Journal of the
Royal Statistical Society
SERIES C
Applied
Statistics

Appl. Statist. (2010)
59, *Part* 5, *pp.* 781–804

A flexible model for estimating price dynamics in on-line auctions

Wolfgang Jank, Galit Shmueli and Shu Zhang

University of Maryland, College Park, USA

[Received July 2009. Revised March 2010]

Summary. The path that the price takes during an on-line auction plays an important role in understanding and forecasting on-line auctions. Price dynamics, such as the price velocity or its acceleration, capture the speed at which auction information changes. The ability to estimate price dynamics accurately is especially important in realtime price forecasting, where bidders and sellers must make quick decisions or react to changes in market conditions. Existing models for estimating price paths from observed bid data suffer from issues of non-monotonicity, high variability or computational inefficiency. We propose a flexible two-parameter beta model which adequately captures a wide range of auction price paths. The model is computationally efficient and has several properties that make it especially advantageous in the on-line auction context. We compare the beta model with non-parametric and parametric alternatives empirically, when used in a variety of forecasting models. Using bidding data from eBay auctions, we find that the beta model leads to fast and high accuracy price predictions. This behaviour is consistent across various forecasting models and data sets. The implication for practice is the usefulness of the beta model for obtaining accurate and realtime bidding and selling decisions in on-line markets.

Keywords: Beta distribution; eBay; Forecasting; Functional data; Growth models; Kullback–Leibler distance; Price path; Price velocity

1. Introduction

On-line auctions have become an important marketplace, as they allow consumers and businesses to sell, buy and bid on a variety of goods. On-line auctions differ from their off-line counterparts in their longer duration (typically several days), the anonymity of participants, the low barriers of entry, their global reach and around-the-clock availability. These conditions lead to a highly dynamic environment, where bidders engage in competitive behaviour that is motivated by both psychological effects and economic reasoning. The longer duration of auctions allows bidders to adjust their behaviour on the basis of the progress of the auction thus far (as well as the progress of competing auctions), contributing to dynamic changes in price.

Empirical research on on-line auctions has been flourishing in recent years, in part because of the important role that these auctions play in the marketplace, and in part because of the availability of large amounts of high quality bid data from Web sites such as eBay, Yahoo!, OnSale and uBid. eBay (`www.eBay.com`) is one of the major on-line marketplaces and currently the biggest consumer-to-consumer on-line auction site. eBay makes publicly available a vast

Address for correspondence: Wolfgang Jank, Center for Complexity in Business, Department of Decisions, Operations and Information Technologies, Robert H. Smith School of Business, University of Maryland, Van Munching Hall, College Park, MD 20742-1815, USA.
E-mail: wjank@rhsmith.umd.edu

 0035–9254/10/59781

amount of rich bidding data that include all the temporal bidding information as well as information about the bidders, the seller and the product being auctioned.

Two of the main objectives in on-line auctions research are

(a) to understand better and to explain the dynamics of competing bidding strategies and complex bidder–seller interactions during an auction and
(b) to forecast the outcome of an on-going or future auction.

Although in this paper we illustrate our methodology in the context of forecasting, it is equally valuable for explaining auction behaviour.

1.1. Realtime forecasting

One stream of research has focused on realtime forecasting of on-line auctions. Early estimates of the auction's closing price have various advantages for auction participants. Bidders can use such information to make more informed (and perhaps even automated) bidding decisions. For instance, Jank and Zhang (2008) developed a system in which price forecasts can be used to select an auction automatically (among hundreds of competitors) and to decide when and how much to bid on that auction. They showed that such an automated bidding system can maximize the *surplus* that is extracted by bidders. Sellers can use price predictions to identify times when the market is more favourable to sell their products and to evaluate the value of their inventory better. Sellers and auction houses can also use price forecasts that arrive during an on-going auction to take corrective action, such as making the auction more salient, actively inviting new bidders or even alerting for fraud. Price forecasts can also serve as a basis for creating seller insurance (see, for example, Ghani and Simmons (2004)).

The literature contains various approaches to predicting the price of an on-going auction. Ghani and Simmons (2004) proposed models that use only the information that is available *at the start* of the auction to produce price forecasts. Such information may contain the length of the auction, its starting price, information about the product (e.g. used *versus* new products) or about the seller (e.g. the seller's trustworthiness). However, one drawback of such a static approach is that it does not take into account the dynamic information that arrives *during* the auction, such as the changing number of bidders or price updates. More recent research has shown that better predictions can be obtained by making use of all the information, until the time of prediction. Jap and Naik (2008) estimated dynamic bidding models in the context of on-line reverse auctions. One particularly challenging feature of on-line reverse auction data is partial observability, where the buyer observes a bid from only one of the bidders at any given point in time, but not bids from the rest of the participating bidders. To overcome this challenge of unobserved bids, Jap and Naik (2008) utilized a state space model in conjunction with Kalman filtering.

In our context (i.e. auctions such as those on eBay), at any given time during the auction all the bids that have been placed until that point are observable (and all bids except the highest bid are revealed). This lends itself to an alternative approach for capturing the dynamic price information via functional data analysis (Ramsay and Silverman, 2005). In a functional approach, one first reconstructs the underlying continuous price path and then integrates the path information into a forecasting model. Wang *et al.* (2008a) pioneered realtime forecasting models for on-going auctions based on functional data analysis. Their model uses as input estimates for the price path and its dynamics. By price path, we mean realtime information about the price up to the time of decision making; by price dynamics, we mean the velocity (and sometimes also the acceleration) of price. Price velocity (and acceleration) can be obtained via the first (and second) derivative of the price path. Wang *et al.* (2008a) showed that the inclusion of the dynamic information significantly improves the predictive accuracy compared with models that do not make use of such information. Whereas Wang *et al.* (2008a) illustrated the power of functional forecasting models in eBay auctions, Dass *et al.* (2009) used similar forecasters to predict the price in on-line auctions for contemporary Indian art. Whereas both Wang *et al.* (2008a) and Dass *et al.* (2009) used regression-like models, Zhang *et al.* (2009) have proposed a new approach via a functional K-nearest-neighbour algorithm to integrate dynamic information into the forecaster. And, in contrast with the previous studies which all focused exclusively on the information from within the focal auction, Jank and Zhang (2008) also incorporated information from simultaneously competing auctions in their model. They showed that the dynamics of auctions that sell the same or similar items during the same period of time further improve predictive power.

Common across all these dynamic models is that they use information that is derived from continuous price curves as *input* into the realtime forecaster. In all these cases, the price path and the associated price dynamics play an important role and have been shown to improve the resulting forecasting accuracy. The reason behind the informativeness of the price path lies in the dynamic on-line auction environment, where bidders make their decisions on the basis of more than simply economic considerations. Terms such as 'auction fever' and 'auction heat' capture the psychological atmosphere, the competitiveness and the often illogical bidding behaviour (Ariely and Simonson, 2003; Heyman *et al.*, 2004). The price path and dynamics capture some of this action.

1.2. Price dynamics and auction heterogeneity

Besides their value for price forecasting, price dynamics can also play an important role for understanding and explaining competing bidding strategies and complex bidder–seller interactions during an auction. Empirical studies have found that price dynamics can be very heterogeneous, even for auctions for the same product. For instance, Jank and Shmueli (2008a) have shown this for auctions of new Palm personal digital assistant handheld devices sold on eBay; Dass and Reddy (2006) found similar behaviour in on-line auctions for contemporary Indian art. Jank and Shmueli (2008) further segmented auctions on the basis of price dynamics and found three types of auctions: 'steady auctions' are those with constant dynamics, 'low energy auctions' are those with late dynamics and 'bazaar auctions' see mostly early activity.

In terms of the bid placement behaviour, Shmueli *et al.* (2007) developed a three-stage non-homogeneous Poisson process for capturing bid timing and showed its flexibility in capturing the bid timing across various items or auction durations. Finally, Wang *et al.* (2008b) proposed a class of functional differential equation models to capture the wide range of auction price paths and dynamics.

1.3. Estimating price dynamics

The contribution of this paper is the introduction of the beta model, which is a parsimonious, flexible and computationally efficient parametric model for capturing the price path and its dynamics. Because price dynamics play a key role in realtime forecasting, we focus on the usefulness of the beta model in improving the predictive accuracy of various forecasting models. Price dynamics models are typically used to produce estimates of the price dynamics, which in turn are used as input into forecasting models. Although the literature includes various forecasting models (e.g. Ghani and Simmons (2004), Jap and Naik (2008), Wang *et al.* (2008a), Dass *et al.* (2009), Zhang *et al.* (2009) and Jank and Zhang (2008)), the purpose of this paper is to show that the way in which price dynamics are estimated significantly affects the forecast accuracy in general. We compare the beta model with several existing alternatives and show that the beta model can lead to significant improvements in terms of producing fast accurate forecasts for a variety of forecasting methods.

The remainder of the paper is organized as follows: Section 2 describes the eBay data that are used in this study, and the on-line auction mechanism that generates these data. In Section 3 we present existing models for capturing price paths and dynamics in on-line auctions. Section 4 introduces our beta model and describes its properties, estimation and advantages. We then compare the beta model with several competitors empirically in Section 5, in terms of run time and forecast accuracy (when used as input into a forecasting model). In Section 6 we discuss further applications of the beta model within on-line auction research.

2. On-line auctions: data and generating mechanism

2.1. Auction mechanism

In on-line auctions, participants bid for products or services over the Internet. The opening price is set by the seller or the auction house, and bidders submit bids on line. There are various auction formats: on-line auctions can be ascending (e.g. in eBay auctions) or descending (e.g. in Dutch flower auctions where the price is bid down); first price or second price (i.e. whether the final price is equal to the highest or second-highest bid); with a hard or soft closing time (i.e. whether the auction duration extends with the arrival of new bids); for single items or bundles. On eBay, most auctions are second-price ascending auctions for single items, with a hard closing time. The seller sets the opening price, and bidders place ascending bids until the auction end time is reached. At that time, the winner is the highest bidder, and she or he pays the second highest bid (plus an increment).

2.2. Auction data

The data that are used in this study are based on the complete bidding records of all digital camera auctions that were listed on eBay between April 2007 and January 2008. The data were obtained directly from eBay and the raw data contain over 4 million bidding records.

We selected, from the raw data, the bidding records for all *Canon Point & Shoot SD1000 camera* auctions for which a single camera was listed for sale (some auctions had multiple items for sale), the item was sold through an auction process (and not via a fixed price listing) and with particular specifications that are the most common for this type of camera (optical zoom 3; resolution 7.1; digital zoom 4 times; memory card format MultiMedia card). This selection was performed so that items sold are as comparable as possible. Moreover, we excluded unusual auctions (those with more than one rechargeable battery, a lens cleaning kit, a tripod, an extra memory card or a camera bag). The resulting sample yielded 698 auctions, all selling the same product; see also Tables 1 and 2.

Fig. 1 illustrates the information overload that bidders face. It shows for each individual auction the *live price curve*, which reflects the price that bidders see at any given time during the on-going auction. We can see that the information can be quite overwhelming: the large number of concurrent auctions, the variation in prices and the fact that some auctions are only in their early stages, whereas others are about to end, all challenge the processing of the given information. Moreover, we see that prices increase unevenly throughout most auctions. They increase quickly in some auctions, but much slower in others. We refer to the movement of price throughout the auction as *price dynamics*, which will play an important factor in our subsequent modelling efforts.

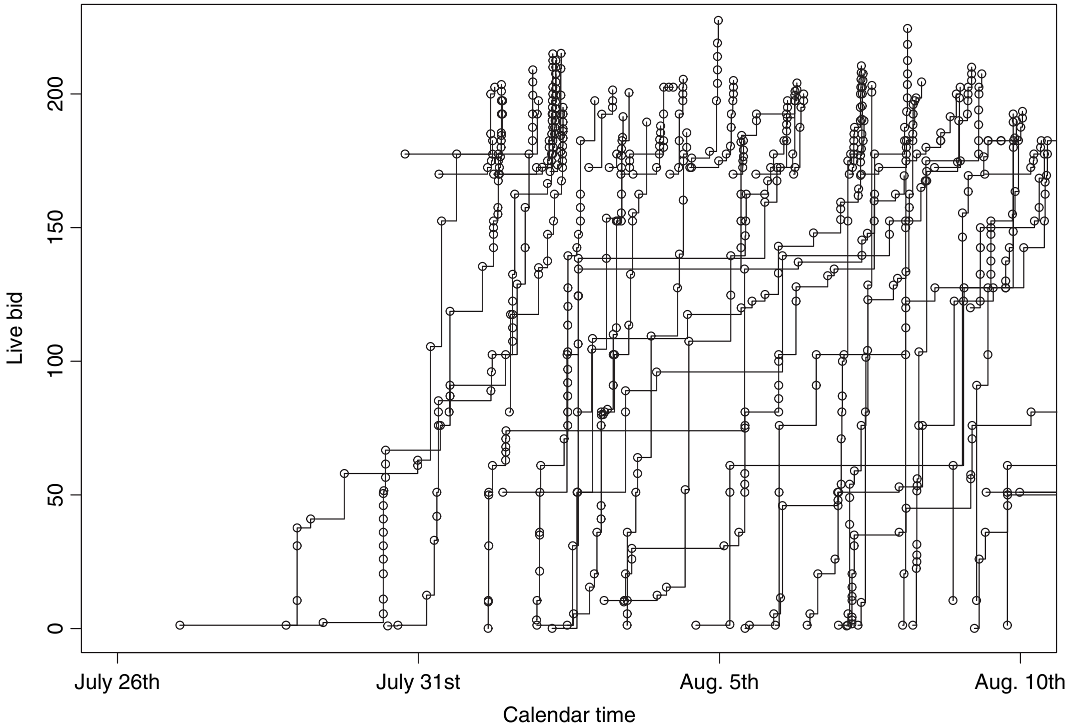


Fig. 1. Snapshot of the live price curves during eBay digital camera auctions: ○, bids; ——, |, price seen during the on-going auction

Table 1. Descriptive statistics for continuous variables for the digital camera auctions

Variable	*Minimum*	*Maximum*	*Mean*	*Median*	*Standard deviation*
Duration	1	10	3.21	3	2.44
Number of bids	1	52	15.94	15	9.57
Number of bidders	1	21	8.09	8	3.98
Closing price	30.99	709	179.49	180.39	36.86
Starting price	0.01	219.99	52.64	1	71.21
Seller feedback	0	12190	760.92	154.5	1547.31

3. Models for auction price paths

Given a smooth curve or function, the dynamics (e.g. velocity or acceleration) are typically computed as the first or second derivative of that function. However, when considering the price path in an on-line auction, observed bids create a non-decreasing step function with jumps at the times of bids (see, for example, Fig. 1). Since we cannot take derivatives at these discrete steps, a smooth representation of the price path is needed. There have been two general approaches to obtaining smooth auction price paths from observed bid data: non-parametric and parametric. Using a functional data analysis approach, Jank and Shmueli (2008) employed penalized smoothing splines (P-splines) to generate smooth curves. An alternative to P-splines

Table 2. Descriptive statistics for discrete variables for the digital camera auctions

Variable	*Yes (%)*	*No (%)*
Condition	93.84	6.16
Picture	62.89	37.11
Reserve price	3.58	96.42
Buy it now	29.23	70.77
Bold	86.1	13.9
Highlight	4.01	95.99

are monotone splines (Ramsay, 1998) which guarantee the monotonicity of the resulting curves. Finally, Hyde *et al.* (2008) proposed a parametric family of four distributions that capture the most typical price path shapes. Each of these three approaches yield smooth price curves, and then price dynamics are computed by taking derivatives of the smooth curves. The first derivative captures price velocity (i.e. how fast the price is moving at any point in time); the second derivative captures price acceleration, and so forth.

Conceptually, going from observed bids to a smooth price path leads to an abstraction of the auction process. By doing so, we assume that the auction process is not entirely deterministic: as the same product may be offered across several hundred or thousand simultaneous auctions, there is an element of chance that causes a bidder to 'detect' a particular auction. Once detected, chance may also play a role in the bidder's decision whether or not to enter and to place a bid in that auction. For instance, if 10 bidders detect the same auction at the same time, the 10th bidder may decide to abstain and not to place a bid. In addition, although the number of bidders entering an auction is not entirely deterministic, so is the timing and the amount of each individual's bid as bidders react to each other's moves; reaction occurs at random moments as most bidders do not have the time or energy to monitor auction Web sites continuously. All in all, the process of bidding involves patterns but it also involves noise. In that sense we can think of the smooth price path as an approximation to that pattern, eliminating noise in the process.

In what follows we describe each of these three approaches and discuss their strengths and weaknesses.

3.1. Penalized smoothing splines

Penalized smoothing splines (P-splines) (Simonoff, 1996) fit a polynomial of order p. To control the smoothness of the fitted curve, a penalty is imposed on the estimating function. Let $\tau_1, \tau_2, \ldots, \tau_L$ be a set of knots; then a polynomial spline of order p is given by

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

where $u+ = u\, I(u \geqslant 0)$ is the positive part of the function u. Many functions of this type tend to fit the data too closely (and thus model noise in addition to the signal); therefore, a *roughness penalty* approach is often employed which takes into account the trade-off between data fit (i.e. minimizing $f(t) = \Sigma_j \{y_j - f(t_j)\}^2$) and function smoothness. A popular measure of roughness, which measures the degree of departure from a straight line, is of the form

$$\mathrm{PEN}_m(t) = \int \{\mathrm{D}^m f(t)\}^2 \,\mathrm{d}t, \tag{2}$$

where $D^m f$, $m=1,2,3,\ldots,$ denotes the mth derivative of the function f. A highly variable function will yield a high value of $\mathrm{PEN}_m(x)$. If the highest derivative of interest is m, then using $m+2$ as the polynomial order will assure m continuous derivatives. The penalized smoothing spline f minimizes the penalized squared error

$$\mathrm{PENSSE}_{\lambda,m} = \int \{y(t) - f(t)\}^2 + \lambda\, \mathrm{PEN}_m(t)\, \mathrm{d}t. \tag{3}$$

When the roughness parameter is set to $\lambda=0$, the penalized squared error drops out, and the function fits the data perfectly. Larger values of λ penalize the function for being curvy and, as $\lambda \to \infty$, the fitted curve approaches a linear regression model.

Penalized smoothing splines are widely used in functional data analysis (e.g. Ramsay and Silverman (2005)). They are advantageous in terms of their flexibility, which results in good data fit, in terms of their ease of obtaining derivatives (i.e. dynamics), and in terms of their computational efficiency. However, there are several challenges when applying penalized smoothing splines to the on-line auction context. Firstly, although bidding data are non-decreasing over time, smoothing splines do not necessarily result in monotonically non-decreasing curves. Hence, they may not truly reflect the monotonic nature of the auction price process. Second, the fitted curves are often very variable, especially at their ends, even with a heavy smoothness penalty. This is problematic in the auction context, where the opening and closing prices are of special importance. Moreover, in a forecasting context it is crucial to obtain precise estimates at the final stages of the auction (Wang *et al.*, 2008a). Finally, smoothing splines require the specification of many nuisance parameters (such as the smoothing parameter, the number and position of knots, and the polynomial order) which are often determined in an *ad hoc* fashion. Although one can estimate some of these parameters from the data (e.g. by using cross-validation), their optimal choice is not always guaranteed.

3.2. Monotone splines

Monotone splines (Ramsay, 1998) are a natural alternative to smoothing splines in the on-line auction context, since they guarantee a monotone path of the resulting price process. The idea behind monotone smoothing is that monotonously 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

$$D^2 f(t) = w(t)\, D f(t). \tag{4}$$

Here, $w(t) = D^2 f(t)/D f(t)$, which is the ratio between acceleration and velocity. The differential equation has the solution

$$f(t) = \beta_0 + \beta_1 \int_{t_0}^{t} \exp\left\{ \int_{t_0}^{v} w(v)\, \mathrm{d}v \right\} \mathrm{d}u \tag{5}$$

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

$$f(t) = \beta_0 + \beta_1 \exp(wt) \tag{6}$$

and estimate β_0, β_1 and $w(t)$ from the data. Since $w(t)$ has no constraints it may be defined as a linear combination of K known basis functions (i.e. $w(t) = \Sigma_k c_k\, \phi_k(t)$). The penalized least squares criterion is thus

$$\text{PENSSE}_\lambda = \sum_i \{y_i - f(t)\}^2 + \lambda \int_0^T \{w^2(t)\}^2 \, dt. \tag{7}$$

For capturing on-line auction price paths and dynamics, monotone smoothing indeed solves the excess variability of penalized smoothing splines and their non-monotonicity problems. The resulting curves are better representations of a continuous non-decreasing price path, and dynamics can be computed via curve derivatives. However, some challenges remain and new ones arise. First, monotone smoothing is computationally intensive, as it relies on an iterative fitting process where several passes have to be made through the data. Therefore, fitting a data set of even tens or hundreds of auctions can take a long time. Second, like smoothing splines, there are many nuisance parameters to be determined (the number and location of knots and the smoothing parameter). Hence, although conceptually monotone splines are preferable over penalized smoothing splines, they can be slow to implement even with medium-sized data sets.

3.3. Parametric growth models

To overcome the disadvantages of smoothing splines and monotone splines, Hyde *et al.* (2008) proposed a family of four growth models for representing the price process in on-line auctions. They found that the shape of auction price paths can be categorized into four main types: exponential growth, logarithmic growth, logistic growth and reflected logistic growth. These four models not only provide a parametric fit of monotone data, but they also have appealing interpretations and are easy to estimate. We briefly describe each of the four models next.

3.3.1. 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 exponential growth, the rate of growth is proportional to a function's current magnitude, i.e. growth follows the differential equation

$$Y'(t) = r\,Y(t), \tag{8}$$

or the equivalent equation

$$Y(t) = A \exp(rt), \tag{9}$$

where t denotes 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. In an on-line auction context, exponential growth describes a price process with gradual price increases until mid- to late auction, and a heavy price jump towards the end.

3.3.2. Logarithmic model

Logarithmic growth is technically the inverse of the exponential function,

$$Y(t) = \frac{1}{r} \ln\left(\frac{t}{A}\right). \tag{10}$$

The resulting curves are reflections of exponential growth over the line $x = y$. In the on-line auction context, such behaviour occurs when early bidding quickly increases the price during the opening stages of the auction but, because of market constraints (e.g. the existence of a market value or budget constraints), the price flattens out for the remainder of the auction. This type of price behaviour tends to be rare, as most bidders do not wish to reveal their valuations early in the auction. However, inexperienced bidders who may not completely understand eBay's proxy bidding mechanism might place high early bids (see `http://pages.ebay.com/help/buy/automatic-bidding.html`).

3.3.3. *Logistic model*

Logistic growth is useful for describing processes which reach a limit or a 'carrying capacity'. In the context of auction prices, in many cases there are competing on-line and 'brick-and-mortar' markets for the auctioned item, thereby creating a 'market value' for the item.

The logistic model is given by

$$Y(t) = \frac{L}{1 + C\exp(rt)}, \tag{11}$$

and the differential equation is

$$Y'(t) = r\,Y(t)\left\{\frac{Y(t)}{L} - 1\right\}, \tag{12}$$

where L is the carrying capacity, t is time, r is the growth rate and C is a constant. Logistic growth can also be explained in the auction context as a stretched-out 'S'-shaped curve, where the price slowly increases early on, then jumps up during mid-auction and levels off towards the end of the auction. The resulting closing price is analogous to the carrying capacity L in the logistic growth function.

3.3.4. *Reflected logistic model*

The fourth growth model is a reflected S-shaped curve. Such behaviour can be captured by the inverse of logistic growth, or reflected logistic growth, given by the function

$$Y(t) = \frac{\ln(L/t - 1) - \ln(C)}{r}. \tag{13}$$

In the on-line auction context, this type of growth occurs when there is some early bidding 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 approaches its close. In particular, price spikes near the end may be caused by 'sniping' (i.e. last moment bidding).

3.3.5. *Four-member growth model family*

The set of four growth models is used to approximate price paths as follows: for a data set of auctions, each of the four models is fitted to each auction. Then, for each auction, the four estimated models are compared in terms of fit, and the best-fitting model is chosen (for more on the fitting process, see Hyde *et al.* (2008)). Hence, the fitting process is a two-stage process.

Since the family is entirely parametric, no nuisance parameters require determination. Moreover, since the family is monotonic, it is well suited for capturing auction price processes.

The four-member family of growth models is computationally efficient compared with monotone splines, and ordinary least squares can be used for estimation.

The main disadvantage of the four-model family is its limitation to only four basic shapes—exponential, logarithmic, logistic and reflected logistic—which may be overly simplistic for some auction scenarios. Moreover, because the four models are not nested within a single model, comparing the fit (for choosing the best model) is non-trivial. Finally, when fitting the exponential and logistic models via least squares, the models minimize error in the bid *amount* space. In contrast, when fitting the two reflected models (logarithmic and reflected logistic growth) by using least squares, the error minimization is done in the bid *time* space. A comparison is therefore more complicated.

3.4. Comparison

To illustrate the difference between penalized splines, monotone splines and the four-member family of growth models, consider Fig. 2, which displays the fit of the three methods to two sample auction price paths. The full curves represent *P*-splines; the broken curves represent monotone splines; in both cases we chose the smoothing parameter via cross-validation. The broken curves represent the best fit of the four-member growth family and we selected the 'best' family member via the iterative process that is outlined in Hyde *et al.* (2008). We see that, although *P*-splines can fit the data very well, they are very variable and do not capture the monotonic nature of the price path. Although the four-member growth family results in a monotonous price path, it does not fit the data well. Monotone splines appear to provide the best fit in this example—at least visually; however, it takes on average almost 7 s to fit a single monotone spline (compared with 0.02 s for one *P*-spline and 0.04 s for one four-member growth model).

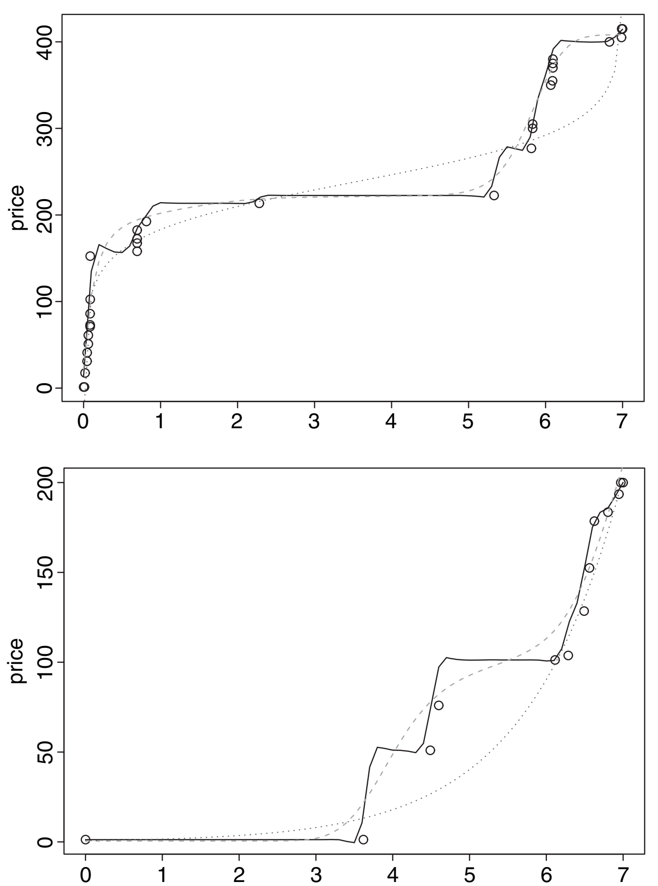


Fig. 2. Illustration of the three existing smoothing methods: ——, *P*-splines; – – –, monotone splines; ·······, growth models

We also point out that we could apply any of the methods that are presented in this paper (including the beta model that is presented in the next section) to the *partial* price path rather than to the complete price path. For example, rather than applying the penalized spline to data from the entire auction, we could apply it only to the information after day 4, thereby eliminating effects from early bidding.

4. New model for auction price paths: the beta model

In light of the shortcomings of existing models for on-line auction price paths and dynamics, we introduce a single parametric model that is flexible yet parsimonious for approximating price paths and their dynamics. Our proposed model is based on the beta cumulative distribution function (CDF). The beta distribution is a continuous probability distribution that is defined on the interval [0,1] with two shape parameters (α, β) that fully determine the distribution. Its CDF can be written as

$$F(x, \alpha, \beta) = \frac{\int_0^x u^{\alpha-1}(1-u)^{\beta-1}\,\mathrm{d}u}{B(\alpha, \beta)}, \tag{14}$$

where $B(\alpha, \beta)$ is the *beta* function $(B(\alpha, \beta) = \int_0^1 u^{\alpha-1}(1-u)^{\beta-1}\,\mathrm{d}u)$, which serves as a normalization constant in the CDF to ensure that $F(1, \alpha, \beta) = 1$.

Using the beta model to capture auction price paths entails standardizing the data. Since auctions can be of varying durations, we standardize the time sequence to a 0–1-scale. Similarly, because auctions close at different prices, we standardize the observed bids to a 0–1-scale. The goal is then to find the values of α and β that most closely fit the standardized data. Finally, the estimates for the price and velocity are obtained by 'unstandardizing' the beta model estimates. An algorithm for achieving this goal is described in Section 4.1.

The beta model is very flexible in the types of curve that it can produce. It includes as special cases the four shapes of the four-member growth model family of Hyde *et al.* (2008). Fig. 3 shows the beta model curves for various values of α and β. The full curve represents the case where price grows rapidly at the beginning of the auction and at the end, but not in the middle, corresponding to logit growth. The broken curve represents the situation where rapid growth occurs only at the end, corresponding to exponential growth. The dotted curve shows early rapid growth, corresponding to logarithmic growth. And, finally, the chain curve captures a rapid increase in price somewhere in the middle of the auction, corresponding to the reverse logit growth pattern.

4.1. *Fitting the beta model*

Fitting the beta model to bid data can be done in a way that results in curves that fit well in two dimensions: bid time and bid amount. In the auction context both dimensions are important. In particular, a good fit in terms of the bid timing is necessary to capture points of different bidding activities accurately. In contrast, a model that adequately captures bid amounts is necessary for generating accurate price forecasts. The only inputs that are required for fitting the beta CDF are the observed bid amounts and their associated time stamps. The resulting price path representation is characterized by only two parameters. The simplicity and parsimony of the beta model distinguish it from alternative approaches. In what follows we describe an algorithm for fitting the beta CDF in a way that minimizes residuals in both bid amount and bid time dimensions simultaneously.

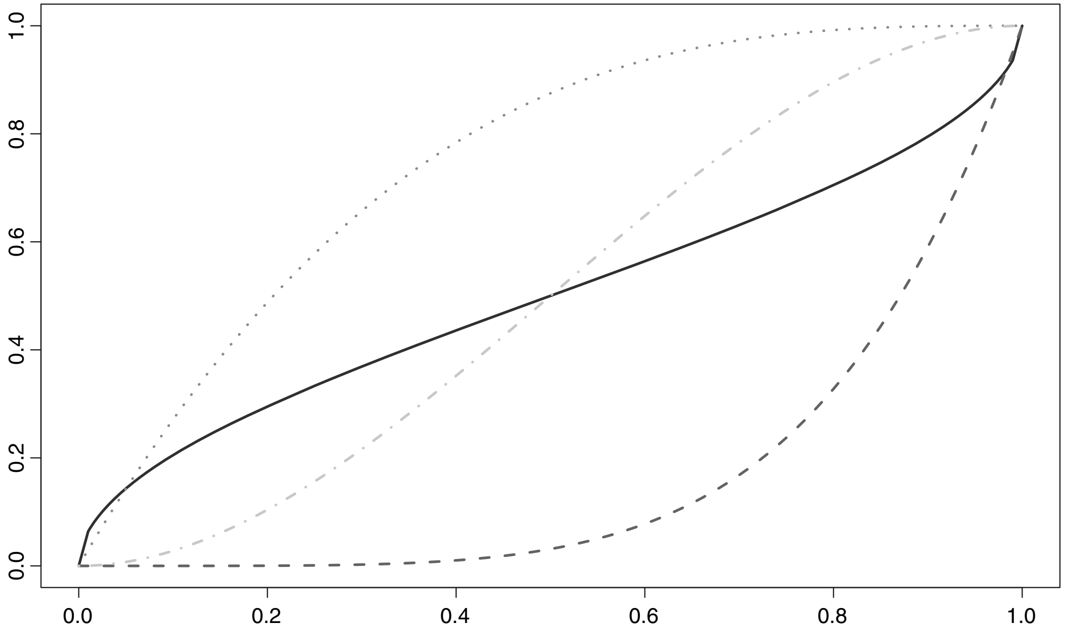


Fig. 3. Beta CDF with various shape parameters (α,β): ———, (0.5,0.5); – – –, (5,1); ·······, (1,3); –·–·–, (2,2)

4.1.1. Beta fitting algorithm

For a given auction, we estimate α and β from the observed bids as follows.

Step 1: standardize bid amounts and bid times. Since the range y as well as the domain x of the beta CDF are [0,1], we first standardize the bid amounts and bid times by the two transformations

$$y \leftarrow \frac{\text{bid} - \min(\text{bid})}{\max(\text{bid}) - \min(\text{bid})}$$

and

$$x \leftarrow \frac{\text{time} - \min(\text{time})}{\max(\text{time}) - \min(\text{time})}.$$

x and y are now bid times and bid amounts standardized within [0,1].

Step 2: compute α_0 and β_0, the initial values of $\hat{\alpha}$ and $\hat{\beta}$. Since we treat x as a beta-distributed random variable, it is reasonable to assume that the empirical average and variance of x are close to their theoretical mean and variance, i.e. $\text{mean}(x) \simeq \alpha/(\alpha+\beta)$ and $\text{var}(x) \simeq \alpha\beta/\{(\alpha+\beta)^2(\alpha+\beta+1)\}$. Therefore, the initial values of α and β are found by solving the minimization problem

$$(\alpha_0, \beta_0) = \{(\alpha^*, \beta^*) : \text{DIST}^{\text{A}}(\alpha^*, \beta^*) = \min\{\text{DIST}^{\text{A}}(\alpha, \beta)\}\},$$

where

$$\text{DIST}^{\text{A}}(\alpha, \beta) = \left\{\text{mean}(x) - \frac{\alpha}{\alpha+\beta}\right\}^2 + \left\{\text{var}(x) - \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}\right\}^2.$$

Step 3: compute $\hat{\alpha}$ and $\hat{\beta}$. To capture both the bid levels as well as the bid times, our model minimizes error in both x- and y-directions simultaneously. Specifically, we choose to minimize the sum of the squared residuals in the x- and y-directions. With the initial values α_0 and β_0 from step 2, we solve for $\hat{\alpha}$ and $\hat{\beta}$ through the minimization problem

$$(\hat{\alpha}, \hat{\beta}) = \{(\alpha^*, \beta^*) : \mathrm{DIST}^{\mathrm{B}}(\alpha^*, \beta^*) = \min\{\mathrm{DIST}^{\mathrm{B}}(\alpha, \beta)\}\}$$

where

$$\mathrm{DIST}^{\mathrm{B}}(\alpha, \beta) := \sum \{y - \mathrm{pbeta}(x, \alpha, \beta)\}^2 + \sum \{x - \mathrm{qbeta}(y, \alpha, \beta)\}^2,$$

and pbeta and qbeta represent the cumulative beta distribution function and the inverse of the cumulative beta distribution function respectively.

Obtaining price estimates at time T is achieved by looking up the estimated beta CDF at standardized time T, and then unstandardizing the estimated CDF by using the inverse transformation for y in step 1, i.e. multiply by max(bid) − min(bid) and add min(bid).

The above algorithm is computationally efficient. It takes, on average, 0.0329 s to fit the beta model to one auction, which compares favourably with the four-member growth family (0.0305 s). Unsurprisingly, penalized smoothing splines are faster (0.0184 s) since they do not require iterations. Conversely, fitting monotone splines, which do require iterative passes through the data, takes 50 times longer to compute (on average, 1.7049 s per auction). All computations are based on an IBM Lenovo T400 laptop computer with 2 Gbytes memory and a dual core 2.26-GHz processor.

Note that $\mathrm{DIST}^{\mathrm{B}}(\alpha, \beta)$ implies an *equal weighting* of both the bid time and the bid amount dimensions. To emphasize one dimension relative to the other, we can generalize the above definition to

$$\mathrm{DIST}^{\mathrm{B}}(\alpha, \beta) = w_y \sum \{y - \mathrm{pbeta}(x, \alpha, \beta)\}^2 + w_x \sum \{x - \mathrm{qbeta}(y, \alpha, \beta)\}^2, \qquad (15)$$

with weights w_x and w_y that satisfy $0 \leqslant w_x, w_y \leqslant 1$ and $w_x + w_y = 1$. For instance, setting $w_x = 0$ and $w_y = 1$ leads to price paths that closely track the bid amounts but ignore their timing. In our context, we choose equal weights since strategizing about when to place a bid is equally important to a bidder as how much to bid. For instance, a bidder who focuses on only the amount of the bid might reveal his or her intentions too early and hence be outbid; conversely, a bidder who strategizes about only the time of the bid might overbid and end up overpaying. One may want to overweight time (i.e. to choose a large w_x) if capturing bid timing is of special interest. One such case is in studying bid shilling, where a seller illegally bids on his or her own auction if the price has not reached a certain level by a certain time (Kauffman and Wood, 2005).

4.2. Properties of the beta model

The beta model shares the main advantages of competing methods (P-splines, monotone splines and the four-family growth model), but it also has several additional properties that set it apart. Like all competing methods, the derivatives of the continuous beta curves can be used to capture price dynamics. The beta model produces monotonically non-decreasing curves, yet it is computationally fast (using the algorithm in Section 4.1). Unlike non-parametric approaches, fitting the beta model does not involve any nuisance parameters.

The beta model has two additional unique properties, which make it especially advantageous in the on-line auction context: first, because both of its dimensions (bid time and bid amount) are derived from a probability function, the beta summary statistics can be used to learn about the bid timing distribution. Although not directly related to realtime forecasting, this property is useful in other on-line auction applications (see Section 6). Second, there is an easy and straightforward way to measure pairwise distances between price paths. The latter is especially useful in the context of pairwise comparisons, clustering and dynamic forecasting. In particular, Zhang *et al.* (2009) have used this property within a functional K-nearest-neighbour forecaster to produce more accurate price forecasts.

We discuss each of the beta model properties in detail next.

4.2.1. Representing price dynamics

The beta CDF representation of the price paths means that price velocity, which is the first derivative of the price curve, is given by the beta probability density function. In particular, at any given time T, the price velocity of an auction with shape parameters α and β can be computed as

$$\text{Vel}(t, \alpha, \beta) = \frac{t^{\alpha-1}(1-t)^{\beta-1}}{B(\alpha, \beta)}, \tag{16}$$

where t is the standardized T on a scale of [0,1] ($t = T/\text{duration}$).

Higher order price dynamics can also be readily obtained by taking higher order derivatives. For example, price acceleration can be computed as

$$\text{Acc}(t, \alpha, \beta) = \frac{t^{\alpha-1}(1-t)^{\beta-1}}{B(\alpha, \beta)} \left(\frac{\alpha-1}{t} - \frac{\beta-1}{1-t} \right).$$

As pointed out earlier, price dynamics carry important information about the auction process. Therefore accurate approximations of price dynamics are beneficial across multiple applications. In Section 5 we show that, in a forecasting context, price dynamics that are generated via the beta model lead to more accurate predictions of the final price compared with competing approaches.

4.2.2. Characterizing growth patterns

Similar to the four-member growth family of Hyde *et al.* (2008), the beta model provides a tool for characterizing price process types. In fact, the four growth models are special cases of the beta model. For example, if both α and β are smaller than 1, then the price curve is similar to the reflected logistic model. Table 3 lists the relationship between the beta parameters and each of the four growth models. The implication of this relationship is that it allows us to characterize auctions easily in terms of their type of price dynamics, without the need for more specialized techniques such as functional clustering (e.g. Jank and Shmueli (2008)) or via laborious visual examination (e.g. Hyde *et al.* (2006)).

Table 3. Correspondence between the beta model and the four growth models

Growth model	*Beta model*	
Exponential	$\alpha = 1$	$\beta < 1$
	$\alpha > 1$	$\beta \leqslant 1$
Logarithmic	$\alpha < 1$	$\beta \geqslant 1$
	$\alpha = 1$	$\beta > 1$
Logistic	$\alpha > 1$	$\beta > 1$
Reflected logistic	$\alpha < 1$	$\beta < 1$

Table 4. Beta distribution summary statistics and their auction meaning

Statistic	*Formula*	*Explanation*
Variance	$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$	Dispersion of the bid arrivals
Mode	$\frac{\alpha-1}{\alpha+\beta-2}$	Peak of the velocity curve; price increases fastest at this point
Skewness	$\frac{2(\beta-\alpha)\surd(\alpha+\beta+1)}{(\alpha+\beta+2)\surd(\alpha\beta)}$	Asymmetry of the bid arrivals

4.2.3. Characterizing bid timing

The estimated beta parameters α and β can be used to compute summary statistics which capture bid timing information. Table 4 gives the formulae for the variance, mode and skewness. The variance gives information about the dispersion of the bid arrivals; the mode, which is the peak of the price velocity curve, tells us about the time during the auction when the price moved fastest. Finally, skewness measures the level of asymmetry in the bid timings. On-line auctions tend to see either high bidding activity at the start and/or at the end.

4.2.4. Measuring distances between price paths

Unlike other models for price paths, the beta model allows us to measure the *distance* between two auction price paths easily. If both price paths are beta curves, then the distance between them can be measured via the Kullback–Leibler (KL) distance (Kullback and Leibler, 1951). The KL distance is a non-commutative measure of the difference between two probability distributions. For two distributions X and Y, it measures how different Y is from X. The KL distance is widely used in the field of pattern recognition for feature selection (Basseville, 1989), and in physics for determining the states of atoms or other particles (Nalewajski and Parr, 2000).

In the case of the beta model, the KL distance is especially simple. Consider X and Y beta distributed with parameters (α, β) and (α', β') respectively. The KL distance between X and Y is then given by a function of the beta parameters alone (Raubera *et al.*, 2008):

$$D_{\mathrm{KL}}(X,Y)=\ln\left\{\frac{B(\alpha',\beta')}{B(\alpha,\beta)}\right\}-(\alpha'-\alpha)\,\psi(\alpha)-(\beta'-\beta)\,\psi(\beta)+(\alpha'-\alpha+\beta'-\beta)\,\psi(\alpha+\beta), \tag{17}$$

where B and ψ are *beta* and *digamma* functions respectively.

The importance of this property is that it allows us to capture the functional structure of curves. Measuring the distance between curves that are estimated from noisy, irregularly sampled data is complicated because there is no single convention on how best to determine the distance between two functional objects (Peng and Müller, 2008). Discretizing the curve and using a pointwise approach have been shown to lead to problems (Abraham *et al.*, 2008). First, the choice of the grid is highly user specific and it may strongly influence the results. Second, after discretizing the curve, one must choose a suitable distance metric and it is not clear whether using the Euclidean distance is better or worse than some other distance metric (e.g. the Manhattan distance). Moreover, using a grid approach does not take into account measurement errors. In fact, using a pointwise approach and measuring the distance via, for example, the Euclidean distance can lead to a loss of information. To illustrate this, consider the four curves in Fig. 3. Note that the shape of beta(0.5,0.5) is very different from that of beta(2,2): when beta(2,2) is *concave*, beta(0.5,0.5) is *convex*, and vice versa. Whereas the KL distance captures this difference, the Euclidean distance does not. Zhang *et al.* (2009) used this property to develop a functional K-nearest-neighbour algorithm and showed that it can increase the forecasting accuracy especially when auctions are very heterogeneous.

5. Empirical comparison of forecast accuracy

In this section we compare the proposed beta model with its three competitors: P-splines, monotone splines and the four-member growth model family. We compare the four methods in terms of their ability to produce accurate estimates for the price and its dynamics, which are then plugged into a forecasting model. As mentioned earlier, recent research has shown that the incorporation of dynamics into realtime forecasting models for the price of an on-line auction can significantly improve forecasting accuracy. However, what has not been investigated to date is whether the method that is used to *estimate* these dynamics has an effect on the forecast accuracy. Recall that by 'dynamics' we refer to the price velocity or the price acceleration. The price velocity measures how fast the price is changing and it can be estimated via the first derivative of the corresponding price curve. Similarly, the price acceleration measures changes in the price velocity and it can be computed via the second derivative of the price curve. It is quite plausible that different estimates for the price curve will yield different estimates for the dynamics and, as a consequence, will result in varying forecasting performance. We investigate this next.

We have already shown that the beta model can be computed very quickly: much faster than monotone splines and almost as fast as P-splines. Hence, the beta model lends itself to realtime forecasting scenarios where updates need to be computed in split seconds for many hundreds or thousands of auctions at a time. What still remains to be shown is whether these fast computing times also lead to accurate forecasts. We have argued earlier that P-splines, despite being very fast to compute, do not capture the monotone auction process and thus may not be as useful for forecasting. In contrast, although monotone splines are, at least in theory, a natural choice for predicting the continuation of an auction process, they are slow to compute. We now compare the beta model against its competitors in terms of predictive accuracy.

We note that the focus of what follows is *not* on the development or comparison of different classes of forecasting models; this has been done in prior literature (e.g. Ghani and Simmons (2004), Jap and Naik (2008), Wang *et al.* (2008a), Dass *et al.* (2009), Zhang *et al.* (2009) and Jank and Zhang (2008)). Rather, our goal is to compare the *input* for such forecasting models, i.e. our goal is to show that the way in which price dynamics are estimated matters and that, given a particular forecasting model, some dynamics estimates lead to more accurate forecasts than others. In that sense, we hold the forecasting model itself constant and only vary its input (i.e. the estimates for the price path and its dynamics), by alternating the way that price dynamics are modelled.

We conduct our investigation in the following way. For a given forecasting model (e.g. a linear model or a non-linear model), we assemble four sets of predictors: one for each of our four price path models. All four sets contain the same input information (for example, they all contain the opening price, the auction duration or the price velocity); their only difference is in *how* we compute estimates for the price velocity. Then, by including each of the four sets of predictors as input into the same forecasting model, we observe the differences in forecasting accuracy (measured on a hold-out set). We can hence conclude that the observed differences must be due to the differences in how we computed the price dynamics. A detailed description of the forecasting models that were used is described next.

5.1. Forecasting models

We compare the effect of price dynamics within four different types of forecasting model—a linear model, a generalized additive model, a neural network and a regression tree. We start by describing a linear forecasting model. Consider an on-going auction at time point T. For instance, for a 7-day auction (which is the most common auction length on eBay), $T=5$ would imply that the auction has been on going for 5 days. Our goal is to forecast, at time T, the final price of auction i; in the above example, our goal would be to forecast at day 5 the final price of auction i by using price path model j. The linear forecaster is given by

$$\text{FinalPrice}_{i,j}(T)=\boldsymbol{\beta}_1'\mathbf{X}_i+\beta_2'\,\text{Price}_{i,j}(T)+\beta_3'\,\text{Velocity}_{i,j}(T), \tag{18}$$

where the design matrix $\mathbf{X}$ includes control variables that characterize the seller, the product and the auction features. Note that the information in $\mathbf{X}$ (the control variables) is identical across all forecasting models. For more information on these control variables, see Appendix A.

Whereas $\mathbf{X}$ is identical and held constant, we vary the information in $\text{Price}_{i,j}(T)$ and $\text{Velocity}_{i,j}(T)$, which denote estimates for the price and its velocity at time T respectively, by using price path model j, $j\in\{$beta model, 4-member growth family, penalized spline, monotone spline$\}$. We therefore obtain, for the linear regression forecasting model, four different forecasts for the final price of each auction, each arising from a different price path model.

A comment on the estimation of the price dynamics is in order. $\text{Price}_{i,j}(T)$ and $\text{Velocity}_{i,j}(T)$ denote the current price and its velocity *at time* T, which is during the on-going auction. To obtain these estimates, we estimate the price path model by using the bids that have been placed only *up to time* T. Moreover, since the beta model is estimated for the *standardized* bid times and amounts, we obtain estimates of price and velocity by unstandardizing the estimates from the beta model using the inverse transformation from step 1 in Section 4.1.

In addition to the linear model (18), we also investigate three non-linear forecasting model formulations. These include a generalized additive model, a neural network and a regression tree. The rationale behind these additional model formulations is that differences in observed forecasting accuracies may be due to non-linearities or higher order interaction terms between the price dynamics and the response, and investigating more flexible modelling alternatives allows us to quantify that difference. In summary, we obtain a 4×4 design with four price path models and four forecasting models.

5.2. Results

Forecasting performances of all forecasters are evaluated on a hold-out set. In particular, we partition the set of auctions into training and hold-out sets, each consisting of 50% of the auctions. Both sets reflect the temporal nature of our data in that the training set contains auctions that transacted earlier and the hold-out set contains later auctions. In that sense, our partitioning reflects the true chronology that would be used in reality, where past and current information is used to predict future auctions. Model parameters are estimated by using the training set, and then the predictive accuracy is measured by computing the mean absolute percentage error MAPE in the hold-out set:

$$\text{MAPE}=\frac{1}{N}\sum_{i=1}^{N}\left|\frac{y_i-\hat{y}_i}{y_i}\right|, \tag{19}$$

where y_i and $\hat{y}_i$ denote the true and estimated final price in auction i (in the hold-out set) respectively.

Fig. 4 shows the results. The four panels correspond to each of the four forecasting models.

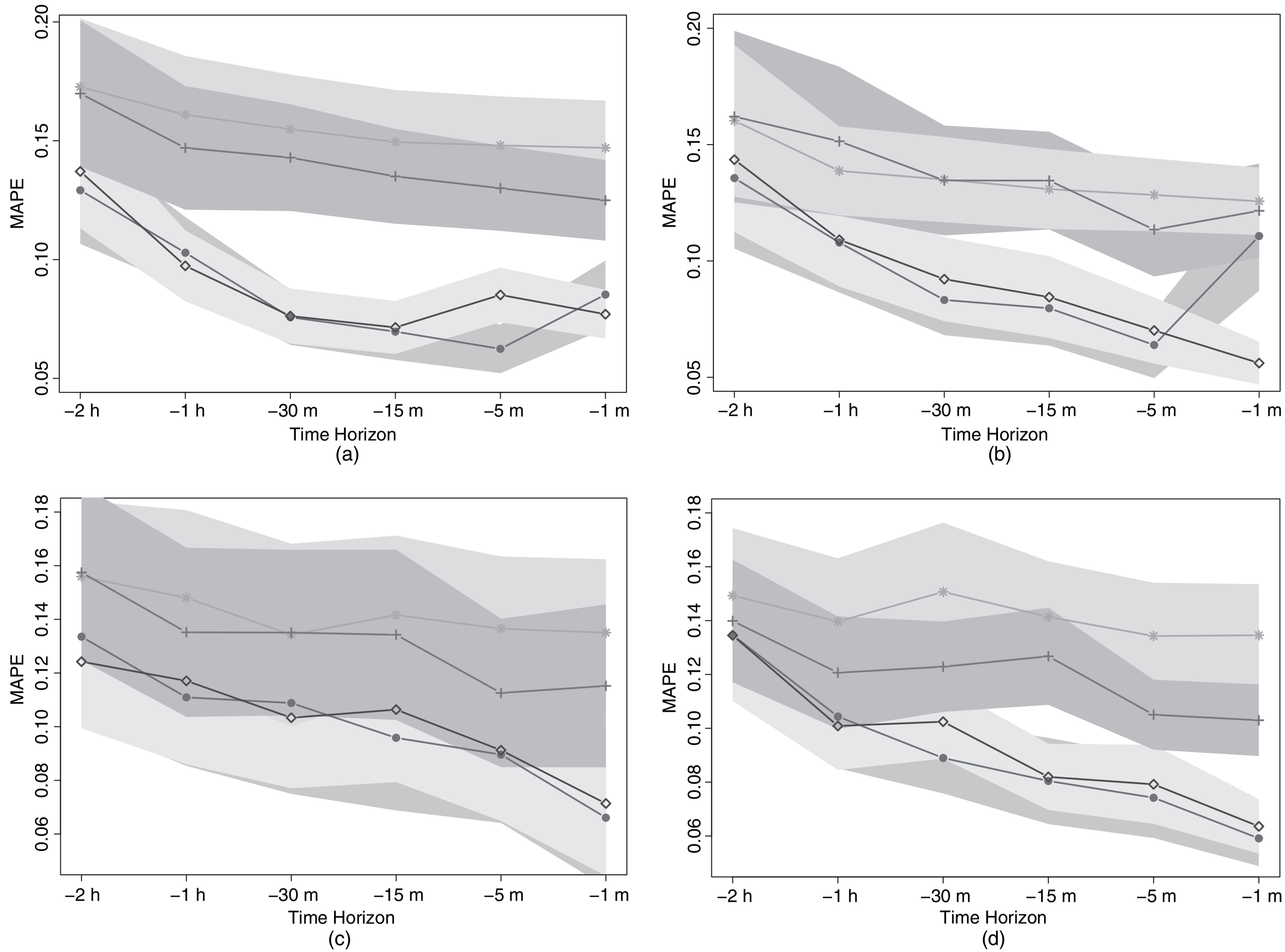


Fig. 4. Forecasting accuracy of various price path models, by forecasting model (the x-axis corresponds to the forecasting window; ■, ■, mean ± 2-standard-error confidence bounds; ●, beta model; +, growth model; ✳, *P*-splines; ◇, monotone splines): (a) linear model; (b) generalized additive model; (c) neural net; (d) regression tree

In each panel, MAPE is shown as a function of the forecasting horizon T, i.e. the amount of time *before the end of the auction* at which we compute the forecast. For instance, $T=-1$ h means that we are forecasting the price 1 h before the auction closes. Fig. 4 shows results for a maximum forecasting horizon of 2 h, since the last moments of an auction are of special importance as many bidders (such as 'snipers' or 'last moment bidders') place bids very late. Although longer forecasting horizons are possible (see, for example, Wang *et al.* (2008a)), the forecasting accuracy usually decreases as we try to predict further into the future.

As expected, we can see that, the longer we wait (i.e. as T approaches 0), the better the forecast. We can also see that, regardless of the forecaster, dynamics based on the beta model and the monotone spline always outperform those based on the penalized spline or the four-member growth model. Whereas this difference is most pronounced in the linear forecaster (Fig. 4(a)), it is statistically insignificant for the neural network (Fig. 4(c)), suggesting that the relationship between dynamics and the response is highly non-linear. Fig. 5 also shows the forecast error distributions of the linear forecaster at $T=-1$ min, using each of the four price path models. Although all four distributions are quite symmetric, the penalized splines and the four-member growth model incur the largest variance. Also, errors are not centred perfectly around zero as there is a general price decrease from the training to the hold-out set.

Finally, we also compare the functional forecaster with two simpler and more common methods for time series data. In particular, we investigate the performance of moving averages and a time series regression model for forecasting the final auction price. The results (which are not reproduced here) suggest that simpler methods perform very poorly. The reason is that in on-line auction data, unlike in standard time series data, the arrival of events (i.e. bids) is user determined and thus happens at irregular and unevenly spaced times which make the application of off-the-shelf forecasting methods difficult.

In summary, although the performances of the beta model and the monotone splines appear very similar, it takes over 50 times more computing time to estimate the monotone spline. For instance, in our data which consist of 698 auctions, it takes only 23 s to fit all beta models, but over 1190 s (about 20 min) to fit all monotone splines. Given that the goal is to perform realtime forecasting and given that forecasts need to be updated 5 min or sometimes 1 min before the close of the auction, a computing time of 20 min is not very practical. We can thus conclude that the beta model is a very attractive alternative to existing methods when the goal is to obtain fast and accurate estimates of the price path.

5.3. Extensions

One direction for extending the beta model is to capture cross-auction information by applying ideas from mixed modelling. The beta model as well as all other price path models that were considered in this paper treat auctions as *independent* of one another. For instance, to fit a penalized smoothing spline to an auction, we minimize the penalized error term in equation (3) independently of all other auctions in our data. Similarly, the beta model (14) does not account for any correlation across auctions. However, two auctions might exhibit some correlation if they sell the same (or similar) product, if they are offered by the same seller or are sought after by the same bidder(s). None of the methods that are considered in this paper take this potential correlation into account.

Accounting for this correlation can help our understanding of *competition* between auctions (Haruvy *et al.*, 2008a). Reithinger *et al.* (2008) proposed a semiparametric mixed model and found a negative correlation between the slopes and intercepts of auction price paths, which implies that auctions for the same product have a tendency to self-correct. Similar ideas could

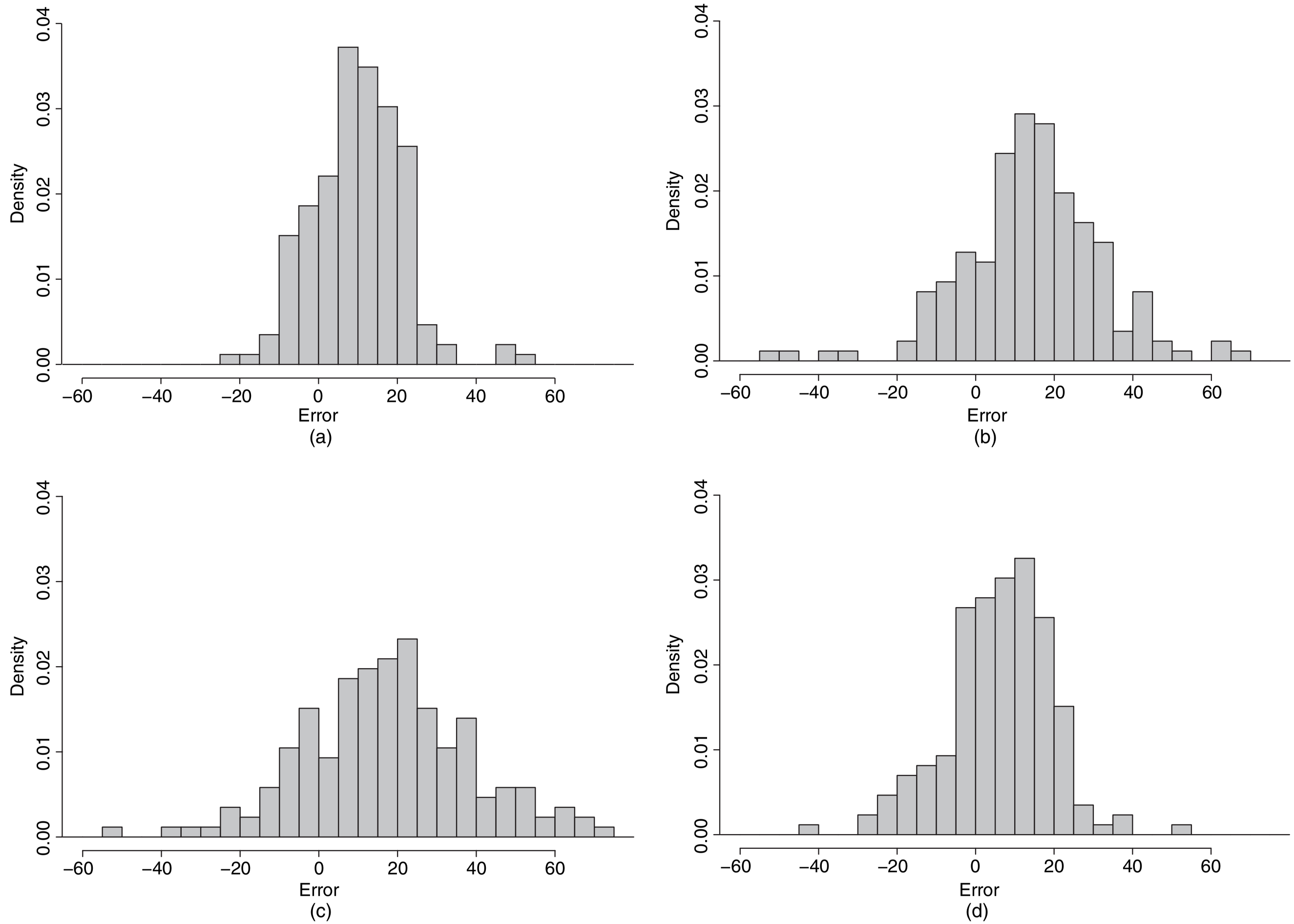


Fig. 5. Error distribution for each of the four path models, based on the linear forecaster (Fig. 4(a)) when forecasting at −1 min: (a) beta model; (b) growth model; (c) *P*-splines; (d) monotone splines

be applied to any of the four price path models that are described here. However, although the conceptual framework of mixed models could, at least in theory, be used to capture correlation across auctions, it is not quite clear how we would estimate a monotone spline with a random effect in equation (5) or how we would apply the concepts of mixed models to the beta model (14). We are hopeful though that our results will inspire future research on this topic.

Another potential improvement to the beta model could be terms of computation, especially when realtime updates are needed. Updating estimates of the price dynamics by using the algorithm that was developed in Section 4.1 requires refitting the beta model in every step. The reason is that the prefitting bid standardization (in step 1) uses the highest current bid for standardizing the bid amounts. When additional bids arrive, the information on the highest bid changes and requires restandardizing the bid amounts accordingly. It could be quite useful to develop a modified algorithm that updates the beta parameter estimates in an on-line fashion.

6. Beyond forecasting

The price path and its dynamics are useful not only in realtime forecasting applications but also for studying and characterizing price profiles, visualizing sets of auction paths and investigating different factors that affect or interact with price paths. The beta model proposed offers a contribution to such studies, due to its flexibility, parsimony and many useful features.

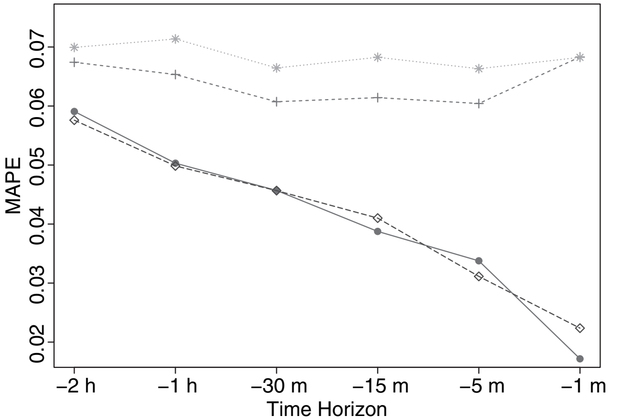
For visualization, it is often useful to plot a set of price paths and/or price dynamics for a set of auctions over calendar time. Rug plots (Hyde *et al.*, 2006) are a useful display for visualizing a set of auctions by describing their entire price evolution, for the purpose of exploring similarities and patterns in concurrent (or partially concurrent) auctions. Once auctions have been plotted, it is useful to investigate types of curve to detect patterns. Hyde *et al.* (2008) proposed the use of the four-member growth curve family for categorizing the different types of curve. The beta model can offer an improvement by offering a single, more parsimonious model for categorizing a wide range of price paths. An interesting extension would be to define distances between partially overlapping auctions by using the beta model.

A second use of price dynamics which is unrelated to price forecasting is identifying types of auction by their price dynamics. Although there has been some work using clustering to identify types of on-line auctions in terms of bidding behaviour (Bapna *et al.*, 2004; Jank and Shmueli, 2008; Peng and Müller, 2008), the main challenge when using functional representations of price paths is defining the distance between curves (Peng and Müller, 2008). The beta model offers the ability to define easily and to compute efficiently a distance measure between two price paths (using the KL distance), which also takes into account the shape of the curves.

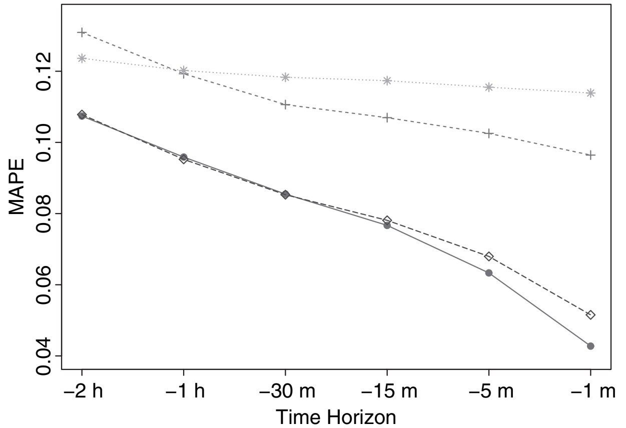
In terms of assessing the effect of various factors on price dynamics, there has been much interest in understanding the effects of product features (e.g. make or model), sellers' characteristics (e.g. the seller's reputation) and auction features (e.g. the length of the auction or its opening bid) on the outcome of an auction (i.e. its final price). However, although all these factors do affect an auction's outcome, they also affect its dynamics, which in turn affect the final price (Bapna *et al.*, 2008). Shmueli and Jank (2008), for instance, illustrated the effect of auction features on the auction dynamics and found that higher opening bids result in lower price dynamics. Jank *et al.* (2008) expanded on this finding and developed model-based regression trees to relate differential equation models to auction features. Thus, a better understanding of

Table 5. Estimated coefficients for the linear forecaster at −1 min

Variable	*Estimate*	*Standard error*	*p-value*
Intercept	50.5997	7.4709	0.0000
Opening price	0.1811	0.5075	0.7213
Duration	−2.0693	0.4286	0.0000
Seller feedback	1.4486	0.5266	0.0062
Condition	8.1408	4.3102	0.0595
Reserve price	17.7990	5.3410	0.0009
Buy it now	2.3612	2.5950	0.3633
Bold	10.1502	3.5172	0.0041
Picture	−3.0576	2.3664	0.1969
Number of bids	1.8106	3.0246	0.5497
Number of bidders	0.2298	3.8884	0.9529
Price_{Ti}	0.6383	0.0211	0.0000
Velocity_{Ti}	0.0011	0.0002	0.0000



(a)



(b)

Fig. 6. Model comparison for (a) personal digital assistants and (b) laptops (in both panels, we employ the linear forecaster from Fig. 4(a)): ●, beta model; +, growth model; ✳, *P*-splines; ◇, monotone splines

price dynamics can also lead to a better understanding of the precise role of auction features or product descriptions and their effect on the outcome of an on-line auction.

Another component that is receiving increased interest is competition (Haruvy *et al.*, 2008b). This includes competition between bidders within the same on-line auction, across multiple auctions, or even beyond the on-line auction marketplace. The price path and its dynamics can reflect unobservable auction behaviour such as the degree of competition between bidders, both within the same auction as well as across different auctions (Dass *et al.*, 2009). Therefore, adequately capturing and modelling the price path can help to understand the effects of competition.

In summary, the price path and the price dynamics are of special interest in on-line auction research for studying multiple phenomena, and therefore developing models that can capture them effectively and accurately is important and useful for advancing research and for practical applications.

Appendix A: Control variables for forecasting models

The design matrix **X** in equation (18) includes all the control variables that are listed in Tables 1 and 2, except for 'Highlight', which is dropped owing to high collinearity with 'Bold'.

Table 5 presents the estimated coefficients from the linear regression forecaster (estimated at $T = -1$ min). The predictors include the 10 control variables and the estimated price and velocity at time T.

Appendix B: Results for additional data sets

Fig. 6 shows the results for two additional bidding data sets. In addition to the digital camera data, we compare the performance of the beta model on 380 auctions for personal digital assistants and 4965 laptop auctions. The laptop auctions are intriguing since they span a wide variety of brands and models and hence test our method's ability to account for variation across products. In contrast, the final prices of the personal digital assistants are less variable compared with the digital cameras and thus allow us to test the effect of variability of price in the presence of product homogeneity.

Fig. 6 shows the results corresponding to the linear forecaster in Fig. 4(a). We can see that the relative prediction accuracy is robust across the various data sets.

References

Abraham, C., Cornillon, P. A., Matzner-Loeber, E. and Molinari, N. (2008) Unsupervised curve clustering using b-splines. *Scand. J. Statist.*, **30**, 581–595.

Ariely, D. and Simonson, I. (2003) Buying, bidding, playing, or competing?: value accessment and decision dynamics in online auctions. *J. Consum. Psychol.*, **13**, 113–123.

Bapna, R., Goes, P., Gupta, A. and Jin, Y. (2004) User heterogeneity and its impact on electronic auction market design: an empirical exploration. *Mangmnt Inform. Syst. Q.*, **28**, 21–43.

Bapna, R., Jank, W. and Shmueli, G. (2008) Price formation and its dynamics in online auctions. *Decsn Support Syst.*, **44**, 641–656.

Basseville, M. (1989) Distance measure for signal processing and pattern recognition. *Signal Process.*, **18**, 349–369.

Dass, M., Jank, W. and Shmueli, G. (2009) Dynamic price forecasting in simultaneous online art auctions. In *Marketing Intelligent Systems using Soft Computing* (eds J. Casillas and F. J. Martinez-Lopez). New York: Springer.

Dass, M. and Reddy, S. K. (2006) Modeling online art auction dynamics using functional data analysis. *Statist. Sci.*, **21**, 179–193.

Ghani, R. and Simmons, H. (2004) Predicting the end-price of online auctions. *Int. Wrkshp Data Mining and Adaptive Modeling Methods for Economics and Management, Pisa*.

Haruvy, E., Popkowski Leszczyc, P., Carare, O., Cox, J., Greenleaf, E., Jank, W., Jap, S., Park, Y.-H. and Rothkopf, M. (2008a) Competition between auctions. *Marktng Lett.*, **19**, 431–448.

Haruvy, E., Popkowski Leszczyc, P., Carare, O., Cox, J., Greenleaf, E., Jap, S., Jank, W., Park, Y. and Rothkopf, M. (2008b) Competition between auctions. *Marktng Lett.*, **19**, 431–448.

Heyman, J., Orhun, Y. and Ariely, D. (2004) Auction fever: the effect of opponents and quasi-endowment on product valuations. *J. Interact. Marktng*, **18**, 4–21.
Hyde, V., Jank, W. and Shmueli, G. (2006) Investigating concurrency in online auctions through visualization. *Am. Statistn*, **60**, 241–250.
Hyde, V., Jank, W. and Shmueli, G. (2008) A family of growth models for representing the price process in online auctions. In *Statistical Methods in eCommerce Research* (eds W. Jank and G. Shmueli). New York: Wiley.
Jank, W. and Shmueli, G. (2008) Studying heterogeneity of price evolution in ebay auctions via functional clustering. In *Handbook of Information Systems Series: Business Computing* (eds G. Adomavicius and A. Gupta). New York: Elsevier.
Jank, W., Shmueli, G. and Wang, S. (2008) Modeling price dynamics in online auctions via regression trees. In *Statistical Methods in eCommerce Research* (eds W. Jank and G. Shmueli). New York: Wiley.
Jank, W. and Zhang, S. (2008) An automated and data-driven bidding strategy for online auctions. *Technical Report*. University of Maryland, College Park. (Available from `http://ssrn.com/abstract=1427212`.)
Jap, S. and Naik, P. (2008) Bid analyzer: a method for estimation and selection of dynamic bidding models. *Marktng Sci.*, **27**, 949–960.
Kauffman, R. J. and Wood, C. A. (2005) The effects of shilling on final bid prices in online auctions. *Electron. Commerce Res. Appl.*, **4**, no. 2, 21–34.
Kullback, S. and Leibler, R. A. (1951) On information and sufficiency. *Ann. Math. Statist.*, **22**, 79–86.
Nalewajski, R. F. and Parr, R. G. (2000) Information theory, atoms in molecules, and molecular similarity. *Proc. Natn. Acad. Sci. USA*, **97**, 8879–8882.
Peng, J. and Müller, H.-G. (2008) Distance-based clustering of sparsely observed stochastic processes, with applications to online auctions. *Ann. Appl. Statist.*, **2**, 1056–1077.
Ramsay, J. O. (1998) Estimating smooth monotone functions. *J. R. Statist. Soc.* B, **60**, 365–375.
Ramsay, J. and Silverman, B. (2005) *Functional Data Analysis*, 2nd edn. New York: Springer.
Raubera, T., Braun, T. and Berns, K. (2008) Probabilistic distance measures of the dirichlet and beta distributions. *Pattn Recogn*, **41**, 637–645.
Reithinger, F., Jank, W., Tutz, G. and Shmueli, G. (2008) Modelling price paths in on-line auctions: smoothing sparse and unevenly sampled curves by using semiparametric mixed models. *Appl. Statist.*, **57**, 127–148.
Shmueli, G. and Jank, W. (2008) Modeling the dynamics of online auctions: a modern statistical approach. In *Economics, Information Systems & Ecommerce Research II: Advanced Empirical Methods* (eds R. Kauffman and P. Tallon). Armonk: Sharpe.
Shmueli, G., Russo, R. P. and Jank, W. (2007) The barista: a model for bid arrivals in online auctions. *Ann. Appl. Statist.*, **1**, 412–441.
Simonoff, J. (1996) *Smoothing Methods in Statistics*, 1st edn. New York: Springer.
Wang, S., Jank, W. and Shmueli, G. (2008a) Explaining and forecasting online auction prices and their dynamics using functional data analysis. *J. Bus. Econ. Statist.*, **26**, 144–160.
Wang, S., Jank, W., Shmueli, G. and Smith, P. (2008b) Modeling price dynamics in ebay auctions using principal differential analysis. *J. Am. Statist. Ass.*, **103**, 1100–1118.
Zhang, S., Jank, W. and Shmueli, G. (2009) Forecasting online auctions via functional k-nearest neighbors. *Int. J. Forecast.*, to be published. (Available from `http://ssrn.com/abstract=1432122`.)