Principal Differential Analysis of Online Auctions

Shanshan Wang

May 18, 2005

RIT Term Paper

1 Introduction

Online auctions display drastic variation in price dynamics both within and across auctions. It is not hard to see why because the auctions differ in all kinds of factors, such as the auctioned item, the opening bid, the seller, the participating bidders and the number of bidders, etc.. Any change in one of the above factors could cause huge variation in the dynamics of the price process of an ongoing auction.

Models for some of the aspects of auctions such as the bidding strategies, the price determinants, the bid arrivals, etc., have been researched so far. For instance, Bapna et al. (2004) develop a stable taxonomy of bidding behavior in online auctions to describe the significant heterogeneity existing in the users. Shmueli et al. (2004) use a class of 3-stage non-homogenenous Poisson processes to model the bid arrival process during auctions. Lucking-Reiley et al. (2000) investigate the determining factors of online auction price using static regression model, while Faraway (1997) and Shmueli and Jank (2004) perform regression analysis for a functional response. Jank and Shmueli (2003) observe that the dynamics of the price change sharply throughout the auction, especially towards the end of the auctions, and Shmueli and Jank (2004) find that different levels of opening bids are associated with different bid dynamics. Besides of online auction data, research of dynamics has also been done for other kinds of functional data. Ramsay et al. (1996) describe techniques that provide useful tools for describing multivariate functional data such as the measurement of speech movements, and functional ANOVA and functional principal components analysis (PCA) are used to reveal the important sources of variation in lip motion.

While work has been done to model the bidding process using functional regression, little effort has been tried to incorporate the bid dynamics into modelling. In particular, fully making use of the relationships among derivatives to model the price process has not been tackled yet.

One wide application of dynamics of functions is differential equation. A differential equation, referred to here as a DIFE, describes processes by finding relationships among the function and its derivatives. Ramsay (2000a) gives a brief introduction to the use of DIFE in statistics and several examples of functional estimation problems, such as simultaneous estimation of a regression model and residual density, monotone smoothing, specification of a link function, differential equation models of data, and smoothing over complicated multidimensional domains. In the FDA book written by Ramsay and Silverman (1997), three chapters are devoted to describe the methods and techniques of principal differential analysis. And Ramsay and Ramsey (2001) and Ramsay (2000b) give examples of applications of these methods and techniques to real data such as the monthly index of goods and a sample of handwriting in Chinese. More applications can be found in Ramsay and Silverman (2002).

This work is devoted to study the dynamics of eBay's online auctions data using principal differential analysis (PDA). The article is organized as follows. In Section 2, we give a brief overview of the data we use in this analysis and the preliminary data processing steps used to make the data prepared for analysis. In Section 3, we describe the differential equation model we use in details and introduce the techniques and explorative explanations related to PDA. Results of applying PDA to our data are given in Section 4. Section 5 describes an explorative study of effects of smoothing parameter and selection of knots on the estimated differential equation. And Section 6 talks about hypothesis test on the price dynamics of auctions in different groups via FITS and pointwise ANOVA. Finally, conclusion is given in Section 7 as well as discussions of open questions and future works.

2 The Data and Preliminary Analyses

Two data sets collected from eBay.com are used in this work for comparison. The first data are the bid histories of 149 closed 7-day auctions for Palm M515 personal digital assistant (PDA) and are so-called "Palm data". And the second data are bid histories of 1221 auctions for a variety of books and magazines. All auctions are chosen to be 7 days in order to be comparable with the results from the Palm data. We call the latter one the "Book data". All auctions in both data resulted in a sale. While Palm data include auctions on only one product, auctions in Book data vary from book to magazines that are from about 6 big categories, i.e., fiction books, nonfiction books, magazine issues, texbookds&education books, children's books, and antiquarian&collectible books. Apparently, the huge variability in the products auctioned in Book data will definitely cause large variability in dynamics of the price processes across auctions. Meanwhile, the variabilities in other characteristics such as the opening bid, the winning price and the seller's rating certainly will be largely increased due to the complexity of the Book data. We're pretty interested in how huge the difference is going to make and how the dynamics will be affected by existence of these variation.

As we know, the above data gathered from the web are in the form of the "proxy bids", the prices that bidders are willing to pay for the auctioned item rather than the real current prices shown over time on the auction page that we call "live bids". As the first step of the preliminary data processing, we transform these proxy bids into live bids using the bidding scheme applied on eBay. Based on the live bids, we create step functions of prices that can accurately reflect the real current price of each auctioned item at every single time point through the auction period. After applying some light smoothing on the data, we could be able to represent the data into a functional object. A package of functions for functional data analysis, called *FDA funs* and available in both S-PLUS/R and MATLAB versions by anonymous ftp from ego.psych.mcgill.ca/pub/ramsay/FDAfuns, made the analyses especially convenient. Smoothing splines and local polynomial estimates of derivatives can be computed using *FDA funs*.

Figure 1 and 2 display the smoothing splines and estimates of the first three derivatives of live bids in both Palm and Book data. It can be noted that dynamics variation exists within and across auctions for both data.

3 The Differential Equation Model

3.1 PDA Model

Let x_i be the observed price function for auction i, $i = 1, \ldots, N$. And let $D^m x_i$ be the m^{-th} derivative of function x_i. In this research, we consider the identification of a linear differential operator (LDO)

$$L = w_0 I + w_1 D + \cdots + w_{m-1} D^{m-1} + D^m \tag{1}$$

that comes as close as possible to satisfying the homogeneous linear differential equation $Lx_i = 0$ for each observation x_i. In another word, we wish the operator L to annihilate the given data functions x_i as nearly as possible. Thus, we seek a linear differential equation model so that our data satisfy

$$D^m x_i = -w_0 x_i - w_1 D x_i - \cdots - w_{m-1} D^{m-1} x_i \tag{2}$$

to the best possible degree of approximation. To carry out PDA, we adopt a least squares approach to the fitting of the differential equation model. The fitting criterion is to minimize the sum of squared norms

$$SSE_{PDA}(L) = \sum_{i=1}^{N} \int [Lx_i(t)]^2 dt \tag{3}$$

over L . Estimating L is equivalent to estimating the m weight functions w_i that define the LDO.

3.2 PDA by pointwise minimization

There are two techniques for estimating the weight functions w_j. The first approach yields a pointwise estimate of the weight functions w_j computable by standard least squares estimation. Define the pointwise fitting criterion

$$PSSE_L(t) = N^{-1} \sum_i (Lx_i)^2(t) = N^{-1} \sum_i [\sum_{j=0}^{m} w_j(t)(D^j x_i)(t)]^2 \tag{4}$$

where $w_m(t) = 1$ for all t.

Define the m-dimentsional coefficient vector as

$$w(t) = (w_0(t), \ldots, w_{m-1}(t))', \tag{5}$$

the $N \times m$ pointwise design matrix as

$$Z(t) = \{(D^j x_i)(t)\}_{i=1,\ldots,N; j=0,\ldots,m-1} \tag{6}$$

and the N-dimensional dependent variable vector as

$$y(t) = \{-(D^m x_i)(t)\}_{i=1,\ldots,N}. \tag{7}$$

Then the pointwise fitting criterion can be expressed in matrix terms as

$$PSSE_L(t) = N^{-1}[y(t) - Z(t)w(t)]'[y(t) - Z(t)w(t)]. \tag{8}$$

Holding t fixed, the least squares solution minimizing $PSSE_L(t)$ with respect to the values $w_j(t)$ is

$$w(t) = [Z(t)'Z(t)]^{-1}Z(t)'y(t). \tag{9}$$

3.3 PDA by basis expansions

The pointwise approach can pose problems in some applications. For instance, once the operator L has been computed by estimating its weight functioins w_j, we are interested in computing a set of m linearly independent basis functions u_j satisfying $Lu_j = 0$. Solving the equation $Lu = 0$ requires that the w_j be available at a fine level of detail, with the required resolution depending on their smoothness. For larger orders m, computing the functions w_j pointwise at a fine resolution level can be computationally intensive. This suggests the need for an approximate solution which can be computed quickly and which is reasonably smooth.

A strategy is to approximate the weight functions w_j by using a fixed set of basis functions $\phi_k, k = 1, \ldots, K$. Let ϕ denote the K-dimensional vector function $(\phi_1, \ldots, \phi_K)'$. We assume that

$$w_j \approx \sum_k c_{jk}\phi_k \tag{10}$$

where the mK coefficients c_{jk} define the approximations. Let the (mK)-vector c contain these coefficients, where index k varies inside index j.

We can approximate the criterion $SSE_{PDA}(L)$ in terms of c as a quadratic form $\hat{F}(c|x)$ that can be minimized by standard numerical algebraic techniques. We have

$$\hat{F}(c|x) = C + c'Rc + 2c's \tag{11}$$

where the constant C does not depend on c, and hence the estimate $\hat{c}$ is given by the solution of the equation $Rc = -s$.

The symmetric matrix R is of order mK, and consists of an $m \times m$ array of $K \times K$ submatrices R_{jk} of the form

$$R_{jk} = N^{-1} \int \phi(t)\phi(t)' \sum_i D^j x_i(t) D^k x_i(t) dt \tag{12}$$

for $0 \leq j, k \leq m-1$. The (mK)-vector s contains m subvectors s_j each of length K, defined as

$$s_j = N^{-1} \int \phi(t) \sum_i D^j x_i(t) D^m x_i(t) dt \tag{13}$$

for $j = 0, \ldots, m-1$.

The integrals involved in these expressions often have to be evaluated numerically such as using the trapezoidal rule over a fine mesh of equally-spaced values of t.

3.4 Assessing fit

Since the objective of PDA is to minimize the norm $||Ly||$ of the forcing function associated with an estimated differential operator, and since the quality of fit can vary over time, it seems appropriate to assess fit in terms of the poinwise error sum of squares $PSSE_L(t)$ as defined in Equation 4. As in linear modelling, the logical baseline against which we should compare $PSSE_L$ is the error sum of squares defined by a theoretical model and its associated weight functions w_j:

$$PSSE_0(t) = \sum_i [\sum_{j=0}^{m-1} w_j(t)(D^j y_i)(t) + (D^m y_i)(t)]^2. \tag{14}$$

Since there is no theoretical model at hand, we may use $w_j = 0$, so that the comparison is simply with the sum of squres of the $D^m y_i$. From these loss functions, we may examine the pointwise squared multiple correlation function

$$RSQ(t) = \frac{PSSE_0(t) - PSSE_L(t)}{PSSE_0(t)} \tag{15}$$

and the pointwise F-ratio

$$FRATIO(t) = \frac{(PSSE_0(t) - PSSE_L(t))/m}{PSSE_0(t)/(N-m)}. \tag{16}$$

3.5 Second-order linear differential equation used in this work

We use B-splines of order 6 to smooth our data, this allows us to have smooth derivatives of up to order 3. So we consider exploring to what extent a second-order linear differential equation can express the price dynamics in our work

$$Lx_i = w_0 x_i + w_1 D x_i + D^2 x_i = 0. \tag{17}$$

Discussions of second-order mechanical systems can be found in most applied texts on ordinary equations and can easily fit to our online auction frame.

The first coefficients w_0 essentially reflects the position-dependent force applied to the system at position x. The second coefficient w_1 indicates influences on the system that are proportional to velocity rather than position and are often internal or external frictional forces or viscosity in mechanical systems. Here we connect it with the influences on the price process when there is change in velocity due to bids placed. The different shapes of weight functions implies different price dynamics.

Any order of linear differential equation can be rewritten into a system of first order equations by substitution. The second-order linear differential equation in 17 can be rewritten into a system of two first order equations by substituting

$$y_1 = x, y_2 = Dx. \tag{18}$$

The matrix form of the second-order linear differential equation is give as

$$\dot{Y} = AY \tag{19}$$

where $\dot{Y} = \begin{bmatrix} \dot{y}_1(t) \\ \dot{y}_2(t) \end{bmatrix}$, $A = \begin{bmatrix} 0 & 1 \\ -w_0 & -w_1 \end{bmatrix}$, $Y = \begin{bmatrix} y_1(t) \\ y_2(t) \end{bmatrix}$.

For linear dynamics system, the matrix A is a crucial tool for analyzing the stability of the system. The stability analysis can be achieved only through looking at the *eigenvalues* of A . The real parts of the eigenvalues determine the stability of the system. The system is stable if all of them are negative, unstable if any one of them is positive, and in equilibrium (or oscillates) if they are all zeros. The imaginary parts of the eigenvalues determine the frequencies of corresponding modes (state variables). In the context of our bidding data, instability would mean exponential growth in the amount of bid price, while stability would imply zero or low acceleration in the price growth, in another word, the price process is in a linear motion. Capability of distinguishing the stable and unstable periods through an auction helps us with distinguishing different dynamics within and across auctions.

4 Application

The techniques described in Section 3 are firstly applied to the Palm data. The weight functions w_0 and w_1 are estimated using the "*pda.fd*" function in R and displayed in the upper panel of Figure 3. We notice that there is not much variability in both functions about the average value that is close to zero. But one does note that w_0 is negative at the beginning of the auction and between day 4 and day 6, and w_1 is negative in the event that $w_0 > 0$ during the auction's closing period. During these periods, the system exhibits instability or exponential growth due to the phenomena of "early bidding" and "bid snipping". While during the middle of the auction, both weight functions are pretty close to zero, indicating that the process is in linear motion, corresponding to $D^2 = 0$ as we can see in Figure 1.

We calculate the eigenvalues of the A matrices for the second-order differential equation on a fine grid. Thus we get an eigenvalue curve for each mode. The upper panel of Figure 4 displays such curves. The black dotted line is the eigenvalue curve for the price (λ_1), and the red dashed line is that for the price velocity(λ_2). λ_1 is positive for the first day and much more positive starting day 4. This indicates that the price system is unstable during the beginning and closing periods of the auction, exhibiting exponential growths over time. This exponential growth is even more intense towards the end of the auction. Between day 1 and day 4, both eigenvalues are negative, implying that the price system is relatively stable during this period, namely, the auction price displays a linear growth trend. The above findings in fact coincidences with what we observe for the changing dynamics through the entire auction.

We compare the price accelerations and the estimated forcing functions of the second-order differential equation applied to Palm data as displayed in the upper panel of Figure 5. It can be noted that the magnitude of the forcing functions are comparable with the accelerations, implying a good approximation of the linear differential equation.

The general PDA procedure of second-order differential equation is applied to the Book data: the weight functions are estimated; the bootstrapped 95% confidence intervals are used to measure whether they are zeros; the eigencurves of the A matrices are obtained to analyze the stability of the price system; the forcing functions are calculated and compared with the accelerations; etc.. And all the above results are compared with those of the Palm data. It is noted that the estimated weight functions look fairly similar, in particular for w_0. The average of w_0 of Book data is lower than that of Palm data by 0.02, though. w_1 decreases as well, while it starts positive other than negative for that of the Palm data. And w_1 for the book data is bumpier than that of the palm data even with a light smoothing step applied. Consequently, the first eigencurve turns out to be not as smooth as that of the palm data. Even though, the eigencurves tell the same story as what has been revealed by those of the palm data: the bidding system exhibits unstability at the beginning and towards the end, and is relatively stable during the middle of the auctions. This conclusion is consistent with finding of the changing dynamics of auctions. There are three main differences exist between those of the book data and palm data. First, which has been addressed already, the first eigencurve for book data is more wiggly; and secondly, for book data, the stable period starts in late day 2 and ends in late day 3, which is shorter than the stable period for palm data; third, during the unstable periods (the beginning and closing periods), the bidding system of book is much more unstable than that of palm. And this is reflected by the magnitude of the positive eigenvalues. The reason is probably due to the fact that book data consists more auctions and these auctions differed greatly in the category, the opening bid and the eventual selling price, which greatly increases the instability of the price system. The palm data only consists of 149 auctions and all are on the same exact product. This greatly reduces the variability in price dynamics.

Since we have seen huge variation in the price dynamics in the Book data induced by the complexity of the data, we apply PDA to different groups of the data. Several variables such as the category, the seller's rating and the opening bid are chosen to apply PDA to see if there exists any difference across different levels of these variables. It would be interesting to see if auctions have different price dynamics across currencies. But since the bid time is temporarily missing from the complete data set which includes auctions in both US currency and non-US currencies, we'll put this spot for future work. There are six categories in the Book data: (1) Fiction Books; (2) Nonfiction Books; (3) Magazine Issues; (4) Textbooks, Education; (5) Children's Books; (6) Antiquarian&Collectible Books. For the variable opening bid, we chose 5 dollar and 20 dollar as the cut-offs for the low opening bid, mid opening bid, and high opening bid. The variable seller's rating is chosen to have two levels, the low and the high, and the cut-off is 1000. Looking at the estimated weight functions of 2nd order differential equations corresponding to different groups partitioned by different variables, we found that the basis shapes of w_0 and w_1 are similar for the 6 categories.

Though, they differ in the magnitudes of these functions. For instance, the magnitudes of w_0 are relatively small for category 4 and 6, while the magnitudes of w_1 are relatively small for category 5 and 6. Apparently, the magnitude of w_1 matters more to the price dynamics since it implies the magnitude of external force applied to the system. The relatively small magnitudes of category 5 and 6 indicate to us that books such as the children's books and antiquarian type books are not prone to cause growth in price as other books are. Seller's rating does not seem to affect the price dynamics much since the weight functions estimated for auctions with low seller's rating and those with high seller's rating look exactly the same both in shapes and magnitudes. The amplitude of w_0 shrinks as opening bid increases, implying impact of opening bid on the dynamic system. While w_1 remains the same as opening bid grows, the influence of w_0 on the entire system is very hard to tell. In summary, comparing the shapes and the magnitudes of the weight functions among groups, we do have the impression that these variables impact the price dynamics more or less. We may hypothesize that there exist different price dynamics for different groups in the Book data. And to test this hypothesis would be the task of Section 5.

5 Explorative study of effects of smoothing parameter and set of knots

In previous work, we have seen that the estimated weight functions for the 7-day auctions in book data possess shapes that are similar with those for the Palm 7-day auctions. However, they're much more wiggly than the latter. This is probably caused by the fact that the book data contains all kinds of products, which increases the variability across auctions. These zig-zags should be smoothed out by adjusting the smoothing parameter λ and the set of knots. To see which one affects the smoothness more, the smoothing parameter or the set of knots, this work is performed in two ways: first, fix the smoothing parameter and vary the set of knots; second, fix the set of knots and let the smoothing parameter change. And there are two findings: (1) reducing the number of knots can greatly increase the smoothness of the weight functions (in particular, for w_1), while increasing the smoothing parameter even to a large scale won't help much; (2) When the number of knots is reduced, change of the smoothing parameter won't affect the general smoothness of the weight functions, but it does affect the scale of the weight functions significantly at the auction start.

We basically compare two sets of knots, one contains 14 knots $(0, 1, 2, 3, 4, 5, 6, 6.25, 6.5, 6.75, 6.8125, 6.875, 6.9375, 7)$ and another on contains 11 knots $(0, 1, 2, 3, 4, 5, 6, 6.25, 6.5, 6.75, 7)$. And when the knots are fixed to each set, let the smoothing parameter λ vary from 0.1 to 50. It can be observed from the following graphs that when the knots are reduced from 14 to 11, the weight functions, in particular w_1, are greatly smoothed out. While the knots are fixed to 11, with other parts maintain the same level, the weight functions at the auction start display huge variability corresponding to the change of the smoothing parameter. For instance, w_0 increases very fast from negative to positive when lambda increased from 0.1 to 1.6, then drops and levels off after that. It changes the sign again around $\lambda = 4$. Similarly, w_1 increases from negative to positive when lambda changes in the small value range, while it keeps increasing when lambda gets larger and larger. The increasing tendency is especially tense for small λ, and then it levels off for larger lambda. The magnitudes of the weight functions for 14 knots are much more stable. They remain the same except for extremely small λ values like 0.1, 0.2 and 0.3. From these results, we may conclude that when large number of knots are used, the weight functions are less smooth and increasing the smoothing parameter won't improve much. However, reducing the number of knots can greatly improve the smoothness of the weight functions, while their values at the auctions start get more sensitive to the value change of the smoothing parameter λ.

6 Hypothesis Testing on the Bidding Dynamics of Auctions in Different Groups

6.1 Hypothesis Test via FITS method

We performed PDA's on auctions from different groups partitioned by different variables such as the category, the open bid and the seller's rating. It was observed that the estimated weight functions have different behavior, in particular for w_0 of groups by category and groups by open bid, indicating to us that there might exists different dynamics in different groups. We need a way to test this null hypothesis: auctions in different groups have the same dynamics (or: w_0 and w_1 are the same for all groups). In the work of James and Sood (2004), the authors propose three different approaches for performing hypothesis tests on the shape of a mean function. The procedures work by testing for patterns in the residuals after estimating the mean function. Instead of working on the mean function, we're working with the weight functions, a small modification is made: we test for patterns in the differences between overall weight functions and group weight functions. Under the null hypothesis, the permuted differences should have a similar distribution to the unpermuted ones. The algorithm is as follows: (1) Compute the least square estimates of the overall weight functions and of the weight functions for each group.(2) Compute the differences $d_{kj}^0 = W_{0j} - W_{0kj}$,$d_{kj}^1 = W_{1j} - W1kj$, $k = 1, \ldots, K$, $j = 1, \ldots, J$. Where K is the number of groups, and J is the number of grid points. (3) Fit $s(t) = b(t)^T \theta$to the differences via least squares $\hat{\theta} = (X^T X)^{-1} X^T d$ where X is a basis matrix with rows corresponding to $b(t_{ij})^T$. $s(t)$ is a q-dim function. (4) Compute the test statistic $T = min(abs(\sum_{l=1}^{q} \hat{\theta}_l^2))$. (5) randomize the differences and refit $s(t)$ to the permuted differences to obtain $\hat{\theta}^{(b)}$. (6) Compute $T^{(b)} = min(abs(\sum_{l=1}^{q} \hat{\theta}^{(b)^2}))$ B times. (7) Repeat Steps (5) and (6) to obtain $T^{(1)}, \ldots, T^{(B)}$. (8) The estimated p-value corresponds to $\frac{1}{B} \sum_{b=1}^{B} I(T \leq T^{(b)})$. We perform the above steps to test the dynamics of groups by the category, the open bid and the seller's rating, respectively. Figure 12, 13 and 14 show the results of the test. Each graph plots the values of $T^{(b)}$ generated from 200 random permutations of the differences along with a red horizontal line which corresponds to the observed value of T . As we hope, Ts are larger than most of 200 $T^{(b)}s$ for w_0 with very small p-values, implying rejection of the null hypothesis. for w_1 are all near the top of the bulk of $T^{(b)}s$ with p-values less than 30%. One confusing thing is that the further away the weight functions seem to be, the higher the p-value. And vise vera. For instance, the differences between w_1 for groups partitioned by categories are the biggest telling from the graph, while the hypothesis test obtains a p-value of 29.5% which is higher than that of other two categorizations.

The FITS algorithm is based on the assumption that the differences are *iid* distributed. It does not account for the correlation in the differences that may exist. Hence, the permuted and unpermuted differences may have slightly different distributions even if the null hypothesis is correct. To adjust for this effect, we make a simple addition to the algorithm as described by James and Sood (2004). Figure 15, 16 and 17 display these adjusted statistics for different categorizations. From the graphs we see that the p-values are increased to around 30% and the confliction between similarity and p-value we show before is adjusted off.

6.2 Hypothesis Test via Pointwise ANOVA

We can also use pointwise ANOVA F-test to test the null hypothesis. In Equation 4, we use $PSSE_L(t)$ to represent the pointwise error sum of squares for the entire auction collection. Assuming that we have K groups, we use $PSSE_{L\_k}$ to represent the pointwise error sum of squares for each group, $k = 1, \ldots, K$. The pointwise sum of squares can then be written in the form

$$SSB(t) = \sum_{k=1}^{K} PSSE_{L\_k}(t) - PSSE_{L\_i}, \tag{20}$$

$$SSE(t) = PSSE_0(t) - \sum_{k=1}^{K} PSSE_{L\_k}(t). \tag{21}$$

Hence, An pointwise ANOVA F-statistic for testing the null hypothesis is given by

$$F(t) = \frac{SSB(t)/(K-1)}{SSE(t)/(N-K)}. \tag{22}$$

Figure 18, 19 and 20 display the F-statistics for testing dynamics by different groups. The .05-level cutoff is marked by a green dotted horizontal line. It is seen that most null hypothesis tests are rejected in pointwise tests at the .05-level except for the last 10 for category and the last 5 for openbid.

7 Conclusion

In this work, we apply principal differential analysis on two data sets from eBay: Palm data and Book data. We use a second-order linear differential equation to fit the data. Weight functions in the differential equation are estimated and analyzed for the empirical meanings under the online auction frame. We found that the changes in values and shapes of the weight functions can be easily related to the varying price dynamics through auctions, in particular, is consistent with the phenomena of "early bidding" and "bid sniping" we observe for online auctions. We compare the results between Palm data and Book data, and Book data show more variation in the price dynamics due to its complexity data structure. To better understand the complicated price dynamics in Book data, we pick some significant variables like the category, the opening bid and the seller's rating to partition the data into different types of groups and apply PDA respectively to each group. It is quite interesting that the price dynamics for different groups do display dissimilarities. To show these dissimilarities are really significant, we try two different methods to test the hypothesis that the price dynamics are the same for different groups: FITS method and pointwise ANOVA F-test. Most hypothesis tests are rejected for both method, indicating to us that there do exist different dynamics for different groups. The more complex the data, the more price dynamics will show up there.

Yet there's still questions left open. For instance, are those tests robust? Are there any other tests that yield more trustful results? What can we do about those different price dynamics? etc.. We'll take our effort to tackle them in future work.

References

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

Faraway, J. J. (1997). Regression analysis for a functional response. *Technometrics*, 39:254–261.

James, G. and Sood, A. (2004). Performing hypothesis tests on the shape of functional data (under review).

Jank, W. and Shmueli, G. (2003). Dynamic profiling of online auctions using curve clustering. Working paper, `"http://www.smith.umd.edu/faculty/wjank/AuctionProfiling.pdf"`.

Lucking-Reiley, D., Bryan, D., Prasad, N., and Reeves, D. (2000). Pennies from ebay: The determinants of price in online auctions. Technical report, University of Arizona, `"http://www.vanderbilt.edu/econ/reiley/papers/PenniesFromEBay.pdf"`.

Ramsay, J. O. (2000a). Differential equation models for statistical functions. *Canadian Journal of Statistics*, 28:225–240.

Ramsay, J. O. (2000b). Function components of variation in handwriting. *Journal of the American Statistical Association*, 95:9–15.

Ramsay, J. O., Munhall, K., Gracco, V., and Ostry, D. (1996). Functional data analysis of lip motion. *Journal of Acoustical Society of America*, 99:3718–3727.

Ramsay, J. O. and Ramsey, J. B. (2001). Functional data analysis of the dynamics of the monthly index of non durable goods production. *Journal of Econometrics*, 107:327–344.

Ramsay, J. O. and Silverman, B. W. (1997). *Functional data analysis.* Springer-Verlag New York, Inc., 1st edition.

Ramsay, J. O. and Silverman, B. W. (2002). *Applied functional data analysis: methods and case studies.* Springer-Verlag New York, Inc., 1st edition.

Shmueli, G. and Jank, W. (2004). Modelling the dynamics of online auctions: A modern statistical approach. *New Empirical Methods in eCommerce (forthcoming), Editor: Kauffman.*

Shmueli, G., Russo, R. P., and Jank, W. (2004). Modelling bid arrivals in online auctions. Working paper, `"http://www.smith.umd.edu/faculty/wjank/BidArrivals.pdf"`.

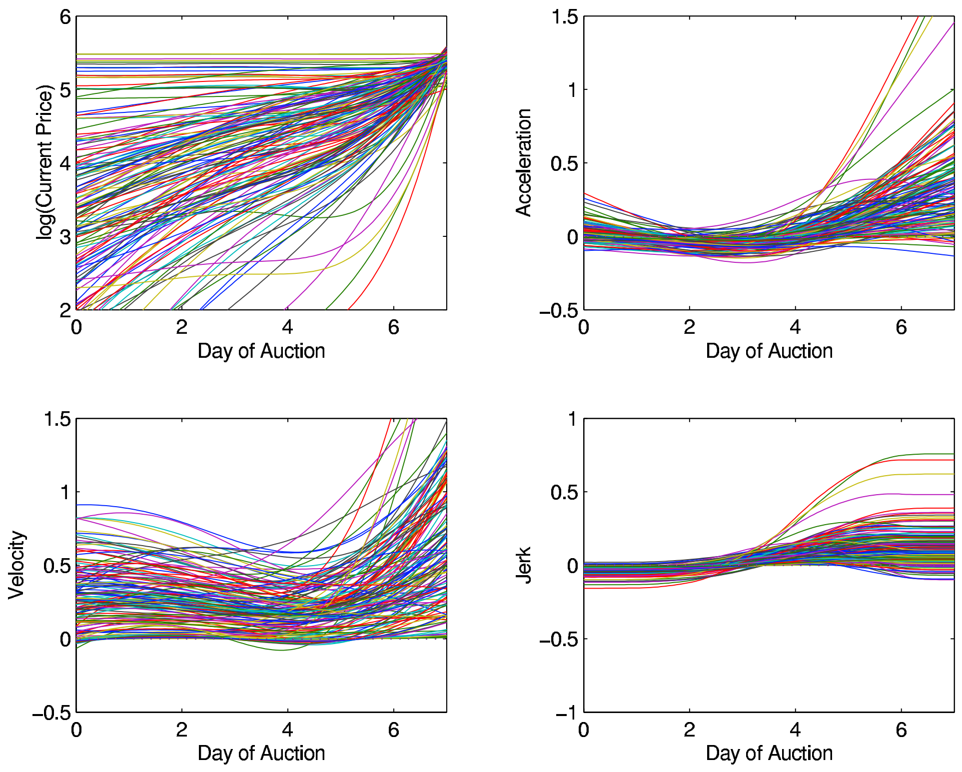

Figure 1: *Smoothed curves of 149 Palm M515 7-day auctions and their estimated first three derivatives.*

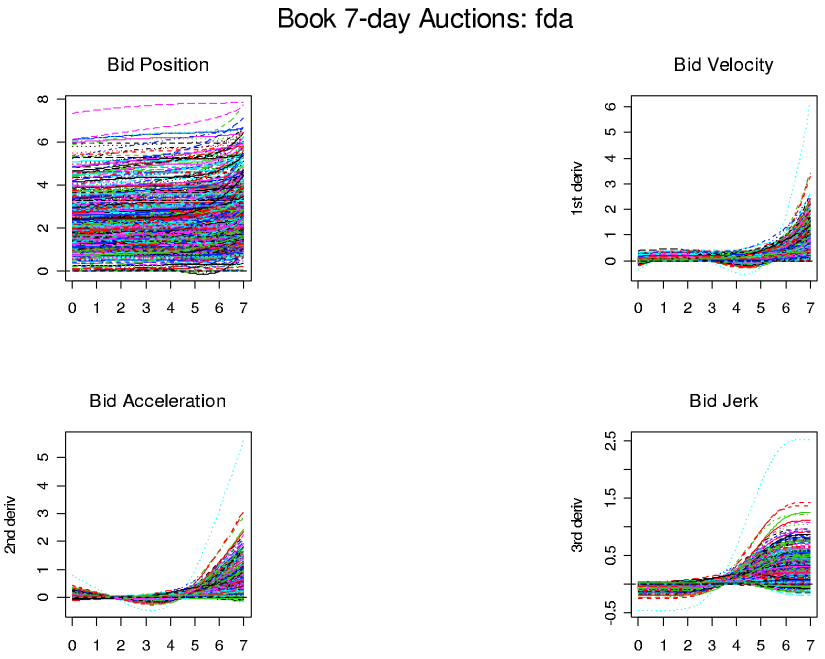

Figure 2: *Smoothed curves of 1221 Book 7-day auctions and their estimated first three derivatives.*

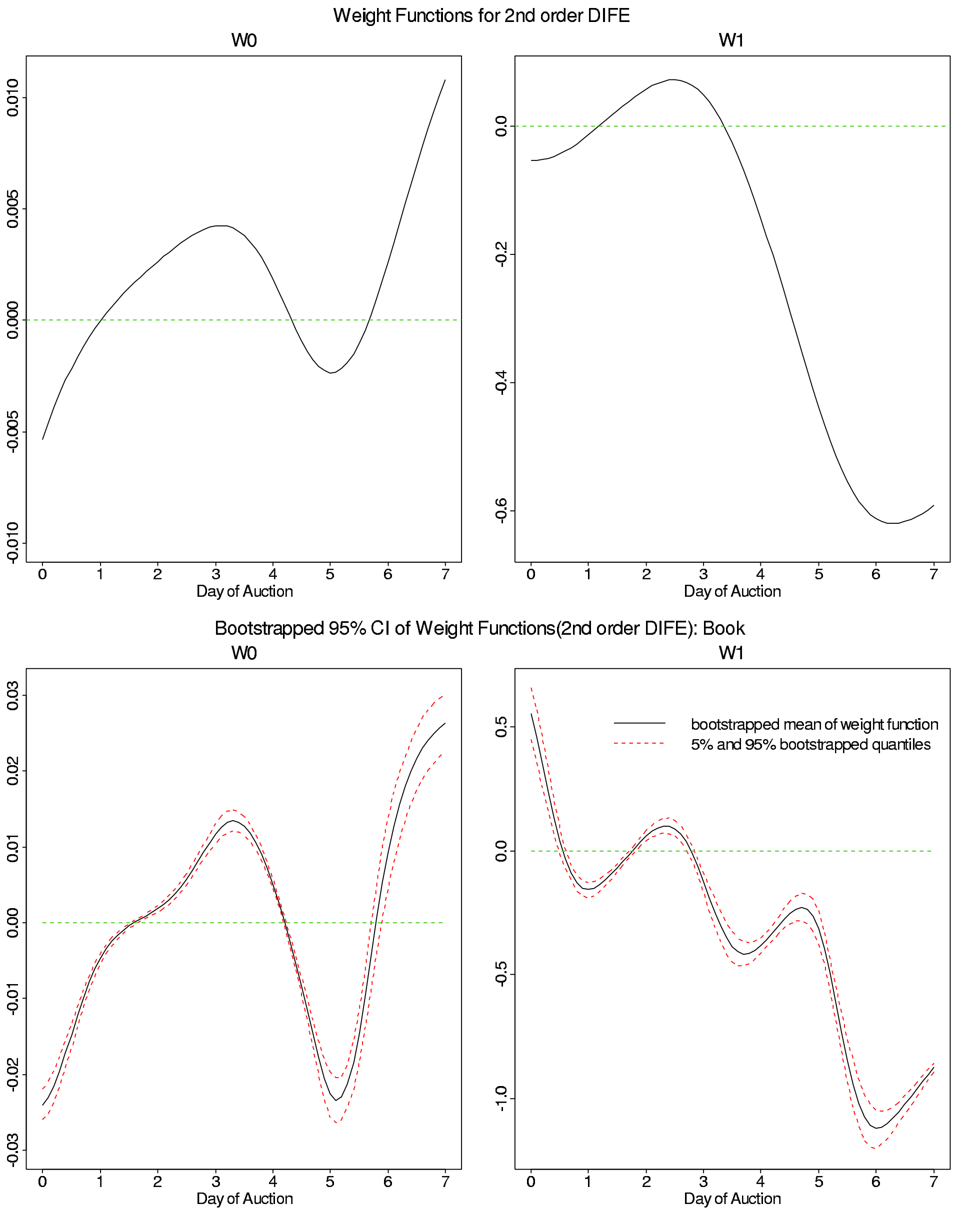

Figure 3: *Estimated weight functions of 2nd order differential equations fitted to online auction data. Upper panel: Palm 149 7-day auctions; lower panel: Book 1221 7-day auctions.*

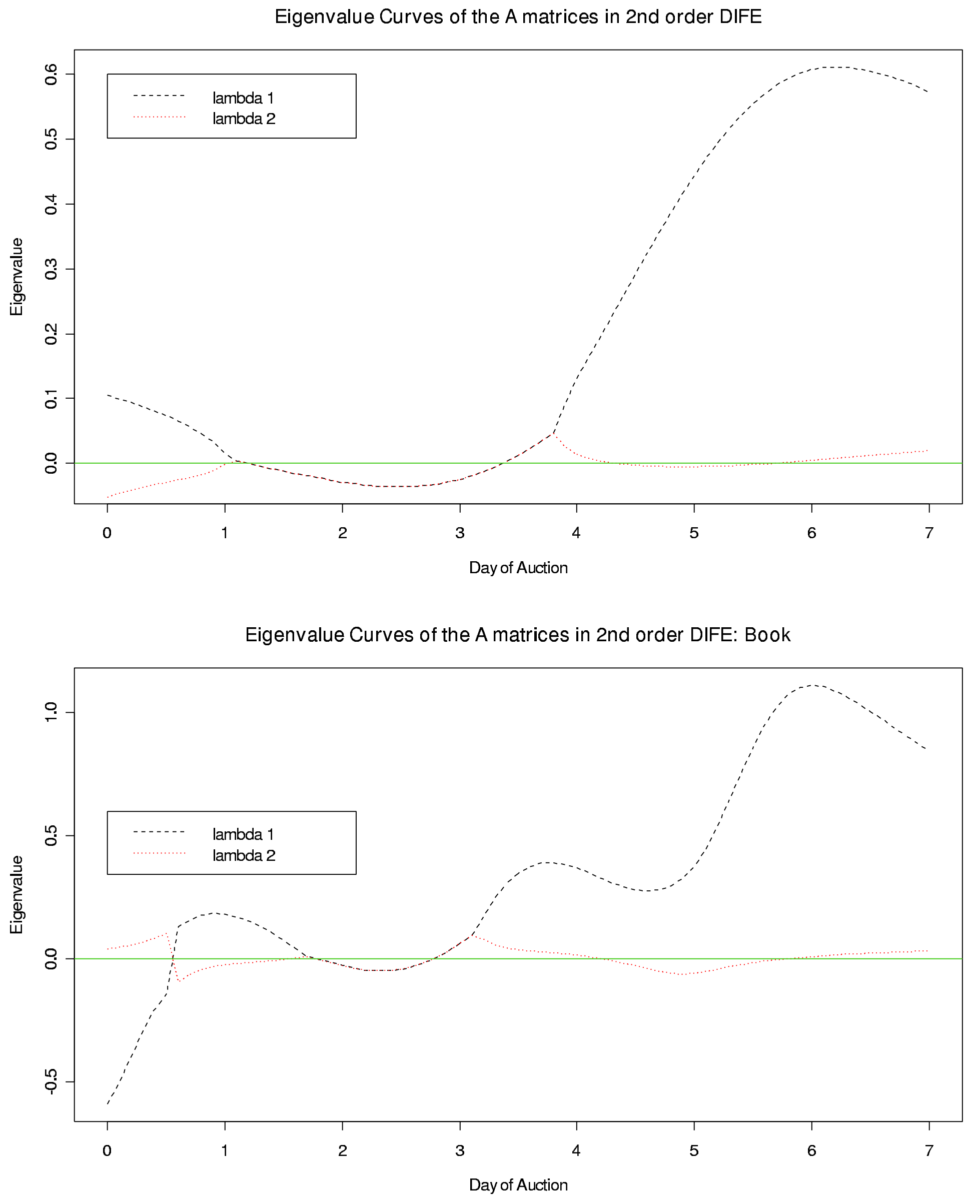

Figure 4: *Comparison of the eigencurves for online auction data. Upper panel: Palm 149 7-day auctions; lower panel: Book 1221 7-day auctions.*

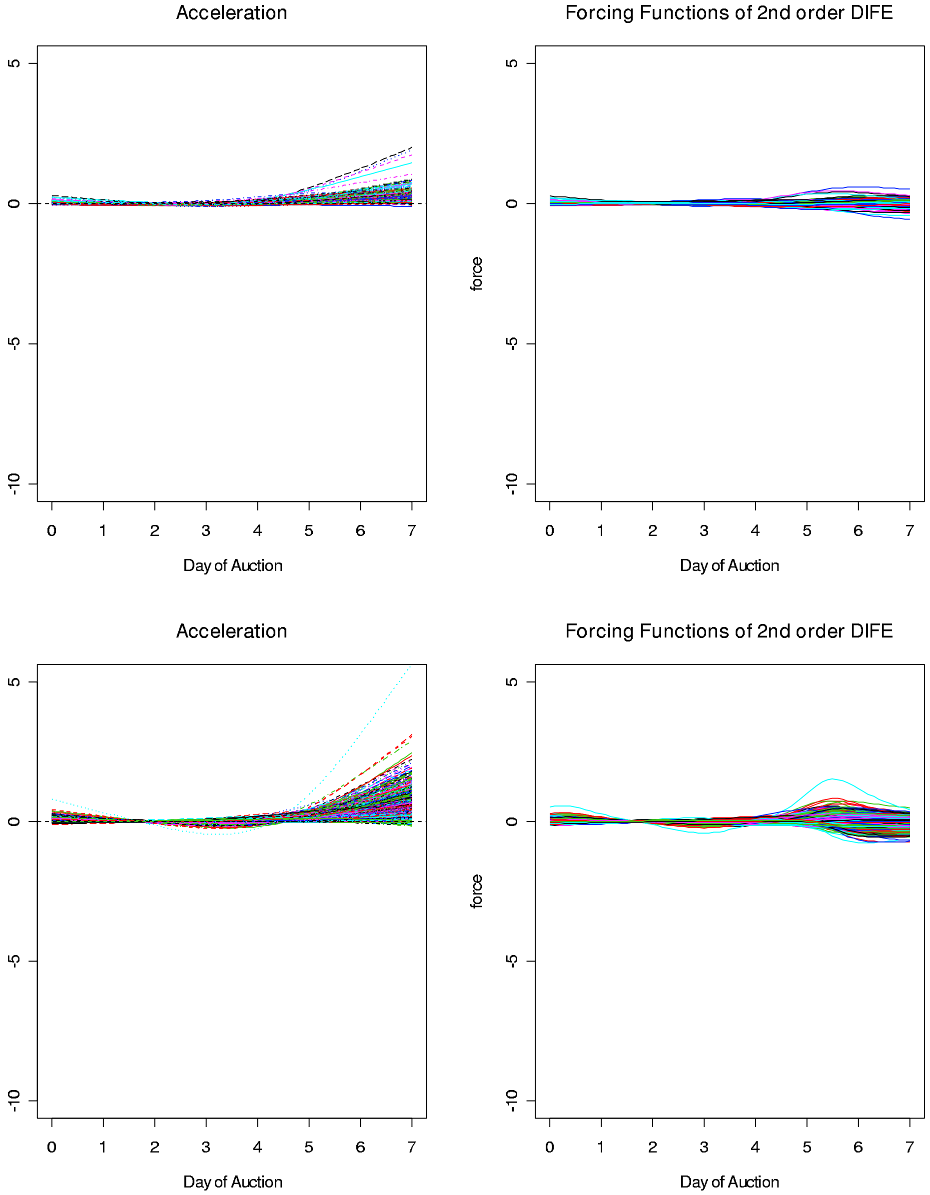

Figure 5: *Comparison between the accelerations (2nd derivatives) and the observed forcing functions for 2nd order differential equation of online auction data. Upper panel: Palm 149 7-day auctions; lower panel: Book 1221 7-day auctions.*

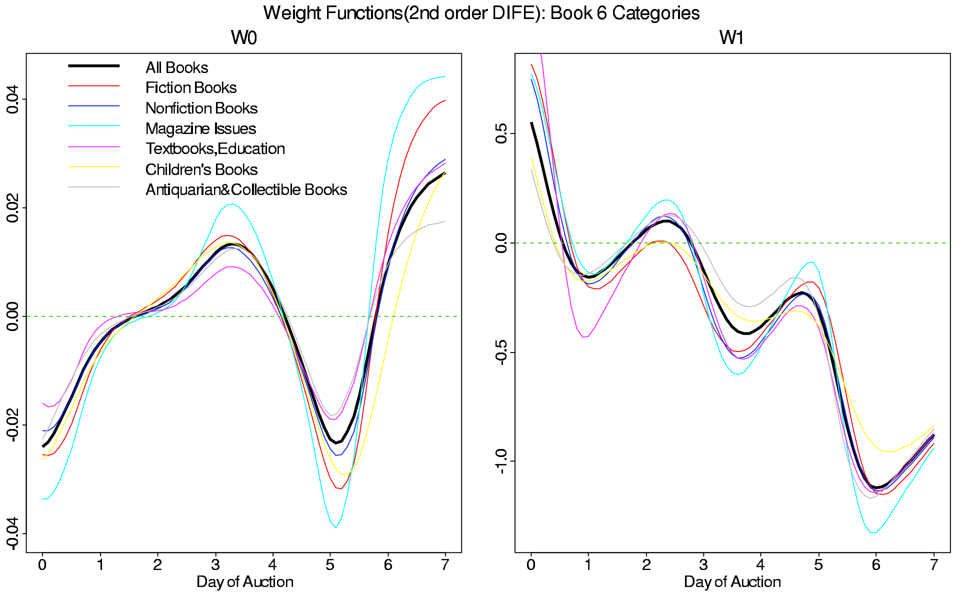

Figure 6: *Estimated weight functions of 2nd order differential equation fitted to auctions from different categories in Book 1221 7-day auction data.*

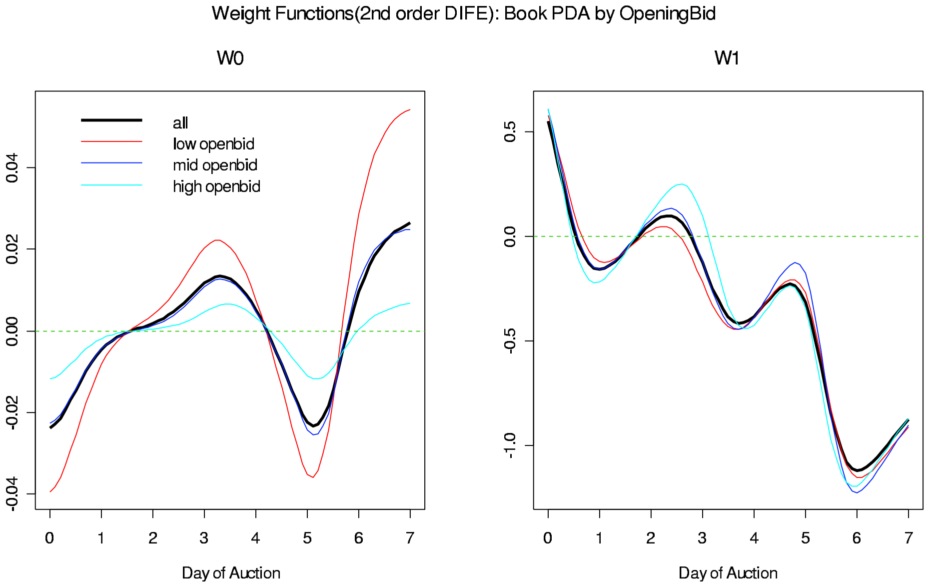

Figure 7: *Estimated weight functions of 2nd order differential equation fitted to auctions with different levels of opening bid in Book 1221 7-day auction data.*

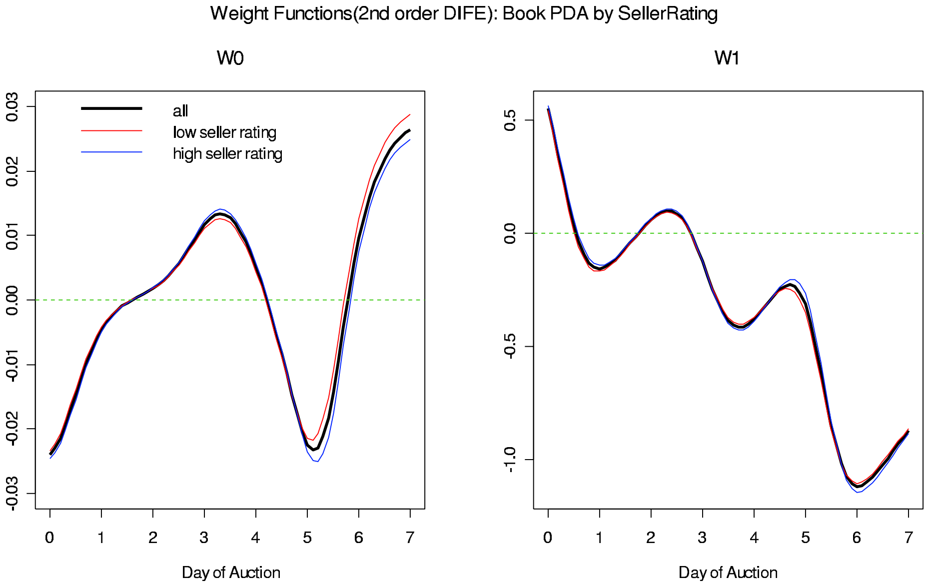

Figure 8: *Estimated weight functions of 2nd order differential equation fitted to auctions from different levels of seller's rating in Book 1221 7-day auction data.*

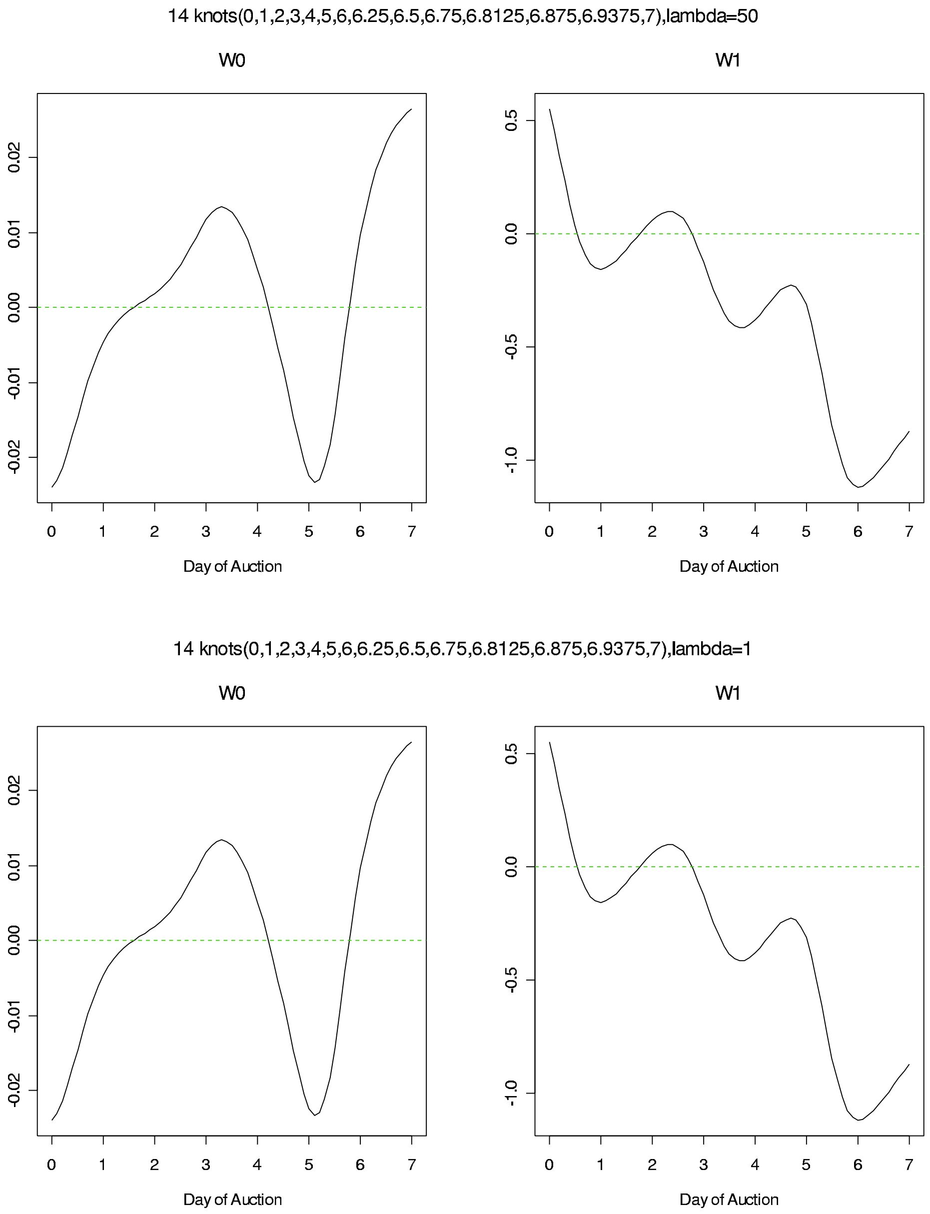

Figure 9: *Comparison of smoothness for different lambda when number of knots = 14 for Book data.*

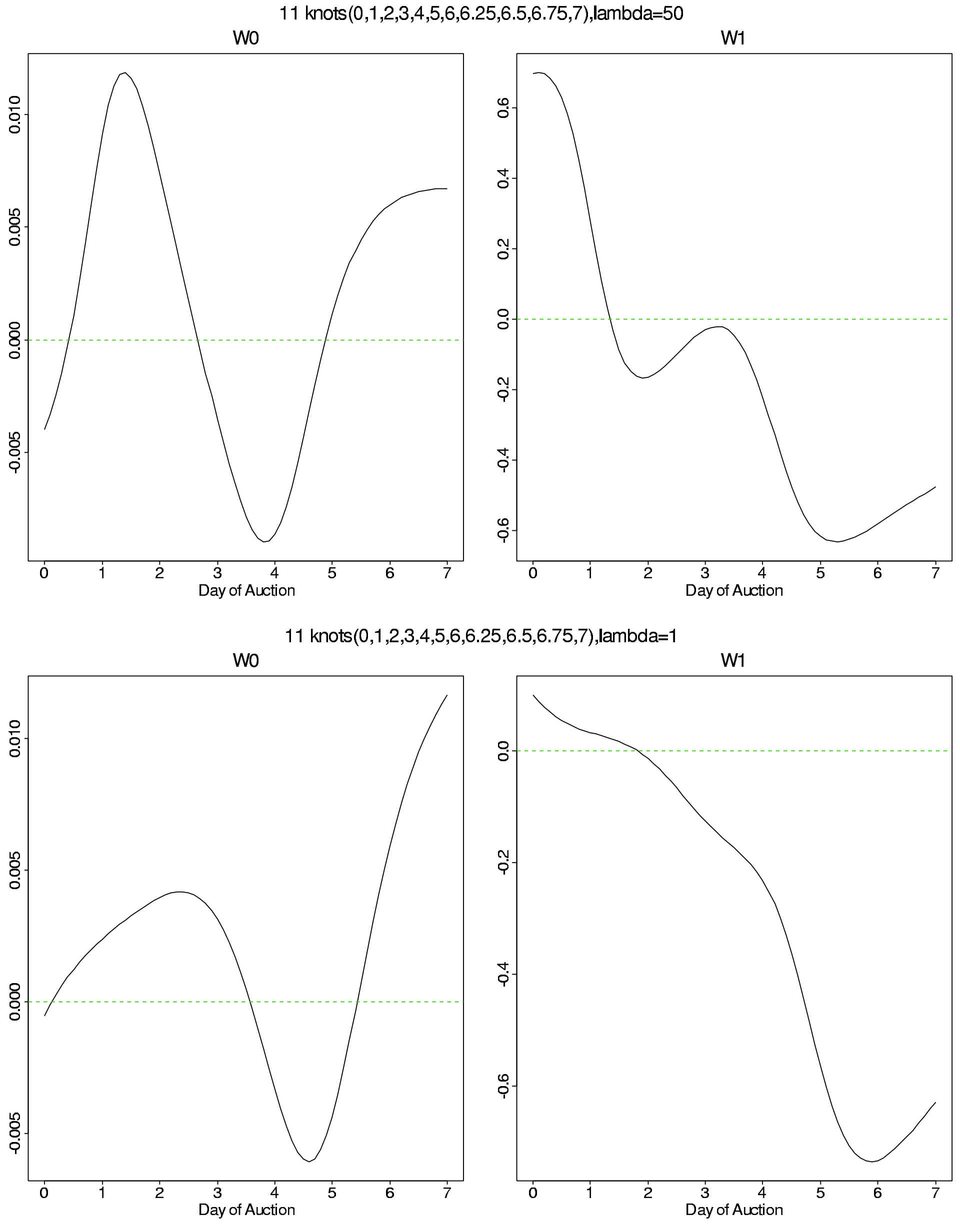

Figure 10: *Comparison of smoothness for different lambda when number of knots = 11 for Book data.*

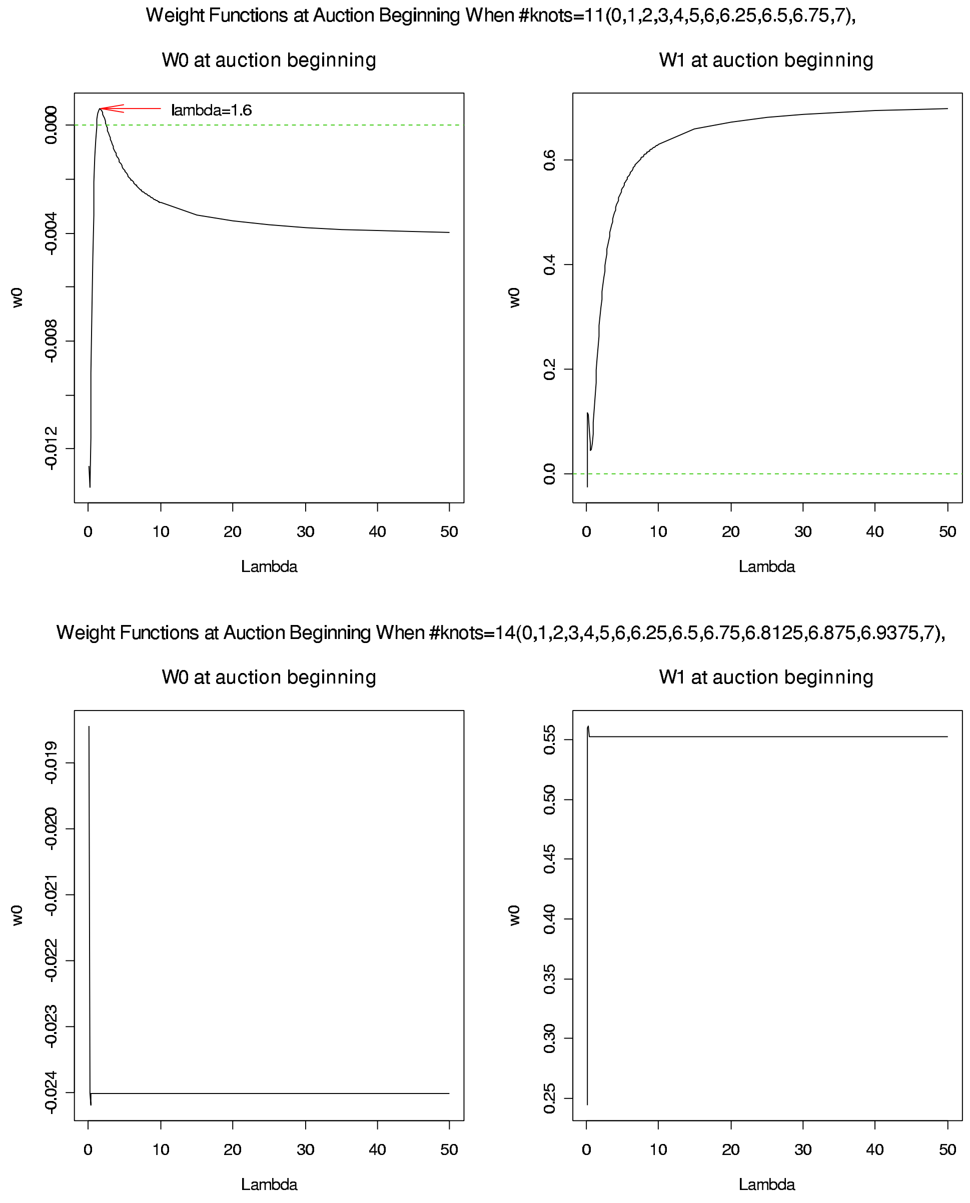

Figure 11: *Behaviors of weight functions at auction start for different sets of knots for Book data.*

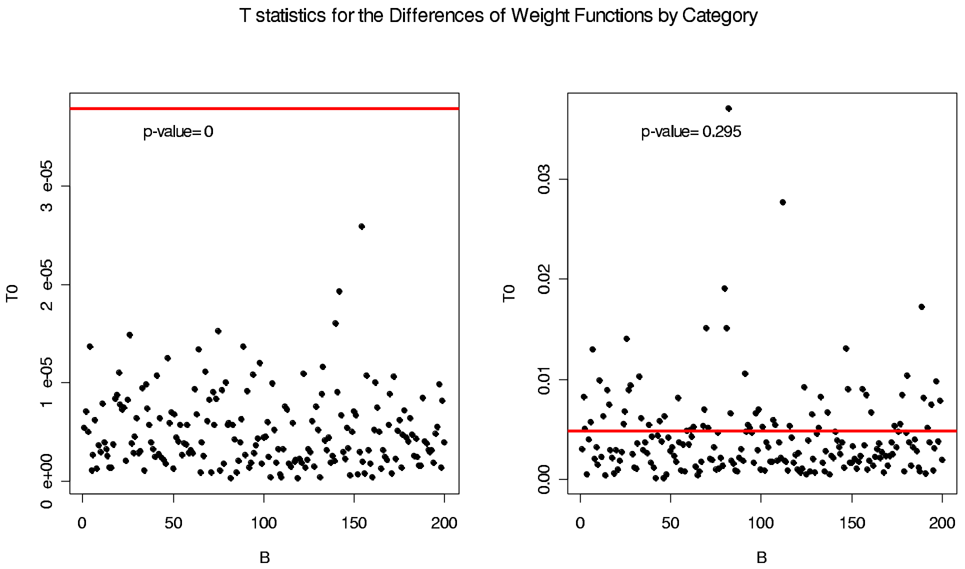
T statistics for the Differences of Weight Functions by Category

Figure 12: *Hypothesis test using T statistics from FITS for the differences of weight functions estimated from auctions from different categories for Book data.*

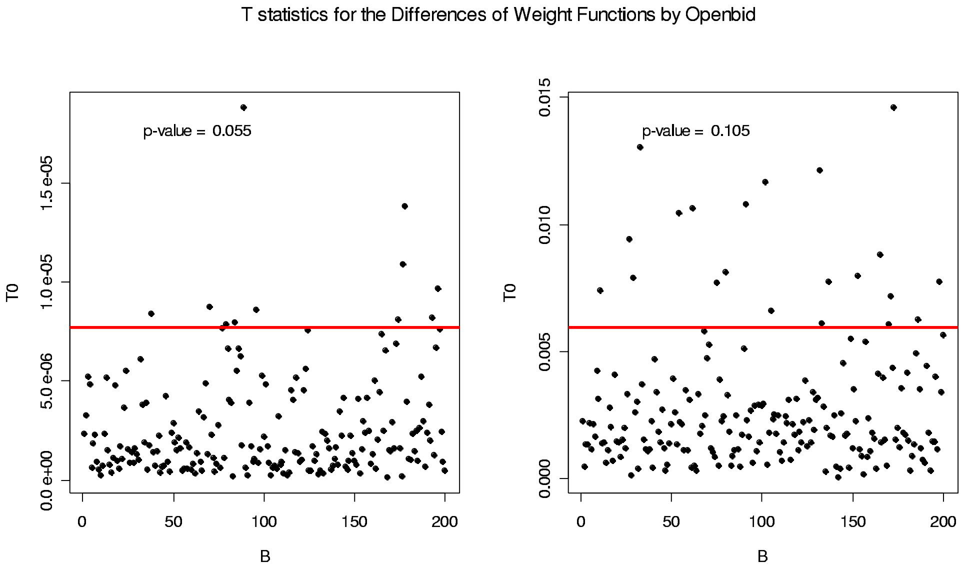
T statistics for the Differences of Weight Functions by Openbid

Figure 13: *Hypothesis test using T statistics from FITS for the differences of weight functions estimated from auctions from different levels of opening bid for Book data.*

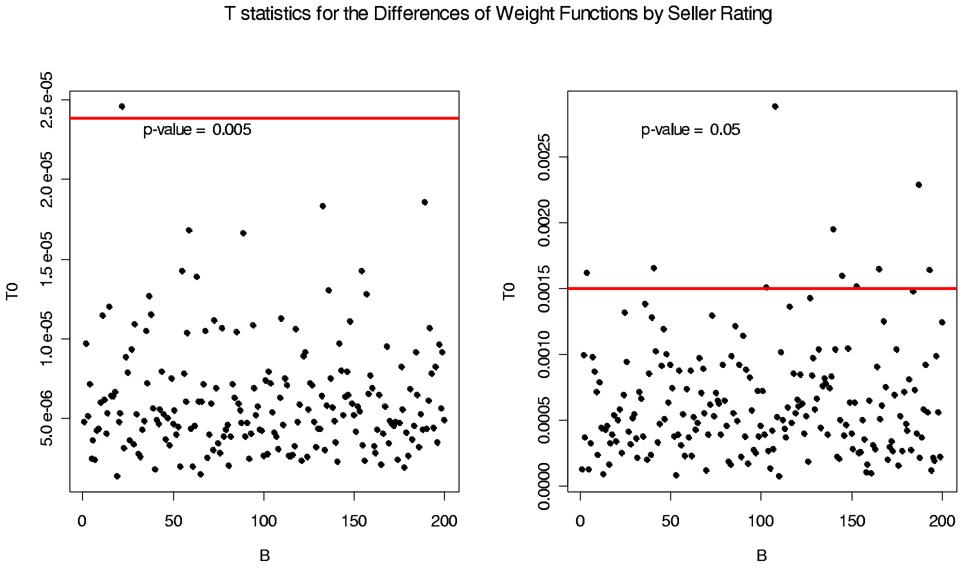

Figure 14: *Hypothesis test using T statistics from FITS for the differences of weight functions estimated from auctions from different levels of seller's rating for Book data.*

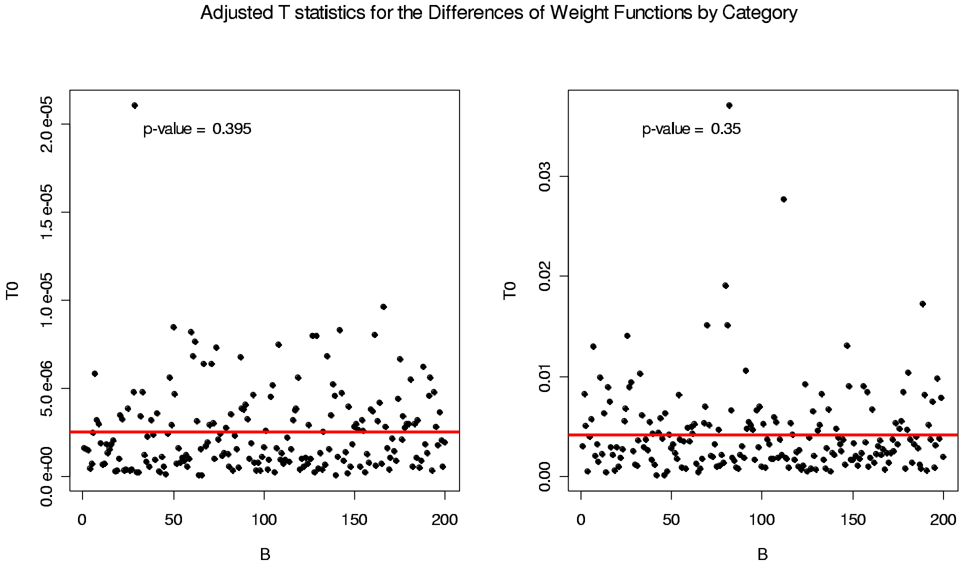

Figure 15: *Hypothesis test using adjusted T statistics from FITS for the differences of weight functions estimated from auctions from different categories for Book data.*

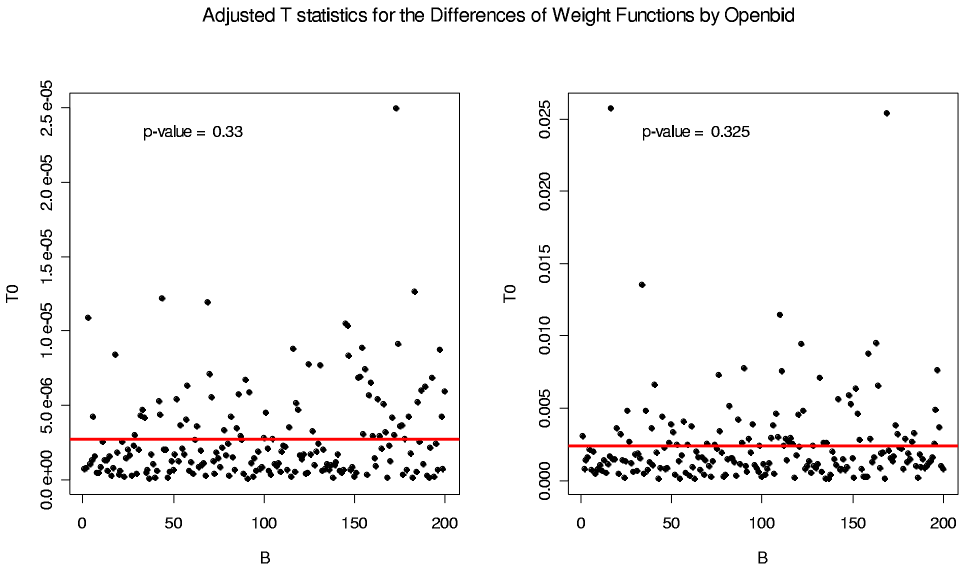

Figure 16: *Hypothesis test using adjusted T statistics from FITS for the differences of weight functions estimated from auctions from different levels of opening bid for Book data.*

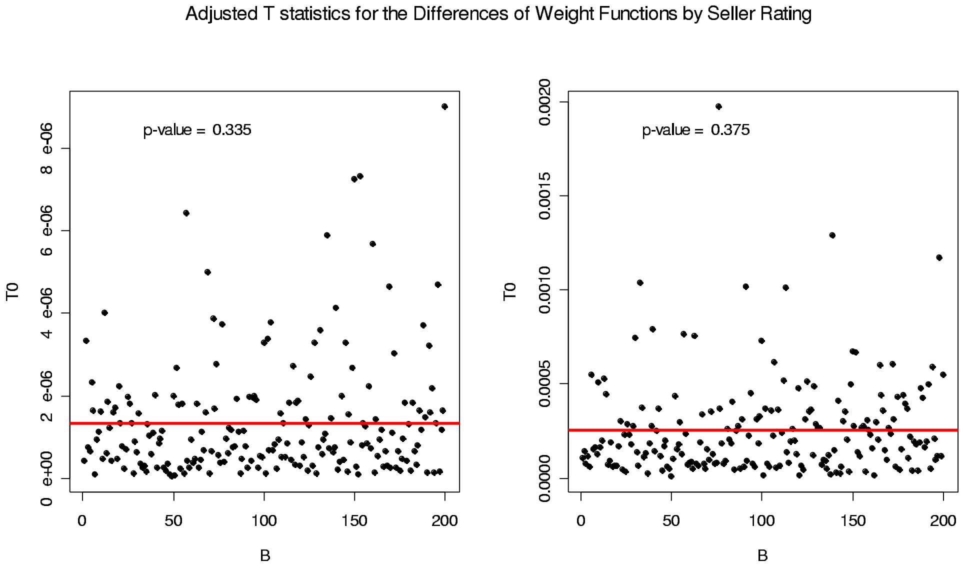

Figure 17: *Hypothesis test using adjusted T statistics from FITS for the differences of weight functions estimated from auctions from different levels of seller's rating for Book data.*

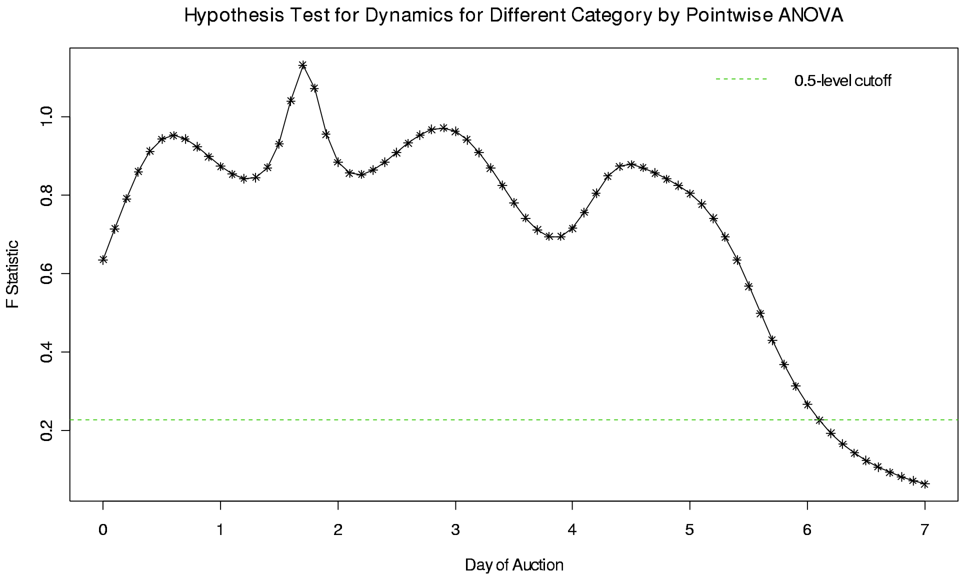

Figure 18: *Hypothesis test using pointwise F-statistics for the differences of weight functions estimated from auctions from different categories for Book data.*

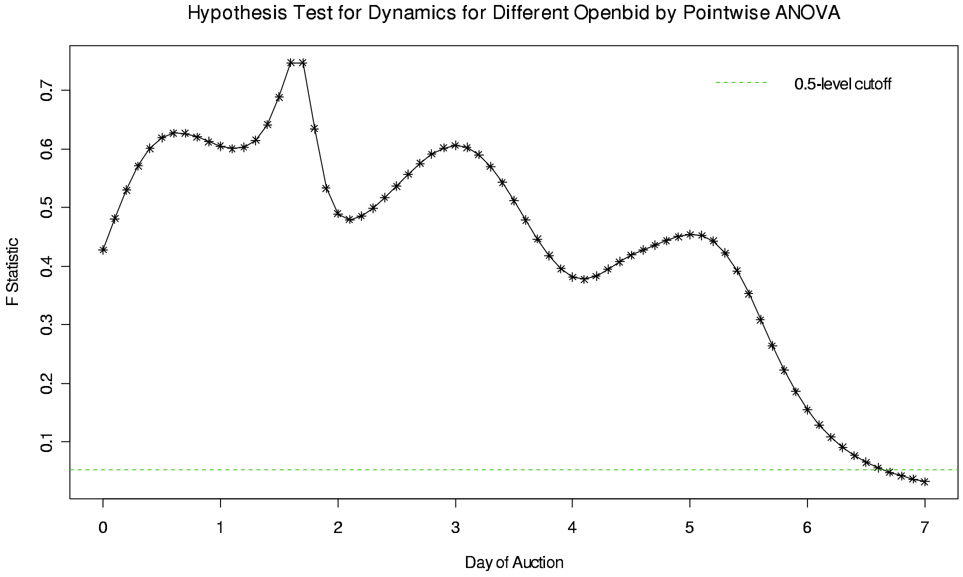

Figure 19: *Hypothesis test using pointwise F-statistics for the differences of weight functions estimated from auctions from different levels of opening bid for Book data.*

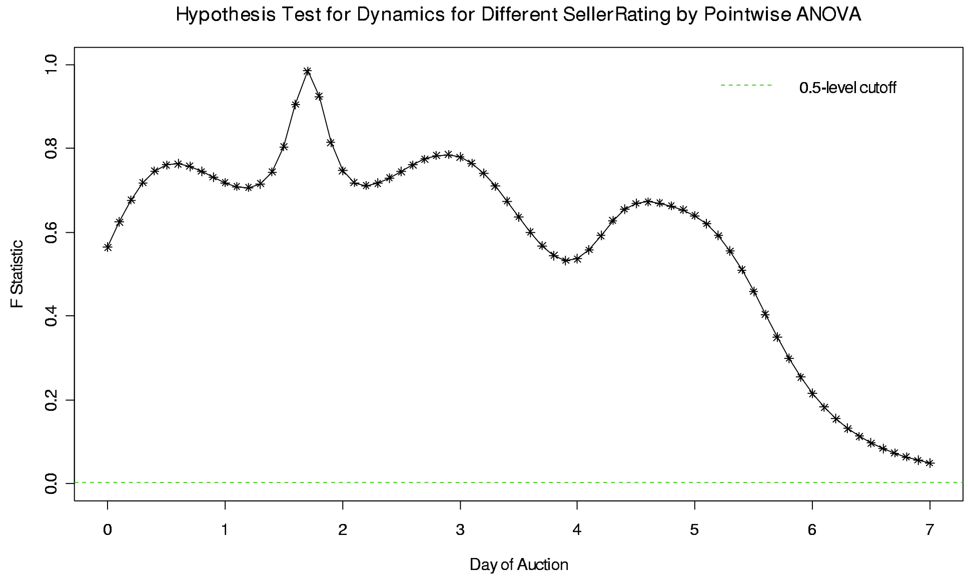

Figure 20: *Hypothesis test using pointwise F-statistics for the differences of weight functions estimated from auctions from different levels of seller's rating for Book data.*