An Integrated Model for Whether, Who, When and How Much in Internet Auctions

Young-Hoon Park and Eric T. Bradlow*

February 2005

An Integrated Model for Whether, Who, When
and How Much in Internet Auctions

Abstract

We develop a general parametric modeling framework for bidding behavior in Internet auctions. Toward this end, we incorporate and model simultaneously four key components of the bidding process under our integrated framework: Whether an auction will have a bid at all, (if so) who has bid, when they have bid, and how much they have bid over the entire sequence of bids in an auction. This integrated framework is based on a single latent time-varying construct of consumer willingness to bid (WTB), which bidders have and update for a particular auction item over the course of the auction duration. Our modeling approach is also based on a simple yet very general bidding premise: The observed bidder's latent WTB at a specific bid is greater than the outstanding bid; yet, WTB is unconstrained for all other potential bidders. In this manner, we impose no structural assumption on bidder rationality or equilibrium behavior; instead, deriving our model using a probabilistic modeling paradigm. We describe in detail the advantages that our reduced-form approach allows us, and the limitations such an approach also entails.

Using a database of notebook auctions from one of the largest Internet auction sites in Korea, we demonstrate that this general (yet parsimonious) model captures the key behavioral aspects of bidding behavior. Furthermore, substantively, through a data-windowing procedure to assess the set of potential bidders for a given auctioned item, we provide a valuable tool for managers at auction sites to conduct their customer relationship management efforts which require them to evaluate the "goodness" (whether) of the listed auction items and the "goodness" (who, when, and how much to bid) of the potential bidders in their Internet auctions.

KEYWORDS: Bayesian Inference, Bidding Behavior, Probability Model

1   Introduction

The recent proliferation of auction sites on the Internet and the growing importance of online auctions as exchange mechanisms have attracted the attention of academic researchers who have studied such issues as the effect of auction formats (Lucking-Reiley 1999), the extent of the winner's curse (Bajari and Hortaçsu 2003), the last-minute bidding phenomenon (Roth and Ockenfels 2002) and the value of seller reputation (Melnik and Alm 2002). However, our understanding is still rather limited in our ability to explain bidding behavior over the entire sequence of bids, as opposed to simply summary outcomes (e.g., final auction prices), in an auction (e.g., Ariely and Simonson 2003, Chakravarti *et al.* 2002).

For instance, while an auction is in progress, participants in the auction will be influenced by various types of value signals (e.g., minimum bid, seller reputation, other participants' bids, number of bids submitted up to that point, etc.) which can, in turn, impact their decision dynamics for the auctioned item (Ariely and Simonson 2003). Recently, the standard assumption of bidder rationality in online bidding behavior has been questioned in a variety of empirical settings. In particular, Dholakia and Soltysinski (2001) reported evidence of herd behavior bias, and Kamins, Drèze and Folkes (2004) found an effect of minimum bid on the final auction price. Furthermore, some fundamental aspects of consumer decision making such as preference construction (Tversky and Kahneman 1986), choice context (Bettman, Luce and Payne 1998) and learning and expertise (Alba and Hutchinson 1987) are likely to apply to auctions just as they do in regular purchase decisions. In this research, therefore, we develop a dynamic parametric stochastic model of bidding behavior, which considerably differs from prior research that has ignored bidding dynamics.

To accomplish this, we derive a probability model for auction behavior by positing the existence of a latent construct which we denote consumer willingness to bid (WTB). This latent construct

is a time-varying stochastic valuation that an individual bidder has and updates for a particular auction item over the course of the auction duration. Based on WTB, we study key determinants of dynamic bidding behavior by developing an integrated (yet parsimonious) model which explicitly captures the critical features of bidding behavior established in the existing literature. Toward this end, we incorporate four key components in building a model of bidding behavior over the entire sequence of bids — whether an auction will have a bid at all, (if so) who has bid, when they have bid and how much they have bid — in one integrated framework. In addition, we include controls for the presence of other bidders in WTB estimation, including both observed bidders in an auction and allowing for the presence of "latent" bidders who might be following the bidding process but do not submit any bid.

We remark that the importance of the assumptions of bidder rationality and equilibrium behavior is evident in the emerging literature of estimating consumer valuations from auction data. Almost all existing empirical papers (e.g., Donald and Paarsch 1996, Guerre, Perrigne and Vuong 2000, Laffont, Ossard and Vuong 1995) rely on theoretical assumptions of auction behavior and equilibria. However, we impose no structural restriction on bidder behavior because there appears to be no theoretical and/or empirical work that addresses in a fully structural equilibrium model bidding dynamics over the entire sequence of bids in an auction. Instead, our approach is based on a very general bidding premise: The observed bidder's WTB at a specific bid is greater than the outstanding bid; yet WTB is unconstrained for all other potential bidders. Therefore, our model is not structural in that it does not result from first-principles of utility or profit maximization for bidders and sellers, respectively. Rather, we develop a model for a dynamic bidding process in which we capture WTB in a reduced-form outcome using a parametric model.

With the lack of equilibrium-generating process directly incorporated into our model, there are of course both advantages it affords as well as limitations it presents. On the one hand, as

2

we demonstrate, our model is parsimonious, fits the data extremely well, and does not require us to model a very complex bidding process that could possibly have to include both underlying dynamics and strategic use of available information among bidders and endogenous choices on the part of the seller. For the purposes of description and prediction, our main goals here, a parametric probabilistic approach is entirely adequate. In fact, it may be less prone to misspecify models as less assumptions need to be made about the underlying process as recent research (Bajari and Hortaçsu 2004) has shown that differences between many structural models of behavior can not be identified from commonly observed data sets (first-price auctions as described here) and in fact "..., we can always reverse engineer a structural model that is consistent with the data."

On the other hand, a plethora of articles (e.g., Bajari and Hortaçsu 2003, Donald and Paarsch 1996, Guerre, Perrigne and Vuong 2000, Laffont, Ossard and Vuong 1995) have clearly demonstrated the existence of strategic behavior among bidders and the ability to perform policy experiments through a structural equilibrium process. Description and prediction may be possible without incorporating strategic behavior (as our research here demonstrates), but a fundamental understanding of the "why" (its source) may still be left unknown. We acknowledge that without such modeling there is a limit to the inferences derived and our model is no exception to this. It is for this reason, that in the concluding section, we make a call for further empirical research that will test the bounds of probabilistic models to fit data generated through equilibrium process. Our belief is that this model is one platform that can address this call.

Our model is built to be an integrated model for the series of whether, who, when and how much for the entire series of bids on an auction item, instead of independently modeling summary features of the auction such as the number of participants, the number of bids submitted, or the ultimate amount of the winning bid. These summary measures are a natural consequence of our model (and an output thereof) but are not modeled directly. We instead use them to validate the

fit of our model. Thus, we present a probability model for dynamic bidding behavior that can be used for exploration, summary and forecasting of Internet auctions. To the best of our knowledge, this is the first attempt to formally model behavioral aspects of bidding behavior for the entire sequence of bids in Internet auctions.

Beyond the specification of the model, per se, our research provides a valuable tool for managers at auction sites to conduct their customer relationship management efforts which require them to evaluate the "goodness" (whether) of the listed auction items and the "goodness" (who, when and how much to bid) of the potential bidders in their Internet auctions. Since our model can infer whether, who, when and how much to bid at each time over the course of the auction duration among the potential bidders, our approach can be considered for developing contact (communication) strategies at auction sites. This is a fundamental marketing problem faced by auction sites; that is, advertising their auction items and recruiting bidders for their auctions (Wang and Montgomery 2003).

Finally, in addressing these important substantive issues, we use a comprehensive database of notebook auctions obtained from an auction company. The database contains auctions of no bids as well as auctions with bids, and information regarding the complete history of bids, features of auction design, bidder and seller characteristics and product specifications of auction items. Hence, this database is indeed a panel data set which allows us to incorporate a set of noble variables (e.g., individual bidder's past bidding experience) under our model. In light of the potential importance of cross-auction behavior, sparse information that exists across auctions and our desire to incorporate heterogeneity, we adopt a Bayesian approach and estimate the models using Markov chain Monte Carlo (MCMC) methods (Gelfand and Smith 1990). Therefore, this paper makes substantive and methodological contributions to both the existing auction and marketing literature.

The remainder of the paper is organized as follows. Section 2 gives an overview of the data

4

and describes summary statistics as a way to begin to understand the nature of bidding behavior in our Internet auction data. In section 3, we present the model specification and discuss our computational approach using a Bayesian framework. In section 4, we apply the proposed model to notebook auctions and validate the proposed model based on the key summary measures described in section 2. We discuss other managerial implications of this research and conclude with directions for future research in section 5.

2   Data Overview

In this section, we describe the data and propose several sets of descriptive statistics derived from the database. This analysis will inform us about the degree of flexibility needed in our modeling approach to investigate the key determinants of the bidding process in online auctions. In addition, it will help demonstrate whether the proposed model properly captures the key behavioral aspects of online bidding behavior, thus acting as measures of model fit.

2.1   Data Description

The data are for notebook computer auctions from one of the largest Internet auction sites in Korea for the time period of July 2001 to October 2001. The auction mechanism used on this site is an ascending first-price auction or English auction in which bids are ascending and the highest bidder wins and pays the amount she bids. The database contains auctions of no bids as well as auctions with bids, and information regarding the complete history of bids, features of auction design, bidder and seller characteristics and product specifications of the auction items. Thus, this database is indeed a panel data set that follows individual bidders over time and thus provides the richness of having multiple bids on each individual in the database (if they exist).

We focus on notebook auctions for the sale of a single item.[1] The total number of notebook

---

[1]Modeling bidding behavior including auctions with multiple items is possible under our framework; yet, we believed that the bidding dynamics would be fundamentally different and hence is an area for future research.

auctions considered here is 2618 items, in which 296 auction items have no bids, and the total number of bids across all the auctions is 21952. On average, there are about 5.8 unique bidders and 8.4 bids per auction. We note that all bids are in Korean currency (won), where 1200 won corresponds approximately to $1.

At the auction site considered in this research, there are five feature variables for (seller) auction design: placement (yes or no) of product images or pictures on the listing page, minimum bid amount (also called a "public" reserve price), "buy-it-now" (BIN) option[2] and its price (a fixed price which allows bidders to prematurely end an auction by exercising an option to buy the item) and auction duration.

Other variables in the dataset include the following. Sellers on this auction site are rated by winning bidders. The rating is in the form of a positive, negative or neutral response after each auction is completed. While this information changes with transaction in which the winning bidder rates the seller, the database we obtained only maintains the cumulative records of these reputation variables at the start of the data period. We also have information on bidder characteristics which include demographics and behavioral characteristics such as the number of previous visits to the site and the number of page views across all product categories. Similar to the seller reputation variables, the auction site only keeps the cumulative information on these variables at the start of the data period. Therefore, these variables are also static in our database, again as of July 2001.

Finally, the data includes information on product features for each auctioned item as listed on the site, which contains the following variables: CPU type (Pentium or Celeron), CPU speed, memory, hard disk, screen size, brand name, and the number of months that the auction item has been used by the seller (0 for a brand-new item). There are three American, three Japanese, and six Korean brands, which account for about 29%, 14%, 52% of the 2618 items, respectively. All of

---

[2]The BIN option at the auction site considered here remains in effect throughout the auction as long as it is not exercised.

the rest of the brands, which we aggregated and grouped into a category "others", accounts for 5% of the items. Table 1 reports summary statistics of each of these variables described in this section. These variables, along with a set of noble time-varying variables (e.g., individual bidder's past bidding experience) to be described, will serve as covariates toward explaining bidding behavior via an individual bidders's dynamic and latent WTB.

Insert Table 1 about here

2.2 Empirical Findings

An interesting feature of the database is that there is wide variation in the number of (observed) bidders and the number of bids per auction. About half of the auctions have three bidders or less. In sharp contrast, about 30% of all the auctions have six bidders or more, and about 40% have six bids or more.

We describe bid timing, unitized to $[0, 1]$ for description and explication (but not in the model), accomplished by dividing each bid time by the auction duration of the item. Bidding activity is concentrated at the end of each auction. About 35% of bids are submitted after 97% of the auction duration has passed (i.e., the last two hours of a three day auction). We observe that winning bids tend to come even later. About 75% of final bids are submitted after 97% of the auction duration has passed. This practice of a last-minute bidding phenomenon has attracted a good deal of attention among academic researchers (e.g., Bajari and Hortaçsu 2003, Roth and Ockenfels 2002).

Besides the timing of bids, bid amounts (or bid increments) are of great importance in Internet auctions since the key decision by potential bidders centers on how much (more) to bid to become the highest current bidder or to win the auction outright. We observe large variation in bid amounts ranging from less than a dollar to more than several thousand dollars (mean = \$762, std. = \$507).

A striking observation is on one of the seller design mechanisms, i.e., BIN feature. In particular, about 90% of the auctions (i.e., 2314 out of 2618 auction items) are designed with this option which turns out to play a major role in influencing bidding behavior. About 40% of the auctions (i.e., 954 out of 2314 auctions) are ended with their corresponding BIN prices. Furthermore, more than half of these auctions are ended at the first bid with their corresponding BIN prices.

The descriptive statistics discussed in this section are useful as a first step to understand the nature of bidding behavior in these notebook auctions. For instance, we show that a last-minute bidding phenomenon is prevalent. We also observe wide variation in bid amounts and a large number of auctions ending with BIN option at the very early stage of the auction duration. These behavioral aspects of online bidding are explicitly explored in section 3 and the descriptive analyses presented here provide us an "inkling" that we need a very flexible parametric model.

3   Model Development

We develop an integrated model of bidding behavior that includes four key modules: Whether an auction will have a bid at all, (if so) who has bid, when they have bid and how much they have bid over the entire sequence of bids in the auction. In constructing the proposed model, we incorporate controls for competition among potential bidders including those participants who we do not directly observe in the auction. As a step to this end, we define an individual bidder's latent WTB at a specific bid in an auction. This is our fundamental construct to explain the bidding process that links all four modules together. We conclude the section with a description of the computational approach for the integrated model.

3.1   Rate for Bid Speed

A consumer's latent WTB is the kernel of our model for online bidding behavior. This latent construct is a time-varying stochastic valuation that an individual bidder has and updates for a

particular auction item over the course of the auction duration. We assume that WTB determines the rate of bid speed which in turn directly governs the whether, who and when models in our integrated framework, and as described the how much model in a more indirect way. More formally, the bid speed, s_{ij}^k, of person j who is a potential bidder at bid k in auction i, is assumed to follow an exponential distribution and is modeled as follows:

$$s_{ij}^k \sim \lambda_{ij}^k \cdot \exp\left\{ -\lambda_{ij}^k \cdot (t_{ij}^k - t_i^{k-1}) \right\} , \tag{1}$$

where

$$\log(\lambda_{ij}^k) = \beta_0 + \beta_1(w_{ij}^k - b_i^{k-1}) + \varsigma_{ij}^k , \quad \varsigma_{ij}^k \sim N(0, \sigma_\varsigma^2) , \tag{2}$$

t_{ij}^k is a random variable denoting person j's timing of bid k in auction i, and w_{ij}^k is person j's WTB at bid k in auction i. Let t_i^{k-1} and b_i^{k-1} denote the observed bid time and amount of the $(k-1)$-st bid in auction i, respectively. Further, it is assumed that $\log(\lambda_{ij}^k)$, the rate that governs one's bid speed, follows a normal distribution with a mean of $\beta_0 + \beta_1(w_{ij}^k - b_i^{k-1})$ and a variance of σ_ς^2. Thus, $(w_{ij}^k - b_i^{k-1})$, person j's surplus at bid k in auction i, is used to compute the deterministic component of person j's rate for bid speed at bid k in auction i. Finally, ς_{ij}^k is a random component of person j's rate for bid speed at bid k in auction i, varying from bid to bid, possibly as a result of unobserved variables.

The exponential distribution used to model the rate of bid speed, based on WTB, merits explicit mention because it is both behaviorally plausible and mathematically desirable. Imagine a bidder whose bid has been outbid, i.e., she is not the highest bidder any more. Thus, her previous bids and effort are now gone and things start conditionally afresh. Her decision now centers on whether or not to continue bidding to be the highest bidder or to win the auction (if so, when and how much to bid). Behaviorally, therefore, the well-known memoryless property of the exponential distribution may hold, conditionally on history, in the auction context.

It is important to note, however, that our choice of the exponential distribution for bid speed comes with a balance of positive and negative aspects. In particular, the positive aspects are that it leads to closed-form solutions to the whether, who and when parts of the model, a feat that should not be minimized. On the other hand, a more general timing model like the Weibull model would be more flexible and may lead to improved model fit; albeit, at the cost of closed-form tractability.

3.2 The Proposed Model

In this section, we first derive the whether model which indicates whether or not any bids are realized over the course of an auction duration. Conditional on the realization of (at least) a bid, we then derive the probabilistic closed-form expressions for the who and when models which are based on the truncated exponential distribution. Finally, in order to consider how much bidders have bid, we employ a parametric family of distributions to capture the observed bid amounts.

In constructing the whether, who and when probability models, the following sequence of stages occurs: We infer WTB at a specific bid, determine which bidders have WTB greater than the outstanding bid, derive the bid rates from the latent WTB construct as given by equations (1) and (2), and calculate the probabilities of whether, who and when. In doing so, we note that if w_{ij}^k is greater than b_i^{k-1} (i.e., outstanding bid or minimum bid for the first bid), person j is in the bidding competition for bid k. Otherwise she is out of the race for bid k in auction i. Therefore, we have a time-varying set of latent bidders, denoted as J_i^k, who will compete for bid k in auction i. Note that $j = 1, \ldots, J_i^k$ will be in the race for bid k and $j = J_i^k + 1, \ldots, J_i$ will be out of the race at bid k in auction i. As a result, the number of latent competing bidders (J_i^k) is a random variable in our system and varies over a series of bids in the auction. Hence, our model can be considered as a way to endogenously incorporate entry behavior in the auction. We note that the number of potential bidders (J_i) is an important determinant of our model and is one we explore extensively

in section 3.3 via a data-windowing procedure. For now, we simply let J_i denote the number of potential bidders who will be in and out of the race over a series of bids in auction i.

3.2.1 The Whether Model

To derive the probability of whether an auction will have a bid at all, we note that one critical aspect of our approach is that the minimum of a set of exponentials is exponentially distributed with rate equal to the sum of the rates (see the Appendix for details). Thus, the whether model is the probability that no bid is realized during the course of the auction duration (T_i) based on the sum of $\lambda_{ij}^{k=1}$ across $J_i^{k=1}$ bidders at the first potential bid. That is, the probability that an auction i will have no bids (or the first is not realized) over its duration T_i is given by:

$$\Pr[\min(t_{ij}^1) > T_i] = \exp\left\{ -\sum_{j=1}^{J_i^1} \lambda_{ij}^1 \cdot T_i \right\}. \tag{3}$$

To get the likelihood function of whether an auction will have a bid at all, equation (3) can be multiplied across the I auctions:

$$\prod_{i=1}^{I} \left\{ \Pr[\min(t_{ij}^1) > T_i] \right\}^{I_i^1}, \tag{4}$$

where I_i^1 is an indicator function in which I_i^1 is 1 if auction i has no bids, and 0 otherwise.

3.2.2 The Who Model

To derive the probability of who has bid at a specific bid on the auction item, we calculate the probability that one exponential random variable is smaller than all the others. That is, person j is the bidder if her exponential bid time drawn from equations (1) and (2) is the smallest among all potential bidders. In other words, at each bid k, there is an exponential race with rates related to WTB, and the bidder with the shortest time wins the race and is the bidder. We note that the observed bidder at bid k, by definition, must be in the race, so her WTB is truncated below by the outstanding bid. For other potential bidders, their WTB could be anywhere, i.e., either less than

11

or greater than the outstanding bid. We thus impose no structural restriction on one's bidding behavior to infer WTB, which we capture in a reduced-form outcome.

For J_i^k latent competing bidders at bid k in auction i, the probability that the k-th truncated exponential bid time by person j is the smallest among a set of bid times by all other bidders is as follows:

$$\Pr[\min(t_{ij}^k) = t_{ij}^k \,|\, t_i^{k-1} < \min(t_{ij}^k) \leq T_i \,, \forall j \in J_i^k] \quad = \quad \frac{\lambda_{ij}^k}{\displaystyle\sum_{\substack{j=1 \\ j \neq j^{k-1}}}^{J_i^k} \lambda_{ij}^k} \,. \tag{5}$$

Equation (5) is based on a set of J_i^k exponentials with rates equal to λ_{ij}^k, $j = 1, \ldots, J_i^k$. We include the detailed derivation of the who model in the Appendix. Since it is not allowed for an individual bidder to bid twice in a row (i.e., outbid herself), it is required to take out the current rate of bid speed by the previous bidder from the denominator in equation (5). The probability of who has bid is proportional to one's bid rate which is in turn a function of her WTB. We remark that while this equation addresses the likelihood of who the bidder is, it does not describe anything related to bid submission time by the bidder.

To get the likelihood function of who has bid, equation (5) can be multiplied across the I auctions, K_i bids, and J_i^k bidders:

$$\prod_{i=1}^{I} \prod_{k=1}^{K_i} \prod_{j=1}^{J_i^k} \left\{ \Pr[\min(t_{ij}^k) = t_{ij}^k \,|\, t_i^{k-1} < \min(t_{ij}^k) \leq T_i \,, \forall j \in J_i^k] \right\}^{I_{ij}^k} , \tag{6}$$

where I_{ij}^k is an indicator function in which I_{ij}^k is 1 if person j bids at bid k in auction i, and 0 otherwise.

3.2.3 The When Model

We derive the probability of when each bid occurs. Since the information of who has bid at bid k in auction i is given in equation (5), we derive the conditional probability of when bid k would be

realized given who has bid. The basic idea is that if we have the J_i^k exponentials as in equations (1) and (2), the bidder is the person with the minimum order statistic of the s_{ij}^k, and the bid submission time is the value of that minimum order statistic. That is, the smallest of the t_{ij}^k is given by:

$$
\begin{aligned}
&\Pr[\min(t_{ij}^k) - t_i^{k-1} = \Delta t_{ij}^k \,|\, t_i^{k-1} < \min(t_{ij}^k) = t_{ij}^k \leq T_i, \forall j \in J_i^k] \\
&= \frac{\displaystyle\sum_{\substack{j=1 \\ j \neq j^{k-1}}}^{J_i^k} \lambda_{ij}^k \times \exp\left\{ -\sum_{\substack{j=1 \\ j \neq j^{k-1}}}^{J_i^k} \lambda_{ij}^k \cdot \Delta t_{ij}^k \right\}}{1 - \exp\left\{ -\displaystyle\sum_{\substack{j=1 \\ j \neq j^{k-1}}}^{J_i^k} \lambda_{ij}^k \cdot (T_i - t_i^{k-1}) \right\}} .
\end{aligned}
\tag{7}
$$

As shown in equation (7), the probability of when bid k occurs does not include λ_{ij}^k, per se, but has the sum of λ_{ij}^k across all the latent competing bidders as its rate. This result, that the minimum of a set of exponentials is exponentially distributed with rate equal to the sum of the rates is a nice property of the exponential distribution. We note that the denominator in equation (7) arises from the truncation that bid k occurs between time t_i^{k-1} and the auction end time T_i. From equations (3), (5), and (7) we can now see the integrated link of WTB, which determines the mean of the bid rates and then simultaneously determines the whether, who and when probabilities.

To get the likelihood function of when the bid is realized, equation (7) can be multiplied across the I auctions, K_i bids and J_i^k bidders:

$$
\prod_{i=1}^{I} \prod_{k=1}^{K_i} \prod_{j=1}^{J_i^k} \left\{ \Pr[\min(t_{ij}^k) - t_i^{k-1} = \Delta t_{ij}^k \,|\, t_i^{k-1} < \min(t_{ij}^k) = t_{ij}^k \leq T_i, \forall j \in J_i^k] \right\}^{I_{ij}^k} .
\tag{8}
$$

3.2.4 The How Much Model

We next describe the distribution of how much is bid at a specific bid in an auction. In order to do so, we first assume a baseline model for latent bid amounts, denoted as d_{ij}^k, of bidder j at bid k in auction i that follows a normal distribution:

$$
d_{ij}^k = w_{ij}^k + \zeta_{ij}^k , \quad \zeta_{ij}^k \sim N(0, \sigma_\zeta^2) ,
\tag{9}
$$

13

with a mean of w_{ij}^k and a variance of σ_ζ^2. Thus, w_{ij}^k (our latent WTB) is the mean component of bidder j's valuation at bid k in auction i and ζ_{ij}^k is a random component.[3] We note, therefore, that the how much model is explicitly linked to the whether, who and the when models through WTB, which consequently leads to a very parsimonious integrated model for online bidding behavior.

In the model of how much to bid, however, we also explicitly incorporate the probability that one bids the BIN price which equation (9) does not. In particular, on an auctioned item designed with BIN, if one's latent bid amount d_{ij}^k is greater than the BIN price, we include a stochastic component to allow for the possibility that she chooses not to exercise BIN.

To formally develop the full model of how much to bid (including the BIN decision), we need to extend equation (9) and thus introduce the following notation. Let b_{ij}^k denote bidder j's bid amount at bid k in auction i and B_i be the BIN price in auction i. We separately consider the two types of auctions: auctions designed with and without BIN. Among the first type of auctions, we further categorize bids based on d_{ij}^k compared to B_i, i.e., the upper bound of the bid amount in auction i. When d_{ij}^k is greater than B_i, bidder j can finish the auction by bidding B_i (i.e., end the auction with BIN price) or bid less than B_i. In order to represent these two cases under the how much model, we employ a logit framework. In particular, when d_{ij}^k is greater than B_i, we denote p_{ij}^k as the probability of bidding BIN price at bid k in auction i:

$$\Pr[b_{ij}^k = B_i] = p_{ij}^k = \frac{\exp\left\{\gamma_0 + \gamma_1(w_{ij}^k - b_i^{k-1}) + \tau_{ij}^k\right\}}{1 + \exp\left\{\gamma_0 + \gamma_1(w_{ij}^k - b_i^{k-1}) + \tau_{ij}^k\right\}} \, , \tag{10}$$

where τ_{ij}^k follows a $N(0, \sigma_\tau^2)$ and $(w_{ij}^k - b_i^{k-1})$ is the surplus as before. Hence, p_{ij}^k is a conditional probability of bidding BIN price if one's valuation allows one to do so.[4]

---

[3]Note that the random component ζ_{ij}^k in the how much model differs from the random component ε_{ij}^k in the rate of bid speed governing the whether, who and when models because the unobserved variables for each model may be different. We assume independence between these error components, yet its impact is an area for future research.

[4]We examined other specifications for incorporating BIN. For example, the bidder adjusts her latent bid amount with p_{ij}^k defined in equation (10) where the mean is now $p_{ij}^k \cdot d_{ij}^k$. This model, whose results are available from the authors, fits worse than the proposed model.

To complete the construction of the how much model, we need to consider two additional cases. First, when d_{ij}^k is less than B_i in auction i designed with BIN, bidder j can choose a bid amount less than B_i. Among auctions not designed with BIN, second, an individual bidder can decide any level of bid amount based on her WTB.

According to our discussion of how much to bid as described above, there are four different probabilistic expressions to capture how much one bids at each round on an auction which is given by:

$$\Pr[b_{ij}^k = b_i^k] = \begin{cases} \text{I.} & \left\{ \Pr[d_{ij}^k \geq B_i] \right\} \times \left\{ \quad p_{ij}^k \right\} \\ \text{II.} & \left\{ \Pr[d_{ij}^k \geq B_i] \right\} \times \left\{ 1 - p_{ij}^k \right\} \times \left\{ \phi(b_i^k, b_i^{k-1}, B_i \mid d_{ij}^k, \sigma_d^2) \right\} \\ \text{III.} & \left\{ \Pr[d_{ij}^k < B_i] \right\} \qquad\qquad \times \left\{ \phi(b_i^k, b_i^{k-1}, B_i \mid d_{ij}^k, \sigma_d^2) \right\} \\ \text{IV.} & \qquad\qquad\qquad\qquad\qquad \left\{ \phi(b_i^k, b_i^{k-1} \mid d_{ij}^k, \sigma_d^2) \right\} \end{cases} \quad (11)$$

where $\phi(z_1, \underline{z}, \overline{z} \mid d_{ij}^k, \sigma_d^2)$ is the normal density with a mean of d_{ij}^k and a variance of σ_d^2, truncated at $\underline{z}$ from below and $\overline{z}$ from above, and $\phi(z_2, \underline{z} \mid d_{ij}^k, \sigma_d^2)$ is the truncated normal density with a mean of d_{ij}^k and a variance of σ_d^2 truncated below by $\underline{z}$. The first three expressions in equation (11) consider three possible cases among auctions designed with BIN: I and II incorporate p_{ij}^k when d_{ij}^k is greater than B_i, and III represents the bid when d_{ij}^k is less than B_i. Finally, IV in equation (11) deals with the bid amount for auctions not designed with BIN in which the bidder can bid any amount above the outstanding bid.

To get the likelihood function of how much to bid, the set of four different probability expressions in equation (11) can be multiplied across the I auctions, K_i bids and J_i^k bidders:

$$\prod_{i=1}^{I} \prod_{k=1}^{K_i} \prod_{j=1}^{J_i^k} \left[\begin{array}{cc} \left\{ \text{I. } \Pr[b_{ij}^k = b_i^k] \right\}^{I_{ij}^k[\text{I}]} & \times \quad \left\{ \text{II. } \Pr[b_{ij}^k = b_i^k] \right\}^{I_{ij}^k[\text{II}]} \\ \times \quad \left\{ \text{III. } \Pr[b_{ij}^k = b_i^k] \right\}^{I_{ij}^k[\text{III}]} & \times \quad \left\{ \text{IV. } \Pr[b_{ij}^k = b_i^k] \right\}^{I_{ij}^k[\text{IV}]} \end{array} \right], \quad (12)$$

where $I_{ij}^k[\cdot]$ is an indicator function in which $I_{ij}^k[\cdot]$ is 1 if $[\cdot]$ happens at bid k by bidder j in auction i, and 0 otherwise.

In this research, we include the bidding behavior concerning auctions designed with BIN, despite the complications it leads to in equation (11) because bidders in this model have greater flexibility and auctions with BIN are common as described in section 2. However, we remark that we incorporate only one form of bidding behavior in BIN auctions through p_{ij}^k among many possible other types that could drive these decisions. For example, List and Lucking-Reiley (2002) reported in a field experiment that cognitive costs influence subjects' bidding behavior, which we leave for future research, and only consider WTB as a driver as described here.

Taken together, the joint likelihood function of the proposed integrated model is given by:

$$L = \text{Equation}(4) \times \left\{ \text{Equation}(6) \cdot \text{Equation}(8) \cdot \text{Equation}(12) \right\}. \tag{13}$$

3.3 Latent Bidders

As described previously, latent bidders play a prominent role in our model specification in two specific ways. First, the number of potential bidders (J_i) determines the maximal set of bidders for which we sum over to compute the total competition set at bid k in auction i. Second, as given in section 3.2, only those bidders J_i^k (out of the potential set J_i) who have positive surplus ($w_{ij}^k > b_i^{k-1}$) are in the race. We describe here in detail both how J_i is selected (including sensitivity analyses to our assumptions) and moreover the assumptions behind the distribution of w_{ij}^k we utilize, and distribution of the set of bidders for which surplus is positive.

The more direct issue that we first address is that of the distribution for w_{ij}^k. Simply put, we desire a set of assumptions that would not be at all restrictive, allowing any potential bidder in the set of J_i to be in the race, yet would be consistent with a reduced-form utility model. To this end, we assumed only that the observed bidder at bid k has latent WTB greater than the outstanding bid b_i^{k-1}, yet for all other potential bidders their WTB could be anywhere, i.e., either less than or greater than the outstanding bid. We note, for example, a restrictive assumption would be that

the latent bids at a specific bid are strictly below the outstanding bid. But it is possible that latent bidders have valuations above the outstanding bid and even above the observed bid, but wait on the sidelines of the bidding process because they are waiting for an opportune moment to enter. In fact, there are a variety of reasons for this waiting behavior, e.g., they would like to enter later in the auction so as to not reveal their preferences or set off a bidding frenzy. While we are agnostic about what these reasons may be, we believe that by choosing the least restrictive assumption we have arrived at a reasonable parametric model.

Second, turning to the choice of J_i, we believe that its choice should satisfy the following criterion: (1) We should include bidders from other auctions who are most likely to want to participate in the focal auction item, but for some reason decide to remain dormant, (2) Our definition needs to be consistent with the design of the auction site considered here, as in many, which lets site visitors/potential bidders sort the auction items by product characteristics (e.g., CPU speed, brand, etc.) and then direct them to auctions containing similar items, (3) Our definition needs to be in line with the strategies auction sites currently use to attract potential bidders (e.g., email a list of 10 auction items to losers in auctions of similar items) and (4) Potential participants are drawn from the set of bidders currently on the site, but at different auctions. We note that what is common across the first three criterion, and is then subsetted using the last one, is the fact that potential bidders may be assumed to arrive from a set of similar items given they are on the site during the duration time of T_i for auction i.

We remark that our choice of the set J_i can also be motivated by consumer behavior theories which suggest that in forming their consideration set, consumers rely on simple heuristics like eliminating alternatives via categorization of items (e.g., Fader and McAlister 1990, Todd and Gigerenzer 2000). Therefore, using the route of scanning concurrent and similar auction items for the definition of latent bidders is consistent both with economic/psychology theory and with the

auction site's navigation tools for bidders.

We also note that a few researchers have recently examined the entry process in auctions with the assumption of symmetric bidders, in which the identity of potential bidders is completely ignored (e.g., Bajari and Hortaçsu 2003, Bradlow and Park 2004). Hence, our approach can be considered as a step to identify potential bidders, in our case, based on exact and partially matching notebook product features.

In order to include latent bidders who are mostly likely to participate in the focal auction item, we utilize a distance metric (Hoch, Bradlow and Wansink 1999) that allows us to compute the similarity of auction items, and therefore the set of potential bidders to include in the set J_i. The way in which our distance metric is computed is as follows. We first turn to any auction that is open concurrently with the focal auction item. For categorical descriptors of notebooks (CPU type, brand names, and categorized CPU speed as listed at the auction site), we next look at the set of items that are an exact match to the given focal item (i.e., lexicographic selectors). For those variables which are continuous in nature (i.e., MEMORY, HDISK, SCREEN and MONTHS, as given in Table 1), we utilize a windowing procedure that allows for auctions that vary by as much as $\pm 10\%$ from the focal item on these features. Extensive empirical testing with smaller ($0\% =$ exact match) and larger windows ($\pm 20\%$) are available upon request however, our findings were quite robust to this specification.

The $\pm 10\%$ windowing procedure lead to on average 9.9 unique potential bidders per auction. That said, alternative definitions that are theoretically sound also exist, e.g., defining latent bidders by looking at budget constraints by focusing on bidders in auctions where the final bid price was less than and closest to the final bid price of the focal item. We believe that this is an exciting area for future research.

From a practical standpoint, furthermore, our approach to assessing the set of potential bidders

can be considered as a useful step for an auction site in developing a better communication (contact) strategy because our model can infer whether, who, when and how much to bid over the entire sequence of bids among the potential bidders. Thus, it provides a way to evaluate the "goodness" (whether) of the listed auctions and the "goodness" (who, when and how much to bid) of the potential bidders on Internet auctions, which is a fundamental marketing problem faced by auction sites in advertising auction items and recruiting bidders for their auctions (Wang and Montgomery 2003). Thus, combining the potential list, with the ability to predict across that list can be a powerful tool.

3.4 The WTB Function

At the heart of our integrated model of bidding behavior lies WTB, w_{ij}^k of person j at bid k in auction i. The WTB governs the four decisions (whether, who, when and how much) where each decision is closely linked to the others based on this behavioral construct that leads to an integrated model. In particular, we assume that each bidder instantaneously decides (renews) w_{ij}^k right after the $(k-1)$-st bid is realized and that w_{ij}^k remains constant until the k-th bid happens.[5]

In order to infer w_{ij}^k, we employ six sets of explanatory variables: (1) features of auction design, (2) seller reputation, (3) product specification of an auction item, (4) bidder characteristics, (5) bid-specific characteristics and (6) an individual bidder's past bidding experience. That is, w_{ij}^k is formally modeled as follows:

$$w_{ij}^k = \alpha_0 + \phi_i + \varphi_j + \alpha_1' \text{AD}_i + \alpha_2' \text{SR}_i + \alpha_3' \text{PS}_i + \alpha_4' \text{BC}_j + \alpha_5' \text{RC}_i^k + \alpha_6' \text{EXP}_{ij}^k + \varepsilon_{ij}^k \, , \qquad (14)$$

where AD_i is a vector of variables for design mechanism in auction i, SR_i is a vector of variables for seller reputation in auction i, PS_i is a vector of variables for product specification (including

---

[5]Prior research has assumed that all the potential bidders are present at the same time and monitoring the auction throughout the entire time while the auction is open, and are able to revise their bids immediately once a certain bid is placed (e.g., Reynolds and Wooders 2003 and references cited therein).

dummy variables for brand names) in auction i, and BC_j is a vector of variables for person j's characteristics. These auction-specific (with subscript i) and bidder-specific (with subscript j) variables are provided in Table 1. The random variables in the WTB function are specified as follows: (1) ϕ_i represents unobserved auction-specific characteristics affecting WTB that are iid across auction items, $\phi_i \sim N(0, \sigma_\phi^2)$, (2) φ_j represents bidder-specific characteristics affecting WTB that are assumed iid across bidders, $\varphi_j \sim N(0, \sigma_\varphi^2)$, and (3) ε_{ij}^k represents the unobservable variation affecting WTB that is iid across auction items, bids and bidders, $\varepsilon_{ij}^k \sim N(0, \sigma_\varepsilon^2)$.

We also include two sets of time-varying variables (with subscript k), RC_i^k and EXP_{ij}^k, in equation (14) because our model is based on the notion that bidder valuations are evolving over a series of bids in an auction. These variables are utilized in order to capture the dynamics of bidding behavior and are described as below:

- RC_i^k is a vector of variables for the k-th bid-specific characteristics in auction i: (1) remaining time to the end of the auction, (2) number of bids submitted before the k-th bid, i.e., $(k-1)$, (3) bid rate operationalized as $(k-1)$ divided by the total elapsed time, and (4) rate of bid increments operationalized as incremental bid amount at the previous round divided by its elapsed time. These bid-specific variables are abbreviated as REMAIN, NUMBID, BIDRATE, and AMTRATE, respectively.[6]

- EXP_{ij}^k is a vector of variables for person j's past bidding experience at bid k in auction i: (1) person j's status (win or loss) on the most recent auction she participated in, (2) total number of auction wins, (3) total number of auction losses, (4) the amount lost by on the most recent auction she participated in (0 for the winner). These variables are abbreviated

---

[6]BIDRATE and AMTRATE at the first bid are not included in the model estimation because they do not exist until the second bid. We also note that a number of variables in the model had moderate correlations (above .5). We assessed whether multi-collinearity was an issue, which was not, by rerunning our model deleting one variable from each collinear pair.

as LWIN, TWIN, TLOSS, AMTLOST, respectively.

We note that RC_i^k describes the impact of other participants' bids on the potential bidders, in terms of the timing of past bids and incremental bid amounts, so that we may understand the time-varying contextual effects of online auctions. That is, does time remaining to end make items appear more attractive? Do large jumps on the most recent bid have an effect on bidder valuations? REMAIN and AMTRATE, respectively, are designed to inform about these dynamics in bidding behavior.

Furthermore, as noted previously, our database is a panel data set which allows us to include an individual bidder's prior bidding experience, i.e., EXP_{ij}^k in equation (14) . We create four variables to capture the impact of experience in terms of recency (LWIN), frequency (TWIN and TLOSS) and monetary value (AMTLOST), a so-called RFM framework. Note that we intentionally make the RFM link in the auction context, as it provides us a rigorous way as how to summarize past behavior of an individual bidder. While empirical research is very limited in studying the role of experience in online auctions (e.g., Wilcox 2000), we take a step further to consider it to uncover whether consumer experience (in terms of recency, frequency and monetary value) drives the bidding process towards certain outcomes.

3.5   Computational Approach

The main advantages of the Bayesian paradigm, utilized here, are to allow for sharing of information across auctions for which there is sparse information and to provide small sample exact p-values not based on asymptotic approximations. Information sharing represents a significant issue for many of the observed auctions due to the large fraction of auctions with few bids as described in section 2. There were a number of possible ways in which we could incorporate shrinkage. We note that all parameters, $(\alpha_0, \alpha_1', \ldots, \alpha_6')$ which govern WTB, $(\beta_0$ and $\beta_1)$ which govern the rate for bid

speed, and (γ_0 and γ_1) which govern the probability of exercising BIN, are given slightly informative but vague priors, $N(0, \text{precision} = .00001)$, to ensure proper posteriors but also to allow the data to primarily govern the inferences. Sensitivity analysis by varying these precisions indicated no impact due to their specific choice due to the large sample sizes that are being used here. All variance components (σ_ζ^2, σ_ϕ^2, σ_φ^2, σ_ε^2, σ_ζ^2, and σ_d^2) are given slightly informative inverse-gamma priors (Gelman *et al.* 1995), with shape and scale set at .01 to ensure proper posteriors.

Inferences under the proposed model were obtained using the freely available software Bayesian Inference Using Gibbs Sampling, WinBUGS, which runs a Markov Chain Monte Carlo (MCMC) sampler to obtain samples from the posterior distribution. Results reported are from the output of three independent chains started from hyper-dispersed starting values, with a burn-in period of 20000 iterations and utilizing the 15000 draws (5000 per chain) thereafter. Convergence was diagnosed both graphically and using the *F*-statistic diagnostic of Gelman and Rubin (1992). We note that as this model was fit using WinBUGS, freely available software, this facilitates with ease the replication (or not) of our findings and its use by practitioners who utilize the software. The code is available upon request from the authors.

4 Empirical Applications

We utilize our data set on notebook auctions from July to October 2001 to provide an empirical demonstration of our model. We use the first half of the auction items to initialize an individual bidder's past bidding experience, EXP_{ij}^k in equation (14) and the second half of the auctions to calibrate (and also validate) the proposed model. To allow for shorter MCMC run times, without loss of generality, we randomly sample 20% of the notebook auctions to obtain model inferences.[7] This sampling resulted in 269 in-sample (calibration) auction items in which 30 notebook auctions

---

[7]We also ran our model for multiple sub-samples of 10% and our findings are entirely robust.

have no bids. The total number of bids in these auctions is 2319 bids.

In order to validate the proposed model with the out-of-sample data, we randomly sample 10% of the notebook auctions which results in 1148 bids across 132 auction items (14 auctions with no bids). An assessment using exploratory methods indicated that both the calibration and validation auction items were representative of the total data set described in Table 1 and section 2.

There are three major areas that we report upon in summarizing our results. First, we describe inferences that can be obtained under the model by looking at summaries of the posterior distribution for $(\alpha_0, \alpha'_1, \ldots, \alpha'_6)$, $(\beta_0$ and $\beta_1)$, and $(\gamma_0$ and $\gamma_1)$. Second, we report findings on assessing the in-sample fit of the model by looking at the posterior predictive distribution (Gelman, Meng and Stern 1996) of various summaries, as described in our exploratory analyses. Finally, we report a set of model validations based on the out-of-sample data (auctions).

4.1 Parameter Inferences

Our initial inferences, as given in Table 2, are those derived directly from the parameter estimates of the model. Presented are findings from the results of three different (plausible) and informative models. Models 1 and 2 use data only on auctions with bids (i.e., ignore auctions with no bids) where Model 1 considers only observed bidders and Model 2 includes latent bidders. That is, as a series of benchmark models, we are interested in whether including latent bidders in the calibration set would fundamentally change our parameter inferences. In contrast, Model 3 is run on data including latent bidders and all auctions with and without bids. We find, after checking in a variety of ways, that estimation results are largely consistent across the three different models. Hence, given the theoretical completeness of Model 3 for having accounted for latent bidders and auctions with no bids, we describe detailed findings from Model 3 only.

Insert Table 2 about here

The major findings suggested are as follows. First, the estimates of most design variables are significantly associated with WTB. In particular, we find that BIN option (BUYNOW) appears to reduce WTB, whereas BIN price (BUYBID) tends to have a significant positive effect on WTB. This is expected because BIN prices on average are lower than the average market prices. We also observe that MINBID has a very significant positive effect on WTB, in accordance with extant literature (e.g., Kamins, Drèze and Folkes 2004) and DURATN is negatively associated with WTB, reflecting that (for our data) the waiting time effect appears to be larger than the arrival process effect.[8]

Second, we find that NEGREP has a very significant negative effect on WTB, whereas POS-REP is not statistically significant. This result is interesting because most research of static bidder valuations of the final auction price reported that the amount of negative (positive) seller reputation is negatively (positively) associated with the sale price (e.g., Houser and Wooders 2003, Melnik and Alm 2002). Third, we note very significant positive effects for both PENTIUM and SPEED, which given their importance in defining notebook computers, is as expected. We also find significant positive effects for MEMORY, HDISK, and SCREEN and a significant negative effect for MONTHS.[9] Finally, we observe that bidder behavioral characteristics play an important role in determining WTB. Among the variables of bidder-specific characteristics, VIEW (the extent of search behavior) is statistically significant. The negative estimate indicates that one's extensive search behavior reduces her WTB, and may suggest strategies for Internet auction sellers towards newer auction participants who have not extensively searched.

---

[8]Two contradicting effects, in theory, exist in online auctions on the basis of the auction duration. First, the waiting time effect implies that a short duration decreases a bidder's disutility of delay. Second, the arrival process effect means that a seller who keeps the auction open longer may be more likely to accumulate bidders and hence get higher bids.

[9]We do not report results of brand names here, yet they are included in the model.

We next discuss the parameter estimates of time-varying variables. First, we find that the estimates of all bid-specific variables except BIDRATE are significantly associated with WTB. REMAIN (i.e., time-to-end) is negatively related with WTB, which is descriptively consistent with a last-minute bidding phenomenon often observed at the end of auctions. The positive estimate of NUMBID is consistent with the notion of herding behavior in the bidding process (e.g., Dholakia and Soltysinski 2001); that is, after controlling for all other factors, the number of bids submitted is associated with one's WTB. These findings indicate that the information in the other participants' bids has an impact on the bidding process, which will influence subsequent bidding decisions among the potential bidders.

Second, we find that an individual bidder's past winning experience (i.e., LWIN and TWIN) are not statistically significant, which may not be surprising given the nature of consumer electronics (notebook) used in this research. However, both variables of an individual bidder's winning experience could be important in repeatedly purchased product categories such as antiques, collectibles, travel packages, and etc. in online auctions. In sharp contrast, an individual's failures (i.e., TLOSS and AMTLOST) in the past are significantly and negatively associated with her WTB. For instance, a negative relationship between AMTLOST and the bidder valuations is found, which implies that the loser at a higher rank (e.g., second highest) on the most recent auction tends to have a higher WTB if she participates in an auction later. The implications of this suggest that when sellers have the opportunity to target previous "losers" at a higher rank, they may want to do so, all else equal.

Turning to the relationship between WTB and the rate of bid speed, we note that the coefficient ($\hat{\beta}_1$) for $(w_{ij}^k - b_i^{k-1})$ is negative, which indicates that if one perceives higher surplus, she is more likely to bid late in the process, supporting our empirical findings of the last-minute bidding phenomenon. Table 2 also presents the parameter estimates for p_{ij}^k under the how much model. We find the coefficient ($\hat{\gamma}_1$) for $(w_{ij}^k - b_i^{k-1})$ is positive, which implies that if one perceives higher surplus,

conditional on someone's valuation being above BIN price, she is more likely to end the auction with BIN price.

4.2 Model Fit

In order to assess whether our proposed model properly captures the key behavioral aspects of online bidding, and to provide an assessment of the improved fit of Model 3 over Models 1 and 2, we report results for both in-sample and out-of-sample fit of the models in a number of different ways. All statistics of the model fit provide strong evidence that the model fits very well in most aspects. We start with two sets of results for in-sample fit and conclude with out-of-sample fit.

One way in which we assess the in-sample fit of the model is to compute the posterior distribution of MAPE (mean absolute prediction error) for the following key measures across the iterations of the MCMC sampler: (1) number of unique bidders per auction, (2) bid submission time, and (3) bid amount. Our findings indicate that the MAPE of the bid amount improves as potential bidders and auctions with no bids are included ($55.63, $52.22, and $50.69 for Models 1, 2 and 3, respectively), reflecting its richer representation. Both the MAPE of the number of bidders per auction (about 1.1) and the MAPE of the bid time (about .26) are largely consistent across the models. These summary features provide us a rough benchmark as to the quality of in-sample fit of the integrated model, especially when compared to the average number of observed and latent bidders (5.8 and 9.9), bid time (3.21 days), and bid amount ($762).

We next check the in-sample hit rate of the whether model to see whether the model inference correctly predicts whether or not the auction will have a bid at all. We find that the hit rate is 85% under Model 3 which is significantly better than chance under random assignment.[10] We also check the in-sample hit rate for auctions with BIN feature. The hit rate metric for BIN auctions

---

[10]Since about 11% (i.e., 296 out of 2618 auctions) have no bids in the database, a hit rate under random assignment is 80%, i.e., $11\% \cdot 11\% + 89\% \cdot 89\%$.

is to see whether the model inference correctly predicts whether or not the bidder bids BIN price if her valuation allows her to do so. We find that the hit rate is 87% under Model 3. We also note the model's performance to predict who the bidder is over the entire sequence of bids in an auction. Our findings indicate a hit rate of 59% for the who part under Model 3. All of these hit rates are a very encouraging sign about the validity and managerial usefulness of the model.

A final way in which we assess the quality of the model fit is to look at the bivariate distribution of predicted values (computed as posterior means across draws) from the model versus observed values in the out-of-sample data. Figure 1 presents scatterplots of the observed versus fitted values of the key measures. The solid line represents a perfect fit, whereas the dotted line is the best fit (least squares) line. Our ability to recover these summaries is generally very high. For the more interesting/important measures (the number of bidders per auction, whether or not an individual bidder ever participated in an auction, and the bid amount of how much the bidder would bid if she participated in the auction), the fit appears to be very good. We note the remarkably high correlations between observed and predicted values: .98 in the number of bidders per auction, .92 in the bid time and .95 in the bid amount under Model 3.

Insert Figure 1 about here

5 Conclusions and Future Research

We have provided a general framework for modeling bidding behavior in Internet auctions. To achieve this goal, we have incorporated four key components of bidding behavior, i.e., whether an auction will have a bid at all, (if so) who has bid, when they have bid and how much they have bid over the entire sequence of auction bids. The method by which we have examined these outcomes is through a latent behavioral construct which we denote WTB. Each decision is closely linked to the others based on this single latent construct that leads to an integrated framework

with mathematically tractable closed-form solutions. While we acknowledge that more complex, and albeit realistic, behavioral models could be built, our integrated model is parsimonious and captures the key behavioral aspects of bidding behavior established in the existing literature on Internet auctions.

Using a database of notebook auctions directly obtained from an auction company, we have demonstrated that this general (yet parsimonious) model shows evidence of remarkable performance in its ability to help us understand and describe online bidding behavior. Beyond the performance of the model, this research can help shed new light on the important topic of contact strategy in which an auction site advertises auctioned items and recruits bidders for those products. While conventional auction theory states that the profitability of an auction is positively related to the number of bidders, Wang and Montgomery (2003) show that too much advertising by the auction site can reduce a bidder's satisfaction and subsequent retention in future auctions, which lowers the profits of the auction site. Since our model can provide a way to evaluate the "goodness" (whether) of the listed auction items and the "goodness" (who, when and how much to bid) of potential bidders on an auctioned item, our research can be considered as a very useful step for deriving better communication strategies for auction sites.

Since this research is among the first attempts to investigate the behavioral aspects of the bidding process over a series of bids in Internet auctions, we have kept the model as simple as possible to highlight the key behavioral phenomena that we have identified. Naturally, there are several limitations in the proposed model that should be acknowledged and perhaps addressed in future research. First, we have estimated our model for notebook auctions in which our results could be reflective of the product category. We hope our model provides a framework for further empirical exploration in other product categories (e.g., collectibles).

Second, as noted in section 2, the data considered here is from ascending first-price auctions.

This choice by the site, in contrast to second-price auctions used at eBay, has been under much discussion in the auction literature (e.g., Lucking-Reiley 1999) and more recently by Bajari and Hortaçsu (2004). With that in mind, the empirical findings reported here should be thought more of as suggestive than in any way definitive and worth further exploration in future research. Yet, we hope our model provides a framework for future empirical exploration. Third, the model developed here is descriptive and exploratory in nature and is not based on theoretical assumptions of auction behavior and equilibria in the extant economic literature applied to auctions (e.g., Donald and Paarsch 1996, Guerre, Perrigne and Vuong 2000, Laffont, Ossard and Vuong 1995). The dynamic model employed here might provide some insight to those working in this area such that even richer theoretical models can be developed.

While the focus of this paper has been on providing an integrated stochastic dynamic model for Internet auctions useful for description and prediction, models such as the one developed here are commonly used for managerial optimization. That is, since many of the variables incorporated in WTB are under the control of the seller, one could try and optimize auction profits as a function of seller's design. While we wholeheartedly agree that this would be a valuable way to apply our model, we do not present such findings here because many of the auction design variables may well be endogenously set on the part of the seller and hence a richer model where the seller's variables may be strategically set is needed.

Nevertheless, all is not lost. One could incorporate into the model knowledge of how a seller designs an auction, thus allowing optimization decisions; albeit, such an approach would require careful work as to the nature of strategic bidding behavior. Fortunately, models which incorporate endogenously set variables directly into the likelihood (Manchanda, Rossi and Chintagunta 2004) are starting to emerge and present an important solution allowing for optimization. Thus, we hope that our research serves as a building block model for which important optimization decisions can

be made in Internet auctions.

Finally, it is important to recognize that while a fundamental understanding of the "why" in bidding behavior is still left somewhat unknown, there are numerous aspects of the research that are novel and extend the auction literature: (1) an integrated probabilistic framework which allows us to forecast all four (whether, who, when and how much) modules of Internet auction behavior simultaneously, (2) a dynamic model that focuses on intermediary outcomes, for which final outcomes are but a result, (3) the inclusion of cross-auction panel variables that allows us to explore the impact of previous wins/losses and their amounts on current bidding behavior, and (4) a parsimonious model which allows the inference of a latent competition set from the observed data. With this general structure in place, our hope is that this paper serves as a call for much more extensive empirical study of the application of probabilistic models to auction data, both first-price and otherwise.

Appendix

Let $t_{i1}^k, \ldots, t_{iJ_i^k}^k$ be independent random variables, with t_{ij}^k having an exponential distribution with rate λ_{ij}^k. One critical aspect of our modeling approach is that the minimum of a set of exponentials is exponentially distributed with rate equal to the sum of the rates:

$$
\begin{aligned}
\Pr[\min(t_{i1}^k, \ldots, t_{iJ_i^k}^k) > t] &= \Pr[t_{i1}^k > t, \ldots, t_{iJ_i^k}^k > t] \\
&= \Pr[t_{i1}^k > t] \cdots \Pr[t_{iJ_i^k}^k > t] \\
&= e^{-\lambda_{i1}^k t} \cdots e^{-\lambda_{iJ_i^k}^k t} \\
&= e^{-(\sum_{j=1}^{J_i^k} \lambda_{ij}^k) t}
\end{aligned}
$$

Without loss of generality, we do not take out the current rate of bid speed by the previous bidder here in the derivation, but do so in the text. The preceding shows that the CDF of $\min(t_{ij}^k, \forall j \in J_i^k)$ is that of an exponential distribution with rate equal to $\sum_{j=1}^{J_i^k} \lambda_{ij}^k$. Thus, the when model can be derived based on the truncated exponential distribution.

We next derive the model of who has bid at bid k in auction i among the total of the J_i^k bidders:

$$
\begin{aligned}
&\Pr[\min(t_{ij}^k) = t_{ij}^k \,|\, t_i^{k-1} < \min(t_{ij}^k) \le T_i \,, \forall j \in J_i^k] \\
=\ &\Pr[t_{ij}^k < t_{ij^c}^k, \forall j^c \ne j \,|\, t_i^{k-1} < \min(t_{ij}^k) \le T_i \,, \forall j \in J_i^k] \\
=\ &\int_{t_i^{k-1}}^{T_i} \left\{ \begin{array}{c} \Pr[t_{ij}^k < t_{ij^c}^k, \forall j^c \ne j \,|\, t_{ij}^k = t, \, t_i^{k-1} < \min(t_{ij}^k) \le T_i \,, \forall j \in J_i^k] \\ \times \\ f(t_{ij}^k = t \,|\, t_i^{k-1} < \min(t_{ij}^k) \le T_i \,, \forall j) \end{array} \right\} dt \\
=\ &\int_{t_i^{k-1}}^{T_i} \Pr[t < t_{ij^c}^k, \forall j^c \ne j] \times \frac{\lambda_{ij}^k e^{-\lambda_{ij}^k t}}{1 - e^{-(\lambda_{i1}^k + \lambda_{i2}^k + \ldots + \lambda_{iJ_i^k}^k) \cdot (T_i - t_i^{k-1})}} \, dt \\
=\ &\int_{t_i^{k-1}}^{T_i} \prod_{j^c=1, j^c \ne j}^{J_i^k} e^{-\lambda_{ij^c}^k t} \times \frac{\lambda_{ij}^k e^{-\lambda_{ij}^k t}}{1 - e^{-(\sum_{j=1}^{J_i^k} \lambda_{ij}^k) \cdot (T_i - t_i^{k-1})}} \, dt \\
=\ &\frac{\lambda_{ij}^k}{1 - e^{-(\sum_{j=1}^{J_i^k} \lambda_{ij}^k)(T_i - t_i^{k-1})}} \int_{t_i^{k-1}}^{T_i} e^{-(\sum_{j=1}^{J_i^k} \lambda_{ij}^k) t} \, dt \\
=\ &\frac{\lambda_{ij}^k}{\sum_{j=1}^{J_i^k} \lambda_{ij}^k} \ .
\end{aligned}
$$

References

Alba, Joseph W. and J. Wesley Hutchinson (1987), "Dimensions of Consumer Expertise," *Journal of Consumer Research*, 13 (4), 113-123.

Ariely, Dan and Itamar Simonson (2003), "Buying, Bidding, Playing or Competing? Value Assessment and Decision Dynamics in Online Auctions," *Journal of Consumer Psychology*, 13 (1), 113-123.

Bajari, Patrick and Ali Hortaçsu (2003), "The Winner's Curse, Reserve Prices, and Endogenous Entry: Empirical Insights from eBay Auctions," *RAND Journal of Economics*, 34 (2), 329-355.

Bajari, Patrick and Ali Hortaçsu (2004), "Are Structural Estimates of Auction Models Reasonable? Evidence from Experimental Data", Working Paper, Department of Economics, Duke University.

Bettman, James R., Mary Frances Luce, and John W. Payne (1998), "Constructive Consumer Choice Processes," *Journal of Consumer Research*, 25 (3), 187-217.

Bradlow, Eric T. and Young-Hoon Park (2004), "Bayesian Estimation of Bid Sequences in Internet Auctions Using a Generalized Record Breaking Model," Working Paper, The Wharton School, University of Pennsylvania.

Chakravarti, Dipankar, Eric A. Greenleaf, Atanu Sinha, Amar Cheema, James C. Cox, Daniel Friedman, Teck H. Ho, R. Mark Isaac, Andrew A. Mitchell, Amnon Rapoport, Michael H. Rothkopf, Joydeep Srivastava, and Rami Zwick (2002), "Auctions: Research Opportunities in Marketing," *Marketing Letters*, 13 (3), 281-296.

Dholakia, Utpal M. and Kerry Soltysinski (2001), "Coveted or Overlooked? The Psychology of Bidding for Comparable Listings in Digital Auctions," *Marketing Letters*, 12 (3), 225-237.

Donald, Stephen G. and Harry J. Paarsch (1996), "Identification, Estimation, and Testing in Parametric Empirical Models of Auctions within the Independent Private Values Paradigm," *Econometric Theory*, 12 (3), 517-567.

Fader, Peter S. and Leigh McAlister (1990), "An Elimination by Aspects Model of Consumer Response to Promotion Calibrated on UPC Scanner Data," *Journal of Marketing Research*, 27 (August), 322-332.

Gelfand, Alan E. and Adrian F. M. Smith (1990), "Sampling-Based Approaches to Calculating Marginal Densities," *Journal of the American Statistical Association*, 85 (410), 398-409.

Gelman, Andrew, John B. Carlin, Hal S. Stern, and Donald B. Rubin (1995), *Bayesian Data Analysis,* London: John Wiley & Sons.

Gelman, Andrew, Xiao-Li Meng, and Hal S. Stern (1996), "Posterior Predictive Assessment of Model Fitness via Realized Discrepancies (with discussion)," *Statistica Sinica*, 6 (4), 733-807.

Gelman, Andrew and Donald B. Rubin (1992), "Inference from Iterative Simulation Using Multiple Sequences," *Statistical Science*, 7 (4), 457-472.

Guerre, Emmanuel, Isabelle M. Perrigne, and Quang Vuong (2000), "Optimal Nonparametric Estimation of First-Price Auctions," *Econometrica*, 68 (3), 525-574.

Hoch, Stephen, J., Eric T. Bradlow, and Brian Wansink (1999), "The Variety of an Assortment," *Marketing Science*, 18 (4), 527-546.

Houser, Daniel and John Wooders (2003), "Reputation in Auctions: Theory, and Evidence from eBay," Working Paper, Department of Economics, University of Arizona.

Kamins, Michael A., Xavier Drèze, and Valerie S. Folkes (2004), "A Field Study of the Effects of Minimum and Reserve Prices on Internet Auction," *Journal of Consumer Research*, 30 (4), 622-628.

Laffont, Jean-Jacques, Herve Ossard, and Quang H. Vuong (1995), "Econometrics of First-Price Auctions," *Econometrica*, 63 (4), 953-80.

List, John A. and David Lucking-Reiley (2002), "Bidding Behavior and Decision Costs in Field Experiments," *Economic Inquiry*, 40 (44), 611-619.

Lucking-Reiley, David (1999), "Using Field Experiments to Test Equivalence Between Auction Formats: Magic on the Internet," *American Economic Review*, 89 (5), 1063-1080.

Manchanda, Puneet, Peter E. Rossi and Pradeep K. Chintagunta (2004), "Response Modeling with Non-Random Marketing Mix Variables," *Journal of Marketing Research*, (forthcoming).

Melnik, Mikhail I. and James Alm (2002), "Does a Seller's E-Commerce Reputation Matter? Evidence from eBay Auctions," *Journal of Industrial Economics*, 50 (3), 337-350.

Reynolds, Stanley S. and John Wooders (2003), "Auctions with a Buy Price," Working Paper, Department of Economics, University of Arizona.

Roth, Alvin E. and Axel Ockenfels (2002), "Last-Minute Bidding and the Rules for Ending Second-Price Auctions: Evidence from eBay and Amazon Auctions on the Internet," *American Economic Review*, 92 (4), 1093-1103.

Todd, Peter, M and Gerd Gigerenzer (2000), "Précis of Simple Heuristics That Make Us Smart," *Behavioral and Brain Sciences*, 23 (5), 727-780.

Tversky, Amos and Daniel Kahneman (1986), "Rational Choice and the Framing of Decisions," *Journal of Business*, 59 (4, Part 2), S251-S278.

Wang, Xin and Alan L. Montgomery (2003), "The Effects of Advertising on Customer Retention and the Profitability of Auctions," Working Paper, Graduate School of Industrial Administration, Carnegie Mellon University.

Wilcox, Ronald T. (2000), "Experts and Amateurs: The Role of Experience in Internet Auctions," *Marketing Letters*, 11 (4), 363-374.

Variables	Abbreviation	Mean	Std.
Auction design			
Product image (Yes = 1, No = 0)	IMAGE	0.6	
Minimum bid (dollars)	MINBID	713	650
BIN option (Yes = 1, No = 0)	BUYNOW	0.9	
BIN price (dollars)	BUYBID	1195	635
Duration (days)	DURATN	5.4	2.7
Seller reputation			
Number of positive ratings	POSREP	1.7	15.5
Number of negative ratings	NEGREP	0.2	3.6
Product specification			
CPU (Pentium = 1, Celeron = 0)	PENTIUM	0.8	
CPU speed (mhz)	SPEED	534	220
Memory (megabytes)	MEMORY	118	60
Hard disk (gigabytes)	HDISK	12	7
Screen size (inches)	SCREEN	12.9	1.2
Number of months for use	MONTHS	9.6	9.9
Bidder characteristics			
Gender (Male = 1, Female = 0)	MALE	0.8	
Age	AGE	31	8
Cumulative site visits	VISIT	99	230
Cumulative page views	VIEW	505	1156

Notes: Brand names are not included.

Table 1: Data Description

	Variables	Model 1	Model 2	Model 3
w_{ij}^k	INTERCEPT	.64*	.72*	.90*
	AD_i			
	IMAGE	.02	.02	.04
	MINBID	.61*	.55*	.39*
	BUYNOW	-1.69*	-1.67*	-1.66*
	BUYBID	.04*	.21*	.31*
	DURATN	-.13*	-.13*	-.15*
	SR_i			
	POSREP	.09	.10	.10
	NEGREP	-.21*	-.13*	-.14*
	PS_i			
	PENTIUM	1.51*	1.43*	1.37*
	SPEED	.99*	.98*	.98*
	MEMORY	.20*	.19*	.19*
	HDISK	.14*	.13*	.13*
	SCREEN	.24*	.23*	.22*
	MONTHS	-.11*	-.08*	-.04*
	BC_j			
	MALE	.03	.02	.02
	AGE	.02	.02	-.04
	VISIT	-.03	-.03	-.03
	VIEW	-.03*	-.03*	-.04*
	RC_i^k			
	REMAIN	-1.20*	-1.31*	-1.53*
	NUMBID	.01*	.01*	.02*
	BIDRATE	.36	.40	.48
	AMTRATE	-.08*	-.11*	-.13*
	EXP_{ij}^k			
	LWIN	-.07	-.07	-.07
	TWIN	-.01	-.01	-.02
	TLOSS	-.03*	-.03*	-.03*
	AMTLOST	-.02*	-.03*	-.03*
$\log(\lambda_{ij}^k)$	INTERCEPT	6.07*	7.35*	9.21*
	$(w_{ij}^k - b_i^{k-1})$	-.67*	-.74*	-.91*
p_{ij}^k	INTERCEPT	-12.24*	-12.01*	-11.97*
	$(w_{ij}^k - b_i^{k-1})$	1.87*	1.57*	1.48*

Notes: 1. Parameter estimates for brand names are not included.
2. * indicates that zero lies outside of the 95% posterior interval.
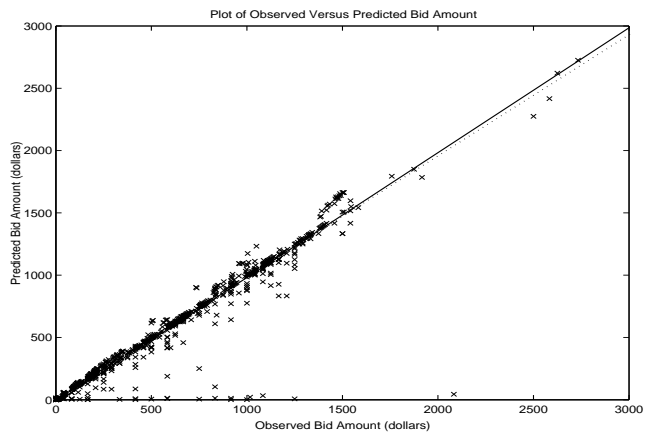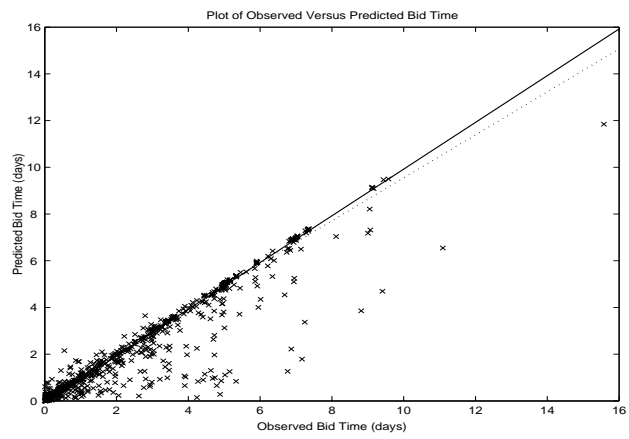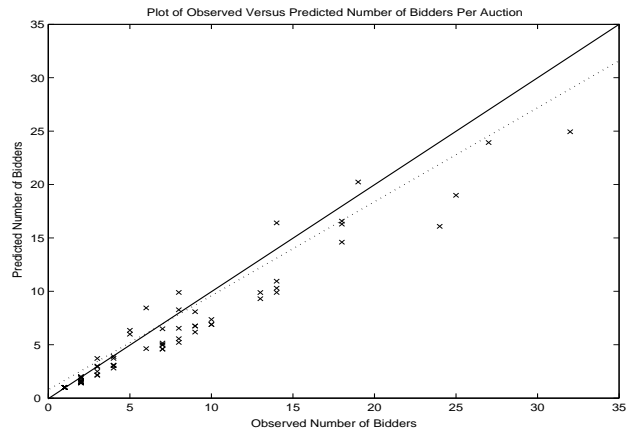
Table 2: Posterior Means of Coefficient Estimates

Figure 1: Plots of Observed Versus Predicted Values