PARIS SCHOOL OF ECONOMICS
ÉCOLE D'ÉCONOMIE DE PARIS

WORKING PAPER N° 2008 - 64

The econometrics of auctions with asymmetric anonymous bidders

Laurent Lamy

JEL Codes: C14, D44
Keywords: Auctions, nonparametric identification, nonparametric estimation, tests for symmetry, unobserved heterogeneity, anonymous bids, uniform convergence rate

PARIS-JOURDAN SCIENCES ECONOMIQUES
LABORATOIRE D'ECONOMIE APPLIQUÉE - INRA

48, BD JOURDAN – E.N.S. – 75014 PARIS
TÉL. : 33(0) 1 43 13 63 00 – FAX : 33 (0) 1 43 13 63 10
www.pse.ens.fr

CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE – ÉCOLE DES HAUTES ÉTUDES EN SCIENCES SOCIALES
ÉCOLE NATIONALE DES PONTS ET CHAUSSÉES – ÉCOLE NORMALE SUPÉRIEURE

The Econometrics of Auctions with Asymmetric Anonymous Bidders *,†

Laurent Lamy‡

Abstract

We consider standard auction models when bidders' identities are not -or partially- observed by the econometrician. First, we adapt the definition of identifiability to a framework with anonymous bids and we explore the extent to which anonymity reduces the possibility to identify private value auction models. Second, in the asymmetric independent private value model which is nonparametrically identified, we generalize Guerre, Perrigne and Vuong's estimation procedure [Optimal Nonparametric Estimation of First-Price Auctions, Econometrica 68 (2000) 525-574] and study the asymptotic properties of our multi-step kernel-based estimator. Third a test for symmetry is proposed. Monte Carlo simulations illustrate the practical relevance of our estimation and testing procedures for small data sets.

Keywords: Auctions, nonparametric identification, nonparametric estimation, tests for symmetry, unobserved heterogeneity, anonymous bids, uniform convergence rate

JEL classification: C14, D44

---

*I would like to thank Philippe Février for very helpful discussions at the early stages of this research strated at CREST-LEI and Bernard Salanié who was the advisor of this part of my PhD. I am also grateful to Han Hong, Isabelle Perrigne, Quang Vuong, Frank Wolak and seminar participants at Stanford Econometric Seminar, at Stanford Econometric Lunch Workshop and at CREST-LEI lunch seminar for stimulating discussions. Finally I am grateful to Véronique Flambard for providing the data for my numerical application. All errors are mine. This paper is based on Chapter V of my Ph.D. dissertation. Part of this research was done while the author was visiting Stanford University.

†The supplementary material of this paper is available upon request.

‡PSE, 48 Bd Jourdan 75014 Paris. e-mail: laurent.lamy@pse.ens.fr

1 Introduction

Motivated by the fact that the identities of the bidders are lacking to the econometrician in some auction data either because this information is confidential or have been lost, or because submissions are structurally anonymous as in internet auctions, we consider a setup where bidders' identities are not -or partially- observed by the econometrician. At first glance, anonymity reduces considerably the scope of the economic analysis and invites the econometrician to assume that bidders are ex ante symmetric.[1] Furthermore, the presence of asymmetries has been the key determinant of many empirical studies of auction data. In Porter and Zona [32, 33] and Pesendorfer [31], the bidding behavior of alleged cartel participants is compared to the ones of non-cartel bidders through reduced form approaches. In Hendricks and Porter [12], neighbor firms are shown to be better informed in auctions for drainage leases. The aim of this paper is to lay the foundations of the econometric of auctions under anonymous data and to show how we can deal with asymmetric models. We adopt the so-called structural approach without any parametric assumptions (see Paarsch and Hong [30]) and focus on private value single-unit auction models.

First, we adapt the definition of identifiability to a framework with anonymous bids by requiring the unique characterization of bidders' primitives up to a permutation of bidders' identities. Then, in the spirit of Laffont and Vuong [16] we explore the extent to which anonymity reduces the possibility to identify private value models in standard auctions with risk neutral buyers. We show in Proposition 3.1 that anonymity prevents the identification of the asymmetric affiliated private value model, contrary to Campo et al. [7]'s analysis when bidders' identities are observed by the econometrician. When the identities of the bidders are not observed, the method that is currently implemented is to assume symmetry as an identifying restriction and to develop Guerre, Perrigne and Vuong [10]'s nonparametric methodology (henceforth GPV). The validity of this method relies on the assumption that bidders are symmetric, an assumption that can not be rejected on any testable restriction without further restrictions if bids are fully anonymous. However, for auction models that explicitly involve asymmetries -e.g. with collusion or shill bidding- or if

[1] See Song [36] and Sailer [34] for eBay auction models with symmetric bidders. Thus those models exclude any shill bidding activity from the seller, a pervasive phenomenon that is analyzed in Lamy [17, 18] respectively for models with pure private values and participation costs and models with interdependent values.

the econometrician knows that the main feature of the underlying market is asymmetries between bidders, this identification route is not appropriate. We propose another identification route. We show in Proposition 3.1 that the asymmetric independent private value (IPV) model is identified. One crucial step in the resolution of this inverse problem is to recover the underlying cumulative distribution functions (CDFs) $(F_{\mathbf{B}_{\mathbf{i}}^*})_{i=1,\cdots,N}$ of each buyers' bids from the CDFs $(F_{\mathbf{B}_p})_{p=1,\cdots,N}$ of the order statistics of the bids. By exploiting independence, the vector of the N bidders CDFs $(F_{\mathbf{B}_{\mathbf{i}}^*})_{i=1,\cdots,N}$ corresponds to the roots of a polynomial of degree N whose coefficients are linear combinations of the CDFs $(F_{\mathbf{B}_p})_{p=1,\cdots,N}$.

Second, we propose a multi-step kernel-based estimation procedure to recover the underlying distributions of bidders' private values. We mainly adapt GPV's nonparametric two-stage estimation procedure.[2] We establish the uniform consistency of our estimator. In the first price auction, this latter reaches the same rate of convergence as the one derived in GPV with nonanonymous bids and that was shown to attain the best rate of uniform convergence for estimating the latent density of private values from observed bids in the symmetric IPV model. In the second price auction, our estimator also reaches the optimal rate of uniform convergence under nonanonymous bids. Our estimation procedure is also tailored to setups where the econometrician may benefit from some additional information as the identity of the winner or the identities of the two highest bidders, e.g. in timber auction data (Li and Perrigne [22], Baldwin et al. [3]). In those latter cases, we know from Athey and Haile [2] that the asymmetric IPV model is identified only through the observation of the highest bid and the identity of the highest bidder. Nevertheless, the existing nonparametric methodology generalizing GPV and that only uses the highest bidding statistics may not perform very well in small data sets. In particular, in the second stage of GPV's estimation procedure, the pseudo-values are computed only for those bids that are not anonymous in such a 'naive' approach. On the contrary, our estimation procedure uses the complete vector of bids at both stages. In particular, we obtain for each bid a pseudo private value according to each possible identities of the bidder.

[2]Our nonparametric estimator can also be useful with regards to parametric procedures, i.e. that specify parametric families of distributions and solve by brute force a maximization program, insofar as it provides a consistent initial point for the maximization. Moreover, the EM-algorithm flavor of our multi-step procedure can be adapted in parametric frameworks -as for maximum likelihood estimation- and thus alleviate the computational burden. See McLachlan and Krishnan [26] for a comprehensive treatment of EM-algorithms.

Then, to estimate the distribution of private values, we should estimate for each bid the probability that it comes from a given bidder.

Third we consider how to test for symmetry. In a first step, we present the theoretical restrictions corresponding to symmetry between some bidders: we apply recent research in applied mathematics [41] which gives an explicit criterion for the determination of the number and multiplicities of the real/imaginary roots of a polynomial based on a Complete Discrimination System, i.e. a set of explicit expressions in terms of the coefficients of the corresponding polynomial. This system provides a basic framework in order to test for any symmetry structure between the bidders. In a second step, we propose a statistical test to test the assumption that all bidders are ex-ante symmetric against the alternative that at least two bidders are not symmetric. The asymptotic theory is developed: our testing statistic converges in Root-N to a Gaussian distribution which is centered at zero (at a strictly positive real number) under the null (resp. under the alternative). We analyze how the power of the test is related to the underlying asymmetry in the data.

In a nutshell, we face two identification routes with fully anonymous data: either to assume symmetry and to apply GPV's method allowing for correlated signals as in Li et al. [23] or to assume independence but not symmetry and to apply ours. Furthermore, with partially anonymous data, our methodology competes with nonparametric alternatives that also assumes independence, in particular 'naive' approaches that throw away the bids that are anonymous. Contrary to those latter approaches, our procedure exploits all bids and also the partial information about bidders' identities. As it is strongly supported by our Monte Carlo simulations, our procedure is a striking improvement, especially for small data sets where 'naive' approaches are useless.

With respect to the econometric literature, our contribution is severalfold. First, whereas Athey and Haile [2] consider nonparametric identification with incomplete sets of bids -which is structurally the case in some auction formats as the Dutch and English auctions, we go further by considering that the observation with respect to a bid itself may also be incomplete insofar as the identity of the bidder may lack to the econometrician. Our analysis is also tailored to the case where the econometrician does not observe a leading discrete covariate that is creating asymmetry between bidders.[3] Second, we propose a nonparametric estimation procedure that corresponds to a natural extension of GPV's procedure and analyze its asymptotic properties according to the same criteria as in GPV. Third a test for symmetry is proposed, which is distribution-free in the case without covariates. The anonymous nature of the data prevents the use of standard symmetry tests (see Hollander and Wolfe [13]).[4] Finally this work can be viewed as belonging to the general problem of unobserved heterogeneity in econometrics. The bulk of the existing works are considering models where a single outcome suffers from two kind of noises: a standard idiosyncatric noise and a noise which corresponds to some underlying unobserved heterogeneity among the individuals and that can receive some direct interpretation. Identification is obtained usually from the combination of some parametric specifications and/or additivity structure as in finite mixture distributions (see Titterington [38]) or in the mixed proportional hazard model (see van den Berg [39]). In the present contribution, the key element for the identification of the unobserved heterogeneity is the observation of multiple outcomes. In this vein, Li and Vuong [24] consider a deconvolution problem with multiple indicators without assuming any parametric assumption on the underlying (continuous) noises in an additive error model which has been applied in the empirical auction literature by Krasnokutskaya [15]. Our model is of a different nature: we impose no restriction on the distribution of the idiosyncratic types conditionally on the unobserved heterogeneity (e.g. no additivity structure is required), however the unobserved heterogeneity is of a discrete nature, the fundamental point which drives identification with multiple indicators.

The paper is organized as follows. In Section 2, we introduce the model and the definition of identifiability under anonymity. In Section 3, we consider nonparametric identification. For the asymmetric IPV model which is identified and allowing for heterogeneity across auctions and variations in the set of participants, we develop a multi-step kernel-based estimator in section 4 where the new caveats resulting from anonymity are presented. In section 5, general testing principles are presented and the statistical theory of a test for full symmetry is derived. In section 6, we establish the asymptotic properties of our estimation procedure. Section 7 illustrates

[3]E.g. the location of the firms in Flambard and Perrigne [8]). A rejection of symmetry with our test would invite the econometrician to collect more covariates.

[4]Testing issues in the auction literature are mainly devoted to test for common values. See Haile et al. [11] for a nonparametric treatment.

the usefulness of our methodology with some Monte Carlo simulations. Section 8 concludes by indicating some future lines of research. Most proofs are relegated to the Appendix.

2 The Model

Consider an auction of a single indivisible good with $n \geq 2$ risk-neutral bidders. We consider the first and second price sealed-bid auctions with no reserve price and when all bids are collected by the econometrician.[5] Nevertheless, if the econometrician can observe the amounts submitted by all bidders, we assume that bids are anonymous, i.e. she can not observe their corresponding identities. Hence, she observes the ordered vector of bids $B = (B_1, \cdots, B_p, \cdots, B_n)$, where B_p denotes the p^{th} order statistic of the vector of bids B. But she does not observe $B^* = (B_1^*, \cdots, B_i^*, \cdots, B_n^*)$, where B_i^* denotes the amount submitted by bidder i. Subsequently, we use the indices i, j for bidders' identities and p, r for bidding order statistics.

We consider the private value paradigm: each participant $i = 1, \cdots, n$ is assumed to have a private value x_i for the auctioned object. Hence, bidder i would receive utility $x_i - p$ from winning the object at price p. In the first and second price auctions, the price p is equal to B_n and B_{n-1}, respectively. Let $F_{\mathbf{X_i}}(.)$ and $F_{\mathbf{X}}(.)$ denote the cumulative distribution functions of X_i and $X = (X_1, \cdots, X_n)$, respectively, which are assumed to be absolutely continuous with probability density functions (PDF) $f_{\mathbf{X_i}}(.)$ and $f_{\mathbf{X}}(.)$ and compact support $[\underline{x}, \overline{x}]$ and $[\underline{x}, \overline{x}]^n$, respectively.[6,7] Each bidder is privately informed about x_i, whereas the common distribution $F_{\mathbf{X}}(.)$ is assumed to be common knowledge among bidders. When we refer to models with *symmetric* bidders we assume that the joint distribution of $\mathbf{X}$ is exchangeable with respect to buyers' indices. On the other hand, when we treat models allowing *asymmetric* bidders we drop the exchangeability assumption. For a generic random variable $\mathbf{S}$

[5] How to extend our methodology with risk-aversed bidders, with binding reserve prices and with incomplete sets of bids is discussed in section 8.

[6] Throughout, uppercase letters are used for distributions, while lowercase letters are used for densities. We also follow the standard notation by using an uppercase letter for a statistic and the corresponding lowercase letter for its realization.

[7] We restrict ourselves to the common-support case that guarantees that almost all bids are 'serious' bids, i.e. win with a strictly positive probability. Otherwise identification is obtained only for 'serious' types. See Lebrun [20] for the analysis of the first-price auction with different supports.

and a class of events $\mathbf{E}$, we denote respectively $F_{\mathbf{S}|\mathbf{E}}(.|e)$ and $f_{\mathbf{S}|\mathbf{E}}(.|e)$ the CDF and PDF of $\mathbf{S}$ conditionally on an event e in $\mathbf{E}$. Our analysis falls into two classes of models:

Independent Private Values (IPV): $F_{\mathbf{X}}(x) = \prod_{i=1}^{n} F_{\mathbf{X_i}}(x_i)$.

Strictly Affiliated Private Value (APV): $\frac{\partial^2 \log f_{\mathbf{X}}(x)}{\partial x_i \partial x_j} \geq \epsilon > 0$ for $i \neq j$ if $f_{\mathbf{X}}(x) > 0$

Assumption A 1 *The joint density $f_{\mathbf{X}}$ is bounded, atomless and strictly positive on $[\underline{x}, \overline{x}]^n$.*

We restrict attention to Bayesian Nash Equilibrium in weakly undominated pure strategies, denoted by $(\beta_1(.), \cdots, \beta_n(.))$, where $\beta_i(.)$ is the bidding function of bidder i. In the equilibrium of the second price auction, buyers are thus bidding their private value. Hence, the link between bids and private types is straightforward:

$$x_i = b_i \equiv \xi_i^{nd}(b_i, F_{\mathbf{B}}) \tag{1}$$

In the first price auction, under assumption (1), Athey [1] guarantees the existence of an increasing pure strategy equilibrium in the IPV and APV models. The link between bids and types for each bidder i is made by a standard rewriting of the first order differential equation derived from bidder i's optimization program:

$$x_i = b_i + \frac{F_{\mathbf{B}^*_{-i}|\mathbf{B}^*_i}(b_i|b_i)}{f_{\mathbf{B}^*_{-i}|\mathbf{B}^*_i}(b_i|b_i)} \equiv \xi_i^{rst}(b_i, F_{\mathbf{B}}), \tag{2}$$

where $\mathbf{B}^*_{-i}$ denotes the maximum of the bids from bidder i's opponents.

Following Laffont and Vuong [16], we extend the literature on identification of private value models to the case where bids are anonymous. On the one hand, if bidders' identities are observed, then identifiability corresponds to the condition that, if two possible underlying distributions $F_{\mathbf{X}}(.)$ and $F'_{\mathbf{X}}(.)$ of private signals lead to the same distribution of bids $F_{\mathbf{B}^*}(.)$, then it follows that $F_{\mathbf{X}}(.)$ and $F'_{\mathbf{X}}(.)$ are equal. On the other hand, the following definition introduces the notion of identifiability that makes sense under anonymity.

Definition 1 (Identifiability under anonymity) *Under anonymous bidding, an auction model is said to be identifiable if for two possible underlying distributions $F_{\mathbf{X}}(.)$ and $F'_{\mathbf{X}}(.)$ of private values leading to the same distribution of bids $F_{\mathbf{B}}(.)$, then it follows that $F_{\mathbf{X}}(.)$ and $F'_{\mathbf{X}}(.)$ are equal up to a permutation of the potential* *buyers, i.e. there exists a permutation* $\pi : [1, n] \to [1, n]$ *such that* $F_{\mathbf{X}}(x_1, \cdots, x_n) = F'_{\mathbf{X}}(x_{\pi(1)}, \cdots, x_{\pi(n)})$ *for almost any vector of types* X*.*

Our definition of identifiability corresponds to the possibility of recovering an anonymous joint distribution of buyers' private values. Note that this information is not sufficient with asymmetric PV models for the computation of the optimal mechanism à la Myerson [28] that requires the knowledge of bidders' identities. Nevertheless, it is sufficient for the computation of the optimal anonymous mechanism or the optimal reserve price in a standard auction.

3 Nonparametric Identification

Anonymity restricts the degree of information of the data and thus it can only reduce the identification possibilities. In particular we show that asymmetric affiliated private value models are not identified on the contrary to Campo et al. [7]'s identification result in a framework where bidders' identities are observed. Nevertheless, we also show in Proposition 3.1 that, for a complete set of bids, either symmetry or independence restores identification. The surprising result is that anonymity does not prevent the identification of asymmetric IPV models. Our proof is constructive as it gives $F_{\mathbf{X}}(.)$ as a function of $F_{\mathbf{B}}(.)$. The empirical counterpart of this construction will then be used in the section devoted to nonparametric estimation. The proof of this result is thus given in the body of the text. The resolution of this inverse problem contains two steps. First we derive the distribution of the bids B_i^* from the distribution of the order statistics B_p, the vector of the bidding order statistics. It is the innovative step: by an appropriate reparametrization, the nonlinear inverse problem we face is reduced to a known one, as it happens the root-finding of well chosen polynomials. The second step is the identification of bidders' private signals from the distribution of B^* and is well-known: it is straightforward in the second price auction, whereas the first price auction has been treated by GPV.

Proposition 3.1 *Under the full observation of any submitted bids and under anonymous bids, in the first price and second price auctions and for* $n \geq 2$*:*

- *The asymmetric APV model is not identified. For any distribution* $F_{\mathbf{X}}(.)$ *from the asymmetric APV model, there exists a continuum of local perturbations* *of $F_{\mathbf{X}}(.)$ that stay in the asymmetric APV model and that are observationally equivalent to $F_{\mathbf{X}}(.)$, i.e. that lead to the same distribution of bids.*

- *The symmetric APV model is identified.*
- *The asymmetric IPV model is identified.*

The second point is immediate since the identification result in Li et al. [23] does not rely on the observability of bidders' identities. For the first point, we construct, as it is done in the appendix, a continuum of local perturbations of the primitives that are observationally equivalent. For any IPV model, the local perturbations constructed in the proof of the first point of Proposition 3.1 break independence, which illustrates the more general point that any unordered (i.e. observable up to a permutation) vector of independent components is observationally equivalent to a model where the components are correlated. In other words, the econometrician has to assume independence in order to identify asymmetry.[8]

The rest of this section is devoted to the proof of the third point. We observe the CDFs $F_{\mathbf{B_p}}$ for any $p = 1, \cdots, n$. As independence implies exchangeability, then we can identify the CDF $F_{\mathbf{B}}^{(r:m)}(u)$, $r \leq m$, that corresponds to the r^{th} order statistic among m bidders that would result by exogenous variation of the number of bidders, by recursive use of the formula (see Athey and Haile [2] p.2128)

$$\frac{m-r}{m} F_{\mathbf{B}}^{(r:m)}(u) + \frac{r}{m} F_{\mathbf{B}}^{(r+1:m)}(u) = F_{\mathbf{B}}^{(r:m-1)}(u), \qquad \forall u, r, m, r \leq m-1, m \leq n. \tag{3}$$

The corresponding induction is initialized by noting that $F_{\mathbf{B}}^{(p:n)} = F_{\mathbf{B_p}}$. In particular, it implies the identification of the CDFs $F_{\mathbf{B}}^{(r:r)}$ for any $r \in [1, n]$. Indeed, the expression of $F_{\mathbf{B}}^{(r:r)}$ corresponds to a linear combination of the CDFs $F_{\mathbf{B_p}}$, for $p = 1, \cdots, n$. Finally, independence allows us to express $F_{\mathbf{B}}^{(r:r)}(b)$ as a function of the vector $\{F_{\mathbf{B_i^*}}(b)\}_{i=1,\cdots,n}$ for any b in the following way.

[8] The nonparametric approaches in the literature that test whether the different components of a vector $X = (x_1, \cdots, x_m) \in \mathbb{R}^m$ are independent, e.g. the Blum et al. [5] test, consider that the statistician observes ordered vectors, i.e. she can distinguish $X = (x_1, \cdots, x_m)$ from $X' = (x_{\pi(1)}, \cdots, x_{\pi(m)})$ where π is a permutation in $[1, m]$. With respect to our setup, those tests are requiring nonanonymous data. Independence can not be fully tested under anonymity. Nevertheless, independence involves some testable restrictions under anonymity: a set of generalized discriminants has to be nonnegative as established in Proposition 5.1. In the same way as we propose tests for symmetry in section 5, partial independence tests could be built.

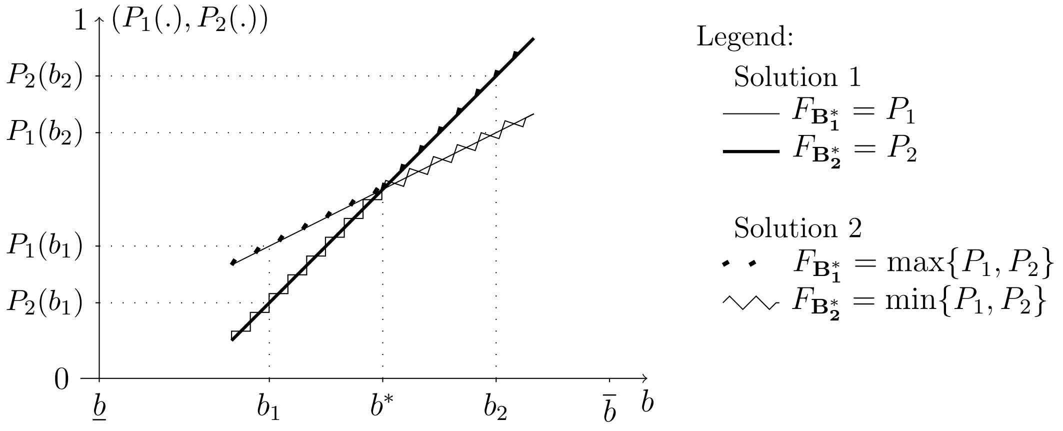
$$
\begin{aligned}
F_{\mathbf{B}}^{(1:1)}(b) =& \frac{1}{n}\cdot & \sum_{i_1=1}^{n} F_{\mathbf{B}^*_{\mathbf{i_1}}}(b) \\
F_{\mathbf{B}}^{(2:2)}(b) =& \frac{1}{n(n-1)}\cdot & \sum_{i_1,i_2,i_1\neq i_2} F_{\mathbf{B}^*_{\mathbf{i_1}}}(b)\cdot F_{\mathbf{B}^*_{\mathbf{i_2}}}(b) \\
& \cdots & \cdots \\
F_{\mathbf{B}}^{(r:r)}(b) =& \frac{1}{n(n-1)\cdots(n-r+1)}\cdot & \sum_{i_1,\cdots,i_r,i_l\neq i_{l'}} \prod_{i_k\in\{i_1,\cdots,i_r\}} F_{\mathbf{B}^*_{\mathbf{i_k}}}(b) \\
& \cdots & .. \\
F_{\mathbf{B}}^{(n:n)}(b) =& \frac{1}{n!}\cdot & \sum_{i_1,\cdots,i_n,i_l\neq i_{l'}} \prod_{i_k\in\{i_1,\cdots,i_n\}} F_{\mathbf{B}^*_{\mathbf{i_k}}}(b)
\end{aligned}
\tag{4}
$$

The right expressions in the system (4) are closely related to the coefficients of the expansion of the polynomial $X \to \prod_{i=1}^{n}(X - F_{\mathbf{B}^*_{\mathbf{i}}}(b))$. The coefficient in front of the monomial X^r for $0 \leq r \leq n-1$ is given by $(-1)^{n-r}\cdot\sum_{i_1<\cdots<i_r}\prod_{i_k\in\{i_1,\cdot,i_r\}} F_{\mathbf{B}^*_{\mathbf{i_k}}}(b)$, which is also equal to $\frac{(-1)^{n-r}}{r!}\cdot\sum_{i_1,\cdots,i_r,i_l\neq i_{l'}}\prod_{i_k\in\{i_1,\cdot,i_r\}} F_{\mathbf{B}^*_{\mathbf{i_k}}}(b)$. From the Fundamental Theorem of Algebra [4], the factorization of a polynomial according to its roots among the complex number $\mathcal{C}$ exists and is unique. Consequently, when b is fixed, the probabilities $(F_{\mathbf{B}^*_{\mathbf{i}}}(b))_{i=1,\cdots,n}$ in the above system of equations correspond exactly to the n roots of the polynomial of degree n:

$$
u \to \sum_{i=0}^{n} a_i(b)\cdot(-1)^{n-i}\cdot u^i, \tag{5}
$$

where $a_n(b) = 1$ and $a_i(b) = \frac{n(n-1)\cdots(i+1)}{(n-i)!}\cdot F_{\mathbf{B}}^{(n-i:n-i)}(b)$, for $i < n$. By continuity of the coefficients of the polynomial as a function of b and since the roots of a polynomial depends continuously on its coefficients (see Theorem 5.12 in [4]), there exists a continuous function $b \to (P_1(b),\cdots,P_n(b))$ mapping the vector of pointwise solutions. What remains to show is the more restrictive condition that the true CDFs $F_{\mathbf{B}^*_{\mathbf{i}}}(b)$, $i = 1,\cdots,n$, are the unique solution up to a permutation. If the n roots of the above polynomial were always distinct for any b in the interior of the bidding support $(\underline{b},\overline{b})$, then, by continuity of the CDFs $F_{\mathbf{B}^*_{\mathbf{i}}}(.)$, $i = 1,\cdots,n$, the only candidate solution would be $(F_{\mathbf{B}^*_{\mathbf{1}}}(.),\cdots,F_{\mathbf{B}^*_{\mathbf{n}}}(.)) = (P_1(.),\cdots,P_n(.))$ (up to a permutation). On the contrary, if the maps $P_i(b)$ cross then the way we construct the continuous selection of the roots $(P_1(.),\cdots,P_n(.))$ is no more unique as it is illustrated in Figure 1 where two candidate solutions are depicted for $n = 2$ when the roots cross at least once.

Figure 1: Identification of the asymmetric IPV model, $n = 2$



Indeed, the sole knowledge of the CDFs $F_{\mathbf{B}}^{(p:m)}$ for any p, m such that $p \leq m \leq n$ can not discriminate between these two possible solutions. Nevertheless, the knowledge of the joint distribution $F_{\mathbf{B}}$ of all order statistics selects a unique solution. For example, consider the case $n = 2$ and a point b^* where $P_1(.)$ and $P_2(.)$ strictly cross as in Figure 1. We consider a point b_2 at the right of the intersection (respectively b_1 at the left of the intersection) such that the derivative of the upper root as a function of b, $P_2'(b_2)$ (resp. $P_1'(b_1)$), is strictly bigger (resp. strictly smaller) than the derivative of the lower root, $P_1'(b_2)$ (resp. $P_2'(b_1)$). Such a point exists in the right (resp. left) neighborhood of b^* since the intersection is strict. Then the two candidate solutions lead to different predictions in term of the joint density of the order statistics: $f_{\mathbf{B}}(b_1, b_2) = f_{\mathbf{B}_1^*}(b_1) \cdot f_{\mathbf{B}_2^*}(b_2) + f_{\mathbf{B}_1^*}(b_2) \cdot f_{\mathbf{B}_2^*}(b_1)$. The difference of the densities $f_{\mathbf{B}}(b_1, b_2)$ between the two depicted solutions is equal to $(P_2'(b_2) - P_1'(b_2)) \cdot (P_2'(b_1) - P_1'(b_1)) \neq 0$. The argument remains valid for any number of bidders and also for more general intersections where the roots may coincide on an interval.

In the subsequent estimation analysis, we will rule out such uncertainties with regards to possible intersections by assuming strict stochastic dominance between bidders' CDFs. It avoids complications concerning the rate of convergence at the intersection points. Without any specific assumption, an appropriate way to decide between the finite number of solutions resulting from the intersections is to choose the solution that maximizes the empirical likelihood. The probability to choose the 'good' candidate converges to 1 as the size of the sample grows.

4 Nonparametric Estimation

In practice the auctioned objects can be heterogeneous and the number and the identities of the participants can vary across auctions. Let $Z_l \in \mathcal{R}^d$ denote the d-dimensional vector of relevant continuous characteristics for the l^{th} auctioned object and I_l (n_{I_l}) the set (number) of participants in the l^{th} auction. The vector (Z_l, I_l) is assumed to be common knowledge among bidders and is also observed by the econometrician.[9] The set of participants (that may vary from an auction to another) is denoted by the letter I and covariates by the letter z. Let $\mathcal{I}$ be the (finite) set of possible values for I. Relative to our previous notation, we will now work with conditional CDFs and PDFs of private values and bids given (Z_l, I_l). E.g. $F_{\mathbf{X_i}|\mathbf{Z},\mathbf{I}}(.|Z_l, I_l)$ denotes the CDF of bidder i's private value X_{il} in the l^{th} auction. Using independence, (1) and (2) can be rewritten as

$$X_{il} = B_{il}^* + \psi_i(B_{il}^*, Z_l, \mathrm{I}_l), \text{ where } \psi_i(.,.,.) \text{ is defined as} \tag{6}$$

$$\psi_i(b, z, I) = \begin{cases} \left[\sum_{j \in I_l, j \neq i} \frac{f_{\mathbf{B}_j^*|\mathbf{Z},\mathrm{I}}(B_{il}^*|Z_l, \mathrm{I}_l)}{F_{\mathbf{B}_j^*|\mathbf{Z},\mathrm{I}}(B_{il}^*|Z_l, \mathrm{I}_l)} \right]^{-1} & \text{in the first price auction} \\ 0 & \text{in the second price auction.} \end{cases}$$

In this section, we adapt GPV's two step estimation procedure to recover the densities of bidders' private values.[10] Two caveats arise. First we can not directly estimate with kernel techniques the ratio $\frac{F_{\mathbf{B}_{-i}^*|\mathbf{B}_i^*,\mathbf{Z},\mathrm{I}}(.|.,.,.)}{f_{\mathbf{B}_{-i}^*|\mathbf{B}_i^*\mathbf{Z},\mathrm{I}}(.|.,.,.)}$ since identities are not observed. Thus we need to convert our estimations of the CDFs and PDFs of B_{pl}, that can be done with the standard kernel estimation techniques as in GPV, into estimations for the CDFs and PDFs of B_{il}^*. Second, if $\frac{F_{\mathbf{B}_{-i}^*|\mathbf{B}_i^*,\mathbf{Z},\mathrm{I}}(.|.,.,.)}{f_{\mathbf{B}_{-i}^*|\mathbf{B}_i^*\mathbf{Z},\mathrm{I}}(.|.,.,.)}$ is suitably estimated, we can apply (6) to define pseudo private values in the first price auction: each pseudo private value being associated to a possible identity of the bidder. With anonymity, an additional step is needed: for a given vector of bidding order statistics

[9] The observation of the identities of the participants by the econometrician may appear in contradiction with our paradigm of anonymous bids. If we could not observe participants' identities, as on eBay, we can adapt our method if we are prepared to make specific assumptions about the identities of the fluctuating bidders (e.g. real bidder versus shill bidder). Anyway, in an asymmetric framework, the exogenous participation assumption that is often made for identification as in Athey and Haile [2] may not be suitable since the expected payoffs in the auction differ across bidders. Considering endogenous participation with a set of participants that is assumed to be known is an alternative.

[10] See Flambard and Perrigne [8] for the the implementation of GPV's procedure in the asymmetric IPV model with nonanonymous bids.

$B_l = (B_{1l}, \cdots, B_{pl}, \cdots, B_{nl})$, we have to estimate the probability that buyer i's bid B^*_{il} is equal to B_{pl} for any $p \in [1, n]$. Then instead of a unique pseudo private value for a given bidder, we obtain a weighted vector of n pseudo private values that is used to estimated nonparametrically buyers' private values PDFs. We also lead in parallel the analysis for the second price auction which is not straightforward as it was with nonanonymous bids and also involves the computation of a vector of pseudo probabilities.

Denote Σ_{I} the set of the $n_{\mathrm{I}}!$ permutations between participants' identities and the order statistics of the bids. Such an assignment of the bids to the participants is denoted $\pi : \mathrm{I} \to [1, n_{\mathrm{I}}]$ where $\pi(i) = p$ means that the p^{th} order statistic of the bids corresponds to bidder i, i.e. $b^*_i = b_p$. To cover both the case where bidders' identities remain fully anonymous with the common case where only the identity of the winner is disclosed, we consider the most general case when the econometrician may have some information linking some submitted bids with the identities of some participants. This information is modeled as a partition of Σ_{I} which may depend both on the vector of bids B and the auction (but not on B^*). Denote by σ_{I_l} this information set at the l^{th} auction. If π is the assignment that match the (observed) vector of bids B_l to the true (unobserved) realization B^*_l, then we know that $\pi \in \sigma_{\mathrm{I}_l}$. $\sigma_{\mathrm{I}_l} = \Sigma_{\mathrm{I}_l}$ corresponds to the case where bids are fully anonymous, whereas the opposite case where σ_{I_l} is always a singleton corresponds to non-anonymous bids.

Our procedure is decomposed in 6 steps, three being already present in GPV.

First step Using the observations $\{(B_{pl}, Z_l, \mathrm{I}_l); p \in \mathrm{I}_l, l = 1, \cdots, L\}$, we estimate the CDFs and the PDFs of the p^{th} ordered statistics of the bids for $p \in [1, n_{I_l}]$ and the PDFs of the covariates. Let x^+ denote $\max\{0, x\}$.

$$\widehat{F}_{\mathbf{B_p},\mathbf{Z},\mathrm{I}}(b, z, \mathrm{I}) = [\min\left\{\frac{1}{Lh^d_{F_{\mathbf{B_p}|\mathbf{Z}}}}\sum_{l=1}^{L}\mathbf{1}(B_{pl} \le b)K_{F_{\mathbf{B_p}|\mathbf{Z}}}\left(\frac{z - Z_l}{h_{F_{\mathbf{B_p}|\mathbf{Z}}}}\right)\mathbf{1}(\mathrm{I}_l = \mathrm{I}), 1\right\}]^+ \quad (7)$$

$$\widehat{f}_{\mathbf{B_p},\mathbf{Z},\mathrm{I}}(b, z, \mathrm{I}) = [\frac{1}{Lh^d_{f_{\mathbf{B_p}|\mathbf{Z}}}}\sum_{l=1}^{L}K_{f_{\mathbf{B_p}|\mathbf{Z}}}\left(\frac{b - B_{pl}}{h_{f_{\mathbf{B_p}|\mathbf{Z}}}}, \frac{z - Z_l}{h_{f_{\mathbf{B_p}|\mathbf{Z}}}}\right)\mathbf{1}(\mathrm{I}_l = \mathrm{I})]^+ \quad (8)$$

$$\widehat{f}_{\mathbf{Z},\mathbf{I}}(z, I) = [\frac{1}{Lh^d_{f_{\mathbf{Z}}}}\sum_{l=1}^{L}\sum_{p=1}^{n}K_{f_{\mathbf{Z}}}(\frac{z - Z_l}{h_{f_{\mathbf{Z}}}}) \cdot \mathbf{1}(I_l = I)]^+. \quad (9)$$

Here $h_{F_{\mathbf{B_p}|\mathbf{Z}}}, h_{f_{\mathbf{B_p}|\mathbf{Z}}}, h_{f_{\mathbf{Z}}}$ are some bandwidths, and $K_{F_{\mathbf{B_p}|\mathbf{Z}}}(.)$, $K_{f_{\mathbf{B_p}|\mathbf{Z}}}(.,.)$ and $K_{f_{\mathbf{Z}}}(.)$ are kernels with bounded supports.

Second step By recursive use of the empirical counterpart of the formula (3), we estimate $\widehat{F}^{(r:r)}_{\mathbf{B},\mathbf{Z},\mathbf{I}}(b,z,I)$ and $\widehat{f}^{(r:r)}_{\mathbf{B},\mathbf{Z},\mathbf{I}}(b,z,I)$ for $r = 1,\cdots,n$, which respectively corresponds (up to a known multiplicative coefficient) to the coefficients and their derivatives with respect to the variable b of a polynomial whose vector of roots is the vector of bidders' bidding distribution $\{F_{\mathbf{B}^*_\mathbf{i},\mathbf{Z},\mathbf{I}}(b,z,I)\}_{i\in I}$.

For $r \le m \le n$, we define $\widehat{F}^{(r:m)}_{\mathbf{B},\mathbf{Z},\mathbf{I}}(b,z,I)$ and $\widehat{f}^{(r:m)}_{\mathbf{B},\mathbf{Z},\mathbf{I}}(b,z,I)$ by recursive use of the formulas: $\forall b,z,r \le m-1$

$$\begin{aligned}\frac{m-r}{m}\widehat{F}^{(r:m)}_{\mathbf{B},\mathbf{Z},\mathbf{I}}(b,z,I) + \frac{r}{m}\widehat{F}^{(r+1:m)}_{\mathbf{B},\mathbf{Z},\mathbf{I}}(b,z,I) &= \widehat{F}^{(r:m-1)}_{\mathbf{B},\mathbf{Z},\mathbf{I}}(b,z,I)\\ \frac{m-r}{m}\widehat{f}^{(r:m)}_{\mathbf{B},\mathbf{Z},\mathbf{I}}(b,z,I) + \frac{r}{m}\widehat{f}^{(r+1:m)}_{\mathbf{B},\mathbf{Z},\mathbf{I}}(b,z,I) &= \widehat{f}^{(r:m-1)}_{\mathbf{B},\mathbf{Z},\mathbf{I}}(b,z,I).\end{aligned} \tag{10}$$

As a weighted sum of the estimators $\widehat{F}_{\mathbf{B_p},\mathbf{Z},\mathbf{I}}$ which are confined in the interval $[0,1]$, the estimators $\widehat{F}^{(r:m)}_{\mathbf{B},\mathbf{Z},\mathbf{I}}(b,z,I)$ are confined in the interval $[0,1]$.

Third step Let $\Upsilon : [0,1]^n \to \mathcal{C}^n$ be the function such that $(\omega_1,\cdots,\omega_n) = \Upsilon(a_{n-1},\cdots,a_0)$ (where $\omega_1 \ge \cdots \ge \omega_n$ according to the lexicographic order in $\mathcal{C}$) is the ordered vector of the roots (possibly complex number) counted with their order of multiplicity of the polynomial $Q(u) = u^n + \sum_{i=0}^{n-1} a_i \cdot (-1)^{n-i}u^i$, i.e. $Q(u) = \prod_{i=1}^n (u-\omega_i)$. Theorem 5.12 in [4] show that Υ is continuous and hence uniformly continuous on the compact $[0,1]^n$. Then, after an immediate generalization of (4) and (5) to our environment with covariates, it would be natural to estimate the CDFs $\widehat{F}_{\mathbf{B}^*_\mathbf{i},\mathbf{Z},\mathbf{I}}(.,.), i \in I$ by

$$(\widehat{F}_{\mathbf{B}^*_{\mathbf{j_1}},\mathbf{Z},\mathbf{I}}(b,z,I),\cdots,\widehat{F}_{\mathbf{B}^*_{\mathbf{j_{n_I}}},\mathbf{Z},\mathbf{I}}(b,z,I)) = \mathcal{R}[\Upsilon(\widehat{a}_{n-1}(b,z,I),\cdots,\widehat{a}_0(b,z,I))], \tag{11}$$

where $\widehat{a}_i(b,z,I) = \frac{n(n-1)\cdots(i+1)}{(n-i)!}\cdot \widehat{F}^{(n-i:n-i)}_{\mathbf{B},\mathbf{Z},\mathbf{I}}(b,z,I)\cdot(\widehat{f}_{\mathbf{Z},\mathbf{I}}(z,I))^{n-i-1}$, $\mathcal{R}[z]$ denotes the real part of the complex vector z and $I = (j_1,\cdots,j_{n_I})$, where $j_1 < \cdots < j_{n_I}$. The rest of this step is devoted to the estimation of $f_{\mathbf{B}^*_\mathbf{i},\mathbf{Z},\mathbf{I}}(b,z,I)$ for $i \in I$. The derivative of the polynomial relation with respect to b leads to:

$$\frac{\partial Q(u)}{\partial b} = \sum_{i=0}^{n_I-1} \frac{\partial a_i}{\partial b}(b,z,I)\cdot(-1)^{n-i}\cdot u^i$$
$$= -\sum_{i\in I}\prod_{j\in I,\ j\neq i}\Big(u - F_{\mathbf{B_j^*},\mathbf{Z},\mathbf{I}}(b,z,I)\Big)\cdot f_{\mathbf{B_i^*},\mathbf{Z},\mathbf{I}}(b,z,I), \forall u,b,z,I,$$

where $\frac{\partial a_i}{\partial b}(b,z,I) = \frac{n(n-1)\cdots(i+1)}{(n-i)!}\cdot f_{\mathbf{B},\mathbf{Z},\mathbf{I}}^{(n-i:n-i)}(b,z,I)\cdot(f_{\mathbf{Z},\mathbf{I}}(z,I))^{n-i-1}$. For a single estimated root, i.e. for i such that $\widehat{F}_{\mathbf{B_i^*},\mathbf{Z},\mathbf{I}}(b,z,I) \neq \widehat{F}_{\mathbf{B_j^*},\mathbf{Z},\mathbf{I}}(b,z,I)$ for any $j\neq i$, we have a natural estimator for the corresponding density:

$$\widehat{f}_{\mathbf{B_i^*},\mathbf{Z},\mathbf{I}}(b,z,I) = \frac{\sum_{s=0}^{n_I-1}\frac{\partial\widehat{a}_k}{\partial b}(b,z,I)\cdot(-1)^{n_I-s+1}\cdot\Big[\widehat{F}_{\mathbf{B_i^*},\mathbf{Z},\mathbf{I}}(b,z,I)\Big]^s}{\prod_{j\in I,\ j\neq i}\Big(\widehat{F}_{\mathbf{B_i^*},\mathbf{Z},\mathbf{I}}(b,z,I) - \widehat{F}_{\mathbf{B_j^*},\mathbf{Z},\mathbf{I}}(b,z,I)\Big)}, \tag{12}$$

where $\frac{\partial\widehat{a}_s}{\partial b}(b,z,I) = \frac{n(n-1)\cdots(s+1)}{(n-s)!}\cdot\widehat{f}_{\mathbf{B},\mathbf{Z},\mathbf{I}}^{(n-s:n-s)}(b,z,I)\cdot(\widehat{f}_{\mathbf{Z},\mathbf{I}}(z,I))^{n-s-1}$. Consider now the case of a multiple estimated root of multiplicity $k>1$, i.e. consider $J = \{j_m,\cdots,j_{m+k-1}\}$ such that for any $i\in J$, $\widehat{F}_{\mathbf{B_i^*},\mathbf{Z},\mathbf{I}}(b,z,I) = \widehat{F}_{\mathbf{B_j^*},\mathbf{Z},\mathbf{I}}(b,z,I)$ if and only if $j\in J$. The derivative of the polynomial relation with respect to b and $k-1$ times with respect to u leads to:

$$\frac{\partial Q(u)}{\partial b(\partial u)^{k-1}} = \sum_{i=0}^{n_I-k}\frac{\partial a_{i+k-1}}{\partial b}(b,z,I)\cdot(-1)^{n-i-k+1}\cdot\frac{i+k-1!}{i!}\cdot u^i$$
$$= -\sum_{i\in I}\frac{\partial\prod_{j\in I,\ j\neq i}\Big(u - F_{\mathbf{B_j^*},\mathbf{Z},\mathbf{I}}(b,z,I)\Big)}{(\partial u)^{k-1}}\cdot f_{\mathbf{B_i^*},\mathbf{Z},\mathbf{I}}(b,z,I), \forall u,b,z,I.$$

For $i\in J$, the expression $\partial\prod_{j\in I,\ j\neq i}\Big(u - F_{\mathbf{B_j^*},\mathbf{Z},\mathbf{I}}(b,z,I)\Big)/(\partial u)^{k-1}$ evaluated at $u = F_{\mathbf{B_i^*},\mathbf{Z},\mathbf{I}}(b,z,I)$ reduces to $\prod_{j\in I,\ j\notin J}\Big(F_{\mathbf{B_i^*},\mathbf{Z},\mathbf{I}}(b,z,I) - F_{\mathbf{B_j^*},\mathbf{Z},\mathbf{I}}(b,z,I)\Big)$. Finally, we have a natural estimator for the corresponding density:

$$\widehat{f}_{\mathbf{B_i^*},\mathbf{Z},\mathbf{I}}(b,z,I) = \frac{\sum_{s=0}^{n_I-k}\frac{\partial\widehat{a}_{s+k-1}}{\partial b}(b,z,I)\cdot(-1)^{n-s-k+1}\cdot\frac{s+k-1!}{s!}\cdot\Big[\widehat{F}_{\mathbf{B_i^*},\mathbf{Z},\mathbf{I}}(b,z,I)\Big]^s}{\prod_{j\in I,\ j\notin J}\Big(\widehat{F}_{\mathbf{B_i^*},\mathbf{Z},\mathbf{I}}(b,z,I) - \widehat{F}_{\mathbf{B_j^*},\mathbf{Z},\mathbf{I}}(b,z,I)\Big)}.$$

For $k=1$, this formula corresponds exactly to (12). Now we have all the elements to estimate the function $\psi_i(.,.,.)$ in the first price auction. In the second price auction the job seems to be done since we have recovered the bid distributions which corresponds to the valuation distributions. However, we still have not used the additional information σ_{I_l} which motivates the three remaining steps where we build a pseudo sample of private values and where a probability is estimated to each private value for each possible identity. Those probabilities are updated according to the Bayesian rule with regards to the additional information σ_{I_l}. For the first price auction, those remaining steps are also generalizing the second step in GPV.

Fourth step In view of (6) and similarly to GPV, it would be natural to construct pseudo private values for each order statistic $p = 1, \cdots, n_{I_l}$ and for each potential bidder $i \in I_l$: $\widetilde{X}_{ipl} = B^*_{pl} + \widetilde{\psi}_i(B^*_{pl}, Z_l, \mathrm{I}_l)$. Unfortunately, as has been emphasized by GPV, the estimator of $\psi_i(.,.,.)$ in the first price auction is asymptotically biased at the boundaries of the support and trimming is required. The same trimming is also needed in the second price auction.

In this aim we first estimate the boundary of the support of the joint distribution of (B, Z, I), which is unknown. Since the support of (Z, I) can be assumed to be known, we focus on the estimation of the support $[\underline{b}(z, I), \overline{b}(z, I)]$ of the conditional distribution of B given (Z, I). On the one hand, we assume that $\underline{b}(z, I)$ does not depend on (z, I) and is estimated by the minimum of all submitted bids. On the other hand, $\underline{b}(z, I)$ should be estimated as in GPV. Let $h_\delta > 0$. We consider the following partition of $\mathbb{R}^d$ with a generic hypercube of side h_δ: $\vartheta_{k_1,\cdots,k_d} = [k_1h_\delta, (k_1+1)h_\delta) \times \cdots \times [k_dh_\delta, (k_d+1)h_\delta)$, where $k_1, \cdots, k_d$ runs over $\mathbb{Z}^d$. This induces a partition of $[\underline{z}, \overline{z}]$. Given a set of participants I and a value z, the estimate of the upper boundary $\overline{b}(z, I)$ is the maximum of those bids for which $I_l = I$ and the corresponding value of Z_l falls in the hypercube $\vartheta_{k_1,\cdots,k_d}(z)$ containing z. Formally, our estimators for the upper and lower boundaries are respectively given by $\widehat{\underline{b}} = \inf\{B_{1l}, l = 1, \cdots, L\}$ and $\widehat{\overline{b}}(z, I) = \sup\{B_{n_Il}, l = 1, \cdots, L; Z_l \in \vartheta_{k_1,\cdots,k_d}(z), I_l = I\}$. Our estimator for $S(F_{\mathbf{B_p},\mathbf{Z},\mathrm{I}})$ is $\widehat{S}(F_{\mathbf{B_p},\mathbf{Z},\mathrm{I}}) = \{(b, z, I) : b \in [\widehat{\underline{b}}, \widehat{\overline{b}}(z, I)], z \in [\underline{z}, \overline{z}], I \in \mathcal{I}\}$.

We now turn to the trimming. It is well known that kernel estimators are asymptotically biased at the boundaries of the support. Following GPV, we have to trim out observations that are close to the boundaries of the support. Because $\underline{b} \le \widehat{\underline{b}} \le \widehat{\overline{b}}(z, I) \le \overline{b}$, $\widehat{f}_{\mathbf{B}_p,Z,I}(.,.,.)$ and thus $\widehat{f}_{\mathbf{B}^*_\mathbf{j},\mathbf{Z},\mathbf{I}}(.,.,.)$ are asymptotically unbiased on $[\widehat{\underline{b}}+\frac{\rho_{f_{\mathbf{B_p}|\mathbf{Z}}} \cdot h_{f_{\mathbf{B_p}|\mathbf{Z}}}}{2}, \widehat{\overline{b}}(z, I)-\frac{\rho_{f_{\mathbf{B_p}|\mathbf{Z}}} \cdot h_{f_{\mathbf{B_p}|\mathbf{Z}}}}{2}]$. This leads to defining the sample of pseudo private values $\{\widehat{X}_{ipl}, i \in I_l; p = 1, \cdots, n_{I_l}; l = 1, \cdots, L\}$ where $\widehat{X}_{ipl}$, the estimate of the private value of bidder i would it be the bidder of the p^{th} order statistic of the vector of bids B_l, is defined by

$$\widehat{X}_{ipl} = \begin{cases} B_{pl} + \widehat{\psi}_i(B^*_{il}, Z_l, \mathrm{I}_l) & \text{if } \ \widehat{\underline{b}} + h_{f_{\mathbf{B_p}|\mathbf{Z}}} \leq B_{pl} \leq \widehat{\overline{b}}(Z_l, \mathrm{I}_l) \\ +\infty & \text{otherwise} \end{cases}, \tag{13}$$

where $\widehat{\psi}_i(b, z, I)$ equals respectively $\left[\sum_{j\neq i} \frac{\widehat{f}_{\mathbf{B^*_j},\mathbf{Z},\mathbf{I}}(b,z,I)}{\widehat{F}_{\mathbf{B^*_j},\mathbf{Z},\mathbf{I}}(b,z,I)}\right]^{-1}$ and 0 in the first and second price auctions.

Fifth step Contrary to GPV, we should not use directly this pseudo sample of private values in a standard kernel estimation to estimate $f_{\mathbf{X_i},\mathbf{Z},\mathbf{I}}(x, z, I)$. Each pseudo values should not be weighted in the same way since for a given order statistic B_p the probability that it results from a given bidder i depends on the identity of this bidder. Thus we have to estimate the corresponding probability weights. Under anonymity, there are at most $n_{I_l}!$ vectors of private values that can rationalize a given vector of bids $(B_{1l}, \cdots, B_{n_I l})$. Denote by $\widetilde{\pi} \in \Sigma_I$ the true permutation that matches a given vector of bidding order statistics $(B_{1l}, \cdots, B_{n_I l})$ with the unobserved vector of bids $(B^*_{1l}, \cdots, B^*_{nl})$. The following expression gives the theoretical probability, denoted by $Prob(\widetilde{\pi} = \pi|(b_1, \cdots, b_{n_I}, z, I))$, that the assignment of bidders to the observed order statistics corresponds to a permutation π:

$$Prob(\widetilde{\pi} = \pi|(b_1, \cdots, b_{n_I}, z, I)) = \frac{\prod_{i\in I} f_{\mathbf{B^*_i},\mathbf{Z},\mathbf{I}}(b_{\pi(i)}, z, I)}{\sum_{\pi'\in\sigma_I} \prod_{i\in I} f_{\mathbf{B^*_i},\mathbf{Z},\mathbf{I}}(b_{\pi'(i)}, z, I)} \cdot \mathbf{1}\{\pi \in \sigma_I\}. \tag{14}$$

Note that we use the information set σ_I to refine our beliefs on $\widetilde{\pi}$. Then the probability, denoted by P_{ip}, that the p^{th} order statistic results from bidder i equals to the sum of the above probabilities for all the permutations that assign i to the p^{th} order statistic, i.e. $P_{ip} = \sum_{\pi\in\Sigma_I \ s.t. \ \pi(i)=p} Prob(\widetilde{\pi} = \pi|(b_1, \cdots, b_n, z, I))$. Its empirical counterpart $\widehat{P}_{ipl}$ is given straightforwardly by means of our previous estimators:

$$\widehat{P}_{ipl} = \sum_{\pi\in\Sigma_{I_l} \ s.t. \ \pi(i)=p} \frac{\prod_{i\in I_l} \widehat{f}_{\mathbf{B^*_i},\mathbf{Z},\mathbf{I}}(B_{\pi(i)l}, Z_l, I_l)}{\sum_{\pi'\in\sigma_{I_l}} \prod_{i\in I_l} \widehat{f}_{\mathbf{B^*_i},\mathbf{Z},\mathbf{I}}(B_{\pi'(i)l}, Z_l, I_l)} \cdot \mathbf{1}\{\pi \in \sigma_{I_l}\}. \tag{15}$$

Sixth step Finally, we use the pseudo sample $\{(\widehat{X}_{ipl}, \widehat{P}_{ipl}, Z_l), i \in I_l, p = 1, \cdots, n_{I_l}, l = 1, \cdots, L\}$ to estimate nonparametrically the densities $f_{\mathbf{X_i}|\mathbf{Z},\mathbf{I}}(x|z, I)$ by $\widehat{f}_{\mathbf{X_i}|\mathbf{Z},\mathbf{I}}(x|z, I) = \widehat{f}_{\mathbf{X_i}|\mathbf{Z},\mathbf{I}}(x, z, I)/\widehat{f}_{\mathbf{Z},\mathbf{I}}(z, I)$, where

$$\widehat{f}_{\mathbf{X_i},\mathbf{Z},\mathbf{I}}(x,z,I) = \frac{1}{Lh^{d+1}_{f_{\mathbf{X_i},\mathbf{Z}}}} \sum_{l=1}^{L} \sum_{p=1,\cdots,n_{I_l}} \widehat{P}_{ipl} \cdot K_{f_{\mathbf{X_i},\mathbf{Z}}}(\frac{x-\widehat{X}_{ipl}}{h_{f_{\mathbf{X_i},\mathbf{Z}}}}, \frac{z-Z_l}{h_{f_{\mathbf{X_i},\mathbf{Z}}}}) \cdot \mathbf{1}(I_l = I). \tag{16}$$

Here $h_{f_{\mathbf{X_i},\mathbf{Z}}}$ are bandwidths and $K_{f_{\mathbf{X_i},\mathbf{Z}}}(.,.)$ are kernels with bounded support.

Summary of the differences with GPV The first step in GPV's approach consists in estimating the maps $\psi_i(.,.,.)$ which requires the estimation of $f_{\mathbf{B_j^*},\mathbf{Z},\mathbf{I}}(B_{pl}, Z_l, I)$ and $F_{\mathbf{B_j^*},\mathbf{Z},\mathbf{I}}(B_{pl}, Z_l, I)$. Instead of being directly estimated in a similar way as in our first step, anonymity requires two additional steps: the second step is a linear reparametrization for which we has thus no reason to be worried about, the third step is a nonlinear reparametrization which is ill-conditioned at the limit where some bidders are symmetric. The fourth step consists as in GPV in the construction of the set of pseudo private values: n_{I_l} pseudo private values are associated to each bid, one for each possible identities of the potential bidders. On the contrary, in GPV, a unique pseudo private value has to be computed for each bid, the one corresponding to the identity of the bidder which is not anonymous. The fifth step is the most interesting step of our estimation procedure and is not linked to the ideas of the identification section: for each bid, we compute the probability that it comes from a given bidder. Finally, as in GPV, the last step computes the CDFs and PDFs from the pseudo sample which does not suffer from anonymity anymore since it includes a consistent estimator of the (unobserved) realized identities. The asymptotic properties as $L \to \infty$ of such a multi-step nonparametric estimator are rigorously derived in section 6. To end this section, we briefly discuss the new error terms resulting from anonymity. We decompose the difference $\widehat{f}_{\mathbf{X_i},\mathbf{Z},\mathbf{I}}(x,z,I) - f_{\mathbf{X_i},\mathbf{Z},\mathbf{I}}(x|z,I)$ into three terms.

$$\widehat{f}_{\mathbf{X_i},\mathbf{Z},\mathbf{I}}(x,z,I) - f_{\mathbf{X_i},\mathbf{Z},\mathbf{I}}(x,z,I) = \varepsilon_1 + \varepsilon_2 + \varepsilon_3, \text{ where} \tag{17}$$

$$\begin{cases} \varepsilon_1 = & \frac{1}{Lh^{d+1}_{f_{\mathbf{X_i},\mathbf{Z}}}} \sum_{l=1}^{L} \sum_{p=1,\cdots,n_{I_l}} (\widehat{P}_{ipl} - P_{ipl}) \cdot K_{f_{\mathbf{X_i},\mathbf{Z}}}(\frac{x-X_{ipl}}{h_{f_{\mathbf{X_i},\mathbf{Z}}}}, \frac{z-Z_l}{h_{f_{\mathbf{X_i},\mathbf{Z}}}}) \cdot \mathbf{1}(I_l = I) \\ \varepsilon_2 = & \frac{1}{Lh^{d+1}_{f_{\mathbf{X_i},\mathbf{Z}}}} \sum_{l=1}^{L} \sum_{p=1,\cdots,n_{I_l}} \widehat{P}_{ipl} \cdot \left(K_{f_{\mathbf{X_i},\mathbf{Z}}}(\frac{x-\widehat{X}_{ipl}}{h_{f_{\mathbf{X_i},\mathbf{Z}}}}, \frac{z-Z_l}{h_{f_{\mathbf{X_i},\mathbf{Z}}}}) - K_{f_{\mathbf{X_i},\mathbf{Z}}}(\frac{x-X_{ipl}}{h_{f_{\mathbf{X_i},\mathbf{Z}}}}, \frac{z-Z_l}{h_{f_{\mathbf{X_i},\mathbf{Z}}}}) \right) \mathbf{1}(I_l = I) \\ \varepsilon_3 = & \widetilde{f}_{\mathbf{X_i},\mathbf{Z},\mathbf{I}}(x,z,I) - f_{\mathbf{X_i},\mathbf{Z},\mathbf{I}}(x,z,I) \end{cases}$$

and where $\widetilde{f}_{\mathbf{X_i},\mathbf{Z},\mathbf{I}}$ is the (infeasible) nonparametric estimator of the density of (X_i, Z, I) using the unobserved values X_{ipl} and the unobserved probabilities P_{ipl}:

$$\widetilde{f}_{\mathbf{X_i},\mathbf{Z},\mathbf{I}}(x,z,I) = \frac{1}{Lh^{d+1}_{f_{\mathbf{X_i},\mathbf{Z}}}} \sum_{l=1}^{L} \sum_{p=1,\cdots,n_{I_l}} P_{ipl} \cdot K_{f_{\mathbf{X_i},\mathbf{Z}}}\left(\frac{x - X_{ipl}}{h_{f_{\mathbf{X_i},\mathbf{Z}}}}, \frac{z - Z_l}{h_{f_{\mathbf{X_i},\mathbf{Z}}}}\right) \cdot \mathbf{1}(I_l = I). \quad (18)$$

The third term ε_3 is standard and corresponds to the usual sampling error if private values were directly observed. When bidders' private value density functions $f_{\mathbf{X_i},\mathbf{Z},\mathbf{I}}(.,.,I)$ have R bounded continuous derivatives, the optimal uniform convergence rate for estimating $f_{\mathbf{X_i},\mathbf{Z},\mathbf{I}}(.,.,I)$ is $(\frac{L}{\log L})^{R/(2R+d+1)}$ (see Stone [37]). The second term ε_2 is the one pointed in GPV in a framework with non-anonymous data: it comes from the discrepancy between the realized (unobserved) private values and the estimated pseudo private values that are estimated from the observed bids and an estimation of the equilibrium equations (1) and (2) for respectively the second first price auctions. In the second price auction, due to the triviality of the strategic interaction, this discrepancy is null and the optimal uniform rate of convergence for estimating private values' densities is thus $(L/\log L)^{R/(2R+d+1)}$ under non-anonymous data. On the contrary, this discrepancy matters in the first-price auction and consequently the above convergence rate can not be attained in GPV but only the rate $(\frac{L}{\log L})^{R/(2R+d+3)}$. The choice of the bandwidth $h_{f_{\mathbf{X_i},\mathbf{Z}}}$ is driven by the trade-off between controlling those two errors terms, the optimal bandwidth being such that the two rates are equal. The optimal estimator involves a larger bandwidth than if bidders' private values were directly observed, i.e. it oversmoothes the pseudo private values in order to average the errors in the estimation of this pseudo sample. Anonymity introduces new caveats that occur in the second, third and fifth steps of our estimation procedure. The second and third steps are making harder the estimation of the pseudo private values. Nevertheless according to the rate of convergence asymptotic criterium, those steps are innocuous since the same rate in any inner closed subset of the bidding support is obtained for the pseudo private values. The fifth step introduces the new error term ε_1 that results from the discrepancy between the true and the estimated probabilities of the different assignments between bids and bidders. We show that the convergence rate of ε_1 does not introduce a new force in the above trade-off in the first price auction. By choosing appropriately the rate

Auction format:	Second-price	First-price
Standard term: ε_3	$(\frac{\log L}{L})^{R/(2R+d+1)}$	$(\frac{\log L}{L})^{R/(2R+d+3)}$
GPV's term: ε_2	0	$(\frac{\log L}{L})^{R/(2R+d+3)}$
Anonymity term: ε_1	$(\frac{\log L}{L})^{R/(2R+d+1)}$	$(\frac{\log L}{L})^{(R+1)/(2R+d+3)}$

Table 1: Decomposition of the error term of the estimator of the density of bidders' private value and their respective rate of convergence in our 'optimal' procedure.

of the bandwidths, this new error term can be maintained such that its convergence rate is strictly bigger than the rates for two other error terms. This discussion is summarized in Table 1.

5 Nonparametric Tests for symmetry

Our procedure to test for symmetry is the same for both the first and second price auctions. We present it with a general structure of covariates as in the previous section but for a fix set of n participants since an important by-product of our analysis is how the power of our test varies with the structure of the data. For example, with two kinds of bidders (Strong versus Weak) distributed according to two given distributions F_S and F_W, we are interested to know whether it is easier to reject symmetry depending not solely on some measure of the degree of asymmetry (e.g. related to the differences $F_{\mathbf{B_i^*},\mathbf{Z}}(.,.) - F_{\mathbf{B_j^*},\mathbf{Z}}(.,.)$) but also on the number of participants and the structure of the asymmetry (e.g. a single strong bidder or as many strong bidders as weak bidders).

5.1 A General Testing Principle

This subsection derives a complete discrimination system to characterize the structure of the roots of polynomials with real roots, i.e. polynomials P such that $P(X) = \prod_{i=1}^{n}(X - x_i)$ where the roots x_i are real numbers. We first define what we call a root structure. Second we introduce the determinant of well chosen matrix.

Definition 2 *A polynomial with real roots $P(X)$ of degree n has the **root structure** $(k_1, \cdots, k_{r(P)})$ where $\sum_{i=1}^{r(P)} k_i = n$ and $k_1 \geq \cdots \geq k_{r(P)} \geq 1$ if $P(X) = \prod_{i=1}^{r(P)}(X - x_i)^{k_i}$ for some $\{x_i\}_{i=1,\cdots,r(P)}$ such that $x_i \neq x_j$ for all i, j. The integer $r(P)$ is the number of distinct roots.*

Definition 3 *The* ***Discrimination matrix*** *of the monic polynomial* $P = X^p + \sum_{i=0}^{p-1} a_i.X^i$ *is the* $(2p+1) \times (2p+1)$ *matrix:*

$$Discr(P) = \begin{bmatrix} 1 & a_{p-1} & a_{p-2} & \cdots & a_0 & & & & \\ 0 & p & (p-1)a_{p-1} & \cdots & a_1 & & & & \\ & 1 & a_{p-1} & \cdots & a_1 & a_0 & & & \\ & 0 & p & \cdots & 2.a_2 & a_1 & & & \\ & & & \cdots & \cdots & & & & \\ & & & \cdots & \cdots & & & & \\ & & & & 1 & a_{p-1} & \cdots & a_0 & \\ & & & & 0 & p & \cdots & a_1 & \\ & & & & & 1 & a_{p-1} & \cdots & a_0 \end{bmatrix}.$$

For $k \in [1, n]$, let $\Delta(P, k)$ denote the determinant of the submatrix formed by the first $2k$ rows and the first $2k$ columns of $Discr(P)$. The numbers $\Delta(P, k)$ are also called generalized discriminants. Denote by $P^{(i)}$ the i^{th} derivative of the polynomial P (with $P^{(0)} = P$). Note that the generalized discriminants $\Delta(P^{(i)}, k)$ are polynomial functions of the coefficients of the primitive polynomial P.

Proposition 5.1 ***(Corollary of Theorem 2.1 in Yang [41])*** *A polynomial with real roots* P *has the root structure* $(k_1, \cdots, k_{r(P)})$ *if and only if, for any* $i \in [0, n-2]$,

$$\begin{cases} \Delta(P^{(i)}, k) > 0 & \text{for } 1 \leq k \leq \rho(i) \\ \Delta(P^{(i)}, k) = 0 & \text{for } \rho(i) < k \leq n - i \end{cases}, \tag{19}$$

where $\rho(i) = n - i - \sum_{j=1}^{r(P)} (k_j - i - 1)^+$.

The number $\rho(i)$ corresponds to the number of distinct roots of the polynomials $P^{(i)}$. Proposition 5.1 provides a complete discrimination system: the generalized discriminants are all positive since the underlying polynomials have only real roots while some are null due to the multiplicity of some roots.[11]

Coming back to our framework, the probabilities $(F_{\mathbf{B}_\mathbf{i}^*, \mathbf{Z}}(b, z))_{i=1,\cdots,n}$ are corresponding exactly to the n roots of the polynomial $P_{(b,z)} : u \to \sum_{i=0}^{n} a_i(b, z) \cdot (-1)^{n-i} \cdot u^i$, where $a_n(b, z) = 1$ and $a_i(b, z) = \frac{n(n-1)\cdots(i+1)}{(n-i)!} \cdot F_{\mathbf{B},\mathbf{Z}}^{(n-i:n-i)}(b, z) \cdot (f_{\mathbf{Z}}(z))^{n-i-1}$ for $i < n$. For a given bid b and a given set of covariates z, the root structure is characterized by the generalized discriminants $\Delta(P_{b,z}^{(i)}, k)$, which can be easily estimated by their sample analogs $\widehat{\Delta}(P_{b,z}^{(i)}, k) = \Delta(\widehat{P}_{b,z}^{(i)}, k)$ where $\widehat{P}_{b,z}$ is the sam-

[11]Note that some constraints in the discrimination system may be redundant since the possible vectors for $(\rho(0), \cdots, \rho(n-1))$ do not cover the full spectrum $[1, n] \times [1, n-1] \times \cdots \times [1, 2]$, e.g. $\rho(i+1) \geq \rho(i) - 1$.

ple analog of the polynomial $P_{b,z}$, i.e. with $a_i(b,z)$ being replaced (for $i < n$) by $\widehat{a}_i(b,z) = \frac{n(n-1)\cdots(i+1)}{(n-i)!} \cdot \widehat{F}^{(n-i:n-i)}_{\mathbf{B},\mathbf{Z}}(b,z) \cdot (\widehat{f}_{\mathbf{Z}}(z))^{n-i-1}$. Then various testing statistics can be build to test for some underlying root structure. Popular examples are:

- Kolmogorov-Smirnov-type tests based on suprema, i.e. on $Sup_{b,z}\widehat{\Delta}(P^{(i)}_{b,z}, k)$.
- Tests based on means, i.e. on weighted expectations of $\widehat{\Delta}(P^{(i)}_{b,z}, k)$.

As shown in Proposition 5.1, such tests would in general rely on multiple nonlinear inequality constraints which would lead to testing statistics that are asymptotically distributed as a weighted sum of chi-squared distributions.[12] The related weights depend on the local nature of the inequality constraints as emphasized by Wolak [40]. The noise associated to the estimation of those weights adds some nuisance for inference. From a theoretical perspective, multiple constraints make the derivation of explicit formulas for the asymptotic distribution difficult. For those reasons and also because it is one of the most fundamental test for symmetry, we consider below how to test for full symmetry against the alternative of some asymmetry. In this case, the discrimination system reduces to a single equation as stated in the following corollary and can thus be easily tested with standard one-sided tests.

Corollary 5.2 *A polynomial with real roots P of degree n has the root structure (n) if and only if $\Delta(P,2) = 0$. If P has some distinct roots, then $\Delta(P,2) > 0$.*

Furthermore, we develop a test based on means. Monte Carlo simulations available upon request have shown that a Kolmogorov-Smirnov-type test based on the statistic $Sup_{b,z}\widehat{\Delta}(P_{b,z}, 2)$ has less power to reject the null.

5.2 A Test based on Means

We develop a test of full symmetry, i.e. $F_{\mathbf{B}^*_1,\mathbf{Z}}(.,.) = \cdots = F_{\mathbf{B}^*_n,\mathbf{Z}}(.,.)$ against some asymmetry, i.e. $F_{\mathbf{B}^*_i,\mathbf{Z}}(b,z) \neq F_{\mathbf{B}^*_j,\mathbf{Z}}(b,z)$, for some i and j on a positive measure of b and z. The discriminant $\Delta(P_{b,z}, 2)$ is equal to $n^2(n-1)\left((F^{(1:1)}_{\mathbf{B},\mathbf{Z}}(b,z))^2 - F^{(2:2)}_{\mathbf{B},\mathbf{Z}}(b,z) \cdot f_{\mathbf{Z}}(z)\right)$. From corollary 5.2, our testing hypothesis can be written as:

$$\begin{array}{lccc} H_0 & \text{(full symmetry)} & : & \mathcal{H} = 0 \\ H_1 & \text{(some asymmetry)} & : & \mathcal{H} > 0 \end{array},$$

[12] See Silvapulle and Sen's comprehensive book [35] on constrained statistical inference. Standard theory applies here since the constraints in the discrimination system (19) are Chernoff-regular.

where $\mathcal{H} = \int\int \left((F^{(1:1)}_{\mathbf{B},\mathbf{Z}}(b,z))^2 - F^{(2:2)}_{\mathbf{B},\mathbf{Z}}(b,z) \cdot f_{\mathbf{Z}}(z) \right) dF^{(1:1)}_{\mathbf{B},\mathbf{Z}}(b,z)$. The sample analog of $\mathcal{H}$ is given by:

$$\widehat{\mathcal{H}} = \frac{1}{Ln} \sum_{l=1}^{L} \sum_{p=1}^{n} \left([\widehat{F}^{(1:1)}_{\mathbf{B},\mathbf{Z}}(B_{pl}, Z_l)]^2 - \widehat{F}^{(2:2)}_{\mathbf{B},\mathbf{Z}}(B_{pl}, Z_l) \widehat{f}_{\mathbf{Z}}(Z_l) \right), \tag{20}$$

where $\widehat{f}_{\mathbf{Z}}$, $\widehat{F}^{(1:1)}_{\mathbf{B},\mathbf{Z}}$ and $\widehat{F}^{(2:2)}_{\mathbf{B},\mathbf{Z}}$ are defined according to the first and second step of our estimation procedure, i.e. from equations (7-10). Next proposition establishes the consistency and asymptotic normality of $\widehat{\mathcal{H}}$.

Proposition 5.3 *Suppose that $K_{F_{\mathbf{B_p}|\mathbf{Z}}}$ and $K_{f_{\mathbf{Z}}}$ are kernels and $h_{F_{\mathbf{B_p}|\mathbf{Z}}}$ and $h_{f_{\mathbf{Z}}}$ converges to zero as $L \to \infty$.*

$$\sqrt{L}.\left(\widehat{\mathcal{H}} - \mathcal{H} \right) \to_d \mathcal{N}(0, \Sigma).$$

Under H_0, we have $\Sigma^2 = E_{Z^t}[[f(Z^t)]^4]/(45n(n-1))$. Without covariates, the expression of Σ^2 is reduced to $\frac{1}{45n(n-1)}$.

Define the test statistic: $t = \frac{\sqrt{L}.\widehat{\mathcal{H}}}{\sqrt{\widehat{\Sigma^2}}}$, where $\widehat{\Sigma^2} = \frac{1}{45n(n-1)} \cdot \frac{1}{L} \sum_{l=1}^{L} [\widehat{f}(Z_l)]^4$ is the sample analog of Σ^2. $\widehat{\Sigma^2}$ is a consistent estimate of Σ^2 converging at the rate $\sqrt{L}$ with some covariates. Without covariates, Σ^2 is known. Symmetry can not be rejected at the level α if and only if the test statistic t is smaller than $q_{1-\alpha}$, the $(1-\alpha)$ quantile of the normal distribution. Without covariates, the distribution of the test statistic is not only asymptotically distribution free but fully distribution free under H_0 as it can be checked directly from its general expression and tests should better rely on the simulated quantiles of the test statistic than on their asymptotic approximations. However, the asymptotic approximation -provided that its accuracy is satisfactory- can be useful for the following exercise: for a given form of asymmetry and for a given level α, we are interested in finding the necessary size of the data to reach a given power of rejection. Asymptotic approximations seem very accurate for small data set as reported in Table 2 for $L = 40$ and $L = 200$ and different values of n. Under the alternative H_1, we have no tractable asymptotic approximation for the standard deviation of the test statistic. However, the median of the test statistic coincides asymptotically with the mean which is known. Finally, we obtain the following corollary about the way to reach the power 50%.[13]

[13]Unreported simulations show that the variance Σ^2 under H_0 is a good approximation of the

Corollary 5.4 *Asymptotically, our test reject symmetry with a probability greater than one half if and only if the variable* $\mathcal{H}$ *is greater than* $\frac{q_{1-\alpha}\Sigma}{\sqrt{L}}$.

Equivalently, it says that for a given degree of asymmetry $\mathcal{H} > 0$, the necessary size L^* of the data to reject symmetry at the level α with probability at least one half is approximately $\left(\frac{q_{1-\alpha}\Sigma}{\mathcal{H}}\right)^2$. Without covariates, the expression simplifies to:[14]

$$L^* = q_{1-\alpha}^2 \cdot \frac{4(n-1)}{45n} \cdot \left[\frac{1}{n^2} \cdot \sum_{i=1}^{n} \sum_{j=1}^{n} E[\left(F_{\mathbf{B}_\mathbf{i}^*}(b) - F_{\mathbf{B}_\mathbf{j}^*}(b)\right)^2]\right]^{-2},$$

where the expectation is for b distributed according to the CDF $\sum_{i=1}^{n} F_{\mathbf{B}_\mathbf{j}^*}(.)/n$. This expression gives some insights on how the structure of the data affect the power of our test. First, if the number of bidders is duplicated while maintaining the same asymmetry structure which is reflected by the term is bracket, then L^* is proportional to $(n-1)/n$ and thus increasing in n: expanding the number of bidders per auction makes the detection of asymmetry more difficult. Two effects are at work. On the one hand, expanding the number bidders per auction expands the data which is beneficial for the accuracy of estimation under nonanonymous data. On the other hand, under anonymous data, more bidders per auction weakens the link between a bid and the identity of the bidder. On the whole, we have shown that it is the second effect that dominates while both effects are canceling each other asymptotically when n is large. Second, the term in bracket shows that asymmetry is easier to detect among 'balanced' panels of bidders. Consider some asymmetry involving two kinds of bidders (Strong and Weak) with the respective CDFs F_S and F_W. Let $k \in [1, n-1]$ be the number of Strong bidders. Then the term in bracket is equal to $\frac{2k(n-k)}{n^2} \cdot \left(E[(F_S(b) - F_W(b))^2]\right)^{-2}$. As a function of k, this term is symmetric with respect to $k = n/2$: it is decreasing from 1 to $n/2$ and then increasing. The 'balanced' panel with $k = [n/2]$ is the best one to reject symmetry.

Numerical application To emphasize that our test has some practical relevance we compute L^* for CDFs that have been estimated nonparametrically by Flambard and Perrigne [8] from auction data for snow removal contracts in Montréal.[15] We

ones under H_1, which suggests a 'practical' approximation for similar computations for any power.

[14]Another expression for $\mathcal{H}$ is $\frac{1}{2n^2(n-1)} \int\int \sum_{i=1}^{n} \sum_{j=1}^{n} \left(F_{\mathbf{B}_\mathbf{i}^*}(b) - F_{\mathbf{B}_\mathbf{j}^*}(b)\right)^2 d\left(\frac{1}{n} \cdot \sum_{k=1}^{n} F_{\mathbf{B}_\mathbf{k}^*}(b)\right)$

[15]Flambard and Perrigne generously provided us with those CDFs for the median covariate. The asymmetry involves two kinds of bidders according to the location of the contract and the location

	$L = 40$			$L = 200$		
n	2	4	6	2	4	6
share of p-values $< 10\%$	0.13	0.13	0.12	0.11	0.11	0.10
share of p-values $< 5\%$	0.06	0.05	0.06	0.05	0.06	0.05

Table 2: Performance of the asymptotic version of the test based on the Means. 5000 replications for the simulated statistics.

obtain $L^* = 190$ ($L^* = 115$) for the 5% level (10% level). Those figures do not vary much when we slightly perturbate the structure of the bidders. For 8 [resp. 4] bidders while 4 [resp. 2] being Strong bidders, we obtain $L^* = 199$ and $L^* = 120$ [$L^* = 171$ and $L^* = 104$] for the 5% and 10% levels.

Contrary to our estimation procedure, our testing procedure does not use additional information on the identity of the bidders. Indeed, a natural procedure (left for further research) would be to implement some standard test for symmetry on the pseudo sample of bids and probabilities which would then use any supplementary information on bidders' identities. Several reasons have motivated our approach. First, the computation of the pseudo-probabilities involve the computation of the bidding densities whereas our approach uses only CDFs of the bidding distributions. Second, the tractability of the asymptotic variance under the null. Finally, without covariates, our approach is distribution free. However, the use of such additional information might be very useful in practice especially if the power is an issue.

6 Asymptotic Properties

6.1 Regularity Assumptions and Key Properties

The next assumptions concern the underlying generating process as well as the smoothness of the latent joint distribution of $(X_{il}, Z_l, \mathrm{I}_l)$ for any $i \in \mathrm{I}_l$.

Assumption A 2 (i) *The $(d+1)$-dimensional vectors $(Z_l, \mathrm{I}_l), l = 1, 2, \cdots$, are independently and identically distributed as $F_{\mathbf{Z},\mathbf{I}}(.,.)$ with density $f_{\mathbf{Z},\mathbf{I}}(.,.)$.*

of the firm in the city. They found evidence for asymmetry between West Firms and Non West Firms for the Western contracts (as depicted in their Figure 2). In their data with $L = 29$, the structure of the bidders contains some variation from one auction to another. We simplify the structure for the computation doing as if all auctions have 6 bidders while 3 being Strong bidders and considering no variation in the covariates.

(ii) *For each l the variables $X_{il}, i \in \mathrm{I}_l$ are independently distributed conditionally upon (Z_l, I_l) as $F_{\mathbf{X_i|Z,I}}(.|.,.)$ with density $f_{\mathbf{X_i|Z,I}}(.|.,.)$, for $i \in \mathrm{I}_l$.*

As in Campo et al. [6], we consider here that the support of buyers' private values does not depend on the (Z, I) to simplify the presentation, while the general case can be fully treated as in GPV. It implies that the lower bound of the support of buyers' bids does not depend on the variables I and Z. Throughout we denote by $S(*)$ and $S^o(*)$ the support of $*$ and its interior, respectively.

Assumption A 3 *For each bidder $i \in I \subset \mathcal{I}$,*

(i) *$S(F_{\mathbf{X_i,Z,I}}) = \{(x, z, I) : z \in [\underline{z}, \overline{z}], x \in [\underline{x}, \overline{x}], I \subset \mathcal{I}\}$; with $\underline{z} < \overline{z}$;*

(ii) *for $(x, z, I) \in S(F_{\mathbf{X_i,Z,I}})$, $f_{\mathbf{X_i|Z,I}}(x|z, I) \geq c_f > 0$, and for $(z, I) \in S(F_{\mathbf{Z,I}})$, $f_{\mathbf{Z,I}}(z, I) \geq c_f > 0$;*

(iii) *for each $I \subset \mathcal{I}$, $F_{\mathbf{X_i|Z,I}}(.|., I)$ and $f_{\mathbf{Z,I}}(., I)$ admit up to $R+1$ continuous bounded partial derivatives on $S(F_{\mathbf{X_i,Z,I}})$ and $S(F_{\mathbf{Z,I}})$, with $R \geq 1$.*

The next assumption is not necessary for identification as established in Proposition 3.1 without heterogeneity across objects. Nevertheless, heterogeneity requires an additional structure to identify the model. Similar intersections as the one in Figure 1 when b varies may arise when the variable capturing heterogeneity Z varies. But the different solutions are observationally equivalent without some mild additional assumptions. Here to preserve identification, we make the assumption that bidding distributions can be ordered according to first order stochastic dominance.[16] Moreover, to simplify our estimation procedure, we also assume that the dominance is strict in the interior of the bidding support.[17]

Assumption A 4 (Strict Stochastic Dominance) *The bid densities $F_{\mathbf{B_i^*|Z,I}}(.|z, I)$ can be strictly ordered according to first order stochastic dominance: $F_{\mathbf{B_i^*|Z,I}}(b|z, I) > F_{\mathbf{B_j^*|Z,I}}(b|z, I)$ if $b \in S^0(f_{\mathbf{B_i^*|Z,I}})$ for any $i, j \in I$ with $j > i$ and any z,I.*

[16] Alternative identification strategies could be to make assumptions on the comparative statics of the bidding distribution according to Z or use the point that, generically, at an intersection, only one candidate solution is differentiable at this point.

[17] Assumption 4 is not on the primitives of the model in the first price auction. With two classes of bidders, Maskin and Riley [25] show that 'conditional stochastic dominance' of private values' distributions is a sufficient condition for first order stochastic dominance of bids' distribution.

A crucial step in deriving uniform rates of convergence in some inverse problem is to study the smoothness of the observables that is implied by the smoothness of the latent distributions of the primitives of the model. Here, relative to GPV, we do not observe the vector of bids B^* but only the vector of bidding order statistic B. Thus we are interested in the smoothness of the densities $f_{\mathbf{B_p}|\mathbf{Z},\mathbf{I}}(.|.,I)$ for $p=1,\cdots,n_I$. This is the purpose of the next proposition. It is the analog of proposition 1 in GPV which derives similar results for the bid densities $f_{\mathbf{B_i^*}|\mathbf{Z},\mathbf{I}}(.|.,.)$.

Proposition 6.1 *Given A3, the conditional distribution* $F_{\mathbf{B_p}|\mathbf{Z},\mathbf{I}}(.|.,I)$, $p=1,\cdots,n_I$ *and* $I\subset\mathcal{I}$, *satisfies for both the first and second price auctions (if not specified):*

(i) *its support* $S(F_{\mathbf{B_p}|\mathbf{Z},\mathbf{I}})$ *is such that* $S(F_{\mathbf{B_p}|\mathbf{Z},\mathbf{I}}) = \{(b,z,I) : z\in[\underline{z},\overline{z}], b\in[\underline{b}(z,I,p),\overline{b}(z,I,p)], I\subset\mathcal{I}\}$ *with* $\overline{b}(z,I,p)>\underline{b}(z,I,p)$ *for any* I,p. *Moreover,* $(\underline{b}(.,I,p),\overline{b}(.,I,p))$ *admit up to* $R+1$ *continuous bounded derivatives on* $[\underline{z},\overline{z}]$ *for each* $I\subset\mathcal{I}$ *and* $p=1,\cdots,n_I$. *We have* $\underline{b}(z,I,p)=\underline{x}$. *In the second price auction,* $\overline{b}(z,I,p)=\overline{x}$. *In the first price auction* $\overline{b}(z,I,n_I)=\overline{b}(z,I,n_I-1)$.

(ii) *for* $(b,z,I)\in\mathcal{C}(B_n)$, $f_{\mathbf{B_p}|\mathbf{Z},\mathbf{I}}(b,z,I)\geq c_{\mathbf{B_p}|\mathbf{Z},\mathbf{I}}>0$, *where* $\mathcal{C}(B_n)$ *is a closed subset of* $S^0(F_{\mathbf{B_n}|\mathbf{Z},\mathbf{I}})$*;*

(iii) *for each* (I,p), $p=1,\cdots,n_I$, $F_{\mathbf{B_p}|\mathbf{Z},\mathbf{I}}(.|.,I)$ *admits up to* $R+1$ *continuous bounded partial derivatives on* $S(F_{\mathbf{B_p}|\mathbf{Z},\mathbf{I}})\setminus(\{\overline{b}(z,I,p)\}_{p=1,\cdots,n_I-1})$*;*

(iv) *in the first price auction, for each* (I,p), $p=1,\cdots,n_I$, *if* $\mathcal{C}(B_p)$ *is a closed subset of* $S^o(F_{\mathbf{B_p}|\mathbf{Z},\mathbf{I}})\setminus(\{\overline{b}(z,I,p)\}_{p=1,\cdots,n_I})$, *then* $f_{\mathbf{B_p}|\mathbf{Z},\mathbf{I}}(.|.,I)$ *admits up to* $R+1$ *continuous bounded partial derivatives on* $\mathcal{C}(B_p)$*;*

(v) *in the second price auction, for each* (I,p), $p=1,\cdots,n_I$, $f_{\mathbf{B_p}|\mathbf{Z},\mathbf{I}}(.|.,I)$ *admits up to* R *continuous bounded partial derivatives on* $S(F_{\mathbf{B_p}|\mathbf{Z},\mathbf{I}})\setminus(\{\overline{b}(z,I,p)\}_{p=1,\cdots,n_I-1})$.

Note that by comparing (iv) and (v), the bid densities in the first price auction are smoother than for the second price auction. Thus $f_{\mathbf{B_p}|\mathbf{Z},\mathbf{I}}(.|.,I)$ can be estimated uniformly at a faster rate, namely $(L/\log L)^{(R+1)/(2R+d+3)}$, in the first price than in the second price auction, namely $(L/\log L)^{R/(2R+d+1)}$. In particular, the optimal bandwidths -that we specify later in assumption A6- are asymptotically smaller for the second price auction than for the first price auction. Nevertheless the optimal uniform convergence rate will be smaller in the first price auction than in the second price auction. This is due to the more indirect nature of the link between observables and latent distributions in the first price auction, see equation (6).

Proposition 6.1 differs from the one appearing in GPV as irregularities of the CDFs of the order statistic may appear in the interior of their support, more precisely we may have $\bar{b}(z, I, p) < \bar{b}(z, I, n)$ for $p \leq n_I - 2$. In the following, to alleviate notation, we make the simplifying assumption A5 that the bidding supports of all bidders coincide, i.e. $\bar{b}(z, I, p)$ does not depend on p. Our uniform consistency results extend provided that the neighborhoods of the bidders' signals than make them bid $\bar{b}(z, I, p)$ are removed. In the same way as the support of bidders' private values is consistently estimated in GPV and that the neighborhoods of the lower and upper bounds of the support are removed with an suitable trimming, we can trim those inner neighborhoods.

Assumption A 5 (Common bidding support) *All bidders have the same bidding support: $\bar{b}(z, I, p)$ does not depend on p.*

6.2 Uniform Consistency

Our main result establishes the uniform consistency of our multistage kernel-based estimators for the first and second price auctions and with their rates of convergence. As a preliminary step, we first set our choice of kernels and bandwidths and then establish in proposition 6.2 the uniform consistency with their rates of convergence of our nonparametric estimators of the upper and lower boundaries $\bar{b}(z, I)$ and $\underline{b}$ and also the rates at which the pseudo private values $\widehat{X}_{ipl}$ and the pseudo probabilities $\widehat{P}_{ipl}$ converge uniformly to their true values. This proposition is the analog of propositions 2 and 3 in GPV.

Assumption A 6 • ***KERNELS***

(i) *The kernels $K_{F_{\mathbf{B_p|Z}}}(.)$, $K_{f_{\mathbf{B_p|Z}}}(.,.)$, $K_{f_{\mathbf{X_i,Z}}}(.,.)$ and $K_{f_{\mathbf{Z}}}(.)$ are symmetric with bounded hypercube supports of length equal to 2 and continuous bounded first derivatives with respect to their continuous argument.*

(ii) $\int K_{f_{\mathbf{Z}}}(z)dz = 1$, $\int K_{F_{\mathbf{B_p|Z}}}(z)dz = 1$, $\int K_{f_{\mathbf{B_p|Z}}}(b, z)dbdz = 1$, *for any* $p = 1, \cdots, n$ *and* $\int K_{f_{\mathbf{X_i,Z}}}(x, z)dxdz = 1$ *for any* $i = 1, \cdots, n$.

(iii) $K_{F_{\mathbf{B_p|Z}}}(.)$, $K_{f_{\mathbf{B_p|Z}}}(.,.)$, $K_{f_{\mathbf{X_i,Z}}}(.,.)$ *and* $K_{f_{\mathbf{Z}}}(.)$ *are of order* $R+1$, $R+1$,R *and* $R+1$ *respectively, i.e. moments of order strictly smaller than the given order vanish.*

- ***BANDWIDTHS***

(i) *In the first price auction, the bandwidths* $h_{F_{\mathbf{B_p|Z}}}, h_{f_{\mathbf{B_p|Z}}}$*, for* $p = 1, \cdots, n$*,* $h_{f_{\mathbf{X_i,Z}}}$ *for* $i = 1, \cdots, n$ *and* $h_{f_{\mathbf{Z}}}$ *are of the form:*

$$h_{F_{\mathbf{B_p|Z}}} = \lambda_{F_{\mathbf{B_p|Z}}}(\frac{\log L}{L})^{\frac{1}{(2R+d+2)}}, \quad h_{f_{\mathbf{B_p|Z}}} = \lambda_{f_{\mathbf{B_p|Z}}}(\frac{\log L}{L})^{\frac{1}{(2R+d+3)}},$$
$$h_{\mathbf{f_{X_i,Z}}} = \lambda_{\mathbf{f_{X_i,Z}}}(\frac{\log L}{L})^{\frac{1}{(2R+d+3)}}, \quad h_{f_{\mathbf{Z}}} = \lambda_{f_{\mathbf{Z}}}(\frac{\log L}{L})^{\frac{1}{(2R+d+2)}},$$

where the λ*'s are strictly positive constants.*

(ii) *In the second price auction, the bandwidths* $h_{F_{\mathbf{B_p|Z}}}, h_{f_{\mathbf{B_p|Z}}}$*, for* $p = 1, \cdots, n$*,* $h_{f_{\mathbf{X_i,Z}}}$ *for* $i = 1, \cdots, n$ *and* $h_{f_{\mathbf{Z}}}$ *are of the form:*

$$h_{F_{\mathbf{B_p|Z}}} = \lambda_{F_{\mathbf{B_p|Z}}}(\frac{\log L}{L})^{\frac{1}{(2R+d)}}, \quad h_{f_{\mathbf{B_p|Z}}} = \lambda_{f_{\mathbf{B_p|Z}}}(\frac{\log L}{L})^{\frac{1}{(2R+d+1)}},$$
$$h_{\mathbf{f_{X_i,Z}}} = \lambda_{\mathbf{f_{X_i,Z}}}(\frac{\log L}{L})^{\frac{1}{(2R+d+1)}}, \quad h_{f_{\mathbf{Z}}} = \lambda_{f_{\mathbf{Z}}}(\frac{\log L}{L})^{\frac{1}{(2R+d+2)}},$$

(iii) *The "boundary" bandwidth is of the form* $h_\delta = \lambda_\delta(\frac{\log L}{L})^{\frac{1}{d+1}}$ *if* $d > 0$ *where the* λ*'s are strictly positive constants.*

As in GPV and for both formats, $h_{F_{\mathbf{B_p|Z}}}$, $h_{f_{\mathbf{B_p|Z}}}$ and $h_{f_{\mathbf{Z}}}$ are corresponding to the standard optimal bandwidths such that the related estimated densities are converging uniformly at the best possible rate.

Proposition 6.2 *Under A1-A6, for any closed subset* $\mathcal{C}$ *of* $S^o(F_{\mathbf{X},\mathbf{Z},\mathrm{I}})$*, we have almost surely* $\sup_{(z,I)\in[\underline{z},\overline{z}]\times\mathcal{I}} |\widehat{\overline{b}}(z,I) - \overline{b}(z,I)| = O(\frac{\log L}{L})^{\frac{1}{d+1}}$ *and* $|\widehat{\underline{b}} - \underline{b}| = O(\frac{\log L}{L})^{\frac{1}{d+1}}$ *for both the first and second price auctions. The pseudo values and pseudo probabilities satisfy almost surely:*

$$(i) \qquad \sup_{i,p,l} \mathbf{1}_{\mathcal{C}}(X_{ipl}, Z_l, \mathrm{I}_l)|\widehat{X}_{ipl} - X_{ipl}| = O\left((\frac{\log L}{L})^{\frac{R+1}{(2R+d+3)}}\right)$$
$$(ii) \qquad \sup_{i,p,l} \mathbf{1}_{\mathcal{C}}(X_{ipl}, Z_l, \mathrm{I}_l)|\widehat{P}_{ipl} - P_{ipl}| = O\left((\frac{\log L}{L})^{\frac{R+1}{(2R+d+3)}}\right)$$

in the first price auction and

$$(i) \qquad \sup_{i,p,l} \mathbf{1}_{\mathcal{C}}(X_{ipl}, Z_l, \mathrm{I}_l)|\widehat{X}_{ipl} - X_{ipl}| = 0$$

$$(ii) \qquad \sup_{i,p,l} \mathbf{1}_{\mathcal{C}}(X_{ipl}, Z_l, \mathrm{I}_l)|\widehat{P}_{ipl} - P_{ipl}| = O\left((\frac{\log L}{L})^{\frac{R}{(2R+d+1)}} \right)$$

in the second price auction.

In the same way as the vector of pseudo private values is not sufficient to estimate the CDFs of each bidders private values (on the contrary to GPV), the estimation of conditional mean, variance or quantiles of a given bidder's private values would also require the joint use of the pseudo private values with the associated vector of pseudo probabilities. We now state our main result. The study of uniform convergence is restricted to inner closed subset of the support to avoid boundary effects.

Proposition 6.3 *Suppose that A1-A6 hold, then* $(\widehat{f}_{\mathbf{X_1}|\mathbf{Z},\mathbf{I}}(.|.,.), \cdots, \widehat{f}_{\mathbf{X_n}|\mathbf{Z},\mathbf{I}}(.|.,.))$ *is uniformly consistent as* $L \to \infty$ *with rate* $(L/\log L)^{R/(2R+d+3)}$ *on any inner compact subset of the support of* $(f_{\mathbf{X_1}|\mathbf{Z},\mathbf{I}}(.|.,.), \cdots, f_{\mathbf{X_n}|\mathbf{Z},\mathbf{I}}(.|.,.))$ *in the first price auction and respectively the rate* $(L/\log L)^{R/(2R+d+1)}$ *in the second price auction.*

In addition to establishing the uniform consistency of our multi-step estimator, we show in the supplementary material that our estimation procedure of the conditional density $F_{\mathbf{X}|\mathbf{Z},\mathbf{I}}(.|.,.)$ in the first and second price auctions under anonymous data reaches the asymptotic optimal rates. At first glance, it seems immediate since the rates derived in proposition 6.3 correspond precisely to the rates derived by GPV which were shown to be optimal when the data is not anonymous. However, the optimality property derived in GPV has been obtained for the symmetric IPV model while we are considering the asymmetric bidders.

Note that if the interest of the econometrician lies in the estimation of the distributions $F_{\mathbf{B}^*|\mathbf{Z},\mathbf{I}}(.|.,.)$, then, in the first price auction, our bandwidths are suboptimal and the same bandwidths as those for the second price auction should be used. We present the proof of Proposition 6.3 as it helps to identify the additional points relative to GPV's procedure and why it does not change the asymptotical rates of convergence.

Proof We have $\widehat{f}_{\mathbf{X_i}|\mathbf{Z},\mathbf{I}}(x|z,I) = \widehat{f}_{\mathbf{X_i},\mathbf{Z},\mathbf{I}}(x,z,I)/\widehat{f}_{\mathbf{Z},\mathbf{I}}(z,I)$. Given the optimal bandwidth choice for $h_{f_{\mathbf{Z}}}$ in assumption A6, we know that $\widehat{f}_{\mathbf{Z},\mathbf{I}}(z,I)$ converges uniformly to $f_{\mathbf{Z},\mathbf{I}}(z,I)$ at the rate $(L/\log L)^{(R+1)/(2R+d+1)}$ on any inner compact of its support. Because this rate is faster than that of the theorem (for both auction formats) and $f_{\mathbf{Z},\mathbf{I}}(z,I)$ is bounded away from 0 by assumption A3-(ii), it suffices to show that $\widehat{f}_{\mathbf{X_i},\mathbf{Z},\mathbf{I}}(x,z,I)$ converges at the rate $(\frac{\log L}{L})^{R/(2R+d+3)}$ and $(\frac{\log L}{L})^{R/(2R+d+1)}$ in the first and second price auctions respectively. We turn back that the way we have decomposed the difference $\widehat{f}_{\mathbf{X_i},\mathbf{Z},\mathbf{I}}(x,z,I) - f_{\mathbf{X_i},\mathbf{Z},\mathbf{I}}(x|z,I)$ in equation (17) and analyze the convergence rate of the three error terms.

In the second price auction, the bandwidth $h_{\mathbf{f}_{\mathbf{X_i}},\mathbf{Z}}$ is optimal and thus leads to a uniform convergence of $\widetilde{f}_{\mathbf{X_i},\mathbf{Z},\mathbf{I}}(x,z,I)$ to $f_{\mathbf{X_i},\mathbf{Z},\mathbf{I}}(x,z,I)$ at the rate $(L/\log L)^{R/(2R+d+1)}$ in any inner compact of its support. In the first price auction, the suboptimal bandwidth leads to the rate $(L/\log L)^{R/(2R+d+3)}$ as in GPV. Thus we are left with the first two terms ε_1 and ε_2, the first one resulting explicitly from the anonymous nature of the bids is new, whereas the second term appears already in GPV.

First consider the second price auction. Since $\widehat{X}_{ipl} = X_{ipl}$, the second term vanishes and we are left with the first term

$$\frac{1}{Lh^{d+1}_{f_{\mathbf{X_i}},\mathbf{Z}}}\sum_{l=1}^{L}\sum_{p=1,\cdots,n_{\mathrm{I}_l}}(\widehat{P}_{ipl}-P_{ipl})\cdot K_{f_{\mathbf{X_i}},\mathbf{Z}}(\frac{x-X_{ipl}}{h_{f_{\mathbf{X_i}},\mathbf{Z}}},\frac{z-Z_l}{h_{f_{\mathbf{X_i}},\mathbf{Z}}})\cdot\mathbf{1}(I_l=I),$$

which is bounded by:

$$\left(\sup_{p,l}\mathbf{1}_{\mathcal{C}}(X_{ipl},Z_l,\mathrm{I}_l)|\widehat{P}_{ipl}-P_{ipl}|\right)\cdot\left[\frac{1}{Lh^{d+1}_{f_{\mathbf{X_i}},\mathbf{Z}}}\sum_{l=1}^{L}\sum_{p=1,\cdots,n_{\mathrm{I}_l}}|K_{f_{\mathbf{X_i}},\mathbf{Z}}(\frac{x-X_{ipl}}{h_{f_{\mathbf{X_i}},\mathbf{Z}}},\frac{z-Z_l}{h_{f_{\mathbf{X_i}},\mathbf{Z}}})|\cdot\mathbf{1}(I_l=I)\right].$$

The above term appearing in the bracket may be viewed as a kernel estimator, and hence converges uniformly on $\mathcal{C}$ to $\sum_{p=1,\cdots,n_{\mathrm{I}}} f_{\mathbf{X_{ip}},\mathbf{Z},\mathbf{I}}(x,z,I)\cdot\int|K_{f_{\mathbf{X_i}},\mathbf{Z}}(x,z)|dxdz$. Thus this term stays bounded almost surely. Finally the difference $\widehat{f}_{\mathbf{X_i},\mathbf{Z},\mathbf{I}}(x,z,I) - f_{\mathbf{X_i},\mathbf{Z},\mathbf{I}}(x,z,I) = O(\log L/L)^{R/(2R+d+1)}$.

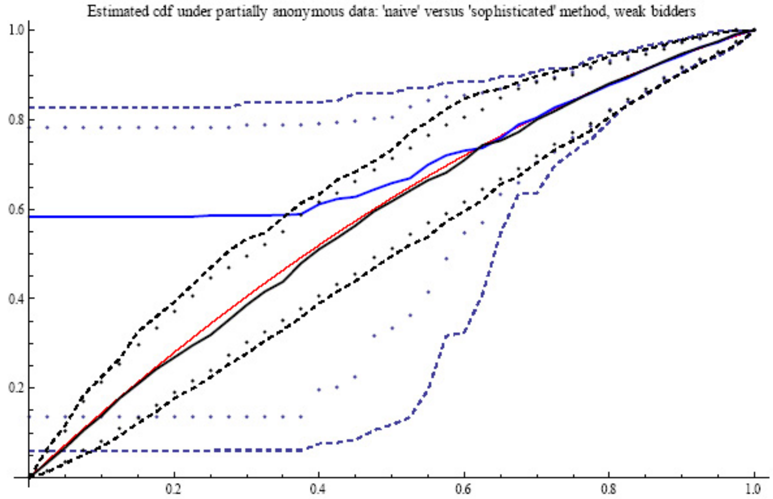
In the first price auction, similarly to GPV, a first-order Taylor expansion establishes that ε_2 has the order $O(\log L/L)^{R/(2R+d+3)}$, whereas the same argument as above establishes that ε_1 has the order $O(\log L/L)^{(R+1)/(2R+d+3)}$. Thus with anonymity, it is still the second error term that results from the gap between estimated and real private values that is the 'binding' term relative to the uniform convergence rate. **CQFD**

7 Monte Carlo Experiments

To illustrate the usefulness of our procedure, we conduct a limited Monte Carlo study.[18] To fit with realistic sizes of auction data, we consider $L = 40$ auctions, each having at most 6 bidders. Our Monte Carlo experiments consist of respectively 200 and 5000 replications for our estimation and testing procedures for the second price auction when the identity of the winner is observed. The true distribution of private values $F_{\mathbf{X}}$ is generated from the densities f_ϵ on the support $[0, 1]$ where $f_\epsilon(x) = (1 + \epsilon \cdot (1 - 2x)) \cdot \mathbf{1}_{0 \leq x \leq 1}$ (ϵ is a parameter in $[-1, 1]$).

Figure 2: asymmetric IPV model

Estimated cdf under partially anonymous data: 'naive' versus 'sophisticated' method, weak bidders



Estimation In Figure 2, which summarizes our results for the estimators of the CDF of the weak bidders, the underlying (true) model is the asymmetric IPV model with 3 (strong) bidders with $\epsilon = -\frac{1}{2}$ and 3 (weak) bidders with $\epsilon = \frac{1}{2}$. The true CDF is displayed in plain red line. For the interval $[0, 1]$, the median (full line), the 5 and 95 percentiles (dashed lines) and the 10 and 90 percentiles (dots) of our estimates of the CDF of the weak bidders are displayed in black. This gives the (pointwise) 80% and 90% confidence intervals. Figure 2 also displays in blue lines the corresponding results under the 'naive' estimation procedure that drops the bids that are anonymous in the data set. In the first-price auction, the 'naive' approach would correspond to treat the data as the one resulting from a Dutch auction which is identified under the independence assumption, see Athey and Haile [2] for identification where results

[18] Practical details on the implementation and additional Monte Carlo simulations are reported in the supplementary material. Programs are written in Mathematica and are available upon request.

from the competing risk literature are applied and Paarsch and Hong [30] p.153-155 for natural estimators that are asymptotically consistent.[19] The results are striking. By keeping only the highest bid, the 'naive' approach can not draw any inference on the lowest tail of the distribution for which bids are practically never recorded with 6 bidders. This is especially true for the weak bidders for which the estimator is too noisy to have any practical interest and is also seriously biased for about one half of the distribution. On the contrary, our estimation procedure does a good job: the median of the estimates perfectly matches the true curve and the 80% confidence intervals are much smaller. In a nutshell, our procedure outperforms the 'naive' approach for the whole support of the distribution, though it is less striking at the upper tail of the distribution.

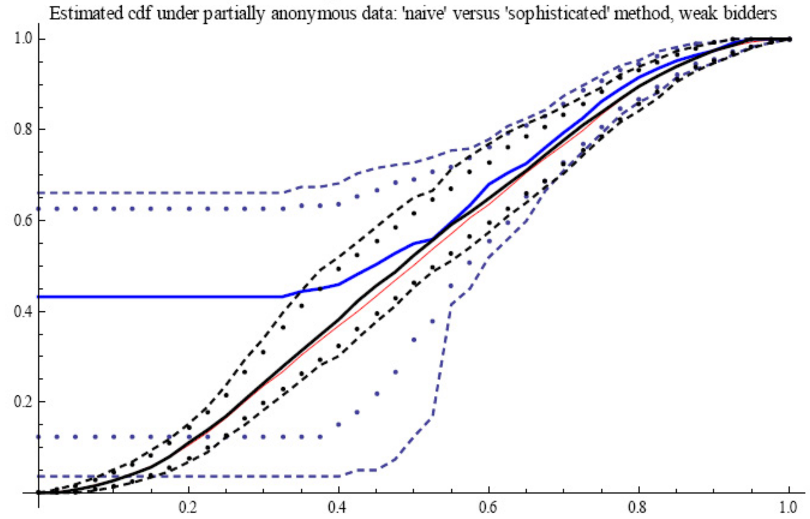
The simulations reported in Figure 3 are devoted to a kind of robustness check. Our 'sophisticated' estimation approach and the 'naive' approach are both relying on the independence assumption. We consider a departure from this assumption: the underlying (true) model is a symmetric correlated PV model with 6 bidders.[20] The legend is the same as for Figure 2. The results in Figure 3 provide another argument in favor of our estimation procedure compared to the 'naive' approach. If we wrongly assume that the sampling scheme is an independent asymmetric model whereas it is indeed a symmetric correlated model, then our procedure leads to accurate unbiased estimates. On the contrary, the 'naive' approach remains flawed: it does not solely fail to give practically useful confidence intervals for the lower tail of the distribution but it is also strongly biased on all the support since it is mislead by the way it exploits the independence assumption -this bias is not a byproduct of the limited sample size as it can be checked with bigger sample sizes. This contrasts with our methodology which implicitly switches to the estimation of the symmetric PV model when bids are positively correlated. By taking the real part of the estimated roots in equation (11), our procedure (at least partially) drops the use of the independence assumption when we estimate complex roots as it happens with positive correlation.[21]

[19] We emphasize that the terminology 'naive' refers to the way anonymous bids are thrown away.

[20] Bidders' values are constructed in the following way. They correspond to a weighted sum between a common shock and an idiosyncratic shock that is associated to each bidder. Shocks are supposed to be independent and uniformly distributed on $[0, 1]$. The weight on the common shock ρ is fixed here to $\rho = 0.25$ such that bidders' values are positively correlated.

[21] A rigorous formalization of this point is left for future research.

Figure 3: symmetric APV model



A reader familiar with the numerical analysis literature which analyzes the sensitivity of the roots of a polynomial with respect to small perturbations to its coefficients could legitimately have serious doubts about the practical relevance of our estimation procedure.[22] Such issues do not seem to prevent the usefulness of our analysis. Note that our application involves polynomials of low degree. Unreported simulations with polynomials of degree 3 show that our methodology still work.

Tests for symmetry For the testing procedure, both the number of bidders and the choice for ϵ vary, e.g. in table 3 the third column '1/2' and the two last rows correspond to 3 bidders: 1 (strong) with $\epsilon = -1$ and 2 (weak) bidders with $\epsilon = 1$. Moreover the simulations are also reported for $L = 200$. The power properties of our test are summarized in Table 3 and are illustrating the theoretical results of section 5: the comparative statics of the power with respect to the data structure that were relying on asymptotic approximations are confirmed by our simulations and the crucial importance of the degree of asymmetry, e.g. from $\epsilon = \pm\frac{1}{2}$ to $\epsilon = \pm 1$, the power goes from 0.20 to 0.78 for $L = 40$.

[22] Wilkinson's polynomial $u \to \prod_{k=1}^{20}(u-k)$ is the classic example where a perturbation of 2^{-23} in the second leading coefficient of a polynomial whose roots are distant from unity leads to first-order perturbations of the roots: the root at $x = 20$ grows to $x \approx 20.8$ and the roots at $x = 18$ and $x = 19$ collide into a double root. See Gautschi [9] and Mosier [27] for more on this topic.

Structure	1/1		1/2		1/3		2/2		3/3	
Range of L	40	200	40	200	40	200	40	200	40	200
Degree of asymmetry										
Distribution, $\epsilon = \pm\frac{1}{2}$										
share of p-values $< 10\%$	0.20	0.44	0.16	0.31	0.15	0.21	0.16	0.31	0.14	0.25
share of p-values $< 5\%$	0.12	0.31	0.09	0.19	0.08	0.12	0.09	0.19	0.07	0.14
Distribution, $\epsilon = \pm 1$										
share of p-values $< 10\%$	0.78	1.00	0.54	0.98	0.30	0.67	0.57	0.99	0.41	0.91
share of p-values $< 5\%$	0.67	1.00	0.39	0.94	0.18	0.50	0.43	0.97	0.27	0.82

Table 3: Monte Carlo Results. Test based on Means. 5000 replications for each experiment.

8 Conclusion

This work has been limited to the IPV model with risk neutral bidders, no reserve price and a complete set of bids. All our analysis of the first-price auction can be adapted to risk averse bidders under a conditional quantile restriction and a parametrization of bidders' utility function following Campo et al. [6]. As in GPV, our analysis can also be adapted to a binding reserve price provided that we are prepared to assume that the number of potential bidders is constant. Naturally, identification is obtained only for the truncated distribution of types that are above the reserve price. More involved is the extension of our methodology with incomplete sets of bids or with an unobserved (exogenous) set of participants, whose developments are left for further ongoing research.[23] E.g. in the second price auction, we can be reluctant to propose identification and estimation methods that are relying on the observation of the complete set of bids, in particular on the observation of the highest bid which may remain unobserved. Moreover, this excludes any direct application for the English auction. Let us briefly precise the different issues: first how to adapt our own estimation methodology whose central step involves the computation, for any x, of the vector $(F_{\mathbf{B}_\mathbf{i}^*}(x))_{i=1,\cdots,n}$ as a function of the vector $(F_{\mathbf{B}}^{(i:i)}(x))_{i=1,\cdots,n}$, a problem which has been shown to be related to the computation of the roots of a polynomial as a function of its coefficients under the key assumption that private values are independently distributed ; second how to deal more generally with identification, estimation and testing using alternatives routes that are exploiting the full joint distribution of the order-statistics $F_{\mathbf{B}}$.

[23]With incomplete sets of bids, assuming independence seems the unique 'natural' identification route for nonparametric approaches. E.g. Theorem 4 of Athey and Haile [2] show that the symmetric affiliated value model is not identified.

According to our methodology, each ordered statistic leads to an equation leading thus to an n equations system, whereas we face n unknowns. Thus the least unobserved bidding statistic breaks the procedure. There are two routes to restore it. First, to impose more symmetry by assuming that some bidders are symmetric: it corresponds to a reduction of the number of unknowns. Second, to exploit some exogenous variations in the number of bidders: it corresponds to an expansion of the number of equations. Under some mild restriction on the asymmetric IPV model, the way we exploit independence could be usefully adapted in further research to obtain identification with an incomplete set of anonymous bids and which goes beyond the symmetric IPV model. However, such additional assumptions are not necessary for identification. Methods that are relying on the joint-distribution of two order-statistics (and that lies outside the scope of this work) allows identification and are providing an alternative route. Nevertheless, doing so is at some cost since it will require the estimation of joint-distributions and add at least one supplementary dimension with respect to the estimation of the order-statistics. On the contrary, our nonparametric procedure under anonymous data does not involve any additional dimension with respect to the standard ones under independent values, i.e. dimension $d+1$ where d is the dimension of the covariates usually reduced to a single dimensional index, as it is reflected by the same convergence rates.[24]

Our approach can also be used for alternative asymmetric auction models with independent private signals as the one developed by Landsberger et al. [19] where the ranking of bidders' private valuations is common knowledge among bidders (but possibly not to the econometrician). A promising avenue for research, which was the initial motivation of this work, is the structural analysis of models with shill bidding as developed by Lamy [17, 18]. With private values, models with shill bidding are strategically equivalent to models with a secret reserve price. It differs only from the econometrician point of view: in the latter, she distinguishes a submitted bid from the reserve price which facilitates the estimation as in Li and Perrigne [22], whereas, in the former, the strategic bidding activity of the seller is indistinguishable from

[24] The idea of our methodology could be also useful for environments with partial anonymity and incomplete sets of bids, e.g. if the identity of the winner is observed and all losing bids are observed anonymously in the second price auction. Komarova [14] shows that the asymmetric IPV model is identified through the identity of the winner and the amount of the highest losing bid and proposes an estimation procedure. If additional losing bids are observed, adaptations of our methodology could be useful to exploit them.

any other bid.

The ideas sustaining our methodology could be useful more generally beyond auction environments for applications as imperfect matching between data set, possible new anonymous designs in experimental economics or the design of surveys for sensitive attributes.

References

[1] S. Athey. Single crossing properties and the existence of pure strategy equilibria in games of incomplete information. *Econometrica*, 69(4):861–890, 2001.

[2] S. Athey and P. Haile. Identification of standard auction models. *Econometrica*, 70(6):2107–2140, 2002.

[3] L. H. Baldwin, R. C. Marshall, and J.-F. Richard. Bidder collusion at forest service timber sales. *J. Polit. Economy*, 105:657–699, 1997.

[4] S. Basu, R. Pollack, and M.-F. Roy. *Algorithms in Real Algebraic Geometry.* Algorithms and Computation in Mathematics. Springer, 2006.

[5] J. Blum, J. Kiefer, and M. Rosenblatt. Distribution free tests of indenpendence based on the sample distributions functions. *Annals of Mathematical Statistics*, 32:485–498, 1961.

[6] S. Campo, E. Guerre, I. Perrigne, and Q. Vuong. Semiparametric estimation of first-price auctions with risk averse bidders. *mimeo*, 2002.

[7] S. Campo, I. Perrigne, and Q. Vuong. Asymmetry in first-price auctions with affiliated private values. *J. Appl. Econ.*, 18:179–207, 2003.

[8] V. Flambard and I. Perrigne. Asymmetry in procurement auctions: Evidence from snow removal contracts. *The Economic Journal*, 116:1014–1036, 2006.

[9] W. Gautschi. On the condition of algebraic equations. *Numer. Math.*, 21:405–424, 1973.

[10] E. Guerre, I. Perrigne, and Q. Vuong. Optimal nonparametric estimation of first price auctions. *Econometrica*, 68:525–574, 2000.

[11] P. Haile, H. Hong, and M. Shum. Nonparametric tests for common values in first-price sealed-bid auctions. *mimeo*, 2004.

[12] K. Hendricks and R. H. Porter. An empirical study of an auction with asymmetric information. *Amer. Econ. Rev.*, 78:865–83, 1988.

[13] M. Hollander and D. Wolfe. *Nonparametric Statistical Methods.* Wiley-Interscience 2nd edition, 1999.

[14] T. Komarova. Nonparametric identification in asymmetric second-price auctions: A new approach. mimeo Northwestern University, January 2008.

[15] E. Krasnokutskaya. Identification and estimation in highway procurement auctions under unobserved auction heterogeneity. *mimeo*, 2004.

[16] J.-J. Laffont and Q. Vuong. Structural analysis of auction data. *Amer.Econ. Rev. Papers and Proceeding*, 86(2):414–420, 1996.

[17] L. Lamy. Competition between auction houses: a shill bidding perspective. *mimeo*, 2008.

[18] L. Lamy. The shill bidding effect versus the linkage principle. *Journal of Economic Theory*, page forthcoming, 2008.

[19] M. Landsberger, J. Rubinstein, E. Wolfstetter, and S. Zamir. First price auctions when the ranking of valuations is common knowledge. *Review of Economic Design*, 3(4):461–480, 2001.

[20] B. Lebrun. Uniqueness of the equilibrium in first-price auctions. *Games Econ. Behav.*, 55:131–151, 2006.

[21] A. J. Lee. *U-Statistics*. Statistics: Textbooks and Monographs. Marcel Dekker, Inc New York and Basel, 1990.

[22] T. Li and I. Perrigne. Timber sale auctions with random reserve price. *Rev. Econ. Statist.*, 85:189–200, 2003.

[23] T. Li, I. Perrigne, and Q. Vuong. Structural estimation of the affiliated private value auction model. *RAND J. Econ.*, 33:171–193, 2002.

[24] T. Li and Q. Vuong. Nonparametric estimation of the measurement error model using multiple indicators. *Journal of Mutivariate Analysis*, 65:139–165, 1998.

[25] E. Maskin and J. Riley. Asymmetric auctions. *Rev. Econ. Stud.*, 67(3):413–438, 2000.

[26] G. McLachlan and T. Krishnan. *The EM Algorithm and Extensions*. Wiley Series in Probability and Statistics, 1997.

[27] R. Mosier. Root neighborhoods of a polynomial. *Mathematics of Computation*, 47(175):265–273, 1986.

[28] R. B. Myerson. Optimal auction design. *Mathematics of Operation Research*, 6(1):58–73, 1981.

[29] W. Newey and D. McFadden. *Large Sample Estimation and Hypothesis Testing*. Handbook of Econometrics, Vol 4. ed. by R. Engle and D. McFadden. North Holland, 1994.

[30] H. Paarsch and H. Hong. *An Introduction to the Structural Econometrics of Auction Data*. The MIT Press, Cambridge, Massachusetts, 2006.

[31] M. Pesendorfer. A study of collusion in first-price auctions. *Rev. Econ. Stud.*, 67(3):381–411, 2000.

[32] R. H. Porter and D. J. Zona. Detection of bid rigging in procurement auctions. *J. Polit. Economy*, 101:518–538, 1993.

[33] R. H. Porter and D. J. Zona. Ohio school milk markets: an analysis of bidding. *RAND J. Econ.*, 30:263–288, 1999.

[34] K. Sailer. Searching the ebay marketplace. *CESifo Working Paper*, 2006.

[35] M. Silvapulle and P. K. Sen. *Constrained Statistical Inference.* Wiley Series in Probability and Statistics, 2005.

[36] U. Song. Nonparametric estimation of an ebay auction model with an unknown number of bidders. *mimeo UBC*, 2004.

[37] C. Stone. Optimal global rates of convergence for nonparametric estimators. *Annals of Statistics*, 10:1040–53, 1982.

[38] D. Titterington, A. Smith, and U. Markov. *Statistical Analysis of Finite Mixture Distributions.* Wiley Series in Probability and Mathematical Statistics, 1985.

[39] G. Van Den Berg. *Duration Models: Specification, Identification and Multiple Durations.* Handbook of Econometrics, Vol. 5. Amsterdam: NorthHolland, 2001.

[40] F. Wolak. The local nature of hypothesis tests involving inequality constraints in nonlinear models. *Econometrica*, 59(4):981–95, 1991.

[41] L. Yang. Recent advances on determining the number of real roots of parametric polynomials. *Journal of Symbolic Computation*, 28:225–242, 1999.

A Appendix

A.1 Proof of Proposition 3.1

Under observability of bidders' identities and in the first price auction, Li et al. [23] shows that the symmetric APV model is identified whereas Campo et al. [7] extends this result to the asymmetric APV model. Let us see why Li et al. [23]'s proof remains valid under anonymity whereas Campo et al. [7]'s proof does not. The main step to obtain identification is the equilibrium equation (2) that express bidder i's private value x_i as a function of his bid b_i and the CDF $F_{\mathbf{B}^*_{-i}|\mathbf{B}^*_i}(.|.)$ of the highest bid among his opponents conditional on his bid. Under observed identities, it is possible to obtain the full distribution of the vector of private valuations X since the CDFs $F_{\mathbf{B}^*_{-i}|\mathbf{B}^*_i}(.|.)$ are identified. Under anonymity, we observe only a weighted average of those CDFs: $\sum_{i=1}^{n} F_{\mathbf{B}^*_{-i}|\mathbf{B}^*_i}(b'|b) \cdot \text{Prob}(\mathbf{B}^*_i = b|\exists j\ \mathbf{B}^*_j = b, \mathbf{B}^*_k \leq b' \text{ for } k \neq j)$, which corresponds to the probability that the bid of the highest opponent of a bidder with an equilibrium bid b is smaller than b'. This prevents an immediate use of the equation (2) in the general case. However, in the symmetric case this average corresponds also to $F_{\mathbf{B}^*_{-i}|\mathbf{B}^*_i}(b'|b) = \frac{1}{n} \cdot \sum_{i=1}^{n} F_{\mathbf{B}^*_{-i}|\mathbf{B}^*_i}(b'|b)$ and the joint distribution of private signals is thus identified as for the second price auction where bids equal private values. Finally the symmetric APV model is identified in both formats.

For any strictly affiliated distribution of bids $F_{\mathbf{B}^*}$, let us construct a continuum of local perturbations $F^{\gamma}_{\mathbf{B}^*}$ that are strictly affiliated, lead to the same observable distribution $F_{\mathbf{B}}$ and that differ (up to a permutation) from $F_{\mathbf{B}^*}$. This will prove our non-identification result for the second price auction. If there were a one-to-one correspondence between signals and bids joint distributions in the first price auction then our non-identification result would extend immediately from the second price to the first price auction. Such a result is not available to the best of our knowledge and the technicalities of the extension of our proof to the first price auction are then relegated in the supplementary material.

Let $\phi(.)$ be a smoothed version of the indicator function on the interval $[0,1]$: $\phi(x) > 0$ if and only if $x \in [0,1]$, $\int \phi = 1$ and ϕ is continuously differentiable. Let x^1, $x^2 > x^1$ in $(\underline{x}, \overline{x})$, take $\epsilon < \min\{x^2 - x^1, x^1 - \underline{x}, \overline{x} - x^2\}$ and define:

$$c(x;\epsilon,i,j) \equiv \left(\phi(\frac{x_i - x^2}{\epsilon})\phi(\frac{x_j - x^1}{\epsilon}) - \phi(\frac{x_j - x^2}{\epsilon})\phi(\frac{x_i - x^1}{\epsilon})\right) \prod_{k \neq i,j} \phi(\frac{x_k - \underline{x}}{\epsilon}).$$

The function c shifts probability weight from some regions to others, in particular $\int\int c = 0$. Define $f^{\gamma}_{\mathbf{X}}(.) \equiv f_{\mathbf{X}}(.) + \gamma \cdot c(.;\epsilon,i,j)$. If γ is sufficiently small, then $f^{\gamma}_{\mathbf{X}}$ is a PDF and the affiliation property still holds ($\frac{\partial^2 log(f^{\gamma}_{\mathbf{X}}(x))}{\partial x_i \partial x_j} = \frac{\partial^2 log(f_{\mathbf{X}}(x))}{\partial x_i \partial x_j} + o(\gamma)$ uniformly on $(\underline{x}, \overline{x})$). Moreover, it leads to the same distribution of bids as the one resulting from $F_{\mathbf{X}}$ since the shift is between regions that are not distinguishable under anonymous bids. Finally, we have to check that $f^{\gamma}_{\mathbf{X}}(.)$ and $f_{\mathbf{X}}(.)$ do not coincide up to a permutation for a continuum of γ. By coincidence, for a given γ, there may exist a permutation π such that $f^{\gamma}_{\mathbf{X}}(x_1, \cdots, x_n) = f_{\mathbf{X}}(x_{\pi(1)}, \cdots, x_{\pi(n)})$ for any x. Our construction is valid for any γ which is sufficiently small, thus an infinite number of γ are potential candidates. On the other hand, there exists only a finite number of permutations and a contradiction is raised if $f^{\gamma}_{\mathbf{X}}(.)$ coincides with $f_{\mathbf{X}}(.)$ up to the same permutation for two different γ's since it would imply that the function

$c(.;\epsilon, i, j)$ is null.

A.2 Proof of Proposition 5.1

The proposition results from lemma A.1, a corollary of Yang [41]'s Theorem 2.1, and lemma A.2 which states how the number of distinct roots of a derivative polynomial $P^{(i)}$ is linked to the root structure of the original polynomial when the original polynomial has only real roots.

Lemma A.1 ***(Corollary of Theorem 2.1 in Yang [41])*** *A polynomial with real roots P has $r(P)$ distinct real roots if and only if $\Delta(P,k) > 0$ for $k \leq r(P)$ and $\Delta(P,k) = 0$ for $k > r(P)$.*

Lemma A.2 *The polynomial with real roots P has the root structure $(k_1, \cdots, k_{r(P)})$ if and only if the number of distinct roots of the polynomials $P^{(i)}$ is given by $\rho(i) = n - i - \sum_{j=1}^{r(P)} (k_j - 1 - i)^+$ for $i = 1, \cdots, n-2$.*

Proof The necessity part comes from Rolle's theorem: derivation creates at least one root between each adjacent roots, while the roots that had a multiplicity k_i strictly greater than one remains a root with multiplicity $k_i - 1$. There is no additional roots other than the ones identified above for the derivative polynomial and the new roots have multiplicity one since the number of identified roots (counted with their multiplicity) is $n-1$. Finally the number of distinct roots of the derivative polynomial is $r(P) - 1 + \sum_{i=1}^{r(P)} \mathbf{1}[k_i > 1]$. Thus we have proved the necessity part for $\rho(1)$. The result for $\rho(2), \cdots, \rho(n-2)$ follows by induction of the above argument.

Consider two polynomials P and Q with two distinct root structures, respectively denoted by $(k_1, \cdots, k_{r(P)})$ and $(k'_1, \cdots, k'_{r(Q)})$. We also use the convention that $k_i = 0$ ($k'_i = 0$) for $i > r(P)$ ($i > r(Q)$). Denote by $i^* = \max\{j | k_j \neq k'_j\}$. Without loss of generality take $k_j > k'_j$. The number of distinct roots of the polynomial $Q^{(k'_{i^*}-1)}$ equals to $n - k'_{i^*} + 1 - \sum_{j=1}^{r(Q)} (k'_j - k'_{i^*})^+$, which is equal to $n - k'_{i^*} + 1 - \sum_{j=1}^{i^*} (k'_j - k'_{i^*})$, which is finally strictly bigger than $n - k_{i^*} + 1 - \sum_{j=1}^{i^*} (k_j - k'_{i^*})$, the number of distinct roots of the polynomial $P^{(k'_{i^*}-1)}$ (after noting that $\sum_{j=1}^{i^*} k_j = \sum_{j=1}^{i^*} k'_j$). Thus we obtain that the number of distinct roots of the $(k'_{i^*} - 1)^{th}$ derivative polynomials differ. **CQFD**

Applying lemma A.1 to the polynomials $\{P^{(i)}\}_{i=1,\cdots,n-2}$, whose numbers of distinct roots are given by lemma A.2, gives Proposition 5.1.

A.3 Proof of Proposition 5.3

To alleviate notations we assume here that $K_{F_{\mathbf{B_p|Z}}} = K_{f_\mathbf{Z}} = K$ and $h_{F_{\mathbf{B_p|Z}}} = h_{f_\mathbf{Z}} = h$. The proof can be immediately adapted to the general case. In the following let $x \vee y$ denote the maximum of x and y. Let $K_l^h(z)$ and K_{l_i,l_j}^h denote respectively $h^{-d}K(\frac{Z_l - z}{h})$ and $h^{-d}K(\frac{Z_{l_j}-Z_{l_i}}{h})$. By carefully plugging the expressions

$$(\widehat{F}_{\mathbf{B},\mathbf{Z}}^{(1:1)}(b,z))^2 = \frac{L^{-2}}{n^2}\sum_{l_2,l_3=1}^{L}\sum_{p_2,p_3=1}^{n}\mathbf{1}(B_{p_2l_2}\le b)\mathbf{1}(B_{p_3l_3}\le b)K_{l_2}^h(z)K_{l_3}^h(z)$$

$$\widehat{F}_{\mathbf{B},\mathbf{Z}}^{(2:2)}(b,z)\widehat{f}_\mathbf{Z}(z) = \frac{L^{-2}}{n(n-1)}\sum_{l_2,l_3=1}^{L}\sum_{\substack{p_2,p_3=1\\p_2\neq p_3}}^{n}\frac{\mathbf{1}(B_{p_2l_2}\vee B_{p_3l_2}\le b)+\mathbf{1}(B_{p_2l_3}\vee B_{p_3l_3}\le b)}{2}K_{l_2}^h(z)K_{l_3}^h(z)$$

into the expression of $\widehat{\mathcal{H}}$ in equation (20), our estimator of the discriminant can be expressed as a U-statistic:

$$\widehat{\mathcal{H}} = L^{-3}\sum_{l_1=1}^{L}\sum_{l_2=1}^{L}\sum_{l_3=1}^{L}m_L(Y_{l_1},Y_{l_2},Y_{l_3})$$

with $Y_l = (B_l, Z_l)$ and where the kernel of the U-statistic $m_L(.,.,.)$ is given by

$$m_L(Y_{l_1},Y_{l_2},Y_{l_3}) = \frac{1}{n^3}\sum_{p_1,p_2,p_3=1}^{n}\mathbf{1}(B_{p_2l_2}\le B_{p_1l_1})\mathbf{1}(B_{p_3l_3}\le B_{p_1l_1})K_{l_1,l_2}^hK_{l_1,l_3}^h$$
$$-\frac{1}{2n^2(n-1)}\sum_{\substack{p_1,p_2,p_3=1\\p_2\neq p_3}}^{n}\left[\mathbf{1}(B_{p_2l_2}\vee B_{p_3l_2}\le B_{p_1l_1})+\mathbf{1}(B_{p_2l_3}\vee B_{p_3l_3}\le B_{p_1l_1})\right]K_{l_1,l_2}^hK_{l_1,l_3}^h.$$

Note that the kernel of the U-statistic depends explicitly on L through the bandwidth h. The limit variance calculation deserves thus additional care.[25] Newey and McFadden [29] show how to deal with such a nuisance for U-statistic with kernels of degree 2. Here the problem is similar up to the point that we face a U-statistic of degree 3 and thus we have to generalize lemma 8.4 in Newey and McFadden [29] for U-statistics with kernels of degree 3. Let $m_L^1(Y_l) = E_{Y_{l_2},Y_{l_3}}[m_L(Y_l,Y_{l_2},Y_{l_3})]$, $m_L^2(Y_l) = E_{Y_{l_1},Y_{l_3}}[m_L(Y_{l_1},Y_l,Y_{l_3})]$, $m_L^3(Y_l) = E_{Y_{l_1},Y_{l_2}}[m_L(Y_{l_1},Y_{l_2},Y_l)]$ and $\mu = E_{Y_{l_1},Y_{l_2},Y_{l_3}}[m_L(Y_{l_1},Y_{l_2},Y_{l_3})]$.

Lemma A.3 ***Generalization of lemma 8.4 in Newey and McFadden [29]***

If $Y_1, Y_2, \cdots$ *are i.i.d. then* $2\mu + L^{-3}\sum_{l_1=1}^{L}\sum_{l_2=1}^{L}\sum_{l_3=1}^{L}m_L(Y_{l_1},Y_{l_2},Y_{l_3}) - \cdots$
$\cdots L^{-1}\sum_{l=1}^{L}m_L^1(Y_l)+m_L^2(Y_l)+m_L^3(Y_l) = \frac{1}{L}.O_L\{(E_{Y_{l_1},Y_{l_2},Y_{l_3}}[||m_L(Y_{l_1},Y_{l_2},Y_{l_3})||^2])^{1/2}$
$\cdots + E_{Y_{l_1},Y_{l_2}}[||m_L(Y_{l_1},Y_{l_1},Y_{l_2})||] + E_{Y_{l_1},Y_{l_2}}[||m_L(Y_{l_1},Y_{l_2},Y_{l_2})||] + \cdots$
$\cdots + E_{Y_{l_1},Y_{l_2}}[||m_L(Y_{l_1},Y_{l_2},Y_{l_1})||] + E_{Y_{l_1}}[||m_L(Y_{l_1},Y_{l_1},Y_{l_1})||]\}$.

[25] See Lee [21] for a detailed exposition of the variance calculation for standard U-statistics.

Proof As in [29], by replacing $m_L(Y_{l_1}, Y_{l_2}, Y_{l_3})$ by $m_L(Y_{l_1}, Y_{l_2}, Y_{l_3}) - \mu$ it can be assumed w.l.o.g. that $\mu = 0$. The term $L^{-3}\sum_{l_1,l_2,l_3=1}^{L} m_L(Y_{l_1}, Y_{l_2}, Y_{l_3}) - L^{-1}\sum_{l=1}^{L}\left(m_L^1(Y_l) + m_L^2(Y_l) + m_L^3(Y_l)\right)$ is equal to $L^{-3}\sum_{l_1,l_2,l_3=1}^{L}\left[m_L(Y_{l_1}, Y_{l_2}, Y_{l_3}) - m_L^1(Y_{l_1}) - m_L^2(Y_{l_2}) - m_L^3(Y_{l_3})\right]$ which can be written as the sum $T_1 + T_2 + T_3$ where the first term T_1 corresponds to the case where the indices l_1, l_2, l_3 in the triple sum are all distinct, the second term T_2 corresponds to the cases where two indices coincide and the third term T_3 corresponds to the cases where $l_1 = l_2 = l_3$. The third term is of the order $O_L\{(E_{Y_{l_1},Y_{l_2},Y_{l_3}}[||m_L(Y_{l_1}, Y_{l_2}, Y_{l_3})||] + E_{Y_{l_1}}[||m_L(Y_{l_1}, Y_{l_1}, Y_{l_1})||])/L^2\}$. The second term is of the order $O_L\{(E_{Y_{l_1},Y_{l_2}}[||m_L(Y_{l_1}, Y_{l_1}, Y_{l_2})||] + E_{Y_{l_1},Y_{l_2}}[||m_L(Y_{l_1}, Y_{l_2}, Y_{l_2})||] + E_{Y_{l_1},Y_{l_2}}[||m_L(Y_{l_1}, Y_{l_2}, Y_{l_1})||] + E_{Y_{l_1}}[||m_L(Y_{l_1}, Y_{l_1}, Y_{l_1})||])/L\}$. As in [29], we work with $E[T_1^2]$.

$$T_1^2 = L^{-6}\sum_{\substack{u_1,u_2,u_3=1\\ u_i\neq u_j}}^{L}\sum_{\substack{v_1,v_2,v_3=1\\ v_i\neq v_j}}^{L}\left(m_L(Y_{u_1}, Y_{u_2}, Y_{u_3}) - m_L^1(Y_{u_1}) - m_L^2(Y_{u_2}) - m_L^3(Y_{u_3})\right)\times$$
$$\times\left(m_L(Y_{v_1}, Y_{v_2}, Y_{v_3}) - m_L^1(Y_{v_1}) - m_L^2(Y_{v_2}) - m_L^3(Y_{v_3})\right).$$

The sum can be divided according to the number of common elements between the sets $\mathfrak{U} = \{u_1, u_2, u_3\}$ and $\mathfrak{V} = \{v_1, v_2, v_3\}$. Let $S(\mathfrak{U}, \mathfrak{V}) = \#(\mathfrak{U} \cap \mathfrak{V})$. If $S(\mathfrak{U}, \mathfrak{V}) \leq 1$, the expectation of the term inside the sum is null. Note that the number of 6-uple $(u_1, u_2, u_3, v_1, v_2, v_3) \in [1, L]^6$ such that $S(\mathfrak{U}, \mathfrak{V}) = 2$ (respectively $= 3$) is of the order $O_L(L^4)$ (resp. $O_L(L^3)$). Note also that both $E_{Y_{u_1},Y_{u_2},Y_{u_3}}[||m_L(Y_{u_1}, Y_{u_2}, Y_{u_3}) - m_L^1(Y_{u_1}) - m_L^2(Y_{u_2}) - m_L^3(Y_{u_3})||^2]$ and $E_{Y_{u_1},Y_{u_2},Y_{u_3},Y_{v_1}}[||(m_L(Y_{u_1}, Y_{u_2}, Y_{u_3}) - m_L^1(Y_{u_1}) - m_L^2(Y_{u_2}) - m_L^3(Y_{u_3}))\times(m_L(Y_{v_1}, Y_{u_2}, Y_{u_3}) - m_L^1(Y_{v_1}) - m_L^2(Y_{u_2}) - m_L^3(Y_{u_3}))||]$ are of the order of $O_L(E_{Y_{u_1},Y_{u_2},Y_{u_3}}[||m_L(Y_{u_1}, Y_{u_2}, Y_{u_3}))||^2])$. The term in $E[T_1^2]$ coming from $S(\mathfrak{U}, \mathfrak{V}) = 3$ is then of the order $E_{Y_{l_1},Y_{l_2},Y_{l_3}}[||m_L(Y_{l_1}, Y_{l_2}, Y_{l_3})||^2]/L^3$ while the term coming from $S(\mathfrak{U}, \mathfrak{V}) = 2$ is then of the order $E_{Y_{l_1},Y_{l_2},Y_{l_3}}[||m_L(Y_{l_1}, Y_{l_2}, Y_{l_3})||^2]/L^2$. We thus obtain that $E[T_1^2] = O_L(E_{Y_{l_1},Y_{l_2},Y_{l_3}}[||m_L(Y_{l_1}, Y_{l_2}, Y_{l_3})||^2]/L^2)$ and then $T_1 = O_L(E_{Y_{l_1},Y_{l_2},Y_{l_3}}[||m_L(Y_{l_1}, Y_{l_2}, Y_{l_3})||^2]/L^2)$. The lemma is obtained by gathering all those terms. **CQFD**

Below, we repeatedly use that $E_{\mathbf{Y_l}}[g(B_l, Z_l)K_l^h(z)] = E_{\mathbf{Y_l}}[g(B_l, Z_l)|Z_l = z] \cdot f_{\mathbf{Z}}(z) + o_L(1)$ for any bounded function g as it is implied from Bochner's lemma.

$E_{Y_{l_1},Y_{l_2}}[||m_L(Y_{l_1}, Y_{l_1}, Y_{l_2})||]$, $E_{Y_{l_1},Y_{l_2}}[||m_L(Y_{l_1}, Y_{l_2}, Y_{l_2})||]$,

$E_{Y_{l_1},Y_{l_2}}[||m_L(Y_{l_1}, Y_{l_2}, Y_{l_1})||]$ and $E_{Y_{l_1}}[||m_L(Y_{l_1}, Y_{l_1}, Y_{l_1})||]$ are all of the same order as $O_L(1)$.[26] We now move to the calculation of $m_L^j(Y_{l_1})$ for $j = 1, 2, 3$ and of μ.

[26]Note that the dependence in the bandwidths is removed by averaging over the full support of the terms that appear in the kernels in the expression of $m_L(.,.,.)$. A term as $E_{Y_{l_1}}[||m_L(Y_{l_1}, Y_{l_1}, Y_{l_1})||]$ may give the wrong impression that we are averaging over a diagonal and not the full support. However, remind that kernels appear only through the variable Z_l.

$m_L^1(Y_l) = E_{Y_{l_2},Y_{l_3}}[m_L(Y_l, Y_{l_2}, Y_{l_3})] = E_{Y_{l_2}}[E_{Y_{l_3}}[m_L(Y_l, Y_{l_2}, Y_{l_3})]]$
$= -\frac{1}{2n^2(n-1)} E_{Y_{l_2}}[\sum_{\substack{i_1,i_2,i_3=1 \\ i_2\neq i_3}}^n \left(\mathbf{1}(B^*_{i_2l_2} \vee B^*_{i_3l_2} \leq B^*_{i_1l}) + F_{\mathbf{B}^*_{\mathbf{i_2}}|\mathbf{Z}}(B^*_{i_1l}|Z_l) \cdot F_{\mathbf{B}^*_{\mathbf{i_3}}|\mathbf{Z}}(B^*_{i_1l}|Z_l)\right) \cdot K^h_{l,l_2} \cdot f_{\mathbf{Z}}(Z_l)]$
$+\frac{1}{n^3} E_{Y_{l_2}}[\sum_{i_1,i_2,i_3=1}^n \mathbf{1}(B^*_{i_2l_2} \leq B^*_{i_1l}) \cdot F_{\mathbf{B}^*_{\mathbf{i_3}}|\mathbf{Z}}(B^*_{i_1l}|Z_l) \cdot K^h_{l,l_2} \cdot f_{\mathbf{Z}}(Z_l)] + o_L(1)$
$= -\frac{1}{2n^2(n-1)} \left(\sum_{\substack{i_1,i_2,i_3=1 \\ i_2\neq i_3}}^n 2 \cdot F_{\mathbf{B}^*_{\mathbf{i_2}}|\mathbf{Z}}(B^*_{i_1l}|Z_l) \cdot F_{\mathbf{B}^*_{\mathbf{i_3}}|\mathbf{Z}}(B^*_{i_1l}|Z_l)\right) \cdot (f_{\mathbf{Z}}(Z_l))^2$
$+\frac{1}{n^3} \left(\sum_{i_1,i_2,i_3=1}^n F_{\mathbf{B}^*_{\mathbf{i_2}}|\mathbf{Z}}(B^*_{i_1l}|Z_l) \cdot F_{\mathbf{B}^*_{\mathbf{i_3}}|\mathbf{Z}}(B^*_{i_1l}|Z_l)\right) \cdot (f_{\mathbf{Z}}(Z_l))^2 + o_L(1)$
$= \frac{1}{n}\sum_{i_1=1}^n (F^{(1:1)}_{\mathbf{B},\mathbf{Z}}(B^*_{i_1l}, Z_l)^2 - F^{(2:2)}_{\mathbf{B},\mathbf{Z}}(B^*_{i_1l}, Z_l) \cdot f_{\mathbf{Z}}(Z_l)) + o_L(1) = \frac{1}{n}\sum_{i_1=1}^n \Delta(P_{B^*_{i_1l},Z_l}, 2) + o_L(1).$

$m_L^2(Y_l) = m_L^3(Y_l) = E_{Y_{l_1},Y_{l_3}}[m_L(Y_{l_1}, Y_l, Y_{l_3})] = E_{Y_{l_1}}[E_{Y_{l_3}}[m_L(Y_{l_1}, Y_l, Y_{l_3})]]$
$= -\frac{1}{2n^2(n-1)} E_{Y_{l_1}}[\sum_{\substack{i_1,i_2,i_3=1 \\ i_2\neq i_3}}^n \left(\mathbf{1}(B^*_{i_2l} \vee B^*_{i_3l} \leq B^*_{i_1l_1}) + F_{\mathbf{B}^*_{\mathbf{i_2}}|\mathbf{Z}}(B^*_{i_1l_1}|Z_{l_1}) \cdot F_{\mathbf{B}^*_{\mathbf{i_3}}|\mathbf{Z}}(B^*_{i_1l_1}|Z_{l_1})\right) \cdot K^h_{l_1,l} \cdot f_{\mathbf{Z}}(Z_{l_1})]$
$+\frac{1}{n^3} E_{Y_{l_1}}[\sum_{i_1,i_2,i_3=1}^n \mathbf{1}(B^*_{i_2l} \leq B^*_{i_1l_1}) \cdot F_{\mathbf{B}^*_{\mathbf{i_3}}|\mathbf{Z}}(B^*_{i_1l_1}|Z_{l_1}) \cdot K^h_{l_1,l} \cdot f_{\mathbf{Z}}(Z_{l_1})] + o_L(1)$
$= -\frac{1}{2n^2(n-1)} \left(\sum_{\substack{i_1,i_2,i_3=1 \\ i_2\neq i_3}}^n E_{Y_{l_1}}[\mathbf{1}(B^*_{i_2l} \vee B^*_{i_3l} \leq B^*_{i_1l_1}) + F_{\mathbf{B}^*_{\mathbf{i_2}}|\mathbf{Z}}(B^*_{i_1l_1}|Z_l) \cdot F_{\mathbf{B}^*_{\mathbf{i_3}}|\mathbf{Z}}(B^*_{i_1l_1}|Z_l)|Z_{l_1} = Z_l]\right) \cdot [f_{\mathbf{Z}}(Z_l)]^2$
$+\frac{1}{n^3} \left(\sum_{i_1,i_2,i_3=1}^n E_{Y_{l_1}}[\mathbf{1}(B^*_{i_2l} \leq B^*_{i_1l_1}) \cdot F_{\mathbf{B}^*_{\mathbf{i_3}}|\mathbf{Z}}(B^*_{i_1l_1}|Z_l)|Z_{l_1} = Z_l]\right) \cdot [f_{\mathbf{Z}}(Z_l)]^2 + o_L(1)$
$= -\left(\frac{1}{2n(n-1)} \sum_{\substack{i_2,i_3=1 \\ i_2\neq i_3}}^n \int_{B^*_{i_2l}\vee B^*_{i_3l}}^{\infty} f^{(1:1)}_{\mathbf{B}|\mathbf{Z}}(b|Z_l)db + \frac{1}{2n}\sum_{i_1=1}^n E_{Y_{l_1}}[F^{(2:2)}_{\mathbf{B}|\mathbf{Z}}(B^*_{i_1l_1}|Z_l)|Z_{l_1} = Z_l]\right) \cdot [f_{\mathbf{Z}}(Z_l)]^2$
$+\frac{1}{n}\sum_{i_2=1}^n \left(\int_{B^*_{i_2l}}^{\infty} F^{(1:1)}_{\mathbf{B}|\mathbf{Z}}(b|Z_l) f^{(1:1)}_{\mathbf{B}|\mathbf{Z}}(b|Z_l)db\right) \cdot [f_{\mathbf{Z}}(Z_l)]^2 + o_L(1)$
$= -\left(\frac{1}{2n(n-1)} \sum_{\substack{i_2,i_3=1 \\ i_2\neq i_3}}^n (1 - F^{(1:1)}_{\mathbf{B}|\mathbf{Z}}(B^*_{i_2l} \vee B^*_{i_3l}|Z_l)) + \frac{1}{2}\int_{-\infty}^{\infty} F^{(2:2)}_{\mathbf{B}|\mathbf{Z}}(b|Z_l) f^{(1:1)}_{\mathbf{B}|\mathbf{Z}}(b|Z_l)db\right) \cdot [f_{\mathbf{Z}}(Z_l)]^2$
$+\frac{1}{2n}\sum_{i_2=1}^n (1 - (F^{(1:1)}_{\mathbf{B}|\mathbf{Z}}(B^*_{i_2l_2}|Z_l))^2) \cdot [f_{\mathbf{Z}}(Z_l)]^2 + o_L(1)$
$= \left(\frac{1}{2n(n-1)} \sum_{\substack{i_2,i_3=1 \\ i_2\neq i_3}}^n F^{(1:1)}_{\mathbf{B}|\mathbf{Z}}(B^*_{i_2l} \vee B^*_{i_3l}|Z_l) - \frac{1}{2n}\sum_{i_2=1}^n (F^{(1:1)}_{\mathbf{B}|\mathbf{Z}}(B^*_{i_2l}|Z_l))^2\right) \cdot [f_{\mathbf{Z}}(Z_l)]^2$
$+\frac{1}{2}\int_{-\infty}^{\infty} F^{(2:2)}_{\mathbf{B}|\mathbf{Z}}(b|Z_l) f^{(1:1)}_{\mathbf{B}|\mathbf{Z}}(b|Z_l)db \cdot [f_{\mathbf{Z}}(Z_l)]^2 + o_L(1)$

$\mu = E_{Y_l}[m_L^1(Y_l)] = E_{Y_l}[m_L^2(Y_l)] = E_{Y_l}[m_L^3(Y_l)] = \mathcal{H} + o_L(1).$

$$\sum_{j=1}^3 m_L^j(Y_l) - 3\mathcal{H} = \underbrace{\left(\frac{1}{n(n-1)} \sum_{\substack{i_2,i_3=1 \\ i_2\neq i_3}}^n F^{(1:1)}_{\mathbf{B}|\mathbf{Z}}(B^*_{i_2l} \vee B^*_{i_3l}|Z_l) - \frac{1}{n}\sum_{i_2=1}^n F^{(2:2)}_{\mathbf{B}|\mathbf{Z}}(B^*_{i_2l}|Z_l) + C(Z_l)\right)}_{K(Y_l), \text{ which is independent of L}} \cdot [f_{\mathbf{Z}}(Z_l)]^2 + o_L(1)$$

where $C(Z_l) = \int_{-\infty}^{\infty} F^{(2:2)}_{\mathbf{B}|\mathbf{Z}}(b|Z_l) f^{(1:1)}_{\mathbf{B}|\mathbf{Z}}(b|Z_l)db - \frac{3\mathcal{H}}{[f_{\mathbf{Z}}(Z_l)]^2}$. After gathering all those terms, we have from lemma A.3: $\sqrt{L}.(\widehat{\mathcal{H}} - \mathcal{H}) = \sqrt{L}.L^{-1}\sum_{l=1}^L K(Y_l)[f_{\mathbf{Z}}(Z_l)]^2 + o_L(1)$. We have $E_{B_l}[K(Y_l)|Z_l = z] = 0$ for all z such that $f_{\mathbf{Z}}(z) > 0$ which gives that $\mathbf{Var}_{Y_l}[K(Y_l)[f_{\mathbf{Z}}(Z_l)]^2] = E_{Z_l}[\mathbf{Var}_{B_l}[K(Y_l)|Z_l] \cdot [f_{\mathbf{Z}}(Z_l)]^4]$. The variable $K(Y_l)$ is not a constant with respect to B_l which guarantees the asymptotic normality of the variable $\sqrt{L}.(\widehat{\mathcal{H}} - \mathcal{H})$ to a centered normal distribution with variance $\Sigma^2 = E_{Z_l}[\mathbf{Var}_{B_l}[K(Y_l)|Z_l] \cdot [f_{\mathbf{Z}}(Z_l)]^4] + o_L(1)$. Finally we are left with the analytical expression for the variance under H_0 which is done by straightforward though tedious calculation. Under H_0 note that we have $F^{(2:2)}_{\mathbf{B}|\mathbf{Z}}(b|z) = [F^{(1:1)}_{\mathbf{B}|\mathbf{Z}}(b|z)]^2$.

$$\begin{aligned}\mathbf{Var}_{B_l}[K(Y_l)|Z_l] &= \underbrace{\mathbf{Var}_{B_l}[\frac{1}{n}\sum_{i_1=1}^{n}(F^{(1:1)}_{\mathbf{B}|\mathbf{Z}}(B^*_{i_1l}|Z_l))^2]}_{=\frac{4}{45n}} + \underbrace{\mathbf{Var}_{B_l}[\frac{1}{n(n-1)}\sum_{\substack{i_1,i_2=1\\ i_1\neq i_2}}^{n} F^{(1:1)}_{\mathbf{B}|\mathbf{Z}}(B^*_{i_1l}\vee B^*_{i_2l}|Z_l)]}_{=\frac{4(n-2)}{45n(n-1)}+\frac{1}{9n(n-1)}} \\ &\underbrace{+2.\mathbf{Cov}_{B_l}[\frac{1}{n}\sum_{i_1=1}^{n}(F^{(1:1)}_{\mathbf{B}|\mathbf{Z}}(B^*_{i_1l}|Z_l))^2, \frac{1}{n(n-1)}\sum_{\substack{i_1,i_2=1\\ i_1\neq i_2}}^{n} F^{(1:1)}_{\mathbf{B}|\mathbf{Z}}(B^*_{i_1l}\vee B^*_{i_2l}|Z_l)]}_{=-\frac{8}{45n}}\,.\end{aligned}$$

Throughout this calculation we use that $\int_{-\infty}^{\infty}(F^{(1:1)}_{\mathbf{B}|\mathbf{Z}}(b|Z_l))^{\alpha}\cdot d[(F^{(1:1)}_{\mathbf{B}|\mathbf{Z}}(b|Z_l))^{\beta}] = \frac{\beta}{\alpha+\beta}$. The calculation of the second term uses also that $E_{B_l}[F^{(1:1)}_{\mathbf{B}|\mathbf{Z}}(B^*_{i_1l}\vee B^*_{i_2l}|Z_l)\cdot F^{(1:1)}_{\mathbf{B}|\mathbf{Z}}(B^*_{i_1l}\vee B^*_{i_3l}|Z_l)] = \frac{7}{15}$. The asymptotic expression of the variance under H_0 is obtained by gathering all those terms: $\Sigma^2 = \frac{E_{Z_l}[[f_{\mathbf{Z}}(Z_l)]^4]}{45n(n-1)}$. In the special case without any covariates it reduces to $\frac{1}{45n(n-1)}$.

A.4 Proof of Proposition 6.1

In their proposition 1, GPV obtains the same properties for the CDFs $F_{\mathbf{B}^*_{\mathbf{i}}|\mathbf{Z},\mathbf{I}}$ instead of $F_{\mathbf{B_p}|\mathbf{Z},\mathbf{I}}$. From (3) and (4), we obtain that any CDF $F_{\mathbf{B_p}|\mathbf{Z},\mathbf{I}}(b,z,I)$ can be expressed as a linear combination of terms which are product of $F_{\mathbf{B}^*_{\mathbf{i}}|\mathbf{Z},\mathbf{I}}(b,z,I)$, i.e. as a continuous function of the CDFs $F_{\mathbf{B}^*_{\mathbf{i}}|\mathbf{Z},\mathbf{I}}$. The CDF $F_{\mathbf{B}^*_{\mathbf{i}}|\mathbf{Z},\mathbf{I}}$ have the desired smoothness properties on the set $S^0(F_{\mathbf{B_{n_I}}|\mathbf{Z},\mathbf{I}})\setminus\{\bar{b}(z,I,i)\}$: on the set $S^0(F_{\mathbf{B}^*_{\mathbf{i}}|\mathbf{Z},\mathbf{I}})$, it comes from GPV, whereas $F_{\mathbf{B}^*_{\mathbf{i}}|\mathbf{Z},\mathbf{I}}$ is equal to 1 above $\bar{b}(z,I,i)$ and is thus C^{∞}. Thus all the regularity properties (iii-v) that are valid for $F_{\mathbf{B}^*_{\mathbf{i}}|\mathbf{Z},\mathbf{I}}$ are still valid for $F_{\mathbf{B_p}|\mathbf{Z},\mathbf{I}}$ if the points $\{\bar{b}(z,I,i)\}_{i\in I}$ have been appropriately removed. The image of a closed interval by a continuous function is a closed interval. Thus (i) holds also for $F_{\mathbf{B_p}|\mathbf{Z},\mathbf{I}}$. Finally we are left with (ii). Note the difference between the similar point in GPV which holds for the whole support and not only for a closed subset of the $S^o(F_{\mathbf{B_p}|\mathbf{Z},\mathbf{I}})$ as above. By deriving (4) and (3), we obtain an another expression of $f_{\mathbf{B_p}|\mathbf{Z},\mathbf{I}}(b|z,I)$ as a function of $F_{\mathbf{B}^*_{\mathbf{i}}|\mathbf{Z},\mathbf{I}}(b|z,I)$ and $f_{\mathbf{B}^*_{\mathbf{i}}|\mathbf{Z},\mathbf{I}}(b|z,I)$:

$$\begin{aligned}f_{\mathbf{B_p}|\mathbf{Z},\mathbf{I}}(b,z,I) =& \frac{1}{(p-1)!(n_I-p-1)!} \\ &\cdot\sum_{\pi\in\Sigma_I}\left[\prod_{k=1}^{p-1}F_{\mathbf{B}^*_{\pi(\mathbf{k})}|\mathbf{Z},\mathbf{I}}(b,z,I)\cdot f_{\mathbf{B}^*_{\pi(\mathbf{p})}|\mathbf{Z},\mathbf{I}}(b,z,I)\cdot\prod_{k=p+1}^{n_I}(1-F_{\mathbf{B}^*_{\pi(\mathbf{k})}|\mathbf{Z},\mathbf{I}}(b,z,I))\right]\end{aligned}$$

Thus we obtain that $f_{\mathbf{B_p}|\mathbf{Z},\mathbf{I}}(b,z,I)$ is strictly positive on $S^o(F_{\mathbf{B_n}|\mathbf{Z},\mathbf{I}})$.[27]

A.5 Proof of Proposition 6.2

We write the proof for the first price auction, the arguments are easily adapted for the second price auction. It is closely related to GPV and uses intensively some rates of uniform convergence derived by GPV. We follow their proof very carefully and focus only on the two new ingredients. First, their proof is based on the uniform rates of convergence for the CDF, the PDF and also the boundaries estimators of the variable B^* that is observed by the econometrician. Here we do not observe B^* but only the vector of order statistics B. Second, the pseudo probabilities are a new ingredient that do not appear in GPV.

The first issue is then to prove that the same uniform rates of convergence are still valid for B^* though it is not observed. Nevertheless, the uniform rates of convergence they obtained for B^* are still valid under anonymity for the variable B that is observed and with our similar choices for the kernels and the bandwidth parameters. Contrary to GPV's analysis which is restricted to a symmetric environment, the observed variable B is here multidimensional: it does not modify their analysis which immediately adapts since our procedure is based only on the estimation of the one dimensional densities $F_{\mathbf{B_p},\mathbf{Z},\mathrm{I}}(b,z,\mathrm{I})$.

First the bidding supports of the bidders are coinciding with the support of the order statistics. Thus all the results for the estimator of the support of B are immediately converted into results for B^*. From GPV (lemma B2), we obtain the following uniform rate of convergence for the kernel estimators $\widehat{F}_{\mathbf{B},\mathbf{Z},\mathrm{I}}(b,z,\mathrm{I})$ and $\widehat{f}_{\mathbf{B},\mathbf{Z},\mathrm{I}}(b,z,\mathrm{I})$ on any inner closed compact subset of the bidding support, denoted by $\mathcal{C}(B)$.

$$\sup_{(b,z,I)\subset\mathcal{C}(B)} ||\widehat{F}_{\mathbf{B_p},\mathbf{Z},\mathrm{I}}(b,z,\mathrm{I}) - F_{\mathbf{B},\mathbf{Z},\mathrm{I}}(b,z,\mathrm{I})||_0 = O(\frac{\log L}{L})^{\frac{R+1}{2R+d+2}}$$

$$\sup_{(b,z,I)\subset\mathcal{C}(B)} ||\widehat{f}_{\mathbf{B_p},\mathbf{Z},\mathrm{I}}(b,z,\mathrm{I}) - f_{\mathbf{B},\mathbf{Z},\mathrm{I}}(b,z,\mathrm{I})||_0 = O(\frac{\log L}{L})^{\frac{R+1}{2R+d+3}}$$

In GPV, the corresponding uniform rates of convergence are obtained for the bidding distributions and densities $\widehat{F}_{\mathbf{B_i^*},\mathbf{Z},\mathrm{I}}(b,z,\mathrm{I})$ and $\widehat{f}_{\mathbf{B_i^*},\mathbf{Z},\mathrm{I}}(b,z,\mathrm{I})$ since bidders' identities are observed. However, we establish that the function mapping the vector

[27] Note that $f_{\mathbf{B_p}|\mathbf{Z},\mathbf{I}}(b,z,I)$ is null at the lower bound $b = \underline{b}(z,I,p)$ for $p > 1$ (respectively at the upper bound $b = \overline{b}(z,I,p)$ for $p < n$).

of the order statistics CDF $(F_{\mathbf{B_p},\mathbf{Z},\mathrm{I}}(b,z,\mathrm{I}))_{p=1,\cdots,n_I}$ into $(F_{\mathbf{B_i^*},\mathbf{Z},\mathrm{I}}(b,z,\mathrm{I}))_{i\in I}$ is continuously differentiable on $\mathcal{C}(B)$ with a Jacobian matrix of full rank. This function is the composition of two functions. First, the function mapping the vector of the order statistics CDF $(F_{\mathbf{B_p},\mathbf{Z},\mathrm{I}}(b,z,\mathrm{I}))_{p=1,\cdots,n_I}$ into $(F^{(r:r)}_{\mathbf{B_p},\mathbf{Z},\mathrm{I}}(b,z,\mathrm{I}))_{p=1,\cdots,n_I}$ is a linear invertible function (the related matrix is triangular with the coefficient 1 on the diagonal). Second, the Jacobian matrix of the function mapping the vector of the order statistics CDF $(F^{(r:r)}_{\mathbf{B_p},\mathbf{Z},\mathrm{I}}(b,z,\mathrm{I}))_{p=1,\cdots,n_I}$ into $(F_{\mathbf{B_i^*},\mathbf{Z},\mathrm{I}}(b,z,\mathrm{I}))_{i=1\in I}$ is well-defined and of full rank on $\mathcal{C}(B)$ as shown below in lemma A.4. We thus conclude that the uniform rate of convergence that holds for $(F_{\mathbf{B_p},\mathbf{Z},\mathrm{I}}(b,z,\mathrm{I}))_{p=1,\cdots,n_I}$ remains valid for $(F_{\mathbf{B_i^*},\mathbf{Z},\mathrm{I}}(b,z,\mathrm{I}))_{i\in I}$.

Lemma A.4 *The Jacobian matrix of* Υ *at any point* $(a_{n-1},\cdots,a_0)$ *such that* $(\omega_1,\cdots,\omega_n)=\Upsilon(a_{n-1},\cdots,a_0)$ *and with* $\omega_1\geq\cdots\geq\omega_n$ *is well-defined if and only if* $\omega_1>\cdots>\omega_n$.

Proof The transpose of the jacobian matrix of the map Υ^{-1} at $\omega=(\omega_1,\cdots,\omega_n)$ is given by:

$$J_\omega=\begin{pmatrix} 1 & \sum_{j_1\neq 1}\omega_{j_1} & \cdot & \cdot & \sum_{j_1,\cdot,j_r,j_k\neq 1,j_k\neq j_{k'}}\prod_{j_k\in\{j_1,\cdot,j_r\}}\omega_{j_k} & \cdot & \cdot & \prod_{j\neq 1}\omega_j \\ 1 & \sum_{j_1\neq 2}\omega_{j_1} & \cdot & \cdot & \sum_{j_1,\cdot,j_r,j_k\neq 2,j_k\neq j_{k'}}\prod_{j_k\in\{j_1,\cdot,j_r\}}\omega_{j_k} & \cdot & \cdot & \prod_{j\neq 2}\omega_j \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & \sum_{j_1\neq l}\omega_{j_1} & \cdot & \cdot & \sum_{j_1,\cdot,j_r,j_k\neq l,j_k\neq j_{k'}}\prod_{j_k\in\{j_1,\cdot,j_r\}}\omega_{j_k} & \cdot & \cdot & \prod_{j\neq l}\omega_j \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & \sum_{j_1\neq n}\omega_{j_1} & \cdot & \cdot & \sum_{j_1,\cdot,j_r,j_k\neq n,j_k\neq j_{k'}}\prod_{j_k\in\{j_1,\cdot,j_r\}}\omega_{j_k} & \cdot & \cdot & \prod_{j\neq n}\omega_j \end{pmatrix}.$$

We then show that the determinant of this matrix is equal to the determinant of the Vandermonde matrix:

$$V_\omega=\begin{pmatrix} 1 & \omega_1 & \cdot & \cdot & \omega_1^{k-1} & \cdot & \cdot & \omega_1^{n-1} \\ 1 & \omega_2 & \cdot & \cdot & \omega_2^{k-1} & \cdot & \cdot & \omega_2^{n-1} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & \omega_l & \cdot & \cdot & \omega_l^{k-1} & \cdot & \cdot & \omega_l^{n-1} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & \omega_n & \cdot & \cdot & \omega_n^{k-1} & \cdot & \cdot & \omega_n^{n-1} \end{pmatrix}.$$

The matrix V_ω and J_ω are also denoted by $V_\omega=[V_1,\cdots,V_n]$ and $J_\omega=[J_1,\cdots,J_n]$. The argument for establishing that $det(J_\omega)=det(V_\omega)$ relies on n successive transformations that leave the determinant invariant and that go from matrix V_ω to matrix J_ω. Denote by S_k the sum $\sum_{j_1,\cdots,j_r,j_k\neq j_{k'}}\prod_{j_k\in\{j_1,\cdot,j_r\}}\omega_{j_k}$ (with the convention $S_0=1$) and respectively by $\mathbf{1}$ and I_ω the vector and the diagonal matrix:

$$\mathbf{1}=\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}, I_\omega=\begin{pmatrix} \omega_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \omega_n \end{pmatrix}.$$

By means of the recursive relation $S_{k-1}\times\mathbf{1} = J_k + I_\omega \times J_{k-1}$, for $k = 1, \cdots, n+1$ (with the convention that J_{n+1} is the null vector), we easily derive a kind of Newton-Girard formula for any $1 \leq k \leq n$:

$$J_k = \sum_{i=1}^{k} (-1)^{i+1} S_{k-i} V_i. \tag{21}$$

From matrix V_ω, if we successively replace the column k (from $k = n$ to $k = 1$) by the column $\sum_{i=1}^{k} (-1)^{i+1} S_{k-i} V_i$, the determinant is preserved at each step whereas equation (21) guarantees that the final matrix is J_ω. The determinant of the Vandermonde matrix is known to be equal to $det(V_\omega) = \prod_{1\leq i<j\leq n} (\omega_i - \omega_j)^2$ (see [4] p. 104-105). We conclude after noting that determinants are invariant by transposition and that the regularity of the Jacobian matrix of a function and its inverse are equivalent. **CQFD**

From equation (10) and (12), we have the following bounds for the densities on $\mathcal{C}(B)$ where, asymptotically, the terms $(\widehat{F}_{\mathbf{B}_\mathbf{i}^*,\mathbf{Z},\mathbf{I}}(b,z,I) - \widehat{F}_{\mathbf{B}_\mathbf{j}^*,\mathbf{Z},\mathbf{I}}(b,z,I)), j \in I \setminus \{i\}$ are bounded away from zero:

$$\begin{aligned} ||\widehat{f}_{\mathbf{B}_\mathbf{i}^*,\mathbf{Z},\mathrm{I}}(b,z,\mathrm{I}) - f_{\mathbf{B}_\mathbf{i}^*,\mathbf{Z},\mathrm{I}}(b,z,\mathrm{I})||_0 &\leq C_1 \cdot ||\widehat{f}_{\mathbf{B}_\mathbf{p},\mathbf{Z},\mathrm{I}}(b,z,\mathrm{I}) - f_{\mathbf{B}_\mathbf{p},\mathbf{Z},\mathrm{I}}(b,z,\mathrm{I})||_0 \\ &+ C_2 \cdot ||\widehat{F}_{\mathbf{B}_\mathbf{i}^*,\mathbf{Z},\mathrm{I}}(b,z,\mathrm{I}) - F_{\mathbf{B}_\mathbf{i}^*,\mathbf{Z},\mathrm{I}}(b,z,\mathrm{I})||_0 \end{aligned}$$

Thus the uniform convergence rate that holds for $\widehat{f}_{\mathbf{B}_\mathbf{p},\mathbf{Z},\mathrm{I}}(b,z,\mathrm{I})$ remains also valid for $\widehat{f}_{\mathbf{B}_\mathbf{i}^*,\mathbf{Z},\mathrm{I}}(b,z,\mathrm{I})$. In any inner compact subset of the support, the pseudo values can be expressed as a continuous differentiable function of $\widehat{f}_{\mathbf{B}_\mathbf{i}^*,\mathbf{Z},\mathrm{I}}$ and $\widehat{F}_{\mathbf{B}_\mathbf{i}^*,\mathbf{Z},\mathrm{I}}$, $i = 1, \cdots, n_I$. Furthermore, it is the rate of convergence of $\widehat{f}_{\mathbf{B}_\mathbf{i}^*,\mathbf{Z},\mathrm{I}}$ which sets the rate of convergence of $\widehat{X}_{ipl}$ to X_{ipl} in any inner compact subset of the support whereas the estimator for $\widehat{F}_{\mathbf{B}_\mathbf{i}^*,\mathbf{Z},\mathrm{I}}$ is converging at a faster rate.

The remaining issues are the consistency and the uniform rates of convergence of $\widehat{P}_{ipl}$. From equations (15), the pseudo probabilities can be expressed as a continuous differentiable function of $\widehat{f}_{\mathbf{B}_\mathbf{i}^*,\mathbf{Z},\mathrm{I}}$ $(i = 0, \cdots, n_I)$ in any inner compact subset of the support (the denominator stays bounded away from zero). Then $\widehat{P}_{ipl}$ is an asymptotically unbiased estimator of P_{ipl} and converges uniformly at the same rate as the one for $\widehat{X}_{ipl}$.