STATISTICAL METHODS IN ECOMMERCE RESEARCH

Chapter: A Family of Growth Models for Representing the Price Process in Online Auctions

Valerie Hyde
Applied Mathematics and Scientific Computation Program, University of Maryland, College Park, MD 20742, USA

Galit Shmueli
Department of Decision and Information Technologies, R. H. Smith School of Business, University of Maryland, College Park, MD 20742, USA

Wolfgang Jank
Department of Decision and Information Technologies, R. H. Smith School of Business, University of Maryland, College Park, MD 20742, USA

WILEY-INTERSCIENCE

A JOHN WILEY & SONS, INC., PUBLICATION

CHAPTER 1

A FAMILY OF GROWTH MODELS FOR REPRESENTING THE PRICE PROCESS IN ONLINE AUCTIONS

Abstract: Bids during an online auction arrive at unequally-spaced discrete time points. Our goal is to capture the entire continuous price-evolution function by representing it as a functional object. Various nonparametric smoothing methods exist to recover the functional object from the observed discrete bid data. Previous studies use penalized polynomial and monotone smoothing splines; however, these require the determination of a large number of coefficients and often lengthy computational time. We present a family of parametric growth curves that describe the price-evolution during online auctions. Our approach is parsimonious and has an appealing interpretation in the online auction context. We also provide an automated fitting algorithm that is computationally fast. Our method is illustrated on several eBay datasets.

1.1 INTRODUCTION AND MOTIVATION

With the advent of the Internet came online auction marketplaces, such as the popular eBay.com, which allow consumers and businesses to sell, buy, and bid on a variety of different goods. Ebay is the largest consumer-to-consumer (C2C) marketplace and touts net revenues that topped $1 billion for the first time for the first quarter of 2005 and were close to $1.4 billion (36% higher) for the first quarter of 2006. On any given day, several million items, dispersed across thousands of categories, are available for sale on eBay. Indeed, eBay's slogan is, "Whatever it is, you can find it on eBay." Buyers and sellers may be located on different continents and still conduct business since the online auction marketplace is always open and available. The wide and growing popularity of online auctions creates enormous amounts of publicly available auction bid data providing an interesting and important topic for research. These data also pose special statistical challenges because of their special structure. In this paper, we focus on the price process during an auction.

Bids during an online auction arrive at unevenly-spaced discrete time points chosen by bidders; We are interested in recovering the underlying continuous price evolution, i.e., the price at any time during the auction. From a visualization point of view, a series of discrete unevenly-spaced bids (e.g. presented as a scatterplot of bid vs. time) loses the conceptual as well as continuous nature of the price process. Furthermore, such plots do not scale to multiple auctions. A better representation is a continuous function with only a few parameters. Such a representation is conceptually more appealing, is more parsimonious, and can be further used for analyses such as clustering or regression models. Finally, the underlying continuous process may depend on derivatives of the function. A smooth price curve allows the calculation of derivatives (i.e., the first derivative is price velocity and the second derivative is price acceleration).

Functional data analysis (FDA) has become popular by the seminal works of Ramsay and Silverman (2002, 2005). At its core, functional data analysis deals with continuous objects such as curves or shapes as the observation of interest. For example, the temperature at a weather station over a period of one day may be considered a functional observation (curve) since it arises from a continuous process (the temperature at any moment in time over that day) even if it is measures at discrete time points (i.e., hourly). The curves representing daily temperature at several different weather stations can be considered a set of functional observations (Ramsay and Silverman, 2002, 2005). Similarly, the price during an online auction is recorded at discrete time points but is inherently continuous. We therefore treat a set of auctions as a set of functional observations.

In FDA, nonparametric smoothing techniques are used to recover a functional object from discrete measurements. Examples are kernel smoothers, polynomial splines, and monotone splines (Ramsay and Silverman, 2005; Ramsay, 1998). Smoothing should be done in such a way that the resulting underlying object adequately represents the continuous process. In the auction setting, we know that the price is always positive and monotonically nondecreasing, so our smoothed price curve must reflect this.

To date, all of the smoothing methods applied to online auctions have proven to involve the specification of many parameters such as the number and position of knots, a roughness penalty parameter, and the polynomial order. Resulting curves capture the price curves reasonably well; however, the fit often requires manual visual inspection (for parameter selection). Smoothing splines, which suffer from edge fitting, often result in a deteriorated fit at the start and end of the price curve and produce curves that are not necessarily monotone. Monotone smoothing splines produce monotone curves; however, they are computationally intensive and require lengthy run times for even a moderate number of auctions. Lastly, nonparametric smoothing does not provide an explanatory model for the price process. This motivated us to explore meaningful parametric representations of the price process. A parametric approach would be more elegant in the sense that it would provide a theoretical explanation of the process, it would potentially be computationally fast, and it would provide a more parsimonious representation.

In Hyde et al. (2006), for example, monotone smoothing is used for recovering the price curves in a set of Palm M515 auctions. They find three distinct shapes of price curves: "j", stretched-out "s", and straight line curves. The most popular shape in their dataset is a concave-up "j"-shaped price curve which represents auctions with gradual price increases until mid-to-late auction and then a price jump towards the end; the second most popular shape is the straight line, where the angle depends on the ratio of the opening and closing prices; and the third typical shape is a stretched-out "s"-shaped curve which reflects auctions where the price increases slowly, jumps up during mid-auction, and slowly increases to the close. These price curve shapes are the result of several bidding styles documented by Bapna et al. (2004b). Determining the number of each type of curve requires visual inspection of every single price curve. Clearly, a better method for grouping curves is needed beyond physical examination.

We can learn a lot about the auction price process by being able to group similar price curves as distinct parametric functions. For example, we can compare the price process distribution for distinct products or for market versus intrinsic values goods. We can also see if the distribution changes over time, which may suggest that bidders are evolving (since the price process is driven by bidders). We can also search for patterns in data sets using this additional information or use the particular growth function as a modeling variable.

The goal of this paper is to introduce, within an FDA framework, a new *parametric* family of growth functions that describe the price processes in online auctions. The paper is organized as follows: Section 1.2 discusses the characteristics of online auction data available on eBay.com, the set of data that we use throughout this work, and the representation of bid data as continuous curves. In Section 1.3, we discuss two nonparametric smoothing methods that have been used to describe auction price evolution and their limitations. In Section 1.4, we describe two popular growth models (exponential and logistic) and two additional useful growth modes (logarithmic and reflected-logistic) and show how our parametric family of growth models is used for representing auction price growth. Section 1.5 introduces an automated selection procedure to choose the best growth model. Section 1.6 compares the quality of curves obtained through nonparametric methods with those from our parametric family of curves. In Section 1.7, we suggest several further uses of growth models. We conclude in Section 1.8 with final remarks and future research.

1.2 DATA FROM ONLINE AUCTIONS

The number of different online auction sites is growing steadily. Despite different formats and rules, there is a common data structure that can be found across most sites. This structure comprises of a time series that describes the bids placed over time (the bid history) and an associated set of features that describe the auction-setting, such as the seller rating, the auction duration, and the item category. We refer to these features as the auction attributes. Figure 1.1. shows a snapshot of a closed auction from eBay.com showing the auction attributes (top) and the bid history (bottom). We see that this is a 7-day auction for a Vintage Rolex Submariner Black Dial Men's Wristwatch. The seller *tutleandwabbit* has a feedback rating of 100 with 100% positive feedback. The closing price is $2600.00, and there are a total of 13 bids from 9 bidders. In this case, eBay has decided to alias the bidders' user names in order to protect them from fake offers. Aliasing is performed quite often for high end merchandize.

To understand the structure of bid history data, it is necessary to understand the auction rules and bidding mechanism. On eBay, the majority of auctions are second-price auctions, which means that the winner is the bidder who placed the highest bid, but s/he pays the second highest price plus an increment. In our auction, Bidder 8 placed the highest bid but paid only Bidder 9's bid. Ebay does not disclose the highest bid (here, by Bidder 8). Furthermore, eBay uses a so-called “proxy bidding" system where bidders place the highest value that they are willing to pay, and then eBay bids on their behalf by increasing the current price by only an increment[1]. During the auction, the “current high bid" displayed is actually the second highest bid at the time plus an increment. Figure 1.2. shows this for the auction in Figure 1.1.. We call the current price “live bid". We see that Bidder 9 placed

[1] For further details see *http://pages.ebay.com/help/buy/proxy-bidding.html*

Table 1.1. Descriptive statistics for 472 completed 7-day eBay wristwatch auctions.

Variable	Mean(Std)	Median	Minimum	Maximum
Closing Price	$2019.00($2,561.89)	$1300.00	$70.00	$24,000.00
Opening Price	$509.70($896.96)	$100.00	$0.01	$6,500.00
Number of Bids	14.65(9.61)	13.00	2.00	57.00
Number of Unique Bidders	7.38(4.05)	7.00	2.00	21.00
Unique Bidders Rating	64.62(179.14)	11.00	-4.00	2,648.00
Seller Rating	571.99(1,505.94)	107.00	-2.00	9,055.00

a proxy bid of $2600 at 19:35:08; however, Bidder 8 placed an unknown higher bid at 18:22:24 so is credited with the highest bid and wins the auction.

1.2.1 Luxury Wristwatch Data

Our data contain information on 472 completed 7-day luxury wristwatch (375 Rolex and 97 Cartier) auctions on eBay.com that transacted between September 15, 2001 and October 27, 2001. Our sample includes a variety of items in terms of make and model, new and used, and closing price. The average selling price for all the auctions is $2019.00, with a median of $1300.00, and a standard deviation of $2561.89. We know from the literature (Bapna et al., 2004b; Wang et al., 2007) that auction attributes such as opening price, seller experience, number of bids, etc. affect not only the closing price of an auction but also the entire price process. Indeed, our sample is varied in all of those attributes. The range of opening prices is $0.01 to $6,500.00, the range of number of bids is 2 to 57, and the average seller experience is 571.99 with a standard deviation of 1505.94. Descriptive statistics are provided in Table 1.1..

The left panel of Figure 1.3. shows the distribution of the number of bids for each day in the auction. Typically, there is some bidding activity at the auction start (first day), followed by a period of very little activity, cumulating in a surge of bidding at the very end of the auction (Shmueli et al., 2004). This last-moment bidding is often referred to as “sniping" (Bajari and Hortascu, 2003; Roth and Ockenfels, 2002). Figure 1.3. shows a similar bid distribution for two additional datasets: Palm Pilot (middle) and Xbox (right) that will be discussed later. For all three products, most of the bidding occurs on the final day of the auction, with the next most bidding occurring on the first day. Clearly, when modeling the price process, the start and end of the auction are of particular importance.

eBay.com Bid History for
Vintage Rolex Submariner Black Dial Mens Wrist Watch (Item # 320068139586)
Listed in category: Jewelry & Watches > Watches > Wristwatches

Winning bid: US $2,600.00
Ended: Jan-10-07 19:43:14 PST
Starting time: Jan-03-07 19:43:14 PST
History: 13 bids
Starting bid: US $0.99

Winning bidder: keefer10 (120)

Seller: tutleandwabbitt (100)
Feedback: 100% Positive
Member: since Jan-06-04 in United States
Item location: Bayside, New York, United States

Bidder	Bid Amount	Date of bid
Bidder 8	US $2,600.00	Jan-10-07 18:22:24 PST
Bidder 9	US $2,600.00	Jan-10-07 19:35:08 PST
Bidder 7	US $2,500.00	Jan-10-07 13:07:05 PST
Bidder 6	US $2,300.00	Jan-10-07 05:46:18 PST
Bidder 5	US $2,250.00	Jan-09-07 01:23:57 PST
Bidder 3	US $2,150.00	Jan-04-07 11:55:31 PST
Bidder 4	US $2,000.00	Jan-04-07 13:43:07 PST
Bidder 3	US $1,500.00	Jan-04-07 11:55:24 PST
Bidder 1	US $1,000.00	Jan-04-07 06:40:26 PST
Bidder 3	US $1,000.00	Jan-04-07 11:55:14 PST
Bidder 3	US $500.00	Jan-04-07 11:55:06 PST
Bidder 3	US $100.00	Jan-04-07 11:54:54 PST
Bidder 2	US $55.56	Jan-04-07 09:38:09 PST

Figure 1.1. Time series and attributes for a men's Rolex wristwatch auction. Note that bids are arranged in descending order by bid amount. This order, however, does not reflect the arrival of the bids. Rather, it reflects the current auction high bid.

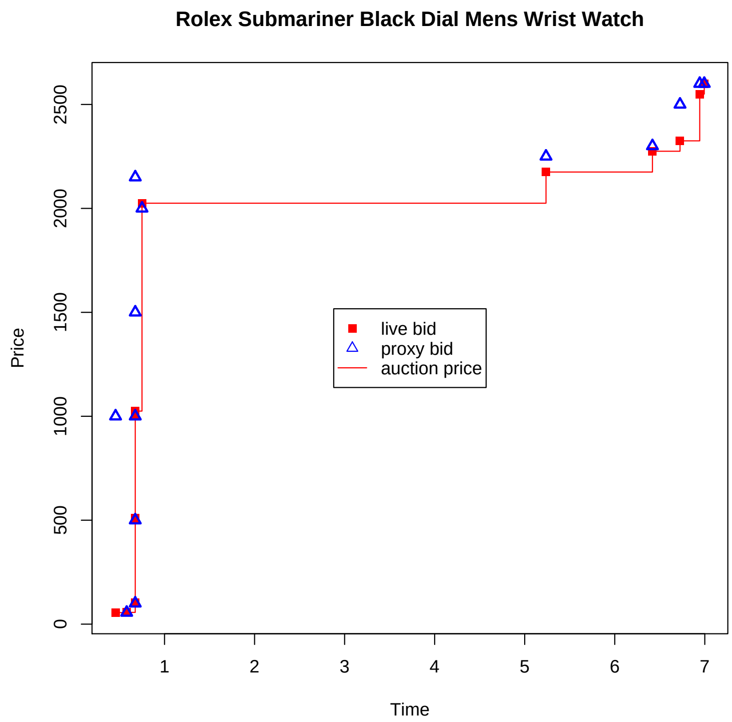


Figure 1.2. Proxy bids (triangles) and live bids (squares) for a men's Rolex wristwatch auction. The line connecting the live bids represents current auction price.

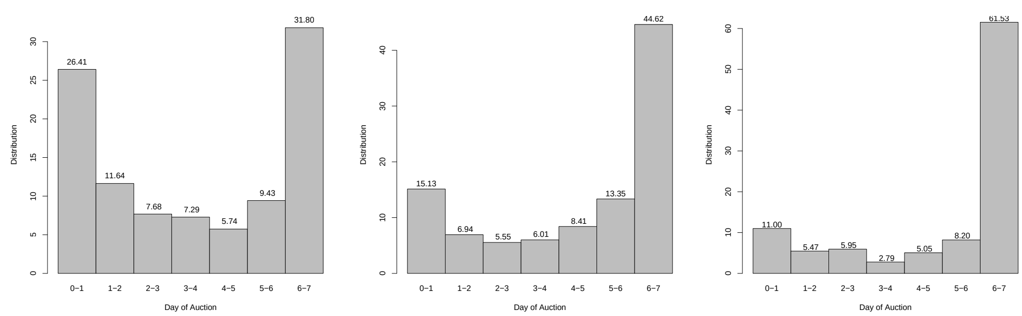


Figure 1.3. Distribution of daily bids over the course of 7-day auctions for 472 luxury wristwatch auctions (left), 134 Palm Pilot auctions (middle), and 85 Xbox auctions (right).

1.2.2 Representing Price Evolution as a Continuous Curve

Bids in online auctions are placed at varying time points. The resulting bid histories are therefore time series that are unevenly-spaced with sometimes very sparse and other times very dense areas. Instead of considering the discrete set of bids in an auction as a vector, we use them to estimate the complete continuous price evolution that takes place during the auction. While we could simply "connect the dots" to obtain the price of the auction at any given time, this would over fit the data (i.e., model the noise), thereby not providing a good representation of the underlying continuous price process. An alternative is to represent the price as a continuous smooth curve. This type of curve representation is prevalent in FDA (Ramsay and Silverman, 2005). The first step is therefore to represent/estimate the continuous price function from the discrete bid data.

Let y_j be the recording of an observation at time $t_j, j = 1, \ldots n$ in an auction with n bids. We convert the raw data into a continuous function, $f(t)$, that allows for the evaluation of the price at any point t during the auction. As with any measurement, there is error, so we have

$$y_j = f(t_j) + e(t_j) \tag{1.1}$$

where $e(t_j)$ is considered white noise (i.e., $e(t_j) \sim (0, \sigma^2)$). Different smoothing methods exist to recover the price function $f(t)$ and will be discussed in Section 1.3. In Section 1.4, we propose a parametric alternative which also produces continuous price curves.

The advantage of the curve representation is that it treats price evolution as a single continuous entity. It captures the complete price evolution in a more compact and easier-to-visualize way than raw bid data. The price process can then be described by a few coefficients. Furthermore, an appealing feature of smooth curves is that we can gauge their derivatives (the first derivative is the price-velocity and the second derivative is the price-acceleration) in order to learn how price dynamics behave during the auction.

1.3 REPRESENTING PRICE EVOLUTION NONPARAMETRICALLY

There have been predominantly two approaches for representing the price process in online auctions. Jank et al. (2007) use penalized polynomial smoothing splines (p-splines), and Hyde et al. (2006) use penalized monotone splines. A comparison of the two for auction data is given in Alford and Urimi (2004). We now describe each of the two methods and their properties.

1.3.1 Smoothing Splines

Each auction has bids placed at different times. Rather than applying smoothers to the raw data directly, we apply them to a derived data set that are sampled on the same set of time points for all auctions as follows. Consider the observed price during an auction, represented by the step function in Figure 1.2., where there is a new step every time a bid is placed. We use this step function to obtain our sampled data by selecting a set of knots at times $\tau_1, \tau_2, \ldots, \tau_L$ and determining the corresponding auction prices at those knots, $y_1^*, y_2^*, \ldots, y_L^*$. The noisy bid data are then discarded and replaced with the observations (τ_i, y_i^*). This transformation then allows us to use the same method to recover the price process in different auctions using the same smoother method.

The polynomial spline (Green and Silverman, 1994) of order p is given by

$$f(t) = \beta_0 + \beta_1 t + \beta_2 t^2 + \ldots + \beta_p t^p + \sum_{l=1}^{L} \beta_{pl}(t - \tau_l)_+^p \tag{1.2}$$

where $u+ = uI(u \geq 0)$ is the positive part of the function u. Many smoothers of this type tend to fit the data too closely (and thus model the noise); therefore, a *roughness penalty* approach is commonly employed. This method takes into account the trade off between data-fit (i.e., minimizing $f(t) = \sum_j (y_j^* - f(t_j))^2$) and function smoothness. A popular measure of roughness, which measures degree of departure from a straight line, is of the form

$$PEN_m(t) = \int [D^m f(t)]^2 dt \tag{1.3}$$

where $D^m f, m = 1, 2, 3, \ldots$ denotes the m^{th} derivative of the function f. A highly variable function will yield a high value of $PEN_m(t)$. If the highest derivative of interest is m, then using $m + 2$ as the polynomial order will assure m continuous derivatives. For online auctions, where the first and second derivatives have been shown to be of interest, we use polynomials of order $m = 4$. The penalized smoothing spline f minimizes the penalized squared error

$$PENSSE_{\lambda,m} = \int (y(t) - f(t))^2 + \lambda PEN_m(t). \tag{1.4}$$

When the roughness parameter is set to $\lambda = 0$, the penalized squared error drops out, and the function fits the data. Larger values of λ penalize the function for being curvy, and as $\lambda \to \infty$, the fitted curves approach a linear regression. Ramsay and Dalzell (1991) and Ramsay (1998) suggest that the smoothing parameter λ can often be chosen by inspection of the smooths or through optimizing metrics such as the generalized cross-validation (GCV).

We have encountered a number of challenges using penalized smoothing splines in the online auction context. First, and most detrimental, is that the created functions are not always monotone nondecreasing, as auction price necessarily must be. Second, the functions are often very wiggly, especially at the ends. This is particularly egregious in the online auction context where the start and end of the auction are of major importance. Third, there tends to be a large number of coefficients to estimate (due to adequate choices of knots and the polynomial order). Finally, there are multiple decisions about smoothing parameters that must be made in advance: the number and position of knots, the polynomial order, and the roughness penalty parameter λ, in order to provide a reasonable fit to the entire set of auctions. An alternative, where each auction is fit separately using a different set of parameters, will lead to confounding in the analysis results (see Jank et al. (2007)).

1.3.2 Monotone Splines

Since the bidding process by nature is nondecreasing, Hyde et al. (2006) use monotone smoothing splines to represent the price process. The idea behind monotone smoothing (Ramsay, 1998) is that monotone increasing functions have a positive first derivative. The exponential function has this property and can be described by the differential equation $f'(t) = w(t)f(t)$. This means that the rate of change of the function is proportional to its size. Consider the linear differential equation

$$D^2 f(t) = w(t)Df(t). \tag{1.5}$$

Here, $w(t) = \frac{D^2 f(t)}{Df(t)}$ which is the ratio of the acceleration and velocity. It is also the derivative of the logarithm of velocity which always exists (because we define velocity to be positive by the equation $Df(t) = e^{w(t)}$). The differential equation has the following solution:

$$f(t) = \beta_0 + \beta_1 \int_{t_0}^{t} exp\left(\int_{t_0}^{v} w(v)dv\right) du \tag{1.6}$$

where t_0 is the lower boundary over which we are smoothing. After some substitutions (see Ramsay and Silverman (2005)), we can write

$$f(t) = \beta_0 + \beta_1 e^{w(t)}. \tag{1.7}$$

and estimate β_0, β_1, and $w(t)$ from the data. Since $w(t)$ has no constraints, as $f(t)$ does in the form of the differential equation, it may be defined as a linear combination of K known basis functions (i.e., $w(t) = \sum_k c_k \phi_k(t)$). Examples of a basis functions are $\phi_k(t) = t$ which represents a linear model or $\phi_k(t) = log(t)$ which is a nonlinear transformation of the inputs. The penalized least squares criterion is thus

$$PENSSE_\lambda = \sum_i [y_i - f(t)]^2 + \lambda \int_0^T [w^2(t)]^2 dt. \tag{1.8}$$

While monotone smoothing solves the wiggly problem of the penalized smoothing splines, some of the same challenges remain and new ones arise. First, monotone smoothing is computationally intensive. The more bids, the longer the fitting process. Second, we still cannot fit the original raw

Table 1.2. Price-evolution, -velocity, and -acceleration functions for growth models.

Growth Model	Price Evolution	Price Velocity	Price Acceleration	Parameters
Exponential	$Y(t) = Ae^{rt}$	$Y'(t) = Are^{rt}$	$Y''(t) = Ar^2e^{rt}$	A, r
Logarithmic	$Y(t) = \frac{\ln(\frac{t}{A})}{r}$	$Y'(t) = \frac{1}{rt}$	$Y''(t) = \frac{-1}{rt^2}$	A, r
Logistic	$Y(t) = \frac{L}{1+Ce^{rt}}$	$Y'(t) = \frac{-LCre^{rt}}{(1+Ce^{rt})^2}$	$Y''(t) = \frac{-LCr^2e^{rt}(1-Ce^{rt})}{(1+Ce^{rt})^3}$	C, r
Reflected-Logistic	$Y(t) = \frac{\ln(\frac{L}{t}-1)-\ln(C)}{r}$	$Y'(t) = \frac{-L}{rt^2(\frac{L}{t}-1)}$	$Y''(t) = \frac{L(L-2t)}{rt^4(\frac{L}{t}-1)}$	C, r

data but rather the derived data (τ_i, y_i^*) (obtained from the step function). Finally, as with smoothing splines, the researcher must determine the number and location of knots and the roughness parameter λ that provide a reasonable fit to the entire set of auctions.

1.4 REPRESENTING PRICE EVOLUTION PARAMETRICALLY

We now introduce a parametric family of four growth models that are able to capture the price evolution in many types of auctions. These are exponential growth, logarithmic growth, logistic growth, and reflected-logistic growth functions. Exponential and logistic growth functions have long been used to model population growth, dissemination of information, spread of disease, and more.

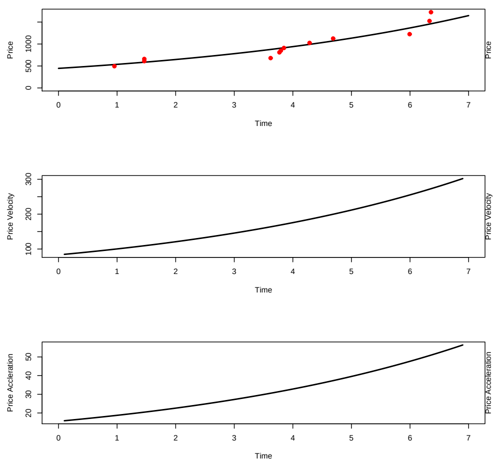
Our parametric approach is elegant, computationally fast, and parsimonious. It allows automated fitting, and there is no need to specify any parameters in advance. We choose models that are theoretically relevant in terms of monotonicity (to accurately reflect the price process) and that provide insight into the price process in online auctions.

Our approach, although motivated by online auctions, actually provides an alternative to the nonparametric smoothing methods that are customary in the field of Functional Data Analysis (FDA). This research opens the door for parametric curves to be the basis of FDA.

In the following, we describe each of the four models. Their functional form, derivative form, and parameters are summarized in Table 1.2..

1.4.1 Exponential Growth

1.4.1.1 Exponential Model Exponential growth has been used for describing a variety of natural phenomena including the dissemination of information, the spread of disease, and the multiplication of cells in a petrie dish. In finance, the exponential equation is used to calculate the value of interest-bearing accounts compounded continuously. The idea behind exponential growth



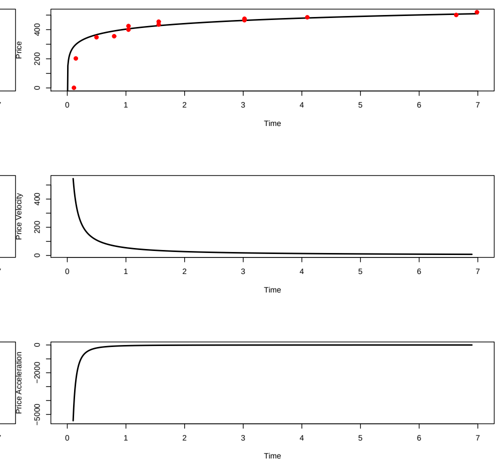


Figure 1.4. Price (top), velocity (middle), and acceleration (bottom) for an auction fit with exponential growth.

Figure 1.5. Price (top), velocity (middle), and acceleration (bottom) for an auction fit with logarithmic growth.

is that the rate of growth is proportional to the function's current size; that is, growth follows the differential equation

$$Y'(t) = rY(t), \tag{1.9}$$

or the equivalent equation

$$Y(t) = Ae^{rt}, \tag{1.10}$$

where t is time, and $r > 0$ is the growth constant. Equivalently, exponential decay, when $r < 0$, can model phenomena such as the half-life of an organic event.

From a theoretical standpoint, exponential growth can describe a price process for auctions where there are gradual price increases until mid-to-late auction and a price jump towards the end. This is reminiscent of the "j"-shaped price curves that Hyde et al. (2006) find. An example of an auction that is captured well with exponential growth is shown in Figure 1.4. (top). The corresponding velocity curve (middle) and acceleration curve (bottom) are proportional to the price curve, as the differential equation implies. The price-velocity and -acceleration are zero or small during most of the auction before spiking at the end.

1.4.1.2 Logarithmic Model We also find that the inverse of the exponential function,

$$Y(t) = \frac{\ln(\frac{t}{A})}{r}, \tag{1.11}$$

which is called logarithmic growth, approximate price processes well. We choose a form of the logarithmic model that is the mapping of the original exponential model over the line $y = x$. This type of growth occurs when early bidding increases the price early in the auction, but because of the existence of a market value, price flattens out for the remainder of the auction (as shown in Figure 1.5.). This type of price behavior tends to be rare, as most bidders do not wish to reveal their bids early in the auction. However, inexperienced bidders who do not understand the proxy bidding mechanism on eBay have been shown to bid high early (Bapna et al., 2004a). In this model, the velocity starts at its maximum and then decays to little or zero velocity as the auction progresses. The acceleration is always negative, since price increases more slowly throughout the auction, and approaches zero at the end of the auction (where very little change in price is occuring).

1.4.2 Logistic Growth

1.4.2.1 Logistic Model While exponential growth often makes sense over a fixed period of time, in many cases, growth can not continue indefinitely. For example, there are only a finite number of people to spread information or disease; the petrie dish can only hold a maximal number of cells. A typical application of the logistic equation is in population growth. In the beginning, there are seemingly unlimited resources and population grows increasingly fast. At some point, competition for food, water, land, and other resources slows down the growth; however, population is still growing. Finally, overcrowding, lack of food, and susceptibiltiy to disease limit the population to some maximal carrying capacity.

In auctions, it is possible that growth starts out exponentially with a big price increase in the middle of the auction, perhaps due to an inexperienced bidder. In the presence of a market value (or “carrying capacity"), the increased price slows competition, as smart bidders will make sure not to overpay for the item, and it is necessary for the price to flatten out near the end of the auction. The closing price that we witness is analogous to the carrying capacity L in the logistic growth function.

The logistic model is given by

$$Y(t) = \frac{L}{1 + Ce^{rt}}, \tag{1.12}$$

and the differential equation is

$$Y'(t) = rY(t)\left(\frac{Y(t)}{L} - 1\right), \tag{1.13}$$

where L is the carrying capacity, t is time, r is the growth rate, and C is a constant. Logistic growth forms a stretched-out “s"-shaped curve, discussed by Hyde et al. (2006), where the price increases slowly, then jumps up during mid-auction, and finally levels off through the end of the auction. This price process and dynamics can been seen in Figure 1.6.. The velocity is small or zero at the beginning

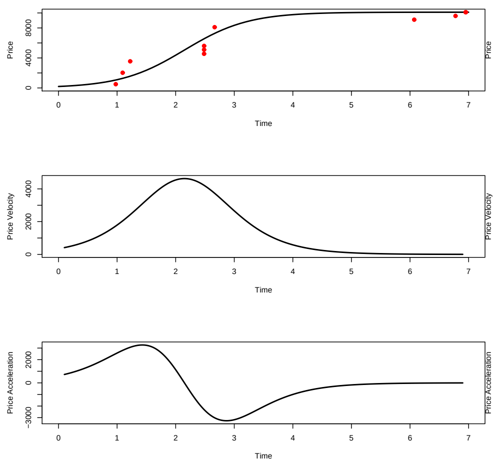


Figure 1.6. Price (top), velocity (middle), and acceleration (bottom) for an auction fit with logistic growth.

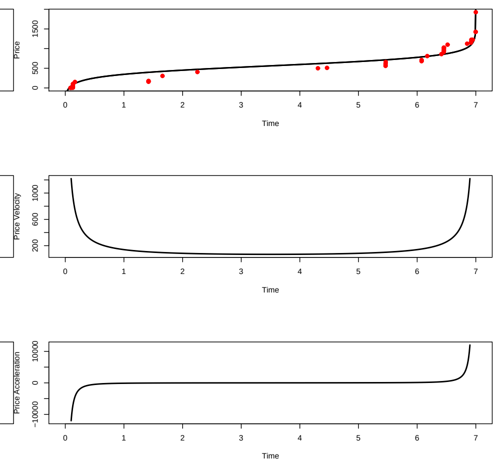


Figure 1.7. Price (top), velocity (middle), and acceleration (bottom) for an auction fit with reflected-logistic growth.

and end of the auction, where there is little change in price, with a peak midauction corresponding to the price increase. The acceleration is zero for most of the beginning and end of the auction. It peaks during the first part of the growth spike, where price is increasing at an increasing rate, followed by a valley during the second part of the growth spike, where price is increasing at a decreasing rate.

1.4.2.2 Reflected-Logistic Model Another common price process in online auctions is the inverse of logistic growth, or reflected-logistic growth, given by the function

$$Y(t) = \frac{\ln(\frac{L}{t} - 1) - \ln(C)}{r}. \tag{1.14}$$

This type of growth occurs when there is some bidding early in the auction that results in a price increase, followed by little to no bidding in the middle of the auction, and then another price increase as the auction progresses towards its close. The early price increases is indicative of early bidding by inexperienced bidders (Bapna et al., 2004a), and the price spike at the end may be caused by sniping (Bajari and Hortascu, 2003; Roth and Ockenfels, 2002). An example of reflected-logistic growth is shown in Figure 1.7.. The velocity has peaks at the start and end of the auction where there are jumps in price and little or zero velocity during the middle of the auction where price does not change significantly. The acceleration curve is similar in shape to the price curve; however, it is negative in the beginning (where price increases at a decreasing rate) and positive at the end (where price increases at an increasing rate).

1.4.3 Fitting Growth Models

Unlike the nonparametric smoothers, we fit the growth models directly to the live bids (t_j, y_j). A simple and computationally efficient method for fitting each of the growth models is by linearizing the function and then performing least squares. Since we are especially interested in obtaining an accurate fit at the beginning and end of the auction, it is usually necessary to add two additional points representing the price at the start and close of the auction: $(t = 0, y = \min(y_j))$ and $(t = l, y = \max(y_j))$ where l [2] is the duration of the auction, and y_j is the value of bid j. Note that $\min(y_j) =$ opening price and $\max(y_j) =$ closing price. It is not necessary to add these extra points for logistic growth, where the maximum price is already incorporated into the function by defining $L = \max(y_j)$. Further, empirical evidence shows that auctions whose underlying price process is logistic tend to start close to zero, where logistic growth must start.

1.4.3.1 Fitting Exponential Growth The exponential growth model from equation (1.10) can be linearized as

$$\ln Y = \ln A + rt. \tag{1.15}$$

We fit this model directly to the live bids with the two additional points in order to constrain the price at the start and end of the auction. In this case, two parameters, A and r are estimated. Although we can fix A as the opening price since $Y(t = 0) = Ae^{0r} = A$, we chose to estimate both parameters for two reasons: first, empirical evidence shows that the two-parameter estimation allows a better fit at the end of the auction, and second, the other three growth models require estimating two parameters. This, it is easier to compare model fit.

The straight line curves that are discussed in Hyde et al. (2006) are a special case of exponential growth. When $r = 0$, we get a horizontal straight line, and when r is close to 0, we get a price curve that resembles a straight line with positive slope.

1.4.3.2 Fitting Logarithmic Growth The logarithmic growth model is given in equation (1.11). As with exponential growth, the two extra points are added to the live bids to ensure a good fit at the start and end of the auction. Notice that we cannot reduce this function to one parameter because when $t = 0$, $\ln(0)$ does not exist. Also, we cannot linearize this function. Therefore, rather than using

[2] If all the auctions have the same duration, say 7-days, then l would be replaced by 7. If the auctions have different durations, we can either vary l accordingly for each auction or standardize the time for all auctions to be between 0 and 1.

optimization methods for parameter estimation where guesses of the initial value are necessary, we fit $T(y) = Ae^{ry}$ and linearize as in the exponential growth case. This time, least squares minimizes over time instead of price.

1.4.3.3 Fitting Logistic Growth The logistic growth model from equation (1.12), where L is the distribution's asymptote, can be linearized as

$$\ln\left(\frac{L}{y} - 1\right) = \ln(C) + rt. \tag{1.16}$$

We know that $\lim_{t\to l} = L$ (since for logistic growth $r < 0$ and $L > 0$). Define $L = \max(price) + \delta$, where $\delta = 0.01$ is needed so that the LHS is defined over all bids y. In this case, there is no need to add the start and closing points to the live bids because defining the asymptote takes care of the fit at the end. Auctions whose underlying price process can be described by logistic growth tend to start out low, so there is also no need to set the start value.

1.4.3.4 Fitting Reflected-Logistic Growth The reflected-logistic function is given in equation (1.14). As with logarithmic growth, we can not linearize this function. We instead fit $T(y) = \frac{L}{1+Ce^{ry}}$, where $L = l + \epsilon$ ($\epsilon = 0.00001$), to obtain the coefficients for C and r. We need $\epsilon << \delta$ since the time range is much smaller than the price range. As with logarithmic growth, least squares minimizes over time instead of price. Note that here, the extra points are $(t = 0.000001, y = \min(y_j))$ and $(t = l, y = \max(y_j))$ so that the LHS is defined over all bid times.

1.5 SELECTING THE BEST GROWTH MODEL

We develop an automated model selection procedure to choose for each auction the best fitting growth model among the four models. The procedure uses a specialized proximity metric that measures the distance between bids and the fitted curve in the two dimensions of time and price. This metric is reminiscent of the Mahalanobis distance. Most model selection criteria only measure the residual distance in the y (price, in this case) dimension; however, we are interested in capturing the best fit in both the price and time dimensions because bid times are informative and are random variables. Furthermore, the fit for the logarithmic growth and reflected-logistic growth models are minimized over the x (time) dimension. If we were to choose between models based simply on the price dimension, we would tend to choose the exponential growth and logistic growth models, even though the reflected models may provide a better representation of the price process (as can be visually seen). While our model selection criteria is primarily aimed at choosing among growth models, the metric may also be used to choose among other methods: growth models, p-splines, monotone smoothing, etc.

1.5.1 Model Selection Metrics

For auction i, let $\{\mathbf{t_i}, \mathbf{y_i}\}$ be the vector of live bids (t_{ij}, y_{ij}) where bid y_{ij} is placed at time t_{ij}, and the number of bids in the auction is n_i. Define a new vector with two additional price points ($n_i^\star = n_i + 2$) that also includes the open and close price of the auction as $\{\mathbf{t_i^\star}, \mathbf{y_i^\star}\} = \{(t = 0, y = \min(y_{ij})), \{\mathbf{t_i}, \mathbf{y_i}\}, (t = l, y = \max(y_{ij}))\}$, where l is the length of the auction. It is important that we examine the fit at the start and end of the auction because that is where most of the bid activity takes place, they are conceptually important, and also because that is where modeling often falls short.

We propose two measures of fit: the weighted sum-of-squares standardized by the range (WSSER) and the weighted sum-of-squares standardized by the variance (WSSEV). Both metrics are weighted averages of fit in the y-direction and fit in the x-direction, using weights w_y and w_x, such that $w_y + w_x = 1$. The WSSER for auction i is defined as

$$WSSER_i = \frac{w_y \sum_{j=1}^{n_i^\star} (y_{ij}^\star - \hat{y}_{ij}^\star)^2}{(\max_j(y_{ij}^\star) - \min_j(y_{ij}^\star))^2} + \frac{w_x \sum_{j=1}^{n_i^\star} (x_{ij}^\star - \hat{x}_{ij}^\star)^2}{(\max_j(x_{ij}^\star) - \min_j(x_{ij}^\star))^2}. \tag{1.17}$$

Notice that the denominator is the squared price range in the y dimension and the squared time range (in our case the auction length l) in the x dimension. The WSSEV for auction i is defined as

$$WSSEV_i = \frac{w_y \sum_{j=1}^{n_i\star} (y_{ij}^\star - \hat{y}_{ij}^\star)^2}{variance_j(y_{ij}^\star)} + \frac{w_x \sum_{j=1}^{n_i\star} (x_{ij}^\star - \hat{x}_{ij}^\star)^2}{variance_j(x_{ij}^\star)}. \tag{1.18}$$

1.5.2 Model Selection Procedure

In this section, we describe how to automatically select between different growth models and choose the best one. The model selection procedure is as follows:

1. Select weights, w_y and w_x, representing the importance of fit in the price and time dimensions, respectively.

2. Fit each of the four growth models to the live bids of an auction.

3. Compute model selection metric(s).

4. Choose model with best fit (minimum WSSE).

For our sample of auctions, we choose equal weights $w_y = w_x = \frac{1}{2}$ since we are equally interested in fit in the price and time dimensions. One may overweight time (large w_x) if capturing bid timing is of special interest. One such case is in studying bid shilling, where a seller may cancel the auction or illegally bid on their own auction if the price has not reached a certain level by a certain time (Kauffman

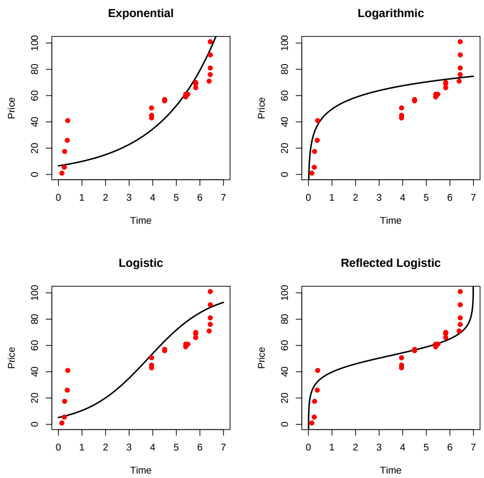


Figure 1.8. Exponential (top left), logarithmic (top right), logistic (bottom left), and reflected-logistic (bottom right) models fit to bids for the same auction. Overweighting in the y-dimension selects exponential growth when reflected-logistic growth should be chosen.

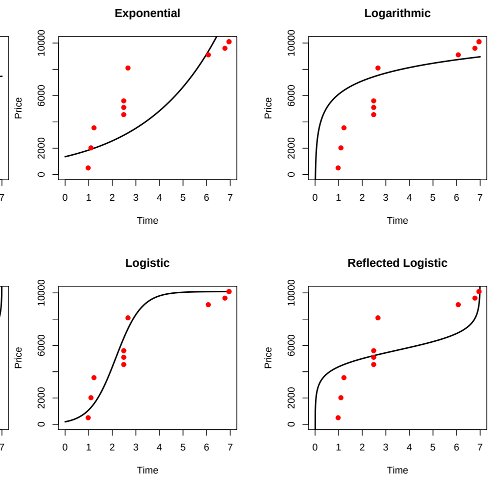


Figure 1.9. Exponential (top left), logarithmic (top right), logistic (bottom left), and reflected-logistic (bottom right) models fit to bids for the same auction. Overweighting in the x-dimension selects reflected-logistic growth when logistic growth should be chosen.

and Wood, 2005). A researcher may overweight price (large w_y) when the focus is on the price level itself (e.g., using this information to make more informed bid decisions.) Note that overweighting price tends to favor exponential and logistic growth models, whereas overweighting time leads to favoring logarithmic and reflected-logistic models. Examples of overweighting are shown in Figures 1.8. and 1.9.. In Figure 1.8., overweighting in the price dimension leads to exponential growth selection, whereas reflected-logistic growth would have been selected had the weights been equal. In Figure 1.9., overweighting in the time dimension leads to reflected-logistic growth selection, whereas logistic growth would have been selected had the weights been equal. In addition to the task at hand, visual inspection of a subset of the auctions can provide an appropriate weighting scheme.

From empirical evidence, we find that both WSSE measures provide very similar results, matching on 453 out of 472 (or 95.97%) of the auctions. In cases where the results do not match, visual inspection shows that the models selected by each metric provide reasonably good results with a slight preference towards WSSER. An example is provided in Figure 1.10.. WSSEV selects logistic growth whereas WSSER selects reflected-logistic growth. Both appear to fit the data reasonably well

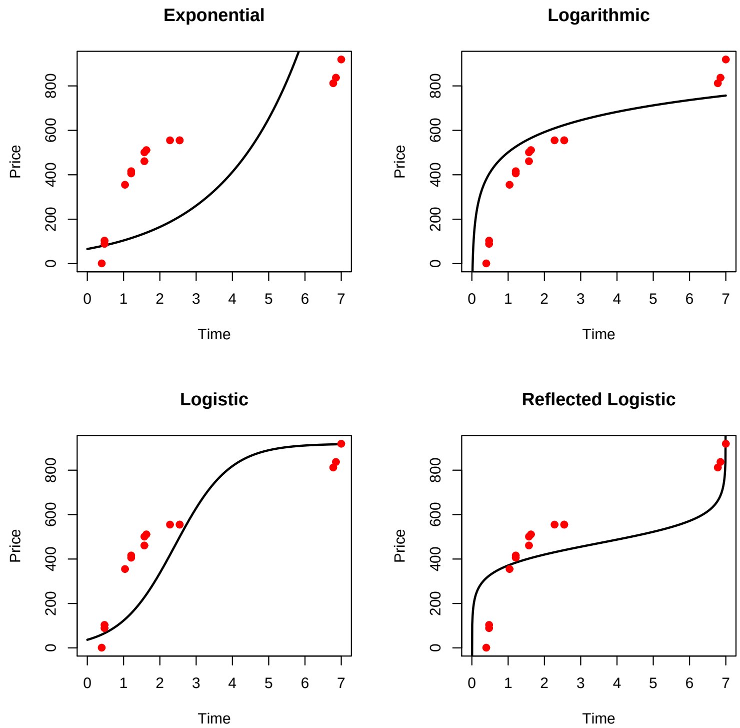


Figure 1.10. Exponential (top left), logarithmic (top right), logistic (bottom left), and reflected-logistic (bottom right) models fit to bids for the same auction. WSSER selects reflected-logistic growth while WSSEV selects logistic growth.

with reflected-logistic growth capturing the process slightly better. We therefore use WSSER in the following.

Figure 1.11. shows live bids and fitted price curves for four different auctions. The top left is best fit with exponential growth, the top right with logarithmic growth, the bottom left with logistic growth, and the bottom right with reflected-logistic growth. For model comparison purposes, the selected model is drawn in black and the three models that are not selected (but fitted) are drawn in grey.

Figure 1.12. provides the distribution of auctions across the four models. As expected, and in accordance with previous empirical evidence, exponential growth best fits the majority of the auctions. The next most frequent model is reflected-logistic growth, which captures the common phenomena of early and late bidding. Logistic growth curves are also selected in many cases. In contrast, logarithmic growth is rarely chosen (2.54% of the auctions). This is most likely because in this set of auctions, early high bidders were rare.

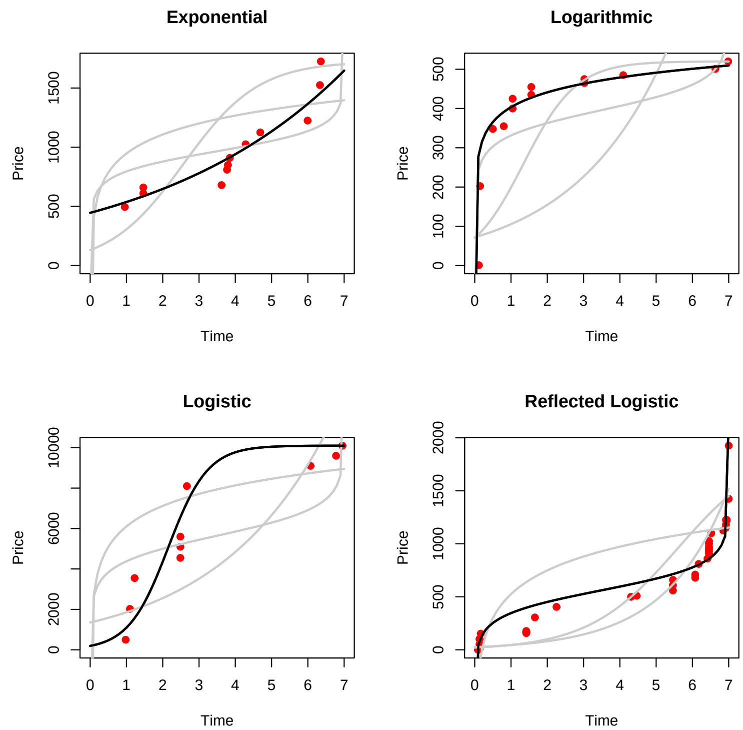


Figure 1.11. Live bids and fitted price curves for four different auctions. The top left is best fit with exponential growth, the top right with logarithmic growth, the bottom left with logistic growth, and the bottom right with reflected-logistic growth.

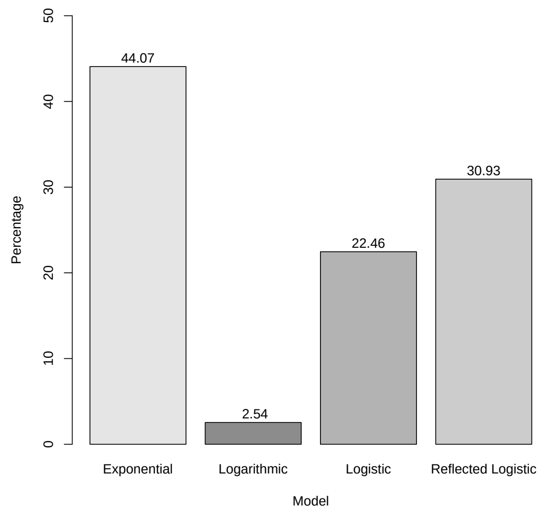


Figure 1.12. Distribution of selected model for 472 luxury wristwatch auctions.

1.6 SMOOTHING METHOD COMPARISON

To assess the differences between our proposed growth models and the nonparametric smoothing methods (p-splines and monotone smoothing splines), we compare them on several dimensions:

Nature of fitted curves - How well does the fitted curve capture the main features of the underlying price process? Specifically, are the estimated curves monotone?

Data fitted - Which data are used for fitting the curve? Can the actual bid data be used, or do we have to sample from the “actual price" step function? In addition, what type of auctions, in terms of the number of bids, can be fit?

Overall fit - How well does the curve fit the actual bid data?

Parsimony - The level of complication in terms of number of parameters

Explanation - The level of informativeness of the fitting mechanisms (model driven vs. data driven)

Automation and computational considerations - The level of user input that is needed for curve fitting and the ability to automate the process. In addition, considerations of computational complexity and run time (especially when considering large auction datasets).

Table 1.3. summarizes the comparison of the parametric and nonparametric curve fitting on all these dimensions.

Table 1.3. Comparison of parametric growth models and nonparametric curve fitting.

	P-splines	Monotone	Growth
Nature of curves	nonmonotone	monotone	monotone
Data fitted	step function	step function	bid data
Overall fit	good	variable	good
Parsimony	many parameters	many parameters	model type + 2 parameters
Explanation	unavailable	unavailable	available
Automation	user specifies parameters	user specifies parameters	no user input
Computation	fast	slow	fast

In terms of fitted curves, growth models have the advantages of fitting monotone curves, fitting directly to the “live bids" (in some cases with the addition of the price at the start and end of the auction), and fitting any number of bids, including single-bid auctions (using the additional start and end prices). The resulting curves fit a variety of bid histories and capture the main features of the price dynamics during the auction. In comparison, nonparametric curves are not fit directly to the bid data but rather to the derived step function that conveys the price seen during the auction. Although monotone splines produce monotone curves, p-splines do not guarantee such monotonicity. In fact, there is a balance between monotonicity and data fit, such that a large smoothing parameter might create monotone curves but create larger deviations between the curve and the data points and vice versa. The wiggliness of the p-splines can be seen in several of the auctions in Figure 1.13.. With respect to the minimal number of bids needed for fitting, monotone splines can only be used to fit auctions with at least two bids. P-splines can be fit to single-bid auctions, but the result will be a wiggly horizonal line.

When comparing goodness-of-fit of the curve to the bid data, growth models appear to provide a good fit without overfitting. To compare goodness-of-fit, we fit each of the three methods (growth models, p-splines, and monotone splines) to each of the 472 auctions in the luxury wristwatch dataset. Using the WSSE metric, we find that p-splines provide the best fit roughly 70% of the time, growth models are selected 25% of the time, and monotone smoothing only 5% of the time (Table 1.4.). However, nearly 90% of the curves fit by p-splines are not monotone (which is verified by our sample in Figure 1.13.). When comparing only monotone smoothing and growth models, we find that growth models are selected 88% of the time.

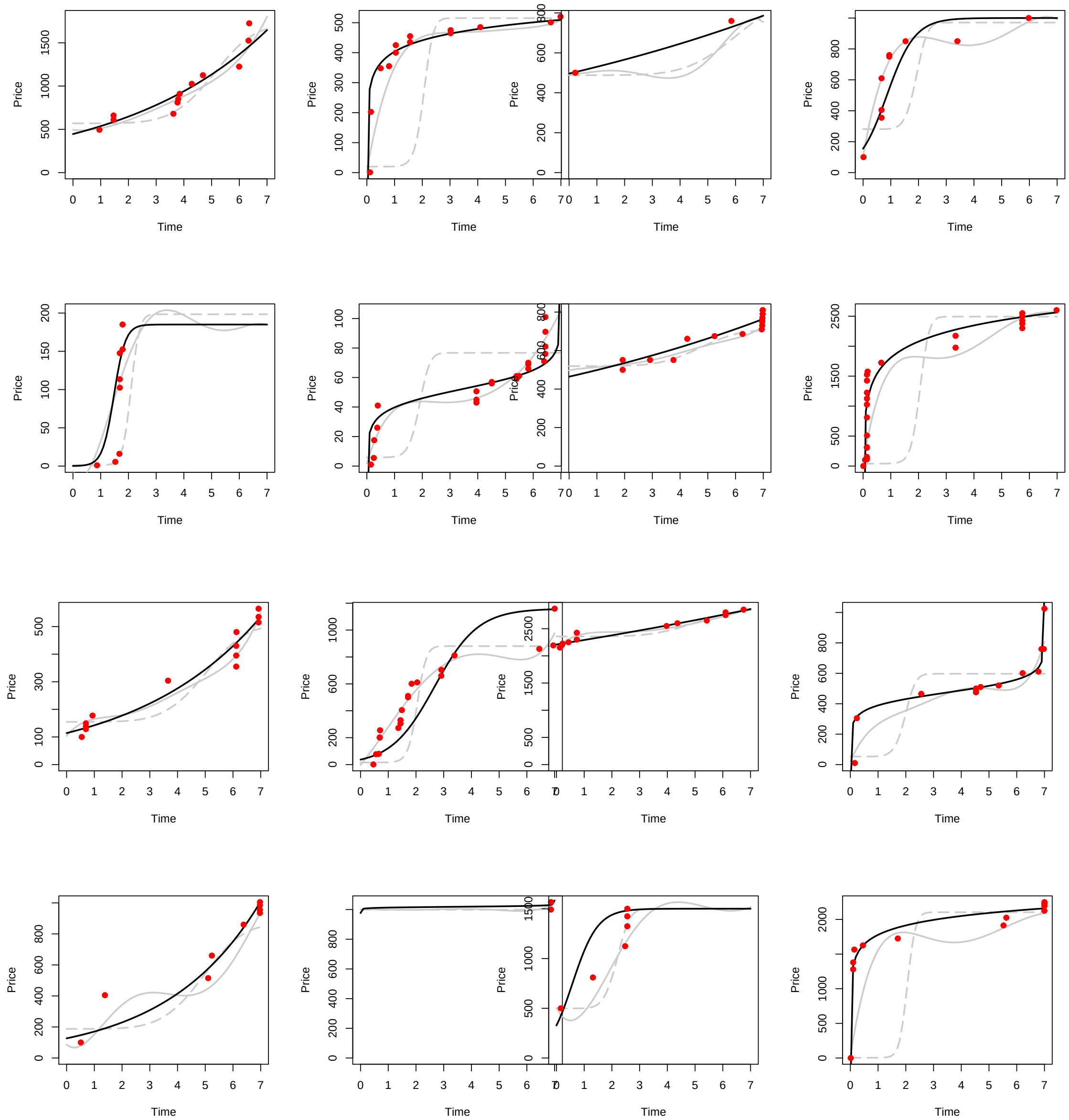


Figure 1.13. Live bids and smoothed price curves for randomly selected 7-day luxury wristwatch auctions. The black line is fit with the growth models, the solid grey line is fit with p-splines, and the grey dashed line is fit with monotone smoothing splines.

Table 1.4. Distribution of chosen price model for 472 completed 7-day luxury wrist watch auctions.

	P-splines	Monotone	Growth
Percent chosen when comparing 3 methods	69.49%	5.08%	25.42%
Percent chosen when comparing 2 methods	-	11.65%	88.35%

With respect to parsimony and explanatory power, the parametric growth models have a clear advantage: they include only two parameters, and the family of four models is able to capture a wide variety of price processes. Furthermore, growth models provide a theoretical basis that describes the price “growth" during the auction and its dynamics. Exponential models are associated with sniping, where the rate of the price increases grows faster and faster. The logistic and reflected-logistic models capture the change in price dynamics associated with early bidding. In contrast, the nonparametric methods are purely data-driven and as such do not provide a theoretical model for price growth. While they do capture the price process and its dynamics, they require a large number of parameters (the polynomial coefficients between each pair of consecutive knots, usually each such polynomial is of order 4, in order to obtain smooth curve derivatives).

Finally, from a computational point of view, fitting the growth models to data is very easy to automate and is reasonably fast, even for a large dataset of auctions. The fitting can be completely automated, and we find that the results of automated fitting are satisfactory. The combination of easy automation and computation time is a major advantage over nonparametric smoothing. When fitting curves nonparametrically, the user is required to specify several parameters in advance: the number and position of knots, the order of the polynomials, and the roughness parameter. The set of knots and roughness parameter that optimally uncover the price process in one auction may not accurately capture the underlying price process of another auction. However, the same number and position of knots and roughness parameter is often used for all auctions in the dataset of interest in order to avoid confounding the curve fitting from other manipulations (see Jank and Shmueli (2007)).

To evaluate computation time, we measure the elapsed time for each of the three smoothing methods on the 472 luxury wristwatch auctions as well as the first 10 auctions (Table 1.5.). It is clear that for even moderate datasets monotone splines require very long run times, whereas p-splines and growth models are much faster.

Table 1.5. Elapsed time (in seconds) to fit 472 and a subset of 10 luxury wristwatch auctions by p-splines, monotone splines, and growth models.

	P-spline	Monotone	Growth
10 Auctions	2	75	4
472 Auctions	6	2082	33

1.7 USING GROWTH CURVES

One of the main advantages of modeling an auction's price process through parametric growth models is that an initial distinction between auctions is directly obtained: each auction is represented (or, labeled) by one of exponential, logarithmic, logistic, or reflected-logistic growth. Knowing the shape of the price curve tells us about the underlying price process. This knowledge is useful in many applications, some of which will be discussed in this section.

1.7.1 Rug Plots

The *rug plot* is a visualization tool, proposed by Hyde et al. (2006), for displaying concurrent processes over a period of time. In the online auction context, the rug plot can display the entire price-evolution of all auctions in the dataset over the period of data collection (calendar time). Specifically, the x-axis is calendar time, the y-axis is price, and each auction's price process is plotted as a curve. Rug plots for datasets of Palm Pilot M515 auctions and Xbox auctions (see Appendix A for descriptions of these datasets) are shown in the middle panels of Figure 1.14.. The final price of each auction is marked with a dot, and the thick black line and grey band are the daily median closing prices and interquartile ranges (IQR), respectively.

The rug plot supports visual exploration of temporal groupings of curves. When curves are fit nonparametrically, it is hard to ascertain types of curves without visual inspection of each curve, which could be a daunting task for even a moderate dataset. Growth models offer an easy solution, by using the WSSE measure to choose the best growth model of the four. The model type can then be easily integrated into the rug plot via color coding. To further improve the information contained in a rug plot, we add color-coded dot plots, where a dot represents an auction that closes on that date (top panels in Figure 1.14.). The dots are jittered to visualize periods where many auctions close on the same day. In addition, we create time-grouped stacked bar charts for the volume of auction closings during the time period, as can be seen in the bottom panel of Figure 1.14. (week for Palm data, day for Xbox

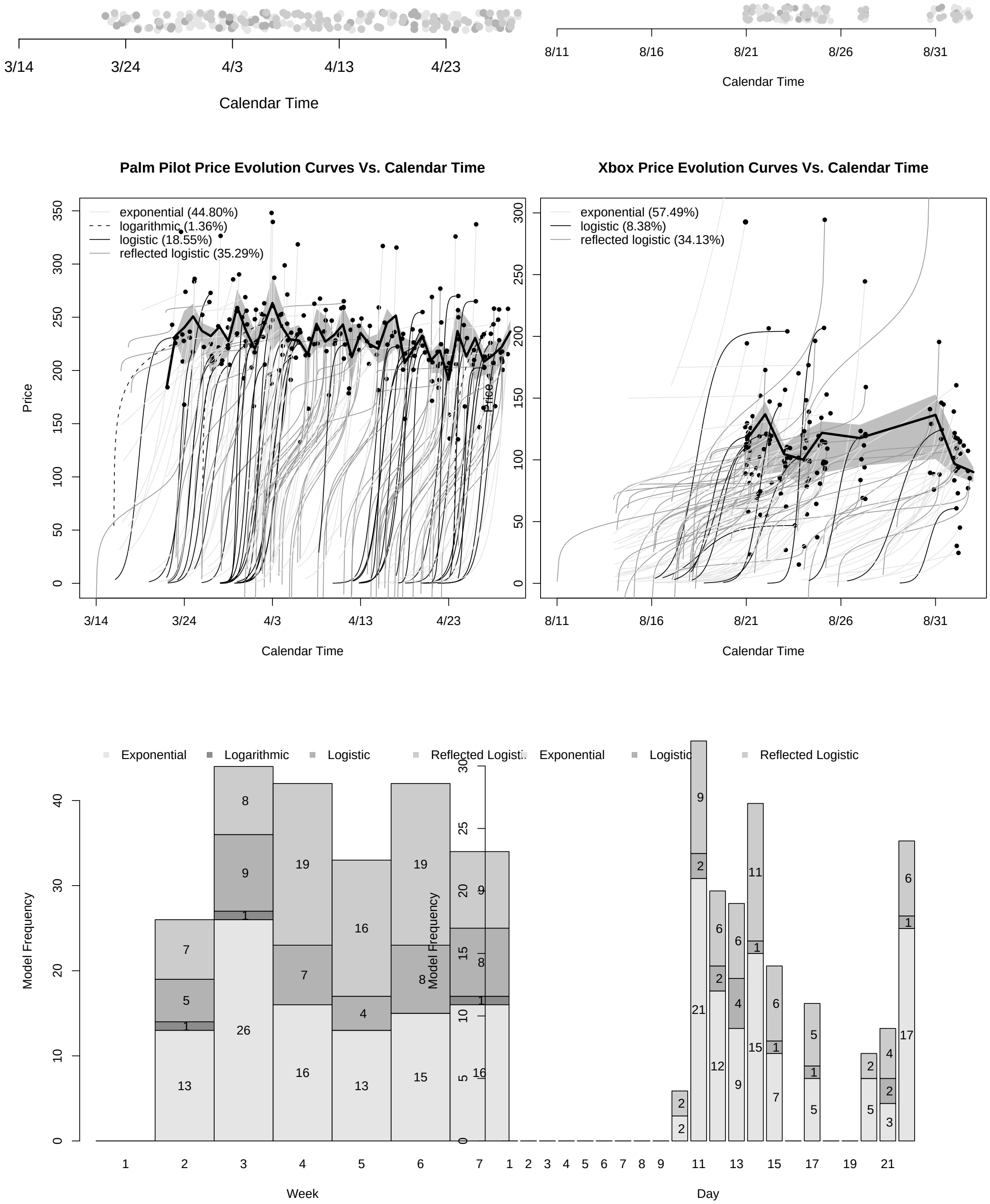


Figure 1.14. Visualizing temporal clustering of price curve types. The left column describes the Palm data; the right column describes the Xbox data. All plots are color-coded by growth model type. The top panels are dot plots (points are jittered for visibility). The middle panels are rug plots. The bottom panels are temporally-aggregated stacked bar charts of auction volume.

data). Using this display we investigate temporal groupings of price processes. For the Palm data, we see a grouping of reflected-logistic price curves over 4/6-4/13 among very few exponential growth curves, whereas during other times (and especially before 4/6) most price curves are exponential. Perhaps the large number of exponential growth curves before 4/6 led bidders in later auctions to believe that most auctions close well over $150, and they therefore bid early in the auction. In the Xbox data, most curves are either exponential or reflected-logistic with no logarithmic price curves. There is also a period with almost no auctions (due to data collection issues). Here we see during the beginning of the period, between 8/14 to 8/18, a clustering of exponential and reflected-logistic curves, with many of the reflected-logistic auctions opening higher than the exponential auctions.

We further explore the relationship of other information, such as auction duration, to the temporal clustering of different price curves. Figure 1.15. is a set of rug plots (for the Palm data), separated by auction duration (10,7,5, and 3 days). We see a temporal clustering of reflected-logistic auctions between days 4/6 and 4/13 in the 7-day auctions. Another observation is sporadic exponential price curves in the 5- and sometimes 7-day auctions.

Another such exploration is the comparison of new vs. used Xbox game consoles (Figure 1.16.). We see temporal groupings of logistic curves and reflected-logistic curves for new Xbox's at the beginning of the calendar whereas for used Xbox's, there does not appear to be such distinct groupings. Because there is still a large volume of auctions in the used dataset, perhaps zooming-in on different calendar dates would reveal more temporal groupings.

1.7.2 Integrating Growth Model Parameters Into Analyses

The parametric growth model representation provides a compact representation of the entire price curve in an auction which is simple and parsimonious – the model type and its two estimated parameters alone. We can then integrate this compact representation into analyses by applying the statistical or data-mining method directly to this representation. This is a classic data mining approach where complicated information is summarized and the summaries are used in the analysis.

One possible application is clustering auctions using the growth model representation and perhaps additional auction related information. Another direction is in distance based methods, where the parametric representation can be used for measuring the distance between auctions. A third example is using classification trees, where the model type and the estimated parameters serve either as predictors (for predicting an outcome of interest), or as the outcome variable. In particular, such a tree could be used for predicting the type of price curve of a new auction as a function of information that is given at the auction start (e.g., opening bid, seller rating, presence of a picture, and closing day). Potential

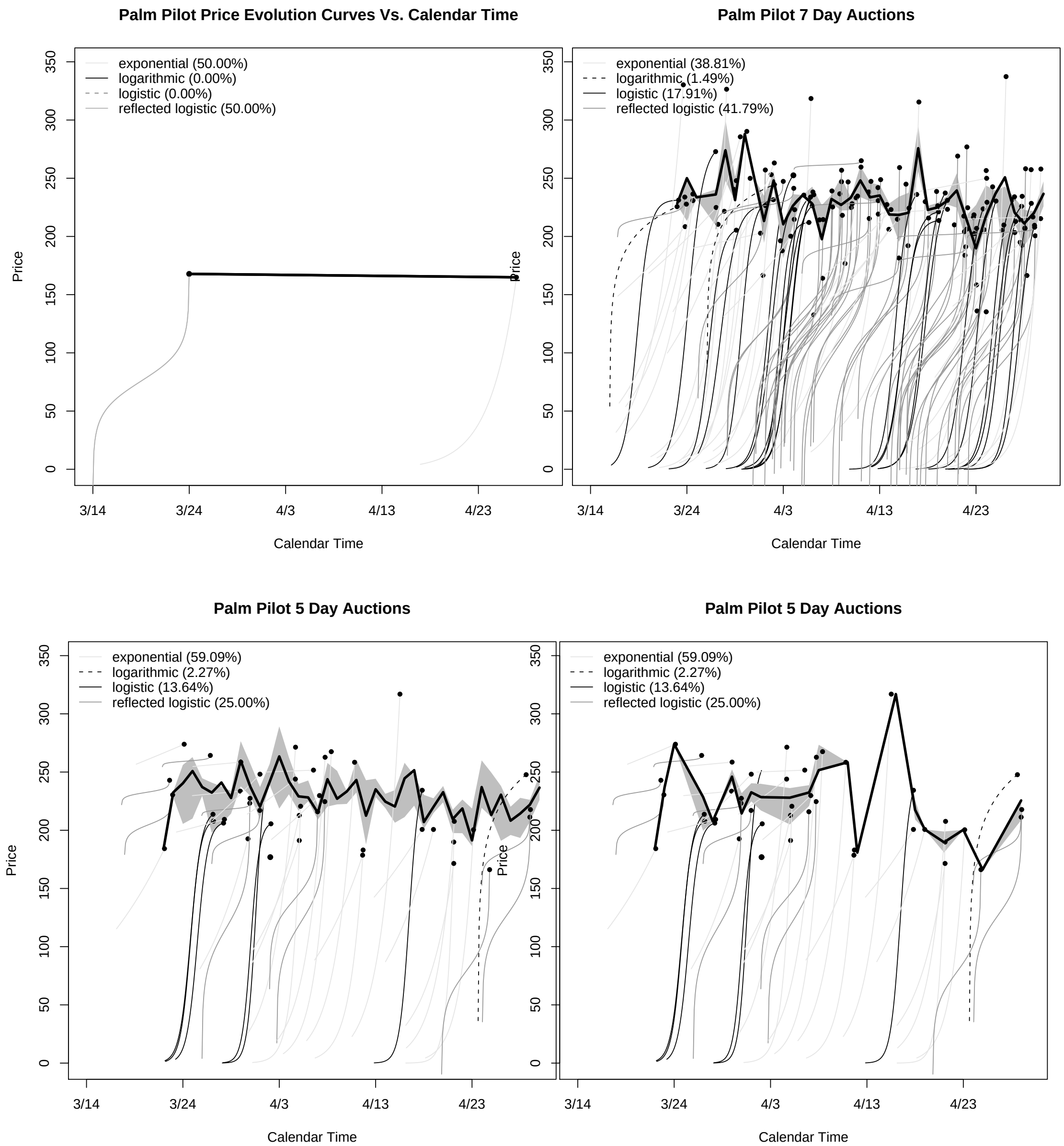


Figure 1.15. Rug plots for Palm Pilot M515 auctions broken down by length: 10-day (top left), 7-day (top right), 5-day (bottom left), and 3-day (bottom right).

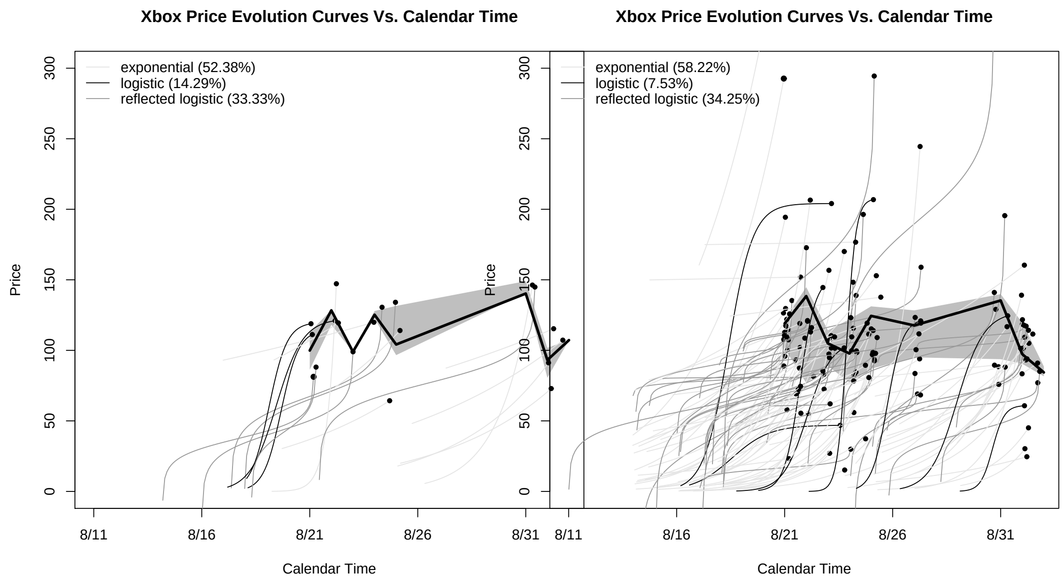


Figure 1.16. Rug plot for Xbox auctions broken down by type:“New" (left) and “Used" (right).

bidders can then use the predicted information in order to decide on which auction to bid, their bid timing, and bid amount.

We describe only a few possible applications here, but obviously the approach is general and parametric representation can be used in almost any type of statistical analysis and/or data mining technique.

1.8 CONCLUSIONS

In this paper, we introduce a family of growth models that describe the underlying continuous price process of online auctions. These are exponential growth, logarithmic growth, logistic growth, and reflected-logistic growth. We also present a metric to choose between models (and more generally, to choose between any type of fitted curves), which allows automation in the data fitting stage.

Our parametric approach is parsimonious, the models are easily fitted to bid data, and they capture a variety of price process shapes. They also provide an appealing theoretical explanation of the price process rather than being purely data-driven. The resulting curve is monotone, as expected for price curves in ascending auctions. Our method is computationally fast and can therefore be applied to large auction datasets. All of these reasons give the parametric approach an advantage over nonparametric smoothing methods.

We fit exponential and logistic models in the price dimension but fit logarithmic and reflected-logistic models in the time dimension. For simplicity and computational efficiency, OLS fitting is performed in the dimension that is linearizable. However, both metrics (WSSER and WSSEV) for selecting the best growth model evaluate fit in both the price and time dimensions simultaneously to avoid overselection of models fit in a certain dimension. We reason that for online auctions, both the time and magnitude of the bid are random variables, so the fit in both dimensions is important. This suggests that our fitting method should also be based on both dimensions simultaneously. Since we can not linearize the growth models in both dimensions, we could employ other optimization techniques, such as steepest ascent or Newton-Raphson, to estimate parameters. One of the reasons we were initially hesitant to employ optimization methods was the simplicity and computational speed of our method. Further, no parameters needed to be set in advance, as is required of the nonparametric smoothing methods. If we employ iterative fitting in both dimensions simultaneously, we may use the estimates obtained via one dimensional OLS as the starting values. This line of research should be expanded and a comparison of parameter estimates as well as computational complexity should be considered.

Our contribution is not limited to the auction setting, but rather proposes the use of parametric functional representations as an alternative to the more popular nonparametric functional objects. We show how the parametric representation provides advantages in data visualization, as well as offers a compact summarization of the price process that can then be used in a variety of statistical analysis and data mining techniques.

One of the limitations of our family of growth models is in the case of auctions with very sparse activity throughout the auction that then changes into very steep price increases at the last moments of the auction. In this case, the exponential growth model, which provides the best fit among the four models, often fails to adequately capture the intense bid activity at the auction end. One solution is to first transform the data, e.g. by moving to log-scale. Another possibility is to heavily weight the data points towards the auction end in the fitting process. And yet another option is to include an additional growth model that describes processes that change little until a peak at the end.

There are many other functions that could potentially be used to model growth, and we provide a few examples. The Chapman-Richards growth function is similar to logarithmic growth but places a limit on growth. The Couttsian growth model is similar to exponential growth except that the growth rate is variable. Different models are popular in different disciplines such as biology, ecology, economics, etc. to desribe a variety of phenomena.

We believe that parametric functional representations enhance the field of FDA and provide additional information for statistical analysis. We hope to spur interest in using theoretically relevant parametric models to decribe continuous processes.

Acknowledgements

The authors would like to thank Shanshan Wang for the Xbox data and Sharad Borle for the luxury wristwatch data.

REFERENCES

Alford, B. and Urimi, L. (2004). An analysis of various spline smoothing techniques for online auctions. Term Paper, AMSC Research Interaction Team. Available at http://www.smith.umd.edu/ceme/statistics.

Bajari, P. and Hortascu, A. (2003). Winner's curse, reserve price and endogenous entry: Empirical insights from ebay. *RAND Journal of Economics*, 34:329–355.

Bapna, R., Goes, P., Gupta, A., and Jin, Y. (2004a). User heterogeneity and its impact on electronic auction market design: An empirical exploration. *MIS Quarterly*, 28(1):21–43.

Bapna, R., Jank, W., and Shmueli, G. (2004b). Price formation and its dynamics in online auctions. Working Paper, Smith School of Business, University of Maryland. Available at http://ssrn.com/abstract=902887.

Green, P. J. and Silverman, B. W. (1994). *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach*. Chapman and Hall, London.

Hyde, V., Jank, W., and Shmueli, G. (2006). Investigating concurrency in online auctions through visualization. *The American Statistician*, 34(3):241–250.

Jank, W. and Shmueli, G. (2007). *Business Computing*, chapter Studying Heterogeneity of Price Evolution in eBay Auctions via Functional Clustering. Handbook of Information Systems Series. Elsevier. Adomavicius and Gupta (Eds.). Accepted.

Jank, W., Shmueli, G., Plaisant, C., and Shneiderman, B. (2007). *Handbook on Computational Statistics on Data Visualization*, chapter Visualizing Functional Data with an Application to eBay's Online Auctions. Springer-Verlag, Heidelberg. Chen, Hardle and Unwin (Eds.). In press.

Kauffman, R. J. and Wood, C. A. (2005). The effects of shilling on final bid prices in online auctions. *Electronic Commerce Research and Applications*, 4:18–31.

Ramsay, J. O. (1998). Estimating smooth monotone functions. *Journal of the Royal Statistical Society B*, 60:365–375.

Ramsay, J. O. and Dalzell, C. J. (1991). Some tools for functional data analysis. *Journal of the Royal Statistical Society B*, 53:539–572.

Ramsay, J. O. and Silverman, B. W. (2002). *Applied Functional Data Analysis: Methods and Case Studies*. Springer-Verlag, New York.

Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis*. Springer-Verlag, New York.

Roth, A. E. and Ockenfels, A. (2002). Last-minute bidding and the rules for ending second-price auctions: Evidence from ebay and amazon auctions on the internet. *The American Economic Review*, 92.

Shmueli, G., Russo, R., and Jank, W. (2004). The barista: A model for bid arrivals in online auctions. Working Paper, Smith School of Business, University of Maryland. Available at http://ssrn.com/abstract=902868.

Wang, S., Jank, W., and Shmueli, G. (2007). Dynamic forecasting of online auction price using functional data analysis. *Journal of Business and Economic Statistics*. In press.

APPENDIX A

ADDITIONAL EBAY DATA

Two additional eBay.com data sets are used to illustrate some applications of growth models. They are for Palm Pilot M515 and Xbox auctions.

A.0.1 Palm Pilot M515 Data

Our data contain information on 221 closed auctions for a brand new Palm Pilot M515. The data were collected between March 11, 2003 and April 20, 2003, roughly a year after the Palm M515 was released to the market. At that time, eBay auctions lasted 3, 5, 7, or 10 days, depending on the length set by the seller. More recently, 1 day auctions were introduced on eBay.

Even though the Palm Pilot has a known market value ($250 at the time of the analysis), auctions do not always close near this value. Low prices can result, for instance, if the box is already open or if the seller has a questionable reputation. Auctions can close high if something special is offered with the product such as an accessory or free shipping. Prices also vary because bidders often get caught up in the excitement of bidding (“auction fever") and pay more than would be expected. The average selling price for all Palm Pilots in our data is $234.00 with a median of $232.50 and standard deviation of $20.86. The least expensive Palm Pilot sold for $172.50 and the most expensive auction closed at $290. Table A.1. provides descriptive statistics for the closing price, opening price, number of bids, number of unique bidders, and unique bidder rating. We also provide descriptive statistics broken down by auction length since we group Palm Pilot auctions by length in Section 1.7. Note that this data set does not contain any seller information.

A.0.2 Xbox Data

Our data contain information on 167 closed auctions for an Xbox game console. The auctions took place between August 11, 2005 and August 30, 2005, and are of fixed duration: 1, 3, 5, 7, or 10 days. While the Xbox product is the same, it may be “new" or “used", and the auction may include extras such as additional games and/or controllers. Therefore, bidders will not have the same valuation for each auction. Descriptive statistics for all the auctions as well as broken down by item condition (new or used) are shown in Table A.2..

This game console is no longer sold in stores (as it is the predecessor of the Xbox 360); however, Amazon.com’s list price was $179.98 at the time the data were collected. The average selling price in our sample is $132.40 with a standard deviation of $62.59, median of $125.00, minimum of $28.00, and maximum of $501.80. The auction that closed at $28.00 is for a damaged console, and the auction that closed at $501.80 is used but includes 84 games.

Table A.1. Descriptive statistics for 221 completed eBay Palm Pilot M515 auctions and broken down by auction length: 3-day (41), 5-day (44), 7-day (134), and 10-day (2) auctions.

Variable	Duration	Mean(Std)	Median	Minimum	Maximum
Closing Price	3 Day	$238.60($24.67)	$232.50	$177.50	$290.00
	5 Day	$233.40($23.08)	$235.00	$183.50	$280.00
	7 Day	$233.60($17.78)	$232.80	$186.50	$283.50
	10 Day	$182.50($14.14)	$182.50	$172.50	$192.50
	Total	$234.00($20.86)	$232.50	$172.50	$290.00
Opening Price	3 Day	$63.46($94.87)	$1.00	$0.01	$259.00
	5 Day	$91.09($97.36)	$35.50	$0.01	$259.00
	7 Day	$41.69($72.47)	$1.00	$0.01	$259.00
	10 Day	$2.51($3.53)	$2.51	$0.01	$5.00
	Total	$63.70($84.05)	$1.00	$0.01	$259.00
Number of Bids	3 Day	17.51(9.85)	19.00	2.00	43.00
	5 Day	16.18(9.18)	17.50	2.00	36.00
	7 Day	20.86(10.25)	21.00	2.00	51.00
	10 Day	20.50(6.36)	20.50	16.00	25.00
	Total	19.33(10.09)	19.00	2.00	51.00
Number of Unique Bidders	3 Day	9.22(5.05)	9.00	2.00	23.00
	5 Day	8.25(4.22)	9.00	2.00	19.00
	7 Day	10.85(4.71)	11.50	1.00	23.00
	10 Day	10.50(3.54)	10.50	8.00	13.00
	Total	10.03(4.77)	10.00	1.00	23.00
Unique Bidders Rating	3 Day	91.14(68.43)	86.00	3.00	204.00
	5 Day	89.93(69.01)	84.00	1.00	217.00
	7 Day	84.56(70.29)	74.00	1.00	217.00
	10 Day	73.81(57.02)	74.00	3.00	167.00
	Total	86.46(69.68)	74.00	1.00	217.00

Table A.2. Descriptive statistics for 167 completed eBay Xbox auctions and broken down by condition: “new" (21) and “used" (146).

Variable	Condition	Mean(Std)	Median	Minimum	Maximum
Closing Price	New	\$121.00(\$14.21)	\$123.20	\$85.00	\$142.50
	Used	\$134.10(\$66.60)	\$125.50	\$28.00	\$501.80
	Total	\$132.40(\$62.59)	\$125.00	\$28.00	\$501.80
Opening Price	New	\$26.19(\$37.33)	\$1.00	\$0.01	\$99.99
	Used	\$40.84(\$43.40)	\$32.49	\$0.01	\$290.00
	Total	\$39.00(\$42.86)	\$25.00	\$0.01	\$290.00
Number of Bids	New	20.95(8.81)	22.00	6.00	38.00
	Used	18.64(11.75)	18.00	2.00	75.00
	Total	18.93(11.43)	18.00	2.00	75.00
Number of Unique Bidders	New	9.19(3.16)	9.00	3.00	14.00
	Used	8.18(3.80)	8.00	1.00	19.00
	Total	8.31(3.73)	8.00	1.00	19.00
Unique Bidders Rating	New	234.70(340.79)	164.00	5.00	1325.00
	Used	299.70(837.90)	36.00	-1.00	5560.00
	Total	291.50(792.29)	44.00	-1.00	5560.00
Seller Rating	New	30.16(69.63)	5.00	0.00	605.00
	Used	42.51(166.44)	5.00	-1.00	2736.00
	Total	40.79(156.63)	5.00	-1.00	2736.00