Supplementary Material to “The Econometrics of Auctions with Asymmetric Anonymous Bidders”∗

Laurent Lamy†

This supplementary material contains four sections. The first is devoted to the proof of the non-identification part of proposition 3.1 for the first price auction. The second is devoted to details on our estimation procedure. The third is devoted to the proof of Proposition 7.1. The last section is devoted to our Monte Carlo simulations.

1 Proof of Proposition 3.1: Complement

We consider an affiliated distribution of private signals $F_X$ and then a corresponding distribution of (non-anonymous) bids $F_B^*$ given by the bidding functions $\beta_i(\cdot)$, $i = 1, \ldots, n$, with the corresponding joint distribution $F_B$ of order statistics. Assumption A1 guarantees that $\beta_i'(x) > 0$ on $[x, \bar{x}]$. Figure A illustrates our construction of the perturbations $F_X^\gamma$ leading to the same distribution $F_B^*$. Below the notation $\beta(S)$ refers to as the set $\{b^*|\exists x$ such that $\beta_i(x_i) = b^*_i, \forall i\}$. In a first step we construct a family of CDFs $F_B^\gamma$ that are perturbations of $F_B^*$. In a second step we show that those perturbations are corresponding to distributions of private signals that differ (up to a permutation) from $F_X$.

Consider two bids $\underline{b}$ and $\bar{b}$ with $\bar{b} > \underline{b}$ and such that there exists $\epsilon$ such that a couple a bidder $(i, j)$ submits bids in the intervals $[\underline{b} - \epsilon, \underline{b} + \epsilon]$ and $[\bar{b} - \epsilon, \bar{b} + \epsilon]$ with positive probability while the remaining bidders are bidding in the interval $[\underline{x}, \underline{x} + \epsilon]$ with positive probability and such that $\underline{x} + \epsilon < \underline{b} - \epsilon$, $\underline{b} - \epsilon < \bar{b} + \epsilon$ and $\bar{b} + \epsilon < \sup_{x, i} \beta_i(x)$. We define a first perturbation:

∗The reference for equations from the paper are put into square brackets while parentheses are used for equations appearing in this supplementary material.
†Paris School of Economics, 48 Bd Jourdan 75014 Paris. e-mail: laurent.lamy@pse.ens.fr
\[ c_1(b^*; x, i, j) = \left( \mathbf{1}\{b_i^* \in [\bar{b}, \bar{b} + \epsilon]\} - \mathbf{1}\{b_i^* \in [\bar{b} - \epsilon, \bar{b}]\} \right) \prod_{k \neq i,j} \mathbf{1}\{b_k^* \in [\bar{x}, \bar{x} + \epsilon]\} \]

\[-(\mathbf{1}\{b_i^* \in [\bar{b}, \bar{b} + \epsilon]\} - \mathbf{1}\{b_i^* \in [\bar{b} - \epsilon, \bar{b}]\} \prod_{k \neq i,j} \mathbf{1}\{b_k^* \in [\bar{x}, \bar{x} + \epsilon]\}\]

This perturbation is depicted in Figure A with the thick plain arrows. We then define the perturbation on private signals that corresponds to \( c \) if all bidders still use the bidding functions \( \beta \): \( c(x; \epsilon, i, j) = \prod_{k=1}^n \beta'_k(x_k) \cdot c(\beta_1(x_1), \ldots, \beta_n(x_n); \epsilon, i, j) \). This function is null except on two distinct cartesian products: \( S_1 = \prod_{k \neq i,j} [x, \beta^{-1}_k(x + \epsilon)] \times [\beta^{-1}_i(\bar{b} - \epsilon), \beta^{-1}_i(\bar{b})] \times [\beta^{-1}_j(\bar{b} - \epsilon), \beta^{-1}_j(\bar{b} + \epsilon)] \) and \( S_2 = \prod_{k \neq i,j} [x, \beta^{-1}_k(x + \epsilon)] \times [\beta^{-1}_i(\bar{b} - \epsilon), \beta^{-1}_i(\bar{b})] \times [\beta^{-1}_j(\bar{b} - \epsilon), \beta^{-1}_j(\bar{b} + \epsilon)] \). We define \( c^*_1(x; \epsilon, i, j) = \prod_{k \neq i} \beta'_k(x_k) \beta'_i(x_i - \beta^{-1}_i(\bar{b})) \cdot c_1(\beta_1(x_1), \ldots, \beta(x_i - \beta^{-1}_i(\bar{b})), \ldots, \beta_n(x_n); \epsilon, i, j) \) the symmetric transformation of the function \( c(x; \epsilon, i, j) \) on the support \( S_1 \) according to the axis \( x_i = \beta^{-1}_i(\bar{b}) \).

The symmetric transformation of the support \( S_1 \) is denoted \( S^*_1 \). From the perturbation \( c^*_1(x; \epsilon, i, j) \) we can come back to perturbations in term of bids on the support \( \beta(S^*_1) \): \( c^*_2(b^*; \epsilon, i, j) \). Then we define \( c^*_2(b^*_{-i,j}, b_j; \epsilon, i, j) = -c^*_2(b^*_{-i,j}, b_j; \epsilon, i, j) \).

![Figure A](image-url)

On the whole we consider the perturbation of the bid distribution \( f_{B^*}(.) = f_{B^*}(.) + \gamma \cdot (\sum_{k=1,2} c_k(.; \epsilon, i, j) + c_k(.; \epsilon, i, j)) \). The perturbations of the bids have been explicitly built such that \( \int f_{B^*} = 1 \). For a sufficiently small \( \gamma > 0 \), \( f_{B^*}(.) > 0 \) and is thus a PDF with the functions \( c \) shifting probability weight from some regions to others.
Moreover the shifts are explicitly built such that they correspond to shifts between regions that are not identifiable with anonymous data: \( f^\gamma_B(.) = f_B(.) \). From \( f^\gamma_B(.) \) we uniquely define \( f^\gamma_X(.) \) by the equilibrium equation (2) of the first price auction which is still strictly increasing if \( \gamma \) is small enough. Finally, it remains to show that \( f^\gamma_X(.) \) differs from \( f_X(.) \) for a continuum of \( \gamma \)'s in the upper neighborhood of \( \gamma = 0 \). The proof is in two steps: first the bidding function of bidder \( j \) is unchanged, second his type distribution is modified by the perturbation. The first step comes from the subtle double perturbation that is not used in the proof for the second price auction. The perturbation \( c^\gamma_i(x;\epsilon,i,j) \) that is added to the original one \( c_i(x;\epsilon,i,j) \) guarantees that the distribution of the bids of his opponents conditional on his type remains unchanged and thus his bidding function is unchanged. The second step follows the same arguments as in the proof for the second price auction: though the modified type distribution could coincide up to a permutation for a finite number of \( \gamma \)'s, it is not possible with an infinite number.

As in the proof for the second price auction, a smoothed version of our perturbations will guarantee that affiliation is preserved.

2 Complements on estimation

Lemma 2.1 In the IPV model and under A1, if \( F_{X_i}(.) \) and \( F_{X_j}(.) \) are strictly distinct then \( F_{B_i^1}(.) \) and \( F_{B_j}(.) \) are strictly distinct.

Proof It is straightforward for the second price auction. Consider the first price auction and suppose that \( F_{B_i^1}(.) \) and \( F_{B_j}(.) \) are not strictly distinct. Then there exists an interval with a positive measure \( I_B \subset [x,\max\{\bar{b}_i,\bar{b}_j\}] \), where \( \bar{b}_k \) denotes the upper bound of the bidding support of bidder \( k \), such that \( F_{B_i^1}(.) = F_{B_j}(.) \) on \( I_B \). Then there exists an interval with a positive measure \( [b_1,b_2] \subset I_B \) such that \( f_{B_i^1}(.) = f_{B_j}(.) \) on \( [b_1,b_2] \). Since bidding distribution are also independent, we obtain that \( \xi_i^{rst}(.,F_B) = \xi_j^{rst}(.,F_B) = \xi(.) \) on \( [b_1,b_2] \). After noting that \( F_{X_k}(x) = F_{X_k}(\xi_i^{-1,rst}(x)) \) for \( k = i,j \), we obtain that \( F_{X_i}(.) = F_{X_j}(.) \) on the interval \( [\xi^{-1}(b_1),\xi^{-1}(b_2)] \) which has positive measure and is included in \( [x,\bar{x}] \), which means that \( F_{X_i}(.) \) and \( F_{X_j}(.) \) are not strictly distinct. Q.E.D.
2.1 Details of the Fourth step: the trimming rule

**Fourth step**  We first estimate the boundary of the support of the joint distribution of \((B, Z)\), which is unknown. Since the support of \(Z\) can be assumed to be known, we focus on the estimation of the support \([\hat{b}(z), \bar{b}(z)]\) of the conditional distribution of \(B\) given \(Z\). On the one hand, we assume that \(\hat{b}(z)\) does not depend on \(z\) and is estimated by the minimum of all submitted bids. On the other hand, \(\bar{b}(z)\) should be estimated as in GPV. Let \(h_d > 0\). We consider the following partition of \(\mathbb{R}^d\) with a generic hypercube of side \(h_d\): \(\vartheta_{k_1, \ldots, k_d} = [k_1 h_d, (k_1 + 1) h_d) \times \cdots \times [k_d h_d, (k_d + 1) h_d)\), where \(k_1, \ldots, k_d\) runs over \(\mathbb{Z}^d\). This induces a partition of \([z, \bar{z}]\). The estimate of the upper boundary \(\bar{b}(z)\) is the maximum of those bids for which the corresponding value of \(Z_l\) falls in the hypercube \(\vartheta_{k_1, \ldots, k_d}(z)\) containing \(z\). Formally, our estimators for the upper and lower boundaries are respectively given by \(\hat{b} = \inf \{B_{1l}, l = 1, \ldots, L\}\) and \(\bar{b}(z) = \sup \{B_{nl}, l = 1, \ldots, L | Z_l \in \vartheta_{k_1, \ldots, k_d}(z)\}\). Our estimator for \(S(F_{B_p, Z})\) is \(\hat{S}(F_{B_p, Z}) = \{(b, z) : b \in [\hat{b}, \bar{b}(z)], z \in [z, \bar{z}]\}\).

We now turn to the trimming for the first price auction. It is well known that kernel estimators are asymptotically biased at the boundaries of the support. Following GPV, we have to trim out observations that are close to the boundaries of the support. Let \(\rho_{f_{B_p|Z}}(z)\) and \(\rho_{F_{B_p|Z}}\) denote respectively the length of the support (i.e., the difference between the maximum and minimum elements in the support) of \(K_{f_{B_p|Z}}(./h_{f_{B_p|Z}}, z)\) and \(K_{F_{B_p|Z}}(./h_{F_{B_p|Z}}, z)\). In particular, \(\hat{f}_{B_p|Z}(./|z)\) and \(\hat{F}_{B_p|Z}(./|z)\) are asymptotically unbiased respectively on \([b + \rho_{f_{B_p|Z}}(z), \bar{b}(z) - \rho_{f_{B_p|Z}}(z)]\) and \([\hat{b} + \rho_{F_{B_p|Z}}, \bar{b}(z) - \rho_{F_{B_p|Z}}]\). This leads to defining the sample of pseudo private values \(\{\hat{X}_{ipl}; i = 1, \ldots, n; p = 1, \ldots, n; l = 1, \ldots, L\}\) where \(\hat{X}_{ipl}\), the estimate of the private value of bidder \(i\) would it be the bidder of the \(p^{th}\) order statistic of the vector of bids \(B_l\), is defined by\(^1\)

\(^1\)The factor in front of the \(\rho\)-coefficients could be only ‘1’ if the boundaries \(\hat{b}\) and \(\bar{b}(z)\) were known. Those bounds are consistently estimated by our estimator and any coefficient strictly greater than ‘1’ would work as in GPV. For simplicity, we chose the coefficient 2.
in the first price auction and \( \hat{X}_{ipl} = B_{pl} \) in the second price auction.

2.2 Detailed formulation of the technical assumptions

**Assumption A3**

(i) The \( d \)-dimensional vectors \( Z_l, l = 1, 2, \cdots \), are independently and identically distributed as \( F_Z(.) \) with density \( f_Z(.) \).

(ii) For each \( l \), the variables \( X_{il}, i = 1, \ldots, n \) are independently distributed conditionally upon \( Z_l \) as \( F_{X|Z}(x|z) \) with density \( f_{X|Z}(x|z) \).

**Assumption A4** For each bidder \( i = 1, \ldots, n \),

(i) \( S(F_{X_i,Z}) = \{(x,z) : z \in [z_l,z_u], x \in [x_l,x_u]\} \) with \( z_l < z_u \);

(ii) for \( (x,z) \in S(F_{X_i,Z}) \), \( f_{X_i|Z}(x|z) \geq c_f > 0 \), and for \( z \in S(F_Z) \), \( f_Z(z) \geq c_f > 0 \);

(iii) \( F_{X_i|Z}(\cdot|\cdot) \) and \( f_Z(\cdot) \) admit up to \( R+1 \) continuous bounded partial derivatives on \( S(F_{X_i,Z}) \) and \( S(F_Z) \), with \( R \geq 1 \).

**Assumption A6**

- **KERNELS**

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

  (ii) \( \int K_{f_{Z}}(z)dz = 1 \), \( \int K_{F_{B_p|Z}}(z)dz = 1 \), \( \int K_{f_{B_p|Z}}(b,z)dbdz = 1 \), for any \( p = 1, \cdots, n \) and \( \int K_{f_{X_i|Z}}(x,z)dxdz = 1 \) for any \( i = 1, \cdots, n \).
(iii) $K_{Fp|z}(\cdot), K_{f_{Bp}|z}(\cdot, \cdot), K_{f_{X_i}|z}(\cdot, \cdot)$ and $K_{f_{Z}}(\cdot)$ are of order $R + 1$, $R + 1$, $R$ and $R + 1$ respectively, i.e. moments of order strictly smaller than the given order vanish.

- **BANDWIDTHS**

  (i) The bandwidths $h_{Fp|z}, h_{f_{Bp}|z}$, for $p = 1, \cdots, n$, $h_{f_{X_i}|z}$ for $i = 1, \cdots, n$, and $h_{f_{Z}}$ are of the form:

  $$h_{Fp|z} = \lambda_{Fp|z} \left( \frac{\log L}{L} \right)^{\frac{1}{2R+d+1}}, \quad h_{f_{Bp}|z} = \lambda_{f_{Bp}|z} \left( \frac{\log L}{L} \right)^{\frac{1}{2R+d+1}},$$

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

  where the $\lambda$’s are strictly positive constants and $f = 0$ in the first price auction while $f = 2$ in the second price auction.

  (ii) The “boundary” bandwidth is of the form $h_\delta = \lambda_\delta \left( \frac{\log L}{L} \right)^{\frac{1}{2R+d+1}}$ if $d > 0$ where the $\lambda$’s are strictly positive constants.

### 3 Proof of Proposition 7.1

We adapt GPV’s proof to the asymmetric framework. To ease the exposition, we consider the case where $n = 2$. The first step is identical to GPV: it is sufficient to prove the proposition by replacing $f_{X|z}$ by $f_{X,z}$. The set $U_e(f_{X,z})$ can also be replaced by any subset $U \subset U_e(f_{X,z})$. Then the second step consists in the construction of a discrete subset $U$ of the form $\{f_{X,z, mk}(\cdot, \cdot), k = 1, \cdots, m^{d+1}\}$, where $m$ is increasing with the sample size $L$, that are suitable perturbations of $f_{X,z}$.

We consider a nonconstant and odd $C_\infty$-function $\phi$, with support $[-1, 1]^{d+1}$, such that

$$\int_{[-1,0]} \phi(b, z)db = 0, \quad \phi(0, 0) = 0, \quad \phi'(0, 0) \neq 0,$$

where $\phi'$ denotes the derivative of $\phi$ according to its first component.

Let $C(B^*)$ be the image of $C(X)$ by the function that maps bidders’ types into observed bids. It is a nonempty inner compact subset of $S(f_{B^*, z})$. Let $(b_k, z_k), k = 1, \cdots, m^{d+1}$ be distinct points in the interior of $C(B^*)$ such that the distance between $(b_k, z_k)$ and $(b_j, z_j), j \neq k$, and the distance between $(b_k, z_k)$ and any point outside
\( C(B^*) \) are larger than \( \lambda_1/m \). Thus, one can choose a constant \( \lambda_2 > 1/\lambda_1 \) such that the \( m^{d+1} \) functions

\[
\phi_{mk}(b, z) = \frac{1}{m^{R+1}} \phi(m\lambda_2(b - b_k), m\lambda_2(z - z_k)) \quad (k = 1, \cdots, m^{d+1})
\]

have disjoint hypercube supports. Let \( C_3 \) be a positive constant (chosen below), for each \( k = 1, \cdots, m^{d+1} \) define:

\[
f_{B^*_1; z, mk}(b, z) = \begin{cases} f_{B^*_1; z}(b, z) & \text{if } i = 1 \\ f_{B^*_2; z}(b, z) - C_3\phi_{mk}(b, z) & \text{if } i = 2. \end{cases}
\]

That is \( f_{B^*_1; z, mk} \) differs from \( f_{B^*_2; z} \) only in the neighborhood of \((b_k, z_k)\). The function \( f_{B^*_2; z, mk}(b, z) \) is a density if \( C_3 \) is small enough (integrates to 1 from (2) and is bounded away from 0) with the same support as \( f_{B^*_1; z}(b, z) \). Now consider the functions \( \xi_{i,mk}(b, z) = b + \frac{f_{B^*_1; z, mk}(b, z)}{f_{B^*_2; z, mk}(b, z)} \) for \( i = 1, 2 \). If \( C_3 \) is small enough, then \( \xi_{i,mk}(b, z) \), \( i = 1, 2 \), is increasing in \( b \) with a differentiable inverse denoted by \( \xi_{i,mk}^{-1}(x, z) \). Then we define for \( i = 1, 2 \)

\[
f_{X_i; z, mk}(x, z) = f_{B^*_1; z, mk}(\xi_{i,mk}^{-1}(x, z), z)/\xi_{i,mk}'(\xi_{i,mk}^{-1}(x, z), z)
\]

\[
= \frac{f_{B^*_1; z, mk}(\xi_{i,mk}^{-1}(x, z), z) \cdot (f_{B^*_3; z, mk}(\xi_{3-i,mk}^{-1}(x, z), z))^2}{2(f_{B^*_3; z, mk}(\xi_{1,mk}^{-1}(x, z), z))^2 - f_{B^*_3; z, mk}(\xi_{3-i,mk}^{-1}(x, z), z)f_{B^*_3; z, mk}(\xi_{i,mk}^{-1}(x, z), z)}
\]

From the above expression, \( f_{X_i; z, mk}(x, z) > 0 \) if and only if \( f_{B^*_1; z, mk}(b, z) > 0 \), where \( b = \xi_{i,mk}^{-1}(x, z) \). This completes the construction of the densities \( f_{X; z, mk}(\cdot, \cdot) \), \( k = 1, \cdots, m^{d+1} \), which composes the set \( U \). Note that the supports of \( f_{X; z, mk}(\cdot, \cdot) \) and \( f_{B^*_2; z, mk}(\cdot, \cdot) \) coincide respectively with the supports of \( f_{X; z}(\cdot, \cdot) \) and \( f_{B^*_2; z}(\cdot, \cdot) \).

Then to adapt GPV’s proof, we need the analog of their lemma B1 where the notation \( f_{mk}(\cdot, \cdot) \) should be replaced by \( f_{X; z, mk}(\cdot, \cdot) \), where the first argument \( x \) is now the vector of bidders’ private values instead of a single uni-dimensional private value. The analog of Lemma B1 gives two points. First, an appropriate asymptotic lower bound is given for the uniform distance between two elements, i.e. the norm \( \|\cdot\|_{C(X)} \), in the set \( U \) as a function of \( \lambda_2 \), \( m \) and \( R \). With this bound we can apply Fano’s lemma exactly in the same way as in GPV: the step 3 in their proof is unchanged. Second, an asymptotic approximation is given for the distance between \( f_{X; z, mk} \) and \( f_{X; z}^0 \) in the norm \( \|\cdot\|_{r,C(X)} \), which guarantees that \( f_{X; z, mk} \) belongs to the
set $U$ if $m$ is large enough.

**Lemma 3.1 (Analog of lemma B1 in GPV)** Given $A3$-$A4$, the following properties hold for $m$ large enough:

(i) For any $k = 1, \ldots, m^{d+1}$, the supports of $f_{X,z,\cdot}^{\cdot}(\cdot, \cdot)$ and $f_{B^*,z,\cdot}^{\cdot}(\cdot, \cdot)$ are $S(f_{X,Z}(\cdot, \cdot))$ and $S(f_{B^*,Z}(\cdot, \cdot))$.

(ii) There is a positive constant $C_4$ depending upon $\phi$, $f_{B^*,z}(\cdot, \cdot)$ and $C(X)$ such that for $j \neq k$,

$$\|f_{X,z,\cdot}^{\cdot} - f_{X,z,\cdot}^{\cdot}\|_{0,C(X)} \geq C_4 \cdot \frac{C_3\lambda_2}{m^R}.$$  

(iii) Uniformly in $k = 1, \ldots, m^{d+1}$, we have

$$\|f_{X,z,\cdot}^{\cdot} - f_{X,z,\cdot}^{\cdot}\|_r = C_3\lambda_2^{r+1}O\left(\frac{1}{m^{R-r}}\right), \quad r = 0, \ldots, R - 1$$

$$\|f_{X,z,\cdot}^{\cdot} - f_{X,\cdot}^{\cdot}\|_R = C_3\lambda_2^{R+1} \cdot O(1) + o(1).$$

where the big $O(.)$ depends upon $\phi$ and $f_{B^*,z}^0$.

Let us detail the proof of (ii) and what has changed relative to GPV’s framework.

Remind that $(b_k, z_k) \in C(B^*)$ implies $(x_k, z_k) \in C(X)$. As in GPV, it then suffices to prove that $|f_{X,z,\cdot}^{\cdot}(x_k, z_k) - f_{X,z,\cdot}^{\cdot}(x_k, z_k)| \geq C_4 \cdot \frac{C_3\lambda_2}{m^R}$, where $x_k = \xi^0(b_k, z_k)$.

From (2), we have: $F_{B_i^*,z,\cdot}^{\cdot}(x_k, z_k) = F_{B_i^*,z}^0(x_k, z_k)$ and $f_{B_i^*,z,\cdot}^{\cdot}(x_k, z_k) = f_{B_i^*,z}^0(x_k, z_k)$ for $i = 1, 2$. The difference is for the expression of $F_{B_i^*,z,\cdot}^{\cdot}(x_k, z_k) - F_{B_i^*,z}^0(x_k, z_k)$ which equals to 0 for $i = 1$ and to $-C_3\frac{\lambda_2}{m^R}\phi'(0,0) \neq 0$ for $i = 2$. Thus $f_{X_2,z,\cdot}^{\cdot}(x_k, z_k) = f_{X_2,z,\cdot}^{\cdot}(x_k, z_k)$ which is bounded away from zero and we are left with the term $f_{X_1,z,\cdot}^{\cdot}(x_k, z_k) - f_{X_1,z,\cdot}^{\cdot}(x_k, z_k)$.

Then, from equation (4), we have:

$$f_{X_1,z,\cdot}^{\cdot}(x_k, z_k) = \frac{f_{B_1^*,z}^0(b_k, z_k) \cdot (f_{B_2^*,z}^0(b_k, z_k))^2}{2(f_{B_1^*,z}^0(b_k, z_k))^2 - F_{B_2^*,z}^0(b_k, z_k)(f_{B_2^*,z}^0(b_k, z_k)) + C_3\lambda_2\phi'(0,0)/m^R}$$

and

$$f_{X_2,z,\cdot}^{\cdot}(x_k, z_k) = \frac{f_{B_2^*,z}^0(b_k, z_k) \cdot (f_{B_2^*,z}^0(b_k, z_k))^2}{2(f_{B_2^*,z}^0(b_k, z_k))^2 - F_{B_2^*,z}^0(b_k, z_k)(f_{B_2^*,z}^0(b_k, z_k))}$$

Now compare (5) and (6). As $\phi'(0,0) \neq 0$ and $F_{B_2^*,z}^0$ is bounded away from zero since $(b_k, z_k)$ are far enough from the boundaries, the desired result (ii) follows. The
proof of (iii) is more involved and follows GPV’s proof with the same modification as above by carefully separating the cases $i = 1$ and $i = 2$. More precisely, we have $||f_{X_1, Z, mk} - f_{X_1}||_r = C_3 \lambda^r + O(\frac{1}{m^{1/2}})$ and $||f_{X_2, Z, mk} - f_{X_2}||_r = O(1)$ and the result follows for the product.

4 Monte Carlo Experiments

To illustrate our nonparametric procedures, we conduct a limited Monte Carlo study. The aim of this section is threefold. In particular, we want to convince the incredulous reader that the third step of our estimation procedure that consists in the estimation of the roots of a polynomial from the estimation of its coefficients -though it increases significantly the variance of our estimator- does not prevent its usefulness for empirical applications. Such concerns are legitimate in regard to the numerical analysis literature (Gautschi (1973), Mosier (1986), Wilkinson (1963)) which analyzes the sensitivity of the roots of a polynomial with respect to small perturbations to its coefficients. Pathologies arise in high-order polynomials and when the roots are relatively close (i.e. the ratio is close to unity). However it does not seem to be relevant in our simulations with two or three different kinds of bidders, i.e with polynomials of degree 2 or 3, and with the leading coefficient fixed to 1. Furthermore, with partially anonymous data -in particular the common case where only the identity of the winning bidder is disclosed- our procedure is competing with the standard ‘naive’ ones where the econometrician keeps only the bids that come from non-anonymous bidders. Both methodologies are not directly comparable though basic intuition suggests that using the complete vector of bids should outperform the ‘naive’ approaches. Our simulations when only the identity of the winner is disclosed confirm this intuition.

4.1 Description of the simulation protocols

Much of the reported results are for the second price auction. This is a pedagogical choice. In the first-price auction, both anonymity and the strategic complexity of the underlying game are making the estimation of the distribution of bidders’ private values from observed bids harder. In the second price auction where bidders are reporting their private values, the strategic dimension is absent and we thus isolate
the issues coming from anonymity (see Table 1 in the paper). To fit with realistic size of auction data, we consider both $L = 40$ and $L = 200$ auctions with two kinds of bidders. Our Monte Carlo experiments consist of 200 replications for our estimation procedure and 5000 replication for our testing procedure. The true distributions $F_X$ of private values are generated from the densities $f$, where $f(x) = (1 + \epsilon \cdot (1 - 2x)) \cdot 1_{0 \leq x \leq 1}$. The choices for $\epsilon$ are reported in Table 2. Except for the simulation 3, private values are generated independently. For each replication, we first generate randomly $L$ private values for each bidder. For the first price auction, an additional step is needed: corresponding equilibrium bids have to be computed. We use Mathematica’s differential equation package for solving the related systems of two or three differential equations.

Next, we apply our estimation procedure for each replication. The computation of the roots of a polynomial which is used extensively is tackled with Mathematica’s standard rootfinding package. On the contrary to the way kernel methods have been applied and theoretically analyzed (including our theoretical analysis) for the analysis of auction data by means of trimming rules, we use a method of boundary correction. Specifically we use the correction proposed first by Zhang et al (1999) and generalized by Karunamuni et al (2005), a kind of generalized reflection involving reflecting a transformation of the data with the reflection parameter $A$ fixed to 0.55 as in Zhang et al (1999). As in GPV and Li et al (2002), we chose the trivariate kernel $\frac{35}{32}(1 - u^2)1_{|u| \leq 1}$, which implicitly assumes that latent densities are once-continuously differentiable. The bandwidths are chosen as $h = C \times \hat{\sigma} \times L^{-1/5}$ where $\hat{\sigma}$ is the estimated standard deviations of the underlying variable whose density is estimated and $C = 2.975 \times 1.06$ as in Li et al (2002).

The estimators used in our simulations differ slightly from the one presented in the paper. In the second price auction, a preliminary estimation of bidders’ CDFs and densities is obtained at the end of the third step. Under full anonymity, there is few benefits to wait from the $4^{th} - 6^{th}$ additional steps and so our simulations report the

---

2 Simulations with three kinds of bidders give similar results and are available upon request.
3 The programs written in Mathematica are available upon request to the author. The execution time of a replication of the estimation procedure lasted between one and three minutes. The grid steps in the computation have been shrunk to 200 equally spaced points on the interval $[0, 1]$ to reduce the computational costs. The testing procedure raises no computational difficulty and a replication lasts a couple of seconds.
4 This work has to be done only once for each Monte Carlo simulations. Time and accuracy were thus not a constraint at this stage such that we do not apply the numerical methods that have been specifically designed for this problem, e.g. Marshall et al (1994) for two kinds of bidders.
estimation obtained at the end of step 3 (which explains which our estimator of the CDF may be outside the interval $[0, 1]$). On the contrary, with partial information about bidders’ identities, the $4^{th} - 6^{th}$ steps are a nice way to use this information in a tractable way in a nonparametric setup. Thus the simulations under full anonymity correspond to the estimator obtained at the end of step 3 and they allow us to gauge the additional noise that results from steps 2-3 which are removed under our theoretical asymptotical criteria and which is a big issue in numerical analysis. Instead of using equation (12) -in the paper- to estimate the densities, we directly derive our estimation of the CDFs jointed with some ad hoc smoothing procedures. Finally, the simulations 1-7 consider a framework with 6 bidders with two kinds of bidders (3 Strong and 3 Weak bidders) and thus do not fit exactly with our theoretical framework which considers that all bidders are different. Such a framework is motivated by real-world auction data (see Flambard and Perrigne (2006)). We thus have to adapt our methodology which is easily done after noting that $F_S + F_W$ and $F_S,F_W$ is uniquely determined by $F_B^{(1:1)}$ and $F_B^{(2:2)}$. The resolution involves thus the computation of the roots of a polynomial of order 2 uniquely based on $F_B^{(i:i)}$ for $i = 1, 2$ instead of $i = 1, \ldots, 6$ as would be required under complete asymmetry.

4.2 Discussion of the simulations

The results of our estimation simulations are summarized in Figures 1-5. In each figure, we display the true value, the median, the 5%, 10%, 90% and 95% percentiles of our estimates. This gives the corresponding 80% and 90% confidence intervals. In some figures, we plot the results obtained with different methodologies: the darker graphs are always corresponding to our new estimation procedure.

The top panels of Figures 1 and 2 compare the estimations of bidders’ CDFs under non-anonymous data and under fully anonymous data in the second price auc-

<table>
<thead>
<tr>
<th>Figures</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Auction format</td>
<td>Second-price</td>
<td>First-price</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of bidders</td>
<td>6</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of replications</td>
<td>200</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$L$</td>
<td>40</td>
<td>200</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distribution, $\epsilon$</td>
<td>$\pm\frac{1}{2}$</td>
<td>-</td>
<td>$\pm1$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Correlation</td>
<td>Independent</td>
<td>Correlated</td>
<td>Independent</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1: Summary of the simulation protocols
tion with two kinds of bidders. In Figure 1, where the sample size is 40 auctions, it appears that the 90%-confidence intervals are roughly doubled when the econometrician lack the knowledge of bidders’ identities. In Figure 2, where the sample size is 200 auctions, it appears that the 90%-confidence intervals are more than tripled and unreported simulations with larger sample sizes confirm this trend: the confidence intervals under full anonymity are shrinking at a much slower pace than the ones under non-anonymous data. It illustrates the point that much noise is added when we go from a noisy estimation of the coefficient of a polynomial to an estimation of its roots. Nevertheless, on the contrary to the pessimistic predictions suggested by a superficial interpretation of the numerical analysis literature, the job is not to bad especially for small data set if the noise is compared to the one resulting from standard sampling errors. Moreover, the bias of the estimator under full anonymity remains very limited and the noise does not appear to be specifically asymmetrically distributed around the true value.

In the case where only the identity of the winner is observed, the bottom panels of Figures 1 and 2 compare our approach with the ‘naive’ one 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)). The results are striking. By keeping only the highest bid, the ‘naive’ approach cannot 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 with a sample size of 200 auctions. On the contrary, our estimation procedure does a good job: the median of the estimates perfectly matches the true curve and confidence intervals are reasonable. For a sample size of 40 auctions, it appears in Figure 1 that our procedure does a better job than the ‘naive’ approach for the whole support of the distribution, in particular also for the upper tail of the distribution for which bids are mainly kept in the estimation procedure since those are winning bids with a very high probability. This last point is not necessary what could be expected: for the upper tail of the distribution, our estimation procedure involves the estimation of the pseudo probabilities which are a source of additional noise compared to the ‘naive’ approach. This phenomena appears
actually for large sample auctions in the top panels of Figure 2 where the ‘naive’ approach slightly outperforms ours for the upper half (respectively upper quarter) of the distribution of the strong (resp. weak) bidder. Anyway, for the lower tail of the distribution, our approach strikingly outperforms the ‘naive’ approach.

Still in the case where the identity of the winner is observed, the median panels of Figures 1 and 2 compare our involved 6-steps approach with the more simple one that stops at the end of step 3. The gain is clearly important: the limited bias under full anonymity disappears completely with the use of this additional information and the shrinks in the confidence intervals lie between one quarter and one third. Remark that our estimator of bidders’ CDFs can lead to ‘infeasible’ values outside the interval [0, 1] at the end of step 3. On the contrary, at the end of step 6, estimated CDFs always lie in the interval [0, 1].

The simulations reported in Figures 3-4 are devoted to a kind of robustness check. Our estimation approach and the ‘naive’ approach are both relying on the independence assumption. We consider thus a (significant) departure from this assumption. Under full anonymity, it appears that assuming wrongly independence to estimate a correlated asymmetric model leads to very close results that the ones obtained if the model were assumed to be symmetric. Basically, as it is illustrated in Figure 3, in about half of the cases, the estimated roots of the polynomial are complex conjugates which leads our procedure to take only the real part (see equation (15) in the paper), which also corresponds to the estimation if we had assumed that bidders are symmetric. In the bottom panel of Figure 3, we compare our estimator with the ‘naive’ approach. In the latter, the estimates of bidders’ CDFs strongly overestimate the true distributions and are very noisy in the lower tail of the distribution. The bad performance of the ‘naive’ estimator is especially true for the weak bidders for which very few observed values are available. In the former, the bias is also strong but has a special symmetric structure such that, on the whole, the estimation of the aggregate distribution of all bidders is not biased.

In Figure 4, we provide another argument in favor of our estimation procedure

---

5 Bidders’ values are constructed in the following way. A random number $y$ in $[0, 1]$ that is the weighted sum between a common shock and idiosyncratic shock is associated to each bidder. Shocks are supposed to be uniformly distributed on $[0, 1]$ and the weight on the common shock $\rho$ is fixed to $\rho = 0.25$ ($\rho = 0$ and $\rho = 1$ are corresponding respectively to independence and full positive correlation). Then for a strong (weak) bidder, the valuation is determined by $F_S^{-1}(y)$ ($F_W^{-1}(y)$). The correlation structure is thus such that bidders’ valuations are positively correlated.
compared to the ‘naive’ approach. If we wrongly assume that the sampling scheme is an independent asymmetric model whereas it is indeed a symmetric correlated model, then our various approaches (under full anonymity or with the knowledge of the identity of the winning bidder) are leading to the same accurate estimation as if bidders’ identities were observed. On the contrary, the ‘naive’ approach remains flawed: it does not estimate significant asymmetries between bidders but is strongly biased on all the support since it is mislead by the way it exploits the independence assumption. Implicitly, by taking the common real part of the estimated roots, our procedure drops the use of the independence assumption when we estimate a multiple root as it happens with positive correlation.

In Figure 5, we report the final results on the densities for the first price auction with two asymmetric bidders. In the top panels, the performance of GPV’s estimator is reported. Compared to GPV, the simulations are run with asymmetric bidders instead of symmetric bidders and also with different distributions. As in GPV, the median of our estimate perfectly matches the true curve in the interior of the support of the distribution. Our boundary correction is a significant improvement: the result is very impressive for the lower bounds of the distribution (for which we do not suffer from the noisy estimation of the function $\psi$ and thus where only one boundary correction is at work). At the upper bound, the estimation suffers from a limited bias but a strong noise (especially for the strong bidders). This issue is of much worry for the upper quintile of bidders’ values. Moreover, it is amplified under anonymous data and infects a larger part of the distribution as illustrated by the bottom panel of Figure 5. Whereas the estimation of the density in the lower half of the support is not biased and with a limited growth of the confidence intervals, the upper half is poorly estimated for the strong bidder. In particular, the upper bound of the strong bidder’s private value is strongly overestimated: the median is above 1.3 and the 95% percentile is at 1.9. Note also that the estimation is significantly deteriorated at the lower boundary of the weak bidder private value’s support (in a much limited way for the strong bidder) compared to the results under non-anonymous data and the related amplification of the noise in the middle of the support. The reason is that the estimation at the lowest boundary now really relies on two boundary corrections.

---

6On the contrary to GPV, we chose densities that peak at the bounds of the support of the distribution. Consequently, without any boundary correction, our estimations suffer from a much higher bias at the boundaries.
and not only one as in GPV as argued above. The noise that should be added is the one with respect to the pseudo probabilities which are relying on the estimation of the density of the bids at the end of the support. It appears that the estimated pseudo probabilities for a bid at the lower boundary overestimate the probability that it comes from the strong bidder. At the upper bound of the support, the above discussion means also than three boundary correction are involved.

On the whole we perceive our simulations as encouraging for the relevance of our procedures for empirical applications, especially in the case where additional information as the identity of the winner is available. The superiority against ‘naive’ approaches is clearly confirmed by our simulations. This is especially true for small samples where nonparametric approaches fail to work with a sufficient amount of bids. Nevertheless, we also emphasize than the details of the kernels estimators should not be neglected, especially the boundary corrections that were not emphasized in the previous literature.

References


Figure 1
Estimated Private Value Distribution in the second price auction with 3 weak and 3 strong bidders with a sample size of 40 auctions, Monte Carlo Results with 200 replications for each estimation. 50 percentile (full line), 5 and 95 percentiles (dashed lines) and 10 and 90 percentiles (dots). The darker plots are corresponding to our methodology.
Figure 2
Estimated Private Value Distribution in the second price auction with 3 weak and 3 strong bidders with a sample size of 200 auctions, Monte Carlo Results with 100 replications for each estimation. 50 percentile (full line), 5 and 95 percentiles (dashed lines) and 10 and 90 percentiles (dots). The darker plots are corresponding to our methodology.
Figure 3
Estimated Private Value Distribution in the second price auction with 3 weak and 3 strong correlated bidders ($\rho=0.25$) with a sample size of 200 auctions, Monte Carlo Results with 100 replications for each estimation. 50 percentile (full line), 5 and 95 percentiles (dashed lines) and 10 and 90 percentiles (dots). The darker plots are corresponding to our methodology.
Figure 4
Estimated Private Value Distribution in the second price auction when the bidders are classified as 3 strong and 3 weak bidders whereas they are 6 symmetric correlated bidders ($\rho=0.25$) with a sample size of 200 auctions, Monte Carlo Results with 100 replications for each estimation. 50 percentile (full line), 5 and 95 percentiles (dashed lines) and 10 and 90 percentiles (dots).
Fig 5: First Price Auction, Non-anonymous and anonymous data
True (thick grey line) and estimated (5, 50 and 95% percentiles) density of bidders' private values in the first price auction with 2 asymmetric bidders with a sample size of 200 auctions, Monte Carlo Results with 200 replications for each estimation. 50 percentile (full line), 5 and 95 percentiles (dashed lines) and 10 and 90 percentiles (dots)