

Ameliorating Buyer's Remorse

Rakesh Agrawal
Search Labs
Microsoft Research
rakesha@microsoft.com

Samuel Jeong
Search Labs
Microsoft Research
saieong@microsoft.com

Raja Velu*
School of Management
Syracuse University
rpvelu@syr.edu

ABSTRACT

Keeping in pace with the increasing importance of commerce conducted over the Web, several e-commerce websites now provide admirable facilities for helping consumers decide *what* product to buy and *where* to buy it. However, since the prices of durable and high-tech products generally fall over time, a buyer of such products is often faced with a dilemma: Should she buy the product now or wait for cheaper prices?

We present the design and implementation of Prodcast, an experimental system whose goal is to help consumers decide *when* to buy a product. The system makes use of forecasts of future prices based on price histories of the products, incorporating features such as sales volume, seasonality, and competition in making its recommendation. We describe techniques that are well-suited for this task and present a comprehensive evaluation of their relative merits using retail sales data for electronic products. Our back-testing of the system indicates that the system is capable of helping consumers time their purchase, resulting in significant savings to them.

Categories and Subject Descriptors

H.2.8 [Database Management]: Database Applications—*Data Mining*; J.4 [Social and Behavioral Sciences]: Economics

General Terms

Experimentation, Algorithms, Economics

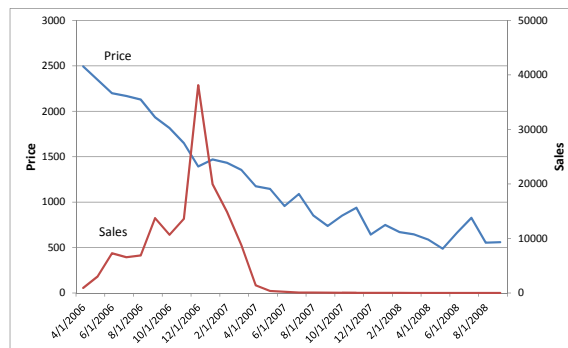
Keywords

Forecasting, Recommendation, Prodcast

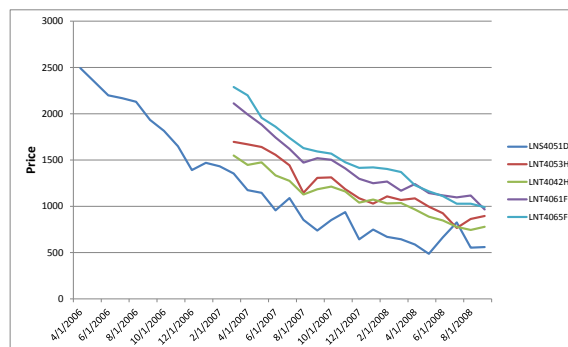
*Work done while author was visiting Search Labs.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

KDD'11, August 21–24, 2011, San Diego, California, USA.
Copyright 2011 ACM 978-1-4503-0813-7/11/08 ...\$10.00.



(a) Price & sales history of Samsung LCD TV LN4051D



(b) Prices of comparable Samsung LCD TVs

Figure 1: Price and sales data of Samsung LCD TVs (Source: NPD)

1. INTRODUCTION

The prices of durable and high-tech products generally fall over time as firms continually introduce products that have enhanced features [13, 17]. As an illustration of how product prices evolve, let us consider the category of LCD TVs, and focus on the TVs made by Samsung between 2006 and 2008. Figure 1(a) shows the price and sales of a particular model, LNS4051D (40 inch Widescreen LCD HDTV). This product first appeared in the market in April 2006. Its sales first increased, peaking in December 2006, and then decreased rapidly until Spring when new models were introduced. Meanwhile, its price declined steadily. The price histories of five comparable models of TVs made by Samsung are shown in Figure 1(b). They also demonstrate similar declining trend.

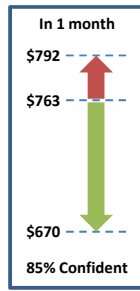


Figure 2: Presenting forecasts to Consumers

Buyers of durable products often suffer from a sense of regret after having made a purchase. While the origin of this remorse can be traced to various factors [5], buyers often agonize whether they could have bought the product for a cheaper price just by waiting for some time. The low-price guarantees advertised by merchants are of little avail because of the loop holes and the hassle costs associated with enforcing the compliance [9].

We present a system that we call Prodcast, which has been designed to ameliorate buyer’s remorse by providing expected price movement information to a consumer interested in buying a particular product. The system forecasts future prices based on price history of the product, incorporating factors such as sales volume, seasonality, and competitive products. This information is encapsulated and presented to the user in the form of a widget as illustrated in Figure 2.

This schematic allows a consumer to learn not only about the direction of future price change, but also distributional aspects of the change. These are important information especially in the context of consumer durables, as prices generally exhibit a downward trend, and the key to making good timing decisions is in knowing the magnitude and likelihood of the change. In this specific example, the current price of the product is at \$763. In one month’s time, the system is 85% certain that prices will be between \$670 and \$792, with the likelihood of a price drop outweighing potential price increase by about 3-to-1.

We describe techniques that are well-suited for this application domain and present a comprehensive evaluation of their relative merits using retail sales data for electronic products. Our back-testing of the system indicates that the system is capable of helping consumers time their purchase, resulting in significant savings to them.

The rest of the paper is organized as follows. We begin with a discussion of related work in Section 2. In Section 3, we give an overview of the architecture of Prodcast, and discuss how the system is evaluated. Focusing on the question of how best to forecast future product prices, we describe a number of methods that we considered in our evaluation in Section 4. We present a large-scale experiment using real-life data to compare the performances of these methods in Section 5. We conclude with the main insights gained from the experiments in Section 6.

2. RELATED WORK

The inspiration for Prodcast came from the the pioneering work of Etzioni *et al.* for making wait or buy recommendation for airline tickets [8]. The authors develop models to

forecast airfare and make recommendations once the consumer reveals the travel date and the points of origin and destination. Price data for particular flights was gathered from the web, and features such as number of hours until departure and the current prices are extracted from the raw data. Various data mining algorithms including Rule Learning, Q-Learning and time series methods are applied to mine the data. The quality of the predictive models is evaluated by quantifying the savings that each model would generate. While all the methods resulted in net savings, a model that combines the forecasts made by the various methods produced maximum savings.

There is a crucial difference in the nature of durable products and airline tickets that makes the two problems very different. In the case of a durable product, a user derives utility from owning and using the product over time, hence if a user defers the purchase, she suffers a *loss of use* while waiting. This cost must be factored into the decision. In contrast, a user derives utility from using an airline ticket for travel. There is no loss in utility if a user obtains the ticket now or later, provided that the ticket is obtained prior to the date of travel. Hence, the problem of deciding when to buy a ticket depends only on whether prices are likely to increase or decrease, whereas the problem of deciding when to buy a durable good depends also on the speed and the magnitude of the changes, as trade-offs between potential price drops and loss of use are made.

Few websites tackle the problem of helping consumers with deciding when to purchase consumer durables. The site Nextag provides just a price history and the current prices with various vendors. A recent site, Gazaro, provides price history for consumer durables and advocates when a product is ready to be purchased. Based on description from its website, its methodology considers factors such as competition, volatility, and price trends from both online and brick and mortar retailers. The problem it addresses is closer in spirit to ours. However, there is no published paper on the techniques or the quality of the recommendations.

There has been a number of studies in the marketing literature on how product prices evolve over time. These studies typically examine the question from the viewpoints of the manufacturers. Prices are believed to evolve in a way related to the lifecycle of a product. The theory of experience curves states that the unit cost of production declines over time as a manufacturer gains experience in the production of a product, and the relationship is non-linear. This model has been successfully applied to model electronic goods prices in [6]. However, the analysis was at a category level and do not directly apply to our setting as our focus is on price evolution at a product level. There has also been work that combines experience curves with demand models [7]. For a review of pricing research in marketing, see [14, 15].

In summary, existing work in marketing research provides evidence that product prices do not behave like random walks, but rather follow definite patterns that depend on the product lifecycles. We believe that by considering prices, sales, and competition, we should be able to produce good price forecasts that can help consumers make more informed decisions on when to buy a product.

There is a large body of related work in statistics on forecasting methods. The classical approach is based on time-series analysis [2]. More recently, Hyndman *et al.* have achieved much success on data sets from forecasting com-

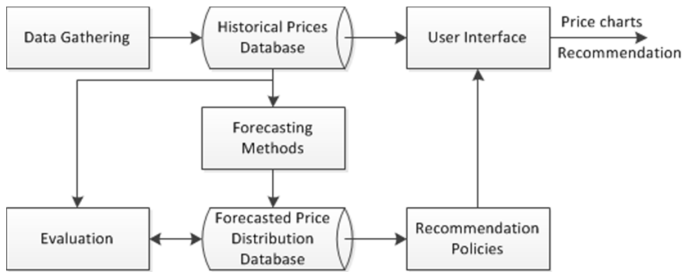


Figure 3: Overview of the Prodcast system

petitions (known as the M-competitions [11]) based on a smoothing-based approach [10]. We compare a number of methods that utilize these approaches in this paper, investigate to what extents product prices can be predicted, and find out what additional information can help with the prediction.

3. SYSTEM DESIGN

We provide an overview of the Prodcast system in Figure 3. The system consists of five major components:

Data Gathering. The data gathering component is responsible for collecting product prices and other information that can help with predicting price changes such as sales volume data. Our present implementation is based on data feeds, and performs basic outlier detection to identify potential data entry errors. An alternative implementation could obtain price data by scraping e-commerce retailer websites. The resulting data are stored in a historical prices database.

Forecasting Methods. The forecasting component is responsible for making forecasts of future prices based on the data gathered about the products. The methods take as input price histories and other information such as sales volume, and produces a distribution over future prices. The results are stored in a forecasted price distribution database, both for evaluation and as an input to the recommendation policies. We focus on this component in this paper, describe a number of methods for this problem and conduct a systematic evaluation of these methods over a real-life data set.

Recommendation Policies. The recommendation component is responsible for making buy-or-wait recommendations to the users based on the forecasted price distribution and user preference parameters. This is a component separate from the forecasting component, as given the same forecasts, the recommendation may be different depending on the user preferences.

User Interface. The user interface component is responsible for providing users with price-related information of the product. In response to a query for a product, the component renders a price chart depicting the price history of the product, and also produces a graphical display of the forecast prices as illustrated in Figure 2. For user interested in receiving a customized recommendation for when to buy the product, the interface further elicits certain preference parameters and provides those as input to the recommendation policies, which can then act as a service that alerts the user when to buy the product.

Evaluation. The evaluation component is responsible for running periodic evaluation of the forecasting methods, and

keeping track of their historical performances. It makes use of the archive data, splits them into training and test sets, and measures the performance of the system. It helps to select the best method for predicting prices for each product.

3.1 Problem Description

Consider a consumer who has selected a particular durable product to buy, and is deciding whether to purchase the product now or at a future time period. The consumer is described by two parameters: a value θ that measures her utility for the product, and a value λ that captures her *loss of use* of the product while she waits. If she purchases the product now, she gains a utility of θ ; if she purchases in the future time period, she gains a utility of $(\theta - \lambda)$. The loss of use parameter captures the important characteristic of durable products that they provide value to the consumers over their lifetime. Its introduction presents a trade-off between buying now and waiting that was not captured by past work in the domain of airfare and tickets.

Let the current price of the product be p . If we know the price of the product in the future time period, say p' , the problem would be simple. We can simply compare between the net utility of buying the product now, $(\theta - p)$, and that of buying in the future, $(\theta - \lambda - p')$, and choose accordingly based on which value is higher. The problem is that we do not know the future price. The role of the forecasting methods is to make use of historical data to make forecasts of the future price.

Let the historical prices of the product be \mathbf{p} , and other relevant historical data, such as sales and competition, be \mathbf{S} . We define the recommendation problem as follows.

Definition 1 (Buy-or-wait Recommendation Problem). Given consumer parameters θ and λ , the current price of the product p , and historical data \mathbf{p} and \mathbf{S} , recommend to the consumer whether she should buy the product now or wait to buy it in the future period.

Given historical prices \mathbf{p} and historical data \mathbf{S} , suppose the forecasting method produces a forecasted distribution of the future prices D . A natural recommendation policy is to take as input consumer parameters θ and λ , the current price p and the forecasted distribution D , and recommends the consumer to buy if

$$\theta - p \geq \theta - \lambda - \mathbb{E}_{p' \sim D}[p'] ,$$

where $\mathbb{E}_{p' \sim D}$ denotes the expectation of p' drawn according to distribution D . This policy tacitly assumes that the consumer is *risk-neutral*, and attempts to maximize the expected utility of the consumer based on the forecasts.

In general, a recommendation policy may take into account additional information such as the risk attitude of the consumers. It may require as input not only the mean forecast of the future price, but also possibly its entire distribution. Thus, we impose a requirement on the forecasting methods to generate distribution of forecasts as their output. Fortunately, as we will see in Section 4.1, there exists a general approach to generate forecast distributions from mean forecasts.

To determine the performance of the recommendation system, we measure the utility consumers have if they follow the recommendations made. Suppose we are given a set of n problem instances, where the i -th instance is given by $(\theta^i, \lambda^i, p^i, \mathbf{p}^i, \mathbf{S}^i)$ and $p^{i'}$ which denotes the realized price in

the future time period. Denote the recommendation of the system for the i -th instance by rec^i . The performance of the system is defined as

$$\sum_{i=1}^n \begin{cases} (\theta^i - p^i) & \text{if } \text{rec}^i \text{ is buy now} \\ (\theta^i - \lambda^i - p^i) & \text{if } \text{rec}^i \text{ is wait} \end{cases} \quad (1)$$

4. FORECASTING METHODS

We now describe a number of forecasting methods that we explore in this study. We first describe the notations for this section and a general technique to produce distribution of forecasts from methods that only produce mean forecasts. We then describe a number of methods that only use past prices for predicting future prices. We next look into how additional signals, such as seasonality, sales, and competition, can be incorporated into the forecasts. Finally, we describe methods that aggregate forecasts.

4.1 Preliminaries

Recall that the input to the forecasting methods are price histories \mathbf{h} and additional historical data \mathbf{S} . For concreteness, we denote the length of the history by s . Hence, $\mathbf{h} = (p_1, p_2, \dots, p_s)$ and $\mathbf{S} = (S_1, S_2, \dots, S_s)$. Thus, the price p_s constitutes the current price according to our problem formulation in Section 3.1, and the objective of the methods is to come up with a distribution of prices for time $(s + 1)$.

In the proceeding, some methods may use only a subset of the history. Let the history up to time t , $t \leq s$, be denoted \mathbf{h}_t and \mathbf{S}_t . For clarity, we denote the overall history analogously by \mathbf{h}_s and \mathbf{S}_s .

Under the framework presented in Section 3.1, forecasting methods are required to output a *distribution* of future prices. We now describe a simple and parsimonious approach to convert methods that only produce mean forecasts to ones that produce distribution of forecasts. We apply this approach to all the methods described in this section.

Given price history \mathbf{h}_s and meta-data \mathbf{S}_s , let the mean forecast according to method A of the price at time $s + 1$ be denoted $\mu_1^A(\mathbf{h}_s, \mathbf{S}_s)$. The historical *in-sample sum of squared errors*, SE_1^A , is defined as

$$SE_1^A(\mathbf{h}_s, \mathbf{S}_s) = \frac{1}{s} \sum_{t=1}^s \left(\mu_1^A(\mathbf{h}_{t-1}, \mathbf{S}_{t-1}) - p_t \right)^2.$$

The distribution at time $s + 1$ is then given by a Gaussian distribution with mean $\mu_1^A(\mathbf{h}_s, \mathbf{S}_s)$ and variance $SE_1^A(\mathbf{h}_s, \mathbf{S}_s)$. This distribution is consistent with the interpretation that method A maximizes the likelihood of the observed data under normally distributed errors.

4.2 Methods that Only Use Prices

We consider two main classes of methods. One class include heuristic methods that are based on smoothing the data. Another class include autoregressive models that are based on time series analysis. They have been used extensively in forecasting [2, 10]. We review these methods in this subsection for completeness.

4.2.1 Smoothing Methods

We consider three smoothing methods in this study. All three methods depend on certain parameters that control the amount of smoothing. We learn these parameters by maximizing likelihood of observing the historical prices.

Moving Average (MA). Given a window size parameter w , for history \mathbf{h}_s , the mean forecast of the price at time $s + 1$ equals

$$\mu_1^{MA}(\mathbf{h}_s) = \begin{cases} (1/w) \sum_{i=0}^{w-1} p_{s-i} & \text{if } s > w \\ \text{undefined} & \text{otherwise} \end{cases}.$$

The window size parameter controls how much smoothing is done. A larger window size increases the amount of smoothing performed on the history.

Exponential Smoothing (Exp). Given a decay parameter α ($0 < \alpha < 1$) and a seed value \hat{p} , for history \mathbf{h}_s , the mean forecast of the price at time $s + 1$ is defined as an exponentially weighted sum of past prices, i.e.,

$$\begin{aligned} \mu_1^{Exp}(\mathbf{h}_s) &= \alpha p_s + (1 - \alpha) \mu_1^{Exp}(\mathbf{h}_{s-1}) \\ &= \alpha \sum_{i=1}^s (1 - \alpha)^{s-i} p_i + (1 - \alpha)^s \hat{p} \end{aligned}$$

and for $s = 0$, $\mu_1^{Exp}(\mathbf{h}_s) = \hat{p}$. The decay parameter controls how much the method weighs recent data compared to past data. Smaller values of α increases the weight placed on historical prices.

Holt's Method (Holts). The Holt's method is a generalization of exponential smoothing. Given a level decay parameter α , a growth decay parameter β , and seed values of level and growth, $\hat{\ell}$ and \hat{b} , the mean forecast of the price at time $s + 1$ is defined by the set of update equations, i.e.,

$$\mu_1^{Holts}(\mathbf{h}_s) = \ell_s + b_s$$

where for $s \geq 1$,

$$\begin{aligned} \ell_s &= \alpha p_s + (1 - \alpha) \mu_1^{Holts}(\mathbf{h}_{s-1}) \\ b_s &= \beta (p_s - \ell_{s-1}) + (1 - \beta) b_{s-1} \end{aligned}$$

and for $s = 0$, $\mu_1^{Holts}(\mathbf{h}_s) = \hat{\ell}$ and $b_s = \hat{b}$. Note that when β and b_1 are set to 0, the method is the same as exponential smoothing.

4.2.2 Autoregressive (AR) Models

Autoregressive models are a family of models that predict future values as linear combinations of past values. The model is specified by the number of past terms the future values depend on (known as its *order*). An autoregressive model of order k , $AR(k)$, is described by the equation

$$p_t = \sum_{i=1}^k \beta_i p_{t-i} + \epsilon_t \quad (2)$$

where ϵ_t are normally distributed errors. The parameters β are estimated by linear regression. Under our framework, for history \mathbf{h}_s , $\mu_1^{AR(k)}(\mathbf{h}_s) = p_{s+1} = \sum_{i=1}^k \beta_i p_{s-i+1}$.

For product prices, one may want to model price changes rather than price levels. This can be achieved by taking the differences between successive values in the series, i.e.,

$$\Delta_t = p_t - p_{t-1}.$$

The differences Δ_t is then modeled by an autoregressive model. We denote an autoregressive model of order k over the differences by $AR\text{-Diff}(k)$.

The order of the model and whether to difference the series is often selected heuristically. In our experiments, we try a number of models to find what works best for our problem.

One advantage of AR models is its extensibility. One can easily incorporate additional signals by adding the signals to the relationship in Equation (2). Such extensions are considered later in this section.

4.3 Modeling Seasonality

Seasonality refers to *cyclical* variations in the sequence of data being modeled. For example, consumer spending often exhibits seasonality (higher in November and December due to holiday shopping).

Classical methods for estimating seasonality require at least two full cycles of data [10]. They do not readily apply to our problem. This is because the length of a cycle for our problem is 12 months, whereas the length of the lifecycle of a typical electronic product is between 18 and 30 months, as estimated using the data set described further in Section 5. Thus, these methods could at best help with forecasts for the last few months of a product about to be discontinued, and are not sufficiently applicable to our problem.

We propose to model seasonality by grouping together products that are similar to the one in question. This allows us to learn seasonality at an aggregate level and apply it to the individual products. Let P denote a set of products and L denote the number of time units per cycle (e.g., 12 months or 52 weeks per year). First, we compute the relative change in prices for each product i in P , $d_t^i = (p_t^i/p_{t-1}^i) - 1$. Next, we compute the average change per period, \bar{d}_t , by averaging over the individual changes d_t^i . We then compute the average change in each time unit within a cycle, i.e., for $j < L$,

$$\text{season}_j = \text{Average}\left(\{\bar{d}_t \mid t \bmod L = j\}\right).$$

Finally, we compute a *seasonal effect* by subtracting the average overall change from s_j , i.e.,

$$\text{season}_j^* = \text{season}_j - \frac{1}{L} \sum_{i=0}^{L-1} \text{season}_i$$

To make use of seasonal effects in forecast, for each product, we first *detrend* the product prices by computing

$$p_t^* = p_t / (1 + \text{season}_{f(t)}^*)$$

where $f(t)$ selects the corresponding season for time t . This has the effect of accounting for the price changes that can be explained by seasonal fluctuations. Forecasting is then performed over the new time series \mathbf{p}^* . To recover the actual value, the mean forecast for time s is multiplied by the seasonal effect $(1 + \text{season}_{f(s)}^*)$.

This procedure leaves open the question of how to select a similar set of products. We explore this question experimentally in Section 5.

4.4 Incorporating Sales Volume

Based on past studies in marketing research, sales volume is an important factor in determining product prices [6]. We propose two approaches to incorporating sales volume in making the forecasts.

In one approach, we add sales volume data to Equation (2) as part of a regression model. In our study, we add the sales volume data from a number of past periods corresponding to the order of the AR model.

In another approach, we use sales volume data to help determine the “stage” of the lifecycle of a product, and learn a separate AR model for each stage. We use autoregressive

trees (AR Trees), proposed in [12], for this purpose. Intuitively, the AR tree approach combines decision trees with AR model. Like a decision tree, each of the internal nodes of the tree splits the data based on some criteria, in this case by sales volume. At the leaf node, an autoregressive model is learned for the data that are classified to the node.

4.5 Incorporating Competition

Competition among products exists in many forms, such as through price or feature differentiation. In this study, we focus on price competition. Given a set of competing products P to the product in question, we propose three approaches to capture the competition. These approaches are based on creating correlated time series that are added to an autoregressive model as additional regression variable(s). They differ in how these time series are generated.

Top k competitors. Create a set of k time series to be added to the AR model by selecting k products from P that minimizes the in-sample sum of squared errors for the product in question. We consider the top 1 competitor in the experiments.

Average competition. Create a single time series by taking the average price over all products in P . The motivation behind this approach is to capture the overall competition from the market.

Principal Component Analysis (PCA). Create a set of k time series by selecting the first k eigenvectors that correspond to the principal component analysis of the prices of all products in P . Formally, form a matrix M where each column corresponds to the prices of a product in P , and each row corresponds to the prices for a given time index. Perform an eigen decomposition of matrix M (i.e., run PCA). This can be interpreted as identifying the variables that best explain the variation in the data [1]. By adding the eigenvectors as regressors, the approach attempts to incorporate signals that help to explain price changes. To make this approach comparable to the other two in experiments, we add only the principal eigenvector so that each method adds only one additional time series in its regression.

4.6 Aggregating Forecasts

A number of studies in statistics and machine learning has found that aggregating forecasts can often lead to better performances than individual forecasts [3, 4, 16]. We consider three approaches to aggregation in this study, all of which has been successfully used in the literature.

Average. Take the simple average of the forecasts to produce the mean of the forecast. This approach tends to be robust to outliers that may skew the distribution.

Regression. Use past data to learn a linear combination of the forecasts that minimizes empirical error. This approach adapts to the data and can serve as means to selecting the best method.

Inverse Variance. Take a weighted average of the forecasts inversely proportional to the variance (and hence, sample error) of the forecasts. This places less weights on forecasts that have high sample errors historically.

5. EXPERIMENTS

To find out what forecasting methods work best for predicting product prices, and how to incorporate additional

sources of signals such as seasonality, sales volume, and competition, we conduct a large-scale empirical study in the electronic goods domain. We start out with using only price data for forecasting, then tested different ways to incorporate each source of signals, and finally consider ways to aggregate the forecasts.

5.1 Data

We obtained monthly price and sales volume data for four categories of electronic products—Televisions, Digital cameras, Camcorders, and Printers—between January 2005 and September 2008. The data is obtained from NPD, a commercial data provider, and is based on sales registered in about ten major retail chains across the US. Prices are computed as the volume-weighted average prices of the products at the checkout registers of these stores.

The evaluation period is from June 2008 to September 2008. Products for which the month-to-month change in prices is greater than 4 times (i.e., more than 300% increase or 75% decrease) are filtered out, as we believe these are due to data entry errors. After this preprocessing step, there are about 1,200 products in the data set. Over the evaluation period, the value of products sold per month was worth about 380 million dollars. The weighted average month-to-month decrease in product prices is about 3.5%, with a standard error of about 5%. Most of the time series are short, with between 18 and 30 data points. These are considered challenging problems in time series analysis due to data sparsity.

5.2 Methodology

We evaluate the methods by measuring the performance of the system as given by Equation (1) in Section 3.1. For each product in the dataset, we generate four problem instances corresponding to each of the four months between June 2008 and September 2008. The consumer parameters for each problem instance are chosen as follows. The value of the product θ is chosen to be 5% higher than the current price when a consumer first gets interested in buying the product. The loss of use λ is expressed as a fraction of the value θ . This fraction, which we call *loss of use ratio*, is chosen to be between 2.5% and 4.5%. These values are consistent with past studies on consumer discount rates for electronic goods, which can be interpreted as proxies for the loss of use [18].

To normalize the results, we report the performance relative to the maximum possible utility that a consumer can obtain with perfect foresight. This maximum is unachievable in practice due to the stochastic nature of price changes. Nonetheless, it helps us gauge how close we are to the best possible result. For the chosen consumer parameters, a 1% difference in relative performance corresponds to about 1.1 million dollars in utility for the data under consideration. A 70% relative performance corresponds to about 77 million dollars in utility.

5.3 Results

5.3.1 Methods that Only Use Prices

We first consider forecasting methods that only use past prices in making a forecast. These include the three smoothing methods presented in Section 4.2.1, with parameters selected by maximizing likelihood (MA-MLE, Exp-MLE, Holts-MLE), and three autoregressive models with different orders

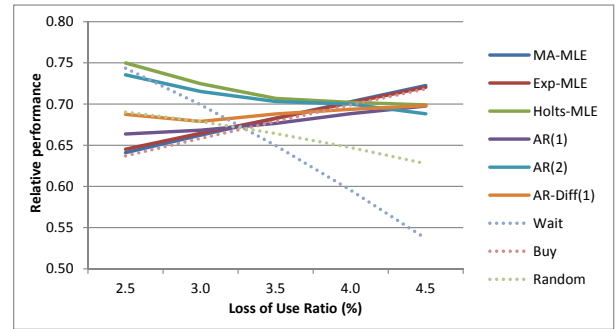


Figure 4: Relative performances of forecasting methods that use only prices, compared to baselines.

and differencing (AR(1), AR(2), AR-Diff(1)). For comparison, we also report the performance obtained when a consumer always buys the product right away (Buy), always waits to buy the product (Wait), and randomly choose between buying now or waiting (Random). The results are presented in Figure 4. The key observations are:

- The best performing methods are Holt’s methods and AR(2). They outperform all three baselines by an average of 5% over the entire range, and with the exception at the extremes when wait is better at a discount rate of 2.5% and buy is better at a discount rate of 4.5%, they are better than all three baselines. Both methods share a similar characteristic in that they consider how prices have *changed* in the past period. This turns out to be important for our application.
- The performances of MA-MLE and Exp-MLE closely mirror that of the buy baseline in the entire range. This is because the forecasts made by both methods are close to the price in the preceding time period, corresponding to a small window parameter w and a large decay parameter α respectively. As a result, the recommendation using either forecast is to buy, except for the rare cases where the methods have identified a definite trend.
- AR(2) outperforms both AR(1) and AR-Diff(1). This is likely because AR(2) uses more data than the other two models, and hence produces more accurate forecasts. However, increasing the orders further to incorporate yet more data could lead to overfitting. Indeed, we observe worse performances for AR models of higher order (not shown in figure to avoid clutter).

5.3.2 Modeling Seasonality

In Section 4, we proposed an approach to model seasonality that depends on how products are grouped together to estimate seasonal trends. We consider two groupings in this section. In one, we group all electronic products together to learn a global seasonal trend. In another, we group together products by their category to learn category-specific seasonal trends. The seasonal effects we learn under these groupings are shown in Figure 5.

We observe that category-specific trends exhibit higher variability compared to the global trend. For all trends, the seasonal effects fluctuate within a fairly small band (between -1.5% and +1.5%) between January and October. There is a

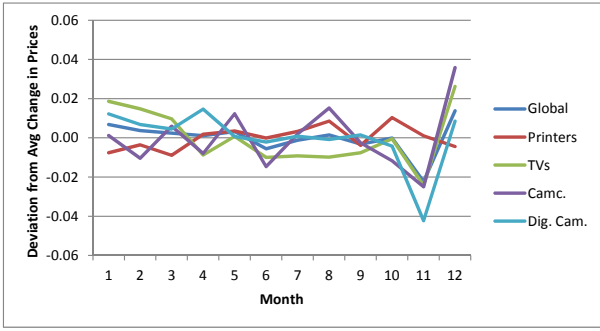


Figure 5: Global and category-specific seasonalities of product prices.

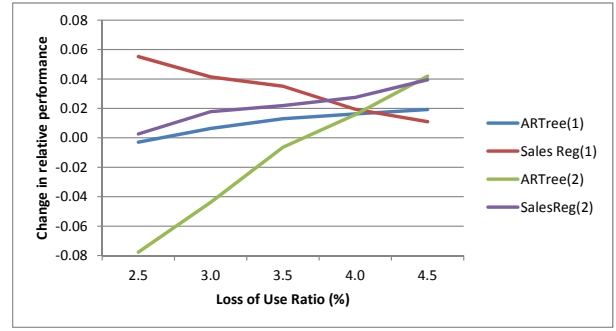


Figure 7: Change in relative performance when sales are incorporated in AR models.

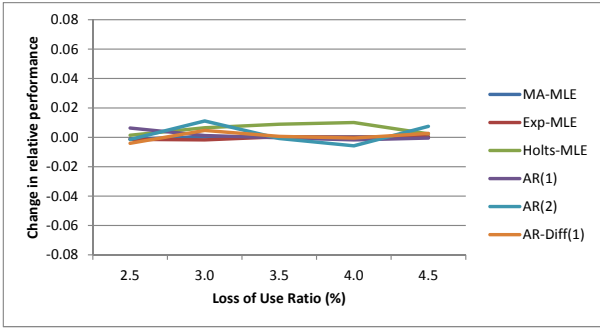


Figure 6: Change in relative performance after correcting for a global seasonal trend. Results for that of category-specific seasonal trends are similar.

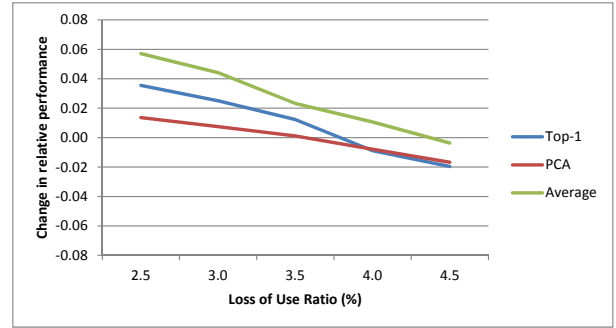


Figure 8: Change in relative performance when competition is incorporated.

large seasonal effect that leads to larger-than-expected price drops in November, follow by smaller-than-expected drops in December. The trends do depend on the category. Printers, for example, exhibit a relatively stable trend throughout.

Next, we incorporate these trends to our forecasting methods. The changes in relative performances after incorporating a global seasonal trend are shown in Figures 6. The results using category-specific trends are similar and omitted for brevity.

We observe that the relative performances for all of the methods remain almost the same after incorporating the seasonal trends. It is not surprising because from June to September, the months over which performance is evaluated, we observe relatively little seasonal effects in Figure 5. We expect a larger gain in performance for November and December. As there are only three full years of data, we were unable to verify the hypothesis due to data sparsity.

5.3.3 Incorporating Sales Volume

In Section 4, we presented two approaches that use sales volume data to help predict future prices. Both methods extend the AR models but differ in how sales data is used. Under the AR Tree approach, sales data is used to segment the AR model into piece-wise linear region. Under the regression approach, sales data are treated as independent variables in a regression against future prices. We present the results of extending the AR(1) and the AR(2) models under both approaches in Figure 7.

We observe that with the exception of extending the AR(2) model using AR Tree, incorporating sales lead to higher rel-

ative performances, in some cases fairly significant improvements. For both models, the AR Tree approach lags behind the regression approach. To understand how the two approaches differ, we compare the recommendations according to the AR models and their extensions. We find that the regression approach disagrees on about 20% of the products. It finds better recommendations in about 75% of the time, explaining its success. On the other hand, the AR tree approach disagrees on about 40% of the products. It finds better recommendations in only about half the time. Hence, the improvement in relative performances is less, and in some cases the relative performance is actually worse.

5.3.4 Incorporating Competition

We presented three different methods to incorporate competition in Section 4. In this experiment, we treat all products within a given category as competition, as we lack detailed product specification to perform more refined groupings of products. The results of extending the AR(1) model are presented in Figure 8.

We observe that taking the average prices of all products within a category gives the largest gain in relative performances among all three methods of incorporating competition. There is a general downward trend for all three methods as discount rates increase. Examining the recommendation in more detail, it appears that including competition often leads to lower price forecast, and hence an increased likelihood to recommend wait to the consumer. This leads to worse relative performances as discount rate increases since waiting is likely to earn less utility at higher discount rates.

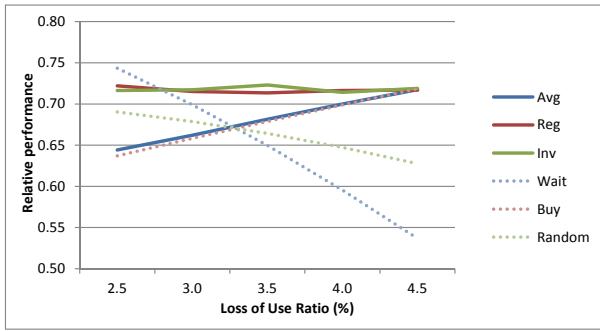


Figure 9: Relative performances of aggregating forecasts from four constituent methods (Holt's-MLE, SalesReg(2), MA-MLE, AR-Diff(1)).

5.3.5 Aggregating Forecasts

We now consider whether aggregating forecasts can improve performances. We select a number of forecasting methods to aggregate based on their performance and how much they differ from one another. Adding more varied forecast may help to complement the shortcomings of one another. Based on the experiments, we selected Holt's methods and sales incorporated through regression (SalesReg(2)) as these are the two best individual methods. From the first set of experiments, we found that both MA-MLE and AR-Diff(1) perform reasonably well, and exhibit qualitatively different behavior as their performances improve as discount rate increases. The results based on the three aggregation methods of averaging (Avg), regression (Reg), and inverse variance (Inv) are presented in Figure 9.

We observe that aggregating forecasts by regression and inverse variance both outperform aggregating forecasts by averaging. The performances of the former two are comparable to the best individual method in the group (SalesReg(2)), but deliver more consistent performances over the entire range of discount rates and outperform all three baselines in almost all cases. We conclude from this experiment that aggregating forecasts is a robust approach to ensuring good performances.

5.3.6 Sensitivity to Value Parameter

For the above experiments, the value of the product θ has been chosen to be 5% higher than the current price when a consumer first gets interested in buying the product. Varying this parameter does not change the relative ordering among the forecasting methods, but it has an effect on the total expected utility the system generates for its users and the relative performances. When θ is 4% higher than the current price, the relative performance of the best method becomes 68%; when θ is 6% higher, the relative performance becomes 76%. This trend is because as θ gets smaller, there is an increased likelihood that a wait recommendation may lead to negative utility. This can be avoided with perfect foresight, hence the relative performance of all methods, including the best one, decreases. Nonetheless, the gap between the best method and the baselines remain consistent at about 5% of relative performance.

6. CONCLUDING REMARKS

We presented Prodcast, an experimental system that is designed to ameliorate buyer's remorse by providing recommendations to consumers on when to buy a durable product. It complements the facilities provided by existing e-commerce websites that help consumers decide what product to buy and where to buy it. We gave an overview of the system architecture, and described a framework for evaluating the system based on the expected utility it generates for the users. Focusing on the question of what forecasting methods work well for this application domain, we systematically compared a large number of forecasting methods based on smoothing and time series analysis. We analyzed additional signals such as seasonality, sales volume, and competition to see to what extent they can help to predict product prices. We also investigated various methods for combining forecasts. Our best method achieves a relative performance of about 72% compared to the optimal decision with perfect foresight. It outperforms the best baseline by about 5% in our dataset. This confirms that the proposed system can generate significant utilities to users shopping for durable products.

We have gained a number of insights from our experimental study using real-life data. First, we found that forecasting methods that use both the change in prices and the past price level, such as Holt's methods and AR(2) models, do well in this domain. Due to data sparsity, methods that use more past data than these methods end up doing worse as a result of overfitting. Second, the relatively short time series present a challenge to traditional approaches to modeling seasonality. By aggregating across multiple products, we discover a trend that is born out by experience, i.e., prices tend to drop more than expected in November. These seasonal effects may help to improve forecasts for the holiday seasons. Third, both sales volume and competition are sources of signals that can be leveraged to improve performance. However, the degree of improvements depends on how these signals are used. Based on our experiment, sales are best incorporated through regression, and competition through averaging. Finally, while we do not see gains in performances by aggregating forecasts compared to the best forecasting method, aggregating forecasts can smooth out the performances across different discount rates, and alleviate the need to explicitly select a specific method.

Some interesting directions for future work are as follows. In this paper, we have considered the problem of recommending whether to buy now or in the next time period. Depending on the length of a time period, we may want to consider the more general problem of buying now or some time in the future several periods from now. This requires extending the forecasting techniques studied to produce multi-period forecasts.

In our experiments, we have relied on data obtained from a commercial vendor. To obtain more timely data, one may want to crawl the data from online retailers. This introduces challenges on how to ensure the quality of the data, and how to aggregate the data, which are at the levels of individual offers, to reflect the prices of the underlying products.

7. REFERENCES

- [1] C. M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.
- [2] G. Box, G. M. Jenkins, and G. Reinsel. *Time Series Analysis: Forecasting and Control*. Prentice Hall, 3rd edition, 1994.
- [3] R. T. Clemen. Combining forecasts: A review and annotated bibliography. *Int. J. Forecasting*, 5:559–583, 1989.
- [4] R. T. Clemen, A. H. Murphy, and R. L. Winkler. Screening probability forecasts: Contrasts between choosing and combining. *Int. J. Forecasting*, 11:133–146, 1995.
- [5] W. H. Cummings and M. Venkatesan. Cognitive dissonance and consumer behavior: A review of the evidence. In M. J. Schlinger, editor, *Advances in Consumer Research Volume 2*, pages 21–32. Association for Consumer Research, 1975.
- [6] R. N. Dino. Forecasting the price evolution of new electronic products. *J. Forecasting*, 4:39–60, 1985.
- [7] R. J. Dolan and A. P. Jeuland. Experience curves and dynamic demand models: Implications for optimal pricing strategies. *J. Marketing*, 45:52–62, 1981.
- [8] O. Etzioni, R. Tuchinda, C. A. Knoblock, and A. Yates. To buy or not to buy: mining airfare data to minimize ticket purchase price. In *Proceedings of KDD*, pages 119–128, 2003.
- [9] M. Hviid and G. Shaffer. Hassle costs: The achilles heel of price-matching guarantees. *J. Economics & Management Strategy*, 8(4), 1999.
- [10] R. J. Hyndman, A. B. Koehler, J. K. Ord, and R. D. Snyder. *Forecasting with Exponential Smoothing*. Springer, 2008.
- [11] S. Makridakis and M. Hibon. The M3-competition: results, conclusions and implications. *Int. J. Forecasting*, 16(4):451–476, 2000.
- [12] C. Meek, D. M. Chickering, and D. Heckerman. Autoregressive tree models for time-series analysis. In *Proceedings of SIAM Int. Conf. on Data Mining*, 2002.
- [13] H. Nair. Intertemporal price discrimination with forward-looking consumers: Application to the US market for console video-games. *Quant. Market. and Econ.*, 5:239–292, 2007.
- [14] S. S. Oren. Comments on “pricing research in marketing: The state of the art”. *J. Business*, 57:61–64, 1984.
- [15] V. R. Rao. Pricing research in marketing: The state of the art. *J. Business*, 57:39–60, 1984.
- [16] R. E. Schapire. A brief introduction to boosting. In *Proceedings of IJCAI*, 1999.
- [17] I. Song and P. K. Chintagunta. A micromodel of new product adoption with heterogeneous and forward-looking consumers: Application to the digital camera category. *Quant. Market. and Econ.*, 1:371–407, 2003.
- [18] R. S. Winer. Discounting and its impact on durables buying decision. *Marketing Letters*, 8:109–118, 1997.