Inferring Ascending Auction Participation from Observed Bidders

Brennan C. Platt*

September 25, 2015

Abstract

Participation in internet auctions goes well beyond those who place a bid. Participants arrive in random order, and if the auction's standing price has already exceeded a participant's valuation, she will not bid. Even so, her unreported valuation is a relevant part of demand for the item, and in an alternate random order, her bid would have been registered. Assuming a Poisson distribution of participants, I provide a method to estimate the average number of participants from the average number of bidders per auction. This enables non-parametric estimation of the distribution of bidder valuations from either the distribution of closing prices or the distribution of all observed bids.

JEL Classifications: C73, D44, D83

Keywords: Auction participation, estimation of number of bidders

1 Introduction

At their introduction, a promising aspect of online auctions was the ability to draw a large volume of bidders worldwide, rather than being confined to small gatherings in auction houses. In practice, this seems to be true for the overall number of bidders in the market, but the number of bidders per auction can be underwhelming. To illustrate this, Table 1 reports statistics on eBay bidders in a sample of 5,718 popular new-in-box products.[1] The average product drew 535 unique bidders into an auction during the year; yet the average auction only had 5.6 unique bidders.

Even so, the number of *participants* in an auction could potentially be much larger than the number of actual *bidders*, where a participant is defined as someone who values the item at more than the seller's reserve price and examines the auction listing while it is active. If the auction were conducted via second-price sealed bid, any participant would find it weakly optimal to place a bid equal to her valuation. However, in an ascending auction like

*Department of Economics, Brigham Young University, brennan_platt@byu.edu

[1] New-in-box products have the virtue of being perfectly substitutable across auctions; thus, one would expect a bidder in one such auction to have interest in other auctions of the same product. Products are identified by sellers from one of several commercial catalogs, much like a UPC or SKU code.

Table 1: eBay Bidder Activity

	N	Min	Median	Max	Mean	Std Dev
Unique bidders per auction	5,718	1.02	5.31	18.93	5.58	2.59
Unique bidders per product	5,718	7	330	29,570	535	959
Listings, by product	5,718	50	83.5	7,601	140.9	253

Notes: Data are drawn from 805,627 eBay auctions of new-in-box products, with at least 50 listings per product during the period October 1st, 2013, to September 30th, 2014. The unit of observation is each unique product, averaging across auctions of each product in the first row.

eBay's standard listings, a participant views the *standing price*, which is the second highest of all previously placed bids, before placing her own bid. Thus, if the standing price has already passed her valuation, the participant will abstain from bidding.

This distinction between bidders and participants has important consequences. First, the number of participants provides a more accurate depiction of the size of the potential market, and thus more clearly depicts the gains from trade in that market. Second, the number of participants enables a more accurate estimate of the distribution of valuations, which is an important input in determining the optimal reserve price or simulating counterfactual mechanisms (such as the inclusion of a "best offer" or "buy-it-now" option). If the number of participants n is known, the distribution of bidder valuations can be recovered from the observed distribution of second highest prices (*i.e.* the closing price of each auction). However, this critically relies on knowing how many participants were beaten to reach that closing price. More participants will increase the likelihood of closing at a high price, even if many of these participants abstain from bidding.

In this paper, I provide a precise characterization of the number of participants based on the observed number of bidders. This can then be used to non-parametrically infer the distribution of bidder valuations, either from the distribution of closing prices or the distribution of all observed bids. I then use Monte Carlo simulations on a known distribution to generate sample auction data, and run the proposed procedure to demonstrate its accuracy.

Beyond taking place in an independent private values setting, my model relies on three assumptions. First, participants enter the market in random order (uncorrelated to their valuation of the product, for instance). This is easy enough to verify in any dataset that includes the order in which bids were received. In particular, the k^{th} bid should be equally likely to beat the current winner as it is to fall short.[2]

[2] Whatever the standing price p is after $k-1$ bids, the current winner's bid and the k^{th} bid (if one is placed) are both random draws from a truncated distribution $F(v|v > p)$. Since these are independent

Second, for each auction, the number of participants is drawn from a fixed distribution. This allows the realized numbers of bidders and of participants to vary across auctions, but the constant mean of the distribution indicates the typical size of the buyer's side of the market. For ease of exposition and computation, I employ the Poisson distribution throughout the paper. Indeed, much of the recent theory on repeated auctions share this assumption that bidders are Poisson distributed,[3] making my participation adjustment immediately applicable to any empirical estimation of their model primitives. Even so, the Appendix indicates how any discrete distribution may be employed instead. Moreover, for any assumed distribution, one can readily test whether the theoretical distribution of bidders under the estimated distribution of participants fits the observed distribution of bidders.

Third, in the analysis below, I proceed as if each bidder places only one bid, equal to her valuation, but this generalizes much further. As explained in de Haan, *et al* (2013), if all bidders are using proxy bidding systems (where the auctioneer automatically enters new bids on ones behalf up to the bidder's specified maximum price), then one may treat the final bid as if it were the only bid, but submitted in the order that the initial bids were placed. In practice, bidders also manually raise their bids; but so long as bidders immediately respond to being outbid (up until their valuation), then this still has the same effect as a proxy bidding system.[4]

The divergence between participants and bidders was first addressed in Song (2004). She shows that the distribution of bidder valuations can be identified from the joint distribution of the second- and third-highest bids, even if the number of participants is unknown and stochastic. In practice, this approach requires a semi-nonparametric method which is sometimes sensitive to the chosen functional form. Also, Song was not focused on estimating the number of participants.[5]

Canals-Cerdá and Pearcy (2013) tackle the same challenge using only the distribution of the second highest bid and the timing of bid arrivals. They assume that participants stochastically arrive according to a Poisson process, but only bid if their valuations exceed

draws, the probability that one exceeds the other evaluates to 1/2.

[3]See Wang (1993); Kultti (1999); Etzion, et al. (2006); Nekipelov (2007); Ingster (2009); Ambrus, et al. (2014); Bodoh-Creed (2013); Hammond (2013); Bauner (2015); Coey, Larsen, and Platt (2015), among others.

[4]Whether a bidder's final bid actually reflects the bidders' true valuation remains a debated question. Zeithammer and Adams (2010) propose five tests of bidding behavior and document that inexperienced bidders consistently violate the predicted behavior, though Hortacsu and Nielsen (2010) point out that these indirect tests do not directly examine the "bid equals valuation" assumption, and offer alternative interpretations of their results which would not reject the assumption. As additional theory develops to explain the behavior of inexperienced bidders, future work may incorporate it into estimations of participation.

[5]Song (2004) briefly presents a strategy to non-parametrically identify the probability of n participants in her Appendix B, but the identification is very weak in practice unless a parametric form is assumed for $Pr(n)$ such as the Poisson distribution.

the standing price. These bid arrivals will systematically slow down if the standing price is higher, so even if the final transaction price is the same, the variation in the standing prices will reveal information about the participants who did not bid. While this approach also does not estimate the number of participants, it does rely on similar structure as my model, in that the total number of participants will be Poisson distributed.

de Haan, *et al* (2013) focus on the reverse, estimating the number of participants but remaining silent on the distribution of valuations. They impose no structure on the distribution of participants, but as a consequence, their predicted relationship between participants and bidders only holds asymptotically. For smaller numbers of bidders, I demonstrate that their approach will systematically overestimate the number of participants (if those are drawn from a Poisson distribution). Indeed, this overestimation is likely to matter in typical empirical applications. With less than 19 bidders per auction — as with every product in Table 1 — the number of participants predicted by de Haan, *et al* is 8% greater than the number predicted in my procedure.

This relates to a broader literature on endogenous auction participation due to the cost of preparing bids or acquiring a signal of the item's value (See Section 6.3 of Athey and Haile, 2007, for a survey). This literature distinguishes between *potential bidders* and *participating bidders*, where the latter are a subset of potential bidders whose expected winnings are large enough to justify the participation cost; all others will abstain. In my setting, all bidders want to participate; abstention occurs only due to bad luck, because the bidder arrived after the standing price had already passed her maximum bid. Thus, all bidders are participating bidders (or *participants* for short), but the unlucky ones are *unobserved* by the econometrician.

This distinction is important in interpreting auction data. If there are n potential bidders, but k decide the cost of participation is too high, then the closing price will be the second highest of $n-k$ valuations.[6] In contrast, if there are n participants, but the standing price blocks k from them from bidding, the closing price is still the second highest of all n valuations. In my model, every participant has a positive probability of placing a bid; hence, all participants are empirically relevant in backing out the distribution of valuations.

The paper proceeds with the development of the estimation method in Section 2. Section 3 then examines the accuracy of the method via Monte Carlo simulations on various sample sizes. Section 4 then concludes.

[6] Of course, the distribution of participating bidders would typically be a truncation of the distribution of potential bidders. In estimating the distribution these valuations, the abstaining bidders can generally be ignored. The main empirical purpose of including them in the model is to estimate participation costs.

2 Analysis

Consider a second-price ascending auction with an unknown number n of participants, where n is Poisson distributed with mean λ. The n participants arrive to the auction in a random order (with all permutations equally likely), which is their only opportunity to bid. If a participant decides to bid, she submits her true valuation, as in a sealed-bid second-price auction, allowing a proxy bidding system to raise her bid up to that submitted valuation if she is outbid. Bidder valuations v are independently drawn from the same continuously differentiable distribution, represented by cumulative distribution function $F(v)$ with compact support $[\underline{v}, \bar{v}]$, where $\underline{v}$ is the opening price of the auction. Of course, this precludes estimating the participation or valuations of those who value the item at less than $\underline{v}$. Note also that the distribution $F(v)$ is independent of the realized number of participants n.

The *standing price* begins at the reserve price: $p_0 = \underline{v}$. After the k^{th} arrival, the standing price is raised to the second highest of all preceding bids: $p_k = V_{(k-1)}$, which is the $k-1^{th}$ order statistic of k draws from the distribution $F(v)$. If a participant arrives after k others and finds that her valuation v is less than p_k, she will abstain from bidding. Otherwise, she places her bid, which will either promote her to the current first place (so p_{k+1} rises to the previous highest bid) or the current second place (so p_{k+1} rises to her v, but she will not win the auction).

2.1 Estimating Participation

Suppose a participant has valuation v and arrives after k of the n other participants. The probability that she places a bid is:

$$Pr(bid|v, n, k) \equiv F(v)^k + kF(v)^{k-1}(1 - F(v)) \tag{1}$$

That is, in order to place the bid, her value must be larger than all of the others (the first term) or larger than all but one of the others (the second term). If two or more valuations exceed her own, then $v < p_k$ and she will abstain. This does not depend on the total number of participants, n, but only the number who have already bid, k. Note that if $k = 0$ or $k = 1$, she places a bid for sure.

Since each permutation is equally likely, this participant has an equal chance of k taking any value in $\{0, 1, \ldots, n\}$. Thus, the probability that a participant will bid, given a valuation

v while facing n opponents, is:

$$Pr(bid|v,n) \equiv \frac{\sum_{k=0}^{n} F(v)^{k-1}\left(k+(1-k)F(v)\right)}{n+1} \tag{2}$$

$$= \frac{2\left(1-F(v)^{n+1}\right)}{(n+1)(1-F(v))} - F(v)^n. \tag{3}$$

In the numerator of Eq. 2, the summation indicates the probability of bidding (from Eq. 1) in any position to which the bidder may be assigned.

Next, recalling that the number of opponents is Poisson distributed, I calculate the ex-ante probability that a participant with valuation v will bid, averaged across all possible realizations of n:

$$Pr(bid|v) \equiv \sum_{n=0}^{\infty} \frac{e^{-\lambda}\lambda^n}{n!} \cdot \left(\frac{2\left(1-F(v)^{n+1}\right)}{(n+1)(1-F(v))} - F(v)^n \right) \tag{4}$$

$$= \frac{2\left(1-e^{-\lambda(1-F(v))}\right)}{\lambda(1-F(v))} - e^{-\lambda(1-F(v))}. \tag{5}$$

In the summation, I multiply the probability of having n opponents times the probability that the participant places a bid (from Eq. 3).

Finally, I average across all possible bidder valuations to get the ex-ante probability that a randomly drawn participant will bid in an auction:

$$Pr(bid) \equiv \int_{\underline{v}}^{\bar{v}} \left(\frac{2\left(1-e^{-\lambda(1-F(v))}\right)}{\lambda(1-F(v))} - e^{-\lambda(1-F(v))} \right) F'(v)dv \tag{6}$$

By change of variables (and recalling that $F(\bar{v}) = 1$ and $F(\underline{v}) = 0$), I obtain:

$$Pr(bid) = \frac{1}{\lambda} \int_0^{\lambda} \left(\frac{2\left(1-e^{-t}\right)}{t} - e^{-t} \right) dt \tag{7}$$

$$= \frac{2}{\lambda} \int_0^{\lambda} \frac{1-e^{-t}}{t} dt - \frac{1-e^{-\lambda}}{\lambda} \tag{8}$$

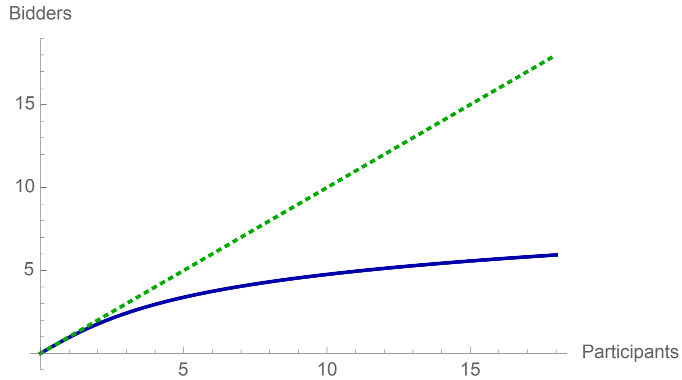
$$= \frac{2}{\lambda} \left(\ln(\lambda) + \gamma + \Gamma(0,\lambda) \right) - \frac{1-e^{-\lambda}}{\lambda}. \tag{9}$$

where $\gamma \approx 0.57721$ is Euler's constant and $\Gamma(0,\lambda) = \int_{\lambda}^{\infty} \frac{e^{-t}}{t} dt$ is the incomplete gamma function, which must be numerically evaluated.

Since there are λ participants on average,[7] each of whom bid with probability $Pr(bid)$,

[7] The distribution of competing participants (as viewed by a participant) and the aggregate distribution of participants (as viewed by the external game theorist) are both Poisson distributed with the same mean λ. This is a general property of Poisson games, as shown by Myerson (1998).

Figure 1: Relationship between Participants and Observed Bidders



Notes: The solid line indicates the average number of observed bidders, $\bar{a}(\lambda)$, for a given number of participants, λ; the dotted line indicates the 45 degree line.

the average number of bidders in a given auction will be:

$$\bar{a}(\lambda) \equiv \lambda \cdot Pr(bid) = 2\left(\ln(\lambda) + \gamma + \Gamma(0, \lambda)\right) - 1 + e^{-\lambda}. \tag{10}$$

The following properties of $\bar{a}(\lambda)$ are easily demonstrated:

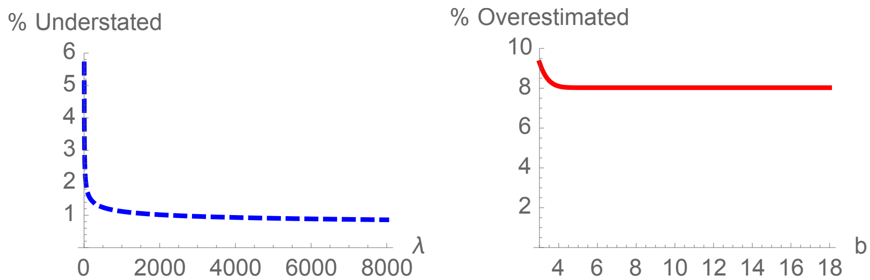
1. $\lim_{\lambda \to 0} \bar{a}(\lambda) = 0$ and $\lim_{\lambda \to \infty} \bar{a}(\lambda) = \infty$.
2. $\bar{a}'(\lambda) = \frac{2\left(1-e^{-\lambda}\right)-\lambda e^{-\lambda}}{\lambda} > 0$ for $\lambda \in (0, \infty)$.
3. $\lim_{\lambda \to 0} \bar{a}'(\lambda) = 1$ and $\lim_{\lambda \to \infty} \bar{a}'(\lambda) = 0$.
4. $\bar{a}''(\lambda) = -\frac{2\left(1-e^{-\lambda}-\lambda e^{-\lambda}\right)-\lambda^2 e^{-\lambda}}{\lambda^2} < 0$ for $\lambda \in (0, \infty)$.

This establishes that the average number of bidders is strictly increasing in (but always below) the true average number of participants. In fact, the rate of increase slows down as the number of participants gets large, yet always continues to grow. Therefore, there is a monotonic transformation between the average number of bidders and the average number of participants.

This relationship is illustrated in Figure 1. Note how quickly the number of bidders underestimates the number of participants: by 50% at $\lambda = 9$, and by 66% by $\lambda = 18$.

By way of comparison, de Haan, *et al* (2013) estimate that λ participants will result in $2\ln(\lambda)$ bidders as λ becomes large. Indeed, this is nearly in agreement with Eq. 10, because $\Gamma(0, \lambda) \longrightarrow 0$ and $e^{-\lambda} \longrightarrow 0$ as $\lambda \longrightarrow \infty$, leaving $2\ln(\lambda) + 2\gamma - 1 \approx 2\ln(\lambda) + 0.154$. Even for smaller values, the discrepancy initially seems minor: for instance, with $\lambda = 12$

Figure 2: Accuracy of the $2\ln(\lambda)$ Approximation



Notes: The dashed line on the left depicts $(\bar{a}(\lambda) - 2\ln\lambda)/(2\ln\lambda)$, the percentage by which the $2\ln\lambda$ approximation understates the expected number of bidders. The solid line on the right depicts $(\bar{a}^{-1}(a) - exp(a/2))/\bar{a}^{-1}(a)$, the percentage by which inverting $a = 2\ln\lambda$ will overestimate the number of participants.

participants on average, the average number of bidders computed here is $\bar{a} = 5.124$, while the expected number of bidders in de Haan, *et al* (2013) is $2\ln(12) = 4.970$, a difference of 3%. The dashed line in the left panel of Figure 2 depicts this percentage difference at each λ. While this discrepancy is decreasing in λ, it is exceedingly slow beyond $\lambda = 100$.

However, the discrepancy becomes more consequential and persistent when used as intended. In practice, one will observe the average number of bidders a and, by inverting the theoretical relationship, will deduce the average number of participants λ. Of course, this must be numerically computed in the case of $\bar{a}^{-1}(a)$. The right panel of Figure 2 reports the percentage difference in the estimated λs. Strikingly, the $2\ln(\lambda)$ approximation consistently overestimates the number of participants by 8% even for very large numbers of bidders. Note that the domain of bidders in the right-hand graph corresponds with the domain of participants on the left-hand graph.

2.2 Estimating Valuations from Closing Prices

Next consider the estimation of participant valuations from observed bids. In a sealed-bid second-price auction, one could take the empirical distribution of all bids (including the winner's) as the estimate for the distribution of valuations. If the highest bid were unavailable (as is often the case), then one could use the empirical distribution of closing prices across multiple auctions as the second highest order statistic of the true distribution of valuations.

However, both of these strategies will produce biased estimates in an ascending auction where the standing price may preclude participants from placing any bid. In this context, the closing price is the second highest valuation of the n participants, not just of the a observed bidders. Using the latter would underestimate the competition which led to that price, and therefore place more density on high valuations than actually exists. Similarly, if one has the full distribution of bids, this is still not representative of the full distribution of participant valuations. Indeed, participants with lower valuations are more likely to be forced into abstention.

Here, I adapt these two approaches to account for the participants who did not bid. Both approaches use the estimated number of participants, λ, as an input, but impose no functional form on the distribution of valuations; rather, this is non-parametrically estimated from the empirical cumulative density of observed bids.

With a known number of bidders a (who are the only participants), the closing price is simply the second highest of the a draws from the distribution $F(v)$, which has a cumulative distribution of $(a + (1 - a)F(v))F(v)^{a-1}$. One could set this equal to the empirical CDF of closing prices, denoted $G(v)$, and solve for $F(v)$. In our setting, however, the closing price is not only the second highest of the a observed bids, but it is also larger than the valuations of the $n - a$ other participants who did not bid. Since the number of participants is a random variable, the CDFs are averaged across the possible realizations of n. The observed distribution of closing prices $G(v)$ will be distributed as follows:

$$\begin{aligned} G(v) &\equiv \left(\sum_{n=2}^{\infty} \frac{e^{-\lambda}\lambda^n}{n!} \cdot (n + (1-n)F(v))F(v)^{n-1}\right) \Bigg/ \left(\sum_{n=2}^{\infty} \frac{e^{-\lambda}\lambda^n}{n!}\right) && (11) \\ &= 1 - \frac{1 - (1 + \lambda(1 - F(v)))e^{-\lambda(1-F(v))}}{1 - (1+\lambda)e^{-\lambda}}. && (12) \end{aligned}$$

The denominator in Eq. 11 accounts for the fact that auctions with less than two participants are excluded, as the second-highest bid is not defined. One can readily verify that $G(v) = 0$ when $F(v) = 0$, $G(v) = 1$ when $F(v) = 1$, and $G(v)$ is strictly increasing in $F(v)$ for any $\lambda > 0$. Therefore, one can uniquely identify $F(v)$ from $G(v)$ at each v. Estimation starts by computing λ from the average number of bidders (by inverting Eq. 10), then concludes by solving for $F(v)$ from the observed distribution of closing prices (by inverting Eq. 12). The latter can be done non-parametrically by specifying a grid of $v_i \in [\underline{v}, \bar{v}]$, and then solving for $F(v_i)$ at each of these grid points. Alternatively, one could impose a functional form on $F(v)$ and use maximum likelihood estimation on $G'(v)$ to determine the parameters of $F(v)$.

The advantage of estimation using $G(v)$ is that, for a given set of participants, the closing price will always be the same, regardless of the permutation of their arrival. The disadvantage is that, with even moderate amounts of participation, very few auctions will close at low prices; thus, the estimate of $F(v)$ will be very imprecise in that range. For example, in the next section, the estimate from simulated data degrades on the lowest 1% of closing prices, but this represents nearly 50% of participant valuations. While infrequently producing a winning bid, these low valuation participants may be of interest to verify whether $\underline{v}$ is an optimal reserve price, for instance.

2.3 Estimating Valuations from All Observed Bids

Alternatively, one can estimate participant valuations across the full domain using the distribution of all observed bids, denoted $H(v)$. Note that in most settings, only standing prices are reported, and so the highest bid is never observed. In fact, even if it were observed, the highest price need not reflect the true valuation of the winner. In a ascending auction, all that is required is to beat the second highest price, so the winner may have only raised his price enough to accomplish this, intending to raise it further only if necessary.

I derive $H(v)$ as the distribution of *non-winning bids*. For this, consider the probability that a participant places a bid but does not win, given that she has valuation v and arrives after k of the n other participants:

$$Pr(bid\,\&\,lose\,|v,n,k) \equiv F(v)^k\left(1-F(v)^{n-k}\right) + kF(v)^{k-1}(1-F(v)). \tag{13}$$

The first term handles situations in which her bid is larger than the preceding k valuations so she becomes the current leader; in order to lose, her valuation must be smaller than at least one of the $n-k$ remaining participants. The second term handles situations in which her valuation is lower than a previously placed bid; she will bid, but will immediately lose.

Averaging across all possible permutations, the probability that a participant with valuation v facing $n \geq 1$ opponents will place a bid but still lose is:

$$\begin{aligned} Pr(bid\,\&\,lose\,|v,n) &\equiv \sum_{k=0}^{n} \frac{F(v)^k\left(1-F(v)^{n-k}\right) + kF(v)^{k-1}(1-F(v))}{n+1} && (14)\\ &= \frac{2\left(1-F(v)^{n+1}\right)}{(n+1)(1-F(v))} - 2F(v)^n. && (15)\end{aligned}$$

Next, in averaging across the possible realizations of the number of opponents, $n=0$ is excluded because, with no other bidders, the second highest price is not defined and no bids are observed. The probability that a participant with valuation v will bid but still lose is:

$$Pr(bid \,\&\, lose\,|v) \equiv \left(\sum_{n=1}^{\infty}\frac{e^{-\lambda}\lambda^n}{n!}\cdot\left(\frac{2\left(1-F(v)^{n+1}\right)}{(n+1)(1-F(v))}-2F(v)^n\right)\right)\Big/\left(\sum_{n=1}^{\infty}\frac{e^{-\lambda}\lambda^n}{n!}\right) \quad (16)$$

$$= \frac{1}{1-e^{-\lambda}}\left(\frac{2\left(1-e^{-\lambda(1-F(v))}\right)}{\lambda(1-F(v))}-2e^{-\lambda(1-F(v))}\right). \quad (17)$$

The cumulative density of observed bids can be derived as:

$$H(w) \equiv \frac{\int_{\underline{v}}^{w} Pr(bid \,\&\, lose\,|v)\cdot F'(v)dv}{\int_{\underline{v}}^{\bar{v}} Pr(bid \,\&\, lose\,|v)\cdot F'(v)dv} \quad (18)$$

$$= \frac{e^{-\lambda}-e^{-\lambda(1-F(w))}+\Gamma(0,\lambda)-\Gamma(0,\lambda(1-F(w)))-\ln(1-F(w))}{\ln(\lambda)+\gamma+\Gamma(0,\lambda)-1+e^{-\lambda}}. \quad (19)$$

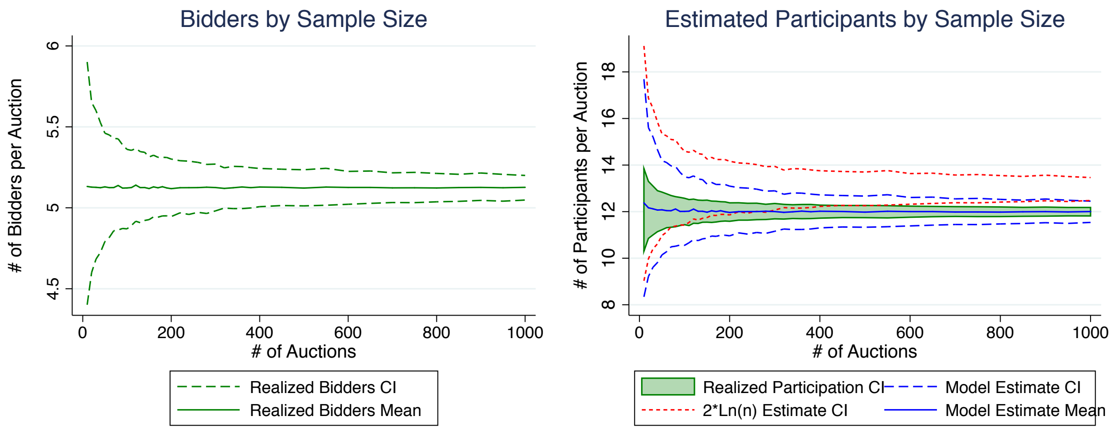
The integral in the denominator rescales the distribution so that $H(\bar{v})=1$. These integrals evaluate nearly in the same way as in $Pr(bid)$. As with $G(v)$, $H(v)=0$ when $F(v)=0$, $H(v)=1$ when $F(v)=1$, and $H(v)$ is strictly increasing in $F(v)$ for all $\lambda>0$. Thus, one can uniquely identify $F(v)$ from $H(v)$ at a given v, whether non-parametrically or imposing a distribution on $F(v)$.

3 Simulation

Simulated data is used to examine the accuracy of the proposed estimation procedures. Note that there is no closed form solution for the average number of participants as a function of the average number of bidders (the inverse of Eq. 10), much less an expression for the variance of this estimate. Instead, data is simulated repeatedly in accordance with the model's auction timing, using a fixed λ and $F(v)$ across all auctions, then the proposed procedures are employed to estimate both λ and $F(v)$, varying the sample size. Specifically, I use a uniform distribution on $[0,1]$ for randomized valuations. Note that the participation estimate, $\bar{a}(\lambda)$, is unaffected by the underlying distribution of valuations. I set $\lambda=12$ resulting in 5.12 bidders on average, in line with the typical eBay auction for new-in-box items.

Each simulation requires three indices: the sample size s, the auction number q, and the iteration i. The sample size indicates how many time the auction is conducted. I increase this from $s=10$ to $s=1000$ auctions, at intervals of 10 through 200, intervals of 20 through 400, and intervals of 50 through 1000. The sequence of simulated auctions is then indexed with $q\in\{1,2,\ldots,s\}$. For every sample size, the simulation is repeated in 1000 iterations to allow the construction of confidence intervals.

Figure 3: Participation Estimate from Simulated Data



3.1 Estimating Participation

First, consider the participation estimate. Each auction is simulated by randomizing anew the number of participants $p^i_{s,q}$ (drawn from a Poisson distribution with mean λ), the valuations of each participant $v^i_{s,q}(k)$ for $k \in \{1, \ldots, p^i_{s,q}\}$ (drawn from the uniform distribution), and their order of arrival $r^i_{s,q}(k)$ (with equal weight on all permutations). I then determine the number of participants who place a bid, denoting this $a^i_{s,q}$, and average this across the sample of auctions: $a^i_s = \sum_q a^i_{s,q}/s$. This process is repeated 1000 times for each s.

The solid line in the left panel of Figure 3 reports for each sample size the average number of bidders per auction across the 1000 iterations: $a_s = \sum_i a^i_s/1000$. This always remains close to 5.12 bidders per auction. Also shown in dashed lines are the 95% and 5% confidence intervals, which is to say that a^i_s lay within this interval in 900 of the 1000 iterations.

From each a^i_s, I then compute the estimated number of participants from my model, denoted $\hat{\lambda}^i_s \equiv \bar{a}^{-1}(a^i_s)$. The solid line in the right panel of Figure 3 reports $\hat{\lambda}^i_s$ at each s, averaged across the 1000 iterations, which the dashed lines denote the 95% and 5% confidence intervals of the estimate. Note that the mean lies almost exactly at the true parameter value of $\lambda = 12$. The confidence interval is initially wide, but is within ± 1 participant of the true value λ by the time $j = 200$ — a number of auctions that can easily be obtained for popular items on eBay.

For a better context in judging the accuracy of the participation estimate, the confidence interval of the average realized number of participants $p^i_s = \sum_q p^i_{s,q}/s$ is also indicated in the shaded area on the right panel of Figure 3. In this simulated data, the realized number of participants in an auction is known (but not used in the estimation). The average number of participants is relevant because, if it were known, it would represent the most accurate estimate of λ. As can be seen in the figure, at each sample size, the confidence interval for realized participation is about 40% of the size of the confidence interval for the estimated participation.

For a comparison to the $2\ln(\lambda)$ estimate provided by de Haan, *et al* (2013), the dotted lines on the right panels of Figure 3 indicate the confidence intervals from $\hat{\lambda}_s^i = e^{a_s^i/2}$. Note that the true value of $\lambda = 12$ lies outside the confidence interval beyond $s = 250$. Moreover, for large sample sizes, this confidence interval does not even intersect the confidence interval for my estimate.

3.2 Estimating Valuations: Closing Prices

A similar evaluation is now presented of how closely the estimation process on the distribution of closing prices can recover the true (uniform) distribution of bidder valuations. This proceeds by computing the empirical cumulative distribution function of closing prices $G_s^i(v)$ for each sample size s and each iteration i. Equation 12 is then used to compute the $\hat{F}_s^i(v)$ that is consistent with the observed $G_s^i(v)$. Since Eq. 12 requires the average number of participants as a parameter, the estimate $\hat{\lambda}_s^i$ is used in its place.

The results are illustrated in Figure 4 for a sample size of 360 auctions. The left panel depicts the observed distribution of closing prices, averaged over 1000 iterations in the solid line and with the 95th and 5th percentiles in the dashed lines. The right panel depicts the estimated distribution of valuations across all bidders, with a solid line for the average and dashed lines for the confidence intervals. For convenience, the actual (uniform) distribution is plotted with the x symbol.

Both graphs are truncated because less than 1% of auctions close at prices below 0.5. Moreover, small changes in $G(v)$ have a disproportionate effect on $F(v)$ in this range of $v < 0.5$. These result in rather noisy estimates; however, this inaccuracy in the low valuation range is not particular to this model, but applies to any estimation of the full distribution from a distribution of an order statistic. Menzel and Morganti (2013) study this problem in depth, concluding that the optimal rates for nonparametric estimation of bidder valuations are much slower than the $\sqrt{n}$ rate. Furthermore, this can slow down the convergence rates of estimates for expected revenue or optimal reserve prices.

Figure 5 illustrates how accuracy improves as the sample size is increased. For each sample size, the range of the confidence interval is averaged across all prices greater than 0.5. For the estimates from closing prices, this average of the confidence interval range is depicted with a dashed line. The accuracy greatly improves through 200 auctions, then

Figure 4: Estimated Distribution of Valuations using Closing Prices

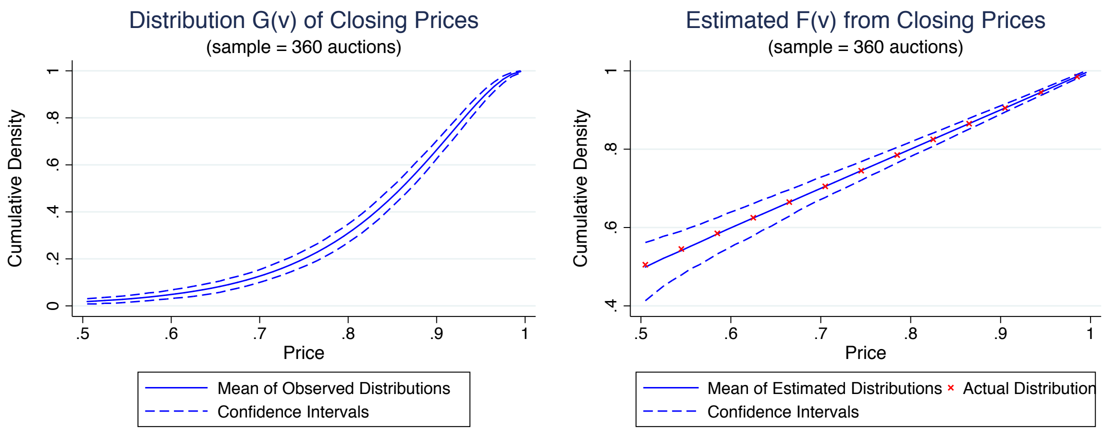


Figure 5: Range of Estimated Distribution Confidence Intervals

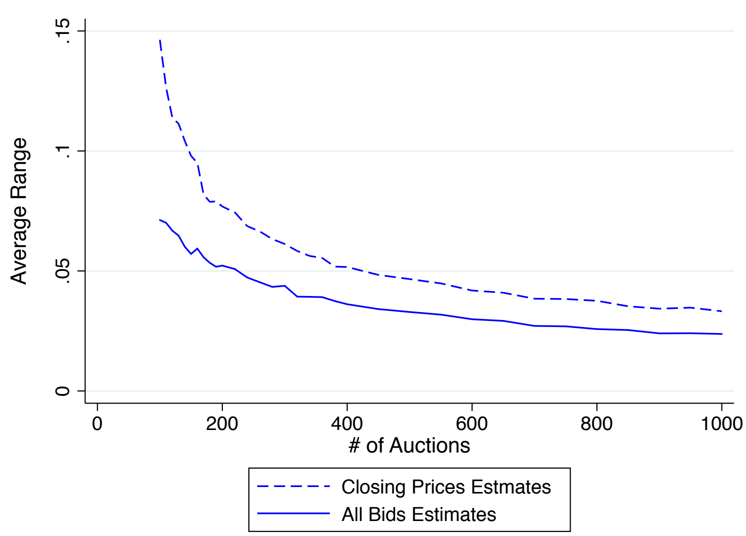
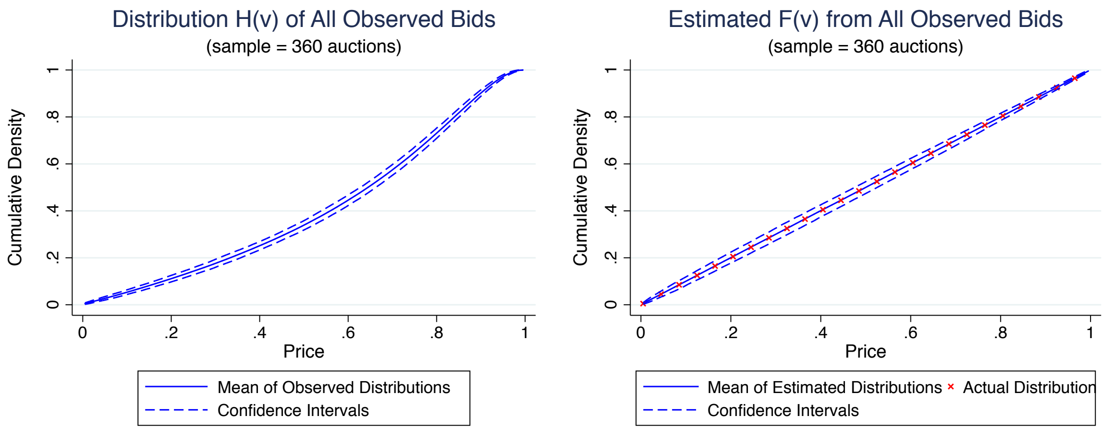


Figure 6: Estimated Distribution of Valuations using All Bids



declines at a slower rate thereafter.

3.3 Estimating Valuations: All Bids

In applications where the estimated distribution is needed even in the lower range of valuations, one will fare much better using the full distribution of observed bids. Again, I compute the empirical cumulative distribution function of all observed bid $H_s^i(v)$ for each sample size s and each iteration i. Equation 19 is then used to compute the $\hat{F}_s^i(v)$ that is consistent with the observed $H_s^i(v)$, using $\hat{\lambda}_s^i$ for the participation parameter.

The resulting fit is illustrated in Figure 6 for a sample of 360 auctions. The left panel indicates the observed distribution of bids $H_s^i(v)$, with the mean distribution denoted in solid and confidence intervals with dashed lines. Note that, if the econometrician had assumed that all participants had been observed, then $H_s^i(v)$ would be the estimate for $F(v)$ — differing dramatically from the uniform distribution. The inversion process recovers the true distribution with tight confidence intervals, however, as depicted in the right panel.

To examine how the fit improves with larger samples, the solid line in Figure 5 indicates the average range in the confidence intervals from the distributions estimated from all bids. Beyond 200 auctions, the estimates from all bids have a range that is consistently 40% smaller than that produced by the closing price estimates. In part, this is explained by the fact that there are typically 5 times more bids than there are closing prices.

4 Conclusion

Auction participation includes more than those who placed a bid. These non-bidders are still part of the market — they are willing to purchase under the right circumstances — and correctly accounting for them could benefit a seller, market designer, or econometrician in identifying potential customers or product demand. Indeed, the typical eBay auction registers only 5.58 bidders; at the very least, one would expect all the losers to consider the next auction of the same item (just over 16 hours away, on average), along with some newcomers. The participation estimate developed here would estimate 15.1 participants per auction, nearly three times as large.

Moreover, an improved participation estimate also enables better estimates of participant valuations, which will be first order stochastically dominated by bidder valuations (since participants are excluded from the latter precisely when their valuations are too low to place a bid). In effect, the winning bidder was really competing against three times the observed number of bidders. If the econometrician ignores this unseen competition, he will mistakenly attribute high closing prices to a skewed distribution of bidder valuations.

The participation estimates developed here can potentially be generalized in several directions. First, the Poisson distribution of participants is analytically convenient, but not strictly necessary. In the appendix, I characterize the probability of a bidders given n participants without any structure on the random process generating n, as well as the average number of bidders for a given n. Thus, one could apply any other distributional assumption on the number of participants and then estimate parameters of that distribution.

Second, the model may be adapted to allow some degree of correlated values. For instance, an incremental step in that direction could be in allowing for auction-specific shocks which raise the valuations of all bidders identically, known to the bidders but not the econometrician. Quint (2015) identifies the distribution of valuations in this environment from the auction closing prices, including in a setting in which the number of participants is not observed but follows a Poisson distribution with a known mean. The methods presented here could potentially estimate the average number of participants in that setting.

A Relaxing the Poisson Assumption

The model has been presented assuming throughout that the number of participants is Poisson distributed. However, the analysis applies more broadly. Note that the analysis up to Eq. 3 does not depend on the the Poisson distribution. Thus, instead of averaging over the possible realizations of n opponents at that point, one can average over valuations v as follows:

$$
\begin{aligned}
Pr(bid|n) &\equiv \int_{\underline{v}}^{\bar{v}} Pr(bid|v,n)F'(v)dv && (20)\\
&= \int_{\underline{v}}^{\bar{v}} \left(\frac{2\left(1-F(v)^{n+1}\right)}{(n+1)(1-F(v))} - F(v)^n \right) F'(v)dv && (21)\\
&= \frac{1-n+2(1+n)H_n}{(1+n)^2}, && (22)
\end{aligned}
$$

where $H_n = \sum_{i=1} n\frac{1}{i}$ are the harmonic numbers.

Thus, given n participants, the expected number of bidders will be:

$$\bar{a}(n) \equiv n \cdot Pr(bid|n-1) = 2H_{n-1} - \frac{n-2}{n}. \tag{23}$$

Here, each of the n bidders face the same ex-ante probability of placing a bid, $Pr(bid|n-1)$. We define $\bar{a}(0) = 0$.

If the econometrician believed that there was a fixed number of participants, he could infer that number n from the average number of bidders by numerically inverting Eq. 23. Alternatively, if he believed the number of participants was stochastically drawn from a fixed, discrete distribution $d(n)$, he could compute the average of $\bar{a}(n)$ in that distribution, and select the distribution parameters to match the observed average. For instance, $\sum_{n=0}^{\infty} \frac{e^{-\lambda}\lambda^n}{n!}\bar{a}(n)$ evaluates to $\bar{a}(\lambda)$ from Eq. 10 (as anticipated, since the order of the expectation operators can be interchanged).

Of course, the distribution $d(n)$ of participants must have a finite mean. Moreover, the number of parameters cannot exceed the cardinality of the set of realized bidders. This, in particular, prevents a non-parametric estimation of $d(n)$ for each n. It is always the case that $a \leq n$, and very rarely will one observe a near n if the latter is moderately large. Even with 10 participants, there is only a $1/10,000$ chance of observing all 10 bidders. Thus, the set of realized number of bidders will often be much smaller than the set of realized numbers of participants.

Despite the need to rely on distributional assumptions on $d(n)$, one can test the goodness of fit post-estimation. To do this, one needs the full probability distribution of realized bidders given a number of participants, $Pr(a|n)$. This can be recursively defined as follows.

$$Pr(1|n) = 1 \text{ iff } n = 1 \tag{24}$$

$$Pr(a|1) = 1 \text{ iff } a = 1 \tag{25}$$

$$Pr(a|n) = \begin{cases} \frac{n-2}{n} Pr(a|n-1) + \frac{2}{n} Pr(a-1|n-1) & \text{if } 1 < a \leq n \\ 0 & \text{otherwise} \end{cases} \tag{26}$$

After computing these probabilities, one can compute the unconditional probability of observing a bidders via $Pr(a) = \sum_n Pr(a|n)d(n)$, using the estimated $d(n)$. This predicted $Pr(a)$ can then be compared to the observed distribution of realized bidders using a χ^2 goodness-of-fit test.

References

Ambrus, A., J. Burns, and Y. Ishii. (2014). "Gradual Bidding in eBay-Like Auctions," Mimeo, Harvard University.

Athey, S., and P. Haile. (2007). "Nonparametric Approaches to Auctions," *Handbook of Econometrics* **6**, 38473965.

Bauner, C. (2015). "Mechanism Choice and the Buy-It-Now Auction: A Structural Model of Competing Buyers and Sellers," *International Journal of Industrial Organization* **38**, 19-31.

Bodoh-Creed, A. (2013). "Optimal Platform of Two-Sided Market Platforms: An Empirical Case Study of eBay," Working Paper #13-22, NET Institute.

Canals-Cerdá, J. and J. Pearcy (2013). "Arriving in Time: Estimation of English Auctions With a Stochastic Number of Bidders," *Journal of Business and Economic Statistics*, **31**, 125-135.

Coey, D., B. Larsen, and B. Platt. (2015) "A Theory of Bidding Dynamics and Deadlines in Online Retail," Mimeo, Stanford University.

Etzion, H., E. Pinker, and A. Seidmann. (2006) "Analyzing the Simultaneous Use of Auctions and Posted Prices for Online Selling," *Manufacturing & Service Operations Management* **8**, 68-91.

de Haan, L., C. G. de Vries, and C. Zhou. (2013) "The Number of Active Bidders in Internet Auctions," *Journal of Economic Theory* **148**, 1726-1736.

Hammond, R. G. (2013). "A Structural Model Of Competing Sellers: Auctions And Posted Prices," *European Economic Review* **60**, 52-68.

Hortacsu, A., and E. Nielsen. (2010) "Do Bids Equal Values on eBay?" *Marketing Science*, **29**, 994-997.

Ingster, A. (2009). "A Structural Study Of A Dynamic Auction Market," Mimeo, New York University.

Kultti, K. (1999). "Equivalence Of Auctions And Posted Prices," *Games and Economic behavior*, **27** 106-113.

Menzel, K., and P. Morganti (2013) "Large Sample Properties for Estimators Based on the Order Statistics Approach in Auctions," *Quantitative Economics*, **4** 329-375.

Myerson, R. (1998) "Population Uncertainty And Poisson Games," *International Journal of Game Theory*, **27** 375-392.

Nekipelov, D. (2007). "Entry Deterrence And Learning Prevention On Ebay," Mimeo, Duke University.

Quint, D. (2015). "Identification in Symmetric English Auctions with Additively Separable Unobserved Heterogeneity," Mimeo, University of Wisconsin.

Song, Unjy. (2004) "Nonparametric Estimation of an eBay Auction Model with an Unknown Number of Bidders," Mimeo, University of British Columbia.

Wang, Ruqu. (1993) "Auctions Versus Posted-price Selling," *The American Economic Review* **83** 838-851.

Zeithammer, R., and C. Adams. (2010) "The Sealed-Bid Abstraction in Online Auctions," *Marketing Science*, **29**, 964-987.