# Estimating Functional Agent-Based Models: An Application to Bid Shading in Online Markets Format

Wei Guo*
Applied Mathematics & Statistics, and Scientific Computation Program, Department of Mathematics, University of Maryland, College Park, Maryland, USA
weiguo@math.umd.edu

Wolfgang Jank
Department of Decision, Operations & Information Technology & Center for Complexity in Business, Robert H. Smith School of Business, University of Maryland, College Park, Maryland, USA
wjank@rhsmith.umd.edu

William Rand
Department of Marketing & Department of Computer Science & Center for Complexity in Business, Robert H. Smith School of Business, University of Maryland, College Park, Maryland, USA
wrand@umd.edu

## ABSTRACT

Bid shading is a common strategy in online auctions to avoid the "winner's curse". While almost all bidders shade their bids, at least to some degree, it is impossible to infer the degree and volume of shaded bids directly from observed bidding data. In fact, most bidding data only allows us to observe the resulting price process, i.e. whether prices increase fast (due to little shading) or whether they slow down (when all bidders shade their bids). In this work, we propose an agent-based model that simulates bidders with different bidding strategies and their interaction with one another. We calibrate that model (and hence estimate properties about the propensity and degree of shaded bids) by matching the emerging *simulated price process* with that of the observed auction data using genetic algorithms. From a statistical point of view, this is challenging because we match functional draws from simulated and real price processes. We propose several competing fitness functions and explore how the choice alters the resulting ABM calibration. We apply our model to the context of eBay auctions for digital cameras and show that a balanced fitness function yields the best results.

## Categories and Subject Descriptors

J.1 [**Administrative Data Processing**]: Marketing; I.6.3 [**Simulation and Modeling**]: Applications

## General Terms

Management, Economics

*The authors are listed in alphabetical order.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
*GECCO'11,* July 12–16, 2011, Dublin, Ireland.
Copyright 2011 ACM 978-1-4503-0690-4/11/07 ...$10.00.

## Keywords

Internet Auctions, Agent Based Modeling, Calibration, Business, Simulation, Genetic Algorithms

## 1. INTRODUCTION

In this paper, we propose a method for calibrating agent-based models to *functional data*. We focus specifically on functional data [19], because ABMs are often used to model processes where emerging phenomena are measured over time. Repeated application of the same ABM then results in multiple replications of the same measurement-process and we can interpret this process as a functional observation. We propose the use of a genetic algorithm (GA) [10] in a functional data analysis framework to calibrate a complex ABM.

We test and apply our methodology in the context of online auctions. In particular, we are interested in quantifying bidders' propensity to shade their bids, that is, their aversion to risk. Bid shading is related to the "winner's curse" [1], that is, bidders, recognizing that winning an auction is conditional on being the most optimistic bidder about an item's worth, respond strategically by lowering their actual bids below their WTP [2]. In an auction setting, bid shading cannot be directly observed and we thus propose the use of ABMs to infer it from observable data.

To that end, we design an agent-based modeling framework with each bidder represented as an agent who has to make several repeated decisions: whether to place a bid, when to place a bid, how much to bid and whether the bidding process should be repeated once another bidder places a higher bid. In a real auction setting, the cause for each of these decisions cannot be observed directly. We hypothesize that the reasons are linked (among other things) to a bidder's willingness to take on risks and hence "equip" our bidding agents with varying levels of bid shading. Using our proposed methodology, we identify the model parameters which most closely match real auction data using several different fitness functions and find that most of the bidders are averse to the risk of overpaying (i.e. they prefer to make conservative bids despite a high willingness to pay).

This paper is organized as follows. In Section 2, we will review related research on agent based modeling, internet auctions and genetic algorithm. Section 3 gives an overview of the data and describes exploratory analyses. In Section 4 we provide a detailed specification of our agent based model. In Section 5, we discuss calibration of the proposed agent based model to real auction data and model identifiability. Section 6 discusses model results and inferences. In Section 7, we conclude with future research directions.

## 2. LITERATURE REVIEW

One of the main reasons for the recent increase in popularity of agent-based models is that it enlarges the set of questions we can explore [17]. In contrast to classical statistical models which rely on restrictive – and at times unrealistic – assumptions (such as linearity, homogeneity, normality and stationarity) which are often imposed for mathematical analysis and proof rather than practical applicability, agent-based models operate within a framework of minimal, simple and very realistic rules. As a result, ABM allows researchers to examine issues that have been avoided previously in theoretical disciplines and for which mathematical and analytical derivation is impossible [3].

We apply the ABM framework to the context of online auctions. The spectacular growth of internet auctions and the availability of massive amounts of auction data has lead to new insights into bidder-seller behavior [4, 11, 8, 25], the impact of the auction format [20], the auction process [27, 24, 1, 5, 13, 22] and its dynamics [18, 6]. However, while most of the extant literature focuses on behavior that is directly measurable from observed data (such as the timing and the magnitude of individual bids), we are interested in behavior that is *unobservable*. In particular, we are interested in bidders' propensity to shade their bids, that is, the extent to which bidders bid below their true willingness to pay. In an auction setting, bid shading can not be directly observed but it is reflected in the amount and timing of their *bid increment*. For instance, a risk averse bidder, who is afraid of overpaying for an item, will bid closer to the minimum required bid amount (despite having a much higher willingness to pay). On the other hand, an aggressive bidder, who cares more about winning than the final price, will bid high (i.e. closer to his true willingness to pay) early to deter other bidders and thus to increase his own chances of winning. The difference between a bidder's willingness to pay and his actual bid is often referred to as the amount of the *shaded bid* and we can view it as a measure of a bidder's riskiness. In this work, we will infer the distribution of shaded bids using ABMs.

We calibrate our ABM via a genetic algorithm (GA) [12, 10] that searches for the degree and volume (i.e. the distribution) of bid shading that best matches the observed data. Weinberg [26] proposed one of the earliest applications of GAs to characterize the parameters of a cell simulation. Later, Miller [16] proposed the use of nonlinear optimization techniques for a variety of model-exploration and -testing tasks, dubbed "Active nonlinear testing" or ANT [16]. In recent years, GA has been used for several parameter search tasks in the context of ABMs in a variety of domains including: ant food foraging [7], consumer retail environments [15] and viral marketing strategies [23].

## 3. ONLINE AUCTION DATA

The dataset used for this study is based on 1104 new Canon SD1000 digital camera auctions sold on eBay.com. eBay is the world's largest internet market place with more than 90 million active users globally. The eBay website uses a proxy bidding mechanism, in which the highest bidder wins and pays the second highest bid amount of the auction plus a minimum increment[1]. One thing worth noting is that the auction mechanism of eBay shows the current second highest bid as the current auction "price" and the system does not reveal bidders' real bid until they are out-bid by others, thus the highest bid of an auction is not revealed on the eBay website. By contrast, our dataset which was obtained directly from eBay contains the real bid of each bidders. Since all the auctions sell the identical product, we are controlling for heterogeneity due to product differences. In addition, there are many repeat sellers ("PowerSellers") and we hence also have very little variation due to sellers. Moreover, most auctions share a similar format. Thus, most of the variability we observe is due to the bidders and their different strategies. This is summarized in Table 1.

**Table 1: Bidder characteristics of the digital camera auctions. A high bid increment is defined as an increment of at least 100% over the previous bid. A low bid increment is defined as an increment of less than 5% increment over the previous bid.**

| | Mean | Std. dev. |
|---|---|---|
| Number of bidders per auction | 8.2 | 4.0 |
| Number of repeat bids per bidder | 0.97 | 3.63 |
| Prop. of early bidders | 15.9% | |
| Prop. of last-minute bidders | 10.8% | |
| Prop. of high bid increments | 13.8% | |
| Prop. of low bid increments | 37.8% | |
| Prop. of bidders with more than 1 bid | 26.6% | |

Table 1 shows that, despite a very homogeneous auction setting (i.e. same product, similar bidder and similar auction format), bidders' behavior ranges vastly. For instance, while only 15.9% of all bidders place bids early (within the first 10% duration of the auction), even fewer (10.8%) place them within the last minute[2]. Moreover, while 13.8% place bids higher than 100% over the previous bid ("high bid increment"), many more bid more conservatively (i.e. 37.8% with a low bid increment). This shows that bidders' strategies vary enormously. In fact, differences in timing and magnitude of bidders' bids suggest that there exist enormous variety in bidders' willingness to assume risk; in other words, the data suggests that some bidders shade their bids much more than others.

While Table 1 suggests that there are differences in bidder's strategies, simple summary statistics cannot capture the *interaction* of bidders' behavior. Bidders compete against one another [11] and they react to each others' moves. In fact, a bidder's strategy might adjust as a result of other bidders' actions. Measuring the reaction of one bidder to

[1]The incremental amount is predetermined based on the current high bid on the item. The incremental table of eBay auctions is available at http://pages.ebay.com/help/buy/bid-increments.html, which is also been adopted to our ABM.

[2]Last-minute bidding is often regarded as one of the most popular bidding strategies [2].

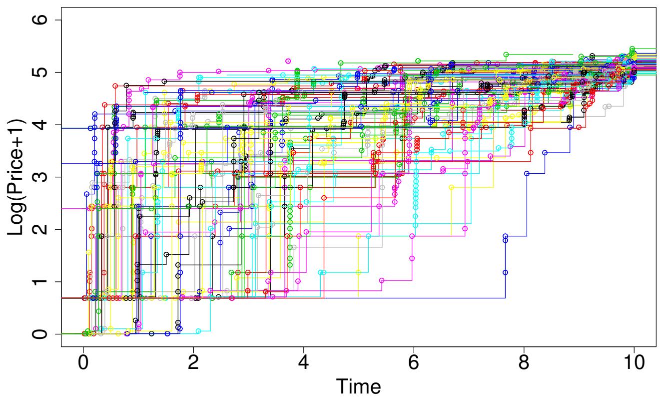


**Figure 1: Price Curves of observed eBay auction data.**

another bidder's action is impossible. However, what we *are* able to measure is the emerging phenomenon, that is, the resulting price and the rate at which price changes. Take a look at Figure 1 which shows prices curves for all auctions in our data. We can see that, despite the homogeneity of the product, the seller and the format, price curves vary drastically, with some curves moving very slow initially, only to speed up towards the end. In contrast, other price curves climb fast during the early stages of the auction and level-off later. This change in dynamics (fast price increase vs. price deceleration) is a result of different bidders and their interaction with one another. In other words, the price curves (and in particular their shape) allow us to capture the action and reaction of bidders to one another and the change in their behavior. By focusing on the shape of the price curves (and their dynamics), we adopt a point of view that borrows ideas from the field of *Functional Data Analysis* [19, 9, 27, 24].

# 4. FUNCTIONAL AGENT-BASED MODEL FOR ONLINE AUCTIONS

As mentioned earlier, our objective is to understand bidders' propensity to shade their bids, which is unobservable. To that end we develop a model which simulates different levels of bid shading. Bidders' actions (including their willingness to assume risk) are interconnected in the sense that the timing and magnitude of one bidder's bid will influence the reaction of all remaining bidders. As pointed out above, the interplay between action and reaction is observable only in the price curve and its shape. To that end, we view the output of the agent based model as a *functional object*, and compare it with the observed price curves. In what follows, we explain how to model this price curve by making assumptions only about the individual bidder-level interactions. In the subsequent section, we will then propose ways for matching this simulated price curve to observed price histories from real auctions data.

## 4.1 Model Parameters

An online auction (such as on eBay.com) consists of several key components: an item to be sold, a seller selling that item, an auction format that describes the rules of the transaction and a set of bidders. In the following, we describe each of these components separately.

### 4.1.1 The Item's Value

The item's value is given by the bidder with the highest willingness to pay (WTP). Different bidders possess different WTPs since information about an item's value is gathered from a variety of sources, including internet search, in store prices and advertisements. As a result, bidders' evaluations tend to fluctuate randomly around the average market value. While the WTP distribution could have a variety of different shapes, we do not have any reasons to assume anything different but a symmetric shape. In fact, it is quite plausible that some bidders value an item above the retail price while it is below market value for others. Theoretical modeling (e.g. [14]) thus often assumes a uniform distribution, typically out of mathematical convenience. We generalize this assumption by allowing for a Normal fluctuation around the item's market value, i.e. we assume a Normal distribution for bidders' WTP:

$$w_{k,i} \sim \text{Normal}(\mu_w, \sigma_w^2).$$

As pointed out earlier, bidders' actual bids might be different from their WTP to the winner's curse and the resulting propensity to shade bids. We will thus model bid shading explicitly further below.

### 4.1.2 The Seller and the Auction

As pointed out above, our data is very homogeneous in terms of the seller characteristics and the auction format. We will hence only allow variation in the starting price and assume other auction parameters are constant in our simulations. In our simulations, we model the variation in starting prices using a re-scaled Beta distribution because it provides the best fit to the observed variation in starting prices. Also, the length of auctions are standardized to 10 days.

### 4.1.3 The Bidders

Most of the dynamics of our simulation are focused on bidders and their interactions with one another. Bidders' strategies are determined by three key elements: (1) the number of competing bidders, (2) the timing and (3) the magnitude of the bid.

**A) The Number of Bidders**

The number of bidders determines the overall level of competition in an auction and more bidders competing for the same item usually results in an increase of the final price. We distinguish between the number of *potential* bidders and *actual* bidders. A potential bidder might be interested in the auction, and she might monitor the auction progress, but she might never decide to place a bid because the current price might be higher than her own willingness to pay. Thus, the number of actual bidders is a subset of the total number of potential bidders. We model the number of potential bidders according to a Poisson distribution, which is a common assumption for bidders' arrival rate [21, 24]. The number of potential bidders of each auction is draw from a Poisson distribution in the beginning of the simulation and this number will not change thereafter.

**B) The Bid-Timing**

Each of the potential bidders engages in a hierarchical decision process. Upon a bidder's arrival, first she decides whether or not to place a bid. She may not place a bid because the current bid is higher than her willingness to pay. Also, if a bidder is revisiting an auction and is still in the lead, she may decide not to bid again (and outbid herself). As a result, not all potential bidders' arrival events will result in actual bids. Thus, the actual bids are a subset of all arrival events. After a bidder decides that she will place a bid, she decides how much to bid based on both the auction price and her own willingness to pay. The details of the bid amount will be discussed further below. At the time of a bidder's first arrival, she decides how often to check back (i.e. revisit). Since some bidders place multiple bids in the same auction, the bidder's first time of visit $t_{k,i,1}$ is modeled separately from potential revisits $t_{k,i,j}, j \geq 2$. The reason we separate each bidder's time of first visit from revisits is that a bidder's time of first visit, $t_{k,i} = t_{k,i} - 0$, could be very different from subsequent revisiting time intervals, $t_{k,i,j} - t_{k,i,j-1}, j \geq 2$. For example, a bidder may notice an online auction very late, but since she is very interested in the auction, she revisits the auction frequently. Thus, this bidder will make her first visit to the auction very late, but subsequently revisit that auction in very short time intervals. On the other hand, there exist bidders who bid very early once but rarely come back to check the auction. Therefore these bidders will have a very early first visit, but longer time intervals for their revisits, and some may never return to the auction.

As a result, we model bidders' first arrivals in the following way. After generating the total number of bidders in an auction, the beta distribution is used to generate a bidders' first arrival time. That is, the first arrival time of bidder $i$ in $k$th auction is given by

$$t_{k,i,1} = T_k \times X_{k,i}, X_{k,i} \sim \text{Beta}(\alpha_t, \beta_t)$$

We then generate a bidder's revisiting time using a Poisson process. For a Poisson process, if the given number of arrivals in $[t_1, t_2]$ is $n$, then the $n$ unordered arrival epochs are i.i.d. uniformly distributed on $[t_1, t_2]$. Following this assumption, each bidder's time epochs of revisit are generated in two steps. First, each bidder's number of revisits $n_{k,i}$ is drawn from a Poisson distribution. Then, bidder $i$'s time of revisits are generated from uniform distribution on the interval $[t_{k,i,1}, T_k]$.

**C) The Bid-Amount**

The bid amount is dependent on the current price: a bidder will decide the amount of her bid after checking the current auction price. More specifically, three factors typically influence how much a bidder will bid: (1) the current auction price, (2) the bidder's own WTP and (3) the bidder's sensitivity to risk (i.e. the amount of bid shading).

We model bid shading in the following way. Let $\rho_{k,i}$, $0 < \rho_{k,i} < 1$, denote the amount by which a bidder bids *below* her WTP, where $\rho_{k,i}$ is drawn from a Beta distribution, $\rho_{k,i} \sim \text{Beta}(\alpha_\rho, \beta_\rho)$. The Beta distribution allows for flexibility in capturing different volume and degree of bid shading. For instance, Figure 2 shows examples of the Beta distribution for six different parameter pairs $(\alpha_\rho, \beta_\rho)$. We can see that $(\alpha_\rho, \beta_\rho) = (1, 1)$ (top left corner) results in a uniform distribution; the implication of that distribution is a pool of bidders with very diverse amounts of bid shading,

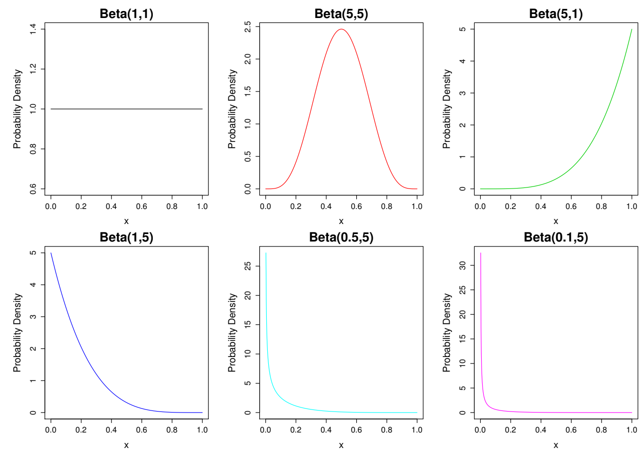


**Figure 2: Probability Density Function of Bid Shading based on the Beta distribution**

**Table 2: List of Variables in the ABM.**

| Notation | Definition | Distribution |
|---|---|---|
| $N_k$ | Number of bidders | Poisson($\lambda_N$) |
| $P_k$ | Starting price | 180·Beta($\alpha_p, \beta_p$) |
| $w_{k,i}$ | Willingness to pay | Normal($\mu_w, \sigma_w^2$) |
| $\rho_{k,i}$ | Bid shading | Beta($\alpha_\rho, \beta_\rho$) |
| $n_{k,i,j}$ | Number of revisits | Poisson($\lambda_{\text{re}}$) |
| $t_{k,i,1}$ | First visit time | 10·Beta($\alpha_t, \beta_t$) |
| $t_{k,i,j}$ | Revisit times $j \geq 2$ | Uniform[$t_{k,i,1}$, 10] |

with some bidders shading their bids almost entirely while others do not shade their bids at all. The bottom right $(\alpha_\rho, \beta_\rho) = (0.1, 5)$ shows an example of a very right-skewed bid shading distribution; in that case, most bidders would shade their bids almost entirely while only few bidders bid close to their WTP; we can think of this scenario as one of very conservative bidders. In contrast, the top right distribution $(\alpha_\rho, \beta_\rho) = (5, 1)$ is highly left-skewed and represents bidders who bid very aggressively (i.e. close to their WTP). We will come back to these bid shading scenarios in subsequent analyses.

Let $P_{k,m}$ denote the current price, and let $Inc(P_{k,m})$ denote the minimum bid increment; recall that $w_{k,i}$ denotes the bidder's WTP. We then model the bidder's $m$th bid in auction $k$ as the current price plus required increment, plus the shaded difference between the WTP and $P_{k,m} + Inc(P_{k,m})$; or:

$$B_{k,m} = P_{k,m} + Inc(P_{k,m}) + (w_{k,i} - P_{k,m} - Inc(P_{k,m})) \times \rho_{k,i}.$$

## 4.2 Simulation Implementation

After generating the simulation variables, we construct an agent-based model that simulates each bidder's arrival event according to a decision-making process. Upon a bidder's arrival at an auction, she first checks whether she is the current winner of the auction. If so she will just leave the auction as is, otherwise she will check the current price of the auction. If the minimum required bid amount is lower than her WTP, she will make a bid, otherwise she will leave the

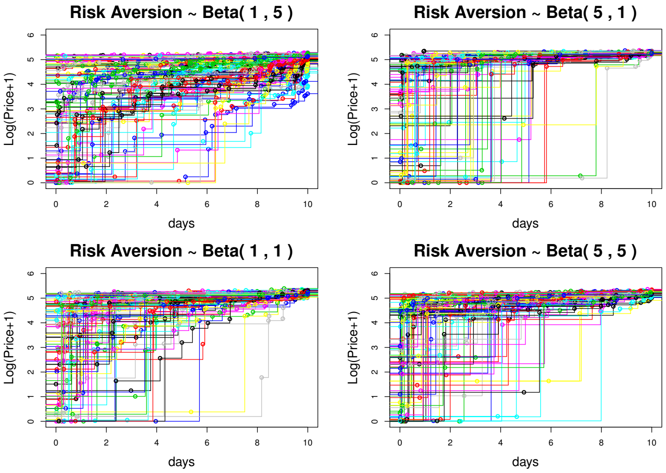


**Figure 3: Simulated Price Paths for different levels of bid shading.**

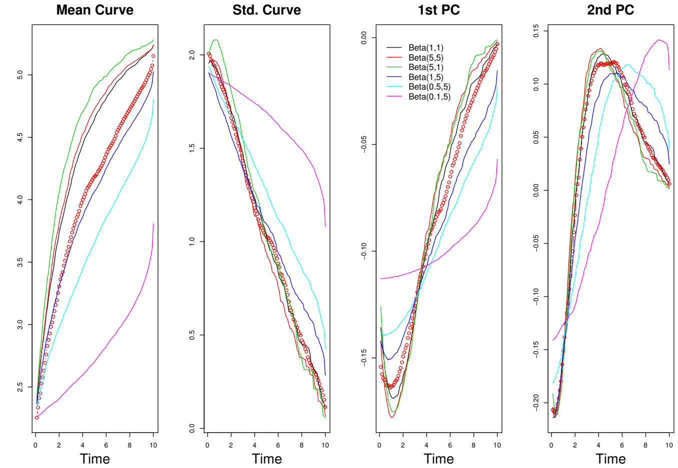


**Figure 4: Visual Comparison of Summary Curves**

auction. Her bid amount, denoted by $B_{k,m}$ is affected by the current price $P_{k,m}$, her own WTP $w_{k,i}$, and bid shading $\rho_{k,i}$. There are two possible scenarios after the bidder bids, since the current price that is shown is only the second highest bid of the auction: (1) If the new bid is higher than the previous highest bid, the bidder becomes the current auction leader and the new bid becomes the highest bid of the auction, or (2) the current bidder is automatically overbid and the bidder of the previous highest bid is still the leader, but the current price of the auction is updated to the current second highest bid plus the minimum bid increment. After the bidder makes a bidding decision (including whether to bid and the bid amount) and the auction price is updated, the simulation goes to the next arrival event.

## 4.3 Visual Parameter Estimation

In order to implement this ABM, we need to calibrate the parameters from Section 4.1. Table 2 lists all the 7 variables and their associated 10 parameters used in our ABM. For most of these variables, we can observe data directly, thus the corresponding parameters are estimated using maximum likelihood.

However, bid shading is unobservable. Figure 3 illustrates the price path of 100 simulated auctions with four different distributional assumptions (see Figure 2) for the bid shading parameter $\rho_{k,i}$, and all other parameters kept their estimation values (will be discussed in Sec 6.2). The top row represents situations in which bidders bid conservatively (left panel) or aggressively (right panel) and the bottom row shows symmetric situations (uniform bid shading in the left panel, bell-shaped bid shading in the right panel). We can see that for conservative bidding, price increases very slowly and there is large variation in the closing price. On the other hand, with mostly aggressive bidders, price increases quickly, the closing price is higher and exhibits less variation. Finally, in the case of symmetric bid shading, there is more variation in the price paths. It is interesting that, at least from a pure visual inspection, the simulated price curves pertaining to the conservative bidders (top left panel in Figure 3) best resembles the observed data in Figure 1. In other words, it suggests that the degree and volume of bid shading in real eBay auctions could be characterized by a Beta distribution with parameters 1 and 5. However, while such a visual matching provides some initial insight, it is not precise. To that end, we now derive a method via the genetic algorithm.

# 5. ESTIMATION VIA GA

In order to use GA to calibrate our auction ABM, the first step is to define a proper fitness function. Note that our task involves matching observed (Figure 1) and simulated (Figure 3) price curves. Since matching a sample of simulated curves to a sample of observed curves is challenging, we propose to match corresponding *summary curves* instead. That is, we propose to match the mean curve of the simulated data ($\mu(t)^{sim}$) to the mean curve of the observed data ($\mu(t)^{obs}$). Similarly, in order to capture variability in the observed price curves, we also match the corresponding standard deviation curves ($\sigma(t)^{sim}$ and $\sigma(t)^{obs}$). And finally, since we have argued earlier that the *shape* of the price curve is of particular importance, we also match the corresponding first and second *principal component curves* ($PC1(t)^{sim}$, $PC1(t)^{obs}$ and $PC2(t)^{sim}$, $PC2(t)^{obs}$), respectively[3].

Figure 4 illustrates these four summary curves for the observed (red circles) and simulated (thin lines) data. We can see e.g. in the leftmost panel that the mean of the simulated curves ($\mu(t)^{sim}$) pertaining to Beta(1,5) (blue line) best tracks the mean of the observed price curves ($\mu(t)^{obs}$). On the other hand, the standard deviation curve ($\sigma(t)^{sim}$) pertaining to Beta(0.1, 5) (pink line) is most dissimilar compared to the corresponding standard deviation curve of the observed data (second panel). Overall, while these summary curves allow some comparison between simulated and observed data, detecting the best match is hard, at least visually.

To that end, we define a fitness function across all four summary curves. In fact, we propose a weighted root means

[3]The first and second principal component curves capture trends (e.g. up/down) or curvature (e.g. concave/convex) of functional objects; see e.g. [9].

squared error (RMSE) criterion of the following form:

$$\begin{aligned} F_{\mathbf{y}}(\mathbf{x}) &= w_1 \cdot \left|\mu(t)^{sim} - \mu(t)^{obs}\right|_{\text{RMSE}} \\ &+ w_2 \cdot \left|\sigma(t)^{sim} - \sigma(t)^{obs}\right|_{\text{RMSE}} \\ &+ w_3 \cdot \left|PC1(t)^{sim} - PC1(t)^{obs}\right|_{\text{RMSE}} \\ &+ w_4 \cdot \left|PC2(t)^{sim} - PC2(t)^{obs}\right|_{\text{RMSE}}, \end{aligned}$$

where the RMSE of two functions is defined as

$$|f(t) - g(t)|_{\text{RMSE}} = \sqrt{\frac{\sum_{j=1}^{100}(f(t_i) - g(t_i))^2}{n}}.$$

Our ABM and GA selection of parameters are implemented in the freely available software R. We implemented the GA using the "genalg" package in R, with a few modifications to better measure the results. The GA is used to optimize 6 parameters of the ABM. Each parameter is specified by a real number, so each individual in the GA is composed of 6 genes that represent a potential parameter set of our ABM. To evaluate the corresponding fitness function of each individual, 1000 simulations are carried out to generate summary price curves. The population size of the GA is 100 and the number of iterations is 100. So, in total there are 1000 simulations $\times$ 100 individuals $\times$ 100 generations $= 10,000,000$ simulated auctions used in the estimation process. In each iteration, the 20 best individuals are kept for the next generation (elitism), while the other 80 pairs of parents are randomly selected (each pair of parents only have one offspring). The crossover rate is 60% and the mutation rate is 5% per gene. The range of each parameter are based on inference from visual inspection and simulation experiments.

# 6. RESULTS AND DISCUSSION

## 6.1 Fitness Function Weighting

First, we investigate the impact of the weights $w_i$ on the fitness function. To that end, we investigate three different scenarios: In scenario 1, we set $w_1 = 0.75$, $w_2 = 0.25$, and $w_3 = w_4 = 0$; this scenario only uses the mean and standard deviation curves and hence ignore the shape of the price curves via PC1 and PC2. Scenario 2 is reversed and puts all weight on the shapes PC1 and PC2 and no weight on the mean and standard deviation curves (i.e. $w_1 = w_2 = 0$, and $w_3 = 0.75$ and $w_4 = 0.25$). And finally scenario 3 uses all four components in a balanced form via $w_1 = 0.35$, $w_2 = 0.15$, $w_3 = 0.35$ and $w_4 = 0.15$.

We first examine these three scenarios on a synthetic data set. That is, we generate data with a known set of parameters and then use GA to extract the original underlying bid shading parameters. The results of this experiment is given in the left panel of Figure 5. We can see that the results are very robust to the choice of the weights on the synthetic data. In other words, regardless of the choice of the weights, the algorithm produces (almost) the identical result.

We repeat the same experiment on the observed data and find that the outcome is more sensitive to the choice of the weights (right panel of Figure 5). While the outcome is more variable, we do observe that, regardless of the choice of the weights, the method estimates a right-skewed distribution for bid shading. Thus we can conclude that while the choice of the weights matters, it does not impact the results too much in that, qualitatively, our overall conclusions remain the same.

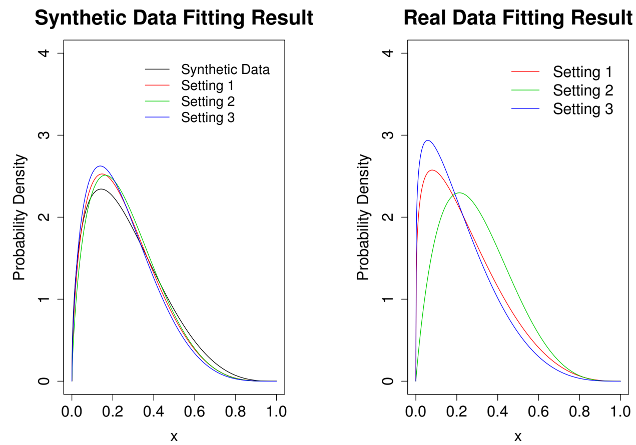


**Figure 5: Different Fitness Function Settings and Resulting Probability Density Function of Beta Distributions of Synthetic and Real Data**

**Table 3: RMSE**

| Weighting | $\mu$ | $\sigma$ | PC1 | PC2 |
|---|---|---|---|---|
| Setting 1 | 0.03696 | 0.07417 | 0.00746 | 0.01451 |
| Setting 2 | 0.25209 | 0.06321 | 0.00488 | 0.01146 |
| Setting 3 | 0.03765 | 0.05276 | 0.00505 | 0.00671 |

Finally, in order to evaluate the difference between the different fitness function weighs, we examine the RMSE of the resultant components of the price curves, i.e, $\mu$, $\sigma$, PC1 & PC2 for the three different settings and the real data. The hypothesis is that if one of the three fit the data better on all four components that would be the best choice for the weighting. Table 3 and Figure 6 shows the results. We can see that scenario 3, which uses both the mean and standard deviation curves as well as the shapes captured by PC1 and PC2, provides the best results on almost all four RMSE values with RMSE of PC1 a little greater than the scenario 2, so for the rest of this paper we will utilize this scenario.

## 6.2 Estimation Result

We now apply our algorithm to the eBay auction data in the following way. As pointed out earlier, we estimate some of the ABM parameters directly from the data (e.g. starting price and each bidder's first visit time) using maximum likelihood estimation. For other parameters, we use maximum likelihood estimates as starting values and then update these values inside the GA. The reason we combine this two method together is that it is only possible to get good initial values of parameters by just using MLE. For example, the average number of different bidders in each auction is a good initial value for the arrival rate of bidders, but we would expected the true parameter is greater then the observed average, because there are potential bidders that checked the auction status but did not bid [6]. Simi-

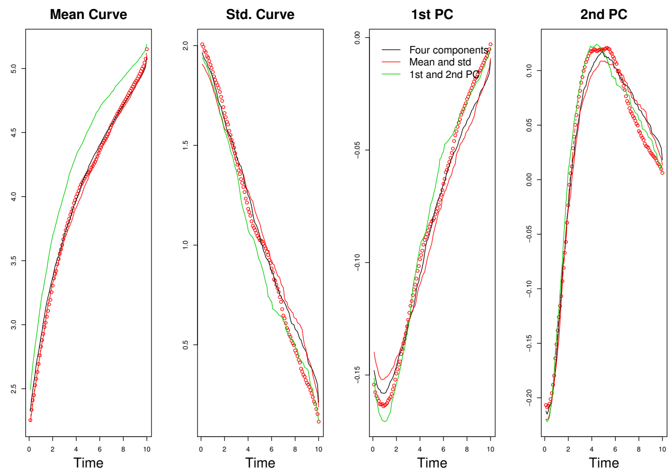


**Figure 6: Visual Comparison of Different Fitness Function Settings' Effect on Price Curves.**

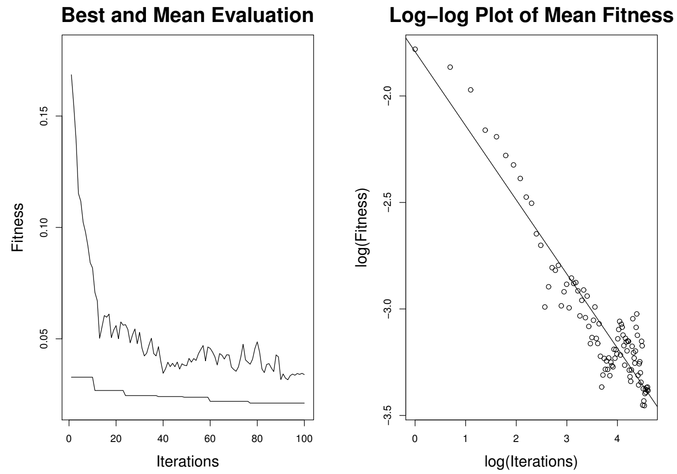


**Figure 7: GA performance**

**Table 4: Estimation Results**

| Variable | Parameter | Method | Estimation |
|---|---|---|---|
| Number of bidders | $\lambda_N$ | MLE&GA | 12.35 |
| Starting price | $\alpha_p, \beta_p$ | MLE | 0.17, 0.59 |
| Willingness to pay | $\mu_w, \sigma_w^2$ | MLE&GA | 174.4, 21.0 |
| Bid shading | $\alpha_\rho, \beta_\rho$ | GA | 1.2,4.2 |
| Number of revisits | $\lambda_{\mathrm{re}}$ | MLE&GA | 0.97 |
| First visit time | $\alpha_t, \beta_t$ | MLE | 0.58, 0.34 |

larly, by just using MLE from observed data, the inference of bidders' revisit rate $\lambda_{\mathrm{re}}$ and WTP $w_{k,i}$ would be biased. So both MLE and GA are used in estimation of parameters. The parameters for the bid shading distribution are estimated entirely from the GA. Table 4 shows the estimation results and the estimation methods.

## 6.3 Computational Error Analysis

Finally, we examine the convergence rate of our algorithm. For simplicity, we focus the investigation only on the bid shading parameters and hold all other parameters constant. The left panel in Figure 7 shows the fitness performance of GA over 100 populations. We can see that the fitness function improves rapidly over the first 20 iterations and slows down subsequently. In the right panel of Figure 7, we fit a linear regression line to the log-transformed fitness (Y) and log-transformed number of iterations (X). This model fits the data very well ($R^2 = 0.88$) and both the intercept and slope are significant with values of -1.79 and -0.35, respectively. Thus, if we let $F$ denote the fitness and $n$ the number of iterations, then the relationship between fitness and number of iterations can be approximated by

$$F = 0.179n^{-0.35}. \quad (1)$$

From the regression, we can see that the mean fitness converges toward 0. Thus, the fitness of the best individual is also convergent to 0, since mean fitness is an upper bound of it. Also, from the regression function we can calculate that from 50 iterations to 100 iterations, the weighted RMSE drops about $22\% = 1 - 100^{-0.35}/50^{-0.35}$. While for an additional step, after 100 iterations, the weighted RMSE drops only $.3\% = 1 - 101^{-0.35}/100^{-0.35}$, which indicates that 100 iterations is a good stop point for the algoritm used in our model.

# 7. CONCLUSIONS AND FUTURE WORK

Understanding the reason why bidders behave a certain way allows invaluable insight into the auction process. In this work, we shed some light on the amount by which bidders shed their bids, i.e. their aversion to taking risks. We do so by combining novel tools from functional data analysis, agent based modeling and genetic algorithm. One obvious benefit of the genetic algorithm is computational simplicity, in this case we save a lot of computation time by not having to fully iterate through the whole search space.

We find that bidders tend to bid rather conservatively, with a bid shading distribution that is right-skewed. This indicates the auction price might not give sufficient information on bidders' willingness to pay. So, if a bidder places a bid early with bid amount close to the current price, it triggers other bidders to overbid her, since the required overbid price tends to stay in the range of their willingness to pay. So sniping (placing bid in last minute) an auction might be a good strategy for bidders in internet auctions to avoid competition from other bidders.

We hope to continue this research in a number of different ways. First of all since our final result and model should be generalizable, we hope to explore whether we can replicate these results in other datasets. Moreover, one limitation of our current work is that bidders do not consider multiple auctions at the same time; we hope to relax this limitation in the future and explore multiple auctions. We also could explore how well our weighted fitness function works compared to a more general multiobjective fitness function. In general, we feel that the technique presented here allows researchers to infer unobservable parameters of a complex system by combining GAs with ABM, while at the same time taking advantage of the underlying temporal nature of the data using functional data analysis.

# 8. REFERENCES


[1] P. Bajari and A. Hortacsu. The winner's curse, reserve prices, and endogenous entry: empirical insights from ebay auctions. *RAND Journal of Economics*, pages 329–355, Summer 2003.

[2] P. Bajari and A. Hortacsu. Economic insights from internet auctions. *Journal of Economic Literature*, XLII:457–486, 06 2004.

[3] S. C. Bankes. Agent-based modeling: A revolution? *PNAS*, 99:7199–7200, 05 2002. suppl. 3.

[4] R. Bapna, P. Goes, A. Gupta, and Y. Jin. User heterogeneity and its impact on electronic auction market design: An empirical exploration. *MIS Quarterly*, pages 21–43, 03 2004.

[5] R. Bapna, W. Jank, and G. Shmueli. Consumer surplus in online auctions. *Information Systems Research*, 19(4):1–17, 2008.

[6] E. T. Bradlow and Y.-H. Park. Bayesian estimation of bid sequences in internet auctions using a generalized record-breaking model. *Marketing Science*, pages 218–229, 03-04 2007.

[7] B. Calvez and G. Hutzler. Automatic tuning of agent-based models using genetic algorithms. *In MABS 2005: Proceedings of the 6th International Workshop on Multi-Agent-Based Simulation*, 2005.

[8] R. Easley and R. Tenorio. Jump bidding strategies in internet auctions. *Management Science*, pages 1407–1419, 10 2004.

[9] N. Foutz and W. Jank. Pre-release demand forecasting for motion pictures using functional shape analysis of virtual stock markets. *Marketing Science*, 29 (3):568–579, 2010.

[10] D. E. Goldberg. *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison-Wesley Professional, Reading, Massachusetts, Jan. 1989.

[11] E. Haruvy, P. T. L. P. Leszczyc, O. Carare, J. C. Cox, E. A. Greenleaf, W. Jank, S. Jap, Y.-H. Park, and M. H. Rothkopf. Competition between auctions. *Marketing Letters*, 19(3-4):431–448, 2008.

[12] J. Holland. *Adaptation in Natural and Artificial Systems*. University of Michigan Press, Ann Arbor, MI, 1975.

[13] V. Hyde, W. Jank, and G. Shmueli. Investigating concurrency in online auctions through visualization. *The American Statistician*, 60(3):241–250, 2006.

[14] V. Krishna. *Auction Theory*. Academic Press, 03 2002.

[15] D. Midgley, R. Marks, and D. Kunchamwar. Building and assurance of agent-based models: An example and challenge to the field. *Journal of Business Research*, 60:884–893, 08 2007.

[16] J. Miller. Active nonlinear tests (ants) of complex simulation models. *Management Science*, pages 820–830, 1998.

[17] S. E. Page. Agent based models. ABS Notes, 09 2005.

[18] Y.-H. Park and E. T. Bradlow. An integrated model for bidding behavior in internet auctions: Whether, who, when, and how much. *American Marketing Association*, pages 470–482, 11 2005.

[19] J. Ramsay and B. W. Silverman. *Functional Data Analysis*. Springer, 2010.

[20] S. Reynolds and J. Wooders. Auctions with a buy price. *Economic Theory*, pages 9–39, 01 2009.

[21] R. P. Russo, G. Shmueli, and W. Jank. The barrista: A model for bid arrivals in online auctions. *The Annals of Applied Statistics*, 1(2):412–441, 2007.

[22] G. Shmueli and W. Jank. Visualizing online auctions. *Journal of Computational and Graphical Statistics*, 14(2):299–319, 2005.

[23] F. Stonedahl, W. Rand, and U. Wilensky. Evolving viral marketing strategies. *Proceedings of the 12th Annual Conference on Genetic and Evolutionary Computation (GECCO '10)*, 07 2010.

[24] S. Wang, W. Jank, and G. Shmueli. Explaining and forecasting online auction prices and their dynamics using functional data analysis. *Journal of Business and Economic Statistics*, 26(2):144–160, 2008.

[25] X. Wang and Y. Hu. The effect of experience on internet auction bidding dynamics. *Market Letter*, pages 245–261, 2009.

[26] R. Weinberg. Computer simulation of a living cell. *PhD thesis, University of Michigan*, 1970.

[27] S. Zhang, W. Jank, and G. Shmueli. Real-time forecasting of online auctions via functional k-nearest neighbors. *International Journal of Forecasting*, 26(4):666–683, 2010.