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. Monte Carlo simulations illustrate the practical relevance of our estimation procedure for small data sets.

Keywords: Auctions, nonparametric identification, nonparametric estimation, unobserved heterogeneity, anonymous bids, uniform convergence rate
JEL classification: C14, D44

*I would like to thank Bernard Salanié, who was my advisor for this part of my PhD, and Philippe Février for very helpful discussions at the early stages of this research which started at CREST-LEI. I am also grateful to Han Hong, Isabelle Perrigne, Quang Vuong, Frank Wolak and seminar participants at the 3th CAPCP conference 2010 at Penn State University, at SITE (Stanford) 2009, at the ESEM 2009 (Barcelona), at Stanford Econometric Seminar, at Stanford Econometric Lunch Workshop and at CREST-LEI lunch seminar for stimulating discussions. I am grateful to two anonymous referees for their comments to improve the paper. 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.

†Paris School of Economics, 48 Bd Jourdan 75014 Paris. e-mail: laurent.lamy@pse.ens.fr

1 Introduction

This paper is motivated by the fact that the 'identities' of the bidders are lacking to the econometrician in some auction data.
First, their formal identities (or names) may be confidential. Sometimes, the seller is reluctant, on the one hand, to disclose the identities of the losing bidders as it is perceived as helping collusion. E.g., only the identity of the winner is publicly disclosed in French timber auctions organized by ONF (French National Forest Service) and analyzed by Li and Perrigne (2003). On the other hand, the seller may be willing to disclose publicly the amounts of some losing bids in order to give to potential entrants a more accurate signal on their expected profit to participate in the auction. In French timber auctions, the amounts of the two highest losing bids are disclosed if the number of submitted bids is greater than five. With less participants, only the highest losing bid or possibly no losing bids are disclosed since it would break the anonymity paradigm -requiring that the identity of the bidder corresponding to a given bid amount can not be traced back- because the identities of the participants are observed.[1,2] Sometimes, it is the buyers themselves that are reluctant to disclose their identities and are recruiting representatives or submit telephone bids as it is common in the auctions organized by auction houses such as Christie's and Sotheby's. Ginsburgh (1998) reports having seen a fairly large number of cases of art sales in which all the bids were entered exclusively by Christie's employees acting for telephone bidders. Furthermore, Cassady (1967) reports that a buyer who is known as an outstanding figure frequently chooses to remain anonymous in the auction process, a strategy for which Hernando-Veciana and Tröge (2010) give a theoretical foundation: an insider, that has better information about a common component of the value of the good for sale than the other bidders, may benefit from protecting the value of her private information by remaining anonymous.[3]

[1]See http://www.ofme.org/documents/ONF/reglementvente for the current auction rules which have been subject to an investigation by the French Competition Authority (see the report at http://www10.finances.gouv.fr/fonds_documentaire/dgccrf/boccrf/05_04/a0040015.htm). The observation of the set of participants seems to result from the physical nature of bids' submissions.

[2]Similarly, Baldwin et al. (1997) report that 'only the identities of the second-highest and highest bidders are recorded by the Forest Service' in US and for a set of auctions that took place in the 70's. (section VI.D.). According to one referee, it is no longer the case since US Forest Service data now contain the identities of all bidders.

[3]Another rationale suggested by Cassady (1967) is when a strong buyer suspects that he may encourage the seller to bid or equivalently to update the reserve price, a rationale that is intertwined with our subsequent 'shill bidding' motivation.

Second, the asymmetry of the auction model may come from some 'group affiliation' among bidders, that make them bid according to different distributions, but that is not observed though their formal identities are observed. We emphasize that, all over the paper, what we call bidder's identity is the 'relevant' identity, i.e. the element which drives his bid distribution, and which does not necessary correspond to the formal identity.[4] In this line, the terminology 'anonymous' refers to the lack of knowledge about bidders' identities, could it be either formal identities or group affiliation. E.g. on eBay, we can distinguish between two kinds of bidders: the 'real bidders' that bid to consume the good and 'shill bidders' that some sellers use to inflate prices by means of false names. Actually, the agent managing a given bidding account on eBay may sometimes bid as a real bidder or as a shill bidder. This latter bidding activity has a structurally anonymous nature since it is often prohibited.[5] The empirical analysis of eBay auction data of Song (2004) and Sailer (2006) consider symmetric bidders and thus exclude any shill bidding activity, a pervasive phenomenon that is not confined to internet auctions[6] and that is analyzed theoretically in Lamy (2009,2010) respectively for models with pure private values and participation costs and models with interdependent values. Group affiliations may also be tailored to the case where the econometrician does not observe a leading discrete covariate that drives the asymmetry between bidders, e.g. the location of the firms[7], an incumbency status with respect to the contract for sale or a seniority status in the market. Finally, we emphasize that in most of the data sets that have been analyzed in the literature,

[4] Formal identities may be vacuous if we have only few observations per bidder such that it is not possible to estimate bidder-specific distributions.

[5] Ockenfels et al. (2006) report that, in Germany, a commercial company provides a service that automates the process of shill bidding. In particular, for a given seller, the 'shill bidder' changes identities across auctions. In the same vein, one of the French largest auction houses, the Hôtel Drouot, has been smeared by a theft scandal among his corporation of packers in 2009. As a by-product, it sheds some light on some questionable services that those packers informally offer and in particular shill bidding due to their special acquaintances with the sellers which result from their official handling responsibilities. In December 2009, they were deprived from their ability to take bidding orders and bid on their own in the auctions. See e.g. Les Mystères de l'Hôtel Drouot, Le Monde, February 02 2010.

[6] A large part of Cassady's (1967) description of how auctions work is devoted to shill bidding and related tactics to prevent oneself against it as the tactic presented in footnote 3. See also Lamy (2010) for recent headlines and more motives for shill bidding. In some auction houses, bids from the vendor are actually perfectly legal and the possibility of such bids is mentioned in their listed terms and conditions (see the discussion in Lamy (2010) on Heritage Auctions Galleries). Nevertheless, auction houses prefer to do it secretly.

[7] In Flambard and Perrigne (2006), firms are competing for snow removal contracts in the city of Montréal. Asymmetry is captured by a binary location covariate (West versus East) which is observed by the authors.

bidders' identities are known. In some cases, an important part of the econometrician's work has been to recover those identities as the location of the firms in Flambard and Perrigne (2006) or the names of the firms that won neighbor tracts in Hendricks and Porter (1988) in auctions for drainage leases. This comes from the fact that the econometrician is usually working for the seller or a regulation agency. On the contrary, if we consider a broader application of the econometrics of auction data where the econometrician may work for some bidders or potential entrants that have access only to limited public information, then the scenario where bidders' identities are not observed is probably the typical one that the econometrician would encounter[8] and then goes much beyond to our previous examples.

We consider thus 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. However, the presence of asymmetries has been the key determinant of many empirical studies of auction data. Furthermore, identifying asymmetric behavior among anonymous bidders is the right perspective for models with shill bidding or with a better informed bidder not solely because those models explicitly involve asymmetries but also since it is precisely the asymmetries between bidders that drive the need for anonymity. Collusion is another example. In Porter and Zona (1993,1999) and Pesendorfer (2000), the bidding behavior of alleged cartel participants is compared to the ones of non-cartel bidders through reduced form approaches. The lack of knowledge about bidders' formal identities would prevent their analysis of collusion. As illustrated by Asker (2010)'s example of a ring of stamp dealers, collusion may actually occur in such anonymous environments as the auctions organized by auction houses. When we observe jointly the bids from several bidders and the number of participants, then our methodology will provide a route to identify some asymmetries between bidders.[9] The aim of this paper is to lay the foundations of the econometrics 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 (2006)) and focus on private value single-unit auction models. We are dealing thus with the identification and the estimation of the distri-

[8]The mere fact that many auction theorists are commonly hired to advise auction participants, e.g. in spectrum auctions, suggests that there could be a high demand for such an activity.

[9]We implicitly assume that only one bidder of the ring bids in the auction. Otherwise, we need to impose some structure on the behavior of the losing bids submitted by the ring.

bution of bidders' valuations. Nevertheless, we emphasize that the main idea of the paper can be applied more generally, e.g. for reduced-form approaches as under the common value paradigm which is not identified in general: the key point being how to recover the distribution of the bids of the participants (up to a permutation) from data where the identities of the bidders are lacking to the econometrician.

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 (1996) 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.'s (2003) 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's (2000) 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 or if 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) $\{F_{\mathbf{B}_\mathbf{i}^*}(.)\}_{i=1,...,N}$ of each buyers' bids from the CDFs $\{F_{\mathbf{B}_\mathbf{p}}(.)\}_{p=1,...,N}$ of the order statistics of the bids. By exploiting independence, for any bid b, the vector of the N bidders CDFs at b, $(F_{\mathbf{B}_\mathbf{1}^*}(b), \ldots, F_{\mathbf{B}_\mathbf{n}^*}(b))$, corresponds to the roots of a polynomial of degree N whose coefficients are linear combinations of $(F_{\mathbf{B}_\mathbf{p}}(b))_{p=1,...,N}$.

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.[10] We establish the uniform consistency of our

[10]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

estimator. In the first price auction, the 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 as in French timber auction data. We actually know from Athey and Haile (2002) 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 'naive' approaches. On the contrary, our estimation procedure uses the complete vector of bids at both stages and also the partial information about bidders' identities. In particular, we obtain for each bid a pseudo private value according to each possible identities of the bidder. Then, to estimate the distribution of private values, we estimate for each bid the probability that it comes from a given bidder. On the whole, 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. As it is strongly supported by our Monte Carlo simulations, our procedure is then a striking improvement, especially for small data sets where 'naive' approaches are useless.

With respect to the econometric literature, our contribution is several-fold. First, whereas Athey and Haile (2002) 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. Second, we propose a nonparametric estimation procedure that corresponds to a natural generalization of GPV's procedure and analyze its asymptotic properties according to the same criteria as in GPV. Finally this work can be viewed as belonging to the general problem of unobserved heterogeneity in economet-

estimation- and thus alleviate the computational burden. See McLachlan and Krishnan (1997) for a comprehensive treatment of EM-algorithms.

rics. The bulk of the existing works are considering models where a single outcome suffers from two kinds 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 (1985)) or in the mixed proportional hazard model (see van den Berg (2001)). 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 (1998) 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 Li et al. (2000) and Krasnokutskaya (2004). 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 paper is organized as follows. In Section 2, we introduce the model and the definition of identifiability under anonymity. In Section 3, we investigate nonparametric identification. For the asymmetric IPV model which is identified and allowing for heterogeneity across auctions, we develop a multi-step kernel-based estimator in section 4 where the new caveats resulting from anonymity are presented. Section 5 illustrates the usefulness of our methodology with some Monte Carlo simulations. In section 6, we establish the asymptotic properties of our estimation procedure and in particular the rates of uniform convergence at which we estimate the latent densities of private values. The optimality of those convergence rates for estimating the densities of private values from observed bids is established in section 7. Section 8 concludes by indicating some future lines of research. Most proofs are relegated to the Appendix. A supplementary material (henceforth Supp. Mat.) is devoted to the exposition of some technical or complementary elements that are less central in the paper and/or have close analogs in GPV.

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.[11] Though 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 $B = (B_1, \ldots, B_p, \ldots, B_n)$, with $B_1 \leq \ldots \leq B_n$, where B_p denotes the p^{th} order statistic of the vector of bids B. But she does not observe $B^* = (B_1^*, \ldots, B_i^*, \ldots, B_n^*)$, 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.[12,13] 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 $\mathbf{S}$ 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 continuous, 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 and where symmetric bidders are using the same bidding function. In the equilibrium

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

[12] 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.

[13] 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 (2006) for the analysis of the first-price auction with different supports.

of the second price auction, buyers are thus bidding their private value. Hence, the link between bids and private types is straightforward:

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

In the first price auction, under assumption A1, Athey (2001) guarantees the existence of an increasing pure strategy equilibrium in the IPV and APV models. Following GPV, the link between bids and types for each bidder i is made by a repameterization of the first order differential equation derived from bidder i's optimization program (see Li and al. (2002) for APV models):

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

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

Following Laffont and Vuong (1996), 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 the standard notion of 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 bids, 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 $\pi : [1, n] \rightarrow [1, n]$ 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 (1981) 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 in contrast to Campo et al.'s (2003) 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. It is the innovative step: by an appropriate reparameterization, the nonlinear inverse problem we face is reduced to a known one, namely 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 A1, 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* $F_{\mathbf{B}}(.)$*.*
- *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. (2002) 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 strictly correlated. In particular, it means that independence can not be fully tested.

In a nutshell, Proposition 3.1 says that the econometrician has to depart from the general affiliated private value model to obtain identification and proposes two identification routes: either to assume symmetry or to assume independence. Those additional assumptions are actually stronger than what is 'needed' to identify the model, i.e. they can be refuted from the observables under some circumstances. On the one hand, for a given asymmetric APV model, we may reject that it comes from a symmetric APV model: under a symmetric affiliated value model, then the variables $B_1, \cdots, B_n$ are affiliated in both the first and second price auctions (Milgrom and Weber (1982)), while $B_1, \cdots, B_n$ may not be affiliated under an asymmetric APV model. The intuition is that asymmetry may be a source of negative correlation between order statistics.[14] De Castro and Paarsch (2009) develop a test of symmetric affiliation under nonanonymous data, a test that is then suitable for our setup with anonymous data and can thus be used to refute the symmetric APV model.[15] On the other hand, for a given asymmetric APV model, we may reject that it comes from an asymmetric IPV model: independence involves some testable restrictions under anonymity and can be thus partially tested even if it can not be fully tested as argued above.[16]

[14]To gain intuition, consider two kinds of bidders (strong versus weak). If B_n raises, then you may attribute a greater probability that it comes from a strong bidder and thus that a given remaining order statistic comes for a weak bidder and is thus lower.

[15]On the contrary, if we do not impose any correlation structure as independence or affiliation, then, in the second price auction, any asymmetric PV model is observationally equivalent to some symmetric PV model: the symmetric PV model where $f_{\mathbf{X}}(x_1, \cdots, x_n) = \frac{1}{n!} \sum_{\sigma \in \Sigma_n} f_{\mathbf{B}}(x_{\sigma(1)}, \cdots, x_{\sigma(n)})$ (with Σ_n denoting the set of permutations among n elements) is always consistent with the observables. In the first price auction, note that the argument is not valid: we may be able to reject the symmetric APV model if the corresponding inverse bidding function $\xi_i^{rst}(b)$ is not strictly increasing when we move from bids to valuations.

[16]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 test proposed by Blum et al. (1961), 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. Under anonymity, independence involves some testable restrictions based on a set of generalized discriminants as shown in Lamy (2008). Partial independence tests could then be built on those discriminants in the same way as the tests for symmetry that were proposed in Lamy (2008).

The rest of this section is devoted to the proof of the third point. From definition 1, saying that the 'asymmetric IPV model is identified' is equivalent to say that if two IPV models $\{F_{\mathbf{X_i}}(.)\}_{i\in[1,n]}$ and $\{F'_{\mathbf{X_i}}(.)\}_{i\in[1,n]}$ lead to the same distribution of bids $F_{\mathbf{B}}$, then there exists a permutation $\pi:[1,n]\to[1,n]$ such that $F_{\mathbf{X_i}}(x)=F'_{\mathbf{X}_{\pi(\mathbf{i})}}(x)$, $i=1,\ldots,n$, for almost any type x. Straightforwardly for the second price auction and from GPV for the first price auction, it is actually sufficient to show that if $\{F_{\mathbf{B_i^*}}(.)\}_{i\in[1,n]}$ and $\{F'_{\mathbf{B_i^*}}(.)\}_{i\in[1,n]}$ lead to the same distribution of bids $F_{\mathbf{B}}(.)$, then there exists a permutation $\pi:[1,n]\to[1,n]$ such that $F_{\mathbf{B_i^*}}(b)=F'_{\mathbf{B}^*_{\pi(\mathbf{i})}}(b)$, $i=1,\ldots,n$, for almost any bid b. Define $F_{\mathbf{B}}^{(r:m)}(.)$ for $r\leq m\leq n$ as the CDF of the r^{th} order statistic among $(B_{1m},\ldots,B_{mm})$ where the latter are independently drawn without replacement from $(B_1,\ldots,B_n)$. Then we can identify the CDF $F_{\mathbf{B}}^{(r:m)}(u)$ by recursive use of the formula (see Athey and Haile (2002) 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 $F_{\mathbf{B}}^{(p:n)}(.)=F_{\mathbf{B_p}}(.)$, where the CDFs $F_{\mathbf{B_p}}(.)$, $p=1,\ldots,n$, are observed. 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 $p=1,\ldots,n$. 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.

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

where the notation $\sum_{i_1,\cdots,i_r,i_l\neq i_{l'}}$ corresponds to $\sum_{i_1=1}^{n}\sum_{\substack{i_2=1\\ i_2\neq i_1}}^{n}\cdots\sum_{\substack{i_m=1\\ i_m\neq i_1,\cdots,i_{m-1}}}^{n}\cdots\sum_{\substack{i_r=1\\ i_r\neq i_1,\cdots,i_{r-1}}}^{n}$. [17]

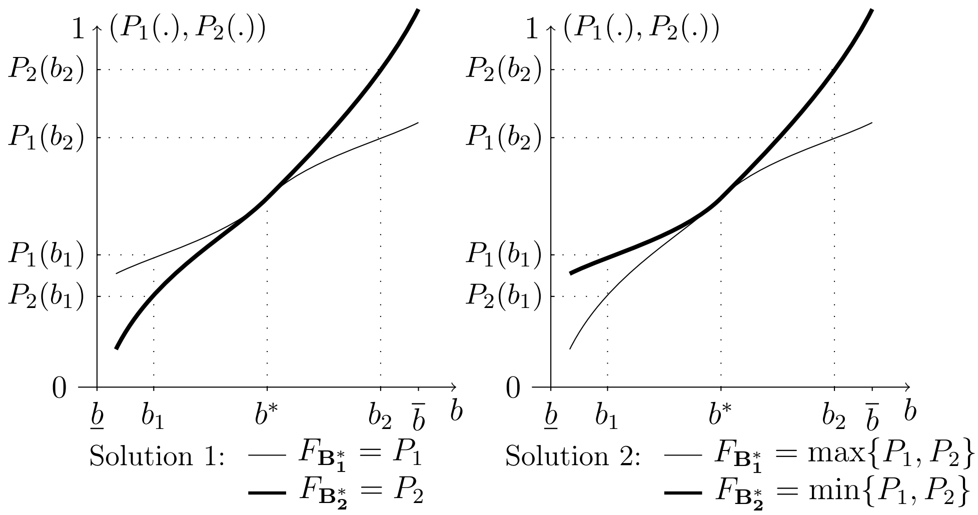
The right expressions in the system (4) are closely related to the coefficients of the expansion of the polynomial $X\to\prod_{i=1}^{n}(X-F_{\mathbf{B}_{\mathbf{i}}^*}(b))$. The coefficient in front of the monomial X^{n-r} for $1\leq r\leq n$ is given by $(-1)^r\cdot\sum_{i_1=1}^{n}\sum_{\substack{i_2=1\\ i_2>i_1}}^{n}\cdots\sum_{\substack{i_r=1\\ i_r>i_{r-1}}}^{n}\prod_{i_k\in\{i_1,\cdots,i_r\}}F_{\mathbf{B}_{\mathbf{i_k}}^*}(b)$ which is also equal to $\frac{(-1)^r}{r!}\cdot\sum_{i_1,\cdots,i_r,i_l\neq i_{l'}}\prod_{i_k\in\{i_1,\cdots,i_r\}}F_{\mathbf{B}_{\mathbf{i_k}}^*}(b)$. From the Fundamental Theorem of Algebra (see Basu, Pollack and Roy (2006)), the factorization of a polynomial according to its roots among the complex numbers $\mathcal{C}$ 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:

$$u\to\sum_{s=0}^{n}a_s\cdot\frac{n!}{(n-s)!s!}\cdot(-1)^{n-s}\cdot u^s, \tag{5}$$

where $a_n=1$ and $a_s=F_{\mathbf{B}}^{(n-s:n-s)}(b)$, for $0\leq s\leq n-1$. 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 Basu et al. (2006)), there exists a continuous function $b\to(P_1(b),\ldots,P_n(b))$ mapping the vector of pointwise solutions. At this stage, we have shown solely that for any bid b, then the true probabilities $F_{\mathbf{B}_{\mathbf{i}}^*}(b)$, $i=1,\ldots,n$, are the unique solution up to a permutation. What remains to show is the more restrictive condition that the true CDFs $F_{\mathbf{B}_{\mathbf{i}}^*}(.)$, $i=1,\ldots,n$, 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 $F_{\mathbf{B}_{\mathbf{i}}^*}(.)$, $i=1,\ldots,n$, the only candidate solution would be $(P_1(.),\ldots,P_n(.))=(F_{\mathbf{B}_{\mathbf{1}}^*}(.),\ldots,F_{\mathbf{B}_{\mathbf{n}}^*}(.))$ (up to a permutation). On the contrary, if the maps $P_i(.)$ cross and in a way such that at some crossing point i,j,b^* with $P_i(b^*)=P_j(b^*)$ we have $P_i'(b^*)=P_j'(b^*)$, then the continuously differentiable selection of the roots $(P_1(.),\ldots,P_n(.))$ 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. In other words, 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.

[17] We stick to this notation all over the paper. In a similar way, the notation $\sum_{i_1,\cdots,i_r,i_l\neq k,i_l\neq i_{l'}}$ will correspond to $\sum_{\substack{i_1=1\\ i_1\neq k}}^{n}\sum_{\substack{i_2=1\\ i_2\neq i_1\\ i_2\neq k}}^{n}\cdots\sum_{\substack{i_m=1\\ i_m\neq i_1,\cdots,i_{m-1}\\ i_m\neq k}}^{n}\cdots\sum_{\substack{i_r=1\\ i_r\neq i_1,\cdots,i_{r-1}\\ i_r\neq k}}^{n}$.

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



Nevertheless, the knowledge of the joint distribution $F_{\mathbf{B}}(.)$ of all order statistics selects a unique solution. To gain intuition, consider for example 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 $(P_2'(b_2) - P_1'(b_2)) \cdot (P_2'(b_1) - P_1'(b_1)) \neq 0$.

We now move to the general argument. The proof below is limited to the case where the CDFs $F_{\mathbf{X_i}}(.)$, $i = 1, \ldots, n$, are all distinct while the general case where some CDFs $F_{\mathbf{X_i}}(.)$, $i = 1, \ldots, n$, may coincide is deferred to the appendix. Define $f_{\mathbf{B}}^{([1,m]:n)}(u_1, \ldots, u_m)$ for $m \leq n$ as the PDF of the vector $(B_{1m}, \ldots, B_{mm})$ where the latter is built from independent draws without replacement from $(B_1, \ldots, B_n)$. Independence gives the following general expression for $f_{\mathbf{B}}^{([1,m]:n)}$:

$$f_{\mathbf{B}}^{([1,m]:n)}(u_1, \ldots, u_m) = \frac{(n-m)!m!}{n!} \sum_{i_1, \cdots, i_m, i_l \neq i_{l'}} \prod_{j=1}^{m} f_{\mathbf{B^*_{i_j}}}(u_j)$$

In particular, we obtain the following system of equations for any u_1, u_2:

$$
\begin{aligned}
f_{\mathbf{B}}^{(1:n)}(u_2) =& \frac{1}{n} \cdot \sum_{i_1=1}^{n} f_{\mathbf{B}^*_{\mathbf{i_1}}}(u_2) \\
f_{\mathbf{B}}^{([1,2]:n)}(u_2, u_1) =& \frac{1}{n(n-1)} \cdot \sum_{i_1=1}^{n} f_{\mathbf{B}^*_{\mathbf{i_1}}}(u_2) \cdot \left(\sum_{\substack{i_2=1 \\ i_2 \neq i_1}}^{n} f_{\mathbf{B}^*_{\mathbf{i_2}}}(u_1) \right) \\
& \cdots\cdots \\
f_{\mathbf{B}}^{([1,r]:n)}(u_2, \underbrace{u_1, \ldots, u_1}_{(r-1)-times}) =& \frac{1}{n \cdots (n-r+1)} \cdot \sum_{i_1=1}^{n} f_{\mathbf{B}^*_{\mathbf{i_1}}}(u_2) \cdot \left(\sum_{i_2, \cdots, i_r, i_l \neq i_1, i_l \neq i_{l'}} \prod_{i_k \in \{i_2, \cdots, i_r\}} f_{\mathbf{B}^*_{\mathbf{i_k}}}(u_1) \right) \\
& \cdots\cdots \\
f_{\mathbf{B}}^{([1,n]:n)}(u_2, \underbrace{u_1, \ldots, u_1}_{(n-1)-times}) =& \frac{1}{n!} \cdot \sum_{i_1=1}^{n} f_{\mathbf{B}^*_{\mathbf{i_1}}}(u_2) \cdot \left(\sum_{i_2, \cdots, i_n, i_l \neq i_1, i_l \neq i_{l'}} \prod_{i_k \in \{i_2, \cdots, i_n\}} f_{\mathbf{B}^*_{\mathbf{i_k}}}(u_1) \right)
\end{aligned} \tag{6}
$$

Note that the left terms are observed. Then, after integrating according to the variable u_2 from the lower bound of the bidding distribution $\underline{x}$ to b_2, we obtain that the products of the form $X \times J_{(f_{\mathbf{B}^*_{\mathbf{1}}}(u_1), \ldots, f_{\mathbf{B}^*_{\mathbf{n}}}(u_1))}$, where $J_{(f_{\mathbf{B}^*_{\mathbf{1}}}(u_1), \ldots, f_{\mathbf{B}^*_{\mathbf{n}}}(u_1))}$ is a matrix defined in appendix A.1 and $X = [F_{\mathbf{B}^*_{\mathbf{1}}}(b_2), \ldots, F_{\mathbf{B}^*_{\mathbf{n}}}(b_2)]$, are identified.

Consider in a first step that there exists b_1 such that the $f_{\mathbf{B}^*_{\mathbf{i}}}(b_1)$'s, for $i = 1, \ldots, n$, are all distinct. Consider a candidate to be a solution, i.e. pick a continuously differentiable selection of the roots $(P_1(.), \ldots, P_n(.))$. Since each solution is defined up to a permutation, we can assume w.l.o.g. that $(P'_1(b_1), \ldots, P'_n(b_1)) = (f_{\mathbf{B}^*_{\mathbf{1}}}(b_1), \ldots, f_{\mathbf{B}^*_{\mathbf{n}}}(b_1))$. For any b_2, the system of equations (6) implies that:

$$
[P_1(b_2), \ldots, P_n(b_2)] \times J_{(f_{\mathbf{B}^*_{\mathbf{1}}}(b_1), \ldots, f_{\mathbf{B}^*_{\mathbf{n}}}(b_1))} = [F_{\mathbf{B}^*_{\mathbf{1}}}(b_2), \ldots, F_{\mathbf{B}^*_{\mathbf{n}}}(b_2)] \times J_{(f_{\mathbf{B}^*_{\mathbf{1}}}(b_1), \ldots, f_{\mathbf{B}^*_{\mathbf{n}}}(b_1))}. \tag{7}
$$

The matrix $J_{(f_{\mathbf{B}^*_{\mathbf{1}}}(b_1), \ldots, f_{\mathbf{B}^*_{\mathbf{n}}}(b_1))}$ is invertible as a corollary of lemma A.1. Finally, we obtain that $(P_1(.), \ldots, P_n(.)) = (F_{\mathbf{B}^*_{\mathbf{1}}}(.), \ldots, F_{\mathbf{B}^*_{\mathbf{n}}}(.))$ is the unique solution.

The existence of b_1 such that the $f_{\mathbf{B}^*_{\mathbf{i}}}(b_1)$'s, for $i = 1, \ldots, n$, are all distinct is not guaranteed in general. However, as shown in the Appendix A.2, there exists more generally an event E such that the probabilities of the events $B_i^* \in E$, $i = 1, \ldots, n$, are all distinct. This is the key element that drives the proof since it is clear that analogs of the system of equations (6) imply analog forms for (7) where $J_{(f_{\mathbf{B}^*_{\mathbf{1}}}(b_1), \ldots, f_{\mathbf{B}^*_{\mathbf{n}}}(b_1))}$ is replaced by $J_{(\text{Prob}(B_1^*\in E),\ldots,\text{Prob}(B_n^*\in E))}$.

4 Nonparametric Estimation

For the estimation procedure, we consider that the set of participants is fixed across auctions but we allow auctioned objects to be heterogeneous. Let $Z_l \in \mathcal{R}^d$ denote the d-dimensional vector of relevant continuous characteristics for the l^{th} auctioned object. The vector Z_l is assumed to be common knowledge among bidders and is also observed by the econometrician. Our methodology can be immediately extended to an environment where the set of participants varies across auctions provided that the set of participants is known to the econometrician. We emphasize that the observation of the set of participants does not correspond to the requirement of the observation by the econometrician of the formal identities of the participants.[18] What we are assuming is that the econometrician can isolate the sets of auctions such that the generating process of bidders' valuations is the same, i.e. where the same set $\{F_{\mathbf{X_i},\mathbf{Z}}(.,.)\}_{i=1,\ldots,n}$ prevails. According to our 'group affiliation' perspective, this knowledge is quite natural in many instances: the econometrician may know that there is a unique 'incumbent' in a procurement that consists in a renewal of a contract, she may also know that there is a unique 'informed' bidder in the auction, a unique shill bidder or a bidder that represents a ring of colluders, while considering that the remaining bidders are all symmetric.[19]

Relative to our previous notation, we will now work with conditional CDFs and PDFs of private values and bids given Z_l. E.g. $F_{\mathbf{X_i}|\mathbf{Z}}(.|Z_l)$ denotes the CDF of bidder i's private value X_{il} in the l^{th} auction. Using independence, (1) and (2) can be

[18] This latter requirement seems to stand in contradiction with our paradigm of anonymous bids. Nevertheless, the set of participants could be observed due to the physical nature of bid submissions. Alternatively and especially in environments where the set of potential participants is limited, one can assume that the same set of potential bidders prevails in all the auctions while variations in the actual set of bidders is explained by a binding reserve price.

[19] In the current form, our methodology works under a model where the econometrician knows for sure that there is a 'special bidder' that bids differently, e.g. a shill bidder in an auction house where this activity is legal (see footnote 6). Actually, in many instances, especially shill bidding, she would only suspect that this 'special' bidder participates with some probability and what we want to estimate would be the bidding distribution of all bidders and the probability to participate of the special bidder. Our methodology would enable us to estimate the distribution of the symmetric bidder plus the distribution of a bidder which should be interpreted as a mixture of the 'special' bidder and the symmetric bidders. Under mild assumptions, e.g. if the support of the two distributions are not equal as it would be naturally the case in models with shill bidding, we can separate the two elements of the mixture and then identify the model.

rewritten as

$$X_{il} = B^*_{il} + \psi_i(B^*_{il}|Z_l), \text{ where } \psi_i(.|.) \text{ is defined as} \tag{8}$$

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

In this section, we adapt GPV's two step estimation procedure to recover the densities of bidders' private values.[20] Two caveats arise. First we can not directly estimate with kernel techniques the ratios $\frac{f_{\mathbf{B}^*_i|\mathbf{Z}}(\cdot|\cdot)}{F_{\mathbf{B}^*_i|\mathbf{Z}}(\cdot|\cdot)}$ since identities are not observed. Thus we need to convert our estimations of the CDFs and PDFs of B_p, that can be done with the standard kernel estimation techniques as in GPV, into estimations for the CDFs and PDFs of B^*_i. Second, if $\frac{f_{\mathbf{B}^*_i|\mathbf{Z}}(\cdot|\cdot)}{F_{\mathbf{B}^*_i|\mathbf{Z}}(\cdot|\cdot)}$ is suitably estimated, we can apply (8) 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 $B_l = (B_{1l}, \cdots, B_{pl}, \cdots, B_{nl})$, we have to estimate the probability that buyer i's bid B^*_{il} is equal to B_{pl} for any $p = 1, \ldots, n$. 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 estimate nonparametrically the PDFs of buyers' private values. 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_n the set of the $n!$ permutations between participants' identities and the order statistics of the bids. Such an assignment of the bids to the participants is denoted $\pi : [1, n] \to [1, n]$ 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_n which may depend both on the vector of bids B and the auction (but not on B^*). Denote by σ_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 $\pi \in \sigma_l$.

[20]See Flambard and Perrigne (2006) for the implementation of GPV's procedure in the asymmetric IPV model with nonanonymous bids.

$\sigma_l = \Sigma_n$ corresponds to the case where bids are fully anonymous, whereas the opposite case where σ_l is always a singleton corresponds to non-anonymous bids.

Our estimation procedure will cover the cases where some bidders are symmetric. More precisely, the estimation procedure will depend on the so-called underlying asymmetry structure.

Definition 2 *An* ***asymmetry structure*** *is a vector of integers* $(d_1, \cdots, d_r)$ *where* $\sum_{k=1}^{r} d_k = n$ *and* $d_1 \geq \cdots \geq d_r \geq 1$. *The integer* r *corresponds to the number of distinct elements in the structure.*

Definition 3 *Two (univariate) CDFs* $F(.)$ *and* $G(.)$ *are called* ***strictly distinct*** *if there is no interval* $\mathfrak{I}$ *with positive measure such that* $F(.) = G(.) \in (0, 1)$ *on* $\mathfrak{I}$.

Definition 4 *The asymmetric IPV model is said to be* ***generated by a given asymmetry structure*** $(d_1, \cdots, d_r)$ *if, for any realization* z *of the covariates, there exists a set of* r *strictly distinct CDFs* $(G_k^{\mathbf{X}}(.))_{k=1,\ldots,r}$ *such that for each* $k = 1, \ldots, r$, *there exists exactly* d_k *bidders whose private value CDFs* $F_{\mathbf{X_i}|Z}(.|z)$ *match the CDF* $G_k^{\mathbf{X}}(.)$.

Under full asymmetry and full symmetry, the asymmetry structures are respectively $(1, \ldots, 1)$ and (n). Next we present our multi-step kernel-based estimation procedure in two stages. First, we consider environments under full asymmetry and then we move to general asymmetry structures. From lemma 2.1 in the Supp. Mat., the asymmetry structure of the private values CDFs is passed on the bids CDFs: if the asymmetric IPV model is generated by a given asymmetry structure $(d_1, \cdots, d_r)$, then for any realization z of the covariates, there exists a set of r strictly distinct CDFs $(G_k^{\mathbf{B}}(.))_{k=1,\ldots,r}$ such that for each $k = 1, \ldots, r$, there exists exactly d_k bidders whose bidding CDFs $F_{\mathbf{B_i}|Z}(.|z)$ match the CDF $G_k^{\mathbf{B}}(.)$.

Furthermore, we consider in our estimation procedure that the asymmetry structure is known to the econometrician:

Assumption A 2 *The asymmetric IPV model is generated by a given asymmetry structure that is known to the econometrician.*

We emphasize that this is just an assumption aimed to shorten the paper. We can actually test independently the form of the asymmetry structure according to the general principles that have been developed in Lamy (2008) (see also footnote 16). If such tests are implemented before the estimation procedure, the right structure will be selected with probability one asymptotically and our statistical results will extend immediately to the general case where the econometrician does not know the true asymmetry structure.

4.1 Estimation under full asymmetry

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

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

$$\widehat{F}_{\mathbf{B_p},\mathbf{Z}}(b,z) = [\frac{1}{Lh^d_{F_{\mathbf{B_p}|\mathbf{Z}}}} \sum_{l=1}^{L} \mathbf{1}(B_{pl} \leq b) K_{F_{\mathbf{B_p}|\mathbf{Z}}} \left(\frac{z - Z_l}{h_{F_{\mathbf{B_p}|\mathbf{Z}}}} \right)]^+ \tag{9}$$

$$\widehat{f}_{\mathbf{B_p},\mathbf{Z}}(b,z) = [\frac{1}{Lh^d_{f_{\mathbf{B_p}|\mathbf{Z}}}} \sum_{l=1}^{L} K_{f_{\mathbf{B_p}|\mathbf{Z}}} \left(\frac{b - B_{pl}}{h_{f_{\mathbf{B_p}|\mathbf{Z}}}}, \frac{z - Z_l}{h_{f_{\mathbf{B_p}|\mathbf{Z}}}} \right)]^+ \tag{10}$$

$$\widehat{f}_{\mathbf{Z}}(z) = [\frac{1}{Lh^d_{f_{\mathbf{Z}}}} \sum_{l=1}^{L} \sum_{p=1}^{n} K_{f_{\mathbf{Z}}}(\frac{z - Z_l}{h_{f_{\mathbf{Z}}}}) \cdot]^+. \tag{11}$$

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.

Then the corresponding CDFs and PDFs conditional on Z are estimated by:

$$\widehat{F}_{\mathbf{B_p}|\mathbf{Z}}(b|z) = \min\{\frac{\widehat{F}_{\mathbf{B_p},\mathbf{Z}}(b,z)}{\widehat{f}_{\mathbf{Z}}(z)}, 1\} \quad \text{and} \quad \widehat{f}_{\mathbf{B_p}|\mathbf{Z}}(b|z) = \frac{\widehat{f}_{\mathbf{B_p},\mathbf{Z}}(b,z)}{\widehat{f}_{\mathbf{Z}}(z)}. \tag{12}$$

Second step By recursive use of the empirical counterpart of the formula (3), we estimate $\widehat{F}^{(r:r)}_{\mathbf{B}|\mathbf{Z}}(b|z)$ and $\widehat{f}^{(r:r)}_{\mathbf{B}|\mathbf{Z}}(b|z)$ for $r = 1, \cdots, n$, which respectively correspond (up to known multiplicative coefficients) 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 $\{F_{\mathbf{B^*_i},\mathbf{Z}}(b|z)\}_{i\in[1,n]}$.

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

$$\begin{aligned}\frac{m-r}{m}\widehat{F}^{(r:m)}_{\mathbf{B}|\mathbf{Z}}(b|z)+\frac{r}{m}\widehat{F}^{(r+1:m)}_{\mathbf{B}|\mathbf{Z}}(b|z)&=\widehat{F}^{(r:m-1)}_{\mathbf{B}|\mathbf{Z}}(b|z)\\ \frac{m-r}{m}\widehat{f}^{(r:m)}_{\mathbf{B}|\mathbf{Z}}(b|z)+\frac{r}{m}\widehat{f}^{(r+1:m)}_{\mathbf{B}|\mathbf{Z}}(b|z)&=\widehat{f}^{(r:m-1)}_{\mathbf{B}|\mathbf{Z}}(b|z).\end{aligned} \tag{13}$$

As a weighted sum of the estimators $\{\widehat{F}_{\mathbf{B_p}|\mathbf{Z}}(b|z)\}_{p=1,\dots,n}$ which are confined in the interval $[0,1]$, the estimators $\widehat{F}^{(r:m)}_{\mathbf{B}|\mathbf{Z}}(b|z)$ 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 $\omega_1 \geq \cdots \geq \omega_n$ 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 $Q(u) = u^n + \sum_{i=0}^{n-1} a_i \cdot \frac{n!}{(n-i)!i!} \cdot (-1)^{n-i}u^i$, i.e. $Q(u) = \prod_{i=1}^n (u-\omega_i)$. Theorem 5.12 in Basu et al. (2006) states 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 $\widehat{F}_{\mathbf{B_i^*}|\mathbf{Z}}(.|.), i \in [1,n]$ by

$$(\widehat{F}_{\mathbf{B_1^*}|\mathbf{Z}}(b|z), \cdots, \widehat{F}_{\mathbf{B_n^*}|\mathbf{Z}}(b|z)) = \mathcal{R}[\Upsilon(\widehat{a}_{n-1}(b|z), \cdots, \widehat{a}_0(b|z))], \tag{14}$$

where $\widehat{a}_s(b|z) = \widehat{F}^{(n-s:n-s)}_{\mathbf{B}|\mathbf{Z}}(b|z)$ and $\mathcal{R}[c]$ denotes the real part of the complex vector c. The rest of this step is devoted to the estimation of $f_{\mathbf{B_i^*}|\mathbf{Z}}(b|z)$. The derivative of the polynomial relation with respect to b leads to:

$$\begin{aligned}\frac{\partial Q(u)}{\partial b} &= \sum_{s=0}^{n-1} \frac{\partial a_s}{\partial b}(b|z) \cdot \frac{n!}{(n-s)!s!} \cdot (-1)^{n-s} \cdot u^s\\ &= -\sum_{i=1}^{n} \prod_{\substack{j=1\\ j\neq i}}^{n} \left(u - F_{\mathbf{B_j^*}|\mathbf{Z}}(b|z)\right) \cdot f_{\mathbf{B_i^*}|\mathbf{Z}}(b|z), \forall u, b, z,\end{aligned}$$

where $\frac{\partial a_s}{\partial b}(b|z) = f^{(n-s:n-s)}_{\mathbf{B}|\mathbf{Z}}(b|z)$. For a single estimated root, i.e. for i such that $\widehat{F}_{\mathbf{B_i^*}|\mathbf{Z}}(b|z) \neq \widehat{F}_{\mathbf{B_j^*}|\mathbf{Z}}(b|z)$ for any $j \neq i$, we have a natural estimator for the corresponding density:

$$\widehat{f}_{\mathbf{B_i^*}|\mathbf{Z}}(b|z) = \frac{\sum_{s=0}^{n-1} \frac{\partial \widehat{a}_s}{\partial b}(b|z) \cdot \frac{n!}{(n-s)!s!} \cdot (-1)^{n-s+1} \cdot \left[\widehat{F}_{\mathbf{B_i^*}|\mathbf{Z}}(b|z)\right]^s}{\prod_{\substack{j=1\\ j\neq i}}^{n} \left(\widehat{F}_{\mathbf{B_i^*}|\mathbf{Z}}(b|z) - \widehat{F}_{\mathbf{B_j^*}|\mathbf{Z}}(b|z)\right)}, \tag{15}$$

where $\frac{\partial \widehat{a}_s}{\partial b}(b|z) = \widehat{f}_{\mathbf{B}|\mathbf{Z}}^{(n-s:n-s)}(b|z)$. Consider now the case of a multiple estimated root of multiplicity $k > 1$, i.e. consider $J = \{j_m, \ldots, j_{m+k-1}\}$ such that for any $i \in J$, $\widehat{F}_{\mathbf{B}_\mathbf{i}^*,\mathbf{Z}}(b,z) = \widehat{F}_{\mathbf{B}_\mathbf{j}^*,\mathbf{Z}}(b,z)$ 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:

$$\begin{aligned}\frac{\partial Q(u)}{\partial b(\partial u)^{k-1}} &= \sum_{s=0}^{n-k} \frac{\partial a_{s+k-1}}{\partial b}(b|z) \cdot \frac{n!}{(n-s-k+1)!s!} \cdot (-1)^{n-s-k+1} \cdot u^s \\ &= -\sum_{i=1}^{n} \frac{\partial \prod_{\substack{j=1 \\ j\neq i}}^{n} \left(u - F_{\mathbf{B}_\mathbf{j}^*|\mathbf{Z}}(b|z)\right)}{(\partial u)^{k-1}} \cdot f_{\mathbf{B}_\mathbf{i}^*|\mathbf{Z}}(b|z), \forall u, b, z.\end{aligned}$$

For $i \in J$ [resp. $i \notin J$], the expression $\partial \prod_{\substack{j=1 \\ j\neq i}}^{n} \left(u - F_{\mathbf{B}_\mathbf{j}^*|\mathbf{Z}}(b|z)\right)/(\partial u)^{k-1}$ evaluated at $u = F_{\mathbf{B}_\mathbf{i}^*|\mathbf{Z}}(b|z)$ reduces to $(k-1)! \prod_{\substack{j=1 \\ j\notin J}}^{n} \left(F_{\mathbf{B}_\mathbf{i}^*|\mathbf{Z}}(b|z) - F_{\mathbf{B}_\mathbf{j}^*|\mathbf{Z}}(b|z)\right)$ [resp. 0]. Finally, we have a natural estimator for the corresponding density:

$$\widehat{f}_{\mathbf{B}_\mathbf{i}^*|\mathbf{Z}}(b|z) = \frac{\sum_{s=0}^{n-k} \frac{\partial \widehat{a}_{s+k-1}}{\partial b}(b|z) \cdot \frac{n!}{(n-s-k+1)!s!} \cdot (-1)^{n-s-k} \cdot \left[\widehat{F}_{\mathbf{B}_\mathbf{i}^*|\mathbf{Z}}(b|z)\right]^s}{k! \cdot \prod_{\substack{j=1 \\ j\notin J}} \left(\widehat{F}_{\mathbf{B}_\mathbf{i}^*|\mathbf{Z}}(b|z) - \widehat{F}_{\mathbf{B}_\mathbf{j}^*|\mathbf{Z}}(b|z)\right)}. \quad (16)$$

For $k = 1$, this formula corresponds exactly to (15).

Remark 4.1 Now we have all the elements to estimate the function $\psi_i(.|.)$ in the first price auction. In particular, we can immediately end our estimation procedure by using the empirical counterpart of the one-to-one mapping between valuations and bids, i.e. $F_{\mathbf{X_i}|\mathbf{Z}}(x|z) = F_{\mathbf{B}_\mathbf{i}^*|\mathbf{Z}}(\psi_i(x|z)|z)$, which would lead to an estimator with the same (optimal) uniform converge rates as our following more complex procedure. Similarly, the job seems to be done in the second price auction since we have recovered the bid distributions which correspond exactly to the valuation distributions. However, we emphasize that the remaining steps we introduce are crucial from a empirical perspective with partially anonymous data as it will be illustrated in our reported Monte Carlo simulations. At this stage, we still have not used the additional information σ_l which mainly motivates the three remaining steps where we build a pseudo sample of private values in the same way as in GPV 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_l.

Fourth step In view of (8) and similarly to GPV, it would be natural to construct a sample of pseudo private values $\{\widehat{X}_{ipl}; i = 1, \cdots, n; p = 1, \cdots, n; l = 1, \cdots, L\}$ where

$\widehat{X}_{ipl}$ would correspond to 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: $\widehat{X}_{ipl} = B_{pl} + \widehat{\psi}_i(B_{pl}|Z_l)$, where $\widehat{\psi}_i(b|z)$ equals respectively $\left[\sum_{\substack{j=1 \\ j\neq i}}^{n} \frac{\widehat{f}_{\mathbf{B}_\mathbf{j}^*|\mathbf{Z}}(b|z)}{\widehat{F}_{\mathbf{B}_\mathbf{j}^*|\mathbf{Z}}(b|z)}\right]^{-1}$ and 0 in the first and second price auctions.

Unfortunately, as already 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 details of this step are relegated in the Supp. Mat.

Fifth step Contrary to GPV, we should not use directly this pseudo sample of private values in a standard kernel estimation to estimate $f_{\mathbf{X_i},\mathbf{Z}}(x,z)$. 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 $n!$ vectors of private values that can rationalize a given vector of bids $(B_{1l},\cdots,B_{nl})$. Denote by $\widetilde{\pi} \in \Sigma_n$ the true permutation that matches a given vector of bidding order statistics $(B_{1l},\cdots,B_{nl})$ with the unobserved vector of bids $(B_{il}^*)_{i\in[1,n]}$. The following expression gives the theoretical probability, denoted by $Prob(\widetilde{\pi}=\pi|(b_1,\cdots,b_n,z))$, that the assignment of bidders to the observed order statistics corresponds to a given permutation π:

$$Prob(\widetilde{\pi}=\pi|(b_1,\cdots,b_n,z)) = \frac{\prod_{i=1}^n f_{\mathbf{B}_\mathbf{i}^*|\mathbf{Z}}(b_{\pi(i)}|z)}{\sum_{\pi'\in\sigma_l}\prod_{i=1}^n f_{\mathbf{B}_\mathbf{i}^*|\mathbf{Z}}(b_{\pi'(i)}|z)} \cdot \mathbf{1}\{\pi\in\sigma_l\}. \qquad (17)$$

Note that we use the information set σ_l 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 bidder i to the p^{th} order statistic, i.e. $P_{ip} = \sum_{\pi\in\Sigma_n \ s.t. \ \pi(i)=p} Prob(\widetilde{\pi}=\pi|(b_1,\cdots,b_n,z))$. Its empirical counterpart $\widehat{P}_{ipl}$ is given straightforwardly by means of our previous estimators:

$$\widehat{P}_{ipl} = \sum_{\pi\in\Sigma_n \ s.t. \ \pi(i)=p} \frac{\prod_{i=1}^n \widehat{f}_{\mathbf{B}_\mathbf{i}^*|\mathbf{Z}}(B_{\pi(i)l}|Z_l)}{\sum_{\pi'\in\sigma_l}\prod_{i=1}^n \widehat{f}_{\mathbf{B}_\mathbf{i}^*|\mathbf{Z}}(B_{\pi'(i)l}|Z_l)} \cdot \mathbf{1}\{\pi\in\sigma_l\}, \qquad (18)$$

where we set $\widehat{P}_{ipl}=0$ if the denominator vanishes, i.e. if $\sum_{\pi'\in\sigma_l}\prod_{i=1}^n \widehat{f}_{\mathbf{B}_\mathbf{i}^*|\mathbf{Z}}(B_{\pi'(i)l}|Z_l) = 0$.

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

$$\widehat{f}_{\mathbf{X_i},\mathbf{Z}}(x,z) = \frac{1}{Lh^{d+1}_{f_{\mathbf{X_i},\mathbf{Z}}}} \sum_{l=1}^{L} \sum_{p=1}^{n} \widehat{P}_{ipl} \cdot K_{f_{\mathbf{X_i},\mathbf{Z}}}\left(\frac{x - \widehat{X}_{ipl}}{h_{f_{\mathbf{X_i},\mathbf{Z}}}}, \frac{z - Z_l}{h_{f_{\mathbf{X_i},\mathbf{Z}}}}\right). \tag{19}$$

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 $f_{\mathbf{B_j^*}|\mathbf{Z}}(B_{pl}|Z_l)$ and $F_{\mathbf{B_j^*}|\mathbf{Z}}(B_{pl}|Z_l)$. Instead of being directly estimated in a similar way as in our first step, anonymity requires two additional steps: our second step is a linear reparametrization for which we has thus no reason to be worried about, our third step is a nonlinear reparametrization which is ill-conditioned at the limit where some bidders are symmetric (Appendix A.1). The fourth step consists as in GPV in the construction of the set of pseudo private values: n 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 $\widehat{f}_{\mathbf{X_i},\mathbf{Z}}(x,z) - f_{\mathbf{X_i},\mathbf{Z}}(x,z)$ into three terms.

$$\widehat{f}_{\mathbf{X_i},\mathbf{Z}}(x,z) - f_{\mathbf{X_i},\mathbf{Z}}(x,z) = \varepsilon_1 + \varepsilon_2 + \varepsilon_3, \text{ where} \tag{20}$$

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

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

$$\widetilde{f}_{\mathbf{X_i},\mathbf{Z}}(x,z) = \frac{1}{Lh^{d+1}_{f_{\mathbf{X_i},\mathbf{Z}}}} \sum_{l=1}^{L} \sum_{p=1}^{n} P_{ipl} \cdot K_{f_{\mathbf{X_i},\mathbf{Z}}}(\frac{x - X_{ipl}}{h_{f_{\mathbf{X_i},\mathbf{Z}}}}, \frac{z - Z_l}{h_{f_{\mathbf{X_i},\mathbf{Z}}}}). \tag{21}$$

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 $f_{\mathbf{X_i},\mathbf{Z}}(.,.)$ have R bounded continuous derivatives, the optimal uniform convergence rate for estimating $f_{\mathbf{X_i},\mathbf{Z}}(.,.)$ is $(\frac{L}{\log L})^{R/(2R+d+1)}$ (see Stone (1982)). 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 and 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 actually 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 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 while the

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' estimation procedure (Assumption 6).

formalization is in section 6.

4.2 Estimation under general asymmetry structures

In the general case, only the third part of the estimation procedure has to be modified and more precisely the estimators of the bids' CDFs in (14). To this end we first introduce a generalization of the function Υ for any asymmetry structure $(d_1, \cdots, d_r)$ which is denoted by $\Upsilon_{(d_1,\cdots,d_r)} : [0,1]^n \to \mathcal{C}^n$ and defined in the following way.

For $k \leq l$, we define $H_k^l : \mathcal{C}^l \to \mathcal{C}^k$ a function that maps to any vector of complex numbers $Y = (y_1, \dots, y_l)$ a vector that consists of k elements of Y such that there is no other subset such that the maximal distance between two elements is strictly smaller. Formally, $H_k^l(Y) = (y_{i_1}, \dots, y_{i_k})$ with $i_l \neq i_s$ for any $l \neq s$ and there is no vector $(y_{j_1}, \dots, y_{j_k})$ with $j_l \neq j_s$ for any $l \neq s$ such that $\max_{l,s\in\{1,\dots,k\}} |y_{j_l} - y_{j_s}| < \max_{l,s\in\{1,\dots,k\}} |y_{i_l} - y_{i_s}|$.

For any vector $Y = (y_1, \dots, y_n)$ and any asymmetry structure $(d_1, \dots, d_r)$, we define the sets $Y_i = (y_1^i, \dots, y_{d_i}^i)$, $i = 1, \dots, r$ by induction in the following way:

$$
\begin{array}{lll}
Y_1 & = H_{d_1}^n(Y) & \text{Initialization Stage} \\
Y_{i+1} & = H_{d_{i+1}}^{n-\sum_{k=1}^i d_k}(Y \setminus \bigcup_{k=1}^i Y_k), i = 1, \dots, n-1 & \text{Induction Stage.}
\end{array}
\tag{22}
$$

Then we consider a reordering denoted by $\{\overline{Y}_i\}_{i=1,\dots,r}$ of the sets $\{Y_i\}_{i=1,\dots,r}$ characterized by $\overline{Y}_i = Y_{\pi(i)}$ for any $i = 1, \dots, r$ where $\pi \in \Sigma_r$ satisfies $d_i = d_{\pi(i)}$ for any $i = 1, \dots, r$ and $\sum_{k=1}^{d_i} y_k^{\pi(i)} \geq \sum_{k=1}^{d_j} y_k^{\pi(j)}$ if $d_i = d_j$ and $i < j$.[21] Finally we define $(a_{d_i-1}^i, \dots, a_0^i)$ the vector of the 'normalized' coefficients of the monic polynomial with the vector of roots $\overline{Y}_i$: $Q(u) = u^{d_i} + \sum_{s=0}^{d_i-1} a_s^i \cdot \frac{n!}{(n-s)!s!} \cdot (-1)^{n-s} u^s$ with $Q(u) = \prod_{s=1}^{d_i} (u - y_s^{\pi(i)})$.

We now have all the ingredients to define $\Upsilon_{(d_1,\cdots,d_r)}$ for any vector $(a_{n-1}, \dots, a_0) \in$

[21] The ranking among the complex numbers is according to the lexicographic order.

$[0,1]^n$ as the vector $(a^1_{d_1-1},\dots,a^1_0,\dots,a^r_{d_r-1},\dots,a^r_0)$ where the vector Y is chosen such that $Y=\Upsilon(a_{n-1},\dots,a_0)$. In particular, we have:

$$u^n+\sum_{i=0}^{n-1}a_i\cdot\frac{n!}{(n-i)!i!}\cdot(-1)^{n-i}u^i=\prod_{k=1}^{r}\left(u^{d_k}+\sum_{s=0}^{d_k-1}a^k_s\cdot\frac{n!}{(n-s)!s!}\cdot(-1)^{n-s}u^s\right).$$

Then, for any asymmetry structure $(d_1,\cdots,d_r)$, let $\Lambda_{(d_1,\cdots,d_r)}:\mathcal{C}^n\to\mathcal{C}^n$ be the function that maps to any vector $(u^1_{d_1-1},\dots,u^1_0,\dots,u^1_{d_r-1},\dots,u^r_0)$ the vector $(\underbrace{u^1_{d_1-1},\dots,u^1_{d_1-1}}_{d_1-times},\dots,\underbrace{u^r_{d_r-1},\dots,u^r_{d_r-1}}_{d_r-times})$.

Finally, with respect to our estimation procedure under full asymmetry, we have only to replace equation (14) by

$$(\widehat{F}_{\mathbf{B^*_1}|\mathbf{Z}}(b|z),\cdots,\widehat{F}_{\mathbf{B^*_n}|\mathbf{Z}}(b|z))=\mathcal{R}[\Lambda_{(d_1,\cdots,d_r)}\left(\Upsilon_{(d_1,\cdots,d_r)}(\widehat{a}_{n-1}(b|z),\cdots,\widehat{a}_0(b|z))\right)],\qquad(23)$$

where $\widehat{a}_i(b|z)=\widehat{F}^{(n-i:n-i)}_{\mathbf{B}|\mathbf{Z}}(b|z)$.

Remark 4.2 In the case of full asymmetry, we have $\Upsilon=\Upsilon_{(1,\dots,1)}$ and $\Lambda_{(d_1,\cdots,d_r)}$ is the identity function such that the estimation procedure corresponds exactly to the one in the previous subsection. In the case of full symmetry, then the estimation procedure corresponds exactly to GPV: $\widehat{F}_{\mathbf{B^*_1}|\mathbf{Z}}=\cdots=\widehat{F}_{\mathbf{B^*_n}|\mathbf{Z}}=\widehat{F}^{(1:1)}_{\mathbf{B}|\mathbf{Z}}$ and then equation (16) reduces to $\widehat{f}_{\mathbf{B^*_1}|\mathbf{Z}}=\cdots=\widehat{f}_{\mathbf{B^*_n}|\mathbf{Z}}=\widehat{f}^{(1:1)}_{\mathbf{B}|\mathbf{Z}}$.

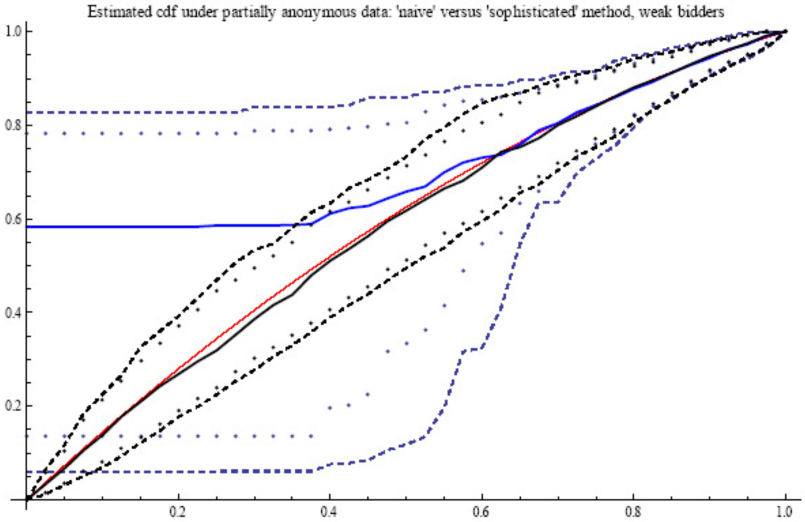
5 Monte Carlo Experiments

To illustrate the usefulness of our procedure, we conduct a limited Monte Carlo study.[22] To fit with realistic sizes of auction data, we consider $L=40$ auctions, each having 6 bidders: 3 of which belonging to a set of strong bidders, while the 3 remaining bidders to a set of weak bidders. Our Monte Carlo experiments consist of 200 replications of our estimation procedure for the second price auction and mainly under the knowledge of the identity of the winner (namely weak or strong), which is labeled as 'partially anonymous' data.

[22]Practical details on the implementation are reported in the Supp. Mat. Programs are written in Mathematica and are available upon request.

Figure 2: asymmetric IPV model

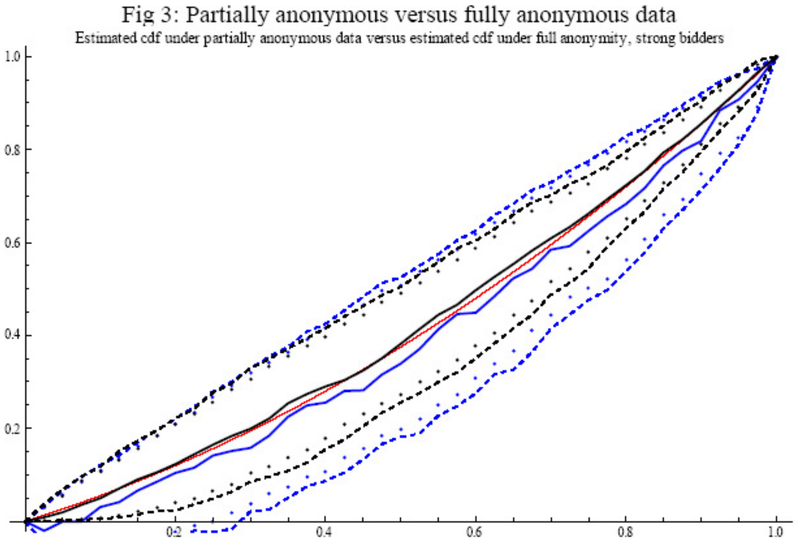
Estimated cdf under partially anonymous data: 'naive' versus 'sophisticated' method, weak bidders



Asymmetric IPV model 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 where the distribution of private values $F_{\mathbf{X}}$ is generated from the densities f_ϵ on the support $[0, 1]$ where $f_\epsilon(x) = (1 + \epsilon \cdot (1 - 2x)) \cdot \mathbf{1}_{0 \leq x \leq 1}$ and where we take $\epsilon = -\frac{1}{2}$ and $\epsilon = \frac{1}{2}$ for respectively the 3 strong and the 3 weak bidders. 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 (2002) for identification where results from the competing risk literature are applied and Paarsch and Hong (2006) p.153-155 for natural estimators that are asymptotically consistent.[23] 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 'sophisticated' estimation procedure does a good job: the median of the estimates perfectly matches the true curve and the 80%

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

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.



With regards to our 'sophisticated' estimation procedure, Figure 3 reports the analogs of Figure 2 (still displayed in black) for the same underlying model but with respect to the strong bidders. The blue lines now report the corresponding results for the preliminary estimator at the end of the third step, i.e. if the three remaining steps that are intended to use the knowledge about the identity of the winner were not implemented and which corresponds roughly to our estimator in the case of fully anonymity. The differences are important and illustrate how additional information on bidders' identities reduces the bias and the variance of our estimator in small samples.

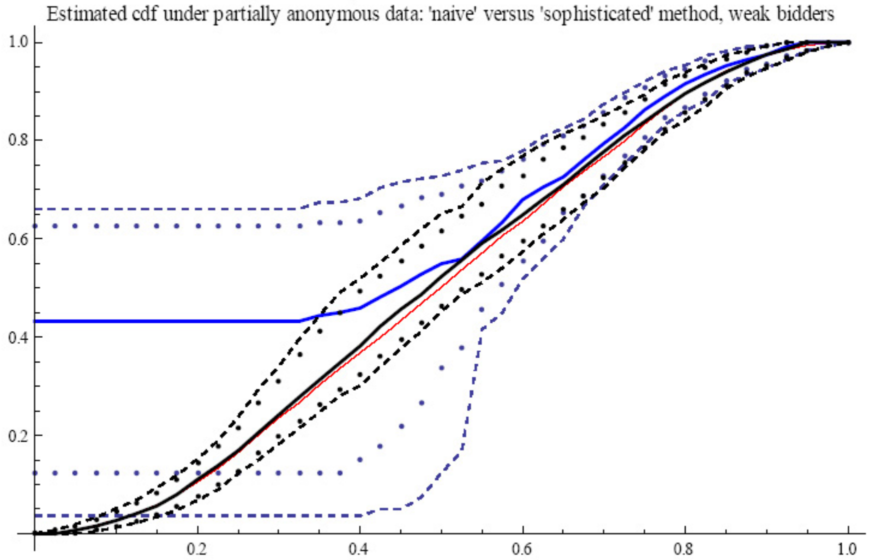
Symmetric APV model The simulations reported in Figure 4 are devoted to a kind of robustness check. Our 'sophisticated' estimation approach and the 'naive' approach are both relying on the independence assumption. We consider a departure from this assumption: the underlying (true) model is a symmetric correlated PV model with 6 bidders.[24] The legend is the same as for Figure 2. The results in Figure 4 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

[24]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.

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 (14), our procedure (at least partially) drops the use of the independence assumption when we estimate complex roots as it happens with positive correlation.[25]

Figure 4: symmetric APV model

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



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.[26] 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 works.

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

[26] 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 (1973) and Mosier (1986) for more on this topic.

6 Asymptotic Properties

Throughout we denote by $S(*)$ and $S^o(*)$ the support of $*$ and its interior, respectively.

6.1 Regularity Assumptions and Key Properties

The two next assumptions concern the underlying generating process as well as the smoothness of the latent joint distribution of (X_{il}, Z_l) for any $i = 1, \ldots, n$. The precise mathematical formulations are relegated in the Supp. Mat. since those assumptions can be found already in GPV (respectively as A1 and A2).

Assumption A 3 *An i.i.d underlying generating process.*

Assumption A 4 *Smoothness of the latent joint distributions which are bounded away from zero on their support* $[\underline{z}, \overline{z}] \times [\underline{x}, \overline{x}]$.[27]

A4 implies in particular that $f_{\mathbf{X_i}|\mathbf{Z}}(.|.)$ admits up to R bounded continuous partial derivatives on $S(F_{\mathbf{X_i}|\mathbf{Z}})$. 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. The analysis of the smoothness of the CDFs and PDFs $F_{\mathbf{B_p}|\mathbf{Z}}(.|.)$ and $f_{\mathbf{B_p}|\mathbf{Z}}(.|.)$ for $p = 1, \cdots, n$ is gathered in Proposition A.3 in the Appendix. The results are roughly the same. In particular the bid densities in the first price auction are smoother than for the second price auction. Thus $f_{\mathbf{B_p}|\mathbf{Z}}(.|.)$ 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 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. Nevertheless, proposition A.3 differs slightly from the one appearing

[27] There is actually a mild difference between A4 and its analog in GPV: as in Campo et al. (2002), we consider that the support of buyers' private values does not depend on Z to simplify the presentation. It implies that the lower bound of the support of buyers' bids does not depend on Z. The general case can be fully treated following GPV.

in GPV as irregularities of the CDFs of the order statistic may appear in the interior of their support. Let $\overline{b}(z;p)$ denote the upper bound of the support of the CDF $F_{\mathbf{B_p}|\mathbf{Z}}(.|z)$. If $\overline{b}(z;p) < \overline{b}(z;n)$ which may occur for $p \leq n-2$, then irregularities will occur at $\overline{b}(z;p)$.[28] In the following, to alleviate the presentation, we will make the simplifying assumption A5(ii) that the bidding supports of all bidders coincide, i.e. $\beta_i(\overline{x};z)$ does not depend on i which implies that $\overline{b}(z;p)$ does not depend on p. Without it, our uniform consistency results would extend provided that the neighborhoods of the bidders' signals that make them bid $\overline{b}(z;p)$, $p=1,\ldots,n$, 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 could trim those inner neighborhoods.

The first part of the next assumption requires some preliminary discussions. First, the natural extension with covariates of our definition of identifiability consists in replacing $F_{\mathbf{X}}(.)$ and $F_{\mathbf{B}}(.)$ by $F_{\mathbf{X}|\mathbf{Z}}(.|.)$ and $F_{\mathbf{B}|\mathbf{Z}}(.|.)$. The presence of covariates is not innocuous for identification since the asymmetric IPV model is not identified with coraviates without further assumptions. This issue is easily illustrated in presence of a binary covariate $z = z_1$ or z_2. Consider two bidders 1 and 2 and two distinct CDFs G_1 and G_2. The two following generating process A and B are observationally equivalent: first, $F^A_{\mathbf{X}|\mathbf{Z}}(x_1,x_2|z_1) = F^A_{\mathbf{X}|\mathbf{Z}}(x_1,x_2|z_2) = G_1(x_1)\cdot G_2(x_2)$ and, second, $F^B_{\mathbf{X}|\mathbf{Z}}(x_1,x_2|z_1) = G_1(x_1)\cdot G_2(x_2)$ and $F^B_{\mathbf{X}|\mathbf{Z}}(x_1,x_2|z_2) = G_1(x_2)\cdot G_2(x_1)$. Nevertheless, $F^A_{\mathbf{X}|\mathbf{Z}}(.|.)$ and $F^B_{\mathbf{X}|\mathbf{Z}}(.|.)$ are not equal up to a permutation of the buyers. Since our model considers continuous covariates and distributions that are smooth with regards to the covariates, then the above example only gives an intuition why heterogeneity requires an additional structure to identify the model. Another way to figure this identification issue is that similar intersections as the one illustrated in Figure 1 when b varies may arise with respect to the variable capturing heterogeneity Z when this variable varies. Here to guarantee identification[29], we make the assumption A5(i) that bidders' CDFs can be ordered according to first order stochastic dominance such

[28] At $\overline{b}(z;p)$ such that $\overline{b}(z;p) < \overline{b}(z;n)$, $f_{\mathbf{B_p}|\mathbf{Z}}(.|z)$ drops discontinuously from a positive constant to zero while the PDFs $(f_{\mathbf{B_k}|\mathbf{Z}}(.|z))_{k>p}$ are also discontinuous. This is due to the discontinuity of the PDFs $f_{\mathbf{X_i}|\mathbf{Z}}(.|z)$ respectively at the points $\beta_i(\overline{x};z)$.

[29] Assumption A5(i) is not necessary for identification. 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. Assuming distinct bidding supports may also be of some help.

that the pointwise solutions of the system (4)[30] for any b and z lead to a unique solution for the CDFs $F_{\mathbf{B_i^*}|\mathbf{Z}}(.|.)$, $i = 1, \ldots, n$. This assumption avoids one caveat: in the presence of crossing points, the set of roots given by the system (4) would not fully characterize the solution and the right solution candidate should be selected, e.g. from the empirical counterpart of some additional restrictions as the ones coming from the system (6).

Assumption A 5 (i) ***Strict Stochastic Dominance***[31]*: For any pair of bidders i, j with $j > i$, the bid densities $F_{\mathbf{B_k^*}|\mathbf{Z}}(.|.)$, $k = i, j$, are either equal or can be strictly ordered according to first order stochastic dominance: we have either $F_{\mathbf{B_i^*}|\mathbf{Z}}(.|z) = F_{\mathbf{B_j^*}|\mathbf{Z}}(.|z)$ for any z or $F_{\mathbf{B_i^*}|\mathbf{Z}}(b|z) > F_{\mathbf{B_j^*}|\mathbf{Z}}(b|z)$ for any $b \in S^0(f_{\mathbf{B_i^*}|\mathbf{Z}})$ and any z.*

(ii) ***Common bidding support****: All bidders have the same bidding support: $\overline{b}(z) \equiv \overline{b}(z;p)$, which does not depend on p.*

Moreover, to simplify our estimation procedure, we have also assumed that the dominance is strict in the interior of the bidding support. This avoids an additional caveat: at points where strictly distinct CDFs are equal, the rate of convergence derived in propositions 6.1 and 6.2 will break down. In particular, our proof fails since the function $\Upsilon_{(d_1,\ldots,d_r)}$ is not differentiable at any point $(a_{n-1}, \ldots, a_1)$ such that the polynomial in (5) has strictly less than r distinct roots. In such cases, our statistical results remain nevertheless true on any inner compact subset of the support of the bidding distribution that contains no crossing point.[32]

6.2 Uniform Consistency

Our main result establishes the uniform consistency of our multi-step kernel-based estimators for the first and second price auctions and with their rates of convergence. As a preliminary step, we first complete the definition of our estimators by choosing

[30] More precisely the straightforward generalization of (4) with covariates.

[31] Assumption A5 is not on the primitives of the model in the first price auction. Under a set of assumptions that is guaranteed under A2-A3, Lebrun (1999) (Corollary 3) show that 'conditional stochastic dominance' of private values' distributions (a restriction that has been first introduced by Maskin and Riley (2000) for two classes of bidders) is a sufficient condition for first order strict stochastic dominance of bidding distributions.

[32] Under the mild restriction that the CDFs of non-symmetric bidders are strictly distinct, then lemma 2.1 in the Supp. Mat. guarantees that such a compact subset can be chosen such that the probability that all bids belong to this set is arbitrary close to one.

optimally the kernels and the bandwidths and then establish in proposition 6.1 the uniform consistency with their rates of convergence of our nonparametric estimators of the upper and lower boundaries $\overline{b}(z)$ 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 *An optimal choice of high-order kernels and bandwidths.*

The complete formulation of A6 is relegated in the Supp. Mat. since those choices correspond to the ones proposed by GPV [Assumptions 3 & 4] for the first price auction and by Härdle (1990) for the second price auction. In particular, we set $h_{f_{\mathbf{B_p|Z}}} = \lambda_{f_{\mathbf{B_p|Z}}}(\frac{\log L}{L})^{\frac{1}{(2R+d+3)}}$ in the second price auction, i.e. the standard optimal bandwidth such that the estimator $\widehat{f}_{\mathbf{B_p,Z}}(.,.)$ converges uniformly at the best possible rate.

Proposition 6.1 *Under A2-A6, for any closed subset $\mathcal{C}$ of $S^o(F_{\mathbf{X,Z}})$, we have almost surely $\sup_{z\in[\underline{z},\overline{z}]}|\widehat{\overline{b}}(z) - \overline{b}(z)| = O(\frac{\log L}{L})^{\frac{1}{d+1}}$ 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:*

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

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

in the first price auction and

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

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

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 subsets of the support to avoid boundary effects.

Proposition 6.2 *Suppose that A2-A6 hold, then* $(\widehat{f}_{\mathbf{X_1|Z}}(.|.),\cdots,\widehat{f}_{\mathbf{X_n|Z}}(.|.))$ *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* $(f_{\mathbf{X_1|Z}}(.|.),\cdots,f_{\mathbf{X_n|Z}}(.|.))$ *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 section 7 that our estimation procedure of the conditional density $F_{\mathbf{X|Z}}(.|.)$ 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 6.2 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 asymmetric bidders. Note that if the interest of the econometrician lies in the estimation of the distributions $F_{\mathbf{B_i^*|Z}}(.|.)$, $i=1,\ldots,n$, 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 6.2 as it helps to identify the additional points relative to GPV's procedure and why the asymptotic rates of convergence remain unchanged.

Proof We have $\widehat{f}_{\mathbf{X_i|Z}}(x|z)=\widehat{f}_{\mathbf{X_i,Z}}(x,z)/\widehat{f}_{\mathbf{Z}}(z)$. Given the optimal bandwidth choice $h_{f_{\mathbf{Z}}}=\lambda_{f_{\mathbf{Z}}}(\frac{\log L}{L})^{\frac{1}{(2R+d+2)}}$ in assumption A6, we know that $\widehat{f}_{\mathbf{Z}}(z)$ converges uniformly to $f_{\mathbf{Z}}(z)$ 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 $f_{\mathbf{Z}}(z)$ is bounded away from 0 by assumption A4, it suffices to show that $\widehat{f}_{\mathbf{X_i,Z}}(x,z)$ converges at the rate $(\frac{L}{\log L})^{R/(2R+d+3)}$ and $(\frac{L}{\log L})^{R/(2R+d+1)}$ in the first and second price auctions respectively. We turn back to the way we have decomposed the difference $\widehat{f}_{\mathbf{X_i,Z}}(x,z)-f_{\mathbf{X_i,Z}}(x|z)$ in equation (20) and analyze the convergence rate of the three error terms.

In the second price auction, the bandwidth $h_{\mathbf{f_{X_i,Z}}}=\lambda_{\mathbf{f_{X_i,Z}}}(\frac{\log L}{L})^{\frac{1}{(2R+d+3)}}$ leads to a uniform convergence of $\widetilde{f}_{\mathbf{X_i,Z}}(x,z)$ to $f_{\mathbf{X_i,Z}}(x,z)$ 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 $h_{\mathbf{f_{X_i,Z}}}=\lambda_{\mathbf{f_{X_i,Z}}}(\frac{\log L}{L})^{\frac{1}{(2R+d+1)}}$ 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

$$\frac{1}{Lh_{f_{\mathbf{X_i},\mathbf{Z}}}^{d+1}} \sum_{l=1}^{L} \sum_{p=1}^{n} (\widehat{P}_{ipl} - P_{ipl}) \cdot K_{f_{\mathbf{X_i},\mathbf{Z}}}(\frac{x - X_{ipl}}{h_{f_{\mathbf{X_i},\mathbf{Z}}}}, \frac{z - Z_l}{h_{f_{\mathbf{X_i},\mathbf{Z}}}}),$$

which is bounded by:

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

The above term appearing in the bracket may be viewed as a kernel estimator, and hence converges uniformly on $\mathcal{C}$ to $\sum_{j=1}^{n} f_{\mathbf{X_j},\mathbf{Z}}(x,z) \cdot \int |K_{f_{\mathbf{X_i},\mathbf{Z}}}(x,z)|dxdz$. Thus this term stays bounded almost surely. Finally from proposition 6.1, we have $\widehat{f}_{\mathbf{X_i},\mathbf{Z}}(x,z) - f_{\mathbf{X_i},\mathbf{Z}}(x,z) = O(\log L/L)^{R/(2R+d+1)}$.

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. **Q.E.D.**

7 Optimal Uniform Convergence Rate

In this section, we adopt a minmax approach to obtain bounds for the rate at which the latent density of private values can be estimated uniformly from observed bids. The next proposition gives an upper bound for the optimal uniform convergence rate for estimating $f_{\mathbf{X}|\mathbf{Z}}(.|.)$ from observed (anonymous) bids. GPV derives the same bound for the symmetric IPV model and nonanonymous bids. Here we extend their result to the asymmetric IPV model. In the following, for a given density function f, denote by $||f||_r$ (resp. $||f||_{r,\mathcal{C}}$) the maximum of f and all its derivatives up to the r^{th} order on $S(F)$ (resp. on $\mathcal{C}$).

Proposition 7.1 *Assume A3-A4 and* $||f^o_{\mathbf{X},\mathbf{Z}}(x,z)||_R < M$. *Let* $\mathcal{C}(X)$ *be an inner compact subset of* $S(f^o_{\mathbf{X}|\mathbf{Z}})$ *with nonempty interior. There exists a constant* $\kappa > 0$

such that

$$\lim_{\epsilon\to 0}\underline{\lim}_{L\to+\infty}\inf_{\widehat{f}_L}\sup_{f\in U_\epsilon(f^o_{\mathbf{X}|\mathbf{Z}})} Prob_f\left[\left(\frac{L}{\log L}\right)^{\frac{R}{(2R+d+3)}}\sup_{(x,z)\in\mathcal{C}(X)}||\widehat{f}_{\mathbf{X}|\mathbf{Z}}(x|z)-f_{\mathbf{X}|\mathbf{Z}}(x|z)||_0>\kappa\right]>0$$

in the first price auction, and

$$\lim_{\epsilon\to 0}\underline{\lim}_{L\to+\infty}\inf_{\widehat{f}_L}\sup_{f\in U_\epsilon(f^o_{\mathbf{X}|\mathbf{Z}})} Prob_f\left[\left(\frac{L}{\log L}\right)^{\frac{R}{(2R+d+1)}}\sup_{(x,z)\in\mathcal{C}(X)}||\widehat{f}_{\mathbf{X}|\mathbf{Z}}(x|z)-f_{\mathbf{X}|\mathbf{Z}}(x|z)||_0>\kappa\right]>0$$

in the second price auction, where the infimums are taken over all possible estimators $\widehat{f}_L$ *of* $f_{\mathbf{X}|\mathbf{Z}}(.|.)$ *based upon* (B_{pl}, Z_l) *for any* $p = 1, \cdots, n$ *and* $l = 1, \cdots, L$ *and where* $U_\epsilon(f^o_{\mathbf{X}|\mathbf{Z}})$ *is a neighborhood of* $f^o_{\mathbf{X}|\mathbf{Z}}$ *defined as*

$$U_\epsilon(f^o_{\mathbf{X}|\mathbf{Z}}) \equiv \left\{ f;\ \sup_{(x,z)\in S(F^o_{\mathbf{X}|\mathbf{Z}})} ||f(x,z)-f^o_{\mathbf{X}|\mathbf{Z}}(x,z)||_0<\epsilon, ||f(.,.)||_R<M \right\},$$

where $M > 0$.

The set of possible estimators based upon anonymous bids is tautologically smaller than those based upon (B^*_{il}, Z_l) for any $i \in [1, n]$ and $l = 1, \cdots, L$. Thus it is sufficient to prove the above proposition with this richer set of estimators. In this latter case, for the second price auction where observed bids correspond exactly to private values, the above result follows from Khas'minskii (1978). In the first price auction, the above proposition has been proved in the symmetric IPV model by GPV who adapts Khas'minskii's (1978) arguments. It seems intuitive that a faster local rate of uniform convergence is not available in the general case with asymmetric bidders. Nevertheless, due to the local nature of such results, the argument is not tautologic. Indeed, since a general asymmetric model with n bidders involves n overlapped differential equations for bidders' distributions, the asymmetric structure may 'smooth' the link between observables and the latent private values.

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. (2002) (see also

Guerre et al. (2009)). 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.[33] 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 $(F_{\mathbf{B}_\mathbf{i}^*}(x))_{i=1,\ldots,n}$ as a function of the vector $(F_\mathbf{B}^{(r:r)}(x))_{r=1,\ldots,n}$, 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 and estimation using alternatives routes that are exploiting the full joint distribution of the order-statistics $F_\mathbf{B}(.)$.

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 restrictions 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) allow 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 di-

[33] With incomplete sets of bids, assuming independence seems the 'natural' identification route for nonparametric approaches. E.g. Theorem 4 of Athey and Haile (2002) shows that the symmetric affiliated value model is not identified.

mension 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. 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 or under the complex disclosure rules of French timber auctions in the first price auction, identification is typically not an issue. Nevertheless, the ideas underlying our methodology could be also useful: adaptations of our methodology could be useful to exploit the full set of observed bids.

Our approach can also be used for alternative asymmetric auction models with independent private signals as models with one informed bidder against a set of non-informed bidders (Engelbrecht-Wiggans et al. (1983), Hernando-Veciana and Tröge (2010)), models with collusion through a ring (Marshall and Marx (2007)) or finally the model developed by Landsberger et al. (2001) where the ranking of bidders' private valuations is common knowledge among bidders (but possibly not to the econometrician). The ideas sustaining our methodology could also be useful more generally beyond auction environments for applications as imperfect matching between data sets, possible new anonymous designs in experimental economics or the design of surveys for sensitive attributes.

References

[1] J. Asker. A study of the internal organisation of a bidding cartel. *American Economic Review*, forthcoming, 2010.

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

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

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

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

[6] 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.

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

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

[9] R. Cassady. *Auctions and Auctioneering*. University of California Press, 1967.

[10] L. I. de Castro and H. J. Paarsch. Testing affiliation in private-values models of first-price auctions using grid distributions. Cei working paper series, Center for Economic Institutions, Institute of Economic Research, Hitotsubashi University, 2009.

[11] R. Engelbrecht-Wiggans, P. Milgrom, and R. J. Weber. Competitive bidding and proprietary information. *J. Math. Econ.*, 11(2):161–169, 1983.

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

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

[14] V. Ginsburgh. Absentee bidders and the declining price anomaly in wine auctions. *Journal of Political Economy*, 106(6):1302–1331, 1998.

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

[16] E. Guerre, I. Perrigne, and Q. Vuong. Nonparametric identification of risk aversion in first price auctions under exclusion restrictions. *Econometrica*, 77(4):1193–1228, 2009.

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

[18] A. Hernando-Veciana and M. Tröge. The insider's curse. *Games Econ. Behav.*, forthcoming, 2010.

[19] W. Härdle. *Applied Nonparametric Regression*. Econometric Society Monograph, 1990.

[20] R. Khas'minskii. A lower bound on the risks of nonparametric estimates of densities. *Theory of Probability and its Applications*, 23:794–798, 1978.

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

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

[23] L. Lamy. The econometrics of auctions with asymmetric anonymous bidders. Pse working papers, 2008.

[24] L. Lamy. The shill bidding effect versus the linkage principle. *Journal of Economic Theory*, 144:390–413, 2009.

[25] L. Lamy. 'upping the ante': how to design efficient auctions with entry? Pse working papers, 2010.

[26] 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.

[27] B. Lebrun. First price auctions in the asymmetric n bidder case. *International Economic Review*, 40(1):125–42, 1999.

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

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

[30] T. Li, I. Perrigne, and Q. Vuong. Conditionally independent private information in ocs wildcat auctions. *J. Econometrics*, 98:129–161, 2000.

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

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

[33] R. Marshall and L. Marx. Bidder collusion. *J. Econ. Theory*, 133:374–402, 2007.

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

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

[36] P. Milgrom and R. Weber. A theory of auctions and competitive bidding. *Econometrica*, 50:1089–1122, 1982.

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

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

[39] A. Ockenfels, D. Reiley, and A. Sadrieh. *in: Terrence Hendershott (Ed.)*, chapter 12 Online Auctions, pages 571–628. Handbook of Economics and Information Systems. Elsevier Science, 2006.

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

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

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

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

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

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

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

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

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

A Appendix

A.1 Some useful algebraic results

Let J_ω denote the following matrix:

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

Lemma A.1 *The matrix J_ω is invertible if and only if $\omega_i \neq \omega_j$ for any $i \neq j$.*

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

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

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 $\sum_{j_1,\cdots,j_r,j_k \neq j_{k'}} \prod_{j_k \in \{j_1,\cdots,j_r\}} \omega_{j_k}$ (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 $S_{k-1} \times \mathbf{1} = J_k + k \cdot I_\omega \times J_{k-1}$, for $k = 1, \cdots, n+1$ (with the convention that J_0 and J_{n+1} are both the null vector), we easily derive a kind of Newton-Girard formula for any $1 \leq k \leq n$:

$$J_k = \sum_{i=1}^{k} (-1)^{i+1} \frac{k!}{(k+1-i)!} S_{k-i} \times V_i. \tag{24}$$

From matrix V_ω, if we successively replace the column k (from $k = n$ to $k = 1$) by the column $\sum_{i=1}^{k} (-1)^{i+1} \frac{k!}{(k+1-i)!} S_{k-i} \times V_i$, the determinant is preserved at each step whereas equation (24) guarantees that the final matrix is J_ω. We conclude after noting that 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 Basu et al. (2006) p. 104-105). **Q.E.D.**

Let $J_\omega^{r \times r} := ([J_\omega]_{i,j})_{\substack{1 \leq i \leq r \\ 1 \leq j \leq r}}$ denote the submatrix of J_ω formed by selecting the r first rows and columns. Similarly, let $V_\omega^{r \times r} := ([V_\omega]_{i,j})_{\substack{1 \leq i \leq r \\ 1 \leq j \leq r}}$ and $V_\omega^{r \times r} = [V_1^{r \times r}, \cdots, V_r^{r \times r}]$.

Lemma A.2 *The matrix $J_\omega^{r \times r}$ is invertible if and only if $\omega_i \neq \omega_j$ for any $i \neq j$ with $i, j \in [1, r]$.*

Proof The proof is similar as the one for lemma A.1 which corresponds actually to the special case with $r = n$. From matrix $V_\omega^{r \times r}$, if we successively replace the column k (from $k = r$ to $k = 1$) by the column $\sum_{i=1}^{k} (-1)^{i+1} \frac{k!}{(k+1-i)!} S_{k-i} \times V_i^{r \times r}$, the determinant is preserved at each step whereas equation (24) guarantees that the final matrix is $J_\omega^{r\times r}$. We conclude after noting that $V_\omega^{r\times r}$ is equal to the Vandermonde matrix with respect to the vector $(\omega_1, \cdots, \omega_r)$. **Q.E.D.**

Definition 5 *Let $P = \sum_{i=0}^{p} a_i X^i$ and $Q = \sum_{i=0}^{q} b_i X^i$ be two real polynomials. The Sylvester Matrix of P and Q is the $(p+q)\times(p+q)$ matrix defined by:*

$$Syl(P,Q) = \begin{bmatrix} b_q & \cdots & \cdots & \cdots & b_0 & 0 & \cdots & \cdots & 0 \\ 0 & \ddots & & & & \ddots & \ddots & & \vdots \\ \vdots & \ddots & \ddots & & & & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & & & & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & b_q & \cdots & \cdots & \cdots & b_0 \\ a_p & \cdots & \cdots & \cdots & \cdots & a_0 & 0 & \cdots & 0 \\ 0 & \ddots & & & & & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & & & & & \ddots & 0 \\ 0 & \cdots & 0 & a_p & \cdots & \cdots & \cdots & \cdots & a_0 \end{bmatrix},$$

Proposition A.1 (Basu et al. (2006)) *The determinant of the Sylvester Matrix is given by:*

$$det(Syl(P,Q)) = a_p^q b_q^p \prod_{i=1}^{p}\prod_{j=1}^{q}(x_i - y_j),$$

where $(x_i)_{i=1,\ldots,p}$ and $(y_i)_{i=1,\ldots,p}$ are the vector of roots counted with their order of multiplicity of respectively P and Q, i.e. $P(u) = \prod_{i=1}^{p}(u - x_i)$ and $Q(u) = \prod_{i=1}^{q}(u - y_i)$. As a corollary, the Sylvester Matrix is invertible if and only if P and Q are coprime.

Let $\mathcal{M}_{d_1,\ldots,d_r} : \mathcal{R}^{d_1} \times \cdots \times \mathcal{R}^{d_r} \to \mathcal{R}^{\sum_{k=1}^{r} d_k}$ be the function that associates to the coefficients of r monic polynomials of respective degrees $d_1, \ldots, d_r$ the coefficients of the product polynomial. If $P_k = X^{d_k} + \sum_{i=0}^{d_k-1} a_i^k X^i$ for $k = 1, \ldots, r$, then $\mathcal{M}_{d_1,\ldots,d_r}(P_1, \ldots, P_r) = (a_{\sum_{k=1}^{r} d_k - 1}, \cdots, a_0)$ where $X^{\sum_{k=1}^{r} d_k} + \sum_{i=0}^{\sum_{k=1}^{r} d_k - 1} a_i X^i = \prod_{k=1}^{r} P_k$. With a slight abuse of notation, we use the notation $\mathcal{M}_{d_1,\ldots,d_r}(P_1, \ldots, P_r)$ for $\mathcal{M}_{d_1,\ldots,d_r}(a_{d_1-1}^1, \ldots, a_0^1, \ldots, a_{d_k-1}^k, \ldots, a_0^k, \ldots, a_{d_r-1}^r, \cdots, a_0^r)$.

Proposition A.2 *The Jacobian matrix of $\mathcal{M}_{d_1,\ldots,d_r}$ at $(P_1, \ldots, P_r)$ is invertible if the polynomials P_k, $k = 1, \ldots, r$, are coprime.*

Let $M_{(P_1,\ldots,P_r)}$ denote the corresponding Jacobian matrix.

Proof The proof is by induction on r.

Initialization Step ($r = 2$): we can check that the Jacobian matrix of $\mathcal{M}_{d_1,d_2}$ at $(a^1_{d_1-1}, \cdots, a^1_0, a^2_{d_2-1}, \cdots, a^2_0)$ is equal to the Sylvester Matrix $Syl(P_1, P_2)$. The result comes then from proposition A.1.

Induction Step: from the recursive relation

$$\mathcal{M}_{d_1,\dots,d_r}(P_1, \dots, P_r) = \mathcal{M}_{d_1,\sum_{k=2}^r d_k}(P_1, \mathcal{M}_{d_2,\dots,d_r}(P_2, \dots, P_r)),$$

we obtain a recursive relation on the Jacobians

$$M_{(P_1,\dots,P_r)} = M_{(P_1,\prod_{k=2}^r P_k)} \times \begin{pmatrix} 1 & & & \\ & \ddots & & \mathbf{0} \\ & & 1 & \\ & \mathbf{0} & & M_{(P_2,\dots,P_r)} \end{pmatrix}.$$

If P_k, $k = 1, \dots, r$, are coprime, then $M_{(P_1,\prod_{k=2}^r P_k)}$ and $M_{(P_2,\dots,P_r)}$ are invertible by the induction hypothesis. After noting that a block diagonal is invertible if each of its blocks are invertible, we obtain that $M_{(P_1,\dots,P_r)}$ is invertible. **Q.E.D.**

Remark A.1 The function $\mathcal{M}_{1,\dots,1}$ coincides with the restriction on $\mathcal{R}^n$ of Υ^{-1}. As a corollary to proposition A.2, we obtain that the Jacobian matrix of the function Υ is well-defined and of full rank at any point such that $\Upsilon(a_{n-1}, \dots, a_0) = (\omega_1, \dots, \omega_n) \in \mathcal{R}^n$ with $\omega_i \neq \omega_j$ for any $i \neq j$.

A.2 Proof of Proposition 3.1: additional elements

Under observability of bidders' identities and in the first price auction, Li et al. (2002) shows that the symmetric APV model is identified whereas Campo et al. (2003) extends this result to the asymmetric APV model. Let us see why Li et al.'s (2002) proof remains valid under anonymity whereas Campo et al.'s (2003) 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: $\sum_{j=1}^n F_{\mathbf{B}^*_{-j}|\mathbf{B}^*_j}(b'|b) \cdot \text{Prob}(\mathbf{B}^*_j = b | \exists l\ \mathbf{B}^*_l = b, \mathbf{B}^*_k \leq b' \text{ for } k \neq l)$, 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 $F_{\mathbf{B}^*_{-i}|\mathbf{B}^*_i}(b'|b) = \frac{1}{n} \cdot \sum_{j=1}^n F_{\mathbf{B}^*_{-j}|\mathbf{B}^*_j}(b'|b)$ 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. The extension of this proof for the first price auction is a bit technical and is then relegated to the Supp. Mat.

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.

Complementary elements for the third point

I. End of the proof in the full asymmetric case We end the identification proof of the asymmetric IPV model in the case where there is full asymmetry. Since the PDFs $f_{\mathbf{B^*_k}}$ are continuous and atomless, there exists a infinite number of bidding values b such that $f_{\mathbf{B^*_i}}(b) \neq f_{\mathbf{B^*_j}}(b)$ for any pair $i,j \in [1,\ldots,]$. Finally, there exists a (finite) family of distinct bidding values $\mathfrak{B} = (\widetilde{b}_{ij})_{i<j}$ in the interior of the bidding support such that $f_{\mathbf{B_i^*}}(\widetilde{b}_{ij}) \neq f_{\mathbf{B_j^*}}(\widetilde{b}_{ij})$ for any $i < j$. Let (b_{inf}, b_{sup}) denote an open subset of the interior of the bidding support that contains the family $\mathfrak{B}$. Let $\delta = \frac{1}{2} \cdot \min_{x,y \in \mathfrak{B} \cup b_{inf} \cup b_{sup}} |x - y|$. Our aim is then to build a subset $\mathfrak{I}^*$ such that the probabilities $\int_{\mathfrak{I}^*} f_{\mathbf{B_i^*}}(u)du$ are all distinct for $i \in [1, \ldots, n]$ which implies that $J_{(\int_{\mathfrak{I}^*} f_{\mathbf{B_1^*}}(u)du, \ldots, \int_{\mathfrak{I}^*} f_{\mathbf{B_n^*}}(u)du)}$ is invertible such that the same argument as the one developed from (7) holds.

In our following construction, we use assumption A1 that guarantees that the bidding PDFs $f_{\mathbf{B_i^*}}(.)$ are bounded away from zero and also bounded (see Proposition A.3) on $[b_{inf}, b_{sup}]$ through the following remark.

Initial remark. The functions $\epsilon \rightarrow \int_{[b-\epsilon, b+\epsilon]} f_{\mathbf{B_i^*}}(u)du$ are continuous differentiable and strictly increasing locally in the right neighborhood of 0, for any $b \in [b_{inf}, b_{sup}]$ and any $i \in [1, \ldots, n]$. At $\epsilon = 0$, its value is zero and the value of the derivative is $2 \cdot f_{\mathbf{B_i^*}}(b)$.

Let run the following iterative procedure.

Initialization step. Pick one element $\widetilde{b}_{ij} \in \mathfrak{B}$. Let $\mathfrak{B}^1 = \mathfrak{B} \setminus \widetilde{b}_{ij}$. With regards to the realized values $f_{\mathbf{B_k^*}}(\widetilde{b}_{ij})$, $k = 1, \ldots, n$, you can define an asymmetry structure $d^1 = (d_1^1, \ldots, d_{r^1}^1)$ that groups the index of the bidders with the same value for $f_{\mathbf{B_k^*}}(\widetilde{b}_{ij})$. Let $G^1 = (G_1^1, \ldots, G_{r^1}^1)$ denote the corresponding equivalence class for the values $f_{\mathbf{B_k^*}}(\widetilde{b}_{ij})$, in particular $\sharp G_k^1 = d_k^1$. Let $T^1 = \sum_{s=1}^{r^1} (d_s^1 - 1)$. First, if the values are all distinct, i.e. $T^1 = 0$, then we are done by setting $\mathfrak{I}^* = [\widetilde{b}_{ij} - \epsilon, \widetilde{b}_{ij} + \epsilon]$ for $\epsilon > 0$ small enough. Second, in any other case, we move to the induction loop and set $\mathfrak{I}^1 = [\widetilde{b}_{ij} - \epsilon, \widetilde{b}_{ij} + \epsilon]$ with $0 < \epsilon < \delta$ and ϵ small enough such that $\int_{\mathfrak{I}^1} f_{\mathbf{B_i^*}}(u)du \neq \int_{\mathfrak{I}^1} f_{\mathbf{B_j^*}}(u)du$ (the 'initial remark' guarantees the existence of such an ϵ).

Induction loop. Take as given $\mathfrak{B}^k \subset \mathfrak{B}$ (the set of remaining '$\widetilde{b}_{ij}$') and $\mathfrak{I}^k$ a current subset of $[b_{inf}, b_{sup}]$ such that if $\int_{\mathfrak{I}^k} f_{\mathbf{B_i^*}}(u)du = \int_{\mathfrak{I}^k} f_{\mathbf{B_j^*}}(u)du$ then $\widetilde{b}_{ij} \in \mathfrak{B}^k$. Define the asymmetry structure $d^{k+1} = (d_1^{k+1}, \ldots, d_{r^{k+1}}^{k+1})$ that groups the index of the bidders with the same value for $\int_{\mathfrak{I}^k} f_{\mathbf{B_i^*}}(b)db$. Let $G^{k+1} = (G_1^{k+1}, \ldots, G_{r^{k+1}}^{k+1})$ denote the corresponding equivalence class for the values $\int_{\mathfrak{I}^k} f_{\mathbf{B_i^*}}(b)db$. Let $T^k = \sum_{s=1}^{r^k} (d_s^k - 1)$. First, if the values are all distinct, i.e. $T^k = 0$, then we are done by setting $\mathfrak{I}^* = \mathfrak{I}^k$ and we exit the induction loop. Second, in any other case, there exists i, j, $i \neq j$ such that $i, j \in G_l^{k+1}$ for some $l \in \{1, \ldots, r^{k+1}\}$ and from our induction hypothesis we are sure that $\widetilde{b}_{ij} \in \mathfrak{B}^k$. Then we set $\mathfrak{B}^{k+1} = \mathfrak{B}^k \setminus \widetilde{b}_{ij}$ and

$\mathfrak{I}^{k+1} = \mathfrak{I}^k \cup [\widetilde{b}_{ij} - \epsilon, \widetilde{b}_{ij} + \epsilon]$ with $\epsilon < \delta$ and ϵ small enough such that the differences $\int_{[\widetilde{b}_{ij}-\epsilon,\widetilde{b}_{ij}+\epsilon]} f_{\mathbf{B}_\mathbf{l}^*}(b)db - \int_{[\widetilde{b}_{ij}-\epsilon,\widetilde{b}_{ij}+\epsilon]} f_{\mathbf{B}_{\mathbf{l}'}^*}(b)db$ for any l, l' are smaller than the minimum of the non-vanishing differences of the form $(\int_{\mathfrak{I}^k} f_{\mathbf{B}_\mathbf{l}^*}(b)db - \int_{\mathfrak{I}^k} f_{\mathbf{B}_{\mathbf{l}'}^*}(b)db)/2$ (such an ϵ exists from the 'initial remark'). The construction of ϵ guarantees that for bidders l and l' such that $\int_{\mathfrak{I}^k} f_{\mathbf{B}_\mathbf{l}^*}(b)db \neq \int_{\mathfrak{I}^k} f_{\mathbf{B}_{\mathbf{l}'}^*}(b)db$, we have $\int_{\mathfrak{I}^{k+1}} f_{\mathbf{B}_\mathbf{l}^*}(b)db \neq \int_{\mathfrak{I}^{k+1}} f_{\mathbf{B}_{\mathbf{l}'}^*}(b)db$. The construction guarantees also that $\int_{\mathfrak{I}^{k+1}} f_{\mathbf{B}_\mathbf{i}^*}(b)db \neq \int_{\mathfrak{I}^{k+1}} f_{\mathbf{B}_\mathbf{j}^*}(b)db$. Finally we have $\int_{\mathfrak{I}^{k+1}} f_{\mathbf{B}_\mathbf{i}^*}(u)du = \int_{\mathfrak{I}^{k+1}} f_{\mathbf{B}_\mathbf{j}^*}(u)du$ then $\widetilde{b}_{ij} \in \mathfrak{B}^{k+1}$ for any $i, j, i \neq j$ and we start again the induction loop.

By construction, if the iterative procedure stops, then we have found a solution for $\mathfrak{I}^*$. The remaining point is to note that at each step of the induction loop the value of T^k (which is initially smaller than $n-1$) decreases of at least one increment while the induction loop stops when $T^k = 0$. Finally, the induction loop will end after a finite number of iterations and find a solution $\mathfrak{I}^*$.

II. End of the proof with general symmetry structures Consider now the general case where some CDFs $F_{\mathbf{X_i}}$, $i = 1, \ldots, n$, may coincide and let r denote the number of distinct CDFs among those latter. Let $(F_{\mathbf{B}_\mathbf{j}^{**}})_{1 \leq j \leq r}$ denote the corresponding distinct CDFs such that for any $i \in [1, n]$, there exists $j \in [1, r]$ such that $F_{\mathbf{B}_\mathbf{i}^*} = F_{\mathbf{B}_\mathbf{j}^{**}}$. Since each solution is defined up to a permutation, we can assume w.l.o.g. that $F_{\mathbf{B}_\mathbf{i}^*} = F_{\mathbf{B}_\mathbf{i}^{**}}$ for $i = 1, \cdots, r$. By considering the r first equalities in (6) and after integrating according to the variable u_2 from $\underline{x}$ to b_2, the right terms are equal to $X \times J^{r \times r}_{(f_{\mathbf{B}_\mathbf{1}^*}(u_1), \ldots, f_{\mathbf{B}_\mathbf{n}^*}(u_1))}$, where $J^{r \times r}_{(f_{\mathbf{B}_\mathbf{1}^*}(u_1), \ldots, f_{\mathbf{B}_\mathbf{n}^*}(u_1))}$ is a matrix defined in appendix A.1 and $X = [k_1 \cdot F_{\mathbf{B}_\mathbf{1}^{**}}(b_2), \ldots, k_r \cdot F_{\mathbf{B}_\mathbf{r}^{**}}(b_2)]$, and are thus identified. Suppose that there exists b_1 such that the $f_{\mathbf{B}_\mathbf{j}^{**}}(b_1)$'s take r distinct values.[34] From lemma A.2, $J^{r \times r}_{(f_{\mathbf{B}_\mathbf{1}^*}(u_1), \ldots, f_{\mathbf{B}_\mathbf{n}^*}(u_1))}$ is then invertible and $[k_1 \cdot F_{\mathbf{B}_\mathbf{1}^{**}}(b_2), \ldots, k_r \cdot F_{\mathbf{B}_\mathbf{r}^{**}}(b_2)]$ is identified. We first identify the vector $(k_j)_{1 \leq j \leq r}$ by fixing $b_2 = \overline{x}$. Second we identify the vectors $[F_{\mathbf{B}_\mathbf{1}^{**}}(b_2), \ldots, F_{\mathbf{B}_\mathbf{r}^{**}}(b_2)]$ for any b_2, which concludes the proof.

A.3 Smoothness of the observables

We are interested in the smoothness of the densities $f_{\mathbf{B_p}|\mathbf{Z}}(.|.)$ for $p = 1, \cdots, n$. 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 $f_{\mathbf{B}_\mathbf{i}^*|\mathbf{Z}}(.|.)$.

[34] Similarly to the full asymmetric case, there exists an event E such that the probabilities of the events $B_i^* \in E$ takes r distinct values such that the logic of the argument is indeed general.

Proposition A.3 *Given A3-A4, the conditional distribution* $F_{\mathbf{B_p}|\mathbf{Z}}(.|.)$, $p = 1, \cdots, n$, *satisfies for both the first and second price auctions (if not specified):*

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

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

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

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

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

Proof In their proposition 1, GPV obtains the same properties for the CDFs and PDFs $F_{\mathbf{B_i^*}|\mathbf{Z}}$ and $f_{\mathbf{B_i^*}|\mathbf{Z}}$ instead of $F_{\mathbf{B_p}|\mathbf{Z}}$ and $f_{\mathbf{B_p}|\mathbf{Z}}$. From (3) and (4), we obtain that any CDF $F_{\mathbf{B_p}|\mathbf{Z}}(b,z)$ can be expressed as a linear combination of terms which are products of $F_{\mathbf{B_i^*}|\mathbf{Z}}(b,z)$, i.e. as a continuous function of the CDFs $F_{\mathbf{B_i^*}|\mathbf{Z}}$. The CDF $F_{\mathbf{B_i^*}|\mathbf{Z}}$ have the desired smoothness properties on the set $S(F_{\mathbf{B_n}|\mathbf{Z}}) \setminus \{\beta_i(\overline{x};z)\}$: on the set $S(F_{\mathbf{B_i^*}|\mathbf{Z}})$, it comes from GPV, whereas $F_{\mathbf{B_i^*}|\mathbf{Z}}$ is equal to 1 above $\beta_i(\overline{x};z)$ and is thus C^∞ on $(\beta_i(\overline{x};z), \infty)$. The regularity property (iii) that is valid for $F_{\mathbf{B_i^*}|\mathbf{Z}}$ is thus still valid for $F_{\mathbf{B_p}|\mathbf{Z}}$ if the points $\{\beta_i(\overline{x};z)\}_{i=1,\cdots,n}$ have been appropriately removed. The same argument works for (iv-v).

The image of a closed interval by a continuous function is a closed interval. Thus (i) holds also for $F_{\mathbf{B_p}|\mathbf{Z}}$. 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 $S^o(F_{\mathbf{B_p}|\mathbf{Z}})$ as above. By deriving (4) and (3), we obtain another expression of $f_{\mathbf{B_p}|\mathbf{Z}}(b|z)$ as a function of $F_{\mathbf{B_i^*}|\mathbf{Z}}(b|z)$ and $f_{\mathbf{B_i^*}|\mathbf{Z}}(b|z)$:

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

Thus we obtain that $f_{\mathbf{B_p}|\mathbf{Z}}(b,z)$ is strictly positive on $S^o(F_{\mathbf{B_p}|\mathbf{Z}})$.[35] **Q.E.D.**

A.4 Proof of Proposition 6.1

We write the proof for the first price auction, the arguments are easily adapted for the second price auction. In a first stage, we first consider full asymmetric structure and then show how to adapt the proof to general asymmetry structures. 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 did not appear in GPV.

Full asymmetric structure

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 that GPV 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 $f_{\mathbf{B_p},\mathbf{Z}}(b,z)$.

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 $\widehat{F}_{\mathbf{B}|\mathbf{Z}}(b|z)$ and $\widehat{f}_{\mathbf{B}|\mathbf{Z}}(b|z)$ on any inner

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

closed compact subset of the bidding support, denoted by $\mathcal{C}$.

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

In GPV, the corresponding uniform rates of convergence are obtained for the bidding distributions and densities $\widehat{F}_{\mathbf{B_i^*}|\mathbf{Z}}(b|z)$ and $\widehat{f}_{\mathbf{B_i^*}|\mathbf{Z}}(b|z)$ since bidders' identities are observed. However, we establish that the function mapping the vector of the order statistics CDF $(F_{\mathbf{B_p}|\mathbf{Z}}(b|z))_{p\in[1,n]}$ into $(F_{\mathbf{B_i^*}|\mathbf{Z}}(b|z))_{i\in[1,n]}$ is continuously differentiable on $\mathcal{C}$ 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 CDFs $(F_{\mathbf{B_p}|\mathbf{Z}}(b|z))_{p\in[1,n]}$ into $(F^{(r:r)}_{\mathbf{B}|\mathbf{Z}}(b|z))_{r\in[1,n]}$ is a linear invertible function (the related matrix is triangular with the coefficient 1 on the diagonal). Second, we consider the function in (14) mapping the vector of the order statistics CDFs $(F^{(r:r)}_{\mathbf{B}|\mathbf{Z}}(b|z))_{r\in[1,n]}$ into $(F_{\mathbf{B_i^*}|\mathbf{Z}}(b|z))_{i\in[1,n]}$. The function $c \to \mathcal{R}[c]$ is a projection operator and is thus continuously differentiable on $\mathcal{C}^n$. We are thus left with the function Υ. The Jacobian matrix of the function Υ is well-defined and of full rank on $\mathcal{C}$ since assumption A5 guarantees that the probabilities $(F_{\mathbf{B}_i^*|Z}(b|z))_{i\in[1,n]}$ are all distinct for any $(b,z)\in\mathcal{C}$ (see remark A.1). We thus conclude that the uniform rate of convergence that holds for $(F_{\mathbf{B_p}|\mathbf{Z}}(b|z))_{p\in[1,n]}$ remains valid for $(F_{\mathbf{B_i^*}|\mathbf{Z}}(b|z))_{i\in[1,n]}$.

From equations (13) and (15), we have the following bounds for the densities on $\mathcal{C}$ where, asymptotically, the terms $(\widehat{F}_{\mathbf{B_i^*}|\mathbf{Z}}(b|z) - \widehat{F}_{\mathbf{B_j^*}|\mathbf{Z}}(b|z)), j\neq i$ are bounded away from zero:

$$\begin{aligned} ||\widehat{f}_{\mathbf{B_i^*}|\mathbf{Z}}(b|z) - f_{\mathbf{B_i^*}|\mathbf{Z}}(b|z)||_0 \leq C_1 \cdot ||\widehat{f}_{\mathbf{B_p}|\mathbf{Z}}(b|z) - f_{\mathbf{B_p}|\mathbf{Z}}(b|z)||_0 \\ + C_2 \cdot ||\widehat{F}_{\mathbf{B_i^*}|\mathbf{Z}}(b|z) - F_{\mathbf{B_i^*}|\mathbf{Z}}(b|z)||_0 \end{aligned} \tag{25}$$

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

The remaining issues are the consistency and the uniform rates of convergence of the pseudo probabilities $\widehat{P}_{ipl}$. From (18), the pseudo probabilities can be expressed as a continuous differentiable function of $\widehat{f}_{\mathbf{B_i^*}|\mathbf{Z}}$ 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 $\widehat{X}_{ipl}$.

General asymmetry structures

We first show that the uniform rate of convergence that holds for $(F_{\mathbf{B_p}|\mathbf{Z}}(b|z))_{p\in[1,n]}$ remains valid for $(F_{\mathbf{B_i^*}|\mathbf{Z}}(b|z))_{i\in[1,n]}$ which comes from the fact that the function mapping the vector of the order statistics CDFs $(F_{\mathbf{B_p}|\mathbf{Z}}(b|z))_{p\in[1,n]}$ into $(F_{\mathbf{B_i^*}|\mathbf{Z}}(b|z))_{i\in[1,n]}$ is continuously differentiable on $\mathcal{C}$. Similarly to the full asymmetric structure, the key point is the continuous differentiability of the function in (23) mapping the vector of the order statistics CDFs $(F^{(r:r)}_{\mathbf{B}|\mathbf{Z}}(b|z))_{r\in[1,n]}$ into $(F_{\mathbf{B_i^*}|\mathbf{Z}}(b|z))_{i\in[1,n]}$. The functions $c \to \mathcal{R}[c]$ and $\Lambda_{(d_1,\ldots,d_r)}$ are continuously differentiable on $\mathcal{C}^n$. We are then left with showing that the function $\Upsilon_{(d_1,\cdots,d_r)}(.,\ldots,.)$ is continuously differentiable at any point $(F^{(1:1)}_{\mathbf{B}|\mathbf{Z}}(b|z),\ldots,F^{(n:n)}_{\mathbf{B}|\mathbf{Z}}(b|z))$ where $(b,z)\in\mathcal{C}$, an inner closed compact subset of the bidding support. First, our construction for $\Upsilon_{(d_1,\cdots,d_r)}$ guarantees that the map $\Upsilon^{-1}_{(d_1,\cdots,d_r)}$ corresponds exactly to $\mathcal{M}_{d_1,\ldots,d_r}$ in some neighborhood of any point $(P_1,\ldots,P_r)$ where $P_k=(X-y^k)^{d_k}$ and $(y^k)_{k=1,\ldots,r}$ is a vector of distinct roots such that $y^i>y^j$ for $i<j$ such that $d_i=d_j$. Second, the jacobian matrix of $\mathcal{M}_{d_1,\ldots,d_r}$ at such a point $(P_1,\ldots,P_r)$ is given by $M_{\mathbf{P_1},\ldots,\mathbf{P_r}}$, a matrix which is invertible if the y^k, $k=1,\ldots,r$ are all distinct as established in proposition A.2. Finally, under assumption A5, we obtain that the Jacobian matrix of the function $\Upsilon_{(d_1,\cdots,d_r)}$ is well-defined and of full rank at any point $(F^{(1:1)}_{\mathbf{B}|\mathbf{Z}}(b|z),\ldots,F^{(n:n)}_{\mathbf{B}|\mathbf{Z}}(b|z))$ where $b\in\mathcal{C}$.

Second, to obtain the same rate of convergence as in the fully asymmetric case, we now have to use equations (13) and (16) [instead of (15)], where asymptotically the terms $(\widehat{F}_{\mathbf{B_i^*}|\mathbf{Z}}(b|z)-\widehat{F}_{\mathbf{B_j^*}|\mathbf{Z}}(b|z))$ for i and j such that $\widehat{F}_{\mathbf{B_i^*}|\mathbf{Z}}(.|.)\neq\widehat{F}_{\mathbf{B_j^*}|\mathbf{Z}}(.|.)$ are bounded away from zero on $\mathcal{C}$. We then obtain exactly the same bound as in (25).