat{U}_i(x_{\tau-1})$ as in the previous claim.

For the final claim, suppose $\tau \in [T]$. From FP, $x_{\tau-1} < 1 - \theta$. We first consider the case where $\gamma^{(\tau)}_{\ell^{(\tau)}} \geqslant 1/8$. In this case, we have for some i from the facts $x_{\tau,m} < 1 - \theta/2$, $x_{\tau,m} \geqslant \nu$ and $\hat{U}_i(x_{\tau-1}) \geqslant \hat{G}_i(\nu) \geqslant (1 - \eta\nu)(\alpha\nu)^{k-1} - \varepsilon_g$:

$$64\left(\frac{\eta}{\alpha}\right)^6 \frac{1}{(\alpha\nu)^{k-1}} \frac{1}{\theta} \sum_{m=1}^{\ell^{(\tau)}} \frac{1}{(1-x_{\tau,m})} \Delta_{i,m}^{(\tau)} \geqslant 16\left(\frac{\eta}{\alpha}\right)^6 \frac{1}{\hat{U}_i(x_{\tau-1})} \sum_{m=1}^{\ell^{(\tau)}} \frac{x_{\tau,m}}{(1-x_{\tau,m})^2} \Delta_{i,m}^{(\tau)} \geqslant \frac{1}{8}$$

Re-arranging the above and applying Lemmas B.3 and B.6 yields for all $j \in [k]$:

$$\begin{aligned} 2\left(\frac{\eta}{\alpha}\right)^4 (U_j^*(x_\tau) - U_j^*(x_{\tau-1})) &\geqslant 2\eta(U_i^*(x_\tau) - U_i^*(x_{\tau-1})) \\ &\geqslant \hat{U}_i(x_\tau) - \hat{U}_i(x_{\tau-1}) \geqslant \frac{1}{1024} \cdot \left(\frac{\alpha}{\eta}\right)^6 \cdot \theta \cdot (\alpha\nu)^{k-1} \end{aligned}$$

proving the claim in this case.

Next, we consider the case where $x_{\tau-1,\ell^{(\tau)}+1} > 2x_{\tau-1}$. In this case, note that as $x_{\tau-1} \geqslant \nu$ and our bound on δ yields $x_{\tau-1} + \ell^{(\tau)}\delta \geqslant \frac{3}{2}x_{\tau-1}$ and the claim follows as:

$$U_i^*(x_\tau) - U_i^*(x_{\tau-1}) = \int_{x_{\tau-1}}^{x_\tau} \sum_{j \neq i} f_j(z) \prod_{m \neq i,j} F_m(z) dz \geqslant (k-1)\alpha \int_0^{\frac{\nu}{2}} (\alpha z)^{k-2} dz \geqslant \frac{(\alpha\nu)^{k-1}}{2^{k-1}}.$$

For the final case, the second claim and MACRO imply $x_T = x_\tau$ with $x_{\tau-1} < 1-\theta$ and $x_\tau + \delta \geqslant 1-\theta/2$. In this case, we note again from our choice of δ that $x_\tau \geqslant 1-(3\theta)/4$. Furthermore, since $x_{\tau-1} < 1-\theta$, we have from Lemma B.5:

$$\begin{aligned} U_i^*(x_\tau) - U_i^*(x_{\tau-1}) &\geqslant U_i^*\left(1-\frac{3\theta}{4}\right) - U_i^*(1-\theta) = \int_{1-\theta}^{1-\frac{3\theta}{4}} \sum_{j \neq i} f_j(z) \prod_{m \neq i,j} F_m(z) dz \\ &\geqslant \alpha \int_{1-\theta}^{1-\frac{3\theta}{4}} \sum_{j \neq i} \prod_{m \neq i,j} F_m(z) dz \geqslant \alpha \int_{1-\theta}^{1-\frac{3\theta}{4}} (k-1) U_i^*(1-\theta) dz \geqslant \frac{\alpha\theta}{8}. \end{aligned}$$

concluding the proof in this case. □

To bound T, we break $[\nu, 1-\theta]$ into three segments and handle each separately:

$$I_1 : \left[\nu, \left(\frac{\alpha}{2\eta}\right)^{32}\right], I_2 : \left[\left(\frac{\alpha}{2\eta}\right)^{32}, 1-\frac{1}{4\eta k}\right], \text{ and } I_3 : \left[1-\frac{1}{4\eta k}, 1-\frac{\theta}{2}\right].$$

Case 1: We start with I_1 and restrict ourselves to the intervals, $[x_{\tau-1}, x_\tau]$ such that $\gamma_{\ell^{(\tau)}}^{(\tau)} \geqslant 1/8$ as as there are at most $\log_2(2/\nu)$ intervals where this doesn't happen (MACRO and Lemma B.2). From the definition of $\gamma_{\ell^{(\tau)}}^{(\tau)}$, there exists some $i \in [k]$ such that:

$$16 \cdot \left(\frac{\eta}{\alpha}\right)^6 \cdot \frac{1}{\hat{U}_i(x_{\tau-1})} \sum_{m=1}^{\ell^{(\tau)}} \frac{x_{\tau,m}}{(1-x_{\tau,m})^2} \cdot \Delta_{i,m}^{(\tau)} \geqslant \frac{1}{8}$$
$$U_i^*(x_{\tau-1}) \geqslant (\alpha x_{\tau-1})^{k-1} \implies x_{\tau-1} \leqslant \frac{U_i^*(x_{\tau-1})^{1/(k-1)}}{\alpha}$$

and as a result, we obtain from Lemmas 3.5 and B.6 and the last claim of Lemma B.2:

$$\frac{1}{8} \leqslant 16 \cdot \left(\frac{\eta}{\alpha}\right)^6 \cdot \frac{1}{\hat{U}_i(x_{\tau-1})} \cdot \frac{2x_{\tau-1}}{(1-x_\tau)} \sum_{m=1}^{\ell^{(\tau)}} \frac{1}{(1-x_{\tau,m})} \cdot \Delta_{i,m}^{(\tau)}$$

$$\leqslant 32 \cdot \left(\frac{\eta}{\alpha}\right)^6 \cdot \frac{1}{\hat{U}_i(x_{\tau-1})} \cdot \frac{1}{\alpha} \cdot \frac{U_i^*(x_{\tau-1})^{1/(k-1)}}{(1-x_\tau)} \sum_{m=1}^{\ell(\tau)} \frac{1}{(1-x_{\tau,m})} \cdot \Delta_{i,m}^{(\tau)}$$
$$\leqslant 512 \cdot \left(\frac{\eta}{\alpha}\right)^8 \cdot \frac{1}{U_i^*(x_{\tau-1})^{(k-2)/(k-1)}} \cdot \left(U_i^*(x_\tau) - U_i^*(x_{\tau-1})\right).$$

Re-arranging the above, two applications of Lemma B.3 with the fact $U_j^*(x) \leqslant (\eta x)^{k-1}$:

$$j \in [k] : U_j^*(x_\tau) \geqslant \left(\frac{2\alpha}{\eta}\right)^{14} \cdot U_j^*(x_{\tau-1})^{(k-2)/(k-1)} \geqslant U_j^*(x_{\tau-1})^{1-1/(2(k-1))}.$$

Defining $S_1 := \{\tau : [x_{\tau-1}, x_\tau] \subset I_1 \text{ and } \gamma_{\ell(\tau)}^{(\tau)} \geqslant 1/8\}$, $T_1 := |S_1|$ and $\tau_1^* := \max S_1$, we get by a recursive application of the above inequality for all $j \in [k]$:

$$e^{-(k-1)} \geqslant U_j^*(x_{\tau_1^*}) \geqslant (U_j^*(x_0))^{\left(1-\frac{1}{2(k-1)}\right)^{T_1}} \geqslant (\alpha\nu)^{(k-1)\left(1-\frac{1}{2(k-1)}\right)^{T_1}}.$$

Iteratively taking logs, $\exp\left\{-\frac{T_1}{2(k-1)}\right\}\log(\alpha\nu) \geqslant 1 \implies T_1 \leqslant 2(k-1)\log\log(1/(\alpha\nu))$. Hence, the number of intervals in I_1 is bounded by $2(k-1)\log\log(1/(\alpha\nu)) + \log_2(2/\nu)$.

Case 2: Similarly, for I_2, we restrict ourselves to intervals $[x_{\tau-1}, x_\tau] \subset I_2$ with $\gamma_{\ell(\tau)}^{(\tau)} \geqslant 1/8$. Noting $1 - x_\tau \geqslant 1/(4\eta k)$, we have from Lemmas 3.5, B.2 and B.6 for some $i \in [k]$:

$$\frac{1}{8} \leqslant 16 \cdot \left(\frac{\eta}{\alpha}\right)^6 \cdot \frac{1}{\hat{U}_i(x_{\tau-1})} \cdot \frac{2x_{\tau-1}}{(1-x_\tau)} \sum_{m=1}^{\ell(\tau)} \frac{1}{(1-x_{\tau,m})} \cdot \Delta_{i,m}^{(\tau)}$$
$$\leqslant 16 \cdot \left(\frac{\eta}{\alpha}\right)^6 \cdot \frac{1}{\hat{U}_i(x_{\tau-1})} \cdot 8\eta k \cdot \sum_{m=1}^{\ell(\tau)} \frac{1}{(1-x_{\tau,m})} \cdot \Delta_{i,m}^{(\tau)}$$
$$\leqslant 1024k \cdot \left(\frac{\eta}{\alpha}\right)^8 \cdot \frac{1}{U_i^*(x_{\tau-1})} \cdot \left(U_i^*(x_\tau) - U_i^*(x_{\tau-1})\right).$$

Re-arranging the above inequality and two applications of Lemma B.3 yield:

$$\forall j \in [k] : U_j^*(x_\tau) - U_j^*(x_{\tau-1}) \geqslant \frac{1}{8192k} \cdot \left(\frac{\alpha}{\eta}\right)^{11} \cdot U_i^*(x_{\tau-1}) \geqslant \frac{1}{8192k} \cdot \left(\frac{\alpha}{\eta}\right)^{14} \cdot U_j^*(x_{\tau-1}).$$

Define $S_2 := \{\tau : [x_{\tau-1}, x_\tau] \subset I_2 \text{ and } \gamma_{\ell(\tau)}^{(\tau)} \geqslant 1/8\}$, $T_2 := |S_2|$ and $\tau_2^* := \max S_2$ as before. As $x_{\tau-1} \geqslant (\eta/2\alpha)^{32}$ for all $\tau \in S_2$, recursively applying the above inequality yields:

$$1 \geqslant U_j^*(x_{\tau_2^*}) \geqslant \left(1 + \frac{1}{8192k} \cdot \left(\frac{\alpha}{\eta}\right)^{14}\right)^{T_2} U_j^*\left(\left(\frac{\alpha}{2\eta}\right)^{32}\right).$$

Again, noting $U_j^*(x) \geqslant (\alpha x)^{k-1}$ and taking logs, the current case follows:

$$1 \geqslant \exp\left\{\frac{T_2}{16384k} \cdot \left(\frac{\alpha}{\eta}\right)^{14}\right\} \cdot \left(\frac{\alpha}{2\eta}\right)^{64(k-1)} \implies T_2 \leqslant 2^{20}k^2 \left(\frac{\eta}{\alpha}\right)^{14} \log(2\eta/\alpha).$$

Case 3: For I_3, we subdivide it into r subintervals $\left\{I_{3,p} := \left[1-2^p\theta, 1-2^{p-1}\theta\right]\right\}_{p=0}^r$. Note that $r \leqslant \log_2(2/\theta)+1$. We now bound the number of intervals in each of these sub-intervals and restrict ourselves to intervals $[x_{\tau-1}, x_\tau] \subset I_{3,p}$ for some p with $\tau < T$. Note this excludes at most $2r+1$ intervals. For one such interval $[x_{\tau-1}, x_\tau] \in I_{3,p}$, note that $\gamma_{\ell(\tau)}^{(\tau)} \geqslant 1/8$ (Lemma B.2). Similarly, we have for some i using Lemmas 3.5, B.2, B.5 and B.6:

$$\begin{aligned}
\frac{1}{8} &\leqslant 16 \cdot \left(\frac{\eta}{\alpha}\right)^6 \cdot \frac{1}{\hat{U}_i(x_{\tau-1})} \sum_{m=1}^{\ell(\tau)} \frac{x_{\tau,m}}{(1-x_{\tau,m})^2} \cdot \Delta_{i,m}^{(\tau)} \\
&\leqslant 16 \cdot \left(\frac{\eta}{\alpha}\right)^6 \cdot \frac{1}{\hat{U}_i(x_{\tau-1})} \sum_{m=1}^{\ell(\tau)} \frac{1}{(1-x_{\tau,m})(2^{p-1}\theta)} \cdot \Delta_{i,m}^{(\tau)} \\
&\leqslant \frac{256}{2^{p-1}\theta} \cdot \left(\frac{\eta}{\alpha}\right)^7 \cdot \left(U_i^*(x_\tau) - U_i^*(x_{\tau-1})\right)
\end{aligned}$$

which yields:

$$U_i^*(x_\tau) - U_i^*(x_{\tau-1}) \geqslant \frac{2^{p-1}\theta}{2048} \cdot \left(\frac{\alpha}{\eta}\right)^7 \implies \forall j \in [k] : U_j^*(x_\tau) - U_j^*(x_{\tau-1}) \geqslant \frac{2^{p-1}\theta}{2048} \cdot \left(\frac{\alpha}{\eta}\right)^{10}.$$

Again, defining $S_{3,p} = \{\tau < T : [x_{\tau-1}, x_\tau] \subset I_{3,p}\}, T_{3,p} := |S_{3,p}|$, we have from the above:

$$\begin{aligned}
T_{3,p} &\leqslant \frac{2048}{2^{p-1}\theta} \cdot \left(\frac{\eta}{\alpha}\right)^{10} \left(U_i^*\left(1-2^{p-1}\theta\right) - U_i^*\left(1-2^p\theta\right)\right) \\
&= \frac{2048}{2^{p-1}\theta} \cdot \left(\frac{\eta}{\alpha}\right)^{10} \cdot \int_{1-2^p\theta}^{1-2^{p-1}\theta} \sum_{j\neq i} f_j(x) \prod_{q\neq i,j} F_q(x)dx \\
&\leqslant \frac{2048}{2^{p-1}\theta} \cdot \left(\frac{\eta}{\alpha}\right)^{10} \cdot (\eta k \cdot 2^{p-1}\theta) \leqslant 2048k \cdot \left(\frac{\eta}{\alpha}\right)^{11}.
\end{aligned}$$

Summing up over p, we get that: $T_3 := \sum_{p=0}^r T_{r,p} \leqslant 4096k \cdot \left(\frac{\eta}{\alpha}\right)^{11} \log_2(2/\theta)$. Finally, summing over the previous three cases concludes the proof of the lemma. □

B.3.1 Miscellaneous Results

Here, we present miscellaneous results used in various parts of our proof. The first lemma shows that the functions, U_i^*, for different i are within a constant factor of each other.

Lemma B.3. *We have $\forall i, j \in [k], x > y \in [0,1]$:*

$$\left(\frac{\alpha}{\eta}\right)^3 \left(U_j^*(x) - U_j^*(y)\right) \leqslant U_i^*(x) - U_i^*(y) \leqslant \left(\frac{\eta}{\alpha}\right)^3 \left(U_j^*(x) - U_j^*(y)\right).$$

Proof. We have:

$$U_i^*(x) - U_i^*(y) = \int_y^x \sum_{l\neq i} f_l(z) \prod_{m\neq i,l} F_m(z)dz \leqslant \left(\frac{\eta}{\alpha}\right)^3 \left(U_j^*(x) - U_j^*(y)\right)$$

where the second inequality follows from Assumption 3.2:

$$\forall l, l', z \in [0,1] : f_l(z) \prod_{m\neq i,l} F_m(z) \leqslant \left(\frac{\eta}{\alpha}\right)^3 \cdot f_{l'}(z) \prod_{m\neq j,l'} F_m(z).$$

□

Our next lemma establishes concentration of $\hat{G}_i$, as empirical approximations of G_i.

Lemma B.4. *For all $i \in [k]$, we have $\|\hat{G}_i - G_i\|_\infty \leqslant \varepsilon_g$ with probability at least $1-\rho$ as long as $n \geqslant \log(2k/\rho)/(2\varepsilon_g^2)$.*

Proof. Define random variables, $\forall i \in [k], j \in [n] : Z_j^i := Y_j \cdot \mathbf{1}\{W_j = i\} + \mathbf{1}\{W_j \neq 1\}$. The CDF of Z_j^i corresponds to G_i while its empirical CDF corresponds to $\hat{G}_i$. The lemma follows from the DKW inequality [Dvoretzky et al., 1956] and a union bound over i. □

The two following lemmas find use in the proof of Lemma 3.9.

Lemma B.5. *For all $i \in [k]$ and for all $x \in \left[1 - \frac{1}{4\eta k}, 1\right]$, we have $U_i^*(x) \geqslant \frac{3}{4}$.*

Proof. Let $X_i \overset{iid}{\sim} F_i$. Then, by the union bound: $\mathbb{P}\left\{\exists i : X_i \geqslant 1 - \frac{1}{4\eta k}\right\} \leqslant \frac{1}{4}$. □

Lemma B.6. *We have, for all $i \in [k]$ and all $\tau \in [T]$,*

$$\forall l \in \ell^{(\tau)} : \left(\hat{U}_i(x_{\tau,l}) - \hat{U}_i(x_{\tau,l-1})\right) \leqslant \frac{1}{1-x_{\tau,l}} \cdot \Delta_{i,l}^{(\tau)} \leqslant 2\left(\hat{U}_i(x_{\tau,l}) - \hat{U}_i(x_{\tau,l-1})\right)$$

$$\forall \hat{U}_i(x_\tau) - \hat{U}_i(x_{\tau-1}) \leqslant \sum_{l=1}^{\ell^{(\tau)}} \frac{1}{1-x_{\tau,l}} \cdot \Delta_{i,l}^{(\tau)} \leqslant 2 \cdot \left(\hat{U}_i(x_\tau) - \hat{U}_i(x_{\tau-1})\right)$$

Proof. For the lower bound, we have $\forall l \in \ell^{(\tau)}$:

$$\begin{aligned}\frac{1}{1-x_{\tau,l}} \cdot \Delta_{i,l}^{(\tau)} &= \frac{1}{n} \cdot \sum_{j=1}^{n} \frac{1}{(1-x_{\tau,l})} \cdot \mathbf{1}\left\{Z_j = i, x_{\tau,l-1} < Y_j \leqslant x_{\tau,l}\right\} \\ &\geqslant \frac{1}{n} \cdot \sum_{j=1}^{n} \frac{1}{(1-Y_j)} \cdot \mathbf{1}\left\{Z_j = i, x_{\tau,l-1} < Y_j \leqslant x_{\tau,l}\right\} = \hat{U}_i(x_{\tau,l}) - \hat{U}_i(x_{\tau,l-1}).\end{aligned}$$

Summing the above inequality over l concludes the proof of the lower bound. Similarly, for the upper bound, we get $\forall l \in \ell^{(\tau)}$:

$$\begin{aligned}&\frac{1}{1-x_{\tau,l}} \cdot \Delta_{i,l}^{(\tau)} - \left(\hat{U}_i(x_{\tau,l}) - \hat{U}_i(x_{\tau,l-1})\right) \\ &= \frac{1}{n} \cdot \sum_{j=1}^{n} \left(\frac{1}{(1-x_{\tau,l})} - \frac{1}{(1-Y_j)}\right) \cdot \mathbf{1}\left\{Z_j = i, x_{\tau,l-1} < Y_j \leqslant x_{\tau,l}\right\} \\ &\leqslant \frac{1}{n} \cdot \sum_{j=1}^{n} \left(\frac{\delta}{(1-x_{\tau,l})(1-Y_j)}\right) \cdot \mathbf{1}\left\{Z_j = i, x_{\tau,l-1} < Y_j \leqslant x_{\tau,l}\right\} \\ &\leqslant \frac{1}{n} \cdot \sum_{j=1}^{n} \left(\frac{1}{2 \cdot (1-Y_j)}\right) \cdot \mathbf{1}\left\{Z_j = i, x_{\tau,l-1} < Y_j \leqslant x_{\tau,l}\right\} = \frac{1}{2}\left(\hat{U}_i(x_{\tau,l}) - \hat{U}_i(x_{\tau,l-1})\right).\end{aligned}$$

Again, re-arranging and summing over l concludes the proof. □

B.4 Proof of Theorem 3.12

Proof. We combine this result with the same binary search from Subsection 2.3 to identify quantiles of the functions $\widehat{F}_i$. In particular, fix $\epsilon > 0$ and let

$$W = \{w_a := \gamma + a \cdot \frac{\epsilon}{2} \mid a \in \mathbb{N} \text{ and } \gamma + a \cdot \frac{\epsilon}{2} \leqslant 1\} \cup \{1\}.$$

An identical argument from the one from the proof of Theorem 2.9 shows that for any $\epsilon > 0$, we can use binary search to find $\hat{z}_{j,a}$ such that

$$\begin{aligned} &|F_j(\hat{z}_{j,a}) - w_a| \leqslant \frac{\epsilon}{2} \text{ for all } j \in [k] \text{ and } a \in [|W|] \\ &\text{w.p.} \qquad 1 - \frac{4k}{\epsilon} \log\left(\frac{4L}{\epsilon}\right) \exp\left\{-\epsilon^2 \gamma n_1 / 192\right\} \end{aligned} \tag{27}$$

using $C \cdot n_1 \cdot k \cdot \log(4L/\epsilon)/\epsilon$ samples for a universal constant C (in particular, see the proof of Theorem 2.9, set $\delta = \epsilon_1 = \epsilon/2$, and use Lemma 3.11 for the pointwise guarantee in place of (5)).

Conditioning on this event, we can thus define the piecewise-constant functions $\widehat{F}_j$ for $j \in [k]$ as

$$\widehat{F}_j(x) = \sum_{a \in [|W|]} \mathbf{1}\left\{x \in [z_{j,a}, z_{j,a+1})\right\} \cdot \left(\gamma + a \cdot \frac{\epsilon}{2}\right)$$

Now, consider any $x \in [p, 1]$ and define $a \in \mathbb{N}$ such that $x \in [\hat{z}_{j,a}, \hat{z}_{j,a+1}]$, so that by construction $\widehat{F}_j(x) = w_{j,a}$. Then:

$$\begin{aligned} F_j(x) &\geqslant F_j(\hat{z}_{j,a}) && \text{(monotonicity of CDF)} \\ &= w_{j,a+1} + (F_j(\hat{z}_{j,a}) - w_{j,a}) + (w_{j,a} - w_{j,a+1}) \\ &\geqslant w_{j,a+1} - \epsilon/2 - \epsilon/2 && \text{(definition of } W \text{ and (27))} \\ &\geqslant \widehat{F}_j(x) - \epsilon && \text{(definition of } \widehat{F}_j\text{)} \end{aligned}$$

Similarly, $F_j(x) \leqslant \widehat{F}_j(x) + \epsilon$.

It remains to handle $x \in [p, \hat{z}_{j,0}]$: note that by (strict) monotonicity of the $\widehat{F}_j$, we must have $\widehat{F}_j^{-1}(\gamma) \leqslant p$. Thus, if $p \leqslant x \leqslant \hat{z}_{j,0}]$,

$$F_i(x) \geqslant F_i(p) \geqslant \gamma = \widehat{F}_i(x), \qquad \text{and} \qquad F_i(x) \leqslant F_i(\hat{z}_{j,0}) \leqslant \gamma + \frac{\epsilon}{2} \leqslant \widehat{F}_i(x) + \epsilon.$$

Thus, $|F_j(x) - \widehat{F}_j(x)| \leqslant \epsilon$ over the entire interval $[p, 1]$ with probability at least $1 - \delta$, as long as

$$n \geqslant \frac{Ck \log(k/\epsilon) \log(L/\epsilon)^2}{\epsilon^3 \gamma},$$

for a universal constant C, completing the proof. ☐