Forecasting Simulated Retail Demand Using Statistical and Data Mining Techniques

by

Ayşe Gül Tunçelli

A Thesis Submitted to the

Graduate School of Engineering

in Partial Fulfillment of the Requirements for

the Degree of

Master of Science

in

Industrial Engineering

Koc University

December 2007

Koc University

Graduate School of Sciences and Engineering

This is to certify that I have examined this copy of a master's thesis by

Ayşe Gül Tunçelli

and have found that it is complete and satisfactory in all respects,

and that any and all revisions required by the final

examining committee have been made.

Committee Members:

Assist. Prof. Özden Gür Ali (Advisor)

Prof. Serpil Sayın (Advisor)

Prof. Selçuk Karabatı

Assist. Prof. Sibel Salman

Assist. Prof. Yalçın Akçay

Date

ABSTRACT

Primary research purpose of this thesis is to evaluate statistical and data mining techniques for demand forecasting in the presence of promotions. A consumer choice model is developed by modifying present models of consumer choice in marketing literature for grocery retail industry. Data generation task is carried out according to the developed consumer choice model. Data is generated in a retail environment which has a single category, multiple products, multiple product attributes and multiple customer segments. Marketing drivers, such as price discounts, advertisement and feature displays are in data generation. Forecasting is performed over generated data by using both traditional statistical techniques, such as exponential smoothing and regression and recent data mining techniques, such as support vector machine regression and regression tree. Consequently, forecasting results of several techniques are compared according to accuracy of the SKU demand forecasts, simplicity in terms of parameter estimation and forecasting performance in new SKU entered situations. Finally, preferred methods for different data conditions are explained.

ÖZET

Bu tezin başlıca amacı, promosyonların bulunduğu bir perakende ortamında istatistik ve veri madenciliği yöntemleri ile talep tahmini yapmaktır. Bu çalışmada, perakende sektörü için pazarlama literatüründe bulunan müşteri tercih modellerine dayanan bir müşteri tercih modeli geliştirilmiştir. Geliştirilen müşteri tercih modeline göre veri üretimi yapılmıştır. Veri üretimi için dikkate alınan perakende ortamı, tek bir kategoriyi, birden fazla ürünü, birden fazla ürün özelliğini ve birden fazla müşteri segmentini içermektedir. Fiyat indirimi, reklam ve özel teşhir gibi pazarlama aktiviteleri göz önünde bulundurulmuştur. Talep tahmini, üretilen data üzerinde hem üstel düzeltme ve lineer regresyon gibi istatistik yöntemleri ile hem de SVM regresyon ve regresyon ağacı yöntemleri gibi veri madenciliği teknikleri ile gerçekleştirilmiştir. Kullanılan yöntemler, SKU talep tahmini doğruluğu, paremetre hesap kolaylığı ve kategoriye yeni giriş yapan SKU'ların tahmin doğruluğu üzerinden kıyaslanmıştır. Sonuç olarak farklı veri durumları için önerilen talep tahmin metodları belirtilmiştir.

ACKNOWLEDGEMENTS

I would like to thank to all those who gave me the possibility to complete this thesis. First, I would like to extend my sincere gratitude to my advisors Assist. Prof. Özden Gür Ali and Prof. Serpil Sayın for their patience, guidance, and encouragement throughout this past two years.

I am also grateful to members of my thesis committee for critical reading of this thesis and for their valuable comments.

I owe thanks to my father for motivating me and my mother for always being there when I need her. To my sister, thank you for the joy and fun you bring to my life. Without the spiritual support of my family, it would have been harder to complete this thesis. I would like to thank to my colleagues Zehra, Seda, Dilek, Pınar, Uğur and Fadime for their encouragements and my officemates Tuğba, Can, Taha and Burak for their intelligent ideas and remarks. Last, but not the least I thank my friend, İlkan Sarıgöl for his invaluable support and guidance.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

INTRODUCTION

Forecasting, the act of making predictions of future events and conditions, is very important in many types of organizations since predictions of future events should be incorporated into decision-making processes. Especially business firms require forecasts of many events and conditions in all phases of their operations. Demand forecasts drive a firm’s production, capacity, and scheduling systems and affect the financial, marketing and personnel planning functions. For example, in marketing departments, reliable forecasts of demand must be available so that sales strategies can be planned and in production scheduling, predictions of demand for each product line are needed in order to plan production schedules and inventory maintenance , according to Bowerman et al. [3].

As Harris and Sollis [25] explained, a forecast can be classified into three categories by the future time horizon that it covers. A short-range forecast has a time span generally up to 3 months and is used for planning purchasing, job scheduling and workforce levels etc. A medium-range forecast generally spans from 3 months to 3 years and is useful in sales planning, production planning and budgeting etc. A long-range forecast has generally 3 years or more in time span and is used in planning for new products, capital expenditures and facility location etc.

Forecasts also can be classified according to their types. According to Bowerman et al. [3], three types of forecasts are: economic forecasts that address the business cycle by predicting inflation rates, money supplies and other planning indicators, technological forecasts that concern with rates of technological progress, which can result in birth of new products, requiring new plants and equipment and demand forecasts that project demand for a company's products or services. Demand forecasts, which are also called sales forecasts, drive a company's production, capacity and scheduling systems and serve as inputs to financial, marketing and personnel planning.

According to Heizer and Render [10], forecasting follows these seven basic steps:

- Determine the use of forecast
- Select the items to be forecasted
- Determine the time horizon for the forecast
- Select the forecasting model(s)
- Gather the data needed to make the forecast
- Make the forecast
- Validate and implement the results.

There are two general approaches to forecasting, these are quantitative analysis and qualitative approach. Quantitative forecasts use a variety of mathematical models that rely on historical data and/or causal variables to forecast demand. Subjective or qualitative forecasts incorporate factors like decision maker's intuition, emotions, personal experiences and value system in reaching a forecast. Some firms use one approach, some use the other and some firms prefer to use a combination of the two approaches, according to Heizer and Render [10].

Quantitative forecasting methods can be collected under two headings: statistical methods and data mining methods. Two approaches of statistical forecasting are time-series methods and causal methods. Forecasting methods are explained in detail in Chapter 2.

Forecasting demand is one of the critical tasks of retail operations. Retailers use SKU demand forecasts in their ordering policy. Therefore the quality of demand forecast has a direct impact on the inventory levels and on stock-outs, i.e. lost sales, both of which impact the retailer's operational cost.

Primary research purpose of this thesis is demand forecasting in a generated time series data for retail industry using statistical and machine learning techniques. Considered retail environment is a single category with multiple SKUs (stock keeping unit), multiple SKU attributes, multiple customer segments and marketing-mix instruments.

The generated data is a weekly sales data of multiple SKUs in black tea category along with associated drivers, such as price discounts, promotions and feature displays, for forecasting demand for multiple SKUs in a retail environment. Sales data is obtained by developing a data generation model on the basis of consumer choice models in marketing literature. A consumer choice model attempts to represent how consumers use and combine information about alternatives in order to make a choice among them. Our data generation model is built on Fader and Hardie (1996)'s study. Fader and Hardie (1996) built a consumer choice model that includes all of the distinguishing attributes that characterizes a particular product category's set of SKUs. Authors explain the attribute as the physical characteristics that uniquely identify every item available on the store shelf. For example, brand and size are the most common SKU attributes that can be available for almost all categories in a grocery retail store.

Data generation model has input parameters like marketing-mix variables, base preferences and response vectors. Marketing-mix variables of this study are temporary price reduction, display and advertisement actions. Base preferences indicate households' preferences toward SKU attributes when there is no marketing-mix activity. Response vectors are the responses of households toward marketing-mix instruments. These inputs are estimated after interviews with experts from industry. According to information obtained from experts, a marketing-mix instruments pattern is obtained and other input parameters are calibrated.

The forecasting objectives of this study are high accuracy, model simplicity and good performance on new SKU prediction. In order to experiment with the performance of new SKU forecasting, two additional data sets that include new SKU introductions are generated. Forecasting experiments are performed on three different data sets. Beside this, data sets are handled in two approaches as individual data set forecasting and combined data set forecasting. Individual forecasting is performed to predict each SKU independently and combined forecasting is performed to predict the sales of whole category by using a single data set and model.

In this thesis, forecasting is performed by using statistical techniques like multiple linear regression and exponential smoothing techniques, as well as data mining techniques like regression tree and support vector regression. A comparison of these techniques in different data conditions are given considering the forecasting objectives.

The rest of the thesis is organized as follows. In Chapter 2, literature review on consumer choice models, statistical forecasting techniques and predictive data mining techniques are given. Chapter 3 describes the data generation model development on the basis of consumer choice models. Also in this chapter, black tea category characteristics, expert opinions on the black tea category and input parameters calibration are explained. In Chapter 4, results of forecasting experiments with statistical and data mining techniques for different data sets are given. Chapter 5 summarizes what has been done so far, discusses forecasting results and gives future research directions.

Chapter 2

LITERATURE REVIEW

2.1 Introduction

In this study, retail demand is forecasted by using both statistical and data mining techniques. Forecasted data is a generated weekly sales data of the black tea category in a grocery retail store. In order to generate such data, a data generation model that is explained in Chapter 3 is developed on the basis of consumer choice models in the marketing literature. While developing the data generation model, marketing literature related with consumer choice models is investigated. In addition to this, literature is also reviewed for statistical and data mining forecasting techniques. In a recent study, Gür Ali et al. [17] compare exponential smoothing and data mining techniques for various SKU-store combinations using a real grocery retail data. In this study, data is pooled across sub categories in a store and across stores. Authors found that, data mining techniques give better accuracy for promotional time period, whereas exponential smoothing performance is same or better than data mining techniques for non promotional time period. In our study, similar forecasting techniques are used on experiments, but with a generated time series data.

In this chapter, literature review is handled under these three headings: consumer choice models, statistical forecasting techniques and data mining forecasting techniques.

2.2 Consumer Choice Models

There is an extensive marketing literature on SKU demand prediction that includes many models estimating the response to marketing mix effects, such as price discounts, feature and display. The marketing literature explains customer demand behavior with models that receive consumer demand behavior as input. The data is used in order to estimate the parameters of these models.

Bucklin and Gupta [26] examine the practitioner community's view of the use of scanner data and compare these views with academic research. Authors discuss following six areas of the marketing mix: pricing, trade promotion, consumer promotion, advertising, product distribution and retail management. They described exponential smoothing with lift adjustment technique, which is the commonly used method of forecasting sales in the practitioner community. In this study, advertising effects on sales and prices are discussed as long term and short term effects of advertising.

Foekens et al. [7] state that, because time series observations typically summarize market shares and marketing activities at the brand level, no distinction could be made between the different types, sizes or flavors in which product items are available to consumers. Thus, researchers required to place restrictions on the nature of competition in consumer response models. Authors propose hierarchical models of item dimensions such as brand name and package size and compare different possible hierarchies against each other and against other nonhierarchical specifications in order to examine the competition in a market. In a hierarchical model, consumers may differentiate items according to dimensions like brand name, package size and price. In hierarchical structures, consumers may first choose a brand name and, according to this decision, choose a particular item. However, as the number of considered dimensions increases, the hierarchical structure becomes increasingly complex. For example, with a category having 4 distinct dimensions, there are 24 possible hierarchies representing every factorial combination of dimensions. If heterogeneity in hierarchy structure is allowed, the modeling effort would be unmanageable. Because of this reason, a hierarchical model is not preferred in this thesis.

Cadeaux [13] states that supermarket retailers offer assortments in response to the supply and demand conditions that are specific to the categories and subcategories. From a macro and industry-wide perspective, this paper represents alternative theories relating the size of supermarket subcategory demand to brand and SKU dimensions of assortments.

Bell et al. [6] show that SKU-level parameters can be recovered by calculation from estimated attribute-level parameters, eliminating the need for direct estimation of the more complex SKU-level model. Attributes can be considered as the physical characteristics that uniquely identify every item on the store shelf. In this study, brand, flavor, function, form and size are the five SKU attributes of toothpaste category. Authors calibrate the store data market share model using 98 weeks of data for ten brands and 168 SKUs of toothpaste category. In this article authors construct a consumer choice model which is composed of these two parts: fixed effects and covariate effects. Basic model with fixed effects does not include marketing mix effects such as price, feature and display. It is developed for a single category with multiple products considering these factors: SKU, attribute, attribute level and time. Market share of an SKU j at a time t is obtained by SKU-specific utility relative to an outside good. SKU-specific fixed effects and attribute-specific fixed effects are obtained directly from ordinary least squares estimation on the SKU-level and attribute-level market share equations, respectively. Consequently, a demonstration of how to recover marketing mix effects (i.e., price, feature and display) is given. This study also gives a comparison of market share model parameters estimated at different levels of aggregation.

Guadagni and Little [20] demonstrate that household level logit models provide a powerful method for identifying and measuring the effects of marketing variables on product choice. This study is among the first to acknowledge the role of SKU attributes by recognizing the importance of package sizes and brand names. Authors model the process of the brand choice of a customer's possible purchasing activity on a shopping trip to a supermarket in a given category. The calibrated model predicts the behavior of a hold-out sample of customers of the regular ground coffee market. In this study, explanatory variables include store actions, such as price and promotion, and customer characteristics such as brand and size loyalty. In a more recent study Guadagni and Little [21] extended the formulation of [20] to include the decision of category selection on a shopping trip. In this study, a nested model, which is a generalization of the multinomial logit, is developed to model the brand choice in coffee category. Guadagni and Little describe the multinomial logit model as the selection of brand and size given that the customer makes a category purchase. The category choice introduced further variables like household inventory, category price, and the attractiveness of purchasing a product now as opposed to later. According to the authors, the combined model, incorporating both category and product choice, achieves the calculation of sales response including both category expansion and brand switching. In both of these studies of Guadagni and Little, brand and size are the considered SKU attributes.

Despite the precedent study of Guadagni and Little [20], relatively few researchers have used additional SKU attributes in choice models; most tend to consider brand as the only SKU attribute. Although the approach of Pedrick and Zufryden [12] is to pool items across an attribute by creating a composite average, Fader and Lattin [23] and Horsky, Misra and Nelson [5] treat items with any different SKU characteristics as distinct brands.

Another study that uses multinomial logit model while modeling the consumer choice is Fader and Hardie [23]'s. In this paper, authors discuss how a set of discrete attributes (e.g., brand name, package size type) can be used to characterize a large set of SKUs. Also the authors develop a modeling approach that includes all of the distinguishing attributes that define an SKU set of a particular product category. The model is shown to be substantially superior to a more traditional framework that does not emphasize the complete use of SKU attribute information. In this study, modeling approach for product choice with scanner data includes a single category of a retail store with multiple products and multiple segments.

Most consumer product categories have hundreds of SKUs and it is a challenge to estimate the consumer choice models when the models consist of product specific parameters that are at least as large as the number of items in the categories, according to Ho and Chong [30]. An approach reducing the number of product specific parameters is focusing on a subset of the SKUs. For example Fader, Lattin and Little [24] and Siddarth, Bucklin and Morrison [28] discarded all purchase incidences of the least frequently bought SKUs. In another approach, Guadagni and Little [20] aggregate the level of analysis to a higher level from SKU to brand-size combination. Fader and Hardie [23] overcame the challenge of estimating product-specific parameters by assuming that a product category comprises remarkable attributes and that each remarkable attribute has different levels. An SKU derives its main value from the attribute levels it acquires. The product specific parameters become sums of attribute level values. We use this approach in our simulation model of consumer behavior. In our generated black tea category data there are 35 SKUs and sum of attribute levels is 17.

According to Fader and Hardie [23], a consumer choice model should cover the preferences over the levels of each SKU attribute, not the direct preferences for each SKU. There are two main advantages of focusing on the SKU attributes: considering fewer preference parameters and covering all SKUs in a category. In order to find probability of choosing an item in the category, given that the customer makes a category purchase, authors use a multinomial logit model. Our data generation model is developed by modifying the consumer choice model in the study of Fader and Hardie [23]. The detail of data generation model is explained in Chapter 3.

2.3 Statistical Forecasting Techniques

Statistical forecasting techniques can be classified as time series methods and causal methods.

Time-series models predict on the assumption that the future is a function of the past. In the other words, time-series models look at what has happened over a period of time and use a series of past data to make a forecast. If a forecaster is predicting weekly sales of mobile phones, the forecaster uses the past weekly sales for mobile phones when making forecasts. Most common time-series methods are naïve approach, moving averages and exponential smoothing. Naïve approach assumes that the demand in the next period will be equal to demand in the most recent period. A moving average forecast uses a number of historical actual data vales to generate a forecast. Exponential smoothing is a sophisticated weighted moving-average forecasting method. In exponential smoothing, the latest estimate of demand is equal to the old estimate adjusted by a fraction of the difference between the last period’s actual demand and old estimate Hanke and Wichern [11].

Exponential smoothing has proven to be a robust forecasting method and is probably the most used of the statistical approaches to forecasting demand [31]. These models are widely used in practice, are easy to understand and give respectable results in accuracy. They are also able to adapt to new patterns and smooth out the noise.

Unlike time-series forecasting, causal forecasting models usually consider several variables that are related to the quantity being predicted. Once these related variables have been found, a statistical model is built and used to forecast the item of interest. This approach is more powerful than the time-series methods that use only the historical values for the forecasted variable. The most common quantitative causal forecasting model is linear-regression analysis, according to Heizer and Render [10].

Linear regression and multiple regression methods are also used to forecast the future value of a dependent variable from the value of another variable (independent) and the values of other variables (independent) respectively. Linear regression and multiple linear regression methods are broadly used in literature. For example, [14] compared multiple linear regression method with feed forward artificial neural networks in milk production forecasting. The details of these classical models can be found, for instance, in [35].

2.4 Data Mining Forecasting Techniques

An alternative approach relevant to SKU forecasting is data mining. Data mining strives to learn the patterns in the data, without restricting the structure of the relationship based on neither behavioral theory nor statistical estimation, as Roiger and Geatz [27] explained.

Data mining is an inclusive term for various types of analyses designed to discover relationships within a set of data. Data mining includes associations, sequences, grouping of data though classification and clustering and forecasting [8]. Most common data mining techniques used for forecasting are regression trees, neural networks and support vector machine regression.

Decision tree is one of the most widely used and practical methods for inductive inference. It is a method for approximating discrete-valued functions that is robust to noisy data and capable of learning disjunctive expressions. Decision tree is called as regression tree when the target variable is continuous, according to Mitchell [29]. Salford Systems [36] defines the regression tree as a predictive model which is structured as a sequence of simple questions, and the answers to these questions trace a path down the tree. The end point reached determines the numerical forecast made by the model.

A neural network is a special kind of model that relates inputs to outputs and does like the organization of the human brain. The network consists of an interlinked set of simple processing nodes, which represent complex relationships among inputs and outputs [29].

Support Vector Machines are the most popular new development in the data mining literature in the last ten years after Neural Networks. Since the creation of Support Vector Machine (SVM) theory by Vapnik of the AT&T Bell Laboratories [32], [33], there have been intensive studies on SVM for both classification and prediction tasks (e.g. [16] and [1]). Support vector machines construct nonlinear features of the input variables like the neural networks, however they differ in that they have the ability to generalize, strive to minimize model complexity, and they perform a global optimization to fit the parameters. This is largely due to the structural risk minimization principle in SVM, which has greater generalization ability and is superior to the empirical risk minimization principle employed in neural networks. In SVM, the results guarantee global minima, whereas empirical risk minimization can only locate local minima. For example, in the training process of neural networks, the results give out any number of local minima that are not guaranteed to include global minima. Furthermore, SVM is adaptive to complex systems and robust in dealing with corrupted data. This feature offers SVM a greater generalization ability than the neural network approach [34].

The rapid development of SVMs in statistical learning theory encourages researchers to actively apply SVM to various research fields. Traditionally, many studies focus on the application of SVM on classification and pattern recognition. Recently, the application of SVM to prediction, called Support Vector Regression (SVR), has also shown many breakthroughs such as forecasting of financial market [9], forecasting of electrical load [15] and travel-time prediction [4]. In addition to these studies, [19] demonstrate that although the statistical models cannot easily capture the nonlinear patterns, support vector machines are successfully applied in solving nonlinear regression estimation problems. This study proposes a hybrid methodology that exploits the unique strength of the statistical models and the SVMs model in stock price problems. Likewise, [18] demonstrate the use of support vector regression for prediction of gold prices. Since there are many successful results of time varying applications with SVR prediction, it motivates our research in using SVR for retail demand forecasting. A detailed description of SVR formulation is given in Appendix I.

Chapter 3

A MODEL FOR DATA GENERATION

3.1 Introduction

The purpose of this study is forecasting sales of a single category in a store of a grocery retail chain. Forecasting demand is one of the most critical tasks in retail industry. As mentioned before, our forecasting objectives are high accuracy, model simplicity and good performance on new SKU prediction. New SKU entrance to a category is a common issue in retail environment. Performing analyses on different data sets that include different number of new SKUs motivated us to develop a data generation simulation. Using the developed data generation model three different data sets are obtained for forecasting experiments.

Data generation model is developed on the basis of the consumer choice models in the marketing literature. Researchers in marketing field use consumer choice models in order to predict the market shares and sales of SKUs in the retail environment. In this study consumer choice models are used to develop the data generation model.

Consumers typically choose among SKUs on the basis of a set of product attributes, which tend to be discrete and tangible. For most consumer packaged goods, it is convenient to describe SKUs in terms of a set of categorical attributes that are physical characteristics that uniquely identify every item available on the store shelf. Beyond the attributes, the attribute level is used to denote each of the distinct options that together constitute an attribute. For instance, A, E and B are a subset of the levels under the brand attribute in the black tea category. According to Fader and Hardie [23] an SKU attribute must be easily observable, there should be no ambiguity about each item's precise level for each SKU attribute and every SKU attribute must be applicable to every SKU.

Because of large number of SKUs in most grocery product categories, it is a challenge to estimate the parameters of the consumer choice models when the models contain product specific parameters that are at least as large as the number of items in the categories. Fader and Hardie [23] overcame the challenge of estimating product-specific parameters by assuming that a product category comprises remarkable attributes and that each remarkable attribute has different levels. An SKU derives its main value from the attribute levels it acquires.

Our data generation model is based on Fader and Hardie [23]'s model, because not only it requires less parameters but also it captures all SKU attributes in the category, consumer heterogeneity and marketing-mix effects at the same time.

In this chapter, firstly the data generation model that is developed to obtain weekly sales data is described, secondly the considered retail category and the details acquired from experts are explained, then the parameter setting analyses are given and finally generated data is explained.

3.2 Data Generation Model

Consider household h, who visits a store to buy an SKU in a particular product category. The product category has many SKUs indexed by i. The household evaluates the product category by a set of R SKU attributes indexed by r. Each SKU attribute r has L_r attribute levels indexed by j. According to an example in Bell, Bonfrer, Chintagunta [6], the toothpaste category can be evaluated by attributes like brand, flavor, form, function and size. The possible attribute levels for toothpaste flavor are mint, extra fresh, regular, and so on. An SKU offers an attribute-level combination such as "Colgate, mint, gel, tartar control, medium". Beside this, there are multiple customer segments, $s = 1,...,S$, and multiple customers, $h = 1,...,H$, in the category. Each household belongs to a single customer segment deterministically. Because each household h belongs to a single customer segment s, heterogeneity modeling is captured in this model. That gives the opportunity of considering different consumer segments preferences in a retail category.

We defined the black tea category as having 35 SKUs, 5 SKU attributes which are brand, package type, tea aroma and price category, and 3 customer segments which are bulk tea users, tea-bag users and teapot bag users.

The household makes product purchase on multiple occasions. On each purchase occasion, the household decides which SKU i to purchase, given all SKU's marketing-mix activities. The goal of the model is to estimate which SKU i is purchased by the household h on a purchase occasion given the purchase history. In our model a household's utility for an attribute level changes over time, because the consumer accumulates a consumption experience for the chosen attribute level.

Inputs of Data Generation Model

1) Customer Arrivals

Because we are generating data according to a consumer choice model, customers' purchase activities would be an input for data generation task. In order to obtain purchase activities of customers we need their arrivals for shopping on the store. We assumed that the number of households visiting the store is 10,000 in a week. Let

$C_h^t = 1$, if customer h arrives at the store at week t,

$C_h^t = 0$, otherwise

C_h^t values will be obtained by simulation. In the simulation, customers' arrival frequency in each segment is considered as once every two weeks.

2) Marketing Mix Effects

Temporary price reduction (TPR), display and advertisement are considered as marketing mix variables. In this study, marketing mix variables are SKU specific, they do not depend on SKU attributes. We define:

$X_i^{price,t} = 1$ if SKU i has a TPR at week t, 0 otherwise

$X_i^{disp,t} = 1$ if SKU i has a display at week t, 0 otherwise

$X_i^{adv,t} = 1$ if SKU i has an advertisement at week t, 0 otherwise

These marketing mix variables will be initialized according to information acquired from the experts in industry.

3) Attribute Levels

L_r^i is an integer and corresponds to the attribute level of SKU i for attribute r.

4) Base Preferences

$a_r^{h,0}$ is a vector of size L_r, that shows the base preference of household h for attribute r.

5) Marketing Mix Scalar Multipliers

$\beta_{price}^h, \beta_{disp}^h, \beta_{adv}^h$ are scalars that will be used in calculation of response vectors that indicate the response of customers towards marketing mix instruments such as advertisement and price reduction. Multiplying marketing mix scalars with base preference vectors gives response vectors. For instance, $b_{price,r}^{h,t} = \beta_{price}^h . a_r^{h,t}$ is the equation that will be used for TPR response vector calculation.

6) Attribute Loyalty Scalar Multipliers

α_r^h is a scalar that corresponds to attribute r' s attribute loyalty scalar multiplier for customer h.

7) Attribute Specific Constants (Lambda)

λ_r is a smoothing constant which is attribute specific, and it is used in the calculation of attribute loyalty vectors.

The standard approach to modeling product choice involves the use of multinomial logit model (MNL). Building on the work of Fader and Hardie (1996), the standard multinomial logit function commonly used in consumer choice model literature is modified by adding an outside good effect. By adding such an outside good effect a customer is enabled to choose none of the SKUs in the category in a specific shopping trip. The MNL in Fader and Hardie (1996)'s study is modified as in the following structure to allow the consumer choice of not buying anything:

$$p_i^{h,t} = \frac{\exp(v_i^{h,t})}{1+\sum_i \exp(v_i^{h,t})} \qquad 3.1$$

$p_i^{h,t}$ is the probability that household *h* chooses SKU *i* in week *t*. $v_i^{h,t}$ is the utility of SKU *i* for household *h* in week *t*. Utility ($v_i^{h,t}$) is estimated only at choice points (if $C_h^t = 1$), for SKU i, customer h and week t. The outside good effect is given by:

$$outgood_i^{h,t} = \frac{1}{1+\sum_i \exp(v_i^{h,t})} \qquad 3.2$$

The $v_i^{h,t}$ can be viewed as having two components on the household's choice behavior. These components are a preference component which represents the household's base preference toward SKU *i* and a marketing mix component which represents the effects of marketing mix variables (price reduction, display and advertisement).

$$v_i^{h,t} = \underbrace{\sum_{r=1}^{R} e_{ir}.a_r^{h,t}}_{\substack{Preference\\ component}} + \underbrace{\sum_{r=1}^{R} e_{ir}.b_{price,r}^{h,t}.X_i^{price,t} + \sum_{r=1}^{R} e_{ir}.b_{disp,r}^{h,t}.X_i^{disp,t} + \sum_{r=1}^{R} e_{ir}.b_{adv,r}^{h,t}.X_i^{adv,t}}_{Marketing-mix\ component} \qquad 3.3$$

e_{ir} is an elementary row vector of size L_r.

$$e_{ir} = \begin{bmatrix} e_{ir}^1 \\ \vdots \\ e_{ir}^{L_r} \end{bmatrix} \qquad e_{ir}^j = \begin{cases} 1 & if\ \ SKU\ i\ has\ level\ \ j\ of\ \ attribute\ r \\ 0 & otherwise \end{cases}$$

$a_r^{h,t}$ is the vector of preferences over the L_r levels of attribute r. $a_r^{h,t}$ is composed of two components: base preferences and attribute loyalties. Base preferences ($a_r^{h,0}$) can be considered as a vector of segment specific attribute-level intercept terms. In order to consider additional heterogeneity across consumers, these intercept terms are augmented by a set of attribute specific loyalties for attribute r, $ATTLOY_r^{h,t}$. The vector of preferences ($a_r^{h,t}$) has the following structure:

$$a_r^{h,t} = (a_r^{h,0} + \alpha_r^h ATTLOY_r^{h,t}) \qquad 3.4$$

where α_r^h are scalars that correspond to the attribute loyalty scalars of household h for attribute r.

$ATTLOY_r^{h,t}$ is a vector of size L_r, that shows attribute loyalty vector for attribute r at week t. For initialization, attribute loyalty vectors, $ATTLOY_r^{h,t}$**,** are considered as 0 vectors. $ATTLOY_r^{h,t}$ is a column vector and number of its elements is equal to the number of attribute r' s levels. For a household, $attloy_{jr}^h(t+1) = \lambda_r attloy_{jr}^h(t) + (1-\lambda_r)\sum_i \delta_{it}$, where $\delta_{it} = 1$ if SKU i was purchased at time t, 0 otherwise, and $\sum_i$ is over all the SKUs possessing level j of attribute r. $ATTLOY_r^{h,t}$ vectors evolve by the purchasing experiments of households. Each purchasing activity updates the preference vector ($a_r^{h,t}$) by $ATTLOY_r^{h,t}$ vectors.

$b_{price,r}^{h,t}, b_{disp,r}^{h,t}, b_{adv,r}^{h,t}$ are the response vectors that indicate the household *h*'s responses for attribute *r* in week *t* toward price reduction, display and advertisement marketing mix instruments respectively. Response vectors in Fader and Hardie (1996)'s study are composed of scalars that are distinct from preference vectors. In this thesis, response vectors of Fader and Hardie (1996) are modified by multiplying the response vectors by response scalars. By this modification, response vector of a household *h* for marketing mix instruments became dependent on preferences of that household.

The response vectors ($b_{price,r}^{h,t}, b_{disp,r}^{h,t}, b_{adv,r}^{h,t}$) are obtained by the multiplication of the marketing-mix scalar multipliers ($\beta_{price}^{h}, \beta_{disp}^{h}, \beta_{adv}^{h}$) and the preference vectors ($a_r^{h,t}$).

$$\begin{aligned} b_{price,r}^{h,t} &= \beta_{price}^{h}.a_r^{h,t} \\ b_{disp,r}^{h,t} &= \beta_{disp}^{h}.a_r^{h,t} \\ b_{adv,r}^{h,t} &= \beta_{adv}^{h}.a_r^{h,t} \end{aligned} \tag{3.5}$$

Output of the Data Generation Model

In this model $p_i^{h,t}$ gives the probability that household *h* chooses SKU *i* in week *t*. In other words, the output of this data generation model is SKU shares for customer *h* in week *t*. Using this share the total demand for SKU *i* in week *t* denoted by $demand_i^t$ is obtained.

$demand_i^t = \sum_h Z_i^{h,t}$, where $Z_i^{h,t}$ indicates the purchase of household *h* of SKU *i* at week *t*.

$$Z_i^{h,t} = \begin{cases} 1, if\ C_h^t = 1\ and\ "SKU\ simulation"\ gives\ SKU\ i\ \ for\ customer\ h\ at\ week\ t \\ \\ 0, otherwise \end{cases}$$

where “SKU simulation” is a task that either gives out the selected SKU for household *h* at week *t* or shows that household *h* does not buy any product in the visit at week *t*. In “SKU simulation”, all SKU shares for household *h* at week *t*, $p_i^{h,t}$, are calculated initially. These shares are interpreted as purchase probabilities and a uniform random number in the interval (0,1) for household *h* at week *t* is generated. Generated random number is used for determining the selected SKU *i* for household *h*. While cumulatively summing up SKU purchase probabilities for household *h*, generated random number is compared with cumulative sum of SKU purchase probabilities at that moment. For instance, assume that there are 4 SKUs in our category and we found all $p_i^{h,t}$ values, for *t*=1 and *h*=1 we will cumulatively sum up purchase probabilities for 4 SKUs until we reach the generated random number. This can be expressed as follows:

If $p_1^{1,1}$>=random number, then household 1 at week 1 selects SKU 1.

Else if $p_1^{1,1} + p_2^{1,1}$>=random number> $p_1^{1,1}$, household 1 at week 1 selects SKU 2.

Else if $p_1^{1,1} + p_2^{1,1} + p_3^{1,1}$>=random number> $p_1^{1,1} + p_2^{1,1}$, household 1 at week 1 selects SKU 3.

Else if $p_1^{1,1} + p_2^{1,1} + p_3^{1,1} + p_4^{1,1}$>=random number> $p_1^{1,1} + p_2^{1,1} + p_3^{1,1}$, household 1 at week 1 selects SKU 4.

Else household 1 at week 1 buys nothing, or purchases the outside good in accordance with the probabilities of our model.

3.3 Black Tea Category

The decision of the retail category selection is made upon the interviews with two experts from industry. One of the experts we interviewed has worked as a product manager for many different products for ten years in an FMCG company and the other expert is the planning and category manager in one of the biggest retail chains in Turkey for many years.

While selecting the category, we looked for characteristics such as existence of multiple customer segments that present different shopping preferences for the category, multiple SKU attributes, multiple marketing-mix instruments and multiple competitor brands in the category. Because black tea category includes all these characteristics and it is the most consumed drink in Turkey, it is decided to perform the forecasting study on the black tea category. Besides this, customer segments differentiate according to preferred tea types. Many retail categories include customer segments that have same preferences but their buying power differentiate them. For example, in detergents category, all customers want to purchase from same product group, but price of the product becomes the most important factor on their purchasing activities. However, Black tea category customers in each segment are loyal to their tea tastes, not necessarily to price of the product.

From the experts, mainly the marketing-mix instrument features like promotion type, frequency and effects on sales volume and details of customer segments in the black tea category are obtained. Besides this, the brands and SKU attributes of the category are also obtained from the experts. During the interview a form, which can be seen in Appendix II, including the questions about the category specifications is filled in order to better understand the category. In addition to this, the web store of a Turkish grocery retail chain is analyzed in order to better obtain the SKUs and SKU attributes in the black tea category.

In this part, firstly the characteristics of SKU attributes are described and then the information obtained about marketing-mix instruments and their relation with customer segments is explained. Finally, information about customer segments in the black tea category obtained from the experts is given.

3.3.1 SKU Attributes

According to the internet store we referred to, black tea category includes 60 SKUs. The most salient attributes of the black tea category are brand, package type and tea aroma. As explained in the data generation model, each SKU attribute is composed of attribute levels. Brand attribute has 5 levels which are A, B, C, D and E, aroma attribute has 5 levels which are rize, filiz, ceylon, earl grey and breakfast, and 3 levels of package type attribute are bulk tea, tea bag and teapot bag.

Price of the product is one of the important decisive factors for a purchasing activity. Because of this reason, we introduced price as an SKU attribute. Each SKU in the category has distinct price values, therefore assuming these distinct values as levels of price attribute would be very complicated to the model. In order to get rid of such a complicated and unmanageable size of model parameters, the price attribute is decided to indicate the price category of the SKU. SKUs in the black tea category are clustered according to their prices to obtain the price category that each SKU belongs to. While the levels of brand, aroma and package type attributes are easily observable for an SKU, levels of price attribute needs a computational approach in order to be recognized.

The computational approach for obtaining price attribute levels was as follows.

- Firstly, price/gram values are calculated for each SKU. Then price groups are obtained by clustering the price/gram values using K-means algorithm. Weka 3.5 program is used for K-means application.

- According to experiments with several cluster numbers (K) these results are found: while K=2, instance percentage in each cluster is 72% and 28% respectively, while K=3, instance percentage in each cluster is 48%, 27% and 25% respectively, while K=4, instance percentage in each cluster is 47%, 15%, 25% and 13% respectively, while K=5, instance percentage in each cluster is 22%, 15%, 27%, 13%, and 23% respectively, while K=6, instance percentage in each cluster is 22%, 18%, 25%, 12%, 18% and 5% respectively.
- Within cluster errors are as seen below:

Number of Clusters	Within cluster sum of squared errors
K=2	0.977
K=3	0.476
K=4	0.201
K=5	0.182
K=6	0.179

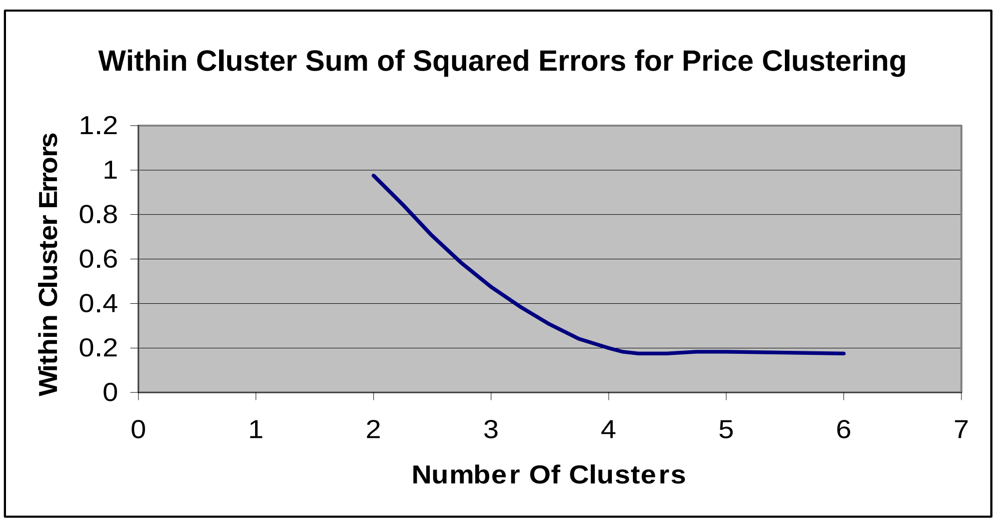


Figure 3.1: Within Cluster Sum of Squared Errors for Price Clustering

As seen from Figure 3.1, within cluster sum of squared errors reach almost a stable value after K=4. Because of this reason, the number of price groups is selected as 4 (K=4).

While setting each SKU's price attribute level, all SKUs are sorted according to their price/gram values in ascending order and price attribute levels from 1 to 4 are assigned for each SKU considering the mean values and instance numbers of each cluster. So attribute levels are obtained for each of the brand, package type, aroma and price attributes.

Fader and Hardie (1996) explain that each SKU should represent a unique combination of attributes. In accordance with this explanation, our SKUs in black tea category should be uniquely identified by their attribute levels. When 60 SKUs in the black tea category are identified by the attribute levels, it is understood that some SKUs have same attribute level combination mainly because of the added price attribute. However the data generation model is developed for SKUs with uniquely identified by attribute levels. To obtain this, 25 of 60 SKUs are eliminated in the black tea category. From here on, the study includes 35 SKUs that are uniquely identified by attribute levels in the black tea category.

Attribute levels of brand, package type, aroma and price attributes, which are inputs for data generation model, are obtained as a result of this computational approach. Table 3.1 shows attribute levels for brand, package type, tea aroma, and price attributes. In this study, original brand names are not used. Table 3.2 shows 35 SKUs and their attribute levels in the black tea category.

Attribute	**Level1**	**Level2**	**Level3**	**Level4**	**Level5**
Brand	A	B	C	D	E
Aroma	Rize	Filiz	Ceylon	Earl Grey	Breakfast
Package type	Bulk tea	Tea-bag	Teapot bag		
Price	1	2	3	4	

Table 3.1: SKU Attributes and Attribute Levels in the Black Tea Category

SKU	Brand	Tea Aroma	Package Type	Price Category
SKU 1	A	Rize	Bulk	1
SKU 2	A	Filiz	Bulk	1
SKU 3	A	Earl Grey	Bulk	2
SKU 4	B	Rize	Tea Bag	3
SKU 5	B	Rize	Tea Bag	4
SKU 6	B	Rize	Pot Bag	2
SKU 7	B	Earl Grey	Pot Bag	2
SKU 8	B	Earl Grey	Tea Bag	3
SKU 9	B	Earl Grey	Tea Bag	4
SKU 10	B	Filiz	Bulk	1
SKU 11	B	Rize	Bulk	1
SKU 12	C	Rize	Tea Bag	3
SKU 13	C	Rize	Pot Bag	2
SKU 14	C	Earl Grey	Bulk	2
SKU 15	C	Earl Grey	Tea Bag	3
SKU 16	C	Filiz	Bulk	1
SKU 17	C	Rize	Bulk	1
SKU 18	C	Breakfast	Tea Bag	3
SKU 19	D	Rize	Tea Bag	2
SKU 20	D	Rize	Tea Bag	3
SKU 21	D	Rize	Pot Bag	1
SKU 22	D	Filiz	Bulk	1
SKU 23	D	Rize	Bulk	1
SKU 24	E	Rize	Pot Bag	2
SKU 25	E	Rize	Bulk	1
SKU 26	E	Earl Grey	Bulk	2
SKU 27	E	Earl Grey	Tea Bag	4
SKU 28	E	Earl Grey	Pot Bag	2
SKU 29	E	Breakfast	Tea Bag	4
SKU 30	E	Ceylon	Tea Bag	4
SKU 31	E	Ceylon	Bulk	3
SKU 32	E	Ceylon	Pot Bag	2
SKU 33	E	Filiz	Tea Bag	4
SKU 34	E	Filiz	Bulk	1
SKU 35	E	Filiz	Pot Bag	2

Table 3.2: Attribute Levels for 35 SKUs in the Black Tea Category

3.3.2 Marketing Mix Instruments

Information about marketing mix instruments in the black tea category are obtained from the interviewed experts. This information is used in the calibration of model parameters related with marketing mix instruments. According to the information we obtained, the most observed marketing mix instruments in black tea category are price discount, display and advertisement. Companies use these instruments while they make promotions like big campaign, local campaign or tasting. During big campaigns all price discount, display and advertisement instruments, during local campaigns price discount and display instruments are applied. During tasting promotions only display instrument is applied. Table 3.3 shows our experts' opinion on frequency, length and impact of these promotions on the sales volumes.

Promotion Type	**Frequency**	**Length**	**Impact on Sales**
Big Campaign	Once in a year	1 month	100% increase
Local Campaign	4 times in a year	1 week	55% increase
Tasting	Once in two months	1 week	20% increase

Table 3.3: Promotion Types and Their Characteristics in the Black Tea Category

According to the category management expert we consulted with, grocery retail chains apply promotions in a pattern in that brands are ordered and it is important not to make promotions on one brand's similar SKUs at the same time. Besides this, promotions are typically applied on weekly basis. For instance if brand A's rize aroma, bulk tea is promoted in a week, then brand A's rize aroma tea-bag is not promoted.

Figure 3.2 shows the market shares of brands in the black tea category in Turkey. This figure is obtained from Akşam newspaper [37].

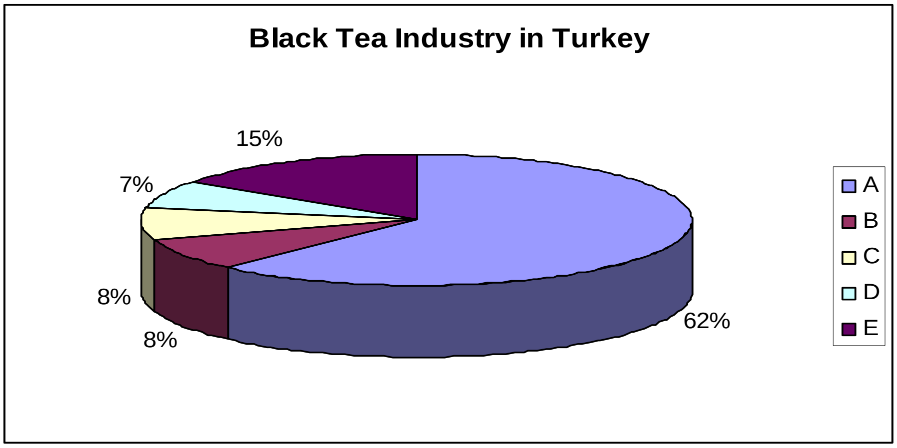


Figure 3.2: Market Shares in Black Tea Category in Turkey

3.3.3 Customer Segments

Information about characteristics of customer segments in black tea category is obtained from the product management expert. Below statements are acquired from the expert opinion:

- Customer loyalty is strong in black tea category towards both tea tastes and tea package types (bulk tea, tea-bag and teapot bag).
- There are three customer segments in the black tea category, which are bulk tea users, tea-bag users and teapot bag users. These segments differ in their sales volumes and responses to promotions.

Customer Segment	**Main Characteristic**	**Volume**	**Response to promotions**
1	Bulk tea users	80%	Very price sensitive
2	Teapot bag users	15%	Somewhat price sensitive
3	Tea-bag users	5%	Not so price sensitive

Table 3.4: Customer Segments and Their Characteristics in the Black Tea Category

- Customer loyalties towards tea tastes (brands) and package types are very strong. A customer may switch his/her brand if a competitor firm has applied a price discount. Especially bulk tea consumers are very sensitive to price discounts.
- A customer may switch the type of tea package once he/she experiments another package type. The percentage of switching package type after testing another package type is 10%. Bulk tea consumers do not leave bulk tea packages totally. If they were purchasing 100% of their tea need as bulk tea, after testing another package type, they would purchase 80% of their tea need as bulk tea and 20% of their tea need as tea-bag or teapot bag.
- Because most of the consumers (bulk tea users) are very price sensitive and may switch a small proportion of their tea package type usage, substitution can be seen mostly during promotions and rarely after promotions in this category.

3.4 Parameter Setting for Black Tea Category

In this part input parameter calibration for black tea category is explained. In order to capture the expert statements given in the previous part, experiments are needed on obtaining parameters like, base preferences ($a_r^{h,0}$), marketing mix scalar multipliers ($\beta_{price}^h, \beta_{disp}^h, \beta_{adv}^h$), attribute loyalty scalar multipliers (α_r^h) and lambda (λ_r).

Firstly an optimization is performed in Excel Solver, in order to obtain base preferences. Then, a grid search is performed to obtain marketing mix scalar multipliers. And finally attribute loyalty scalar multipliers and lambda are set.

Base preference vectors for each segment ($a_r^{s,0}$) should be set for the black tea category. In the data generation model, base preference vectors are household based ($a_r^{h,0}$). As explained before, each household *h* deterministically belongs to a single customer segment *s*, so the households belonging to the same customer segment have the same base preference vectors. However, their preferences can and do change over time with their purchase activity.

A base preference vector ($a_r^{s,0}$) for attribute *r* is a vector of size L_r. The attributes in the black tea category are brand attribute with A, B, C, D and E levels; aroma attribute with rize, filiz, ceylon, earl grey and breakfast levels; package type attribute with bulk tea, tea-bag and teapot bag levels and price group attribute with group 1, group 2, group 3 and group 4 levels. Considering all attributes with their attribute levels, there are 17 attribute levels in total in the black tea category. Households in a customer segment have same base preferences for each of the attribute levels. Because there are three customer segments in this category, in total there are 51 base preferences of attribute levels in the category.

Base preference vectors ($a_r^{h,0}$) are input parameters for the data generation model. The only expert information that would be used in obtaining the base preferences is the market shares of brands in the category. It is too complicated to obtain 51 parameters arbitrarily while aiming to capture five brands' market shares. This problematic situation brought the idea of using an optimization approach in order to obtain 51 base preference parameters.

Market shares of black tea brands are used as target values and the optimization approach aims to get closer to these values, while setting 51 base preference parameters. Because optimization is assumed to minimize the sum of squared errors between target brand share values and estimated brand share values, it forms a nonlinear optimization model. Solving such a model using Excel Solver may not give globally optimal solutions, but in any case the solutions will be useful for our problem. Because of this reason, conditions for global optimization are not checked.

As explained, target brand shares are the real market shares of brands in the black tea category in Turkey. The target brand shares can be seen in the following table:

Brand	Target brand share
A	0.62
B	0.08
C	0.08
D	0.07
E	0.15

Table 3.5: Target Brand Shares for Black Tea Category

Because promotions may increase the overall sales of the category, to capture such increases, we allow category growth. We assume that total of the market shares in the category as 0.8. Therefore, category growth is enabled and total of the market shares can reach up to 1.0. While assuming total market share is 0.8, all brand shares should be multiplied by 0.8. New target brand shares are found as seen from the below table:

Brand	Target Brand Share	Target Brand Share*0.8
A	0.62	0.496
B	0.08	0.064
C	0.08	0.064
D	0.07	0.056
E	0.15	0.120

Table 3.6: New Target Brand Shares for Black Tea Category

Data generation model simulation starts with an initialization part for one week, in order to obtain first attribute loyalty vectors (ATTLOY) for the black tea category. It is assumed that at the initialization there is no marketing mix instrument and all long term attribute loyalty effects (ATTLOY) are zero. By running the model in initialization, attribute loyalties are constituted only considering the base preferences toward SKU attributes. Then simulation continues for 176 weeks including marketing mix instruments and evolving attribute loyalty vectors with promotions. Base preference vectors for each segment ($a_r^{s,0}$) are estimated considering the initialization part of the simulation.

Because there is no long term attribute loyalty effect (ATTLOY) in initialization, preference vectors would be equal to base preference vectors for each customer segment.

$$a_r^{s,t} = a_r^{s,0} \tag{3.6}$$

There is no promotion effect in initialization, so SKU utilities would include only the preference component.

$$v_i^{s,t} = \sum_{r=1}^{R} e_{ir} a_r^{s,t} \tag{3.7}$$

The probability of a customer from segment *s* purchases SKU *i* is as seen below:

$$p_i^{s,t} = \frac{\exp(v_i^{s,t})}{1 + \sum_i \exp(v_i^{s,t})} \tag{3.8}$$

Because only the initialization part of the simulation model is considered, there is no need to use time indices (t). So, preference vector, utility and probability of purchase can be written as below,

$$a_r^s = a_r^{s,0} \qquad v_i^s = \sum_{r=1}^{R} e_{ir} a_r^s \qquad p_i^s = \frac{\exp(v_i^s)}{1 + \sum_i \exp(v_i^s)}$$

The nonlinear optimization model for base preference parameter estimation can be written in the following form:

$$Min \sum_j \left\{ \left(\sum_{\substack{i \\ all\ SKU\ i's\ belong \\ to\ brand\ j}} \frac{\exp(\sum_{r=1}^{R} e_{ir} a_r^s)}{1 + \sum_i \exp(\sum_{r=1}^{R} e_{ir} a_r^s)} - Target\ brand\ share_j \right)^2 \right\} \tag{3.9}$$

subject to

$$\forall\ a_{r,l}^s \geq 0$$

where
r *indicates the attributes* (*brand*, *variant*, *package type and price*)
e_{ir} *is a unit vector*,
j *indicates the brand* ($j = 1...5$),
$a^{s}_{r,l}$ *indicates* l^{th} *element of vector* a^{s}_{r}, *which is the base preference of attribute* r*'s* l^{th} *level for segment* s

There are 17 $a_r^{s,0}$ (customer segment base preference) parameters for each customer segment in our problem and in total we have 51 parameters. Besides this, we have 5 new target brand share values, to which we want our estimated brand shares to be considerably close. The parameter optimization has too many degrees of freedom. Because of these reasons target market shares of brands will be allocated to each customer segment. Here, the aim is obtaining finer results from the nonlinear model. The allocated new target brand share values can be seen below:

	Allocated New Target brand shares			
	S1	S2	S3	
A	0.5950	0.1100	0.0820	0.4966
B	0.0410	0.1460	0.1750	0.0634
C	0.0440	0.1350	0.1550	0.0631
D	0.0440	0.1000	0.1200	0.0560
E	0.0770	0.3000	0.2800	0.1210
			SUM	0.8000

Table 3.7: Allocated New Target Brand Shares for Black Tea Category

In this table S1, S2 and S3 indicates the customer segments of bulk tea users, teapot bag users and tea-bag users respectively. The allocation of brand shares to customer segments is performed considering the relative volumes of customer segments. The relative volumes of the customer segments are: 80% of the category is bulk tea users, 15% of the category is teapot bag users and 5% of the category is tea bag users.

When the nonlinear model in 3.9 is solved using Excel Solver, market shares seen in Table 3.8 are found.

Brand	S1	S2	S3
A	0.5954	0.1099	0.0821
B	0.0411	0.1460	0.1750
C	0.0440	0.1347	0.1549
D	0.0440	0.0999	0.1200
E	0.0771	0.3001	0.2801

Table 3.8: Market Shares Found by Excel Solver

Squared errors between found market shares and target market shares are given in Table 3.9.

	Squared errors			
	S1	S2	S3	
A	2.02E-11	5.29E-13	2.16E-12	2.29E-11
B	9.76E-10	1.85E-14	5.53E-13	9.77E-10
C	8.04E-12	3.25E-13	7.25E-12	1.56E-11
D	1.28E-12	4.87E-12	4.83E-13	6.63E-12
E	5.58E-12	3.65E-18	7.27E-13	6.31E-12
			SUM	1.03E-09

Table 3.9: Sum of Squared Errors Found by Excel Solver

The base preference parameters ($a_r^{s,0}$) are given in Table 3.10. These base preference parameters may not be globally optimal, but they are still acceptable for the data generation model. Since, there were considerable degrees of freedom when optimizing parameters (51 parameters with 15 data points). As seen from table 3.10, base preferences for brand attribute are in accordance with market shares of these brands.

	Brand					Variant				
Segment	A	B	C	D	E	Rize	Filiz	Ceylon	Earl Grey	Breakfast
S1(Bulk)	3.510403	0	0.178768	0.466576	0.216018	0.044359	0.094381	0.044359	0	0
S2(Pot)	0.790039	0.292129	0.348852	0	0.536626	0.618176	0.684696	0.835564	0.213219	0
S3(Bag)	0.247594	0.179347	0	0.200638	0.264831	0.402154	0.607849	0.670105	0.459072	0

	Package type			Price			
Segment	Bulk	Bag	Pot	1	2	3	4
S1(Bulk)	0.126694	0	0	0.099952	1.25E-05	3.59E-05	0
S2(Pot)	0.105332	0	0.670597	0.759019	0.168245	0.278973	0
S3(Bag)	0.441026	0.26754	0	0.492228	0.767415	1.12115	0

Table 3.10: Base Preferences Found by Excel Solver

After estimating the base preference parameters, response parameters ($\beta^s_{price}, \beta^s_{disp}, \beta^s_{adv}$) towards promotions should be determined considering the strong loyalty towards tea tastes (brand) and package types in this category. Marketing mix effects are considered in three ways: display only, price reduction and display, price reduction and display & and advertisement. While setting the response parameters, sales increase amounts obtained from experts are aimed to be captured. The response parameters should adhere to the following reactions normally observed in the market:

- Display only promotions increase the sales amount 20%.
- Display and price reduction only promotions increase the sales amount 55%.
- Display, price reduction and advertisement promotions at the same time increase the sales amount 100%.

Because it is known that how a display only promotion will affect the sales amount, firstly β^s_{disp} parameter are set. Then according to obtained β^s_{disp} values, β^s_{price} values are set. Finally, according to obtained β^s_{disp} and β^s_{price} values, β^s_{adv} parameters are calibrated.

While obtaining response parameters the aim is:

$$\begin{aligned}
&\textit{Minimizing} \sum_j (\textit{Sales impact}_j - \textit{Desired sales impact}_j)^2 \\
&\textit{when decision variables are}: \beta^s_{price}, \beta^s_{disp}, \beta^s_{adv} \textit{ respectively} \\
&\textit{subject to}: \textit{calibrated } a^s_r \textit{ values} \\
&\textit{where} \\
&\textit{Sales impact}_j = \left(\frac{(\textit{Share of brand } j \textit{ in promotion week}) - (\textit{Share of brand } j \textit{ in non promotional week})}{(\textit{Share of brand } j \textit{ in non promotional week})} \right)
\end{aligned} \tag{3.10}$$

A simulation procedure similar to grid search algorithm is used, while obtaining $\beta^s_{disp}, \beta^s_{price}, \beta^s_{adv}$ parameters respectively. After a few experiments, it is observed that $\beta^s_{disp}, \beta^s_{price}, \beta^s_{adv}$ cannot be larger than 0.5. Also it is known that $\beta^s_{disp}, \beta^s_{price}, \beta^s_{adv}$ values should be greater than or equal to 0. Simulation is realized for $\beta^s_{disp}, \beta^s_{price}, \beta^s_{adv}$ between 0.05 and 0.5 in steps of 0.05. Thus, 1000 experiments are obtained in every turn. Firstly β^s_{disp} parameters are obtained in the first turn, β^s_{price} and β^s_{adv} parameters are obtained in the second and third turns respectively.

According to grid search results, following response parameters ($\beta^s_{disp}, \beta^s_{price}, \beta^s_{adv}$) give the closer sales increase amounts to the targeted sales increases.

	Display	**Price**	**Advertisement**
S1(Bulk)	0.15	0.35	0.35
S2(Pot)	0.25	0.3	0.4
S3(Bag)	0.3	0.35	0.3

Table 3.11: Response Parameters Found by Grid Search

Lambda and alpha parameters (λ_r, α^h_r) that indicate the strength of consumers' past purchase behavior (long term attribute loyalty) should be selected as large as possible (near to 1) because of the strong consumer loyalty towards tea tastes and package types.

- Because customer loyalty is very strong in black tea category lambda (λ_r) parameters for aroma, package type and price attributes, are set to 0.9. In order to consider the switching between brands during promotions lambda (λ_r) for brand attribute is set as 0.1.
- α_r^h parameters are set to 0.4 for all attributes and all segments to show not only the customer loyalty but also the switching between customer segments.

3.5 Generated Data

Data for black tea category is obtained by using the base preferences ($a_r^{h,0}$), response parameters ($\beta_{price}^h, \beta_{disp}^h, \beta_{adv}^h$), lambda and alpha parameters (λ_r, α_r^h) that indicate the strength of consumers' past purchase behavior (long term attribute loyalty).

Data generation simulation started with an initialization part in which all marketing mix instruments were off and attribute loyalty vectors (ATTLOY) were equal to zero. In initialization first ATTLOY vectors are obtained and then simulation is performed for 176 weeks. Finally, weekly SKU by SKU sales data, for 176 weeks and 35 SKUs, is obtained. Because at the beginning of time period attribute loyalties were zero, first 20 weeks of data of 176 weeks is eliminated in order to obtain attribute loyalties different from zero for most of the households. Next 156 weeks of data is used in forecasting experiments in two sets: 104 weeks as training set and 52 weeks as test set. It is assumed that 10,000 households visit the store each week with a probability of 0.5. Data is generated with 15% noise level in order to obtain more realistic data. In simulation, noisy data is obtained by multiplying utility $(v_i^{h,t})$ with a noise factor as seen in Equation 3.11, in which distribution of random number is normal(0,1).

$$noise(15\%) = random\ number * (0.15)$$
$$\hat{v}_i^{h,t} = v_i^{h,t} * (1 + noise(15\%)) \quad (3.11)$$

As explained before, attribute loyalties evolve with purchasing activities of a household. After attribute loyalty vectors are calculated, they are added to preference vectors. In other words, because of attribute loyalty vectors preference vectors evolve with purchasing experiments. Figure 3.3 shows the preferences towards each of the 17 attribute levels in the preference vector of a specific household. As seen from this figure, intense increases in the preferences indicate the purchasing activity for the introduced attribute level and slow decreases indicate that considered household does not continue to buy items with the same attribute level.

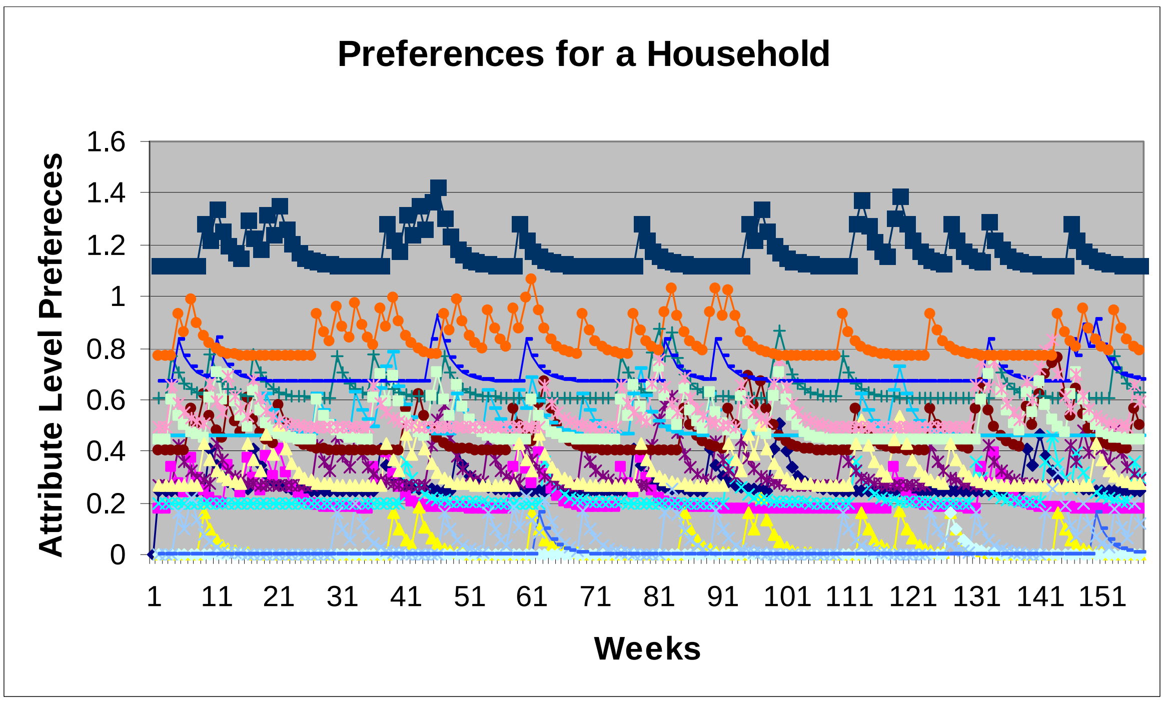


Figure 3.3: Preferences for a Specific Household

Because new SKU entrance is a common issue in retail industry, many forecasts need to be performed with less data for the new SKU than for other SKUs. Our aim is to compare forecasting methods according to accuracy, model simplicity and new SKU forecasting performance. The data generation model simulation gives the flexibility of obtaining data sets that have different characteristics.

In this study, three different sales data sets are generated in order to investigate experiments on data sets with different number of new SKUs. The first data set is explained above. The second data set includes a new SKU entering to the black tea category. 35^{th} SKU enters the category at week 101 as a new SKU. The third data set includes four new SKU entries in the category. The 10^{th}, 17^{th}, 33^{rd} and 35^{th} SKUs are new entering SKUs to the category at weeks 11, 71, 86 and 101 respectively.

First data set is the sales data of 35 SKUs for 156 weeks; second data set is the sales data of 34 SKUs for 156 weeks and 1 SKU for 56 weeks; third data set is the sales data of 31 SKUs for 156 weeks, 1 SKU for 146 weeks, 1 SKU for 86 weeks, 1 SKU for 71 weeks and 1 SKU for 56 weeks.

In the following chapter performed forecasting experiments for all three data sets are explained.

Chapter 4

FORECASTING

4.1 Introduction

We wish to forecast black tea category demand using statistical and machine learning techniques in the presence of marketing mix instruments. Forecasted sales data is a generated data and obtained by using the data generation model explained in Chapter 3. Sales data is generated for 35 SKUs and a time period of 176 weeks. Household purchases affect attribute loyalties. At the beginning of the time period it is assumed that there is no attribute loyalty, and attribute loyalties are obtained by the purchasing activities of households as explained in the data generation model in Chapter 3. Because of these reasons, first 20 weeks of data is eliminated and next 156 weeks of data is used for forecasting purposes in order to obtain a situation in which attribute loyalties are clearer. First 104 weeks of 156 weeks sales data is considered as training data set and remaining 52 weeks of data is considered as test data set. In addition to these, data is generated with a noise level of 15%.

New SKU entrance to a category is a common issue for retail industry. Because of this reason, new SKU entrances to category is examined by working on three different data sets that have different numbers of new SKU entrances. First data set is obtained for a black tea category with no new SKU entrance, second data set includes only one new SKU entrance to category and third data set includes four new SKU entrances to category. On the first data set, only preliminary experiments are run by using statistical techniques. On the second and third data sets, comprehensive experiments are performed with both statistical and data mining techniques.

In this study, exponential smoothing with lift adjustment, linear regression, regression tree and support vector regression with several kernel types are used as forecasting techniques. These techniques are compared according to the accuracy of total category, convenience (simplicity) of the model and performance in forecasting new SKU demand. In order to handle convenience issue, forecasting is implemented in two ways. In the first approach, individual data sets for each SKU in the category are used in order to obtain different forecasting models for each SKU. In the second approach, a combined data set including required data for all SKUs is used in order to obtain a single forecasting model.

Weka 3.5 is used as the data mining tool in this study. Linear regression, SMO reg and Rep tree functions of Weka 3.5 are used for support vector regression and regression tree applications respectively. SMO reg function implements Smola and Scholkopf [1]'s sequential minimal optimization algorithm for training a support vector regression model. Rep tree function is a decision tree learner and builds a decision/regression tree.

In section 4.2, data set with no new SKU is investigated and linear regression and exponential smoothing with lifts are applied on the individual data sets of each SKU.

Section 4.3 includes broader experiments. Data set with one new SKU entrance is examined and linear regression, regression tree and support vector regression are used as forecasting techniques. In this section sales data of each SKU is normalized between 0 and 1. Some SKUs in the category have comparatively bigger sales volumes than the other SKUs. While obtaining a forecasting model for the combined data set, forecasting methods give very high weights to SKUs with high sales volume and very low weights to SKUs with low sales volume, with the consequence that the lower sales volume SKUs end up with poor accuracy. In order to avoid such a situation the normalization is applied on sales amounts and each SKU's sales amounts are transformed into a value between 0 and 1.

$$\hat{Y}_i(t) = \frac{\left(Y_i(t) - \min\limits_{t=1\ldots104}(Y_i(t))\right)}{\left(\max\limits_{t=1\ldots104}(Y_i(t)) - \min\limits_{t=1\ldots104}(Y_i(t))\right)} \tag{4.1}$$

Here, $\hat{Y}_i(t)$ denotes SKU$_i$'s normalized sales amount in week t. Minimum and maximum values of Y_i(t) are taken from training set and applied to whole sales data of SKU$_i$. In addition to this, in section 4.3, forecasting experiments on selected 5 SKUs are explained and then experiments on the whole category are performed on both individual and combined data sets. In experiments, past sales data is included in data sets in these two ways: sales with one week lag and sales with four weeks lag.

Section 4.4 contains similar experiments on the data set with four new SKU entrances. Data is normalized between 0 and 1, and linear regression, regression tree and support vector regression are used as forecasting techniques. In this section, past sales data is included only by one week lag. Besides this, forecasting is performed on both individual and combined data sets.

In section 4.5, a brief summary of this chapter and comments on forecasting results are given.

4.2 Preliminary Forecasting Experiments for Sales Data Including No New SKU

Linear regression and exponential smoothing are the most typical ways of forecasting demand. In this section a data set for 35 SKUs and 156 weeks of black tea category is analyzed by linear regression and exponential smoothing with lift adjustment which are considered as statistical techniques in this study. Data is generated with a noise level of 15% using the data generation model explained in Chapter 3. Data of first 104 weeks of 156 weeks is considered as the training data set and last 52 weeks is considered as the test data set. Figure 4.1 shows the total sales amounts of the black tea category for the time period of 156 weeks. The aim of this section is to compare the performance of two selected statistical techniques on the data set with no new SKU and provide a benchmark performance for the data mining methods.

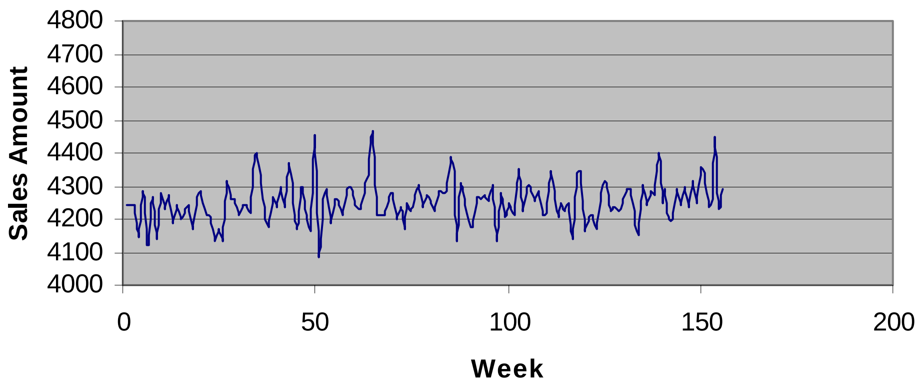


Figure 4.1: Total Sales Amounts for Black Tea Category for 156 weeks

In exponential smoothing experiments promotions are considered by obtaining lift values for each promotional variable. For example a promotional activity of price reduction, advertisement and display are available in week 21 for SKU 6. Figure 4.2 shows the sales amounts of SKU 6 for the training time period. The lift factor would indicate the increase in the sales amount because of a promotional activity.

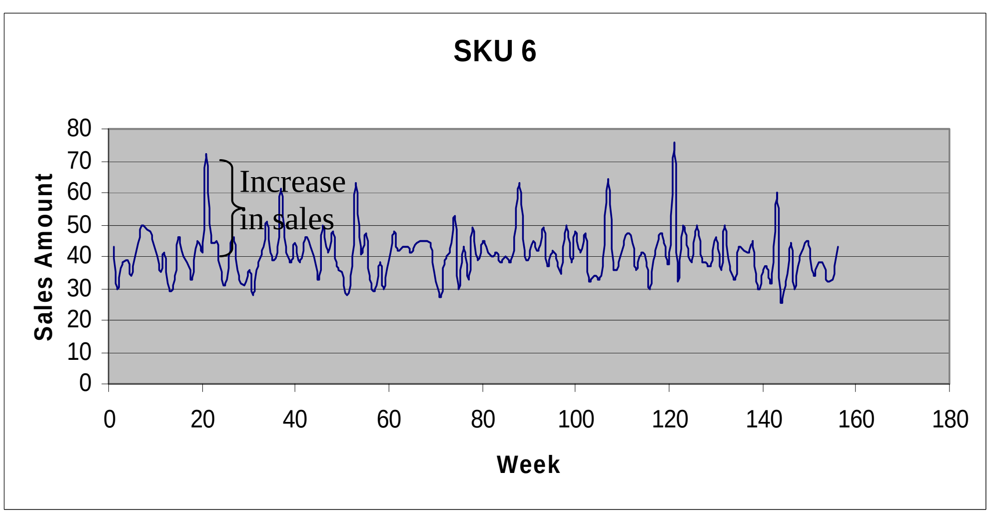


Figure 4.2: Sales Amounts of SKU 6 in the Training Period

Lift factors are estimated for each SKU and promotion type combination. A data set that includes all sales and promotion data for all SKUs is considered as the combined data set that includes many variables. Because exponential smoothing with lifts would require lift estimations for each SKU and promotion variable combination, there would be a huge computation in experiments with exponential smoothing. In order to avoid such a situation, only individual data sets that include only the considered SKU's sales and promotion data are used in this section. Because data is obtained only in individual form, there is no need to normalize sales data. Figure 4.3 shows the template of individual data set in this section.

$Y_i(t)$	$Disp_i$	$PriceDisp_i$	$PriceDispAdv_i$	$Y_i(t\text{-}1)$
TRAINING DATA (104 weeks)				
TEST DATA (52 weeks)				

Figure 4.3: Template of Individual Data Set for Experiments in Section 4.2

Here, Y_i indicates the sales amount of SKU_i, $Disp_i$ indicates if the display marketing mix effect occurs or not for SKU_i in the considered week, $PriceDisp_i$ indicates if both of the display and price discount marketing mix effects occur or not for SKU_i in the considered week, $PriceDispAdv_i$ indicates if all of display, price discount and advertisement marketing mix effects occur for SKU_i in the considered week. $Disp_i$, $PriceDisp_i$ and $PriceDispAdv_i$ are binary variables. Y_i(t-1) indicates the sales amount of SKU_i for week t-1. While obtaining an individual data, promotional data that belong to the considered SKU is included in the data set. In section 4.3 and section 4.4 individual data sets include promotional data of not only the considered SKU but also all SKUs in the category. As seen in Figure 4.3, each SKU has a training data set of size 104×5 and a test data set of size 52×5.

4.2.1 Experiments with Linear Regression

Because sales amounts in our data are affected by promotional activities, we need to use a multiple linear regression function in order to capture all promotional activities. Besides this, the other variable of the multiple linear regression would be the sales data of week t-1. The following linear regression is performed by using the stepwise regression function of Minitab software.

$$\hat{Y}(t) = \beta_0 + \beta_1 Y(t-1) + \beta_2 Disp(t) + \beta_3 \Pr iceDisp(t) + \beta_4 \Pr iceDispAdv(t) \tag{4.2}$$

where $Y(t-1)$ indicates the sales amount in week t-1 for the considered SKU and $\hat{Y}(t)$ denotes the predicted sales amount in week t. Disp, PriceDisp and PriceDispAdv indicate the marketing mix effects of display only; display and price reduction together; and display, price reduction and advertisement together, respectively. This function is run for all 35 SKUs in the category.

For each SKU in the category, an individual linear regression model is developed. Coefficients of the variables in the regression are found by Minitab and MAPE (Mean Absolute Percentage Error) values are estimated according to these coefficients. Table 4.1 shows average sales volumes of SKUs for 156 weeks which is total of training and test time period, obtained regression coefficients and MAPE values for each SKU. In addition, the coefficient of determination (R^2) values are obtained from experiments with Minitab and shown in Table 4.1. The coefficient of determination indicates the percentage of the variability in the sales volume explained by the model.

	Average Sales	Multiple Linear Regression Estimated Parameters					R^2	MAPE
		Constant Coefficient	Disp Coefficient	PriceDisp Coefficient	PriceDispAdv Coefficient	Y(t-1) Coefficient		
SKU1	989.045	965.356	190.644	883.144	0.000	0.000	87.82	0.033
SKU2	1035.019	1013.426	296.574	894.074	0.000	0.000	92.80	0.032
SKU3	812.859	813.558	0.000	0.000	0.000	0.000	68.52	0.073
SKU4	35.385	35.297	0.000	10.703	34.703	0.000	42.63	0.097
SKU5	26.910	26.745	0.000	5.755	50.812	0.000	39.25	0.167
SKU6	40.929	40.564	0.000	22.436	29.436	0.000	44.52	0.145
SKU7	32.763	31.931	0.000	6.069	30.069	0.000	44.18	0.172
SKU8	31.135	23.924	0.000	8.349	20.452	0.022	53.63	0.152
SKU9	22.840	27.395	0.000	0.000	0.000	-0.020	21.24	0.150
SKU10	53.397	51.941	0.000	24.059	56.059	0.000	77.30	0.119
SKU11	50.077	58.855	0.000	28.952	53.172	-0.019	68.53	0.143
SKU12	38.128	31.034	0.000	19.019	37.589	0.016	26.56	0.123
SKU13	45.141	44.564	0.000	20.936	46.436	0.000	48.60	0.139
SKU14	37.205	36.673	0.000	18.827	32.327	0.000	50.62	0.174
SKU15	33.391	33.475	0.000	16.525	28.525	0.000	41.11	0.186
SKU16	58.038	57.574	0.000	28.926	54.426	0.000	80.83	0.117
SKU17	55.929	55.119	0.000	26.881	53.881	0.000	82.86	0.117
SKU18	27.276	27.048	0.000	11.973	24.067	0.000	35.69	0.188
SKU19	37.763	37.567	0.000	0.000	0.000	0.000	40.00	0.220
SKU20	42.436	42.186	6.231	24.814	40.962	0.000	56.00	0.153
SKU21	58.256	58.471	10.243	32.529	57.014	0.000	82.00	0.105
SKU22	60.917	59.804	11.351	30.696	61.216	0.000	71.20	0.152
SKU23	57.686	58.177	9.452	29.824	0.000	0.000	57.28	0.111
SKU24	52.635	52.178	8.567	29.822	61.822	0.000	68.74	0.121
SKU25	64.058	61.400	15.600	35.600	76.600	0.000	79.39	0.099
SKU26	41.103	39.990	7.242	19.510	43.010	0.000	50.55	0.149
SKU27	28.788	28.128	3.935	13.373	26.924	0.000	37.86	0.167
SKU28	41.096	41.079	0.000	8.421	17.921	0.000	52.64	0.174
SKU29	25.526	25.433	0.000	0.000	0.000	0.000	35.29	0.184
SKU30	35.756	35.812	6.345	16.188	31.188	0.000	65.43	0.170
SKU31	60.968	59.490	16.510	34.510	51.510	0.000	78.87	0.115
SKU32	58.846	57.545	0.000	31.955	83.455	0.000	80.87	0.134
SKU33	35.904	35.644	4.624	17.845	37.094	0.000	60.21	0.140
SKU34	67.885	65.822	0.000	40.678	67.178	0.000	76.23	0.111
SKU35	56.115	55.270	12.730	38.730	58.730	0.000	80.08	0.144

Table 4.1: Multiple Linear Regression Coefficients and MAPE values

4.2.2 Experiments with Exponential Smoothing with Lifts

Simple exponential smoothing with lift adjustment is used as an alternative forecasting method for the generated data. It is applied on individual data sets of each SKU and lifts are estimated for each SKU and variable combination. Lift factor indicates the increase in the sales volume that occurs by the promotion. In order to apply exponential smoothing, promotional increases on sales amounts are eliminated in two different approaches. After eliminating promotional increases, lift factors are estimated for the training data set. Using lift factors on test set, forecasted amounts and MAPE values are estimated.

In the first promotional increase elimination approach, the sales amounts in promotional weeks are replaced with a new calculated amount, no promotional sales, in order to obtain the lift factors of promotions.

If there exist a promotion in week t for SKU_i, then sales amount of SKU_i in week t, is replaced by $\frac{Y_i(t-1)+Y_i(t+1)}{2}$. So the effect of the promotion on sales amount would be cancelled.

Applying this replacement in all promoted weeks, a sales data with promotion effects removed is obtained. These sales amounts are forecasted by using exponential smoothing method with $\alpha = 0.2$.

Sales amounts increase in promotional weeks. In order to obtain the magnitude of this increase a “lift” factor is calculated by dividing the sales amount in the promotional week by exponentially smoothed non promotional sales amount for that week. Finally average lift factors are obtained for all promotions (Disp, PriceDisp and PriceDispAdv). This methodology is performed as explained in Figure 4.4. According to this figure, for an SKU firstly promotional weeks are defined, then non promotional sales are obtained by the replacement explained in this figure and then exponential smoothing is applied on the sales amounts with promotional effects removed and finally lift factors are estimated.

Promotion	**Sales Amounts**	**Non Promotional Sales**	**Exponential smoothing for sales with promotional effects removed (alpha=0.2)**	**Lift Factor**
	$Y(t=1)$	$S(t=1)$	$\hat{S}(t=1)$	
	$Y(t=2)$	$S(t=2)$	$\hat{S}(t=2)$	
	$Y(t=3)$	$S(t=3)$	$\hat{S}(t=3)$	
	$Y(t=4)$	$S(t=4)$	$\hat{S}(t=4)$	
✓	$Y(t=5)$	$S(t=5)=(Y(t=4)+Y(t=6))/2$	$\hat{S}(t=5)$	$Y(t=5)/S(t=5)$
	$Y(t=6)$	$S(t=6)$	$\hat{S}(t=6)$	
	$Y(t=7)$	$S(t=7)$	$\hat{S}(t=7)$	
	$Y(t=8)$	$S(t=8)$	$\hat{S}(t=8)$	
	$Y(t=9)$	$S(t=9)$	$\hat{S}(t=9)$	
✓	$Y(t=10)$	$S(t=10)=(Y(t=9)+Y(t=11))/2$	$\hat{S}(t=10)$	$Y(t=10)/S(t=10)$
	$Y(t=11)$	$S(t=11)$	$\hat{S}(t=11)$	
	$Y(t=12)$	$S(t=12)$	$\hat{S}(t=12)$	
				Average Lift

Figure 4.4: Lift Estimation for Exponential Smoothing

In Figure 4.4, $Y(t)$ indicates the sales of considered SKU in week t and $S(t)$ indicates the sales with no promotional effect. $\hat{S}(t)$ denotes the predicted sales with no promotional effect. Lift factor indicates the ratio between real sales amount and sales amount with no promotional effect.

$$\hat{S}(t+1) = 0.2 * S(t) + (1-0.8) * \hat{S}(t)$$
$$Lift\ Factor = Y(t)/S(t) \tag{4.3}$$

Average lift factors of promotions (Disp, PriceDisp and PriceDispAdv) are found for 35 SKUs in training data. Then, using these average lift factors forecasted sales amounts are found and MAPE values are calculated for test data sets. Lift factors for marketing mix elements and MAPE values that are found according to exponential smoothing ($\alpha = 0.2$) results can be seen in Table 4.2.

In the second promotional increase elimination approach, obtained lifts are called as “Real Lifts”. Real lifts for an SKU are estimated by generating a new data set for 35 SKUs and 156 weeks assuming the promotions for considered SKU are off and promotions for other SKUs are on. This new generated data set would not include sales amount increase because of promotions. Sales amounts of the considered SKU in new generated data set are used as the non promotional sales shown in Figure 4.4. In order to obtain lift factors for all SKUs in the category, new data sets are generated for each SKU. Real lift factors and MAPE values obtained by second approach of promotional increase elimination are given in Table 4.3.

Estimated lift factors in Table 4.2 and real lift factors in Table 4.3 are within 5% of each other. Beside this, MAPE values in these tables are very close. This similarity indicates that first approach used in exponential smoothing for eliminating increased sales amounts is as good as obtaining new data set with no promotional increase. Because we generate data by the data generation simulation, we have an opportunity of comparing two approaches for lift estimation. In Table 4.2 and Table 4.3 there is no significant difference between two approaches.

MAPE values found by multiple linear regression experiments in Table 4.1 are smaller than MAPE values found by exponential smoothing in Table 4.2 and Table 4.3. Because of this reason, multiple linear regression is used for the experiments in next sections as the only statistical forecasting technique. It is found that exponential smoothing gives out worse accuracy levels than linear regression.

	Average Disp Lift	Average PriceDisp Lift	Average PriceDispAdv Lift	MAPE
SKU1	1.241	1.623	0.000	0.035
SKU2	1.293	1.703	0.000	0.033
SKU3	0.000	0.000	0.000	0.079
SKU4	0.000	1.337	1.725	0.167
SKU5	0.000	1.376	1.670	0.184
SKU6	0.000	1.594	1.757	0.148
SKU7	0.000	1.331	2.185	0.175
SKU8	0.000	1.471	1.738	0.168
SKU9	0.000	0.000	0.000	0.179
SKU10	0.000	1.480	2.111	0.141
SKU11	0.000	1.435	2.349	0.147
SKU12	0.000	1.647	1.982	0.137
SKU13	0.000	1.536	2.083	0.142
SKU14	0.000	1.346	1.612	0.180
SKU15	0.000	1.345	1.577	0.187
SKU16	0.000	1.427	1.862	0.125
SKU17	0.000	1.528	1.956	0.131
SKU18	0.000	1.432	1.872	0.190
SKU19	0.000	0.000	0.000	0.224
SKU20	1.127	1.599	0.000	0.161
SKU21	1.145	1.393	0.000	0.106
SKU22	1.124	1.491	0.000	0.162
SKU23	1.122	1.487	0.000	0.114
SKU24	1.140	1.566	2.272	0.124
SKU25	1.347	1.587	2.213	0.103
SKU26	1.123	1.528	1.846	0.160
SKU27	1.145	1.443	1.836	0.170
SKU28	0.000	1.474	1.554	0.175
SKU29	0.000	0.000	0.000	0.205
SKU30	1.145	1.299	1.798	0.171
SKU31	1.328	1.679	1.776	0.123
SKU32	0.000	1.527	2.123	0.144
SKU33	1.103	1.297	1.763	0.148
SKU34	0.000	1.668	1.910	0.115
SKU35	1.269	1.632	1.966	0.145

Table 4.2: Exponential Smoothing Lift Factors and MAPE values

	Avg D Lift	Avg PD Lift	Avg PDA Lift	MAPE
SKU1	1.299	1.584	0.000	0.037
SKU2	1.317	1.682	0.000	0.034
SKU3	0.000	0.000	0.000	0.075
SKU4	0.000	1.412	1.826	0.142
SKU5	0.000	1.451	1.773	0.172
SKU6	0.000	1.580	1.720	0.152
SKU7	0.000	1.372	2.363	0.177
SKU8	0.000	1.516	1.704	0.160
SKU9	0.000	0.000	0.000	0.162
SKU10	0.000	1.512	2.062	0.126
SKU11	0.000	1.454	2.268	0.149
SKU12	0.000	1.616	2.014	0.133
SKU13	0.000	1.516	2.125	0.144
SKU14	0.000	1.387	1.724	0.176
SKU15	0.000	1.362	1.625	0.187
SKU16	0.000	1.453	1.834	0.124
SKU17	0.000	1.547	2.012	0.122
SKU18	0.000	1.472	1.912	0.191
SKU19	0.000	0.000	0.000	0.225
SKU20	1.132	1.562	0.000	0.159
SKU21	1.182	1.431	0.000	0.107
SKU22	1.135	1.521	0.000	0.160
SKU23	1.124	1.493	0.000	0.113
SKU24	1.130	1.530	2.301	0.126
SKU25	1.332	1.613	2.119	0.110
SKU26	1.163	1.532	1.813	0.154
SKU27	1.123	1.424	1.824	0.172
SKU28	0.000	1.503	1.735	0.177
SKU29	0.000	0.000	0.000	0.209
SKU30	1.153	1.324	1.773	0.172
SKU31	1.322	1.625	1.824	0.122
SKU32	0.000	1.543	2.012	0.142
SKU33	1.109	1.342	1.823	0.149
SKU34	0.000	1.624	1.925	0.114
SKU35	1.284	1.593	1.992	0.147

Table 4.3: Exponential Smoothing Real Lift Factors and MAPE values

4.3 Comprehensive Forecasting Experiments for Sales Data Including 1 New SKU

Because new SKU entrance is a common issue in retail industry, we investigate this situation in our study. In this section, forecasting is performed on sales data that includes one new SKU entrance. New SKU enters the category later than the other SKUs which are in the category in all weeks of the considered time period. Because of this reason, new SKU would have less training data then the other SKUs in the category.

Sales data is obtained using the data generation model simulation that is explained before. Sales data is generated for 35 SKUs and a time period of 156 weeks. First 104 weeks of data is considered as training data set and other 52 weeks of data is considered as test data set. In addition to these, data is generated with a noise level of 15%. Obtained sales data is normalized between 0 and 1 as explained in section 4.1.

The thirty fifth SKU is a new entering SKU to the category. At the beginning of the time period there are 34 SKUs in the category and at the 101st week a new SKU enters the category as the 35th SKU. In total we have a sales data of 35 SKUs for 156 weeks. So the 35th SKU has only 4 weeks of training data. Graphs that show sales amounts for 35 SKUs can be found in Appendix III.

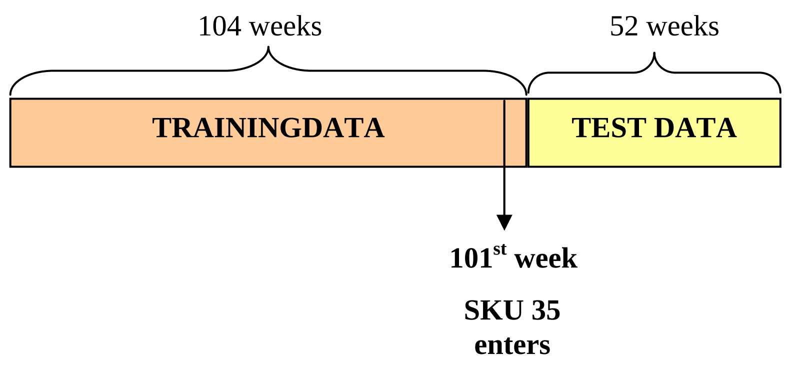


Figure 4.5: Time period of the Data Set with 1 New SKU

4.3.1 Experiments on Selected 5 SKUs

We firstly performed forecasting experiments on the selected 5 SKUs that have different characteristics and then further analysis are performed on the whole black tea category.

5 SKUs are selected according to their average sales volume and average promotion frequency. Forecasting methods are applied firstly on these selected 5 SKUs in order to understand for which sales volume and promotion frequency combination a better accuracy can be obtained with various techniques.

These five SKUs and their attributes can be seen in Table 4.4.

SKU NO	Brand	Variant	Package Type	Price Group
2	Caykur	Filiz	Bulk	1
7	Deren	Earl Grey	Pot Bag	2
23	Label Private	Rize	Bulk	1
25	Lipton	Rize	Bulk	1
27	Lipton	Earl Grey	Tea Bag	4

Table 4.4: Attributes of Selected 5 SKUs

Average sales volumes and promotion frequencies for these SKUs during the considered time period of 156 weeks are given below. These values also can be found in Table 4.10.

SKU NO	Average Sales Volume	Promotion Frequency
2	1035.02	5
7	32.76	7
23	57.69	5
25	64.06	9
27	28.79	9

Figure 4.6 shows these 5 SKU's average sales volumes and average promotion frequencies relative to each other.

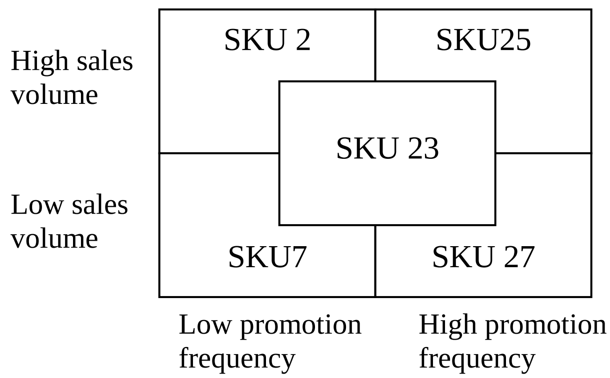


Figure 4.6: Relative Promotion Frequencies and Sales Volumes of Selected 5 SKUs

Linear regression, regression tree, support vector regression with polynomial kernel in degree 1, support vector regression with polynomial kernel in degree 2 and support vector regression with RBF kernel are used to forecast selected 5 SKUs individually and combined. Also different C values are experimented with mentioned kernels.

Individual Data

While applying the linear regression, regression tree and support vector regression, template of the forecasting data of an individual SKU is as seen in Figure 4.7.

Y_i	**$Disp_i$**	**$PriceDisp_i$**	**$PriceDispAdv_i$**	**…**	**$Disp_N$**	**$PriceDisp_N$**	**$PriceDispAdv_N$**	**$Y_i(t-1)$**
TRAINING DATA (104 weeks)								
TEST DATA (52 weeks)								

Figure 4.7: Template for the Individual Data Set

Here, Y_i indicates the normalized sales amount of SKU_i and Y_i(t-1) indicates the normalized sales amount of SKU_i for week t-1. $Disp_i$, $PriceDisp_i$, $PriceDispAdv_i$ are the same with the ones explained in section 4.2. While obtaining an individual data set, promotional inpus that belong to all of the SKUs in the category are included and normalized sales data of only the considered SKU is included.

Because sales data is normalized between 0 and 1, we report MAE (Mean Absolute Error) as the measure of forecast accuracy. Below table shows MAE values obtained from forecasting selected SKUs in individual data set format.

In experiments we followed two approaches for the considered past data. In the fist approach only Y_i(t-1) is considered as the past data, and in the second approach we considered four weeks of lagged data. We need to see if an improvement will occur when we add more past data to our data set. So we added four weeks lagged promotion and sales data. $Disp_i$(t-1), $PriceDisp_i$(t-1), $PriceDispAdv_i$(t-1), $Disp_i$(t-2), $PriceDisp_i$(t-2), $PriceDispAdv_i$(t-2), $Disp_i$(t-3), $PriceDisp_i$(t-3), $PriceDispAdv_i$(t-3), $Disp_i$(t-4), $PriceDisp_i$(t-4), $PriceDispAdv_i$(t-4) attributes for own SKU are added to our data as four weeks lagged promotion data and $Sales_i$(t-2), $Sales_i$(t-3) and $Sales_i$(t-4) attributes are added to our data set as sales amounts in weeks t-2, t-3 and t-4 respectively.

Table 4.5 shows the MAE values for applied forecasting techniques for selected SKUs in individual data sets. In this table, SVR Poly (deg1) (C=1) indicates that the considered forecasting technique is support vector regression with polynomial kernel in degree 1 with C value 1. Also, SVR RBF (C=1) indicates that the considered technique is support vector regression with radial basis function kernel with C value 1.

	Data Set with 1 Week Lagged Sales					Data Set with 4 Weeks Lagged Sales				
	SKU2	SKU7	SKU23	SKU25	SKU27	SKU2	SKU7	SKU23	SKU25	SKU27
Linear Regression	0.0197	0.1744	0.1039	0.0764	0.1672	0.0204	0.1746	0.1045	0.0801	0.1702
Regression Tree	0.0252	0.1655	0.1043	0.0929	0.1610	0.0254	0.1648	0.1052	0.0952	0.1714
SVR Poly (deg1) (c=1)	0.0213	0.1688	0.1072	0.0799	0.1788	0.0217	0.1701	0.1073	0.0927	0.1816
SVR Poly (deg2) (c=1)	0.0210	0.1714	0.1095	0.0789	0.1893	0.0224	0.1752	0.1102	0.0912	0.1942
SVR RBF (c=1)	0.0382	0.1587	0.1012	0.0887	0.1694	0.0388	0.1592	0.1132	0.1142	0.1788
SVR Poly (deg1) (c=0.5)	0.0212	0.1687	0.1070	0.0797	0.1786	0.0216	0.1695	0.1059	0.0945	0.1903
SVR Poly (deg2) (c=0.5)	0.0210	0.1710	0.1096	0.0788	0.1891	0.0212	0.1736	0.1094	0.0984	0.1924
SVR RBF (c=0.5)	0.0386	0.1591	0.1016	0.0889	0.1701	0.0387	0.1612	0.1128	0.1011	0.1805
SVR Poly (deg1) (c=10)	0.0213	0.1688	0.1072	0.0799	0.1788	0.0214	0.1698	0.1103	0.0941	0.1895
SVR Poly (deg2) (c=10)	0.0219	0.1716	0.1102	0.0793	0.1897	0.0221	0.1742	0.1112	0.0932	0.1906
SVR RBF (c=10)	0.0327	0.1598	0.1019	0.0894	0.1704	0.0329	0.1614	0.1142	0.1121	0.1896

Table 4.5: MAE Values for Selected 5 SKUs in Individual Data Sets

According to Table 4.5, MAE values for data set with lag 4 are worse than MAE values for data set with lag 1 for all methods applied to each considered SKU. Because of this reason, data set with lag 4 will not be used as a comparison tool for forecast results anymore. It is decided not to add more past data to our data set for the experiments ahead.

Table 4.5 shows that best accuracy levels for SKU 2 and SKU 25 are obtained by using linear regression. SKU 7 and SKU 23 are best predicted by SVR with RBF kernel. SKU 27 obtains the minimum MAE when the forecasting is performed with SVR with polynomial kernel in degree 1. Up to this point, it is obvious that SKUs with high volume are better predicted by linear regression and SKUs with low and middle volume are best predicted by SVR techniques.

Combined Data

As explained before, forecast data is formed in two ways which are obtaining individual data sets for each SKU and obtaining a combined data set that includes all SKUs. Up to now, the reported experiments considered individual data sets. Use of combined data set for forecasting leads to simpler models. Besides, we will be able to see if more accurate results can be derived this way.

The combined data set is considered as the combination of individual data sets and dummy variables that show which instance belongs to which SKU. The template for the combination of individual data sets of each SKU and dummy variables can be seen in Figure 4.8.

Y	$Disp_i$	$PriceDisp_i$	$PriceDispAdv_i$	…	$Disp_N$	$PriceDisp_N$	$PriceDispAdv_N$	Y(t-1)	SKU_i	SKU_{i+1}	…	SKU_N
TRAINING DATA of SKU_i (104 weeks)												
TEST DATA of SKU_i (52 weeks)												
Y	**$Disp_i$**	**$PriceDisp_i$**	**$PriceDispAdv_i$**	**…**	**$Disp_N$**	**$PriceDisp_N$**	**$PriceDispAdv_N$**	**Y(t-1)**	**SKU_i**	**SKU_{i+1}**	**…**	**SKU_N**
TRAINING DATA of SKU_{i+1} (104 weeks)												
TEST DATA of SKU_{i+1} (52 weeks)												
⋮												
Y	**$Disp_i$**	**$PriceDisp_i$**	**$PriceDispAdv_i$**	**…**	**$Disp_N$**	**$PriceDisp_N$**	**$PriceDispAdv_N$**	**Y(t-1)**	**SKU_i**	**SKU_{i+1}**	**…**	**SKU_N**
TRAINING DATA of SKU_N (104 weeks)												
TEST DATA of SKU_N (52 weeks)												

Figure 4.8: Template for the Combined Data Set

Attributes in a combined data set are similar to attributes in an individual data set. Here, $Disp_i$, $PriceDisp_i$, $PriceDispAdv_i$ are the same attributes with the ones explained for individual data set template. Y and Y(t-1) denote the normalized amount of SKUs in week t and t-1 respectively. In Figure 4.8, N indicates the total number of SKUs in the category which is 35 for our black tea category. SKU_i, …, SKU_N are the dummy variables that denote if the considered instance belongs to SKU_i or not. The number of dummy variables will be equal to the number of SKUs to be forecasted. By using dummy variable attributes, we are indicating which instance belongs to which SKU in the combined data set. As in the individual data sets, the combined data set contains all promotional attributes ($Disp_i$, $PriceDisp_i$, $PriceDispAdv_i$) that belong to all SKUs to be forecasted. Table 4.6 shows MAE values obtained from forecasting the combined data set for the selected 5 SKUs.

	MAE FROM COMBINED DATA SET				
	SKU2	**SKU7**	**SKU23**	**SKU25**	**SKU27**
Linear Regression	0.0443	0.1614	0.0996	0.0908	0.1581
Regression Tree	0.0470	0.1862	0.1114	0.1067	0.1640
SVR Poly (deg1) (c=1)	0.0398	0.1552	0.0978	0.0906	0.1672
SVR Poly (deg2) (c=1)	0.0220	0.1655	0.1107	0.0815	0.1802
SVR RBF (c=1)	0.0397	0.1582	0.0992	0.0921	0.1626
SVR Poly (deg1) (c=0.5)	0.0396	0.1551	0.0980	0.0905	0.1673
SVR Poly (deg2) (c=0.5)	0.0222	0.1656	0.1106	0.0817	0.1800
SVR RBF (c=0.5)	0.0397	0.1580	0.0988	0.0917	0.1625
SVR Poly (deg1) (c=10)	0.0398	0.1552	0.0988	0.0908	0.1672
SVR Poly (deg2) (c=10)	0.0224	0.1661	0.1109	0.0820	0.1809
SVR RBF (c=10)	0.0408	0.1586	0.0998	0.0928	0.1632

Table 4.6: MAE Values for Selected 5 SKUs in Combined Data Set

Considering Table 4.5 and Table 4.6 that shows MAE obtained from individual data sets and combined data set respectively, it is seen that when C values are changed for several kernel types in support vector regression, MAE values do not change distinctly. On average a change value less than 0.001 is seen when C value is changed to 0.5 or 10. Considering this result, C value is decided to be taken as 1 for all support vector regression applications.

MAE values for both individual and combined data sets give an important insight about the chosen forecasting methods and SKUs. If we put MAE values of each method to the schema that is drawn for relative promotion frequencies and sales volumes for the selected 5 SKUs, it is recognized that SKUs in high volume are better predicted than SKUs in low volume for a given level of promotion frequency. Also for most of the chosen forecasting methods, SKUs having low promotion frequency are better forecasted than SKUs having high promotion frequency for a given level of sales volume.

In general, SKUs with low sales volume benefit from combined data set models. In order to declare such forecasting insights for the whole black tea category and look for new SKU prediction performance, experiments that cover the whole category are realized.

4.3.2 Experiments for Black Tea Category on Individual Data Sets

Individual training and test data sets are obtained for all of the 35 SKUs in the black tea category. An individual data set for an SKU_i includes Y_i, $Disp_i$, $PriceDisp_i$, $PriceDispAdv_i$ and $Y_i(t-1)$ attributes which are explained in Section 4.3.1.

Because N = 35 for our category, for an individual data set there are 1 normalized sales (Y), 3*35=105 promotional (Disp, PriceDisp, PriceDispAdv) and 1 normalized sales of week t-1 (Y(t-1)) attributes. While the size of an individual SKU training data set is 104×107 except for 35th SKU, the size of an individual SKU test data set is 52×107. In both of these data sets an instance indicates a week, so row numbers of these data sets denote the time period considered. Because of this reason, size of training data set of 35th SKU is 4×107.

Based on past outputs reported in Section 4.3.1, these results were obtained to be used in the forecast of the whole category:

1) For promotional data there will be no lag, for sales amounts data lag will be selected as 1.
2) C values will be taken as 1 for all support vector regression applications.

Considering these results, linear regression, regression tree, support vector regression with polynomial kernel in degree 1, support vector regression with polynomial kernel in degree 2, support vector regression with RBF kernel are used as forecasting techniques.

As explained before, because sales amounts are normalized between 0 and 1 before the forecasting, MAE is used as the measure of forecast accuracy. MAE values of the forecasting techniques for each individual SKU can be seen in Table 4.7.

According to Table 4.7, linear regression gives the minimum MAE on average for the whole category. Second best accuracy level on average for the whole category is obtained by SVR with polynomial kernel in degree 1.

Technique / SKU	Linear Regression	Regression Tree	SVR Poly deg1	SVR Poly deg2	SVR RBF
SKU1	0.0201	0.0448	0.0215	0.0228	0.0405
SKU2	0.0197	0.0252	0.0213	0.021	0.0382
SKU3	0.0769	0.0755	0.0744	0.0782	0.0926
SKU4	0.0969	0.1067	0.0938	0.0964	0.1025
SKU5	0.1863	0.1876	0.1919	0.2005	0.1799
SKU6	0.1427	0.1586	0.143	0.1427	0.1483
SKU7	0.1744	0.1655	0.1688	0.1714	0.1587
SKU8	0.1495	0.1602	0.1728	0.1978	0.1542
SKU9	0.1545	0.1546	0.1696	0.1775	0.155
SKU10	0.0754	0.0916	0.0786	0.084	0.0878
SKU11	0.0742	0.0986	0.0792	0.0781	0.0931
SKU12	0.1041	0.0988	0.1026	0.1165	0.0988
SKU13	0.1318	0.1383	0.1395	0.1462	0.1364
SKU14	0.1778	0.1777	0.1794	0.1864	0.176
SKU15	0.1408	0.157	0.1382	0.1401	0.1486
SKU16	0.1165	0.1227	0.121	0.1206	0.117
SKU17	0.1044	0.1248	0.105	0.1119	0.1104
SKU18	0.1811	0.1818	0.1992	0.2294	0.185
SKU19	0.2289	0.2198	0.2257	0.2253	0.2196
SKU20	0.1689	0.1663	0.1716	0.1693	0.1702
SKU21	0.0946	0.1019	0.0968	0.1061	0.1028
SKU22	0.1434	0.1577	0.1496	0.1518	0.1494
SKU23	0.1039	0.1043	0.1072	0.1095	0.1012
SKU24	0.0976	0.0989	0.1087	0.1084	0.1004
SKU25	0.0764	0.0929	0.0799	0.0789	0.0887
SKU26	0.1347	0.1261	0.14	0.1409	0.1181
SKU27	0.1672	0.161	0.1788	0.1893	0.1694
SKU28	0.1816	0.1884	0.1799	0.1852	0.1786
SKU29	0.144	0.1439	0.1463	0.1469	0.143
SKU30	0.1585	0.152	0.1674	0.1694	0.1508
SKU31	0.0904	0.1104	0.0928	0.0917	0.104
SKU32	0.0965	0.115	0.0935	0.0932	0.106
SKU33	0.1382	0.1369	0.1464	0.1428	0.148
SKU34	0.0611	0.0762	0.0607	0.0638	0.0728
SKU35	0.1473	0.1481	0.1541	0.1539	0.1542
AVERAGE MAE	0.1246	0.1306	0.1285	0.1328	0.1286

Table 4.7: MAE Values for Individual Data Sets with 1 New SKU Introduction

4.3.3 Experiments for Black Tea Category on Combined Data Set

The combined data set includes all normalized sales amounts and promotional data for all of the SKUs in the category. By using the combined data set, only one forecasting model would be obtained for each forecasting technique. Whereas, experiments with individual data sets require 35 individual models for each forecasting technique. The combined data set contains Y, $Disp_i$, $PriceDisp_i$, $PriceDispAdv_i$, Y(t-1) variables and SKU_i dummy variables which are explained in section 4.3.1. Because there are 35 SKUs in our category, for the combined data set there are 1 normalized sales (Y), 3*35=105 promotional (Disp, PriceDisp, PriceDispAdv), 1 normalized sales of week t-1 (Sales(t-1)) and 35 dummy variables yielding 142 variables in total.

While the size of the combined SKU training data set is 3540×142 (there are 104 training instances for each of first 34 SKUs and 4 training instances for 35^{th} SKU), the size of the combined SKU test data set is 1820×142. In both of these data sets an instance indicates a week and SKU combination, so numbers of observations in these data sets denote the SKU number multiplied by the considered time period length.

MAE values on the combined set for each forecasting technique for each SKU can be seen in Table 4.8. According to Table 4.8, SVR with polynomial kernel in degree 2 gives the best accuracy level on average for the whole category and second best method on average is linear regression.

In section 4.3.4, results of Table 4.7 and Table 4.8 are analyzed in detail.

Technique / SKU	Linear Regression	Regression Tree	SVR Poly deg1	SVR Poly deg2	SVR RBF
SKU1	0.0483	0.0339	0.0445	0.0225	0.0451
SKU2	0.0442	0.0454	0.0400	0.0220	0.0403
SKU3	0.0869	0.0903	0.0920	0.0758	0.0898
SKU4	0.1036	0.1091	0.1026	0.0932	0.1043
SKU5	0.1774	0.1853	0.1754	0.1979	0.1759
SKU6	0.1581	0.1561	0.1565	0.1457	0.1570
SKU7	0.1614	0.1722	0.1571	0.1653	0.1565
SKU8	0.1622	0.1742	0.1592	0.1769	0.1570
SKU9	0.1587	0.1665	0.1566	0.1713	0.1556
SKU10	0.0866	0.0941	0.0814	0.0769	0.0839
SKU11	0.0933	0.1211	0.0916	0.0807	0.0916
SKU12	0.1012	0.1104	0.1008	0.1056	0.0990
SKU13	0.1379	0.1478	0.1401	0.1355	0.1439
SKU14	0.1725	0.1674	0.1711	0.1755	0.1696
SKU15	0.1526	0.1487	0.1496	0.1366	0.1499
SKU16	0.1173	0.1319	0.1160	0.1255	0.1167
SKU17	0.1186	0.1213	0.1195	0.1037	0.1189
SKU18	0.1825	0.1852	0.1806	0.1991	0.1768
SKU19	0.2126	0.2220	0.2111	0.2205	0.2117
SKU20	0.1602	0.1674	0.1634	0.1732	0.1645
SKU21	0.0967	0.1185	0.0993	0.1015	0.0981
SKU22	0.1533	0.1539	0.1447	0.1519	0.1437
SKU23	0.0994	0.0993	0.0981	0.1109	0.0990
SKU24	0.0994	0.1110	0.1000	0.1088	0.0992
SKU25	0.0907	0.1319	0.0897	0.0817	0.0920
SKU26	0.1232	0.1340	0.1191	0.1393	0.1149
SKU27	0.1582	0.1584	0.1662	0.1800	0.1618
SKU28	0.1772	0.1904	0.1728	0.1857	0.1752
SKU29	0.1389	0.1480	0.1373	0.1503	0.1373
SKU30	0.1501	0.1529	0.1488	0.1706	0.1490
SKU31	0.1081	0.1137	0.1066	0.0887	0.1088
SKU32	0.1121	0.1138	0.1050	0.0878	0.1056
SKU33	0.1389	0.1439	0.1482	0.1488	0.1479
SKU34	0.0779	0.0812	0.0811	0.0627	0.0792
SKU35	0.1391	0.1542	0.1575	0.1361	0.4495
AVERAGE MAE	0.1288	0.1359	0.1289	0.1281	0.1363

Table 4.8: MAE Values for Combined Data Set with 1 New SKU Introduction

Linear regression equation for combined data with 1 new SKU is found by Weka 3.5 linear regression function as seen below:

$$Y = -0.1674 * PD1 + (-0.087) * D2 + (-0.1263) * PD2 + (-0.0563) * PDA25 + (-0.1353) * SKU1 + (-0.135) * SKU2 + 0.3779 * SKU3 + (-0.0314) * SKU4 + 0.0887 * SKU5 + (-0.0437) * SKU6 + 0.0941 * SKU8 + 0.0399 * SKU9 + (-0.0744) * SKU11 + (-0.0528) * SKU12 + 0.0647 * SKU13 + 0.1182 * SKU14 + 0.1004 * SKU15 + (-0.0532) * SKU16 + (-0.0309) * SKU17 + 0.1946 * SKU18 + 0.0737 * SKU19 + 0.0699 * SKU20 + (-0.0681) * SKU21 + 0.1459 * SKU23 + (-0.1026) * SKU24 + (-0.1094) * SKU25 + 0.0782 * SKU26 + 0.1142 * SKU27 + 0.0361 * SKU29 + 0.1154 * SKU30 + (-0.1505) * SKU31 + (-0.1327) * SKU32 + 0.1017 * SKU33 + (-0.129) * SKU34 + 0.5681 * SKU35 + 0.3764$$

According to this linear regression equation, promotional effects of only SKU 1, SKU2 and SKU 25 have impact on sales. These SKUs have much higher sales volumes than other SKUs in the category. Because of this reason, promotional effects of price reduction, feature display and advertisement of these SKUs are effective on the whole category.

Similar results are found for regression tree which can be found in Appendix IV. It is not possible to see the results for SVR as in regression equation because of kernel operations. However, Weka 3.5 SMO reg function outputs support vector numbers. SVR with polynomial kernel in degree 2 and SVR with RBF kernel have 3515 and 3524 support vectors respectively for combined data with 1 new SKU.

4.3.4 Remarks on Experiments with Data Set with 1 New SKU Entrance

In order to evaluate the performance of the forecasting methods, a goodness measure is developed. For each SKU, methods with minimum MAE value and methods that have MAE within 5% deviation of this best MAE are called good methods for the considered SKU. Table 4.9 shows which method is good and which method is not good for each SKU when individual data sets and combined data set are used in forecasting. The value “1” indicates a good method and “0” otherwise. Table 4.9 is developed on the basis of Table 4.7 and Table 4.8.

According to total of goods in Table 4.9, linear regression is the most successful model for forecast with individual data sets: its accuracy is within 5% of the best MAE for 31 of the 35 SKUs. When the combined data set is used in forecasting there is no significantly best method for the whole category. We will no perform further analyzes in order to better understand which methods perform well for which data conditions.

In the experiments with the selected 5 SKUs in Section 4.3.1, it is found that the sales volumes and promotion frequencies of SKUs are effective on the accuracy of the forecast. To find the effects of sales volumes and promotion frequencies on accuracy, 35 SKUs in the black tea category are assigned to clusters according to their average sales volumes and promotion frequencies. 10 clusters are obtained for different volume levels and promotion frequency levels. These clusters can be seen in Table 4.10.

Performance of forecasting methods needs to be analyzed for different data conditions. Accuracy levels of forecasting methods are different for SKUs with different sales volumes and promotion frequencies. In order to better analyze forecasting results, average of goodness measures are found and graphed for each of 10 SKU clusters.

Average of goodness measures for each method for both the individual data sets and the combined data set are obtained for each of these clusters and graphs of goodness measures for the individual data sets and combined data set can be found in Figure 4.9 and Figure 4.10 respectively. M1, M2, M3, M4 and M5 in Figures 4.9 and 4.10 indicates the linear regression, regression tree, support vector regression with polynomial kernel degree 1, support vector regression with polynomial kernel degree 2 and support vector regression with RBF kernel, respectively.

In Table 4.10, promotion frequency values show how many times a marketing mix effect is available for the SKU during total of the training and test periods.

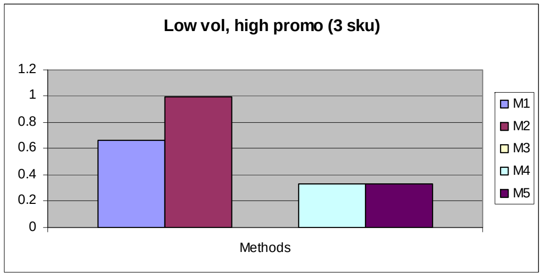
Technique / SKU	INDIVIDUAL DATA SETS					COMBINED DATA SET				
	Linear Reg	Reg Tree	SVR Poly deg1	SVR Poly deg2	SVR RBF	Linear Reg	Reg Tree	SVR Poly deg1	SVR Poly deg2	SVR RBF
SKU1	1	0	0	0	0	0	0	0	1	0
SKU2	1	0	0	0	0	0	0	0	1	0
SKU3	1	1	1	0	0	0	0	0	1	0
SKU4	1	0	1	1	0	0	0	0	1	0
SKU5	1	1	0	0	1	0	1	0	0	0
SKU6	1	0	1	1	1	0	0	0	1	0
SKU7	0	1	0	0	1	1	0	1	1	1
SKU8	1	0	0	0	1	1	1	1	0	1
SKU9	1	1	0	0	1	1	1	1	0	1
SKU10	1	0	1	0	0	0	0	0	1	0
SKU11	1	0	0	0	0	0	0	0	1	0
SKU12	0	1	1	0	1	1	0	1	0	1
SKU13	1	1	0	0	1	1	0	1	1	0
SKU14	1	1	1	0	1	1	1	1	1	1
SKU15	1	0	1	1	0	0	0	0	1	0
SKU16	1	0	1	1	1	1	0	1	0	1
SKU17	1	0	1	0	0	0	0	0	1	0
SKU18	1	1	0	0	1	1	0	1	0	1
SKU19	1	1	1	1	1	1	1	1	0	1
SKU20	1	1	1	1	1	1	0	1	0	1
SKU21	1	0	1	0	0	1	0	1	1	1
SKU22	1	0	1	0	1	0	0	1	0	1
SKU23	1	1	0	0	1	1	0	1	0	1
SKU24	1	1	0	0	1	1	0	1	0	1
SKU25	1	0	1	1	0	0	0	0	1	0
SKU26	0	0	0	0	1	0	0	1	0	1
SKU27	1	1	0	0	0	1	1	0	0	1
SKU28	1	0	1	1	1	1	0	1	0	1
SKU29	1	1	1	1	1	1	1	1	0	1
SKU30	0	1	0	0	1	1	0	1	0	1
SKU31	1	0	1	1	0	0	0	0	1	0
SKU32	1	0	1	1	0	0	0	0	1	0
SKU33	1	1	0	1	0	1	0	0	0	0
SKU34	1	0	1	0	0	0	0	0	1	0
SKU35	1	1	0	0	0	1	0	0	1	0
Total	31	17	19	12	19	19	7	18	18	18

Table 4.9: Good Methods for Individual and Combined Data Sets with 1 New SKU Introduction

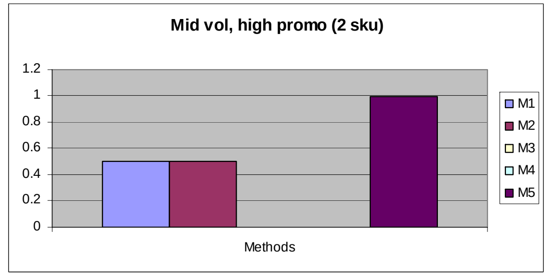
SKU NO	Average Sales Volume	Promotion frequency	Volume Level	Promotion Frequency Level
25	64.06	9	High Volume	High Promotion Frequency
31	60.97	9		
35	56.12	9		
24	52.63	9	Middle Volume	High Promotion Frequency
26	41.10	9		
33	35.90	9	Low Volume	High Promotion Frequency
30	35.76	9		
27	28.79	9		
34	67.88	7	High Volume	Middle Promotion Frequency
32	58.85	7		
16	58.04	7		
17	55.93	7		
10	53.40	7	Middle Volume	Middle Promotion Frequency
11	50.08	7		
13	45.14	7		
28	41.10	7		
6	40.93	7		
12	38.13	7	Low Volume	Middle Promotion Frequency
14	37.21	7		
4	35.38	7		
15	33.39	7		
7	32.76	7		
8	31.13	7		
5	26.91	7		
2	1035.02	5	High Volume	Low Promotion Frequency
1	989.04	5		
22	60.92	5		
21	58.26	5		
23	57.69	5		
20	42.44	5	Middle Volume	Low Promotion Frequency
3	812.86	0	High Volume	No Promotion
19	37.76	0	Low Volume	No Promotion
18	27.28	0		
29	25.53	0		
9	22.84	0		

Table 4.10: SKU clusters in the Black Tea Category

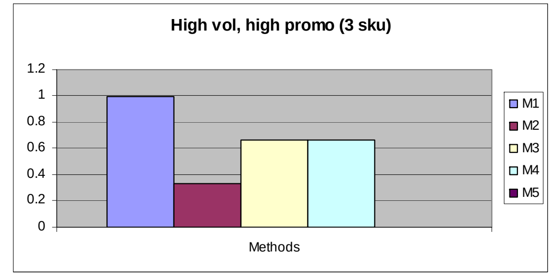
Best MAE: 0.1369 Avg MAE: 0.1498



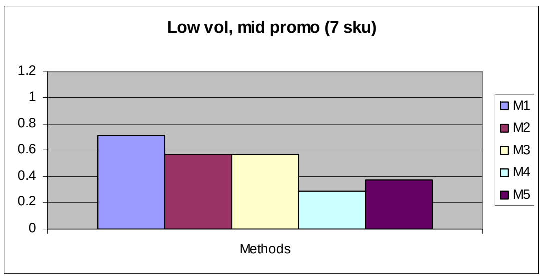
Best MAE: 0.0976 Avg MAE: 0.1038



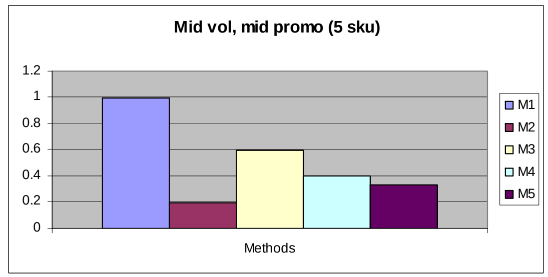
Best MAE: 0.0764 Avg MAE: 0.1009



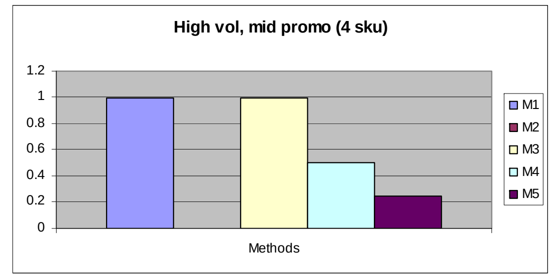
Best MAE: 0.0938 Avg MAE: 0.1450



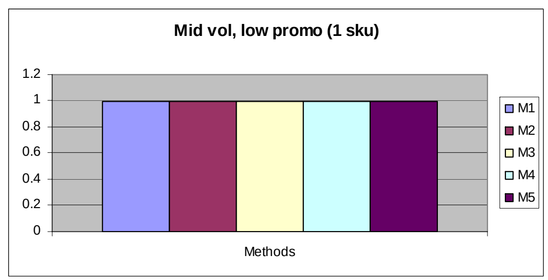
Best MAE: 0.0742 Avg MAE: 0.1383



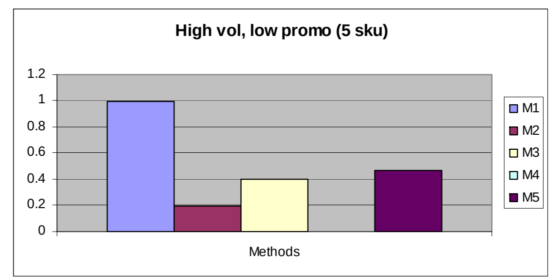
Best MAE: 0.0607 Avg MAE: 0.0990



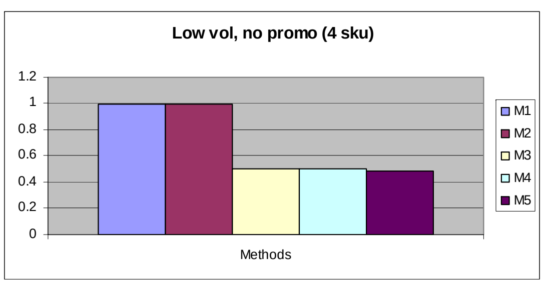
Best MAE: 0.1663 Avg MAE: 0.1693



Best MAE: 0.0201 Avg MAE: 0.0983



Best MAE: 0.1439 Avg MAE: 0.1785



Best MAE: 0.0744 Avg MAE: 0.0756

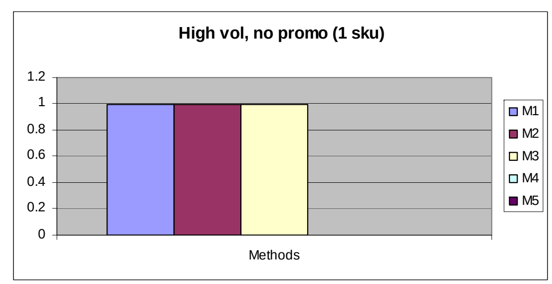
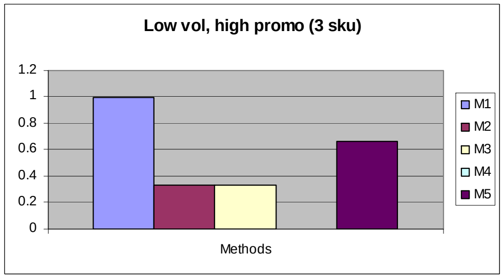
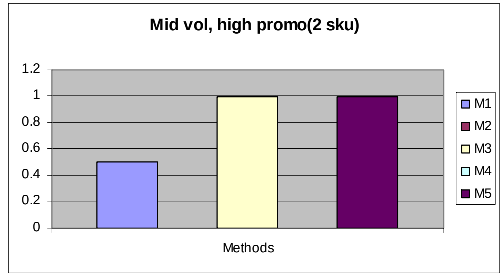


Figure 4.9: Average of Goodness Measures for Each Cluster for Individual Data Sets with 1 New SKU

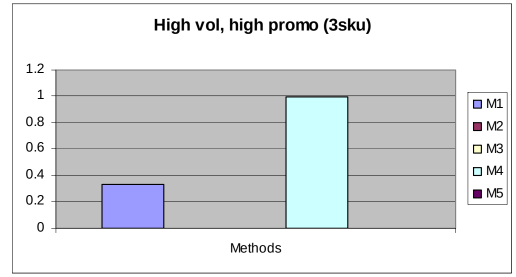
Best MAE: 0.1389 Avg MAE: 0.1533



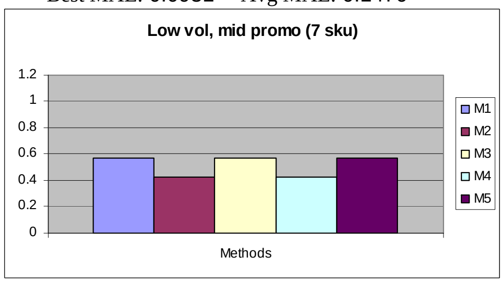
Best MAE: 0.0990 Avg MAE: 0.1062



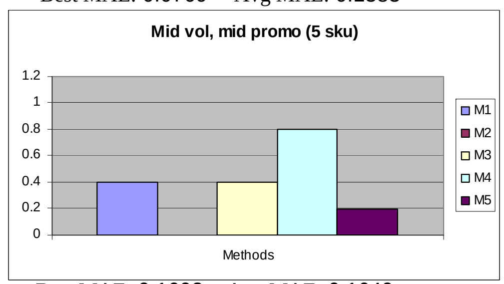
Best MAE: 0.0815 Avg MAE: 0.1042



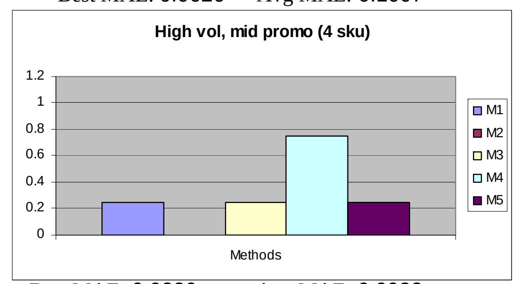
Best MAE: 0.0931 Avg MAE: 0.1476



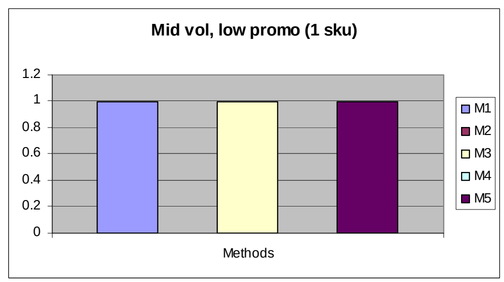
Best MAE: 0.0769 Avg MAE: 0.1383



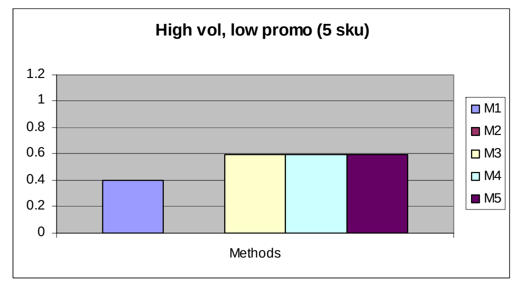
Best MAE: 0.0626 Avg MAE: 0.1007



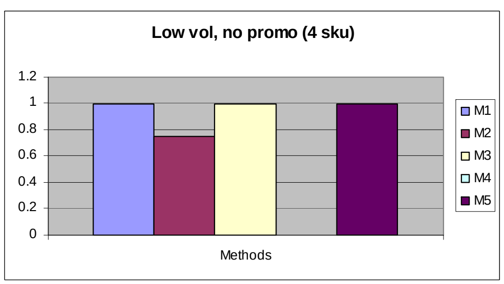
Best MAE: 0.1602 Avg MAE: 0.1640



Best MAE: 0.0220 Avg MAE: 0.0932



Best MAE: 0.1366 Avg MAE: 0.1712



Best MAE: 0.0757 Avg MAE: 0.0757

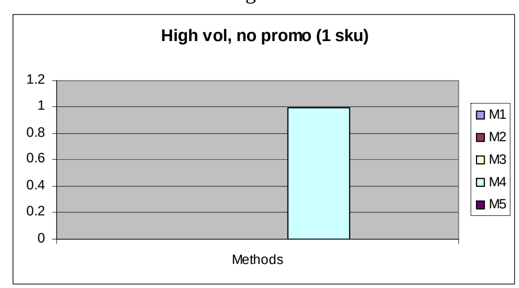


Figure 4.10: Average of Goodness Measures for Each Cluster for the Combined Data Set with 1 New SKU

Best MAE and Avg MAE, written on top of each cluster's graph, in Figures 4.9 and 4.10 indicate the minimum MAE and average of MAE values in the considered cluster respectively.

According to Figures 4.9 and 4.10, these results are found:

- Keeping promotion frequency constant, as volume increases Best MAE and Avg MAE decreases for each cluster. In the other words, SKUs with high sales volume are better predicted than the SKUs with low sales volume.
- Keeping sales volume constant, as promotion frequency decreases Best MAE and Avg MAE decreases for most of the clusters. Middle volume and middle promotion frequency, high volume and no promotion frequency, low volume and no promotion frequency clusters are outlier clusters to this pattern.

There are some outlier clusters that do not obey these results. One of these clusters is the one with middle volume and middle promotion frequency. Because this cluster includes only 1 SKU, it is acceptable to ignore this cluster while setting the above results. The other two outlier clusters contain SKUs with no promotions. SKUs with no promotion do not have a maximum sales amount as high as an SKU with promotion. Considering the normalization equation, (4.1), an SKU with no promotion would have much bigger normalized sales amounts than an SKU with promotion. Because the normalized sales amounts of an SKU with no promotion would be high, MAE for the same SKU would also be high. Therefore, the cluster of SKUs with no promotion would have higher MAE than a cluster of SKUs having promotion. This result makes acceptable to consider clusters containing SKUs with no promotions as outlier clusters.

Accuracy Difference

We define and analyze another measure that we call accuracy difference. Accuracy difference ($AccDiff(SKU_i, M_j)$) can be explained as the price that is paid for preferring the combined data against individual data for SKU_i and M_j combination. Accuracy difference graph shown in Figure 4.11 would be helpful to observe which methods are working well in which SKU clusters.

Accuracy differences are found for each SKU in the following way:

$$AccDiff(SKU_i, M_j) = MAE_i(M_j, Combined\ Data) - MAE_i(best\ in\ Individual\ Data)$$

Where,

$i = 1 \cdots 35$

$j = \{linear\ regression, regression\ tree, SVR\ with\ polynomial\ kernel\ in\ degree\ 1,$
$SVR\ with\ polynomial\ kernel\ in\ degree\ 2, SVR\ with\ RBF\ kernel\}$

M_j denotes the j^{th} forecasting method.

$AccDiff(SKU_i, M_j)$ denotes the accuracy difference for SKU_i when $Method_j$ is used.

$MAE_i(M_j, Combined\ Data)$ denotes the MAE value of M_j for SKU_i obtained by using the combined data.

$MAE_i(best\ in\ Individual\ Data)$ denotes the minimum MAE value for SKU_i obtained by using the individual data.

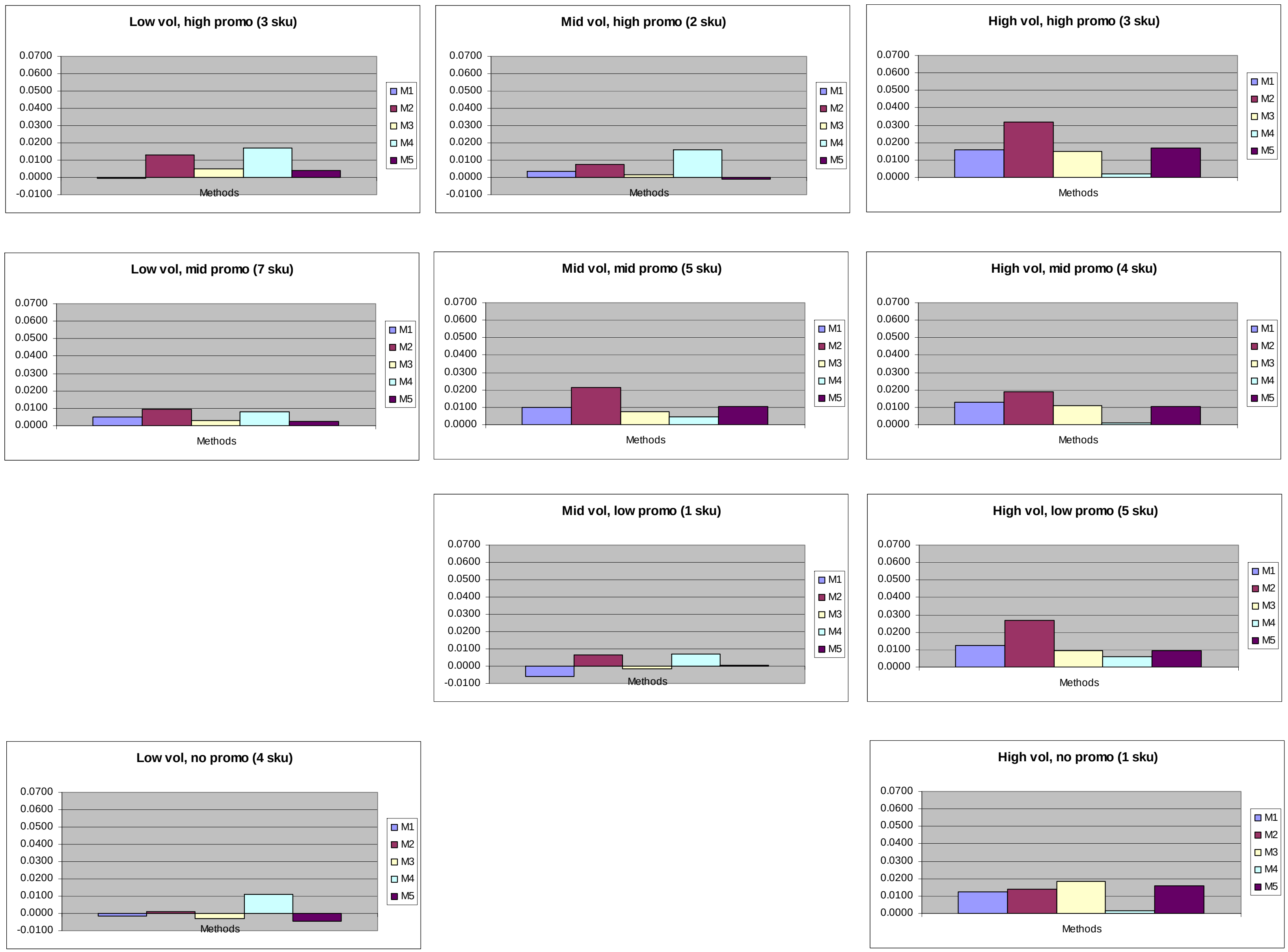


Figure 4.11: Accuracy Difference between Results of Combined and Individual Data Sets with 1 New SKU

According to Figure 4.9 showing good methods in each cluster, All SKUs having high sales volume, all SKUs with no promotion, SKUs with middle sales volume and middle promotion and SKUs with middle sales volume and low promotion are forecasted best with linear regression when individual data sets are used.

According to Figure 4.10, SVR with polynomial kernel in degree 2 is the best method for all SKUs having high sales volume, SKUs with middle sales volume and middle promotion frequency; linear regression is the best method for all SKUs with low volume and SKUs with middle sales volume and low promotion frequency; both of SVR with polynomial kernel in degree 1 and SVR with RBF kernel are the best methods for SKUs with middle sales volume and high promotion frequency, SKUs with middle sales volume and low promotion frequency and SKUs with low sales volume and no promotion.

Besides their accuracy, this study aims to compare forecasting methods according to their convenience (simplicity) and new SKU forecasting performance. When individual data sets are used in forecasting, individual forecast models are developed for each of the SKUs in the category. However, when a combined data set is used in forecasting, only one forecast model is developed. Although individual data sets have fewer attributes and instances than the combined data set has, handling 35 different data sets and obtaining 35 different forecast models is more complicated than building a single data set and obtaining a single forecast model.

As mentioned before, the only new SKU in the considered data is the 35th SKU. Considering Tables 4.7 and 4.8, when individual data sets are used, new SKU is predicted best with linear regression and when the combined data set is used, new SKU is predicted best with SVR with polynomial kernel in degree 2. According to average MAE values in these tables, linear regression is the best forecast model for the whole category when individual data sets are used and SVR with polynomial kernel in degree 2 is the best forecast model for the whole category when the combined data set is used.

Final Comments:

In order to find the best forecast method for the black tea category that has high accuracy, model simplicity and good performance on new SKU prediction, we realized many experiments as explained. To achieve model simplicity, one should forecast with the combined data set. As MAE values for both of the overall category and the new SKU show, SVR with polynomial kernel in degree 2 method gives the best accuracy with the combined data set. Simplicity, overall accuracy and good new SKU prediction performance objectives would be achieved when a combined data set is predicted with SVR with polynomial kernel in degree 2.

Because SVR with polynomial kernel in degree 2 did not perform well for all SKUs that have no promotion, using only a single forecast method would not be a good approach for the whole category. Individual data sets can be obtained for SKUs with no promotion and linear regression, which is the best method for SKUs with no promotion, would predict these SKUs with best accuracy.

4.4 Comprehensive Forecasting Experiments for Sales Data Including 4 New SKUs

In order to investigate affects of new SKU entrance better, a new data set that includes 4 New SKUs is generated. Experiments are performed for individual and combined data sets again and comments for these experiments are given in the end of section.

Sales data is generated for 35 SKUs and a time period of 156 weeks. First 104 weeks of data is considered as training data set and other 52 weeks of data is considered as test data set. In addition to these, data is generated with a 15% noise level.

The 10^{th}, 17^{th}, 33^{rd} and 35^{th} SKUs are new entering SKUs to the category. At the beginning of the time period there are 31 SKUs in the category. At the 11^{th} week SKU 10 enters the category as a new SKU, at the 71^{st} week SKU 17 enters the category as a new SKU, at the 86^{th} week SKU 33 enters the category as a new SKU and at the 101^{st} week

SKU 35 enters the category as a new SKU. In total we have a sales data of 35 SKUs for 156 weeks.

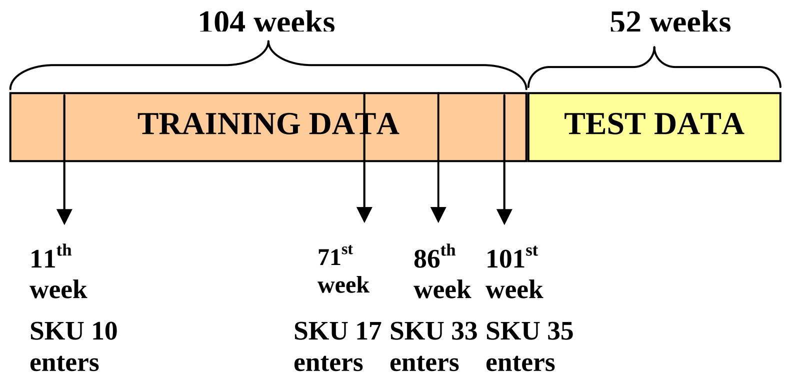


Figure 4.12: Time period of the Data Set with 4 New SKUs

4.4.1. Forecasting with Individual and Combined Data Sets

Individual training and test data sets are obtained for the sales data with 4 new SKUs in a similar way to the data set with 1 new SKU. Again linear regression, regression tree, SVR with polynomial kernel in degree 1, SVR with polynomial kernel in degree 2, SVR with RBF kernel are used as forecasting tools. Because the sales data is normalized between 0 and 1 before the forecast, MAE is used as the accuracy measure. MAE values of the forecasting techniques for each individual SKU can be seen in Table 4.11. The combined data set is obtained for the sales data with 4 new SKUs as explained for the data with 1 new SKU. MAE values resulting from the forecasting methods applied to the combined data set can be seen in Table 4.12.

SKU \ Technique	Linear Regression	Regression Tree	SVR Poly deg1	SVR Poly deg2	SVR RBF
SKU1	0.0240	0.0287	0.0262	0.0268	0.0417
SKU2	0.0231	0.0240	0.0226	0.0226	0.0389
SKU3	0.1003	0.1024	0.1098	0.1111	0.1287
SKU4	0.1217	0.1226	0.1267	0.1303	0.1222
SKU5	0.1539	0.1539	0.1790	0.2003	0.1577
SKU6	0.1108	0.1066	0.1110	0.1111	0.0998
SKU7	0.1330	0.1455	0.1425	0.1442	0.1433
SKU8	0.1361	0.1421	0.1386	0.1385	0.1345
SKU9	0.1837	0.1661	0.1991	0.2034	0.1628
SKU10	0.0776	0.0681	0.0682	0.0782	0.0730
SKU11	0.0891	0.1103	0.1049	0.0993	0.1059
SKU12	0.1584	0.1588	0.1703	0.1631	0.1681
SKU13	0.1000	0.1072	0.1079	0.1064	0.1050
SKU14	0.1224	0.1175	0.1316	0.1347	0.1203
SKU15	0.1460	0.1443	0.1534	0.1611	0.1430
SKU16	0.0701	0.0780	0.0716	0.0700	0.0742
SKU17	0.1305	0.1012	0.1220	0.1330	0.1404
SKU18	0.1555	0.1629	0.1623	0.1597	0.1520
SKU19	0.1749	0.1524	0.1862	0.1938	0.1567
SKU20	0.1288	0.1342	0.1490	0.1420	0.1364
SKU21	0.1481	0.1549	0.1556	0.1664	0.1470
SKU22	0.1370	0.1554	0.1475	0.1598	0.1445
SKU23	0.1379	0.1400	0.1596	0.1552	0.1476
SKU24	0.0840	0.0855	0.0756	0.0794	0.0729
SKU25	0.0795	0.0943	0.0839	0.0845	0.0920
SKU26	0.1965	0.2104	0.2127	0.2034	0.2010
SKU27	0.2023	0.2027	0.1981	0.2038	0.1977
SKU28	0.1179	0.1196	0.1195	0.1259	0.1142
SKU29	0.1602	0.1479	0.1569	0.1599	0.1418
SKU30	0.1137	0.1024	0.1361	0.1411	0.1136
SKU31	0.1071	0.1161	0.1103	0.1068	0.1131
SKU32	0.0924	0.1028	0.0976	0.1005	0.0998
SKU33	0.1317	0.1314	0.1384	0.1367	0.1574
SKU34	0.0882	0.1077	0.1013	0.1074	0.1107
SKU35	0.1590	0.1603	0.1633	0.1727	0.1823
AVERAGE MAE	0.1224	0.1235	0.1294	0.1324	0.1334

Table 4.11: MAE Values for Individual Data Sets with 4 New SKU Introductions

SKU \ Technique	Linear Regression	Regression Tree	SVR Poly deg1	SVR Poly deg2	SVR RBF
SKU1	0.0517	0.0521	0.0444	0.0271	0.0433
SKU2	0.0540	0.0288	0.0433	0.0240	0.0431
SKU3	0.1154	0.1184	0.1215	0.1112	0.1156
SKU4	0.1265	0.1378	0.1251	0.1298	0.1228
SKU5	0.1504	0.1571	0.1510	0.1800	0.1503
SKU6	0.1070	0.1242	0.1042	0.1117	0.1054
SKU7	0.1444	0.1703	0.1426	0.1440	0.1439
SKU8	0.1339	0.1467	0.1374	0.1372	0.1334
SKU9	0.1601	0.1731	0.1631	0.1905	0.1642
SKU10	0.0741	0.0709	0.0725	0.0739	0.0772
SKU11	0.1108	0.1290	0.1061	0.1040	0.1082
SKU12	0.1728	0.1595	0.1683	0.1665	0.1667
SKU13	0.1091	0.1073	0.1107	0.1075	0.1087
SKU14	0.1215	0.1249	0.1145	0.1313	0.1156
SKU15	0.1413	0.1461	0.1385	0.1541	0.1390
SKU16	0.0785	0.0800	0.0810	0.0692	0.0795
SKU17	0.0997	0.0986	0.0889	0.0968	0.0891
SKU18	0.1516	0.1553	0.1502	0.1610	0.1508
SKU19	0.1529	0.1526	0.1498	0.1851	0.1488
SKU20	0.1335	0.1358	0.1276	0.1503	0.1299
SKU21	0.1598	0.1679	0.1464	0.1582	0.1482
SKU22	0.1509	0.1437	0.1419	0.1565	0.1394
SKU23	0.1408	0.1348	0.1376	0.1561	0.1382
SKU24	0.0836	0.0851	0.0735	0.0760	0.0747
SKU25	0.1022	0.1021	0.1008	0.0847	0.0993
SKU26	0.2094	0.2234	0.2034	0.2125	0.2060
SKU27	0.1943	0.2032	0.1929	0.2057	0.1923
SKU28	0.1172	0.1158	0.1171	0.1232	0.1180
SKU29	0.1433	0.1437	0.1420	0.1587	0.1384
SKU30	0.1053	0.1060	0.1041	0.1366	0.1027
SKU31	0.1173	0.1303	0.1145	0.1092	0.1138
SKU32	0.1014	0.1095	0.1004	0.0977	0.1041
SKU33	0.0741	0.0799	0.0740	0.0738	0.1000
SKU34	0.1129	0.1193	0.1079	0.1018	0.1072
SKU35	0.1366	0.1344	0.1366	0.1301	0.4292
AVERAGE MAE	0.1289	0.1276	0.1225	0.1272	0.1317

Table 4.12: MAE Values for the Combined Set with 4 New SKU Introductions

4.4.2 Remarks on Experiments with Data Set with 4 New SKUs Entrance

In order to compare performance of different methods on both individual and combined data sets, a goodness measure, introduced in Section 4.3.4, is computed. Good methods for each SKU are considered as 1 and methods that are not good are considered as 0. Table 4.13 shows which method is good and which method is not good for each SKU when individual data sets and combined data set are used in forecasting.

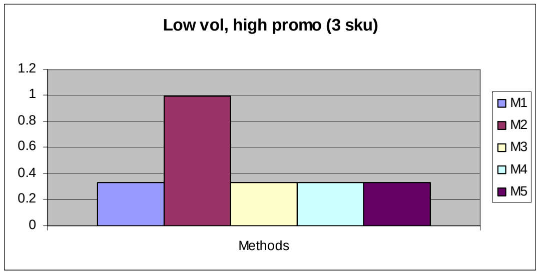
According to the counts in Table 4.13, linear regression is the most successful model for forecasting with individual data sets. When the combined data set is used in forecasting, SVR with polynomial kernel in degree 1 is the best method for the whole category.

Average goodness measure graphs for individual and combined data sets and accuracy difference graphs are obtained in this section as explained in Section 4.3.4. These graphs can be found in Figures 4.12, 4.13 and 4.14 respectively.

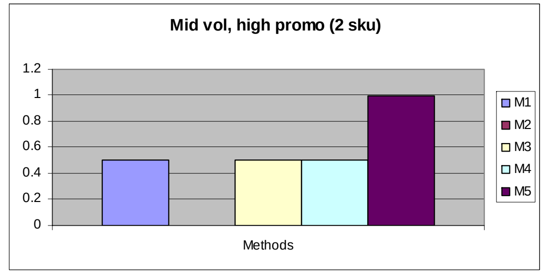
Technique / SKU	INDIVIDUAL DATA SETS					COMBINED DATA SET				
	Linear Reg	Reg Tree	SVR Poly deg1	SVR Poly deg2	SVR RBF	Linear Reg	Reg Tree	SVR Poly deg1	SVR Poly deg2	SVR RBF
SKU1	1	0	0	0	0	0	0	0	1	0
SKU2	1	0	1	1	0	0	0	0	1	0
SKU3	1	1	0	0	0	1	0	0	1	1
SKU4	1	1	1	0	1	0	0	1	1	1
SKU5	1	1	0	0	1	1	0	1	0	1
SKU6	0	0	0	0	1	1	0	1	1	1
SKU7	1	0	0	0	0	1	0	1	1	1
SKU8	1	0	1	1	1	1	0	1	1	1
SKU9	0	1	0	0	1	1	0	1	0	1
SKU10	0	0	0	0	1	1	0	1	0	1
SKU11	1	0	0	0	0	0	1	1	1	0
SKU12	1	1	0	1	0	1	1	1	1	1
SKU13	1	0	0	0	1	1	0	1	1	1
SKU14	1	1	0	0	1	0	0	1	0	1
SKU15	1	1	0	0	1	1	0	1	0	1
SKU16	1	0	1	1	0	0	0	0	1	0
SKU17	0	1	0	0	0	0	1	0	0	0
SKU18	1	0	0	0	1	1	0	1	0	1
SKU19	0	1	0	0	1	1	1	1	0	1
SKU20	1	1	0	0	0	0	0	1	0	1
SKU21	1	0	0	0	1	0	0	1	0	1
SKU22	1	0	0	0	0	0	1	1	0	1
SKU23	1	1	0	0	0	1	1	0	0	1
SKU24	0	0	1	0	1	0	0	1	1	1
SKU25	1	0	0	0	0	0	0	0	1	0
SKU26	1	0	0	1	1	1	0	1	1	1
SKU27	1	1	1	1	1	1	1	1	1	1
SKU28	1	1	1	0	1	0	1	1	0	0
SKU29	0	1	0	0	1	1	0	1	0	1
SKU30	0	1	0	0	0	1	1	1	0	1
SKU31	1	0	1	1	0	0	0	1	1	1
SKU32	1	0	0	0	0	1	0	1	1	1
SKU33	0	1	0	0	0	0	1	0	0	0
SKU34	1	0	0	0	0	0	0	1	1	1
SKU35	1	0	1	0	0	0	0	0	1	0
Total of Goods	26	16	9	7	17	18	10	26	19	26

Table 4.13: Good Methods for Individual and Combined Data Sets with 4 New SKU Introductions

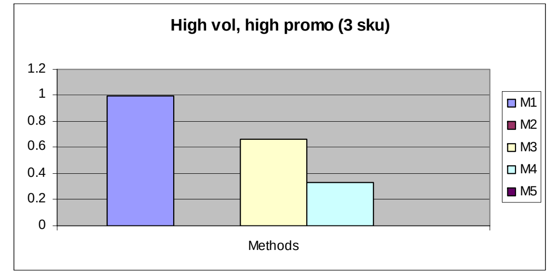
Best MAE: 0.1224 Avg MAE: 0.1845



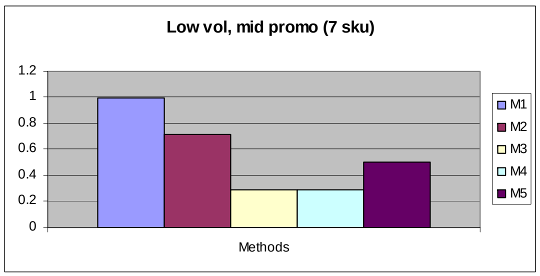
Best MAE: 0.0908 Avg MAE: 0.1499



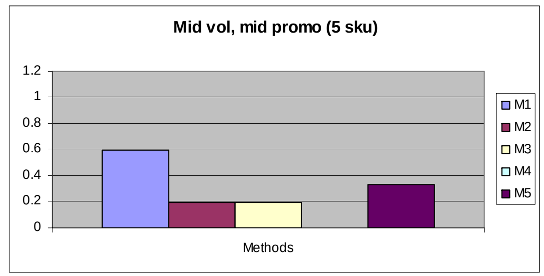
Best MAE: 0.0795 Avg MAE: 0.1009



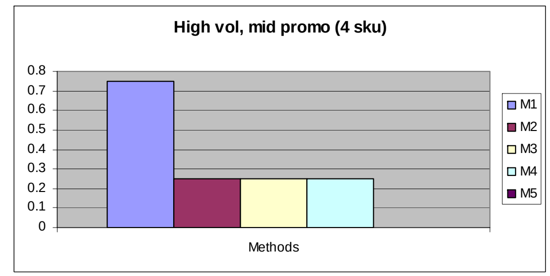
Best MAE: 0.1175 Avg MAE: 0.1387



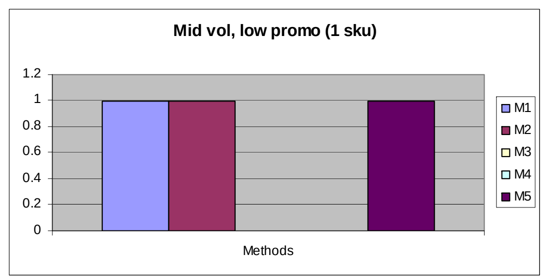
Best MAE: 0.0891 Avg MAE: 0.1081



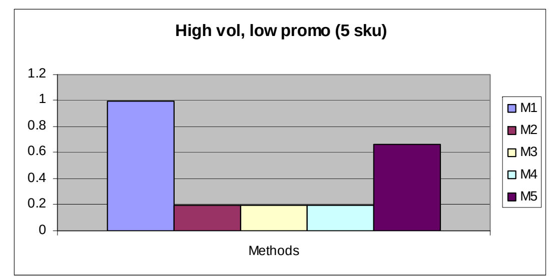
Best MAE: 0.07 Avg MAE: 0.0785



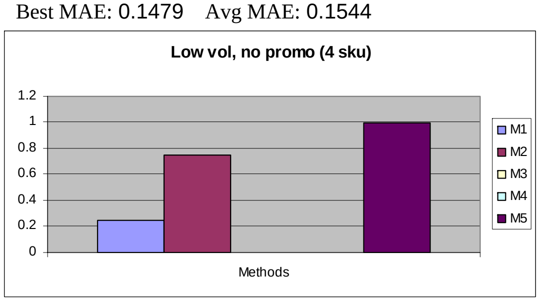
Best MAE: 0.1288 Avg MAE: 0.1315



Best MAE: 0.0226 Avg MAE: 0.0891



Best MAE: 0.1479 Avg MAE: 0.1544



Best MAE: 0.1003 Avg MAE: 0.1014

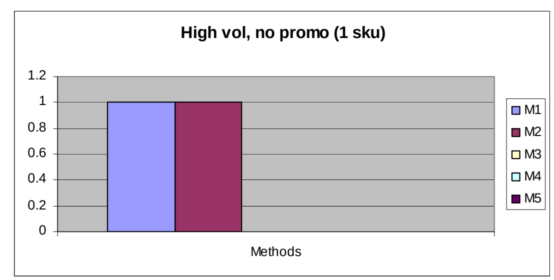
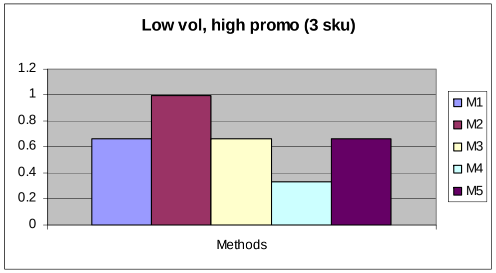
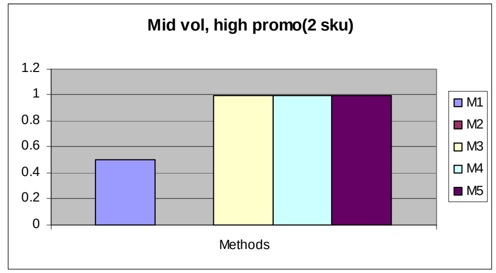


Figure 4.13: Average of Goodness Measures for Each Cluster for Individual Data Sets with 4 New SKUs

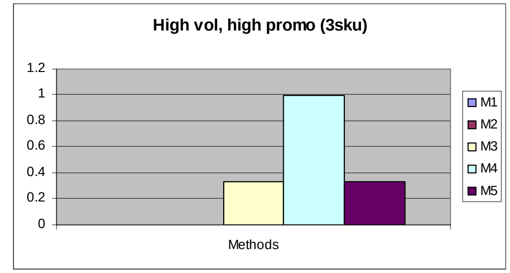
Best MAE: 0.1212 Avg MAE: 0.1572



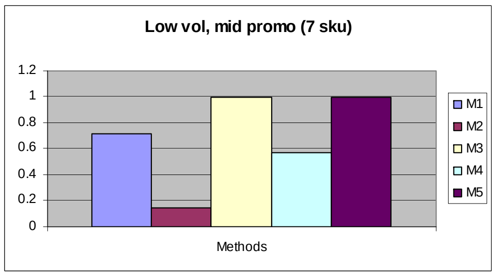
Best MAE: 0.0940 Avg MAE: 0.1533



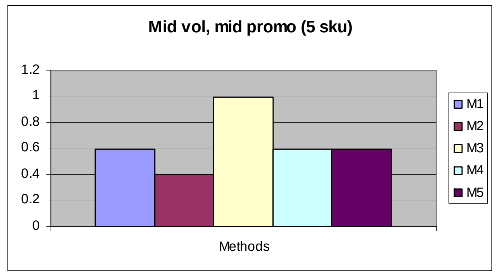
Best MAE: 0.0830 Avg MAE: 0.1245



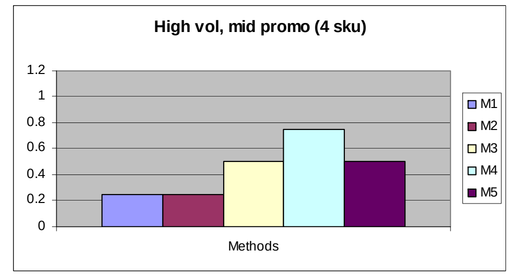
Best MAE: 0.1161 Avg MAE: 0.1414



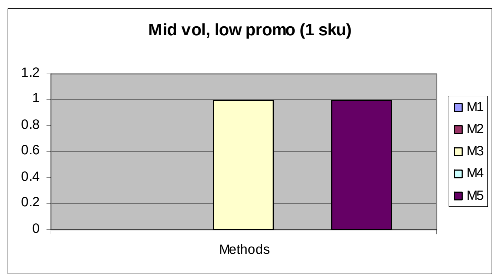
Best MAE: 0.0840 Avg MAE: 0.1089



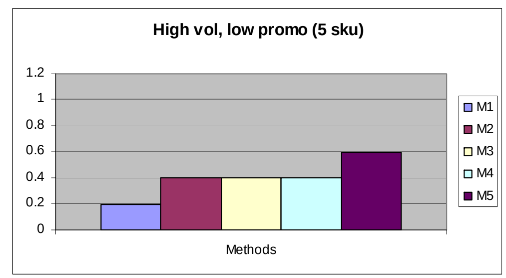
Best MAE: 0.0700 Avg MAE: 0.1176



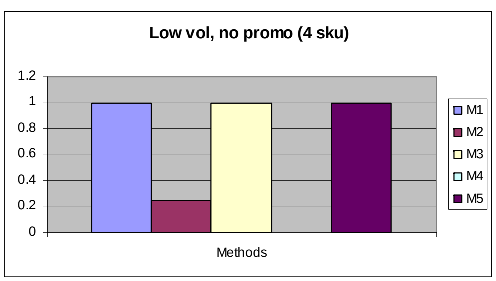
Best MAE: 0.1322 Avg MAE: 0.1351



Best MAE: 0.0242 Avg MAE: 0.1090



Best MAE: 0.1357 Avg MAE: 0.1523



Best MAE: 0.1123 Avg MAE: 0.1143

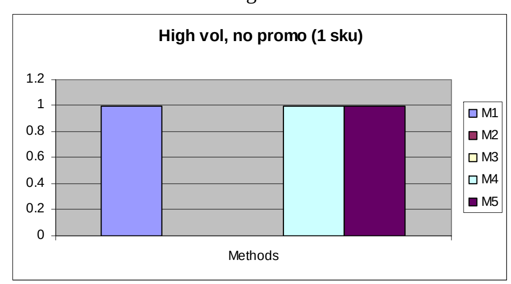


Figure 4.14: Average of Goodness Measures for Each Cluster for the Combined Data Set with 4 New SKUs

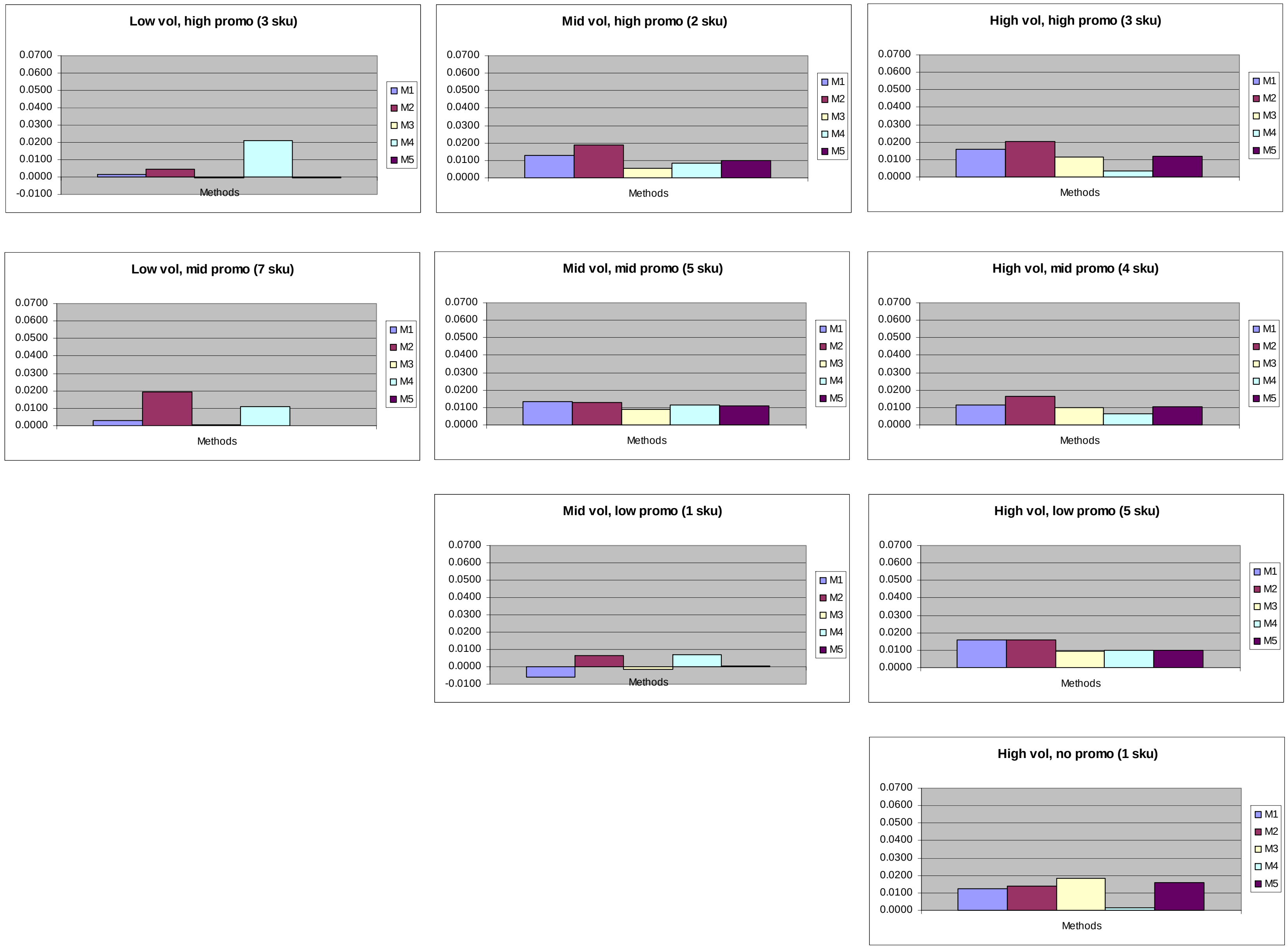


Figure 4.15: Accuracy Difference between Results of Combined and Individual Data Sets with 4 New SKUs

Considering the best MAE values written on top of each cluster in Figures 4.12 and 4.13, similar patterns to the one new SKU introduction case for average sales volume and promotion frequency combinations are found. These patterns are:

- When promotion frequency is constant, as volume increases Best MAE and Avg MAE decreases for each cluster.
- When sales volume is constant, as promotion frequency decreases Best MAE and Avg MAE decreases for most of the clusters. Middle volume and middle promotion frequency, high volume and no promotion frequency, low volume and no promotion frequency clusters are outlier clusters to this pattern. Because the outlier clusters are the same ones in both of the two data analyses, explanations for the outlier clusters are same.

According to Figure 4.12, showing good methods in each cluster, all SKUs having high sales volume, SKUs with low sales volume and middle promotion frequency and SKUs with middle sales volume and low promotion frequency are forecasted best with linear regression when individual data sets are used.

According to Figure 4.13, SVR with polynomial kernel in degree 1 is the best method for all SKUs having middle sales volume, SKUs with low sales volume and middle promotion frequency and SKUs with low sales volume and no promotion. Linear regression is the best method for all SKUs with no promotion. SVR with polynomial kernel in degree 2 is the best method for SKUs with high sales volume and high promotion frequency, SKUs with high sales volume and middle promotion frequency and SKUs with high sales volume and no promotion. SVR with RBF kernel is the best method for all SKUs with no promotion, SKUs with middle sales volume and high promotion frequency, SKUs with middle sales volume and low promotion frequency, SKUs with low sales volume and middle promotion frequency. Regression tree is the best method for SKUs with low sales volume and high promotion frequency.

Because new SKU prediction performance is one of the comparison tools in this study, MAE values of each of the new SKUs should be analyzed. When individual data sets are used SKU 10, SKU 17 and SKU 33 are predicted best by regression tree and SKU 35 is best predicted by linear regression. When the combined data set is used, SKU 10 is best predicted by regression tree, SKU 17 is best predicted by SVR with polynomial kernel in degree 1 and SKU 33 and SKU 35 are best predicted by SVR with polynomial kernel in degree 2. The whole category is best predicted by linear regression when individual data sets are used and by SVR with polynomial kernel in degree 1 when the combined data set is used.

For three new SKUs, the combined data model does provide a considerable advantage. This explains that, sales amount of an SKU is affected by sales amount of other SKUs in the category. A combined data model captures all competitive effects, whereas an individual data model can not.

4.5 Remarks on Forecasting Experiments

In this chapter, forecasting experiments are grouped under three headings according to the analyzed data sets.

First data set including no new SKUs is forecasted by multiple linear regression and exponential smoothing with lifts. According to results of these experiments, it is decided not to use exponential smoothing in further experiments.

Second data set was including 1 new SKU entrance. Forecasting experiments are performed using linear regression, regression tree and support vector regression on this data set. Combined data set is preferable because of the models simplicity objective, as explained before. According to results with combined data set for this data set in Section 4.3, the overall category is forecasted best with SVR with polynomial kernel in degree 2.

The last data set with 4 new SKU entrances is experimented with the same techniques used for data set with 1 new SKU. According to results with combined data set in Section 4.4, the category is best forecasted by SVR with polynomial kernel in degree 1.

Although combined data set is preferable because of its advantage on model simplicity, experiments on individual data sets for both second and third data sets gave better results especially with linear regression. Because of this reason, additional individual forecasting models can be developed for especially SKUs with no promotion, in order to obtain a better accuracy level in overall.

Considering Tables 4.7, 4.8, 4.11 and 4.12, as the number of new entering SKUs is increased the performance of combined data model is also increased. According to experiments in sections 4.3 and 4.4, linear regression performed well when individual data sets are used and machine learning techniques performed well when the combined data set is used. In section 4.4, 3 of new entering 4 SKUs are much better predicted by a combined forecasting model with machine learning technique. An improvement of 27% on average is observed for these 4 new SKUs when combined data sets are used. However, there is no single machine learning method which is best in all combined or individual data set conditions.

Chapter 5

CONCLUSION

A firm's primary goal is to make profit. To achieve this goal, the firm tries to increase the amount and value of purchases customers make over time. In this sense, forecasting becomes one of the most critical issues of a firm. In this study, we performed forecasting using a generated sales data for a particular category in a retail setting.

We considered the retail environment for a single category in a store with multiple customer segments, multiple SKUs, multiple SKU attributes. In addition to these, price reduction, feature display and advertisement are considered as the marketing mix instruments in the retail setting.

Generated sales data is obtained by the data generation model that is developed on the basis of the consumer choice model in the study of Fader and Hardie (1996). In this study authors developed their model that does not emphasize the products themselves but the product attributes whose unique combination indicates a single product in the category.

Forecasting is performed by both statistical techniques like multiple linear regression and exponential smoothing; and data mining techniques like regression tree and support vector regression. Our forecasting objectives were accuracy, model simplicity and new SKU prediction performance. To compare forecasting techniques regarding model simplicity, combined and individual data sets are obtained. A combined data set includes all SKUs' sales and promotional data, while an individual data set includes the considered SKU's sales data and all SKUs' promotional data. Because of this, using a combined data set requires only one forecasting model for the whole category, while using individual data sets require N (number of SKUs in the category) forecasting models for the whole category. In order to handle our new SKU prediction performance objective, three different data sets that include different number of new SKUs are investigated. The main advantage of using a data generation model is the flexibility of obtaining data sets that have different characteristics.

The first data set that includes no new SKUs is analyzed by preliminary experiments with linear regression and exponential smoothing with lifts. Lift values are estimated in order to obtain the effects of promotions on the baseline sales levels. Individual data sets for each SKU are obtained for experiments with exponential smoothing. Same data sets are experimented with multiple linear regression. It is found that multiple linear regression gives better accuracy levels than the exponential smoothing. Because of this reason, exponential smoothing is not used as an alternative forecasting technique on the other data sets.

Second data set including 1 new SKU entrance, is forecasted by linear regression, regression tree and SVR. Combined data set is preferable because of the model simplicity objective. Using the combined data set, SVR with polynomial kernel in degree 2 gave the best accuracy result for the overall. In addition to this, it is advisable to obtain individual data sets for SKUs that have no promotion to be forecasted by linear regression.

Third data set, that includes 4 new SKU entrances, is forecasted by linear regression, regression tree and SVR too. Combined data set in 4 new SKU entered situation, SVR with polynomial kernel in degree 1 gave the best result on average for the whole category. It is also advisable for this data set, to obtain individual forecasting models for SKUs with no promotion.

Another important result of analyses on both data set with 1 new SKU and data set with 4 new SKUs is the following pattern that shows a relationship between SKU sales volume, promotion frequency and forecasting accuracy.

- Keeping promotion frequency constant, as volume increases Best MAE and Avg MAE decreases for each cluster.
- Keeping sales volume constant, as promotion frequency decreases Best MAE and Avg MAE decreases for most of the clusters. Middle volume and middle promotion frequency and no promotion frequency clusters are outlier clusters to this pattern.

According to our forecasting experiments, to forecast the black tea category, a combined data set is preferable to obtain model simplicity objective. In addition to this, SVR performs well for combined data set for both new SKUs in the category and the whole category. No single method appears as best in all different situations, but for new SKU entered data sets SVR gave the best accuracy on average for the combined data sets. Because of this reason, SVR is advisable for forecasting a single category especially that includes new SKU entrances. Besides this, individual data sets for some specific SKUs can be obtained in order to increase the overall accuracy in the category.

While SKUs with high sales volume are better predicted when individual data sets are used, SKUs with low sales volume are better predicted when the combined data set is used. In 1 new SKU entered and 4 new SKU entered situations, respectively 78% and 81% of forecasting results of SKUs with high and low sales volumes obey these results according to Tables 4.7, 4.8, 4.11 and 4.12.

When Table 4.9 and Table 4.13, that show number of good methods for data sets including 1 new SKU and 4 new SKUs respectively, are compared, it is seen that SVR with polynomial kernel in degree 1 is good for 26 times for combined data set in Table 4.13 and 18 times for combined data set in Table 4.9. SVR with polynomial kernel is good for 19 times in combined set of data with 4 new SKUs and 18 times for combined data with 1 new SKU and SVR with RBF kernel is good for 26 times for combined set of data with 4 new SKUs and 18 times for data with 1 new SKU. Methods other than SVR gave better results for combined data set with 1 new SKU than combined data set with 4 new SKUs and all methods, linear regression, regression tree and SVR, performed better for individual data sets with 1 new SKU than individual data sets with 4 new SKUs according to Table 4.9 and 4.13. Considering these results, SVR performs best for combined data sets with 4 new SKUs. However, there is no single best kernel for SVR for combined data sets.

Depending on the experiments in Chapter 4, combined data set provides advantage for new entering SKUs. 4 new SKUs are 27% better predicted on average, when combined data set is used. We obtained that as number of new SKUs in the category increase, combined data set performs better than individual data set.

We performed forecasting analyzes only on three different data sets that have none, one and four new SKUs respectively. According to our forecasting results, there is no single forecasting method that is dominantly better in all conditions. Additionally, SVR is not a common forecasting methodology for retail industry. However, much more experiments with SVR and several kernels could be a future research area in order to better criticize performance of SVR on different data conditions.

In this study, while obtaining combined data sets, we assumed that all SKUs in the category belong to a singe cluster. Depending on this assumption, obtained combined data sets include data for all SKUs in the category. Although SKUs belong to the same product category, they have different characteristics like brand, price, sales volume etc. If SKUs were partitioned to clusters according to their characteristics and a combined data set was obtained for each cluster, forecasting results would be different. Clustering SKUs for obtaining combined data sets before forecasting a category’s demand would be another future research topic. Zotteri et. al. [38] examines the aggregation of SKU demand across stores and across sizes of the same brand flavor.

Our data generation simulation differs from a real grocery retail environment in some aspects. We generated the simulated data with a 15% noise level in order to obtain the randomness but this is not enough to capture all effects in a real retail situation. Also, we did not add substitution or out of stock to our data generation model, we assumed that all SKUs are available in the store in all weeks. Temporary price reduction amount is not considered although TPR as a dummy variable is considered. Advertisement is considered as a marketing mix instrument itself, but we did not look for different channels in advertisement like TV, radio or magazine. Because of these reasons, results of analysis on simulated data may be different from analysis on real data. Gür Ali et al. [17] show that for promotional real grocery retail data, data mining techniques perform better up to 65% than exponential smoothing for promotional weeks. However, in our study we did not measured forecasting performance of promotional and non promotional weeks separately and we did not obtain such a distinct performance improvement. Improving the data generation simulator to include the factors that differentiate our data from real situation could be a future research.

Customer loyalty is very strong in black tea category and in a short time period customer preferences are not affected much by promotional activities. As seen in Gür Ali et al. [17], simple methods perform well for stationary categories that are not affected much by promotional activities, although data mining methods increase forecasting performance of dynamic categories that are easily affected by promotions. In our study, we found that linear regression performs well for many data conditions. If forecasting experiments are applied on a different category in which customer attribute preferences are changeable by promotional activities, linear regression may not be enough.

APPENDIX I

Support Vector Regression

The basic idea of Support Vector Regression is described here, a more detailed description can be found in [1], [16], [32], [33] and [34].Given a training set of data $\{(x_1, y_1), \ldots, (x_l, y_l)\}$, where each $x_i \subset \Re$ denotes the input space of the sample and has a corresponding target value $y_i \subset \Re$ for $i = 1, \cdots, l$ where l corresponds to the size of the training data. In ε-insensitive SVR [32], the goal is to find a function $f(x)$ that has at most ε deviation from the actually obtained targets y_i for all the training data, and at the same time is as flat as possible.

In the case of linear functions, f takes the form

$$f(x) = \langle w \cdot x \rangle + b \text{ with } x \in X, b \in \Re \tag{A1.1}$$

Flatness in the case of (1) means that a small w is searched out. This problem can be written as a convex optimization problem:

$$\begin{aligned} & \textit{minimize} \quad \frac{1}{2}\|w\|^2 \\ & s.t. \quad \begin{cases} y_i - \langle w \cdot x_i \rangle - b \le \varepsilon \\ \langle w \cdot x_i \rangle + b - y_i \le \varepsilon \end{cases} \end{aligned} \tag{A1.2}$$

The “soft margin” loss function was used in support vector machines by [2], introducing slack variables $\xi_i, \hat{\xi}_i$. The formulation stated in [32]

$$\begin{aligned} & \textit{minimize} \quad \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l}(\xi_i + \hat{\xi}_i) \\ & s.t. \quad \begin{cases} (\langle w \cdot x_i \rangle + b) - y_i \le \varepsilon + \xi_i \\ y_i - (\langle w \cdot x_i \rangle + b) \le \varepsilon + \hat{\xi}_i \\ \xi_i, \hat{\xi}_i \ge 0, \ i = 1, 2, \ldots, l. \end{cases} \end{aligned} \tag{A1.3}$$

where, C is a constant which can be seen to control the size of $\|w\|$ for a fixed training set.

The corresponding dual problem can be derived as:

$$
\begin{aligned}
&maximize \quad \sum_{i=1}^{l}(\hat{\alpha}_i-\alpha_i)y_i-\varepsilon\sum_{i=1}^{l}(\hat{\alpha}_i+\alpha_i)y_i-\frac{1}{2}\sum_{i,j=1}^{l}(\hat{\alpha}_i-\alpha_i)(\hat{\alpha}_j-\alpha_j)\langle x_i\cdot x_j\rangle \\
&s.t. \quad 0\le\alpha_i,\hat{\alpha}_i\le C,\ i=1,\dots,l, \\
&\qquad \sum_{i=1}^{l}(\hat{\alpha}_i-\alpha_i)=0,\ i=1,\dots,l.
\end{aligned}
\tag{A1.4}
$$

Only the nonzero values of Lagrange multipliers α_i and $\hat{\alpha}_i$ are useful in forecasting the regression line and are known as support vectors. For all points inside the ε tube, the Lagrange multipliers equal to zero do not contribute to the regression function. Only if the requirement $|f(x)-y|\ge\varepsilon$ is fulfilled, lagrange multipliers may be nonzero values and used as support vectors.

The corresponding Karush-Kuhn-Tucker conditions are

$$
\begin{aligned}
&\alpha_i(\langle w\cdot x_i\rangle+b-y_i-\varepsilon-\xi_i)=0, \quad i=1,\dots,l, \\
&\hat{\alpha}_i(\langle w\cdot x_i\rangle+b-y_i-\varepsilon-\hat{\xi}_i)=0, \quad i=1,\dots,l, \\
&\xi_i\hat{\xi}_i=0,\ \alpha_i\hat{\alpha}_i=0, \quad i=1,\dots,l, \\
&(\alpha_i-C)\xi_i=0,\ (\hat{\alpha}_i-C)\hat{\xi}_i=0 \quad i=1,\dots,l.
\end{aligned}
\tag{A1.5}
$$

The support vector algorithm can also be nonlinear. This, could be achieved by simply preprocessing the training patterns x_i by a map $\phi: X\to F$ into some feature space F and then applying the standard SVR algorithm. An example given in [32] considers the map $\phi:\Re^2\to\Re^3$ with $\phi(x_1,x_2)=(x_1^2,\sqrt{2}x_1x_2,x_2^2)$. It is understood that the subscripts in this case refer to the components of $x\in\Re^2$. Training a linear support vector machine on the preprocessed features would yield a quadratic function. This approach becomes computationally infeasible for both polynomial features of higher order and higher dimensionality [1].

Implicit mapping via kernels is a feasible and computationally cheaper way. By using the trick of kernel function, one lets the kernel function be the inner product of mapping function, $K(x_i, x_j) = \langle \phi(x_i) \cdot \phi(x_j) \rangle$. Therefore, one only needs to specify a kernel function without considering the mapping function or the feature space explicitly.

The name kernel is derived from integral operator theory, which supports much of the theory of the relation between kernels and their corresponding feature spaces. An important consequence of the dual representation is that the dimension of the feature space need not affect the computation. As one does not represent the feature vectors explicitly, the number of operations required to compute the inner product by evaluating the kernel function is not necessarily proportional to the number of features. The use of kernel makes it possible to map the data implicitly into a feature space and to train in such a space. The only information used about the training examples is the kernel matrix in the feature space [16].

Three common kernel functions include:

Linear function: $K(x_i, x_j) = \langle x_i \cdot x_j \rangle$

Polynomial function (with parameter d): $K(x_i, x_j) = \left(\langle x_i \cdot x_j \rangle + 1\right)^d$

Radial Basis Function (RBF): $K(x_i, x_j) = \exp\left(-\gamma \left\| x_i - x_j \right\|^2\right)$

In our experiments, we used polynomial with d=1 and d=2 and RBF kernels.

APPENDIX II

Questions for Data Generation Simulation

We want to select a product category of nonperishable products. Multiple customer segments and marketing mix instruments like price reduction, feature display and advertisement should be available for the category. We want to look for different demand and response dynamics in the category.

About the Category

➢ How many SKUs are there ?

➢ What kind of SKU attributes does this category includes? (e.g. brand, price, packet type etc). How many attribute levels can be found in each attribute?

No	Attribute Name	Attribute Level1	Attribute Level2	Attribute Level3	Attribute Level4	Attribute Level5	....
1							
2							
3							
4							
....							

➢ What is the average SKU life cycle for this category? Do new SKUs enter to category frequently?

➢ How many customer segments are there in the category? What kind of differences do these categories have? Do these categories give different responses to promotional activities? What is the arrival to shop frequency for these categories?

Customer Segment No	Arrival Frequency	Response to promotional activity (for instance, best response to advertisement, worst response to feature display)	Distinct response to product categories? (Purchases same brand always or purchases same tea aroma always etc.)
1			
2			
3			
4			
....			

➢ What are the sizes of customer segments relative to each other?

➢ Is there a market leader for this category? Which SKUs or brands are leader in the category?

Base Preferences of Customers ($a_r^{h,0}$ Base Preferences)

➢ What would be the preferences for product attributes when there is no price reduction, feature display and advertisement marketing mix instrument. For example, can you give expressions like 70% of segment A prefer big sized packets or 90% of segment B prefer SKUs with high price etc. Can you order product attributes according to the preferences of customers?

	Base preferences for Attribute1		Base preferences for Attribute2				Base preferences for Attribute3		
	L1	L2	L1	L2	L3	L4	L1	L2	L3
Segment1									
Segment2									
Segment3									

Marketing Mix Instruments

➢ What is the frequency of price reduction for this category? Is price reduction applied on brand level or SKU level? What is the typical price reduction percentage? Are there great differences between campaigns?

➢ What is the frequency of feature display for this category? Is feature display applied on brand level or SKU level?

➢ What is the frequency of advertisement for this category? Is advertisement applied on brand level or SKU level?

➢ Is there a distinct time pattern for these marketing mix instruments (price reduction, feature display and advertisement)? Can more than one marketing mix instruments be applied at the same time?

➢ If a company applies a marketing mix instrument, do competitors apply marketing mix instruments too in this category?

Marketing Mix Instruments ($\beta^h_{price}, \beta^h_{disp}, \beta^h_{adv}$ Marketing Mix Scalar Multipliers)

Do all customer segments response in the same way when a marketing mix instrument like price reduction, feature display or advertisement is applied? Does sales amount increase according depend on the marketing mix instrument? For instance, sales amounts of customer segment A increase at most when price reduction is applied, lowest increase is seen when feature display is applied.

	Price Reduction	Feature Display	Advertisement
Segment1			
Segment2			
Segment3			

Effect of Product Testing on Customer Loyalty (λ_r Lambda)

➢ Do base preferences of a customer change, when the customer purchases a product that he/she did not prefer before? We think, for instance, a customer who buys product with small package size may change his/her preferences after testing a product with big size. For which attributes do you think this kind of a preference change would be available?

	Attribute1(e.g. Brand)	Attribute2(e.g. Price)	Attribute3(e.g. Aroma)	Attribute4	etc.
All customers					

Customer Loyalties for Product Attibutes (α_r^h Attribute Loyalty Scalar Multipliers)

➢ Can you explain customer loyalties for product attributes like brand, packet type etc.? Considering sales amounts of each customer segment, can you make explanations like below?

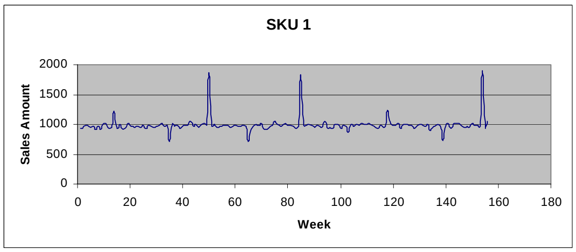
Customers in segment A purchase mostly the same brand but they purchase products with various package and aroma types.

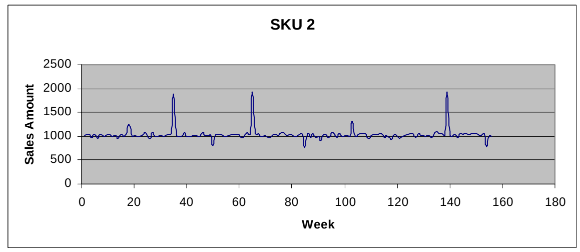
Customers in segment B, do not prefer a specific brand but mostly purchase middle sized packages.

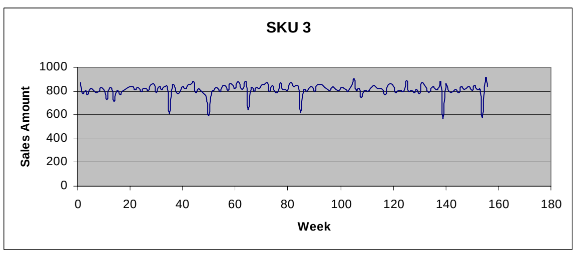
	Attribute 1	Attribute 2	Attribute 3	Attribute 4	etc
Segment1					
Segment2					
Segment3					
etc					

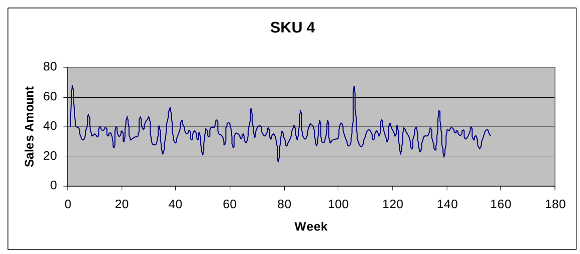
APPENDIX III

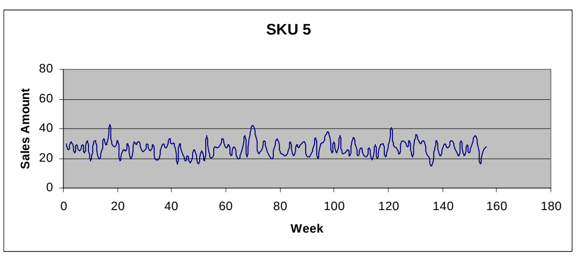
SKU Sales Volumes and Promotion Pattern for 1 New SKU Entered Situation

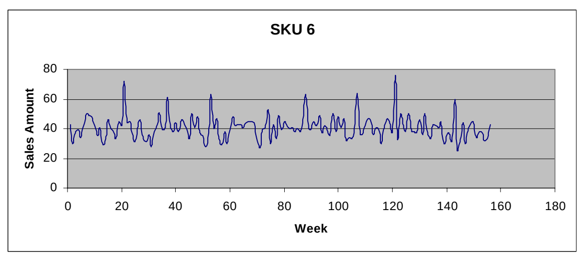


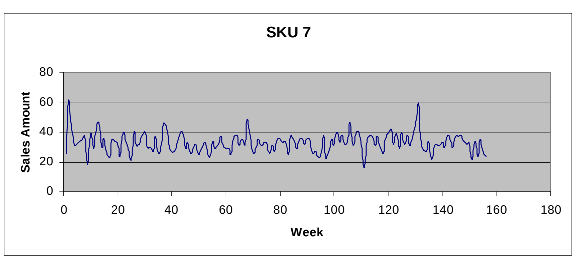


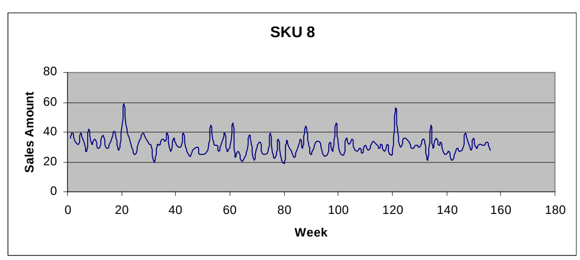


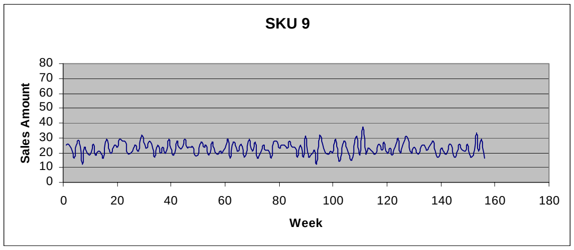


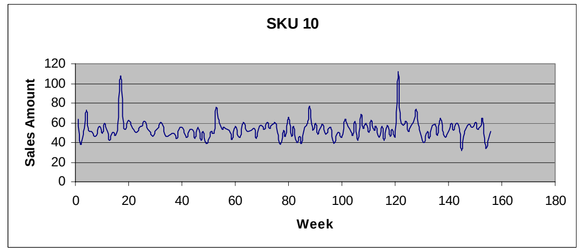


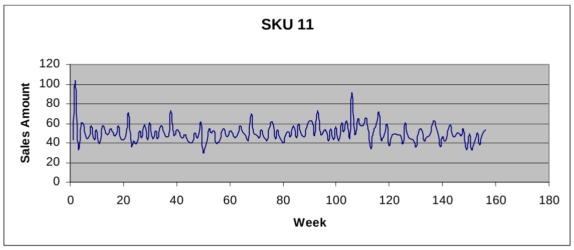


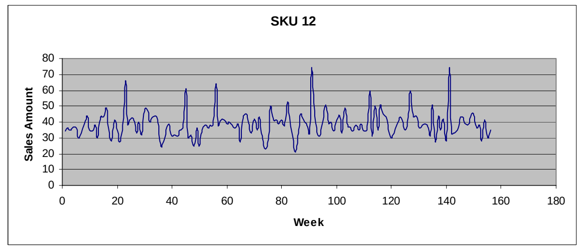


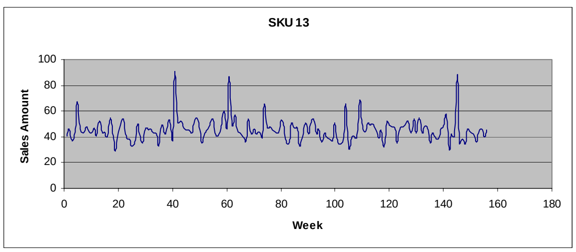


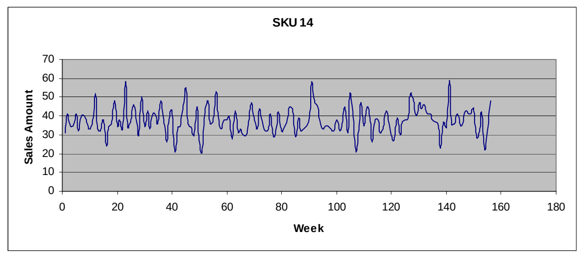

SKU 9
Sales Amount
Week

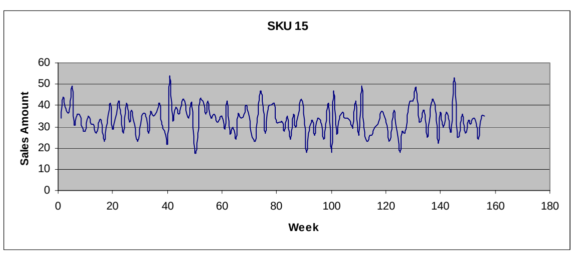

SKU 10
Sales Amount
Week

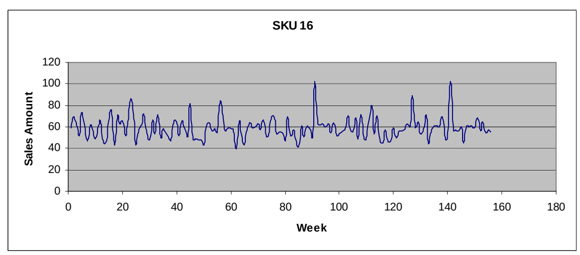

SKU 11
Sales Amount
Week

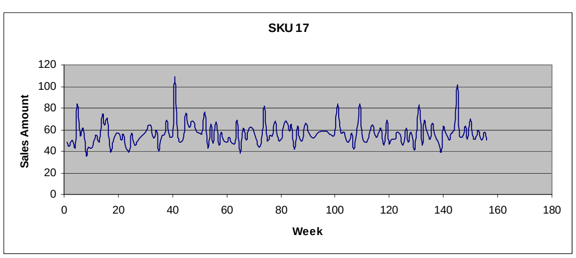

SKU 12
Sales Amount
Week

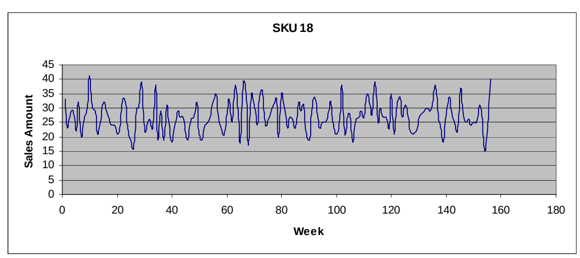

SKU 13
Sales Amount
Week

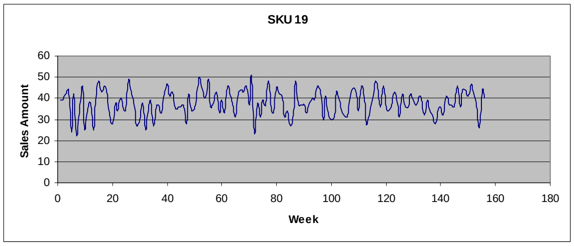

SKU 14
Sales Amount
Week

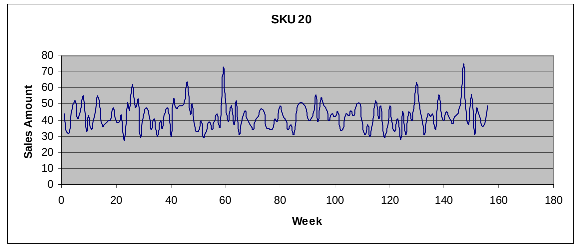

SKU 15
Sales Amount
Week

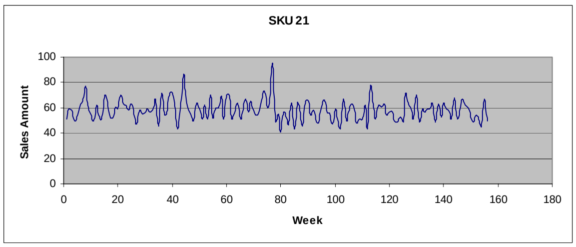

SKU 16
Sales Amount
Week

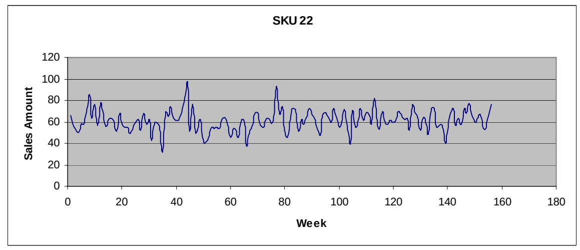

SKU 17
Sales Amount
Week

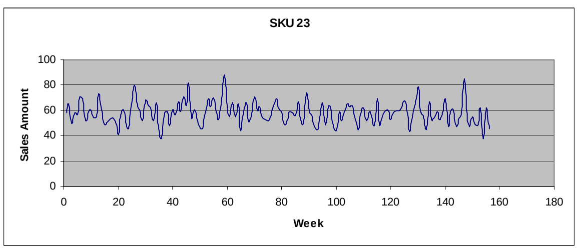

SKU 18
Sales Amount
Week

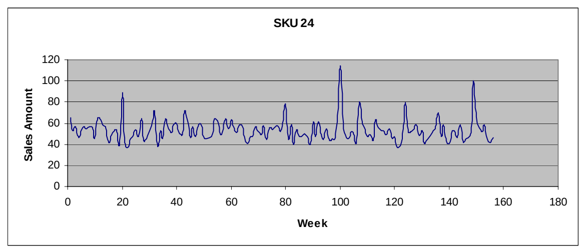

SKU 19
Sales Amount
Week

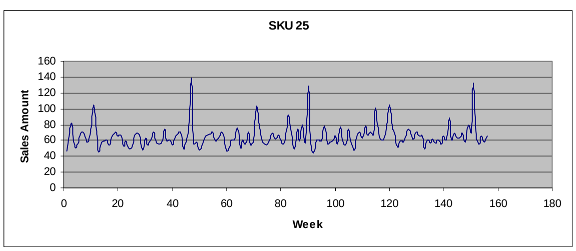

SKU 20
Sales Amount
Week

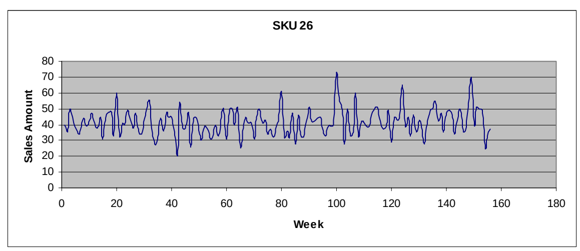

SKU 21
Sales Amount
Week

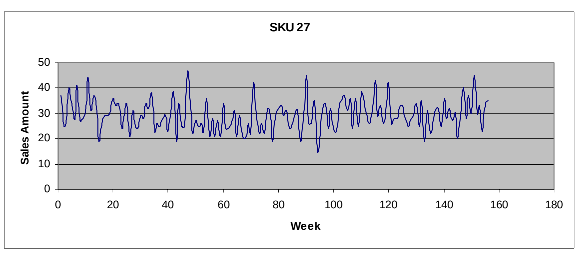

SKU 22
Sales Amount
Week

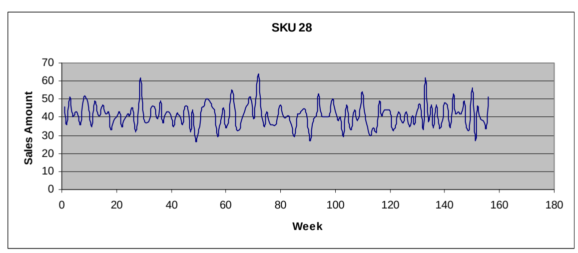

SKU 23
Sales Amount
Week

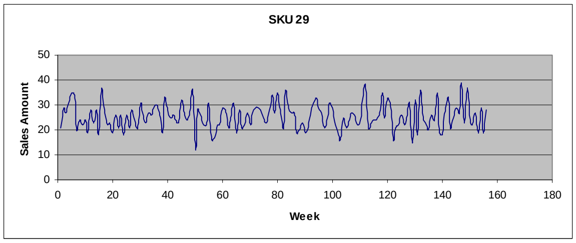

SKU 24
Sales Amount
Week

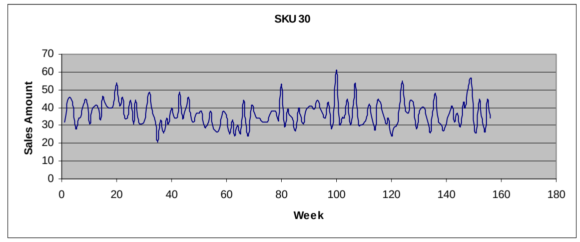

SKU 25
Sales Amount
Week

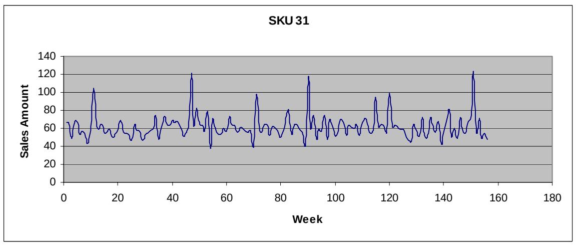

SKU 26
Sales Amount
Week

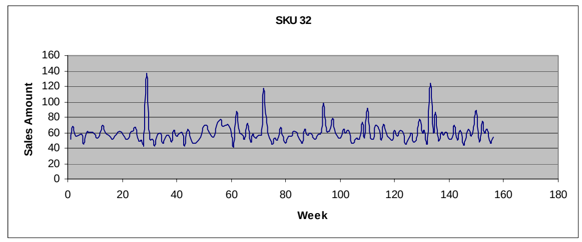

SKU 27
Sales Amount
Week

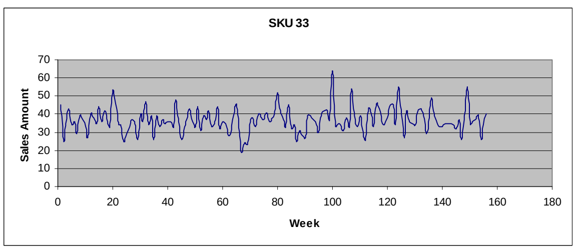

SKU 28
Sales Amount
Week

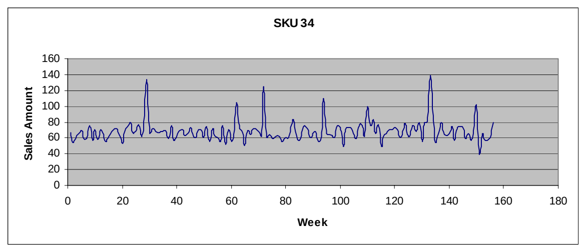

SKU 29
Sales Amount
Week

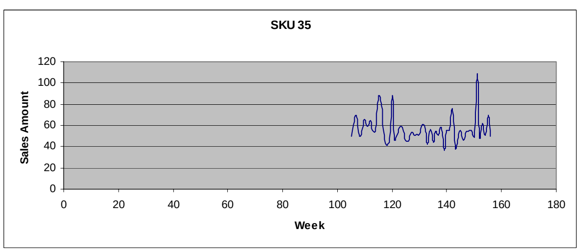

SKU 30
Sales Amount
Week


SKU 31
Sales Amount
Week


SKU 32
Sales Amount
Week


SKU 33
Sales Amount
Week


SKU 34
Sales Amount
Week


SKU 35
Sales Amount
Week

	Weeks of Promotional Activities		
	Display	**Price and Display**	**Price and Display and Advertisement**
SKU 1	14, 118	50, 85, 154	-
SKU 2	18, 103	35, 65, 139	-
SKU 3	-	-	-
SKU 4	-	38, 68, 86, 116, 137	2, 106
SKU 5	-	53, 98, 130, 152	17, 70 , 121
SKU 6	-	37, 53, 88, 107, 143	21, 121
SKU 7	-	13, 38, 68, 106	2, 131
SKU 8	-	53, 61, 88, 99, 134	21, 121
SKU 9	-	-	-
SKU 10	-	4, 53, 88, 128	17, 121
SKU 11	-	22, 38, 68, 116	2, 106
SKU 12	-	23, 45, 56, 112, 127	91, 141
SKU 13	-	5, 74, 104, 109	45, 61, 145
SKU 14	-	12, 23, 45, 56, 127	91, 141
SKU 15	-	5, 74, 101, 111, 131	41, 145
SKU 16	-	23, 45, 56, 112, 127	91, 141
SKU 17	-	5, 74, 101, 109, 131	41, 145
SKU 18	-	-	-
SKU 19	-	-	-
SKU 20	26, 46, 130	59, 147	-
SKU 21	8, 113	44, 77, 149	-
SKU 22	8, 113	44, 77, 149	-
SKU 23	26, 46, 130	59, 147	-
SKU 24	32, 43, 136	20, 80, 107, 124	100, 149
SKU 25	83, 142	11, 71, 115, 120	11, 71, 115
SKU 26	32, 43, 136	20, 80, 107, 124	100, 149
SKU 27	83, 142	11, 71, 115, 120	11, 71, 115
SKU 28	-	62, 94, 110, 150	29, 72, 133
SKU 29	-	-	-
SKU 30	32, 43, 136	20, 80, 107, 124	100, 149
SKU 31	83, 142	11, 71, 115, 120	11, 71, 115
SKU 32	-	62, 94, 110, 150	29, 72, 133
SKU 33	32, 43, 136	20, 80, 107, 124	100, 149
SKU 34	-	62, 94, 110, 150	29, 72, 133
SKU 35	83, 142	11, 71, 115, 120	11, 71, 115

This table shows the time pattern of marketing mix instruments for 35 SKUs for 1 new SKU entered situation. Numbers in the table indicate the week in which the marketing mix instrument is applied.

APPENDIX IV

Regression Tree for Combined Data Set with 1 New SKU

```
SKU3 < 0.5
|   Sales(t-1) < 0.29
|   |   SKU15 < 0.5
|   |   |   SKU2 < 0.5
|   |   |   |   SKU32 < 0.5
|   |   |   |   |   SKU31 < 0.5
|   |   |   |   |   |   SKU1 < 0.5
|   |   |   |   |   |   |   SKU34 < 0.5
|   |   |   |   |   |   |   |   SKU24 < 0.5
|   |   |   |   |   |   |   |   |   PD1 < 0.5 : 0.37 (520/0.03) [276/0.04]
|   |   |   |   |   |   |   |   |   PD1 >= 0.5 : 0.13 (8/0.01) [6/0.01]
|   |   |   |   |   |   |   |   SKU24 >= 0.5
|   |   |   |   |   |   |   |   |   Sales(t-1) < 0.17
|   |   |   |   |   |   |   |   |   |   Sales(t-1) < 0.08 : 0.32 (2/0) [1/0]
|   |   |   |   |   |   |   |   |   |   Sales(t-1) >= 0.08
|   |   |   |   |   |   |   |   |   |   |   Sales(t-1) < 0.16 : 0.23 (10/0) [6/0.01]
|   |   |   |   |   |   |   |   |   |   |   Sales(t-1) >= 0.16 : 0.13 (3/0) [0/0]
|   |   |   |   |   |   |   |   |   Sales(t-1) >= 0.17 : 0.29 (29/0.01) [12/0.05]
|   |   |   |   |   |   |   SKU34 >= 0.5
|   |   |   |   |   |   |   |   Sales(t-1) < 0.28 : 0.25 (39/0.01) [30/0.03]
|   |   |   |   |   |   |   |   Sales(t-1) >= 0.28 : 0.15 (4/0.01) [2/0.01]
|   |   |   |   |   |   SKU1 >= 0.5
|   |   |   |   |   |   |   PD1 < 0.5 : 0.22 (75/0) [24/0]
|   |   |   |   |   |   |   PD1 >= 0.5 : 0.97 (2/0) [0/0]
|   |   |   |   |   SKU31 >= 0.5 : 0.23 (55/0.02) [30/0.01]
|   |   |   |   SKU32 >= 0.5 : 0.22 (52/0.01) [25/0.02]
|   |   |   SKU2 >= 0.5 : 0.24 (70/0.01) [30/0.02]
|   |   SKU15 >= 0.5 : 0.5 (15/0.04) [3/0.03]
|   Sales(t-1) >= 0.29
|   |   SKU18 < 0.5
|   |   |   Sales(t-1) < 0.46
|   |   |   |   SKU25 < 0.5
|   |   |   |   |   SKU27 < 0.5
|   |   |   |   |   |   SKU31 < 0.5
|   |   |   |   |   |   |   SKU24 < 0.5
```

```
| | | | | | | | SKU34 < 0.5
| | | | | | | | | D2 < 0.5
| | | | | | | | | | SKU16 < 0.5
| | | | | | | | | | | SKU21 < 0.5
| | | | | | | | | | | | SKU11 < 0.5
| | | | | | | | | | | | | SKU20 < 0.5
| | | | | | | | | | | | | | PDA12 < 0.5
| | | | | | | | | | | | | | | SKU12 < 0.5
| | | | | | | | | | | | | | | | Sales(t-1) < 0.46
| | | | | | | | | | | | | | | | | Sales(t-1) < 0.3 : 0.46 (16/0.07) [6/0.06]
| | | | | | | | | | | | | | | | | Sales(t-1) >= 0.3
| | | | | | | | | | | | | | | | | | SKU22 < 0.5
| | | | | | | | | | | | | | | | | | | PDA13 < 0.5
| | | | | | | | | | | | | | | | | | | | SKU26 < 0.5
| | | | | | | | | | | | | | | | | | | | | PD1 < 0.5
| | | | | | | | | | | | | | | | | | | | | | SKU32 < 0.5
| | | | | | | | | | | | | | | | | | | | | | | SKU19 < 0.5
| | | | | | | | | | | | | | | | | | | | | | | | PD4 < 0.5
| | | | | | | | | | | | | | | | | | | | | | | | | SKU23 < 0.5
| | | | | | | | | | | | | | | | | | | | | | | | | | PD28 < 0.5
| | | | | | | | | | | | | | | | | | | | | | | | | | | Sales(t-1) < 0.45
| | | | | | | | | | | | | | | | | | | | | | | | | | | | PDA5 < 0.5
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | PD2 < 0.5
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | D20 < 0.5
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | PDA24 < 0.5
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | PD5 < 0.5
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | SKU6 < 0.5
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | SKU17 < 0.5
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | SKU7 < 0.5
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | PD13 < 0.5
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | PDA4 < 0.5
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | SKU4 < 0.5
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | SKU10 < 0.5 : 0.44 (167/0.03) [106/0.04]
```

| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | SKU10 >= 0.5 : 0.36 (33/0.01) [13/0.01]
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | SKU4 >= 0.5
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | Sales(t-1) < 0.4 : 0.35 (18/0.02) [10/0.02]
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | Sales(t-1) >= 0.4
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | Sales(t-1) < 0.42 : 0.24 (2/0) [2/0.04]
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | Sales(t-1) >= 0.42 : 0.33 (5/0.03) [1/0]
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | PDA4 >= 0.5 : 0.37 (5/0.03) [1/0.15]
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | PD13 >= 0.5 : 0.56 (5/0.01) [2/0.01]
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | SKU7 >= 0.5 : 0.36 (17/0.03) [9/0.03]
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | SKU17 >= 0.5 : 0.33 (21/0.02) [10/0.02]
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | SKU6 >= 0.5 : 0.28 (17/0.02) [9/0.02]
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | PD5 >= 0.5 : 0.53 (5/0.03) [5/0.03]
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | PDA24 >= 0.5 : 0.53 (4/0.1) [2/0.01]
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | D20 >= 0.5 : 0.45 (8/0.02) [2/0.07]
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | PD2 >= 0.5 : 0.39 (4/0.03) [2/0.15]
| | | | | | | | | | | | | | | | | | | | | | | | | | | | PDA5 >= 0.5 : 0.61 (2/0.15) [0/0]
| | | | | | | | | | | | | | | | | | | | | | | | | | | Sales(t-1) >= 0.45
| | | | | | | | | | | | | | | | | | | | | | | | | | | | PD12 < 0.5 : 0.26 (5/0.01) [1/0]
| | | | | | | | | | | | | | | | | | | | | | | | | | | | PD12 >= 0.5 : 0.34 (2/0) [0/0]
| | | | | | | | | | | | | | | | | | | | | | | | | | PD28 >= 0.5 : 0.26 (7/0.03) [5/0.01]

| | | | | | | | | | | | | | | | | | | | | | | | | SKU23 >= 0.5 : 0.51 (21/0.01) [16/0.04]
| | | | | | | | | | | | | | | | | | | | | | | | PD4 >= 0.5 : 0.59 (7/0.08) [2/0.12]
| | | | | | | | | | | | | | | | | | | | | | | SKU19 >= 0.5 : 0.46 (14/0.03) [7/0.09]
| | | | | | | | | | | | | | | | | | | | | | SKU32 >= 0.5 : 0.29 (16/0.05) [7/0.01]
| | | | | | | | | | | | | | | | | | | | | PD1 >= 0.5 : 0.3 (6/0.03) [3/0.03]
| | | | | | | | | | | | | | | | | | | | SKU26 >= 0.5 : 0.5 (20/0.03) [15/0.03]
| | | | | | | | | | | | | | | | | | | PDA13 >= 0.5 : 0.56 (7/0.04) [2/0.2]
| | | | | | | | | | | | | | | | | | SKU22 >= 0.5 : 0.37 (14/0.05) [8/0.05]
| | | | | | | | | | | | | | | | Sales(t-1) >= 0.46 : 0.32 (18/0.03) [4/0.01]
| | | | | | | | | | | | | | | SKU12 >= 0.5 : 0.3 (24/0.02) [15/0.02]
| | | | | | | | | | | | | | PDA12 >= 0.5 : 0.51 (4/0.05) [2/0.09]
| | | | | | | | | | | | | SKU20 >= 0.5
| | | | | | | | | | | | | | PD20 < 0.5 : 0.42 (26/0.05) [11/0.04]
| | | | | | | | | | | | | | PD20 >= 0.5 : 0.86 (2/0.02) [0/0]
| | | | | | | | | | | | SKU11 >= 0.5
| | | | | | | | | | | | | PD4 < 0.5 : 0.29 (30/0.01) [9/0.01]
| | | | | | | | | | | | | PD4 >= 0.5 : 0.55 (2/0.01) [0/0]
| | | | | | | | | | | SKU21 >= 0.5 : 0.3 (27/0.01) [11/0.03]
| | | | | | | | | | SKU16 >= 0.5 : 0.29 (23/0.02) [18/0.02]
| | | | | | | | | D2 >= 0.5 : 0.18 (5/0.01) [3/0.04]
| | | | | | | | SKU34 >= 0.5 : 0.23 (16/0.01) [9/0.01]
| | | | | | | SKU24 >= 0.5 : 0.28 (24/0.01) [15/0.04]
| | | | | | SKU31 >= 0.5
| | | | | | | PD21 < 0.5 : 0.15 (10/0.01) [4/0]
| | | | | | | PD21 >= 0.5 : 0.38 (2/0.01) [0/0]
| | | | | SKU27 >= 0.5 : 0.5 (19/0.04) [12/0.04]
| | | | SKU25 >= 0.5 : 0.22 (18/0.01) [14/0.01]
| | | Sales(t-1) >= 0.46
| | | | PD2 < 0.5
| | | | | SKU35 < 0.5
| | | | | | PD1 < 0.5
| | | | | | | SKU28 < 0.5
| | | | | | | | SKU23 < 0.5
| | | | | | | | | SKU31 < 0.5
| | | | | | | | | | SKU25 < 0.5
| | | | | | | | | | | SKU7 < 0.5

```
| | | | | | | | | | | | Sales(t-1) < 0.95
| | | | | | | | | | | | | SKU4 < 0.5
| | | | | | | | | | | | | | SKU22 < 0.5
| | | | | | | | | | | | | | | D20 < 0.5
| | | | | | | | | | | | | | | | SKU6 < 0.5
| | | | | | | | | | | | | | | | | SKU17 < 0.5
| | | | | | | | | | | | | | | | | | PDA12 < 0.5
| | | | | | | | | | | | | | | | | | | SKU16 < 0.5
| | | | | | | | | | | | | | | | | | | | SKU32 < 0.5
| | | | | | | | | | | | | | | | | | | | | SKU9 < 0.5
| | | | | | | | | | | | | | | | | | | | | | PDA25 < 0.5
| | | | | | | | | | | | | | | | | | | | | | | PDA24 < 0.5
| | | | | | | | | | | | | | | | | | | | | | | | SKU29 < 0.5
| | | | | | | | | | | | | | | | | | | | | | | | | PD24 < 0.5
| | | | | | | | | | | | | | | | | | | | | | | | | | SKU12 < 0.5 : 0.47 (347/0.04) [170/0.04]
| | | | | | | | | | | | | | | | | | | | | | | | | | SKU12 >= 0.5 : 0.35 (10/0.03) [5/0.02]
| | | | | | | | | | | | | | | | | | | | | | | | | PD24 >= 0.5 : 0.6 (9/0.09) [4/0.09]
| | | | | | | | | | | | | | | | | | | | | | | | SKU29 >= 0.5 : 0.42 (22/0.03) [13/0.05]
| | | | | | | | | | | | | | | | | | | | | | | PDA24 >= 0.5 : 0.57 (3/0.11) [2/0.06]
| | | | | | | | | | | | | | | | | | | | | | PDA25 >= 0.5 : 0.32 (6/0.03) [0/0]
| | | | | | | | | | | | | | | | | | | | | SKU9 >= 0.5 : 0.41 (32/0.03) [12/0.03]
| | | | | | | | | | | | | | | | | | | | SKU32 >= 0.5 : 0.28 (3/0) [0/0]
| | | | | | | | | | | | | | | | | | | SKU16 >= 0.5 : 0.31 (10/0.02) [4/0.02]
| | | | | | | | | | | | | | | | | | PDA12 >= 0.5 : 0.51 (5/0.07) [4/0.08]
| | | | | | | | | | | | | | | | | SKU17 >= 0.5 : 0.36 (9/0.03) [9/0.02]
| | | | | | | | | | | | | | | | SKU6 >= 0.5 : 0.36 (17/0.02) [1/0.01]
| | | | | | | | | | | | | | | D20 >= 0.5 : 0.29 (5/0.03) [0/0]
| | | | | | | | | | | | | | SKU22 >= 0.5 : 0.37 (20/0.05) [12/0.03]
| | | | | | | | | | | | | SKU4 >= 0.5 : 0.3 (10/0.01) [3/0.03]
| | | | | | | | | | | | Sales(t-1) >= 0.95 : 0.34 (27/0.02) [7/0.03]
| | | | | | | | | | | SKU7 >= 0.5 : 0.35 (13/0.03) [11/0.03]
| | | | | | | | | | SKU25 >= 0.5 : 0.25 (5/0.01) [1/0.05]
| | | | | | | | | SKU31 >= 0.5 : 0.13 (3/0.01) [0/0]
| | | | | | | | SKU23 >= 0.5 : 0.53 (40/0.02) [18/0.01]
```

```
|  |  |  |  |  |  |  SKU28 >= 0.5 : 0.33 (13/0.01) [12/0.06]
|  |  |  |  |  |  PD1 >= 0.5 : 0.21 (15/0.03) [7/0.03]
|  |  |  |  |  SKU35 >= 0.5 : 0.93 (2/0) [1/0.01]
|  |  |  |  PD2 >= 0.5
|  |  |  |  |  SKU7 < 0.5
|  |  |  |  |  |  Sales(t-1) < 0.64 : 0.23 (6/0.01) [3/0.06]
|  |  |  |  |  |  Sales(t-1) >= 0.64 : 0.31 (2/0) [5/0.02]
|  |  |  |  |  SKU7 >= 0.5 : 0.02 (2/0) [0/0]
|  |  SKU18 >= 0.5 : 0.56 (72/0.04) [24/0.05]
SKU3 >= 0.5 : 0.75 (70/0.02) [34/0.03]
```

BIBLIOGRAPHY

[1] A.J. Smola, B.S. Schölkopf, A Tutorial on Support Vector Regression, Statistics and Computing, 14 (2004), 199-222.

[2] C. Cortes, V.N. Vapnik, Support Vector Networks, Machine Learning, 20 (1995), 273-297.

[3] B.L. Bowerman, R.T. O'Connell, Forecasting and Time Series: an Applied Approach, Duxbury Press (1993).

[4] C.H. Wu, J.M. Ho, D.T. Lee, Travel-Time Prediction With Support Vector Regression, IEEE Trans. Intelligent Transportation Systems, 5(2004), 276-281.

[5] D. Horsky, S. Misra, P. Nelson, Observed and Unobserved Preference Heterogeneity in Brand Choice-Models, Marketing Science, 25 (2006), 322-335.

[6] D.R. Bell, A. Bonfrer, P.K. Chintagunta, Recovering Stockkeeping-Unit-Level Preferences and Response Sensitivities from Market Share Models Estimated on Item Aggregates, Journal of Marketing Research, 42 (2005), 169-182.

[7] E.W. Foekens, P.S.H. Leeflang, D.R. Wittink, Hierarchical versus other market share models for markets with many items, Intern. J. of Research in Marketing, 14 (1997), 359-378.

[8] G.L. Lilien, A. Rangaswamy, Marketing Engineering, Prentice Hall (2003).

[9] H. Yang, L. Chan, I. King, Support Vector Machine Regression for Volatile Stock Market Prediction, Proc. Intelligent Data Engineering and Automated Learning, 2412 (2002), 391-396.

[10] J. Heizer, B. Render, Operations Management, Pearson Prentice Hall (2006).

[11] J.E. Hanke, D.W. Wichern, Business Forecasting, Pearson Prentice Hall (2005).

[12] J.H. Pedrick, F.S. Zufryden, Evaluating The Impact of Advertising Media Plans: A Model of Consumer Purchase Using Single-Source Data, Marketing Science, 10 (1991), 111-130.

[13] J.M Cadeaux, Category Size and Assortment in US Macro Supermarkets, Int. Rev. of Retail, Distribution and Consumer Research, 9 (1999), 367-377.

[14] L. Sanzogni, D. Kerr, Milk Production Estimates Using Feed Forward Artificial Neural Networks, Computers and Electronics in Agriculture,32 (2001), 21-30.

[15] M. Espinoza, J.A.K. Suykens, B. De Moor, Fixed-Sized Least Squares Support Vector Machines: A Large Scale Application in Electrical Load Forecasting, Computational Management Science, 3 (2006), 113-129.

[16] N. Cristianini, J. Shawe-Taylor, An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods, Cambridge University Press, (2000).

[17] Ö. Gür Ali, S. Sayın, T. Van Woensel, and J. Fransoo, Pooling Information Across SKUs for Demand Forecasting with Data Mining, working paper, (2007).

[18] P. Ongsritrakul, N. Soonthornphisaj, Applying Decision Tree and Support Vector Regression to Predict the Gold Price, Neural Networks, Proc. International Joint Conference, 4 (2003), 2488- 2492.

[19] P.F. Pai, C.S. Lin, A Hybrid ARIMA and Support Vector Machines Model in Stock Price Forecasting, Omega The International Journal of Management Science, 33 (2005), 497-505.

[20] P.M. Guadagni, J.D.C. Little, A Logit Model of Brand Choice Calibrated on Scanner Data, Marketing Science, 2 (1983), 203-238.

[21] P.M. Guadagni, J.D.C. Little, When and What to Buy: A Nested Logit Model of Coffee Purchase, Journal of Forecasting, 17 (1998), 303-326.

[22] P.S. Fader, J.M. Lattin, Accounting for Heterogeneity and Nonstationarity in A Cross Sectional Model Of Consumer Purchase Behavior, Marketing Science, 12 (1993), 304-317.

[23] P.S. Fader, B.G.S. Hardie, Modeling Consumer Choice Among SKUs, Journal of Marketing Research, 33 (1996), 442-452.

[24] P.S. Fader, J.M. Lattin, J.D.C. Little, Estimating Nonlinear Parameters in The Multinomial Logit Model, Marketing Science, 11 (1992), 372-385.

[25] R. Harris, R. Sollis, Applied Time Series Modeling and Forecasting, Wiley (2003).

[26] R.E. Bucklin, S. Gupta, Commercial Use of UPC Scanner Data: Industry and Academic Perspectives, Management Science, 18 (1999), 247-273.

[27] R.J. Roiger, M.W. Geatz, Data Mining: A Tutorial Based Primer, Pearson Education, (2003).

[28] S. Siddarth, R.E. Bucklin, D.G. Morrison, Making The Cut: Modeling and Analyzing Choice Set Restriction in Scanner Panel Data, Journal of Marketing Research, 32 (1995), 255-266.

[29] T. Mitchell, Machine Learning, McGraw Hill, (1997).

[30] T.H. Ho, J.K. Chong, A Parsimonious Model of Stockkeeping-Unit Choice, Journal of Marketing Research, 40 (2003), 351-365.

[31] T.R. Willemain, C.N. Smart, H.F. Schwarz, A New Approach to Forecasting Intermittent Demand for Service Parts Inventories, International Journal of Forecasting, 20 (2004).

[32] V.N. Vapnik, The Nature of Statistical Learning Theory, New York: Springer, (2000).

[33] V.N. Vapnik, Statistical Learning Theory, New York: Wiley, (1998).

[34] V.N. Vapnik, An Overview of Statistical Theory, IEEE Trans. Neural Networks, 10 (1999), 988-999.

[35] W.L. Winston, S.C. Albright, Practical Management Science, Duxbury Press, (2001).

[36] Salford Systems, http://www.salfordsystems.com/112.php

[37] Akşam Newspaper,
http://www.aksam.com.tr/arsiv/aksam/2005/03/16/ekonomi/ekonomi2.html

[38] Zotteri G, Kalchschmidt M and Caniato F, The impact of aggregation level on forecasting performance, International Journal of Production Economics, Vols. 93-94 (2005), pp. 479-491.