Estimation of Auction Models with Shape Restrictions*

Joris Pinkse and Karl Schurter†

March 15, 2021

Abstract

We introduce several new estimation methods that leverage shape constraints in auction models to estimate various objects of interest, including the distribution of a bidder's valuations, the bidder's ex ante expected surplus, and the seller's counterfactual revenue. The basic approach applies broadly in that (unlike most of the literature) it works for a wide range of auction formats and allows for asymmetric bidders. Though our approach is not restrictive, we focus our analysis on first–price, sealed–bid auctions with independent private valuations. We highlight two nonparametric estimation strategies, one based on a least squares criterion and the other on a likelihood criterion. We establish several theoretical properties of our methods to guide empirical analysis and inference. In addition to providing the asymptotic distributions of our estimators, we identify ways in which methodological choices should be tailored to the objects of interest. For objects like the bidders' ex ante surplus and the seller's counterfactual expected revenue with an additional symmetric bidder, we show that our input–parameter–free estimators achieve the semiparametric efficiency bound. For objects like the bidders' inverse strategy function, we provide an easily implementable boundary–corrected kernel smoothing and transformation method in order to ensure the squared error is integrable over the entire support of the valuations. An extensive simulation study illustrates our analytical results and demonstrates the respective advantages of our least–squares and maximum likelihood estimators in finite samples. Compared to estimation strategies based on kernel density estimation, the simulations indicate that the smoothed versions of our estimators enjoy a relatively large degree of robustness to the choice of an input parameter.

*We thank Yanqin Fan, Paul Grieco, Brent Hickman, Kei Hirano, Sung Jae Jun, Laurent Lamy, Ruixuan Liu, Orville Mondal, Peter Newberry, and participants of the Penn State brown bag workshop, the Lancaster auctions conference, the EARIE conference in Barcelona, and the PennCorn conference on econometrics and industrial organization in Ithaca, and the Greater New York Metropolitan Area Econometrics Colloquium for their helpful comments.

†corresponding author keschurter@gmail.com

1 Introduction

We develop several new estimators that impose shape constraints implied by the bidder's incentive–compatibility condition in auction models and illustrate how these estimators can be tailored to the ultimate target of the estimation. Compared with existing methods that are specific to the study of first–price auctions, our approach generalizes to other auction mechanisms, has the same optimal nonparametric rates of convergence, and attains the semiparametric efficiency bound in the estimation of some objects of direct interest to applied researchers. Moreover, our methodology achieves these semiparametric efficiency bounds without having to choose any input parameters, such as a kernel bandwidth. Practitioners can therefore leverage shape constraints to costlessly avoid searching for a parsimonious parametric approximation, formally defining a sieve space, or selecting a bandwidth, as well as the associated risks of choosing wrongly.

The key insight behind our approach is that, despite the great diversity of auction formats we might consider, the fundamental nature of a bidder's decision problem is the same. Specifically, given the strategies of a bidder's competitors and the distribution of their private values, the bidder chooses her bid to optimally trade off the probability of winning with her expected payment to the seller. Though the details of this trade–off as a function of the bid are complicated and depend on the specifics of the auction rules, the equilibrium expected payment to the seller is a convex function of the probability with which the bidder expects to win in equilbrium.[1] Moreover, the first–order condition of the bidder's problem requires that the slope of the equilibrium expected payment function at the optimally chosen win probability is equal to her private valuation. Thus, the derivative of the convex payment function is equivalently viewed as the inverse strategy function, which maps optimally chosen win–probabilities to values. These facts have been used to establish the revenue equivalence theorem (Myerson, 1981; Milgrom and Weber, 1982) and were subsequently invoked as a generic nonparametric identification strategy (Larsen and Zhang, 2018). Our paper further exploits this change of variables from bids to win–probabilities in order to relate the literature on nonparametric estimation in auctions to the large literature on nonparametric estimation under shape constraints.

The main benefits of this change of variables are threefold. First, by reformulating the target of estimation as the slope of a convex function, we open the door to a variety of well–known estimation strategies such as (shape–)constrained (nonparametric) least squares and (a new version of) nonparametric maximum likelihood estimation (MLE),[2] as well as some more adventurous estimators like a jackknife estimator. Second, this framework generalizes the large and growing toolkit for nonparametric estimation and testing in first–price auctions to generic auction mechanisms. And, finally, it allows the econometrician to easily impose the structure of symmetric equilibria, namely that the marginal distribution of a bidder's optimally chosen win–

[1] In first price sealed bid auctions, convexity of the expected payment is equivalent to monotonicity of the bid function. Lemma 12 in appendix A proves convexity of in any feasible auction mechanism (as defined by Myerson, 1981) with independent private values.

[2] See e.g. Brunk (1955) for an early example of nonparametric estimation subject to shape constraints.

probability is a known function that only depends on the number of bidders. Importantly, the distribution function does *not* depend on the unknown distribution of the bidders' private valuations. This a priori knowledge of the win–probability distribution yields sizable improvements in the asymptotic distribution of our estimators.

The latter observation is especially useful when we consider the estimation of objects that are less primitive than the value density but may be of more direct interest to the researcher. To provide an example, we introduce the notation $e(p)$ for a bidder's expected payment when she bids so as to win with probability p in equilibrium and $\alpha(p)$ for its derivative. The bidder's ex ante expected surplus can then be expressed as an integral of $\alpha(p)p - e(p)$ with respect to the distribution of the bidder's optimally chosen win–probability, and the mean of the bidder's valuations and the seller's expected profit can similarly be expressed as a functional of e and α. One can construct "plug–in" estimators for these objects by combining the estimated α with a precise estimate of the win–probability distribution. In the symmetric case, however, we show that one can significantly reduce the asymptotic variance by substituting the win–probability distribution which is known to prevail in any symmetric equilibrium rather than an estimate thereof.

Moreover, we show that objects such as the bidder's surplus can be estimated at a square–root rate of convergence even when α is replaced with an unsmoothed estimate. The reason is that although the first step estimator of α converges at a cube–root rate, it has little bias.[3] The act of integration in the second stage acts as an average and hence reduces the asymptotic variance. Moreover, the resultant estimators achieve the semiparametric efficiency bound when one fully exploits the symmetric structure of the equilibrium. If the researcher does not assume bidders are symmetric or only observes one competitor bid per auction, the semiparametric efficiency bound on the asymptotic variance is larger, but we again show that the unsmoothed plug–in estimators for the mean valuation and bidder's surplus attain the efficiency bound.

Thus, optimally smoothing the estimate of α does not necessarily improve the asymptotic performance of the desired object. Indeed, optimal pointwise estimation of α can be detrimental. For example, in order to achieve square–root consistency of the mean valuation using a smoothed estimate of the inverse strategy function, one would need to "undersmooth" by choosing an input parameter to slow down the pointwise rate of convergence of the inverse strategy function and reduce its bias. However, one must avoid too much undersmoothing using methods that do not impose monotonicity a priori (e.g. Marmer and Shneyerov (2012, MS), Ma et al. (2019b, Ma19b), and Guerre et al. (2000, GPV)), because letting the bandwidth go to zero for a fixed sample size would yield an inconsistent estimator for the inverse bid function and produce an inconsistent estimator of the mean valuation. In contrast, there is no risk of too much undersmoothing using our approach, because our undersmoothed and unsmoothed estimators of objects like the mean valuation are asymptotically equivalent and attain the efficiency bound. This asymptotic efficiency result for both our

[3] Indeed, the asymptotic distribution is centered at zero.

unsmoothed and undersmoothed estimators therefore provides a large degree of robustness in the choice of bandwidths relative to existing methods.

Our paper relates to recent work on identification in trading mechanisms and estimation of monotone bidding strategies in first–price auctions. Larsen and Zhang (2018) operationalize a similar change of variables to prove generic nonparametric identification results in settings where the researcher does not observes the rules of the mechanism and may not directly observe the agent's actions, either, but is willing to assume the data are generated in a Bayes–Nash equilibrium. Thus, their analysis begins one step behind ours in the sense that they estimate the mapping from actions (e.g. bids) to outcomes (payments and allocations) in a first stage. Not surprisingly, their simulations demonstrate their approach suffers from a large loss of precision compared to estimation strategies that take advantage of prior knowledge of the auction mechanism. We therefore view our respective contributions as complementary advances in identification and estimation of auctions and auction–like mechanisms under shape constraints.[4]

Three recent papers have also considered shape–constrained estimation in first–price auctions. Henderson et al. (2012) impose monotonicity on a nonparametric estimator of the inverse bidding strategy—which is equivalent to convexity of the expected payment function—by 'tilting' the empirical distribution of the bids, Luo and Wan (2018, LW) consider an alternative approach that imposes convexity of the integrated quantile function of the bidders' valuations, and Ma19b apply a rearrangement technique to the first step estimator in GPV. The constrained least squares estimator (LS) in this paper may be viewed as an extension of LW to more general auction models with possibly asymmetric bidders, which we achieve by considering the equilibrium expected payment instead of the integrated quantile function. Indeed, both our constrained least squares estimator and the one in LW can be characterized as the (slope of) greatest convex minorants (GCM), albeit of different functions.[5] In the case of first–price auctions with two symmetric bidders, the integrated quantile function coincides with the equilibrium expected payment function, and our constrained least squares estimator is numerically equivalent to the LW estimator. More generally, however, the integrated quantile function differs from the expected payment function when there are more than two bidders, and it need not be convex when bidders are asymmetric, because a bidder with a valuation equal to the τ–quantile of its distribution will generally not submit a bid equal to the τ–quantile of its highest competing bid. Thus, the LW approach does not apply to asymmetric auctions. The approach pursued in Ma19b uses the bids instead of the probabilities and hence does not readily extend to other auction mechanisms.

Like the estimator proposed in LW, neither our constrained least squares estimator nor our nonparametric MLE requires the choice of an input parameter. Both of our unsmoothed estimators of the equilibrium expenditure function e converge as a process to the same (tight) Gaussian limit process. The inverse strategy function α, if the choice variable is the probability of winning, is the derivative of e. Both of our unsmoothed

[4]In appendix C we discuss how our approach can be applied to other mechanisms.

[5]Note that our LS estimator yields the GCM of the expected payment function.

estimators of α converge at a $\sqrt[3]{T}$ rate to the same Chernoff–distributed limit, where T is the number of auctions. If bidders are symmetric, our estimators have the same Chernoff limiting distribution as the estimator in LW. Computation of both of our unsmoothed estimators is simple: the constrained least squares estimator can be computed using an off–the–shelf algorithm and we show that our nonparametric maximum likelihood estimator can be easily computed using a simple pool–adjacent–violators algorithm, also.

Although it is not our primary objective, in the interest of completeness and to facilitate comparison with earlier work, we provide estimators of the quantile function, the distribution function, and the density f_v of valuations. The quantile function and distribution functions can be estimated using routine operations (such as the delta method) on our estimates of α. As noted by GPV and others, estimating f_v requires nonparametric derivative estimation and the optimal convergence rate is a leisurely $T^{2/7}$.[6] We provide estimation results for the derivative α' of α, which indeed converges at the $T^{2/7}$ rate. There are two ways of estimating f_v using our approach: a two–step procedure in the spirit of GPV and a one–step procedure like MS.[7] We do not see any reason to prefer either the one–step or two–step procedure. We provide asymptotic linear expansions of our first step estimators to allow readers to make up their own mind.

We analyze the performance of our estimators in a fairly extensive simulation study. In general, we find that our estimators perform well and exhibit considerable robustness, both with respect to the design of the simulation study and to the choice of input parameter. However, no clear winner emerges and our various methods differ in systematic ways that are consistent with our asymptotic theory and with intuition. Hence, we do not offer empirical researchers a specific recommendation; rather, we provide a collection of tools, asymptotic results, and general insights that can be applied on a case–by–case basis.

Though we aim to provide a thorough foundation of asymptotic results, this paper conspicuously lacks discussion of risk–aversion, affiliated private valuations, unobserved auction–level heterogeneity, endogenous entry, and other extensions to the basic approach of GPV (Levin and Smith, 1994; Li et al., 2002; Li, 2005; Campo et al., 2003; Guerre et al., 2009; Li and Zheng, 2009; Krasnokutskaya, 2011; Marmer et al., 2013; Roberts, 2013; Gentry and Li, 2014, see, for example,). We restrict attention to asymmetric auction models with independent private valuations and leave the many interesting extensions for future work.

Our paper is organized as follows. In section 2 we describe the auction model. We then describe the shape–constrained estimators of the equilibrium expenditure function e and its derivative α in section 3. In section 4 we introduce and provide results for the smoothed versions of our estimates, including a boundary correction and transformation scheme. We next show how the preceding estimates can be modified and combined to estimate several objects of potential interest in section 5. We then present simulation results in

[6]To obtain a fourfold improvement requires a data set that's 128 times as large, compared to 32 for typical nonparametric estimators of univariate objects and 16 for parametric estimators.

[7]LW raise the interesting possibility of using a first step unsmoothed estimator as an input to the second stage of GPV, but do not provide asymptotic results. It is likely that consistency obtains, but the f_v convergence rate and indeed the asymptotic distribution are unknown.

section 6 and conclude in section 7. Several supplementary results and variations on our methodology are presented in appendix B. All proofs are in appendix A.

2 First–price auction model with independent private values

Let $i = 1, \dots, n$ index the bidders competing for an object in a first–price, sealed–bid auction. A bidder's value v_i is drawn from a distribution F_i, which takes support on a compact interval in the nonnegative reals. We assume the seller sets a nonbinding reserve price of zero. We further assume each distribution is absolutely continuous with a density f_i that is bounded away from zero on its support.

Each risk–neutral bidder chooses her bid to maximize her expected surplus taking her competitors' strategies as given. We will take bidder one (1) to be the bidder whose value is to be recovered and use a subscript c to denote her competitors. Thus, bidder one solves

$$\max_b \left\{ G_c(b)(v_1 - b) \right\}, \tag{1}$$

where $G_c(b)$ denotes the probability that bidder one's competitors all bid no more than b.[8]

We can equivalently formulate the bidder's problem as a choice of her equilibrium probability of winning:

$$\max_p \{ p v_1 - e_1(p) \}, \tag{2}$$

where $e_1(p) = Q_c(p)p$ is bidder i's equilibrium expected payment to the seller with $Q_c = G_c^{-1}$ the quantile function of the maximum rival bid.

The well–known fact that e_1 must be convex in p can be seen as a consequence of monotonicity of the equilibrium strategies or incentive compatibility of the direct revelation selling mechanism that implements the Bayes–Nash equilibrium of the first–price auction (Maskin and Riley, 2000). In any case, the solution to bidder one's problem is illustrated in figure 1. As noted by Larsen and Zhang (2018) and Milgrom and Weber (1982), bidder one's indifference curves in (p, e)-space are represented by straight lines with a slope equal to v_1. The optimal expected surplus is therefore attained where $\alpha_1(p) = e_1'(p) = v_1$. From here on we drop the subscript from the function e_1 and write e to mean e_1.

3 Nonsmooth estimation of e and α

In order to eliminate conditioning variables in our notation for the competing distribution of bids and equilibrium expected payment function, we assume valuations are independent across bidders, there is no

[8] In principle one can accommodate dependence among bidder one's competitors' bids by treating groups of bidders as individual bidders. Doing so requires additional assumptions to ensure monotonicity of equilibrium strategies.

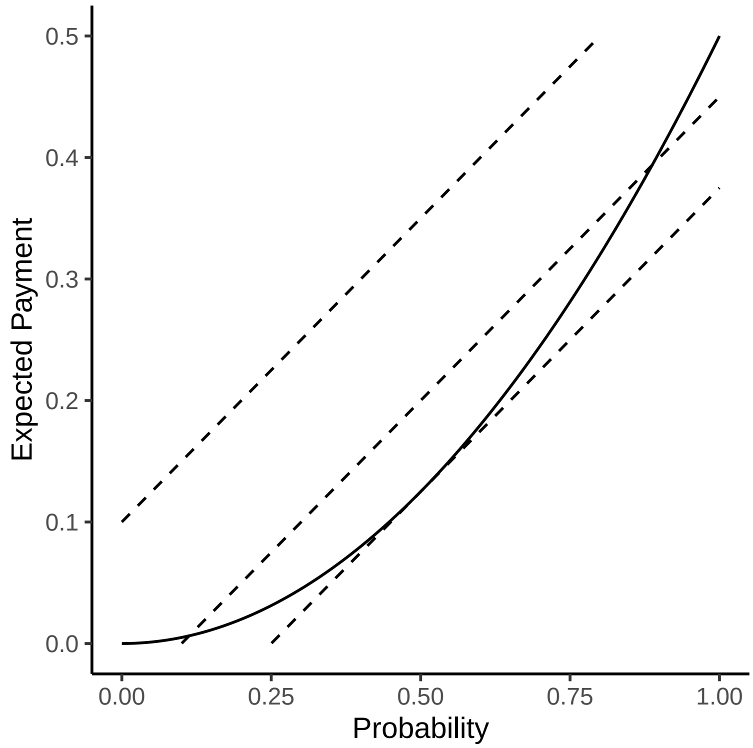

Figure 1: A risk–neutral bidder with indifference curves represented by the dashed lines would optimally submit a bid that will win with a probability of 1/2 and expect to pay 1/8 to the seller (unconditional on winning the auction).

auction–level heterogeneity, and the same set of bidders compete in each auction.

Under the assumption that bidders' valuations are independent across auctions, each auction is an independent realization of the same game. Therefore, the probability of winning and the expected payment as a function of b can be estimated by $G_{cT}(b)$ and $G_{cT}(b)\, b$, where G_{cT} is a suitable estimate of the distribution function G_c of the maximum of bidder one's competitors' bids.

Though a piecewise linear function e_T whose graph contains

$$\big\{\big(G_{cT}(b),\ b\, G_{cT}(b)\big) :\ b = b_1, \dots, b_T\big\} \tag{3}$$

converges to e at a $\sqrt{T}$–rate, it is generally non–convex in finite samples, and the slope of the menu between two nearby points is a poor approximation of its derivative.[9] We will show that the greatest convex minorant of e_T can be used to estimate the expected payment function and its derivative in a single step.[10] Moreover,

[9] $\sqrt{T}\{e_T(\cdot) - e(\cdot)\}$ converges weakly to a Gaussian limit process.

[10] LW consider a greatest convex minorant estimator of a different function. Under independence and symmetry assumptions. their estimator coincides with ours for two bidders whereas with more than two bidders the estimators are different but the limit distributions coincide. This follows from lemma 2, stated later, and some elbow grease. LW do not deal with the asymmetric and

we show that this estimator can be justified by a least–squares criterion on α and estimated by isotonic regression.[11]

3.1 Convexification

To motivate the least–squares criterion, suppose that a differentiable estimate of the quantile function for bidder one's highest competing bid were available. Multiplying this hypothetical estimator by p would yield a differentiable, though possibly non–convex, estimator, e_T. A shape–constrained estimate of the derivative of the expected payment function could then be obtained by solving the following problem

$$\min_{\alpha\in\mathscr{A}}\left(\frac{1}{2}\int_0^1\left(\alpha(p)-e'_T(p)\right)^2\,\mathrm{d}p\right),$$

where $\mathscr{A}$ is the set of nondecreasing nonnegative functions defined on $[0,1]$. This least–squares objective can be rewritten as

$$\frac{1}{2}\int_0^1\left(\alpha(p)-e'_T(p)\right)^2\,\mathrm{d}p=\frac{1}{2}\int_0^1\alpha^2(p)\,\mathrm{d}p-\int_0^1\alpha(p)\,e'_T(p)\,\mathrm{d}p+\frac{1}{2}\int_0^1 e'_T(p)^2\,\mathrm{d}p\,.$$

The last term does not depend on α and may therefore be dropped from the criterion without affecting the shape–constrained estimator. The problem becomes

$$\min_{\alpha\in\mathscr{A}}\left(\frac{1}{2}\int_0^1\alpha^2(p)\,\mathrm{d}p-\int_0^1\alpha(p)\,\mathrm{d}e_T(p)\right),\tag{4}$$

which can be solved for any e_T, differentiable or not, provided that the second integral in (4) exists.

In a first-price auction, we use an unconstrained estimate of the empirical quantile function for bidder one's highest competing bid and set $e_T(p)=Q_{cT}(p)\,p$ in (4), where $Q_{cT}(p)=\inf\{b\,:\,G_{cT}(b)\geq p\}$ for $p>0$ and $Q_{cT}(0)$ is taken to be the minimum rival bid. As we noted above, this e_T will generally be non–convex in finite samples and piecewise linear. If $Q_{cT}(p)$ is the empirical quantile function of the highest rival bid, e_T will be discontinuous at t/T for $t=1,\ldots,T$, and the least–squares criterion can be rewritten as

$$\frac{1}{2}\int_0^1\alpha^2(p)\,\mathrm{d}p-\sum_{t=1}^{T}\int_{\frac{t-1}{T}}^{\frac{t}{T}}\alpha(p)\,Q_{cT}(p)\,\mathrm{d}p-\sum_{t=1}^{T}\alpha\left(\frac{t-1}{T}\right)\frac{t-1}{T}\left\{Q_{cT}\left(\frac{t}{T}\right)-Q_{cT}\left(\frac{t-1}{T}\right)\right\}.\tag{5}$$

where the second integral in (4) exists because α is bounded and increasing, and e_T is left–continuous.

Given this representation, we show that the minimizer of (5) over all $\alpha\in\mathscr{A}$ is a right–continuous step–function (lemma 13).

dependent cases.

[11] Applying a least–squares criterion to e would also yield a piecewise–linear convex estimator, but the solution to this least–squares problem is more difficult to characterize and compute.

The least–squares problem can then be further simplified to

$$\min_{\alpha_{T1} \le \cdots \le \alpha_{Tt}} \sum_{t=1}^{T} \Big(\frac{1}{2} \alpha_{Tt}^2 - Q_{cTt} \alpha_{Tt} - \big(Q_{cTt} - Q_{cT,t-1} \big)(t-1) \alpha_{Tt} \Big), \tag{6}$$

where $\alpha_{Tt} = \alpha\{(t-1)/T\}$ and $Q_{cTt} = Q_{cT}(t/T)$. The optimization problem in (6) is then recognizable as a classic isotonic regression problem. We can integrate up α_T to obtain a convex estimator $\breve{e}_T$ of e,

$$\breve{e}_T(p) = \int_0^p \alpha_T(u) \, \mathrm{d}u.$$

As it turns out, $\breve{e}_T$ is simply the greatest convex minorant of e_T, and α_T is its left–derivative: see figure 2. Note that $\breve{e}_T$ is both piecewise linear and continuous by construction.

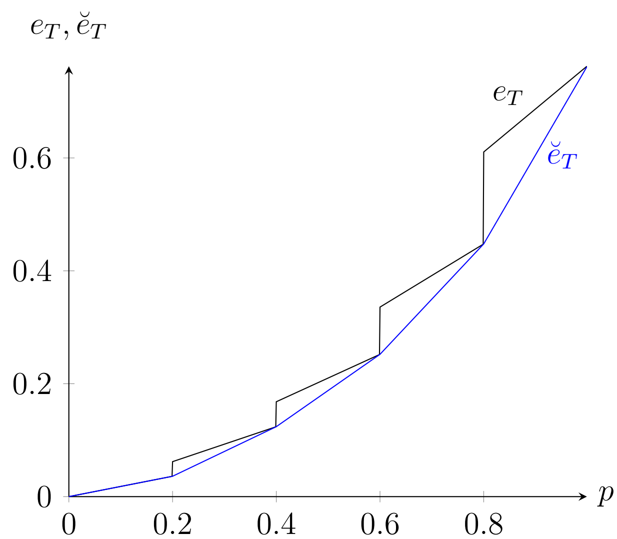

Figure 2: Convexification algorithm illustrated for $T = 5$

If bidders are symmetric then a more efficient unconstrained estimator for the expected payment function is given by $pQ_T\big(p^{1/(n-1)}\big)$, where Q_T is the empirical quantile function for the pooled sample of bids $\{b_\ell\}$ for $\ell = 1, \dots, nT$. In this case, the solution to the least–squares problem in (4) is found via a weighted isotonic regression of $\big\{b_{(\ell)}\ell^{n-1} - b_{(\ell-1)}\big(\ell - 1\big)^{n-1}\big\}/\big\{\ell^{n-1} - (\ell-1)^{n-1}\big\}$ on $\big\{(\ell-1)/nT\big\}^{n-1}$ with weights given by $(\ell/nT)^{n-1} - \big\{(\ell-1)/nT\big\}^{n-1}$, where $b_{(\ell)}$ denotes the ℓ–th order statistic. The corresponding constrained estimator for e is the GCM of $pQ_T\big(p^{1/(n-1)}\big)$, as before.

3.2 Asymptotics for the LS estimator

We now develop some asymptotic estimation results for our convex estimator $\breve{e}_T$. Before we do so, we will make several assumptions and discuss conditions under which they would hold.

Assumption A. *The private values* $v_{t1}, \ldots, v_{tn}$ *in each auction* t *are independent and drawn from continuous distributions* $F_1, \ldots, F_n$*, respectively. The distributions have bounded convex supports* $[\underline{v}, \bar{v}]$ *and their density functions* $f_1, \ldots, f_n$ *are continuous and nonzero on* $(\underline{v}, \bar{v}]$.[12] *There is independence across auctions.* □

Assumption A is a standard assumption in the auctions literature. It is sufficient to ensure the existence of monotone equilibrium bid functions (Lebrun, 2006). The common support assumption embedded in assumption A is unnecessary for the existence of a monotone equilibrium, but is imposed to make our analysis more wieldy.[13] We note that assumption A is stronger than we need: we do not use independence among the competitors' valuations in the proofs of any of our theorems. The assumption can be relaxed provided that bidder one's unique best reply to its competitors' bids is a monotone pure strategy and e_T consistently estimates bidder one's expected payment.[14]

Assumption B. *Bidders are risk neutral and bid according to the (strictly monotonic) Bayes–Nash equilibrium strategies.* □

Assumption B is also stronger than we need, because our results only require bidder one's decision problem to be of the form in (2); other bidders' attitudes toward risk do not affect our inferences related to bidder one's bidding strategy or valuations. Thus, assumptions A and B are merely one set of sufficient conditions on the primitives of the model under which our results can be proven.

A consequence of the assumptions made thus far is that Q_c' is continuous on any closed interval that does not contain zero. Indeed, the first order condition corresponding to (2) implies that

$$Q_c'(p) = \frac{v - Q_c(p)}{p}.$$

We now make a high level assumption on the convergence of an estimator of the bid distribution functions and develop conditions under which it is known to hold.

Assumption C. *The estimator* G_{cT} *of the maximum rival bid distribution* G_c *satisfies* $\sqrt{T}\{G_{cT}(\cdot) - G_c(\cdot)\} \rightsquigarrow \mathbb{G}^*$*, where* $\mathbb{G}^*$ *is a Gaussian process with covariance kernel* H^* *and* $\rightsquigarrow$ *denotes weak convergence.* □

[12] It is standard to model a binding reserve price as an atom at the low end of the value distribution.

[13] If there is a binding reserve price or are more than two bidders with heterogeneous supports, $p = 0$ may be a corner solution to the bidder's problem. That is to say, any bidder with a valuation below the reserve price or a bidder with a value below $\underline{b}$ would prefer to effectively abstain from the auction by submitting a bid that will lose for sure. In the heterogeneous support case, we can redefine $\alpha_i(0)$ as the value of bidder i for which i is indifferent to participating. In the case of a binding reserve price, we can simply treat the reserve price as another competing bid (that has a degenerate distribution) when estimating Q_{cT}. In both cases, the usual caveat applies: the researcher cannot nonparametrically identify any function of the primitives that depends on f_i below the screening value. In particular, these unidentifiable objects include the mean of bidder i's valuation and the counterfactual equilibrium in an auction that decreases the equilibrium screening value. Objects that would still be identified include bidder i's expected surplus and the counterfactual equilibrium in an auction that increases the equilibrium screening value. A second caveat is that the supports of some bidder's optimally chosen win–probability might be bounded above zero or below one, in which case estimates of the support would be needed for the boundary correction schemes in section 4. Hence, we use "wieldy" in part to mean that we will avoid repeating these caveats in the remainder of this paper.

[14] For example, this allows for dependence among rival bids, though independence between bidder one and its rivals.

Assumption C is a relatively weak assumption, and as previously discussed, is the starting point for our analysis. It is for instance satisfied if we take G_{cT} to be the empirical distribution function of the maximum competitor bid, in which case

$$H^*\{Q_c(p), Q_c(p^*)\} = \min(p, p^*) - pp^*. \tag{7}$$

It would also be satisfied if, instead, we assumed symmetry and took $G_{cT} = G_T^{n-1}$, i.e. the empirical bid distribution estimated off all bids raised to the power $n - 1$, in which case[15]

$$H^*\{Q_c(p), Q_c(p^*)\} = \frac{(n-1)^2}{n}\left\{\min(p, p^*)^{1/(n-1)} - (pp^*)^{1/(n-1)}\right\}(pp^*)^{(n-2)/(n-1)}. \tag{8}$$

A final example is one in which there is asymmetry plus independence and all bids are observed in which case[16]

$$H^*\{Q_c(p, Q_c(p^*))\} = \sum_{i=2}^{n}\left(\frac{1}{G_i[Q_c\{\max(p, p^*)\}]} - 1\right)pp^*, \tag{9}$$

where G_i is the bid distribution of bidder i.[17] Note that for $n = 2$ (9) collapses to (7) divided by two since bidder one's bids can also be used in the case of symmetry. We make assumption C to avoid having to hard–wire a particular set of distributional assumptions into the problem.

In order to address the issue of auction–specific covariates, consider the case in which values v_{tj}^* in auction t are such that $v_{tj}^* = v_{tj} + x_t\mu$, where v_{tj} is independent of x_t.[18] In equilibrium, the observed bids will then be additively separable, as well: $b_{tj}^* = b_{tj} + x_t\mu$.[19] Our proposed estimator for μ fits with the rest of our paper in the sense that it is estimated simultaneously with α rather than in a first step (c.f. Haile and Tamer, 2003).

To this end, define $\alpha_{Tt}^{(\tilde{\mu})}$ as α_{Tt} but defined on the residualized bids $\{b_{ct}^* - x_t\tilde{\mu}\}$, such that $\alpha_{Tt}^{(\tilde{\mu})} = \alpha_{Tt}$ if $\tilde{\mu} = \mu$. Now, define

$$\hat{\mu} = \underset{\tilde{\mu}}{\operatorname{argmin}} \sum_{t=1}^{T}\left(\alpha_{Tt}^{(\tilde{\mu})}\right)^2. \tag{10}$$

[15]Note that $\sqrt{T}(G_T - G)$ converges to a Gaussian limit process with covariance kernel $[G\{\min(b, b^*)\} - G(b)G(b^*)]/n$. Hence $\sqrt{T}(G_T^{n-1} - G^{n-1})$ converges to a Gaussian limit process with covariance kernel $G^{n-2}(b)G^{n-2}(b^*)[G\{\min(b, b^*)\} - G(b)G(b^*)]/n$. Insert $Q_c(p) = Q(p^{1/(n-1)})$ to get the stated result.

[16]Note that

$$\sqrt{T}(G_{cT} - G_c) = \sqrt{T}\left(\prod_{i=2}^{n} G_{Ti} - \prod_{i=2}^{n} G_i\right) \simeq \sum_{i=2}^{n}\sqrt{T}(G_{Ti} - G_i)G_{-i1}, \tag{*}$$

where $G_{-i1} = \prod_{j\neq i,1}^{n} G_j$. The right hand side in (*) converges weakly to a Gaussian limit process with covariance kernel $\sum_{i=2}^{n} G_{-i1}(b)G_{-i1}(b^*)[G_i\{\min(b, b^*)\} - G_i(b)G_i(b^*)]$, which produces (9), after noting that $G_{-i1}\{Q_c(p)\} = p/G_i\{Q_c(p)\}$.

[17]In between the cases considered in (8) and (9), it is also possible that some subsets of bidders are known to have the same valuation distribution. In such cases, one could pool bids among bidders within groups in order to estimate G_c. For example, if there are two groups of bidders, $G_c = G_1^{n_1-1}G_2^{n-n_1}$, where G_1 is the marginal distribution of bids submitted by the n_1 bidders in the same group as bidder 1 and G_2 is the marginal bid distribution for bidders in the other group.

[18]A more general specification in which covariates are arbitrarily correlated with valuations would require estimates of the conditional quantile function (Gimenes and Guerre, 2019).

[19]Multiplicative separability works, also.

The intuition behind $\hat{\mu}$ is that the residualized bids are a convolution of $b_{ct} = b^*_{ct} - x_t\mu$ and $x_t(\mu - \tilde{\mu})$. Because the convolutions of independent random variables with bounded support have a larger support than either of the individual variables, simply minimizing the support of the residualized bids with respect to $\tilde{\mu}$ yields a superconsistent estimator for μ. We show this in appendix B.1 by noting that the behavior of $b_{\text{resid};t}(m;T) = b^*_{ct} - x_t\big(\mu + m/\sqrt{T}\big)$ is different depending on whether $m = 0$ or $m \neq 0$. Indeed, the difference between the maximum and minimum order statistics of $\{b_{\text{resid};t}(m;T)\}$ exceeds the length of the support of b_{ct} with probability approaching one if $m \neq 0$.

Our estimator $\hat{\mu}$ defined in (10) also makes use of a related property of convolutions, namely that the density of $b_{\text{resid};t}(m;T)$ at the top of its support tends to zero as $T \to \infty$ if $m \neq 0$ even though assumptions A and B imply it is greater than zero if $m = 0$: a zero density translates into an unbounded α at the top end, by assumption B, which contradicts assumption A. Letting $Q_c^{(\tilde{\mu})}$ denote the true quantile function of the residualized bids, the population object $\alpha^{(\tilde{\mu})}(p)$ defined as $Q_c^{(\tilde{\mu})}(p) + Q_c^{(\tilde{\mu})\prime}(p)p$ is larger for values of p near one if $\tilde{\mu} \neq \mu$ because $Q_c^{(\tilde{\mu})}(1) > Q_c^{(\mu)}(1)$ and because $\infty = Q_c^{(\tilde{\mu})\prime}(1) > Q_c^{(\mu)\prime}(1)$. Thus, both the enlarged support of the residualized bids and the zero density when $\tilde{\mu} \neq \mu$ contribute to more extreme estimates of $\alpha_T^{(\tilde{\mu})}$ at the endpoints.[20] As an example, figure 3 graphs the estimated $\alpha_T^{(\tilde{\mu})}$ for $\tilde{\mu} = \mu$ and $\tilde{\mu} = \mu + 0.05$ for the same simulated sample of bids.

We are now ready to state our first theorem.

Theorem 1. *Let assumptions A to C hold. Then $\breve{e}_T$ satisfies $\sqrt{T}\{\breve{e}_T(\cdot) - e(\cdot)\} \rightsquigarrow \mathbb{G}$, on $[0,1]$, where $\mathbb{G}$ is a centered Gaussian process with covariance kernel $H(p,p^*) = \zeta(p)\zeta(p^*)H^*\big\{Q_c(p), Q_c(p^*)\big\}$, where $\zeta(p) = Q_c'(p)p$ for $p \neq 0$ and $\zeta(0) = 0$.*

Moreover, for any fixed $0 < p < 1$, for any closed interval $\mathscr{P} \subset (0,1)$, if α is continuously differentiable then $\sqrt[3]{T}\max_{p\in\mathscr{P}}|\alpha_T(p) - \alpha(p)| = O_p(1)$. □

It should be noted that weak convergence of quantile processes for distributions with compact support is usually stated on $(0,1)$; see e.g. van der Vaart (2000, p308). The reason is that Hadamard–differentiability only obtains on the interior of $[0,1]$. We prove that that weak convergence of our quantile–related process in fact obtains on $[0,1]$.

Cube–root–T convergence of α_T is not surprising in view of e.g. Kim and Pollard (1990). Indeed, $\sqrt[3]{T}\{\alpha_T(p) - \alpha(p)\}$ has a Chernoff limit distribution at each fixed p. Further, equations (15) and (16) in Jun et al. (2015) suggest that

$$\sqrt[3]{T}\{\alpha_T(p) - \alpha(p)\} \xrightarrow{d} \alpha'(p)\arg\max_{t\in\mathbb{R}}\{\mathbb{G}^\circ(t) - \alpha'(p)t^2/2\}, \tag{11}$$

[20] The density of the residualized bid is also zero at the left boundary of its support, but this does not impact the estimates of $\alpha^{(\tilde{\mu})}$ because $\alpha_T^{(\tilde{\mu})}(0)$ is equal to the lowest order statistic of the residualized bids, by construction.

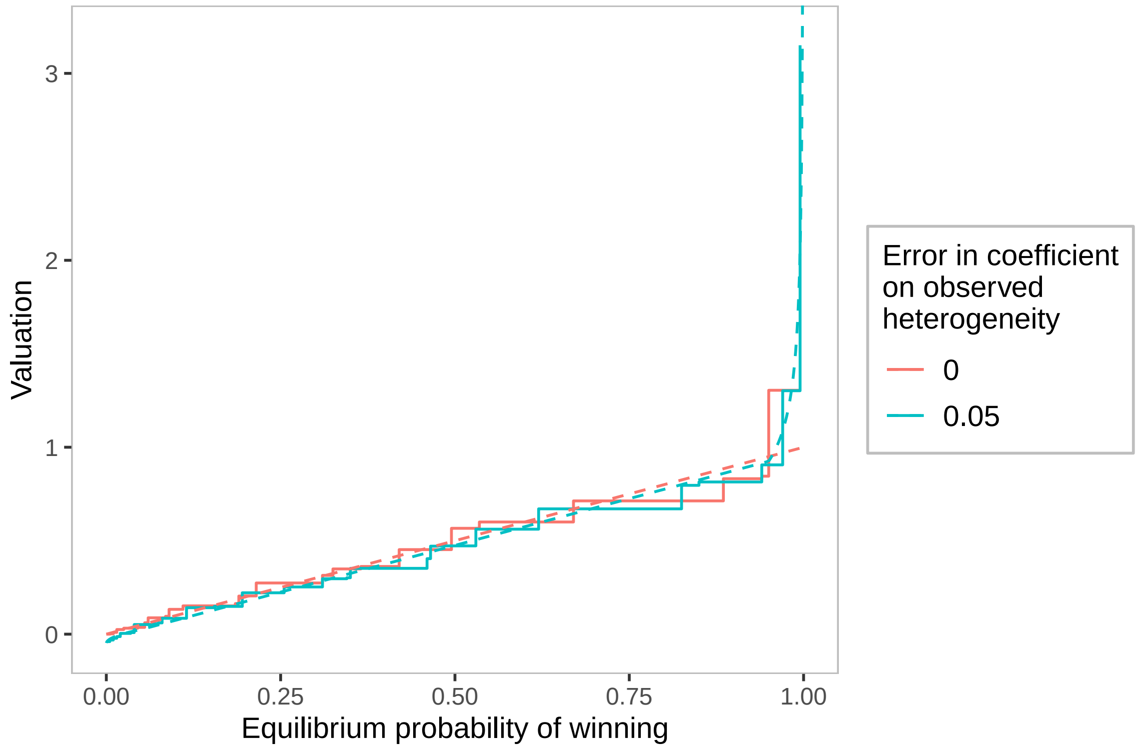

Figure 3: The estimated $\alpha_T^{(\tilde{\mu})}$ for $\tilde{\mu} - \mu = 0$ and 0.05 using the same sample of 200 bids generated by a symmetric two–bidder auction where V_1, V_2, and X are independent standard uniform random variables. The dashed lines represent $Q_c^{(\tilde{\mu})}(p) + Q_c^{(\tilde{\mu})\prime}(p)p$, where $Q_c^{(\tilde{\mu})}$ is the true quantile function of the residualized bids.

where $\mathbb{G}^\circ$ is a Gaussian process. A justification and description of the properties of $\mathbb{G}^\circ$ can be found in appendix A.1. If (7) holds then (11) simplifies to

$$\sqrt[3]{T}\{\alpha_T(p) - \alpha(p)\} \xrightarrow{d} \sqrt[3]{4\zeta^2(p)\alpha'(p)}\,\mathbb{C}\,, \tag{12}$$

where $\mathbb{C}$ is a standard Chernoff–distributed random variable. Equation (12) is also justified in appendix A.1.

Our result for α_T in theorem 1 extends the convergence rate result to uniform convergence. Note that this is in contrast to e.g. nonparametric kernel regression or density estimation where uniform convergence only obtains at a slower rate. Note that uniformity is in p: all results are for a fixed e.[21]

3.3 Nonparametric maximum likelihood

3.3.1 Definition of the estimator

To this point we have relied on the least–squares criterion to motivate the constrained estimator for e and the isotonic regression estimator for its derivative. The LS estimator has the feature that it may be broadly

[21] So, we do not claim uniformity in e.

applied in any auction or auction–like setting as long as an appropriate unconstrained estimate e_T is available. In this section, we develop an estimator based upon a nonparametric likelihood criterion that specifically exploits the structure of a first–price auction. The nonparametric maximum likelihood estimator (MLE) has the same asymptotic distribution as the least–squares estimator, but the two estimators differ in systematic ways in finite samples. In particular, the MLE for α exhibits less finite sample bias than the least–squares estimator when the true expected payment function is relatively convex, as may be anticipated when the bidder is strong relative to its maximum of its rivals. In contrast, when many roughly symmetric bidders compete in an auction, the equilibrium expected payment function is less convex and the least–squares estimator may be preferred.

For ease of exposition and notation, we assume that only the competitors' maximum bid is used to construct the likelihood, though we note how more data may be used to produce a more efficient estimate in section 3.3.2.

We rearrange the familiar formula for a bidder's inverse strategy function

$$\alpha\{G_c(b)\} = b + \frac{G_c(b)}{g_c(b)},$$

in order to relate the density of a bidder's highest competing bid to her expected payment function:

$$g_c(b) = \frac{e\{G_c(b)\}/b}{\alpha\{G_c(b)\} - b}.$$

The loglikelihood of an independent sample of highest competing bids may then be written as

$$\mathscr{L}(\tilde{\alpha}, \tilde{e}) = \sum_{t=1}^{T} \left\{ \log \tilde{e}_t - \log b_t - \log(\tilde{\alpha}_t - b_t) \right\}, \tag{13}$$

where the b_t's are the maximum competitor bid and the shorthand forms $\tilde{e}_t$ and $\tilde{\alpha}_t$ are candidate values for $e\{G_c(b_t)\}$ and the left–derivative of e evaluated at $G_c(b_t)$, respectively. Before (13) may be used as the basis for a nonparametric maximum likelihood estimator, a few remarks are in order. First, nondecreasing convex real–valued functions defined on $[0, 1]$ are continuous on $[0, 1)$,[22] which implies that e is uniquely determined on $[0, 1)$ by its left-derivative $\tilde{\alpha}$. We will therefore replace $\tilde{e}$ with a function of $\tilde{\alpha}$ in what follows. Second, for a given $\tilde{e}$, the implied G_c will be a proper distribution function if $\tilde{e}$ is convex and $\tilde{e}(1)$ equals the highest order statistic among the rivals' observed bids. Third, because the loglikelihood contribution of b_t is increasing in $\tilde{e}_t$ and decreasing in $\tilde{\alpha}_t$, the shape–constrained MLE should be piecewise linear in order to minimize the density at values of b between realizations of the competitors' bids while maximizing the density at the observed bids. In particular, kinks in the MLE $\check{e}_T^{\text{MLE}}$ occur precisely where $\check{e}_T^{\text{MLE}}(p)/p = b_t$ for some observed bid b_t. We

[22] A nondecreasing convex function defined on a compact interval can jump discontinuously at the right boundary.

can therefore maximize (13) by searching over the space of left–continuous, nondecreasing step functions $\tilde{\alpha}$ defined on the unit interval.

Using these observations, the Lagrangian for the isotonic maximum likelihood problem becomes[23]

$$\max_{\{\tilde{\alpha}_t\}_{t>2}} \left(\sum_{t=1}^{T} \left\{ (t-2)\log(\tilde{\alpha}_{(t)} - b_{(t)}) - (t-1)\log(\tilde{\alpha}_{(t)} - b_{(t-1)}) \right\} + \lambda_2(\tilde{\alpha}_{(2)} - b_{(2)}) + \sum_{t=4}^{T} \lambda_t(\tilde{\alpha}_{(t)} - \tilde{\alpha}_{(t-1)}) \right), \quad (14)$$

where $b_{(t)}$ is the t-th lowest order statistic. This problem can be solved using an efficient pool–adjacent–violators algorithm (PAVA), which divides the large optimization problem into a sequence of at most $T-3$ one-dimensional optimizations. Theorem 2 guarantees the algorithm converges to a global maximum.

Theorem 2. *The final iterate of the PAVA is the nonparametric MLE α_T^{mle} for α.* □

Additional details of the derivation of the likelihood criterion and the PAVA can be found in appendix B.4. The key steps use piecewise–linearity and continuity of the expected payment function to replace e_t in (13) with a function of $\{\alpha_{(s)}\}_{s>t}$ and b_T. Cancelling terms and adding the constraints $\alpha_{(2)} > b_2$ and $\alpha_{(t)} \geq \alpha_{(t-1)}$ yields the result above.

We next obtain the MLE of the equilibrium payment function by substituting α_T^{mle} into equation (56). Figure 4 depicts the maximum likelihood estimator for e in comparison with $\breve{e}_T$ for a sample of five rival bids. In larger samples, the differences in the estimators for e are not visually apparent.

As can be seen in figure 4, the MLE is invariably above the LS estimator. This is no coincidence. The nodes of the GCM are positioned at integer multiples of $1/T$ by construction, whereas the MLE can move the position of the nodes as well as the values at the nodes. The MLE can therefore achieve both convexity and proximity to the original nonconvex estimator without having to duck below the original estimator everywhere.

3.3.2 Alternative characterization of the MLE

An estimate $\beta_T(v)$ of bidder one's bid function at v can be obtained as the minimizer of[24]

$$\mathbb{S}_T(b, v) = \sum_{t=1}^{T} \left(\frac{t-2}{v - b_{(t)}} - \frac{t-1}{v - b_{(t-1)}} \right) \mathbb{1}(b_{(t)} \leq b).$$

The estimator $\beta_T(v)$ can be viewed as an inverse monotone regression of something akin to a score function. This method works because, for fixed b, $\mathbb{S}_T$ crosses zero exactly one time from above on (b, ∞) (see lemma 6). The inverse of β_T yields an estimator α_T^{mle} of α at $p = G_c(b)$. Indeed, it turns out that both $\sqrt[3]{T}\{\beta_T(v) - \beta(v)\}$

[23] We omit the constraint $\tilde{\alpha}_{(3)} > \tilde{\alpha}_{(2)}$ from the Lagrangian because this constraint is always slack.

[24] To our knowledge, this is the first direct estimator of the bid function in a first–price auction. Although the inverse bid function estimated using alternative methods can be inverted after monotonizing the estimate (if necessary), this simple, input–parameter–free estimator of the bid function may be more convenient in some applications.

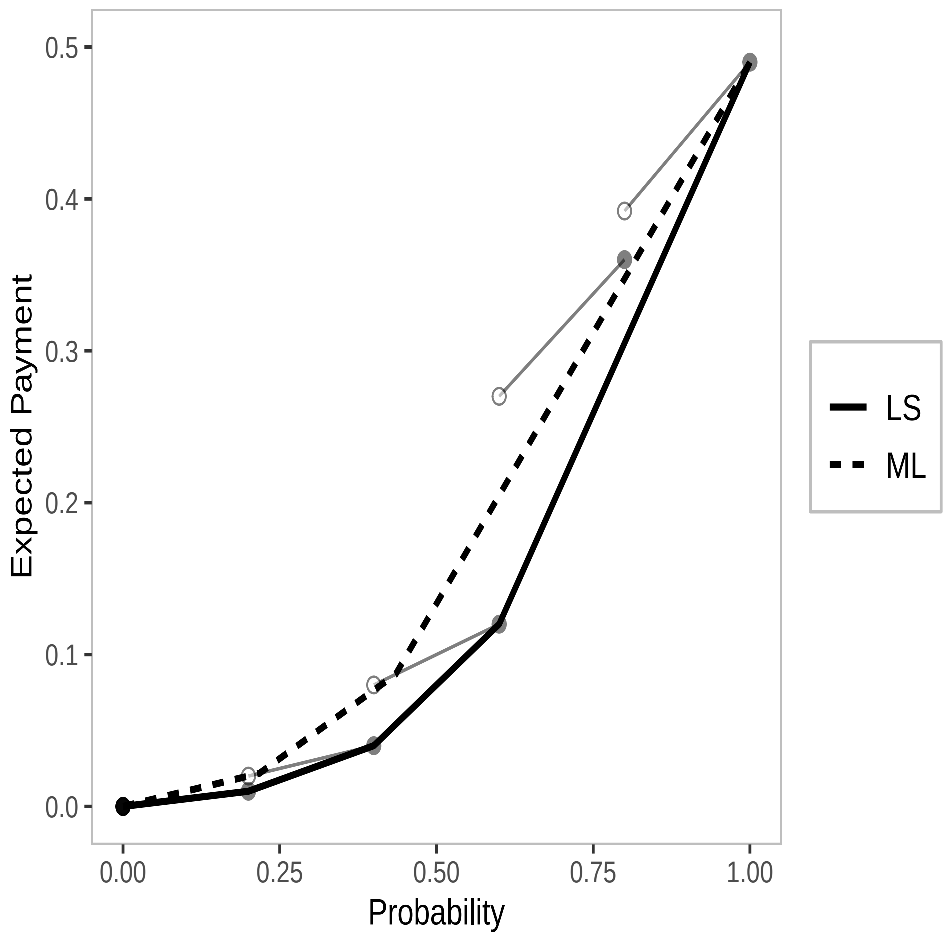

Figure 4: An illustrative example of the maximum likelihood estimator (dashed) for the expenditure function compared to the greatest convex minorant $\breve{e}_T$ (solid) of the unconstrained estimator e_T (grey).

and $\sqrt[3]{T}\{\alpha_T^{\text{mle}}(p) - \alpha(p)\}$ have limiting Chernoff distributions. Indeed, we have

$$\forall 0 < p < 1 : \sqrt[3]{T}\{\alpha_T^{\text{mle}}(p) - \alpha(p)\} \xrightarrow{d} \sqrt[3]{4\zeta^2(p)\left\{2Q_c'(p) + Q_c''(p)p\right\}}\mathbb{C}, \tag{15}$$

where $\mathbb{C}$ is a standard Chernoff distribution. A sketch of the proof and a derivation of the limit distribution can be found in appendix A.2. The limit distribution in (15) is the same as that in (12) under (7). So the MLE and LS approaches have the same (first order) limit behavior under the same assumptions, but the LS approach covers more cases.[25]

In the case of n symmetric bidders, the likelihood of the pooled sample of bids is obtained from $(n-1)g(b) = G(b)/(v-b)$, which becomes $\{e(p)/p\}^{1/(n-1)}/\{\alpha(p) - Q(p^{1/(n-1)})\}$ after a change of variables. The MLE

[25]We only consider first order asymptotics in this paper.

can then be computed by applying the PAVA to the objective

$$\sum_{\ell=1}^{nT}\left\{\frac{\ell-n}{n-1}\log\big(\alpha_{(\ell)}-b_{(\ell)}\big)-\frac{\ell-1}{n-1}\log\big(\alpha_{(\ell)}-b_{(\ell-1)}\big)\right\}.$$

The maximum likelihood estimator for the bid function at a fixed v is then given by the minimizer over b of

$$\sum_{\ell=1}^{nT}\left\{\frac{\ell-n}{(n-1)(v-b_{(\ell)})}-\frac{\ell-1}{(n-1)(v-b_{(\ell-1)})}\right\}\mathbb{1}(b_{(\ell)}\leq b).$$

A formal derivation of this estimator's limit distribution would follow the same outline as the derivation of (15) in appendix A.2. Similar to the case in which only the highest rival bid is used, we conjecture that the limit distribution coincides with that of the least–squares estimator in the symmetric case. The maximum likelihood estimator using the full vector of bids in the asymmetric case is more complicated, because the mapping from α to the marginal densities of the other bidders' bids is not one-to-one. Rather than derive the maximum likelihood of observing an independent sample of rivals' bids for a candidate α, we can apply the PAVA to the objective

$$\begin{aligned}&\int \log e(G_c(b))-\log b-\log\big(\alpha(G_c(b))-b\big)\,\mathrm{d}G_{cT}(b)\\&\quad=\sum_{\ell=1}^{(n-1)T}\left\{\big[2G_{cT}\big(b_{(\ell-1)}\big)-G_{cT}\big(b_{(\ell)}\big)\big]\log\big(\alpha_{(\ell)}-b_{(\ell)}\big)-G_{cT}\big(b_{(\ell-1)}\big)\log\big(\alpha_{(\ell)}-b_{(\ell-1)}\big)\right\}.\end{aligned}$$

where the unconstrained estimator for the highest rival bid is equal to the product of the marginal distributions of the other bidders' bids: $G_{cT}=\prod_{j>1}G_{jT}$. The maximizer of this objective is the monotone inverse strategy function α for which the implied distribution of the maximum rival bid is closest (in the sense of Kullback-Leibler divergence) to the unconstrained MLE for G_c given an independent sample of rivals bids. When there are $n=2$ bidders, the above criterion reduces to the criterion in (13).

4 Smoothing, transformations, and boundary correction

The unsmoothed estimates of the preceding section are sufficient to optimally estimate several objects, such as the bidders' ex ante expected surplus or the counterfactual expected revenue in an auction with a different number of bidders. In these cases, the reader may skip the results in this section and proceed to section 5.

On the other hand, if the researcher is ultimately interested in the valuation distribution, counterfactual expected revenue under a different reserve price, or other objects that depend on the pointwise rate of convergence of the estimator for α, then the smoothing methods in this section are needed to achieve the optimal nonparametric rate of convergence. Indeed, if α is twice continuously differentiable then the $\sqrt[3]{T}$ convergence rate of the unsmoothed estimators of α can be improved to the standard nonparametric $T^{2/5}$ rate. In this section, we first introduce our basic smoothing method, which replaces the unsmoothed estimator with its local weighted average and is similar to the smoothing method in LW. We then develop two important enhancements—boundary correction and transformation—that are needed to achieve the optimal rates of convergence for $\alpha(p)$ at the boundaries $p \in \{0, 1\}$. Our analysis and derivations use the LS estimator as the basis for our results, but that is inessential.

As noted by Hickman and Hubbard (2015, HH), boundary correction can be important in the estimation of auction models, especially if the objective is to estimate the density of valuations. The reason is that the bid distribution (in HH) or the distribution of win probabilities (here) has compact support and it is well–known that, absent a boundary correction, most nonparametric density estimators are inconsistent at the boundaries. The situation is more favorable in our case since we know that probabilities vary from zero to one whereas the top of the bid distribution must be estimated, albeit that this can be done super–consistently. We provide two distinct boundary correction methods, one based on boundary kernels, and one on a boundary correction scheme in the spirit of HH. As expected, both methods yield vast improvements on the performance of our uncorrected estimators near the boundary. In developing these methods, we have identified an improvement in the choice of the bandwidth sequence recommended in KZ, which improves the performance of HH's version of the GPV estimator substantially. This improvement is described in a separate paper, Pinkse and Schurter (2019).

Our smoothed estimators for α can be further improved by applying a transformation ψ to the win–probabilities as part of the smoothing method. Indeed, we show that such transformations ψ can improve the first order asymptotic mean square error, though not the pointwise convergence rate, of our smoothed estimators of α. In addition, because the inverse strategy and inverse bid functions can have unbounded derivatives at the left, boundary, our transformation method is required to ensure the mean squared error is integrable over $[0, 1]$.[26] The effect of such transformations on first order asymptotics help explain the feature noted in Ma et al. (2019a, Ma19a) that the asymptotic variance of the MS quantile–based–estimator of f_v is often greater than that of the corresponding GPV estimator. These transformation methods are complements, not substitutes, to our boundary correction methods.

[26]The transformation is similar in spirit to the change of variable employed in GPV to accomodate the unbounded bid density in the presence of a binding reserve price. When there is a binding reserve price and bidders are symmetric, the bid density behaves like $1/\sqrt{b-r}$ when b is close to the reserve price r. In which case, a consistent estimate can be obtained from a kernel density estimate of a known transfomation of the data. In this paper, however, we do not know the rate at which α'' diverges as p approaches zero. We also note that α'' is bounded in the presence of a binding reserve price. Hence, transformations are only potentially needed when the reserve price does not bind.

4.1 Smoothing

The main limitation of our method above is that α_T converges at a $\sqrt[3]{T}$ rate. This is due to the fact that α_T is discontinuous and hence that $\breve{e}_T$ is kinky. To obtain convergence at the typical nonparametric $n^{2/5}$ rate, we can replace $\breve{e}_T$ with a smoothed version $\hat{e}_T$, defined by,

$$\hat{e}_T(p) = \frac{1}{h}\int_{-\infty}^{\infty} \breve{e}_T(s)\, k\Big(\frac{s-p}{h}\Big)\, \mathrm{d}s,$$

where k is a twice continuously differentiable kernel with compact support for which $\int_{-\infty}^{\infty} k(s)s^2\, \mathrm{d}s = 1$,[27] and $h = h_T$ is a bandwidth such that $\Xi = \lim_{T\to\infty} \sqrt{Th^5} < \infty$. The restriction on the bandwidth sequence is not necessary for consistency of $\hat{e}_T$: unlike kernel–estimators employed by others, $\hat{e}_T$ is a consistent estimator of e for all bandwidth sequences that converge to zero and the same is true for $\hat{\alpha}_T = \hat{e}'_T$.

This definition of $\hat{e}_T$ requires modification near the boundaries because $\breve{e}_T$ is not defined outside $[0, 1]$. We address this issue in section 4.3. Before providing results for smoothed estimates of α evaluated away from the boundary, we need one further assumption.

Assumption D. *Q_c is thrice continuously differentiable on any closed interval $\mathscr{P}_Q \subset (0, 1]$.* □

Assumption D is essentially equivalent to assuming that $g_c = G'_c$ is twice continuously differentiable, which is implied by continuous differentiability of the value densities (GPV). Thus, assuming one more continuous derivative in assumption A is sufficient for assumption D. Assuming that a density is twice continuously differentiable is standard in the nonparametric kernel estimation literature. The compact subset requirement is needed since $g_c(0)$ may be zero.

Theorem 3. *Under assumptions A to D, $\hat{e}_T$ is convex, has the same limit properties as $\breve{e}_T$ on a closed interval $\mathscr{P}$ contained in $(0, 1)$, and*[28]

$$\forall p \in \mathscr{P} \,:\, \sqrt{Th}\big\{\hat{\alpha}_T(p) - \alpha(p)\big\} \xrightarrow{d} N\{a''(p)\Xi/2, \mathscr{V}\},$$

where $\mathscr{V}(p) = \lim_{h\to 0} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathscr{H}_h(p, s, \tilde{s})k'(s)k'(\tilde{s})\, \mathrm{d}\tilde{s}\, \mathrm{d}s$, *with*

$$\mathscr{H}_h(p, s, \tilde{s}) = \{H(p + sh, p + \tilde{s}h) - H(p + sh, p) - H(p, p + \tilde{s}h) + H(p, p)\}/h. \tag{16}$$

□

[27] $\int k(s)s^2\, \mathrm{d}s = 1$ is a normalization.

[28] The function $\hat{\alpha}_T$ converges as a process in an h–neighborhood of p, but not on any interval of positive length. Regular kernel density and regression function estimates have the same lack–of–tightness property. The local tightness result is not in this paper. A consequence of the lack of tightness is that $\hat{\alpha}_T(p)$ and $\hat{\alpha}_T(p^*)$ are asymptotically independent for fixed distinct p, p^*.

The variance formula in theorem 1 is intimidating, but in many cases it simplifies substantially. First, as noted following assumption C, if G_{cT} is taken to be the empirical distribution function of the maximum rival bid then (7) holds.

Lemma 1. *If* (7) *holds then*

$$H(p, p^*) = \zeta(p)\zeta(p^*)\{\min(p, p^*) - pp^*\}, \tag{17}$$

and $\mathscr{V}(p)$ *in theorem 3 simplifies to* $\mathscr{V}(p) = \zeta^2(p)\kappa_2$, *where* $\kappa_2 = \int_{-\infty}^{\infty} k^2(s)\,\mathrm{d}s$. □

Since k is chosen, $\mathscr{V}$ is easy to estimate.

A second simplification obtains under full symmetry, i.e. when (8) holds.

Lemma 2. *If* (8) *holds then*

$$H(p, p^*) = \frac{1}{n}\big\{\min(p, p^*)^{1/(n-1)} - (pp^*)^{1/(n-1)}\big\}pp^* Q'(p^{1/(n-1)})Q'(p^{*\,1/(n-1)}). \tag{18}$$

Further,

$$\lim_{h\to 0} \mathscr{H}_h(p, s, \tilde{s}) = \frac{p^{n/(n-1)}Q'^2(p^{1/(n-1)})|\mathrm{Med}(s, \tilde{s}, 0)||}{n(n-1)},$$

and $\mathscr{V}$ *simplifies to* $p^{n/(n-1)}Q'^2(p^{1/(n-1)})\kappa_2/\{n(n-1)\}$. □

Finally, we provide a result for the asymmetric IPV case with n bidders.

Lemma 3. *If* (9) *holds then*

$$H(p, p^*) = \zeta(p)\zeta(p^*)pp^* \sum_{i=2}^{n} \left(\frac{1}{G_i[Q_c\{\max(p, p^*)\}]} - 1\right). \tag{19}$$

Further,

$$\lim_{h\downarrow 0} \mathscr{H}_h(p, s, \tilde{s}) = \zeta^2(p)Q_c'(p) \sum_{i=2}^{n} G_{-i1}^2\{Q_c(p)\}g_i\{Q_c(p)\}|\mathrm{Med}(s, \tilde{s}, 0)|, \tag{20}$$

where G_{-i1} *means the distribution of the maximum bid of all bidders other than* i *and* 1 *with* $G_{-i1} = 1$ *if there are only two bidders. Finally,* $\mathscr{V}$ *equals* $\kappa_2\zeta^2(p)Q_c'(p)\sum_{i=2}^n G_{-i1}^2\{Q_c(p)\}g_i\{Q_c(p)\}$. □

Note that for $n = 2$, the result in lemma 3 reduces to that in lemma 1. For $n > 2$, the function H in lemma 3 is generally more favorable, i.e. it is more efficient to estimate each rival bid distribution separately than to estimate the distribution of the maximum rival bid using only the maximum rival bids.[29]

A more interesting comparison is that of the formulas for $\mathscr{V}$ in lemmas 2 and 3 if there is symmetry. Indeed, the ratio of variances is $(n-1)/n$ in favor of exploiting symmetry. This result is intuitive since

[29]For the case in which rival distributions happen to coincide but this fact is not used in the estimation, $\mathscr{V}$ in lemma 3 reduces to $\kappa_2\zeta^2(p)p^{(n-2)/(n-1)}$ which equals $\mathscr{V}$ in lemma 1 if $n = 2$ or $p \in \{0, 1\}$ but is otherwise less. More generally, note that $\mathscr{V}$ in lemma 3 is $\kappa_2\zeta^2 \sum_{i=2}^n G_{-i1}^2 g_i / \sum_{i=2}^n G_{-i1} g_i$ which is equal to $\kappa_2\zeta^2$ and hence to $\mathscr{V}$ in lemma 1 if $G_{-i1} = 1$, i.e. if $n = 2$ or $p = 1$.

exploiting symmetry means that one can also use the bids of bidder one to estimate G_c: one then uses data on n bids per auction instead of $n-1$.

One limitation of theorem 3 compared to theorem 1 is that theorem 3 does not extend to all of $[0,1]$ due to boundary effects. A second issue is that the bias of $\hat{\alpha}_T$ can be large for small values of p as the following example illustrates, even though the variance is vanishingly small near $p=0$.[30]

Example 1. *Consider the symmetric case with F_v a standard uniform and $n=3$. Then $e(p)=Q_c(p)p=2p^{3/2}/3$ and $e'''(p)=-p^{-3/2}/4 \to -\infty$ as $p \downarrow 0$.* □

We address each of these limitations in turn, beginning with the large bias for p close to zero.

4.2 Transformations

Because the bias in our estimate of α can explode at the left boundary while the variance is vanishingly small, a natural solution is to use a smaller bandwidth near the left boundary than elsewhere.[31] Because we also know a little about the shape of α near zero, e.g. that $Q_c'(p)p$ converges to zero, rather than use a variable bandwidth, we prefer to introduce an estimator that uses transformations: a variant of this estimator can be found in appendix B.5.1. Let ψ be an increasing function such that for $j=1,2,3$, $e^{(j)}(p)/\psi'^j(p)$ and $\psi^{(j)}(p)/\psi'^j(p)$ are uniformly bounded on $(0,1]$ and for which $\lim_{p\downarrow 0}$ of each of these functions is finite, also.[32] Then define

$$\bar{\alpha}_{T\psi}(p)=\frac{1}{h}\int_{-\infty}^{\infty}\psi'(s)\alpha_T(s)k\Big(\frac{\psi(p)-\psi(s)}{h}\Big)\,\mathrm{d}s. \tag{21}$$

To see how (21) can alleviate the exploding bias near zero, we use the change of variables $z=[\psi(p)-\psi(s)]\,/\,h$ and note that the second–order term in a Taylor expansion of $\alpha\{\psi^{-1}[\psi(p)-zh]\}$ around $\alpha(p)$ involves

$$\frac{\alpha''(p)}{\psi'^2(p)}-\frac{\alpha'(p)\psi''(p)}{\psi'^3(p)}. \tag{22}$$

A standard argument in kernel based estimation then implies the bias is proportional to (22) if k is a second order kernel. If ψ is chosen such that ψ' in the denominator is large enough, the asymptotic bias will be finite near zero. The asymptotic bias is zero if one chooses $\psi=\alpha$, which is unfortunately infeasible because α is unknown.[33] A feasible choice is illustrated in example 2.

Example 2. *Recall example 1. If one uses $\psi(p)=\log p$ then $\psi'(p)=1/p$, $\psi''(p)=-1/p^2$, and $\psi'''(p)=2/p^3$. This yields for instance $\alpha''(p)/\psi'^2(p)=-\sqrt{p}$, which is well–behaved near zero. One can verify that*

[30] The GPV estimator also has the unbounded bias at zero problem.

[31] Interestingly, this contradicts the typical practice of using a larger bandwidth near the boundary when estimating a density with boundary kernels.

[32] By ψ'^j we mean the derivative of ψ raised to the jth power.

[33] Perhaps a two–step procedure in which an inefficient estimate of α is obtained in a first stage to inform ψ is feasible. We do not explore such a procedure here.

the other ratio in (22) is equally well–behaved.

Note that our solution does not only work for $n = 3$. Indeed, in the symmetric case with arbitrary n, $e(p) = (1 - 1/n)p^{n/(n-1)}$, such that $\alpha''(p)/\psi'^2(p) \sim p^{1/(n-1)}$.

The same goes for the situation in which there are $n - 1$ stronger rivals. It really does not matter since each derivative of e removes a power of p and each negative power of ψ' restores one. □

Example 2 shows one choice of ψ, indeed a conservative choice. In fact, it is guaranteed to eliminate the bias at $p = 0$ in any feasible auction mechanism, because $\alpha'(p)p$ and $\alpha''(p)p^2$ must converge to zero as p approaches zero in order for α to be finite for all p in a neighborhood of zero. We note, however, that this choice makes the variance larger near zero and possibly unbounded.

4.3 Boundary correction

When computing $\hat{e}_T$ at values of p near the boundary or using a kernel with infinite support, the locally weighted average of $\breve{e}_T(s)$ attempts to put positive weight on values of $\breve{e}_T$ for which $\breve{e}_T$ is undefined. If one does not make adjustments to the kernel k or the definition of $\breve{e}_T$ outside of $[0, 1]$ then $\alpha(1)$ will not be consistently estimated, as the following example illustrates for $\bar{\alpha}_{T\psi}$.

Example 3. *The estimator $\bar{\alpha}^{\text{bad}}_{T\psi}(p) = h^{-1} \int_0^1 \psi'(s)\alpha_T(s)k\big((\psi(p) - \psi(s))/h\big)\,\mathrm{d}s$ is inconsistent at $p = 1$. To see this, note that*

$$\bar{\alpha}^{\text{bad}}_{T\psi}(1) = \frac{1}{h}\int_0^1 \psi'(s)\alpha_T(s)k\Big(\frac{\psi(1) - \psi(s)}{h}\Big)\,\mathrm{d}s = \\ \frac{1}{h}\int_0^1 \psi'(s)\alpha(s)k\Big(\frac{\psi(1) - \psi(s)}{h}\Big)\,\mathrm{d}s + o_p(1) = \alpha(1)\int_{-\infty}^0 k(-s)\,\mathrm{d}s + o_p(1) = \frac{\alpha(1)}{2} + o_p(1),$$

by consistency of α_T and substitution of $s \leftarrow \{\psi(s) - \psi(p)\}/h$. This is the well–known boundary bias problem of nonparametric kernel density estimators. □

If the lower end of the valuation's support is zero then $\alpha(0) = 0$ and the estimator $\bar{\alpha}^{\text{bad}}_{T\psi}(0)$ *is* a consistent estimator of $\alpha(0) = 0$. Note that the problem is true whether $\psi(p) = p$ or not.

There are many solutions to this problem. The traditional approach is to use a 'boundary kernel,' i.e. a kernel that scales the kernel to make up for the lost mass if a function is estimated near a boundary. We discuss this possibility in section 4.3.1. A second possibility is to make use of techniques similar to those espoused in Karunamuni and Zhang (2008, KZ) in order to "make up" values of e and α beyond $[0, 1]$. This approach is investigated in appendix B.5.3.[34]

[34] Gimenes and Guerre (2019) smooth the quantile function using a local polynomial approach. The problem studied therein is otherwise unrelated.

4.3.1 Boundary kernels

The boundary bias problem can be addressed by the use of boundary kernels. We replace (21) with

$$\bar{\alpha}_{T\psi}(p) = \frac{1}{h}\int_0^1 \psi'(s)\alpha_T(s)k_{\psi h}\Big(\frac{\psi(p)-\psi(s)}{h} \wr p\Big)\,\mathrm{d}s, \tag{23}$$

where, for each $p \in [0,1]$, the function $k_{\psi h}(\cdot \wr p)$ is a boundary kernel defined now. An alternative to (23) is (60) in appendix B. Let $\bar{\upsilon}_\psi = \{\psi(1)-\psi(p)\}/h$ and $\underline{\upsilon}_\psi = \{\psi(0)-\psi(p)\}/h$. Then we require $k_{\psi h}$ to be such that for all $p \in [0,1]$, $j = 0,1,2$,

$$\lim_{h\downarrow 0}\int_{\underline{\upsilon}_\psi}^{\bar{\upsilon}_\psi} s^j k_{\psi h}(-s \wr p)\,\mathrm{d}s = |1-j|, \tag{24}$$

where the requirement for $j=2, p\in\{0,1\}$ is replaced with boundedness. The requirement that the kernel integrate to one is to ensure consistency in view of example 3. We also want it to integrate to zero if multiplied by s to kill the 'h term' in a bias expansion. The cut–out for $j=2$ and $p\in\{0,1\}$ in the requirements for the boundary kernel is there because the requirements on $\int s^2 k_{\psi h}(-s)$ only affect the 'bias' in the asymptotic distribution and because it simplifies the formula for the boundary kernel. Such boundary kernels are easy to construct from second–order kernels, as lemma 14 demonstrates using the Gaussian kernel ϕ. A formula for a boundary kernel that satisfies (24) for $j=2$ and all $p\in[0,1]$ is provided in lemma 15 in appendix A.

Assumption E. *The transformation ψ is thrice continuously differentiable on $(0,1]$ with ψ' positive. Further the function $\Psi(s) = \alpha\{\psi^{-1}(s)\}$ is twice continuously differentiable.* □

The extra assumption on Ψ in assumption E only has bite at $s=\psi(0)$. The concern is that $g_c(0)=0$ in which case $Q_c'(0)=\infty$ and derivatives of α at zero can be infinite. For instance, for power distributions $Q_c(p) = p^\gamma$ for some $\gamma>0$, such that $\alpha(p) = (\gamma+1)p^\gamma$. Taking $\psi(p)=\log p$ yields $\Psi(s) = (\gamma+1)\exp(s\gamma)$, which satisfies assumption E. Less concave choices of ψ may work better but require some knowledge of the curvature of Q_c near $p=0$.

Theorem 4. *Suppose that $k_{\psi h}$ is constructed as in lemma 14, that assumptions A to E are satisfied. Then $\forall p\in(0,1] : \sqrt{Th}\{\bar{\alpha}_{T\psi}(p)-\alpha(p)\} \xrightarrow{d} N\big\{\bar{\mathscr{B}}_\psi(p), \mathscr{V}_\psi(p)\big\}$, where $\bar{\mathscr{B}}_\psi(p) =$ expression (22) $\times\, \Xi/2$ and*

$$\mathscr{V}_\psi(p) = \psi'^2(p)\lim_{h\to 0}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\phi'(s)\phi'(\tilde{s})\mathscr{H}_h\big\{p, s/\psi'(p), \tilde{s}/\psi'(p)\big\}\,\mathrm{d}\tilde{s}\,\mathrm{d}s,$$

where $\mathscr{H}_h$ was defined in (16). For $p=1$, $\bar{\mathscr{B}}_\psi, \mathscr{V}_\psi$ are finite. Under (7), $\mathscr{V}_\psi$ simplifies to $\kappa_2\zeta^2(p)\psi'(p) = \zeta^2(p)\psi'(p)/\sqrt{\pi}$. □

Finally, lemma 4 illustrates that the integral in (23) is easy to compute because α_T is a step–function.

Lemma 4. *The formula for* $\tilde{\alpha}_{T\psi}$ *simplifies to* $\tilde{\alpha}_{T\psi}(p) = \sum_{j=1}^{T} \alpha_{Tj} \Lambda_{\psi j}(p)$, *where* $\Lambda_{\psi j}(p) = K_{\psi h}\{-\upsilon_{j-1}(p) \wr p\} - K_{\psi h}\{-\upsilon_j(p) \wr p\}$, *with* $K_{\psi h} = \int k_{\psi h}$ *and* $\upsilon_j(p) = \{\psi(j/T) - \psi(p)\}/h$. *For* $k_{\psi h}$ *as constructed in lemma 14,* $K_{\psi h}(s) = \omega_{\psi 1}\Phi(s) + \omega_{\psi 2}\phi(s)$. □

4.4 Derivative estimators

As we will see in section 5, the density of the value distribution depends on α', not on α itself. Although the primary objective in our paper concerns estimation of derived objects like the bidder surplus, we include results for the value density in the interest of completeness. For that purpose, we derive some results for an estimator of α', both away from (theorem 5) and near the boundary (theorem 15 in appendix B).

Theorem 5. *Let (i) assumptions A to C be satisfied; (ii)* Q_c *be four times continuously differentiable on any compact subset of* $(0,1)$*; (iii) k be the Epanechnikov kernel; (iv)* $\lim_{T\to\infty} \sqrt{Th^7} = \Xi_d < \infty$. *Then, at any fixed* $0 < p < 1$,

$$\sqrt{Th^3}\{\tilde{\alpha}'_{T\psi}(p) - \alpha'(p)\} \xrightarrow{d} N\big(\mathscr{B}^R(p), \mathscr{V}^R(p)\big),$$

with $\mathscr{B}^R(p) = \Xi_d(\alpha'''\psi'^2 - 3\alpha''\psi''\psi' - \alpha'\psi'''\psi' + 3\alpha'\psi''^2)/(10\psi'^4)$, *where all* α, ψ*'s are evaluated at p, and*

$$\mathscr{V}^R(p) = \frac{9}{4}\psi'^4(p) \lim_{h\downarrow 0} \int_{-1}^{1}\int_{-1}^{1} \mathscr{H}_h\big\{p, s/\psi'(p), \tilde{s}/\psi'(p)\big\}\, \mathrm{d}s\, \mathrm{d}\tilde{s}.$$

If (7) *holds then* $\mathscr{V}^R(p)$ *simplifies to* $(3/2)\psi'^3(p)\zeta^2(p)$. *If* (8) *holds then* $\mathscr{V}^R(p)$ *simplifies to*

$$\frac{3\psi'^3(p)}{2n(n-1)} p^{n/(n-1)} Q'^2(p^{1/(n-1)}). \tag{25}$$

Simplifying expressions for F_p, α *and their first three derivatives in the symmetric case can be found in lemma 9 in appendix A.5.* □

Observe that the optimal convergence rate is the same as that for nonparametric kernel derivative estimators, namely $T^{2/7}$ for $h \sim T^{-1/7}$, as expected. Note further that, like before, the scale of ψ and the bandwidth h are interchangeable. Again, the optimal yet infeasible choice of ψ in terms of the asymptotic bias is $\psi \propto \alpha$.

5 Derived objects

Applied researchers typically are not directly interested in the private values that rationalize a particular sample of bids, but may estimate these so–called pseudo values in order to construct other estimates. For instance, the sample of pseudo values may be used to obtain estimates of the bidder's expected surplus.

The same comment applies to the density of the private value distribution: because the marginal value distributions are the primitives of the model, i.e. any counterfactual outcomes or other objects of interest may be computed using the private value distribution, estimating the (density of the) pseudo values accurately is considered a goal itself in a good chunk of the literature. This intermediate step may be unnecessary or undesirable when the ultimate target of estimation can be written in terms of higher level objects or when the distribution of equilibrium win–probabilities is known.

In each of the examples below, the object of interest can be expressed as $\theta(\alpha, F_p)$, where θ is a known function, and we estimate the object by plugging in some combination of estimates of α and F_p. There are two overarching themes in the following discussion. First, the asymptotic derivations are greatly simplified by the fact that F_p is known in any symmetric equilibrium; we can expect significant improvements in finite–sample (and often also asymptotic) performance when we plug in the true F_p as opposed to an estimated distribution and pool bids across bidders to more accurately estimate the rival bid distributions. Second, we may expect the plug–in estimator for θ to be $\sqrt{T}$–consistent and asymptotically unbiased[35] when θ takes the form $\theta(\alpha, F_p) = \int \theta_1(\alpha)\, \mathrm{d}F_p$ for an appropriately differentiable function θ_1, as is often the case when integrating over nonparametrically estimated objects.

We now turn to a discussion of individual objects to be estimated, first assuming the researcher exploits symmetry both by substituting

$$F_p(p) = p^{1/(n-1)} \quad \text{and} \quad f_p(p) = \frac{1}{n-1} p^{(2-n)/(n-1)} \tag{26}$$

and by pooling bids in the estimation of Q_c so that (8) holds. We later revisit some of the results under asymmetry. Although not the primary objective in our exercise, we briefly discuss how to estimate the value distribution function, quantiles, and density function in section 5.1. We then turn to some objects of greater interest to us, namely the bidder surplus, the mean of the value distribution, profit as a function of the number of bidders, and profit as a function of a hypothetical reserve price.

5.1 Value distribution

There are different attributes of the value distribution that can be estimated. The easiest object to recover is the quantile function. Letting Q_v, Q_b, and Q_p denote the quantiles of bidder one's values, bids, and win–probabilities, note that since $v = \alpha(p) = \alpha\{G_c(b)\}$,

$$Q_v(\tau) = \alpha\{Q_p(\tau)\} = \alpha[G_c\{Q_b(\tau)\}], \qquad \tau \in [0, 1],$$

which simplifies to $\alpha(\tau^{n-1})$ in the symmetric case. The functions G_c, Q_b can be estimated $\sqrt{T}$–consistently, but not so for α as our results thus far have shown. So even though we are estimating quantiles, namely quantiles of the value distribution, these quantiles cannot be estimated at the parametric rate because the

[35] By 'asymptotically unbiased' we mean that the limit distribution has mean zero.

values are not observed. Indeed, the limit distribution of an estimator $\hat{Q}_v(\tau)$ of $Q_v(\tau)$ is simply the limit distribution of $\alpha_T(\tau^{n-1})$ in theorem 3 with the variance derived in lemma 2. Likewise, the value distribution function is simply

$$F_v(v) = F_p\{\alpha^{-1}(v)\} = \alpha^{-1}(v)^{1/(n-1)}.$$

By the delta method, the limit distribution is given by $\alpha^{-1\prime}(v)\{\alpha^{-1}(v)\}^{(2-n)/(n-1)}/\{(n-1)\}$ times the limit distribution of α_T. Note that the delta method is only valid for $v \neq 0, \bar{v}$, which is of little consequence since we already know the values of $F_v(0), F_v(\bar{v})$, albeit that uniformity arguments would suggest that the implied asymptotic distribution would not reflect the finite sample performance near 0 and 1, although the convergence rate is still $T^{2/5}$ for the same reason that $\breve{e}_T$ converges at the $\sqrt{T}$ rate on the entire interval $[0,1]$: see the comments in the paragraph following theorem 1.

To estimate the valuation density in the symmetric case, we note that the value density function is

$$f_v(v) = (f_p/\alpha')\{\alpha^{-1}(v)\} = \frac{\{\alpha^{-1}(v)\}^{(2-n)/(n-1)}}{(n-1)\alpha'\{\alpha^{-1}(v)\}},$$

and hence requires an estimate of α', which we provided in section 4.4. Again by the delta method, the aymptotic distribution of the value density estimator is $-[\alpha^{-1}(v)]^{(2-n)/(n-1)}/[(n-1)\alpha'\{\alpha^{-1}(v)\}^2]$ times the limit distribution of the estimator for α'. From (25) it follows that the bias and variance of our estimator of the value density at $v = \alpha(p)$ are given by

$$\mathscr{B}_f^{\text{symm}}(p) = -\frac{p^{(2-n)/(n-1)}}{(n-1)\alpha'^2(p)}\mathscr{B}^R(p),$$

and

$$\begin{aligned}\mathscr{V}_f^{\text{symm}}(p) = \frac{3(n-1)^5\psi'^3(p)}{2n^5} & \frac{p^{\frac{3n-4}{n-1}}Q'^2(p^{\frac{1}{n-1}})}{\left\{Q'(p^{\frac{1}{n-1}}) + p^{1/(n-1)}Q''(p^{\frac{1}{n-1}})/n\right\}^4} \\ &= \frac{3(n-1)^5\psi'^3\{G^{n-1}(b)\}}{2n^5}\frac{G^{3n-4}(b)g^{10}(b)}{\{g^2(b) - G(b)g'(b)/n\}^4}.\end{aligned}$$

For $\psi(p) = Q_c(p)$ (or indeed a suitable estimate thereof) our variance coincides with that of MS, theorem 2. Ma19a note that the variance of the MS estimator exceeds that of GPV for the same choice of kernel and bandwidth if one undersmooths, i.e. if one chooses a bandwidth which makes the bias disappear faster than the variance. Ma19b point out that the variance of the GPV estimator exceeds that of their estimator, again using bandwidth sequences that undersmooth both estimators. We recommend against undersmoothing for the purpose of estimating α and note that $\psi = Q_c$ is not optimal.[36]

[36]In the nonparametric kernel estimation literature, undersmoothing is sometimes used as a device to obtain asymptotic validity of an inference procedure. For the purpose of estimation, however, undersmoothing is unwise since estimators then converge more

5.2 Bid function

Note that the bid function at v is simply $Q_c\{\alpha^{-1}(v)\}$, and that Q_c can be estimated $\sqrt{T}$–consistently. Hence the limit distribution of our bid function estimate is simply Q_c'/α' times the limit distribution of the estimate of α used. Since the bid function estimate uses an estimate of the inverse of α, the estimate of α had better be monotonic: this is yet another advantage of imposing monotonicity from the outset.

5.3 Bidder surplus

We now turn our attention to estimation of the bidder's expected surplus, which will be of interest to the researcher who seeks to quantify the distribution of the total value of trade between bidders and the seller. The surplus for bidder one is given by

$$\mathrm{BS} = \mathbb{E}\{(V_1 - B_1)\mathbb{1}(B_c \le B_1)\} = \int_0^1 A(p) f_p(p)\,\mathrm{d}p, \tag{27}$$

where $A(p) = \alpha(p)p - e(p) = Q_c'(p)p^2$.

One would expect $\sqrt{T}$–consistency despite the presence of nonparametric objects in the definition of BS. This is a common theme in the semiparametric econometrics literature (see e.g. Robinson, 1988; Powell et al., 1989). Even though nonparametric estimators, other than e.g. the empirical distribution function, typically converge at a rate slower than $\sqrt{T}$, integrating them often restores the parametric $\sqrt{T}$ rate. The reason is that integrating is like averaging and hence reduces the variance, which opens up the possibility of undersmoothing to make the bias vanish at a rate faster than $\sqrt{T}$. Note that if the unsmoothed estimator α_T is used, no adjustment of smoothing parameters is needed at all since no smoothing is conducted in the first place. It does not appear to matter for the asymptotic distribution of our estimator of BS whether or not a smoothed estimator of BS is used, as long as it is undersmoothed.

Bidder symmetry simplifies the asymptotic theory since integration by parts and (26) turns (27) into

$$\mathrm{BS} = \frac{e(1)}{n-1} - \int_0^1 e(p)\{f_p'(p)p + 2f_p(p)\}\,\mathrm{d}p = \frac{e(1)}{n-1} - \frac{n}{n-1}\int_0^1 e(p)p^{\frac{2-n}{n-1}}\,\mathrm{d}p,$$

slowly. Second, the linear expansions (31) provided in section 5.6 suggests that the observation in Ma19a is due to the use of a one–step instead of a two–step estimator. Finally, Ma19a do not employ transformations like ψ, which can yield a smaller variance. Indeed, for the *infeasible* choice $\psi(p) = c\alpha(p)$ for $c > 0$ one obtains a bias of zero and a variance equal to

$$\frac{c^3 K_1 G^2(b)g(b)}{n^2(n-1)\{g^2(b) - g'(b)G(b)/n\}},$$

which can be made small by choosing c small. Thus, any gains one obtains from doing a two–step procedure can be obtained by making a different choice of ψ and kernel or bandwidth. But bear in mind that since this is for an infeasible choice of ψ, one cannot simply implement the optimal choice of ψ and drive c to zero in order to kill the bias and get superconsistency.

which can be estimated by

$$\widehat{\mathrm{BS}}^{\mathrm{symm}} = \frac{\breve{e}_T(1)}{n-1} - \frac{n}{(n-1)^2}\int_0^1 \breve{e}_T(p) p^{\frac{2-n}{n-1}}\,\mathrm{d}p.$$

The asymptotic theory for $\widehat{\mathrm{BS}}^{\mathrm{symm}}$ is trivial in view of theorem 1.

Theorem 6. *Under the assumptions of theorem 1,* $\sqrt{T}(\widehat{\mathrm{BS}}^{\mathrm{symm}} - \mathrm{BS}) \xrightarrow{d} N(0, \mathscr{V}_{\mathrm{BS}}^{\mathrm{symm}})$, *where*

$$\mathscr{V}_{\mathrm{BS}}^{\mathrm{symm}} = \frac{n^2}{(n-1)^4}\int_0^1\int_0^1 H(p,p^*)\,(pp^*)^{\frac{2-n}{n-1}}\,\mathrm{d}p\,\mathrm{d}p^*. \qquad \square$$

From (8) it follows that

$$\begin{aligned}\mathscr{V}_{\mathrm{BS}}^{\mathrm{symm}} &= \frac{n}{(n-1)^4}\int_0^1\int_0^1 Q'\left(p^{\frac{1}{n-1}}\right)Q'\left(p^{*\frac{1}{n-1}}\right)(pp^*)^{\frac{1}{n-1}}\{\min(p,p^*)^{1/(n-1)} - (pp^*)^{1/(n-1)}\}\,\mathrm{d}p^*\,\mathrm{d}p \\ &= \frac{n}{(n-1)^2}\int_0^1\int_0^1 Q'(p)Q'(p^*)(pp^*)^{n-1}\{\min(p,p^*) - pp^*\}\,\mathrm{d}p^*\,\mathrm{d}p,\end{aligned}$$

which equals

$$\frac{n}{(n-1)^2}\int_0^{\bar b}\int_0^{\bar b} G^{n-1}(b)G^{n-1}(b^*)[G\{\min(b,b^*)\} - G(b)G(b^*)]\,\mathrm{d}b\,\mathrm{d}b^*, \tag{28}$$

where we provide (28) if readers would like to compare it to a future GPV–based estimator.

5.4 Mean of the value distribution

We next consider the mean valuation as an example of a functional of the value distribution that one can estimate at the typical parametric rate of convergence. The mean valuation might be of direct interest to the researcher or used to compute test statistics as in Haile et al. (2003).

The mean of the value distribution of bidder one is

$$\mathrm{MV} = \int_0^{\bar v} v f_v(v)\,\mathrm{d}v = \int_0^1 \alpha(p) f_p(p)\,\mathrm{d}p \tag{29}$$

Theorem 7. *Under the assumptions of theorem 1,* $\sqrt{T}\int_0^1\{\alpha_T(p) - \alpha(p)\}\,\mathrm{d}F_p(p) \xrightarrow{d} N(0,\mathscr{V}_{\mathrm{MV}}^{\mathrm{symm}})$, *where* $\mathscr{V}_{\mathrm{MV}}^{\mathrm{symm}} = (n-2)^2(n-1)^{-4}\int_0^1\int_0^1 H(p,p^*)(pp^*)^{(3-2n)/(n-1)}\,\mathrm{d}p^*\,\mathrm{d}p$, *which under (8) simplifies to*

$$\frac{(n-2)^2}{(n-1)^2 n}\int_0^1\int_0^1 Q'(p)Q'(p^*)\{\min(p,p^*) - pp^*\}\,\mathrm{d}p^*\,\mathrm{d}p,$$

which can alternatively be expressed as $[(n-2)^2/\{(n-1)^2 n\}]\int_0^{\bar b}\int_0^{\bar b}[G\{\min(b,b^*)\} - G(b)G(b^*)]\,\mathrm{d}b^*\,\mathrm{d}b.$ $\square$

Note that the asymptotic variance in theorem 7 equals zero if $n = 2$. This is intuitive since then the mean of the value distribution is simply $\bar{b}$, which can be estimated super–consistently. Naturally, this property evaporates when we examine the asymmetric case in section 5.6.4.

The symmetry assumption is also easy to exploit without using our machinery, because $G_c(b)/g_c(b) = G(b)/\{(n-1)g(b)\}$ and

$$\mathrm{MV} = \int_0^{\bar{b}} b \, \mathrm{d}G(b) + \int_0^{\bar{b}} \frac{G(b)}{n-1} \, \mathrm{d}b = \frac{\bar{b} + (n-2)\mathbb{E}b}{n-1}.$$

Since the upper bound of the bid distribution can be estimated at a rate faster than $\sqrt{T}$, estimation of MV by replacing $\bar{b}$ and $\mathbb{E}b$ with their sample counterparts would work, also. So for the purpose of estimating the mean of the value distribution in the symmetric case, our methodology is probably overkill because estimating the inverse strategy function is not a necessary intermediate step.

5.5 Profit

Estimating the seller's profit is a trivial exercise (if the seller's valuation is zero as we assume throughout) since profit is simply the sum of the winning bids. A more interesting object is profit as a function of a hypothetical reserve price r, $\mathrm{PR}(r)$, or number of bidders $\mathrm{PR}^*(n)$.

In the asymmetric case, this is a complicated endeavor. Indeed, using the machinery described earlier in the paper we can recover the value distributions for each bidder. However, there is generally no analytical solution for the bid function in the asymmetric case like there is in the symmetric case. We must therefore numerically solve for the counterfactual equilibrium bid distributions in order to compute the counterfactual revenue. Because this method does not depart from the existing literature, we limit our discussion to the symmetric case, where we do have suggestions for how to exploit the symmetry assumption in the counterfactual analysis. Since the theoretical results here are similar to those obtained earlier in terms of method of proof, we state the results in the text instead of stating them formally.

5.5.1 Counterfactual number of bidders

We have repeatedly made use of the fact that F_p is a known function of n, the number of bidders. F_p is still a known function of any counterfactual number of bidders m, possibly different from n. In addition, the counterfactual expected payment function is a known function of the factual expected payment function if the distribution of valuations is held fixed. Specifically, the equilibrium α for a given number of bidders n satisfies $\alpha(\tau^{n-1}; n) = Q_v(\tau)$ for all n, τ, which implies that for $\xi = \xi_{mn} = (n-1)/(m-1)$, $\alpha(p; m) = \alpha(p^{\xi}; n)$

for all $p \in [0,1]$ and $n, m \geq 2$.[37]

The counterfactual expected payment function is then $e(p;m) = \int_0^p \alpha(t^\xi)\,\mathrm{d}t$ and the expected revenue is given by

$$\mathrm{PR}^*(m) = m\int_0^1 e(p;m)\,\mathrm{d}F_p(p;m) = m\int_0^1\int_0^p \alpha(t^\xi)\,\mathrm{d}t\,\mathrm{d}F_p(p;m) = m\int_0^1 \alpha(p^\xi)\{1 - F_p(p;m)\}\,\mathrm{d}p = \\ m\int_0^1\Big(\frac{\chi_2+1}{\xi}p^{\chi_2} - \frac{\chi_1+1}{\xi}p^{\chi_1}\Big)e(p)\,\mathrm{d}p,$$

where $\chi_j = (m-2n+j)/(n-1)$.

Following the previous examples, a plug–in estimator for $\mathrm{PR}^*(m)$ in which we substitute an estimate of e and the known $F_p(\cdot;m)$ converges at a $\sqrt{T}$–rate. Indeed, the limit is a mean–zero normal distribution with variance

$$m^2\int_0^1\int_0^1 H(p,p^*)\Big(\frac{\chi_2+1}{\xi}p^{\chi_2} - \frac{\chi_1+1}{\xi}p^{\chi_1}\Big)\Big(\frac{\chi_2+1}{\xi}p^{*\chi_2} - \frac{\chi_1+1}{\xi}p^{*\chi_1}\Big)\,\mathrm{d}p^*\,\mathrm{d}p.$$

5.5.2 Counterfactual reserve prices

By (1) in Jun and Pinkse (2019), we have $\mathrm{PR}(r) = \bar{v} - rF_v^n(r) + \int_r^{\bar{v}}\big\{F_v^n(v) - nF_v^{n-1}(v)\big\}\,\mathrm{d}v$. Using the substitution $p = F_v(v)$ the problem then entails finding $p^* = F_v^{n-1}(r)$ for which

$$\mathrm{PR}\{\alpha(p^*)\} = n\bigg(p^*\alpha(p^*)\big(1 - p^{*1/(n-1)}\big) + \frac{1}{n-1}\int_{p^*}^1\int_{p^*}^p \alpha(u)\,\mathrm{d}u\,p^{\frac{2-n}{n-1}}\,\mathrm{d}p\bigg) \tag{30}$$

It should be apparent from our earlier discussion that since α is estimated at a slower–than–parametric rate, the first right hand side term in (30) is estimated at a rate less than $\sqrt{T}$ but the second right hand side term in (30) can be estimated at the typical parametric rate. The asymptotic distribution is hence determined by the estimation of $F_v(r)$, which was already discussed in section 5.1. The choice of bandwidth should therefore be made with an eye toward the precise value or range of values of counterfactual reserve prices under consideration.

To conduct inference on the optimal reserve price would require that we derive the limiting process of $\sqrt[3]{T}(\alpha_T - \alpha)$. Although we establish $\sqrt[3]{T}$ uniform convergence in theorem 1, we do not pursue such a result here. Instead, since we know the convergence rate of the estimator of the optimal resrve price to be $\sqrt[3]{T}$, one could use subsampling. Also note that our assumptions on the value distribution are not sufficient to guarantee a unique globally optimal reserve price. One can proceed by restricting the value distribution or by focusing on the minimal solution when multiple exist as in Jun and Pinkse (2019).

[37] Note that the bid distribution changes with n but the value distribution remains constant. Indeed, $Q_v(\tau) = Q(\tau;n) + \tau Q'(\tau;n)/(n-1)$ defines $Q(\cdot;n)$ as an implicit function of Q_v, indeed $Q(\tau;n) = \int_0^1 Q_v(t^{1/(n-1)}\tau)\,\mathrm{d}t$.

5.6 Asymmetry

When bidders are asymmetric, we must substitute (7) or (9) for (8). Moreover, when the object of interest is $\sqrt{T}$–consistent, we must also consider the contribution of the estimation error in the win–probability distribution. In this section, we first discuss an estimate of the win–probability distribution then provide a set of results analogous to the earlier theorems proved in section 5.

5.6.1 An estimate of the win–probability distribution

To estimate F_p, we note that, in a high–bid auction with bidders whose valuation distributions are not identically distributed, the equilibrium distribution of win–probabilities for bidder is $F_p(p) = G\{Q_c(p)\}$.[38] The distribution F_p can then be estimated in a straightforward fashion as $F_{pT}(p) = G_T\{Q_{cT}(p)\}$, where G_T and Q_{cT} are the empirical distribution of bidder 1's bid and an estimate of the quantile function of its highest competing bid. The weak convergence of this process on (0, 1) is closely related to the extensively studied "copula process" and the fact that the marginal bid densities are strictly positive on their compact support.

Theorem 8. $\sqrt{T}(F_{pT} - F_p) \rightsquigarrow \mathbb{G}_p$*, where* $\mathbb{G}_p$ *is a Gaussian process with covariance kernel* $F_p\{\min(p, p^*)\} - F_p(p)F_p(p^*) + f_p(p)f_p(p^*)H^*\{Q_c(p), Q_c(p^*)\}$. □

Recall that $H^*\{Q_c(p), Q_c(p^*)\}$ can be as simple as $\min(p, p^*) - pp^*$ in case only the maximum competitor bid is used: see (7).

5.6.2 Value Distribution

Because the valuation distribution and quantiles converge more slowly than F_{pT}, the asymptotic distribution of these estimators will be the same as in the symmetric case if we substitute F_{pT} for F_p.

The density estimator merits further discussion for the sake of comparison with other methods and because there are two ways to estimate the value density: one–step and two–step. For the one–step estimator, one substitutes an estimate of the win–probability density evaluated at $\alpha^{-1}(v)$ for f_p. By the delta method, the asymptotic distribution of this value density estimator is again $-(f_p/\alpha'^2)\{\alpha^{-1}(v)\}$ times the asymptotic distribution of the estimator for α'. We caution that f_p may be unbounded near $p = 0$, which could lead to poor finite–sample performance at small values of v.

With the two–step estimator, one first generates valuation estimates by doing e.g. $\hat{v}_t = \bar{\alpha}_{T\psi}\{\hat{G}_{cT}(b_{t1})\}$ and then plugs those estimates into a nonparametric kernel density estimator. The two–step estimator is analogous to GPV except that our first step is different. Since the second step of both estimators is the same, we only need to compare the first step of the two estimators. It can be shown[39] that both the first step in GPV and

[38] A high–bid auction is one in which the highest bidder wins with probability one.

[39] Derivation not provided here.

our smoothed estimates of α (first step of the two–step version of our estimator) permit asymptotic linear expansions of the estimator minus its expectation at $b = Q_c(p)$,

$$\begin{cases} -\left\{ \frac{1}{Th} \sum_{t=1}^{T} \frac{G_c(b)}{g_c^2(b)} k\left(\frac{b_{ct} - b}{h} \right) - \text{its expectation} \right\} & \text{(GPV)}, \\ -\left\{ \frac{1}{Th} \sum_{t=1}^{T} \frac{G_c(b)}{g_c(b)} k\left(\frac{G_c(b_{ct}) - p}{h} \right) - \text{its expectation} \right\} & \text{(ours)}, \\ -\left\{ \frac{1}{Th} \sum_{t=1}^{T} \psi'\{G_c(b)\} \frac{G_c(b)}{g_c(b)} k\left(\frac{\psi\{G_c(b_{ct})\} - \psi(p)}{h} \right) - \text{its expectation} \right\} & \text{(ours with } \psi), \end{cases} \tag{31}$$

The first two formulas in (31) are similar, but note the different arguments to the kernel and the fact that one denominator has a square on g_c and the other one does not. The formula with ψ simplifies to the one without for $\psi(p) = p$ and to the GPV expansion for $\psi(p) = Q_c(p)$. However, the bias of our estimator with $\psi = Q_c$ does not coincide with that for the first step GPV bias: either can be greater.

5.6.3 Bidder Surplus

A natural generic estimator of BS in the absence of a symmetry assumption is

$$\widehat{\text{BS}} = \int_0^1 \{\alpha_T(p)p - \check{e}_T(p)\} \, \mathrm{d}F_{pT}(p). \tag{32}$$

Naturally, α_T can be replaced with a smoothed version in which case it would be advisable to replace $\check{e}_T$ with the estimator of e corresponding to the smoothed estimate of α, also.

Theorem 9. *Under the assumptions of theorem 1, if G, G_c are estimated using different data and G_T is the empirical distribution function of bids of bidder one then $\sqrt{T}(\widehat{\text{BS}} - \text{BS}) \xrightarrow{d} N(0, \mathscr{V}_{\text{BS}}^a)$, where*

$$\mathscr{V}_{\text{BS}}^a = \int_0^1 \int_0^1 \Big[\Gamma_1(p)\Gamma_1(p^*) H^*\{Q_c(p), Q_c(p^*)\} + \Gamma_2(p)\Gamma_2(p^*) H_1^*\{Q_c(p), Q_c(p^*)\} \Big] \, \mathrm{d}p^* \, \mathrm{d}p, \tag{33}$$

with $H_1^*(q, q^*) = G\{\min(q, q^*)\} - G(q)G(q^*)$, $\Gamma_2(p) = \alpha'(p)p$, and $\Gamma_1(p) = Q_c''(p)p^2 f_p(p) + Q_c'(p)\{p^2 f_p'(p) + 4pf_p(p)\}$. Under (7),[40] the asymptotic variance becomes

$$\mathbb{V} \frac{G_c^2(b)}{g_c(b)} + \mathbb{V} \left(\frac{G_c^2(b_c) g(b_c)}{g_c^2(b_c)} + 2 \int_0^{b_c} \frac{G_c(t) g(t)}{g_c(t)} \, \mathrm{d}t \right), \tag{34}$$

which is the semiparametric efficiency bound for estimators of BS *which only use bids and maximum rival bids for estimation.* □

[40] I.e. if one uses an estimator based on the maximum rival bid.

The asymptotic variance in (33) is intimidating but it simplifies considerably in an important special case, as (34) illustrates. The fact that our estimator achieves the semiparametric efficiency bound should come as no surprise since our estimator is asymptotically linear and imposing shape restrictions is well–known not to help in reducing the asymptotic variance in many cases.[41]

Note that the semiparametric efficiency bound is defined only relative to the amount of information available. For instance, if one uses all bids instead of only the maximum rival bid then (33) is less than (34) but still achieves the semiparametric efficiency bound. We do not show this. Nevertheless, it is reassuring that no other regular estimator exists with a smaller asymptotic variance under the same assumptions.

It is not immediately obvious that the variance in theorem 9 is worse than that in theorem 6, albeit that the fact that a more efficient estimate of G_c can be used should tip the balance. We provide a comparison in the least favorable case for symmetry, namely that of two bidders.[42]

Example 4. *Suppose that $n = 2$ bidders are symmetric and $F_v(v) = v^{1/\gamma}$ for some $\gamma > 0$. Then $\bar{b} = 1/(1+\gamma)$, $G(b) = G_c(b) = \{(1+\gamma)b\}^{1/\gamma}$, $g\{Q_c(p)\} = (1+\gamma)p^{1-\gamma}/\gamma$. $g'\{Q_c(p)\} = (1+\gamma)^2(1-\gamma)p^{1-2\gamma}/\gamma^2$, $Q(p) = Q_c(p) = p^\gamma/(1+\gamma)$, $Q' = Q_c' = \gamma p^{\gamma-1}/(1+\gamma)$, $Q'' = Q_c'' = \gamma(\gamma-1)p^{\gamma-2}/(1+\gamma)$, and $e(p) = p^{\gamma+1}/(1+\gamma)$. Thus, from (28) it follows using some tedious calculus that*

$$\mathscr{V}_{\text{BS}}^{\text{symm}} = \frac{2\gamma^2}{(1+\gamma)^2(2+\gamma)^2(3+2\gamma)},$$

which equals $1/90$ for a uniform value distribution. To obtain $\mathscr{V}_{\text{BS}}^a$ note that $\Gamma_2(p) = \gamma p^\gamma$, $\Gamma_1(p) = \gamma(3+\gamma)p^\gamma/(1+\gamma)$, and $H^*\{Q_c(p), Q_c(p^*)\} = H_1\{Q_c(p), Q_c(p^*)\} = \min(p, p^*) - pp^*$ which (after some tedious calculus) yields

$$\mathscr{V}_{\text{BS}}^a = \frac{2\gamma^2(5+4\gamma+\gamma^2)}{(1+\gamma)^2(2+\gamma)^2(3+2\gamma)},$$

which equals $1/9$ in the uniform F_v case. The variance in the asymmetric case is $5+4\gamma+\gamma^2$ times as large as in the symmetric case. Since assuming symmetry speeds up convergence of $\breve{e}_T$ and *obviates the need to estimate F_p, it was clear that the ratio would exceed two. But in the uniform F_v case the factor is ten!*[43] □

More generally, the variance relatively large $\mathscr{V}_{\text{BS}}^a$ is due not only to the estimation error in F_{pT}, which is reflected by the term involving Γ_2 in (33), but also to the fact that Q_c is estimated off the maximum rival bids as opposed to combining estimates of the marginal bid distributions for each rival bidder. Indeed, using (9) with $n > 2$, the variance reduction will be more modest than under (7). Nevertheless, the conclusion from example 4 must be that symmetry should be imposed whenever reasonable.

[41] See Newey (1990, page 106) for a discussion and Tripathi (2000) for results on the semiparametric efficiency bound subject to shape restrictions in the partially linear model of Robinson (1988).

[42] With more than two bidders, the gain in efficiency of estimating G_c is greater.

[43] That's not a factorial.

As it turns out, smoothing does not improve the asymptotic distribution, and too much smoothing can introduce an asymptotic bias and slow down convergence. The most important consideration is that the implied estimator of e converges (after norming and scaling) to the same Gaussian limit process as $\breve{e}_T$ which for the smoothed estimator simply requires that $h \to 0$ fast enough as $T \to \infty$. We state the theorem for $\bar{\alpha}_{T\psi}$ but the result applies with minor modifications to any estimators which satisfy the afore–mentioned desiderata.

Theorem 10. *Suppose that the assumptions of theorem 4 are satisfied. Then the same limit distribution obtains if one replaces $\alpha_T, \breve{e}_T$ in theorem 9 with $\bar{\alpha}_{T\psi}, \bar{e}_{T\psi}$ and chooses a bandwidth h which tends to zero faster than $T^{-1/4}$.* □

Note that the bandwidth can tend to zero arbitrarily fast since a bandwidth of zero simply takes us back to the unsmoothed estimator. This is in sharp contrast to other approaches, e.g. one based on the estimator of the inverse bid function in GPV, where taking the bandwidth to zero *before* taking the sample size to infinity would blow up the asymptotic variance: letting $h \downarrow 0$ with GPV does not produce a consistent estimator of g_c, g, α. Consequently, it is not clear a priori that using a second order kernel and undersmoothing GPV would produce a consistent estimator of BS, let alone a $\sqrt{T}$–consistent estimator. We have no theoretical results on this, though our simulation results suggest that letting the bandwidth go to zero in a GPV–based estimator of BS would break $\sqrt{T}$–consistency.

5.6.4 Mean Valuation

In the asymmetric case, we estimate the mean valuation by $\widehat{\mathrm{MV}} = \int_0^1 \alpha_T(p)\,\mathrm{d}F_{pT}(p)$. The asymptotics for $\widehat{\mathrm{MV}}$ are similar to those for $\widehat{\mathrm{BS}}$ and the proof is therefore mercifully short.

Theorem 11. *Under the assumptions of theorem 1, $\sqrt{T}(\widehat{\mathrm{MV}} - \mathrm{MV}) \overset{d}{\to} N(0, \mathscr{V}^a_{\mathrm{MV}})$, where $\mathscr{V}^a_{\mathrm{MV}}$ is like $\mathscr{V}^a_{\mathrm{BS}}$ with Γ_1, Γ_2 divided by p.* □

6 Monte Carlo simulations

We provide a simulation study to compare the performance of our estimators for a representative set of objects. In the category of $\sqrt{T}$–consistent estimators, we consider the estimation of the bidder's expected surplus and the mean of the valuations. Because the bidder's surplus is $\int \{\alpha(p)p - e(p)\}\,\mathrm{d}F_p(p)$, the behavior of the estimator for α near $p = 1$ is relatively more important than for small values of p (the seller's expected revenue in a symmetric equilibrium with a counterfactual number of bidders is similar in this regard). By contrast, the behavior of α near $p = 0$ plays a more important role in the expected valuation $\int \alpha(p)\,\mathrm{d}F_p(p)$, especially when the bidder is strong relative to its competition, because the win–probability density is then unbounded near zero. In the category of $\sqrt{Th}$–consistent estimators, we examine the behavior of estimators of the valuation distribution function and its quantiles, including the maximum valuation. As a "primitive" of the model, the valuation distribution is an input to many other objects of interest. Hence, these objects derive their asymptotic properties from the estimator of the valuation distribution. For example, the dominant terms in the asymptotic expansion of the seller's counterfactual expected revenue with a higher reserve price depend only on the behavior of the estimator for the valuation distribution at the reserve price. The quantile function, particularly at the endpoints, demonstrates the behavior of the boundary correction routines. And finally, in the category of $\sqrt{Th^3}$–consistent estimators, we consider the estimation of the density of the valuations. As discussed in section 5, the dominant terms in the asymptotic expansion come from the estimator for the derivative of α, but the valuation density also highlights the finite–sample benefit of substituting the true win–probability density f_p when it is known rather than an estimate, especially when $f_p(0)$ is unbounded (i.e. there are more than two symmetric bidders). Our goal is not to crown a winner but to highlight systematic ways in which various methodological choices impact the bias and mean squared error of the estimators.

6.1 Simulation parameters

We parameterize bidder one's maximum competitor bid distribution as $G_c(b) \propto (\theta/b + \gamma - \theta)^{-1/\theta}$ for $b \in [0, \bar{b}]$ with $\bar{b} = 2\big/\big(1 + \theta + \sqrt{4\gamma + (\theta - 1)^2}\big)$ and $\gamma, \theta > 0$. Bidder one's inverse bid function is then $\beta^{-1}(b) = (1 + \theta)b + (\gamma - \theta)b^2$. Note that β^{-1} is strictly increasing on $[0, \bar{b}]$, which implies convexity of the expected payment function $e(p) = \theta c p^{\theta+1}\big/\big\{1 + c(\gamma - \theta)p^{\theta}\big\}$, where c is a constant that depends on γ and θ.

The maximum competitor bid distribution is chosen such that the support of bidder one's valuations is $[0, 1]$ regardless of the values of γ and θ. We can then fix bidder one's valuation distribution and independently vary the maximum competitor bid distribution to achieve various shapes of the inverse strategy function α and competitor bid density, which are the respective targets of estimators based on our approach or estimators based on the inverse bid function (IBF) like GPV. Figures 5 and 6 plot these functions. If $\gamma = \theta$ then the competitor bid distribution is a power distribution and the inverse bid function is simply linear. As θ approaches zero, the competitor bid distribution approaches a truncated Fréchet distribution and the inverse bid function is a convex quadratic.

For every combination of $\gamma = 3/2,\ 3/4,\ 1/3,\ 1/7$ and $\theta = 3/4, 1/3, 1/7, 1/9$, we draw independent and identically distributed samples of $T = 100$, 250, and 500 maximum competitor bids. Thus, T represents the number of auctions as well as the number of bids used to estimate bidder one's expected payment function and inverse strategy. We then independently sample T draws of bidder one's valuations according to a power distribution $F_{v_1}(v) = v^{3/2}$, compute her optimal bid, and apply our various methodologies to estimate several objects of interest.

Similar to our approach to the analysis in the preceding sections, we remain agnostic about the underlying

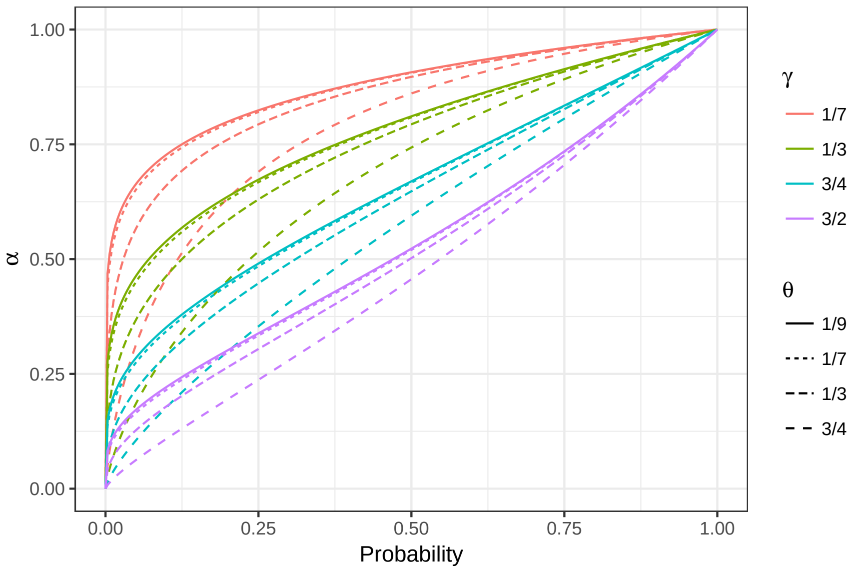

Figure 5: The inverse strategy function for various parameter values.

data generating processes that might have produced the simulated sample of maximum competing bids by directly parameterizing bidder one's competitor(s)'s bid distribution(s). This choice serves two purposes. First, we more directly control the shapes of the immediate targets of the nonparametric estimators than if we fixed the number of bidders and varied their valuation distributions. Second, we do not confound the two distinct ways of leveraging symmetry to improve the estimates: pooling bids to more precisely estimate α and using the true distribution of win–probabilities that must prevail in any symmetric equilibrium of a high–bid auction. We isolate the latter benefit by always using only one bid per auction to estimate α but report some simulation results in which we substitute the true win–probability distribution when estimating the object of interest.[44]

We compare our estimators with an estimator based on an approach similar to GPV in which only the independent sample of highest competitor bids are used to estimate the inverse bid function. Specifically, we combine the empirical distribution and kernel density estimator to estimate the inverse bid function $b + G_c(b)/g_c(b)$. This estimator is labeled "IBF" to indicate the estimates of the various objects were constructed from a nonparametric estimate of the inverse bid function. The IBF estimator does not perform any boundary correction or trimming, hence cannot be expected to perform well near the boundary. To be clear, this is not a critique of GPV: the goal in their paper is to estimate the inverse bid function and valuation density at an optimal rate in the interior of the support of the valuations. Here, we use the estimator for the

[44] We verified that there exists at least one data generating process that could have produced the data. For example, the bids could result from a two-bidder auction in which the other bidder's valuation distribution is $F_v = G_c(\beta_2(v))$, where $\beta_2(v) = 3v/5$ if $\gamma = \theta$ and $\beta_2(v) = \{6v(\gamma - \theta) - 5(\theta + 1) + \sqrt{16(1+\theta)^2 + 9[1 + \theta + 2(\gamma - \theta)]^2}\}/[16(\gamma - \theta)]$ if $\gamma \neq \theta$ is the other bidder's optimal bid function. If $\gamma = \theta = 1/(n-1)$, the sample of highest competing bids also could have been generated by a symmetric n-bidder auction.

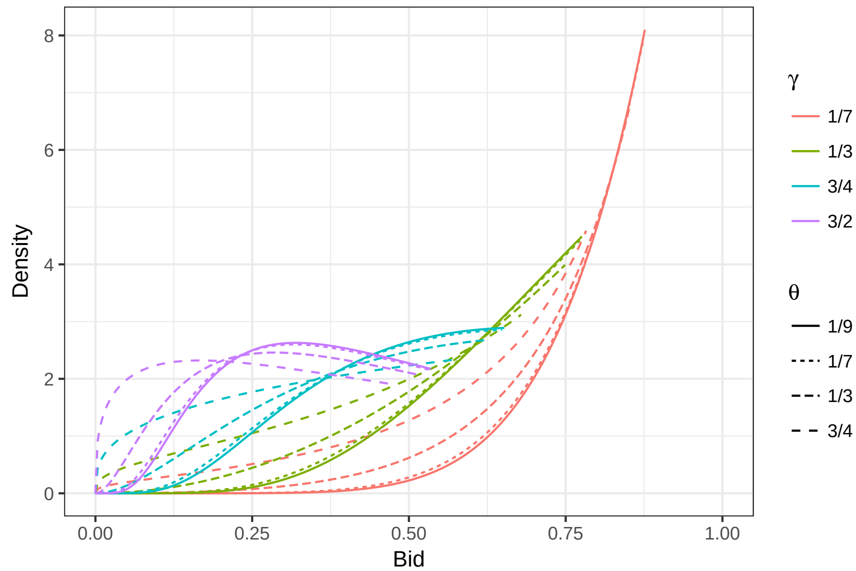

Figure 6: The competitor bid density for various parameter values.

inverse bid function as an input into other objects of interest. A fairer comparison with our boundary–corrected estimators can be found in "IBF–BC." This estimator uses a boundary correction routine similar to HH and is hence better suited for estimating objects that require integration over the entire support of the bids or valuations.

All simulations employ an Epanechnikov kernel and a rule–of–thumb bandwidth sequence, multiplied by an additional scaling factor of $1/5, 1/2, 1$, or $3/2$ in order to explore sensitivity to the choice of bandwidth. We use a Gaussian reference distribution for choosing bandwidths for methods that use nonparametric kernel density estimators and $\alpha(p) = \bar{v}p^{\gamma}$ as our reference function for bandwidths using our procedure. Specifically, we use the sample mean and variance of $\alpha_T\{G_{cT}(b_1)\}$ to estimate the parameters $\bar{v}$ and γ in the parametric reference model, then choose the bandwidth that would minimize the mean integrated squared error of the estimator for α under the reference model. This optimal bandwidth also depends on ψ.[45]

We consider five different choices of ψ. The first is the identity transformation $\psi_1(p) = p$ and the second is the infeasible zero–bias transformation $\psi_2(p) = \alpha(p)$. The next transformation $\psi_3(p) = \log(p)$ ensures that the asymptotic bias is vanishingly small for p close to zero, though the asymptotic variance can be large. The transformation $\psi_4 = \sqrt{p}$ minimizes the MISE in α if α is a power function with exponent greater than $1/2$.[46] Finally, $\psi_5(p) = \sqrt[5]{p}$ balances the integrated asymptotic bias and variance of the estimator for α when α is a power function, regardless of the exponent.

For ψ_2, the rule of thumb suggests that an infinite bandwidth would minimize the integrated MSE because

[45] For some choices of ψ, the squared error is not integrable on $[0, 1]$. In these cases, our rule of thumb minimizes the integrated squared error on $[0.05, 1]$.

[46] If α is a power function and the exponent is less than $1/2$, the optimal ψ would be the infeasible choice ψ_2.

the first–order bias is always zero. A better rule of thumb would suggest a bandwidth sequence on the order of $T^{-1/7}$, resulting in a faster rate of convergence than the other estimators. For the sake of comparison, we do not take this route and instead use the same bandwidth as we do for ψ_5. For the undersmoothed estimates, we simply multiply the rule–of–thumb bandwidths by $T^{-2/15}$ so that the sequences are on the order of $T^{-1/3}$.

For the boundary corrected estimates that use reflection—IBF–BC and $\bar{\alpha}_T^R$—the auxiliary bandwidth is proportional to the main bandwidth. In Pinkse and Schurter (2019), we show that choosing bandwidths converging at a rate of $T^{-1/7}$ would be optimal for both estimators if α (equivalently g_c) has three continuous derivatives near the boundary. In this paper, we do not choose a bandwidth sequence to capitalize on this extra smoothness because doing so would put the reflection methods at an advantage relative to the boundary kernel estimators. Unlike the reflection methods, our boundary kernel method does not involve any auxiliary input parameters that could be modified to take advantage of this extra smoothness. That said, we note that if the researcher is willing to strengthen the smoothness assumption, the reflection method or a boundary kernel method that takes advantage of the extra smoothness could be more attractive in practice precisely for this reason.[47]

All simulations use a thousand replications.

6.2 Simulation results: $\sqrt{T}$–consistent estimators

We first review the simulation results for the integrated objects MV and BS. Figure 7 illustrates the relative root mean squared error (RMSE) of our unsmoothed and smoothed, boundary–kernel–based estimators along with the IBF estimators. The bandwidths are chosen proportional to $T^{-1/3}$ so that the resulting $\sqrt{T}$–consistent estimator is asymptotically unbiased. For lack of a better rule, the constant of proportionality in the bandwidth sequence is simply the rule–of–thumb constant multiplied by our additional scale factor. The various estimators are arranged in columns, and each row represents a different combination of the target object, bandwidth scaling factor, γ, θ, and T. The value in each cell is colored to reflect the value of the RMSE divided by the minimum RMSE across the columns. The lightest green indicates the best performing estimator, while the darkest purple indicates the RMSE was at least three times as large as that of the best performing estimator. The color scale is top–coded because some estimators performed extremely poorly.

The unsmoothed MLE consistently performs well across a variety of parameter values, while the unsmoothed isotonic regression estimator (LS) has difficulty for some parameter values because it suffers from finite–sample bias for values of p close to one. Intuitively, this bias arises from the fact that the GCM, by definition, must lie below the estimate of the true expected payment function, which leads to an upward bias in its slope near $p = 1$. This bias is more pronounced when γ and θ are both relatively large, because the

[47] In fact, if the target of the estimation were the valuation density, the researcher might assume three derivatives of g_c, anyway, in order to attain the typical $T^{2/7}$ rate of convergence.

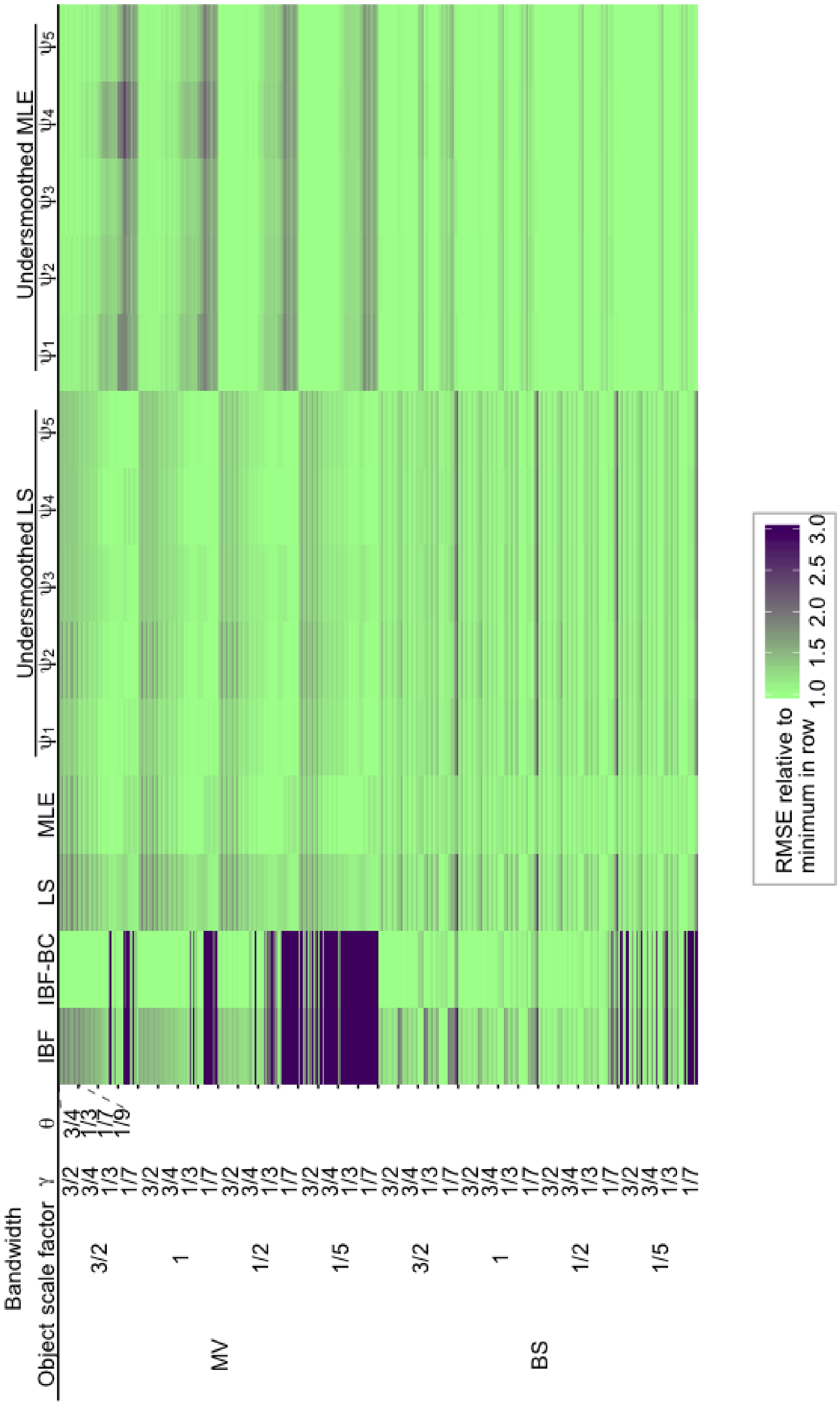

Figure 7: Relative RMSE of $\sqrt{T}$–consistent estimators, BS and MV. T is the most frequently repeating parameter (in decreasing order) along the vertical axis. The first row reflects estimates of the mean valuation when the rule-of-thumb bandwidths are scaled by 3/2, $\gamma = 3/2$, $\theta = 3/4$, and $T = 500$.

Table 1: RMSE of estimators for the bidder's expected surplus using $\psi_5(p) = p^{1/5}$ and $T = 500$ auctions relative to the minimum RMSE across all combinations of estimators and bandwidth scaling factors. Relative values are multiplied by 1000.

	Bandwidth scaling factor	(γ, θ)							
		(1/7, 1/9)	(1/7, 1/7)	(1/7, 1/3)	(1/7, 3/4)	(1/3, 1/9)	(1/3, 1/7)	(1/3, 1/3)	(1/3, 3/4)
IBF–BC	0.2	1671	3596	1741	1000	1153	1103	1848	1008
IBF–BC	0.5	1306	1510	1026	1037	1036	1041	1000	1067
IBF–BC	1	1244	1050	1041	1050	1025	1026	1008	1084
IBF–BC	1.5	1212	1000	1049	1059	1020	1020	1018	1093
LS	0	1465	1262	1000	1010	1104	1115	1326	1000
MLE	0	1148	1070	1059	1059	1041	1044	1094	1064
Sm. LS	0.2	1434	1201	1008	1014	1093	1103	1264	1007
Sm. LS	0.5	1399	1177	1016	1018	1085	1095	1208	1014
Sm. LS	1	1386	1128	1024	1024	1088	1097	1237	1019
Sm. LS	1.5	1369	1107	1029	1032	1088	1097	1229	1022
Sm. MLE	0.2	1053	1128	1078	1074	1008	1007	1044	1097
Sm. MLE	0.5	1022	1167	1086	1078	1000	1000	1035	1103
Sm. MLE	1	1014	1184	1093	1084	1005	1002	1074	1107
Sm. MLE	1.5	1000	1209	1097	1091	1005	1003	1071	1109
Min. value		$5.43 \cdot 10^{-3}$	$2.1 \cdot 10^{-3}$	$4.57 \cdot 10^{-2}$	$9.25 \cdot 10^{-2}$	$4.35 \cdot 10^{-2}$	$4 \cdot 10^{-2}$	$6.12 \cdot 10^{-3}$	$7.44 \cdot 10^{-2}$
		(3/4, 1/9)	(3/4, 1/7)	(3/4, 1/3)	(3/4, 3/4)	(3/2, 1/9)	(3/2, 1/7)	(3/2, 1/3)	(3/2, 3/4)
IBF–BC	0.2	1075	1062	1083	1316	1074	1074	1080	1140
IBF–BC	0.5	1014	1016	1021	1056	1021	1021	1024	1045
IBF–BC	1	1003	1003	1004	1006	1005	1005	1005	1009
IBF–BC	1.5	1000	1000	1000	1000	1000	1000	1000	1000
LS	0	1075	1077	1104	1382	1086	1087	1094	1150
MLE	0	1042	1043	1059	1208	1063	1064	1069	1112
Sm. LS	0.2	1069	1070	1096	1327	1079	1080	1087	1140
Sm. LS	0.5	1064	1066	1091	1308	1077	1076	1084	1137
Sm. LS	1	1066	1067	1090	1299	1077	1076	1082	1131
Sm. LS	1.5	1067	1069	1092	1311	1077	1077	1083	1132
Sm. MLE	0.2	1012	1012	1017	1061	1028	1028	1031	1051
Sm. MLE	0.5	1009	1010	1014	1064	1026	1025	1029	1049
Sm. MLE	1	1012	1012	1015	1074	1028	1027	1029	1047
Sm. MLE	1.5	1014	1014	1017	1086	1029	1028	1031	1049
Min. value		$1.07 \cdot 10^{-1}$	$1.07 \cdot 10^{-1}$	$8.88 \cdot 10^{-2}$	$1.45 \cdot 10^{-2}$	$1.77 \cdot 10^{-1}$	$1.78 \cdot 10^{-1}$	$1.76 \cdot 10^{-1}$	$1.28 \cdot 10^{-1}$

true expected payment is more convex near the right boundary. In contrast, the unsmoothed MLE is not as badly biased when γ and θ are large, because the graph of the MLE for e does not have to lie below the unconstrained estimator. The finite sample bias in estimates of α for large values of p more negatively affects the relative performance in estimating the bidder's expected surplus because the values of $\alpha(p)$ for large p are weighted relatively more in the integral formula for BS than for MV. We expect these differences in the unsmoothed estimators to vanish as T increases because all our estimators in figure 7 are asymptotically equivalent.

When γ is small, the undersmoothed, boundary–corrected IBF estimator for BS and MV appears to under–perform in small samples, but otherwise has a relatively small RMSE. As expected, however, the relative performance of the IBF approach is sensitive to the scale of the bandwidth sequence. We cannot conclude from figure 7 that our approach is robust to the choice of bandwidth, however, because it does not compare the relative performance of different bandwidth scaling factors. Table 1 makes this comparison in the estimation of the bidder's surplus using $T = 500$ auctions. The results demonstrate that the asymptotic behavior of our undersmoothed estimators is fairly similar across bandwidth sequences, whereas the IBF–BC estimator can be the best performing for some parameter values or worst performing estimator depending on the choice of bandwidth. We consider this robust performance, and indeed not having to choose an input parameter, a valuable characteristic of our approach.

The IBF–BC estimator for BS performs particularly well when γ equals θ, in which case the maximum competitor bid is a power distribution and the inverse bid function is linear. Even without undersmoothing, the asymptotic bias in the bidder's expected surplus would be zero when $\gamma = \theta = 1$ or $\gamma = \theta = 1/2$ and is fairly small at intermediate values. This fact helps explain why the RMSE for BS using LS and MLE relative to IBF–BC is larger, for example, in the $(1/3, 1/3)$ column compared to the columns on either side.

Table 1 represents less than 6% of the information contained in figure 7. Many more tables (available online here) provide further quantitative comparisons of the RMSE, as well as the bias of these and other estimators discussed below.

6.3 Simulation results: distribution and quantile functions

The next set of results compare the root mean integrated squared error (RMISE) of the value distribution and quantile function. Based upon the results in the first two columns, the IBF approach without boundary correction is (unsurprisingly) dominated by the boundary–corrected estimator. Next, comparing the second column with the columns to the right, we find that our estimators are again more robust to the choice of bandwidth and tend to outperform the boundary–corrected IBF approach when γ and θ are small, i.e. the highest competing bid is stronger. Echoing our earlier observations, we again see that the smoothed MLE has a smaller RMSE than the smoothed LS estimator when γ and θ are larger so that α exhibits less curvature near at the left boundary and is steeper at the right boundary.

Although all bandwidth sequences are proportional to $T^{-1/5}$, the finite–sample behavior of our smoothed estimators is less (negatively) impacted by a small bandwidth. This finding is related to the fact that our estimator for α is consistent even when the bandwidth tends to zero. Again, robustness is a virtue.

Comparing the rows for which either γ or θ is small, we also find that our transformation method significantly reduces the RMISE in the estimate of the quantile function. In particular, the IBF–BC and the "no transformation" (ψ_1) estimators have greater RMISE compared to the estimators that employ the transformations ψ_3, ψ_4, or ψ_5. Looking across all columns, the smoothed least–squares estimator in conjunction with the transformations ψ_3 or ψ_5 appears to consistently perform best or near the best when the highest competing bid is relatively stronger (γ and θ are small), while the the IBF–BC estimator and the smoothed MLE estimator with ψ_5 or ψ_3 perform better in terms of RMISE when the highest competing bid is relatively weak. We note that in auctions with three or more bidders one would expect the highest competing bid to be relatively strong and hence the smoothed least squares estimator to outperform the alternatives.

The differences between the estimators of the value distribution are less striking. This is due in part to the fact that some of the differences in the estimates of the quantile function are driven by the behavior near the left boundary. Using the delta method, one can show that the asymptotic distribution of the estimators for F_v are scaled by the value density at v. Under our simulation design, f_v approaches zero for small v. The relative differences in the estimators are dampened as a result. This also explains the fact that the log–transformation ψ_3 tends to be the best in terms of RMISE(F_v) for small values of γ and θ. Under these parameters, α'' diverges as p approaches zero, which produces a large bias for small values of p absent a transformation. The log–transformation ensures the asymptotic bias vanishes at the low end, while the fact that $f_v(v)$ is small mitigates the detrimental effects on the asymptotic variance. The transformation ψ_5 yields similar results, though it does not reduce the bias and increase the variance by as much.

Zooming in on a subset of these simulation results in tables 2 and 3, we see that our estimator using the ψ_5 transformation does a good job of estimating the quantiles compared with the IBF–BC estimator. For rows using the rule–of–thumb bandwidths, i.e. when the bandwidth scaling factor is one, there are no columns in which the IBF–BC estimator outperforms both the smoothed LS and smoothed ML estimators, and there are several columns in which it does much worse. In contrast, the IBF–BC estimator for the value distribution function is always nearly the best if not the best.

6.4 Simulation results: density function

We now compare the various methods of estimation of the density of bidder one's valuations. For the indirect methods, a bandwidth proportional to $T^{-1/5}$ is used in both the first and second steps. For the direct method, α' is estimated using a bandwidth proportional to $T^{-1/7}$. We compare the direct method using an estimate of f_p as well as the true f_p, which the econometrician would know under the assumption the data were generated in a symmetric equilibrium. The boundary-corrected kernel density estimate f_p is obtained from the sample of $p_t = G_{cT}(b_{1t})$ using a bandwidth proportional to $T^{-1/5}$. Note, however, that the true density f_p is unbounded near $p = 0$ in a symmetric equilibrium with more than two bidders, which may result in poor performance of the density estimate for small values of $p > 0$. Thus, even though the pointwise rate of convergence of our estimate of f_p is faster than the rate of convergence of our estimate of α', we would expect this estimator to perform poorly in finite samples. Indeed, the simulation results indicate that the direct method combined with the true f_p compares favorably with the indirect estimates, but the direct method combined with an estimate of f_p can be relatively poor. In such cases, better results might be achieved by estimating the density of an

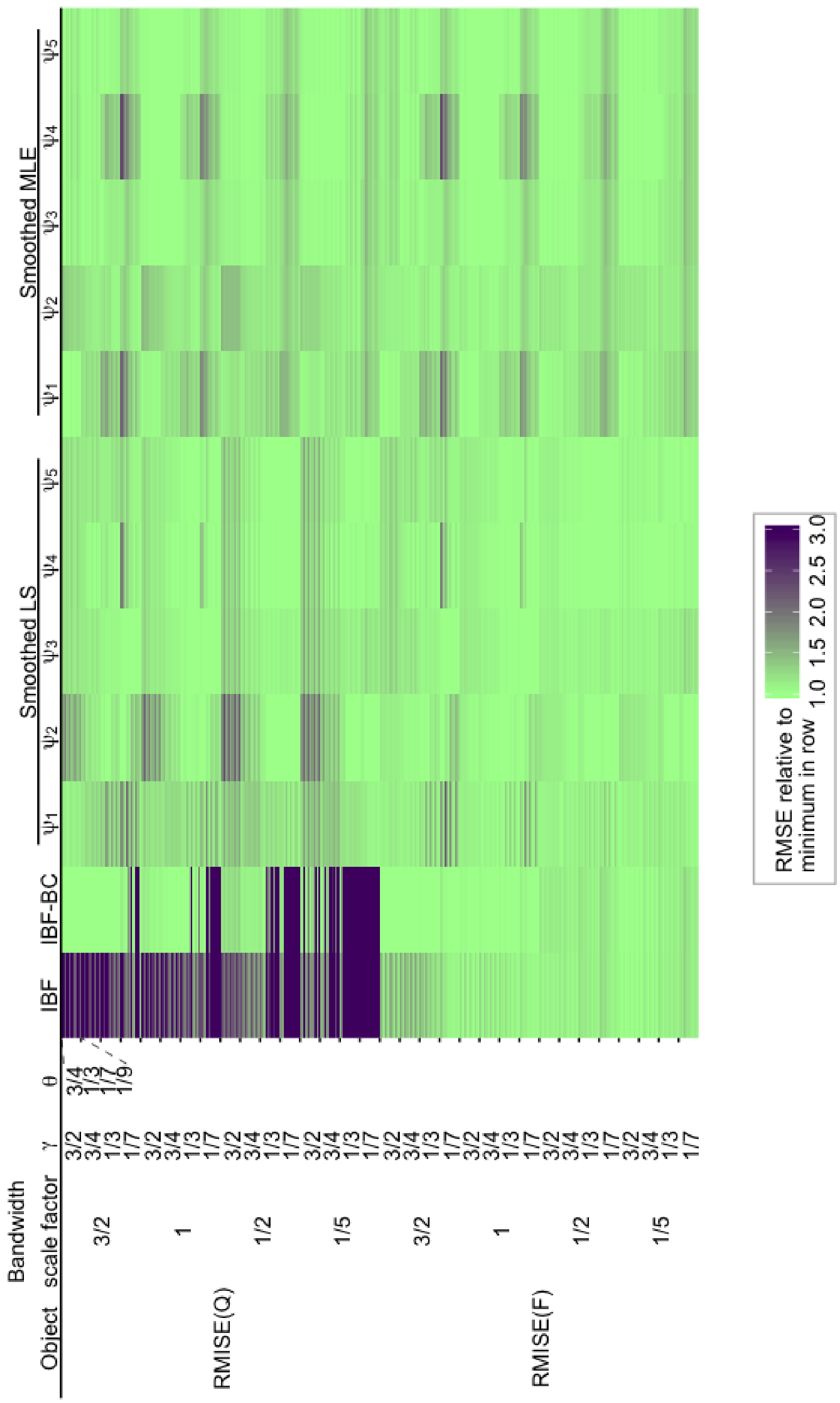

Figure 8: Relative RMISE of estimators of value distribution and quantile function. T is the most frequently repeating parameter (in decreasing order) along the vertical axis. The first row reflects estimates of the mean valuation when the rule-of-thumb bandwidths are scaled by 3/2, $\gamma = 3/2$, $\theta = 3/4$, and $T = 500$.

Table 2: RMISE of estimators for F_v using $\psi_5(p) = p^{1/5}$ and $T = 500$ auctions relative to the minimum RMISE across all combinations of estimators and bandwidth scaling factors. Relative values are multiplied by 1000.

	Bandwidth scaling factor	(γ,θ)							
		$(1/7,1/9)$	$(1/7,1/7)$	$(1/7,1/3)$	$(1/7,3/4)$	$(1/3,1/9)$	$(1/3,1/7)$	$(1/3,1/3)$	$(1/3,3/4)$
IBF-BC	0.2	1211	1238	1360	1345	1214	1224	1298	1404
IBF-BC	0.5	1135	1153	1235	1268	1127	1136	1195	1303
IBF-BC	1	1072	1084	1098	1076	1025	1027	1061	1111
IBF-BC	1.5	1043	1050	1067	1083	1001	1000	1020	1044
Sm. LS	0.2	1094	1094	1036	1214	1085	1095	1183	1338
Sm. LS	0.5	1000	1000	1000	1000	1000	1000	1032	1120
Sm. LS	1	1051	1069	1260	1019	1081	1063	1036	1064
Sm. LS	1.5	1120	1145	1368	1202	1262	1225	1131	1111
Sm. MLE	0.2	1311	1345	1415	1264	1173	1185	1236	1325
Sm. MLE	0.5	1291	1319	1387	1148	1109	1114	1125	1141
Sm. MLE	1	1250	1282	1369	1168	1048	1049	1050	1044
Sm. MLE	1.5	1222	1256	1352	1242	1085	1060	1000	1000
Min. value		$2.56 \cdot 10^{-2}$	$2.61 \cdot 10^{-2}$	$2.61 \cdot 10^{-2}$	$2.55 \cdot 10^{-2}$	$2.27 \cdot 10^{-2}$	$2.29 \cdot 10^{-2}$	$2.3 \cdot 10^{-2}$	$2.45 \cdot 10^{-2}$
		$(3/4,1/9)$	$(3/4,1/7)$	$(3/4,1/3)$	$(3/4,3/4)$	$(3/2,1/9)$	$(3/2,1/7)$	$(3/2,1/3)$	$(3/2,3/4)$
IBF-BC	0.2	1245	1248	1269	1297	1249	1252	1255	1253
IBF-BC	0.5	1183	1193	1206	1261	1236	1249	1261	1284
IBF-BC	1	1046	1047	1054	1076	1079	1086	1089	1105
IBF-BC	1.5	1000	1000	1000	1000	1000	1000	1000	1000
Sm. LS	0.2	1184	1189	1222	1289	1245	1250	1265	1295
Sm. LS	0.5	1101	1107	1108	1145	1183	1190	1201	1209
Sm. LS	1	1147	1143	1127	1123	1227	1232	1222	1212
Sm. LS	1.5	1284	1272	1217	1170	1341	1340	1300	1254
Sm. MLE	0.2	1163	1168	1197	1257	1201	1204	1220	1253
Sm. MLE	0.5	1057	1063	1056	1095	1093	1101	1113	1128
Sm. MLE	1	1045	1040	1019	1018	1105	1109	1094	1084
Sm. MLE	1.5	1171	1155	1080	1022	1272	1264	1203	1130
Min. value		$2.51 \cdot 10^{-2}$	$2.53 \cdot 10^{-2}$	$2.65 \cdot 10^{-2}$	$2.94 \cdot 10^{-2}$	$3.12 \cdot 10^{-2}$	$3.12 \cdot 10^{-2}$	$3.24 \cdot 10^{-2}$	$3.52 \cdot 10^{-2}$

appropriate transformation of p and using a change of variable formula to recover an estimate of f_p.

6.5 Simulation results: boundary correction methods

Figure 10 illustrates the simulation results for estimates of the inverse strategy function at the right boundary. The reflection–based boundary correction methods tend to perform better when the target of smoothing—α or g_c—is relatively flat and linear near the boundary. In this case, we would expect the error in the estimate of the auxiliary parameter $\hat{d}$ to be relatively small. On the other hand, the boundary kernel method tends to perform better when α' is large near the boundary, which would be more likely to happen if the number of bidders is small.

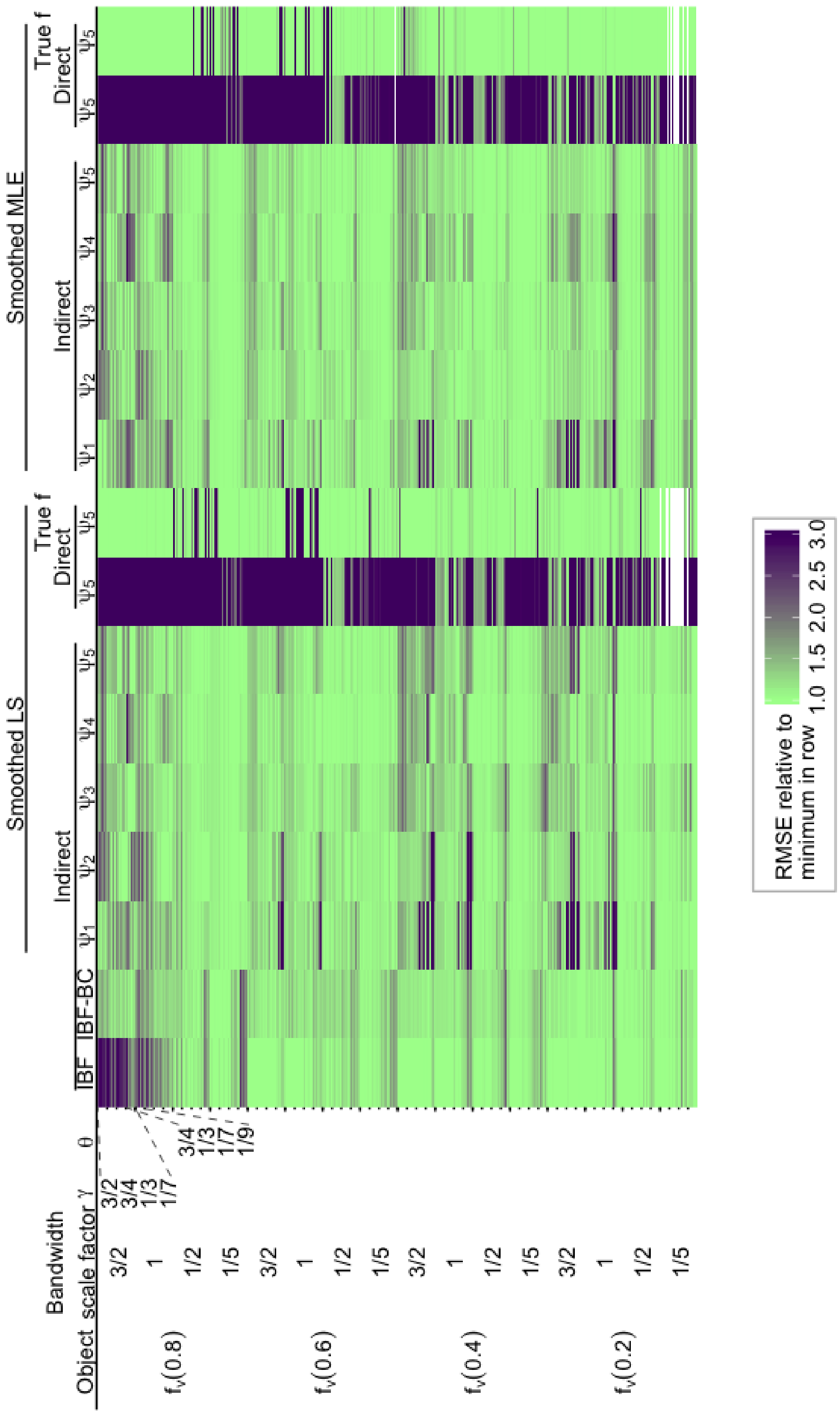

Figure 9: Relative MSE of estimators of value density. T is the most frequently repeating parameter (in decreasing order) along the vertical axis. The first row reflects estimates of the mean valuation when the rule-of-thumb bandwidths are scaled by 3/2, $\gamma = 3/2$, $\theta = 3/4$, and $T = 500$.

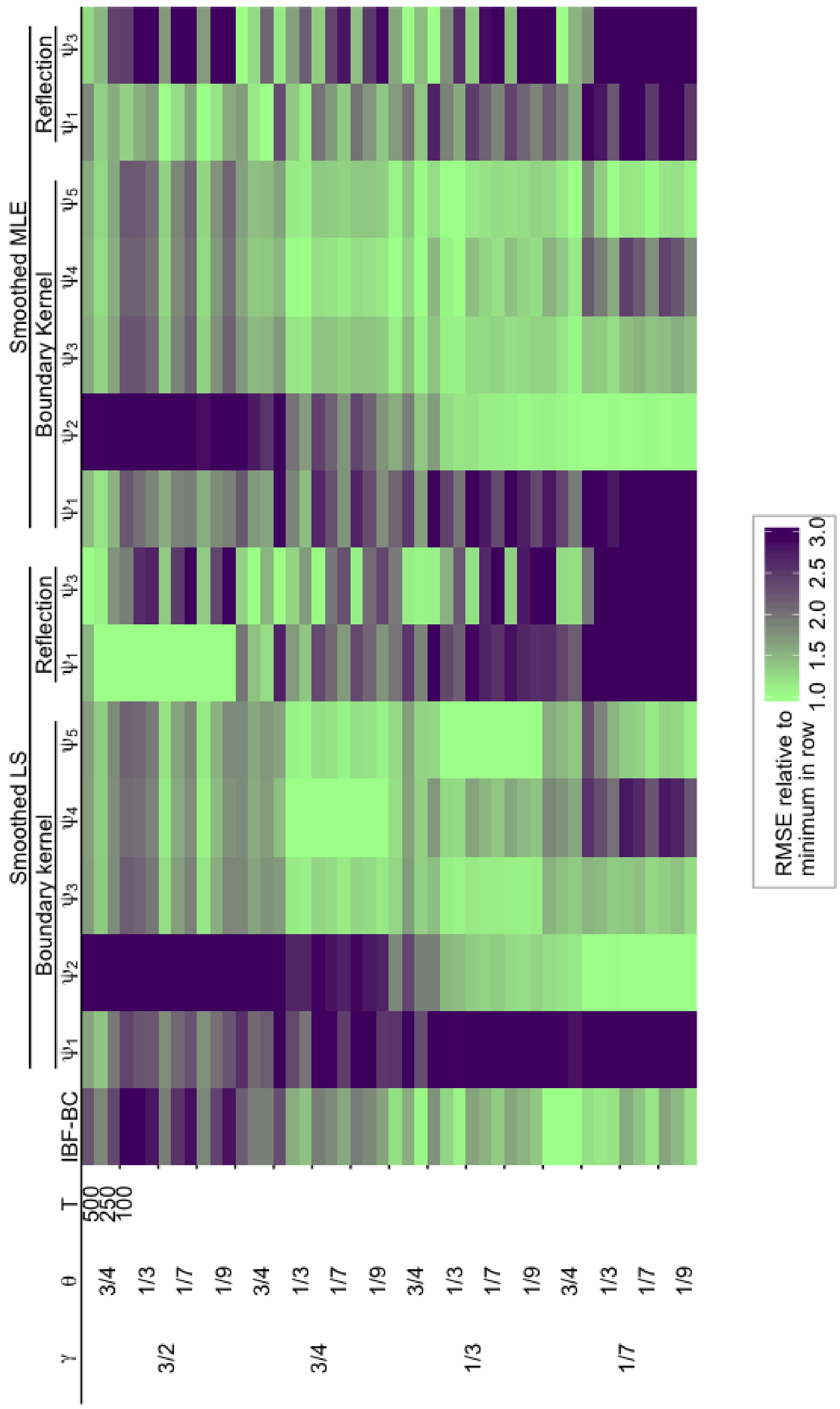

Figure 10: Relative RMSE of the estimator for the maximum valuation using boundary kernel and generalized reflection methods.

Table 3: RMISE of estimators for Q_v using $\psi_5(p) = p^{1/5}$ and $T = 500$ auctions relative to the minimum RMISE across all combinations of estimators and bandwidth scaling factors. Relative values are multiplied by 1000.

	Bandwidth scaling factor	(γ, θ)							
		(1/7, 1/9)	(1/7, 1/7)	(1/7, 1/3)	(1/7, 3/4)	(1/3, 1/9)	(1/3, 1/7)	(1/3, 1/3)	(1/3, 3/4)
IBF-BC	0.2	40155	219413	783522	123300	52007	56878	675179	1432
IBF-BC	0.5	85850	15733	61140	1227	3343	5603	2571	1288
IBF-BC	1	5930	166249	7220	1050	1233	8974	1036	1098
IBF-BC	1.5	12365	28752	1046	1064	1004	1001	1002	1032
Sm. LS	0.2	1078	1080	1015	1205	1112	1122	1205	1439
Sm. LS	0.5	1000	1000	1000	1000	1000	1000	1023	1179
Sm. LS	1	1047	1066	1253	1037	1067	1049	1010	1080
Sm. LS	1.5	1107	1133	1360	1273	1248	1214	1118	1128
Sm. MLE	0.2	1299	1335	1422	1290	1224	1238	1286	1396
Sm. MLE	0.5	1279	1307	1368	1176	1150	1157	1166	1192
Sm. MLE	1	1236	1269	1331	1158	1067	1073	1070	1059
Sm. MLE	1.5	1219	1258	1328	1223	1082	1065	1000	1000
Min. value		$2.62 \cdot 10^{-2}$	$2.69 \cdot 10^{-2}$	$2.77 \cdot 10^{-2}$	$2.53 \cdot 10^{-2}$	$2.26 \cdot 10^{-2}$	$2.29 \cdot 10^{-2}$	$2.32 \cdot 10^{-2}$	$2.36 \cdot 10^{-2}$
		(3/4, 1/9)	(3/4, 1/7)	(3/4, 1/3)	(3/4, 3/4)	(3/2, 1/9)	(3/2, 1/7)	(3/2, 1/3)	(3/2, 3/4)
IBF-BC	0.2	2309	19684	10833	1448	26947	1502	1511	1545
IBF-BC	0.5	1205	1216	1233	1295	1293	1299	1314	1330
IBF-BC	1	1058	1059	1068	1091	1104	1109	1113	1126
IBF-BC	1.5	1000	1000	1000	1000	1000	1000	1000	1000
Sm. LS	0.2	1371	1382	1450	1586	1598	1627	1670	1750
Sm. LS	0.5	1148	1161	1192	1301	1305	1307	1351	1413
Sm. LS	1	1116	1115	1115	1150	1189	1195	1200	1223
Sm. LS	1.5	1247	1236	1190	1157	1286	1283	1247	1204
Sm. MLE	0.2	1285	1294	1342	1431	1388	1409	1441	1496
Sm. MLE	0.5	1101	1111	1120	1192	1151	1153	1184	1225
Sm. MLE	1	1038	1035	1021	1028	1062	1064	1057	1061
Sm. MLE	1.5	1141	1127	1059	1000	1207	1198	1141	1067
Min. value		$2.39 \cdot 10^{-2}$	$2.41 \cdot 10^{-2}$	$2.51 \cdot 10^{-2}$	$2.78 \cdot 10^{-2}$	$2.95 \cdot 10^{-2}$	$2.96 \cdot 10^{-2}$	$3.08 \cdot 10^{-2}$	$3.38 \cdot 10^{-2}$

6.6 Simulation results: symmetric auctions

Finally, we simulate 250 symmetric two–bidder auctions in which the bidders' valuations are independently drawn from a mixture of independent beta distributions and estimate the bidders' expected surplus and mean valuation using the pooled sample of bids. In table 4, we isolate and quantify the additional benefit of exploiting symmetry by substituting the known win–probability distribution as in section sections 5.3 and 5.4 compared with the indirect method, which computes the mean valuation and expected surplus from averages of the pseudo sample of values. For the bidder's surplus the indirect estimator had a standard deviation that was between 1.32 and 4.29 times as large as the standard deviation of the estimator in section 5.3. For the mean valuation, the relative standard deviation of the estimators ranged from 1.01 to 1.75.

Table 4: The standard deviations of the indirect estimators for the bidder's surplus and mean valuation relative to the estimators in section 5.3 and section 5.4. The bidder's valuations are distributed as $0.1B(\theta_1, \theta_2) + 0.9B(\theta_3, \theta_4)$, where $B(\theta_i, \theta_j)$ are independent Beta–distributed random variables with shape parameters θ_i and θ_j.

	$(\theta_1, \theta_2, \theta_3, \theta_4)$					
	(1, 1, 1, 1)	(2, 1, 1, 2)	(1, 1, 2, 2)	(1, 1, 2, 3)	(1, 1, 1, 3)	(2, 1, 1, 3)
BS	3.03	3.63	1.42	1.32	4.16	4.29
MV	1.07	1.28	1.01	1.02	1.75	1.68

7 Conclusion

This paper reformulates the empirical analysis of auction models as an isotonic estimation problem by treating the probability of winning as the choice variable in the bidders' decision problem. The nonparametric least–squares and nonparametric maximum likelihood estimators for a bidder's inverse strategy function are shown to converge at the optimal nonparametric rate. As a complementary set of results, we prove the asymptotic behavior of two boundary correction methods that can be combined with transformation to better control the bias–variance tradeoff in the kernel smoothed versions of our estimators. While these smoothing methods are important when estimating some objects of potential interest to the researcher, smoothing is not necessary for others. We prove that using our unsmoothed estimator as an input to a simple plug–in estimator of parameters such as the bidder's expected surplus achieves the semiparametric efficiency bound.

Because auction models can be identified when bidders are asymmetric and when only a subset of the bids is observed (Athey and Haile, 2002; Campo et al., 2003), we provide separate results depending on assumptions made about the bids observed by the econometrician. The reason for this flexibility is that, in a first–price auction, bidder one's equilibrium expenditure function e only depends on the distribution of the maximum competitor bid. We therefore assume that the data are sufficient to obtain an estimate of this distribution and use this (unconstrained) estimator as the starting point for our analysis. If bidders are symmetric and their bids are independent, such an estimator could be G_T^{n-1}, where G_T is the empirical distribution of the bids. If bidders are asymmetric then the product of the competitors' marginal empirical bid distributions would be a natural choice. If there is possible dependence among the competitors' bids, either arising from dependence in the competitor's valuations or coordination in their bids, then the empirical distribution of the maximum of the competitors' bids can be used. We show how the asymptotic properties of our constrained estimators of e and α improve as we add independence and symmetry assumptions: such improvements can be substantial and depend on the object being estimated.

When bidders are symmetric, the benefits of pooling bids in order to estimate e and α using more data are conceptually clear and obvious. Another benefit is that the symmetric monotone equilibrium of a high–bid auction implies the distribution of optimally chosen win–probabilities is a known function of the number of bidders. We show that using the true distribution in combination with estimates of e and α can significantly reduce the variance of estimated objects compared to two–step methods, which use the empirical distribution of win–probabilities.

Our paper addresses many issues and is fairly exhaustive in several dimensions. Nevertheless, there are several issues that we do not address in the paper. First, we ignore the potential presence of a (binding) reserve price. A binding reserve price would affect identification of certain objects,[48] but for many other objects, allowing for a reserve price would pose a minor, not especially interesting (from an econometric perspective), nuisance. Further, we do not allow for endogenous entry. Although endogenous entry can be an important concern in empirical work and raises interesting modeling and identification questions (Levin and Smith, 1994; Li and Zheng, 2009; Marmer et al., 2013; Gentry and Li, 2014), there are many ways of modeling this and it would be beyond the scope of this paper. The same comment applies to possible risk aversion, albeit that risk aversion would likely pose a tougher problem because nonparametric identification of the bidders' utility functions requires an exclusion restriction (Guerre et al., 2009). Relaxing the independence assumption to allow for affiliation among the bidders' valuations would similarly require an extension that combines the insight of Campo et al. (2003) with our approach to estimate bidder 1's expected payment conditional on its private valuation. Finally, there can be auction–level heterogeneity. We propose a new method for correcting for observable auction–specific heterogeneity; unobserved heterogeneity might be addressed using methods similar to Krasnokutskaya (2011) or Roberts (2013). We leave these questions for future work.

References

Athey, S. and Haile, P. A. (2002). Identification of Standard Auction Models. *Econometrica*, 70(6):2107–2140.

Brunk, H. D. (1955). Maximum likelihood estimates of monotone parameters. *The Annals of Mathematical Statistics*, pages 607–616.

Campo, S., Perrigne, I., and Vuong, Q. (2003). Asymmetry in first-price auctions with affiliated private values. *Journal of Applied Econometrics*, 18(2):179–207.

Carolan, C. and Dykstra, R. (2001). Marginal densities of the least concave majorant of brownian motion. *Annals of Statistics*, 29(6):1732–1750.

Gentry, M. and Li, T. (2014). Identification in Auctions With Selective Entry. *Econometrica*, 82(1):315–344.

Gimenes, N. and Guerre, E. (2019). Quantile regression methods for first–price auctions. Technical report, University of Kent.

Grenander, U. (1956a). On the theory of mortality measurement. *Scandinavian Actuarial Journal*, pages 70–96.

[48] For instance, the value distribution below the reserve price.

Grenander, U. (1956b). On the theory of mortality measurement: Part ii. *Scandinavian Actuarial Journal*, pages 125–153.

Guerre, E., Perrigne, I., and Vuong, Q. (2000). Optimal Nonparametric Estimation of First-Price Auctions. *Econometrica*, 68(3):525–574.

Guerre, E., Perrigne, I., and Vuong, Q. (2009). Nonparametric Identification of Risk Aversion in First-Price Auctions Under Exclusion Restrictions. *Econometrica*, 77(4):1193–1227.

Haile, P. A., Hong, H., and Shum, M. (2003). Nonparametric Tests for Common Values in First-price Sealed-bid Auctions.

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

Hanson, M. A. (1981). On Sufficiency of the Kuhn-Tucker Conditions. *Journal of Mathematical Analysis and Applications*, 80:545–550.

Hanson, M. A. (1999). Invexity and the Kuhn-Tucker Theorem. *Journal of Mathematical Analysis and Applications*, 236:594–604.

Henderson, D. J., List, J. A., Millimet, D. L., Parmeter, C. F., and Price, M. K. (2012). Empirical implementation of nonparametric first–price auction models. *Journal of Econometrics*, 168(1):17 – 28.

Hickman, B. R. and Hubbard, T. P. (2015). Replacing sample trimming with boundary correction in nonparametric estimation of first-price auctions. *Journal of Applied Econometrics*, 30(5):739–762.

Jun, S. J. and Pinkse, J. (2019). An information–theoretic approach to partially identified auction models.

Jun, S. J., Pinkse, J., and Wan, Y. (2015). Classical Laplace estimation for $\sqrt[3]{n}$–consistent estimators: Improved convergence rates and rate-adaptive inference. *Journal of econometrics*, 187(1):201–216.

Karunamuni, R. J. and Alberts, T. (2005). On boundary correction in kernel density estimation. *Statistical Methodology*, 2(3):191–212.

Karunamuni, R. J. and Zhang, S. (2008). Some improvements on a boundary corrected kernel density estimator. *Statistics & Probability Letters*, 78(5):499–507.

Kim, J. and Pollard, D. (1990). Cube root asymptotics. *Annals of Statistics*, 18(1):191–219.

Krasnokutskaya, E. (2011). Identification and Estimation of Auction Models with Unobserved Heterogeneity. *The Review of Economic Studies*, 78(1):293–327.

Larsen, B. and Zhang, A. L. (2018). A mechanism design approach to identification and estimation.

Lebrun, B. (2006). Uniqueness of the equilibrium in first-price auctions. *Games and Economic Behavior*, 55(1):131–151.

Levin, D. and Smith, J. L. (1994). Equilibrium in Auctions with Entry. *American Economic Review*, 84(3):585–599.

Li, T. (2005). Econometrics of first-price auctions with entry and binding reservation prices. *Journal of*

Econometrics, 126(1):173–200.

Li, T., Perrigne, I., and Vuong, Q. (2002). Structural Estimation of the Affiliated Private Value Auction Model. *The RAND Journal of Economics*, 33(2):171–193.

Li, T. and Zheng, X. (2009). Entry and Competition Effects in First-Price Auctions: Theory and Evidence from Procurement Auctions. *Review of Economic Studies*, 76(4):1397–1429.

Luo, Y. and Wan, Y. (2018). Integrated–quantile–based estimation for first–price auction models. *Journal of Business & Economic Statistics*, 36(1):173–180.

Ma, J., Marmer, V., and Shneyerov, A. (2019a). Inference for first–price auctions with guerre, perrigne, and vuongs estimator. *Journal of Econometrics*, 211:507–538.

Ma, J., Marmer, V., Shneyerov, A., and Xu, P. (2019b). Monotonicity–constrained nonparametric estimation and inference for first–price auctions. Technical report.

Marmer, V. and Shneyerov, A. (2012). Quantile–based nonparametric inference for first-price auctions. *Journal of Econometrics*, 167(2):345–357.

Marmer, V., Shneyerov, A., and Xu, P. (2013). What model for entry in first-price auctions? A nonparametric approach. *Journal of Econometrics*, 176(1):46–58.

Maskin, E. and Riley, J. (2000). Equilibrium in sealed high bid auctions. *Review of Economic Studies*, 67(3):439–454.

Milgrom, P. R. and Weber, R. J. (1982). A theory of auctions and competitive bidding. *Econometrica*, 50(5):1089–1122.

Myerson, R. B. (1981). Optimal auction design. *Mathematics of operations research*, 6(1):58–73.

Newey, W. K. (1990). Semiparametric efficiency bounds. *Journal of applied econometrics*, 5(2):99–135.

Pinkse, J. and Schurter, K. (2019). Improved bandwidth selection for boundary correction using the generalized reflection method. Penn State working paper.

Powell, J. L., Stock, J. H., and Stoker, T. M. (1989). Semiparametric estimation of index coefficients. *Econometrica*, pages 1403–1430.

Roberts, J. W. (2013). Unobserved heterogeneity and reserve prices in auctions. *RAND Journal of Economics*, 44(4):712–732.

Robinson, P. M. (1988). Root-n-consistent semiparametric regression. *Econometrica*, pages 931–954.

Tripathi, G. (2000). Local semiparametric efficiency bounds under shape restrictions. *Econometric Theory*, 16(5):729–739.

van der Vaart, A. W. (2000). *Asymptotic statistics*, volume 3. Cambridge university press.

A Proofs of results in the main text

The proofs are ordered in the same way as the technical results in the main text. Additional results can be found in appendix B.

A.1 Least squares estimator

Proof of theorem 1. We first establish the results for $\breve{e}_T$, then uniform consistency of α_T. Convexity of $\breve{e}_T$ follows by construction since α_T is restricted to be monotonic.

Trivially extending the arguments in (van der Vaart, 2000, lemma 21.4 and the dicussion at the top of p308) about the empirical quantile process to $\hat{e}_T$, $\sqrt{T}\{e_T(\cdot)-e(\cdot)\}$ has the asserted limit process on $(0,1)$. We now extend this to $[0,1]$. Note that $e_T(0)-e(0)=0$ and $\sqrt{T}\{e_T(1)-e(1)\}=o_p(1)$, so we have convergence of finite marginals and tightness on $(0,1)$. We thus only need to extend tightness to $[0,1]$.

We show the argument at one, where the argument at zero follows analogously. Let $p_T = 1-1/T$ and $\Delta_T(p)=\sqrt{T}(e_T(p)-e(p))$. Then for any sequence $\delta_T = o(1)$ by the triangle inequality,

$$\sup_{1>p>1-\delta_T}\left|\Delta_T(1)-\Delta_T(p)\right| \leq \sup_{1>p>1-\delta_T}\left|\Delta_T(p_T)-\Delta_T(p)\right| + \left|\Delta_T(p_T)-\Delta_T(1)\right|. \tag{35}$$

The first right hand side term in (35) is $o_p(1)$ by tightness on $(0,1)$. The second right hand side term in (35) is $o_p(1)$ since the second highest order statistic converges at rate T.

By Carolan and Dykstra (2001), the limit process is identical for the greatest convex minorant provided that e is *strictly* convex, which is implied by assumption B since $\alpha = e'$ is the inverse bid function composed with Q_c.

Finally, uniform convergence of α_T. Let $t_T = 1/\sqrt[3]{t}$. Then, by the monotonicity of α and α_T,

$$\max_{t_T\leq p\leq 1-t_T}\{\alpha_T(p)-\alpha(p)\} \leq \max_{t_T\leq p\leq 1-t_T}\frac{\breve{e}_T(p+t_T)-\breve{e}_T(p)-e(p+t_T)+e(p)}{t_T} + \max_{p\in\mathscr{P}}\left(\frac{e(p+t_T)-e(p)}{t_T}-\alpha(p)\right)$$

The first right hand side term is $O_p(\sqrt{1/t_T T}) = O_p(t_T)$. The second right hand side term is bounded above by $\alpha(p+t_T)-\alpha(p) = O_p(t_T)$, also. The minimum can be dealt with analogously. □

Justification of (11) and (12). We provide a sketch of the proof and a derivation of the limit distribution. A full proof would be more careful, especially about issues pertaining to uniformity. However, there is nothing special about the present scenario and a full rigorous proof would be lengthy but routine.

Our justification follows two steps. In the first step, we derive a limit result for the inverse problem, i.e. the estimation of α^{-1}. In the second step we then apply equations (15) and (16) of Jun et al. (2015) to obtain the limit distribution of α_T itself.

We first establish asymptotics for the 'inverse isotonic regression'–type estimator and then take its inverse to obtain asymptotics for α_T. Note that for $\xi = \alpha(p)$, $\operatorname{argmin}_{\tilde{p}}\{e(\tilde{p}) - \xi\tilde{p}\} = \alpha^{-1}(\xi) = p$. Let $\hat{p}$ be its sample equivalent, such that

$$\begin{aligned}\sqrt[3]{T}(\hat{p}-p) = \sqrt[3]{T}\left\{\operatorname*{argmin}_{\tilde{p}}\{\breve{e}_T(\tilde{p}) - \xi\tilde{p}\} - p\right\} = \sqrt[3]{T}\left\{\operatorname*{argmin}_{\tilde{p}}\{\breve{e}_T(\tilde{p}) - \breve{e}_T(p) - \xi(\tilde{p}-p)\} - p\right\} = \\ \sqrt[3]{T}\left\{\operatorname*{argmin}_{\tilde{p}}\Big(\{\breve{e}_T(\tilde{p}) - \breve{e}_T(p) - e(\tilde{p}) + e(p)\} + \{e(\tilde{p}) - e(p) - \xi(\tilde{p}-p)\}\Big) - p\right\} \simeq \\ \sqrt[3]{T}\left\{\operatorname*{argmin}_{\tilde{p}}\Big(\{\breve{e}_T(\tilde{p}) - \breve{e}_T(p) - e(\tilde{p}) + e(p)\} + \alpha'(p)(\tilde{p}-p)^2/2\Big) - p\right\} = \\ \operatorname*{argmin}_{t}\Big(T^{2/3}\{\breve{e}_T\big(p + t/\sqrt[3]{T}\big) - \breve{e}_T(p) - e\big(p + t/\sqrt[3]{T}\big) + e(p)\} + \alpha'(p)t^2/2\Big) \\ \xrightarrow{d} \operatorname*{argmin}_{t}\big(\mathbb{G}^\circ(t) + \alpha'(p)t^2/2\big) \sim \operatorname*{arg\,max}_{t}\big(\mathbb{G}^\circ(t) - \alpha'(p)t^2/2\big).\end{aligned}$$

where $\mathbb{G}^\circ$ is a Gaussian process with covariance kernel[49]

$$\lim_{T\to\infty} \sqrt[3]{T}\left\{H\big(p + t/\sqrt[3]{T}, p + s/\sqrt[3]{T}\big) - H\big(p + t/\sqrt[3]{T}, p\big) - H\big(p, p + s/\sqrt[3]{T}\big) + H(p,p)\right\},$$

which under (7) simplifies to $\zeta^2(p)|\mathrm{Med}(s,t,0)|$: in other words, $\mathbb{G}^\circ$ is then $\zeta(p)$ times a standard two–sided Brownian motion $\mathbb{G}^B$, such that then by a change of variables,

$$\operatorname*{arg\,max}_{t}\big(\mathbb{G}^\circ(t) - \alpha'(p)t^2/2\big) \sim \left\{2\zeta(p)/\alpha'(p)\right\}^{2/3} \operatorname*{arg\,max}_{t}\big(\mathbb{G}^B(t) - t^2\big)$$

From equations (15) and (16) in Jun et al. (2015) it then follows that

$$\sqrt[3]{T}\{\alpha_T(p) - \alpha(p)\} \xrightarrow{d} \sqrt[3]{4\zeta^2(p)\alpha'(p)}\operatorname*{arg\,max}_{t}\big(\mathbb{G}^B(t) - t^2\big).$$

A.2 NPMLE

Definition 1 (Invex function)**.** A function $f : S \subseteq \mathbb{R}^n \to \mathbb{R}$ is invex at $u \in S$ if there exists a $\mathbb{R}^n$-valued function η such that $f(x) - f(u) \geq \eta(x) \cdot \nabla f(u)$ for all $x \in S$, where ∇f denotes the gradient vector of f. Such a function η is known as an invexity kernel.

A function $g : S \subseteq \mathbb{R}^n \to \mathbb{R}$ is type I invex at $u \in S$ if there exists a $\mathbb{R}^n$-valued function η such that $-g(u) \geq \eta(x) \cdot \nabla g(u)$ for all $x \in S$.

Theorem 12 (Hanson (1999) theorem 2.1)**.** *Consider the problem* $\min_{x\in S} f(x)$ *subject to* $g(x) \leq 0$ *for some functions* $f : \mathbb{R}^n \to \mathbb{R}$ *and* $g : \mathbb{R}^n \to \mathbb{R}^m$ *that are differentiable on* S*. For* $u \in S$ *to be optimal, it is*

[49] This fact can be most easily seen by thinking in terms of a Bahadur representation.

sufficient that the KKT conditions are satisfied at u and f and g satisfy $f(x) - f(u) \geq \eta(x) \cdot \nabla f(u)$ and $-g_t(u) \geq \eta(x) \cdot \nabla g_t(u)$ for every active component g_t of g for some common invexity kernel $\eta : \mathbb{R}^n \to \mathbb{R}^n$. □

Proof. See Hanson (1999), theorem 2.1. □ □

Lemma 5. *The KKT conditions (58) are necessary and sufficient for the isotonic maximum likelihood problem (14).*

Proof. The proof proceeds as follows. First, we establish that $-\mathscr{L}$ and the isotonicity constraints are invex functions, a generalization of convex functions (see definition 1) which may be equivalently characterized as the collection of differentiable functions for which every stationary point is a global minimum. Second, we show that these functions are invex with respect to the same invexity kernel using Gale's theorem of the alternative, as suggested by Hanson (1981). Finally, we conclude that the KKT conditions for the constrained minimization of $-\mathscr{L}$ are sufficient by a direct application of theorem 2.1 of Hanson (1999).

First, $-\mathscr{L}$ is invex because every stationary point is a global minimum. To see this, we note that $-\mathscr{L}$ is convex at its (unique) stationary point because

$$-\frac{\partial^2 \mathscr{L}}{\partial \alpha_{(t)}^2} = \frac{t-2}{(\alpha_{(t)} - b_{(t)})^2} - \frac{t-1}{(\alpha_{(t)} - b_{(t-1)})^2} = \frac{t-1}{(\alpha_{(t)} - b_{(t-1)})(\alpha_{(t)} - b_{(t)})} - \frac{t-1}{(\alpha_{(t)} - b_{(t-1)})^2} > 0\,,$$

where the second equality follows by substitution using the first-order condition, and the inequality follows from $b_{(t)} > b_{(t-1)}$. Thus, there is a unique local minimum. Finally, there are no minima at the boundaries because $-\partial_{\alpha_{(t)}} \mathscr{L}$ is eventually positive as $\alpha_{(t)}$ tends to infinity and $-\mathscr{L}$ diverges to infinity as $\alpha_{(t)}$ approaches $b_{(t)}$. Thus, every stationary point is a global minimum. Hence, $-\mathscr{L}$ is invex. Then, by definition of invexity, there exists a vector-valued invexity kernel η such that $L(\breve{\alpha}_T^{\text{MLE}}) - L(\tilde{\alpha}) \geq \eta(\tilde{\alpha}) \cdot \nabla L(\tilde{\alpha})$ for all $\tilde{\alpha}$. The constraints on $\tilde{\alpha}_{(t)}$ are linear and therefore invex, as well.

Second, we must further show that there exists a *common* invexity kernel with respect to which $-\mathscr{L}$ and the (active) constraint functions $c_t(\tilde{\alpha}) = -\tilde{\alpha}_{(t)} + \tilde{\alpha}_{(t-1)}$ are (type I) invex at the solution. The existence of such an invexity kernel is implied by the existence of a solution to the linear system $A\eta(\tilde{\alpha}) \leq C(\tilde{\alpha})$ for all $\tilde{\alpha}$, where A is the $\mathbb{R}^{T-2} \times \mathbb{R}^{T-2}$ Jacobian matrix $(-\nabla \mathscr{L}; \nabla c_4; \dots; \nabla c_T)$ evaluated at $\breve{\alpha}_T^{\text{MLE}}$ and $C(\tilde{\alpha}) = (\mathscr{L}(\breve{\alpha}_T^{\text{MLE}}) - \mathscr{L}(\tilde{\alpha}), -c_4(\tilde{\alpha}), \dots, -c_T(\tilde{\alpha}))$. By Gale's theorem of the alternative, a solution to this system of inequalities exists if and only if there does not exist a vector $y \geq 0$ such that $y'A = 0$ and $C(\tilde{\alpha})'y = -1$.

Because A is of the form

$$\begin{pmatrix} -\partial_{\breve{\alpha}^{\text{MLE}}_{(3)}}\mathscr{L} & -\partial_{\breve{\alpha}^{\text{MLE}}_{(4)}}\mathscr{L} & -\partial_{\breve{\alpha}^{\text{MLE}}_{(5)}}\mathscr{L} & -\partial_{\breve{\alpha}^{\text{MLE}}_{(6)}}\mathscr{L} & \cdots \\ 1 & -1 & 0 & 0 & \cdots \\ 0 & 1 & -1 & 0 & \\ 0 & 0 & 1 & -1 & \\ \vdots & & & \ddots & \ddots \end{pmatrix}$$

and because the stationarity condition implies $\lambda_t = \sum_{s=3}^{t-1} -\partial_{\breve{\alpha}^{\text{MLE}}_{(s)}}\mathscr{L}$, we can show that $y'A = 0$ has a solution only if y is a scalar multiple of $(1, \lambda_4, \dots, \lambda_T)$. But then $C(\tilde{\alpha})'y = -1$ does not have a solution for any $\tilde{\alpha}$ because y and $C(\tilde{\alpha})$ are nonnegative vectors for all $\tilde{\alpha}$. Thus, the objective and active constraints are (type I) invex with respect to some shared invexity kernel.

Finally, any solution to the KKT conditions is a global minimum by theorem 12. □ □

Lemma 6. *The derivative of* $\mathscr{L}_j(\alpha) = \sum_{s=t_j}^{t_{j+1}-1} \left\{(s-2)\log(\alpha - b_{(s)}) - (s-1)\log(\alpha - b_{(s-1)})\right\}$ *with respect to* α *is zero exactly once on* $(b_{t_{j+1}-1}, \infty)$ *and crosses zero from above.*

Proof. Multiplying the stationarity condition $\mathscr{L}_j'(\alpha) = 0$ by $\alpha - b_{(t_{j+1}-1)}$ and collecting terms, we find that α is a stationary point if and only if

$$(t_j - 1)y_{t_j-1}(\alpha) + 2\sum_{s=t_j}^{t_{j+1}-2} y_s(\alpha) = t_{j+1} - 3, \qquad \text{where } y_s(\alpha) = \frac{\alpha - b_{(t_{j+1}-1)}}{\alpha - b_{(s)}}.$$

For all s, $y_s(b_{(t_{j+1}-1)}) = 0$ and $y_s(\alpha)$ is continuous and increasing in α. The left side of the equation is equal to zero at $b_{(t_{j+1}-1)}$ and approaches $t_j - 1 + 2(t_{j+1} - 1 - t_j) = 2t_{j+1} - t_j - 3 > t_{j+1} - 3$ as α increases. There exists an α that solves the equation above by the intermediate value theorem, and the solution is unique because the left side is strictly monotonic in α.

Finally, $\mathscr{L}_j'$ crosses zero from above because $\mathscr{L}_j'$ diverges to positive infinity as α approaches $b_{(t_{j+1}-1)}$. □ □

Lemma 7. *If* $\tilde{\alpha}^{(k-1)}_{(t_j)}$ *and* $\tilde{\alpha}^{(k-1)}_{(t_{j+1})}$ *are the values of* $\tilde{\alpha}$ *in the two blocks that are pooled together in the* k*-th step, then the new value is* $\tilde{\alpha}^{(k)}_{(t_j)}$ *between* $\tilde{\alpha}^{(k-1)}_{(t_j)}$ *and* $\tilde{\alpha}^{(k-1)}_{(t_{j+1})}$.

Proof. Without loss of generality, assume $\tilde{\alpha}^{(k-1)}_{(t_{j+1})} < \tilde{\alpha}^{(k-1)}_{(t_j)}$.The zero of $\mathscr{L}_j' + \mathscr{L}_{j+1}'$ must be greater than $\tilde{\alpha}^{(k-1)}_{(t_{j+1})}$, because both $\mathscr{L}_j'$ and $\mathscr{L}_{j+1}'$ are positive to the left of $\tilde{\alpha}^{(k-1)}_{(t_{j+1})}$ by lemma 6. On the other hand, the zero of $\mathscr{L}_j' + \mathscr{L}_{j+1}'$ must be less than $\tilde{\alpha}^{(k-1)}_{(t_j)}$ because both derivatives are negative to the right of $\tilde{\alpha}^{(k-1)}_{(t_j)}$ by lemma 6. □ □

Lemma 8. *PAVA for the NPMLE is a dual active set method, i.e. PAVA satisfies stationarity, complementary slackness, and dual feasibility at every step of the algorithm, but does not satisfy primal feasibility until the final iterate.*

Proof. The PAVA algorithm clearly satisfies the stationarity and complementary slackness conditions at every step, and satisfies primal feasibility at the last step (primal feasibility is the stopping criterion). It remains to show that the Lagrange multipliers are nonnegative at every step.

Let

$$\tilde{\lambda}_t^{(k)} = \sum_{s=3}^{t-1} \partial_{\tilde{\alpha}_{(s)}} \mathscr{L}(\tilde{\alpha}) = \sum_{s=3}^{t-1} \left(\frac{s-2}{\tilde{\alpha}_{(s)}^{(k)} - b_{(s)}} - \frac{s-1}{\tilde{\alpha}_{(s)}^{(k)} - b_{(s-1)}} \right)$$

denote the Lagrange multipliers implied by the stationarity conditions, where the superscripts (k) indicate the value of the variable after the k-th step of the algorithm. Initially, $\tilde{\lambda}_t^{(0)} = 0$ for all t.

We will proceed by induction on k. Let t_j and t_{j+1} be the starting points of the adjacent blocks pooled together in the k-th step for some $k > 0$. Assume $\tilde{\lambda}_t^{(j)} \geq 0$ for all t and $j < k$.

Suppose by way of contradiction that a negative Lagrange multiplier is introduced in the k-th step. The negative multiplier must apply to one of the active constraints in the two most recently merged blocks of constraints, because the Lagrange multipliers on constraints outside these two blocks are unaffected: the multipliers are all zero for the slack constraints on the singleton blocks to the right, and clearly $\tilde{\lambda}_t^{(k)}$ is unaffected for all $t \leq t_j$.

A negative multiplier on one of the constraints in the most recently merged blocks implies that there exists a constraint within this chain of equalities that can be slackened and increase the loglikelihood. We will show that this leads to a contradiction because slackening any one of the constraints and moving in the direction that would increase the loglikelihood will necessarily violate primal feasibility.

Suppose we slacken the constraint $\tilde{\alpha}_{(s)} \geq \tilde{\alpha}_{(s-1)}$. There are two cases to consider. First, suppose the slackened constraint belongs to the left pre-merged block, i.e. s is such that $t_j \leq s < t_{j+1}$, and let $\tilde{\alpha}_{(t_j)}^{\prime(k)}$ denote the new solution to the stationarity condition in the sub-block to the left of s. Let $\tilde{\alpha}_{(s)}^{\prime(k-1)}$ denote the solution for $\tilde{\alpha}$ in the block beginning with s and ending $t_{j+1} - 1$. Then $\tilde{\alpha}_{(t_j)}^{\prime(k)}$ must be greater than the value of $\tilde{\alpha}_{(s)}^{\prime(k-1)}$ in the right sub-block, otherwise relaxing this constraint would have been feasible and improved the loglikelihood in an earlier iterate, thereby contradicting our assumption that the k-th step is the first that introduces a negative Lagrange multiplier. In addition, $\tilde{\alpha}_{(t_j)}^{\prime(k)} > \tilde{\alpha}_{(t_j)}^{(k-1)} > \tilde{\alpha}_{(s)}^{\prime(k-1)}$ by lemma 7. Finally, we invoke lemma 7 again to conclude that the value of $\tilde{\alpha}_{(t_{j+1})}^{\prime(k)}$ is less than $\tilde{\alpha}_{(t_j)}^{\prime(k)}$. Hence, none of the constraints in the left block can be removed while maintaining primal feasibility.

On the other hand, we may suppose the objective would by improved by making one of the constraints in the right block slack. By a similar argument, this too would violate primal feasibility. Therefore, none of the constraints in the merged block can be removed without violating primal feasibility or decreasing the objective. Therefore, none of the Lagrange multipliers are negative after the k-th iterate. By induction, there are no negative Lagrange multipliers in any step of the algorithm. □ □

Proof of theorem 2. The final iterate of the PAVA algorithm satisfies the KKT conditions by lemma 8, which are sufficient for the global maximum by lemma 5. □

Justification of (15). We provide a sketch of the proof and a derivation of the limit distribution. A full proof would be more careful, especially about issues pertaining to uniformity. However, there is nothing special about the present scenario and a full rigorous proof would be lengthy but routine.

We remind the reader that $\alpha_T^{\text{mle}}(p) = \alpha_T^{\text{mle}}\{G_c(b)\}$ can be characterized in terms of the inverse of the solution to an 'inverse regression' problem, namely to find the solution $\beta_T(v)$ that minimizes

$$\mathbb{S}_T(b,v) = \frac{1}{T}\sum_{t=2}^{T}\left(\frac{t-2}{v-b_{(t)}} - \frac{t-1}{v-b_{(t-1)}}\right)\mathbb{1}(b_{(t)} \le b),$$

over a region of b's for which $v-b$ is bounded away from zero. Note that β_T is an estimate of the bid function $\beta(v)$. We first obtain the limit distribution of $\sqrt[3]{T}\{\beta_T(v)-\beta(v)\}$. We then invoke the results of Jun et al. (2015) to obtain properties of the inverse.

We first obtain an approximation of the form $\mathbb{S}_T(b,v) \simeq \mathbb{S}_T^*(b,v)+\mathbb{S}_T^\circ(b)+\mathbb{S}(b,v)$, for functions $\mathbb{S}_T^*, \mathbb{S}_T^\circ, \mathbb{S}$ introduced below, where $\simeq$ means that the omitted terms are negligible. Taking $\sqrt[3]{T}$–consistency as given, we then argue that $T^{2/3}\left[\mathbb{S}_T\{\beta(v)+x/\sqrt[3]{T}, v\} - \mathbb{S}_T\{\beta(v), v\}\right]$ converges to a limiting Gaussian process plus a quadratic in x, whose minimizer as a function of x has a (scaled) Chernoff distribution. Applying equations (15) and (16) in Jun et al. (2015) then yield the stated limit distribution.

Noting that $\max_t |b_{(t)} - Q_{ct}| = O_p(T^{-1/2})$ for $Q_{ct} = Q_c(t/T)$, we have (uniformly in b, v),

$$\begin{aligned}
\mathbb{S}_T(b,v) = \frac{1}{T}\sum_{t=2}^{T}\left(\frac{t-2}{v-b_{(t)}} - \frac{t-1}{v-b_{(t-1)}}\right)\mathbb{1}(b_{(t)} \le b) = \\
\frac{1}{T}\sum_{t=2}^{T}\frac{t}{v-b_{(t)}}\mathbb{1}(b_{(t)} \le b < b_{(t+1)}) - \frac{1}{T}\sum_{t=2}^{T}\frac{2}{v-b_{(t)}}\mathbb{1}(b_{(t)} \le b) \\
= \underbrace{\frac{1}{T}\sum_{t=2}^{T}\frac{t}{v-Q_{ct}}\mathbb{1}(b_{(t)} \le b < b_{(t+1)})}_{\text{I}} - \underbrace{\frac{1}{T}\sum_{t=2}^{T}\frac{2}{v-Q_{ct}}\mathbb{1}(b_{(t)} \le b)}_{\text{II}} \\
+ \underbrace{\sum_{t=2}^{T}\frac{t}{T}\frac{b_{(t)}-Q_{ct}}{(v-b_{(t)})^2}\mathbb{1}(b_{(t)} \le b < b_{(t+1)})}_{\text{III}} - \underbrace{\frac{2}{T}\sum_{t=2}^{T}\frac{b_{(t)}-Q_{ct}}{(v-b_{(t)})^2}\mathbb{1}(b_{(t)} \le b)}_{\text{IV}} + O_p(T^{-1}).
\end{aligned}$$

Now, term I is

$$\frac{1}{T}\sum_{t=2}^{T}\frac{t}{v-Q_{ct}}\mathbb{1}\left(\frac{t}{T}\leq G_{cT}(b)<\frac{t+1}{T}\right)\simeq\frac{G_{cT}(b)}{v-Q_c\{G_T(b)\}}\simeq 2\frac{G_{cT}(b)-G_c(b)}{v-b}+\frac{G_c(b)}{v-b}, \tag{36}$$

where $\simeq$ means that asymptotically negligible terms were omitted. Further, term II is

$$\simeq\int_0^{G_{cT}(b)}\frac{2}{v-Q_c(p)}\,\mathrm{d}p\simeq 2\int_0^b\frac{g_c(\tilde b)}{v-\tilde b}\,\mathrm{d}\tilde b+2\frac{G_{cT}(b)-G_c(b)}{v-b}\simeq$$
$$2\frac{G_c(b)}{v-b}-2\int_0^b\frac{G_c(\tilde b)}{(v-\tilde b)^2}\,\mathrm{d}\tilde b+2\frac{G_{cT}(b)-G_c(b)}{v-b} \tag{37}$$

Term III is

$$\simeq\sum_{t=2}^{T}\frac{t}{T}\frac{b_{(t)}-Q_{ct}}{(v-Q_{ct})^2}\mathbb{1}(Q_{ct}\leq b<Q_{c,t+1})\simeq G_c(b)\frac{Q_{cT}\{G_c(b)\}-b}{(v-b)^2}\simeq-\frac{G_{cT}(b)-G_c(b)}{(v-b)}. \tag{38}$$

Finally, term IV is

$$\simeq\frac{2}{T}\sum_{t=2}^{T}c_t(v)(b_{(t)}-Q_{ct})\mathbb{1}(Q_{ct}\leq b)\simeq 2\int_0^{G_c(b)}\frac{Q_{cT}(p)-Q_c(p)}{\{v-Q_c(p)\}^2}\,\mathrm{d}p\simeq-2\int_0^b\frac{G_{cT}(\tilde b)-G_c(\tilde b)}{(v-\tilde b)^2}\,\mathrm{d}\tilde b. \tag{39}$$

Adding the right hand sides in (36) and (38) and subtracting the right hand sides in (37) and (39) from the sum yields after integration by parts on one of the nonstochastic terms

$$\mathbb{S}_T(b,v)\simeq\underbrace{-\frac{G_{cT}(b)-G_c(b)}{v-b}}_{\mathbb{S}_T^*(b,v)}+\underbrace{2\int_0^b\frac{G_{cT}(s)-G_c(s)}{(v-s)^2}\,\mathrm{d}s}_{\mathbb{S}_T^\circ(b,v)}+\underbrace{\int_0^b\frac{G_c(s)-g_c(s)(v-s)}{(v-s)^2}\,\mathrm{d}s}_{\mathbb{S}(b,v)}. \tag{40}$$

Now, by assumption C,

$$\sqrt{T}(\mathbb{S}_T^*+\mathbb{S}_T^\circ)\rightsquigarrow\underbrace{2\int_0^b\frac{\mathbb{G}^*(s)}{(v-s)^2}\,\mathrm{d}s}_{\mathbb{S}_1^R}-\underbrace{\frac{\mathbb{G}^*(b)}{v-b}}_{\mathbb{S}_2^R},$$

as a process indexed by (b,v). Thus,

$$\begin{cases}T^{2/3}\left[\mathbb{S}_1^R\{\beta(v)+t/\sqrt[3]{T},v\}-\mathbb{S}_1^R\{\beta(v),v\}\right]=o_p(1),\\ T^{2/3}\left\{\mathbb{S}_2^R\left(\beta(v)+t/\sqrt[3]{T},v\right)-\mathbb{S}_2^R\{\beta(v),v\}\right\}\rightsquigarrow\sqrt{g_c(b)}\{v-\beta(v)\}^{-1}\mathbb{G}^B,\end{cases}$$

where $\mathbb{G}^B$ is a standard two–sided Brownian motion.

Further, $T^{2/3}\left\{\mathbb{S}(b+t/\sqrt[3]{T},v)-\mathbb{S}(b,v)\right\}=\mathbb{S}''(b,v)t^2/2+o(1)$, where the derivatives are taken with respect to b. Putting everything together suggests that under (7) for $b = \beta(v)$,

$$\sqrt[3]{T}\{\beta_T(v) - \beta(v)\} \rightsquigarrow \operatorname*{argmin}_x \left(\frac{\sqrt{g_c(b)}}{v-b} \mathbb{G}^B(x) + \frac{S''(b)}{2} x^2 \right) = \left(\frac{4g_c(b)}{(v-b)^2\{\mathbb{S}''(b)\}^2} \right)^{1/3} \mathbb{C},$$

where $\mathbb{C}$ is a standard Chernoff. Note that when $\mathbb{S}''$ is evaluated at $\beta(v)$, we get $\mathbb{S}''(b) = \{2g_c^2(b) - g_c'(b)G_c(b)\}g_c(b)/G_c^2(b)$. By equations (15) and (16) of Jun et al. (2015) we have that for $b = Q_c(p)$,

$$\sqrt[3]{T}\{\alpha_T^{\text{mle}}(p) - \alpha(p)\} \xrightarrow{d} \left(2 - \frac{G_c(b)g_c'(b)}{g_c^2(b)}\right)\left(\frac{4g_c(b)}{(\alpha(p)-b)^2\{S''(b)\}^2}\right)^{1/3} \mathbb{C} =$$
$$\left\{ \frac{4G_c^2(b)}{g_c^3(b)} \left(2 - \frac{G_c(b)g_c'(b)}{g_c^2(b)}\right) \right\}^{1/3} \mathbb{C} = \sqrt[3]{4\zeta^2(p)\{2Q_c'(p) + Q_c''(p)p\}}\mathbb{C},$$

as claimed. □

A.3 Smoothing

Proof of theorem 3. First convexity. Substitution of $t = (s-p)/h$ yields $\hat{e}_T(p) = \int_{-\infty}^{\infty} \breve{e}_T(p+sh)k(s)\,\mathrm{d}s$. Thus, for any $0 < \lambda < 1$ and any $p_\ell < p_h$,

$$\hat{e}_T\{\lambda p_\ell + (1-\lambda)p_h\} = \int_{-\infty}^{\infty} \breve{e}_T\{\lambda p_\ell + (1-\lambda)p_h + sh\}k(s)\,\mathrm{d}s \le$$
$$\int_{-\infty}^{\infty} \{\lambda\breve{e}_T(p_\ell + sh) + (1-\lambda)\breve{e}_T(p_h + sh)\}k(s)\,\mathrm{d}s = \lambda\hat{e}_T(p_\ell) + (1-\lambda)\hat{e}_T(p_h).$$

Now convergence. We have

$$\sqrt{T}\{\hat{e}_T(p) - e(p)\} = \sqrt{T}\left(\int_{-\infty}^{\infty} \frac{1}{h}\breve{e}_T(s)k\left(\frac{p-s}{h}\right)\mathrm{d}s - e(p)\right) = \sqrt{T}\left(\int_{-\infty}^{\infty} \hat{e}_T(p+sh)k(s)\,\mathrm{d}s - e(p)\right) =$$
$$\underbrace{\sqrt{T}\int_{-\infty}^{\infty}\{\breve{e}_T(p+sh) - \breve{e}_T(p) - e(p+sh) + e(p)\}k(s)\,\mathrm{d}s}_{\text{I}}$$
$$+ \underbrace{\sqrt{T}\{\breve{e}_T(p) - e(p)\}}_{\text{II}} + \underbrace{\sqrt{T}\int_{-\infty}^{\infty}\{e(p+sh) - e(p)\}k(s)\,\mathrm{d}s}_{\text{III}}$$

Term II is what we want left over. Term III is $o(1)$ by a standard kernel bias expansion and assumption D. Finally, term I is $o_p(1)$ by theorem 1.

We end by establishing convergence of $\hat{\alpha}_T$. Note that

$$\sqrt{Th}\big\{\hat{\alpha}_T(p)-\alpha(p)\big\} = -\sqrt{Th}\left(\frac{1}{h^2}\int_{-\infty}^{\infty}\breve{e}_T(s)k'\left(\frac{s-p}{h}\right)\mathrm{d}s+\alpha(p)\right) = -\sqrt{Th}\left(\frac{1}{h}\int_{-\infty}^{\infty}\breve{e}_T(p+sh)k'(s)\mathrm{d}s+\alpha(p)\right) =$$

$$\underbrace{-\sqrt{\frac{T}{h}}\int_{-\infty}^{\infty}\big\{\breve{e}_T(p+sh)-\breve{e}_T(p)-e(p+sh)+e(p)\big\}k'(s)\,\mathrm{d}s}_{\mathrm{I}}$$

$$\underbrace{-\sqrt{\frac{T}{h}}\big\{\breve{e}_T(p)-e(p)\big\}\int_{-\infty}^{\infty}k'(s)\,\mathrm{d}s}_{\mathrm{II}}\underbrace{-\sqrt{\frac{T}{h}}\int_{-\infty}^{\infty}e(p+sh)k'(s)\,\mathrm{d}s-\sqrt{Th}\alpha(p)}_{\mathrm{III}}$$

Term II is zero by the assumptions on the kernel. Further, using a standard kernel derivative bias expansion, term III is $e'''(p)\Xi/2+o(1)$. Finally, term I has by the continuous mapping theorem the same asymptotic distribution as $-h^{-1/2}\int_{-\infty}^{\infty}\big\{\mathbb{G}(p+sh)-\mathbb{G}(p)\big\}k'(s)\,\mathrm{d}s$, which converges to a normal distribution with variance

$$\lim_{h\to 0}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{1}{h}\big\{H(p+sh,p+\tilde{s}h)-H(p+sh,p)-H(p,p+\tilde{s}h)+H(p,p)\big\}k'(s)k'(\tilde{s})\,\mathrm{d}\tilde{s}\,\mathrm{d}s,$$

which equals $\mathscr{V}(p)$, as asserted. □

Proof of lemma 1. Suppose first that $\tilde{s}\geq s\geq 0$. Then,

$$\begin{cases} H(p+sh,p+\tilde{s}h)=\zeta(p+sh)\zeta(p+\tilde{s}h)\{p(1-p)+(1-p)sh-p\tilde{s}h-s\tilde{s}h^2\},\\ H(p+sh,p)=\zeta(p+sh)\zeta(p)\{p(1-p)-psh\},\\ H(p,p+\tilde{s}h)=\zeta(p+\tilde{s}h)\zeta(p)\{p(1-p)-p\tilde{s}h\},\\ H(p,p)=\zeta^2(p)p(1-p).\end{cases}$$

Thus, $\lim_{h\to 0}\mathscr{H}_h(p,s,\tilde{s})=\zeta^2(p)s$. Repeating the argument for the other cases yields $\lim_{h\to 0}\mathscr{H}_h(p,s,\tilde{s}) = \zeta^2(p)|\mathrm{Med}(0,s,\tilde{s})|$. Now,

$$\begin{aligned}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}&|\mathrm{Med}(0,s,\tilde{s})|k'(s)k'(\tilde{s})\,\mathrm{d}\tilde{s}\,\mathrm{d}s=\\ &-\int_{-\infty}^{0}\int_{-\infty}^{0}\max(\tilde{s},s)k'(s)k'(\tilde{s})\,\mathrm{d}\tilde{s}\,\mathrm{d}s+\int_{0}^{\infty}\int_{0}^{\infty}\min(\tilde{s},s)k'(s)k'(\tilde{s})\,\mathrm{d}\tilde{s}\,\mathrm{d}s=\\ &-2\int_{-\infty}^{0}k'(s)s\int_{-\infty}^{s}k'(\tilde{s})\,\mathrm{d}\tilde{s}\,\mathrm{d}s+2\int_{0}^{\infty}k'(s)s\int_{s}^{\infty}k'(\tilde{s})\,\mathrm{d}\tilde{s}\,\mathrm{d}s=-2\int_{-\infty}^{\infty}k'(s)sk(s)\,\mathrm{d}s=\kappa_2,\end{aligned}$$

where the last equality follows using integration by parts. □

Proof of lemma 2. Let $p>0$ and $z=p^{1/(n-1)}$. To obtain (18), simply note that $\zeta(p)=Q_c'(p)p=zQ'(z)/(n-$

1). Suppose first that $\tilde{s} \geq s \geq 0$. Then,[50]

$$
\begin{cases}
H(p+sh, p+\tilde{s}h) \simeq \dfrac{(n-1)z(1-z) + z^{2-n}sh - z^{3-n}sh - z^{3-n}\tilde{s}h}{n(n-1)}\{p^2 + p(s+\tilde{s})h\}\times \\
\qquad \{Q'^2(z) + Q''(z)Q'(z)z^{2-n}(s+\tilde{s})h/(n-1)\}, \\
H(p+sh, p) \simeq \dfrac{(n-1)z(1-z) - z^{3-n}sh}{n(n-1)}(p^2 + psh)\{Q'^2(z) + Q''(z)Q'(z)z^{2-n}sh/(n-1)\}, \\
H(p, p+\tilde{s}h) \simeq \dfrac{(n-1)z(1-z) - z^{3-n}\tilde{s}h}{n(n-1)}(p^2 + p\tilde{s}h)\{Q'^2(z) + Q''(z)Q'(z)z^{2-n}\tilde{s}h/(n-1)\}, \\
H(p,p) = \dfrac{(n-1)z(1-z)}{n(n-1)}p^2 Q'^2(z),
\end{cases}
$$

where $\simeq$ means that $o(h)$ terms are omitted. Thus, $\lim_{h\to 0} \mathscr{H}_h(p, s, \tilde{s}) = z^n Q'^2(z)s/\{n(n-1)\}$. Repeating the same argument for other $s, \tilde{s}, 0$ orderings we get $\lim_{h\to 0} \mathscr{H}_h(p, s, \tilde{s}) = z^n Q'^2(z)|\operatorname{Med}(s, \tilde{s}, 0)|/\{n(n-1)\}$. Now use the integration argument from the proof of lemma 1 to obtain the claimed result. ☐

Proof of lemma 3. The proof follows the same path as that of lemma 2 and is hence omitted. ☐

A.4 Boundary kernels

Proof of theorem 4. Consider $\bar{\alpha}_{T\psi}$. We have

$$
\begin{aligned}
\sqrt{Th}\{\bar{\alpha}_{T\psi}(p) - \alpha(p)\} &= \sqrt{Th}\left(\int_0^1 \psi'(s)\alpha_T(s)k_{\psi h}\left(\frac{\psi(p)-\psi(s)}{h} \,\Big\wr\, p\right) \mathrm{d}s - \alpha(p)\right) \\
&= \underbrace{\sqrt{\frac{T}{h}}\int_0^1 \psi'(s)\{\alpha_T(s) - \alpha(s)\}k_{\psi h}\left(\frac{\psi(p)-\psi(s)}{h} \,\Big\wr\, p\right) \mathrm{d}s}_{\text{I}} \\
&\qquad + \underbrace{\sqrt{\frac{T}{h}}\left(\int_0^1 \psi'(s)\alpha(s)k_{\psi h}\left(\frac{\psi(p)-\psi(s)}{h} \,\Big\wr\, p\right) \mathrm{d}s - \alpha(p)\right)}_{\text{II}}. \qquad (41)
\end{aligned}
$$

Recall that $\Psi(s) = \alpha\{\psi^{-1}(s)\}$. Then term II in (41) becomes by substitution of $s \leftarrow \{\psi(p) - \psi(s)\}/h$ and a standard kernel bias expansion,

$$
\begin{aligned}
\sqrt{Th}\left(\int_{\underline{v}_\psi}^{\bar{v}_\psi} \Psi\{\psi(p) + sh\}k_{\psi h}(-s \wr p)\,\mathrm{d}s - \alpha(p)\right) &= \frac{\Xi\Psi''\{\psi(p)\}}{2}\int_{\underline{v}_\psi}^{\bar{v}_\psi} s^2 k_{\psi h}(-s \wr p)\,\mathrm{d}s + o(1) \\
&= \frac{\Xi\Psi''\{\psi(p)\}}{2}\lim_{h\to 0}\frac{(\Omega_{\psi 0} + \Omega_{\psi 2})^2 - \Omega_{\psi 1}(3\Omega_{\psi 1} + \Omega_{\psi 3})}{\Omega_{\psi 0}^2 + \Omega_{\psi 0}\Omega_{\psi 2} - \Omega_{\psi 1}^2} + o(1). \qquad (42)
\end{aligned}
$$

[50]The expansions below are informal to reduce length, but we have verified that they obtain if they are conducted more rigorously.

The limit in the right hand side in (42) equals one for all $0 < p < 1$ and equals $(\pi - 4)/(\pi - 2)$ for $p = 0, 1$. Since $\Psi'' = (\alpha'')/(\psi'^2) - (\alpha'\psi'')/(\psi'^3)$, we get the asserted asymptotic bias.

Now term I in (41). Integration by parts produces

$$\underbrace{\sqrt{\frac{T}{h}}\psi'(1)\{\breve{e}_T(1) - e(1)\}k_{\psi h}\Big(\frac{\psi(p) - \psi(1)}{h}\,\wr\, p\Big)}_{\text{I}}$$
$$+\underbrace{\sqrt{\frac{T}{h^3}}\int_0^1 \psi'^2(s)\{\breve{e}_T(s) - e(s) - \breve{e}_T(p) + e(p)\}k'_{\psi h}\Big(\frac{\psi(p) - \psi(s)}{h}\,\wr\, p\Big)\,\mathrm{d}s}_{\text{II}}$$
$$+\underbrace{\sqrt{\frac{T}{h^3}}\{\breve{e}_T(p) - e(p)\}\int_0^1 \psi'^2(s)k'_{\psi h}\Big(\frac{\psi(p) - \psi(s)}{h}\,\wr\, p\Big)\,\mathrm{d}s}_{\text{III}}$$
$$-\underbrace{\sqrt{\frac{T}{h}}\int_0^1 \psi''(s)\{\breve{e}_T(s) - e(s) - \breve{e}_T(p) + e(p)\}k_{\psi h}\Big(\frac{\psi(p) - \psi(s)}{h}\,\wr\, p\Big)\,\mathrm{d}s}_{\text{IV}}$$
$$-\underbrace{\sqrt{\frac{T}{h}}\{\breve{e}_T(p) - e(p)\}\int_0^1 \psi''(s)k_{\psi h}\Big(\frac{\psi(p) - \psi(s)}{h}\,\wr\, p\Big)\,\mathrm{d}s}_{\text{V}} \quad (43)$$

Term I in (43) vanishes because $\breve{e}_T(1)$ converges faster than $\sqrt{T/h}$. Term III equals

$$\sqrt{\frac{T}{h}}\psi'(0)\{\breve{e}_T(p) - e(p)\}k_{\psi h}\Big(\frac{\psi(p) - \psi(0)}{h}\,\wr\, p\Big)$$
$$-\sqrt{\frac{T}{h}}\psi'(1)\{\breve{e}_T(p) - e(p)\}k_{\psi h}\Big(\frac{\psi(p) - \psi(1)}{h}\,\wr\, p\Big)$$
$$+\sqrt{\frac{T}{h}}\{\breve{e}_T(p) - e(p)\}\int_0^1 \psi''(s)k_{\psi h}\Big(\frac{\psi(p) - \psi(s)}{h}\,\wr\, p\Big)\,\mathrm{d}s. \quad (44)$$

For fixed $0 < p < 1$, (44) is $o_p(1)$ by the conditions on $k_{\psi h}$ and at $p = 1$, the superconsistency of $\breve{e}_T(p)$ takes care of the problem.

By a change of variables, terms IV and V in (43) are $o_p(1)$. Finally, term II in (43). Consider fixed $0 < p < 1$. Let $\Psi(p; sh) = \psi^{-1}\{\psi(p) + sh\}$. Substitute $s \leftarrow \{\psi(s) - \psi(p)\}/h$ to obtain

$$\sqrt{\frac{T}{h}}\int_{\underline{v}_\psi}^{\bar{v}_\psi} \psi'\{\Psi(p; sh)\}\Big(\breve{e}_T\big\{\Psi(p; sh)\big\} - e\big\{\Psi(p; sh)\big\} - \breve{e}_T(p) + e(p)\Big)k'_{\psi h}(-s\,\wr\, p)\,\mathrm{d}s$$
$$\simeq \frac{\psi'(p)}{\sqrt{h}}\int_{\underline{v}_\psi}^{\bar{v}_\psi}\Big\{\mathbb{G}\Big(p + \frac{sh}{\psi'(p)}\Big) - \mathbb{G}(p)\Big\}k'_{\psi h}(-s\,\wr\, p)\,\mathrm{d}s,$$

which has a limiting mean zero normal distribution with covariance kernel

$$\psi'^2(p) \lim_{h \to 0} \int_{\underline{\upsilon}_\psi}^{\bar{\upsilon}_\psi} \int_{\underline{\upsilon}_\psi}^{\bar{\upsilon}_\psi} k'_{\psi h}(-s \wr p) k'_{\psi h}(-\tilde{s} \wr p) \mathscr{H}_h\{p, s/\psi'(p), s'/\psi'(p)\} \, \mathrm{d}\tilde{s} \, \mathrm{d}s.$$

Note that $\underline{\upsilon}_\psi \to -\infty$, $\bar{\upsilon}_\psi \to \infty$, $k'_{\psi h} \to \phi'$ as $h \to 0$. □

Proof of lemma 4. We have

$$\bar{\alpha}_{T\psi}(p) = \frac{1}{h} \int_0^1 \psi'(s) \alpha_T(s) k_{\psi h}\Big(\frac{\psi(p) - \psi(s)}{h} \wr p\Big) \, \mathrm{d}s = \frac{1}{h} \sum_{j=1}^T \alpha_{Tj} \int_{\frac{j-1}{T}}^{\frac{j}{T}} \psi'(s) k_{\psi h}\Big(\frac{\psi(p) - \psi(s)}{h} \wr p\Big) \, \mathrm{d}s$$

$$= \sum_{j=1}^T \alpha_{Tj} \int_{\upsilon_{j-t}(p)}^{\upsilon_j(p)} k_{\psi h}(-s \wr p) \, \mathrm{d}s = \sum_{j=1}^T \alpha_{Tj} \Lambda_{\psi j}(p).$$

The second statement in lemma 4 is a natural property of the normal distribution. □

A.5 Derivatives

Lemma 9. *In the symmetric case,*

$$F_p(p) = p^{1/(n-1)}, \quad f_p(p) = \frac{F_p^{2-n}(p)}{n-1}, \quad f_p'(p) = \frac{2-n}{(n-1)^2} F_p^{3-2n}(p), \quad f_p''(p) = \frac{(2-n)(3-2n)}{(n-1)^2} F_p^{4-3n}(p).$$

Further,

$$\begin{cases} \alpha(p) = \dfrac{1}{n-1}[F_p(p)Q'\{F_p(p)\} + (n-1)Q\{F_p(p)\}], \\ \alpha'(p) = \dfrac{F_p^{2-n}}{(n-1)^2}(nQ' + F_pQ''), \\ \alpha''(p) = \dfrac{F_p^{3-2n}}{(n-1)^3}\{(2-n)nQ' + 3F_pQ'' + F_p^2Q'''\}, \\ \alpha'''(p) = \dfrac{F_p^{4-3n}}{(n-1)^4}\{(3-2n)(2-n)nQ' + (12-4n-n^2)F_pQ'' + (8-2n)F_p^2Q''' + F_p^3Q''''\}. \end{cases} \tag{45}$$

Proof. Trivial, but messy, calculus. □ □

Proof of theorem 5. Note that

$$\alpha'_{T\psi}(p) = \frac{\psi'(p)}{h^2} \int \psi'(s) \alpha_T(s) k'\left(\frac{\psi(p) - \psi(s)}{h}\right) \mathrm{d}s. \tag{46}$$

First, note that by integration by parts and substitution we have

$$\frac{\psi'(p)}{h^2}\int \psi'(s)\alpha(s)k'\left(\frac{\psi(p)-\psi(s)}{h}\right)\mathrm{d}s \simeq \frac{\psi'(p)}{h}\int \alpha'(s)k\left(\frac{\psi(p)-\psi(s)}{h}\right)\mathrm{d}s = \\ \psi'(p)\int \frac{\alpha'}{\psi'}\big[\psi^{-1}\{\psi(p)+sh\}\big]k(s)\,\mathrm{d}s \simeq \alpha'(p)+\frac{h^2}{2}\psi'(p)\Big[\frac{\alpha'}{\psi'}\{\psi^{-1}\}\Big]''\{\psi(p)\}\int k(s)s^2\,\mathrm{d}s,$$

which yields the asserted bias after noting that the Epanechnikov kernel is a density with variance $1/5$. Finally, skipping some steps that repeat steps taken in the proofs of earlier theorems and using integration by parts and substitution plus the properties of the Epanechnikov kernel,

$$\frac{\psi'(p)}{h^2}\int \psi'(s)\{\alpha_T(s)-\alpha(s)\}k'\left(\frac{\psi(p)-\psi(s)}{h}\right)\mathrm{d}s \simeq \\ \frac{\psi'(p)}{h^3}\int \{\psi'(s)\}^2\{\breve{e}_T(s)-e(s)-\breve{e}_T(p)+e(p)\}k''\left(\frac{\psi(p)-\psi(s)}{h}\right)\mathrm{d}s \\ \simeq \frac{\psi'^3(p)}{h^2}\int_{-1}^{1}\Big\{\breve{e}_T\Big(p+\frac{sh}{\psi'(p)}\Big)-e\Big(p+\frac{sh}{\psi'(p)}\Big)-\breve{e}_T(p)+e(p)\Big\}k''(-s)\,\mathrm{d}s \\ = -\frac{3\psi'^3(p)}{2h^2}\int_{-1}^{1}\Big\{\breve{e}_T\Big(p+\frac{sh}{\psi'(p)}\Big)-e\Big(p+\frac{sh}{\psi'(p)}\Big)-\breve{e}_T(p)+e(p)\Big\}\,\mathrm{d}s \\ \sim -\frac{3\psi'^3(p)}{2h^2\sqrt{T}}\int_{-1}^{1}\Big\{\mathbb{G}\Big(p+\frac{sh}{\psi'(p)}\Big)-\mathbb{G}(p)\Big\}\,\mathrm{d}s,$$

which has the asserted limit distribution. □

A.6 Distribution of win–probabilities

Proof of theorem 8. Note that

$$\sqrt{T}\big[G_T\{\hat{Q}_c(p)\}-G\{Q_c(p)\}\big] = \sqrt{T}\big[G_T\{Q_{cT}(p)\}-G_T\{Q_c(p)\}-G\{Q_{cT}(p)\}+G\{Q_c(p)\}\big] \\ +\sqrt{T}\big[G_T\{Q_c(p)\}-G\{Q_c(p)\}\big]+\sqrt{T}\big[G\{Q_{cT}(p)\}-G\{Q_c(p)\}\big]. \quad (47)$$

The first right hand side term in (47) is $o_p(1)$since $\sqrt{T}(G_T-G)$ converges weakly to a Gaussian process and Q_{cT} is consistent for Q_c. The third right hand side term expands as $g\{Q_c(p)\}\{Q_{cT}(p)-Q_c(p)\}$ plus terms of (uniformly) lesser order. Noting that Q_{cT} converges superconsistently at the boundaries, the stated result then follows from the independence of G_T and Q_{cT}. □

A.7 Derived objects

Proof of theorem 6. Using integration by parts we get

$$\sqrt{T}(\widehat{\mathrm{BS}}-\mathrm{BS}) = \sqrt{T}\int_0^1 \big[\{\alpha_T(p)-\alpha(p)\}p-\{\breve{e}_T(p)-e(p)\}\big]f_p(p)\,\mathrm{d}p =$$

$$\sqrt{T}\{\breve{e}_T(1)-e(1)\}f_p(1)-\sqrt{T}\int_0^1\{\breve{e}_T(p)-e(p)\}\{f_p'(p)p+2f_p(p)\}\,\mathrm{d}p=$$
$$\sqrt{T}\int_0^1\{\breve{e}_T(p)-e(p)\}\frac{n}{(n-1)^2}p^{(2-n)/(n-1)}\,\mathrm{d}p+o_p(1).$$

Apply theorem 1 and lemma 2. □

Lemma 10. $\alpha_T(1)=O_p(1)$.

Proof. Since $\mathbb{G}(1)=0$ a.s., we have that for any $C>\alpha(1)$,

$$\lim_{T\to\infty}\Pr\{\alpha_T(1)>3C\}=\lim_{T\to\infty}\Pr\{\alpha_T(1)-\alpha(1)>2C\}=$$
$$\lim_{T\to\infty}\Pr\bigl[T\{\breve{e}_T(1)-\breve{e}_T(1-1/T)-e(1)+e(1-1/T)\}+T\{e(1)-e(1-1/T)\}-\alpha(1)>2C\bigr]\leq$$
$$\lim_{T\to\infty}\Pr\bigl\{-\sqrt{T}\mathbb{G}(1-1/T)>C\bigr\}=\lim_{T\to\infty}\Phi\biggl(-\frac{C}{\sqrt{TH(1-1/T,1-1/T)}}\biggr)=\Phi(-cC),$$

for some $c<\infty$ independent of C. Take $C\to\infty$ to make the right hand side zero. □ □

Lemma 11. $\int_0^1\{\alpha_T(p)-\alpha(p)\}^2f_p(p)\,\mathrm{d}p=o_p(1)$.

Proof. We have uniform convergence of α_T by theorem 1 except at the boundaries. Since α_T,α are nondecreasing and nonnegative, we only have to worry about α_T near one. Now, let $I_m=\int_0^{\bar{p}}\{\alpha_T(p)-\alpha(p)\}^2f_p(p)\,\mathrm{d}p$ and $I_r=\int_{\bar{p}}^1\{\alpha_T(p)-\alpha(p)\}^2f_p(p)\,\mathrm{d}p$ for a $0<\bar{p}<1$ to be manipulated later. Now, for any $\epsilon>0$ and $0<C<\infty$,

$$\Pr(I_m+I_r>2\epsilon)\leq\Pr(I_m>\epsilon)+\Pr\{I_r>\epsilon,\alpha_T(1)\leq C\}+\Pr\{\alpha_T(1)>C\}\tag{48}$$

Take $C=\epsilon/(1-\bar{p})$. Then the second right hand side probability in (48) equals zero. Then take $\limsup_{T\to\infty}$ in (48), followed by $C\to\infty$ to obtain the stated result. □ □

Proof of theorem 9. Note that

$$\sqrt{T}(\widehat{\mathrm{BS}}-\mathrm{BS})=\underbrace{\sqrt{T}\int_0^1[\{\alpha_T(p)-\alpha(p)\}p-\{\breve{e}_T(p)-e(p)\}]f_p(p)\,\mathrm{d}p}_{\mathrm{I}}+$$
$$\underbrace{\int_0^1A(p)\,\mathrm{d}\mathbb{G}_{Tp}(p)}_{\mathrm{II}}+\underbrace{\int_0^1\{\alpha_T(p)-\alpha(p)\}\,\mathrm{d}\mathbb{G}_{Tp}(p)}_{\mathrm{III}}-\underbrace{\int_0^1\{\breve{e}_T(p)-e(p)\}\,\mathrm{d}\mathbb{G}_{Tp}(p)}_{\mathrm{IV}},\tag{49}$$

where $\mathbb{G}_{Tp}=\sqrt{T}(F_{pT}-F_p)$. First, note that $\mathbb{G}_{Tp}\rightsquigarrow\mathbb{G}_p$ by theorem 8. Thus, since the class of right–continuous step functions is Donsker and by lemma 11, van der Vaart (2000, lemma 19.24) implies that term

III in (49) is $o_p(1)$.[51] Further, term IV is $o_p(1)$ by theorem 8.

Now, term I in (49) is using integration by parts equal to

$$\sqrt{T}\{\breve{e}_T(1) - e(1)\} f_p(1) - \sqrt{T} \int_0^1 \{\breve{e}_T(p) - e(p)\}\{p f_p'(p) + 2 f_p(p)\}\,\mathrm{d}p. \tag{50}$$

Note that the first term in (50) is $o_p(1)$. Term II in (49) can likewise be written as

$$-\int_0^1 \alpha'(p) p \mathbb{G}_{Tp}(p)\,\mathrm{d}p. \tag{51}$$

Now, combining the above results with the proof of theorem 8, it follows that

$$\sqrt{T}(\widehat{\mathrm{BS}} - \mathrm{BS}) = -\int_0^1 \left\{p^2 f_p'(p) + 2p f_p(p) + \alpha'(p) p g\{Q_c(p)\}\right\} \sqrt{T}\{Q_{cT}(p) - Q_c(p)\}\,\mathrm{d}p \\ - \int_0^1 \alpha'(p) p \sqrt{T}[G_T\{Q_c(p)\} - G\{Q_c(p)\}]\,\mathrm{d}p + o_p(1),$$

which has a mean zero normal limit with variance $\mathscr{V}_{BS}^a$, where we have used the fact that G, G_c are estimated using different data such that G_T and Q_{cT} are independent.

To establish (34), consider $\int_0^1 \int_0^1 \Gamma_2(p)\Gamma_2(p^*) H_1\{Q_c(p), Q_c(p^*)\}\,\mathrm{d}p^*\,\mathrm{d}p$, which we now show to equal the first term in (34): showing that the remainder of (33) is equal to the second term in (34) follows the same path, but is messier.

Let $I_j = \int_0^{\bar{b}} \{G_c^2(b)/g_c(b)\}^j g(b)\,\mathrm{d}b$. First, use integration by parts to obtain

$$\int_0^1 \Gamma_2(p) G\{Q_c(p)\}\,\mathrm{d}p = Q_c'(1) - I_1. \tag{52}$$

Then, $\int_p^1 \Gamma_2(t)\,\mathrm{d}t = Q_c'(1) - Q_c'(p)p^2$, whence

$$\int_0^1 \int_0^1 \Gamma_2(p)\Gamma_2(p^*) G[Q_c\{\min(p, p^*)\}]\,\mathrm{d}p^*\,\mathrm{d}p = 2\int_0^1 \Gamma_2(p) G\{Q_c(p)\} \int_p^1 \Gamma_2(t)\,\mathrm{d}t\,\mathrm{d}p = \\ Q_c'^2(1) - 2Q_c'(1) I_1 + I_2. \tag{53}$$

Subtract the square of (52) from (53) to obtain $I_2 - I_1^2$, as promised. To see that (34) is in fact the semiparametric efficiency bound note that for any hypothetical parameter vector θ indexing g, g_c,

[51] Lemma 19.24 in van der Vaart (2000) is stated specifically for empirical processes, but its proof relies merely on continuity properties and the fact that F_{pT} is not an empirical distribution function is hence immaterial (F_{pT} is the empirical distribution function of estimated p's, not of the p's themselves).

$$
\begin{aligned}
\partial_\theta \mathrm{BS} = \partial_\theta \int_0^{\bar b} \frac{G_c^2(b) g(b)}{g_c(b)} \, \mathrm{d}b = \\
\int_0^{\bar b} \frac{G_c^2(b)}{g_c(b)} \partial_\theta g(b) \, \mathrm{d}b - \int_0^{\bar b} \frac{G_c^2(b) g(b)}{g_c^2(b)} \partial_\theta g_c(b) \, \mathrm{d}b - 2 \int_0^{\bar b} \int_0^b \frac{G_c(t) g(t)}{g_c(t)} \, \mathrm{d}t \partial_\theta g_c(b) \, \mathrm{d}b = \\
\mathbb{E}\Big(\frac{G_c^2(b)}{g_c(b)} \partial_\theta \log g(b) \Big) - \mathbb{E}\Big\{ \Big(\frac{G_c^2(b_c) g(b_c)}{g_c^2(b_c)} + 2 \int_0^{b_c} \frac{G_c(t) g(t)}{g_c(t)} \, \mathrm{d}t \Big) \partial_\theta \log g_c(b_c) \Big\} \\
= \mathbb{E}\Big\{ \Big(\frac{G_c^2(b)}{g_c(b)} - \frac{G_c^2(b_c) g(b_c)}{g_c^2(b_c)} - 2 \int_0^{b_c} \frac{G_c(t) g(t)}{g_c(t)} \, \mathrm{d}t \Big) \partial_\theta \log\{ g(b) g_c(b_c) \} \Big\},
\end{aligned}
$$

which yields the stated bound by the arguments in Newey (1990, page 106). We have ignored the possibility that the upper bound can depend on θ but that is irrelevant since the upper bound can be estimated at a rate faster than $\sqrt{T}$. □

Proof of theorem 10. The proof is largely a repeat of that of theorem 9. The main difference concerns

$$
\begin{aligned}
&\sqrt{T} \int_0^1 \big[\{ \hat\alpha_{T\psi}(p) - \alpha(p) \} p - \{ \hat e_{T\psi}(p) - e(p) \} \big] f_p(p) \, \mathrm{d}p \\
&= \sqrt{T} \{ \hat e_{T\psi}(1) - e(1) \} f_p(1) - \frac{\sqrt{T}}{h} \int_0^1 \int_{-\infty}^{\infty} \{ e(s) - e(p) \} \psi'(s) k \Big(\frac{\psi(p) - \psi(s)}{h} \Big) \, \mathrm{d}s p f_p'(p) \, \mathrm{d}p \\
&\quad - \frac{\sqrt{T}}{h} \int_0^1 \int_{-\infty}^{\infty} \{ \breve e_T(s) - e(s) \} \psi'(s) k \Big(\frac{\psi(p) - \psi(s)}{h} \Big) \, \mathrm{d}s p f_p'(p) \, \mathrm{d}p \\
&\quad = o_p(1) - \int_0^1 \sqrt{T} \{ \breve e_T(p) - e(p) \} p f_p'(p) \, \mathrm{d}p,
\end{aligned}
$$

where we have omitted a few steps entailing nothing more than substitution and simple expansions, including a nonparametric kernel bias expansion. □

Proof of theorem 7. Note that

$$
\begin{aligned}
\sqrt{T} \int_0^1 \{ \alpha_T(p) - \alpha(p) \} \, \mathrm{d}F_p(p) &= \sqrt{T} \{ \breve e_T(1) - e(1) \} f_p(1) - \sqrt{T} \int_0^1 \{ \breve e_T(p) - e(p) \} f_p'(p) \, \mathrm{d}p \\
&\xrightarrow{d} \int_0^1 \mathbb{G}(p) f_p'(p) \, \mathrm{d}p,
\end{aligned}
$$

by theorem 1. Note that $f_p'(p) = (2-n) p^{(3-2n)/(n-1)} / (n-1)^2$. □

Proof of theorem 11. The proof follows with minor adjustments by repeating the steps in the proof of theorem 9. □

B Technical results and alternative methods

This appendix contains further details, supplementary results, and variants of the estimators defined earlier. The results are organized in parallel with the structure of the main text.

B.1 Covariates

We assume the covariates are independent of the bidders' idiosyncratic valuations and enter each bidder's utility of winning in an additively separable manner. The bidders' equilibrium bids are therefore separable in $\beta(v_t)$ and x_t, as well. Let $b_{ct}^{(\tilde{\mu})} = b_{ct}^* - x_t\tilde{\mu}$, where b_{ct}^* is the observed maximum competing bid in auction t. Note that the above discussion implies that $b_{ct}^{(\mu)}$ is independent of x_t.

Suppose without loss of generality that the support of x_t includes 0 and 1 as extremities and that b_{ct} has support $[0, 1]$. We will only consider values $\tilde{\mu}$ in a $1/\sqrt{T}$ neighborhood of μ.

Thus, let $m_T = \tilde{\mu} - \mu = m/\sqrt{T}$ for a fixed m. We consider the case $m \geq 0$ where the case $m < 0$ follows analogously. Let O_t^* denote the t–th order statistic of $b_{ct}^{(\tilde{\mu})}$.[52]

Note that if $m = 0$ then $O_T^* - O_1^*$ must be between 0 and 1. We now show that if $m > 0$ then $O_T^* - O_1^* > 1$ with probability approaching one. Let G_c^*, g_c^* denote the distribution and density function of $b_{ct}^{(\tilde{\mu})}$. Thus,

$$\Pr(O_T^* - O_1^* \leq 1) = T(T-1)\int_{-m_T}^{1}\int_{u}^{\min(1+u,1)} g_c^*(y)\{G_c^*(y) - G_c^*(u)\}^{T-2} g_c^*(u)\,\mathrm{d}y\,\mathrm{d}u =$$

$$T\int_{-m_T}^{0}\{G_c^*(1+u) - G_c^*(u)\}^{T-1}g_c^*(u)\,\mathrm{d}u + T\int_0^1\{1 - G_c^*(u)\}^{T-1}g_c^*(u)\,\mathrm{d}u$$

$$= T\int_{-m_T}^{0}\{G_c^*(1+u) - G_c^*(u)\}^{T-1}g_c^*(u)\,\mathrm{d}u + \{1 - G_c^*(0)\}^T. \quad (54)$$

Note that for $b \geq 0$, $G_c^*(b) = \int_0^1 G_c(b + m_T x)\,\mathrm{d}F_x(x) \simeq G_c(b) + g_c(b)m_T\mathbb{E}x$. Thus, $\{1 - G_c^*(0)\}^T \simeq \left\{1 - g_c(0)m\mathbb{E}x/\sqrt{T}\right\}^T = o(1)$. Now consider the first right hand side term in (54). Consider $-m_T \leq u \leq 0$. Note that for $v = u/m_T$, we get

$$G_c^*(1 + vm_T) - G_c^*(vm_T) = \int_0^1\left[G_c\{1 + (v+x)m_T\} - G_c\{(v+x)m_T\}\right]\mathrm{d}F_x(x) =$$

$$\int_0^{-v} G_c\{1 + (v+x)m_T\}\,\mathrm{d}F(x) + \int_{-v}^1\left[1 - G_c\{(v+x)m_T\}\right]\mathrm{d}F_x(x) \simeq$$

$$1 + m_T g(1)\int_0^{-v}(v+x)\,\mathrm{d}F(x) - m_T g_c(0)\int_{-v}^1(v+x)\,\mathrm{d}F_x(x) \leq 1 - cm_T,$$

for a constant c independent of v. Consequently, the first right hand side term in (54) is $o(1)$, also.

[52] The first order statistic is the lowest value.

B.2 Convexity

Lemma 12. *The expected payment function is convex in the probability of winning the auction in any Bayes–Nash equilibrium of a feasible auction mechanism with independent private values.* □

Proof. Let v be bidder i's valuation and, by an abuse of notation, let $e(v)$ denote i's equilibrium expected payment when bidder i submits a bid equal to its equilibrium strategy evaluated at v, and let $p(v)$ denote i's associated probability of winning. Incentive compatibility of the direct revelation mechanism that implements the Bayes–Nash equilibrium implies $p(v+\epsilon)v - e(v+\epsilon) \le p(v)v - e(v)$ and $p(v)(v+\epsilon) - e(v) \le p(v+\epsilon)(v+\epsilon) - e(v+\epsilon)$ for all ϵ. Combining these inequalities, we have

$$e(v+\epsilon) - e(v) - [p(v+\epsilon) - p(v)]\epsilon \le [p(v+\epsilon) - p(v)]v \le e(v+\epsilon) - e(v)\,.$$

Letting $\epsilon > 0$ proves $p(v)$ is nondecreasing in v (see lemma 2 of Myerson, 1981). Letting ϵ tend to zero from above and below, we find that $e_+(v) - e_-(v) = [p_+(v) - p_-(v)]v$, where the subscripts denotes the limit of the function from the right (+) and left (−). In other words, e is discontinuous in v if and only if p is discontinuous. Moreover, the ratio of the discontinuous jumps is v. Monotonicity of $p(v)$ then implies the expected payment as function of p is convex wherever $e(p)$ is defined, although the graph of the expected payment function has "gaps" at any points of discontinuity in p, e.g. when the seller sets a binding reserve price. However, $p(v)$ is Riemann integrable (Myerson, 1981), so $p(v)$ is continuous almost everyone with respect to the Lebesgue measure.

□

B.3 Least squares

Lemma 13 characterizes the solution to the least squares problem as an intermediate step toward proving the solution is an isotonic regression function.

Lemma 13. *If e_T is piecewise linear in p then the minimizer of the least–squares criterion* (4) *among nondecreasing, nonnegative functions is a right–continuous step–function.* □

Proof. Because α has to be nondecreasing, the optimizer must be constant between $(t-1)/T$ and t/T. The minimizer is right–continuous with possible discontinuities at t/T, because the objective is minimized if the "jumps" in α coincide with discontinuities in the derivative of e_T. In other words, the value of α should be smaller anywhere to the left of the discontinuity in order to minimize the first integral in (4). Any jumps in α should be "timed" to take advantage of the negative contribution to the least–squares criterion that comes from the discontinuities in the derivative of e_T, as illustrated in figure 11. □

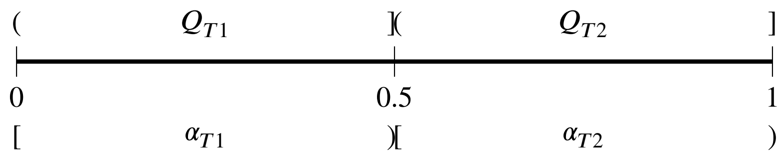

Figure 11: Illustration of the characterization of the α_{Tt}'s for $T = 2$.

An alternative way to arrive at an estimator α_T^* and thence an estimator $\breve{e}_T^*$ is by defining the problem as an inverse isotonic regression problem. The idea is to define

$$\Theta_T(\tilde{\alpha}) = \inf \operatorname*{argmin}_p \sum_{t=1}^{T} \left\{ e_T\left(\frac{t}{T}\right) - e_T\left(\frac{t-1}{T}\right) - \frac{\tilde{\alpha}}{T} \right\} \mathbb{1}\left(\frac{t-1}{T} \leq p\right), \tag{55}$$

The rationale for (55) is that (55) essentially imposes monotonicity of the derivative of e: it is the natural analog to inverse isotonic regression estimators for the current context.[53] An alternative way of thinking about it is that the population objective function corresponding to (55) is

$$\frac{1}{T} \sum_{t=1}^{T} \left[T\left\{ e\left(\frac{t}{T}\right) - e\left(\frac{t-1}{T}\right) \right\} - \tilde{\alpha} \right] \mathbb{1}\left(\frac{t-1}{T} \leq p\right) \simeq \frac{1}{T} \sum_{t=1}^{T} \left\{ \alpha\left(\frac{t-1}{T}\right) - \tilde{\alpha} \right\} \mathbb{1}\left(\frac{t-1}{T} \leq p\right),$$

which is optimized at the value of $p = (t-1)/T$ for which $\alpha\{(t-1)/T\}$ is the largest value less than $\tilde{\alpha}$.

Returning to (55), an estimator α_T^* can be defined as $\alpha_T^*(p) = \sup\{\tilde{\alpha} : \Theta_T(\tilde{\alpha}) \leq p\}$. There is no a priori reason to prefer α_T^* to α_T or vice versa, albeit that α_T may be easier to compute. In fact, they are numerically equivalent because $\Theta_T(\tilde{\alpha}) = \sup\{p : \alpha_T(p) < \tilde{\alpha}\}$.

B.4 NPMLE

To facilitate the numerical optimization of the maximum likelihood objective in (13), we use the fact that $\breve{e}_T^{\text{MLE}}$ is linear on the interval $[e_{(t-1)}/b_{(t-1)}, e_{(t)}/b_{(t)}]$ to express $e_{(t)}$ in terms of its derivative and $e_{(t-1)}$

$$e_{(t)} = e_{(t-1)} + \alpha_{(t)}\left(\frac{e_{(t)}}{b_{(t)}} - \frac{e_{(t-1)}}{b_{(t-1)}}\right) = e_{(t-1)} \frac{\alpha_{(t)}/b_{(t-1)} - 1}{\alpha_{(t)}/b_{(t)} - 1}.$$

Combining this recursive relationship with the constraint that $e_{(T)} = b_{(T)}$, we may write $e_{(t)}$ as

$$e_{(t)} = b_{(T)} \prod_{s=t+1}^{T} \frac{\alpha_{(s)}/b_{(s)} - 1}{\alpha_{(s)}/b_{(s-1)} - 1}, \tag{56}$$

[53] An inverse isotonic regression estimator can for a given m be characterized as a minimizer x of $\sum_{i=1}^{n} (y_i - m)\mathbb{1}(x_i \leq x)$.

where the product $\prod_{s=t+1}^{T} a_s$ is defined equal to one for $t = T$ for any sequence $\{a_s\}$. Using this expression to replace $e_{(t)}$ in (13), the loglikelihood becomes

$$\mathscr{L}(\tilde{\alpha}; b) = \sum_{t=1}^{T} \left(\log b_{(T)} + \sum_{s=t+1}^{T} \left\{ \log(\tilde{\alpha}_{(s)}/b_{(s)} - 1) - \log(\tilde{\alpha}_{(s)}/b_{(s-1)} - 1) \right\} - \log b_{(t)} - \log(\tilde{\alpha}_{(t)} - b_{(t)}) \right). \quad (57)$$

By inspection of the above display, the MLE must satisfy $\alpha_{(1)} = b_{(1)}$ and $\alpha_{(t)} \geq b_{(t)}$. Problematically, this implies an unbounded density at the lower end of the competitor bids' support, which in turn implies that the solution to the maximum likelihood problem is not unique, since the loglikelihood criterion is infinite for any α with $\alpha_{(1)} = b_{(1)}$ and $\alpha_{(2)} > b_2$. Nonetheless, one maximizer of the likelihood distinguishes itself from the rest because, for a fixed $\alpha_{(1)}$ and $\alpha_{(2)}$ with $b_{(1)} < \alpha_{(1)}$, $b_{(2)} < \alpha_{(2)}$ and $\alpha_{(2)} < 2\, b_{(3)} - b_{(2)}$, the solution for $\{\alpha_{(t)}\}$ for $t = 3, \ldots, T$ is unique. Furthermore, this unique solution does not depend on the values of $\alpha_{(1)}$ and $\alpha_{(2)}$ because the loglikelihood is additively separable in $\alpha_{(t)}$ and the monotonicity constraints on $\alpha_{(t)}$ do not bind for $t = 1$, 2, and 3. Thus, we may first maximize $\mathscr{L}(\tilde{\alpha}; b_1, \ldots, b_T) - \log(\tilde{\alpha}_{(1)} - b_{(1)})$ over $\{\tilde{\alpha}_{(t)} : t > 2\}$. We may then separately define $\breve{\alpha}_{T,(1)}^{\mathrm{MLE}} = b_{(1)}$ and choose any $\alpha_{(2)} \in (b_{(2)}, \breve{\alpha}_{T,(3)}^{\mathrm{MLE}}]$. Within this (shrinking) interval, the likelihood contribution of the second–lowest observed competitor bid is strictly decreasing in $\tilde{\alpha}_{(2)}$. In practice, we suggest defining the MLE equal to the boundary value $\breve{\alpha}_{T,(2)}^{\mathrm{MLE}} = b_{(2)}$.

B.4.1 Pooled–adjacent–violator algorithm (PAVA) for MLE

By adding and subtracting $\log b_{(s)}$ and $\log b_{(s-1)}$ and canceling terms in the inner summation of (57), we can rewrite the loglikelihood as

$$\begin{aligned}
&\sum_{t=1}^{T} \Bigg[\log b_{(T)} + \\
&\qquad \sum_{s=t+1}^{T} \left\{ \log(\tilde{\alpha}_{(s)} - b_{(s)}) - \log b_{(s)} - \log(\tilde{\alpha}_{(s)} - b_{(s-1)}) + \log b_{(s-1)} \right\} - \log b_{(t)} - \log(\tilde{\alpha}_{(t)} - b_{(t)}) \Bigg] \\
&\qquad\qquad = \sum_{t=1}^{T} \left[\sum_{s=t+1}^{T} \left\{ \log(\tilde{\alpha}_{(s)} - b_{(s)}) - \log(\tilde{\alpha}_{(s)} - b_{(s-1)}) \right\} - \log(\tilde{\alpha}_{(t)} - b_{(t)}) \right] \\
&\qquad\qquad\qquad = \sum_{t=1}^{T} \left\{ (t-2) \log(\tilde{\alpha}_{(t)} - b_{(t)}) - (t-1) \log(\tilde{\alpha}_{(t)} - b_{(t-1)}) \right\}.
\end{aligned}$$

The Karush–Kuhn–Tucker (KKT) conditions for the nonparametric maximum likelihood problem are

$$\begin{cases} \dfrac{t-2}{\tilde{\alpha}_{(t)} - b_{(t)}} - \dfrac{t-1}{\tilde{\alpha}_{(t)} - b_{(t-1)}} + \lambda_t - \lambda_{t+1} = 0, \\ \lambda_t \geq 0, \\ \tilde{\alpha}_{(t)} - \tilde{\alpha}_{(t-1)} \geq 0, \\ \lambda_t(\tilde{\alpha}_{(t)} - \tilde{\alpha}_{(t-1)}) = 0. \end{cases} \tag{58}$$

Let t_j be the subsequence of starting points of "blocks" for which the nondecreasing constraint binds. By construction, $\tilde{\alpha}_{(t_j-1)} < \tilde{\alpha}_{(t_j)} = \cdots = \tilde{\alpha}_{(t_{j+1}-1)} < \tilde{\alpha}_{(t_{j+1})}$. Complementary slackness then implies $\lambda_{t_j} = \lambda_{t_{j+1}} = 0$. Within each block j, the value $\tilde{\alpha}$ that satisfies the KKT conditions can then be found by solving for $\tilde{\alpha}$ in

$$\begin{aligned} 0 = \sum_{t=t_j}^{t_{j+1}-1} \left(\frac{t-2}{\tilde{\alpha} - b_{(t)}} - \frac{t-1}{\tilde{\alpha} - b_{(t-1)}} + \lambda_t - \lambda_{t+1} \right) &= \sum_{t=t_j}^{t_{j+1}-1} \left(\frac{t-2}{\tilde{\alpha} - b_{(t)}} - \frac{t-1}{\tilde{\alpha} - b_{(t-1)}} \right) \\ &= -\frac{t_j - 1}{\tilde{\alpha} - b_{t_j - 1}} - \frac{2}{\tilde{\alpha} - b_{(t_j)}} - \frac{2}{\tilde{\alpha} - b_{(t_j+1)}} - \cdots - \frac{2}{\tilde{\alpha} - b_{(t_{j+1}-2)}} + \frac{t_{j+1} - 3}{\tilde{\alpha} - b_{(t_{j+1}-1)}} \end{aligned}$$

To find the solution to the constrained NPMLE, we initially assign each $\tilde{\alpha}_t$ to its own block and set $\{\tilde{\alpha}_t\}$ equal to the unconstrained solution $\tilde{\alpha}_{(t)} = (t-1)b_{(t)} - (t-2)b_{(t-1)}$ for $t > 1$ and $\tilde{\alpha}_{(1)} = b_{(1)}$. This initial guess satisfies the constraint $\alpha_{(t)} \geq b_{(t)}$ but might not produce a monotonic sequence. Beginning with $t = 4$, the PAVA proceeds sequentially by finding the smallest t such that $\tilde{\alpha}_t < \tilde{\alpha}_{t-1}$. If such a t exists, we pool $\tilde{\alpha}_t$ together with the left adjacent block and recalculate $\tilde{\alpha}$ in the above first–order condition for that block. We then set $\tilde{\alpha}_{(s)} = \tilde{\alpha}$ for all s in the block and repeat until no further violations are found. This algorithm will converge in no more than $T - 3$ steps, because exactly one more of the $T - 3$ monotonicity constraints are made to bind with equality in each step and no constraints are ever made slack again.

Importantly, every iterate satisfies the dual feasibility KKT condition $\lambda_t \geq 0$ because violations of the primal feasibility condition $\tilde{\alpha}_{(t)} \geq \tilde{\alpha}_{(t-1)}$ are resolved by imposing $\tilde{\alpha}_{(t)} = \tilde{\alpha}_{(t-1)}$. Though primal feasibility may also be satisfied, for instance, by setting $\tilde{\alpha}_{(t)} = \tilde{\alpha}_{(t+1)}$, these deviations from the PAVA algorithm typically lead to a violation of dual feasibility unless the PAVA algorithm would have pooled these values in a later iteration. Thus, the final iterate of $\{\tilde{\alpha}_t\}$ will satisfy the KKT conditions. Lemma 8 formally establishes this claim in an appendix. Moreover, lemma 5 demonstrates that the KKT conditions are both necessary and sufficient for the constrained global maximum of the loglikelihood objective. Thus, the algorithm converges to the MLE for α.

B.5 Alternative smoothing methods

B.5.1 Alternative transformation

As an alternative method of applying a transformation ψ in conjunction with smoothing, one can define

$$\hat{\alpha}_{T\psi}(p) = \frac{\psi'(p)}{h} \int_{-\infty}^{\infty} \alpha_T(s) k\Big(\frac{\psi(p) - \psi(s)}{h}\Big)\, \mathrm{d}s, \tag{59}$$

which can also be boundary corrected using boundary kernels:

$$\hat{\alpha}_{T\psi}(p) = \frac{\psi'(p)}{h} \int_{0}^{1} \alpha_T(s) k_{\psi h}\Big(\frac{\psi(p) - \psi(s)}{h} \,\Big\vert\, p\Big)\, \mathrm{d}s. \tag{60}$$

Theorem 13. *Suppose that $k_{\psi h}$ is constructed as in lemma 14 and that assumptions A to E are satisfied. Then*

$$\forall p \in [0,1] : \sqrt{Th}\{\hat{\alpha}_{T\psi}(p) - \alpha(p)\} \xrightarrow{d} N\big\{\mathscr{B}_\psi(p), \mathscr{V}_\psi(p)\big\},$$

where for $0 < p < 1$, $\mathscr{B}_\psi(p) = \text{expression (66)} \times \Xi/2$ *and*
$\mathscr{V}_\psi(p) = \psi'^2(p) \lim_{h\to 0} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \phi'(s)\phi'(\tilde{s})\mathscr{H}_h\big\{p, s/\psi'(p), \tilde{s}/\psi'(p)\big\}\, \mathrm{d}\tilde{s}\, \mathrm{d}s$. *where $\mathscr{H}_h$ is as defined in (16). If (7) holds then we obtain the simpler expression $\mathscr{V}_\psi(p) = \kappa_2\zeta^2(p)\psi'(p)$. For $p \in \{0,1\}$, $\mathscr{B}_\psi, \mathscr{V}_\psi$ are finite.* □

Proof. Consider $\hat{\alpha}_{T\psi}$. We have

$$\begin{aligned}\sqrt{Th}\{\hat{\alpha}_{T\psi}(p) - \alpha(p)\} &= \sqrt{Th}\Big(\psi'(p)\int_0^1 \alpha_T(s) k_{\psi h}\Big(\frac{\psi(p) - \psi(s)}{h} \,\Big\vert\, p\Big)\, \mathrm{d}s - \alpha(p)\Big) \\ &= \underbrace{\sqrt{\frac{T}{h}}\psi'(p)\int_0^1 \{\alpha_T(s) - \alpha(s)\} k_{\psi h}\Big(\frac{\psi(p) - \psi(s)}{h} \,\Big\vert\, p\Big)\, \mathrm{d}s}_{\mathrm{I}} \\ &\quad + \underbrace{\sqrt{\frac{T}{h}}\Big(\psi'(p)\int_0^1 \alpha(s) k_{\psi h}\Big(\frac{\psi(p) - \psi(s)}{h} \,\Big\vert\, p\Big)\, \mathrm{d}s - \alpha(p)\Big)}_{\mathrm{II}}. \end{aligned} \tag{61}$$

Let $\Psi(s) = \alpha\{\psi^{-1}(s)\}/\psi'\{\psi^{-1}(s)\}$. Then term II in (61) becomes by substitution of $s \leftarrow \{\psi(p) - \psi(s)\}/h$,

$$\sqrt{Th}\Big(\psi'(p)\int_{\underline{\upsilon}_\psi}^{\bar{\upsilon}_\psi} \Psi\{\psi(p) + sh\} k_{\psi h}(-s \,\vert\, p)\, \mathrm{d}s - \alpha(p)\Big) = \frac{\Xi\psi'(p)\Psi''\{\psi(p)\}}{2} + o(1)$$

for all $0 < p < 1$ by a standard kernel bias expansion. For $p = 0, 1$ the asymptotic bias differs by a multiplicative constant. Note that $\psi'\Psi''$ equals (66), which produces the asserted asymptotic bias.

Now term I in (61). Integration by parts produces

$$\underbrace{\sqrt{\frac{T}{h}}\psi'(p)\{\breve{e}_T(1)-e(1)\}k_{\psi h}\Big(\frac{\psi(p)-\psi(1)}{h}\,\wr\, p\Big)}_{\text{I}} \\ +\underbrace{\sqrt{\frac{T}{h^3}}\psi'(p)\int_0^1 \psi'(s)\{\breve{e}_T(s)-e(s)-\breve{e}_T(p)+e(p)\}k'_{\psi h}\Big(\frac{\psi(p)-\psi(s)}{h}\,\wr\, p\Big)\,\mathrm{d}s}_{\text{II}} \\ +\underbrace{\sqrt{\frac{T}{h^3}}\psi'(p)\{\breve{e}_T(p)-e(p)\}\int_0^1 \psi'(s)k'_{\psi h}\Big(\frac{\psi(p)-\psi(s)}{h}\,\wr\, p\Big)\,\mathrm{d}s}_{\text{III}}. \quad (62)$$

Term I in (62) vanishes because $\breve{e}_T(1)$ converges at rate T. Term III equals

$$\sqrt{\frac{T}{h}}\psi'(p)\{\breve{e}_T(p)-e(p)\}k_{\psi h}\Big(\frac{\psi(p)-\psi(0)}{h}\,\wr\, p\Big)-\sqrt{\frac{T}{h}}\psi'(p)\{\breve{e}_T(p)-e(p)\}k_{\psi h}\Big(\frac{\psi(p)-\psi(1)}{h}\,\wr\, p\Big). \quad (63)$$

For fixed $0\le p\le 1$, (63) is $o_p(1)$ by the conditions on $k_{\psi h}$.

Finally, term II in (62). Consider fixed $0\le p\le 1$. Substitute $s\leftarrow\{\psi(s)-\psi(p)\}/h$ to obtain

$$\sqrt{\frac{T}{h}}\psi'(p)\int_{\underline{\upsilon}_\psi}^{\bar{\upsilon}_\psi}\Big(\breve{e}_T\big[\psi^{-1}\{\psi(p)+sh\}\big]-e\big[\psi^{-1}\{\psi(p)+sh\}\big]-\breve{e}_T(p)+e(p)\Big)k'_{\psi h}(-s\wr p)\,\mathrm{d}s \\ \simeq\frac{\psi'(p)}{\sqrt{h}}\int_{\underline{\upsilon}_\psi}^{\bar{\upsilon}_\psi}\Big(\mathbb{G}\Big(p+\frac{sh}{\psi'(p)}\Big)-\mathbb{G}(p)\Big)k'_{\psi h}(-s\wr p)\,\mathrm{d}s,$$

which by a tedious repetition of the arguments of theorem 3 has a limiting mean zero normal distribution with variance $\lim_{h\to 0}\psi'^2(p)\int_{\underline{\upsilon}_\psi}^{\bar{\upsilon}_\psi}\int_{\underline{\upsilon}_\psi}^{\bar{\upsilon}_\psi}\mathscr{H}_h(p,s/\psi'(p),\tilde{s}/\psi'(p))k'_{\psi h}(-\tilde{s}\wr p)\,\mathrm{d}\tilde{s}\;k'_{\psi h}(-s\wr p)\,\mathrm{d}s$, which under (7) simplifies to

$$\zeta^2(p)\psi'(p)\lim_{h\to 0}\Big(\bar{\upsilon}_{\psi h}k^2_{\psi h}(\bar{\upsilon}_{\psi h}\wr p)-\underline{\upsilon}_{\psi h}k^2_{\psi h}(\underline{\upsilon}_{\psi h}\wr p) \\ -k_{\psi h}(-\bar{\upsilon}_{\psi h}\wr p)\int_0^{\bar{\upsilon}_{\psi h}}k_{\psi h}(-s\wr p)\,\mathrm{d}s-k_{\psi h}(-\underline{\upsilon}_{\psi h}\wr p)\int_{\underline{\upsilon}_{\psi h}}^0 k_{\psi h}(-s\wr p)\,\mathrm{d}s+\int_{\underline{\upsilon}_{\psi h}}^{\bar{\upsilon}_{\psi h}}k^2_{\psi h}(-s\wr p)\,\mathrm{d}s\Big) \\ =\zeta^2(p)\psi'(p)\lim_{h\to 0}\int_{\underline{\upsilon}_{\psi h}}^{\bar{\upsilon}_{\psi h}}k^2_{\psi h}(-s\wr p)\,\mathrm{d}s, \quad (64)$$

as promised. For $0<p<1$, the right hand side in (64) reduces to $\zeta^2(p)\psi'(p)/\sqrt{\pi}$. □

To see how (59) solves the exploding bias near zero problem, consider the following. The reason we needed e to be three times boundedly differentiable in theorem 1 is that its proof contains a second order (bias) expansion of both $e(p+sh)-e(p)$ and $\alpha(p+sh)-\alpha(p)$: the former for $\breve{e}_T$, the latter for $\hat{\alpha}_T$. If one uses $\hat{e}_{T\psi}$ then the corresponding expansions become $e[\psi^{-1}\{\psi(p)+sh\}]-e(p)$ and $\psi'(p)\big(\alpha[\psi^{-1}\{\psi(p)+sh\}]-\alpha(p)\big)$. This makes all the difference since the second derivative of the first difference with respect to s evaluated at $s=0$ is

$$\frac{\alpha'(p)}{\psi'^2(p)}-\frac{\alpha(p)}{\psi'(p)}\frac{\psi''(p)}{\psi'^2(p)} \tag{65}$$

The corresponding expression for the second difference in the preceding paragraph is

$$\frac{\alpha''(p)}{\psi'^2(p)}-3\frac{\alpha'(p)}{\psi'(p)}\frac{\psi''(p)}{\psi'^2(p)}-\alpha(p)\frac{\psi'''(p)}{\psi'^3(p)}+3\alpha(p)\frac{\psi''^2(p)}{\psi'^4(p)}. \tag{66}$$

Note that the asymptotic bias in (66) can be made to equal zero by choosing $\psi=e$. Unfortunately, we do not know e, so making that choice is infeasible.[54]

The bias formula in (66) is somewhat complicated and a downside of the formula for $\hat{\alpha}_{T\psi}$ in (59) is that, depending on the choices of k,ψ, it may require numerical integration. This is an inconvenience more than a serious problem since α_T is piecewise constant. Nonetheless, $\bar{\alpha}_T$ does not suffer either of these drawbacks.

B.5.2 Alternative boundary kernels

In section 4.3.1 we present one boundary correction scheme that uses boundary kernels. The following lemmas illustrate how one can construct a boundary kernel from a Gaussian kernel.

Lemma 14. *Let ϕ,Φ be the standard normal density and distribution functions. Then $k_{\psi h}(s \wr p)=(\omega_{\psi 1}-\omega_{\psi 2}s)\phi(s)$ satisfies the requirements in (24) for*

$$\omega_{\psi 2}=\frac{\Omega_{\psi 1}}{\Omega_{\psi 0}^2+\Omega_{\psi 0}\Omega_{\psi 2}-\Omega_{\psi 1}^2}, \qquad \omega_{\psi 1}=\frac{\Omega_{\psi 0}+\Omega_{\psi 2}}{\Omega_{\psi 0}^2+\Omega_{\psi 0}\Omega_{\psi 2}-\Omega_{\psi 1}^2},$$

where $\Omega_{\psi j}=\Phi^{(j)}(\bar{\upsilon}_\psi)-\Phi^{(j)}(\underline{\upsilon}_\psi)$. □

Proof. Let $\mathscr{I}_j=\int_{\underline{\upsilon}_\psi}^{\bar{\upsilon}_\psi} s^j\phi(s)\,\mathrm{d}s$. Note that $\mathscr{I}_0=\Omega_{\psi 0}$, $\mathscr{I}_1=-\Omega_{\psi 1}$, $\mathscr{I}_2=\Omega_{\psi 2}+\Omega_{\psi 0}$. Thus, we need

$$\begin{cases} \omega_{\psi 1}\Omega_{\psi 0} - \omega_{\psi 2}\Omega_{\psi 1} = 1,\\ -\omega_{\psi 1}\Omega_{\psi 1} + \omega_{\psi 2}(\Omega_{\psi 0}+\Omega_{\psi 2}) = 0. \end{cases}$$

Solve for $\omega_{\psi 1},\omega_{\psi 2}$. □

[54]Recall that choosing $\psi=\alpha$ was infeasible for $\bar{\alpha}_{T\psi}$.

Lemma 15. *Let $\mathscr{I}_j$ be defined as in the proof of lemma 14 and let $k_{\psi h}(s \mid p) = \sum_{j=1}^{3} s^{j-1}\omega_{\psi j}\phi(s)$, where*

$$\begin{bmatrix} \omega_{\psi 1} \\ \omega_{\psi 2} \\ \omega_{\psi 3} \end{bmatrix} = \begin{bmatrix} \mathscr{I}_0 & \mathscr{I}_1 & \mathscr{I}_2 \\ \mathscr{I}_1 & \mathscr{I}_2 & \mathscr{I}_3 \\ \mathscr{I}_2 & \mathscr{I}_3 & \mathscr{I}_4 \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$$

Then $k_{\psi h}$ satisfies the requirements of (24) everywhere, including at the boundaries. □

Proof. Trivial. □

In theorem 4 the kernel used was taken to be the kernel constructed in lemma 14. This is inessential. Indeed, the results go through with ϕ replaced with a second–order kernel k if $k_{\psi h}$ is chosen as $k_{\psi h}(s \mid p) = (\omega_{\psi k1} - \omega_{\psi k2} s)k(s)$ where $\omega_{\psi k1}$ and $\omega_{\psi k2}$

$$\omega_{\psi k1} = \frac{\Omega_{\psi k2}}{\Omega_{\psi k0}\Omega_{\psi k2} - \Omega_{\psi k1}^2}, \quad \omega_{\psi k2} = \frac{-\omega_{\psi k1}\Omega_{\psi k1}}{\Omega_{\psi k2}},$$

with $\Omega_{\psi kj} = \int_{\underline{\upsilon}_\psi}^{\bar{\upsilon}_\psi} u^j k(-u)\,\mathrm{d}u$.

B.5.3 Another boundary correction

A second way of implementing boundary corrections is to create artificial values of $\alpha_T(p)$ for p outside $[0,1]$. Our approach is loosely motivated by the KZ method for kernel estimators, but it is a bit cleaner because of our specific circumstances: we are trying to smooth out an existing estimator which means that we already have values of $\alpha_T(p)$ between zero and one.

Here, we restrict k to be the Epanechnikov kernel  which is a quadratic on $[-1,1]$; indeed it is $3(1-x^2)/4$.[55] Consequently, any boundary correction procedure will be immaterial if the distance between $\psi(p)$ and $\psi(1), \psi(0)$ exceeds h. We focus on correcting estimates near the upper bound, $p = 1$. Impose the scale and location normalizations $\psi(1) = 0$ and $\psi'(1) = 1$.

Define

$$\alpha(1+s) = \alpha\big[1 - \rho\big\{\psi(1+s)\big\}\big]\rho'\big\{\psi(1+s)\big\}, \quad s > 0, \tag{67}$$

where $\rho(s) = s + ds^2 + \{d^2 - \psi''(1)d/6\}s^3$, with $d = \alpha'(1)/\alpha(1)$. Then it is straightforward but unpleasant to verify that the thus extended version of α is twice continuously differentiable at one. We extend α_T analogously to (67) using a suitable estimator $\hat{d}$ in lieu of d, defining $\hat{\rho}$ to be like ρ but with $\hat{d}$ replacing d.

[55] Earlier, we had taken $\int k(s)s^2\,\mathrm{d}s$ to equal one, which is not true for an Epanechnikov kernel. We adjust the asymptotic bias expression accordingly.

We can then obtain a smoothed estimate of α by defining

$$\bar{\alpha}^R_{T\psi}(p) = \frac{1}{h}\int_{-\infty}^{\infty} \alpha_T(s)\psi'(s)k\Big(\frac{\psi(p)-\psi(s)}{h}\Big)\,\mathrm{d}s, \tag{68}$$

where the superscript R stands for 'reflection.' As noted, away from the boundary, the behavior of $\bar{\alpha}^R_{T\psi}$ is no different than that of the estimator without boundary bias correction. So we only analyze its behavior in an h–neighborhood of the boundary, as formulated in theorem 14.

Theorem 14. *Let (i) assumptions A to D be satisfied; (ii) k be the Epanechnikov kernel; (iii)* ψ *be twice continuously differentiable at 1 with* $\psi(1) = 0$ *and* $\psi'(1) = 1$*; (iv)* $\hat{d}$ *converge to d at a rate no slower than* $\sqrt[5]{T}$*. Then,*

$$\sqrt{Th}\{\bar{\alpha}^R_{T\psi}(1-th) - \alpha(1-th)\} + \frac{\sqrt{Th^3}}{8}\alpha(1)(1-t)^3(t+3)(\hat{d}-d) \xrightarrow{d} N\big(\mathscr{B}^R(t), \mathscr{V}^R(t)\big), \tag{69}$$

where $\mathscr{B}^R(t) = \{\alpha''(1) - \alpha'(1)\psi''(1)\}\Xi/10$ *and* $\mathscr{V}^R(t) = \lim_{h\to 0}\int_0^{1+t}\int_0^{1+t} \mathring{k}_t(s)\mathring{k}_t(\tilde{s})\mathscr{H}_h(1,-s,-\tilde{s})\,\mathrm{d}s\,\mathrm{d}\tilde{s}$*, where* $\mathring{k}_t(s) = (3/2)\{(t-s)\mathbb{1}(1-t \le s \le 1+t) - 2s\mathbb{1}(0 \le s \le 1-t)\}$ *and* $\mathscr{H}_h$ *was defined in (16). If (7) holds then we obtain the simpler expression* $\mathscr{V}^R(t) = 3\zeta^2(1)\big\{2 - t^2(t^3 - 5t + 5)\big\}/5$*.* □

Below, let ϵ_{dT} denote the convergence rate of $\hat{d}$.

Lemma 16. *Suppose that for* $0 \le \hat{a}, a \le 1$*,* $\hat{a} - a = O_p(\epsilon_T)$*. Then*

$$\breve{e}_T(\hat{a}) - e(a) = \breve{e}_T(a) - e(a) + e(\hat{a}) - e(a) + O_p\Big(\sqrt{(-\epsilon_T \log \epsilon_T)/T}\Big).$$

Proof. This is just a rearrangement of $\breve{e}_T(\hat{a}) - e(\hat{a}) - \breve{e}_T(a) + e(a) = O_p\Big(\sqrt{(-\epsilon_T \log \epsilon_T)/T}\Big)$, which is Levy's modulus of continuity theorem. □ □

Lemma 17. *Let* ψ *satisfy condition (iii) in Theorem 14. For any* $0 \le C < \infty$*,* $\sup_{0\le s\le C}|\hat{\rho}\{\psi(1+sh)\} - \rho\{\psi(1+sh)\}| = O_p(h^2\epsilon_{dT})$*.*

Proof. Follows immediately from writing out noting that $\hat{\rho}, \rho$ are third degree polynomials and noting that $\psi(1) = 0$. □ □

Lemma 18. *Let* ψ *satisfy condition (iii) in Theorem 14 and* ρ *defined as in Appendix B.5.3. For any* $0 \le C < \infty$*,* $\sup_{0\le s\le C}|\rho\{\psi(1+sh)\} - sh| + \sup_{0\le s\le C}|\hat{\rho}\{\psi(1+sh)\} - sh| = O_p(h^2)$*.*

Proof. Follows from the mean value theorem and the fact that $\hat{d}$ is bounded in probability. □ □

Lemma 19.

$$\breve{e}_T[1 - \hat{\rho}\{\psi(1 + sh)\}] - e(1 - sh) =$$
$$\{\breve{e}_T(1 - sh) - e(1 - sh)\} + \{e[1 - \hat{\rho}\{\psi(1 + sh)\}] - e(1 - sh)\} + o_p(\sqrt{h/T}).$$

Proof. Follows directly from lemmas 16 to 18. □ □

Lemma 20. *For any* $0 \le C < \infty$,

$$\sup_{0 \le t \le C} \left| \breve{e}_T(1 + th) - 2\breve{e}_T(1) + \breve{e}_T(1 - th) - e(1 + th) + 2e(1) - e(1 - th) - \mathring{\mathscr{B}}_T^R(t) \right| = o_p\big(\sqrt{h/T}\big),$$

where $\mathring{\mathscr{B}}_T^R(t) = 0$ *for* $t \le 0$ *and for* $t > 0$ *it is*

$$\mathring{\mathscr{B}}_T^R(t) = \frac{e[1 - \hat{\rho}\{\psi(1 + th)\}] - e[1 - \rho\{\psi(1 + th)\}]}{\psi'(1 + th)} -$$
$$h \int_0^t \Big(e[1 - \hat{\rho}\{\psi(1 + sh)\}] - e[1 - \rho\{\psi(1 + sh)\}] \Big) \frac{\psi''(1 + sh)}{\psi'^2(1 + sh)} \, \mathrm{d}s.$$

Proof. Using integration by parts we get

$$\breve{e}_T(1 + th) = \breve{e}_T(1) + \int_0^{th} \alpha_T(1 + s) \, \mathrm{d}s =$$
$$\breve{e}_T(1) + \int_0^{th} \alpha_T\big[1 - \hat{\rho}\{\psi(1 + s)\}\big] \hat{\rho}'\{1 + \psi(1 + s)\} \, \mathrm{d}s =$$
$$2\breve{e}_T(1) - \frac{\breve{e}_T[1 - \hat{\rho}\{\psi(1 + th)\}]}{\psi'(1 + th)} - h \int_0^t \breve{e}_T[1 - \hat{\rho}\{\psi(1 + sh)\}] \frac{\psi''(1 + sh)}{\psi'^2(1 + sh)} \, \mathrm{d}s. \quad (70)$$

Now, by lemma 19 we have

$$\breve{e}_T[1 - \hat{\rho}\{\psi(1 + sh)\}] = \breve{e}_T(1 - sh) - e(1 - sh) + e[1 - \hat{\rho}\{\psi(1 + sh)\}] + o_p\big(\sqrt{h/T}\big),$$

uniformly in $0 \le s \le C$. Thus, (70) is

$$2\breve{e}_T(1) - \{\breve{e}_T(1 - th) - e(1 - th)\} - \frac{e[1 - \hat{\rho}\{\psi(1 + th)\}]}{\psi'(1 + th)} -$$
$$h \int_0^t e[1 - \hat{\rho}\{\psi(1 + sh)\}] \frac{\psi''(1 + sh)}{\psi'^2(1 + sh)} \, \mathrm{d}s + o_p\big(\sqrt{h/T}\big). \quad (71)$$

Repeat (70) for e in lieu of $\breve{e}_T$ and subtract from (71). □

□

Lemma 21. *For* $\mathring{\mathscr{B}}_T^R$ *defined in lemma 20 and any* $0 \le C < \infty$, $\sup_{0 \le s \le C} |\mathring{\mathscr{B}}_T^R(s) + \alpha(1)(\hat{d} - d)s^2h^2| = o_p(h^2\epsilon_{dT} + h^3)$.

Proof. Note that by the mean value theorem and the definitions of $\rho, \hat{\rho}, \psi$,

$$e[1 - \hat{\rho}\{\psi(1 + sh)\}] - e[1 - \rho\{\psi(1 + sh)\}] = -\alpha(1)\{\hat{\rho}''(0) - \rho''(0)\}\frac{s^2h^2}{2} + o_p(h^3).$$

Further, $\rho''(0) = 2d$ and $\hat{\rho}''(0) = 2\hat{d}$. The stated result then follows from the fact that $\hat{d} - d = O_p(\epsilon_{dT})$. □

□

Lemma 22. *For $\mathring{\mathscr{B}}_T^R$ defined in lemma 20 and any $0 \le C < \infty$, $\sup_{0\le s\le C}|\mathring{\mathscr{B}}_T^R(s)| = O_p(h^2\epsilon_{dT}) + o_p(h^3)$.*

Proof. This is an immediate consequence of lemma 21. □ □

Lemma 23. *Uniformly in $0 \le t \le 1$,*[56]

$$\frac{1}{h}\int_{-\infty}^{\infty} k\Big(\frac{\psi(1-th) - \psi(s)}{h}\Big)\psi'(s)\{\alpha_T(s) - \alpha(s)\}\,\mathrm{d}s \simeq \\ -\frac{h}{8}\alpha(1)(1-t)^3(t+3)(\hat{d} - d) + \frac{1}{h}\int_t^1 k'(s)[\breve{e}_T\{1-(s-t)h\} - e\{1-(s-t)h\}]\,\mathrm{d}s \\ -\frac{1}{h}\int_{-1}^{t} k'(s)[\breve{e}_T\{1+(s-t)h\} - e\{1+(s-t)h\}]\,\mathrm{d}s, \quad (72)$$

where $\simeq$ means that any omitted terms are asymptotically negligible.

Proof. The left hand side in (72) is by integration by parts equal to

$$\frac{1}{h^2}\int_{-\infty}^{\infty} k'\left(\frac{\psi(1-th) - \psi(s)}{h}\right)\psi'^2(s)\{\breve{e}_T(s) - e(s)\}\,\mathrm{d}s \\ -\frac{1}{h}\int_{-\infty}^{\infty} k\left(\frac{\psi(1-th) - \psi(s)}{h}\right)\psi''(s)\{\breve{e}_T(s) - e(s)\}\,\mathrm{d}s. \quad (73)$$

The first term in (73) dominates the second term, so we deal with the first term only. The first term in (73) is for $\varsigma_{ts}(h) = \psi^{-1}\{\psi(1-th) + sh\}$ equal to

$$-\frac{1}{h}\int_{-1}^{1} k'(s)\psi'\{\varsigma_{ts}(h)\}\big(\breve{e}_T\{\varsigma_{ts}(h)\} - e\{\varsigma_{ts}(h)\}\big)\,\mathrm{d}s \\ \simeq -\frac{1}{h}\int_{-1}^{1} k'(s)\psi'\{\varsigma_{ts}(0) + \varsigma_{ts}'(0)h\}\big(\breve{e}_T\{\varsigma_{ts}(0) + \varsigma_{ts}'(0)h\} - e\{\varsigma_{ts}(0) + \varsigma_{ts}'(0)h\}\big)\,\mathrm{d}s \\ \simeq -\frac{1}{h}\int_{-1}^{1} k'(s)[\breve{e}_T\{1+(s-t)h\} - e\{1+(s-t)h\}]\,\mathrm{d}s. \quad (74)$$

Since $\breve{e}_T(1)$ is a super–consistent estimator of $e(1)$, we get by lemma 20 that

[56] For $t > 1$, we get standard asymptotics.

$$-\frac{1}{h}\int_t^1 k'(s)[\breve{e}_T\{1+(s-t)h\}-e\{1+(s-t)h\}]\,\mathrm{d}s \simeq$$

$$-\frac{1}{h}\int_t^1 k'(s)\Big(\mathring{\mathscr{B}}_T^R(s-t)-[\breve{e}_T\{1-(s-t)h\}-e\{1-(s-t)h\}]\Big)\,\mathrm{d}s \overset{\text{lemma 21}}{\simeq}$$

$$h\alpha(1)(\hat{d}-d)\int_t^1 k'(s)(s-t)^2\,\mathrm{d}s+\frac{1}{h}\int_t^1 k'(s)[\breve{e}_T\{1-(s-t)h\}-e\{1-(s-t)h\}]\,\mathrm{d}s. \quad (75)$$

Since k is the Epanechnikov kernel, the right hand side in (75) simplifies to

$$-\frac{h}{8}\alpha(1)(1-t)^3(t+3)(\hat{d}-d)+\frac{1}{h}\int_t^1 k'(s)[\breve{e}_T\{1-(s-t)h\}-e\{1-(s-t)h\}]\,\mathrm{d}s. \qquad \square$$

Lemma 24. *Uniformly in* $0\le t\le C$ *for given* $0<C<\infty$,

$$\frac{1}{h}\int_{-\infty}^{\infty} k\left(\frac{\psi(1-th)-\psi(s)}{h}\right)\psi'(s)\{\alpha(s)-\alpha(1-th)\}\,\mathrm{d}s=\{\alpha''(1)-\alpha'(1)\psi''(1)\}\frac{h^2}{10}+o(h^2).$$

Proof. Follows directly from a standard kernel bias expansion followed by an application of the mean value theorem, noting that $\int_{-1}^1 k(s)s^2\,\mathrm{d}s=1/5$. $\square$ $\square$

Proof of theorem 14. The asymptotic bias is derived in lemma 24. The second term in (69) corresponds to the first right hand side term in (72). Thus, the only issue remaining is to show that $\sqrt{Th}$ times the sum of the second and third terms in (72) have a zero mean limiting normal distribution with variance equal to $\mathscr{V}^R(t)$.

Now, using the shorthand $\Upsilon_T=\breve{e}_T-e$ and noting that k is the Epanechnikov kernel, the second right hand side term in (72) can be written as

$$-\frac{3}{2h}\int_t^1 s\Upsilon_T\{1-(s-t)h\}\,\mathrm{d}s=-\frac{3}{2h}\int_0^{1-t}(s+t)\Upsilon_T(1-sh)\,\mathrm{d}s. \quad (76)$$

The last right hand side term in (72) is then

$$\frac{3}{2h}\int_{-1}^t s\Upsilon_T\{1+(s-t)h\}\,\mathrm{d}s=\frac{3}{2h}\int_0^{1+t}(t-s)\Upsilon_T(1-sh)\,\mathrm{d}s. \quad (77)$$

Summing (76) and (77) yields

$$-\frac{3}{2h}\int_0^{1-t}2s\Upsilon_T(1-sh)\,\mathrm{d}s+\frac{3}{2h}\int_{1-t}^{1+t}(t-s)\Upsilon_T(1-sh)\,\mathrm{d}s. \quad (78)$$

Recall that $\mathring{k}_t(s)=(3/2)\{(t-s)\mathbb{1}(1-t\le s\le 1+t)-2s\mathbb{1}(0\le s\le 1-t)\}$. Then (78) becomes $h^{-1}\int_0^{1+t}\mathring{k}_t(s)\Upsilon_T(1-sh)\,\mathrm{d}s$, which leads to the asserted limit distribution using techniques similar to the ones employed in proofs of e.g. theorem 4 above.

Now the simplification of $\mathscr{V}^R(t)$ if H^* has the indicated form. Note first that $\mathscr{H}_h(1,-s,-\tilde{s}) = \zeta^2(1)\min(s,\tilde{s})$. Hence

$$\mathscr{V}^R(t) = \zeta^2(1)\int_0^{1+t}\int_0^{1+t} \mathring{k}_t(s)\mathring{k}_t(\tilde{s})\min(s,\tilde{s})\,\mathrm{d}\tilde{s}\,\mathrm{d}s =$$
$$2\zeta^2(1)\Bigl(\int_0^{1-t}\mathring{k}_t(s)s\int_s^{1-t}\mathring{k}_t(\tilde{s})\,\mathrm{d}\tilde{s}\,\mathrm{d}s + \int_0^{1-t}\mathring{k}_t(s)s\int_{1-t}^{1+t}\mathring{k}_t(\tilde{s})\,\mathrm{d}\tilde{s}\,\mathrm{d}s + \int_{1-t}^{1+t}\mathring{k}_t(s)s\int_s^{1+t}\mathring{k}_t(\tilde{s})\,\mathrm{d}\tilde{s}\,\mathrm{d}s\Bigr). \quad (79)$$

Now,

$$\begin{cases} \int_s^{1+t}\mathring{k}_t(\tilde{s})\,\mathrm{d}\tilde{s} = 3\{(t-s)^2-1\}/4, & \text{if } 1-t \le s \le 1+t, \\ \int_s^{1-t}\mathring{k}_t(\tilde{s})\,\mathrm{d}\tilde{s} = 3\{s^2-(1-t)^2\}/2, & \text{if } 0 \le s \le 1-t, \end{cases}$$

which implies that (79) equals

$$2\zeta^2(1)\Bigl(\frac{3}{5}(1-t)^5 + 3t(1-t)^4 + \frac{3}{10}t^2(-9t^3+30t^2-35t+15)\Bigr) = \frac{3}{5}\zeta^2(1)\bigl\{2-t^2(t^3-5t+5)\bigr\},$$

as claimed. □

As noted, the conditions on ψ are normalizations: without them $\psi(1), \psi'(1)$ would pop up in various places. Our assumption of the Epanechnikov kernel is not essential but the proofs do make use of the fact that the kernel has bounded support. Moreover, the polynomial portion of the second term in (69) would be more complicated.

Since $\hat{d}$ is essentially a nonparametric kernel derivative estimator, achieving a $T^{1/5}$ rate is feasible under assumption D.[57] If one assumes Q_c to have one more derivative at 1 then α is thrice differentiable at 1, which would imply that picking a bandwidth h_d for $\hat{d}$ that converges faster than $T^{-1/10}$ and slower than $T^{-1/5}$ would make the second term in (69) disappear: $h_d \sim T^{-1/7}$ would be optimal.

So, here we advocate picking a bandwidth for $\hat{d}$ which tends to zero more slowly than $T^{-1/5}$ whereas KZ advocates making the bandwidth go to zero faster than $T^{-1/5}$. In a separate note (Pinkse and Schurter, 2019) we show that there is a bug in both Karunamuni and Alberts (2005) and KZ and that there one needs to assume the existence of an extra derivative and choose a bandwidth that converges more slowly in order to obtain their claimed results.

Near the left boundary, we apply an analogous reflection method based upon

$$\alpha(s) = \alpha\Bigl[\rho_0\Bigl(\frac{\psi(0)-\psi(s)}{\psi'(0)}\Bigr)\Bigr]\rho_0'\Bigl(\frac{\psi(0)-\psi(s)}{\psi'(0)}\Bigr),$$

[57] If the function whose derivative is estimated is twice differentiable then it is well–known that the bias is $O(h_d)$ and the variance $O(1/Th_d^3)$, where h_d is the bandwidth used for the estimation of d. Here, α is the function whose derivative is to be estimated, which is twice differentiable under assumption D since $\alpha'' = Q_c'''p + 3Q_c''$.

where $\rho_0(s) = s - d_0 s^2 + \left\{d_0^2 - d_0\psi''(0)/[6\psi'(0)]\right\}s^3$ and $d_0 = \alpha'(0)/\alpha(0)$. The formula is messier simply because we had already normalized the location and scale of ψ at $p = 1$ to simplify the expressions near the right boundary.

B.5.4 Estimating α' near the boundary

The following result provides an example of how one can estimate the derivative of α local to the boundary.

Theorem 15. *Let (i) assumptions A to C be satisfied; (ii) Q_c be four times continuously differentiable on any compact subset of $(0,1)$; (iii) k be the Epanechnikov kernel; (iv) ψ be thrice continuously differentiable at 1 with $\psi(1) = 0$ and $\psi'(1) = 1$; (v) $\hat{d} - d = O_p(T^{-2/5})$; (vi) $\lim_{T\to\infty} \sqrt{Th^7} = \Xi_d < \infty$. Then for any $0 \le t \le 1$,*

$$\sqrt{Th^3}\{\bar{\alpha}^R_{T\psi}{}'(1-th) - \alpha'(1-th)\} - \sqrt{\frac{T}{h}}\frac{\alpha(1)}{2}(1-t)^3(\hat{d}-d) \xrightarrow{d} N\left(\mathscr{B}^{Rd}(t), \mathscr{V}^{Rd}(t)\right),$$

where

$$\mathscr{B}^{Rd}(t) = \frac{\Xi_d}{80}\Big(8\left\{\alpha'''_\uparrow(1) + 3\alpha'(1)\psi''^2(1) - \alpha'(1)\psi'''(1) - 3\alpha''(1)\psi''(1)\right\} + (4+t)(1-t)^4\{\alpha'''_\uparrow(1) - \alpha'''_\downarrow(1)\}\Big)$$

with $\alpha'''_\uparrow$, $\alpha'''_\downarrow$ denoting left and right derivatives,and

$$\mathscr{V}^{Rd}(t) = \frac{9}{4}\lim_{h\downarrow 0}\int_{1-t}^{1+t}\int_{1-t}^{1+t}\mathscr{H}_h(1,-s,-\tilde{s})\,\mathrm{d}\tilde{s}\,\mathrm{d}s.$$

If (7) holds then the asymptotic variance simplifies to $\mathscr{V}^{Rd}(t) = 3\zeta^2(1)t^2(3-t)$. □

Lemma 25. *Uniformly in $0 \le t \le 1$,*

$$\frac{1}{h^2}\int_{-\infty}^{\infty} k'\Big(\frac{\psi(1-th)-\psi(s)}{h}\Big)\psi'(s)\{\alpha_T(s) - \alpha(s)\}\,\mathrm{d}s \simeq \frac{\alpha(1)}{2}(1-t)^3(\hat{d}-d) + \frac{3}{2h^2}\int_t^1[\check{e}_T\{1-(s-t)h\} - e\{1-(s-t)h\}]\,\mathrm{d}s - \frac{3}{2h^2}\int_{-1}^t[\check{e}_T\{1+(s-t)h\} - e\{1+(s-t)h\}]\,\mathrm{d}s, \quad (80)$$

Proof. The line of proof is the same as lemma 23 but with k' instead of k, noting that k'' is constant whereas k' is odd. □ □

Lemma 26. *Uniformly in $0 \le t \le C$ for given $0 < C < \infty$,*

$$\frac{\psi'(1-th)}{h^2}\int_{-\infty}^{\infty} k'\Big(\frac{\psi(1-th)-\psi(s)}{h}\Big)\psi'(s)\alpha(s)\,\mathrm{d}s = o(h^2) + \alpha'(1-th) + \frac{h^2}{80}\Big(8\big\{\alpha_{\uparrow}'''(1) + 3\alpha'(1)\psi''^2(1) - \alpha'(1)\psi'''(1) - 3\alpha''(1)\psi''(1)\big\} + (4+t)(1-t)^4\{\alpha_{\uparrow}'''(1) - \alpha_{\downarrow}'''(1)\}\Big), \quad (81)$$

where $\alpha_{\uparrow}''', \alpha_{\downarrow}'''$ denote the third left and right derivatives, respectively.

Proof. Let $z_{th}(s) = \psi^{-1}\{\psi(1-th)+sh\}$. Then $z_{th}(0) = 1-th$, $z_{th}'(0) = h/\psi'(1-th)$, $z_{th}''(0) = -h^2\psi''(1-th)/\psi'^3(1-th)$. The left hand side in (81) is

$$\frac{\psi'(1-th)}{h}\int_{-\infty}^{\infty} k\Big(\frac{\psi(1-th)-\psi(s)}{h}\Big)\alpha'(s)\,\mathrm{d}s = \psi'\{z_{th}(0)\}\int_{-1}^{1} k(s)\frac{\alpha'\{z_{th}(s)\}}{\psi'\{z_{th}(s)\}}\,\mathrm{d}s. \qquad (82)$$

Now, for $|s| \le 1$ we have, uniformly in s,

$$\psi'\{z_{th}(0)\}\frac{\alpha'\{z_{th}(s)\}}{\psi'\{z_{th}(s)\}} - \alpha'\{z_{th}(0)\} = o(h^2) + hs\Big(\frac{\alpha''}{\psi'} - \frac{\alpha'\psi''}{\psi'^2}\Big) + \frac{h^2s^2}{2}\big\{\alpha_{\uparrow}'''(1) - 3\alpha''\psi'' - \alpha'\psi''' + 3\alpha'\psi''^2\big\} + \mathbb{1}(s>t)\frac{h^2(s-t)^2}{2}\{\alpha_{\downarrow}'''(1) - \alpha_{\uparrow}'''(1)\},$$

where all omitted arguments of the α, ψ functions are $1-th$. Hence the right hand side in (82) is

$$o(h^2) + \alpha'(1-th) + \frac{h^2}{2}\kappa_2^*(\alpha_{\uparrow}''' + 3\alpha'\psi''^2 - \alpha'\psi''' - 3\alpha''\psi'') + \frac{h^2}{2}(\alpha_{\downarrow}''' - \alpha_{\uparrow}''')\int_t^1 k(s)(s-t)^2\,\mathrm{d}s,$$

where the α's and ψ's are evaluated at 1. Finally, observe that for the Epanechnikov kernel, $\kappa_2^* = 1/5$ and,

$$\int_t^1 k(s)(s-t)^2\,\mathrm{d}s = \frac{(4+t)(1-t)^4}{40}. \qquad \square$$

Proof of theorem 15. Lemma 26 provides the formula for the asymptotic bias. For the asymptotic distribution, we start from lemma 25. Take Υ_T to have the same meaning as in the proof of theorem 14. Note that the sum of the last two terms in (80) equals

$$\frac{3}{2h^2}\int_0^{1-t}\Upsilon_T(1-sh)\,\mathrm{d}s - \frac{3}{2h^2}\int_0^{1+t}\Upsilon_T(1-sh)\,\mathrm{d}s = \frac{3}{2h^2}\int_{1-t}^{1+t}\Upsilon_T(1-sh)\,\mathrm{d}s,$$

which produces the promised asymptotic distribution.

Under (7) the asymptotic variance simplifies to

$$\frac{9}{4}\zeta^2(1)\int_{1-t}^{1+t}\int_{1-t}^{1+t}\min(s,\tilde{s})\,\mathrm{d}\tilde{s}\,\mathrm{d}s = \frac{9}{2}\zeta^2(1)\int_{1-t}^{1+t} s\int_{s}^{1+t}\mathrm{d}\tilde{s}\,\mathrm{d}s = \frac{9}{2}\zeta^2(1)\int_{1-t}^{1+t} s(1+t-s)\,\mathrm{d}s$$
$$= \frac{3}{4}\zeta^2(1)\big[3(1+t)\{(1+t)^2-(1-t)^2\} - 2\{(1+t)^3-(1-t)^3\}\big] = 3\zeta^2(1)t^2(3-t),$$

as asserted. □

Although the asymptotic variances are formulated differently, the asymptotic distributions in the two theorems coincide if one takes $t = 1$ in theorem 15. Indeed, if $t = 1$ then the correction via $\hat{d}$ becomes immaterial since there is no boundary bias concern then. Note that if $h \sim T^{-1/7}$ then the convergence rate is still $T^{2/7}$ irrespective of the value of t.

A perhaps puzzling finding is that the asymptotic variance is zero if $t = 0$. However, note that this is not the asymptotic variance of $\bar{\alpha}_T^{R\prime}(1)$ itself. Indeed, the (variation in the) asymptotic distribution of $\bar{\alpha}_T^{R\prime}(1)$ is then determined by the estimation of d. To get the asymptotic distribution of $\bar{\alpha}_T^{R\prime}(1)$ itself requires us to commit to a specific estimator of $\hat{d}$ and derive the joint distribution. This is neither difficult nor interesting.

B.5.5 Preserving monotonicity

One caveat to our boundary kernel estimators and 'reflection' procedure is that they can undo monotonicity near the boundaries in finite samples, although for different reasons. The boundary kernels are nonpositive near the boundary and are therefore capable of producing nonmonotonic estimates when α_T is relatively flat near the boundary. On the other hand, the transformation–and–reflection procedure in (68) continuously extends α and its first two derivatives such that $\alpha(1+s)$ is generally decreasing in s for large enough $s > 0$. Indeed, this is inevitable when α' is close to zero and α'' is negative. In any case, we may easily remedy this by redefining the smoothed estimator for α as the "cumulative maximum" of the objects defined in (60), (23), and (68), for example $\bar{\alpha}_{T\psi}(p) = \max\big\{[\psi'(p)/h]\int_0^1 \alpha_T(s)k_{\psi h}\{(\psi(p)-\psi(s))/h \wr p\}\,\mathrm{d}s,\ \sup_{q<p}\hat{\alpha}_{T\psi}(q)\big\}$.

Alternatively, in the case of the transformation-and-reflection procedure, we may apply this monotonization device to the definition of the extended α_T. The kernel–smoothed estimator of the resulting monotonic function will then be increasing on [0, 1] because k is a nonnegative kernel. Such a procedure will continuously extend α' and α'' at one, but may introduce a discontinuity in α'' at a point $p > 1$ for which $\alpha'(p) = 0$. We tolerate this discontinuity, however, because $d > 0$ and a finite α'' imply that the discontinuity is at a location bounded away from one. As a result, theorem 14 does not require any modification.

B.6 Jackknife estimators

Theorem 1 and lemma 1 motivate still more estimators of α. Note that $\alpha(p) = Q_c'(p)p + Q_c(p) = \zeta(p) + Q_c(p)$. Since Q_c can be estimated at a rate of $\sqrt{T}$, its estimation is of secondary concern. But $\zeta(p)$ enters the variance formulas in theorem 1 and lemma 1.

We will assume for the purpose of this discussion that G_c is estimated using the empirical distribution function of the maximum rival bid, such that $H(p, p^*) = \zeta(p)\zeta(p^*)\{\min(p, p^*) - pp^*\}$ and the conditions of lemma 1 are satisfied.

We present two versions, one based on theorem 1 and one on lemma 1:

$$\begin{cases} \breve{\alpha}_{TJ}(p) = \sqrt{\dfrac{(T-1)\sum_{t=1}^{T}\{\breve{e}_T(1) - \breve{e}_T(p) - \breve{e}_{T,-t}(1) + \breve{e}_{T,-t}(p)\}^2}{p(1-p)}} + \hat{Q}_c(p), \\ \hat{\alpha}_{TJ}(p) = \sqrt{\dfrac{h(T-1)\sum_{t=1}^{T}\{\hat{\alpha}_T(p) - \hat{\alpha}_{T,-t}(p)\}^2}{\kappa_2}} + \hat{Q}_c(p), \end{cases}$$

where the $-t$ subscripts denote leave–one–out estimators, i.e. the identical estimator without using observation t. Note that $\breve{\alpha}_{TJ}$ is only defined on $0 < p < 1$ albeit that it can be defined to equal zero at zero and one. This is precisely the reason for having $\breve{e}_T(1) - \breve{e}_{T,-t}(1)$ in the numerator even though it could be left out without affecting the result for fixed $0 < p < 1$.[58]

We inserted a generic estimator $\hat{Q}_c$ into the definitions of $\breve{\alpha}_{TJ}, \hat{\alpha}_{TJ}$. Its form is largely immaterial, but natural choices would be respectively $\breve{e}_T(p)/p$ and $\hat{e}_T(p)/p$ for $p > 0$ and zero for $p = 0$.

There are three downsides to the use of these jackknife estimators. The first issue is that in their current incarnation it is assumed that H^* has a specific form. But the formulas can be generalized or derived for other forms of H^*. Second, the jackknife estimators are costlier to compute since each estimator has to be computed $T + 1$ times. This may be of little practical relevance since computation of $\hat{\alpha}_T$ is fast. Finally, the jackknife estimators are not guaranteed to be monotonic. This is a property they share with other estimators, including GPV, and which can be addressed by the use of a monotonization procedure, which is not difficult but admittedly cumbersome.[59] We do not study the asymptotic properties of jackknife estimators in this paper.

C Other mechanisms

The general estimation strategy is based on the convexity of the expected payment function in the Bayes–Nash equilibrium of a trading mechanism. Because the convexity of the expected payment function does not depend on the mechanism—it is a consequence of the incentive compatibility condition of the direct revelation mechanism that implements a Bayes–Nash equilibrium of the trading mechanism—the approach generalizes beyond first–price auctions. The asymptotic results in this paper are specific to first–price auctions, but they may be adapted in natural ways to other auctions or auction–like settings. A trivial example is the second–price auction, in which a bidder's expected payment to the seller is $e(p) = \int_0^p Q_c(q)\,\mathrm{d}q$ when players adopt undominated strategies. Thus, if a bidder optimally bids $b = Q_c(p)$ so as to win with a probability of p

[58] $\sqrt{T}\{\breve{e}_T(1) - e(1)\} = o_p(1)$.

[59] See Ma19b for a monotonization procedure of the GPV estimator.

in equilibrium, its private valuation must be $v = b = Q_c(p) = e'(p)$. Of course, we do not need to estimate the expected payment function and its derivative in order to rediscover the bidder's weakly dominant strategy.

In an all–pay auction, our approach suggests estimating the inverse strategy function with the reciprocal of Grenander estimator for the density of the highest competing bid (Grenander, 1956a,b). To see this, note that the equilibrium expected payment function is $e(p) = Q_c(p)$. Depending on which bids are obervable, various unconstrained estimators for the competing quantile function can be computed. Our estimator for the inverse strategy function, i.e. the slope of the greatest convex minorant of the unconstrained estimator for the expected payment function, is equal to the reciprocal of the slope of the least concave majorant of the competing bid distribution.

Another example is the k–double auction, in which a buyer and a seller simultaneously submit price offers b and s, respectively. If the seller's offer is less than the buyer's, trade occurs at a price of $kb + (1 - k)s$; otherwise, there is no transaction. In a given equilibrium of this game, the buyer's expected payment function is therefore $e_B(p) = kpQ_s(p)+(1-k)\int_0^p Q_s(q)\,\mathrm{d}q$, where Q_s is the quantile function of the seller's equilibrium offer and p is the probability of trade. When $k = 1$, the price is solely determined by the seller's offer and the buyer's offer merely determines whether trade occurs. Accordingly, the expected payment function resembles that of the second–price auction and convexity of the expected payment function will be satisfied whenever e_B is estimated by substituting a nondecreasing estimate of Q_s into the previous formula. When $k = 0$, the buyer's decision problem resembles its choice of p in a first–price auction; while for k between 0 and 1, the expected payment function is a convex combination of the two. From the seller's perspective, the relevant object is the expected revenue $-e_S(p) = k\int_{1-p}^1 Q_B(q)\,\mathrm{d}q + (1 - k)pQ_b(1 - p)$, which must be convex in equilibrium.

The example of the k–double auction also serves to highlight the difference between our approach and the general identification strategy of Larsen and Zhang (2018). Larsen and Zhang illustrate how one can estimate the seller's valuations from data that only include the seller's offer, the transaction price, and an indicator for whether trade occurred. The buyer's offer and the parameter k are unobserved. Whereas their approach is agnostic about the mechanism that generated these data, the methods in this paper would take advantage of the econometrician's prior knowledge of k and use these same data to estimate the buyer's inverse strategy function, not the seller's valuations.