The Institute for Integrating Statistics in Decision Sciences

Technical Report TR-2017-5

**Modeling First Bid in Retail Secondary Market Online Auctions:
A Bayesian approach**

Babak Zafari
Mathematics and Science Division
Babson College, USA

Refik Soyer
Department of Decision Sciences
The George Washington University, USA

Modeling First Bid in Retail Secondary Market Online Auctions: A Bayesian Approach

Babak Zafari[a] , Refik Soyer [b]

[a]Mathematics and Science Division, Babson College, Babson Park, MA 02457, USA - bzafari@babson.edu

[b]Department of Decision Sciences, School of Business, The George Washington University, Washington, DC 20052, USA

Abstract

We propose a Bayesian framework to model bid placement time in retail secondary market online B2B auctions. In doing so, we propose a Bayesian beta regression model to predict the first bidder and time to first bid, and a dynamic probit model to analyze participation. In our development, we consider both auction-specific and bidder-specific explanatory variables. While we primarily focus on the predictive performance of the models, we also discuss how auction features and bidders' heterogeneity could affect the bid timings as well as auction participation. We illustrate implementation of our models by applying to actual auction data and discuss additional insights provided by the Bayesian approach which can benefit auctioneers.

Keywords: Dynamic Probit, Beta Regression, B2B Auction, Bidder Behavior

1 Introduction

E-commerce has transformed trading markets for both businesses and consumers, mainly by how online auctions are conducted. Beyond changes to business-to-consumer (B2C) and business-to-business (B2B) auctions, advances in e-commerce have enabled users to sell their merchandise directly to other users (consumer-to-consumer, or C2C, auctions). The literature has extensively explored challenges with optimal auction design, optimal bidder behaviors, and the role of major factors in these processes. In this paper, we study two critical events in the online auction process: bidders' first bid and their decision to participate. Specifically, we investigate whether, and how, different strategies are adopted over the time and whether future bidding behaviors can be predicted.

The first bid, both through its size and its arrival time, plays a crucial role in an auction. Its influence on the number of bidders and final price was noted by Bajari and Hortacsu (2003). These findings were echoed by Simonsohn and Ariely (2008) who pointed out that early bidding is a "necessary condition for herding" and that the first bid -by virtue of the offered price and time - can influence the number of bids in an auction. Ku et al. (2006) noted how lower starting prices lower entry barriers to attract more bidders, finding that early bidders tend to bid more in an auction, ultimately increasing the final price. Li et al. (2009), meanwhile, discussed how early bids signal information about the item's value, which may increase the final price and the broader auction's recovery rate. Most of these studies, however, are in the context of B2C or C2C auction platforms. Our study focuses on the formation and timing of first bids in the retail secondary market of online B2B auctions.

The B2B auction market is an important component of the retail industry's cost-recovery efforts for returned and liquidated goods. According to National Retail Federation (2015) Consumer's Return Report, total merchandise returns accounted for almost $260.5 billion (8% of total sales) in lost sales for U.S. retailers in 2015. While some retailers sell their store returns to a small group of brokers, others have launched their own online marketplaces to liquidate returns or rely on wholesale logistics companies to liquidate their products. These companies post information on their online platform about a pallet's item quantity and retail value, offering few details on the contents. Given the nature of the products, they are sold "as-is" with no defined return policy. This ultimately creates uncertainty about the items' condition, only compounded by a lack of seller feedback.

In marketplaces where buyers are uncertain about an item's valuation, they are likely to rely on other auctions' outcomes and other bidders' behavior (Bajari and Hortacsu (2003), Li et al. (2009)). These signals can help reduce bidder uncertainty and prove influential in bids through similar auctions. In Pilehvar et al. (2016), the authors investigated the impact of bidders' internal and external reference prices on the first bid and explored how it is moderated by bidders' heterogeneity, which is closely related to their experience and participation level in concurrent auctions. Empirical analyses of online auction marketplaces such as eBay have similarly shown that experienced bidders tend to place their bid closer to the beginning or end of the auction (Borle et al. (2006), Wilcox (2000)), while less-experienced bidders may over-value an auction and suffer from the "winner's curse," paying more than an item's true value (Bajari and Hortacsu, 2003). This could be because less-experienced bidders are still learning about these auctions, while experienced bidders have adopted a strategy. At the same time, bidders' experience may influence them to "shade" their bids - offer less than they value the item - or avoid early bidding, staving off others in an effort to keep the final price lower. The lack of information on pallets' value and contents in the secondary market increases the importance of this experience.

Modeling bid times in B2B auctions is an important step in studying auction dynamics. As mentioned, the effect of the first bid on the final outcome of the auctions has been studied in multiple works in the literature. However, most of the studies are in the context of B2C auctions. Given the unique characteristics of the secondary B2B auction platform, the first bid only grows in importance in this environment. In Pilehvar et al. (2016), authors studied first bid in B2B platforms and how it is associated with the final price of the auction. Consistent with the literature, they showed how first bid, through its arrival time and its size, can influence auction's dynamic and final price. But as also discussed in their paper, arrival time and the size of the bids are both functions of the bidder's characteristics and auction's features. Furthermore, facing the secondary market conditions, auctioneers can benefit from knowing who will participate and who will start an auction; these factors can change the dynamics and final outcome. Developing a greater understanding of these dynamics also carries important managerial implications. In general, an auctioneer is selling many of these items under a consignment agreement under which the liquidator must move as quickly as possible to avoid inventory problems and be able to pay retailers. Secondary market auctions also tend to have a low recovery rate [1] ($0.261 in our dataset), thus a quick sale at the highest possible final price mitigates the impact of unsuccessful auctions.

In this work, we address the central issues of who will place the first bid at what time and who will participate in a given auction. More specifically, we answer the following questions: How does information available across auctions and bidders' past activities influence their strategy? How can we use this information to make predictions about future auctions? In so doing, we test for changes in the significance and the dynamics of the factors over the time. Ultimately, this can help auctioneers revise operational strategies and design more efficient marketplaces that adapt with, and benefit from, these changes.

A number of works in the literature consider the use of Bayesian methods in modeling online auctions. A key advantage to this methodology is that it allows for information sharing across auctions, which proves useful when some auctions have received few bids. In Park and Bradlow (2005), the authors proposed a Bayesian framework to capture key behavioral components of an Internet auction such as who bids, when they bid, and how much they bid at the auction level. In an extension to this work, Bradlow and Park (2007) considered a latent competing set of bidders, using a Bayesian data augmentation method to develop a Bayesian record-breaking model for bidders' changing auction valuations from bid to bid. Li et al. (2009) implemented a hierarchical Bayesian framework based on the developed model of Park and Bradlow (2005) to study how different auction features influenced bidders' perception toward the auction's uncertainty. The authors have suggested a need for studying the dynamics of bidding behavior over time to develop stronger models.

A central contribution of this paper is the development of Bayesian models to study auctions at the bid level, incorporating participant bidding history and real-time activity. In this work, we extend the use of Bayesian beta regression and dynamic models to understand and predict bidder behavior. Few, if any, papers to the best of our knowledge have studied auctions and bidding behaviors over time using these models. The developed models also

[1] Recovery rate is defined as the final price, divided by the pallet's retail value.

can help managers identify bidders with a low or high probability of participation or help direct bidder traffic to less-attractive auctions. Using these models also helps managers identify potential first bidders and obtain a more accurate estimation of first-bid value and consequently final outcome of the auction.

The rest of the paper is organized as follows. In section 2, we propose a Bayesian beta regression model to describe bid arrival times and to predict the first bidder. We also discuss the studied auction platform, the data, and the results of the model. In section 3, we consider a Bayesian dynamic probit model for bidder participation. In section 4, we conclude by discussing limitations of our proposed approaches and possible extensions.

2 Modeling Time to First Bid

In view of our previous discussion on the major role that the first bid plays in auction dynamics, here we propose a model to describe first time to bid in an auction. In doing so, we model (potential) bidders' time to their first bid and identify the first bidder by comparing the distributions of bidders' bid time. In defining the time of the first bid, it is more appropriate to treat time relative to the total auction duration. Thus,the bid time is defined as a ratio of the time passed (since the auction has started) to the total time of the auction, that is, $\frac{TimePassed}{TotalTime}$. We refer to this relative bid time, which takes values in $(0, 1)$, as the bid time in our development.

We let random variable Y_{ij} denote the time of first bid of bidder i in auction j. Since Y_{ij} takes values in $(0, 1)$, we assume that it follows beta density given by

$$p(y_{ij|a_{ij},b_{ij}} \mid a_{ij}, b_{ij}) = \frac{\Gamma(a_{ij} + b_{ij})}{\Gamma(a_{ij})\Gamma(b_{ij})} y_{ij}^{a_{ij}-1}(1 - y_{ij})^{b_{ij}-1}, 0 < y_{ij} < 1$$

where $a_{ij} > 0$, $b_{ij} > 0$ and $\Gamma(.)$ is the gamma function. We denote the beta density as $Y_{ij}|a_{ij}, b_{ij} \sim Beta(a_{ij}, b_{ij})$ and reparametrize it as $a_{ij} = \phi\mu_{ij}$, $b_{ij} = \phi(1 - \mu_{ij})$ where

$$E[Y_{ij}] = \mu_{ij}, \text{ and } V[Y_{ij}] = \frac{\mu_{ij}(1 - \mu_{ij})}{1 + \phi},$$

are the mean and variance of Y_{ij}, respectively. Thus, we can denote beta density as $Y_{ij}|\mu_{ij}, \phi \sim Beta(\mu_{ij}, \phi)$.

The reparamerization of the beta density provides us with flexibility in considering bidder and/or auction-specific covariates in our model. Specifically, we can write

$$h(\mu_{ij}) = \beta X_{ij},$$

using a link function $h(.)$ where $\beta = (\beta_1, ..., \beta_k)$ is a vector of k regression parameters and $X_{ij} = (X_{1,ij}, ..., X_{k,ij})$ is a vector of k explanatory variables corresponding to bidder i in j^{th} auction. In our development, we consider both *probit* and *logit* link functions and compare their performance. We refer to this beta regression model with fixed precision parameter ϕ as "Model 1". In Model 1 we assume a multivariate normal prior on β and a gamma prior on the precision ϕ where β and ϕ are assumed independent apriori. Using the independence assumption of the observations, the likelihood function will have the following form:

$$\mathcal{L}(\beta, \phi) = \prod_{i=1}^{n} \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1-\mu)\phi)} y_i^{\mu\phi-1}(1-y_i)^{(1-\mu)\phi-1}$$

We can use Gibbs sampling to generate MCMC samples from poster distribution of $p(\beta, \phi \mid Y)$ by repeatedly sampling from full conditional distribution of $p(\beta \mid \phi, Y) \propto L(\beta, \phi)p(\beta)$ and $p(\phi \mid \beta, Y) \propto L(\beta, \phi)p(\phi)$. In this development, we use normal priors for β's and gamma prior for ϕ.

As an alternative model, we relax the homogeneity assumption on the precision parameter by allowing it to change with the covariates. Specifically, we reparameterize the precision as ϕ_{ij} and use *log* link function to write

$$log(\phi_{ij}) = \gamma X_{ij},$$

where $\gamma = (\gamma_1, ..., \gamma_k)$ is a vector of k parameters. We refer to the heterogeneous precision model as "Model 2" in our development. Specification of Model 2 is completed by assuming independent multivariate normal priors on β and γ vectors.

In both models, the vector of covariates consists of two groups of variables: one group provides information about the auction and the second group explains bidders' heterogeneity. The bidder-specific variables can be divided into two categories: variables related to bidders' prior bidding activities and variables capturing their activities around the time of the focal auction. The dataset and list of the variables will be discussed in the next section.

2.1 Predicting the First Bidder

For each of the models, the next step is then to predict the first bidder in a given auction. In other words, the objective is to identify the bidder that places his/her bid before other bidders among a given set of bidders of an auction.

Let's assume we have a set of n bidders for a focal auction. This is the list of bidders who saw the auction and participated in it. In section 3, we propose a model for bidders' participation where we also look at potential bidders; bidders who saw the focal auction but did not participate in it.

Using the beta regression model of last section, we can obtain the posterior predictive distribution for each bidder's time to first bid in each auction. In other words, we obtain $P(y_{ij}|D)$ where D denotes all observed data. Having obtained all n potential bidders' distributions, we calculate the probability of bidder i having the minimum bid time as following:

$$P(y_{ij_i} < y_{kj_k}, \text{for all } k \neq i \mid D)$$

In other words, this is the probability that bidder i places his/her first bid before everybody else, i.e. probability of being the first bidder. Also, note that $\sum_{i=1}^{n} P(y_{ij_i} < y_{kj_k}, \text{for all } k \neq i \mid D) = 1$ which means that the sum of these probabilities over all bidders of the auction is 1. For a given auction and in each run of MCMC simulations, we label the bidder with lowest bid time as the first bidder. Then by going all MCMC simulation runs, the probability of being the first bidder (for each bidder in a given auction) is estimated by averaging the number of times that the bidder was labeled as the first bidder.

In addition to predicting the first bidder, the use of posterior probabilities ,$P(y_{ij}|D)$, can be extended to obtain the probability distribution of the time of the first bid for each auction. This distribution can be estimated by considering minimum value of $P(y_{ij}|D)$ among the set of bidders (i) of a given auction (j). So while the above-mentioned model help auctioneers identify the first bidder, this model enables them to estimate time of the first bid. Without the loss of the generality, both of the models can also be extended to study next bids in auctions.

2.2 Dataset and Variables

We consider the proprietary dataset used by Pilehvar et al. (2016). A major logistics company specializing in wholesale returns and liquidations runs the studied marketplace, auctioning uninspected returns, open-box items, and excess and salvage consumer electronics that major North American electronics retailers send in pallets to its warehouse. Most of the company's customers (bidders) are re-sellers such as off-price retailers and members of eBay's PowerSeller program, which incentivizes high-volume sellers. The data collected from this auction platform span five years, from 2003 to 2008. Observations are at the bid level, encompassing bid information from 2,000 unique bidders who have participated in more than 11,000 auctions. The average number of bidders per auction is 5.18 and the average number of bids is 8.02. The market's busiest season is the first quarter because of holiday-season returns that begin in late November and continue into January. The rich panel dataset enables us to track bidders' activity over time. In other words, we can measure bidders' overall experience, defined in terms of their number of participated auctions and number of wins or losses. We also can see cross-bidding activities in similar open or recently closed auctions at the time of the bid, which allows us to study their effect on subsequent bidding decisions.

For the purposes of this study, we consider auction- and bidder-related variables. Auction-specific characteristics include per-pallet item quantity and a normalized version of the pallet's declared retail value (i.e., per-unit price of pallet items, determined by dividing retail value by quantity). We also consider the duration of the auction, typically two to three days depending on the day posted. Other auction-specific variables are item condition (return or salvage), day and time of day the auction starts, and retailer brand. We should note that these auctions have no reserve price, with the starting price largely the same.

Our bidder-related variables include their past and current bid activity. These constructs can be divided into two sub-categories. The first is related to the bidder's past activities, where we measure variables such as participation volume, number of wins/losses, time between the focal auction and prior bid, etc. The bidder's experience can be defined in terms of overall participation, number of wins, or number of losses. We define bidder experience as one's number of wins or losses in the last three months. This moving-window approach, which is based on the internal reference price literature in marketing, best represents experience at auction time and accounts for bidder inactivity over a long period of time. The second group of bidder-specific variables is related to bidder activity around the time of the focal auction. These cross-bidding activity variables measure several factors, including the number of similar, overlapping auctions in which the bidder is participating and the number of similar, just-ended auctions in which he or she has bid or won. In order to define overlapping and similar auctions, similar to the work of Chan et al. (2007): we pool all auctions running when the focal auction opens and calculate their mean quantity and retail value. Next, we select auctions within one standard deviation of the focal auction's quantity and retail value. We define these auctions as the "similar auctions" superset, from which the bidder gleans influential market price information. For one, the price and dynamics of these auctions better inform the bidder about the focal auction, which has yet to receive its first bid. Loss or win experience in just-finished similar auctions also might influence bidder participation in similar, overlapping auctions.

In selecting the final set of covariates, we include the following instrumental auction-specific variables: quantity (number of items in the auctioned pallet), value (average monetary value of each item calculated by dividing the pallet estimated retail value divided by its quantity), weekend (a variable showing whether the auction is posted on Friday/Saturday), winter (a variable showing whether the auction is posted on first quarter) and simultaneous auctions (number of similar auctions that are posted at the same time with the focal auction). In addition, we include the following bidder-specific covariates: average TOFB (average time of the first bid placement in the participated auctions in the last three months), first bidder rate (proportion of bidder's participated auctions in the last three months in which the bidder has placed the first bid), winning rate (proportion of bidder's participated auctions in the last three months in which the bidder has won the auction) and just-finished auctions (number of wins in similar auctions that ended after the start of the auction and before placement of the first bid).

2.3 Implementation and Results - First Bidder

Before we move with the analysis, we remove all single-bid auctions (i.e., those with only one participant). This avoids a potential false increase in the model's predictive performance, given that the bidder's probability of being first in these cases is always 1. At each run, we begin the process by sampling 550 auctions, only considering participants' first bids. We then split this sample into two subsamples of 500 and 50 auctions, fitting the model with the larger sample and measuring models' predictive performance with the smaller, holdout sample. We repeat this sampling and testing process 10 times each on models 1 and 2.

In our analysis, we use proper but diffused priors for all parameters in the model. More specifically, for the model with constant precision, we specify gamma priors on ϕ with both shape and scale parameters of 0.01 for each. We assume all covariate coefficients β and γ in both models have independent normal priors with mean 0 and precision 0.01. In modeling $\mu_i's$ using covariates, we tested both probit and logit link functions. Given the functions' very similar performance and our use of probit in the model's first essay, we choose the probit model to stay consistent throughout the paper. Each run of the iterative process for the model with common precision takes seven minutes, which includes simulations and calculation of fit and predictive measures. Each run of the second model takes 12 minutes. All results are based on running a Gibbs sampler with a burn-in sample of 10,000 iterations and collecting 10,000 posterior samples after thinning by five. No convergence issues emerged during the run of the models. In addition to the iterative sampling and processing of the models, we conduct two separate models using the full auction sets, with their convergence diagnostics plots and tables shown in the appendix. The p-values associated with the Geweke diagnostic are large, while the Raftery and Lewis dependence factors are all smaller than 5, suggesting no convergence issues.

Table 1 compares model 1 (i.e., the model with the regression structure for mean parameters and constant precision) and model 2 (i.e., the model with the regression structure for both mean and precision parameters) over the 10 randomly constructed datasets. In doing so, in addition to the Bayesian fit measures, we also compare the models in terms of their accuracy [2]. As the results suggest, model 2 performs better over all the runs, with

[2] Accuracy is defined as the ratio of the auctions in which the model correctly identifies the first bidder.

considerably smaller values of $\overline{D}$ and DIC and higher accuracy in every iteration. In addition, p_D has higher values in model 2 due to extra regression parameters used in modeling the precision parameters. The average accuracy across all runs is 0.36 for model 1 and 0.38 for model 2. Given that the average number of bidders in these auctions is 5.4, the results suggest both models effectively predict the first bidder, though model 2 performs better.

	Model 1				**Model 2**			
Run.	**Dbar**	**pD**	**DIC**	**Accuracy**	**Dbar**	**pD**	**DIC**	**Accuracy**
1	-3783.11	9.96	-3773.14	0.361	-3876.15	17.84	-3858.31	0.372
2	-3720.05	9.92	-3710.13	0.369	-3842.5	17.86	-3824.64	0.380
3	-3599.04	10.03	-3589.01	0.368	-3748.29	17.77	-3730.51	0.382
4	-3438.22	10.04	-3428.17	0.370	-3577.3	17.88	-3559.42	0.385
5	-3673.28	10.08	-3663.2	0.372	-3799.12	17.76	-3781.36	0.387
6	-3739.25	10.04	-3729.21	0.372	-3860.23	17.86	-3842.37	0.387
7	-3917.33	10.02	-3907.31	0.356	-4023.83	17.82	-4006.01	0.370
8	-3738.02	10.00	-3728.02	0.383	-3887.92	17.79	-3870.13	0.399
9	-3693.43	9.98	-3683.45	0.364	-3803.5	17.81	-3785.68	0.374
10	-3747.83	9.97	-3737.85	0.361	-3847.68	17.87	-3829.8	0.374

Table 1: Model Comparisons - Fit Measures over Main Datasets

The next two tables, show the predictive log likelihood and accuracy (Table 2) and log Bayes factor in favor of Model 2 (Table 3) over the holdout sets. Consistent with the fit measure over the main datasets, model 2 outperforms model 1, with higher predictive log likelihood values and higher accuracy. The log predictive Bayes factor values in favor of model 2, moreover, show evidence in support of the model. In terms of accuracy, results are comparable to the results of Table 1, where higher accuracy is achieved by modeling both mean and precision parameters using explanatory variables. Note that the average number of bidders in holdout auctions is 5.5.

	Model 1		**Model 2**	
Run.	**Pred Log Likelihood**	**Accuracy**	**Pred Log Likelihood**	**Accuracy**
1	202.39	0.358	211.59	0.366
2	185.34	0.386	191.62	0.401
3	180.36	0.423	181.36	0.441
4	201.32	0.359	208.38	0.377
5	218.88	0.372	226.89	0.389
6	227.39	0.396	235.13	0.404
7	135.54	0.380	140.61	0.393
8	151.36	0.387	157.22	0.409
9	200.28	0.343	205.54	0.353
10	185.62	0.318	188.74	0.326

Table 2: Model Comparisons - Predictive Measures over Holdout Datasets

Run	1	2	3	4	5	6	7	8	9	10
Log Bayes Factor	9.20	6.28	1.00	7.06	8.01	7.74	5.07	5.86	5.26	3.12

Table 3: Log Predictive Bayes Factor in Favor of Model 2

In order to illustrate how the model predicts the first bidder, let's choose an auction with two participants. Figure 1 shows posterior distributions of the timing of bidders' first bid. The plot on the right is generated by plugging posterior mean of estimated shape parameters. The posterior mean of parameters for bidder1 (shown in red) are $a = 1.19$ and $b = 0.51$ (i.e., $\mu = 0.69$ and $\phi = 1.71$) and $a = 3.67$ and $b = 0.64$ for bidder2 (i.e., $\mu = 0.85$ and $\phi = 4.32$). As the results suggest, bidder1 is more likely to place the first bid earlier than bidder2 ($\mu_1 = 0.69 < \mu_1 = 0.85$). Using the model, the probability of bidder 1 placing the first bid before bidder2 is estimated at 0.65. The dataset shows the actual bid time of bidder1 is 0.88 and bidder2 bid's time is 0.96, showing bidder1 was, in fact, first. As discussed at the end of section 2.1, the use of these estimations can be extend to obtain the probability of TOFB for the auction. In doing so, at each point, we consider the minimum value of the curves to obtain the overall probability of the TOFB for this auction.

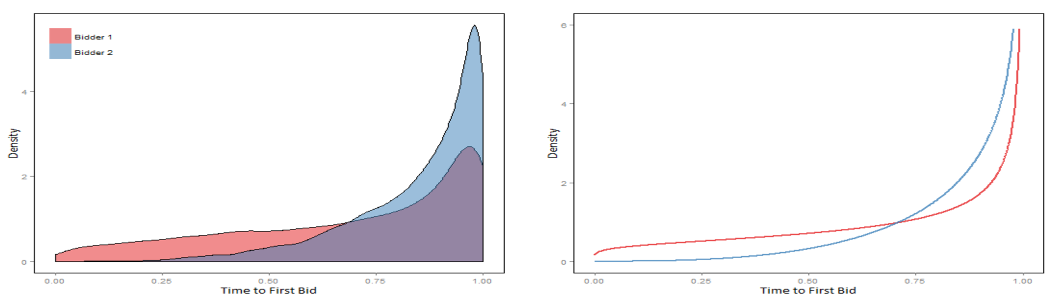

Figure 1: Posterior distribution of TOFB for the two participated bidders of the auction (left) and posterior predictive distribution of their TOFB (right)

We also run both models over the full set of auctions (i.e. no sampling) to fully capture the effect of different variables. Table 4 shows posterior parameter summaries of model 2. As the results suggest, the two most important factors that affect (i.e., delay) first-bid arrival time are the bidder's average time of first-bid placements in the last three months and the number of wins in just-finished auctions. The average time over bidding history suggests a behavior some bidders practice in bidding late or early; an observation consistent with prior findings and our discussions regarding bid placement difference among experienced and inexperienced bidders. Further, bidders who just won in some auctions show less interest in starting another right away. We also find that the more a bidder has placed the first bid in prior auctions, the earlier he/she tends to place the first bid in the focal auction. This bidding behavior again demonstrates the value of studying bidders over time. In addition, the number of wins and participation volume in other open auctions do not seem to have a significant effect on bid arrival time.

Parameter	Mean	StDev.	95% CCI	Parameter	Mean	StDev.	95% CCI
$\beta_{Ave.TOFB}$	1.035	0.016	(1.003,1.067)	$\gamma_{Ave.TOFB}$	0.392	0.026	(0.342,0.445)
$\beta_{FB.Rate}$	-0.623	0.029	(-0.681,-0.565)	$\gamma_{FB.Rate}$	-0.084	0.041	(-0.168,-0.006)
β_{NJFA}	1.147	0.177	(0.800,1.498)	γ_{NJFA}	1.261	0.341	(0.571,1.921)
β_{NOA}	-0.158	0.067	(-0.291,0.025)	γ_{NOA}	0.050	0.115	(-0.173,0.273)
β_{Q}	-0.032	0.070	(-0.106,0.168)	γ_{Q}	-0.790	0.105	(-0.995,-0.585)
$\beta_{Weekend}$	0.115	0.038	(-0.040,0.191)	$\gamma_{Weekend}$	-0.046	0.019	(-0.084,-0.007)
β_{Win3}	0.053	0.039	(-0.024,0.130)	γ_{Win3}	-0.051	0.070	(-0.188,0.089)
β_{Winter}	-0.062	0.012	(-0.087,-0.036)	γ_{Winter}	0.011	0.019	(-0.025,0.051)
$\beta_{RetailValue}$	0.205	0.056	(0.153,0.342)	$\gamma_{RetailValue}$	-0.604	0.084	(-0.774,-0.439)

Table 4: Posterior Summaries of Model 2

As for auction features, pallet items' average retail value appears to be the most important factor. Auctions with higher-value items discourage bidders from entering auctions earlier, potentially because they are trying to avoid bidding wars that may increase final price and trigger the "winner's curse." In addition, first-bid arrival times are longer for auctions posted over the weekend, which may be due to greater inactivity in that period. On the other hand, bidders tend to enter auctions earlier during the market's busy first-quarter season. Item quantity, meanwhile, does not seem to have a significant effect on this process. Table 4 shows posterior summaries of the different γ parameters, which are coefficients of the regression model that estimates the precision parameters. Similar to β parameters, the average time of previous first bids and number of just-finished auctions have the largest effect (i.e., they have positive values, which leads to higher precision and lower variance in bid-time distribution). On the other hand, item quantity and retail value add more uncertainty to bidders' first-bid time. Knowing how bidders' and auctions' heterogeneity influence timing of bid arrivals, in the following section, we propose a new model to study how bidders enter the auctions.

3 Modeling Participation in Auctions

In the previous section, we proposed a model to assess time of first bid and to identify the first bidder among a set of bidders who participate in the selected auctions. In this section, we propose a dynamic model to study bidders' participation in auctions. The auction literature includes a number of models to study participation. Li and Zheng (2009) proposed an empirical model based on a semiparametric Bayesian framework to study entry and bidding based on only auctions' heterogeneity. But in general, a bidder's decision to participate in an auction is endogenous. Pilehvar et al. (2016) noted that the entry decision is derived from an underlying decision process, with the current marketplace structure (i.e., multiple similar auctions run simultaneously, which may affect bidders' entry decisions) adding to the modeling complexity. In studying the effect of market price and bidders' heterogeneity on the value of the first bid, they modeled entry to the auction through a discrete choice model. But their suggested static model does not take the time-varying effect of these information on bidders' decisions. In our model, however, we consider data on multiple bidders across a set of different auctions and study how their decision process changes over time using a dynamic approach. The proposed model is quite robust and has good predictive performance as it is based on behavior of a set of heterogeneous bidders.

3.1 A Dynamic Probit Model for Participation

Modeling categorical longitudinal data using time-varying coefficients have been studied by authors such as Carlin and Polson (1992) and more recently by Soyer and Sung (2013). In our setup we consider a Bayesian dynamic probit model and develop Bayesian inference for it following the work of Soyer and Sung (2013).

The binary random variable Y_{it} takes value 1(0) if bidder i participates (does not participate) in the $t-th$ auction. Given $K \times 1$ a vector of covariates X_{it} we denote the probability of participation by $Pr\{Y_{it} = 1 \mid X_{it}\} = \pi_{it}$ and assume a probit model as

$$\pi_{it} = \Phi(\beta_t X_{it}),$$

where $t = 1, \ldots, T$, $i = 1, \ldots, n$ and Φ is the cumulative normal distribution function. We note that the $1 \times K$ vector of regression parameters β_t is time-dependent. We assume that conditional on (π_{it}, X_{it}) Y_{it} is independent of Y_{jt} and Y_{is} for all $i \neq j$ and $t \neq s$. In our development, the dynamic structure of the model will be reflected by the state equation

$$\beta_t = \beta_{t-1} + w_t \text{ with } \omega_t \sim N(0, W)$$

where w_t's are uncorrelated multivariate normal error vectors with mean 0 and covariance matrix W. This specific evolution of β_t is referred to as the *steady model* in the literature; see for example, West and Harrison (1999). Note that a static probit model can be obtained by assuming $\beta_t = \beta$ in the above by setting $W = 0$ (i.e., 0 matrix).

An important concept in developing Bayesian inference for the dynamic probit model is the use of latent variables. Albert and Chib (1993) have introduced a a data augmentation approach using this latent structure within Gibbs sampling to develop Bayesian analysis of the static probit models. Soyer and Sung (2013) proposed a dynamic version of Albert and Chib (1993) by combining it with the Forward Filtering Backward Sampling (FFBS) approach of Frühwirth-Schnatter (1994) to analyze dynamic probit models. This approach provides an exact Gibbs sampler where all the full conditional distributions can be easily obtained using conjugate Bayesian results. This is achieved by specifying a multivariate normal prior for β_0 as $\beta_0 \sim N(m_0, C_0)$ and a Wishart prior for W^{-1}, the inverse of the covariance matrix (that is, the precision matrix) in the steady model; see Soyer and Sung (2013) for details of the Gibbs sampler.

3.2 Numerical Implementation

3.2.1 Data

We start by creating a set of participated and potential auctions for each bidder. A participated auction is an auction in which bidder i has placed a bid ($Y_{it} = 1$). On the other hand, potential auctions are auctions in which we assume the bidder has seen and has shown interest in but has not participated, that is, $Y_{it} = 0$. We do so by searching the data set during running time of that auction to see whether the bidder has placed a bid in a similar auction during that time. For participated auctions, we consider his/her characteristics at the time of the first bid. For potential auctions (during which the bidder has placed multiple observed bids in other comparable auctions), we consider bidder's characteristics at the time of his/her last observed bid. We should note that bidder's characteristic is referring to bidder's time-specific (i.e. time-varying) covariates such as number of open auctions he/she is participating, number of just finished auctions he/she has won which are specific to the time each bid is placed. The process is repeated for other bidders to construct the set of their participated and potential auctions. We then randomly select a set of n bidders and for each bidder, we choose T of his/her consecutive auctions activity which can either be a participation in an auction or being a potential bidder of an auction[3]

Having built this dataset (which has $n \times T$ records), we go through the following steps for both the dynamic (A_1) and the static (A_2) models:

Step One: Divide the data into two time frames: main period (from time/auction 1 to time/auction s) and a prediction period (from auction $s+1$ to auction T) and run the model on the main period to calculate the parameters up to the time/auction s (i.e., β_t for $t = 1, ..., s$).

Step Two: Calculate fit measures for the s data points and make a one-step ahead prediction for each bidder's participation in his/her upcoming auction $\pi_{i(s+1)} = Pr\{Y_{i(s+1)} = 1 \mid X_{i(s+1)}\} = \Phi(\beta_{s+1} X_{i(s+1)})$.

Step Three: Calculate predictive measures in terms of bidders' future participation and store $\pi_{i(s+1)}$'s as the posterior predictive values for the i^{th} bidder's participation status in his/her $(s+1)^{\text{th}}$ auction. The predictive measure for each model A_i is the one-step ahead log predictive likelihood $PL_{A_i}(s+1) = p(Y_{s+1} \mid Y_s, A_i)$.

Step Four: Update both data frames by moving the just-forecasted auctions from the prediction period to the main period. Based on this rolling basis approach, the main period is now from auction 1 to $s+1$ and prediction period is from $s+2$ to T. If the prediction dataset is empty we stop, otherwise go to **Step Two**.

[3] T is the same for every bidder. We should note that index t refers to a bidder's t^{th} auction. In other words, one bidder's first auction is not necessarily another bidder's first auction and so on.

3.2.2 Results

In running and comparing the results of static and dynamic models, we randomly select 50 ($n = 50$) bidders along with 210 ($T = 210$) consecutive auctions activities for each. We choose bidders' first 200 ($s = 200$) auctions as the main dataset, while their next 10 auctions represent the holdout set. Following the discussed algorithm, we run each model 10 times, making a one-step-ahead prediction at each iteration, after which we update the datasets and move to the next iteration. As for covariates, we select two distinct categories. The first specifies auction characteristics, including item quantity and retail value (value per item); a binary indicator to show whether the auction is posted over the weekend; a binary indicator to show whether the auction is running in the first quarter; and the number of similar, overlapping auctions running simultaneously. The second category explains bidders' characteristics and includes participation volume in the past three months; win experience (i.e., ratio of wins in participated auctions over the past three months); number of open auctions in which the bidder has placed a bid; number of wins in auctions that have finished after the start of the focal auction but before his or her bid placement; number of days since bidder's last bid (i.e., a measure of inactivity); and, finally, bidder's participation rate (i.e., the ratio of participation in potential auctions over the past three months).

In the dynamic setting, the Wishart prior on W^{-1} has $r = K$ degrees of freedom with the scale matrix $R = diag(1, ..., 1)$, which has $K \times K$ dimension ($K = 13$ includes all the covariates plus the intercept term). For the analysis, we use R and WinBUGS on a personal computer with an INTELI i7-2600 CPU 3.40Ghz processor and 16GB RAM memory. This dynamic setting requires 41 minutes for each run of the iterative process, which includes simulations and calculation of fit and predictive measures. Inferences are made based on 5,000 posterior samples after burn-in sample of 10,000 iterations and thinning by five.

For model fit comparisons, we use the Deviance Information Criterion (DIC) of Spiegelhalter et al. (2002), a criterion that accounts for both model fit and complexity. Having deviance defined as $D = -2log\mathcal{L}(\Theta)$ with Θ representing unknown parameters of the model, DIC in its general form is defined as DIC $= \overline{D} + p_D$, where $\overline{D}$ is the posterior mean of the deviance and $p_D = \overline{D} - D(\hat{\Theta})$, where $D(\hat{\Theta})$ is a point estimate of the deviance obtained by substituting in the posterior means for Θ. In DIC formulation, p_D is a penalty measure for the complexity of the model by its effective number of parameters. At each run of the model, DIC is calculated only over the main, not prediction, period. Referring to Table 5, we calculate iteration 1 fit measures over the first 200 data points, iteration 2 the first 201, and so on. We find the dynamic model outperforms the static model in all of the iterations by having smaller values of $\overline{D}$ and DIC.

	Static			**Dynamic**		
Iter	**Dbar**	**pD**	**DIC**	**Dbar**	**pD**	**DIC**
1	10876.1	13.0	10889.0	10372.3	377.3	10749.6
2	10943.9	13.3	10957.2	10429.5	379.0	10808.6
3	11001.0	13.0	11014.0	10473.5	379.2	10852.7
4	11055.5	13.1	11068.6	10515	382.8	10897.8
5	11108.3	13.0	11121.4	10561.6	385.5	10947.2
6	11167.8	12.8	11180.6	10624.2	384.1	11008.4
7	11220.3	13.0	11233.3	10673	387.1	11060.1
8	11269.8	12.8	11282.7	10731.4	389.7	11121.1
9	11323.5	13.3	11336.8	10771	394.2	11165.2
10	11369.6	13.1	11382.7	10802.8	397.6	11200.4

Table 5: Fit comparisons of static and dynamic models over the main dataset in each of 10 iterations

We also evaluate predictive performance in terms of bidders' future participation. The one-step-ahead predictions are on a rolling basis because we use observations up to time s to make a prediction for observation at time $s+1$, then include the $(s+1)^{\text{th}}$ observation to predict observation at $s+2$, and so on. In evaluating models' predictive quality, we follow the work of Geweke and Amisano (2010) by obtaining one-step-ahead log predictive likelihoods and calculating the cumulative log predictive Bayes factors. Given two competing models A_1 (dynamic) and A_2 (static), the log Bayes factor may be decomposed as:

$$log\left[\frac{p(Y_T \mid Y_s, A_1)}{p(Y_T \mid Y_s, A_2)}\right] = \sum_{t=s+1}^{T} log\left[\frac{PL_{A_1}(t)}{PL_{A_2}(t)}\right]$$

where $PL_A(s+1) = p(Y_{s+1} \mid Y_s, A)$ and PL_{A_1}/PL_{A_2} is the predictive Bayes factor in favor of A_1 over A_2 for observation t. This decomposition shows how individual observations contribute to the evidence in favor of one model over another. In our setup, at each time t we are making 50 one-step-ahead predictions to account for all 50 bidders' next-step prediction. In other words, we have: $PL_A(s+1) = p(Y_{s+1} \mid Y_s, A) = \prod_{i=1}^{n} p(I_{i,s+1} \mid Y_s, A)$.

We conduct evaluations according to Kass and Raftery (1995), scoring rules where a log Bayes factor value between 0 and 1 is "not worth more than a bare mention," values between 1 and 3 show positive evidence, values between 3 and 5 show strong evidence, and values greater than 5 show very strong evidence in favor of the dynamic model over the static model. Table 6 shows the results of the one-step-ahead predictions (i.e., classification) in each iteration. The first two columns show the log predictive likelihood of models, and the third column shows the log predictive Bayes factor in support of the dynamic model over the static model in that iteration. The cumulative log Bayes factor is $14.03 > 5$, which shows a very strong support for the dynamic model.

	Log Predicitve Likelihood		
Iter.	**Static**	**Dynamic**	**Log Predictive Bayes Factor**
1	-33.830	-30.490	3.340
2	-28.797	-23.614	5.184
3	-27.223	-23.712	3.511
4	-26.483	-23.649	2.834
5	-30.859	-31.575	-0.716
6	-27.177	-26.995	0.182
7	-26.937	-28.545	-1.609
8	-26.617	-26.933	-0.316
9	-23.149	-21.174	1.975
10	-24.062	-24.411	-0.349

Table 6: Log Predictive Likelihood and Log Bayes Factor Over 10 Iterations

Figure 2 compares the estimated posterior probabilities of 50 bidders' participation for their 210^{th} potential auctions (i.e., $s = 209$, the last run of the model) versus their actual values. We can see that both models slightly underestimate participation probabilities but the dynamic model predictions (right) are generally better than the static ones (left). In comparing both models, we also evaluate their predictive classification performances, as shown in Figure 3. The plot on the left compares the average (over iterations) accuracy of both models for different cutoff values. The plot on the right is a comparison of their receiver operating characteristic (ROC) curves by plotting their average true positive rates versus false positive rates for different cutoff values.[4] In addition, a cutoff sensitivity analysis of the dynamic model over the whole dataset suggests that by considering a cutoff value between 0.3 and 0.35, we can achieve reasonably high accuracy of about 70%, a high true positive rate of about 66%, and a low false positive rate of about 25%. This finding is similar to those in Figure 2 (i.e., an overall underestimation of posterior probabilities).

In addition to prediction performances, we analyzed regression parameters in terms of their significance and possible change over the study period. Figure 4 plots posterior means

[4]Each boxplot shows the spread of the estimated measure over 10 iterations at that cutoff point.

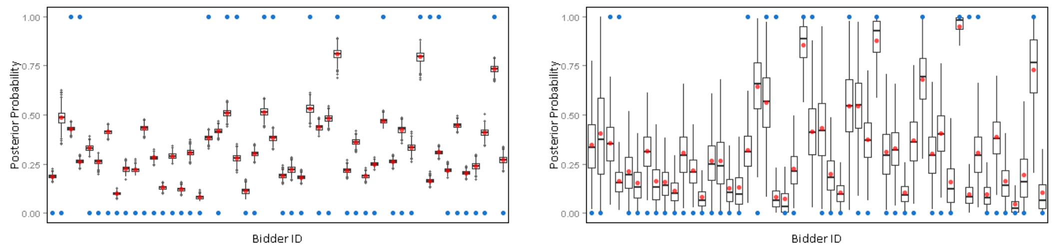

Figure 2: Posterior Probability of the 50 Bidders' Participation in their 210$^{\text{th}}$ Auction - Static Model (left) vs. Dynamic Model (right) - (blue points are actual participation status (1's and 0's) and red points are posterior means)

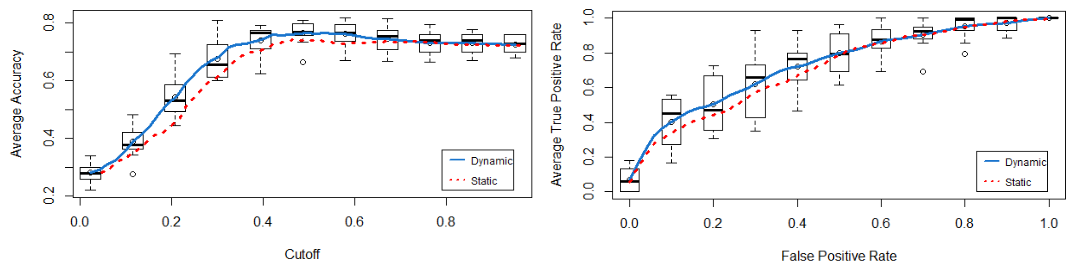

Figure 3: Average Accuracy and ROC Performance Curves Over Prediction Iterations

of four of the regression parameters in the dynamic model, where we can see the change in the parameters over the course of auctions. This is yet another indication of the dynamic model's appropriateness. While some parameters such as "past auction participations," "quantity of items," "retail value per item," and "days since last bid" show downward or upward trends, others do not clearly exhibit these patterns. The plots suggest quantity has a negative effect on participation that strengthens over time. The retail value of auction items, meanwhile, has a positive effect on bidders' entry decision, which strengthens over time. In other words, bidders are more willing to participate in an auction with a smaller pallet size (fewer items) and more valuable items. As they participate in more auctions, these features grow in importance. The same positive relation holds between the number of days since bidder's last activity and his or her participation probability. In other words, a bidder has a higher probability of participation in a potential auction after a longer inactivity period. The results suggest bidders participating in more similar auctions also are more willing to participate in the focal auction. On the other hand, the participation rate decreases when the auction platform is busy. This illustrates the negative effect of

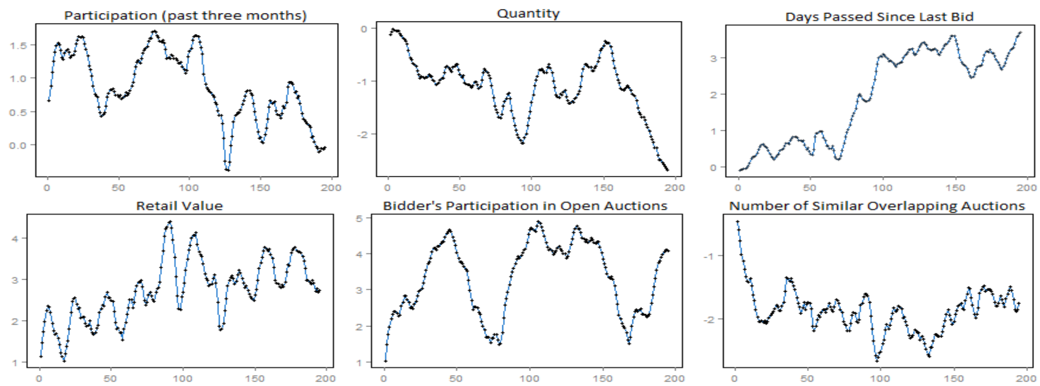

Figure 4: Posterior Mean of the Regression Parameters

multiple auction postings - a key practice in this marketplace - on auction dynamics through key factors such as bid arrival and time, bidder participation, and number of bids. The effect of other covariates - bidder win rate in the past three months, auction opening day, auction time of year - do not appear significant. The bidder's participation rate in the past three months (ratio of participation in potential auctions), meanwhile, has a mostly positive effect on the entry decision.

4 Concluding Remarks

In this paper we proposed two models to study bidders' timing of their first bid and their participation by adapting to some dynamic aspects of online secondary retail auction market. We explored different auction- and bidder-specific variables and their significance to bidders' time of the bid and participation, particularly how their effect may increase or decrease as bidders participate in more auctions over time. We did so by presenting a Bayesian beta regression model based on two different setups to study bidders' first-bid timing. Our analysis showed the effect of bidders' heterogeneity, namely how those whose past behavior differed from the time of the auction tend to place their first bid at different times. We also proposed a Bayesian dynamic probit approach for modeling bidders' entry into auctions. The analysis of the tested models revealed that the performance is vastly improved by using the dynamic modeling approach. The proposed dynamic probit approach uses the full Gibbs sampler presented in Soyer and Sung (2013), which is based on using data augmentation and sequential updating method of forward filtering-backward sampling in the context of Bayesian dynamic models (West and Harrison (1999)). To the best of our knowledge, we are the first to use Bayesian dynamic and beta regression models in studying auctions over time to predict bidders' behavior in upcoming auctions.

As discussed, secondary market auctions tend to have low recovery rate and high inventory costs. Auction designers and managers, however, could implement the developed models to achieve quicker sales at higher final prices to generate more revenue and reduce the number of unsuccessful auctions. Both models, we found, could be integrated to be part of a predictive model to identify the first bidder in a given auction. While the first model estimates bid arrival times and identifies the first bidder, the second model identifies the participants. The developed models can help auctioneers identify bidders with a low or high probability of participation or help direct bidder traffic to less-attractive auctions. Using these models also helps managers identify potential first bidders and obtain a more accurate estimation of first-bid value and final auction outcome. Beyond applying these models to study the first bid, they could be generalized separately or together to study other bids and their arrival times throughout the auction. Modeling first-bid arrival time, however, remains instrumental as the arrival of subsequent bids will be truncated below by the time of the first.

We note a number of possible limitations in our study, some of which we identified in the course of our analysis. The first stems from the nature of our dataset, which is at the bid level. As such, we only have information on placed bids, including auction and bidder characteristics associated with that bid. A major part of this analysis is based on the definition of available bidders (i.e., online status) at auction time. In our work we have defined potential bidders as those who placed a bid in similar auctions during the time of a given auction. A more robust model would could be developed around a dataset that shows whether bidders are watching an auction before they bid or whether they leave auctions after watching them, and so on. This would significantly increase the prediction power of the model since lots of already-included potential bidders (0's in the dataset) will be removed. In other words, the model will only include actual potential bidders of the auctions.

The one-step-ahead forecasting and updating procedure in the dynamic models represents another limitation. The iterative process of a prediction and model re-run can be impractical and computationally expensive for a large dataset and a large volume of predictions. Given that possible extension of our method for practical applications is crucial, we will explore further extensions of the model by using particle filtering. This is a more efficient way to updating the model in real time rather than running it with past data, which could prove computationally expensive.

References

Albert, J. H. and S. Chib (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American statistical Association 88*(422), 669–679.

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

Borle, S., P. Boatwright, and J. B. Kadane (2006). The timing of bid placement and extent of multiple bidding: An empirical investigation using ebay online auctions. *Statistical Science*, 194–205.

Bradlow, E. T. and Y.-H. Park (2007). Bayesian estimation of bid sequences in internet auctions using a generalized record-breaking model. *Marketing Science 26*(2), 218–229.

Carlin, B. P. and N. G. Polson (1992). Monte carlo bayesian methods for discrete regression models and categorical time series. *Bayesian statistics 4*, 577–586.

Chan, T. Y., V. Kadiyali, and Y.-H. Park (2007). Willingness to pay and competition in online auctions. *Journal of Marketing Research 44*(2), 324–333.

Frühwirth-Schnatter, S. (1994). Data augmentation and dynamic linear models. *Journal of time series analysis 15*(2), 183–202.

Geweke, J. and G. Amisano (2010). Comparing and evaluating bayesian predictive distributions of asset returns. *International Journal of Forecasting 26*(2), 216–230.

Kass, R. E. and A. E. Raftery (1995). Bayes factors. *Journal of the american statistical association 90*(430), 773–795.

Ku, G., A. D. Galinsky, and J. K. Murnighan (2006). Starting low but ending high: A reversal of the anchoring effect in auctions. *Journal of Personality and Social Psychology 90*(6), 975.

Li, S., K. Srinivasan, and B. Sun (2009). Internet auction features as quality signals. *Journal of Marketing 73*(1), 75–92.

Li, T. and X. Zheng (2009). Entry and competition effects in first-price auctions: theory and evidence from procurement auctions. *The Review of Economic Studies 76*(4), 1397–1429.

Park, Y.-H. and E. T. Bradlow (2005). An integrated model for bidding behavior in internet auctions: Whether, who, when, and how much. *Journal of Marketing Research 42*(4), 470–482.

Pilehvar, A., W. J. Elmaghraby, and A. Gopal (2016). Market information and bidder heterogeneity in secondary market online b2b auctions. *Management Science 63*(5), 1493–1518.

National Retail Federation (2015). Consumer returns in the retail industry. `https://nrf.com/sites/default/files/Images/MediaCenter/NRFRetailReturnFraudFinal_0.pdf`. [Online; accessed 10-October-2017].

Simonsohn, U. and D. Ariely (2008). When rational sellers face nonrational buyers: evidence from herding on ebay. *Management Science 54*(9), 1624–1637.

Soyer, R. and M. Sung (2013). Bayesian dynamic probit models for the analysis of longitudinal data. *Computational Statistics & Data Analysis 68*, 388–398.

Spiegelhalter, D. J., N. G. Best, B. P. Carlin, and A. Van Der Linde (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 64*(4), 583–639.

West, M. and J. Harrison (1999). *Bayesian Forecasting and Dynamic Models*. Springer Series in Statistics. Springer New York.

Wilcox, R. T. (2000). Experts and amateurs: The role of experience in internet auctions. *Marketing Letters 11*(4), 363–374.