



TI 2005-031/2

Tinbergen Institute Discussion Paper

Auctions with Numerous Bidders

Silvia Caserta^{1,2}

Casper G. de Vries^{1,3}

¹ Faculty of Economics, Erasmus Universiteit Rotterdam,

² LCH.Clearent,

³ Tinbergen Institute.

Tinbergen Institute

The Tinbergen Institute is the institute for economic research of the Erasmus Universiteit Rotterdam, Universiteit van Amsterdam, and Vrije Universiteit Amsterdam.

Tinbergen Institute Amsterdam

Roetersstraat 31

1018 WB Amsterdam

The Netherlands

Tel.: +31(0)20 551 3500

Fax: +31(0)20 551 3555

Tinbergen Institute Rotterdam

Burg. Oudlaan 50

3062 PA Rotterdam

The Netherlands

Tel.: +31(0)10 408 8900

Fax: +31(0)10 408 9031

Please send questions and/or remarks of non-scientific nature to driessen@tinbergen.nl.

Most TI discussion papers can be downloaded at <http://www.tinbergen.nl>.

Auctions with Numerous Bidders

Silvia Caserta*

Erasmus University Rotterdam and LCH.Clearnet

Casper G. de Vries

Erasmus University Rotterdam and Tinbergen Institute

March 2005

Abstract

We study auctions in which the number of potential bidders is large, such as in Internet auctions. With numerous bidders, the expected revenue and the optimal bid function in a first price auction result in complicated expressions, except for a few simple distribution function for the bidders' valuations. We show that these expressions can be well approximated using extreme value theory without assuming a particular distribution function. The theory is applied to data from Internet auctions.

Key words: Auctions; Numerous bidders; Extreme values; Internet auctions

JEL classification: D44

*The views expressed are those of the authors and do not necessarily represent those of LCH.Clearnet. Corresponding author: S. Caserta, Faculty of Economics, H14-01, Erasmus University Rotterdam, P.O. Box 1738, 3000 DR Rotterdam, the Netherlands. Tel. +31 (0)10 4081286. E-mail caserta@few.eur.nl

1 Introduction

Auctions have historically been seen as (games of) strategic interactions between a few agents. Recently auctions on the Internet have become a very popular way to sell to the widest audiences the most disparate objects ranging from collection items to airplane tickets. The Internet has made widespread auction participation feasible due to low participation costs as the physical presence of attendees is not needed. As a consequence, the number of auction participants, both in terms of sellers and buyers, has risen. E-bay, the leading on-line marketplace for the sale of goods such as collectibles, computers and autos has as of today more than a hundred million registered users around the world¹.

Auctions with numerous bidders have not received much attention, neither theoretically nor empirically. The main reason for this negligence is that the computational complexity for many of the interesting expressions quickly becomes prohibitive as the number of bidders increases. For example, calculating the expected revenue as the expectation of the second highest order statistics of valuations is computationally difficult except for a few simple distributions of valuations. This is so because the cumulative distribution function of the second highest valuation depends on the distribution of valuations $F(x)$ and the number of bidders n , through the term $F^n(x)$. An analytic expression for the latter is usually cumbersome to obtain already for small values of n . The same applies to an explicit derivation of the equilibrium bid function $b(x)$ in a first-price auction. For n large McAfee and McMillan (1987a), pp. 710, noted that: "Finding the Nash-equilibrium bid in the first-price or Dutch auction is a nontrivial computational problem". The limiting case with an infinite number of bidders has been discussed in Holt (1979) and Harris and Raviv (1981). In these papers it was shown that indefinitely increasing the number of bidders, increases the seller's revenue, and in the limit the seller just fetches a price equal to the highest possible valuation.

In this paper we extend the theory of the standard auctions under the Independent Private Values Paradigm (IPVP) to the case when there is a large but nevertheless finite number of bidders. As we show, the theory of extremes can be used to derive approximate expressions for the expected revenue and the equilibrium bid function in a first-price auction when there are numerous bidders, without requiring that the distribution of valuations $F(x)$ is known in detail. This sounds like magic, but drawing the analogy

¹Company Overview <<http://www.ebay.com>>.

between the limit laws from extreme value theory and the central limit law for averages is insightful. The central limit law states that normed averages from an unknown distribution are well approximated by the normal distribution if the sample becomes large. Similarly, albeit less well known, in large samples the limit laws from extreme value theory provide good approximations to the distributions of appropriately normed higher order statistics, such as the maximum and the second highest order statistics.

Of course there are many auctions for which the assumption of a large bidder audience is not appropriate, but the recent phenomenon of Internet auctions is a case that fits our framework quite well. Apart from the theoretical extension to the numerous bidder case, we also provide the necessary estimators to be able to apply the theory to real data.

The objective in most studies is to test the validity of the IPVP. In our application, we take the IPVP for granted and ask which price level one can expect to prevail in a repetition of a similar auction. For practical users we believe this application is the more interesting usage of auction theory. Surprisingly little attention has been given to this problem. This is perhaps due to the fact that standard statistical approaches can not be used to analyze the statistical properties of the extreme order statistics. We develop the necessary estimators to address this question by extending statistical extreme value theory. The methodology we use introduces a different approach to the literature in empirical auctions as we do not pool data from different auctions, nor do we assume a specific functional form for the distribution of valuations. The few papers that deal with empirical auction data tend to assume a specific distribution function for the valuations and pool data from different auctions of comparable items; see Laffont et al. (1995), Paarsch (1992), Donald and Paarsch (1993). An exception to the parametric approach is Guerre et al. (2000), which uses kernel estimators to estimate the distribution of observed bids (still employing cross-sectional data), and Haile and Tamer (2003). But this approach is not informative regarding the distribution of the highest bids and the expected revenue, on which we focus.

The paper is structured as follows. In the next section we develop the extension to the theory of the standard auctions with risk neutral bidders within the IPVP for the case where the number of bidders is large. By means of extreme value theory we obtain expressions for the expected revenue and the equilibrium bid in a first-price auction, which are not sensitive to the exact distribution of valuations $F(x)$.

In the third section we confront the theory with data from Internet auctions. We use bids from second-price auctions, so that we can identify bids

with valuations. We first show graphically that the largest bids fit the limit distribution suggested by extreme value theory quite well. Subsequently, we estimate the expected revenue and the largest valuation with the help of appropriate estimators. As the focus of this paper is not the statistical properties of the estimators, we relegate the derivation of these estimators to an appendix. The final section concludes.

2 Auctions with Many Bidders

In this section we extend the theory for the standard auctions to the case when there is a large but finite number of bidders. First we derive approximations to the expected highest valuation and the expected revenue for the four standard auctions². Then we look at the optimal bid function for first price auctions.

Consider the IPVP according to which each bidder knows the value of the object to herself, while this valuation is not known to the other bidders and the auctioneer (which we always identify with the seller). All bidders and the auctioneer are assumed to be risk neutral. All bidders' valuations are independent draws from the same probability distribution function $F(x)$, which is known to the bidders. And last, each bidder knows the number of her competitors. This last assumption might not be realistic for auctions with large audiences. Nevertheless, the arguments used to prove the Revenue Equivalence Theorem (RET) are independent of whether each bidder knows the number of competitors before she bids; see Klemperer (1999), pp. 240. This fact is important, since it allows us to use the conclusion of the RET when we consider Internet auctions. In Internet auctions a bidder is typically not informed about the exact number of potential bidders, but can safely assume that there are many of them. Also the other assumptions of the IPVP limit the applicability of the theory. The private values assumption is, however, a reasonable assumption for auctions of consumer goods that are bought by individuals for their private use. This was already pointed out by Milgrom and Weber (1982), pp. 1093. In the applications we have restricted ourselves to these type of goods. The Internet auctions are moreover mostly second price auctions, hence the bidders' optimal strategy is dominant so that the assumption of risk neutrality is not restrictive.

Consider a sample of n independent and identically distributed random

²By the RET, all the standard auctions yield an expected revenue to the auctioneer equal to the expected value of the second highest valuation; see e.g. McAfee and McMillan (1987a).

variables $\{X_1, \dots, X_n\}$ with distribution function $F(x)$ representing the bidders' valuations in a particular auction. The descending order statistics of this sample are

$$M_n = \max\{X_1, \dots, X_n\} = X_{(1)} \geq X_{(2)} \geq \dots \geq X_{(n)},$$

where the appropriate rank is indicated between the bracketed subscript. Under the IPVP and if the auction awards the item to the highest bidder, auction theory shows that the expected selling price is $E[X_{(2)}]$. Under very mild regularity conditions extreme value theory provides approximate expressions for the distributions of M_n and $X_{(2)}$; see Caserta (2002) for a complete list of references. These can then be used to derive approximate expressions for the expectations $E[M_n]$ and $E[X_{(2)}]$ as we show below.

2.1 Expected highest valuation: $E[M_n]$

Under the assumption that valuations are independent and identically distributed, the distribution of the maximum M_n is $F^n(x)$. When we drive n to infinity, this distribution becomes degenerate because

$$\lim_{n \rightarrow \infty} P\{M_{(n)} \leq x\} = \lim_{n \rightarrow \infty} F^n(x) = \begin{cases} 1 & \text{if } x = z \\ 0 & \text{if } x < z \end{cases} \quad (1)$$

where z is the upper endpoint of the distribution $F(x)$. In similarity with the central limit law, the question extreme value theory answers is for which linear rescaling of x the distribution $P\{M_{(n)} \leq x\}$ converges to a non-trivial limit distribution, and how this limit distribution looks like. The appropriate limit laws are found by the Extreme Value Theorem (EVT). Below we report the EVT for distributions with bounded support ($z < \infty$), because in the auction literature it is often assumed that the support of $F(x)$ has a finite upper endpoint z . This assumption is harmless when we consider non-durable consumption goods. In other cases, such as valuable arts, there might not be a well defined upper endpoint because the valuation of such items depends on the future value of the object itself. The EVT can also be stated for distributions with unbounded support ($z = \infty$). For the general statement of the EVT and its proof, see Embrechts et al. (1997), Resnick (1987) and de Haan (1970).

Theorem 1 (EVT) *For a cumulative distribution function $F(x)$ such that $z < \infty$, if there exist norming constants $s_n > 0$ and $u_n \in \mathbb{R}$ such that*

$$\lim_{n \rightarrow \infty} P\{s_n(M_n - u_n) \leq x\} = \lim_{n \rightarrow \infty} F^n(s_n^{-1}x + u_n) = G(x), \quad (2)$$

where the convergence is in distribution to a non degenerate $G(x)$, then only two possibilities (up to location and scale parameters) arise for $G(x)$:

$$G(x) = \Psi_\alpha(x) \equiv \begin{cases} \exp\{-(-x)^\alpha\} & x \leq 0 \\ 1 & x > 0 \end{cases} \quad \alpha > 0, \quad (3)$$

or

$$G(x) = \Lambda(x) \equiv \exp\{-e^{-x}\} \quad x \in \mathbb{R}. \quad (4)$$

The distributions $\Psi_\alpha(x)$ and $\Lambda(x)$ are respectively the Weibull and Gumbel extreme value distributions. The EVT says that under certain conditions regardless the specific form of $F(x)$, the scaled maximum admits a limit distribution as the number of bidders tends to infinity. It follows that in finite but large samples the maximum M_n is approximately Weibull $\Psi_\alpha(x)$, or it is approximately Gumbel $\Lambda(x)$ distributed. Then we say that $F(x)$ is in the (maximum) domain of attraction of an extreme value distribution, and we indicate³ it by $F \in \text{MDA}(\Lambda)$ or $F \in \text{MDA}(\Psi_\alpha)$ with tail index α .

For reasons that will become clear in the estimation procedure⁴ and to keep notation easy, from now on we concentrate on the Weibull limit law (3), that is the case $F \in \text{MDA}(\Psi_\alpha)$. For the case of $F \in \text{MDA}(\Lambda)$ see Caserta (2002). For a specific distribution $F \in \text{MDA}(\Psi_\alpha)$, the theoretical values for the norming constants in (2) are given by

$$u_n = z, \quad s_n = \frac{1}{u_n - F^{\leftarrow}(1 - 1/n)} \quad (5)$$

where $F^{\leftarrow}(p)$ is the generalized inverse function, i.e. $F^{\leftarrow}(p) \equiv \inf\{x : F(x) \geq p\}$, for $0 < p < 1$ and $F^{\leftarrow}(1) \equiv z$. The intuition for the norming constants in (5) is that we need to use the upper endpoint and the difference between this and the realized maximum in order to avoid a trivial limit law as in (1).

The idea is to compute the expected maximum valuation and the expected revenue to the auctioneer by using the result in the EVT. This is not directly possible, however, since the EVT result only establishes convergence in distribution. Under a mild extra condition we can guarantee that

³An example of $F \in \text{MDA}(\Psi_\alpha)$ is given by the uniform distribution, while the normal distribution is in the domain of attraction of the Gumbel.

⁴The functional form of the tail of $F(x)$ determines which of the two limit laws is appropriate. There are sufficient conditions, such as the Von Mises tail conditions, for the distribution to be in the domain of attraction of a particular limit law; see Embrechts et al. (1997) and Resnick (1987). Also, there are simple statistical devices such as the QQ-plot, which we discuss in Section 3, to determine which case is relevant in practice.

convergence in moments is achieved. In particular, we have to control the behavior of the left tail of $F(x)$ in order to prevent an improbable large event from disrupting moment convergence (given that the EVT only refers to the right tail). The following theorem is adapted to the case we are interested in, and we refer the reader to Resnick (1987) for its general statement and proof.

Theorem 2 *Suppose $F(x) \in \text{MDA}(\Psi_\alpha)$, and $F(x)$ has finite upper endpoint z . If for some $m \in \mathbb{N}^+$*

$$\int_{-\infty}^z |x|^m F(dx) < \infty \quad (6)$$

then, for u_n and s_n as in (2), we have⁵

$$\lim_{n \rightarrow \infty} E[(s_n(M_n - u_n))^m] = \int_{-\infty}^0 |x|^m \Psi_\alpha(dx) = (-1)^m \Gamma(1 + m/\alpha). \quad (7)$$

Condition (6) holds for distributions with a finite lower endpoint, such as the uniform. Since prices and valuations of auctioned objects can not have negative values, we can safely assume that (6) is satisfied by the distribution of valuations. Subsequently, the result in (3) can be used to calculate $E[M_n]$.

Corollary 3 *Under the condition of Theorem 2, if the number of bidders n is large then*

$$\begin{aligned} E[M_n] &\approx \frac{-\Gamma(1 + \alpha^{-1})}{s_n} + u_n \\ &= \{1 - \Gamma(1 + \alpha^{-1})\}z + \Gamma(1 + \alpha^{-1})F^{\leftarrow}(1 - 1/n). \end{aligned} \quad (8)$$

Proof. Use (7) and take $m = 1$. This gives the first line in (8). Subsequently substitute (5) for the norming constants. ■

The corollary gives a formula for finding the expected maximum valuation for a sample of n valuations. Recall, for example, that in theory a sample of valuations is provided by the bid book of a second-price auction. Hence, the formula can be directly used for those auctions.

⁵The Gamma function is $\Gamma(\lambda) = \int_0^\infty x^{\lambda-1} e^{-x} dx$, for $\lambda > 0$.

2.2 Expected revenue: $E[X_{(2)}]$

A result similar to the above can be obtained for the expected revenue. Consider a second price sealed bid auction. As Vickrey argued the optimal strategy in this auction is to bid one's valuation: it neither pays to overbid or to shade one's bid. Given the second price nature the expected revenue is

$$E[X_{(2)}] = \int_0^z x dH(x). \quad (9)$$

The law $H(x)$ is given by

$$\begin{aligned} H(x) &= nF^{n-1}(x)[1 - F(x)] + F^n(x) \\ &= nF^{n-1}(x) - (n-1)F^n(x). \end{aligned} \quad (10)$$

As pointed out in the introduction, even when $F(x)$ is known, the computation of $H(x)$ is already cumbersome for moderate values of n , except for some specific distributions. In addition, the auctioneer does not know $F(x)$. The auctioneer has to infer the distribution from data, and assuming an *a priori* functional form for $F(x)$ appears restrictive. On both accounts the seller would be helped if $E[X_{(2)}]$ could be found without going through the cumbersome process of having to estimate $F(x)$ and computing $H(x)$.

Consider again the case of an infinitely large set of bidders. The effect on the seller's revenue of increasing the number of bidders at infinitum was first analyzed by Holt (1979) and Harris and Raviv (1981). They consider the limiting case of an infinite number of participants. In this case it is quite intuitive that the seller receives a price equal to the highest possible valuation. This happens because also the second highest valuation converges to the upper endpoint of the distribution of valuations

$$\begin{aligned} \lim_{n \rightarrow \infty} P\{X_{(2)} \leq x\} &= \lim_{n \rightarrow \infty} \{nF^{n-1}(x)[1 - F(x)] + F^n(x)\} = \\ &= \begin{cases} 1 & \text{if } x = z \\ 0 & \text{if } x < z \end{cases} \quad a.s., \end{aligned} \quad (11)$$

because when $x < z$ we know that $F(x) < 1$ and therefore both $F^n(x)$ and $nF^{n-1}(x)$ converge to zero for n tending to infinity. Consequently, the effect of an unlimited increase in the number of bidders is to increase the revenue to the seller as the sales price tends to z .

Hence, when the number of bidders is large one might either directly compute the seller's expected revenue or approximate it by the upper endpoint of the distribution of valuations. Both suggestions are, however, of

little practical use. Direct computation is extremely difficult to work with, unless a particularly easy distribution function of valuations applies (that is why the auction literature in most cases considers the uniform distribution). And the approximation by the upper endpoint results in a large approximation error.

Nevertheless, if the distribution of valuation is such that the EVT holds, then a non-trivial limit exists also for $X_{(2)}$. This is given by

$$\begin{aligned} \lim_{n \rightarrow \infty} P\{s_n(X_{(2)} - u_n) \leq x\} &= [1 + (-x)^\alpha] \exp(-(-x)^\alpha) \\ &\equiv \Psi_{\alpha,2}(x) \quad x \leq 0, \quad \alpha > 0 \end{aligned} \quad (12)$$

for $s_n > 0$, $u_n \in \mathbb{R}$ as in the EVT⁶.

In order to measure the expected revenue to the auctioneer, we prove that a result similar to Theorem 2 holds for the second highest order statistic as well. Theorem 4 below is the main technical result of this paper. The general proof of the theorem can be obtained from the authors, while below we give details for the simple case of positive random variables, such as bidders' valuations.

Theorem 4 *Under the condition of Theorem 2 and for positive random variables, we have*

$$\lim_{n \rightarrow \infty} E[(s_n(X_{(2)} - u_n))] = \int_{-\infty}^0 |x| d\Psi_\alpha^{(2)}(x) = (-1)\Gamma(2 + \alpha^{-1}). \quad (13)$$

Proof. The first equality intuitively follows upon realizing that the variable $s_n(X_{(2)} - u_n)$ is always lower than the upper endpoint z . As a consequence, by the Dominated Convergence Theorem the limit and expectation operators can be interchanged and the first equality in (13) is established. The second equality follows by using the density of $\Psi_{\alpha,2}(x)$. ■

Corollary 5 *Under the condition of Theorem 2, if the number of bidders n is large then*

$$\begin{aligned} E[X_{(2)}] &\approx \frac{-\Gamma(2 + \alpha^{-1})}{s_n} + u_n \\ &= \{1 - \Gamma(2 + \alpha^{-1})\}z + \Gamma(2 + \alpha^{-1})F^{\leftarrow}(1 - 1/n). \end{aligned} \quad (14)$$

⁶When the original distribution is not in the domain of attraction of the Weibull, the limit in (12) is incorrect. For the general theorem concerning the limit of the second highest order statistic and its proof see Leadbetter et al. (1983).

Proof. As valuations are positive random variables, for n large one can use the approximation derived by (13), and subsequently substitute (5) for the norming constants. ■

Formula (14) is an essential contribution of the paper, and its usage is twofold. First, in those cases in which $F(x)$ is known but difficult to work with, the computation of the expected revenue $E[X_{(2)}]$ is simplified by the use of (14), when the number of bidders is large, compared to the necessity of calculating the expectation of a random variable with distribution as in (10). The examples below will further demonstrate this usefulness. Second, when we deal with empirical problems and we can not assume a specific functional form for $F(x)$, formula (14) is the only possibility we have to estimate the expected revenue.

2.3 Equilibrium bid function

The other theoretical contribution for the case of large bidder audiences we make concerns the approximation of the equilibrium bid function in first price auctions. Under the IPVP auction theory shows, see McAfee and McMillan (1987a), that in a first-price sealed-bid auction the symmetric Nash equilibrium bidding strategy as a function of the valuation x is

$$b(x) = \frac{\int_0^x y dF^{n-1}(y)}{F^{n-1}(x)} = E[X_{(2)} | M_n = x]. \quad (15)$$

The optimal equilibrium strategy is to bid the conditional expected second highest valuation, given that the bidder's valuation is the highest valuation. In forming her bid, each bidder makes the harmless assumption of having the highest valuation and determines how much she can shade down her bid towards the bidder with the second highest valuation and still win the auction⁷. Extreme value theory provides also an approximation to the bid function (15).

Proposition 6 *If $F(x) \in MDA(\Psi_\alpha(x))$ with bounded support $[0, z]$, then the equilibrium bid function (15) can be approximated by*

$$b(x) \approx x - \exp(s_n^\alpha(z-x)^\alpha) \frac{1}{\alpha s_n} \int_{s_n^\alpha(z-x)^\alpha}^\infty y^{\frac{1}{\alpha}-1} e^{-y} dy. \quad (16)$$

⁷If the bidder does not have the highest valuation this assumption is harmless since only the winner has to pay.

Proof. Given the bounded support, it is well known that partial integration of (15) gives

$$b(x) = x - \frac{1}{F^{n-1}(x)} \int_0^x F^{n-1}(y) d(y), \quad (17)$$

where the last part signifies the amount by which the bidder shades down her valuation in the bidding process. Since by assumption $F(x)$ is in the maximum domain of attraction of the Weibull extreme value distribution $\Psi_\alpha(x)$, we may again approximate $F^{n-1}(x)$ with the aid of the EVT. This means $F^{n-1}(x) \approx \exp(-s_n^\alpha(z-x)^\alpha)$, for $x \leq z$. Substituting this last expression into (17) gives, after some calculus manipulations, expression (16). ■

The problem in (17) is again the presence of the terms $F^{n-1}(x)$ and $F^{n-1}(y)$ which can be well approximated using the EVT. The resulting expression for $b(x)$ involves the incomplete gamma integral. Fortunately, the incomplete gamma integral in (16) has been extensively tabulated. Thus, expression (16) provides in general a significant simplification to (15).

2.4 Examples

To show the relevance of our theoretical results, we build some examples for specific distributions of valuations. We start with the example most commonly used in auction theory.

Uniform distribution Consider the uniform distribution $F(x) = x$, for $x \in [0, 1]$. The distribution of the maximum valuation M_n from a sample of n uniform random variables is $F_{M_n}(x) = x^n$. Direct integration yields $E[M_n] = 1 - 1/(n+1)$. The expected revenue follows by using the density $h(x) = n(n-1)(1-x)x^{n-2}$ from (10) and integrating over x , resulting in $E[X_{(2)}] = 1 - 2/(n+1)$.

Alternatively, we can use the EVT to compute the distribution of the maximum and subsequently approximate $E[M_n]$ and $E[X_{(2)}]$. The maximum from a sample of uniformly distributed random variables is in the domain of attraction of the Weibull distribution with $\alpha = 1$. In addition, one shows that $s_n = n$ and $u_n = z = 1$. Because the lower endpoint of $F(x)$ is finite, the condition for moment convergence is satisfied, and thus from the expressions in (8) and (14) we obtain

$$E[M_n] \approx 1 - \frac{1}{n} \quad \text{and} \quad E[X_{(2)}] \approx 1 - \frac{2}{n}.$$

Clearly for large n , the approximate maximum valuation and expected revenue are comparable to their corresponding theoretical values. Note that the approximate expected revenue is always lower than the theoretical one, although it still converges to one for n tending to infinity. As an example of the quality of the approximation, we compute the theoretical and the approximated expected revenue for $n = 40$. These are 0.9513 and 0.95 respectively, and the error is only 0.13 percent. If we would use the upper endpoint, equal to one, to approximate the seller's expected revenue, the error would be of 4.87 percent.

The bid function for the first-price sealed-bid auction is

$$b(x) = \frac{\int_0^x y(n-1)y^{n-2}dy}{x^{n-1}} = \frac{n-1}{n}x.$$

Using Proposition 6, we can compute the approximate optimal bid function as

$$b(x) \approx x - \exp(n(1-x)) \frac{1}{n} \int_{n(1-x)}^{\infty} e^{-y} dy = x - \frac{1}{n}.$$

Note that both the approximate equilibrium bid function and the exact equilibrium bid function converge to the bidder's valuation x when the number of bidders tends to infinity. As numerical example, consider $x = 0.9$ and $n = 40$. Then, the theoretical optimal bid is 0.877 and the approximate one is 0.875 with 0.2 percent error.

The uniform distribution example, although instructive, does not give full credit to the approximation technique, since the simplicity of the uniform distribution provides us with exact results for $F^n(x)$. Perhaps for this very reason the uniform distribution is often used in auction theory. In the next example, we consider a class of distribution functions which includes the uniform but which is quite difficult to deal with in general.

Power law behavior at the finite right endpoint Consider the beta type distribution functions

$$F(x) = 1 - K(z-x)^\alpha, \quad \text{for } z - K^{-1/\alpha} \leq x \leq z, \quad K, \alpha > 0 \text{ and } z < \infty.$$

It is easy to see that for $K = \alpha = z = 1$ we obtain the uniform distribution. For ease of comparison with the latter, let us fix $K = z = 1$, so that $F(x) = 1 - (1-x)^\alpha$ for $x \in [0, 1]$. These distributions are characterized by the fact that the higher α the faster they grow to one in comparison with the uniform, so that high values carry a lower probability the higher is α . This

set-up is fitting in situations where high valuations have low probability of occurrence.

If we want to compute the expected revenue to the auctioneer, we need to compute $F^n(x) = (1 - (1 - x)^\alpha)^n$ which is a polynomial⁸ of order αn . This is a computationally demanding task for large n , particularly when K and z are different from one. However, the EVT provides us with an approximation for the expected revenue. The distributions in this example are in the domain of attraction of the Weibull with norming constants $s_n = (nK)^{1/\alpha} = n^{1/\alpha}$ and $u_n = z = 1$; see Embrechts et al. (1997). Because the lower endpoint of $F(x)$ is finite, the condition for moment convergence is also satisfied and we easily obtain from expressions in (8) and (14)

$$E[M_n] \approx \frac{-\Gamma(1 + \alpha^{-1})}{n^{1/\alpha}} + 1 \quad \text{and} \quad E[X_{(2)}] \approx \frac{-\Gamma(2 + \alpha^{-1})}{n^{1/\alpha}} + 1. \quad (18)$$

For $\alpha = 1$ these expressions coincide with what we had in the previous example of the uniform distribution. For $\alpha > 1$ the expected revenue in (18) is lower than the expected revenue found for $\alpha = 1$. Consider, as in the other example $n = 40$ and suppose that $\alpha = 3$. Then, the theoretical expected maximum valuation and the theoretical expected revenue are respectively equal to 0.740 and 0.653. The approximate values are easily found via (18) and are 0.738 and 0.651, giving errors of 0.2 percent. In case the auctioneer uses the upper endpoint of the distribution, equal to one, as a quick approximation she would make a relatively large error of more than 30 percent.

The corresponding exact equilibrium bid function is

$$b(x) = \frac{\int_0^x y dF^{n-1}(y)}{F^{n-1}(x)} = \frac{\int_0^x \alpha(n-1)y(1-y)^{\alpha-1}(1-(1-y)^\alpha)^{n-2} dy}{(1-(1-x)^\alpha)^{n-1}}.$$

The integral is difficult to evaluate for large n . But, the approximate equilibrium bid function is

$$b(x) \approx x - \exp(n(1-x)^\alpha) \frac{1}{\alpha n^{1/\alpha}} \int_{n(1-x)^\alpha}^{\infty} y^{\frac{1}{\alpha}-1} e^{-y} dy.$$

Both the exact and approximate expressions coincide with the corresponding ones for the case of the uniform distribution when $\alpha = 1$. As in that example, fix $x = 0.9$, $n = 40$ and choose $\alpha = 3$. Then the exact equilibrium bid is

⁸The expression for the power of the polynomial is provided by the Newton's Binomial formula $(a + b)^n = \sum_{h=0}^n \binom{n}{h} a^h b^{n-h}$.

0.730 and the approximate equilibrium bid is 0.731, with only a 0.1 percent error.

The results obtained so far are valid only for the distributions in the domain of attraction of the Weibull extreme value distribution. If we consider auctions for artwork or collectable or even shares of private companies⁹, the hypothesis of the distribution of valuations having a finite upper endpoint might be restrictive. Even though prices are always finite, we can imagine that there is no fixed value they might converge to. Consider, for example, artwork and collectibles bought for (eventual) resale. Then, people's valuation might depend on estimates of future prices, which in turn depend on future valuations. This dependence makes it inappropriate to fix a finite upper endpoint. To see how the theory works when the distribution of valuations has an infinite upper endpoint, we refer the reader to Caserta (2002).

3 On-line Auction Application

In this section we use the theory developed above to answer to the question "what can one learn from previous held auctions of similar items before participating oneself?". Surprisingly little attention has been given to this question. Our analysis below analyzes which price level one can expect to prevail in a repetition of a similar auction. In contrast to most applied work, it is not our aim here to test auction theoretic paradigms. We take the IPVP for consumer goods as given, although we do test the often made assumption that the distribution of valuation has a finite upper endpoint.

The large size of the on-line auction audience makes these auctions amenable to extreme value analysis. We can use data from Internet auctions, assuming that the collected bids are representative of the valuations of the participants. Such an assumption appears reasonable because the second-price sealed-bid auction has the unique symmetric dominant equilibrium strategy whereby each bidder's best action is to submit a bid equal to his own valuation.

In order to study our Internet data, we develop the necessary estimators extending statistical extreme value theory. The derivation of the estimators is provided in Appendix B. The methodology we develop can be applied to the data collected from first and second price auctions as well as English auctions. It can not be applied to the data from a single Dutch auction,

⁹Initial public offerings of shares are sometimes run via auctions.

because only one bid (the top one) is observed in such a case. However, also for traditional first price auctions (as procurement auctions) and English auctions (as artwork auctions), data collection is complex and, in addition, the number of participating bidders is often relatively low. This is not the case for on-line second-price auctions, which have lately been booming due to their use on the Internet¹⁰, see Lucking-Reiley (2000).

On-line auction houses post all the bids for any auction and these data can be easily downloaded. We collected information and bids for three auctions from Yahoo!Auctions, and a summary of the auctions characteristics is contained in Table 1. We only considered auctions which had no reserve price and had automatic extension. For a detailed description of the auctions we refer the reader to Appendix A. Although we can not avoid the situation

Table 1: Auction description

	Auc.1	Auc.2	Auc.3
Total Bids	132	139	138
Selling Price \$	311.50	2210.01	825.00

that a bid might not completely reveal a bidder's true valuation, we use the bids to test whether the distribution of the maximum valuation conforms to a Weibull extreme value distribution, and to compute the seller's expected revenue.

The theory developed in Section 2 assumes that the bidders' valuations are drawn from a distribution with a finite upper endpoint. This is a generally made assumption when dealing with non-durable consumer goods. Such assumption is economically easy to justify, as it would not make sense for items such as mobile telephones or personal computers to be valued above the prices set by the most expensive outlets. However, to the best of our knowledge the boundedness of valuations for consumer goods is never tested for. The existence of a finite upper endpoint is vital as it (partly) determines the behavior of the highest statistics and therefore of the expected revenue. Hence, we start by verifying that the Internet data are drawn from a distribution which is in the domain of attraction of the Weibull extreme value distribution. To verify this, we first use the exploratory data analysis technique of the QQ-plot; see Appendix B. We reproduce the double-log plot for the data of the three auctions in Figure 1.

¹⁰Most on-line auction houses run second-price auctions. This is possibly so because of the simplicity of the optimal bidding strategy.

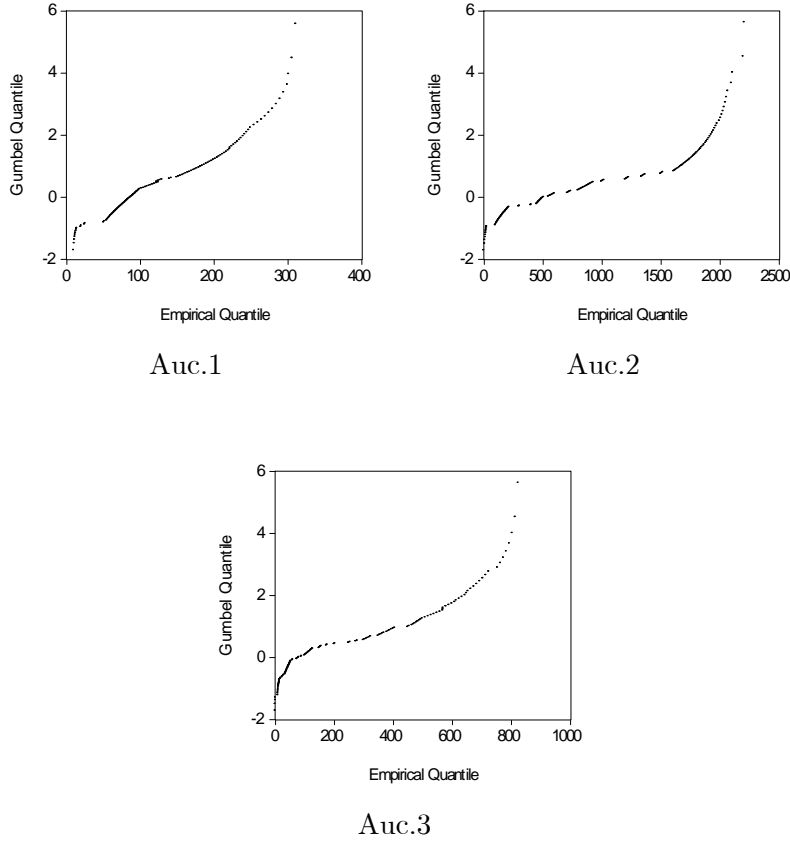


Figure 1: QQ-plots bids against Gumbel quantile

From these plots we deduce that the supposition of the data being generated from a distribution whose maximum values are in the domain of attraction of a Weibull extreme distribution with finite upper endpoint, is tenable, since all three curves are sloping upward at the highest quantiles. Equivalently, this confirms the existence of an upper bound for the valuations and the tail index estimates reported below provide further evidence.

With such support for our assumption, we can move on to estimate the expected revenue and the endpoint of the distribution. We start by estimating the tail index α needed in (14), which gives further information about the boundary of the distribution. We use the DEDH estimator described in Appendix B. For the three auctions, Figure 2 plots the DEDH estimates

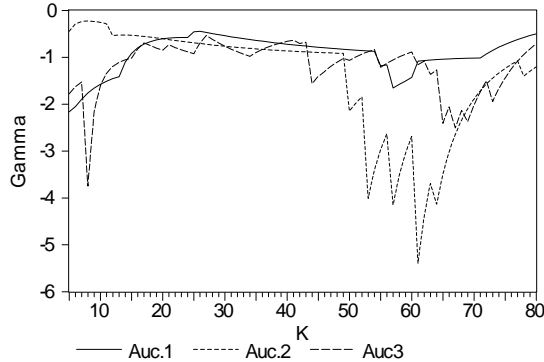


Figure 2: Gamma $\hat{\gamma} = 1/\hat{\alpha}$ estimates for the three auctions

of $\hat{\gamma} \equiv -1/\hat{\alpha}$ against the number k of higher bids taken into account (for k between 5 and 80). In the central part, the graphs are quite stable with values between -1 and -0.5 . Note that the estimated values are consistently negative¹¹ for all values of k , further confirming the hypothesis that the series are from a distribution whose maximum is in the domain of attraction of a Weibull. In turn, this implies that the distribution has a finite upper endpoint.

Similar plots are made to extract estimates for the expected revenue¹² and the upper endpoint. On the basis of these plots we choose the number of order statistics k employed in the estimation procedure. The results are presented in Table 2. While it needs not be the case, it turns out that for all three auctions our expected revenue estimates are below the actual selling price. This suggests the sellers were lucky in these particular auctions.

As a verification for the validity of our estimates we use the QQ-plot of the original series against the quantiles of the corresponding estimated Weibull extreme value distributions. The plots in Figure 3 seem to be sufficiently straight in the North-East part of the plane, to be satisfied with the

¹¹In addition, because the DEDH estimator for the tail index γ is asymptotically normal, see Dekkers et al. (1989), we obtain that at a 90 percent (one-sided) confidence level the tail index estimates for the three auctions are negative and in particular they are less than -0.18 , -0.05 and -0.22 for Auc.1, Auc.2 and Auc.3 respectively.

¹²It is important to understand that the selling prices reported in Table 1 are naturally the empirical maximums in each of the three data sets. But in practice, for the very structure of a second-price auction they represent the second highest bid, because the maximum bid is never shown. Hence, we will estimate the expected revenue to the auctioneer using the estimator for the maximum, i.e. using (25) and not using (24).

Table 2: Auction estimates

	γ	$\$ z$	$\$ E[\text{rev}]$
Auc.1	-0.58	321.70	307.73
	$k = 22$	$k = 22$	$k = 22$
Auc.2	-0.52	2242.20	2179.47
	$k = 14$	$k = 14$	$k = 14$
Auc.3	-0.82	820.79	815.35
	$k = 16$	$k = 14$	$k = 16$

estimates.

Now that the auctioneer has an estimate of her expected revenue, she might want to capitalize on this estimate. For example, if the auctioneer were to auction another Sony Pcg-xg29k Vaio Notebook (like the one in Auc.2) she could use the estimated value as a reference for some of the facilities provided by on-line auction houses, such as the "Early Close Time" and the "Buy Price". The first is employed when a seller wants to terminate the auction as soon as the current bid reaches a satisfactory level. The second facility can be used to set an amount and simply sell the item to the first person who posts that price.

Internet auction houses clearly explain what a buy price or an early close time are and that they might be used to shorten the time it takes to sell the object. However, they do not explicitly suggest how to set the one or the other. In some cases, they do suggest to navigate around in search of auctions for comparable items, to see what price was reached in those auctions. Hence, in our example the seller could follow the evolution of a new auction and terminates it as soon as the current bid equals the estimated expected revenue equal to \$2179.47 in Auc.2. Similarly, she could use the estimated value as a "Buy Price".

4 Conclusions

This paper is dedicated to the analysis of theoretical and empirical auctions with numerous bidders. In the past few years Internet auction houses have become one of the most popular marketplaces for the sale of goods such as collectibles and second-hand items. If we want to analyze these auctions, we need to adapt the classical theory of auctions to a large number of bidders. In the auction literature the effect on the seller's revenue of increasing the number of bidders has been analyzed only from a theoretical point of view

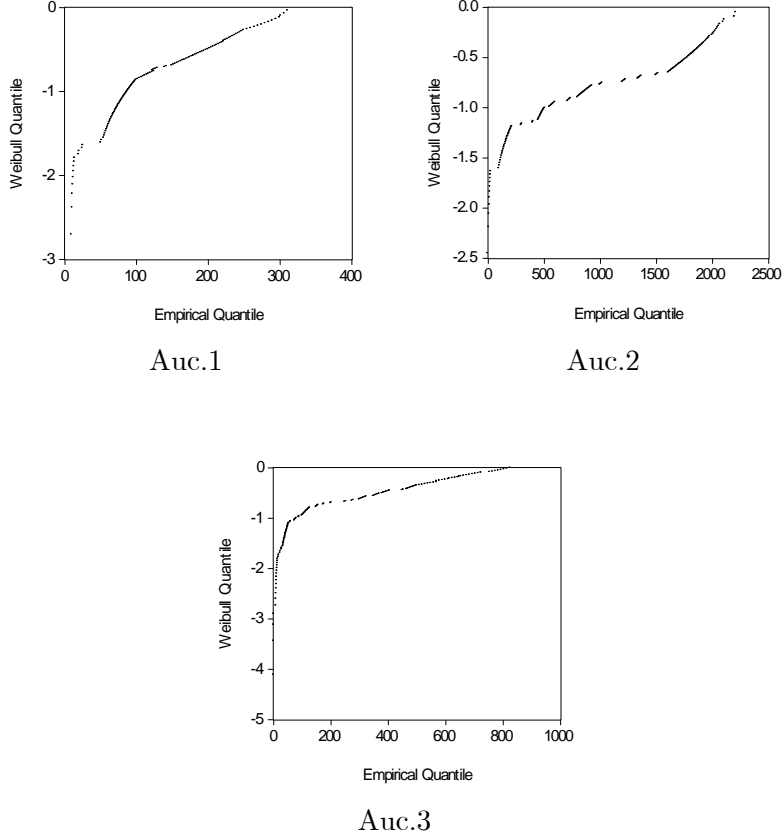


Figure 3: QQ-plots bids against Weibull quantile for estimated gammas

when the number becomes unbounded.

In the limit case of an infinite number of bidders, the seller gains a price equal to the highest possible valuation. Hence, when the number of bidders is large one might either directly compute the seller's expected revenue or approximate it by the upper endpoint of the distribution of valuations. Both suggestions are, however, of little practical use. Direct computation is extremely difficult to work with, unless a particularly easy distribution function of valuations applies. And the approximation by the upper endpoint results in a large approximation error. With the aid of extreme value theory, we have shown how to obtain a simple approximate expression for the seller's expected revenue, and for the equilibrium bid function in a first-price

sealed-bid auction.

In the last section we have utilized on-line second-price auction data to estimate the expected revenue to the auctioneer, relying only on data from a single auction. In this way we avoided the necessity, typical for the empirical auction literature, of pooling data from different auctions of comparable items.

We have also shown that in order to estimate the expected revenue to the auctioneer, we do not need to assume a specific distribution function for the bidders' valuations. We only needed to assume that the distribution of valuations is in the domain of attraction of an extreme value distribution. Such hypothesis is confirmed by the empirical results. More precisely, we have confirmed that for the three data sample we used, the distributions of the valuations (bids) is in the domain of attraction of the Weibull extreme value distribution. In this way we have also verified that the distribution of valuations possess a finite upper endpoint, as it is often assumed in auction literature.

It is important to note that our procedure can also be used directly to estimate the distributions of the highest bids for other auction formats. We chose second-price auctions because in that format the optimal bids are the valuations. Nonetheless, our method is also applicable to (large sets of) data from first-price auctions to estimate the distribution of the largest bids. Our technique is, therefore, complementary to the one used in Guerre, Perrigne and Vuong (2000), where kernel density estimators are used to study the distribution of observed bids from first-price auctions. As they point out, the use of kernel estimators is, by its own nature, problematic for the largest bids, which are those we focus on and for which our technique has a comparative advantage.

As it often happens in the analysis of complex problems, we have made a number of assumptions. Some of these assumptions (such as the use of distributions of valuations with finite upper endpoint) are neither necessary nor restrictive. Others might lead to discrepancies between the theory and the applications. For example, the study of dependent valuations and valuations with different distribution functions should be considered. Another interesting problem would be to look at auctions with simultaneously many bidders and many sellers.

Appendix A

In this appendix we briefly describe on-line auctions and provide details for those we used. It is fair to say that Internet auctions are a mixture between the English auction and the second-price auction. A bidder can choose to monitor the bidding process herself and bid as in an English auction. But since many of the auctions run over multiple days, a bidder can also submit her maximum valuation like in the second-price auction, and leave the bidding, called proxy bidding, to the computer program run by the Internet auction house.

Judging by the answers to the FAQ sections for different Internet auction houses, a lot of emphasis is placed on explaining the idea that bidding one's valuation, through the proxy bidding system, is a dominant strategy for the participants. Hence, we can assume that most bids are placed as in a second-price auction, and therefore are representative of the bidders' valuations.

On-line auctions are run over a fixed time period, traditionally equal to one week. A practical issue we need to consider is that in some Internet auction setups, there exists an incentive for under-reporting one's true valuation due to the fixed time period over which the auction runs. Agents who do not want to reveal their private information bid at the very last moment, so that others will not be able to exploit this information by over-bidding. Some auction houses preclude this so called "sniping" effect by automatically extending the time horizon of the auction at the last bid.

Yahoo!Auctions allows for auctions in which the closing time is automatically extended by five extra minutes whenever a bid is posted within the last five minutes of the auction expiration. This option may or may not be available. In order to avoid sniping and its influences, which have been analyzed in Roth and Ockenfels (2000), we have chosen data from items sold on Yahoo!Auctions with an automatic extension.

For any auctioned item, Yahoo!Auctions reports an extended bid history after the auction has finished. This history contains the bids and the proxy bids¹³ of the participating bidders. We use the total collection of bids as recorded. However, these records may contain different bids posted by the same person, if for example the bidder misunderstood the way the system works. Hence, the same bidder may have posted different bids at different times, which can be detected because the complete bid history contains as identification the E-mail addresses from the participants. We have verified

¹³In the estimation procedure we implicitly assume that these automated bids represent the bids entered by other bidders.

that our estimates are robust to the presence of multiple bids by the same bidder¹⁴.

The first auction (Auc.1) concerned a ticket for the Backstreet Boys concert (Minneapolis). It was started on 2000, December 07 at 18:12, and it ended on December 12 at 17:12. The total number of submitted bids was 69, to which 63 proxy bids were added by the system, resulting in a total amount of 132 collected bids, as listed in the extended bids history¹⁵. The ticket was sold at \$311.5 (to be compared with the official price of \$136.50).

The second auction (Auc.2) was for a laptop computer, a new Sony Pcg-xg29k Vaio Notebook. It was started on 2000 November 01 at 11:41, and it ended on November 4 at 11:41. The total number of entered bids was 73 and the item was sold at \$2210.01 (to be compared with the official price of \$3599.99). The extended bids history provides us with a total list of 139 bids, both proxy and non-proxy.

The third auction (Auc.3) was for a laptop, a Dell latitude Notebook. It was started on 2001, April 11 at 15:51, and it ended on April 19 at 20:06. The total number of submitted bids was 83 and the selling price was \$825.00. The extended bids history provides us with a total list of 138 bids, both proxy and non-proxy.

Appendix B

In this appendix we extend extreme values statistics to obtain estimators for the expected highest and second highest order statistics. Surprisingly as it may sound, estimation of the expected extreme order statistics has not been dealt with in the statistics of extreme values. A lot is known about estimating the tail index and the norming constants of the limit distribution but the expected values have not been provided with estimators.

We concentrate on the question of how to estimate the expected revenue from a single auction (a complete treatment of all the statistical issues involved is outside the scope of the present paper, and we focus only on the estimators). As shown in Corollary 5, the expected revenue depends on the coefficients α , u_n and s_n . Estimators for these coefficients exist, but now need to be combined to obtain an estimator for the expected revenue.

Several estimators for the tail index have been proposed in the extreme

¹⁴ Details can be obtained on request from the authors.

¹⁵The official extended bids history contained 261 data. However, many of these bids are ties due to the computerized system. We, therefore, removed the multiple bids in order to avoid artificial ties.

values literature; see Embrechts et al. (1997), Reiss and Thomas (2001) and references therein. Most estimators use a semi-parametric approach whereby only information from the upper tail of the distribution function $F(x)$ is used, because only in the tail area the approximation by the limit laws for the highest order statistics is appropriate. In those cases in which more information is available, we could estimate a fully parametric model and infer the EVT parameters. But in our case we do not have such information available. A fully non-parametric method is also not advisable since typically in the tail area few observations are present, and hence neglecting the parametric information from the limit distribution is inefficient.

One particular class of semi-parametric estimators is the logarithmic moment estimators initiated by Hill (1975). The Hill estimator is restricted to distributions with infinite upper endpoint. Dekkers, Einmahl and de Haan (1989) extended the idea of the Hill estimator to encompass the limit laws with finite upper endpoint. An alternative for the moment based estimator of the tail index is the maximum likelihood estimator (MLE) developed in Smith (1984). However, the use of MLE procedure is valid only when $\alpha > 2$. In particular, when $\alpha \leq 1$ the likelihood function has no global maximum, so that MLE estimates are non-existent or not consistent. Since we can not exclude the possibility of α being lower than 2, we will not rely on the MLE approach.

Although the MLE approach has been improved in the meantime, we concentrate on the logarithmic moment estimator proposed by Dekkers et al. (1989), which we indicate by DEDH. This estimator is a function of the sample size n and of the number k of upper order statistics $X_{(1)} \geq X_{(2)} \geq \dots \geq X_{(k)} \geq \dots \geq X_{(n)}$ to be used. Define $\alpha \equiv -1/\gamma$, then the DEDH estimator is

$$\hat{\gamma}_n = H_n^{(1)} + 1 - \frac{1}{2} \left(1 - \frac{(H_n^{(1)})^2}{H_n^{(2)}} \right)^{-1} \quad \text{and} \quad \hat{\alpha}_n = -\frac{1}{\hat{\gamma}_n} \quad (19)$$

where

$$H_n^{(1)} = \frac{1}{k} \sum_{j=1}^k (\log X_{(j)} - \log X_{(k+1)}) \quad (20)$$

and

$$H_n^{(2)} = \frac{1}{k} \sum_{j=1}^k (\log X_{(j)} - \log X_{(k+1)})^2. \quad (21)$$

We indistinctly indicate by DEDH both $\hat{\alpha}_n$ and $\hat{\gamma}_n$, whenever no confusion is possible. The name of moment estimator stems from the fact that (20) and (21) can be thought of as empirical moments of the variable $\log X$.

The DEDH estimator applies to extreme value problems in general. When the distribution function is in the domain of attraction of the Weibull extreme value distribution, $F(x) \in MDA(\Psi_\alpha)$, then for the estimated tail index it has to hold $\hat{\gamma}_n < 0$. The other two limit laws are covered by the situations $\hat{\gamma}_n = 0$ and $\hat{\gamma}_n > 0$. Thus, the sign of the estimate can be taken as an indication of the appropriate limit law. This fact was used in the application to test for the assumption of bounded support for the valuations.

We also employed QQ-plots as an alternative means for checking this presumption. In a QQ-plot the empirical quantiles, i.e. the ordered data, are plotted against the quantiles of a theoretical distribution. If the data were drawn by a distribution of the type of the theoretical distribution, then the QQ-plot should depict a straight line. If the QQ-plot does not resemble a straight line, the deviations from linearity may provide an indication of the nature of the mismatch. See Embrechts et al. (1997) for a concise and clear discussion on the QQ-plot.

Suppose we use a QQ-plot of a data sample against the Gumbel extreme value distribution¹⁶, the so called double-log plot. If the data were generated by a distribution in the domain of attraction of a Gumbel, then the QQ-plot should be approximately linear in the area where the highest values are plotted, that is in the North-East part of the plane. On the contrary, if the generating distribution is in the domain of attraction of a Weibull, then the plot curves upward, while in the case of a Fréchet law, it curves downward, as shown in Figure 4. This figure contains the theoretical QQ-plot for the extreme value Weibull (a) and Fréchet (b) distributions against the Gumbel distribution. Thus, the curvature is indicative of the limit law.

We now move on to the problem of finding estimators for the sequences s_n and u_n . A first natural choice is to look at estimators which correspond to the possible theoretical values for s_n and u_n . Recall that when $F(x) \in MDA(\Psi_\alpha)$ appropriate theoretical values for the norming constants are given by

$$s_n = \left(z - F^{\leftarrow} \left(1 - \frac{1}{n} \right) \right)^{-1} \quad \text{and} \quad u_n = z$$

where z is the upper endpoint of the distribution, and $F^{\leftarrow} (1 - 1/n)$ is the inverse quantile function computed at the $1 - 1/n$ probability level. Hence,

¹⁶We can not use the Weibull extreme value distribution at this stage because we would have to fix a particular value for its tail index.

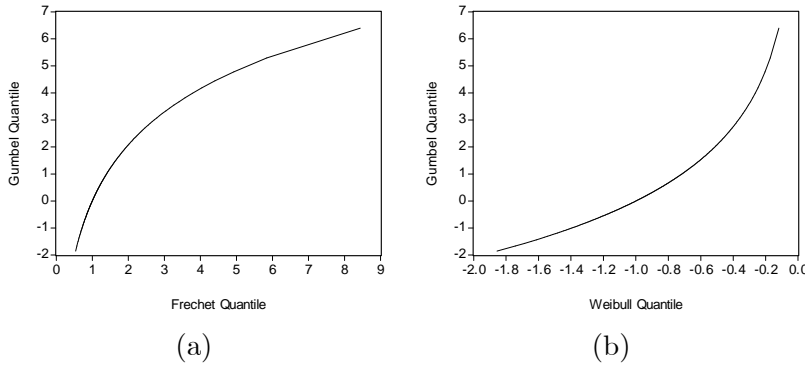


Figure 4: QQ-plots

estimators for the upper endpoint and the $1 - 1/n$ quantile can be used as estimators for the norming sequences. We use the estimators developed in Dekkers et al. (1989). As a result, natural choice estimators for the norming sequences are

$$\hat{s}_n = \left(\hat{z} - F^{\leftarrow} \left(\widehat{1 - \frac{1}{n}} \right) \right)^{-1} = \left(k^{\hat{\gamma}_n} \left(1 - \frac{1}{\hat{\gamma}_n} \right) H_n^{(1)} X_{k+1} \right)^{-1} \quad (22)$$

and

$$\hat{u}_n = \hat{z} = \left(1 - \frac{1}{\hat{\gamma}_n} \right) H_n^{(1)} X_{k+1} + X_{k+1}. \quad (23)$$

For further details on the derivation of the estimators \hat{s}_n and \hat{u}_n , we refer the reader to Caserta (2002) and the references therein.

On the basis of the above estimators we obtain estimators for $E[M_n]$ and $E[X_{(2)}]$. Substituting the three estimators (19), (22) and (23) into expressions (8) and (14) yields the following estimators.

Definition 7 Suppose $z \equiv \sup\{x : F(x) < 1\} > 0$, and $F(x) \in MDA(\Psi_\alpha)$ then for $k(n) \rightarrow \infty$ and $k(n)/n \rightarrow 0$, the estimators for $E[X_{(2)}]$ and $E[M_n]$

are respectively

$$\begin{aligned} \widehat{E}[X_{(2)}] &= \widehat{u}_n - \frac{\Gamma(2 + \widehat{\alpha}^{-1})}{\widehat{s}_n} = \\ &= \left[1 - k^{\widehat{\gamma}_n} \Gamma(2 - \widehat{\gamma}_n)\right] \left(1 - \frac{1}{\widehat{\gamma}_n}\right) H_n^{(1)} X_{k+1} + X_{k+1}. \end{aligned} \quad (24)$$

$$\begin{aligned} \widehat{E}[M_n] &= \widehat{u}_n - \frac{\Gamma(1 + \widehat{\alpha}^{-1})}{\widehat{s}_n} \\ &= \left[1 - k^{\widehat{\gamma}_n} \Gamma(1 - \widehat{\gamma}_n)\right] \left(1 - \frac{1}{\widehat{\gamma}_n}\right) H_n^{(1)} X_{k+1} + X_{k+1}. \end{aligned} \quad (25)$$

All extreme value estimators depend on the choice of the number k of upper order statistics X_{k+1} . The number k varies with the sample size and it is, in general, different for the different estimators. Moreover, the number of order statistics has to satisfy two conditions: 1) $k \rightarrow \infty$ and 2) $k/n \rightarrow 0$. These conditions state that when the sample size increases the number of order statistics increases as well, although at a lower rate. In fact, if too few upper order statistics are used the variance dominates, but if too many upper order statistics are used a considerable bias appears. The optimal choice for k can be achieved by balancing the bias squared and variance through a minimization of the asymptotic mean squared error of the estimates. Often, the choice for k is made with the aid of a so called estimates plot; see Embrechts et al. (1997), Reiss and Thomas (2001) and references therein for a detailed exposition. This is a plot which contains on the x -axis the number k and on the y -axis the value of the estimator for the corresponding k . The choice for k is where the graph becomes relatively stable but before wandering off in one or the other direction due to the bias. We use this plot in the applications¹⁷, since data limitations prevent us from using methods like the bootstrap or sub-sample bootstrap advocated in Danielsson et. al. (2001).

References

- Caserta, S. (2002). Extreme Values in Auctions and Risk Analysis. PhD thesis. Tinbergen Institute, Erasmus University Rotterdam. Research Series 287.

¹⁷See Figure 2 for an example.

- Danielsson, J., L de Haan, L. Peng and C.G. de Vries (2001). Using a bootstrap method to choose the sample fraction in tail index estimation. *Journal of Multivariate Analysis* **76**, 226–248.
- Dekkers, A.L.M., J.H.J. Einmahl and L. de Haan (1989). A moment estimator for the index of an extreme-value distribution. *Annals of Statistics* **17**, 1833–1855.
- Donald, S. and H. Paarsch (1993). Piecewise pseudo-maximum likelihood estimation in empirical models of auctions. *International Economic Review* **34**, 121–148.
- Embrechts, P., C. Klueppelberg and T. Mikosch (1997). *Modelling Extremal Events for Insurance and Finance*. Springer-Verlag. Berlin.
- Guerre, E., I. Perrigne and Q. Vuong (2000). Optimal nonparametric estimation of first-price auctions. *Econometrica* **68**(3), 525–574.
- Haan, L. de (1970). On Regular Variation and its Application to the Weak Convergence of Sample Extremes. PhD thesis. Mathematical Centre Amsterdam. Mathematical Centre Tracts 32.
- Haile, P.A. and E. Tamer (2003). Inference with an incomplete model of english auctions. *Journal of Political Economy* **111**, 1–51.
- Harris, M. and A. Raviv (1981). Allocation mechanisms and the design of auctions. *Econometrica* **49**(6), 1477–1499.
- Hill, B. M. (1975). A simple general approach to inference about the tail of a distribution. *Annals of Statistics* **35**, 1163–1173.
- Holt, C.A. (1979). Uncertainty and the bidding for incentive contracts. *American Economic Review* **69**(4), 697–705.
- Klemperer, P. (1999). Auction theory: A guide to the literature. *Journal of Economic Surveys* **13**(3), 227–286.
- Laffont, J., H. Ossard and Q. Vuong (1995). Econometrics of first-price auctions. *Econometrica* **63**(4), 953–980.
- Leadbetter, M.R., G. Lindgren and H. Rootzen (1983). *Extremes and Related Properties of Random Sequences and Processes*. Springer-Verlag. New York.

- Lucking-Reiley, D. (2000). Vickrey auctions in practice: From nineteenth-century philately to twenty-first-century e-commerce. *Journal of Economic Perspectives* **14**(3), 183–192.
- McAfee, R.P. and J. McMillan (1987a). Auctions and bidding. *Journal of Economic Literature* **xxv**, 699–738.
- Milgrom, P.R. and R.J. Weber (1982). A theory of auctions and competitive bidding. *Econometrica* **50**(5), 1089–1122.
- Paarsch, H.J. (1992). Deciding between the common and the private value paradigms in empirical models of auctions. *Journal of Econometrics* **51**, 191–215.
- Reiss, R.D. and M. Thomas (2001). *Statistical Analysis of Extreme Values*. 2nd ed.. Birkhauser. Basel.
- Resnick, S.I. (1987). *Extreme Values, Regular Variation, and Point Processes*. Springer-Verlag. New York.
- Roth, A. E. and A. Ockenfels (2000). Last minute bidding and the rules for ending second-price auctions: Theory and evidence from a natural experiment on the internet. *National Bureau of Economic Research Working Paper*.
- Smith, R.L. (1984). Threshold methods for sample extremes. In: *Statistical Extremes and Applications* (J. Tiago de Oliveira, Ed.). Reidel. Dordrecht. pp. 621–638.