Optimal Design of the Online Auction Channel: Analytical, Empirical and Computational Insights

Ravi Bapna[1]
(rbapna@sba.uconn.edu)

Paulo Goes[1]
(paulo@sba.uconn.edu)

Alok Gupta[2]
(gupta037@umn.edu)

Gilbert Karuga[1]
(karuga@sba.uconn.edu)

1- Dept. of Operations & Information Management,
U-41 IM, School of Business Administration,
University of Connecticut,
Storrs, CT 06269

2- Information and Decision Sciences Department
Carlson School of Management
University of Minnesota
3-365 Carlson School of Management
321 - 19th Avenue South
Minneapolis, MN 55455

Acknowledgement

Alok Gupta's research is supported by NSF CAREER grant #IIS-0092780, but does not necessarily reflect the views of the NSF. Partial support for this research was also provided by TECI - the Treibick Electronic Commerce Initiative, OPIM/SBA, University of Connecticut.

Abstract

The focus of this study is on business-to-consumer (B2C) online auctions made possible by the advent of electronic commerce over an open-source, ubiquitous Internet Protocol (IP) computer network. This work presents an analytical model that characterizes the revenue generation process for a popular B2C online auction, namely Yankee auctions. Such auctions sell multiple identical units of a good to multiple buyers using an ascending and open auction mechanism. The methodologies used to validate the analytical model range from empirical analysis to simulation. A key contribution of this study is the design of a partitioning scheme of the discrete valuation space of the bidders such that equilibrium points with higher revenue structures become identifiable and feasible. Our analysis indicates that the auctioneers are, most of the time, far away from the optimal choice of key control factors such as the bid increment, resulting in substantial losses in a market with already tight margins. With this in mind, we put forward a portfolio of tools, varying in their level of abstraction and information intensity requirements, which help auctioneers maximize their revenues.

Keywords: emerging supply chain channels, online auctions, simulation

1. Introduction

Online auctions, in the absence of spatial, temporal and geographic constraints, provide an alternative supply chain channel for the distribution of goods and services. This channel differs from the common posted-price mechanism that is typically used in the retail sector. In consumer-oriented markets, buyers can now experience the thrill of 'winning' a product, potentially at a bargain, as opposed to the typically more tedious notion of 'buying' it. Sellers, on the other hand, have an additional channel to distribute their goods, and the opportunity to liquidate rapidly aging goods at greater than salvage values. The primary facilitator of this phenomenon is the widespread adoption of electronic commerce over an open-source, ubiquitous Internet Protocol (IP) based network.

In this paper, we concentrate on optimizing the design of an emerging business-to-consumer (B2C) distribution channel known as Yankee auctions. Such auctions sell multiple identical units of a good to multiple buyers using an ascending and open auction mechanism, which has its roots in the English auction, yet are significantly different.

This work presents an analytical model that characterizes the revenue generation process of Yankee auctions. To validate the analytical model and to gain a better understanding of the revenue generation process of such auctions, we analyze real-world empirical data collected by a software agent that tracked these auctions round the clock. An interesting by-product of the auction data collection process is our ability to construct empirical demand curves for the auctioned goods. Consumer demand information is an important input in supply chain management. It provides the feedback necessary for making key decisions downstream in the supply chain. We demonstrate, using the capabilities of the Internet, how dynamic pricing mechanisms such as online auctions can provide opportunities for integration of demand information into the mechanism design process. This enhances the mechanisms in two ways. First, by appropriately setting the online auction parameters, auctioneers can maximize their returns. Secondly, by recognizing the demand implication and visualizing the trading process *a priori,* the eventual allocation can be more equitable thus resulting in higher welfare for both consumers and the auctioneer.

To validate and complement the analytical model, we also introduce a flexible simulation model, thus offering auctioneers a portfolio of tools, varying in their level of abstraction and information intensity requirements, to help them maximize their revenues. In summary, this portfolio of decision-making tools provides a relatively risk-free and cost-effective approach to managing this new, web-based dynamic pricing distribution channel prevalent in the online setting, namely the Yankee auction.

The rest of this paper is organized as follows. In section 2 we describe the revenue generation process of the Yankee auction mechanism. Our understanding of the revenue generation process of Yankee auctions directs us to focus our attention to the combinatorial dynamics of the penultimate rounds of the auction, which forms the basis of our theoretical model in this paper, presented in section 3. In section 4 we show how the theoretical results can be applied to actual Yankee auctions by using consumer demand estimates derived from real collected data. Section 5 provides a validation approach to the theoretical analysis using a simulation tool under varying degree of information abstraction. We conclude in section 6 with an overview of our approach and a summary of the findings of the paper.

2. **The Yankee Auction Mechanism**

The Yankee auction is a special case of multi-unit English auction. Here, multiple units of the same product are sold to multiple bidders. It is well known [Rothkopf and Harstad (1994a)] that single-item results (a vast majority of auction theory studies fall under this category) do not carry over in multiple-unit settings. The multi-unit and discrete nature of these mechanisms renders the traditional analytic framework of game theory intractable [Nautz and Wolfstetter (1997)].

The auction is progressive in nature; however, each new bid does not have to be strictly greater than the previous bid since there are multiple units available. The set of winning bids consists of the top N bids, where N is the number of units up for auction. A new bid either has to be equal to the minimum bid that is among the winning bids (if the set of winning bids has a cardinality of less than N) or it has to be at least equal to minimum winning bid plus a pre-specified minimum bid increment. With multiple identical items on offer, it is possible to observe several winning bids that are equal. Once the consumers have bid for the entire lot size, a new bid will have to be greater than the smallest winning bid. When such a bid is submitted, the winner with the smallest winning bid is replaced by the new bid. If several offers are equal and at the minimum winning bid level, a time priority is applied to determine the bid to be displaced when a new and higher offer is received. The last bid at the minimum winning bid level becomes the first to leave the auction winners' list. This process continues until the auction closes. At this point the auction winners are determined. The auction terminates on or after a pre-announced closing time and each of the winning bidders pay the amount they last bid to win the auction. Most auctions have a going, going, gone period such that the auction terminates after the closing time has passed and no further bids are received in the last five minutes. Note that in multi-unit settings this often leads to discriminatory pricing with consumers paying different amounts for the same item. Such auctions are used on a variety of auction sites on the WWW, pioneered by Egghead.com's Surplus Auctions (now defunct) and now popularized by Ubid.com.

The key factors that auctioneers can control in Yankee auctions are: (i) the lot size; (ii) the bid increment; (iii) the auction duration; and (iv) the opening bid.

Ignoring monitoring costs for the present, we assume that customers maximize their net value and hence always bid at the current ask price, provided that the current ask price does not exceed their valuation of the item. Rothkopf and Harstad (1994b), in their single item analysis, characterize this as the *pedestrian* approach to bidding. Such a strategy is consistent with the rational, net worth maximizing assumption for consumers. Easley and Tenorio, 1999, extend this result to Yankee auctions, conditioning it on the absence of any cost of preparing and submitting a bid. Notably, such a strategy could involve active manual participation or could be undertaken using a programmed software agent that bids the minimum required bid at any stage during the auction. Both have been observed in practice. In adopting this strategy bidders choose to be no more aggressive than necessary to continue competing.

2.1 Bid Increment and Auction Revenue

In prior research Bapna (1999) uses a regression model to show that amongst all the control factors mentioned in the previous section, the bid increment is the only factor significant in explaining variations of auction revenues. Additionally, anecdotal evidence suggests that auctioneers realize the importance of bid increments. We routinely observed that similar items are auctioned, at different times, using different bid increments.

While most of existing theory [see McAfee and McMillan, (1987), Milgrom, (1989), Milgrom and Weber (1982), for a detailed overview] analyzes auctions under either the private or the common value setting, the online context in which these *B2C* auctions take place, makes such a strict classification inaccurate. A close observation of the types of goods sold in these liquidation kinds of auctions indicates that most of the items (such as computer hardware and consumer electronics) have both idiosyncratic (private) and common value elements. This is more so given the presence of imperfect substitutes and price-comparison agents that provide information regarding the alternative comparable products and their posted prices. Thus, based on the general model of Milgrom and Weber (1982), the multi-unit *B2C* auctions lie in the continuum between the private and common value models. See Paarsch (1992) on how to decide between the common values and private value paradigms in auctions. The presence of price-comparison agents creates a mass of consumer valuation at or around the prevailing market price. Consequently, we would expect that in progressive online auctions, such as the Yankee auction, such bid levels would be realized towards the end of the auctions, rather than in the beginning or intermediate stages. This forms the motivation behind our attention to the combinatorial dynamics of the penultimate auction rounds.

Assuming that bidders are rational and follow a pedestrian bidding strategy, at the final stage of the auction, at most two bidding levels are observed [see, Bapna, Goes, and Gupta (2000)]. At a minimum, all bids would be at the lower level, while the other extreme would be to have all bidders at the higher level. The difference between the two bidding levels is equal to the bid increment. From an auctioneer's perspective the larger the number of bidders at the higher bid level, the greater the revenue. *Intuitively, the process of determining the optimal bid increment is to create a partition in the discrete valuation space of the bidders such that the higher bid level becomes feasible to the maximum possible number of bidders.* It follows that such a partitioning policy would optimize the expected revenue.

In the next section we present our analytical model that characterizes the expected revenue for Yankee auctions.

3. The Theoretical Model

Prior research [Bapna (1999) and Bapna, Goes, and Gupta (2000, 2001a)], based on a multi-variate regression analysis of multi-unit Yankee auctions, revealed that, to a large extent, the valuation of the *marginal consumer* and the *bid increment* set by the auctioneer, determine the range of the auction revenues. The standard practice in the auction literature is to define the marginal consumer as either the highest unsuccessful bidder or the lowest successful bidder [Bulow and Roberts (1989)]. Both definitions characterize the price-setting consumer and are equally useful in examining the structural characteristics of these auctions. We consider in this paper the marginal bidder as the *highest unsuccessful bidder*.

In the derivation of a theoretical model for the auctioneer's expected revenue we assume that all bidders employ a pedestrian strategy, i.e., at any point in time during the course of the auction, they will not bid higher than the required minimum to make it to the winner's list. The implication of this assumption is that the winning bids display at most two values, say B and $B + k$, where k is the bid increment.

Let there be N identical items for sale, and I bidders, each with a value V_i, $i = 1,..., I$, for the product. It is assumed that $I > N$, otherwise bidders who have a clear objective of maximizing their surplus would not bid any higher than the opening bid level, and the opening bid would be binding. With $I > N$, the bidders with the highest N bids (note that, winners may not be the bidders with N highest valuations) will win the items being auctioned. The variations in revenue arise from the combinatorial dynamics of the penultimate round and the last bid of marginal customer as elaborated below.

If all the bidders employ the pedestrian approach to bidding, the highest unsuccessful bidder or marginal consumer coincides with the consumer who, at worst, has the $(N+1)$th highest valuation. Even though the results could be generalized to more than $N+1$ bidders eligible to bid in last 2 rounds, for expositional purposes we analyze the behavior of the $N+1$ highest valued bidders being in contention during the last 2 rounds of the auction.

Let B_0 denote the bid level, which all the $N+1$ final bidders can achieve. Let p denote the probability that a bidder's valuation is greater than B_0+k. Let p_1 be the probability that a person has a valuation greater than B_0+2k, p_2 be probability that a person has a valuation greater than B_0+3k, and so on. Figure 1 depicts an example distribution for these top $N+1$ valuations and the probabilities that a randomly chosen bidder among these top $N+1$ bidders will have a valuation greater than B_0+k, B_0+2k, and B_0+3k, respectively. Note that $p \geq p_1 \geq p_2$.

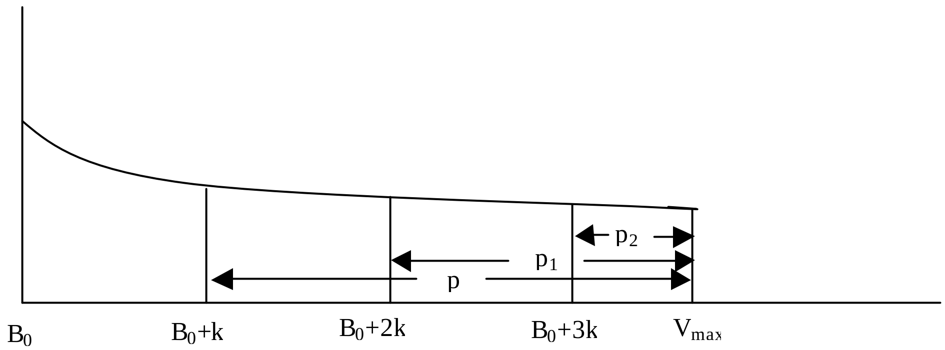

Figure 1 – The Distribution of Critical Fractile of Top *N+1* Valuations

Let us first consider that the maximum achievable bidding level for the auction is B_0+k. This will happen if the marginal customer cannot bid at that level. Let us derive what we term a *first order approximation* for the expected revenue. In this case, we know that the auction stops with all the bids either at B_0 or B_0+k. The expected number of bids that reach the maximum level B_0+k can be estimated by examining the sequence of bids that were first materialized at level B_0 (the penultimate round), and in particular the temporal position in which the marginal customer placed her bid at that level. Consider the scenarios below, which represent bidding at the level B_0. M is the marginal bidder with the *(N+1)th* highest value, which in this case is below B_0+k. W denotes the eventual winners and $N=5$.

Scenario	Customer who did not bid at B_0	Sequence of winning bids at B_0	No. bids that reach level B_0+k	Scenario probability
1	**M**	WWWWW	0	$1-p$
2	**W**	**M**WWWW	1	$(1-p)p$
3	**W**	**WM**WWW	2	$(1-p)p^2$

4	**W**	**WWM**WW	3	*(1-p)p*^{3}
5	**W**	**WWWM**W	4	*(1-p)p*^{4}
6	**W**	**WWWWM**	5	*(1-p)p*^{5}

In scenario 1, the marginal customer never had a chance to bid at B_0. Since his valuation is lower than B_0+k, the auction will stop with all winning bids at level B_0. The probability of this scenario occurring is equivalent to the probability that a bidder with valuation less than B_0+k is drawn at the "outside" position, which is $1-p$. In scenario 2, the marginal customer is the last one to bid at level B_0, and will be the first one to be displaced by a winner who can bid at B_0+k. Once out, M cannot bid again and the auction stops with only one bid at the higher level. The probability of scenario 2 occurring is equivalent to drawing two customers, one with valuation less than B_0+k and the other with valuation greater than B_0+k: $(1-p)p$. In scenario 3, M is the second last to bid at B_0. She will be displaced after a winner, who is out, displaces the last bidder, who in turn comes back and displaces M, with the final outcome being 2 bids at the higher level. Scenario 3 happens with probability $(1-p)p^2$. Similar reasoning can be applied to the other scenarios. Basically, we are computing probabilities of the geometric series bolded in the second and third columns of the table above. Hence the expected revenue for this first order approximation is

$$E(R) = NB_0 + k\sum_{i=0}^{N} ip^i(1-p), \tag{1}$$

where the second term indicates the expected additional revenue generated by those winners that were forced to move from B_0 to B_0+k. In a *second order approximation*, we incorporate the possibility that all *N+1* bidders have valuation greater than B_0+k, but at least one has valuation less that B_0+2k, so the maximum realizable bidding level is B_0+2k. The expected revenue is then derived as

$$E(R) = NB_0 + (1-p^{(N+1)})k\sum_{i=0}^{N} ip^i(1-p) + p^{(N+1)}\left(Nk + k\sum_{i=0}^{N} ip_1^i(1-p_1)\right) \tag{2a}$$

In a *third order approximation*, we consider the possibility of going to the next bidding level B_0+3k, and obtain

$$E(R) = NB_0 + (1-p_1^{N+1})\left[(1-p^{(N+1)})k\sum_{i=0}^{N} ip^i(1-p) + p^{(N+1)}\left(Nk + k\sum_{i=0}^{N} ip_1^i(1-p_1)\right)\right] + p_1^{N+1}\left(N*2K + k\sum_{i=0}^{N} ip_2^i(1-p_2)\right) \tag{2b}$$

We can theoretically derive higher order expressions. All of them will have three terms, with the second term getting more complex. If we represent each *(n-1)*-th order approximation by $R_{n-1} = E(R_{n-1}) - NB_0$, we can write the recursive expected revenue function as

$$E(R_n) = NB_0 + (1-p_{n-2}^{N+1}) * R_{n-1} + p_{n-2}^{N+1}\left(N*(n-1)k + k\sum_{i=0}^{N} ip_{n-1}^i(1-p_{n-1})\right) \tag{2c}$$

As *p*'s are getting smaller, the last term is getting to 0, and the first term probabilities $(1-p_*^{N+1})$ are going to 1. In addition, as *N* becomes larger p_*^{N+1} tends to zero. Therefore, in the limit, we can work with the first order approximation presented in equation 1, which can be rewritten as:

$$E(R) = NB_0 + k\sum_{i=0}^{N} ip^i(1-p) = NB_0 + k\left[\frac{p(1-p^N)}{(1-p)} - Np^{(N+1)}\right] \quad (3)$$

In other words, with $N+1$ bidders left, we only consider the possibility of one additional round of bids. The above expression represents a geometric series that, as illustrated in the table above, takes into account the temporal position of the marginal bidder, a key revenue determinant. Let us now motivate the combinatorial aspects of the revenue generation using a numerical example.

Numerical Example 1 **–** Let us assume N = 3 and p = ½, i.e., the probability that a person will have value > B_0+k = ½. We thus consider $N+1$ = 4 in the analysis of the last two rounds of the auction. There can be 5 cases possible: 1) None have value > B_0+k; 2) 1 of the 4 has value > B_0+k; 3) 2 of the 4 have values > B_0+k; 4) 3 of the 4 have values > B_0+k; 5) all 4 have values > B_0+k. Total numbers of states for this case are $4!*2^{4.}$ Let A, B, C, and D represent the 4 bidders. The underlined character means that the person has value > B_0+k. Assume that the first person is currently not winning the auction while the other 3 are, and the second character represent the person that is replaced first. So if the state of the world is <u>B</u>C<u>A</u>D then B bids at B_0+k and replaces C (since B has valuation > B_0+k) and the new state is C<u>A</u>D<u>B</u>. The auction stops at this stage because C cannot bid at B_0+k.

Consider the case that there are 3 bidders with valuation > B_0+k and (yet) the auction stops with 1 bid at B_0+k. This can happen in 24 different ways shown in the table below:

3 persons > B_0+k			
<u>A</u>D<u>BC</u>	<u>A</u>C<u>BD</u>	<u>A</u>B<u>CD</u>	<u>B</u>A<u>CD</u>
<u>A</u>D<u>CB</u>	<u>A</u>C<u>DB</u>	<u>A</u>B<u>DC</u>	<u>B</u>A<u>DC</u>
<u>B</u>D<u>AC</u>	<u>B</u>C<u>AD</u>	<u>C</u>B<u>AD</u>	<u>C</u>A<u>BD</u>
<u>B</u>D<u>CA</u>	<u>B</u>C<u>DA</u>	<u>C</u>B<u>DA</u>	<u>C</u>A<u>DB</u>
<u>C</u>D<u>AB</u>	<u>D</u>C<u>AB</u>	<u>D</u>B<u>AC</u>	<u>D</u>A<u>BC</u>
<u>C</u>D<u>BA</u>	<u>D</u>C<u>BA</u>	<u>D</u>B<u>CA</u>	<u>D</u>A<u>CB</u>

Table 1 – Example of Bidding Stopping With 1 Bid at B_0+k

Of course to compute the overall likelihood of the auction stopping at B_0+k we need to *enumerate all* the cases when the marginal bidder in the penultimate round has a valuation $> B_0+k$ and the lowest winning bidder has valuation $< B_0+k$.

In the example these are operationalized where the first letter is underlined but the second is not, i.e., the person not winning has valuation $> B_0+k$ but person displaced has valuation $< B_0+k$.

This would mean that we have to need to also consider the states when 1 and 2 bidders have valuations $> B_0+k$ respectively, and the temporal ordering is such that marginal bidder in the penultimate round has a valuation $> B_0+k$ and the lowest winning bidder has valuation $< B_0+k$. Through a similar enumeration of these cases as in Table 1, there are 24 such cases when 1 bidder has valuation $> B_0+k$ and 48 such cases when 2 bidders have valuation $> B_0+k$.

Therefore, in all, there are 96 such cases, implying that the likelihood of the expected revenue being NB_0+k = 96/384= ¼. This is of course identical to probability from our expected revenue (1a) function $p^1*(1-p)$ with p = ½. .

From the expected revenue expression in equations 1-3, it is clear that the bid increment k is a key determinant of the auction revenue. In this paper we seek to establish calibration mechanisms for the bid increment that optimize the expected auction revenue. A critical parameter necessary to optimize the bid increment is the value of the probability p that a bidder will be able to bid at the next higher bid level above B_0. To estimate p the auctioneer has to have some information on the bidders valuation, a non-trivial task. Using such information, for any given bidding level, the auctioneers can infer the number of bidders who may have valuations for the product that are equal to or higher than the next feasible bid level.

Before we describe the empirical estimation tools we would like to provide further intuition into the optimization of the expected revenue by setting the optimal bid increment using a numerical example:

Numerical Example 2 – Suppose an auction has five items on sale, and the valuations for the highest six bidders is as follows:

Bidders	**1**	**2**	**3**	**4**	**5**	**6**
Valuation	110	111	112	115	121	132

Let B_0 = \$110. Without knowing the actual valuation of each bidder, it would be sufficient if the auctioneer knew the number of bidders with a valuation above a certain bidding level. With the set of bidders above, the following table summarizes the relevant information for the auctioneer.

Bid Levels	**110**	**115**	**120**	**125**	**130**	**135**
No. of Bidder	5	2	2	1	1	0

Therefore, if the marginal bid is 110, a bid increment can be determined that gives the optimum revenue.

K	**1**	**2**	**3**	**4**	**5**	**6**	**11**	**22**	**>22**
No. of Bidder with valuation greater than or equal to $B_0 + k$	5	4	3	3	3	2	2	1	0
P	5/6	4/6	3/6	3/6	3/6	2/6	2/6	1/6	0
Expected revenue *E(R)*	551.31	552.59	552.67	553.56	554.45	552.94	555.40	554.39	550

From the table above, it is clear that setting the bid increment at $ 11 would yield the optimal expected revenue.

The example presented above, assumes that the auctioneer knows the distribution of bidders' valuations *a priori.* It is common in auction theory to assume some known continuous distribution to which consumer valuations are said to belong. In this study we make use of automated software data collecting agents to track real online auctions, and in doing so build historical repositories of bid patterns that permit the empirical estimation of *p* values, as well as for making informed distributional assumptions regarding the bidders' valuations. In the next section we explain how we collect data from real online auctions, and use it for deriving empirical distributions of the *p* values. We are also able to fit uniform distributions to the critical fractile of bidders' valuations obtained from the empirical observations. This, in turn, allows us to analytically obtain the optimal bid increment and revenue.

4. **Applying Analytical Results to Real Auctions**

4.1 Data Collection

An automatic agent was programmed to capture, directly from the website, the html document containing a particular auction's product description, minimum required bid, lot size and current high bidders at frequent intervals of 5-15 minutes. A parsing module developed in Visual Basic was utilized to condense all the information pertinent to a single auction, including all the submitted bids, into a single spreadsheet. We tracked over 150 auctions, however, complete bidding data was available for 65 auctions. The screening process was designed to ensure: a) that there was no sampling loss (due to occasional server breakdowns), and b) that there was sufficient interest in the auction itself, given that some auctions did not attract any bidders. Data collection lasted over a period of 6 months so as to guarantee a large enough sample-size (> 20) for each of the levels of bid increment chosen ($10, and $20). From the data collected we can construct the bidding history of each bidder who participated in each of the 65 auctions.

4.2 Obtaining empirical valuation data

Based on the final bid of each bidder, a valuation is generated, for that bidder, by adding a random number drawn from $U(0, k)$. For the losing bidders using a pedestrian bidding strategy, this is a rational estimate because the final bid offer can be considered a tight lower bound on the consumer's valuation. If the consumer's real valuation is greater than one bid increment above the final bid, then the bidder should have been able to constitute a new bid and either be in the winners list or propel the auction to a higher bidding level. For the winners of the auction, these estimates are conservative, because as we saw in section 3, that auction can stop with the N winners not necessarily bidding all the way to close to their real valuation. However, for the purpose of estimating the effect on auctioneers' revenue they may be adequate. Table 2 below gives a list of bids on one of the auctions that we observed, and the consumer valuations inferred from these bids.

Auction 10-36 (*all data coded*)										
Lot Size = 17, Opening bid = 9, Total no. of bidders = 40, Bid increment = 10										
	Winners									
Bidder No.	1	2	3	4	5	6	7	8	9	10
Final Bid	109	109	109	109	99	99	99	99	99	99
Valuations	119	116	115	113	109	108	107	106	105	104
								Losers		
Bidder No.	11	12	13	14	15	16	17	18 - 27	28 - 39	40-45
Final Bid	99	99	99	99	99	99	99	89	79	69
Valuations	104	103	103	102	101	100	99	98-89	88-79	74

Table 2: Estimating bidder valuations from empirically observed final bids

4.3. Optimizing auction revenues using empirical distributions

For computing the revenue using the analytical model (equation 1) at a given bid increment (k), we need to estimate the probability p that a bidder's valuation is greater than $B_0 + k$ where B_0 represents the marginal bid level. Consider the example provided in Table 2, where the top $N+1$ valuations range from \$98-\$119, with the marginal bid level $B_0 = 98$. A good estimator of the probability $\hat{p}$ that a bidder's valuation is greater than $B_0 + k$, can be obtained by taking the ratio of n the number of bidders having valuations greater than $B_0 + k$ to the number of bidders having valuations greater than B_0:

$$\hat{p} = \frac{n}{N+1} \tag{4}$$

In this example for a bid increment $k = 3$, $n = 15$. Since, the lot size in this example is 17, $\hat{p}$ = 15/18 = 0.8. The expected revenue that corresponds to this bid increment can now be determined using equation 1. This exercise is carried out for all feasible integer values of k to determine the optimal bid increment. We do not consider the fractional bid increments because a dollar is usually considered as the minimum difference in successive bid increments. The results of this exercise for this example are provided in Table 3.

K ($)	1	2	3	4	**5**	6	7	8	9	10	11
$B_0 + k$ **($)**	99	100	101	102	**103**	104	105	106	107	108	109
N	17	16	15	14	**14**	11	9	8	7	6	5
$\hat{p}$	0.94	0.89	0.83	0.78	**0.78**	0.61	0.50	0.44	0.39	0.33	0.28
Revenue ($)	1,666	1,675	1,678	1,679	**1,682**	1,675	1,673	1,672	1,671	1,671	1,670
K ($)	12	13	14	15	16	17	18	19	20	21	22
$B_0 + k$ **($)**	110	111	112	113	114	115	116	117	118	119	120
N	4	4	4	4	3	3	2	2	2	1	0
$\hat{p}$	0.22	0.22	0.22	0.22	0.17	0.17	0.11	0.11	0.11	0.06	0.00
Revenue ($)	1,669	1,669	1,670	1,670	1,669	1,669	1,668	1,668	1,668	1,667	1,666

Table 3: Expected revenue computation for various bid increment levels

Observe that the optimal bid increment for this example is $5.

4.4. Optimizing auction revenues using inferred uniform distribution

Our optimization approach is based on the knowledge of the critical fractile of the *N+1* highest valuations, which is contained in the upper-tail of the overall consumer value distribution. Since most of the theoretical distributions are relatively flat in the tail, we conjecture that it may be possible to approximate the critical fractile distribution with a uniform distribution.

From the data we collected, we can compute the number of distinct bidders that placed a bid at or above a given bid level, and thus can construct an implicit demand curve of the auction participants by calculating the number of people that will be willing to purchase a product at any given price.

Since a uniformly distributed value distribution produces a demand curve that is a linearly descending straight line, we can test our hypothesis that the distribution in the critical fractile is uniformly distributed by testing whether the demand curve is indeed a linearly descending straight line. We performed this test by taking the tail data of bids (data for last few bidding cycles) for each tracked auction. We then constructed the demand curves by computing the number of distinct individuals that placed a bid at a given bid level or higher. Figure 2 shows a representative demand curve based on the data for auction represented in Table 2 earlier. As the figure indicates we get a fairly straight-line representation, with acceptable distortions in real data.

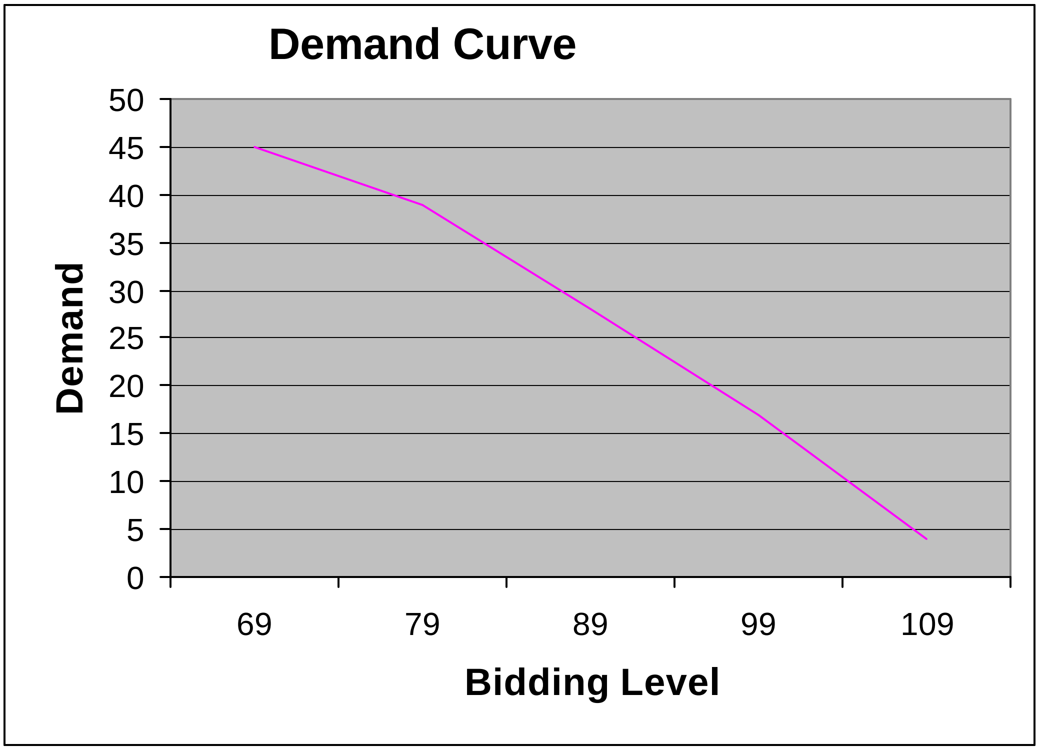

Figure 2 – Downward sloping demand curve

We tested the fit of the straight line by using a simple linear regression with demand as dependent variable and bid level as independent variable as follows:

$$\text{Demand (Consumers willing to purchase at a given price)} = \alpha + (\beta * \text{price}) + \varepsilon \qquad (5)$$

For the particular example described above we obtained a $\beta = -1.4^{***}$ and a R-square of 98.5%. We deployed the above-mentioned procedure for all the auctions that we tracked and found extremely good support for our contention that the critical fractile valuations can be approximated by a uniform distribution. Table 4 below presents the R-Square values of tests performed on a percentile basis. As is evident, more than 94 percent of the auctions had an R-square of 90 and above and close to 60% of the auction had a fit with an R-square of 96 and above.

R-SQUARE	PERCENT ABOVE	R-SQUARE	PERCENT ABOVE
88	100.0%	96	59.3%
90	94.2%	98	36.0%
92	83.7%	99	15.1%
94	72.1%		

Table 4– Regression results from testing for uniformly distributed tail-end valuations

The regression equation (5) representing a downward sloping demand curve can be used to estimate $\hat{p}$. We can also compute the parameters of interest such as B_0, and V_{max} using the regression equation. For example, to determine the highest valuation of V_{max}, we can solve for the bid level at which the demand is zero and, similarly, the valuation of the marginal bidder B_0 can be determined by solving for the bid level at which the demand is equal to the $N+1$ where N represents the lot size.

Using a Uniform[B_0, $V_{\max}$] distribution for the critical fractile, for a given bid increment k, the probability that a bidder will have a valuation greater than $B_0 + k$ is given by:

$$p = \frac{V_{\max} - B_0 - K}{V_{\max} - B_0} \text{ and } 1 - p = \frac{K}{V - B} \tag{6}$$

Substituting (6) and differentiating equation (1), $E(R)$ with respect to k, we obtain

$$k^* = \frac{2\sum_{i=0}^{N} i\left(V_{\max} - B_0 - k\right)^i}{\sum_{i=0}^{N} i^2\left(V_{\max} - B_0 - k\right)^{i-1}} \tag{7}$$

Therefore, by estimating the critical fractile using an uniform distribution, we are able to use equation (7) and interactively determine the optimal value of the bid increment and the associated maximized revenue.

Equation (7) also provides insights into the structure of optimal bid increment with respect to the number of items on sale. This result is formalized in the following proposition:

Proposition 1: *Assuming the tail end of consumers' valuations are distributed $U[B_0, V_{max}]$, as suggested by the empirical evidence, the optimal bid increment decreases as the lot size increases .*

Proof – As N increases, $\sum_{i=0}^{N} i^2$ grows at a much larger rate than $\sum_{i=0}^{N} i$. Since $\sum_{i=0}^{N} i^2$ is in the denominator of equation (5), k decreases as N increases. Q.E.D.

5. Results and Validation

5.1 Validation via simulation

Bapna et al. (2001b) developed a robust simulation tool for replicating Yankee auction runs. The input to the simulation is the list of all final bids placed by all bidders in a real auction. A valuation is generated for each bidder by using the process described in section 4.2. The simulation program reads these valuations from a file, and for any given bid increment it can generate a specified number of runs representing different possibilities of bidding outcome as represented by the combinatorial possibilities illustrated in Section 3.

This simulation tool has proven to be very robust; comprehensive tests showed that the bidding streams, generated by the simulation, are samples of the same distribution as the real auction occurrence. It provides a very effective means to assess the impact of the choice of auction parameters. For example, in Bapna et al (2001b), we discovered that, with respect to the bid increment, the revenue function is multi-modal, i.e., one can often obtain optimal values for the revenue at multiple choices of the bid increment. We also observed that in most cases the auctioneers chose bid increments that were far from the optimal choices. Due to its flexibility and robustness, we chose to use this simulation tool to validate the theoretical results of this paper. In particular, we wish to compare two theoretically estimated measures of optimal revenues with two simulated scenarios.

The two theoretically estimated quantities are:

- The optimal revenue $R(k_{TO})$ obtained from expression (3), when the bid increment k_{TO} is determined based on the estimated $\hat{p}$ represented in equation (4) using the empirical data directly. This is done by numerically evaluating equation (3) for all feasible choices of bid increments.
- The optimal revenue $R(k_{TU})$ obtained from expression (3), when the bid increment k_{TU} is chosen by using equation (7), i.e., the underlying assumption is that the critical fractile comes from a uniform distribution.

Each one of these are compared to the following two scenarios obtained from using the simulation tool:

- The maximum revenue $R_S(k_{SO})$ obtained by using the simulation tool to computationally evaluate each feasible bid increment and identifying the optimal one (k_{SO}). This comparison is intended to measure the accuracy of our analytical approach in deriving the optimal revenue.
- The optimal revenue generated by the simulation tool using the corresponding optimal bid increment, $R_S(k_{TO})$ or $R_S(k_{TU})$. This comparison test is intended to measure the impact of using the optimal bid increment derived by each theoretical model.

Recall that the auctioneers' primary interest is in maximizing revenue, and we wish to test the accuracy of the analytical model using our simulation tool under two different levels of information intensity (empirical and Uniform). Our focus is on the examination of the equivalence of the revenue structures. We do this in a two-stage process beginning first by exploring simple percentage deviations between the theoretical and simulated revenues, followed by a rigorous trace drive simulation validation procedure suggested by Kleijnen et al (1996, 1998).

5.2 Exploratory Data Analysis

Based on the data acquired from real-world auctions as discussed in section 4, we ran 31 replications for each auction at each of the three revenue cases: $R_S(k_{SO})$, $R_S(k_{TO})$, and $R_S(k_{TU})$. We used the average of the 31 runs for each of the 65 auctions under each revenue case to compare to the theoretical cases $R(k_{TO})$ and $R(k_{TU})$. The results for the

65 auctions are presented in the following table in terms of the mean percentage deviation from the corresponding simulated case.

$abs(R(k_{TO}) - R_S(k_{SO})) / R_S(k_{SO})$	0.48%
$abs(R(k_{TO}) - R_S(k_{TO})) / R_S(k_{TO})$	0.47%
$abs(R(k_{TU}) - R_S(k_{SO})) / R_S(k_{SO})$	5.85%
$abs(R(k_{TU}) - R_S(k_{TU})) / R_S(k_{TU})$	3.74%

Table 5 – Mean Percentage Deviations: Theoretical vs. Simulated Revenues

The first row indicates that when using real valuations in the theoretical model, we achieve an average deviation of 0.48% from the simulation results using the best possible bid increment. This leads us to form an initial belief that given the empirically derived probability *p*, the theoretically optimal revenue is structurally similar to the simulated maximum revenue.

The theoretically determined optimal bid increment is often different from the bid increment at which the simulated revenue is maximized. Therefore, another issue worth exploring is whether the simulated revenue at the theoretically optimal bid increment is significantly lower than the estimated theoretical revenue. In row two of Table 5 we observe that while applying the theoretically derived bid increment in the simulation, the deviation is 0.47%. This result further adds a measure of robustness to our initial belief that the analytically determined bid increment is the one that maximizes the auctioneers' revenue. Both rows 1 and 2 of Table 5 were examined under the high information intensity case when the auctioneer had empirical distributions of consumer's valuations.

However, in many instances an auctioneer may have a more limited set of information regarding the consumer valuations. For example, they may only have estimates of lower and upper bounds on expected prices that consumers might be willing to pay. Based on our findings in Section 4.4, the revenue comparisons in rows 3 and 4 of Table 5 are conducted using a uniform distribution for the consumer valuations. It is clear that with lower information intensity, as compared to the empirical knowledge case, the accuracy of the theoretical model is not as high as the mean percentage deviation increases to 5.85%, but is still within a very reasonable range.

5.3 Trace Driven Validation

To more rigorously validate our analytical results against our simulation model, we adopted a trace-driven validation technique, as proposed by Kleijnen et al (1996, 1998). The idea is to compare results of two temporal streams, in our case a theoretical stream of auction revenue values, and a simulated stream. We are able to do this because our analysis captures the temporal dimensions of the auctions we tracked. They varied for each auction along the dimensions of lot size, price magnitude, and number of participating bidders. *The trace driven validation approach tests if the two streams have identical means and variances*. While details of the approach are beyond the scope of this paper, we present the main methodology below.

The validation procedure for comparing the revenues using the 65 auctions involves using the following 2 regression models.

Regression 1 is of the form $R_S = b_0 + b_1 R_T$, where R_S is the revenue estimates from the simulation model and R_T is the revenue predicted from the theoretical model. If the theoretical model is a good predictor of the revenue then we should expect a regression $\boldsymbol{R^2}$ of close to 1 and the slope estimate b_1 should be close to 1.

Regression 2 is of the form $R_S - R_T = b_0 + b_1(R_S + R_T)$. If the variances of the two estimates are the same then we should expect the regression $\boldsymbol{R^2}$ to be close to 0. Essentially, the first regression is a test of equivalence of means, while the second one is a test of equivalence of variances [see Kleijnen et al (1996, 1998) for details]. Collectively, these 2 regression models provide support for the hypothesis that "*The mean and variance of the distribution produced by theoretical estimation are the same as those produced by the simulation model*." The following table summarizes the results of the regression tests for the four cases we have:

$\mathbf{R_S}$	$\mathbf{R_T}$	***R^2 from Regression 1***	***b_1 from Regression 1***	***R^2 from Regression 2***
$R_S(k_{SO})$	$R(k_{TO})$	0.999	1.001	0.008
$R_S(k_{TO})$	$R(k_{TO})$	0.999	1.001	0.006
$R_S(k_{SO})$	$R(k_{TU})$	0.995	1.026	0.064
$R_S(k_{TU})$	$R(k_{TU})$	0.997	1.031	0.173

Table 6 – Results of the 2-regression validation approach of Kleijnen et al

There is overwhelming support for the equivalence of means for all revenue comparisons. There is also overwhelming support for the equivalence of variances for the comparisons involving theoretical model with empirical observed valuations. The support goes down (R^2 goes up in the fourth column of Table 6) for the comparisons of the theoretical model with assumed uniform valuations. This is not surprising because under the uniform assumption we have more loss of information since individual valuations are not used, and therefore more variability. As Kleijnen et al. (1996, 1998) point out this is the case when the two distributions do not have equivalent variances. In other words, though the central tendencies are equivalent, the distributions themselves are not. Kleijnen et al. (1998) suggest that when the results of second regression are non conforming to the hypothesis of equal variance, individual tests or pairwise t-test should be used to further investigate the properties.

Therefore, at the individual auction level, we test whether the theoretical revenue was equivalent to the mean of the 31 simulation runs for that auction. We used a standard *t*-test of difference of means. The following is the number of individual auctions in which we failed to reject revenue equivalence (theoretical vs. simulated) at the 10% significance level, for the same comparisons above:

$\mathbf{R_S}$	$\mathbf{R_T}$	*No. auctions we failed to reject revenue equivalence at 10% level*
$R_S(k_{SO})$	$R(k_{TO})$	64/65
$R_S(k_{TO})$	$R(k_{TO})$	61/65
$R_S(k_{SO})$	$R(k_{TU})$	45/65
$R_S(k_{TU})$	$R(k_{TU})$	44/65

Table 7 – Individual auction validation

Table 7 indicates that in a large number of cases even with limited information (assuming uniform critical fractile valuations) we can at least predict the expected revenue. Overall, the results indicate that if the auctioneer has access to empirical consumer valuations, the theoretical model can be used with a high degree of accuracy to determine the optimal bid increment for the auction. If these valuations are not available, but the customer demand of the critical fractile can still be inferred through a uniform distribution, the bid increment determined by the theoretical model can still yield very good results. From Table 7, revenues generated by using such bid increments can be very accurate 2/3 of the time. From the results in Table 5, on the average, these the revenues obtained by using these bid increments are off by less than 4%.

The overall significance of our findings from the exploratory data analysis and the trace driven validation is that there appears to be strong support for the use of the theoretically computed bid increment in the design of multi unit online auction. *By choosing the analytically determined bid increment auctioneers can expect to maximize their revenues.* There is also an interesting cost-benefit tradeoff between the cost of acquiring the information necessary to plug into the model (the *p* estimate) and the corresponding revenue benefit.

7. Concluding Remarks

In this paper we provide approaches at various levels of modeling and data abstractions to address the problem of optimal auction design. The actual applicability of each approach might depend upon the specific product and/or market information availability. Figure 3 provides a conceptual framework for integrating the insights provided in this paper into the design of online Yankee auctions. Beginning with tracking real online auctions to obtain estimates of consumer valuations and the demand curve, we develop a theoretical model of the revenue generation process. The optimization of this model requires estimates of the probabilities that bidders will bid at the next higher bid level, above the marginal bid.

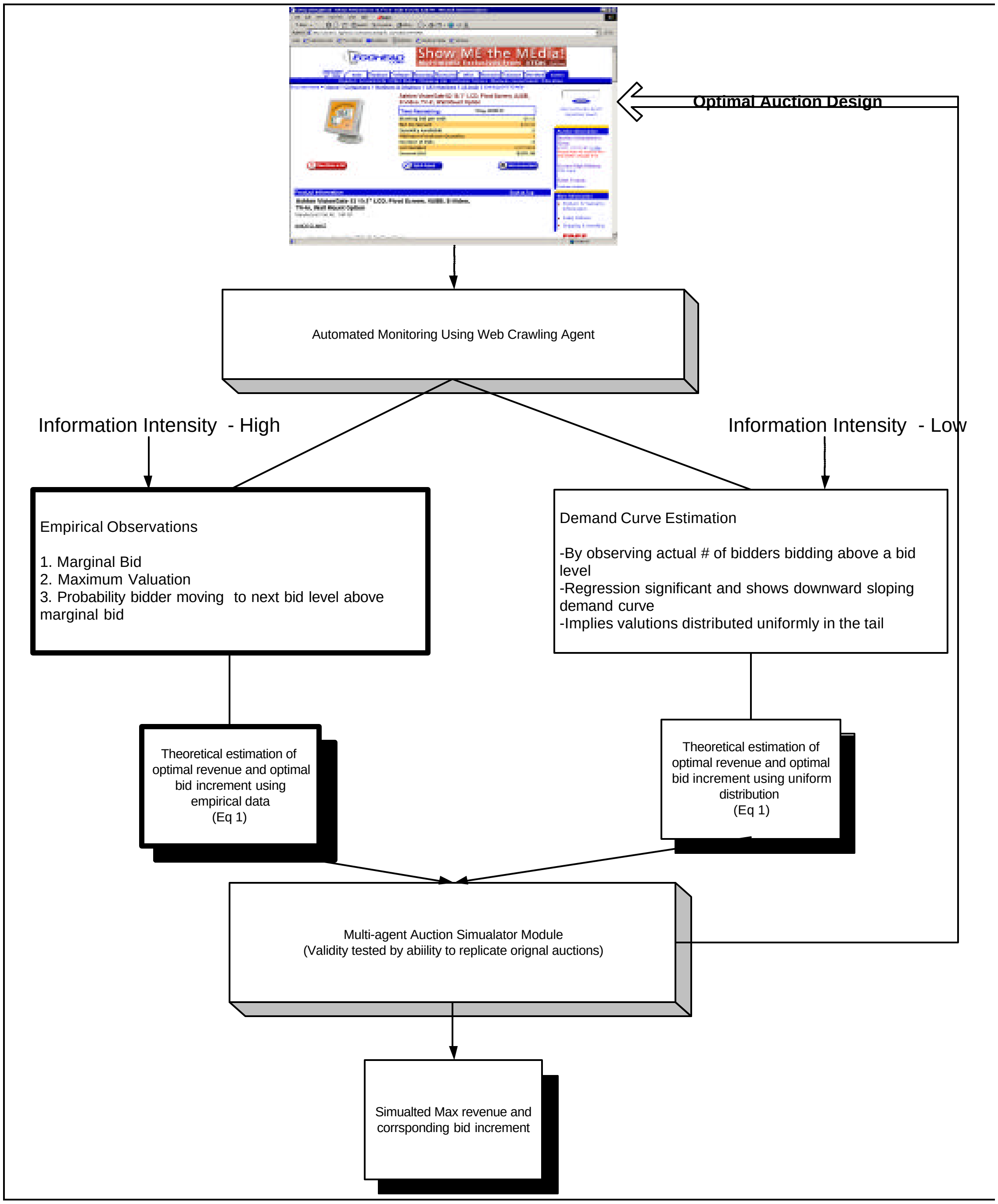

Figure 3 - Research Metascape Depicting Varying Levels of Abstractions and Information Intensity

To summarize, we developed a theoretical basis for determining the optimal bid increment setting for online Yankee auctions, an important emerging channel in the supply chain. The theory relies on the information that the auctioneer has regarding the bidder valuations. On the one extreme, relying on the empirical distribution of bidder valuations, our theoretical model can be used to determine optimal bid increment setting. The results of the theoretical model were tested using a simulation tool and we observed that the theoretically optimal bid increment yields maximal revenues. Similar, but not as strong results are obtained if we rely on a Uniform distributional characterization of the bidder valuation in the critical fractile, as opposed to an empirical one. It should be noted from an information cost perspective, the latter is far easier to ascertain than the former, hence the corresponding slight weakness in the result. In other words, depending on the amount of information available to the auctioneers a low or high information intensity track could be pursued and optimal bid increments can be derived. The simulation tool can be used to explore a variety of scenarios and policies and can be used as a test bed to improve the design and/or parameterization of a given auction.

From a practical perspective we expect auctioneers to have some prior estimates of the marginal (B_0) and the maximum valuations that consumers could be expected to have for a product being auctioned. These could be obtained from historical distributional data, price comparison agents and other 3rd party sources. They could use these initial estimates to compute the optimal bid increment and initiate the auction. As the auction progresses, it may be necessary that the original estimates may need to be revised, or quite simply a more accurate estimation of B_0 may become available to the auctioneers. The revised estimate would capture the dynamics of that particular auction and the bidding strategies being employed in it. It would then behold the auctioneer to reapply our, computationally efficient, optimization procedure with the revised parameter estimates, and dynamically adjust the auction parameters, as the auction progresses. For instance, an auction could start with a bid increment of $20 and switch to a lower bid increment of say $10 as it appears to be closing.

In future work we will focus our attentions to other auction parameters in the design of online auctions that require optimization. These include, but are no limited to, the auction duration and the lot size among others.

References

1. Banks, J. and Carson, J. S. (1984). *Discrete-Event System Simulation*. Prentice-Hall, Inc., Englewwod Cliffs, New Jersey.
2. Bapna, R. (1999). *Economic and Experimental Analysis and Design of Auction Based On-line Mercantile Processes*. Unpublished doctoral dissertation, The University of Connecticut, Storrs, CT.
3. Bapna, R., Goes, P., and Gupta, A. (2000). A Theoretical and Empirical Investigation of Multi-item On-line Auctions. *Information Technology and Management* 1(2), 1-23.
4. Bapna, R., Goes, P., and Gupta, A. (2001a). Online Auctions: Insights and Analysis. *Communications of the ACM* 44(11), 42-50.
5. Bapna, R., Goes, P., and Gupta, A. (2001b). Simulating Online Yankee Auctions to Optimize Sellers Revenue. In *Proceedings of the HICCS-34 Conference*, CD-ROM.

6. Bulow, J., and Roberts, J. (1989). The Simple Economics of Optimal Auctions. *Journal of Political Economy* 7(5), 1060-1090.

7. Easley, R. and Tenorio R. (1999). Bidding Strategies in Internet Yankee Auctions. *Working Paper,* Notre Dame University.

8. Kleijnen, J. P.C., Bettonvil, B. and van Groenendaal, W. (1996). Validation of Simulation Models: Regression Analysis Revisited. In *Proceedings of the Winter Simulation Conference,* 352-359.

9. Kleijnen, J. P.C., Bettonvil, B. and van Groenendaal, W. (1998). Validation of Trace-Driven Simulation Models: a Novel Regression test. *Management Science* 44(6), 812-819.

10. Lummus, R. R. and Vokurka R. J. (1999). Managing the demand chain through managing the information flow: Capturing the "Moments of Information. *Production and Inventory Management Journal* 40(1), 16-20.

11. McAfee, R. P., and McMillan, J. (1987). Auctions and Bidding. *Journal of Economic Literature* 25, 699-738.

12. Milgrom, P. (1989). Auctions and Bidding: A Primer. *Journal of Economic Perspectives* 3, 3-22.

13. Milgrom, P. and Weber, R. (1982). A Theory of Auctions and Competitive Bidding. *Econometrica* 50, 1089-1122.

14. Nautz, D. and Wolfstetter, E. (1997). Bid Shading and Risk Aversion in Multi-unit Auctions with Many Bidders. *Economics Letters* 56(2), 195-200.

15. Paarsch, H. J. (1992). Deciding Between Common Values and Private Value Paradigms in Empirical Models of Auctions. *Journal of Econometrics*, 51, 191 – 215.

16. Rothkopf, M. H., and Harstad, R. M. (1994a). Modeling Competitive Bidding: A Critical Essay. *Management Science* 40(3), 364-384.

17. Rothkopf, M. H., Harstad, R. M. (1994b). On the Role of Discrete Bid Levels in Oral Auctions. *European Journal of Operations Research* 74, 572-581.