Explaining and Forecasting Online Auction Prices and Their Dynamics Using Functional Data Analysis

Shanshan WANG
Department of Mathematics, University of Maryland, College Park, MD 20742 (*shanshan@math.umd.edu*)

Wolfgang JANK
Department of Decision and Information Technologies, Robert H. Smith School of Business, University of Maryland, College Park, MD 20742 (*wjank@rhsmith.umd.edu*)

Galit SHMUELI
Department of Decision and Information Technologies, Robert H. Smith School of Business, University of Maryland, College Park, MD 20742 (*gshmueli@rhsmith.umd.edu*)

Online auctions have become increasingly popular in recent years, and as a consequence there is a growing body of empirical research on this topic. Most of that research treats data from online auctions as cross-sectional, and consequently ignores the changing dynamics that occur during an auction. In this article we take a different look at online auctions and propose to study an auction's price evolution and associated price dynamics. Specifically, we develop a dynamic forecasting system to predict the price of an ongoing auction. By dynamic, we mean that the model can predict the price of an auction "in progress" and can update its prediction based on newly arriving information. Forecasting price in online auctions is challenging because traditional forecasting methods cannot adequately account for two features of online auction data: (1) the unequal spacing of bids and (2) the changing dynamics of price and bidding throughout the auction. Our dynamic forecasting model accounts for these special features by using modern functional data analysis techniques. Specifically, we estimate an auction's price velocity and acceleration and use these dynamics, together with other auction-related information, to develop a dynamic functional forecasting model. We also use the functional context to systematically describe the empirical regularities of auction dynamics. We apply our method to a novel set of Harry Potter and Microsoft Xbox data and show that our forecasting model outperforms traditional methods.

KEY WORDS: Autoregressive model; Dynamics; Electronic commerce; Exponential smoothing; Functional data; Nonparametric; Online auctions; Smoothing spline.

1. INTRODUCTION

Electronic commerce, and in particular online auctions, has created much public interest in recent years. One of the main drivers of this interest is eBay (www.eBay.com). On any given day, there are several million items, dispersed across thousands of categories, for sale on eBay. eBay's popularity among the public is evidently quantified in the following numbers: In 2004 alone, \$34.2 billion was reported in gross merchandise volume, up from \$23.8 billion in 2003; the cumulative confirmed registered users totaled a record 135.5 million, which was a 43% increase over the 94.9 million users reported in 2003; and eBay hosts approximately 254,000 stores worldwide, with approximately 161,000 stores hosted in the United States alone. According to the Forrester Technographics survey, close to 30% of all U.S. households had bid in an eBay online auction in 2004.

The popularity of online auctions is also one reason for an increasing amount of scholarly research. Most of this research has been published in the economics, marketing, and information systems literature (e.g., Lucking-Reiley, Bryan, Prasad, and Reeves 2000; Roth and Ockenfels 2002; Ba and Pavlou 2002; Bajari and Hortacsu 2003, 2004; Bapna, Goes, Gupta, and Jin 2004). We find it surprising, though, that the statistics community has hardly been involved to date, because online auctions arrive with a huge amount of new and different data-related challenges and problems. In particular, a main feature of online auction data is the change in dynamics of the bidding process. Most approaches to date tend to ignore this dynamic information and treat data as cross-sectional, by aggregating over the temporal dimension. However, such approaches lead to a great loss in information. Moreover, studies investigating bidding regularities also tend to be limited to reporting summary statistics (such as the percent of bids placed in the last minute of the auction). A first step beyond those limitations is via the use of advanced data visualization. For instance, Shmueli and Jank (2005) introduce graphical methods such as profile plots and statistical zooming, and Shmueli, Jank, Aris, Plaisant, and Shneiderman (2006) develop a tool for interactive visualization of bidding data, a tool that couples temporal with cross-sectional data. Another step is to model bidding regularities directly. Shmueli, Russo, and Jank (in press), for instance, take a probabilistic approach for modeling the bid arrival process and proposed the BARISTA (Bid ARrivals In STAges) model, a class of three-stage nonhomogeneous Poisson processes that captures different bidding dynamics, including the self-similarity property that has been empirically observed in Roth and Ockenfels (2002). And, finally, taking a different approach, Jank and Shmueli (2006) propose

Journal of Business & Economic Statistics
April 2008, Vol. 26, No. 2
DOI 10.1198/073500106000000477

state-of-the-art functional data methodology for directly modeling temporal bidding information and its dynamic change. For example, Jank and Shmueli (in press) use functional data analysis to find that the price dynamics change quite sharply over the course of an auction, even for auctions for the same item. In fact, functional data analysis proves to be a very suitable tool for the analysis of online auction data, and more discussion on the versatility of this toolset in the broader context of electronic commerce research can be found in Jank and Shmueli (2006). We build upon the general versatility of functional data analysis in this work.

In this article we develop a dynamic forecasting model to predict price in online auctions. By dynamic, we mean a model that operates during the live auction and forecasts price at a future time point of the ongoing auction and, as a by-product, also at the auction end. This is in contrast to static forecasting models that predict only the final price and that take into consideration only information available at the start of the auction. Such information may involve the length of the auction, its opening price, product characteristics, or the seller's reputation and may be modeled using standard least squares regression analysis. However, a static approach cannot account for information that becomes available after the start of the auction (e.g., the amount of competition or current price level), and it cannot incorporate such information "on the fly." As we explain throughout this article, we find functional data analysis a very suitable tool for developing dynamic price predictions.

Forecasting price in online auctions can have benefits to different auction parties. For instance, price forecasts can be used to dynamically score auctions for the same (or similar) item by their predicted price. On any given day, there are several hundreds, or even thousands of open auctions available, especially for very popular items such as Apple iPods or Microsoft Xboxes. Dynamic price scoring can lead to a ranking of auctions with the lowest expected price. Such a ranking could help bidders focus their time and energy on only a few select auctions, that is, those that promise the lowest price. Auction forecasting can also be beneficial to the seller or the auction house. For instance, the auction house can use price forecasts to offer insurance to the seller. This is related to the idea by Ghani and Simmons (2004) who suggested offering seller insurance that guarantees a minimum selling price. In order to do so, however, it is important to correctly forecast the price, at least on average. Whereas Ghani and Simmons' method is static in nature, our dynamic forecasting approach could potentially allow more flexible features such as an "Insure-It-Now" option, which would allow sellers to purchase insurance either at the beginning of the auction or during the live auction (with a time-varying premium). Price forecasts can also be used by eBay-driven businesses that provide services to buyers or sellers. Recently, the authors were contacted by a company that provides brokerage services for eBay sellers, about using the dynamic forecasting system to create a secondary market for eBay-based derivatives.

While there has been some work related to price forecasting in online auctions, our approach is novel particularly because of its dynamic nature (see also Hortacsu and Cabral 2005, for the dynamics of seller reputation). As pointed out earlier, Ghani and Simmons (2004), using data-mining methods, also predict the end-price of online auctions, however, that method is static and cannot account for newly arriving information in the live auction. Structural models to recover the bid distribution (Bajari and Hortacsu 2003), although able to more explicitly account for mechanism design, are also focused on the final price. The dynamic nature of our forecasting approach is founded within the framework of functional data analysis (FDA). In FDA, the center of interest is a set of curves, shapes, or objects or, more generally, a set of *functional observations*. This is in contrast to classical statistics where the interest centers around a set of data vectors. Although the concept of FDA has been around for a longer time, the field has gained fast momentum lately due to the work of Ramsay and Silverman (2000). There is quite a bit of interest in the statistics literature to generalize classical models and methods to the functional context. Examples are regression analysis for a functional response (Faraway 1997), functional logistic regression (Escabias, Aguilera, and Valderrama 2004), and generalized linear models for functional data (James 2002).

Our forecasting approach presents several methodological additions to this stream of literature. First, to the best of our knowledge, forecasting functional data is a topic that has not been sufficiently addressed in the FDA literature to date. In fact, the use of functional data analysis presents several practical and conceptual advantages for online auction data. Traditional methods for forecasting time series, such as exponential smoothing or moving averages, cannot be applied in the auction context, at least not directly, because of the special data structure. Traditional forecasting methods assume that data arrive in evenly spaced time intervals such as every quarter or every month. In such a setting, one trains the model on data up to the current time period t and then uses this model to predict at time $t+1$. Implied in this process is the important assumption that the distance between two adjacent time periods is equal, which is the case for quarterly or monthly data. Now consider the case of online auctions. Bids arrive in very unevenly spaced time intervals, determined by the bidders and their bidding strategies, and the number of bids within a short period of time can sometimes be very sparse and other times be extremely dense. In this setting, the interval between bids can sometimes be more than a day and at other times only seconds. Second, online auctions, even for the same product, can experience price paths with very heterogeneous *price dynamics*. By price dynamics, we mean the speed at which price travels during the auction and the rate at which this speed changes. Traditional models do not account for instantaneous change and its effect on the price forecast. This calls for new methods that can measure and incorporate this important dynamic information.

Unevenly spaced data have always been very challenging for traditional forecasting methods. Some of the more recent and influential work in this area includes Engle and Russel (1998) who propose the use of the so-called autoregressive conditional duration model for unevenly spaced transaction data, such as high-frequency asset price data. Similarly, Shmueli et al. (in press) develop a class of three-stage nonhomogeneous Poisson processes, the BARISTA, to model the changing bid arrival process. In both of these cases, the underlying assumption is that the discrete observations are manifestations of a continuous process. Following this idea, we take a functional data approach. Specifically, we assume an underlying price curve that describes the price increase, or *price evolution*, of an online auction. However, limitations in human perception and measurement capabilities allow us to observe this curve only at discrete time points. Following functional principles, we recover this curve from the observed data using appropriate smoothing techniques.

Another appeal of the functional data framework is the observation that the price dynamics change quite significantly over the course of an auction (Jank and Shmueli in press). By treating auction price as a functional object and recovering the underlying price curve, we obtain reliable estimates of the price dynamics via derivatives of the smooth functional object, and we can consequently incorporate these dynamics into the forecasting model. This results in a novel and potentially very powerful forecasting system. Although one may also approximate dynamics differently, for example, by using the first forward difference or the central difference, such an approach is likely to be much less accurate, especially for applications with very unevenly spaced data (as in the case of an online auction), and even more so for approximating higher order derivatives.

The article is organized as follows. In Section 2 we briefly introduce the auction mechanism on eBay, its data availability, and the data used in this work. In Section 3 we provide a systematic description of the empirical regularities in online bidding dynamics. Section 4 develops the forecasting model, and we apply the method to our data in Section 5. Section 6 concludes with final remarks.

2. EBAY'S ONLINE AUCTIONS

2.1 Auction Mechanism and Data Availability

The dominant auction format on eBay is a variant of the second-price sealed-bid auction (Krishna 2002) with "proxy bidding." This means that individuals submit a "proxy bid," which is the maximum value they are willing to pay for the item. The auction mechanism automates the bidding process to ensure that the person with the highest proxy bid is in the lead of the auction. The winner is the highest bidder and pays the second highest bid (plus an increment). Unlike other auctions, eBay has strict ending times, ranging between 1 and 10 days from the opening of the auction, as determined by the seller. eBay posts information on closed auctions for a duration of at least 15 days on its web site (see *http://listings.ebay.com/pool1/listings/list/completed.html*). These publicly available postings make eBay an invaluable source of rich bidding data.

A typical bid history for a closed auction (see Fig. 1) includes information about the magnitude and time when each bid was placed. Additional information that is made available includes information about the seller and the bidders (e.g., user name, feedback ratings), information about the item sold (e.g., name, description), and information about the auction format (e.g., auction duration, magnitude of the opening bid).

Every day on eBay, there are several million items for sale, which means that a large amount of data are publicly available. Although such data can be collected manually, simply by browsing through individual web pages, in practice this is very time consuming. Therefore, data are often collected automatically using web agents or web crawlers. Web crawlers are software programs that visit a number of pages automatically and extract required information. In that way, high-quality information on a large number of auctions can be gathered in a short period of time.

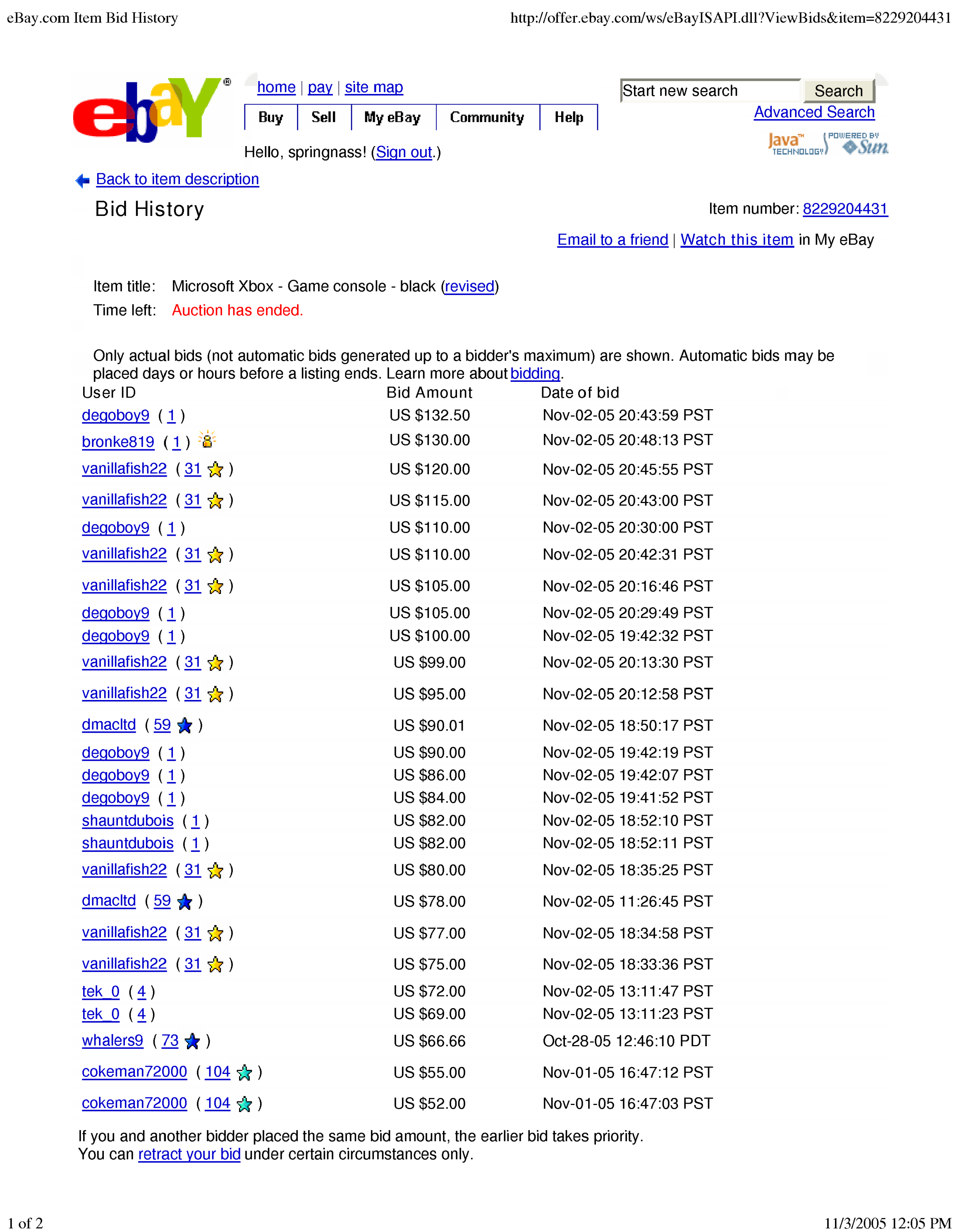
2.2 Data Used in This Study

The data used in this study are 190 7-day auctions of *Microsoft Xbox* gaming systems and *Harry Potter and the Half-Blood Prince* books obtained via a web crawler during the months of August and September of 2005. At the time of writing, Xbox systems were very popular items on eBay and sold for about $179.98 (based on *Amazon.com*). Harry Potter books were also very popular items and sold for about $27.99 on *Amazon.com*. We can, thus, consider Xbox systems high-valued items and can compare the performance of our method relative to the lower valued Harry Potter books.

For each auction in our dataset, we collected the bid history, which reveals the temporal order and magnitude of bids and which forms the basis of the functional forecasting model. Figure 2 shows a scatterplot of the bid history for a typical auction. We can see that bids arrive at very unevenly spaced time intervals. Although the number of incoming bids is sparse during some periods of the 7-day auction (especially in the middle), it can be very dense at other times such as at the very beginning and especially at the end of the auction. Figure 3 shows the scatter of bids, aggregated over all 190 auctions. Notice that most of the bids arrive in the last minutes of the auction, which, as we have pointed out earlier, is a typical feature of eBay's auctions.

We also collected information on the auction format, the product characteristics, and bidder and seller attributes. This information is summarized in Tables 1 and 2. Unsurprisingly, the high-valued items (Xbox) have, on average, a higher opening bid and a higher final auction price. However, it is noteworthy that the high-valued items see, on average, more competition (i.e., a larger number of bids), but feature auction participants with a lower average bidder and seller rating. Only few auctions in our dataset had made use of the secret reserve price option so the (numerical) difference between high- and low-valued items with respect to this feature may not be of practical importance. More interestingly though, most of the high-valued items are used (over 90%), which compares to only 46% used Harry Potter books. Table 2 also shows the distribution for the two variables early bidding and jump bidding. Note that these two variables are not directly observed but are derived from the bid history. We comment on how we derived these two variables next.

Early Bidding. The timing of a bid plays an important role in bidders' strategic decision making. For example, Roth and Ockenfels (2002) find evidence that many bidders place their bids very late in the auction, resulting in what is often called "bid sniping." Shmueli et al. (in press) found that an auction often consists of three relatively distinct parts: an early part with *some* bidding activity, a middle part with *very little* bidding, and a final part with *intense* bidding. In particular, they found that the early bidding part of the auction typically extends until about day 1.5 of a 7-day auction. Bapna et al. (2004) characterized bidders' strategies by, among other things, the timing



eBay.com Item Bid History http://offer.ebay.com/ws/eBayISAPI.dll?ViewBids&item=8229204431

home | pay | site map

Buy | Sell | My eBay | Community | Help

Start new search Search

Advanced Search

Hello, springnass! (Sign out.)

Back to item description

Bid History

Item number: 8229204431

Email to a friend | Watch this item in My eBay

Item title: Microsoft Xbox - Game console - black (revised)

Time left: Auction has ended.

Only actual bids (not automatic bids generated up to a bidder's maximum) are shown. Automatic bids may be placed days or hours before a listing ends. Learn more about bidding.

User ID	Bid Amount	Date of bid
degoboy9 (1)	US $132.50	Nov-02-05 20:43:59 PST
bronke819 (1)	US $130.00	Nov-02-05 20:48:13 PST
vanillafish22 (31)	US $120.00	Nov-02-05 20:45:55 PST
vanillafish22 (31)	US $115.00	Nov-02-05 20:43:00 PST
degoboy9 (1)	US $110.00	Nov-02-05 20:30:00 PST
vanillafish22 (31)	US $110.00	Nov-02-05 20:42:31 PST
vanillafish22 (31)	US $105.00	Nov-02-05 20:16:46 PST
degoboy9 (1)	US $105.00	Nov-02-05 20:29:49 PST
degoboy9 (1)	US $100.00	Nov-02-05 19:42:32 PST
vanillafish22 (31)	US $99.00	Nov-02-05 20:13:30 PST
vanillafish22 (31)	US $95.00	Nov-02-05 20:12:58 PST
dmacltd (59)	US $90.01	Nov-02-05 18:50:17 PST
degoboy9 (1)	US $90.00	Nov-02-05 19:42:19 PST
degoboy9 (1)	US $86.00	Nov-02-05 19:42:07 PST
degoboy9 (1)	US $84.00	Nov-02-05 19:41:52 PST
shauntdubois (1)	US $82.00	Nov-02-05 18:52:10 PST
shauntdubois (1)	US $82.00	Nov-02-05 18:52:11 PST
vanillafish22 (31)	US $80.00	Nov-02-05 18:35:25 PST
dmacltd (59)	US $78.00	Nov-02-05 11:26:45 PST
vanillafish22 (31)	US $77.00	Nov-02-05 18:34:58 PST
vanillafish22 (31)	US $75.00	Nov-02-05 18:33:36 PST
tek_0 (4)	US $72.00	Nov-02-05 13:11:47 PST
tek_0 (4)	US $69.00	Nov-02-05 13:11:23 PST
whalers9 (73)	US $66.66	Oct-28-05 12:46:10 PDT
cokeman72000 (104)	US $55.00	Nov-01-05 16:47:12 PST
cokeman72000 (104)	US $52.00	Nov-01-05 16:47:03 PST

If you and another bidder placed the same bid amount, the earlier bid takes priority.
You can retract your bid under certain circumstances only.

1 of 2 11/3/2005 12:05 PM

Figure 1. Partial bid history of an eBay Palm-515 auction. On the leftmost side of the table, we can see the bidders' user names, followed by their rating. The stars indicate that an eBay member has achieved 10 or more feedback points. The amount and time of the bids appear on the right.

of their bid and found that bidders of the "early evaluators" type place their first bid on average on day 1.4 of a 7-day auction. Following this empirical evidence, we define an auction as characterized by *early bidding* if the first bid is placed within the first 1.5 days. Table 2 shows the distribution of auctions with early bidding. We can see that among the high-valued auctions (Xbox), over 50% experience early bidding, whereas this number is much lower (28%) for the low-valued items (Harry Potter). It may well be that bidders for high-valued items are more inclined to bid early in order to establish a time priority, because, in the case of two bidders with identical bids, the bidder with the earlier bid wins the auction.

Jump Bidding. We also include information about jump bidding. To that end, one has to define what exactly determines

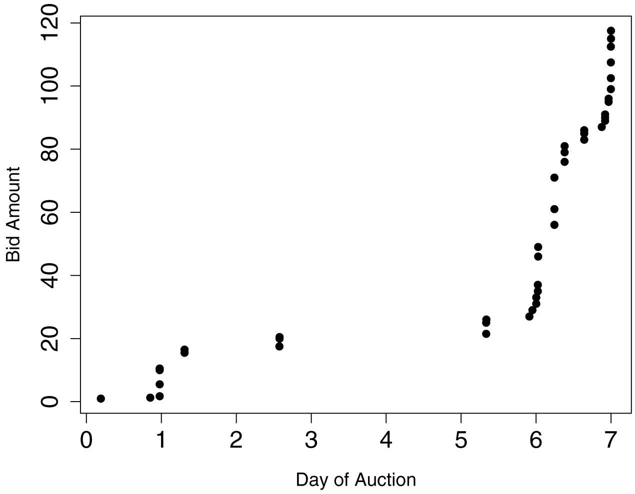


Figure 2. Bids placed in auction 75 of a Microsoft Xbox auction. The horizontal axis denotes time (in days); the vertical axis denotes bid amount (in \$).

a "jump bid," that is, what magnitude of difference between two consecutive bids constitutes an unusually high increase in the bidding process. There exists only little prior investigation on that topic. For instance, Easley and Tenorio (2004) study jump bidding as a strategy in ascending auctions and defined jump bids as bid increments that are larger than the minimum increment required by the auctioneer (see also Isaac, Salmon, and Zillante 2002; Daniel and Hirshleifer 1998). Bid increments larger than the minimum increment are relatively common on eBay (see Fig. 4). We, therefore, focus here on increments that result in a very unusual "jump." In order to define "unusual," we take the following approach. For all auctions in our dataset, we first examine all differences in bid magnitudes between pairs of consecutive bids. The difference in consecutive bids leads to a step function of bid increments. Figure 4 shows this step function for all Xbox and Harry Potter auctions. We can see that most auctions are characterized by only very small bid increments (i.e., only very small steps). But we can also see that the relevance of a jump depends on the scale (i.e., the value of the item) and should be considered *relative* to this value.

The distribution of the *relative* jumps, relative to the average final price, is displayed in Figure 5. We see that the distribution for both high- and low-valued items is very skewed. In addition, the great majority of relative jumps, regardless of item value, are smaller than 30% [see Fig. 5(c)]. We, therefore, define a *jump bid* as a bid that is at least 30% higher than the previous bid. We define a corresponding indicator variable for auctions that have at least one jump bid (i.e., the variable jump bidding in Table 2 equals 1 if and only if the auction has at least one jump bid). Table 2 shows that over 25% of the low-valued auctions (Harry Potter) see jump bidding, whereas this number is only about 9% for the high-valued auctions.

3. FUNCTIONAL REGRESSION AND AUCTION DYNAMICS

To understand the motivation for our forecasting model, it is useful to first take a closer look at eBay auction data. We have pointed out earlier that the data are characterized by rapidly changing price dynamics. We illustrate this phenomenon in this section by investigating the relationship between eBay's auction dynamics and other auction-related information. This will also lay the ground for the forecasting model that we describe in the next section.

We investigate the empirical regularities in eBay's auction dynamics using *functional regression analysis*. Functional regression analysis is similar to classical regression in that it relates a response variable to a set of predictors. However, in contrast to classical regression where the response and the predictors are vector valued, functional regression operates on *functional objects*, which can be a set of curves, shapes, or objects. In our application, we refer to the continuous curve that describes the price evolution between the start and the end of the auction as the functional object. More details on functional regression can be found in Ramsay and Silverman (2005).

Functional regression analysis involves two basic steps. In the first step, the functional object is recovered from the observed data. We describe this step in the next section. After

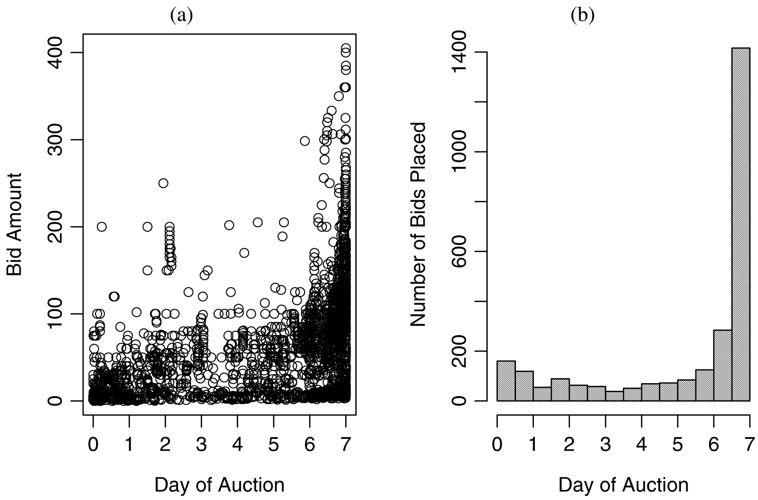


Figure 3. Data for the 190 7-day auctions. Panel (a) shows the amount of the bid versus the time of the bid, aggregated across all auctions. Panel (b) shows a histogram of the distribution of the bid arrivals over 7 days. Each bin corresponds to a 12-hour time interval.

Table 1. Summary statistics for all continuous variables

Variable	Item	Count	Mean	Median	Min	Max	Standard deviation
Opening bid	Xbox	93	36.22	24.99	.01	175.00	37.96
	Harry Potter	97	4.13	4.00	.01	10.99	3.26
Final price	Xbox	93	134.58	125.00	28.00	405.00	66.03
	Harry Potter	97	11.56	11.50	7.00	20.50	2.40
Number of bids	Xbox	93	20.01	19.00	2.00	75.00	12.76
	Harry Potter	97	8.47	8.00	2.00	24.00	4.30
Seller rating	Xbox	93	232.04	49.00	.00	4,604.00	613.07
	Harry Potter	97	325.99	126.00	.00	9,519.00	995.78
Bidder rating	Xbox	93	30.33	4.00	−1.00	2,736.00	135.06
	Harry Potter	97	83.21	14.00	−1.00	2,258.00	226.21

recovering the functional object, we model the relationship between a response object and a predictor object in a way that is conceptually very similar to classical regression. We describe that step in Section 3.2.

3.1 Recovery of the Functional Object

Functional data consist of a collection of continuous functional objects such as the price, which increases in an online auction. Despite their continuous nature, limitations in human perception and measurement capabilities allow us to observe these curves only at discrete time points. Moreover, the presence of error results in discrete observations that are noisy realizations of the underlying continuous curve. In the case of online auctions, we observe only bids at discrete times, which can be thought of as realizations from an underlying continuous price curve. Thus, the first step in every functional data analysis is to recover, from the observed data, the underlying continuous functional object. This is typically done using smoothing techniques.

The recovery stage is often initiated by some data preprocessing steps (e.g., Ramsay 2000). We denote the time that the ith bid was placed, $i = 1, \dots, n_j$, in auction j, $j = 1, \dots, N$, by t_{ij}. Note that because of the irregular spacing of the bids, the t_{ij}'s vary for each auction. In our data, $N = 190$ and $0 < t_{ij} < 7$. Let $y_i^{(j)}$ denote the bid placed at time t_{ij}. To better capture the bidding activity, especially at the end of the auction, we transform bids into log scores. To account for the irregular spacing, we linearly interpolate the raw data and sample it at a common set of time points t_i, $0 \le t_i \le 7$, $i = 1, \dots, n$. Then we can represent each auction by a vector of equal length

$$\mathbf{y}^{(j)} = \left(y_1^{(j)}, \dots, y_n^{(j)}\right), \tag{1}$$

where $y_i^{(j)} = y^{(j)}(t_i)$ denotes the value of the interpolated bid sampled at time t_i.

One typically recovers the underlying functional object using smoothing techniques (see Ramsay and Silverman 2005). One very flexible and computationally efficient choice is the

Table 2. Summary statistics for all categorical variables

Variable	Item	Case	Count	Proportion
Reserve price	Xbox	Yes	4	4.30%
		No	89	95.70%
	Harry Potter	Yes	1	1.03%
		No	96	98.97%
Condition	Xbox	New	8	8.60%
		Used	85	91.40%
	Harry Potter	New	52	53.61%
		Used	45	46.39%
Early bidding	Xbox	Yes	53	56.99%
		No	40	43.01%
	Harry Potter	Yes	28	28.87%
		No	69	71.13%
Jump bidding	Xbox	Yes	9	9.68%
		No	84	90.32%
	Harry Potter	Yes	25	25.77%
		No	72	74.23%

NOTE: "Case" is the category for the particular variable.

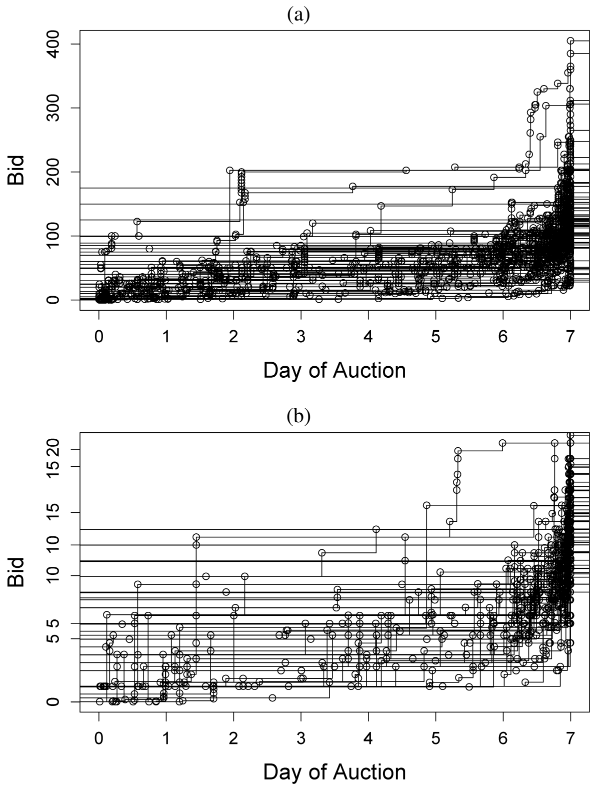


Figure 4. Step function of bid increments for Microsoft Xbox systems (a) and Harry Potter books (b).

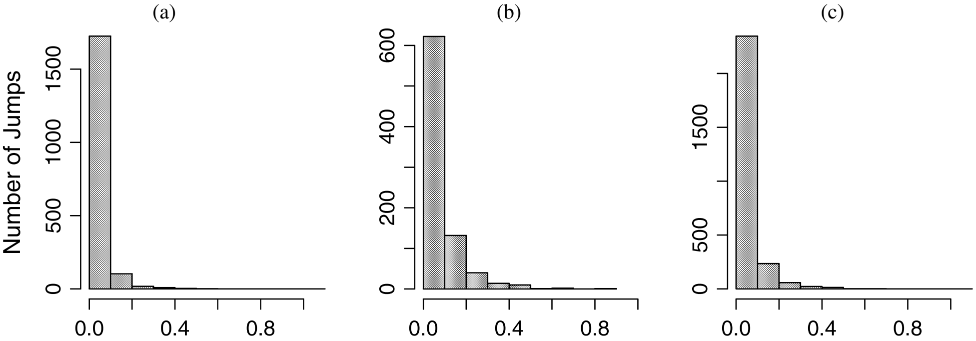


Figure 5. Distribution of relative jumps for Xbox alone (a), Harry Potter alone (b), and both combined (c).

penalized smoothing spline (e.g., Simonoff 1996). Consider a polynomial spline of degree p:

$$f(t) = \beta_0 + \beta_1 t + \beta_2 t^2 + \cdots + \sum_{l=1}^{L} \beta_{pl}(t - \tau_l)_+^p, \qquad (2)$$

where $\tau_1, \ldots, \tau_L$ is a set of knots and u_+ denotes the positive part of a function u. The choices of L and p strongly influence the departure of f from a straight line. The degree of departure can be measured by the roughness penalty $PEN_m = \int D^m f(s)^2\, ds$. The penalized smoothing spline minimizes the penalized residual sum of squares; that is the jth smoothing spline $f^{(j)}$ satisfies

$$PENSS^{(j)}_{\lambda,m} = \sum_{i=1}^{n} \bigl(y_i^{(j)} - f^{(j)}(t_i)\bigr)^2 + \lambda PEN_m^{(j)}, \qquad (3)$$

where the smoothing parameter λ controls the trade-off between data fit and the smoothness of $f^{(j)}$.

We base the selection of the knots on the bid arrival distribution. Consider again Figure 3, which shows that over 60% of the bids arrive during the last day of the auction. Moreover, the phenomenon of bid sniping (Roth and Ockenfels 2002) suggests that auctions should be sampled more frequently at their later stages. Also, Shmueli et al. (in press) found that the bid intensity changes significantly during the last 6 hours. Motivated by this empirical evidence, we place a total of 14 knots and distribute the first 50% equally over the first 6 auction days. Then, we increase the intensity by placing the next 3 knots at every 6 hours, between day 6 and day 6.75. We again increase the intensity over the final auction moments by placing the remaining 4 knots every 3 hours, up to the end of the auction. This results in a total set of smoothing spline knots given by $\Upsilon = \{0, 1, 2, 3, 4, 5, 6, 6.25, 6.5, 6.75, 6.8125, 6.875, 6.9375, 7\}$.

We use smoothing splines of order $m = 5$ because this choice allows for a reliable estimation of at least the first three derivatives of f (Ramsay and Silverman 2005). Note that our results are robust to changes in the knot allocation and with respect to the choice of λ (see the App. for a sensitivity study).

Figure 6 shows the recovered functional object for a typical auction. Figure 6(a) shows the curve pertaining to the price evolution $f(t)$ on the log-scale (solid line), together with the actual bids (crosses), and Figures 6(b)–6(d) show the first, second, and third derivatives of $f(t)$, respectively. The price evolution shows that price, as expected from an auction, increases monotonically toward the end. However, the rate of increase does not remain constant. Although the price evolution resembles almost a straight line, the finer differences in the change of price increases can be seen in the price velocity $f'(t)$ [the first derivative of $f(t)$] or in the price acceleration $f''(t)$ (its second derivative). For instance, whereas the price velocity increases

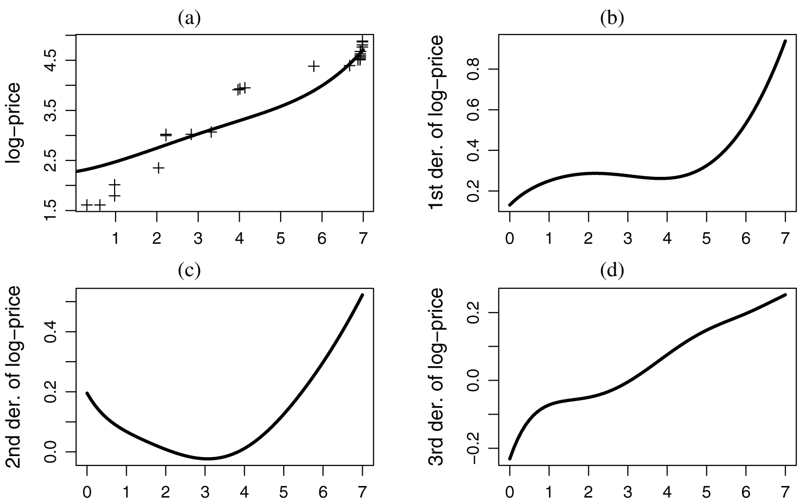


Figure 6. Price dynamics for Xbox auction 10. (a) Price evolution; (b) price velocity; (c) price acceleration; (d) price jerk.

at the beginning of the auction, it stalls after day 3 and remains low until the end of day 5, only to rise again and to sharply increase toward the end of the auction. Acceleration precedes velocity, and we can see that a price deceleration over the first day is followed by a decline in price velocity after day 1. In a similar fashion, the third derivative $[f'''(t)]$ measures the change in the second derivative. The third derivative is also referred to as the "jerk," and we can see that the jerk increases steadily over the entire auction duration, indicating that price acceleration is constantly experiencing new forces that influence the dynamics of the auction. Similar changes in auction dynamics have also been noted by Jank and Shmueli (in press).

3.2 The Mechanics of Functional Regression Models

In this section we briefly review the general mechanics of functional regression models for a functional response variable. For a more detailed description, see Ramsay and Silverman (2005, chap. 11).

Our starting point is an $N \times 1$ vector of functional objects $\underline{\mathbf{y}}(t) = [y_1(t), \ldots, y_N(t)]$, where N denotes the sample size, that is, the total number of auctions in this case. We use the symbol $y_j(t)$ in a rather generic way. To model the price evolution of an auction, we set $y_j(t) \equiv f_j(t)$. However, one of the advantages of functional data is that we also have estimates of the dynamics. For instance, to model an auction's price acceleration, we set $y_j(t) \equiv f_j''(t)$, and so forth. Classical regression models the response as a function of one (or more) predictor variables and that is no different in functional regression. Let $\underline{\mathbf{x}}_i = [x_{i1}, \ldots, x_{iN}]$ denote a vector of p predictor variables, $i = 1, \ldots, p$. x_{ij} can represent the value of the opening bid in the jth auction or, alternatively, its seller rating. Time-varying predictors can also be accommodated in this setting. For instance, $x_{ij}(t)$ can represent the number of bids in the jth auction *at time t*, which we refer to as the *current number of bids at t*. Operationally, one can include such a time-varying predictor into the regression model by discretizing it over a finite grid. Let x_{ijt} denote $x_{ij}(t)$ evaluated at t, for a suitable grid $t = t_1, \ldots, t_G$. We collect all predictors (time varying and time constant) into the matrix $\mathbf{X}$. Typically, this matrix has a first column of ones for the intercept. Also, we could write $\mathbf{X} = \mathbf{X}(t)$ to emphasize the possibility of time-varying predictors, but we avoid it for ease of notation. We then obtain the functional regression model

$$\underline{\mathbf{y}}(t) = \mathbf{X}^T \underline{\boldsymbol{\beta}}(t) + \underline{\boldsymbol{\varepsilon}}(t), \qquad (4)$$

where the regression coefficient $\underline{\boldsymbol{\beta}}(t)$ is time dependent, reflecting the potentially varying effect of a predictor at varying stages of the auction.

Estimating the model (4) can be done in different ways (see Ramsay and Silverman 2005 for a description of different estimation approaches). We choose a pointwise approach, that is, we apply regular least squares to (4) for a fixed $t = t^*$, and repeat that process for all t on a grid, $t = t_1, \ldots, t_G$. By smoothing the resulting sequence of parameter estimates $\hat{\beta}(t_1), \ldots, \hat{\beta}(t_G)$, we obtain the time-varying estimate $\hat{\boldsymbol{\beta}}(t)$.

Although functional regression is, at least in principle, very similar to classical least squares regression, attention has to be paid to the interpretation of the estimate $\hat{\boldsymbol{\beta}}(t)$. We re-emphasize that because the response is a continuous curve, so is $\hat{\boldsymbol{\beta}}(t)$. This makes reporting and interpreting the results different from classical regression and slightly more challenging. We show how this is done in the next section.

3.3 Empirical Application and Results

We fit the functional regression model (4) to our data and investigate two different models: The first model investigates the effect of different predictor variables on the *price evolution*; that is, we set $y_j(t) \equiv f_j(t)$. The results are shown in Figure 7. The second model investigates the effect of the same set of predictors on the *price velocity*; that is, $y_j(t) \equiv f_j'(t)$. Those results are shown in Figure 8. For both models, we use the nine predictor variables described in Tables 1 and 2. Figures 7 and 8 show the estimated parameter curves $\hat{\boldsymbol{\beta}}(t)$ (solid lines) together with 95% confidence bounds (dotted lines) indicating the significance of the individual effects. The confidence bounds are computed pointwise by adding ± 2 standard errors at each point of the parameter curve (Ramsay and Silverman 2005).

Interpretation of the parameter curves has to be done with care. At any time point t, $\hat{\boldsymbol{\beta}}(t)$ evaluated at t indicates the sign and strength of the relationship between the response (i.e., price in Fig. 7; velocity in Fig. 8) and the corresponding predictor variable. The time-varying curve underlines the time-varying nature of this relationship. The confidence bounds help in assessing the statistical significance of that relationship.

The insight from Figures 7 and 8 is summarized as follows.

Mechanism Design. We see that the choices that a seller makes regarding the opening bid and inclusion of a secret reserve price affect price according to what auction theory predicts: Higher opening bids and inclusion of a secret reserve price are associated with higher price, at any time during the auction (see Fig. 7). What has not been shown in previous studies, though, is the fact that this relationship, for both predictors, holds throughout the auction, rather than only at the end. Even more interesting is the observation that high opening bids and usage of a reserve price influence the price dynamics negatively toward the end of the auction by depressing the price velocity (see the negative coefficients in Fig. 8). In both cases, this is most likely because price has already been inflated by the high opening bid and/or the driving bidding force of the unobserved reserve price. We describe each of these two effects in more detail below.

Opening bid. The coefficient for opening bid in the regression on price evolution curves is shown in the middle-left panel in Figure 7. Throughout the entire auction, the coefficient is positive, indicating a positive relationship between the opening bid and the price at any time during the auction. However, the coefficient does decrease toward the end of the auction, indicating that although the positive relationship between opening bid and price is strong at the start of the auction, it weakens as the auction progresses. One possible explanation is that at the start of the auction, in the absence of other bids, auction participants derive of information from the opening bid about their own valuation. As the auction progresses, this source of information decreases in importance, and participants increasingly look to other sources (e.g., number of competitors, number of bids and their magnitude, communication with the seller, etc.) for decision making. In addition, the coefficient for opening bid in the regression on price velocity (Fig. 8) is negative throughout the auction and strongest at the start

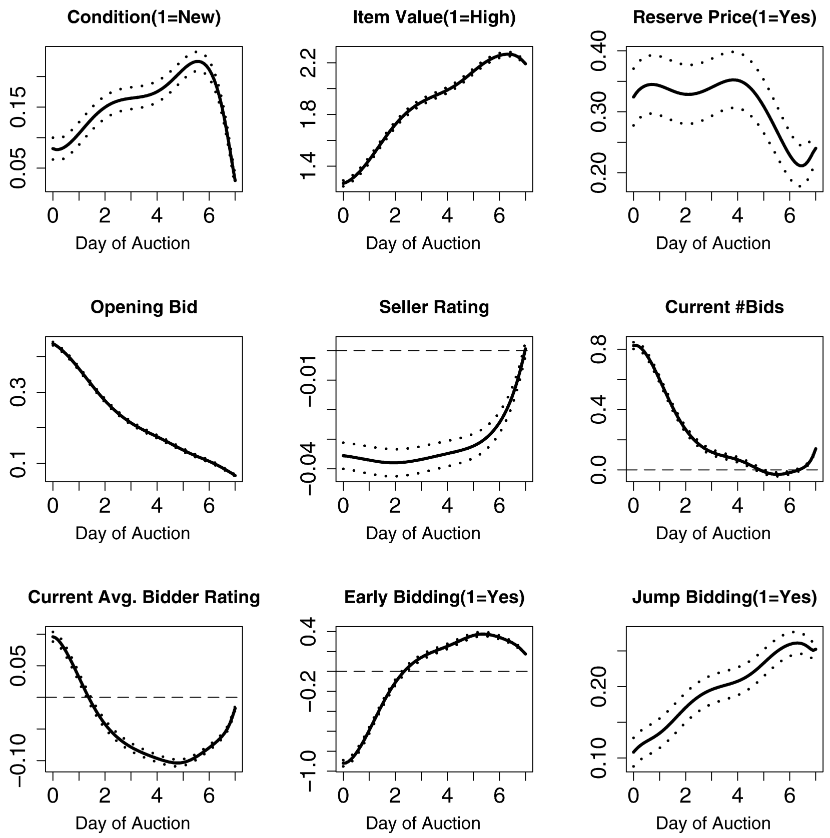


Figure 7. Estimated parameter curves based on functional regression on the price evolution. The x axis denotes the time of the 7-day auction. The dotted lines correspond to 95% pointwise confidence bounds.

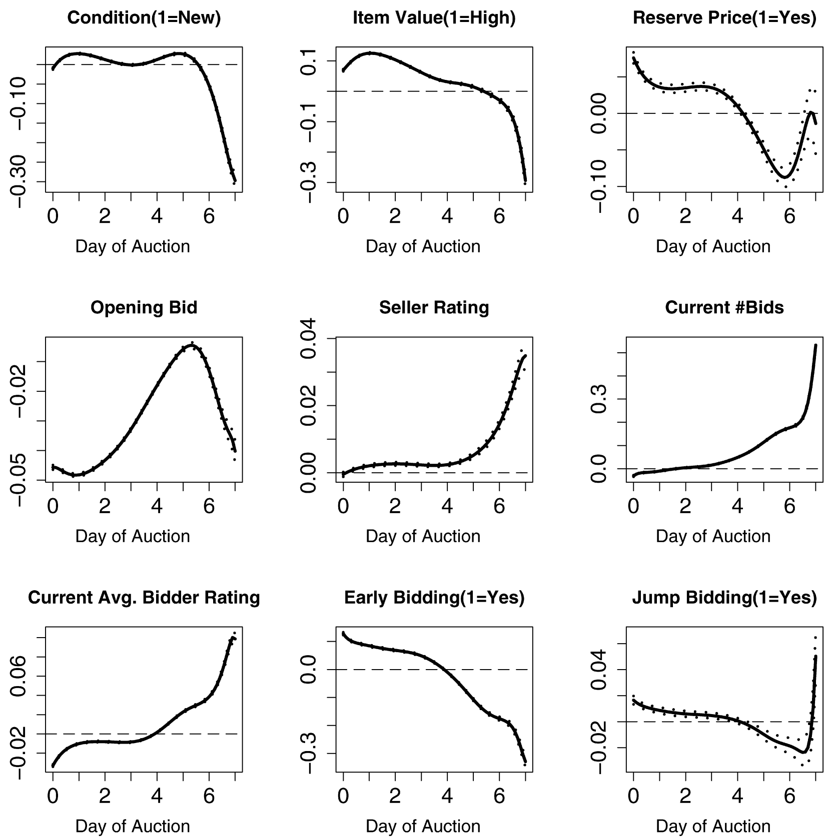


Figure 8. Estimated parameter curves based on functional regression on the price velocity. The x axis denotes the time of the 7-day auction. The dotted lines correspond to 95% pointwise confidence bounds.

and end of the auction. This indicates that higher opening bids depress the rate of price increase, especially at the start and end of the auction. Thus, although higher opening bids are generally associated with higher prices at any time in the auction (Fig. 7), the auction dynamics are slowed down by high opening bids. In some sense, higher opening bids leave a smaller gap between the current price and a bidder's valuation and, therefore, less incentive to bid.

Reserve price. Auctions with a secret reserve price tend to have a higher price throughout the auction, but a reserve price, similar to the high opening bid, appears to negatively influence price dynamics.

Seller Characteristics. The anonymity of the Internet makes it hard to establish trust. A seller's rating is typically the only sign that bidders look to in order to evaluate a seller's trustworthiness (Dellarocas 2003).

Seller rating. Empirical research has shown that higher seller ratings are associated with higher final prices (e.g., Lucking-Reiley et al. 2000; Ba and Pavlou 2002). Figures 7 and 8, though, show that higher seller ratings are associated with lower prices during the entire auction, except for the end of the auction. Moreover, higher seller ratings are associated with faster price increases, but again only toward the end of the auction.

Item Characteristics. The items in our dataset are characterized by condition (used vs. new) and by value (high for Xbox and low for Harry Potter).

Used/new condition. Overall, new items achieve higher prices, which is not surprising. However, the relationship between item condition and price velocity is negative. The price of new items appears to increase faster than used items *earlier*, but slows down *later* in the auction when used item prices increase at a faster pace. Perhaps the uncertainty associated with used items leads bidders to search for more information (such as contacting the seller or waiting for other bids to be placed), thereby leading to delays in the price spurts.

Item value. As one might expect, high-value items see higher prices than low-value items throughout the auction, and this gap increases as the action proceeds. More interesting, the price dynamics are very similar in both low- and high-value items until about day 6, but then price increases much faster for low-value items. This is indicative of later bidding on low-value auctions, a phenomenon that we observed in the exploratory analysis. Bidders are more likely to bid early on high-value items, perhaps to establish their time priority.

Bidding Characteristics. Our data contain four variables that capture effects of bidders and bidding, namely, the current number of bids as a measure of level of competition, the current average bidder rating as a measure of bidder experience, and early and jump bidding as a measure of different bidding strategies. All variables share the feature that their impact changes sometime during the auction, thereby creating two phases in how they affect the price evolution and the price velocity. In some cases, different strategies (such as early vs. late bidding) lead to direct impacts on the price, but to more subtle effects on the price dynamics. For instance, the current number of bids affects the price evolution directly during the first part of the auction, whereas during the second part of the auction, it affects price only through the price dynamics. The opposite phenomenon occurs with early bidding. Thus, functional regression reveals that (1) bidding appears to have two phases and (2) price can be affected either directly or indirectly by the bidding process.

Current number of bids. This factor influences the price evolution during the first part of the auction, with more bids resulting in higher prices. However, this effect decreases toward the end of the auction where it only influences price through increasing price dynamics.

Early bidding. The effect of this factor switches its direction between the first and second part of the auction: At first, auctions with early bidding have higher dynamics but lower price evolution, but later this effect is reversed. This means that early bidding manifests itself as early increased price dynamics, which later turn into higher price curves.

Jump bidding. Auctions with jump bidding tend to have generally higher price curves, and especially high price dynamics close to the end of the auction. The jump bidding obviously causes the price curve to jump and the price velocity to peak at the time of the jump bid. When averaging over the entire set of auctions, the effect of a jump bid has its highest impact on price at the end of auction. This is not necessarily in contrast to Easley and Tenorio (2004), who examined the timing of jump bidding (rather than the time of its highest impact). They found that jump bidding is more prevalent early in the auction, which they explained by the strategic value of jump bidding for bidders.

Current average bidder rating. It appears that higher rated bidders are more likely to bid when the price at the start of the auction is high, compared to lower rated bidders (as reflected by the positive coefficient during the first day). But then they are able to keep the price lower throughout the auction (the coefficient turns negative). Toward the end of the auction, though, participation of high-rated bidders leads to faster price increases, which reduces the final price gap due to bidder rating.

4. DYNAMIC AUCTION FORECASTING VIA FUNCTIONAL DATA ANALYSIS

We now describe our dynamic forecasting model. We have shown in the previous section how unequally spaced data can be overcome by moving into the functional context and also that online auctions are characterized by changing price dynamics. Our forecasting model consists of four basic components that capture price dynamics, price lags, and information related to sellers, bidders, and auction design. First, we describe the general forecasting model, which is based on the availability of price dynamics. Then, we describe how to obtain forecasts for the price dynamics themselves.

4.1 The General Forecasting Model

Our model combines all information that is relevant to price. We group this information into four major components: (1) static predictor variables, (2) time-varying predictor variables, (3) price dynamics, and (4) price lags.

Static predictor variables are related to information that does not change over the course of the auction. This includes the opening bid, the presence of a secret reserve price, the seller rating, and item characteristics. Note that these variables are known at the start of the auction and remain unchanged over the duration of the auction. Time-varying predictor variables are different in nature. In contrast to static predictors, time-varying predictors *do* change during the auction. Examples of time-varying predictors are the number of bids at time t or the number of bidders and their average bidder rating at time t. Price dynamics can be measured by the price velocity, the price acceleration, or both. Finally, price lags also carry important information about the price development. Price lags can reach back to price at times $t-1$, $t-2$, and so on. This corresponds to lags of order 1, 2, and so forth.

We obtain the following dynamic forecasting model. Let $y(t|t-1)$ denote the price at time t, given all information observed until $t-1$. For ease of notation, we write $y(t) \equiv y(t|t-1)$. Our forecasting model can then be formalized as

$$y(t) = \alpha + \sum_{i=1}^{Q} \beta_i x_i(t) + \sum_{j=1}^{J} \gamma_j D^{(j)} y(t) + \sum_{l=1}^{L} \eta_l y(t-l), \quad (5)$$

where $x_1(t), \ldots, x_Q(t)$ is the set of static and time-varying predictors, $D^{(j)}y(t)$ denotes the jth derivative of price at time t, and $y(t-l)$ is the lth price lag. The resulting h-step-ahead prediction, given information up to time T, is then

$$\tilde{y}(T+h|T) = \hat{\alpha} + \sum_{i=1}^{Q} \hat{\beta}_i x_i(T+h|T) + \sum_{j=1}^{J} \hat{\gamma}_j \tilde{D}^{(j)} y(T+h|T) + \sum_{l=1}^{L} \hat{\eta}_l \tilde{y}(T+h-1|T). \quad (6)$$

The model (6) has two practical challenges: (1) price dynamics appear as coincident indicators and must, therefore, be forecasted *before* forecasting $\tilde{y}(T+h|T)$; and (2) the static predictor variables among the x_i's do not change their value over the course of the auction and must, therefore, be adapted to represent time-varying information. We explain these two challenges in more detail in the following discussion and present some solutions.

4.2 Forecasting Price Dynamics

The price dynamics $D^{(j)}y(t)$ enter (6) as coincident indicators. This means that the forecasting model for price at time t uses the dynamics from the same time period! However, because we assume that the observed information extends only until $t-1$, we must obtain forecasts of the price dynamics before forecasting price. This process is described next.

We model $D^{(j)}y(t)$ as a polynomial in t with autoregressive (AR) residuals. We also allow for covariates x_i. The rationale for these covariates is that dynamics are strongly influenced by certain auction-related variables such as the opening bid (see Fig. 8). This results in the following model for the price dynamics:

$$D^{(j)}y(t) = \sum_{k=0}^{K} a_k t^k + \sum_{i=1}^{P} b_i x_i(t) + u(t), \qquad t = 1, \ldots, T, \quad (7)$$

where $u(t)$ follows an autoregressive model of order R:

$$u(t) = \sum_{i=1}^{R} \phi_i u(t-i) + \varepsilon(t), \qquad \varepsilon(t) \sim \mathrm{iid\,N}(0, \sigma^2). \quad (8)$$

To forecast $D^{(j)}y(t)$ based on (7), we first estimate the parameters $a_0, a_1, \ldots, a_K$, $b_1, \ldots, b_P$ and estimate the residuals. Then, using the estimated residuals $\hat{u}(t)$, we estimate $\phi_1, \ldots, \phi_R$. This results in a two-step forecasting procedure: Given information until time T, we first forecast the next residual via

$$\tilde{u}(T+1|T) = \sum_{i=1}^{R} \tilde{\phi}_i u(T-i+1) \quad (9)$$

and then use this forecast to predict the corresponding price derivative

$$D^{(j)}\tilde{y}(T+1|T) = \sum_{k=0}^{K} \hat{a}_k (T+1)^k + \sum_{i=1}^{P} \hat{b}_i x_i(T+1|T) + \tilde{u}(T+1|T). \quad (10)$$

In a similar fashion, we can predict $D^{(j)}y(t)$ h steps ahead:

$$D^{(j)}\tilde{y}(T+h|T) = \sum_{k=0}^{K} \hat{a}_k (T+h)^k + \sum_{i=1}^{P} \hat{b}_i x_i(T+h|T) + \tilde{u}(T+h|T). \quad (11)$$

4.3 Integrating Static Auction Information

The second structural challenge that we face is related to the incorporation of static predictors into the forecasting model. Take, for instance, the opening bid. The opening bid is static in the sense that its value is the same throughout the auction, that is, $x(t) \equiv x$, $\forall\, t$. Ignoring all other variables, we can rewrite model (5) as

$$y(t) = \alpha + \beta x. \quad (12)$$

Because the right-hand side of (12) does not depend on t, the least squares estimates of α and β are confounded!

The problem outlined previously is relatively uncommon in traditional time series analysis because it is usually only meaningful to include a predictor variable in an econometric model if the predictor variable itself carries time-varying information. However, the situation is different in the context of forecasting online auctions and may merit the inclusion of certain static information. The opening bid, for instance, may, in fact, carry valuable information for predicting price in the ongoing auction. Economic theory suggests that sometimes bidders derive information from the opening bid about their own valuation, but the impact of this information decreases as the auction progresses. What this suggests is that the opening bid can influence bidders' valuations and, therefore, also influence price.

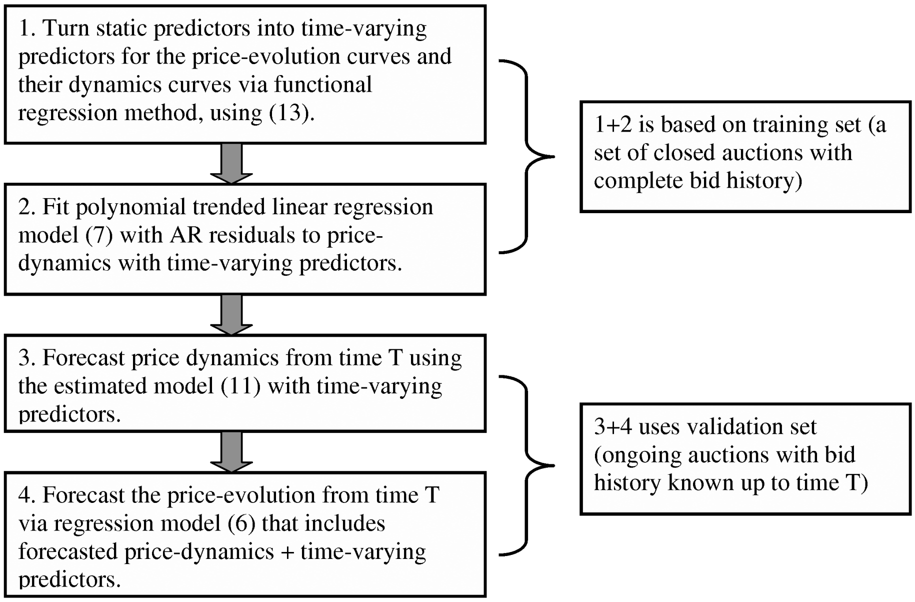


Figure 9. Flowchart of dynamic forecasting model.

What this also suggests is that the opening bid's impact on price does not remain constant but should be discounted gradually throughout the auction.

One way of discounting the impact of a static variable x is via its influence on the price evolution. That is, if x has a stronger influence on price at the beginning of the auction, then it should be discounted less during that period. On the other hand, if x only barely influences price at the end of the auction, then its discounting should be larger at the end of the auction. One way of measuring the influence of a static variable on the price curve is via functional regression analysis, as described in Section 3.2. Let $\tilde{\beta}(t)$ denote the slope coefficient from the functional regression model $y(t) = \alpha(t) + \beta(t)x + \varepsilon$, similar to (4); thus, $\tilde{\beta}(t)$ quantifies the influence of x on $y(t)$ at any time t. We combine x and $\tilde{\beta}(t)$ and compute the *influence-weighted* version of the static variable x as

$$\tilde{x}(t) = x\tilde{\beta}(t). \tag{13}$$

$\tilde{x}(t)$ now carries time-varying information and can, consequently, be included as a time-varying predictor variable.

As pointed out earlier, our dynamic forecasting model consists of two basic parts: one part forecasts the price dynamics, and the other part uses these forecasted dynamics as input into the price forecaster. A flowchart of our algorithm is shown in Figure 9.

5. EMPIRICAL APPLICATION AND FORECASTING COMPARISON

We apply the forecasting methodology to our dataset of 190 eBay auctions. Model fitting and prediction are implemented using modules of the R software package. We randomly partition our data into a training set (70%, or 130 auctions) and a validation set (30%, or 60 auctions). We use the training set to estimate the model, and we test the method on the validation set. For testing, we first remove all price information from the last auction day and then compare our results with the true price.

5.1 Model Estimation

Estimation of the model is done in two steps. We first estimate model (7) and then use the forecasted dynamics as inputs in model (6).

5.1.1 Modeling Price Dynamics. Model (7) is fitted iteratively. This leads to a best fitting model with a quadratic trend ($K = 2$) and three predictors ($P = 3$), where x_1, x_2, and x_3 are the influence-weighted variants of the opening bid, the item value, and jump bidding, respectively. The resulting residuals are AR(1), that is, $R = 1$ in (8). Figure 10 shows the significance of x_1, x_2, and x_3 over the last auction day in the form of *significance curves*. Because we use x_1, x_2, and x_3 to predict

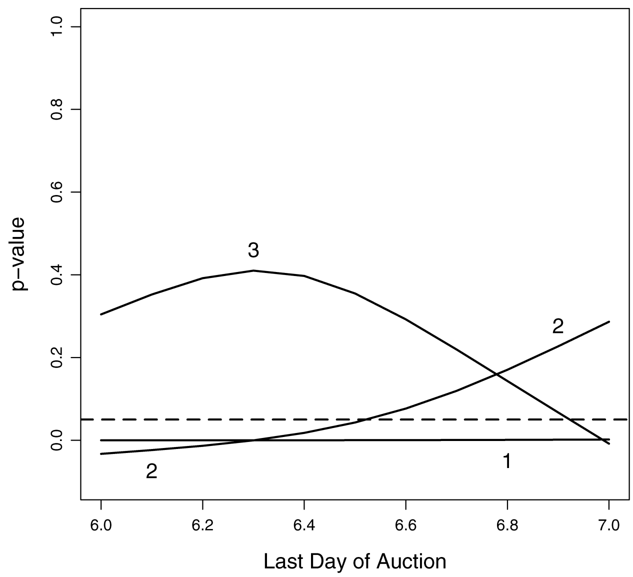


Figure 10. p-value curves for x_1, x_2, and x_3 over the last auction day. Consistent with the three predictors, we denote 1 = opening bid, 2 = item value, and 3 = jump bidding. The dotted horizontal line marks the 5% significance level.

price dynamics for all time points between days 6 and 7, the significance of individual predictors may be different at different time points. Indeed, Figure 10 shows that while jump bidding (line 3 in the graph) is insignificant during the beginning of day 6 (with a huge spike around day 6.3), it becomes significant toward the end of the auction. The opposite is true for the item value, which becomes insignificant at the end of the auction. On the other hand, the opening bid remains significant throughout the last day. This change in significance suggests that the "burden of prediction" does not remain equally distributed over all three predictors. In fact, the burden is heavier on item value at the beginning of the auction and then shifts to jump bidding at the end of the auction. Meanwhile, the opening bid carries the same prediction burden throughout the last day.

Figure 11 illustrates the forecasting performance on the holdout sample. We chose four representative auctions and compared the true price velocity over the last day (solid line) with its prediction based on model (7) (broken line). We see that the model captures the true price dynamics very well.

5.1.2 Modeling Price. We estimate model (5) using the following 11 predictor variables (grouped by their type):

Influence-weighted static predictors. Opening bid, Reserve price, Seller rating, Item condition, Item value, Early bidding, and Jump bidding
Time-varying predictors. Current number of bids and Current average bidder rating
Price dynamics. Price velocity
Price lags. Price at time $t-1$.

Figure 12 shows the significance curves for all 11 predictors. Interestingly, reserve price, seller rating, current number of bids, and current average bidder rating are insignificant at the start of the auction. Whereas the significance of the latter two increases toward the end of the auction, seller rating becomes even more insignificant. On the other hand, whereas reserve price becomes highly significant at the end of the auction, item condition, which is significant at the start, becomes insignificant at the end. All remaining predictors remain at (or below) the 5% significance mark throughout the entire auction.

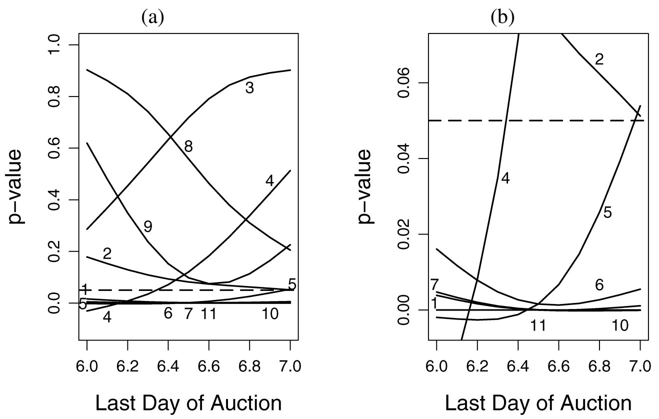


Figure 12. p-value curves for all 11 predictors over the last auction day. 1 = opening bid, 2 = reserve price, 3 = seller rating, 4 = item condition, 5 = item value, 6 = early bidding, 7 = jump bidding, 8 = current number of bids, 9 = current average bidder rating, 10 = price velocity, and 11 = price at time $t-1$. The dotted horizontal line denotes the 5% significance level. Panel (b) shows the information from panel (a) "zoomed in" for p values between 0 and .08.

5.2 Price Forecasting

After estimating the model using the training set, we apply it to the validation set to obtain forecasts for the price on the last day. Because we removed all price information from the last auction day, we can measure prediction accuracy by comparing the true price with our forecast.

Figure 13 illustrates the forecasting method for four sample auctions. Each of the four graphs in Figure 13 contains three separate pieces of information: (1) the actual current auction price (a step function), (2) the functional price curve, and (3) the forecasted price curve. The actual current auction price is the

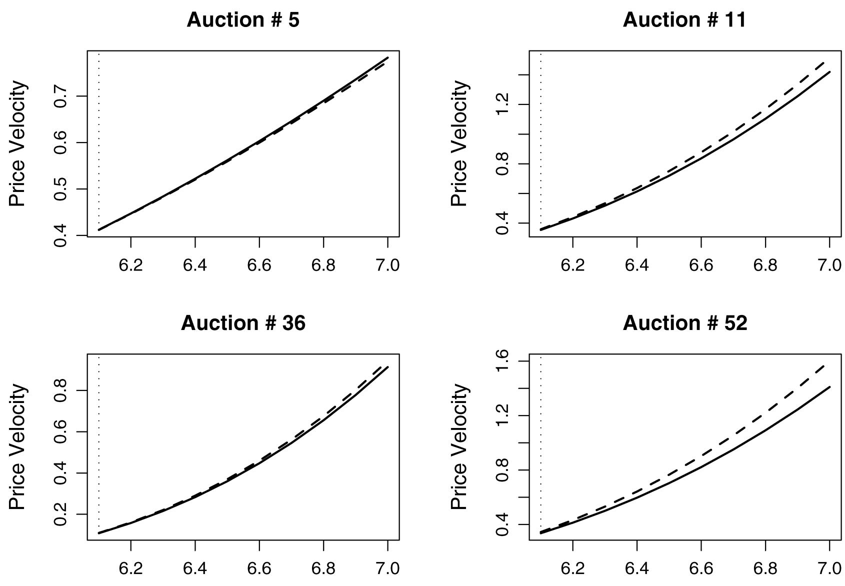


Figure 11. Forecasting performance of model (7) over the last auction day for four sample auctions (—— true price velocity; – – forecast price velocity).

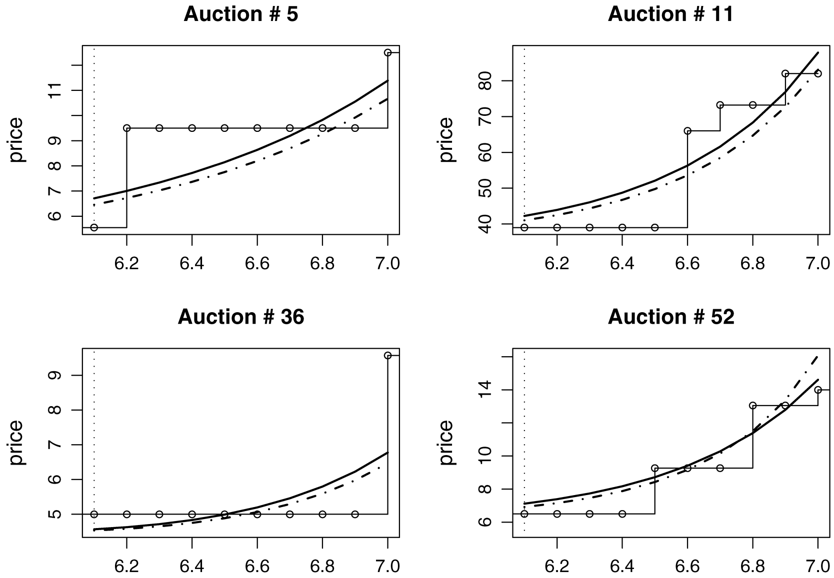


Figure 13. Dynamic forecasting results of last-day price for four sample auctions (—o— current auction price; —— functional price curve; · – · forecast price curve). The x-axis denotes time; the y-axis denotes price in \$.

price observed during the live auction. The functional price curve is the smoothed functional object based on the observed prices. And the forecasted price curve is our forecast based on model (5).

Notice that Figure 13 reveals two levels of "truth." The first is on the functional level, which compares true and forecasted *curves*. Our forecasting method operates on the functional objects and predicts the price curves. In that sense, the closer the forecasted curve is to the functional price curve, the better its functional prediction performance. Indeed, Figure 13 shows that the functional and forecasted curves are generally very close. However, the functional price curve is merely an approximation of the live auction price. Therefore, a second level of truth is revealed by comparing the forecasted price curve with the actual current auction price. On this level, the discrepancy is larger, which is not surprising: The quality of the forecasting output is only as good as its input. If the quality of the input is poor (i.e., functional objects that do not approximate the current auction price well), then not much can be expected of the forecasted output. This underlines the importance of generating high-quality functional objects. The most reliable way of checking the quality of the functional objects is via visualization. Jank, Shmueli, Plaisant, and Shneiderman (in press) proposed several ways of inspecting functional data visually. Another way of guaranteeing the quality of the results is via sensitivity studies with respect to the allocation of knots and the choice of the smoothing parameter (see the App.).

5.2.1 Forecast Accuracy. We measure forecast accuracy on the validation set using the mean absolute percentage error (MAPE). We compute the MAPE in two different ways, similar to Figure 13, once between the forecasted curve and the true functional curve (MAPE_1) and then between the forecasted curve and the actual current auction price (MAPE_2). The result is shown in Figure 14.

Naturally, MAPE_2 is higher than MAPE_1, because it is harder to reach the second level of "truth" compared with the first level. MAPE_1 is, at least on average, less than $+5\%$ for the entire prediction period (i.e., over the last day), implying that our model has a very high forecasting accuracy. MAPE_2 is a bit larger in magnitude because of the inevitable variation in fitting smoothing splines to the observed data. The widths of the confidence bounds underline the heterogeneity across all auctions in our dataset.

5.2.2 Forecast Accuracy by Auction Characteristics. Forecast accuracy can lead to new insight on the empirical regularities of bidding when breaking it down by different auction characteristics. We, therefore, compare forecast accuracy for different levels of the opening bid, secret reserve price, item condition and value, seller reputation, bidder experience, competition, and early and jump bidding. Table 3 shows the results. We find that the error is generally relatively small, no larger than 20% of the true functional price curve, and no larger than 36% of the actual final auction price. But there are subtle differences across the different variables: The error is larger when forecasting new items as compared to used items. High-value items, on the other hand, have a smaller error than low-value items, which could be attributed to the fact that, when the stakes are higher, bidders spend more time researching the

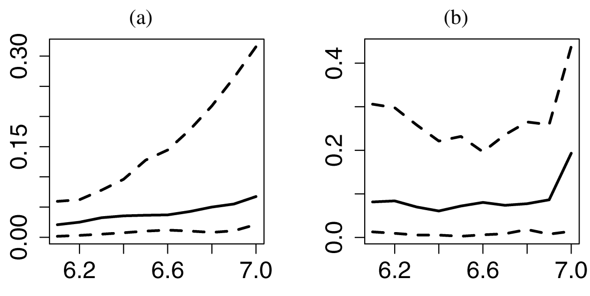


Figure 14. Mean absolute percentage errors (MAPEs). MAPE_1 (a) is the error between the forecasted price curve and the true functional price curve; MAPE_2 (b) is the error between the forecasted price curve and the actual current auction price. The dotted lines correspond to the 5th and 95th percentiles.

Table 3. Mean absolute percentage errors (MAPEs) broken up by different variables

		$MAPE_1$		$MAPE_2$	
Variable	Case	Mean	Standard error	Mean	Standard error
Reserve policy	Yes	.08	N/A	.16	N/A
	No	.12	.02	.23	.02
Condition	New	.17	.05	.31	.05
	Used	.09	.01	.19	.02
Item value	High	.07	.01	.14	.01
	Low	.16	.04	.31	.04
Opening bid	High	.06	.01	.14	.02
	Low	.20	.04	.36	.04
Seller rating	High	.14	.04	.26	.04
	Low	.09	.02	.20	.03
Average bidder rating	High	.15	.04	.30	.04
	Low	.09	.02	.17	.02
Number of bids	High	.13	.03	.24	.03
	Low	.10	.02	.22	.03
Early bidding	Yes	.11	.02	.22	.03
	No	.12	.03	.24	.04
Jump	Yes	.09	.01	.27	.03
	No	.13	.03	.21	.03

NOTE: $MAPE_1$ is the error between the forecasted final price and the functional final price; $MAPE_2$ is the error between the forecasted final price and the actual final price. The standard error of reserve price is "N/A" because there is only one auction with a reserve price in the validation set.

item and, thus, price dispersion is lower. Not surprisingly, auctions with a high opening bid have a smaller forecasting error because, when the opening bid is high and the item's value is relatively well known as in our situation, there is less uncertainty about the possible outcomes of the auction. Lower seller reputation results in more accurate forecasts. This may be due to the fact that higher seller ratings often elicit price premiums (Lucking-Reiley et al. 2000), thereby increasing the price variance. Bidder experience has a similar impact on forecasting accuracy. As for bidding competition (captured by the number of bids), higher competition results in larger variation in the forecast errors. It is also interesting to note that early bidding has barely any effect on the predictability of an auction; this again is different for jump bidding.

5.2.3 Comparison With Exponential Smoothing. To benchmark the performance of our method, we compare it to double exponential smoothing. Double exponential smoothing is a popular short-term forecasting method that assigns exponentially decreasing weights as the observations become less recent and also takes into account a possible (changing) trend in the data. This method cannot be applied directly to the raw bid data because of their uneven spacing. Functional objects once again come to the rescue; that is, we apply double exponential smoothing to a grid of evenly spaced values from the functional curve. The dashed lines in Figure 15 show the performance of exponential smoothing for the same four auctions as in Figure 13. We see that the predictions based on exponential smoothing are very far from the true auction price and even far from the true functional price curve. Table 4 compares our forecasting system with exponential smoothing in terms of MAPE. We find that the forecast error of exponential smoothing is more than twice the error of our forecasting system.

6. CONCLUSION AND FUTURE DIRECTIONS

In this article we propose a dynamic forecasting model for price in online auctions. We set up the forecasting problem in the context of functional data analysis by treating the price evolution in an auction as a functional object. This leads to a novel

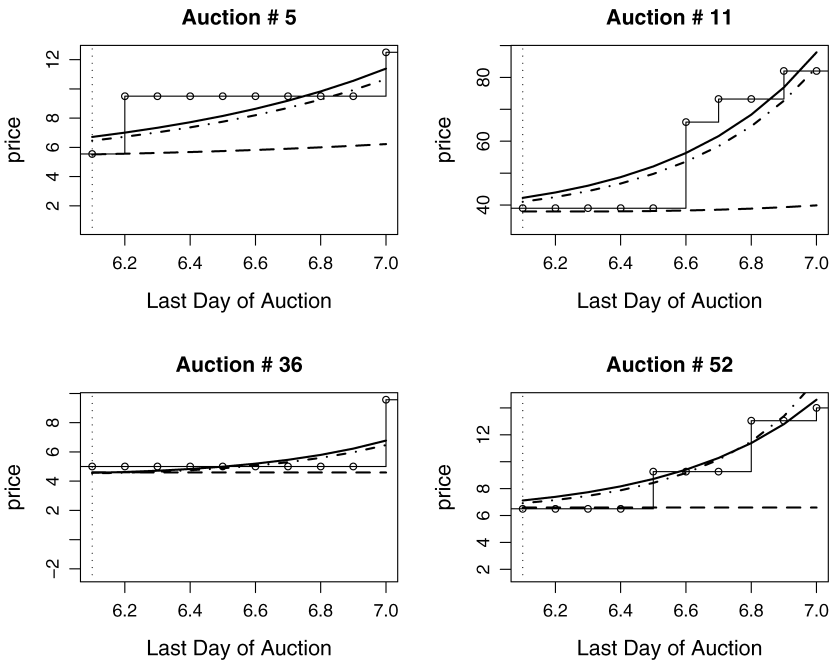


Figure 15. Comparison of forecasting results of last-day price evolution of individual auctions between using the exponential smoothing method and our dynamic forecasting method (–o– current auction price; —— functional price curve; – – exp. smoothing forecast; · – · dynamic forecast).

Table 4. Comparison of forecasting accuracy between our dynamic forecasting model and exponential smoothing

	MAPE$_1$		MAPE$_2$	
Method	Mean	Standard error	Mean	Standard error
Dynamic forecasting	.12	.02	.23	.02
Exponential smoothing	.42	.03	.49	.03

NOTE: The forecasting accuracy is measured by the mean absolute percentage error (MAPE).

use of FDA for forecasting, which has not been considered in the literature to date. It is also new in that it allows dynamic forecasting of an ongoing auction. The functional setup allows us to (1) represent the extremely unevenly spaced series of bids in a compact form; (2) estimate price dynamics via the derivatives of the smooth functional objects and integrate this dynamic information into the forecaster; and (3) incorporate both static and time-varying information about the auction into the forecasting system. Combining the dynamics with the static and time-varying information enables forecasting the price in ongoing live auctions for different types of products. The functional approach allows us also to investigate regularities of the bidding dynamics as a function of relevant auction dimensions.

We apply our forecasting system to real data from eBay on a diverse set of auctions and find that the combination of static and time-varying information creates a powerful forecasting system. The model produces forecasts with low errors, and it outperforms standard forecasting methods, such as double exponential smoothing, that severely underpredict the price evolution. This also shows that online auction forecasting is not an easy task. Whereas traditional methods are hard to apply, they are also inaccurate because they do not take into account the dramatic change in auction dynamics. Our model, on the other hand, achieves high forecasting accuracy and accommodates the changing price dynamics well.

This work can be extended in several ways. In this article we focus on auctions of the same duration. The lessons learned from this work can be used to extend the model to auctions of different length. Combining auctions of different durations is challenging because it involves registration of misaligned curves (see, e.g., Ramsay and Silverman 2005; James 2004). However, in the auction context, the misaligned curves are of different lengths which poses additional difficulties. Another extension is to incorporate a concurrency component. In online auctions, bidders have the option to inspect and follow multiple auctions at the same time. This places new challenges for modeling, especially in the functional framework. In a related series of articles (see Jank and Shmueli 2007; Hyde, Jank, and Shmueli 2006), we propose some solutions via visualization of concurrent functional objects and modeling of concurrent final prices. Finally, further research is required to better understand the exact role of price dynamics and their impact on economic theory. One possible avenue is the exploration of functional differential equation models in the auction context (see Jank and Shmueli in press).

ACKNOWLEDGMENTS

This research was funded in part by NSF grant DMI-02-05489. The authors gratefully acknowledge support by the Center of Electronic Markets and Enterprises, University of Maryland.

APPENDIX: SENSITIVITY ANALYSIS

The choice of our smoothing parameters is governed by reasonable fit. Because there is a wide range of choices that lead to reasonable curve approximations, we investigate the sensitivity of the forecasting accuracy to different choices of the knots and smoothing parameter λ. Table A.1 shows the forecasting accuracy in terms of MAPE$_1$ (between the forecasted price and the functional curve) and MAPE$_2$ (between the forecasted curve and the actual current auction price) for three different sets of knots. Similarly, Table A.2 shows the sensitivity to the choice of λ. In both cases, we see that the magnitude of the MAPE values remains in the area of 10–30%, with very little change in the standard errors.

Table A.2. Sensitivity analysis of λ selection (knots fixed to $\Upsilon 2$)

	MAPE$_1$		MAPE$_2$	
λ	Mean	Standard error	Mean	Standard error
.1	.28	.04	.32	.04
.3	.23	.03	.28	.03
.5	.21	.03	.28	.03
.7	.18	.02	.26	.03
.9	.16	.02	.25	.02
1	.16	.03	.27	.03
5	.15	.03	.27	.03
10	.12	.02	.24	.02
15	.12	.02	.23	.02
20	.12	.02	.23	.02
25	.12	.02	.23	.02
30	.12	.02	.23	.02
40	.11	.02	.23	.02
50	.11	.02	.23	.02

[Received December 2004. Revised May 2006.]

Table A.1. Sensitivity analysis of knot selection based on different knot scenarios

		MAPE$_1$		MAPE$_2$	
Set	Knots	Mean	Standard error	Mean	Standard error
$\Upsilon 1$	0, 1, 2, 3, 4, 5, 6, 6.25, 6.5, 6.75, 6.8750, 7	.18	.05	.29	.04
$\Upsilon 2$	0, 1, 2, 3, 4, 5, 6, 6.25, 6.5, 6.75, 6.8125, 6.8750, 6.9375, 7	.12	.02	.23	.02
$\Upsilon 3$	0, .5, 1, 1.5, 2, 3, 4, 5, 6, 6.25, 6.5, 6.75, 6.8125, 6.8750, 6.9375, 7	.26	.04	.31	.03

NOTE: $\Upsilon 2$ is the one used in this article.

REFERENCES

Ba, S., and Pavlou, P. A. (2002), "Evidence of the Effect of Trust Building Technology in Electronic Markets: Price Premiums and Buyer Behavior," *MIS Quarterly*, 26, 269–289.

Bajari, P., and Hortacsu, A. (2003), "The Winner's Curse, Reserve Prices and Endogenous Entry: Empirical Insights From eBay Auctions," *Rand Journal of Economics*, 3, 329–355.

——— (2004), "Economic Insights From Internet Auctions," Working Paper w10076, National Bureau of Economic Research.

Bapna, R., Goes, P., Gupta, A., and Jin, Y. (2004), "User Heterogeneity and Its Impact on Electronic Auction Market Design: An Empirical Exploration," *MIS Quarterly*, 28, 1.

Daniel, K. D., and Hirshleifer, D. (1998), "A Theory of Costly Sequential Bidding," technical report, Northwestern University, Kellogg Graduate School of Management.

Dellarocas, C. (2003), "The Digitization of Word-of-Mouth: Promise and Challenges of Online Reputation Mechanisms," *Management Science*, 49, 1407–1424.

Easley, R. F., and Tenorio, R. (2004), "Jump Bidding Strategies in Internet Auctions," *Management Science*, 50, 1407–1419.

Engle, R. F., and Russel, J. R. (1998), "Autoregressive Conditional Duration: A New Model for Irregularly Spaced Transaction Data," *Econometrica*, 66, 1127–1162.

Escabias, M., Aguilera, A. M., and Valderrama, M. J. (2004), "Modelling Environmental Data by Functional Principal Component Logistic Regression," *Environmetrics*, 16, 95–107.

Faraway, J. J. (1997), "Regression Analysis for a Functional Response," *Technometrics*, 39, 254–261.

Ghani, R., and Simmons, H. (2004), "Predicting the End-Price of Online Auctions," in *Proceedings of the International Workshop on Data Mining and Adaptive Modelling Methods for Economics and Management*, held in conjunction with the *15th European Conference on Machine Learning* (ECML/PKDDD) [on-line], available at *http:/www.accenture.com/xdoc/en/services/technology/publications/priceprediction.pdf/*.

Hortacsu, A., and Cabral, L. (2005), "Dynamics of Seller Reputation: Theory and Evidence From eBay," working paper, University of Chicago.

Hyde, V., Jank, W., and Shmueli, G. (2006), "Investigating Concurrency in Online Auctions Through Visualization," *The American Statistician*, 60, 241–250.

Isaac, M., Salmon, T. C., and Zillante, A. (2002), "A Theory of Jump Bidding in Ascending Auctions," *Journal of Economic Behavior and Organization*, 62, 144–164.

James, G. M. (2002), "Generalized Linear Models With Functional Predictors," *Journal of the Royal Statistical Society*, Ser. B, 64, 411–432.

——— (2004), "Curve Alignment by Moments," unpublished manuscript.

Jank, W., and Shmueli, G. (in press), "Studying Heterogeneity of Price Evolution in eBay Auctions via Functional Clustering," submitted to *Handbook on Information Series: Business Computing*, eds. G. Adomavicius and A. Gupta, Elsevier.

——— (2006), "Functional Data Analysis in Electronic Commerce Research," *Statistical Science*, 21, 155–166.

——— (2007), "Modeling Concurrency of Events in Online Auctions via Spatio-Temporal Semiparametric Models," *Journal of Royal Statistical Society*, Ser. C, 60, 1–27.

Jank, W., Shmueli, G., Plaisant, C., and Shneiderman, B. (in press), "Visualizing Functional Data With an Application to eBay's Online Auctions," in *Handbook on Computational Statistics on Data Visualization*, eds. C.-H. Chen, W. Härdle, and A. Unwin, Heidelberg: Springer-Verlag.

Krishna, V. (2002), *Auction Theory*, San Diego: Academic Press.

Lucking-Reiley, D., Bryan, D., Prasad, N., and Reeves, D. (2000), "Pennies From eBay: The Determinants of Price in Online Auctions," technical report, University of Arizona.

Ramsay, J. O. (2000), "Function Components of Variation in Handwriting," *Journal of the American Statistical Association*, 95, 9–15.

Ramsay, J. O., and Silverman, B. W. (2000), *Functional Data Analysis* (1st ed.), New York: Springer-Verlag.

——— (2005), *Functional Data Analysis* (2nd ed.), New York: Springer-Verlag.

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

Shmueli, G., and Jank, W. (2005), "Visualizing Online Auctions," *Journal of Computational and Graphical Statistics*, 14, 299–319.

Shmueli, G., Jank, W., Aris, A., Plaisant, C., and Shneiderman, B. (2006), "Exploring Auction Databases Through Interactive Visualization," *Decision Support Systems*, 42, 1521–1538.

Shmueli, G., Russo, R. P., and Jank, W. (in press), "The Barista: A Model for Bid Arrivals in Online Auctions," *Annals of Applied Statistics*.

Simonoff, J. S. (1996), *Smoothing Methods in Statistics*, New York: Springer-Verlag.