A data driven framework for early prediction of customer response to promotions

Full papers

Utkarsh Shrivastava
Muma College of Business
University of South Florida
utkarsh1@mail.usf.edu

Wolfgang Jank
Muma College of Business
University of South Florida
wjank@usf.edu

Abstract

The study proposes a framework based on functional analysis of transaction data to predict customer spending during promotional events. Retailers face the challenge of considering an extant number of variables and of accounting for methodological constraints while modeling customer response to promotions. Noise and uneven distribution of spending further introduces error into traditional models' estimates. We represent each customer's spending as a continuous curve that accounts for heterogeneity due to the purchase cycle as well as cross sectional differences. In order to obtain an optimal functional representation we utilize a data driven iterative procedure. Statistical information from the collection of spending curves is drawn using functional data analysis (FDA). Analysis of a real customer dataset from a North American retail chain shows that the dynamics of information captured by optimal functional representation of transaction data significantly improves predictions for out of sample observations. Framework provides a guideline for modeling fast moving, heterogeneous and noisy customer related data to assist managerial decision making during specialized events such as promotions.

Keywords

Retail promotions, customer spending, functional data analysis, and functional principle component analysis.

Introduction

Retailers invest a substantial amount of their budget on promotions and often end up losing billions in potential revenues because of out-of-stock items (Jing and Lewis, 2011). In 2014 alone marketers in the United States spent around $179 billion in marketing activities (Cmocouncil.org, 2014). Promotional events attract customers to the store or web portal but their success is a function of the retailer's ability to influence quantity and variety of purchases at the store. Out of stock items, have both short term (revenue loss) and long-term (store switching) implications for the retailer (Jing and Lewis 2011). Expected customer spending during a promotional event is central to in-store planning and optimal distribution of resource across spatially located outlets / warehouses of a retail chain. The perishable nature of the products and shorter life cycle makes some industries such as the grocery industry more vulnerable to sales miscalculation during such events.

Predicting customer spending during promotions is challenging for retailers. Around 60% of customer spending is unplanned (Inman and Winer 1999) and during promotions customers show a significant change in their spending pattern (Neslin et al. 1985; Venkatesan and Farris 2012). Product availability, customer demographics, and store layout are some of the factors that influence in-store spending (Lam et al., 2001). Customers also consider product inventory at home (for perishable products) and budget constraints while making a purchase decision (Dreze et al. 2004). As an example, if the level of inventory at home for a customer is low and the quota of monthly spending is not exhausted prior to the promotional event, then higher spending by that customer can be expected. Customers are also heterogeneous in their expectations and preferences for promotions. They often manipulate their
spending prior to the promotion in order to purchase their favorite products during promotion at a lesser price. Retailers face the challenge of accounting for an extant number of factors and considering methodological constraints while modeling customer response to promotions.

Data obtained from technologies such as scanners, shopper cards, RFIDs and distributed transactional databases provide means to evaluate and strategize for promotions. This data is usually a combination of longitudinal (purchase history) as well as cross sectional (customer attributes) information. Statistical models that operate on customer demographics and short-term purchase history cannot accommodate all temporal information that is present in this rich data (Jank and Shmueli 2006). They are also not efficient enough to represent and analyze high dimensional, fast moving longitudinal data. In addition, parametric factor models (regression based) or stochastic process models (time series models), may not be suitable for modeling customer spending because of the temporal nature of heterogeneity in customers’ preferences and the relationship with the store over time. Noise and uneven distribution of customer purchases also introduces error into traditional models’ estimates. How can retailers parsimoniously represent and incorporate most of the temporal and cross sectional information in the customer purchase data to predict promotional spending?

Drawing from prior predictive studies in business and IS that modeled similar data structure in auctions (Jank and Zhang, 2011), product penetration (Sood et al. 2009) and the virtual stock market (Foutz and Jank, 2010) we model each customer’s spending with the retailer as a continuous process (functional curve). We obtain a functional curve for each customer that represents the evolution of her relationship with the retailer and accounts for heterogeneity and noise in the spending. The intrinsic process that governs spending at stores is constrained by inventory and budget limitations that are specific to each household and leads to certain spending patterns, providing an opportunity to learn from them. Promotions act as an external stimulus that tries to disturb the dynamics (rate of change) of the spending process. We relate the process and dynamics information of each customer to predict their spending during promotional events.

In this research context, functional data analysis (FDA) is appropriate for drawing statistical information from the collection of spending curves using multivariate analysis techniques. FDA is a branch of statistic that deals with the analysis of continuously varying information represented as curves or shapes (Ramsay and Silverman 2005). Representing spending as functional curve allows the estimation of dynamics through derivatives and the incorporation of majority of temporal information in data using dimension reduction techniques. FDA is advantageous than other statistical methods in analyzing customer data because a) it accounts for noise and disparity in the observations b) provides reliable estimates of parameters for statistical modeling through smooth functions c) makes no parametric assumptions, and d) incorporates majority of temporal information.

The central assumption of the FDA is the existence of a differentiable smooth function that generates observed data such as discrete purchase history in the study’s context. Smoothness property should effectively apply to the data to utilize the advantages of the FDA over multivariate analysis. We demonstrate that suitable representation of spending data can help in generating representative smoothed functions. We also address the issue of over fitting or under fitting the smoothed functional form by calibrating the smoothness parameter using an iterative algorithm. Finally, the study shows that a customer’s spending dynamics can be useful predictor of its future spending. In summary, the study presents a data driven framework for early prediction of expected spending of a customer during a promotional event. We test the framework using customer data from a promotional campaign organized by a North American retail chain. Results show that including dynamics into the model significantly improves prediction.

**Literature Review**

Technologies like shopper cards accumulate huge amount of customer transaction data every day. This data drives customer analytics initiatives to address organizational issues such as inventory distribution across stores, revenue or sales projection, and promotion planning (Lam et al. 2001). Customer demographics, psychographic variables, credit histories and purchase pattern helps in building sophisticated models to understand customer behavior. However, past studies have shown that there is significant difference in customer behavior during promotions (Venkatesan and Farris 2012; Dreze et al. 2004). Customers often accelerate their purchases during promotion because: a) promotions have an
Data driven framework for predicting customer response to promotions

expiration date and customer tend to shorten their inter purchase time b) features associated to price promotions, and c) uncertainty in availability of future discounts. Customized coupons being unexpected rewards can increase purchase in other categories through halo effect. Promotions are more dynamic and flexible in online retail. Information related to regency, frequency and monetary value of past purchases can be useful for understanding customer response to promotions (Cui et al. 2006; Van Diepen et al. 2009).

Studies that categorized customers based on spending during promotions predominantly used statistical models such as logistic regression, discriminant analysis, or regression trees (Berger and Magliozzni 1992; Rao and Steckel, 1995; Cui et al. 2006). The objective of these classification studies was to predict if the customer spending during promotions would cross a threshold or not. Timing of purchase or the response to a promotion was predicted using hierarchical bayes models, bayesian networks or hidden markov model (Allenby et. al. 1999; Cui et al. 2006; Netzer et al. 2008). This class of models are appropriate for predicting probability of arrival, coupon redemption, or discrete state of customer during the promotion. However, variable selection, prior distribution assumption, and endogeniety bias problems can complicate the estimation and prediction with these models. Most importantly, predictors used in these models are obtained by pooling or aggregating a subset of temporal information in the purchase data. Similarly, time series stochastic models are a class of process model considered as the natural choice for forecasting variables such as sales, revenues or spending (Huang et al. 2014; Ali et al. 2009). These models too consider only a subset of past information and cannot incorporate dynamics of spending.

Recent advances in data generation and storage capabilities are leading to a variety of new data structures as an example, data related to three-dimensional motion of an object, digitalized images, and click stream in e-commerce. FDA is a tool set that generalizes the ideas of statistics for the data structured in the form of curves, shapes or functional observations (Jank and Shmueli 2006). FDA visualizes data in functional form and the object of analysis is functional in nature. Recent studies in FDA generalized statistical methods like principle component analysis (FPCA), clustering, ANOVA, and regression for functional observations (James 2000; James and Sugar 2003; Guo, 2002; Ratcliffe et al. 2002). Functional approach to data analysis predicted product penetration better than traditional Bass models (Sood et al. 2009). Predictions were 55% better than traditional OLS in some studies (Ullah and Finch, 2010). Data evolution curves allow capturing information that might be ignored by other factor models. FDA treats all observations as one functional curve and makes no parametric assumptions about time effects.

Two critical steps in data representation using FDA approach are 1) functional curve / objects generation from the raw discrete data 2) exploratory statistical analysis of functional curves hence obtained. Roughness penalty or regularization methods based on optimizing the fitting criteria are most versatile techniques to represent raw data in the functional form (Ramsay and Silverman 2005). Choice of fitting criteria is subjective and data dependent. This flexibility comes at the expense of over-fitting or under-fitting the functional form that introduces bias into predictions. Parsimonious representation of temporal information in curves is accomplished by dimension reduction techniques such as functional principle component analysis (FPCA).

Problem Definition

Earlier sections describe the need and challenges related to the prediction of a customer’s spending during promotional events. The problem becomes more intriguing when the retailer targets new customers with a promotion. Suppose the retailer has targeted a set of customers \( C_1 \) for promotion \( P_1 \) from time \( t_1 \) to \( t_1 + n \) where \( n \) is the duration of the promotion. At some time, say \( t_2 \) later a new set of customers \( C_2 \) are targeted with similar promotion \( P_2 \) for the first time. Given the purchase history for customers in set \( C_1 \) during and before the promotion \( P_1 \), we utilize the spending information of the new set of customers \( C_2 \) to predict their spending in promotion \( P_2 \). Figure (1) and Figure (2) bellow show the spending of a customer leading to a 8-week promotion. Visualizing weekly spending as cumulative spending (Figure 2) generates functional smoothed curve that better represents raw spending data. Sections below demonstrate how the information in the shape of the spending curve in Figure (2) is incorporated into the predictive model to obtain spending ‘\( Y \)’ for each customer during the promotion.
Data driven framework for predicting customer response to promotions


Predictive Framework

Suppose for the promotion \( P_i \) the retailer has the transaction data on the total spending \( Y_{it} \) during the promotion and the discrete spending vector \( X_{it} \) for a customer \( i \) up to time \( t_n \) before the promotion. \( X_{it} \) is measured every time the customer makes purchase at the store. As store visits are unplanned, \( X_{it} \) can be sparse for some range of \( t \) and dense during some others. The objective is to utilize the spending history \( X_{it} \) of the customer set \( C_1 \) up to time \( t_n \) before the start of the promotion \( P_1 \) and relate \( Y_{it} \) and dynamics of \( X_{it} \). In this study by dynamics, we refer to rate of change of cumulative spending functional curve \( F(t) \) or its first order derivative \( F'(t) \) (velocity) and second derivative \( F''(t) \) (acceleration). We model the total spending \( \sum X_{it} \) of a customer \( i \) as:

\[
Y_{it} = f(F(t), F'(t), F''(t))
\]

(1)

Step 1: Cumulative Representation: In the first step, raw daily purchase data \( X_{it} \) of customers in the set \( C_1 \) is represented as cumulative spending \( S_{it} \). Such that for any time \(-t \) before the promotion

\[
S_{it} = \sum_{j=-t}^{0} X_{ij}
\]

(2)

where, \(-t \leq t \leq -1 \) for all \( i \) in the set \( C_1 \). Figure (1) and (2) above shows \( X_{it} \) and \( S_{it} \) respectively for a customer \( i \) for \( t = [-30, -1] \) i.e discrete spending 30 days before the start of the campaign. Operating with cumulative spending results in monotonic functional curves that represent the variability in the data points much better figure (2).

Step 2: Generate functional curves: In this step we obtain a function \( F(t) \) for each customer \( i \) in \( C_1 \) that is giving rise to the observed data \( S_{it} \). For the FDA to be useful and capture most of the information the function \( F(t) \) should be such that adjacent values in \( S_{it} \) are linked together and are not much different from each other. \( F(t) \) can be generated using basis expansion (Fourier basis), kernel smoothing, polynomial fitting techniques etc. These techniques offer continuous control of the smoothness in the approximation but are not the optimal solution to the problem. On the contrary, the roughness penalty and other regularization methods optimize a fitting criterion and the extent of smoothing can be determined. The spline method estimates the smoothed function / curve \( F(t) \) such that \( S_{it} = F(t) + \varepsilon_i \) by making two conflicting goals in the function estimation. First goal is to minimize the residual sum of squares \( \sum_{t}^{n} |S_{it} - F(t)|^2 \) and second goal is to avoid over fitting or high local variability. An unbiased estimate of \( F(t) \) will exactly fit the observed values and have high variance which is attributed to rapid local variations in the curve. By imposing smoothness penalty in the curve, we effectively are borrowing information from the neighboring data values. This local pooling of information makes function more stable at the expense of some increase in the bias. The procedure penalizes roughness of...
the functional curve. A measure of the function’s roughness \( R \) is given by the integral of second order derivative:
\[
R = \int |F''(t)|^2 dt
\]

(3)

Highly variable functions will have higher values of \( R \). To penalize the over-fit or roughness, the fitting criteria for smoothing spline with roughness penalty is to minimize (4) for a given value of \( \lambda \)
\[
\sum_{i=1}^{n} (S_i - F_i(t))^2 + \lambda \int |F''(t)|^2 dt
\]

(4)

The choice of \( \lambda \) is subjective to the researcher often depends on the variability in the data. As \( \lambda \) approaches \( \infty \) the curvature is heavily penalized and \( F_i(t) \) is equivalent to a straight line and as \( \lambda \) approaches 0 the function is a unbiased fit to data. In this step firstly, we arbitrary select \( \lambda \) as close to 0 and subsequently keep on increasing it until the framework objective criterion is satisfied. Step (5) discusses the procedure in detail.

Step 3: Feature Extraction: After obtaining twice-differentiable functional curve \( F_i(t) \) for each customer \( i \) for an initial value of \( \lambda \) in the earlier step, we used functional principal component analysis (FPCA) to project curves \( F_i(t), F_i'(t) \) and \( F_i''(t) \) along the dimensions that are orthogonal and represent most variability across curves. FPCA is the generalization of PCA from multivariate statistics. It represents most of the information related to the purchase history. Generally, top \( n \) (\( n < t_n \)) principle components (PC) will able to explain the most of the variation across \( F_i(t), F_i'(t) \) and \( F_i''(t) \). These PC’s would select the most crucial features of the spending data \( S_{it} \). Temporal information in the spending path \( F_i(t) \), its velocity \( (F_i'(t)) \) and acceleration \( (F_i''(t)) \) for each customer \( i \) is represented as the vector of the best \( n \) principle component scores \( P \cdot PCS_1 - P \cdot PCS_n, V \cdot PCS_1 - V \cdot PCS_n \) and \( A \cdot PCS_1 - A \cdot PCS_n \) respectively. These scores are cross products of functional observations and the corresponding principle component. The choice of \( n \) is dependent upon the variations explained by various PCs. In our dataset, top five principle components were able to explain more than 90% of variation across all observations. Please refer (Foutz and Jank 2010; Ramsay and Silverman 2005) for more details on this procedure.

Step 4: Model Building: From the equation (1) we have,
\[
Y_{it} = f(P \cdot PCS_1 - P \cdot PCS_n, V \cdot PCS_1 - V \cdot PCS_n, A \cdot PCS_1 - A \cdot PCS_n)
\]

(5)

\( Y_{it} \) is modeled as the linear combination of PCS of three curves obtained in the earlier step consistent with (Foutz and Jank, 2010). We select the scores that contribute significantly to the equation (5). Stepwise regression procedure with AIC as model selection criteria is used as the variable selection procedure for equation (5). Next, we validate the model (5) using k fold cross validation on the training set \( C_1 \) of the customers to get Mean Squared Errors (MSE) over all folds associated to roughness penalty \( \lambda \) chosen in step 2. The validation procedure randomly divides the set \( C_1 \) in \( k \) equal samples and uses \( k-1 \) samples to train the model and the remaining 1 sample to validate the model using MSE.

Step 5: Optimal roughness penalty selection: In this step, a new value of the roughness penalty \( \lambda \) obtained by incrementing it with a parameter \( p \) is selected and steps 2-4 are repeated again to obtain a new MSE\(_{\lambda} \) until a local minima for MSE\(_{\lambda} \) is obtained. The \( \lambda \) for which MSE\(_{\lambda} \) is the minimum is later used to generate the functional objects \( F_i(t) \) that best represent process and dynamics across set \( C_1 \) and contribute most information from the prediction perspective. We expect that there would only be one minima because, as \( \lambda \) increases from 0 to a higher value the fit of \( F_i(t) \) on the observed data \( S_{it} \) will vary from over-fit to under-fit.

Figure (3) below represents the iterative framework to get optimal \( \lambda \) for obtaining functional curves \( F_i(t) \) for cumulative representation of purchase history data. Starting from \( \lambda = 0 \) which is an unbiased fit of functional curves over the data, if the minimum MSE is not obtained at the end of iteration, \( \lambda \) is incremented by the calibration parameter (p).
Figure (3) Optimal λ selection procedure

The computational λ is estimated using smooth.spline() of R. For simplicity it is represented as monotone function of the smoothing parameter λ′. λ′ is a linear function of logarithm of λ. λ′ typically lies in the interval (0,1] (Stat.ethz.ch, 2015). Hence in the application below we start the iteration with λ′ = 0.1 and increment it by 0.1 (p) until local minima is obtained.

Data and Application

We utilize weekly purchase data of 272 customers of a North American grocery store chain for 102 weeks starting from 03-2005 until 2007. In the dataset, 227 customers (set C₁) are targeted with an eight weeks customized promotion (P₁) that starts on the 59th week since the beginning of collecting observations from 03-2005. Customized coupons directly mailed to these customers are redeemable between 59th and 66th weeks. In order to build the predictive model we utilize the purchase history of each customer in C₁ until 30th week (tₚ = 30) before the start of the promotion P₁. This set of customers contribute to model building and calibrating roughness penalty to get optimal functional curves Fₜ(t). Remaining 45 (set C₂), customers comprise out of sample prediction set. These customers are targeted with a similar promotion (P₂) later in the 72th week in the observation sample and are different from those in set C₁. Prediction comparison is done using mean absolute percentage errors (MAPE) measure. Customers in set C₁ shopped in over four different stores of the retail chain while customers in set C₂ spent across two separate stores of the same retail chain. This arrangement ensures the generalizability of results across the stores and avoid any bias because of store or location effect. In both P₁ and P₂, 10-12 coupons are directly mailed to the customers. Further customers in both sets are not exposed to any other targeted promotion until 30th week before P₁ and P₂ which ensures elimination of variations in dynamics due to any other retailer’s initiative. Table (1) below describes the demographics of the customers in the sets C₁.
and $C_2$. Table (2) details the spending of the customers during and before two promotions. It can be observed that during promotion $P_1$ the standard deviation of spending increases by ≈25% while during $P_2$ average spending increases by approximately 20% indicating contrasts in spending during promotions.

**Table 1. Customer demographics details in training ($C_1$) and prediction set ($C_2$)**

<table>
<thead>
<tr>
<th>Variable</th>
<th>$C_1$ (227)</th>
<th>$C_2$ (45)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Min.</td>
</tr>
<tr>
<td>Income</td>
<td>61348</td>
<td>15000</td>
</tr>
<tr>
<td>Age</td>
<td>44.9</td>
<td>22</td>
</tr>
<tr>
<td>Kids</td>
<td>0.511</td>
<td>0</td>
</tr>
<tr>
<td>Family Size</td>
<td>2.15</td>
<td>1</td>
</tr>
<tr>
<td>Married</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 2. Customer spending details in training ($C_1$) and prediction set ($C_2$)**

<table>
<thead>
<tr>
<th></th>
<th>$C_1$ (227)</th>
<th>$C_2$ (45)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>Avgs*</td>
<td>57.0</td>
<td>38</td>
</tr>
</tbody>
</table>

*Avgs -> Average Spending

**Functional Representation**

In the next step, we obtain functional curves for discrete spending data $S_{it}$ by penalizing roughness optimally. For selecting the optimal roughness penalty we choose the starting smoothing parameter $\lambda'$ as 0.1 and estimate $MSE_{\lambda'}$ using 5 fold cross validation for the training set $C_t$ using steps 2-4. We choose 5 fold cross validation because the procedure will split 227 customers in training set into five equal parts of 45 each which is equal to the sample size of the prediction set. In the second iteration we increment $\lambda'$ by 0.1 (p) to estimate $MSE_{\lambda'+0.1}$. If the criteria in step 5 is satisfied then the iteration is stopped, otherwise the loop continues incrementing $\lambda'$ by 0.1. The objective of this procedure is to find a $\lambda'$ that leads to the minimum $MSE_{\lambda'}$ for the training set. Due to computational limitations, we select this iterative approach. $MSE_{\lambda'}$ is expected to decrease with increasing $\lambda'$ because lower values of $\lambda'$ penalize the curvature of the function less and the function $F_t(t)$ over-fits the observed data $S_{it}$ and for higher values of $\lambda'$, $F_t(t)$ under fits the observed data. It should be noted that for obtaining final value of $\lambda'$ for the observations we perform FPCA (step 3) and model building (step 4) steps which are discussed later. In this study best 5 PCs for all curves were able to represent more than ≈99% of the variations in all iterations thereby allowing to use only top five (n) principle component scores (PCS) to use for model building (4). Figure (4) below demonstrate sharp reduction in MSE as $\lambda'$ increases from 0.1 until 1.5. For $\lambda' = 0.8$ the model attains minimum MSE value of 1525 for the training set. Smoothing parameter 0.8 is therefore selected for obtaining the functional curves form of the discrete data and building the prediction model. Results described in the later sections are with respect to $\lambda' = 0.8$ that satisfies selection criteria and is obtained after iterating through the framework (figure 3) for different values of $\lambda'$. 

Data driven framework for predicting customer response to promotions

Feature Extraction and Model Building

Further, we use FPCA to reduce the dimension and extract relevant features from customer specific functional curves $F_i(t)$, $F'_i(t)$ and $F''_i(t)$. Principle component curves (PCs) thus obtained weight past observations based on variations explained across all customers. Figure (5) above shows top three PCs of the functional curve $F_i(t)$. PC1 gives higher weight (irrespective of sign) to the observations prior to the promotion indicating higher variations in customer spending before promotion. It also explains 98.5% of the variations across all of the spending curves. This is consistent with our argument that customers tend to alter their spending pattern as promotion approaches. The table (4) details variation explained by best 5 PCs of path, velocity, and acceleration curves. Top 5 components collectively represent ≈99% of variations thereby capturing most of the historical information across the customers. We use total 15 principle component scores (PCS), 5 for each of three functional curves ($F_i(t)$, $F'_i(t)$, $F''_i(t)$) for model building.

Table 4. Proportions of variations explained by PCs

<table>
<thead>
<tr>
<th>Spending Curve $F_i(t)$ Components</th>
<th>PC1</th>
<th>PC2</th>
<th>PC3</th>
<th>PC4</th>
<th>PC5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proportion of Variance</td>
<td>0.985</td>
<td>0.0128</td>
<td>0.000169</td>
<td>0.00023</td>
<td>0.0000249</td>
</tr>
<tr>
<td>Cumulative Proportion</td>
<td>0.985</td>
<td>0.998</td>
<td>0.999</td>
<td>0.99997</td>
<td>0.99998</td>
</tr>
<tr>
<td>Velocity Curve $F'_i(t)$ Components</td>
<td>PC1</td>
<td>PC2</td>
<td>PC3</td>
<td>PC4</td>
<td>PC5</td>
</tr>
<tr>
<td>Proportion of Variance</td>
<td>0.907</td>
<td>0.0798</td>
<td>0.0106</td>
<td>0.00212</td>
<td>0.000328</td>
</tr>
<tr>
<td>Cumulative Proportion</td>
<td>0.907</td>
<td>0.9869</td>
<td>0.9975</td>
<td>0.99959</td>
<td>0.999920</td>
</tr>
<tr>
<td>Acceleration Curve $F''_i(t)$ Components</td>
<td>PC1</td>
<td>PC2</td>
<td>PC3</td>
<td>PC4</td>
<td>PC5</td>
</tr>
<tr>
<td>Proportion of Variance</td>
<td>0.511</td>
<td>0.307</td>
<td>0.128</td>
<td>0.0385</td>
<td>0.00859</td>
</tr>
<tr>
<td>Cumulative Proportion</td>
<td>0.511</td>
<td>0.818</td>
<td>0.946</td>
<td>0.9844</td>
<td>0.99303</td>
</tr>
</tbody>
</table>

The framework incorporates historical spending process ($F_i(t)$) and its dynamics ($F'_i(t)$,$F''_i(t)$) information for each customer through the PCSs for the top n PCs. We select best five path ($P_i.PCS_1 \sim P_i.PCS_5$), velocity ($V_i.PCS_1 \sim V_i.PCS_5$), and acceleration ($A_i.PCS_1 \sim A_i.PCS_5$) PCSs for each customer in the training sample. Best five PCs explain ≈99% variation in all of the three curves (table 4). To remove scores that do not contribute much information to the prediction we use stepwise regression procedure with the AIC value as the selection criteria. The procedure eliminates $P_i.PCS_2$, $P_i.PCS_3$, $P_i.PCS_4$, $P_i.PCS_5$. 

Figure (4) Cross validation results    Figure (5) Principle components of F(t)
Path information

Velocity information

Acceleration information

\[
Y_{t1} = f ([P \cdot PCS_1, P \cdot PCS_2], [V \cdot PCS_1, V \cdot PCS_2, V \cdot PCS_3, V \cdot PCS_4, V \cdot PCS_5], [A \cdot PCS_1, A \cdot PCS_2, A \cdot PCS_4, A \cdot PCS_5])
\]

Figure (6)

Predictive Accuracy

Table (5) below compares the proposed model (6) with 8 other predictive models. The model 1 that utilizes only customer demographics to predict spending during promotion leads to the worst prediction. Introducing average spending (avgs) and total spending into the model (1) improves the results. Model (5) that incorporates only path / process information from the functional curve \( F(t) \) obtained using optimal roughness penalty further improves the forecast. Model (5) and (6) incorporate spending dynamics information in the form of PCSs for velocity \( (F'(t)) \) and acceleration \( (F''(t)) \) curves respectively.

Table 5. Model Comparison

<table>
<thead>
<tr>
<th>Model</th>
<th>Description</th>
<th>MAPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>( Y = f(\text{demographics}) )</td>
<td>103%</td>
</tr>
<tr>
<td>2</td>
<td>( Y = f(\text{demographics}, \text{average spending}) )</td>
<td>44 %</td>
</tr>
<tr>
<td>3</td>
<td>( Y = f(\text{demographics}, \text{total spending}) )</td>
<td>43 %</td>
</tr>
<tr>
<td>4</td>
<td>( Y = f(F(t)) \ [\text{path information (} \lambda' = 0.8 \text{)]} )</td>
<td>31.5%</td>
</tr>
<tr>
<td>5</td>
<td>( Y = f(F(t), F'(t)) \ [\text{path + velocity information (} \lambda' = 0.8 \text{)]} )</td>
<td>13.5%</td>
</tr>
<tr>
<td>6</td>
<td>( Y = f(F(t), F'(t), F''(t)) \ [\text{path + velocity + acceleration information (} \lambda' = 0.8 \text{)]} )</td>
<td>6.22%</td>
</tr>
<tr>
<td>7</td>
<td>Linear Growth Model</td>
<td>64.8%</td>
</tr>
<tr>
<td>8</td>
<td>Exponential Smoothing State Space Model (Hyndman et al. (2002))</td>
<td>33.2%</td>
</tr>
</tbody>
</table>
The complete model (6) gives best predictions for out of sample data. We also compare the results with linear trend and time series process models (ARIMA and Exponential Smoothing State Space models (ESSSM). The best specification for these models is selected using AIC and BIC criteria. Proposed model (6) outperforms both of them.

**Early Prediction**

In the earlier sub sections, we utilize the purchase history data starting from the week just before until 30th week prior to the start of the 8 week promotion in order to predict the total spending (Y) for each customer. In this section, we apply the framework to predict Y using observations from 1 to 5 weeks prior to the start of the promotional event until 30th week. Early predictions (1-5 weeks earlier) give the retailer more time to plan and strategize for the promotion. The procedure remains the same only the information included in the final model (6) decreases, and we also retain the smoothing parameter value (λ’ = 0.8) as it generates best functional form for the observed customer data. Figure (7) below demonstrates the analytical framework.

![Figure (7) Framework for early prediction](image)

### Table 6. Predictions until 5 weeks before promotion

<table>
<thead>
<tr>
<th>Model</th>
<th>MAPE (Prediction Set)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Weeks (-1)</td>
</tr>
<tr>
<td>Model (6)</td>
<td>7.65</td>
</tr>
<tr>
<td>Linear (7)</td>
<td>69.7</td>
</tr>
<tr>
<td>ESSSM (8)</td>
<td>36.5</td>
</tr>
<tr>
<td>ARIMA (9)</td>
<td>43</td>
</tr>
</tbody>
</table>

Table (6) above shows the MAPE of various models with information starting from 1 to 5 weeks prior to the start of the promotion. MAPE of other demographics based model remain approximately same because their predictors do not vary with time (table 5). As expected, due to lesser amount of recent information MAPE gradually increases from 7.65 to 16.1% for 1 and 5 weeks early forecast but the proposed model (6) still outperforms other models by huge margins. Table (6) outlines the stability and consistency of the predictions with lesser recent information. In summary, the framework signifies the importance of suitable representation of customer transaction data to extract rich dynamics information that is not directly visible.
Limitations and Discussion

The small sample size (272) for training and testing the framework should not affect comparison results, however testing for a larger sample would be interesting. There are other approaches that utilize survey or experimental data rather than just historical purchase history to predict customers’ response. In the current business environment, large and heterogeneous customer base and high competition make it impossible to collect all the relevant and recent information related to customers and draw inferences from it through surveys or experimental methods. Transaction data collected regularly is a better indicator of the recent market trend and the dynamics of customer spending process. The proposed model does not incorporate RFID and other in-store information because of data unavailability and intent to keep model simplistic. The variability in dynamics of spending before the promotion may also be due to some other external sources such as a better promotional offer by a competitive store, arrival of guests in the family or any other urgent requirement. However, we argue that any sudden change in the spending will be accompanied by some correction in the near future due to budget and inventory constraints thereby stabilizing the spending process. In addition, customers in this study are not exposed to any other promotion or intervention by the retailer during the observation period, but in general variety of marketing initiatives try to simultaneously alter the customers’ spending at the store.

Future studies in this direction can further explore the impact of other marketing interventions such as loyalty reward points, in-store product display etc. on the spending dynamics and response. Functional representation can also be used to investigate and predict cyclic patterns in customer spending and formulate decision rules for profit maximization through selectively targeting customers in an appropriate phase of cycle for the promotion. As majority of promotions now span over certain period of time predicting the intensity of spending throughout the duration of the promotion will also be of strategic interest to the retailers. The approach can also be replicated to other similar contexts, for example to understand the relationship between the dynamics of website visits of customers and their online purchases.

Conclusion

The findings of the study show that the framework based on the FDA that utilizes an optimal smoothing parameter to obtain functional curves gives the best early estimates of customer spending during the promotion. This information will help managers to strategize and plan for promotional events while optimizing available resources. Functional representation of customer transaction data provides extra information in the form of spending dynamics and allows modeling response while accounting for heterogeneity due to purchase cycle factors as well as cross sectional differences. The data driven model significantly improves the out of sample predictions parsimoniously with minimal statistical constraints. Given the large amount of fast moving longitudinal data collected in electronic commerce, the framework can act as a guideline for modeling response to specialized events such as promotions. Our study contributes to the existing literature by demonstrating how to extract an optimal functional curve from noisy customer data without over-fitting it, and by demonstrating the significance of spending dynamics information in predicting customer response. In summary, visualizing fast moving discrete data across various contexts in business as functional curves can help in identifying interesting relationships and improving predictive power.

REFERENCES


