Estimating Auction Models from Transaction Prices with Extreme Value Theory

Paolo Riccardo Morganti †
ITAM

March 14, 2011

Abstract

In this paper we reexamine some results recently appeared in the empirical auction literature: we show that, when only transaction prices are observed (Athey-Haile,(2002)), the distribution of private valuations is *irregularly identified*, in the sense that its estimate does not converge at the usual parametric rate. The sample bias produced by nonparametric estimators will affect all functionals of practical interest. We present here a different approach to estimation, theoretically justified by extreme value theory. Compared to existing approaches this approximation method has several advantages: it produces more accurate results, is computationally easy to perform, and does not require strong assumptions about the unobserved distribution of bidders' valuations.

1 Introduction

The analysis of auctions has inspired over the years one of the most successful marriages between theoretical and econometric models. Theorists, since the seminal work of Vickrey (1961), have elaborated a rich framework to map private valuations into bids. Econometricians, in their attempt to identify and estimate the distribution of these private values, have adopted the results from the theory as restrictions to place on the data (the bids).

†Contact: paolo.morganti@itam.mx. I want to thank Nicola Persico, Joerg Stoye, Konrad Menzel, Juan José Fernandez, Alex Horenstein, Jorge O. Moreno, Moussa Blimpo, Francisco Barillas, for precious suggestions, guidance, corrections and support.

The general approach to nonparametric identification in auction models relies on this *theoretical mapping* between the distribution of bidders' valuations - the object of interest - and the distribution of observed bids - the data. Given the latter, we can obtain the former by inverting the mapping.

When the econometrician has access to limited data - for instance, only to the transaction price - it is still possible, under certain conditions (Athey and Haile (2002), Haile and Tamer (2003)) to recover the object of interest using a *statistical mapping*, a relationship between the distribution of any order statistics and the underlying distribution of the data. The use of such mapping is justified by the observation that transaction prices are an order statistics of the bids, as explicitly described by the rules of the auction (for instance, in a second price auction, the transaction price is equal to the second highest bid). Given the distribution of any order statistics, it is possible to invert the statistical mapping to back out the underlying distribution[1].

However, as pointed out in Menzel and Morganti (2010), even though the statistical mapping preserves the consistency of the nonparametric estimator, the inversion problem is badly behaved. Due to the particular behavior of the inverted mapping at the extremes of the support, convergence of the estimated distribution to the true one fails to reach the standard parametric rate. Moreover, the convergence rate is affected by the number of bidders, N, up to the point that when the number of bidders diverges, the rate converges to zero and the magnitude of the sample size becomes irrelevant. In contrast to models where all the information about existing bids is available (Guerre, Perrigne, Vuong (2000)), when only transaction prices (or more generally, subsets of the data) are observable, the nonparametric estimator of the distribution of the data will converge slowly to the true one, affecting all successive computations (Menzel, Morganti (2010)).

Since the econometrician observes just an extreme (or a function of an extreme) of the parental distribution, the dataset will be *unbalanced*: observations on the lower part of the support are undersampled, and observations on the higher portion of the support are oversampled. As a consequence, inverting the distribution of the extreme imposes a downward bias around the left end of the support, and an upward

[1] We want to stress here the different roles taken by the *theoretical mapping* and the *statistical mapping* mentioned above. While the theoretical mapping links bids to individual valuations, the statistical mapping concerns the link between transaction prices (order statistics) and bids. From now on we are going to abstract from the first to focus on the second. The inversion problem that we are going to refer to is the one that goes from distribution of transaction prices to distribution of unobserved bids.

bias on the right end. All the quantiles are pushed to the right, and all the estimates based on them will suffer from that.

The problem is particularly evident when the number of participants in an auction grows up to infinity: the distribution of transaction prices collapses to a degenerate one, with mass point at the upper extreme of the support. Monte Carlo experiments show that, even when N is finite and small, the bias remains significative even in the presence of large samples.

While in principle it is possible to attenuate the problems on the right tail by smoothing the nonparametric estimators with an appropriate Nearest Neighborhood Estimator, in practice this will be difficult and time consuming. The problem on the left tail, on the other hand, cannot be solved by any method (parametric or nonparametric) unless we bring in more data: nothing can be learned where there is no information.

Given these considerations, we suggest an alternative, practical approach based on Extreme Value Theory (EVT): this parametric method relies on the well known convergence results about the extremes of a distribution (Gnedenko (1943)). Under very mild assumptions, the distribution of such extremes - appropriately normalized - converges uniformly to one of three possible distributions: the so called Extreme Value Distributions (EVD). When we rely on these results, it is possible to obtain estimates of functionals of practical interest such as the expected revenue, or the optimal reserve price, in two steps: first, we estimate the two normalizing constants by minimizing the distance between the normalized empirical distribution of the extreme and the corresponding EVD. Second, by applying a simple change of variable to the integral that expresses the expected revenue of the auction, we can rewrite everything in terms of EVDs and their transformations. Extreme value theory also implies a natural approximation for the underlying distribution of bids: the approximating distribution should be a Generalized Pareto.

We present results from Monte Carlo simulations, and show that the approximation method performs better than the nonparametric one, even in cases where the convergence of the extreme distribution to the limiting one happens at very slow rates[2]. Even though this extreme value estimator and its functionals suffer from the same limitations on the left tail as the nonparametric ones, they appear to be more robust. Moreover, since its relative advantage seem to hold also for those distributions for which the extreme value approximation is known to be poor, we can count with some confidence on the generality of this ap-

[2]This is, for instance, the case of the Normal distribution. The rate of convergence for extremes drawn from a normal distribution is of order $O(1/\log N)$

proach. In addition, this approach gets better as N increases, making the extreme value estimator more precise precisely when the nonparametric estimator gets worse. Finally, we observe that computation time is minimal, making this approach particularly attractive for applied works. Whenever we are observing auctions with limited data, it is important to realize that no estimator is going to perform well over this range: with this understanding, it is better to adopt an approach that is relatively more robust, general and theoretically justified. The extreme value estimator has the additional advantage of being computationally easy to perform.

EVT provides a general framework that can be adopted to all models where an order statistics is observed. For instance, an interesting extension to financial markets is the estimation of the unobserved distribution of valuations in multiunit auctions with uniform price.

The paper is structured in the following way: in Section 2 we are going to present the nonparametric estimator and discuss its behavior in the tails of the distribution. Section 3 introduces basic and general results from Extreme Value Theory. In Section 4 we are going to apply EVT to the auction framework, showing how it is possible to obtain results of interest relying only on EVDs and their transformations. Finally, Section 5 shows results from Monte Carlo simulations.

2 Nonparametric Identification and Estimation

We restrict our attention to symmetric independent private value (IPV) auction models, where only the transaction price is observed. For expositional purposes we are only going to present the case of second price auctions. This will allow us to focus on what we called the *statistical mapping* [3], and on the problems induced by the *statistical inversion*.

The typical dataset will consist of observations from n identical and independent auctions, where exactly N bidders have participated. We are going to use the capital letter N to denote the number of bidders, while lower case n will denote the size of the sample. We assume that N is exogenous: this condition is necessary in order to apply the statistical mapping and to establish consistency of the nonparametric estimator (Athey and Haile (2002), Haile and Tamer (2003)).

Every bidder $i = 1, \cdots, N$, submits an offer, b_i: the bid depends on

[3] In second price auction, it is an optimal strategy for the bidders to bid exactly their value. Therefore, the bids already provide the private values and no *theoretical* inversion is needed.

her own private value for the item, v_i, on the format of the auction, and on the game she is playing with all the other bidders. Private valuations are independently drawn from a common distribution, V. The distribution of the bids is denoted by F. The econometrician observes only the transaction price from each auction: this transaction price will be equal to an extreme of the parental distribution. For instance, in a second price auction, the transaction price corresponds to the $N-1$th order statistics (the second maximum) [4].
The kth order statistic of N independent bids $\{b_1, \cdots, b_N\}$ has distribution $G^{k:N}(z)$[5] , where

$$G^{k:N}(z) = \frac{N!}{(N-k)!(k-1)!} \int_0^{F(z)} t^{k-1}(1-t)^{N-k} \mathrm{d}t$$

Athey and Haile (2002) show that the mapping implicitly described above is always invertible: therefore it is possible to obtain the distribution of the bids, $F(z) = \phi(G^{k:N}(z), N)$, whenever we can estimate the distribution of the transaction prices, $G^{k:N}$ (*statistical inversion*). A simple nonparametric estimator for the distribution of the transaction prices is

$$\hat{G}^{k:N}(z) = \frac{1}{n} \sum_{j=1}^{n} \mathbb{I}[P_j \leq z]$$

which[6], by Glivenko-Cantelli theorem, converges almost-surely uniformly to the true distribution.
Following Haile and Tamer (2003) (Appendix A, *proof of Theorem 3*), the Continuous Mapping Theorem gives

$$\phi(\hat{G}^{k:N}, N) - \phi(G^{k:N}, N) = \phi(\hat{G}^{k:N}, N) - F(z) = o_p(1)$$

The convergence of the last quantity is also uniform in z: since the mapping ϕ is continuous over a compact space, it is also uniformly continuous. This establish uniform convergence. However, the mapping is not Lipschitz continuous: its derivative is unbounded at critical

[4] We define the kth order statistics in the following way: given a set of N bids, we order them starting from the smallest and ending with the largest. The set $\{b_1, \ldots, b_N\}$ denotes the ordered list. The first element of the list is the first order statistics, and corresponds to the minimum of the set. The Nth order statistics is the last element of the list, and corresponds to the maximum. The kth order statistics is simply the element in the kth position of the list.

[5] so that, for instance, the distribution of the second maximum, (or, the $(N-1)$th order statistics) is $G^{N-1:N}(z) = N(N-1)\left[\frac{F(z)^{N-1}}{N-1} - \frac{F(z)^N}{N}\right] = NF(z)^{N-1} - (N-1)F(z)^N$.

[6] The symbol $\mathbb{I}[A]$ denotes the indicator function, which assumes value equal to 1 when A is true, and equal to 0 when A is false. P_j denotes the transaction price from the jth auction.

points of the support, $\{\underline{z}, \overline{z}\}$[7].
To exemplify this point, suppose that the bidders are participating to a Second Price auction: the dominant strategy for every player is to submit her private valuation. The transaction price is therefore the second maximum of the private values. The mapping ϕ is defined implicitly by

$$G = N(N-1)\left[\frac{\phi(G,N)^{N-1}}{N-1} - \frac{\phi(G,N)^{N}}{N}\right]$$

By Implicit Function Theorem we can obtain its derivative

$$\phi'(G,N) = 1/\left\{N(N-1)\phi(G,N)^{N-2}[1-\phi(G,N)]\right\}$$

which is unbounded on the lower tail of the distribution, where G goes to zero, and on the right end, where G goes to 1.
This creates a serious problem in the estimation, since even small biases will be magnified in the neighborhoods around these points. It is possible to see that the problem on the lower tail becomes more severe as N increases, while it is attenuated on the right end of the support. The convergence of the estimated distribution to the true one will be slow and dependent on the number of bidders: when $N = 1$, the problem is not different from usual ones. However, when N grows to infinity, identification is lost: the distribution of the extreme degenerates to a mass point at the upper bound of the support , and the rate of convergence becomes equal to zero. This means that when the number of bidders is high we should expect nonparametric estimates to be a poor description of the behavior of the lower tail of the distribution. The distribution of the bids is *irregularly identificatied*, in the sense that the parametric rate of the preliminary estimator is lost. Similar problems have been analyzed by Khan and Tamer (2009).

Remark 1 *The rate of convergence of the nonparametric estimator $\phi(\hat{G}^{k:N}(z), N)$, with $F(z) \in (0,1)$, decreases in N, and approaches the value of zero as N goes to infinity.*

Proof Remark 1 For the kernel estimator defined above,

$$\sqrt{n}[\hat{G}^{k:N}(z) - G^{k:N}(z)] \longrightarrow \mathcal{N}(0, \sigma^2(z))$$

where $\sigma(z)^2 = F(z)[1-F(z)]$. Then, using the Delta Rule

$$\sqrt{n}[\phi(\hat{G}^{k:N}(z), N) - \phi(G^{k:N}(z), N)] \longrightarrow \mathcal{N}(0, \sigma^2(z)[\phi'(G^{k:N}(z), N)]^2)$$

[7]For $k = 1$, the derivative is unbounded only at the left end of the support.

We need to show that $\phi'(G,N)$ diverges to infinity as N increases. From the implicit definition of the mapping, we obtain

$$\phi'(G,N) \equiv \frac{\partial \phi}{\partial G} = \frac{1}{\frac{N!}{(N-k)!(k-1)!}\phi(G,N)^{k-1}(1-\phi(G,N)^{N-k})}$$

we restrict our attention to the class of problems where $k/N \longrightarrow 1$[8]
First we show that $G(z,N)$ is decreasing in N in the lower tail of the distribution[9] . When $k/N \longrightarrow 1$, we can denote $\frac{N!}{(N-k)!(k-1)!}$ as $P(N, q+1)$, a polynomial in N of degree $q+1$, where $q = N-k$. Then

$$\frac{\partial G}{\partial N} = \tilde{P}(N,q)\int_0^F t^{k-1}(1-t)^{N-k}\mathrm{d}t +$$

$$+P(N,q+1)\int_0^F t^{k-1}(1-t)^{N-k}\ln(1-t)\mathrm{d}t < 0$$

The limit of the sequence is zero. In fact,

$$G^{k:N}(z) \leq P(N,q+1)\int_0^{F(z)} t^{\left(\frac{k-1}{N}\right)N} dt$$

The argument of this integral is continuous over a compact set, therefore it is uniformly continuous. Riemann integrability applies to the limit of the sequence,

$$\lim_{N\longrightarrow\infty}\int_0^{F(z)} t^{\left(\frac{k-1}{N}\right)N} dt = \int_0^{F(z)} \lim_{N\longrightarrow\infty} t^{\left(\frac{k-1}{N}\right)N} dt = 0$$

for z such that $F(z) < 1$, and for $k/N \longrightarrow 1$. The integral falls to zero fast and dominates the explosive effect of the polynomial.
Since $G(z,N)$ falls to zero as N increases to infinity when z belongs to a lower tail of the distribution, $\phi(G,N)$ must fall to zero as well, in order to balance expression (1). This makes the derivative ϕ' unbounded.

■

[8]We focus on the higher extremes of the distribution: the first maximum, the second maximum and so on. We do not consider the lower extremes of the distribution: the minimum, the second minimum... This assumption is consistent with the framework that we are using: auctions models will be involved with the former type of extremes.

[9]What we mean by lower tail of the distribution depends on the particular extreme that we are considering: for instance, if what we are considering is the maximum, the relevant range becomes the full support of the distribution, excluding the upper extreme.

A secondary problem affects the precision of nonparametric estimators $\hat{G}(z)$: the typical dataset is necessarily unbalanced. Higher values of the support are oversampled while lower values are undersampled to the point that entire portions of the lower tail might not even be observed in finite samples. All the measures based on our estimates will be distorted accordingly: for instance, both expected revenue of the auction and reservation price will be systematically upward biased. This problem becomes worse as N grows but it should fade as sample size increases. However, Monte Carlo simulations show that the increase in N dominates the effects of an increase in n: as we show in the next chapters, even with $N = 5$ and $n = 50,000$ the bias stays significant.
An appropriate Nearest Neighborhood estimator could be used to estimate the function. However, we should expect the size of the Neighborhood to increase towards the left of the support, eventually becoming infinitely large. Even though it could be possible to correct for the oversmoothing of high values, in general we will not be able to compensate for the extreme undersmoothing that might occur around the minimum of the support.
Nonparametric estimators perform poorly on both tails of the distribution: the bias fades slowly, and in general affects all the measures of interest. Appropriate smoothing procedures may help reducing the bias in the upper tail, but in general they will not solve the more pressing problem that occur in finite samples. Moreover, such procedures require an appropriate calibration of a width parameter in order to produce effective results, and this can be difficult and time consuming. In the next sections we are going to introduce a new approach to estimation that will require minimum computation time: we will show that such parametric method produces better results than the nonparametric one. But in order to discuss the method, we need to introduce some basic concepts about Extreme Value Theory.

3 Extreme Value Theory

The fundamental insight of EVT is given by the following observation. If the distribution of the maximum of N independent draws from F, appropriately normalized, converges to a distribution function G as N goes to infinity, then G must be one of the following three:

$$
\begin{aligned}
G_1(z) &= \exp(-z^{-\alpha}), \quad z > 0 \quad \text{(Frechet)} \\
G_2(z) &= \exp(-(-z)^{\alpha}), \quad z \leq 0 \quad \text{(Weibull)} \\
G_3(z) &= \exp(-e^{-z}), \quad z \in \mathbb{R} \quad \text{(Gumbel)}
\end{aligned}
$$

Formally, let $\mathcal{P}$ be a probability measure with distribution function F. Denote with $Z_{i:N}$ the ith order statistics for the sample of size N.

Theorem 1 (Gnedenko (1943)) *If there exist real numbers $a_N > 0$ and b_N, such that $\mathcal{P}^N\left(\frac{Z_{N:N}-b_n}{a_n} \leq z\right)$ [10]tends to some nondegenerate limit $G(z)$ then, either $G = G_1$, or $G = G_2$, or $G = G_3$*

If it is possible to find a shifting parameter and a scaling parameter, such that the normalized distribution of the maximum converges, then the limiting distribution belongs to the Extreme Value family. The theorem grants a natural parametric approximation for the distribution of the maximum, up to two normalizing parameters.
Gnedenko (1943) also gave necessary and sufficient conditions for F to belong to the domain of attraction of any of the above limits (denoted $F \in \mathcal{D}(G_h)_{h=1,2,3}$). Von Mises (1936) derived a set of sufficient conditions which are more easily testable. Assume that F has a positive derivative, f, over its support; then F belongs to the domain of attraction of G_1, G_2 or G_3, if conditions 1, 2 or 3 are satisfied, respectively.

$$\lim_{z \longrightarrow \infty} \frac{zf(z)}{1-F(z)} = \alpha \tag{1}$$

$$\lim_{z \longrightarrow \overline{z} \in \mathbb{R}} \frac{(\overline{z}-z)f(z)}{1-F(z)} = \alpha \tag{2}$$

and

$$\int_{-\infty}^{\overline{z}} (1-F(u))\mathrm{d}u < \infty$$

$$\lim_{z \longrightarrow \overline{z}} f(z) \int_{z}^{\overline{z}} (1-F(u))\mathrm{d}u/(1-F(z))^2 = 1 \tag{3}$$

A useful sufficient condition for condition 3 is the following

$$\exists z \in (0,\infty) \quad \lim_{z \longrightarrow \overline{z}} \tfrac{f(z)}{1-F(z)} = c$$

It is easy to verify that the uniform distribution satisfies the first Von Mises condition: therefore the maximum of N independent draws from a uniform distribution converges to G_1. Similarly, we can show that the normal and the exponential distributions belong to the domain of attraction of G_3. More generally, it is possible to show that the class

[10] $\mathcal{P}^N$ denotes the N-fold independent product of $\mathcal{P}$

of distributions that satisfy the Von-Mises conditions is wide, and includes all known analytical distributions.
More interestingly for our purposes, Falk and Marohn (1993) rewrite the von Mises conditions in terms of convergence of the underlying distribution to a corresponding Generalized Pareto Distribution ($gPds$). The gPd family plays an important role in Extreme Value Theory: convergence of the maximum to an EVD is equivalent to convergence of the underlying distribution to a Pareto distribution. We will discuss this point later in the paper.
Falk (1985) shows that the von Mises conditions imply pointwise convergence of the density f^N to g^N as N goes to infinity [11]. This, by virtue of Scheffé's Lemma, in turns entails its uniform convergence over all Borel sets (convergence in Total Variation). In the Appendix we present uniform bounds for the largest order statistics.

We want to capture the attention of the wandering reader on the following crucial fact: the rate of convergence of $\sup_x |F^N(a_N x + b_N) - G(x)|$ to zero depends crucially on the particular distribution F at hands: for instance, it is of order $O(1/N)$ for the negative exponential case , and of order $O(1/\log N)$ for the normal case (see the Appendix) [12]. The fastest possible convergence rate is actually of order $O(1/N)$ and is achieved by members of the gPd family. Since we don't have knowledge of the underlying distribution we can only make conjectures about the quality of the approximation: since the normal is known to converge at low rates, we will use it as a "lower reference" for our simulations. We obtain satisfactory results under the gaussian assumption, as a consequence we are optimistic about the robustness of the estimator.
The results presented so far are not exclusive of the first maximum of a sample of independent draws: in fact, they extend to the whole joint distribution of the extremes. Define $m = N - k + 1$; if F satisfies one of the Gnedenko conditions, then $F^{k:N}(z)$ converges uniformly to $G_h^{(m)}(z) = G_h(z) \sum_{i=0}^{m-1} \frac{1}{i!}[-\log G_h(z)]^i$, where $h = 1, 2, 3$ indicates the appropriate limiting EVD. For instance, for the case of the second maximum ($k = N - 1$, or, equivalently, $m = 2$), the limiting distribution becomes

$$G_h^{(2)}(z) = G_h(z)[1 - \log G_h(z)]$$

[11] The result presented in Falk (1985) extends to the generic kth order statistics. We denote by $G^{k:N}$ the Extreme Value limit distribution for the kth order statistics. Then, if one of the von Mises conditions is satisfied, $f^{k:N}$ converges pointwise to $g^{k:N}$, for any possible k.

[12] Finding the normalizing constants a_N, b_N is not a straightforward task. In practice, for $F \in \mathcal{D}(G_3)$, we might start with the following guess: b_N that solves $F(b_N) = 1 - 1/N$.

Falk (1989) shows that the best rate of convergence of extremes is of order $O(m/N)$, and is attained by distributions belonging to the Generalized Pareto family.

4 Extreme Value Theory in the Estimation of Auction Models

We can make use of the results presented in the previous section, and approximate the distribution of the extreme with the appropriate EVD. As discussed by Falk (1985), this corresponds to approximating the parental unobserved distribution with a gPd. We are going to show that objects of interest such as Expected Revenue and Reservation Price can be easily obtained through a simple transformation.
From now on we are going to assume that F possesses a derivative f. The expected revenue for First Price and Second Price auctions, corresponding to the expectation of the second maximum valuation, is given by the following integral (see, for instance, Krishna (2002))

$$\mathbb{E}[R|N] = \int_0^{\overline{w}} N(N-1)xF(x)^{N-2}[1-F(x)]f(x)\mathrm{d}x$$

We want to stress here that, for the simple case we are considering, it is not necessary to compute the integral in order to obtain the expected revenue of the auction: for this purpose it is enough to find the expected value of the transaction prices. This expected value does not suffer from bias and should therefore be preferred in estimation. However, for expositional purposes, we are going to refer to the integral as a benchmark for the heavy bias that affects the nonparametric estimator. More generally, recovery and use of the distribution F, and computation of the integral, will be required in order to compute the optimal reservation price and to perform counterfactual analysis. For this reason we believe to be important to understand how and with what magnitude the nonparametric estimator can affect our analysis. Since F is unknown we cannot evaluate the integral. For simplicity, we will focus on Second Price auctions, so that the distribution of the bids corresponds to the distribution of the private values. We are going to show that the integral can be expressed and estimated in terms of EVDs, with no significant loss in precision.

Theorem 2 (Expected Revenue) *If there exists $a_N > 0$ and b_N*

such that $\mathcal{P}\left(\frac{Z_{N:N}-b_N}{a_N} \leq z\right)$ *converges to* $G(z)$, *then*

$$\mathbb{E}[R|N] \approx \int_{-\frac{b_N}{a_N}}^{\frac{\overline{w}-b_N}{a_N}} (N-1)(a_N t + b_N)[-\log G(t)^{\frac{1}{N}}]g(t)\,dt$$

For instance, for the class of distributions $F \in \mathcal{D}(G_3)$, the expression becomes

$$\mathbb{E}_3[R|N] \approx \int_{-\frac{b_N}{a_N}}^{\frac{\overline{w}-b_N}{a_N}} (N-1)(a_N t + b_N)\frac{e^{-2t-e^{-t}}}{N}\mathrm{d}t$$

We construct the proof through a sequence of simple Lemmas that follow from Falk (1985).

Lemma 1 $F^{N-2}(a_N t + b_N) \approx G(t)$

This is simply a rewriting of the assumption of the theorem.

■

Lemma 2 $[1 - F(a_N t + b_N)] \approx -\log G(t)^{\frac{1}{N}} + O(h(N))$

***Proof*:** if F belongs to the domain of attraction of G then

$$\begin{aligned} F^N(a_N t + b_N) &\longrightarrow G(t) \iff \\ N \log F(a_N t + b_N) &\longrightarrow \log G(t) \iff \\ N[F(a_N t + b_N) - 1] &\longrightarrow \log G(t) \iff \\ N[1 - F(a_N t + b_N)] &\longrightarrow -\log G(t) \iff \\ \frac{1 - F(a_N t + b_N)}{-\log G(t)^{\frac{1}{N}}} &\longrightarrow 1 \end{aligned}$$

■

Lemma 3 $a_N f(a_N t + b_N) \approx \frac{1}{N}\frac{g(t)}{G(t)}$

***Proof*:** Since F possesses a derivative f near the right end of the support, the previous condition implies

$$\frac{a_N f(a_N \theta + b_N)}{\frac{1}{N}\frac{g(\theta)}{G(\theta)}} = \frac{F(a_N t + b_N) - F(a_N y + b_N)}{[-\log G(t)^{\frac{1}{N}}] - [-\log G(y)^{\frac{1}{N}}]} \longrightarrow 1$$

for some $\theta \in (t, y)$.

■

Proof of Theorem 2 The proof of the theorem is concluded by performing a simple change of variable in the original integral, $t = (x - b_N)/a_N$, and applying the approximations presented in the previous lemmas.

$$\mathbb{E}[R|N] = \int_{-\frac{b_N}{a_N}}^{\frac{\overline{w}-b_N}{a_N}} N(N-1)(a_N t + b_N)F(a_N t + b_N)^{N-2} * \\ *[1 - F(a_N t + b_N)]f(a_N t + b_N)a_N dt \approx$$

$$\approx \int_{-\frac{b_N}{a_N}}^{\frac{\overline{w}-b_N}{a_N}} N(N-1)(a_N t + b_N)G(t)[-\log G(t)^{\frac{1}{N}}]\frac{g(t)}{NG(t)}\mathrm{d}t =$$

$$= \int_{-\frac{b_N}{a_N}}^{\frac{\overline{w}-b_N}{a_N}} (N-1)(a_N t + b_N)[-\log G(t)^{\frac{1}{N}}]g(t)\mathrm{d}t$$

■

The approximation allowed by this theorem does not depend on the unknown distribution F, except through an appropriate choice of normalizing constants: the new expression depends entirely on the normalizing constants a_N, b_N and on the EVD, G. Procedures that test for the particular type of Extreme Value distribution to be used have long existed in the literature. We present here a new and more natural approach, that allows us to simultaneously estimate the normalizing parameters and test for the appropriate approximation.

4.1 Estimation

The normalizing constants can be estimated through some standard minimum distance (MD) criterion[13]. A widely used criterion is the Cramér-von-Mises, which uses the integral of the squared difference between the empirical and the estimated distribution functions. Among

[13]Let $\{P_\theta\}$ be a family of probabilities indexed by θ, and let μ be a metric between probabilities. Let $\hat{\theta}(P)$ be the corresponding *minimum distance functional*, i.e., the solution to

$$\mu(P, P_{\hat{\theta}}) = \min_{\theta} \mu(P, P_\theta)$$

The MD functional is consistent and robust over μ-neighborhoods (Rao-Schuster-Littel (1975), Parr-Schucany (1979), Donoho-Liu (1988)) the estimators based on non-Hilbertian[14] metrics, the most common is the Kolmogorov-Smirnof

$$Q_{n,N}(a_N,b_N)=\sup_{x_n}\left|\hat{F}^{k:N}\left(\frac{x_n-b_N}{a_N}\right)-G^{(m)}(x_n)\right|$$

$$\{\hat{a}_N,\hat{b}_N\}=\arg\min_{a_N,b_N}\sup_{x}\left|\hat{F}^{k:N}\left(\frac{x-b_N}{a_N}\right)-G^{(m)}(x)\right|$$

where $m=N-k+1$. It is well known that Kolmogorov-Smirnof distance immediately provides a test for goodness of fit. So, in order to find the relevant approximating distribution, we simply need to compare the Kolmogorov-Smirnof distance under the three EVD and test for the best approximation. This procedure is simple and avoids having to compute the maximum likelihood estimator of the generalized extreme value distribution.
Consistency arises naturally thanks to Gnedenko's theorem (see for instance Hayashi (2000)).

Assumption 1 The normalizing constants $(\hat{a}_N,\hat{b}_N)$ belong to a compact space, $\Theta\subset\mathbb{R}^2$

Let $Q_{n,N}(a_N,b_N)=\hat{F}^{k:N}\left(\frac{x_n-b_N}{a_N}\right)$ be the objective function: since $Q_{n,N}(a_N,b_N)$ is a continuous and measurable function in (a_N,b_N) for all the data, a measurable function exist that solves the minimization above.
Let $Q_0(a,b)=\lim_{N\longrightarrow\infty}F^{k:N}\left(\frac{x_n-b_N}{a_N}\right)=G^{(m)}(x_n)$.

Assumption 2 Gnedenko Theorem holds. As an implication of that

1. (*identification*) Q_0

4.2 The Optimal Reserve Price

Using a similar approach we can estimate the optimal Reserve Price (RP) of the auction, given a specific Value for the seller, x_0[15]: it is

[14]By *Hilbertian* we mean based on a quadratic measure of deviation

[15]The expected revenue with reserve price is equal to

$$\max_{\theta}\mathbb{E}[R|N,RP]=\int_{RP}^{\overline{w}}N(N-1)xF(x)^{N-2}[1-F(x)]f(x)dx+$$
$$+x_0F(RP)^N+N(RP)[1-F(RP)]F(RP)^{N-1}$$

possible to perform a numerical search over the parameter $\theta = \frac{RP - b_N}{a_N}$ that maximizes the expected revenue

$$\max_{\theta} \mathbb{E}[R|N,\theta] = \int_{\theta}^{\frac{\overline{w}-b_N}{a_N}} (N-1)(a_N t + b_N)[-\log G(t)^{\frac{1}{N}}]g(t)dt +$$
$$+ x_0 G(\theta) + N(a_N\theta + b_N)[\log G(\theta)^{\frac{1}{N}}]G(\theta)$$

As a final remark about the extension to which Extreme Value Theory can be used to replace Nonparametric estimation, a close look at Lemma 2 suggests the possibility to approximate the right tail of the distribution[16] F with a Generalized Pareto distribution (see Pickands (1975)). The class of gPd is composed by the following three normalized families,

$$\begin{aligned} P_1(x) &= 1 - x^{-\alpha}, && x \geq 1, \alpha > 0 \\ P_2(x) &= 1 - (-x)^{\alpha}, && x \in [-1,0], \alpha > 0 \\ P_3(x) &= 1 - e^{-x}, && x \geq 0 \end{aligned}$$

It is easy to see that the uniform distribution belongs to the first family and the negative exponential belongs to the third family. Lemma 2 justifies the following approximation,

$$F(a_N t + b_N) \simeq 1 + \log G_h(t)^{\frac{1}{N}} == \begin{cases} P_1(tN^{\frac{1}{N}}) & \text{for } h = 1 \\ P_2(tN^{-\frac{1}{N}}) & \text{for } h = 2 \\ P_3(t + \log N) & \text{for } h = 3 \end{cases}$$

5 Monte Carlo Simulations

In this section we are going to present some results from Monte Carlo simulations in support of the theory presented in the previous chapters. In order to simplify the discussion we are going to focus on the case of Second Price auctions: this implies that the bids drawn are also the valuations of the bidders. We are going to show and discuss the results for two distributions, chosen for their opposite N-asymptotic behavior: the first distribution is a Normal, with parameters $\mu = 10$ and $\sigma = 2$[17], while the second distribution is a Negative Exponential with parameter $\lambda = 0.2$. As discussed above, extremes of a normal distribution converge at a slow rate to the Gumbel family, while the negative exponential possesses the highest possible rate of convergence. We are considering asymptotic behavior by letting both N,

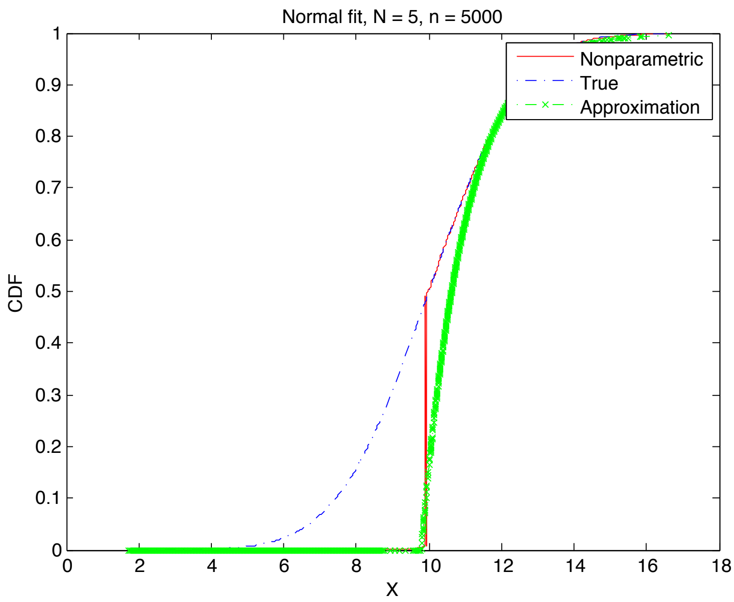
[16]The relative magnitude of this *right tail* depends on N and on the particular parental distribution F.

[17]The specific choice of the parameter does not affect the results

the number of bidders, and n, the number of auctions (i.e. the sample size), increase. In particular, N can take values 5, 50, 100, 150; while n values 5,000 and 50,000.
We estimated the normalizing constants using the Kolmogorov-Smirnof measure. We compared them with estimates obtained with the Cramér-von Mises criterion and found no significant differences. A useful outcome of The Kolmogorov-Smirnof criterion is the availability of a test for the goodness of fit: in all simulations, the normalized empirical distribution is not significally different from the corresponding EVD. We produce standard errors for expected revenue and reserve price through a Bootstrapping procedure.
Figure 1 up to Figure 8 provide a graphical representations of the goodness of fit of the nonparametric estimator and of the estimator based on Extreme Value Theory: since EVT offers an approximate result, we are going to call this estimator the Approximate distribution [18]. The Approximate distribution is represented graphically by the green curve; the red curve represents the nonparametric estimator. The blue dotted curve is the true CDF. The figures immediately illustrate three points: first, as N, the number of bidders, increases the bias of the nonparametric estimator rises. Second, the size of the dataset seems to have very little effect on the quality of the estimates. Finally, while for the case of the Negative Exponential the Approximate estimator performs incredibly well , when we analyze the case of the normal distribution the fit is much less satisfactory : as the number of bidders increases, EVT delivers better results than the nonparametric estimator, but the bias in the lower tail still appears to be relevant.

Next, we are going to show how the different approaches perform in terms of prediction of the expected revenue from the auction: as before, we are presenting results only for the cases of a negative exponential and of a normal distribution (see Table 1 and Table 2, respectively). As the number of bidders increases from 5 to 150, the expected revenue from the auction increases correspondingly: this is intuitive, since the expectation of receiving a higher bid increases with the number of participants in the auction. EVT provides a good estimate of the expected revenue: the bias from the Approximation is high for small number of bidders, but it rapidly decreases. The sample size affects the precision of the estimation of the normalizing constants, $\hat{a}_N, \hat{b}_N$, and, with them, the precision of the fit. The nonparametric estimator however is severely affected by the number of bidders: for both cases it starts form 60% and increases above 1,000 % for the

[18] As discussed in the previous section, the Approximate distribution is an appropriately normalized gPd



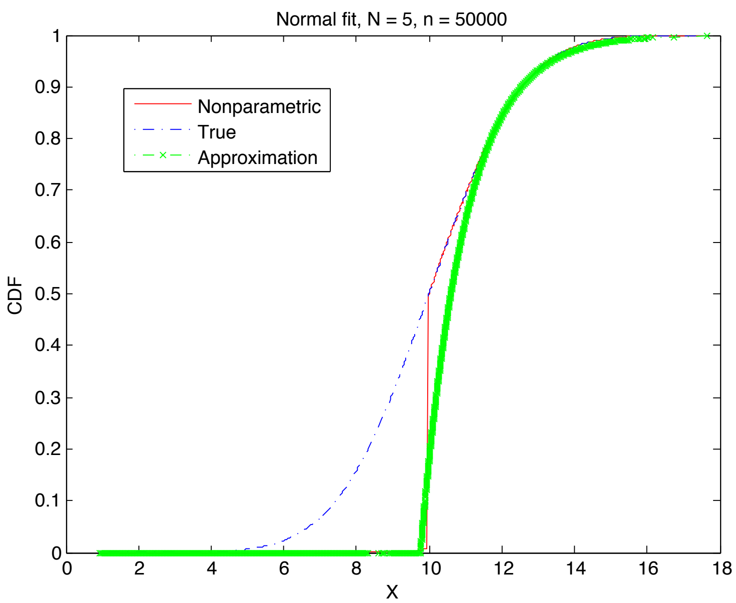
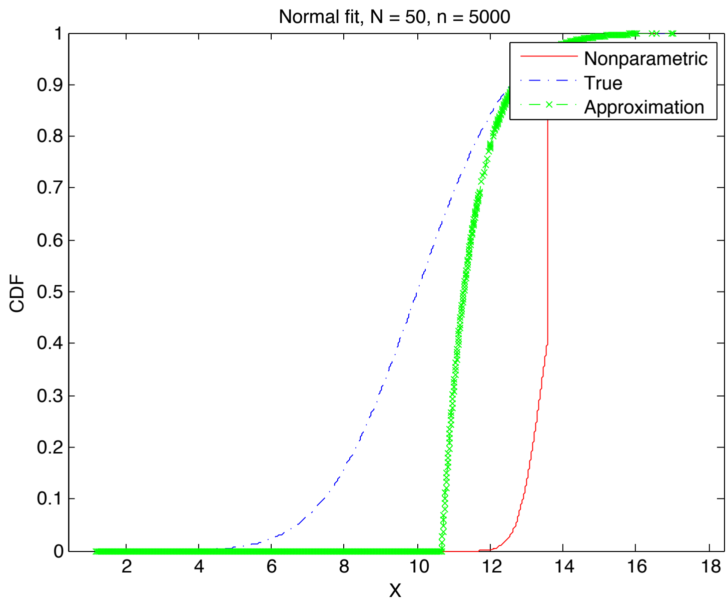


Figure 1: CDF estimation: Normal, $\mu = 10, \sigma = 2$, N = 5.



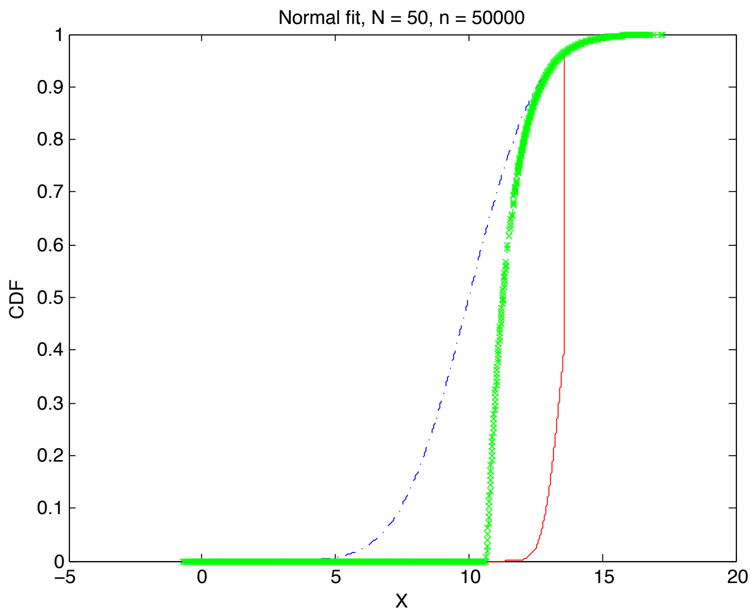


Figure 2: CDF estimation: Normal, $\mu = 10, \sigma = 2$, N = 50.

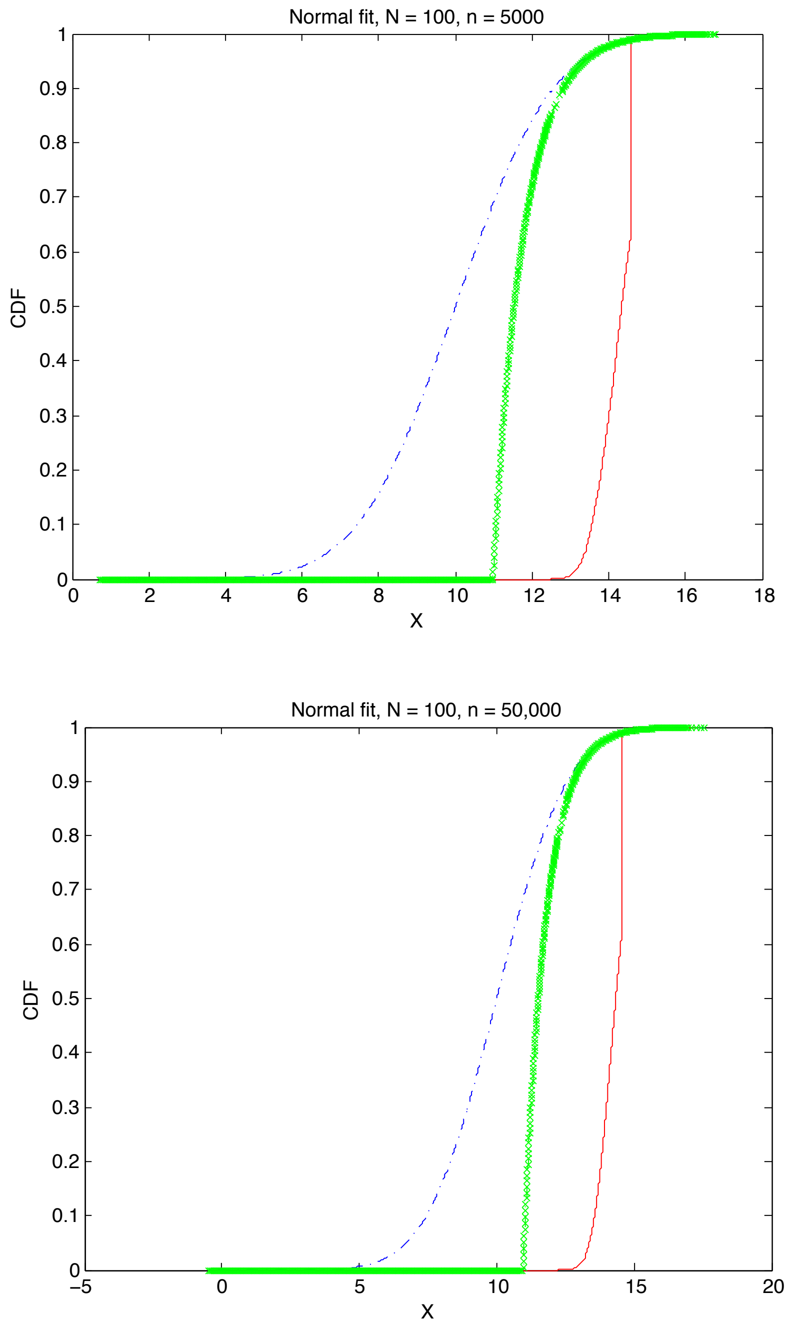


Figure 3: CDF estimation: Normal, $\mu = 10, \sigma = 2$, N = 100.

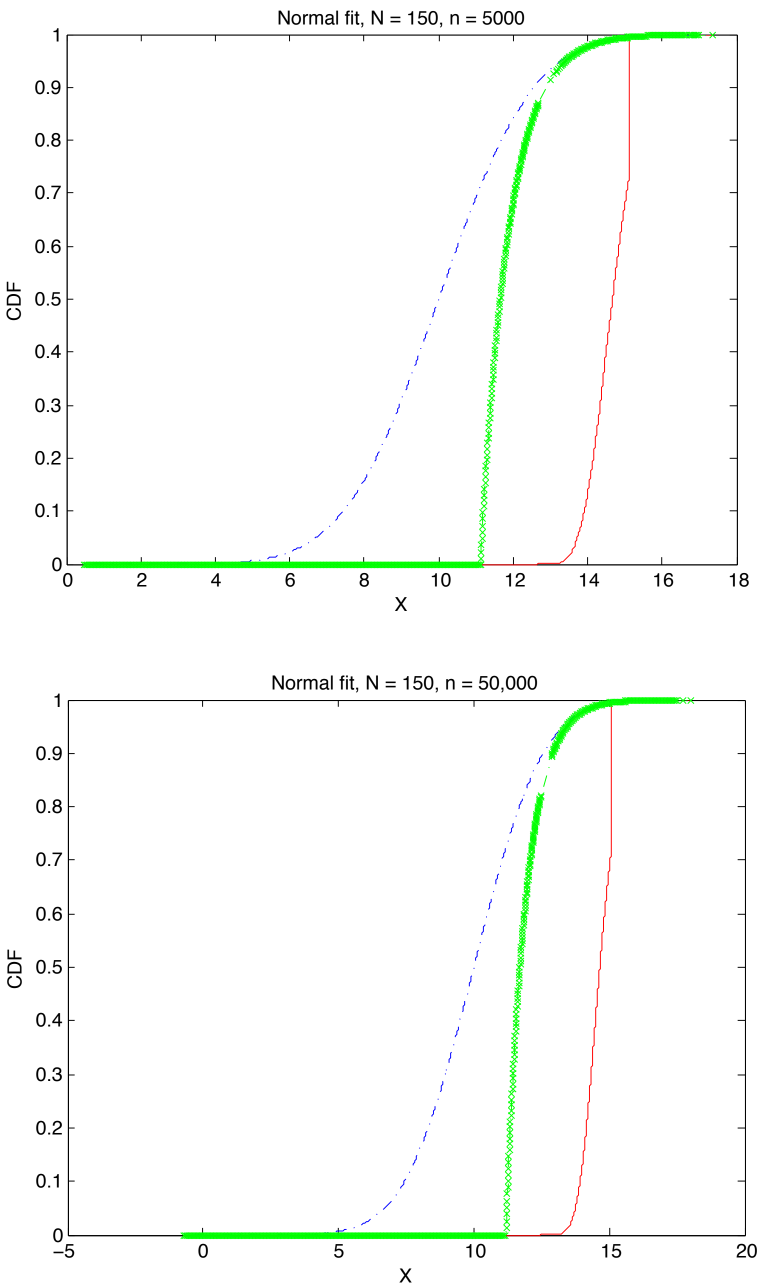


Figure 4: CDF estimation: Normal, $\mu = 10, \sigma = 2$, N = 150.

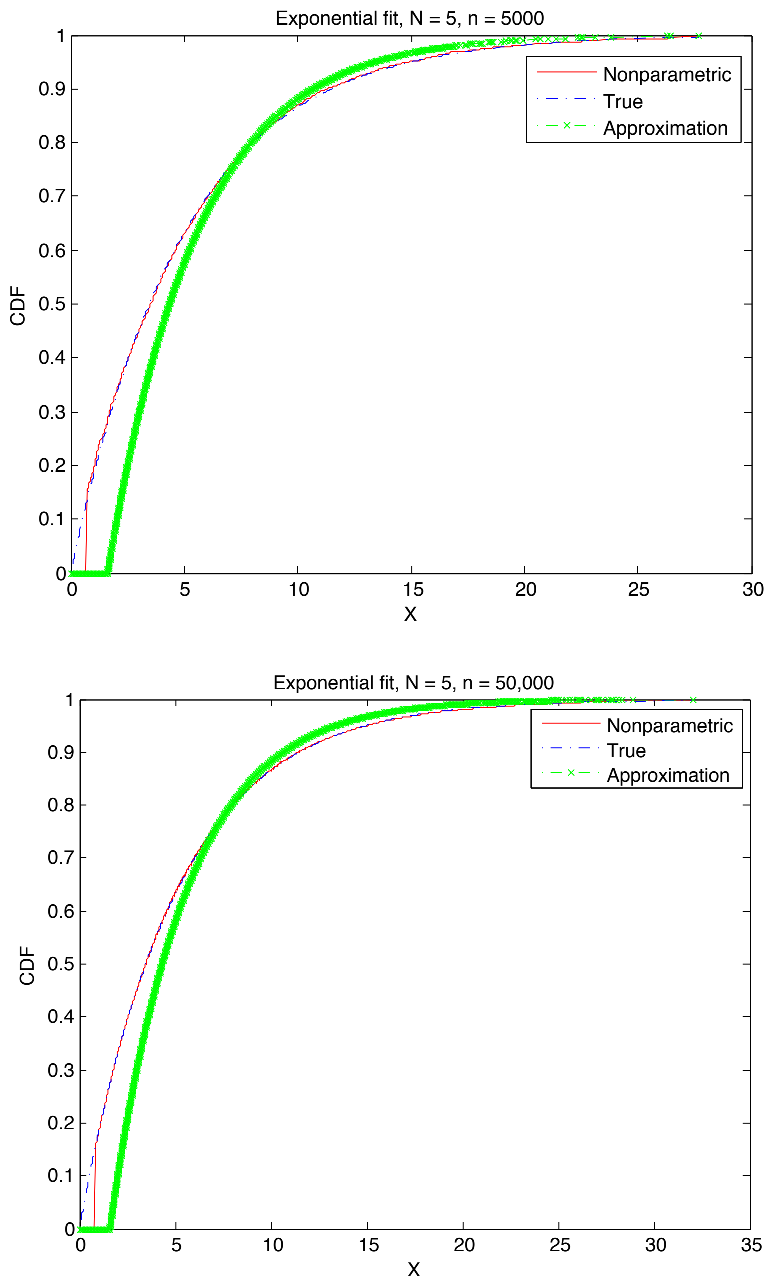
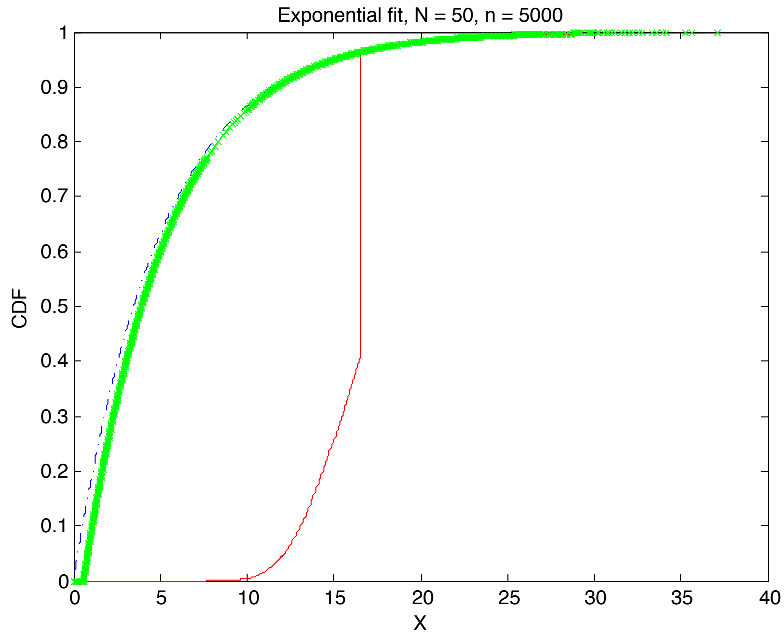


Figure 5: CDF estimation: Negative Exponential, $\lambda = 0.2$, N = 5.



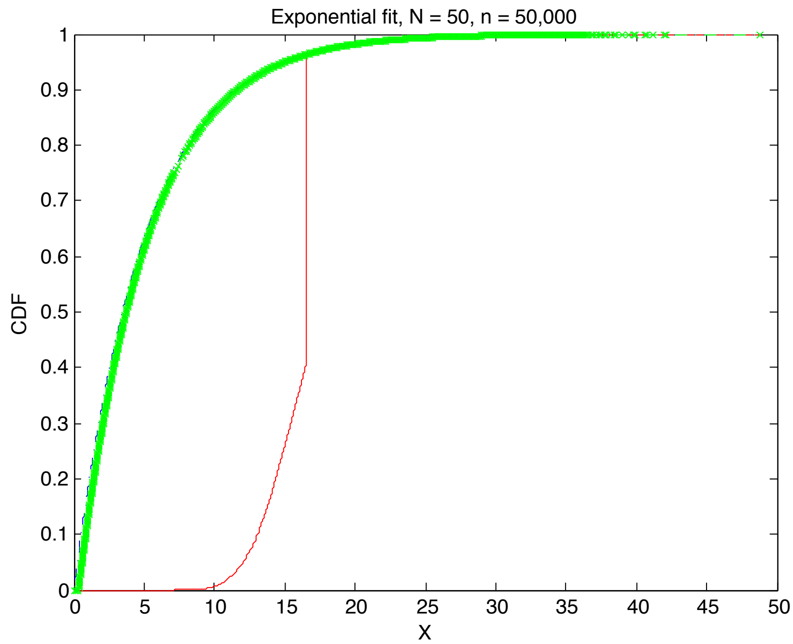


Figure 6: CDF estimation: Negative Exponential, $\lambda = 0.2$, N = 50.

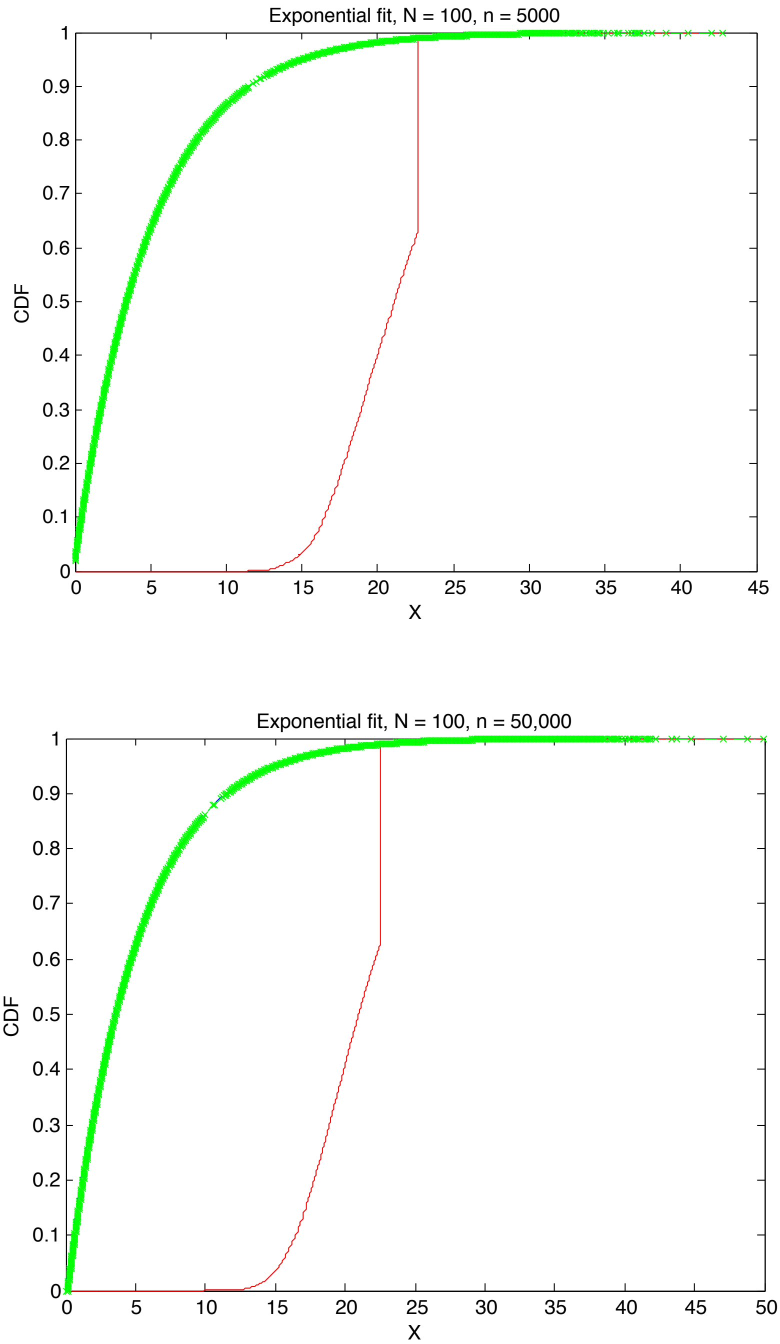


Figure 7: CDF estimation: Negative Exponential, $\lambda = 0.2$, N = 100.

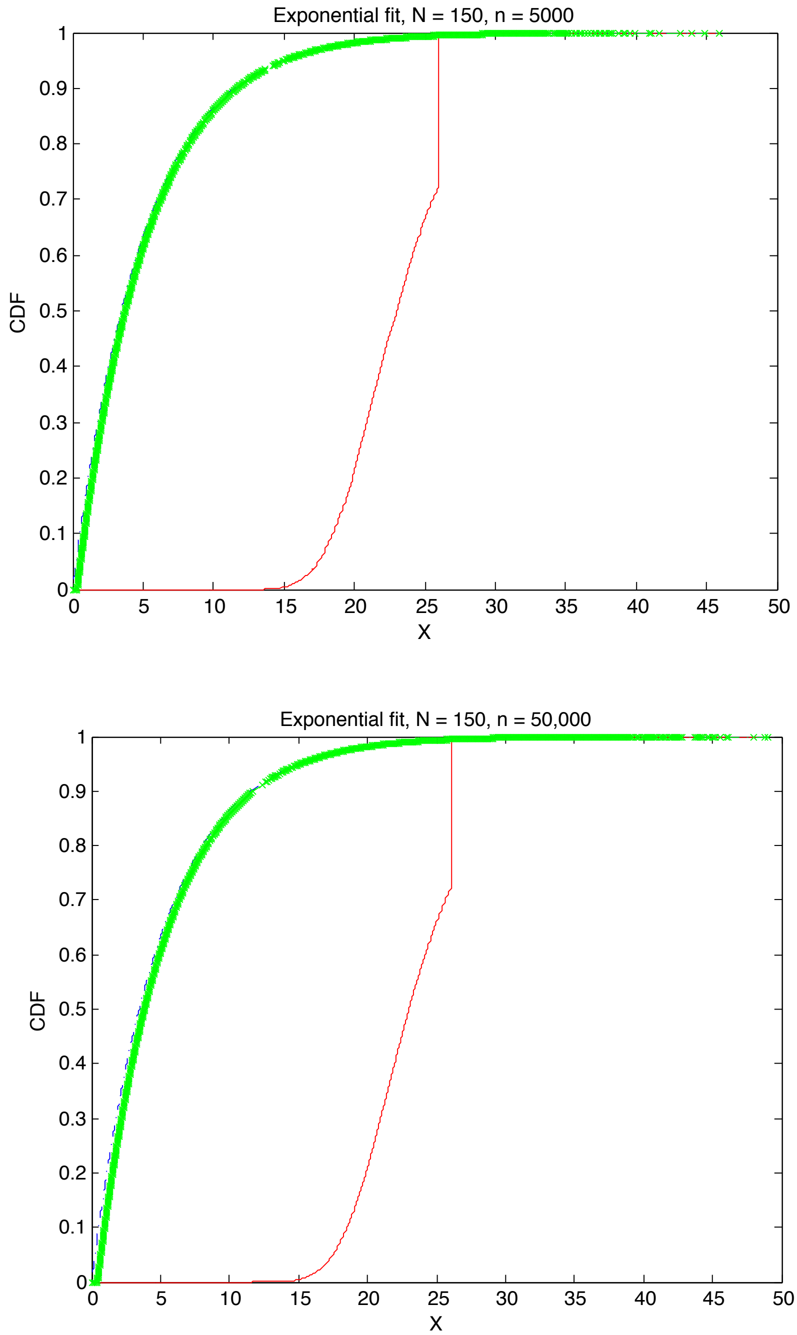


Figure 8: CDF estimation: Negative Exponential, $\lambda = 0.2$, N = 150.

Num. bidders	n. Auctions		True E. Rev.	Bias Approx.	Bias Nonp.
5	200		6.3519	-22.54%	66.29%
50	200		17.4549	-3.01%	1,244%
100	200		20.8246	-1.90%	2,511%
5	5000		6.3519	-19.56%	63.49%
50	5000		17.4549	-1.84%	912.67%
100	5000		20.8246	-0.88%	936.58%

Table 1: Negative Exponential, $\lambda = 0.2$: Prediction of the Expected Revenue

Num. bidders	n. Auctions		True E. Rev.	Bias Approx.	Bias Nonp.
5	200		11.0016	-22.48%	62.12%
50	200		13.718	-2.36%	1,185%
100	200		14.2864	-1.90%	2,440%
5	5,000		11.0016	-22.13%	60.88%
50	5,000		13.718	-2.04%	822.45%
100	5,000		14.2864	-1.03%	866.72%

Table 2: Normal Distribution, $\mu = 10, \sigma = 2$: Prediction of the Expected Revenue

exponential, and 920% for the normal. Increasing the sample size up to 50,000 observations seems to have a minor benefit on the estimates. Again, EVT performs slightly better when the parental distribution is the negative exponential, but the difference in the fit is small. The nonparametric approach favors distributions with slow rate of convergence, like the normal one; but still drastically underperforms compared to EVT.

Last, we are going to focus on the optimal Reserve Price of the auction when the seller has an outside value equal to x_0 (we assume $x_0 = 1.25$ for the negative exponential case, and $x_0 = 10.8$ for the normal case). Tables 3 and 4 present results for the two distributions. A theoretical result from Auction Theory states that the true reserve price is not affected by the number of bidders, nor by the sample size: within the boundaries of numerical computation, the Monte Carlo exercise supports the result. For comparison purposes, we have fixed the true reserve price to an average of the computed values.
The number of bidders however affects the optimal reserve price computed under the two approaches: when the Approximate distribution is used, the precision increases with N. As a consequence, the bias of the *induced* expected revenue, the true expected revenue given the

Num. bidders	n. Auctions	True Res. Val	App. Res. Val	Nonp. Res.Val	True E. Rev.	Bias Approx.	Bias Nonp.
5	5000	2.5306	2.286 [0.0013]	2.5581 [0.1003]	4.9354	-0.11%	-0.26%
50	5000	2.5306	2.3474 [0.0180]	3.2848 [0.357]	5.3238	-0.00%	-4.10%
100	5000	2.5306	2.458 [0.0012]	3.3704 [0.5032]	5.4149	<10^-3	-8.79%
150	5000	2.5306	2.5256 [0.0042]	3.4527 [0.6789]	5.9499	<10^-3	-10.00%
5	50000	2.5306	2.289 [0.0013]	2.5127 [0.074]	4.9354	-0.11%	-0.25%
50	50000	2.5306	2.3196 [0.0139]	3.1894 [0.2319]	5.3238	-0.00%	-4.03%
100	50000	2.5306	2.458 [0.0012]	3.2263 [0.4721]	5.4149	<10^-3	-7.34%
150	50000	2.5306	2.5256 [0.0042]	3.2915 [0.5089]	5.9499	<10^-3	-10.16%

Table 3: Exponential Distribution, $\lambda = 0.2, x_0 = 1.25$: Prediction of the Expected Revenue with Optimal Reserve Price

Num. bidders	n. Auctions	True Res. Val	App. Res. Val	Nonp. Res.Val	True E. Rev.	Bias Approx.	Bias Nonp.
5	5000	12.0832	13.1854 [0.0020]	12.9484 [0.2136]	11.454	-1.56%	-0.76%
50	5000	12.0832	12.8005 [0.0281]	13.8545 [0.3486]	12.909	-0.08%	-0.89%
100	5000	12.0832	12.5411 [0.0134]	13.975 [0.2945]	14.9269	-0.00%	-1.12%
150	5000	12.0832	11.9286 [0.0222]	14.1249 [0.4486]	14.3471	<10^-3	-1.36%
5	50000	12.0832	13.079 [0.0022]	12.9123 [0.2214]	11.454	-1.56%	-0.64%
50	50000	12.0832	12.7532 [0.0239]	13.7297 [0.2486]	12.909	-0.07%	-0.72%
100	50000	12.0832	12.3295 [0.0214]	13.929 [0.2130]	14.9269	-0.00%	-1.09%
150	50000	12.0832	11.8672 [0.0189]	14.0142 [0.3465]	14.3471	<10^-3	-1.22%

Table 4: Normal Distribution, $\mu = 10, \sigma = 2; x_0 = 10.8$: Prediction of the Expected Revenue with Optimal Reserve Price

estimated reserve price, falls to zero immediately. On the other hand, when the nonparametric approach is used, the precision decreases with N. However, the loss in precision is not as drastic as it was when only the expected revenue with no reserve price was considered. When $N = 150$, the bias in the *induced* expected revenue is only 10%, for the case of the exponential distribution, and -1.22%, for the case of the normal distribution. The reason is that the lower tail of the support is ignored by virtue of the outside value: by eliminating the region where most of the bias is located, the nonparametric method is able to provide satisfactory results.
Finally, as we have seen above, the magnitude of the sample size affects only slightly the precision of the estimates: this confirms the argument that convergence occurs slowly.

We have derived results from other distributions, such as uniform, lognormal and mixed distributions for which there is no analytical expression, and the evidence seems consistent. The approach based on EVT systematically provides better estimates than the nonparametric approach. It is to be noted that the approximation method is computationally easier to perform, since it breaks down to the estimation of only two normalizing constants: all the subsequent steps can be solved analytically, using the appropriate gPd or EVD.

6 Conclusions

The results presented in the previous section provide support to the theory advanced in this paper: the parental distribution of bidders' valuations cannot be estimated at the regular rate, and therefore is irregularly identified. Monte Carlo simulations show that even when the sample size contains as much as 50,000 observations, the bias stays relevant and is not significantly lower than the one present when the sample size counted only 5,000 observations. The number of bidders strongly affects the precision of the estimates, and dominates every sample size effect.
The approximating distribution derived through EVT suffers from lack of precision on the left tail as well: however, EVT provides uniform bounds that ensure that the bias fades fast with N. The approximate distribution performs better than its nonparametric counterpart, even when the approximation is known to occur slowly, such as the case of the normal distribution. Increasing the value of N makes the EVT estimates more precise, and, simultaneously, the nonparametric estimates worse. The n-asymptotics predict an increase in efficiency for the nonparametric estimator, given a fixed N: however, the rate of

convergence of such estimator is so low that, realistically, no available dataset is able to satisfy the data requirement.
Even though the form of the approximating distribution is analytical, the set of assumptions that justify its use are very mild and we could reasonably expect most of existing distributions to satisfy them. The practical advantage of adopting analytical formulas relies on saving computational time, making the computation of the relevant measures a minor task.

Appendix

Appendix A: Extreme Value Theory: Convergence Rates for First Order Statistics

We are going to show here that the convergence rate for the case of the negative exponential is of order $O(1/N)$ (Hall and Wellner (1979)), and of order $O(1/\log N)$ for the normal case (Hall (1979)).

Normal Distribution Consider a standard normal random variable. The guesses for the normalizing constants, a_N, b_N are such that

$$1 - \Phi(b_N) = \frac{1}{N}$$

and $a_N = 1/b_N$.
We are going to make use of the approximation

$$\frac{1 - \Phi(x)}{\phi(x)} = \frac{1}{x}$$

to rewrite the condition for b_N

$$1 - \Phi(b_N) = \frac{1}{b_N}\frac{1}{\sqrt{2\pi}}e^{-\frac{b_N^2}{2}} = \frac{1}{N} \Longrightarrow b_N = N\frac{1}{\sqrt{2\pi}}e^{-\frac{b_N^2}{2}}$$

In order to solve for the root of this function, we first log-linearize it

$$f(b_N) = 2\log N - b_N^2 - 2\log b_N - \log 2\pi$$

and then we apply Newton Method: the first order approximation is given by

$$f(b_N) = 0 \Longrightarrow b_N^2 = 2\log N - 2\log b_N - \log 2\pi$$

$$b_{N,0} = \sqrt{2\log N - 2\log b_N - 2\log 2\pi} \approx \sqrt{2\log N}$$

Then,

$$f(b_{N,0}) = 2\log N - 2\log N - 2\log\sqrt{2\log N} - 2\log\pi =$$

$$= -\log 2\log N - \log 2\pi = -\log\log N - \log 4\pi$$

and the first derivative

$$f'(b_N) = -2b_N - 2\frac{2}{b_N} \Longrightarrow$$

$$f'(b_{N,0}) = -2\sqrt{2\log N} - \frac{2}{\sqrt{2\log N}} \approx -2\sqrt{2\log N}$$

Finally,

$$b_N = \sqrt{2\log N} - \frac{1}{2}\frac{\log\log N + \log 4\pi}{\sqrt{2\log N}}$$

Now it is possible to use the result presented in Lemma 3 [19]

$$\frac{1}{b_N}\phi\left(\frac{x}{b_N} + b_N\right) = \left(\frac{\sqrt{2\pi}}{Ne^{-\frac{b_N^2}{2}}}\right)\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2b_N^2}-\frac{b_N^2}{2}-2x} =$$

$$= \underbrace{\exp(-x - \log_N)}_{=\frac{1}{N}\frac{g(x)}{G(x)}}\ \underbrace{e^{-\frac{x}{2b_N^2}}}_{=1+h_N(x)}$$

Since, $e^x \approx 1 + x$, and using the first order approximation for $b_N = \sqrt{2\log N}$

$$h_N(x) = e^{-\frac{x^2}{2b_N^2}} - 1 \approx -\frac{x^2}{2b_N^2} \approx -\frac{x^2}{4\log N}$$

The rate of convergence for a normal distribution is of order $O(1/\log N)$

Exponential Distribution Consider the distribution function for a negative exponential distribution, with parameter λ.

$$F(x) = 1 - e^{-\lambda x}$$

As normalizing constants we are going to use, $a_N = \frac{1}{\lambda}$ and $b_N = \frac{1}{\lambda}\log N$ [20]. From Lemma 3

$$\frac{1}{\lambda}\lambda\frac{e^{-x}}{N} = \frac{e^{-x}}{N} = \frac{1}{N}\frac{g(x)}{G(x)}[1 + h_N(x)]$$

Since $\frac{1}{N}\frac{g(x)}{G(x)} = \frac{e^{-x}}{N}$, we have

$$\frac{e^{-x}}{N} = \frac{e^{-x}}{N}[1 + h_N(x)]$$

[19] We recall here the result, for brevity

$$a_N f(a_N t + b_N) = \frac{1}{N}\frac{g(t)}{G(t)}[1 + h_N(x)]$$

with $h(N)$ going to zero at the Convergence rate.

[20] It is easy to see that this choice gives exact convergence to a Gumbel distribution. In fact

$$F(a_N x + b_N)^N = \left[1 - e^{-\lambda(a_N x + b_N)}\right]^N = \left[1 - \frac{e^{-x}}{N}\right]^N \longrightarrow e^{-e^{-x}}$$

which implies that

$$\frac{h_N(x)}{N} \longrightarrow 0$$

Therefore the rate of convergence is of order N

Appendix B: Uniform Bounds for The Largest Order Statistics

The results presented in the paper have made use of the convergence of the order statistics to the corresponding EVD. The actual type of convergence implied by theory is stronger than the one used in the paper. We are going to show in this section that the order statistics converge uniformly, we are going to provide uniform bounds for such converge, and we will argue that the fastest possible rate of convergence is exactly of order $O(1/N)$.

Going back to the original representation, if

$$F^N(a_N x + b_N) \longrightarrow G(x)$$

then

$$\frac{dF^N(a_N x + b_N)}{dx} = N a_N f(a_N x + b_N) F^{N-1}(a_N x + b_N) \longrightarrow G(x)p(x) = g(x)$$

were $p(x)$ is the density of the corresponding gPd. Convergence of the densities implies uniform convergence, by Scheffé's Lemma

$$\lim_{N \longrightarrow \infty} \sup_{B \in \mathcal{B}} \left| P^N \left(\frac{X_{N:N} - b_N}{a_N} \in B \right) - G(B) \right| = 0$$

We are going to make use of two results implied by Reiss (1981). The first,

Lemma 4 *Let $N \in \mathbb{N}$. There exists a constant $C \in \mathbb{R}^+$ such that*

$$\sup_{B \in \mathcal{B}} \left| U^N \left(\{N[X_{N:N} - 1]\} \in B \right) - G(B) \right| \leq \frac{C}{N}$$

where U is the uniform distribution over $(0,1)$.

and the second,

Lemma 5 *Let $\mathcal{P}_1, \mathcal{P}_2$ be two probability measures on a measurable space, $(X, \mathcal{B})$, dominated by a σ measure μ. Denote by p_1, p_2 the*

$\mu-$densities of $\mathcal{P}_1$ and $\mathcal{P}_2$ respectively. Then,

$$\sup_{B\in\mathcal{B}} |P_1(B) - P_2(B)| \leq \sqrt{1 - \left(\int \sqrt{\frac{p_1}{p_2}} dP_2\right)^2} \leq$$
$$\leq \sqrt{1 - \exp\left\{\int \log \frac{p_1}{p_2} dP_2\right\}}$$

Now we can state the theorem,

Theorem 3 *Suppose there exists $a_N > 0, b_N$, such that, for the probability distribution $\mathcal{P}$,*

$$\mathcal{P}^N\left(\frac{X_{N:N} - b_N}{a_N} \leq x\right) \longrightarrow G(x)$$

then,

$$\sup_{B\in\mathcal{B}} \left|\mathcal{P}^N\left(\frac{X_{N:N} - b_N}{a_N} \in B\right) - G(B)\right| \leq \frac{C+2}{N} + \sqrt{\int \left(\frac{h_N(x)}{2}\right)^2 G(dx)}$$

<u>Proof:</u> Consider the inverse of F, $F^{-1}(t) := \inf\{x \in \mathbb{R} | F(x) \geq t\}$, then

$$\mathcal{P}^N(Z_{N:N}) = U^N(F^{-1}(Z_{N:N}))$$

From Lemma 4,

$$\sup_{B\in\mathcal{B}} \left|\mathcal{P}\left(\frac{X_{N:N} - b_N}{a_N} \in B\right) - \mathcal{E}\left(\frac{F^{-1}\left[1 - \frac{x}{N}\right] - b_N}{a_N} \in B\right)\right| \leq \frac{C}{N}$$

where $\mathcal{E}$ denotes the negative exponential distribution, $E(x) = 1 - e^{-x}$.

Define the measure $\mu/\mathcal{B}$ as

$$\mu(B) := \mathcal{E}\left(\frac{F^{-1}\left[1 - \frac{x}{N}\right] - b_N}{a_N} \in B\right)$$

with associated Lebesque density

$$m(x) = N a_N f(a_N x + b_N) e^{-N(1 - F(a_N x + b_N))}$$

Similarly,

$$\tilde{\mu} := \frac{\mu}{1 - e^{-N}}$$

is a probability measure with density $\tilde{m} := \frac{m}{1-e^{-N}}$. Then,

$$\sup_{B\in\mathcal{B}} |\tilde{\mu}(B) - \mu(B)| \le e^{-N} \le \frac{1}{N}$$

The uniform bound that we are looking for is given by

$$\sup_{B\in\mathcal{B}} |\mathcal{P}^N - G| \le \sup_{B\in\mathcal{B}} |\mathcal{P}^N - \mu| + \sup_{B\in\mathcal{B}} |\mu - \tilde{\mu}| + \sup_{B\in\mathcal{B}} |\tilde{\mu} - G|$$

Next, we apply Lemma 5 and write

$$\sup_{B\in\mathcal{B}} |\tilde{\mu}(B) - G(B)| \le \sqrt{1 - \exp\left\{\int \log \frac{\tilde{h}}{g} dG\right\}} =$$

$$= \sqrt{1 - \exp\left\{\int \log \frac{h}{g} - \log(1-e^{-N})\right\} dG} \le$$

$$\le \sqrt{1 - \exp\left\{\int -N(1 - F(a_N x + b_N)) + \log(N a_N f(a_N x + b_N)) - \log(G(x)p(x))G(dx)\right\}}$$

Isolate the first term inside the integral, then, applying Fubini's Theorem,

$$\implies \int 1 - F(a_N x + b_N)G(dx) = \int \int_{a_N x + b_N}^{\infty} f(y)d(y)G(dx) =$$

$$= \int a_N \int_x^{\infty} f(a_N x + b_N) dy G(dx) = a_N \int f(a_N y + b_N)G(y)dy =$$

$$= a_N \int \frac{f(a_N y + b_N)}{p(y)} dy$$

Finally, this allows us to rewrite

$$\sup_{B\in\mathcal{B}} |\tilde{\mu}(B) - G(B)| \le$$

$$= \sqrt{1 - \exp\left\{\int -N \frac{a_N f(a_N x + b_N)}{p(x)} + \log N \frac{a_N f(a_N x + b_N)}{p(x)} - \log G(x)G(dx)\right\}} \approx$$

$$\approx \sqrt{\int -N \frac{a_N f(a_N x + b_N)}{p(x)} - 1 + \log N \frac{a_N f(a_N x + b_N)}{p(x)} G(dx)} =$$

$$= \sqrt{\int \frac{\frac{g(x)}{G(x)}}{p(x)}[1 + h_N(x)] - 1 - \log \frac{\frac{g(x)}{G(x)}}{p(x)}[1 + h_N(x)]G(dx)} =$$

$$= \sqrt{\int h_N(x) - \log[1 + h_N(x)]G(dx)} \approx \sqrt{\int \frac{h_N(x)^2}{2} G(dx)}$$

since $x-1-\log x \approx \frac{(x-1)^2}{2}$

$\blacksquare$

Since $h_N(x)$ is at most of order $O(1/N)$ for the gPd family, it must be that the tightest possible uniform bound is also of order $O(1/N)$.

Extending the proof, it is possible to show that Theorem 3 extends to the whole joint distribution of the order statistics.

Theorem 4 *Suppose there exists $a_N > 0, b_N$ such that,*

$$\mathcal{P}^N\left(\frac{X_{N:N}-b_N}{a_N} \leq x\right) \longrightarrow G(x)$$

then, for any $N \in \mathbb{N}$ and $k = 1, \cdot, N$

$$\sup_{B \in \mathcal{B}^k}\left|\mathcal{P}^N\left(\left\{\frac{X_{N-i+1:N}-b_N}{a_N}\right\}_{i=1}^k \in B\right) - G^{(k)}(B)\right| \leq \frac{C+2}{N}k + $$
$$+\sqrt{\int \frac{h_N(x)^2}{2}\sum_{i=1}^k g_{(i)}(x)dx}$$

References

1. Athey, S., AND P.A. Haile (2002), "Identification of Standard Auction Models ", *Econometrica*, 70 (6), 2107-2140
2. Donoho, D.L., AND R.C. Liu, (1988), "The "Automatic" Robustness of Minimum Distance Functionals ", *The Annals of Statistics*, 16 (2), 552-586
3. Falk, M. (1985), "Rates of Uniform Convergence of Extreme Order Statistics ", *Annals of the Institute of Statistical Mathematics*, 38 (1), 245-262
4. Falk, M. (1989), "A Note on Generalized Pareto Distributions and the k Upper Extremes", *Probability Theory and Related Fields*, 85 (4), 499-503
5. Falk, M., AND F. Marohn, (1993), "Von Mises Conditions Revisited", *The Annals of Probability*, 21 (3), 1310-1328
6. Gnedenko, B. (1943), "Sur la distribution limite du terme maximum d'une série aléatorie ", *Annals of Mathematics*, 44, 423-453
7. Guerre, E., I. Perrigne, AND Q. Vuong (2000), "Optimal Nonparametric Estimation of First-Price Auctions", *Econometrica*, 68 (3), 525-574
8. Haile, P.A., AND E. Tamer (2003), "Inference with Incomplete Models of English Auctions", *Journal of Political Economy*, 111 (1), 1-51
9. Hall, P. (1979), "On the Rate of Convergence of Normal Extremes", *Journal of Applied Probability*, 16 (2), 434-439
10. Hall, W.J. AND J.A. Wellner, (1979), "The Rate of Convergence in Law of the Maximum of an Exponential Sample" , *Statistica Neerlandica*, 33 (3), 151-154
11. Hayashi, F. (2000) "Econometrics", Princeton University Press.
12. Khan, S., AND E. Tamer, (2009), "Irregular Identification, Support Conditions, and Inverse Weight Estimation", Working Paper
13. Krishna, V. (2002), "Auction Theory", Academic Press
14. Menzel, K. AND P.R. Morganti (2010) "Estimation of Auction with Incomplete Bidding Data", Working Paper
15. von Mises, R. (1936), "La distribution de la plus grande de n valeurs" [Reprinted (1954) in Selected Papers II, pp. 271-294], Amer. Math. Soc., Providence, RI

16. Parr, C.W., AND W.R. Schucany, (1979), "Minimum Distance and Robust Estimation", *Technical Report No. 136 Department of Statistics ONR Contract*

17. Pickands, J. III (1975), "Statistical Inference Using Extreme Order Statistics", *Annals of Statistics*, 3 (1), 119-131

18. Rao, P.V., E.F. Schuster AND R.C. Littell, (1975) "Estimation of Shift and Center of Symmetry Based on Kolmogorv-Smirnov Statistics", *The Annals of Statistics*, 3 (4), pp. 862

19. Reiss, R. D., (1981) "Uniform Approximation to Distribution of Extreme Order Statistics", *Advances in Applied Probability*, 13, pp. 533-547

20. Vickrey, W. (1961) "Counterspeculation, Auctions, and Competitive Sealed Tenders", *Journal of Finance*, 16 (1), pp. 8-37