Designing Promotions: Consumers' Surprise and Perception of Discounts

Wei Sun, Pavankumar Murali, Anshul Sheopuri, Yi-Min Chee

Abstract: This paper proposes a behavioral pricing model that enhances traditional pricing algorithms by incorporating concepts from mathematical psychology and information theory on how consumers perceive discounts. We propose a framework that systematically incorporates the effect of quoted discounts and historical promotions on consumers' valuations and helps marketers determine the optimal discount strategy. We apply our framework on a publicly available data set from an online retailer. The data set consists of transactional and customer data. Our experiments reveal that the behavior pricing model can lead to very different pricing decisions compared to the traditional pricing model. For some product groups, we observe that the behavior model suggests offering lower discounts than the traditional pricing model to capture the *thrill* and *surprise* of a deal without sacrificing the profit margin to a significant extent. On the other hand, for certain product groups, while the traditional pricing model recommends not giving discounts at all, the behavioral pricing model suggests offering a smaller discount to entice customers to make purchases.

Introduction

Promotions are an important aspect of competitive dynamics in the retail sector. Retailers use promotion techniques such as typical price promotions, deep discount deals, feature advertising, and in-store displays to attract consumers. According to a Nielsen study [1], global advertising expenditure reached $557 billion in 2012. Due to the sheer volume of promotions and the dollars spent in running them, there has been a sizeable amount of work done on designing promotions, understanding how consumers react to price changes, and determining the optimal product prices.

Determining the optimal promotion strategy is complex as consumer purchase decisions depend not only on the price of the product and the profile of the customer, but also on softer factors such as whether the pricing strategy influences the customer's behavior and psychology. Levy et al. [2] summarize six factors that must be taken into consideration to determine optimal prices: price sensitivity, substitution effects, effect of price promotions over time, segment-based pricing, cross-category effects, and retailer costs. There is a vast body of knowledge in management science and marketing literature that propose econometric and discrete-choice models to address most of these factors. However, most of the academic literature and industry practice assumes that sales during a promotion are independent of past pricing activity and its effect on consumer behavior. Winer [3] demonstrates that consumers evaluate retail prices for items relative to certain internal reference prices which, in turn, could be influenced by past prices, brand promotion, and store type. This, in turn, influences how customers perceive discounts and the *thrill* and *surprise* they experience. Although this behavioral aspect of customer purchase has been studied in recent past by researchers in marketing under experimental settings, there is an absence of a quantitative framework that retailers could use to factor in consumer behavior. The importance of these soft factors is displayed in some recent examples in the space of promotion pricing. JCPenney, an American department store chain [4], adopted a new promotions strategy

that substituted non-stop store promotions with "everyday low prices". Customers at JCPenney derived a "thrill" out of collecting coupons and getting a great deal, even if it was an illusion. The store experienced its sales dropping by 25% in 2012.

Past research work in pricing has focused on applying econometric models, for example, discrete-choice models to capture customer preferences. For example, Dube et al. (2008) consider the evolution of consumer brand loyalty in determining optimal prices over time. The authors implement a flexible model to measure loyalty while allowing for a highly non-normal distribution of customer heterogeneity. There is a growing stream of literature in marketing and economics that models consumers as Bayesian learners. For implementation purposes, the learning framework is embedded within a discrete choice setting that is calibrated on consumer choice data. Examples include Ackerberg (2003), Mehta et al. (2003) etc. Shin et al. (2012) mention that the discrete choice model with Bayesian learning is data intensive and makes it difficult to distinguish between preference heterogeneity and state dependence. In other words, customer learning is not fully identified from revealed choice data. As a result, the initial conditions, prior to learning, are difficult to pin down. To circumvent this problem, they estimate a logit-based Bayesian learning model where the learning parameters are augmented by the survey information available on consumer preferences and familiarities. A common deficiency in the pricing literature discussed above is that they do not offer a programmatic approach to incorporate customer emotion into choice models while determining optimal prices during a promotion. Findings in behavior economics and psychology suggest that taking behavior and cognitive factors into account often leads to different performance predictions. Some questions that retailers would like to answer are - how do we incorporate customers' perception on promotion into a decision-making tool that systematically determines the right discount, how do we evaluate and adjust the system recommendation to suit the risk profile of the store?

To this end, we develop an adaptive pricing system which utilizes theories from mathematical psychology, machine learning, econometrics, information theory, and operations research to decide the discount that would be profitable to the retailer and, at the same time, increase surprise and thrill for the customer. The central idea is based on understanding human behavior such as the excitement or thrill from getting a deal or surprise associated with prices that deviate from what consumers expect, personalized to the profile of the customer (for example, income), channel of interaction (e-mail, social media, brick and mortar, etc.), product assortment, and promotion duration and budget. In particular, to reflect the emotion-driven aspect of decision-making process that consumers undergo while they shop, we model customer's perception of discount. It captures the thrill and excitement in using discount coupons or hunting for deals. We borrow the concept of Bayesian surprise from information theory to measure the novelty of a pricing strategy with respect to past promotions. Bayesian surprise assumes that a customer would have a prior distribution of expected discount levels based on their experience, and measures the change using the posterior distribution of expected discount levels after the new discount is revealed. A truly novel promotion would be reflected by a larger degree of change in their beliefs. To the best of our knowledge, there is very little prior work that incorporates these concepts from consumer psychology and behavior economics to identify optimal prices.

The remainder of this paper is organized as follows. Firstly, we explain the key components of the behavior pricing system and introduce the main concepts. Next, we provide an illustration of

this system by applying it to the actual transaction data. Lastly, we present a few examples to highlight the differences in the discount strategies compared to the traditional pricing model.

System Overview

The overview of the adaptive pricing system is given in Figure 1. The system takes sales data, customer data as well as clickthrough rate data to estimate the likelihood of a purchase using a discrete choice model which models purchase probability as a function of not only price and discount but also customers' perception of discounts. We also formulate a nonlinear optimization model that incorporates the output from the discrete choice model, along with business constraints such as the promotion budget and the duration of the promotion to determine the optimal discount that maximizes the objective. Lastly, this discount strategy is evaluated in a surprise model to measure its novelty with respect to the historical promotions.

Discrete choice model
Discrete choice models have been widely used to model consumer demand in marketing, transportation research, agricultural economics, etc. They statistically relate the choice made by a person to the attributes of the person and the attributes of all available choices. The underlying assumption is that every individual has a utility function which allows him to rank the alternatives in a consistent and unambiguous manner.

The most popular choice model out of the discrete choice model family is the multinomial logit (MNL) model. To derive this model, consider a consumer, labeled n, faces J alternatives. The utility that the consumer obtains from alternative j is decomposed into a part labeled V_{nj} that is known, and an unknown part ε_{nj} that is treated as random: $U_{nj} = V_{nj} + \varepsilon_{nj}$ for all j. Each ε_{nj} is assumed to be distributed independently, identically extreme value. The distribution is known as Gumbel, and its cumulative distribution is given by $F(\varepsilon_{nj}) = e^{-e^{-\varepsilon_{nj}}}$.

With some algebraic manipulation, it can be shown that the probability that person n choose alternative i can be expressed as a succinct, closed-form expression:

(1) $\qquad P_{ni} = \dfrac{e^{V_{ni}}}{\sum_j e^{V_{nj}}}.$

The deterministic part of the utility is usually specified to be linear in parameters: $V_{nj} = \beta^{\mathrm{T}} x_{nj}$ where x_{nj} is a vector of observed variables relating to alternative j, β is a vector of coefficients which can be the same for all alternatives or alternative-specific, and the symbol T represents transpose,

Despite its popularity, the logit-based model suffers from several limitations (Train [9]). The most notable aspect is its independence from irrelevant alternatives (IIA) property. A number of extensions (e.g., nested logit models and mixed logit models) have been proposed to relax some of the limitations by allowing correlation over alternatives and more general substitution patterns (Hausma and McFadden [10], Hensher and Greene [11]).

Perception of discount

To describe a decision process, which can be potentially influenced by emotions, we incorporate an attribute in the utility function of the discrete choice model to measure a customer's perception of discount, or "the thrill of a deal".

Mathematical psychology offers a quantitative approach to study human response to a stimulus. It comprises mathematical modeling of perceptual, cognitive and motor processes, and establishment of law-like rules that relate quantifiable stimulus characteristics to quantifiable behavior. Here, we use this mathematical approach with the goal of deriving hypotheses that are more exact and thus yield stricter empirical validations.

One of the most widely discussed concepts in mathematical psychology is the Weber-Fechner's law, which provides a functional relationship between a stimulus and behavior responses. It is stated as follows,

(2) $\qquad R = K \log \frac{S}{S_0},$

where

$\qquad$ R = the magnitude of perception for a stimulus S,
$\qquad$ K = the constant of proportionality,
$\qquad$ S = the magnitude of a stimulus, and
$\qquad$ S_0 = a stimulus threshold below which no change in response is detected.

The law states that the magnitude of human perception of a stimulus follows a logarithmic relationship to the magnitude of a given stimulus. This result has been validated for human perceptions of sight and sound, as well as numerical cognition. While its application in the pricing domain has been debated for several decades, many studies [12-17] indicate that there is ample evidence to support the plausibility of the Weber-Fechner's Law applying within a pricing context.

In our context, we denote R as customers' perception of the discount, the stimulus S as the discount, and S_0 as the discount threshold or the expected discount level. Note that when $S=S_0$, the thrill R is reduced to 0, i.e., customers do not get excited over sales that they have been used to (a phenomenon known as "promotion fatigue" in marketing). Meanwhile, the logarithmic relationship also implies diminishing return of discounts. This means that adding an additional 10% on top of an existing, say, 40% discount on a product is less noticeable than adding the same 10% discount to a 20% discount.

Surprise model

Surprise has been hailed as one of the most effective marketing tools to increase sales potential and improve customer satisfaction (Bagozzi et. al [18], Vanhamme [19]). Surprise is used to quantify the novelty of a deal according to a customer's belief based on past promotions.

We base our surprise computation with the information-theoretic framework proposed by Itti and Baldi [20]. Its mathematical formulation is given as follows: Denote the model describing the phenomenon observed as M and a prior belief (i.e., a prior probability distribution) on the model as $P(M)$. In our context, M is the set of discounts that have been historically offered and, hence,

is something that is known apriori to the customer. *P(M)* is the probability density function over this set based on the discount the customer expects to see in future. Upon receiving a measurement *D* (such as a new discount previously unobserved by the customer), the prior is updated to obtain a posterior belief on model space *P(M|D)* using Bayes law, $P(M|D) = \frac{P(D|M)}{P(D)} P(M)$.

Surprise is defined as the change in the beliefs upon observing the new observation *D*. It is measured by using the relative entropy or Kullback-Leibler (*KL*) divergence, which is defined as the expectation of the logarithmic difference between the posterior and the prior, where the expectation is taken using the posterior distribution *P(M |D)*:

(3) $S(D, M) = KL\big(P(M|D), P(M)\big) = \int_{M} P(M|D) \log \frac{P(M|D)}{P(M)} dM.$

It follows that if the posterior is the same as the prior, there is zero surprise. Conversely, the new discount observed, *D*, is surprising if the posterior belief resulting from observing *D* significantly differs from the prior belief.

In the retail context, a consumer who has been exposed to promotions in the past forms a prior belief on the sales *P(M)*, based on the magnitude and the frequency of the sales. When she receives a new deal *D*, she updates her belief on the promotions, *P(M|D)*. Measuring the difference between the two distributions reveals how surprising is the new deal *D* to the customer. Note that the notion of surprise in (2) only measures the difference between the two beliefs. It does not differentiate a pleasant surprise from an unpleasant shock. Take J.C Penny as an example, drastically eliminating promotions at a store which once relied on sales and coupons came as an unpleasant surprise which upset many of its customers. Thus, this metric can also be interpreted as measure on the risk of marketing strategy.

Illustrative solution: Tomorrow's Pricing Today

Overview
The demo, "Tomorrow's Pricing Today", is an illustration of the behavior pricing system by applying it to actual retail transaction data. It is one of the two demos from IBM Research that were showcased at the IBM booth at the National Retail Federation in January 2014. The interface for the demo is shown in Figure 2. During the demo, a user (e.g., a marketer who is planning the next promotion) first specifies a set of products to be included in the analysis. Next, he enters information related to the segment whom the promotion is targeting at. Lastly, he specifies constraints related to the promotions such as its duration and the total promotion budget. The system calibrates the discrete choice model with discount perception and the output (i.e., predicted probabilities with respect to price and discount) is fed into a nonlinear optimization model, along with the business constraints. The output strategy of the optimization model is then compared to the historical promotions in terms of Bayesian surprise and can be fine-tuned to suit the risk appetite of the user.

Data

We use the publicly available KDD (Knowledge Discovery and Data Mining)-Cup 2000 dataset, which contains three months of transaction data from an online legwear store, totaling about 3,465 orders, 4,540 transactions, and 1,831 customers (Kohavi et al. [21]). The newly launched store had run many promotions so as to gain market share. These promotions affect traffic to the site, the type of customers, their purchasing behavior, etc.

The dataset contains two categories of information: customer and order information. Customer information includes customer ID, registration information, registration form questionnaire responses, etc. Order information consists of order date and time, assortment ID, price, quantity, product category, discount, tax, shipping cost, etc.

The bestselling category in the dataset is labeled as "main brands". After some data pre-processing, we selected the ten products with the highest support in this category to be included in the analysis (more information on data pre-processing and the IDs for the ten products can be found in the Appendix).

Data pre-processing
To construct a discrete choice model, we define a choice as the purchase of a single product within the choice set by a customer. In the KDD-Cup dataset, when a transaction shows that *m* units of the same item were bought, we replicate that transaction to represent that *m* such choices were made.

The assortment IDs for the ten items with the highest support in the main category are 9093, 11659, 11667, 11859, 19859, 19913, 19921, 29725, 35887 and 35931.

Attributes of the choice model
We have discussed earlier that a discrete choice mode can be represented by Equation (1), where the deterministic part of the utility, V_{nj}, is expressed as linear combination of attributes. In the model, we consider $V_{nj} = \beta_j^{\mathrm{T}} x_{nj}$, where β_j is a vector of alternative-specific coefficients. This means that there will be a separate coefficient on each independent attribute for each alternative. In other words, the effect of the independent variables will vary across all of the choices.

With this specification, the choice probabilities can be written as

(4) $\qquad P_{ni} = \dfrac{e^{\beta_i x_{ni}}}{\sum_j e^{\beta_j x_{nj}}}.$

The first of the attributes in the utility function is the regular price of the item in the absence of promotions:

$\qquad$ *Price*$_{ni}$ = Regular price of item *i* at the time of customer *n*'s purchase.

In the KDD-Cup dataset, 84% of the orders used discounts. As the discount is recorded at the order level, we normalize it by the entire order amount prior to discount and shipping cost:

$Discount_{ni}$ = Discount of item i in percentage at the time of customer n's purchase.

Another attribute related to the discount is the thrill of the deal:

$Thrill_{ni}$ = Customer n's thrill (perception of discount) derived from item i.

The Weber-Fechner's law in Equation (2) provides a functional form that relates the perception to its stimulus. We rewrite (2) as $R = K \log S - K \log S_0$, where $K \log S_0$ is a constant which is unique to an alternative. We do not need to explicitly specify the discount threshold S_0 as this term is included in the intercept during the estimation.

We investigated several structural forms to model the stimulus S, e.g., discount in percentage, absolute savings in dollars, etc. We compared the performance of the resulting discrete choice models in terms of their prediction accuracies and selected

$$Thrill_{ni} = \log(100 \times Discount_{ni} + 1).$$

We also incorporate attributes that depend on the characteristics of the customer. While the data has customer information, missing entries limit the usefulness of many of these attributes. For example, customers often chose to skip some questions in the registration questionnaire. We select two customer-level attributes with sufficient support in this analysis. One of the attributes concerns the income level of a customer.

$$Income_n = \begin{cases} 1, \text{ if customer } n\text{'s annual household income exceeds \$55,000,} \\ 0, \text{ otherwise.} \end{cases}$$

The original data set specified 9 income levels. In view of the small data size, we aggregate the information and create a binary indicator, where the cutoff value approximates the median family income in 2000.

Another customer-level attribute is based on the customer's response to the question "How did you hear about us?" in the registration questionnaire. We aggregate the responses and define a categorical variable which indicates one of the four channels through which a user was acquired.

$$Channel_n = \begin{cases} 1, & \text{Friends/family} \\ 2, & \text{Email marketing} \\ 3, & \text{Direct mail, print ad} \\ 4, & \text{Others (including missing entires)} \end{cases}$$

Mathematically, we represent the deterministic component of the utility function, V_{ni}, as the following,

(4) $V_{ni} = \beta_{0i} + \beta_{1i}Price_{ni} + \beta_{2i}Discount_{ni} + \beta_{3i}Thrill_{ni} + \beta_{4i}Income_n + \beta_{4i}Income_n + \beta_{5i}(Channel_n = 2) + \beta_{6i}(Channel_n = 3) + \beta_{7i}(Channel_n = 4).$

Calibration and accuracy
We calibrate the model with multinomial logit regression which uses maximum likelihood estimation. A sample regression output for an assortment of three products (product ID 9093, 11659 and 11859) is shown in Table 1. Note that product 11859 is used as the reference product in the regression, i.e., its coefficients are 0. The signs for coefficients on price, discount and thrill are expected, i.e., demand decreases with price (negative sign) and increases with discount and the thrill (positive sign). Coefficients on price and perception of discount are statistically significant at 5% and 1% level respectively.

We evaluated the predictive performance of multinomial logit model using 5-fold cross validation, by fitting the model to 4 folds of the data and then evaluating the likelihood on the remaining fold. While multinomial logistic regression does compute correlation measures to estimate the strength of the relationship (pseudo R square measures), these correlations measures generally do not indicate much about the accuracy or errors associated with the model. A more useful measure to assess the accuracy is classification accuracy, which compares predicted choice in terms of purchase product based on the predicted probabilities of the logistic model to the actual choice.

A benchmark to characterize a multinomial logistic regression model as useful is a 25% improvement over the rate of accuracy achievable by chance alone [22, 23]. The accuracy rate by chance alone has two definitions, depending on different applications: namely, the proportional by chance accuracy rate and the maximum by chance accuracy rate.

The classification matrix for this assortment of three products is shown in Table 2. The proportional by chance accuracy rate was computed by summing the squared proportion of each alternative in the sample, i.e., $0.423^2 + 0.28^2 + 0.297^2 = 34.6\%$. In order to have a 25% improvement, the criteria on proportional by chance accuracy is $1.25(34.6\%) = 43.2\%$. Our model achieves an accuracy of 73.2%, thus satisfies the criterion. Meanwhile, the maximum by chance accuracy rate, which refers to the size proportion of the product with the largest population, was 42.3% as shown in Table 2. A 25% improvement corresponds to an accuracy of 52.9%. Our model also satisfies this criterion.

Optimization and adjustment with surprise
Given an output of the discrete choice model which predicts the choice probabilities as functions of the attributes, the expected profit can be computed. Maximizing the expected profit (or revenue) with respect to the discount yields the optimal discount, or maximizing over the discount and the product prices simultaneously yields a complete pricing strategy. While the objective function is not concave in general, a local maximum can be found using standard numerical optimization techniques.

Once an optimal discount is identified, a user can compare it to the historical promotions to evaluate the surprise metric of this strategy. To do so, we first obtain the prior distribution on promotions by constructing a histogram of discount from the sales data. We then augment this distribution with the discount strategy determined by the optimization model according to the promotion duration and sales frequency. We quantify the surprise metric as the KL divergence between the two histograms according to Equation (3).

Figure 3 illustrates how surprise and the expected profit are related to discounts for the assortment trio (9093, 11659 and 11859). In the KDD-Cup dataset, a significant number of orders became "free" after discounts as the store deployed several aggressive promotions. Meanwhile, 26% of orders did not use discounts. Figure 3 indicates that 100% and 0% discounts are among the least surprising strategies. As noted earlier, surprise is affected by the typical discounts that a customer has historically observed. Since the KDD-Cup dataset contained several instances of products being given out as a free addition with a purchase of another product, a 100% discount was identified as being least surprising.

Comparison with the baseline
To illustrate the results of the behavior pricing model, we report the predicted probabilities and the expected profit under two product assortment scenarios shown in Figure 4 and 5 respectively. Under both assortment scenarios, when we compute the expected profit, we focus on the population with an annual income level below $55K and were acquired through Channel 1 (i.e., friends/family). We have also included the corresponding output from a baseline model which does not incorporate the psychology components (e.g., thrill and surprise) to highlight the difference between the resulting discount strategies.

Figure 4 shows for an assortment trio of products (9093, 11659 and11859), the behavior pricing model suggests 14% as the optimal discount as opposed to 25% given by the baseline model. To gain some intuition, in the graphs of the predicted probabilities, we observe that changes in the perception are more drastic for low discount level (<20%) and become more gradual as discount increases. The phenomenon reflects the diminishing return on discount suggested by the Weber-Fechner's law. On the other hand, for the base model, the rate of change in predicted purchase probabilities is steady across all discount levels. Thus, the behavior model suggests that a low discount level is sufficient to capture "the thrill of deal" and maintain profitability.

Consider another assortment of three products (9093, 11659 and 19913, note that only one product from the trio is replaced) shown in Figure 5. The baseline model recommends no discount while the behavior model advocates a small discount about 8%. In this scenario, the behavior model suggests that a smaller discount is better than none, as a small discount induces higher aggregate profitability. Another observation is that the purchase probabilities for products 11659 and 19913 decrease with an increase in discount. This could mean that since the three products are substitutable and since discount can be applied to any product in the portfolio, customers tend to prefer 9093 (explained by an increase in purchase probability with increase in discount).

Discussion and Conclusion

We proposed a framework that systematically incorporates the effect of discount and historical promotions on consumers' valuations and helps marketers determine the optimal discount strategy. We tested our framework on a publicly available data set from an online retailer. Our

experiments revealed that the behavior pricing model can lead to very different pricing decisions compared to the traditional pricing model.

To date, we have collaborated with a national retail chain and conducted a case study on behavior pricing, based on two years of sales data from both online and brick-and-mortar stores. We focused on one product category which consists of approximately a thousand products and had frequent sales. As the data contains aggregate sales information (i.e., weekly sales per store) as opposed to transactions by individuals, we modified the discrete choice model and the estimation procedure (we refer the reader to [24] for more information on discrete choice model estimation with aggregate data).

The analysis validated the behavior pricing model and showed observations that are consistent with our earlier findings. In particular, for products that are sensitive to sales, the behavior pricing model tends to suggest actions that are quite different from the baseline model. For example, for a subset of products, we show that when they are advertised on store circulars, all things being equal, the behavior model will recommend a higher discount than the baseline model. On the other hand, when they are not advertised, the behavior model recommends a discount which is about 50% lower than the baseline model. The observation implies that the exposure from being featured in the advertisement not only affects the purchase decision, it also influences customers' perception of promotions.

The study generated a lot of interests and discussions from the retailer. For example, some executives were concerned that the discount suggested by the behavior model could be too high and would adversely affect their profit margins. We recommended the retailer to start experimenting with product categories where the behavior model suggests lower discounts than their existing system, as one way to mitigate the risk. Some feedbacks from the retailer also provided us with directions for future work. For instance, one major concern is the risk of a downward spiral of high discounts to induce surprise. Although the model optimizes the discount with the goal to maximize the objective as opposed to maximize surprise, it is well-known that frequent promotions could "train" customers to hold back purchases as they anticipate sales in the future. Therefore, a multi-period dynamic model with behavior pricing could be more appropriate as the model yields the current period pricing strategy by combining information from the past and the updated demand prediction about the future.

References

[1] Nielson case study. "Global Ad Spend Grows 3.2% in 2012". Available: http://nielsen.com/us/en/newswire/2013/global-ad-spend-grows-3.2-percent-in-2012.html
[2] M. Levy, D. Grewal, P. K. Kopalle and J. Hess. "Emerging trends in retail pricing practice: Implications for research", *Journal of Retailing*, 80(3), pp. 1151 – 1165, 2004.
[3] R. S. Winer. "A reference price model of demand for frequently purchased products", *Journal of Consumer Research*, 13, pp. 250 – 256, 1986.
[4] S. Clifford and C. Rampell, "Sometimes, we want prices to fool us", The New York Times, Available: http://www.nytimes.com/2013/04/14/business/for-penney-a-tough-lesson-in-shopper-psychology.html?pagewanted=all&_r=0

[5] D. Ackerberg. "Advertising, learning, and consumer choice in experienced good markets: An empirical examination", *International Economics Review*, 44, pp. 1007 – 1040, 2003.

[6] N. Mehta, S. Rajiv and K. Srinivasan. "Price uncertainty and consumer search: a structural model of consideration set formulation", *Marketing Science*, 22, pp. 58 – 84, 2003.

[7] J. –P. Dube, G. Hitsch, P. E. Rossi and M. Vittorino, "Category pricing with state-dependent utility", *Marketing Science*, 27, pp. 417 – 429, 2008.

[8] S. Shin, S. Misra and D. Horsky. "Disentangling preferences and learning in brand choice models", *Marketing Science*, 31, pp. 115 – 137, 2012.

[9] K. E. Train. 2007, *Discrete Choice Models with Simulation*. New York: Cambridge University Press, 2007.

[10] J. Hausman, J., and D. McFadden, "Specification tests for the multinomial logit model," *Econometrica: Journal of the Econometric Society*, pp. 1219-1240, 1984.

[11] D. A. Hensher, and W. H. Greene, "The mixed logit model: the state of practice," *Transportation,* vol. 30, no. 2, pp. 133-176, 2003.

[12] P. Fouilhi, "The Subjective Evaluation of Price: Methodological Aspects," *Pricing Strategy*. Princeton, N.J.: Brandon/Systems Press, 1970.

[13] J. M. Kamen, and R. J. Toman,"Psychophysics of Prices," *Journal of Marketing Research*, vol. 7, no. 2, pp. 27-35, 1970.

[14] K. Monroe, "Psychophysics of Prices: A Reappraisal," *Journal of Marketing Research*, vol. 8, no. 5, pp. 248-250, 1971.

[15] A. Gabor, C. Granger, and A. Sowter, "Comments on 'Psychophysics of Prices, " *Journal of Marketing Research*, vol. 8, no. 5, pp. 251-252, 1971.

[16] S. Dehaene, and J. F. Marques, "Cognitive Euroscience: scalar variability in price estimation and the cognitive consequences of switching to the Euro," *Quarterly Journal of Experimental Psychology*, vol. 55, no. 3, pp. 705–731, 2002.

[17] M. Chang, and W. Chiou, "Psychophysical methods in study of consumers' perceived price change for food products," *Psychological Reports*, vol. 100, no. 2, pp. 643–652, 2007.

[18] R. P. Bagozzi, M. Gopinath, and P. U. Nyer, "The role of emotions in marketing," *Journal of the Academy of Marketing Science*, vol. 27, no. 2, pp. 184-206, 1999.

[19] J. Vanhamme, "The link between surprise and satisfaction: an exploratory research on how best to measure surprise," *Journal of Marketing Management,* vol.16, no. 6, pp. 565-582, 2000.

[20] L. Itti, and P. F. Baldi, "Bayesian surprise attracts human attention." *Advances in neural information processing systems*. 2005.

[21] R. Kohavi, C. E. Brodley, B. Frasca, L. Mason, and Z. Zheng, "KDD-Cup 2000 organizers' report: peeling the onion," *SIGKDD Explorations,* vol. 2, no. 2, pp. 86-98, 2000.

[22] D. W Hosmer Jr, S. Lemeshow, and R. X. Sturdivant. *Applied Logistic Regression*. Wiley. 2013.

[23] A. BAYAGA. "Multinomial logistic regression: usage and application in risk analysis." *Journal of applied quantitative methods*, vol. 288, 2000.

[24] L. G. Cooper and M. Nakanishi. *Market-share analysis: Evaluating competitive marketing effectiveness*. Vol. 1. Springer, 1988.
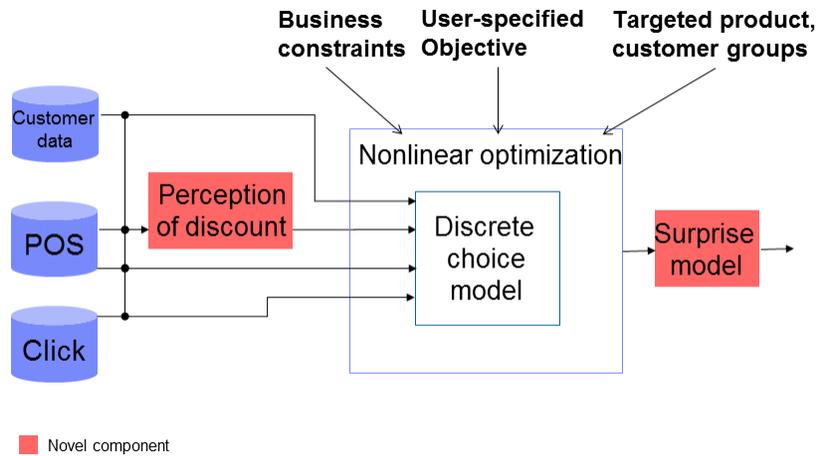
Appendix

<u>**Figures and Tables**</u>

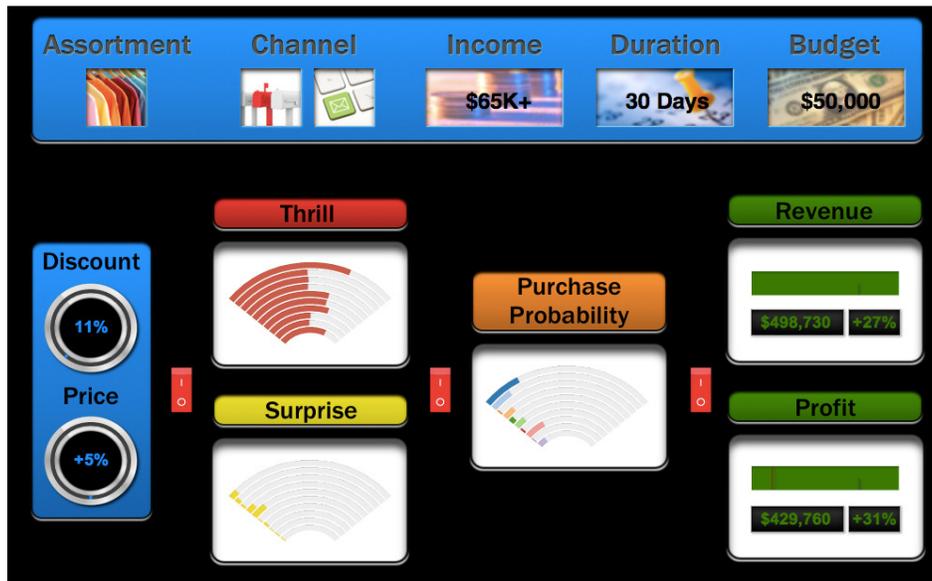Figure 1: Schematic view of the behavior pricing system

Figure 2: Interface for the illustrative solution "Tomorrow's pricing today"
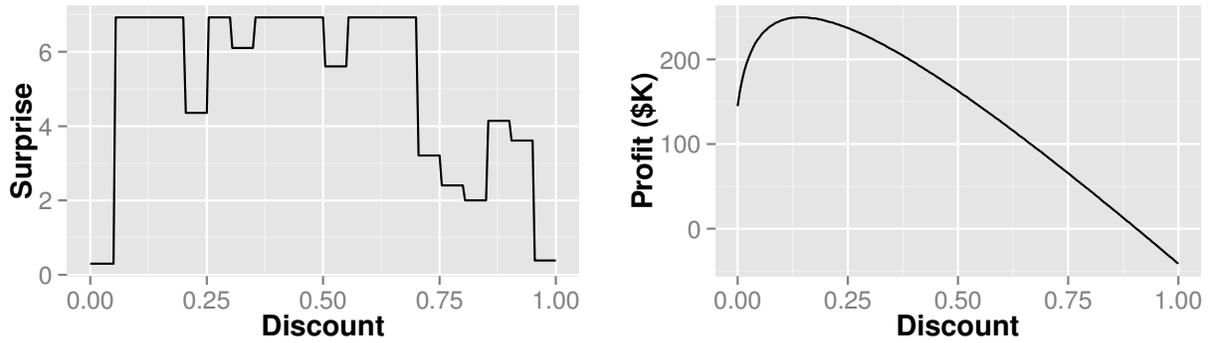
Figure 3: Surprise and the expected profit with respect to discount for an assortment scenario (Product 9093, 11659 and 11859)
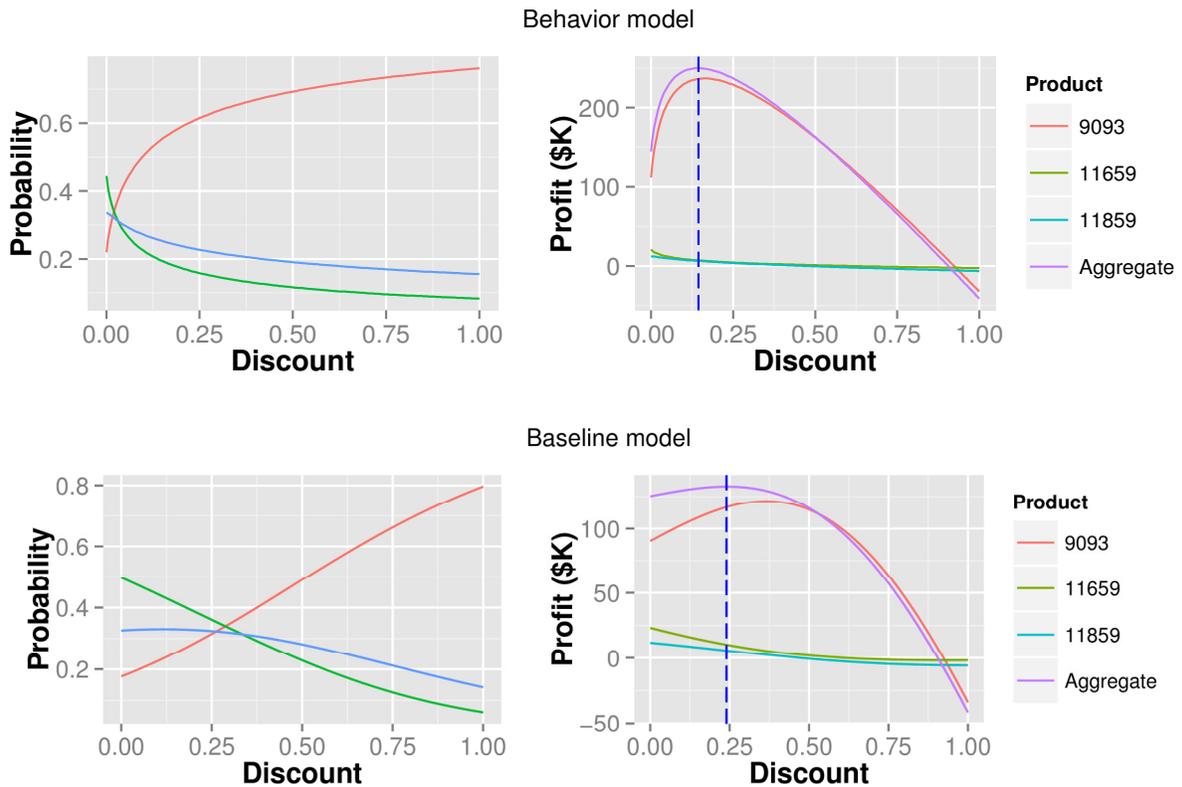

Figure 4: A comparison between the behavior model and the baseline model for an assortment scenario (Product 9093, 11659 and 11859). The dashed line represents the optimal discount that maximizes the expected profit for the given model.
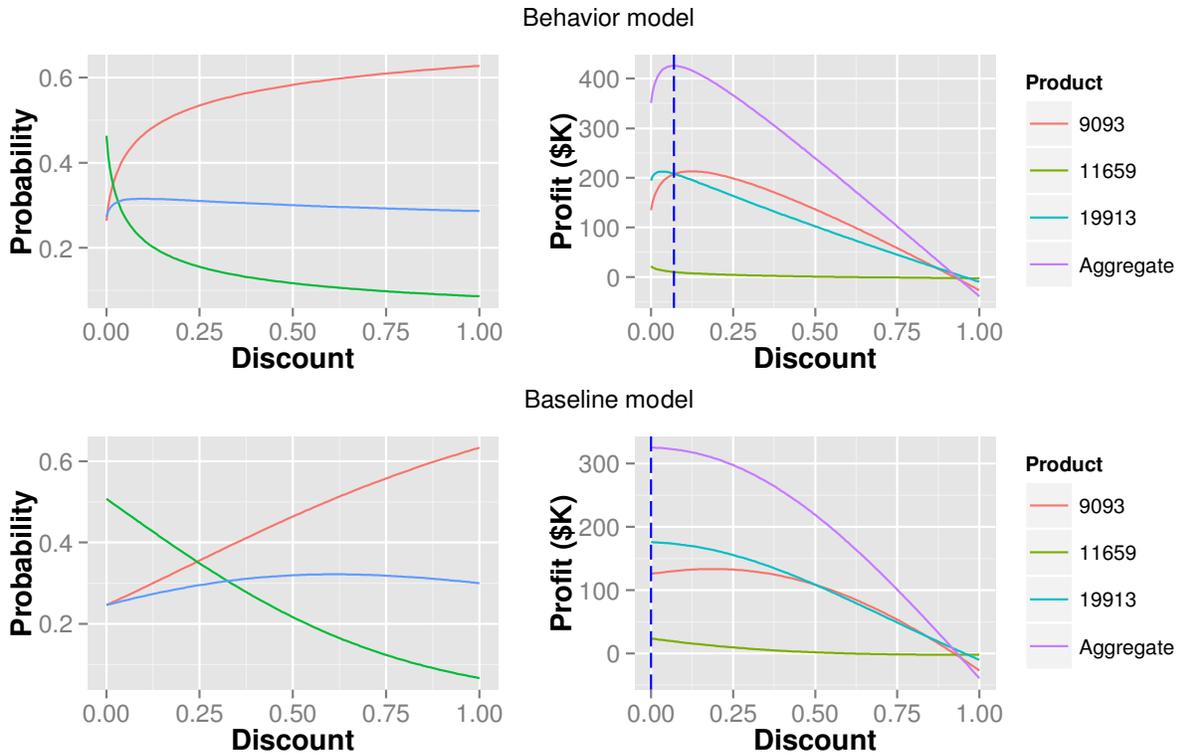
Figure 5: A comparison between the behavior model and the baseline model for an assortment scenario (Product 9093, 11659 and 19913). The dashed line represents the optimal discount that maximizes the expected profit for the given model.

	Intercept	Price	Discount	Thrill
ID: 9093	-1.315**	-0.076**	15.027***	2.358***
	(0.785)	(0.091)	(2.507)	(0.497)
ID: 11659	-1.419**	-0.143*	8.435***	1.294***
	(0.699)	(0.083)	(2.174)	(0.447)

	Income = T	Channel = 2	Channel = 3	Channel = 4
ID: 9093	1.193***	0.534	0.537	-0.823**
	(0.320)	(0.488)	(0.600)	(0.428)
ID: 11659	0.071	0.049	0.419	0.629*
	(0.276)	(0.491)	(0.616)	(0.353)

Table 1: Regression coefficients for the multinomial logit model for products 9093, 11659 and 11859. Standard errors are reported in parentheses. *, **, *** indicates significance at the 90%, 95%, and 99% level, respectively.

Observed	Predicted			
	9093	11659	11859	Percent correct
9093	136	24	17	76.8%
11659	19	90	8	76.9%
11859	25	19	80	64.5%
Overall percentage	42.3%	28.0%	29.7%	73.2%

Table 2: Classification matrix for products 9093, 11659 and 11859 based on the calibrated multinomial logit regression model.

Biographical sketches

Wei Sun *IBM Research Division, Thomas J. Watson Research Center, P.O. Box 218, Yorktown Heights, New York 10598 (sunw@us.ibm.com).* Dr. Sun is a Research Staff Member in the Industry Solutions department at the Thomas J. Watson Research Center. She received her Ph.D. in Operations Research from Massachusetts Institute of Technology (MIT) in 2012. She holds a M.S. degree in Computational Design and Optimization from MIT and a B.Eng. in Electrical and Computer Engineering from National University of Singapore. Her paper on congestion pricing for service industries was awarded Best Student Paper at INFORMS (Institute for Operations Research and Management Sciences) - Service Science in 2011. Since Dr. Sun joined IBM in 2012, she has used optimization, game theory and machine learning theories to achieve process improvement in areas such as commerce and human resources.

Pavankumar Murali *IBM Research Division, Thomas J. Watson Research Center, P.O. Box 218, Yorktown Heights, New York 10598 (pavanm@us.ibm.com).* Dr. Murali is a Research Staff Member in the Industry Solutions department at the IBM T J Watson Research Center. Pavan received his Ph.D. in Operations Research from the University of Southern California (USC) in 2010 and a Bachelors in Mechanical Engineering from the Indian Institute of Technology, Madras. His research expertise lies in the areas of mathematical optimization, predictive analytics and data mining. His current research involves applying these techniques to problems in areas such as marketing and service science, for which he has received IBM Research Division Awards.

Anshul Sheopuri *IBM Research Division, Thomas J. Watson Research Center, P.O. Box 218, Yorktown Heights, New York 10598 (sheopuri@us.ibm.com).* Dr. Sheopuri is a manager of a team of researchers in Industry Solutions Research Department at the IBM T J Watson Research Center. He is passionate about creating innovative customer experience analytics in partnership with the C-suite of clients. Dr. Sheopuri was featured in Fortune CNN as IBM's Face of the Future and his work has been highlighted in an IBM investor briefing. He is the Research Relationship Manager for IBM Global Business Services Human Resources and co-lead of the Customer Insight and Marketing World Wide sub-strategy. For his leadership and contributions in these areas leading to measurable financial impact, he has received the IBM Corporate Technical Award, an Outstanding Innovation Award and three Outstanding Technical Achievement Awards. His work has been accepted or published in *Operations Research*, *Management Science*, *European Journal of Operations Research* and *Interfaces*. He has served as an Adjunct Assistant Professor with New York University's Leonard N. Stern School of Business and a Guest Professor at the McCombs School of Business at the University of Texas at Austin. He received his Ph.D. in Operations Management from New York University's Leonard N. Stern School of Business and a B.Tech. in Mechanical Engineering from the Indian Institute of Technology, Madras.

Yi-Min Chee *IBM Research Division, Thomas J. Watson Research Center, P.O. Box 218, Yorktown Heights, New York 10598 (ymchee@us.ibm.com).* Mr. Chee is a Senior Technical Staff Member in the Industries & Solutions department at the IBM TJ Watson Research Center. His current research interests include the areas of services computing, collective intelligence, and tools and environments for social collaboration, service delivery, and software development. Since joining IBM Research in 1991, Mr. Chee has worked in a variety of areas, ranging from incremental compilers and programming environments for C++, to interfaces and standards for pen-based computing, high-performance computing applications for game processors, and design tools and delivery environments for software architects and consultants. He has contributed to a number of IBM products, and has received an IBM Corporate Award, a Best of IBM Award, and several Outstanding Technical Achievement Awards for his work. He received his bachelor's degree in Electrical Engineering & Computer Science from the Massachusetts Institute of Technology (MIT) and Master's degree in Computer Science from Columbia University.