23 Models for Bid Arrivals and Bidder Arrivals in Online Auctions

Ralph P. Russo, Galit Shmueli, Wolfgang Jank, and Nariankadu D. Shyamalkumar

23.1 Introduction

The arrival process of bidders and bids in online auctions is important for studying and modeling supply and demand in the online marketplace. Whereas bid arrivals are observable in online auction data, bidder behavior is typically not. A popular assumption in the online auction literature is that a homogeneous Poisson bidder arrival process is a reasonable approximation. This approximation underlies statistical models and simulations used in field studies. However, empirical research has shown that the process of bid arrivals is far from homogeneous, as it features early and last-moment bidding, as well as a self-similar structure. In this chapter we discuss two types of models: Descriptive models for bid arrivals that were derived based on features of real bid data in online auctions, and models for bidder arrivals that lead to the bid arrival process. The model for the bid arrival process, called *the BARISTA process*, can generate different intensities at different stages of the auctions. We discuss its properties, show how to simulate bid arrivals from it, and how to estimate its parameters. We then describe two bidder behavior models that lead to versions of the BARISTA process. Model adequacy and performance are illustrated via simulation and real data from eBay.

23.2 Motivation

Online auctions and their empirical research have been flourishing in recent years due to the important role that these auctions play in the marketplace, and the availability of large amounts of high-quality bid data from websites such as eBay, Yahoo!, Amazon, and uBid. Many theoretical results derived for traditional (offline) auctions have been shown to fail in the online setting for reasons such as globalism, computerized bidding, longer auction durations, more flexibility in design choice by the seller, anonymity of buyers and sellers, and issues of trust. A central factor underlying many important results is the number of bidders participating in the auction. This number is usually assumed to be fixed and known [18] or fixed but unknown [16]. In online auctions, however, the number of bidders and bids is not predetermined, and is affected by the auction design and its dynamics. Thus, in both the theoretical and empirical domains the number of bidders and bids plays an important role.

Economists are typically interested in bidder behavior in terms of bidder arrival and departure and bidding strategies. However, creating models for bidder behavior from available bid data is challenging because such behavior is not recorded. Information on bid arrivals, on the other hand, is recorded and often publicly available. eBay, for example, posts the temporal sequence of all bids placed over the course of an auction, yet the time when bidders first arrive at an auction is unobservable from the bid history. Bidders can browse an auction without placing a bid, thereby not leaving a trace or revealing their interest in that auction. In fact, it is likely that bidders first browse an auction and only later place their bid. The lag between a bidder's arrival time and his initial bid placement is not recorded by the auction website, and thus one cannot infer the former from the latter. Another issue is that most online auctions allow bid revision, and therefore many bidders place multiple bids. This further adds to the obscurity of defining the bidder arrival-departure process. Our approach is therefore to first create a descriptive statistical model for the bid arrival process, which is based on empirical evidence, and then to create bidder activity models that lead to the aggregate bid arrival processes. We therefore start by introducing the model of Shmueli et al. [22] for the bid arrival process, and later describe several models for bidder behavior proposed by Russo et al. [21].

Although models for bidder behavior are of more direct interest to economists, descriptive models for bid arrivals have many potential uses, both for online auction research as well as for practical applications. First, it is useful for the purpose of simulation. Researchers in the online auction arena often use simulated bid arrival data to validate their results. For example, Bapna et al. [3] used simulated bid arrival data to validate their model on a bidder's willingness to pay, while Gwebu et al. [9] designed a complex simulation study to analyze bidders' strategies using assumptions about bidder as well as bid arrival rates.

Second, modeling the bid arrival process rather than the bidder arrivals promises to produce more reliable results, as bid placements are often completely observable from the auction's bid history, whereas bidder arrivals are not.

Finally, a statistical model for bid arrivals can be useful for designing and improving online auction applications such as automated bidding agents or enhancing automated electronic negotiation by monitoring auction server performance. For instance, Menasce and Akula [17] study the connection between bid arrivals and auction server performance. They find that the commonly encountered "last minute bidding" creates a huge surge in the workload of auction servers and degrades their performance. They then suggest a model to improve a server's performance through auction rescheduling using simulated bid arrival data.

In the following, we start by addressing the bid arrival process. Before introducing the model, we describe two special features of bid data that have been observed in practice (Section 23.3). We then discuss in Section 23.4 the BARISTA model of Shmueli et al. [23], a flexible statistical model for bid arrivals. We discuss model properties, estimation, and simulation. In Section 23.5 we move to bidder arrivals, and describe several bidder behavior models that are consistent with the BARISTA and related bid arrival processes.

23.3 Features of Bid Arrivals

In this section we consider prominent features of online auction bid arrival data. Constructing statistical models for bid arrivals requires capturing the main features of the data. The current literature reports two dominant features of online auction bid arrivals: (1) a nonhomogenous intensity that possesses two or three distinct stages, and (2) a self similarity effect in the distribution of bid arrival times. We examine each of these properties next.

23.3.1 Multistage Arrival Intensities

Several researchers have noted deadline effects in internet auctions [2,5,14,20,25]. In many of these studies it was observed that a significant percent of bids arrive at the very last minute of the auction. This phenomenon, called "bid sniping" has received much attention, and numerous explanations have been suggested to account for it. Empirical studies of online auctions have also reported an unusual amount of bidding activity at the auction start followed by a longer period of little or no activity [5,12]. Bapna et al. [4] refer to bidders who place a single early bid as "evaluators." Finally, "bid shilling," a fraudulent practice whereby the seller places dummy bids to drive up the price, is associated with early high bidding [13]. The existence of these bid-timing phenomena are important factors in determining outcomes at the auction level as well as at the market level. They have therefore received much attention from the research community.

23.3.2 Self-Similarity (and Its Breakdown)

While both offline and online environments share deadline and earliness effects, the bid arrival process in the online environment appears to possess the additional property of *self-similarity*.[1] Self similarity refers to the "striking regularity" of shape that can be found among the distribution of bid arrivals over the intervals $[t, T]$, as t approaches the auction deadline T. Self similarity is central in applications such as web, network and ethernet traffic. Huberman and Adamic [11] found that the number of visitors to websites follows a universal power law. Liebovitch and Schwartz [15] reported that the arrival process of email viruses is self-similar. However, this has also been reported in other online environments. For instance, Aurell and Hemmingsson [1] showed that the times between bids in the interbank foreign exchange market follow a power law distribution.

Several authors reported results that indicate the presence of self-similarity in the bidding frequency in online auctions. Roth and Ockefels [20] found that the arrival of last bids by bidders during an online auction is closely related to a self-similar process. They approximated the CDF of bid arrivals in "reverse time" (i.e., the CDF of the elapsed times between the bid arrivals and the auction deadline) by the power functional form $F_T(t) = (t/T)^\alpha$ ($\alpha > 0$), over the interval $[0, T]$, and estimated α from the data using OLS. This approximates the distribution of bids over intervals that range from the last 12 hours to the last 10 minutes, but accounts for neither the final minutes of the auction nor the auction start and middle. Yang et al. [26] found that the number of bids and the number of bidders in auctions on eBay and on its Korean partner (auction.co.kr) follows a power law distribution. This was found for auctions across multiple categories.

The self-similar property suggests that

[1] An exception to non-self-similarity in the offline environment can be found in the process of bargaining agreements, as described in Roth and Ockefels [20].

the rate of incoming bids increases steadily as the auction approaches its end. Empirical investigations have indeed found that many bidders wait until the last possible moment to submit their final bid. By doing so, they hope to increase their chance of winning the auction, as the probability that competitors will successfully bid higher before closing diminishes. This common bidding strategy of "bid sniping," (or, "last minute bidding") would suggest a steadily increasing flow of bid arrivals towards the auction end.

However, empirical evidence from online auction data indicates that bid times over the last minute or so of hard-close auctions tend to follow a uniform distribution [19]. This has not been found in soft-close, or "going-going-gone" auctions, such as those on Amazon, Yahoo!, or uBid.com, where the auction extends after the last bid is placed. Thus, in addition to the evidence for self similarity in online auctions, there is also evidence of its breakdown during the very last moments of a hard-close auction. Roth and Ockefels [20] note that the empirical CDF plots for intervals that range between the last 12 hours of the auction and the last 1 minute all look very similar except for the last 1-minute plot. Being able to model this breakdown is essential, since the last moments of the auction (when sniping takes place) are known to be crucial to the final price. The BARISTA model accommodates the changes in bidding intensity from the very start to the very end of a hard-close auction.

We illustrate the self-similarity in the bid arrival process in online auctions by examining data on 189 7-day auctions (with a total of 3651 bid times) on eBay.com for new Palm M515 Personal Digital Assistants.[2] Figure 1 displays the empirical CDFs for the 3651 bid arrivals at several resolutions, "zooming-in" from the entire auction duration (of 7 days) to the last day, last 12 hours, 6 hours, 3 hours, 5 minutes, 2 minutes, and the very last minute. We observe that (1) the 7-day curve is different from the other curves in that it starts out concave, (2) the last day through last 3-hour curves are all very similar to each other, and (3) the last minutes curves gradually approach the 1-minute curve which is nearly uniform. These visual similarities are confirmed by the results of two-sample Kolmogorov–Smirnoff tests comparing all the pairs of distributions, resulting in similarities only among the curves within each of the three groups.

These observations replicate the results in [20] where self-similarity was observed in the bid time distributions of the last 12-hour, 6-hour, 3-hour, 1-hour, 30-minute, and 10-minute periods of the auction, and where this self-similarity breaks down in the last minute of the auction to become a uniform distribution. However, we examine a few additional time resolutions which give further insight: First, by looking at the last 5-minutes and last 2-minutes bid distributions we see that the self-similarity gradually transitions into the 1-minute uniform distribution. Second, our inspection of the entire auction duration (which was unavailable in the study by Roth and Ockefles [20]) reveals an additional early-bidding stage. Self-similarity, it appears, is not prevalent throughout the entire auction duration. Such a phenomenon can occur if a bid placed during the final moments of the auction has a positive probability of not getting registered. There are various factors that may cause a bid not to register: the time it takes to manually place a bid (Roth and Ockefles [20] found that most last minute bidders tend to place their bids manually rather than through available sniping software agents); hardware difficulties, internet congestion, unexpected latency, and

[2] The data are available at *http://www.rhsmith.umd.edu/faculty/gshmueli/web/html/101.html*

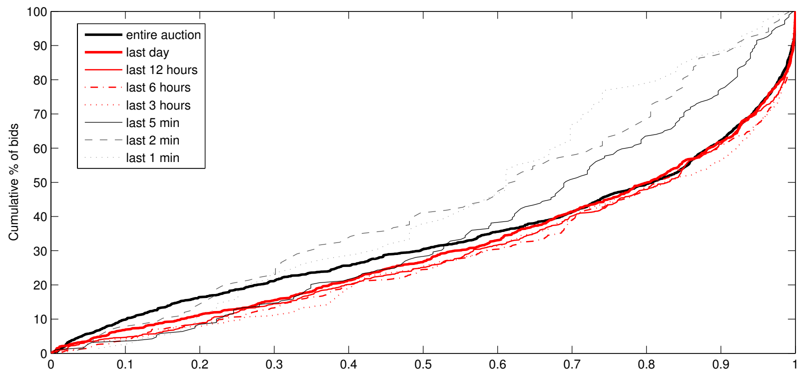


Figure 1: Empirical CDFs of number of bids in 189 Palm M515 auctions overlaid.

server problems on eBay (see, for example, *www.auctionsniper.com*). Clearly, the closer to the end the auction gets, the higher the likelihood that a bid will not get registered successfully.

This increasing likelihood of an unsuccessful bid counteracts the increasing flow of last minute bids. The result can be approximated by a uniform bid arrival process that "contaminates" the self-similarity of the arrivals until that point.

In the next section we describe a flexible nonhomogeneous Poisson process that captures the empirical phenomenon described above. In addition to the self-similarity, it also accounts for the two observed phenomena of "early bidding" and "last minute bidding."

23.4 The BARISTA: A Three-Stage Nonhomogeneous Poisson Process

In this section we describe our model for bid arrivals. The model captures the two main features of bid arrivals in online auctions described earlier: the three different stages and the self-similarity (with its breakdown). We call this process the BARISTA (**B**id **AR**rivals **I**n **STA**ges), because it generates different intensities of activity[3]:

Stage 1 - The auction start, characterized by an early burst of activity,

Stage 2 - Mid-auction bid arrivals, characterized by increasing bid intensity and self similarity that is gradually contaminated as the third stage approaches , and

Stage 3 - The last moments of the auction, characterized by very intense activity dampened possibly by bids that are not successfully registered.

23.4.1 Model Formulation

Let $N(s)$, $0 \leq s \leq T$, be a nonhomogeneous Poisson process with intensity func-

[3]We also call the stages the "espresso stage" (short and intense), "macchiato stage" (stained), and "ristretto stage" (extra intense), and hence the BARISTA.

tion

$$\lambda(s) = \begin{cases} c\left(1-\frac{d_1}{T}\right)^{\alpha_2-\alpha_1}\left(1-\frac{s}{T}\right)^{\alpha_1-1} \\ \quad \text{for } 0 \le s \le d_1 \\ \\ c\left(1-\frac{s}{T}\right)^{\alpha_2-1} \\ \quad \text{for } d_1 \le s \le T-d_2 \\ \\ c\left(\frac{d_2}{T}\right)^{\alpha_2-\alpha_3}\left(1-\frac{s}{T}\right)^{\alpha_3-1} \\ \quad \text{for } T-d_2 \le s \le T, \end{cases} \tag{1}$$

where $c > 0$, $\alpha_j > 0$ for $j = 1, 2, 3$, T is a known constant, and $0 \le d_1 < T - d_2 \le T$. Note that λ is continuous, so there are no jumps at times d_1 and $T - d_2$. The random variable $N(s)$, which counts the number of arrivals until time s, follows a Poisson distribution with mean

$$m(s) = \int_0^s \lambda(x)\,dx. \tag{2}$$

Given that $N(T) = n$, the collection of arrival times are equivalent to the order statistics of a random sample of size n from the distribution having distribution function $F(s) = m(s)/m(T)$ and density function $f(s) = \lambda(s)/m(T)$.

We expect α_3 to be close to 1 (uniform arrival of bids at the end of the auction) and $\alpha_1 > 1$ to represent the early surge in bidding.

23.4.2 Properties of the BARISTA Process

The process described by (1) has two properties that lead to a wide family of processes, and that can be useful in practice. We describe each property and its implications below.

An additive property. If N_k, $1 \le k \le m$, are independent BARISTA processes having c parameters $c_1, ..., c_m$ and common parameters $(\alpha_1, \alpha_2, \alpha_3)$, (d_1, d_2) and T, then the aggregated process $N = \sum_{1 \le k \le m} N_k$ is a BARISTA with parameters $(\alpha_1, \alpha_2, \alpha_3)$, (d_1, d_2), T, and $c = \sum_{1 \le k \le m} c_k$.

Thus, the bid arrival times from several auctions may be aggregated and treated as though they were generated by a single auction, provided that each original auction can be regarded as producing a BARISTA process with (nearly) the same parameters. The advantage of aggregation is more accurate parameter estimation.

A regenerative property. An observer who counts only the bid arrivals occurring after time βT, for some $0 \le \beta < 1$, sees the process

$$N_\beta(s) := N(s) - N(\beta T),\ \beta T \le s \le T.$$

N_β is an $NHPP$ with intensity function $\lambda_\beta = \lambda$, restricted to the interval $[\beta T, T]$. It is shown in Shmueli et al. [23] that N_β is a BARISTA process with βT the new "zero," a new deadline $(1-\beta)T$, a new time unit, and new parameters. This regenerative property allows us to use the model to approximate bid arrivals in an ongoing auction, not only in a completed auction. One application where this is useful is in real-time forecasting of future bid times, e.g., for the purpose of optimizing server performance.

23.4.3 Special Cases

In empirical studies, d_1 appears to be small (1–2 days) and d_2 very small (a few minutes) compared to T (several days). Thus, most of the BARISTA process is realized in the second stage, during which the process can be regarded as having *contaminated self-similarity.* The contamination is caused by the bid arrivals in the third stage, and increases as $s \to T - d_2$.

When $d_1 = d_2 = 0$, the BARISTA process reduces to a single-stage process (NHPP_1) with an intensity function $\lambda(s) = c(1 - \frac{s}{T})^{\alpha-1}$ and associated distribution function $F(s) = 1-(1-\frac{s}{T})^{\alpha}, \quad 0 \le s \le T$. For $(\theta, t) \in [0,1] \times (0,T]$ we have

$$\frac{1-F(T-t\theta)}{1-F(T-t)} = \theta^{\alpha} \; (independent\ of\ t),$$

and thus we have a *pure self-similar* process. The joint MLE of (α, c) is obtainable in this case (see Appendix A in Ref. 23):

$$\begin{aligned} \widehat{\alpha} &= -N(T)\left[\sum_{i=1}^{N(T)} \ln\left(1-\frac{X_i}{T}\right)\right]^{-1}, \\ \widehat{c} &= \frac{N(T)\widehat{\alpha}}{T}. \end{aligned}$$

Because $X \sim F \implies -\ln\left(1-\frac{X}{T}\right) \sim \exp(rate = \alpha)$, and $\lim_{c\to\infty} \Pr(N(T) \to \infty) = 1$, a conditioning argument on $N(T)$ yields an asymptotic result:

$$\sqrt{N(T)}\left(\frac{\alpha}{\widehat{\alpha}} - 1\right) \xrightarrow{D} N(0,1) \quad as \;\; c \to \infty,$$

where $N(0,1)$ indicates a standard normal distribution.

When $d_1 = 0$, the BARISTA process reduces to a two-stage process (NHPP_2), with a single *changepoint* at $T - d_2$. This process is useful for modeling bid arrivals in auctions that lack the initial surge of early bidding. For further technical details on these special cases see Shmueli et al. [23].

23.4.4 Simulating and Fitting the BARISTA to Data

We now describe simulating a BARISTA process and fitting observed bidding data to a BARISTA model. Simulating bid arrivals is useful in field experiments, in evaluation of model fit, and for quantifying sampling error. Our proposed simulation method is simple to program and computationally efficient. Fitting the BARISTA process to data requires estimating the two change-points and three α parameters. We discuss three estimation methods that range in their computational intensiveness and accuracy.[4]

Process simulation. To simulate a BARISTA process, simulate first a variable $N(T)$ having a Poisson($m(T)$) distribution, and then simulate $N(T)$ independent $U(0,1)$ variables $U_1, \ldots, U_{N(T)}$. The order statistics of the collection $F^{-1}(U_1), \ldots, F^{-1}(U_{N(T)})$ represent the ordered arrival times of the BARISTA process.

The algorithm for generating n arrivals $(x_1, \ldots, x_n)$ is then:
(1) Generate n uniform variates $u_1, \ldots, u_n$.
(2) For $k = 1, \ldots, n$ set

$$x_k = \begin{cases} T - T\Big\{1 - \frac{u_k \alpha_1}{CT} \\ \qquad \times \left(1-\frac{d_1}{T}\right)^{\alpha_1-\alpha_2}\Big\}^{1/\alpha_1} \\ \qquad \text{if } u_k < F(d_1) \\ T - T\Big\{\frac{\alpha_2}{CT}(F_3(d_1) - u_k) \\ \qquad + \left(1-\frac{d_1}{T}\right)^{\alpha_2}\Big\}^{1/\alpha_2} \\ \qquad \text{if } F(d_1) \le u_k < F(T-d_2) \\ T - T\left\{\frac{\alpha_3}{CT} u_k \left(\frac{d_2}{T}\right)^{\alpha_3-\alpha_2}\right\}^{1/\alpha_3} \\ \qquad \text{if } u_k \ge F(T-d_2). \end{cases} \tag{3}$$

Parameter estimation. We describe three different estimation methods, each having a different tradeoff between computational intensity and accuracy, and with varying amounts of required user input.

Quick & crude (CDF-based) estimation. The estimation of the α parameters depends on the changepoints d_1, $T-d_2$ and vice-versa. As a crude start, we choose

[4] Matlab code for the simulation and estimation procedures is available at *http://lib.stat.cmu.edu/aoas/117/*.

three intervals of the form $[T-t, T-s]$ that we are confident lie in the first, second, or third stages, and use those for estimating the α parameters. We then use the α estimates to obtain estimates for the changepoints.

In both cases the estimates are based on writing the parameters as a function of the CDF, and then substituting the empirical CDF to obtain estimates. To estimate the α parameters, we use the relationship (see Shmueli et al. [23] for details)

$$\begin{aligned}\alpha_j &= 2[\ln\{F(T-t) - F(T-\sqrt{st})\} \\ &\quad - \ln\{F(T-\sqrt{st}) - F(T-s)\}] \\ &\quad \times[\ln t - \ln s]^{-1}. \qquad (4)\end{aligned}$$

and estimate α_j by substituting F with the empirical CDF $F_e(t) = N(t)/N(T)$ in the approximation.

For α_3 we can use the exact relation

$$\alpha_3 = \frac{\ln[R(t_3)/R(t_3')]}{\ln[(T-t_3)/(T-t_3')]}, \qquad (5)$$

where $R(t) = 1 - F(t)$ and t_3, t_3' are within $[T-d_2, T]$. To estimate α_3 we choose reasonable values of t_3, t_3' and use the empirical survival function $R_e = 1 - F_e$.

Obtaining standard errors for these estimators can be done by bootstrapping (see Ref. 7 for details), due to the low computational effort involved in this estimation method.

To assess this method we simulated 5000 random observations from the BARISTA process on the interval $[0, 7]$ with parameters $\alpha_1 = 3, \alpha_2 = 0.4, \alpha_3 = 1$ and the changepoints $d_1 = 2.5$ (defining the first 2.5 days as the first stage) and $d_2 = 5/10080$ (defining the last 5 minutes as the third stage). The intensity function for these data is shown in Figure 2, and parameter estimates with their standard errors are given in Table 1.

Shmueli et al. [23] study the robustness of the α estimators to the choice of t and s and find that the estimates are relatively insensitive to the exact interval choices as long as they are reasonable.

To estimate the changepoints, we again use the relationship with the CDF:

$$d_1 = T - T\left\{\frac{\alpha_1}{\alpha_2}\frac{F(t_1)}{F(t_2)-F(t_2')} \times\frac{(1-t_2'/T)^{\alpha_2} - (1-t_2/T)^{\alpha_2}}{1-(1-t_1/T)^{\alpha_1}}\right\}^{\frac{1}{\alpha_2-\alpha_1}}, \qquad (6)$$

$$d_2 = T\left\{\frac{\alpha_3}{\alpha_2}\frac{1-F(t)}{F(t_2)-F(t_2')} \times\frac{(1-t_2'/T)^{\alpha_2} - (1-t_2/T)^{\alpha_2}}{(1-t_3/T)^{\alpha_3}}\right\}^{\frac{1}{\alpha_2-\alpha_3}} \qquad (7)$$

and then estimate d_1 and d_2 by selecting "safe" values for t_1, t_2', t_2, and t_3 (which are confidently within the relevant interval) and using the empirical CDF at those points.

Using this method we estimated d_1 and d_2 for the simulated data. We used the true values of the α parameters and the "safe" values $t_1 = 1, t_2' = 3, t_2 = 6$, and $t_3 = 7 - 2/10080$. The estimates and their (bootstrap) standard errors are reported in Table 1. Shmueli et al. [23] show that the d estimates are also relatively robust to the choice of t and s, as long as those are reasonable.

Finally, estimating c depends on the α and d estimates and on the observed number $N(T)$ of bids placed on $[0, T]$. Define $g(\theta; s) = \lambda(s)/c$, $0 \le s \le T$, where λ is the function in (1). We have

$$\begin{aligned}N(T) &\approx E[N(T)] = c\int_0^T g(\theta; s)\,ds \\ &\approx c\int_0^T g(\widehat{\theta}; s)\,ds.\end{aligned}$$

Solving for c, we obtain the estimate

$$\widehat{c} = \frac{N(T)}{\int_0^T g(\widehat{\theta}; s)\,ds}.$$

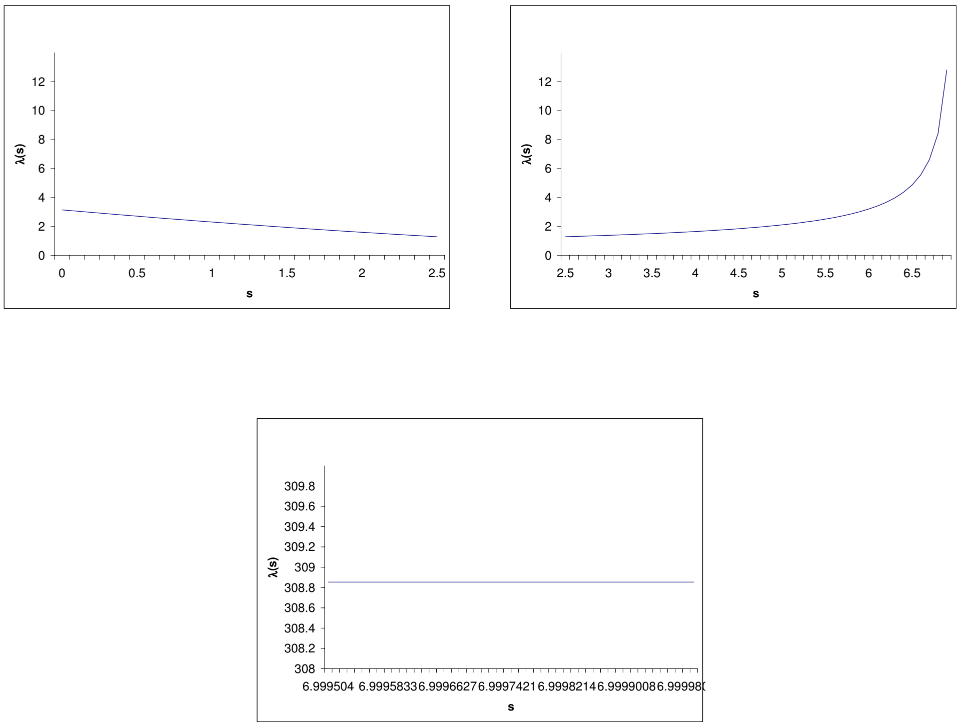


Figure 2: Intensity function $\lambda(s)$ for simulated BARISTA data, with $\alpha_1 = 3, \alpha_2 = 0.4, \alpha_3 = 1, d_1 = 2.5, d_2 = 7 - 5/10080$, and $c = 1$. Note the different time scale for the last 5 minutes (bottom panel).

If $\widehat{\theta}$ is an MLE of θ, then $\widehat{c}$ is an MLE of c.

Maximum Likelihood Estimation. Conditional on $N(T) = n$, the BARISTA likelihood function is given by

$$\begin{aligned}\mathcal{L}(x_1, \ldots, x_n | \alpha_1, \alpha_2, \alpha_3, d_1, d_2) \\ = n \ln C + n_1(\alpha_2 - \alpha_1) \ln\left(1 - \frac{d_1}{T}\right) \\ + n_3(\alpha_2 - \alpha_3) \ln \frac{d_2}{T} + (\alpha_1 - 1) S_1 \\ + (\alpha_2 - 1) S_2 + (\alpha_3 - 1) S_3, \end{aligned} \tag{8}$$

where n_1 is the number of arrivals before time d_1, n_3 is the number of arrivals after $T - d_2$, $S_1 = \sum_{i:x_i \leq d_1} \ln\left(1 - \frac{x_i}{T}\right)$, $S_2 = \sum_{i:d_1 < x_i < T-d_2} \ln\left(1 - \frac{x_i}{T}\right)$, and $S_3 = \sum_{i:x_i > T-d_2} \ln\left(1 - \frac{x_i}{T}\right)$. We note that the form of the MLE (for the α and d parameters) is the same whether $N(T) = n$ is fixed, or $N(T)$ is random (see Ref. 23, Appendix A).

In order to estimate $\alpha_1, \alpha_2, \alpha_3$ for given values of d_1, d_2, the following three equations must be solved (equating the first derivatives in $\alpha_1, \alpha_2, \alpha_3$ to zero).

$$S_1 = n_1 \ln\left(1 - \frac{d_1}{T}\right) - \frac{n}{C}\frac{\partial C}{\partial \alpha_1}, \tag{9}$$

$$S_2 = -n_1 \ln\left(1 - \frac{d_1}{T}\right) - n_3 \ln \frac{d_2}{T} - \frac{n}{C}\frac{\partial C}{\partial \alpha_2}, \tag{10}$$

$$S_3 = n_3 \ln \frac{d_2}{T} - \frac{n}{C}\frac{\partial C}{\partial \alpha_3}, \tag{11}$$

where

$$\begin{aligned}\frac{\partial C}{\partial \alpha_1} \\ = \frac{C^2 T}{\alpha_1^2}\left(1 - \frac{d_1}{T}\right)^{\alpha_2} \\ \times \left[\left(1 - \frac{d_1}{T}\right)^{-\alpha_1} \right. \\ \left. \times \left(1 + \alpha_1 \ln(1 - \frac{d_1}{T})\right) - 1\right], \end{aligned} \tag{12}$$

$$\begin{aligned}\frac{\partial C}{\partial \alpha_2} \\ = \frac{C^2 T}{\alpha_1 \alpha_3 \alpha_2^2}\left\{\alpha_3 \left(1 - \frac{d_1}{T}\right)^{\alpha_2} \right. \\ \times \left[\alpha_2 \ln\left(1 - \frac{d_1}{T}\right) \right. \\ \times \left(\alpha_2 - \alpha_1 + \alpha_2\left(1 - \frac{d_1}{T}\right)^{-\alpha_1}\right) \\ \left. -\alpha_1 \right] + \alpha_1 \left(\frac{d_2}{T}\right)^{\alpha_2} \\ \left. \times \left[\alpha_3 + \alpha_2 \ln \frac{d_2}{T}(\alpha_2 - \alpha_3)\right]\right\} \\ = \frac{C^2 T}{\alpha_2^2}\left[\left(\frac{d_2}{T}\right)^{\alpha_2} - \left(1 - \frac{d_1}{T}\right)^{\alpha_2} \right. \\ -\frac{\alpha_2^2}{\alpha_1}\left(1 - \frac{d_1}{T}\right)^{\alpha_2} \ln\left(1 - \frac{d_1}{T}\right) \\ \times \left(1 - (1 - \frac{d_1}{T})^{-\alpha_1}\right) \\ -\alpha_2 \left(\frac{d_2}{T}\right)^{\alpha_2} \ln \frac{d_2}{T} + \alpha_2\left(1 - \frac{d_1}{T}\right)^{\alpha_2} \\ \left. \times \ln\left(1 - \frac{d_1}{T}\right) + \frac{\alpha_2^2}{\alpha_3}\left(\frac{d_2}{T}\right)^{\alpha_2} \ln \frac{d_2}{T}\right], \end{aligned} \tag{13}$$

$$\frac{\partial C}{\partial \alpha_3} = \frac{C^2 T}{\alpha_3^2}\left(\frac{d_2}{T}\right)^{\alpha_2}. \tag{14}$$

Since the equations are nonlinear in the parameters, an iterative gradient method can be used such as Newton–Raphson or the Broyden–Fletcher–Goldfarb–Powell (BFGP) method, which is a more stable quasi-Newton method that does not require the computation and inversion of the Hessian matrix (see, for example, Ref. 6). If the changepoints d_1 and d_2 are unknown and we want to estimate them from the data, then search algorithms such as genetic algorithms can be more efficient, more stable, and more easily programmable for finding a solution. Otherwise the likelihood needs to be computed for a grid of $d_1 \times d_2$ values. Our experience suggests that gradient methods tend to be unstable for solving this maximization problem. In short, an exhaustive search over a reasonable grid of the parameter space or a stochastic search algorithm (using the quick & crude estimates as starting points) are good practical solutions. Table 1 displays the ML estimates using a genetic algorithm. See Shmueli et al. [23] for details on implementing genetic algorithms for BARISTA model estimation.

Model selection Although three-stage models appear to be most suitable for describing the bid arrival process in online auctions, it is possible to extend the estimation process to include model selection. To allow for a more flexible family of distributions, we consider the family of one-stage (NHPP_1), two-stage (NHPP_2), and 3-stage (BARISTA) models. Since the first two are nested within the BARISTA model, we can choose the best model using likelihood ratios. To compare a 3-stage with a 2-stage model, for instance, we use the statistic

$$-2\left\{\mathcal{L}(NHPP_2) - \mathcal{L}(BARISTA)\right\}, \tag{15}$$

where $\mathcal{L}(i)$ is the log-likelihood for model i. Under the null hypothesis that the models are equivalent in their ability to fit the data (i.e., the NHPP_2 is sufficient), the statistic follows a $\chi^2(p)$ distribution with $p = 5 - 3 = 2$ degrees of freedom (the difference in the number of parameters of the two models). If the p-value is sufficiently small, then it is reasonable to choose the 3-stage model, whereas a large p-value would indicate the use of the 2-stage model. A similar statistic can be designed to test the difference between the 1-stage and 2-stage models, which would also follow a χ^2 distribution, again with $p = 3 - 1 = 2$ degrees of freedom. This test statistic can be used in conjunction with any of the estimation methods that we described. The most comprehensive and computationally intensive option is to find the "best" 1-stage, "best" 2-stage, and "best" 3-stage models (in the sense of the highest likelihood values), and compare them using the likelihood-ratio test. A more practical alternative is to combine the model selection with a stochastic search algorithm. For an illustration of applying model selection to various products in eBay auctions see Section 5 in Reference 23.

23.5 Relating Bidder Arrivals and Bid Arrivals

The online auction literature is rich with papers that assume an ordinary homogenous Poisson bidder arrival process. This assumption underlies various theoretical derivations, is the basis for the simulation of bid data, and is often used to design field experiments. Bajari and Hortascu [2] specify and estimate a structural econometric model of bidding on eBay, assuming a Poisson bidder arrival process. Etzion et al. [8] suggest a model for segmenting consumers at dual channel online merchants. Based on the assumption of Poisson arrivals to the website, they model consumer choice of channel, simulate consumer arrivals and actions, and compute relationships between auction duration, lot size, and the constant Poisson arrival rate λ. Zhang et al. [27] model

Table 1: True and estimated values (with standard errors) for the BARISTA model parameters, by method.

	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}_3$	$\hat{d}_1$	$\hat{d}_2$ (minutes)
Simulated (true) values	3	0.4	1	2.5	5
Quick & Crude	2.85 (0.06)	0.443 (0.001)	0.954 (0.0132)	2.5 (0.0036)	4.7 (0.13)
Genetic Algorithm	2.88 (0.007)	0.387 (0.005)	0.997 (0.009)	2.63 (0.0044)	4.6 (0.42)

the demand curve for consumer products in online auctions based on Poisson bidder arrivals, and fit the model to bid data. Pinker et al. [18] and Vakrat and Seidmann [24] use a Poisson process for modelling the arrival of bidders in going-going-gone auctions. They use the intensity function $\lambda(t) = \lambda_a e^{-t/T}, \leq t \leq T$, where T is the auction duration, and λ_a is the intensity of website traffic into the auction. This model describes the decline in the number of new bidders as the auction progresses. Haubl and Popkowski Leszczyc [10] design and carry out an experiment for studying the effect of fixed-price charges (e.g., shipping costs) and reserve prices on consumer's product valuation. The experiment uses simulated data that are based on Poisson arrivals of bidders. These studies are among the many that rely on a Poisson arrival process assumption.

In online auctions, however, bidder arrivals are unobserved as we pointed out earlier. Therefore it is not straightforward to study their distribution. On the other hand, bid arrivals *are* observed. In the following we investigate the relationship between bidder arrivals and bid arrivals more carefully.

The **General Poisson Bid Process** is a special case of the General Bid Process discussed in Russo et al. [21]. Suppose that m bidders participate in an auction that starts at time 0 and ends at time T. The variable m will generally be random (for example, bidders may enter the auction in accordance with a Poisson process). With bidder k, $1 \leq k \leq m$, associate a random triple $\theta_k = (X_k, \pi_k, H_k)$ comprised of a time of entry X_k of that bidder into the auction, a continuous function π_k mapping $[0,T]$ into $(0,1]$, and a distribution function H_k that satisfies $H_k(0) = 0 = 1 - H_k(T)$ and also possesses a derivative $h_k(s)$ on $[0,T]$. Set $b_{k,0} = X_k$. Bidder k places an initial bid at time $b_{k,1}$ where

$$\Pr(b_{k,1} \leq t) = \frac{H_k(t) - H_k(b_{k,0})}{1 - H_k(b_{k,0})}.$$

That is, at time $b_{k,0} = X_k$ the distribution of the bidder's initial bid is H_k restricted to the interval $[X_k, T] = [b_{k,0}, T]$. With probability $\pi(b_{k,1})$ he departs the auction (makes no further bids), or otherwise places a second bid at time $b_{k,2}$ where

$$\Pr(b_{k,2} \leq t) = \frac{H_k(t) - H_k(b_{k,1})}{1 - H_k(b_{k,1})}$$

so that $b_{k,2} \sim H_k$ restricted to $[b_{k,1}, T]$. With probability $\pi(b_{k,2})$ he departs the auction, or otherwise places a third bid at time $b_{k,3}$, etc., ultimately placing a random number of bids during $[0,T]$. Since π_k is a continuous function on a compact set, it achieves a minimal value $\pi_{k,\min}$ there, which (since the function maps into $(0,1]$) must be strictly positive. Therefore, the number of bids placed by bidder k during $[0,T]$ is stochastically bounded by a geometric($\pi_{k,\min}$) variable, and is hence finite. Suppose that

$$\begin{aligned}\theta_1 &= (X_1, \pi_1, H_1),\\ \theta_2 &= (X_2, \pi_2, H_2), \ldots,\\ \theta_m &= (X_m, \pi_m, H_m)\end{aligned}$$

is an i.i.d. sequence, independent of m. Under this model, each bidder acts independently in the manner outlined above. However, dependence among the components of θ_k is allowable, so that a bidder who tends to bid infrequently can be modelled as one who (for example) also tends to bid late.

Let $p\,(s \mid \theta_k)$ denote the probability that bidder k is *active* (still in the auction) at time $s \in [0, T]$. Note that a bidder who has not yet entered the auction is considered active under this definition. It was shown in Russo et al. [21] that

$$\begin{aligned}&p\,(s \mid \theta_k)\\ &= 1_{x<s} \exp\left[-\int_x^s \frac{\pi_k(t) h_k(t)}{1-H_k(t)}\, dt\right]\\ &\quad +1_{x \geq s}.\end{aligned}$$

Let $E_k(a, b)$ denote the event that bidder k places a bid during the interval $(a, b]$. The intensity of k's bid placement at time s is given by

$$\begin{aligned}&\lambda_k(s \mid \theta_k)\\ &= \lim_{\delta \to 0} \frac{P(E_k(s, s+\delta) \ \mid \ \theta_k)}{\delta}\\ &= \lim_{\delta \to 0} p\,(t \mid \theta_k)\, \frac{H_k(s+\delta) - H_k(s)}{\delta(1 - H_k(s))}\\ &= p\,(s \mid \theta_k)\, \frac{h_k(s)}{1 - H_k(s)}.\end{aligned}$$

The combined bidding intensity of all bidders at time s is thus given by

$$\begin{aligned}&\lambda(s \mid m, \theta_1, \ldots, \theta_m)\\ &= \sum\nolimits_{k=1}^m p\,(s \mid \theta_k)\, \frac{h_k(s)}{1 - H_k(s)}\end{aligned}$$

It is shown in Russo et al. [21] that, for given θ_k, the bid time sequence associated with bidder k is a nonhomogeneous Poisson process on $[x_k, T]$ that is stopped at a random time (the time of the final bid placement of bidder k).

If $\sup_{0 \leq s \leq T} \pi_k(s) < 1 - \delta$ for some $\delta > 0$ then for $s > x_k$,

$$\begin{aligned}&\lambda_k(s \mid \theta_k)\\ &\geq \exp\left[-(1-\delta) \int_0^s \frac{h_k(t)}{1 - H_k(t)}\, dt\right]\\ &\quad \times \frac{h_k(s)}{1 - H_k(s)}\\ &= \frac{h_k(s)}{(1 - H_k(s))^{\delta}}.\end{aligned}$$

Thus, if for some $k \in \{1, ..., m\}$, the function $\pi_k(s)$ maps into $(0, 1)$, and $\liminf_{s \to T} h_k(s)\,(1 - H_k(s))^{-\delta} > 0$ for all $\delta > 0$, we have $\lambda(s \mid m, \theta_1, ..., \theta_m) \to \infty$ as $s \to T$ (an exploding bid intensity as the auction deadline approaches). For self-similarity (with and without its breakdown at the auction end) we require some further restrictions as indicated in the examples below.

23.5.1 Example 1: One-Stage BARISTA, Pure Self-Similarity

Fix $r > 0$. Suppose $p_1, p_2, ..., p_m$ is an i.i.d. sequence of variables on $(0, 1)$, and that for $1 \leq k \leq m$ we have $\pi_k(s) = p_k$ (upon each bid placement, bidder k departs the auction with invariant probability p_k), and that

$$H_k(s) = 1 - \left(1 - \frac{s}{T}\right)^{r/p_k}.$$

The above equality ties the selection function H_k of bidder k to his invariant departure probability p_k. The greater this probability, the later in the auction bidder k tends to place his next bid. We have in this case

$$\begin{aligned} p(s \mid \theta_k) &= \exp\left[-p_k \int_{x_k}^{s} \frac{h_k(t)}{1-H_k(t)} dt\right] \\ &= \left[\frac{1-H_k(s)}{1-H(x_k)}\right]^{p_k} \\ &= \left(1-\frac{x_k}{T}\right)^{-r}\left(1-\frac{s}{T}\right)^{r} \end{aligned}$$

so that

$$\lambda(s \mid m, \theta_1, ..., \theta_m) = c\left(1-\frac{s}{T}\right)^{r-1}, \quad \max(x_1, ..., x_m) \le s \le T$$

where

$$c = \frac{r}{T}\sum_{k=1}^{m} \frac{1}{p_k}\left(1-\frac{x_k}{T}\right)^{-r}.$$

That is, conditional on the number of bidders m and their corresponding types $\theta_1, ..., \theta_m$ the intensity function λ has a pure self-similar form over the interval $[\max(x_1, ..., x_m), T]$.

Simulation 1. Suppose that $m = 1000$, $T = 7$, $X \sim U(0, 5.6)$, $\Pr(p_k < t) = 4t^2$ for $t \in (0, 1/2)$ and $r = 2/5$. In our simulation, 1000 bidders arrive uniformly during the first 5.6 days of a 7-day auction, their departure probabilities are individually invariant and constitute a random sample from a triangular distribution on $(0, 1/2)$. The parameter $r = 2/5$ guarantees that h increases as t approaches $T = 7$. The top panel in Figure 3 displays the empirical cumulative distribution functions of several sets of normalized left-truncated bid times based on our simulation. The bottom panel displays the same curves based on the palm eBay data.

Note the similarity of the curves (excluding the 5-minute curve) for the simulated and real data. The top display shows (as expected when the intensity has a self-similar form) the 5-minute curve to be of the same shape as the others. In the bottom display in Figure 3 (real data) we see the 5 minute curve breaking away from the self-similar form of the other curves.

23.5.2 Example 2: Two-Stage BARISTA

Fix $\beta > 0$ and $d \in (0, T)$. Suppose that we have the same set-up as in example 1 with the following exception: during the interval $[T-d, T]$ the departure probabilities of all active bidders are magnified by the common factor β so that

$$\pi_k(s) = p_k 1_{0 \le s < T-d} + \beta p_k 1_{s \ge T-d} \quad \text{for } 1 \le k \le m.$$

In this case

$$p(s \mid \theta_k) = \begin{cases} \exp\left[-p_k \int_{x_k}^{s} \frac{h_k(t)}{1-H_k(t)} dt\right] & x_k \le s \le T-d \\ \exp\left[-p_k \int_{x_k}^{T-d} \frac{h_k(t)}{1-H_k(t)} dt - \beta p_k \int_{T-d}^{s} \frac{h_k(t)}{1-H_k(t)} dt\right] & T-d \le s \le T \end{cases}$$

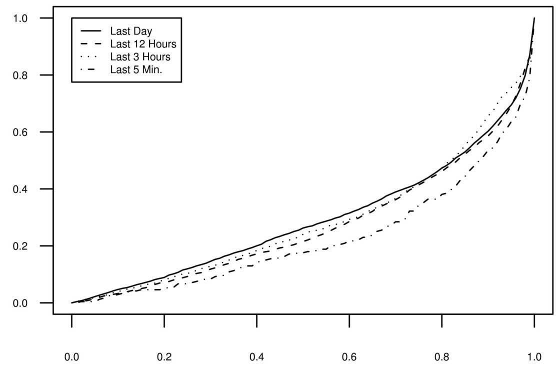
$$= \begin{cases} \left(1-\frac{x_k}{T}\right)^{-r}\left(1-\frac{s}{T}\right)^{r} & x_k \le s \le T-d \\ \left(1-\frac{x_k}{T}\right)^{-r}\left(\frac{d}{T}\right)^{r-r\beta}\left(1-\frac{s}{T}\right)^{r\beta} & T-d \le s \le T \end{cases}$$

so that with c as in Example 1 we have

$$\lambda(s \mid m, \theta_1, ..., \theta_m) = \begin{cases} c\left(1-\frac{s}{T}\right)^{r-1} & \max(x_1, ..., x_m) \le s \le T-d \\ c\left(\frac{d}{T}\right)^{r-\beta r}\left(1-\frac{s}{T}\right)^{r\beta-1} & T-d \le s \le T. \end{cases}$$

Note that the above λ has the form of the intensity function of a two-stage BARISTA process on $[\max(x_1, ..., x_m), T]$ with $d_1 = 0$, $d_2 = d$, $\alpha_2 = r$, and $\alpha_3 = \beta r$.

Simulation 2. We maintain the same set-up as in Simulation 1, except that we set $d = 1/10080$ and $\beta = 2$ thereby doubling the departure probabilities of all active bidders during the final minute of the



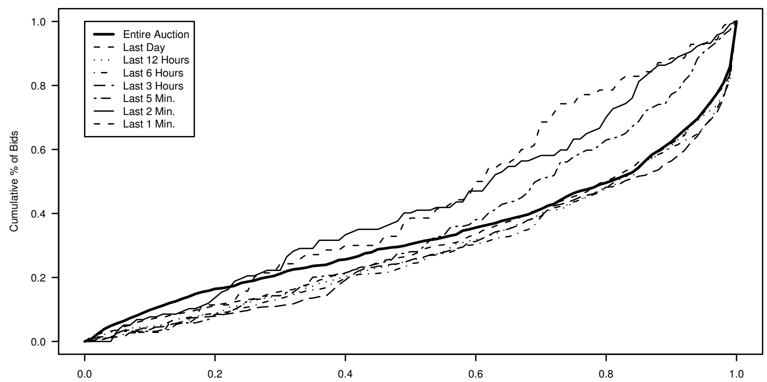


Figure 3: Empirical CDF functions of left-truncated bid times, based on Simulation 1 (top) and Palm data (bottom).

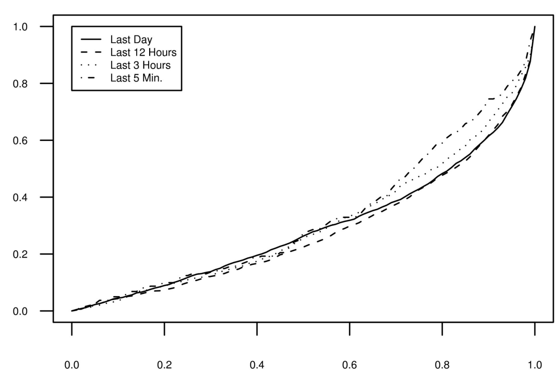


Figure 4: Empirical CDF functions of left-truncated bid times based on Simulation 2.

auction. Figure 4 displays the same collection of empirical cumulative distribution functions as in Simulation 1. Note the similarity of all the curves (including the 5 minute curve) to those in the bottom (real data) display in Simulation 1.

Example 2 demonstrates the existence of a plausible bidder arrival and behavior model that is able to replicate most of the prominent features of online auction bid data. As in the above simulation, the model can be used to generate bid arrival data for the various purposes that we have outlined in Section 23.2.

References

1. Aurell, E. and Hemmingsson, J. (1997). Bid frequency analysis in liquid Market. *TRITA-PDC Report*. ISRN KTH/PDC/R–97/3–SE. ISSN 1401-2731 (*http://www.pdc.kth.se/ payam/pub/AurellHemmingsson970208.ps*).
2. Bajari, P. and Hortacsu, A. (2000). Winner's curse, reserve price and endogenous entry: Empirical insights from eBay auctions. *Working paper*, Department of Economics, Stanford University.
3. Bapna R., Goes, P. and Gupta, A. (2003). Analysis and design of business-to-consumer online auctions. *Manage. Sci.*, **49**, 85–101.
4. Bapna R., Goes, P., Gupta, A., and Karuga, G. (2002). Optimal design of the online auction channel: Analytical, empirical, and computational insights. *Decision Sci.*, **33**, 557–577.
5. Borle S., Boatwright, P. and Kadane, J. B. (2006). The timing of bid placement and extent of multiple bidding: an empirical investigation using eBay online auctions. *Stat. Sci.*, **21**, 194–205.
6. Dennis, J. E. and Schnabel, R. B. (1983). *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*. Prentice Hall, Englewood Cliffs, NJ.
7. Efron, B. and Tibshirani, R. (1993). *An Introduction to the Bootstrap*. Number 57 in Monographs on Statistics and Applied Probability. Chapman & Hall, New York.
8. Etzion, H., Pinker, E., and Seidmann, A. (2003). Analyzing the simultaneous use of auctions and posted prices for on-line selling. *Working Paper CIS-03-01*, Simon School, University of Rochester.
9. Gwebu, K., Wang, J., and Shanker, M. (2005). A simulation study of online auctions: Analysis of bidders' and bid-takers' strategies. In *Proceedings of the Decision Science Institute*.
10. Haubl, G. and Popkowski Leszczyc, P. T. L. (2003). Minimum prices and product valuations in auctions. *Marketing Science Institute Reports*, Issue 3, No. 03-117, 115–141.
11. Huberman, B. A. and Adamic, L. A. (1999). Growth dynamics of the World-Wide Web. *Nature*, **401**, 131.
12. Jank, W. and G. Shmueli (2005). Profiling price dynamics in online auctions using curve clustering. *Robert H. Smith School Research Paper No. RHS-06-004*, Available at SSRN: *http://ssrn.com/abstract=902893*.
13. Kauffman, R. J. and Wood, C. A. (2000). Running up the bid: modeling seller opportunism in internet auctions. In *Proceedings of the 2000 Americas Conference on Information Systems*.
14. Ku, G., Malhorta, D., and Murnighan, J. D. (2004). Competitive arousal in live and internet auctions. *Working paper*, Northwestern University.
15. Liebovitch, L. S. and Schwartz, I. B. (2003). Information flow dynamics and timing patterns in the arrival of email viruses. *Phys. Rev. E*, **68**, 017101-1–017101-4.
16. McAfee, R. P., and McMillan, J. (1987). Auctions with stochastic number of bidders. *J. Econ. Theory*, **43**, 1–19.
17. Menasce, D. A. and Akula, V. (2004). Improving the performance of online auction sites through closing time reschedul-

ing. In *The First International Conference on Quantitative Evaluation of Systems (QEST'04)*, pp. 186–194,

18. Pinker, E., Seidmann, A., and Vakrat, Y. (2003). The design of online auctions: Business issues and current research. *Manage. Sci.*, **49**, 1457–1484.

19. Roth A.E., Murninghan, J. K. and Schoumaker, F. (1998). The deadline effect in bargaining: some experimental evidence. *Am. Econ. Rev.*, **78**, 806–823.

20. Roth A. E. and Ockefels, A. (2000). Last minute bidding and the rules for ending second-price auctions: theory and evidence from a natural experiment on the internet. *NBER Working Paper*, #7729.

21. Russo, R. P., Shmueli, G., and Shyamalkumar, N. D. (2008). Models of bidder activity consistent with self-similar bid arrivals. In *Statistical Methods in eCommerce Research*, W. Jank and G. Shmueli, eds., Wiley, New York.

22. Shmueli, G., Russo, R. P. and Jank, W. (2004). Modeling bid arrivals in online auctions. *Robert H. Smith School Research Paper No. RHS-06-001*, Available at SSRN: *http://ssrn.com/abstract=902868*.

23. Shmueli, G., Russo, R. P., Jank, W., (2007). The BARISTA: A model for bid arrivals in online auctions. *Ann. App. Stat.*, **1**, 412–441.

24. Vakrat Y. and Seidmann, A. (2000). Implications of the bidders' arrival process on the design of online auctions. In *Proceedings of the 33rd Hawaii International Conference on System Sciences*, pp. 1–10.

25. Wilcox, R. T. (2000). Experts and amateurs: The role of experience in internet auctions. *Market. Lett.*, **11**, 363–374.

26. Yang, I., Jeong, H., Kahng, B., and Barabasi, A.-L. (2003). Emerging behavior in electronic bidding. *Phys. Rev. E*, **68**, 016102.

27. Zhang, A., Beyer, D., Ward, J., Liu, T., Karp, A., Guler, K., Jain, S., and Tang, H. K. (2002). Modeling the price-demand relationship using auction bid data. *Hewlett-Packard Labs Technical Report HPL-2002-202* (*http://www.hpl.hp.com/techreports/2002/HPL-2002-202.eps).*