GENERALIZED LINEAR MODELLING FOR ASSESSING PROMOTIONS

by

Sergül Acar

B.S., Chemical Engineering, İstanbul Technical University, 2010

Submitted to the Institute for Graduate Studies in
Science and Engineering in partial fulfillment of
the requirements for the degree of
Master of Science

Graduate Program in Industrial Engineering
Boğaziçi University
2013

ACKNOWLEDGEMENTS

I have learned a lot and really enjoyed while working on this thesis.

I would like to express my deepest gratitude to Assoc. Prof. Wolfgang Hörmann, my thesis supervisor for all his support, patience and most importantly his guidance. I feel honoured to have the chance to work with him.

I would like to thank to Assoc. Prof. Aybek Korugan and Prof. Aslı Sencer, the members of my thesis committee, for their valuable comments and suggestions.

I have been supported by TUBITAK domestic masters degree fellowship program for two years. I am also thankful to TUBITAK for providing me financial aid during my graduate study.

I would like to extend my thanks to my closest friends and colleagues for their support and confidence in me.

I would like to offer my special thanks to İsmail Bertuğ Kocabıçak for making me feel strong and like I am not alone and nor I will be. Without him I would not be able to find my light in the life course.

I would also like to express my love and gratitude to my parents for their support, understanding and endless love during my studies. Their encouragement is invaluable for me.

ABSTRACT

GENERALIZED LINEAR MODELLING FOR ASSESSING PROMOTIONS

Price promotions are very important and widely used in marketing and pricing strategy. Therefore, understanding and quantifying the influences of price promotions on demand become interesting. In this study we modelled the demand as a Poisson distributed random variate and built a model using exponential demand function in microeconomics, which considers the variables that have an influence on demand, such as price, yearly seasonality, week days, competition etc. We also generated the necessary data and used our different data sets in the study. Since our main aim is to measure the impacts of price promotions on product demand and forecast the demand, we used generalized linear models, more specifically Poisson regression. GLMs are used for relating random responses to the linear combination of predictor variables. Poisson regression is a GLM for which the response variable is modelled by the Poisson distribution. We found that GLM is very helpful to quantify the effects of price promotions on product demand. If the mean of the daily demand is small, Poisson regression is needed to be used, otherwise ordinary log-level linear regression can be applied instead. We investigated that the results depend very much on the data and number of observations. In addition, we observed that when the price decreases by 20 %, demand of a product increases about 18 %. We checked our "no interaction between covariates" assumption with the help of AIC-Criterion and found out that the simplest demand model is the best model with the smallest AIC_Value. We realized that once having reliable and usable real data, the models we suggested in this thesis can be used to quantify the effects of price promotions on demand. These models can help us to plan promotions, forecast the effects of promotions and make revenue maximization. Throughout this thesis R-Software environment has been used for all kind of computations.

ÖZET

GELİŞMİŞ İSTATİSTİKSEL YÖNTEMLERLE FİYAT PROMOSYONLARININ TALEBE ETKİLERİNİN ÖLÇÜLMESİ

Fiyat promosyonları pazarlama ve fiyatlandırma stratejilerinde sıklıkla kullanılan önemli bir araçtır. Bu nedenle fiyat promosyonlarının talebe etkisinin anlaşılması ve ölçülmesi oldukça ilgi çekicidir. Bu çalışmada talebi Poisson dağılan bir rassal değişken olarak modelledik ve mikroekonomideki üssel talep fonksiyonundan yararlanarak talep üzerindeki, fiyat, yıllık mevsimsellik, haftanın günleri, rekabet gibi etkileri inceleyen bir model geliştirdik. Ayrıca gerekli olan datayı ürettik ve farklı data setlerimizi çalışmamızda kullandık. Temel amacımız fiyat promosyonlarının talep üzerindeki etkilerini ölçmek ve talebi tahmin etmek olduğundan başta Poisson regresyonu olmak üzere genelleştirilmiş lineer modellerden yararlandık. GLM, bağımlı rassal değişkenleri tahmin edici değişkenlerin lineer kombinasyonu olarak ilişkilendirmede kullanılan yöntemlerin genel adıdır. Bir çeşit GLM olan Poisson regresyonu ise Poisson dağılan bağımlı değişkenler söz konusu olduğunda kullanılır. Çalışmamızda gördük ki GLM, fiyat promosyonlarının talebe olan etkilerini ölçmede oldukça yararlıdır. Eğer günlük talep küçük ise Poisson regresyonu kullanılmalıdır, aksi takdirde yerine basit regresyon da kullanılabilir. Sonuçların eldeki dataya ve gözlem sayısına çok bağlı olduğunu tespit ettik. Fiyattaki %20'lik düşüşün talebi % 18 civarında arttırdığını gözlemledik. Bağımsız değişkenlerin birbiri ile etkileşimleri olmadığı varsayımımızı AIC-Kriteri kullanarak test ettik ve en basit modelin en düşük AIC değeri ile en iyi model olduğunu bulduk. Güvenilir ve kullanılabilir gerçek datanın varlığında bu tezde önerdiğimiz modellerin promosyonların talebe etkisini ölçmede kullanabileceğimizi gördük. Bu modellerin promosyon planlamada, promosyon etkilerinin tahmin edilmesinde ve kar maksimizasyonu konularında kullanılabileceği sonucuna vardık.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS

a	Constant of the exponential demand function
b	Constant of the exponential demand function
c	Constant that adjusts the degree of the yearly linear trend
E	Expectation
g^{-1}	Link function
I_i	Week day indicator
m	Number of week days
n	Number of weeks
k	Shape parameter of gamma distribution
p	Product price
p_c	Price of the competitor product
p_s	Price of the substitute product
p_{cg}	Price of the complementary good
q	Product quantity
t	Yearly trend
var	Variance function
X	Covariate
Y	Response variable
α_i	Week day factors
β_0	Intercept of the demand function
β_0'	Logarithm of the intercept
β_1	Coefficient of price in the demand function
Γ	Density function of gamma distribution
δ	Increase in price
ϵ	Error term
η	Linear predictor
θ	Scale parameter of gamma distribution

μ	Mean
μ_{LN}	Mean of lognormal distribution
σ	Standard deviation
σ_{LN}	Standard deviation of lognormal distribution
ϕ	Over-dispersion coefficient
ψ	Dispersion parameter

LIST OF ACRONYMS/ABBREVIATIONS

ANOVA	Analysis of Variance
ANCOVA	Analysis of Covariance
GLM	Generalized Linear Models
AIC	Akaike's Information Criterion
CV	Coefficient of Variation
CI	Confidence Interval
H_0	Null Hypotesis

1. INTRODUCTION

Price promotions are frequently used in retailing. They involve reducing the price of a product or service temporarily to attract customers and develop loyalty for increased future revenue and profit. A retailer's main target is to improve his profitability and his competitive situation in the marketplace, therefore price promotions are the most important and widely used tool in order to reach this target.

Based on different definitions of promotions, Blattberg and Neslin (1990) develop their own definition of price promotion. Their definition states:

"Price promotion is an action-focused marketing event whose purpose is to have a direct impact on the behaviour of the firm's customers."

The application of promotional pricing is relatively simple. Retailers and manufacturers first come to the conclusion that more attention is needed on a label, item or company. They then look at the typical market value of the item or service. They calculate how much of a discount they can afford and decide how long the discount should last.

The important thing to consider here is how to make effective promotions, which result in revenue and profit increase. Although there are some other reasons for making promotions, such as attracting customers and developing loyalty, it is shown that, the primary function of promotions is to increase retailer sales and, consequently, retailer profit. As a consequence retailers always try to find new ways to increase the effectiveness of promotions. Therefore, information about the effects of retail price promotions on the sales of products and competitor performance is crucial. Due to these reasons and since price promotions constitute a critical part of marketing and pricing strategy, understanding and quantifying the impacts of promotions on sales, and in fact on the demand of a product are becoming more and more valuable.

Due to the importance and effectiveness of the price promotions, exploring the effects of promotions on the demand of a product is evolving into an interesting area. Related to this, various questions may come in to mind. What can be other impacts on the demand of a product? Is it possible to quantify the impacts on demand, especially price promotions? How should the data look like in order to quantify the effects? etc. Motivated by these questions, we aim to improve our knowledge in this field, analyse and quantify the effects of price promotions on demand by considering other aspects as well, such as yearly seasonality of demand, week day effects, immeasurable uncertainties and competitive environment.

Questions we want to answer in this study are:

- Has promotion an influence on demand?
- Can we forecast the demand with and without promotions?
- Can we forecast how much the demand is increased by promotions?
- How much were the sales increased by promotion?
- What can be the parameters of the exponential demand function?
- How precise can we estimate the demand model parameters?
- Which statistical techniques can be used to answer these questions and how?

Finding reasonable answers to these questions may lead us to plan promotions in a much better way. We study here a framework model for measuring the impacts of price promotions on product demand. In order to fulfil this aim and accordingly forecast the demand, a statistical approach is necessary. This statistical approach is selected as generalized linear models considering the nature of our response variable, namely demand. In order to build a model, investigate the effects of price promotions on the demand and examine the relationship between demand and other impacts data are needed. Therefore we collected data from two different companies. However the collected data were not useful for our study, because data on the time of promotions and the time of demand are not enough. For this reason we decided to simulate the necessary data by considering the possible aspects and use our simulated data in the rest of the study. Possible models that show the relationship between price and demand are worked on, an exponential demand function is selected in order to define the demand and a model is built which considers the variables that have an influence on demand, such as price, yearly seasonality week days, etc. With the help of the simulated data, influences on demand, related with price promotions are studied precisely by using generalized linear models, more specifically Poisson regression.

It is important to mention here, that in practice observing demand is not possible. Only the demand that ends up with sales can be recorded. Hence, in this study it is assumed that demand is equal to sales. The underlying main assumption under this statement is that there exists enough inventory to fulfil the demand.

This thesis consists of two literal and four analytical chapters. Each of these chapters helps us to reach our goal and understand the impacts on demand precisely. In the analytical chapters all calculations are performed using R.

In Chapter 2 a short literature survey is provided. This part informs about related studies in this area. In Chapter 3, general information about the statistical models used in the study is given. These models include multiple linear regression, analysis of variance (ANOVA), analysis of covariance (ANCOVA), generalized linear models (GLM), more specifically Poisson regression. In Chapter 4 a short review of necessary microeconomic terms, such as price elasticity is given in the context of the used demand model. Moreover the model development steps and related assumptions which are used in modelling the demand are explained in details, including the considered variables which have an impact on demand. The data simulation process is described in details in Chapter 5. This process is examined not only in the single product case, but under competition as well. Chapter 6 discusses the parameter estimation procedure for the simulated demand data. Mainly ANOVA and GLM are utilized in this process. All models are explained in detail and different approaches are established. In all these analysis' it is assumed that the demand is Poisson distributed as this is a realistic assumption for many applications. Moreover, the Poisson distribution is easy to apply due to the fact that its variance is equal to its mean. However, there are other possible distributions to consider as well, such as lognormal and gamma distributions. These alternative models are discussed in this chapter as well. In Chapter 7 some experiments are designed to understand the impacts on demand. Questions about the importance of sample size and mean demand, demand behaviours against different promotion strategies, power of the tests and interactions between covariates are discussed. Chapter 8 contains our conclusions.

2. LITERATURE SURVEY

Due to the importance and complexness of price promotions, many researchers studied the effects of promotions in order to understand their behaviours with different methods and in various aspects.

In all of the summarized papers in this chapter, authors aims are to understand the dynamics behind price promotions with the help of statistics. They use different techniques and various data sets. We selected especially these papers, because they helped us a lot to learn the price promotion environment. They demonstrated us distinct statistical techniques to be used in analysing price promotions. Moreover we realized that generalized linear models are not used frequently in this area.

PAUWELS *et al.* (2004) compare the short- and long-term impact of price promotions and new product introductions on top-line, bottom-line, and stock market. They focus on the automobile industry and apply multivariate time-series models to it. They define the model requirements and create a vector-autoregressive (VAR) model, which suits the requirements well. They consider 41 brands in 6 major product categories. The most stunning result of the study is that, although new product introductions have positive and increasing effects on firm value, sales promotions lower long-term firm value. However, they have positive effects on revenues and on profits in short run.

The main goal of the research of GILBERT and JACKARIA (2002) is to investigate consumer response to the four different promotional deals most commonly used in UK supermarkets, namely coupons, price discounts, samples and "buy-one-get-one-free". Analyzed consumer behaviors are brand switching, brand loyalty, stockpiling, purchase acceleration and product trial. In the study 160 respondents are asked to fill the questionnaire and consequently multi-discriminant analysis is used to check whether there is a relationship between the four consumer promotional approaches and respondents' reported buying behavior, due to the nominal characteristic of the dependent variable. The independent t-test and one-way ANOVA are used to test the results against personal characteristics as well. The results imply that only price discount promotions are statistically significant on consumer's buying behavior. The two most influential variables related to a discount are found to be purchase acceleration and product trial.

PAUWELS *et al.* (2002) explore the long-term effect of price promotions on brand sales, namely category incidence, brand choice and purchase quantity. They use in their research weekly sales data of a perishable and a storable product derived from a scanner panel. Firstly they perform unit-root tests to investigate the evolution of the variables. Then they specify a vector-error correction model (VEC) and vector-autoregressive model (VAR). After that they use impulse-response and multivariate persistence estimates. The most significant results exhibit that, permanent effects of promotions are absent for all sales components. Additionally, they show that the immediate promotional effects are higher for brand choice than for the other two components. Also, the total promotional impact is higher for purchase quantity for storable products, while it is higher for brand choice for perishable products.

MULHERN and PADGETT (1995) try to find out in their study whether shoppers buying promoted items also buy regular price items. They make an in-store survey, which investigates the reason for visiting the store. They collect 412 surveys from two selected stores. 4 hypotheses are defined to determine how price promotions relate to purchase behavior and retail profits. 2 x 2 contingency table analysis and F-tests are performed. The results show that there is a positive relationship between regular price and promotion purchase.

WALTERS (1991) aims to build a framework which investigates the impact of retail price promotions of a store and its competitors on promoted and non-promoted products within the store. Brand substitution effects, the effect of promotion on complementary products and inter-store sales displacement of products are taken into consideration. The product pairs of spaghetti and spaghetti sauce and cake mix and cake frosting are selected to formulate a model describing the promotional effects. A model is developed for each brand in the four product categories. Regression analysis is done and the results demonstrate that retail price promotions create significant complementary and substitution effects within the store. Moreover, the author proves that high prices on one brand mean high sales of substitute brands.

3. MULTIVARIATE STATISTICAL MODELS

There are several multivariate statistical techniques, and some of them are used in this study. It is therefore helpful to explain shortly the multivariate statistical methods used in this work.

3.1. Multiple Linear Regression

In linear regression the aim is to investigate the relationship between deterministic explanatory variables and a single random response variable. Two main goals of a regression model are understanding the influence of the different covariates on the response variable and forecasting the mean response for a given vector of covariates. In building a regression model "the correct model" is not the thing which is looked for, instead a simple model is searched, that is in accordance with the data and fulfils the aims of interest.

The model assumptions of the linear model can be summarized by:

- Linear relationship between mean response and the covariates.
- Constant variance of residuals.
- Independence of residuals.
- Normal distributions of residuals.

In multiple linear regression the model has more than one covariate. The model equation is given below:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n + \epsilon \tag{3.1}$$

where Y is the response variable, X_n is the nth covariate, β_0 is the intercept or regression constant, β_n is the coefficient of the nth covariate, n is the total number of covariates and ϵ is the error term.

The error term ϵ denotes a random vector of independent variates, all with expectation 0 and variance σ^2.

If the total number of variables is not too high a step-wise regression procedure can be started directly. In other words, a regression model is built by either starting from all one-variable models and trying to include more variables, which is called forward stepwise regression, or by starting with a large model and then removing the useless variables, which is called backward stepwise regression. After a not too complicated model is reached, the model assumptions should be checked with the help of diagnostic plots, above all the residual plot. If the constant variance assumption seems to be wrong, transforming the response variable may lead to a better model. If the linear relationship seems wrong, the response variable or covariates may be transformed or quadratic terms can be added to the covariates.

3.2. Analysis of Variance - ANOVA

ANOVA is a statistical tool which is used in comparative experiments, those in which only the difference in outcomes is of interest. It is a generalization of the two-sample t-test to more than two different samples. In these comparative experiments hypothesis tests can be performed, which is useful for decision making. The ANOVA approach can be presented in terms of a linear model. The fundamental difference between regression and ANOVA is that, although both require a response variable with interval scale, ANOVA works with covariates that have a nominal scale, while regression needs covariates with interval scale. The covariates (grouping variables) of ANOVA are called “factors”.

The main assumptions of an ANOVA - test are:

- The observations are independent.
- The distribution of the residuals are normal (Normality assumption).
- The variances of data in groups are equal (Equality of Homoscedasticity assumption).

The most common parameterizastion of ANOVA is to use:

$$y_{ij} = \mu + \alpha_i + \epsilon_{ij} \quad (3.2)$$

where μ denotes the overall mean response, α_i is called treatment effect and ϵ_{ij} is a random error.

In "R" there are two different parameterizations, The first one is the default parameterization and called "contr.treatment". It compares all treatments with a standard treatment. The other one is called "contr.sum", that compares all treatments with the overall mean μ.

3.3. Analysis of Covariance - ANCOVA

Linear models that include both covariates with interval scale and factors with nominal scale are called Analysis of Covariates (ANCOVA). In other words ANCOVA mixes linear regression and ANOVA in order to model the expectation of the response variable.

The main assumptions of an ANCOVA - model are:

• Observations must be randomly sampled from the population and independent from each other.

• There must be a normal distribution of the dependent variable in the population.

• The variances of the dependent variable must be equal for all levels of the categorical independent variable and the covariates.

Different regression models can be generated in ANCOVA. In the simplest case of a single covariate x and a single factor the model then has the form:

$$Y = \beta_i + \gamma x \quad (3.3)$$

In this model it is assumed that all group have the same slopes (γ) but different intercepts (β_i).

3.4. Generalized Linear Models GLM

A large and important class of statistical models is summarized in the framework of generalized linear models (GLM). They are used for relating random responses to the linear combination of predictor variables. These predictor variables may be used for many different types of dependent variables and may possess special error structures. While linear regression models are appropriate for continuous dependent variables; models for rates and proportions, binary, ordinal and multinormal variables and counts can be handled as GLMs. The GLM approach is important and attractive due to the fact that it gives a general theoretical framework for many commonly encountered statistical models. Moreover, it simplifies the implementation of these models with statistical software, because with essentially a single algorithm, estimation, inference and assessing model adequacy for all GLMs can be handled. Generalized linear models involve 4 main parts. These are:

(i) Linear predictor: $\eta = X\beta$

(ii) Link function, which relates the linear predictor to the mean of the response variable.

$$g(\mu) = \eta = X\beta \tag{3.4}$$

$$\mu = g^{-1}(\eta) = g^{-1}(X\beta) \tag{3.5}$$

(iii) Distribution of the response variable. These are usually a member of the exponential family, which includes normal, lognormal, poisson, binomial, gamma and hypergeometric.

(iv) Dispersion parameter ϕ, which is related to the spread of the distribution and depends on the relationship between the mean and the variance. For some distributions, such as Poisson and binomial this parameter is fixed, while for others, such as normal and gamma distribution, it is an additional parameter to be modelled and estimated.

The selected distribution family and the dispersion parameter specify together the variance of the response as a function of the mean response.

3.4.1. Poisson Regression

Poisson regression is the generalized linear model for which the response variable is modelled by the Poisson distribution. Random variables with non-negative integer values, namely counts can be often modelled well by the Poisson distribution. Since the mean value of the Poisson distribution is larger than zero, using Poisson regression in such cases is the most appropriate way. So the logarithm of the mean of the response variable is linked to a linear function of response variables such that $log(\mu(y)) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots$. In other words, in the typical Poisson regression model the log outcome rate is expressed as a linear function of a set of predictors. It is important to note that, for large mean values the Poisson distribution is well approximated by the normal distribution. The main assumptions of the Poisson regression are as such:

- Logarithm of the response variable changes linearly with the covariates.
- Changes in the response variable originating from different are multiplicative.
- Poisson distribution has only a single parameter μ, that is both the mean and the variance.
- Observations are independent.

4. MODELLING THE DEMAND

In order to examine the effects of price promotions on demand, a demand model is necessary; therefore, we start our study with modelling the demand. This generated demand model that represents the impacts on demand, we will use in our study to quantify the influences, mainly of price promotions, on demand.

We assume that demand can be modelled as a random variate and our aim is to model it by considering different aspects. The important point is to decide on the distribution of daily demand. As demand cannot be negative, it can follow, for example, Poisson, lognormal or gamma distribution. Lognormal and gamma distributions are continuous probability distributions, while Poisson distribution is discrete. In cases such as demand of gasoline or vegetables, demand can be a continuous variable, but for most goods it is discrete. For instance, demand of coke in the can is a discrete measure. So, it is substantial to have an appropriate distribution for the demand.

In our study, we consider that our observed demand can be explained by Poisson process. A Poisson process is the stochastic process in which events occur continuously and independently of one another. It is a collection of random variables, which contains the number of events that have occurred between a time interval. Since demand is discrete, independent of the time since the last occurrence and occurs in a fixed interval of time, in our case, it can be explained as the number of events in a Poisson process. Therefore it has non-negative integer values and follows a Poisson distribution. It has only a single parameter μ, that is both the mean and the variance. Thus, if two demand values are both Poisson random variates, the only way that they can have the same variance is if they have the same mean. So, under these three alternatives the Poisson distribution is the easiest one to use. However, there are other possible distributions for demand and they are worthwhile to be examined, which explained in detail in Section 6.4.

After deciding on the distribution of demand, we start modelling the price influences on the daily demand. First the fixed mean of demand case is considered, however, this model is not enough. Therefore, we model the influences on the daily demand by defining a deterministic mean demand function that depends on different available information, so-called covariates. At this point the most important covariate is the price of the product itself. Afterwards, we are going to add other prices, such as competitor product prices, substitutional products prices and complementary good prices into the mean demand model. We also decide that considering only price influence on demand is not enough. Week day influences should also be included in the mean demand model. In addition, there are unknown effects on demand as well, which are added to the mean demand model as random noise. As a result price and week days are defined as covariates in the model and their functional relationships with the mean of the daily demand are modelled by using Poisson regression. We selected the simple model because it contains all assumptions of standard a GLM and parameter estimation is possible with standard software.

4.1. Price and Price Elasticity

As we mentioned before, we start our demand model by considering the price impact only. Therefore, understanding the price structure of the demand of a product is necessary, so we begin with a basic micro economics survey about price elasticity.

In every day life it is impossible to find a market which possesses only one producer, without any competitor. Every day goods have always several competitors or substitute products. However, in measuring the influences of promotions on demand, it is necessary to make some assumptions in order to develop the model. Hence, a simplifying assumption is that, there is only a single product in the market at a specific segment. Such a market is called monopolistic market.

In a monopolistic market, where a specific person or enterprise exists as the only supplier of a particular commodity, there is a lack of economic competition to produce the good or service and a lack of viable substitute goods. In such a market a single supplier produces and sells a given product. In some cases, although there are other suppliers in the industry and/or close substitutes for the goods being produced, the company possesses the main market power. This case can be defined as monopolistic competition. In a monopolistic market or monopolistic competition environment the supplier can determine the sales price of the product he produces and the sales of this product are only influenced by the sales strategy of this single company.

Considering the sales quantity (q) is a function of price (p) it can be expressed for example with the well known demand function $q(p) = ap^b$ where a and b are constants. Then the turnover is equal to $pq(p)$. Since we assume that, in our study the sales are equal to demand quantity, q can be considered as demand. In this situation there are 3 main cases which describe the turnover.

Case 1: $b = -1$

$$Turnover = pq(p) = pap^{-1} = a \tag{4.1}$$

The turnover is constant for different prices.

Case 2: $b > -1$ (e.g. $b = -0.5$)

$$Turnover = pap^{-0.5} = ap^{0.5} \tag{4.2}$$

The turnover increases with increasing price, which indicates inelastic demand.

Case 3: $b < -1$ (e.g. $b = -2$)

$$Turnover = pap^{-2} = ap^{-1} \tag{4.3}$$

The turnover decreases with increasing price, which indicates elastic demand.

At this point it is necessary to specify demand elasticity, elastic and inelastic demand.

Price elasticity: Responsiveness of the quantity demanded of a good or service to a change in its price.

Inelastic demand: An increase in unit price will tend to increase revenue, while a decrease in price will tend to decrease revenue. In other words if a demand is perfectly inelastic, the quantity purchased does not change, no matter what the price is.

Elastic demand: For elastic goods, an increase in unit price will tend to decrease revenue, while a decrease in price will tend to increase revenue.

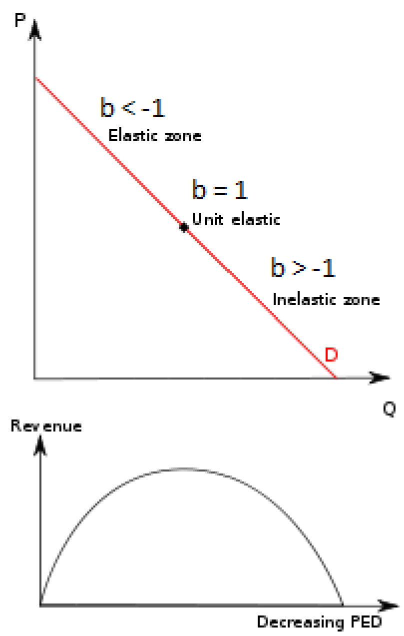

Figure 4.1. Price elasticity of demand and revenue.

In Figure 4.1, the relationship between demand and turnover for a linear demand curve is given. As the price decreases in the elastic zone, turnover increases, while in the inelastic zone turnover decreases. At the quantity where price elasticity of demand is equal to 1, in other words in the unit elastic zone, the total revenue is maximized. Hence, Case 1 corresponds to the unit elastic zone on the graph where the turnover is constant, while Case 2 is identical to the inelastic zone. Due to the fact that, turnover decreases as price increases in Case 3, this situation is consistent with the elastic zone.

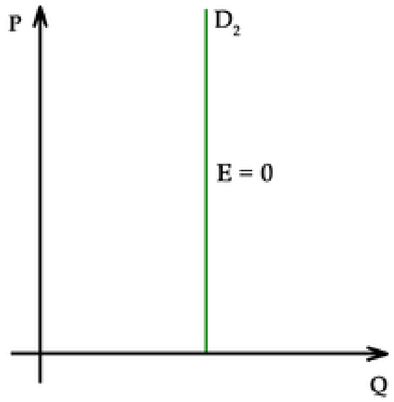

Figure 4.2. Inelastic demand.

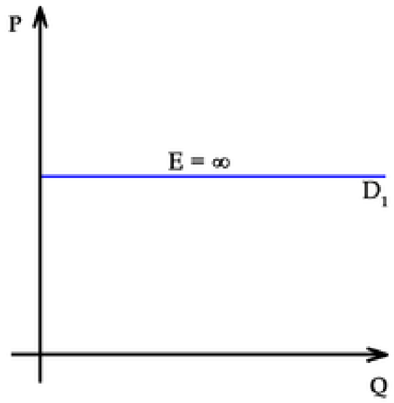

Figure 4.3. Elastic demand.

Fundamentally, considering a monopolistic market, it is reasonable to examine, if the demand of a product is only influenced by the price of the product. At the beginning of the experiments, a very simple model is used, such that only the price of the product influences the demand of this product.

4.2. Demand Depending on a Single Price

Demand is an unknown variable. It changes every day due to different effects, but mainly because of customer preferences. Observing the mean of the daily demand directly is not possible. However, in order to model the daily demand, different assumptions can be made.

4.2.1. Constant Mean Model

The most basic model uses a constant mean. Demand is modelled as a random variate and is assumed to be Poisson distributed. For the Poisson distribution we need only one parameter, namely μ, since $E(Y) = var(Y)$. So it is really easy to assign a constant mean demand value as the parameter of the Poisson distribution and model the daily demand according to this fixed parameter.

In modelling the demand with constant mean, we have to consider the influences of price promotions as well. It is clear that there are not price promotions on each day. Therefore, firstly the existence of price promotions should be defined. Afterwards, the weight of this impact on demand should be determined. For example, a promotion impact weight which is equal to 0.25 means that, in the case of a price promotion, demand of a product is increased by 25 %. These weights are assumed to be constant in this model and can be changed according to the assumptions.

In sum we need a regression model allowing to Poisson distributed daily demand with a constant mean value and considers the existence of price promotions and their impact weights on the daily demand.

Under these assumptions, the new demand model, which assumes demand is Poisson distributed with a constant mean and allows excess demand if there is a price promotion, can be formulated. The existence of promotion is described by the " prom_yes-no" variable, which only contains zeros and ones according to the absence or presence of promotion.

$$Demand = Poisson(\mu_{demand}(1 + (prom_{yes-no})(prom_{weight}))) \tag{4.4}$$

4.2.2. Deterministic Demand Model

Demand can be modelled alternatively and in a more sophisticated way by modelling its mean. The way of modelling the mean demand, which comes to mind first and which is mostly used, is putting the deterministic demand models into use, that are considering the price impact. We can use these functions in order to model the mean demand by using price as a covariate. In this case, for different prices we get different mean values. Various types of these models are expressed below. In Section 4.1, demand was explained by a deterministic exponential demand function. In that section, constants used in the model were expressed as a and b, but from now on we are going to call them β_0, β_1 etc. The value of the parameter β_1 determines the elasticity of the demand, as it was mentioned in Section 4.1. We use logarithmic form of the demand function, in order to get an additive model without interactions.

The main form of the demand function, that is considered in this study is the exponential demand function. The influence of price on mean demand, under the consideration of exponential demand function, can be modelled by:

$$log(\mu_{demand}) = log(\beta_0) + \beta_1 log(p) \tag{4.5}$$

Since β_0 and $log(\beta_0)$ are both constants, we can define a new parameters, such as β_0' which is equal to $log(\beta_0)$. So, the exponential demand function can be expressed in the following way:

Exponential demand function:

$$log(\mu_{demand}) = \beta_0' + \beta_1 log(p) \tag{4.6}$$

Another deterministic demand model uses the rational demand function. In general, it is not very different from the exponential demand function, the only difference is the absence of the coefficient of log(p).

Rational demand function:

$$log(\mu_{demand}) = log(\beta_0) - log(p) \tag{4.7}$$

A linear demand function can be considered as well; however in this case, the log mean demand formulation is changed.

Linear demand function:

$$log(\mu_{demand}) = log(\beta_0 + \beta_1 p) \tag{4.8}$$

Since the mean function of product demand contains only price as a variable, if the price is kept constant, then a constant mean demand, which is totally independently and identically distributed, is obtained. In other words, at the same price levels the same mean demand amount is observed.

4.3. Demand Depending on Several Prices

The existence of the demand depending on several prices can be examined in three cases, namely competition effect, substitution effect and complementary product effect, which have not only separate, but also combined impacts on demand. In all three cases, the only difference between the single price dependent demand case is adding also the product price of each of these elements with a corresponding coefficient $\beta(.)$ into the mean demand function.

4.3.1. Competition Effect

Competition is a complex phenomenon. Firms attempt to choose the group of services and features that maximize profits in an environment where they compete across the entire retail and marketing mix, persuading customers with an attractive set of products, competitive prices, convenient locations, and a host of other services, features, and promotional activities. Therefore, considering an oligopolistic market instead of a monopolistic market is more realistic. In exploring the factors which have an impact on demand, the pricing strategies of the competitors should be evaluated as well. All other influences on demand remain the same, only the price of the competitor (p_c) should be added to the demand function, because prices of the competitors may have an impact on the demand. In Equation 4.9. the log mean function of the demand under competition can be seen:

$$log(\mu_{demand}) = \beta_0' + \beta_1 log(p) + \beta_2 log(p_c) \tag{4.9}$$

4.3.2. Substitution Effect

Besides the competition effects, it is important to consider also the substitution effects. The substitution effect can be defined as the change in demanded quantity due to a change of the price of a good which induces buyers to substitute the purchase of one good by a similar one. When the price of the good increases, the relative prices of substitute goods decreases, and as a result buyers are inclined to buy more of the other substitute goods and less of this good. The result is a decrease in the quantity demanded. Vice versa, when a decrease in price occurs, buyers are inclined to buy less of the other substitute good and more of this good. The result is an increase in the quantity demanded. For this reason, substitution effects should be included in a model which examines the demand of a product. Considering all other important influences on demand, just the price of the substitute product (p_s) is added into the demand function.

Considering the competition and substitution effects the new log mean function of the demand is:

$$log(\mu_{demand}) = \beta_0' + \beta_1 log(p) + \beta_2 log(p_c) + \beta_3 log(p_s). \tag{4.10}$$

4.3.3. Complementary Product Effect

Apart from competition and substitution, complementary products play a serious role in the market. A complementary good has a negative cross elasticity of demand, in contrast to a substitute good. This means that, a good's demand is increased when the price of another good is decreased. Conversely, the demand for a good is decreased when the price of another good is increased.

The model including the complementary effect is given in Equation 4.11. Here p_{cg} indicates the price of the complement.

$$log(\mu_{demand}) = \beta_0' + \beta_1 log(p) + \beta_2 log(p_c) + \beta_3 log(p_s) + \beta_4 log(p_{cg}). \tag{4.11}$$

4.4. Time Influences on the Mean Demand

So far we have only talked about price impact on demand, but the daily demand of a product is not only affected by the price. We therefore also consider the time influences on the mean demand.

Week day may have an important impact on demand. On every single day the consumer preferences may change, but in the big picture it can be often seen that, every week day has a different and specific coefficient, which shows the impact of the day. Even if the price is kept constant during the week, the mean demand of a product often varies during the whole week according to the week day. Sometimes we can see that, although there are no price promotions, the demand of a product can increase due to the week day, for example on weekends. For instance, it is obvious that people do more shopping on weekends than on weekdays. Since this has an important impact on the demand of a product it must be included in the mean demand model as well.

$$log(\mu_{demand}) = \beta_0' + \beta_1 log(p) + \sum_{i}^{m-1}(I_i \alpha_i) \tag{4.12}$$

where m is the number of week days, I_i is the weekday indicator and α_i denotes the week day factors changing according to the day.

Another element that affects the demand of a product can be yearly seasonality. It is hard to assume that, the mean of the daily demand is constant during the whole year. Consumers' demand on a specific product can follow a particular trend. If the trend is linearly increasing, then the demand cannot be the same at the beginning and at the end of the year, even if the sales prices are the same. A specific example of this situation can be demand of ice-cream during a year. Its demand increases during summer time, while it decreases in winter time. Another example would be demand of a newly produced computer. It is high when the new computer model firstly appears in the market, but since technology always improves, new products find places in technology markets every day, the product which has entered the market two months ago, can be categorized as an old model with the entry of a new model. As a consequence, the demand of the previous product is expected to decrease. The same thing is valid for the trend of a newly published book. Its mean at the beginning may be high and even may increase when people start to read it and talk about it with their relatives. However, after some time the demand will decrease, since new books are published continuously. In order to see the effect of trend in the examined data, we add a linear trend to the mean function of the demand.

$$log(\mu_{demand}) = \beta_0' + \beta_1 log(p) + \sum_{i}^{m-1}(I_i\alpha_i) + log(ct) \quad (4.13)$$

where t defines the linear trend and c is the constant that adjusts the degree of trend.

One major assumption is that the promotion does not change the general characteristics of the demand distribution, but it mainly depends on the types of the goods. We tried to model something comparatively stable.

4.5. Unknown Influences on the Mean Demand

As we assume that the demand is Poisson distributed, we already model randomness. However, since the daily mean demand is unobservable, it may be assumed to have randomness due to unknown effects such as weather conditions, disaster etc., which have also influence on the demand. We should add these unknown influences into the mean demand model, in order to include unknown effects. If we ignore these effects, they can cause over-dispersion in Poisson distributed data, which means that the variance of the demand data is larger than the mean demand. As it is already known, the key assumption of the Poisson regression model is that the variance equals the mean:

$$var(Y) = E(Y) = \mu. \tag{4.14}$$

However, count data often exhibit over-dispersion, with a variance larger than the mean. This situation can be formulated by assuming that the variance is proportional to the mean, say

$$var(Y) = \phi E(Y) = \phi\mu \tag{4.15}$$

If $\phi = 1$ then the variance equals the mean. If $\phi > 1$, we have over-dispersion.

In order to manage the over-dispersion problem in the model a random component can be added to the mean demand function. This component can be treated as a nuisance variable, which is not of immediate interest but which must be accounted for in the analysis of those parameters which are of interest. This random variable brings to the model extra randomness, in addition to the natural randomness of the Poisson distributed data. With the help of this parameter over-dispersion in the model can be explained.

Adding randomness to the mean of the Poisson distribution, makes the mean demand model more complicated, since it becomes a stochastic variable model. The random effect added to the mean demand model is typically log-normally distributed. In the model shown in Equation 4.16, which describes the logarithm of the mean of demand, the random noise ϵ follows a normal distribution. Random noise in the mean of the demand function follow the lognormal distribution due to the exponentiation process.

$$log(\mu_{demand}) = \beta_0' + \beta_1 log(p) + \sum_{i}^{m-1}(I_i\alpha_i) + log(ct) + \epsilon. \qquad (4.16)$$

To sum up; we build a model, which considers the daily demand as Poisson distributed and which models the mean demand by using an exponential mean demand function. Our model assumes a functional relationship between mean demand and corresponding covariates, such as price, time, weekly seasonality and unknown influences on demand. With the help of this model we can quantify the impacts on demand and estimate the parameters of the model by using Poisson regression. Since, the logarithm of the mean demand functions are used in order to avoid interactions, the "log - link function" of the GLM (Poisson regression) is used to estimate the parameters.

5. SIMULATION OF THE DEMAND DATA

After we have determined the mean demand model, that we will use in order to quantify the effects of promotions on the demand of a product in Chapter 4, we need demand data to estimate the parameters of our models. Getting real demand data from a company is the first option that comes to the mind. However, finding real and suitable demand data to quantify the effects of price promotions on demand is a difficult task. Data on the time of promotions and the time of demand are not enough. Therefore, it was decided to generate the demand data, which will be used in the rest of our study. Possible effects on demand are considered and demand data are generated accordingly. This chapter explains how we generated the demand data and what kind of path we followed during this process.

5.1. Simulation of Demand Using Mean Demand Model

In order to start our demand data simulation study with an easy model, we consider the case, where price is the only independent variable in the model. So, in the data matrix there should be only price and demand data which are depending on the price.

Here we will use the exponential log mean demand model,

$$log(\mu_{demand}) = \beta_0' + \beta_1 log(p) \tag{5.1}$$

that is explained in Section 4.2.2, in which only the influence of product price on demand is examined. It is important to note here that other versions of the deterministic demand functions, which are explained in Section 4.2.2, can be used easily as well.

In order to obtain a mean demand value, we have to specify a price value for the mean demand function. If we give use than one price value, we get more than one mean demand value. Since we have to test the effects of price promotions on demand, we have to specify different price levels, hence we have to build a price vector, which contains different prices. In our study we considered 4 different price levels, namely 1, 0.9, 0.8, 0.7. The price value 1 is always considered as standard price in our simulation, and other price values are determined according to the assumed price reduction. For example price value 0.9 indicates 10% price reduction, where 0.7 indicates 30% price reduction. Here comes another important point, namely sample size, because the price vector, that will be used in the experiments must have the same length as the demand data, otherwise we cannot generate data. In the experiments n is the number of weeks of the year and m is the number of days of the week in which sales occurs, so our sample size becomes nm. As a result our price vector length should be nm. Therefore, considering that we have 4 different price levels to be tested, we have to replicate this vector $nm/4$ times in order to get a vector of length nm. Obtaining a price vector of length nm is not enough. We want to produce a vector which has a fixed price level always during a full week. We called this vector "pv" in our simulation code. First we produce a matrix, which has our price levels in each of its rows. Then we transpose this matrix in order to get the same price sequentially, so we can have a fixed price during a full week. But still we have a matrix of prices in hand. In order to transform this matrix into a vector of length nm we use "as.vector" command, which outputs the required "pv" vector. The "as.vector()" command of R applied to matrices, returns a vector holding all column vectors of the matrix concentrated to a single vector. So far, we determined the mean demand function, different price levels, the sample size and the corresponding price vector that is used in the calculation. Since we assume that the demand data are Poisson distributed, that information is enough to generate demand data of length nm.

It is critical to remember that, in the log mean demand function β_0' and β_1 are the constants, which determine the mean demand value. If big values are assigned to β_0', then a big mean demand value is observed. The elasticity level of the demand is determined by changing β_1, as explained in Section 4.1.

Below the code we have developed is given. The last line of the code gives the generated data matrix, which contains the demand and the price vectors.

```
generate.demand<- function(beta0_prime,beta1,p, n, m){

#outputs a data matrix which contains demand and price of length n*m
#n... number of weeks
#m... number of weekdays
#p... price levels
#beta0_prime and beta1... constants of the mean demand function

demand <- rep(0, n*m)
#creates a vector of length n*m

pv <- as.vector(t(matrix(rep(p,n*m/length(p)),ncol=n*m/length(p))))
#replicates p vector, so that we obtain a price vector of lenght n*m, with 4 different
#prices

logmu_demand <- beta0_prime+beta1*log(pv)
#deterministic mean demand function (exponential)

demand<- rpois(n*m,exp(logmu_demand))
#generates n*m many Poisson data with mean log(beta0)+beta1*log(pv)

data.frame(demand,pv)
}
m=6
beta0_prime=log(30)
beta1=-1
p=c(1,0.9, 0.8, 0.7)

res <- generate.demand(beta0_prime=log(30),beta1=-1,p=c(1,0.9, 0.8, 0.7), n=52, m=6)
```

5.2. Simulation of Demand Under Time Influences

As we discussed in Chapter 4, price is not the only influence on demand. There are time influences as well, such as yearly seasonality and week day effects. In the data generation process, the mean demand function with yearly seasonality is selected (see Section 4.4):

$$log(\mu_{demand}) = \beta_0' + \beta_1 log(p) + log(ct) \tag{5.2}$$

The yearly trend can behave differently. In our study we assumed that the demand data which holds nm daily observations follows a linear trend. However we want to show another possibility of the behaviour of the yearly trend as well. For example, trend which behaves linearly increasing until the half of the data and then decreases linearly in the other half of the data can be used. For this triangular trend two different groups for the time, namely first half and second half of the data are classified and different regression coefficients are allowed for the two groups. Thus separate analysis of monotonic portions of the curve is used. These two kinds of trends, namely linear and changing trends, are added to the model and effects of trend in relationship with price is detected. In Figures 5.1 and 5.2 those two kinds of trend are illustrated.

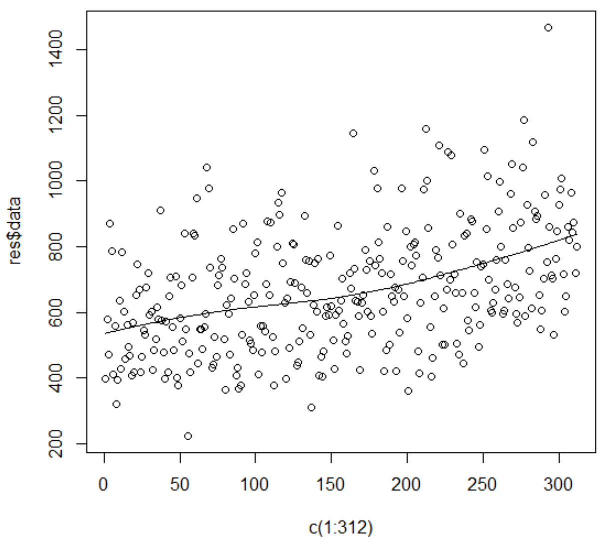

Figure 5.1. Linear trend.

Since we use logarithm of the mean of the demand function, we add the logarithm of the trend vector to the mean demand model in Equation 5.2. For this reason, independently of the shape of the yearly trend, expectation of the logarithm of the linear trend term in the model should be equal to 0, because otherwise we change the

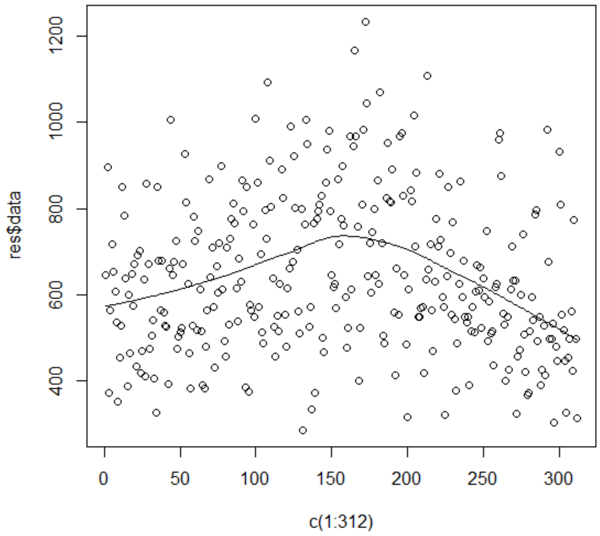

Figure 5.2. Changing trend.

mean of the demand.

In order to create yearly trend a function, namely "trendpattern", that describes the trend is created and a variable which is called "maxchange" is defined in order to determine the start and end points of the trend. For instance, if the "maxchange" value is 0.1, then linear trend vector which has nm points starts with 0.9 and ends with 1.1, giving the expectation equal to 1. Similarly, when "maxchange" is set to 0.5, the trend increases linearly between 0.5 and 1.5.

Clarifying the constant "c" in the trend component of the log mean demand function is also important. We need this constant in order to define the slope of the yearly trend. In general "c" is the increment between the values of the linear trend vector. In other words, we can write "c" in terms of "maxchange". First, we take the difference between the first and the last value of the trend factor vector:

$$(1 + maxchange) - (1 - maxchange) = 2maxchange \quad (5.3)$$

Since we have $n - 1$ data points, where n is equal to sample size in this context, we have to divide this calculated difference over $n - 1$ in order to find the increment value.

So we obtain the relationship between constant c and the parameter "maxchange":

$$c = 2maxchange/(n-1) = maxchange/((n-1)/2) \tag{5.4}$$

Below the function which produces the increasing linear trend vector is seen:

```
trendpattern <- function(n,maxChange){
# outputs a trend factor vector of length n with a linear increasing trend.
# the average of the result vector is 1
# the first value in the result vector is 1-maxChange
# the last value in the result vector is 1+maxChange

1+maxChange*(1:n-(n+1)/2)/((n-1)/2)
}
```

Linear trend is not the only option. We can generate changing trend, which we have explained previously. The logic remains unchanged. Instead of generating only linearly increasing trend, we have linearly increasing values in the first part and linearly decreasing values in the second part. Then we combine these two sub vectors of trend and get a up and down changing trend vector, which is called "tp" in our code below.

```
trendpattern1 <- function(n,maxChange){
# outputs a trend factor vector of length n with a linear increasing trend.
# the average of the result vector is 1
# the first value in the result vector is 1-maxChange
# the last value in the result vector is 1+maxChange

1+maxChange*(1:n-(n+1)/2)/((n-1)/2)
}

trendpattern2 <- function(n,maxChange){
# outputs a trend factor vector of length n with a linear increasing trend.
# the average of the result vector is 1
# the first value in the result vector is 1-maxChange
```

```
# the last value in the result vector is 1+maxChange

1-maxChange*(1:n-(n+1)/2)/((n-1)/2)
}

tp <- c(trendpattern1(156,0.5), trendpattern2(156, 0.5))
```

The second time influence on the demand is the effect of the week days. Here we use the mean demand function which is given in Section 4.4 again, including the week day effects:

$$log(\mu_{demand}) = \beta_0' + \beta_1 log(p) + log(ct) + \sum_{i}^{m-1}(I_i\alpha_i) \tag{5.5}$$

It is assumed that on different week days, demand is different, such that each week day has a specific factor. For instance, since Saturday is weekend, the demand should be more than on Tuesday; or on Friday customers are more likely to buy than on Monday. According to these assumptions a day factor vector, which has m elements due to the considered days in a week, is determined and given as default in the experiment. The important point is that again the average of these factors, as it is mentioned before, should be equal to 1 in order to keep the mean of the demand unchanged. In the experiments day factor vector is selected as dayfact = c(0.7, 0.6, 0.9, 0.8, 1.3, 1.7). The code below shows the demand data simulation under price, yearly seasonality and week day effects:

```
generate.demand<- function(beta0_prime,beta1,p_united, n, m, trendpatt, dayfact){
#outputs a data matrix which contains demand and price of length n*m
#n... number of weeks
#m... number of weekdays
#p_united... price levels
#beta0_prime and beta1... constants of the mean demand function
#trendpatt... output of the function ‘‘trendpattern’’
```

```
#dayfact... different factors of the weekdays

demand <- rep(0, n*m)
#creates a vector of length n*m

day <- 1+(0:(n*m-1))%%m
#modular of n*m many samples between 1 to n*m

pv <- as.vector(t(matrix(rep(p,n*m/length(p)),ncol=n*m/length(p))))
#replicates p vector, so that we obtain a price vector o length n*m, with 4 different
#prices

logmu_demand <- beta0_prime+beta1*log(pv)
#deterministic mean demand function (exponential)

     demand<- rpois(n*m,exp(logmu_demand + log(trendpatt) + log(dayfact[day])))
#generates n*m many Poisson data under price, trend and weekday influences
data.frame(demand,pv,t=1:length(pv), day)
}

m=6
n=52
beta0_prime=log(30)
beta1=-1
p=c(1,0.9, 0.8, 0.7)
ps = c(0.7, 0.8 ,0.9, 1)
#use of sequence b in price timing sequence
p_united =cbind(p,ps)
trendpatt <- trendpattern(n*m, 0.2)
#yearly trend function
dayfact = c(0.7, 0.6, 0.9, 0.8, 1.3, 1.7)

res <- generate.demand(beta0_prime=log(30), beta1=-1,
p_united=c(1, 0.9, 0.8, 0.7,0.7, 0.8 ,0.9, 1),
n=52, m=6, trendpatt=trendpattern(n*m, 0.2), dayfact=c(0.7, 0.6, 0.9, 0.8, 1.3, 1.7))
```

When the time influences are added to the mean demand model, price timing becomes important. Using a correct price timing sequence is necessary in order to distinguish the price effects. This means that, if we do not design the price vector carefully, we can always see for example a different price on the same week day. Consequently distinguishing the effect of this particular week day becomes impossible. Therefore, different price timing sequences can be used and chosen arbitrarily in the simulation code.

Considering the code in Section 5.1, the defined vector "pv" contains the price timing sequence. In this example we used a sequence, which has prices following each other in an ordered way. In other words, the sequence is like that: *p1, p2, p3, p4, p1, p2, p3, p4, p1....* But, since in this code we do not include any time influences, the price timing sequence is not that important. In the codes which are generated under time influences we have to change the sequence in order to distinguish the effects coming from week days better.

One possibility is to use a random price timing sequence. If the random price timing sequence is selected, then, ignoring the random impacts, mainly there is no problem in distinguishing the effects of week days on demand. Or, another sequence can be used, such as: *p1, p2, p3, p4, p4, p3, p2, p1, p1, p2 ....* This sequence can be created arbitrarily, but since we cannot go over all possibilities, we discused only three of them, which are summed up below:

Sequence a : *p1, p2, p3, p4, p1, p2, p3, p4, p1 ...*
Sequence b: *p1, p2, p3, p4, p4, p3, p2, p1, p1, p2 ...*
Sequence c: Random timing for pricing

In order to obtain a "pv" vector which has a price timing of Sequence b, we reformulate the price vector which contains different price levels. In stead of using p=*(p1, p2, p3, p4)* we use p=*(p1, p2, p3, p4, p4, p3, p2, p1)*, as we have shown in the previous simulation code with "p_united" variable. So, in fact, the price timing sequence is related to the design of experiment. It is fundamental to make a design that separates the interactions of variables.

5.3. Simulation of Demand Under Unknown Influences

Besides the randomness of the demand that comes from Poisson distribution, it is sensible to add randomness to the mean of the demand. This leads to a complicated model: The influence of price and time influences and random noise on the mean of the demand data are combined. Generating the demand data under unknown influences is performed by using the mean of the demand function below, which is explained in Section 4.5. Here ϵ indicates the random noise:

$$log(\mu_{demand}) = \beta_0' + \beta_1 log(p) + \sum_{i}^{m-1}(I_i\alpha_i) + log(ct) + \epsilon \tag{5.6}$$

It must be checked that the mean value of the random noise is equal to 0, in order to keep the mean of the demand value unchanged, because we use additive random noise. If we add random noise multiplicatively to the mean demand model, then we have to use 1 as the mean value of the random noise. Only the change in the standard deviation of random noise should affect the results. Hence, different standard deviation values can be applied and their results can be interpreted. However, in our study 2 different standard deviation values are used, 0.1 and 0.2. In the mean demand function in the code shown below, noise is normally distributed; however when the mean demand function is exponentiated, then the random noise becomes lognormally distributed. Since our demand data contain nm observations, our random noise vector must have length nm as well.

```
generate.demand<- function(beta0_prime,beta1,p_united, n, m, munoise=0, sdnoise=0.2,
trendpatt, dayfact, epsilon){

#outputs a data matrix which contains demand and price of length n*m
#n... number of weeks
#m... number of weekdays
```

```
#p_united... price levels
#beta0_prime and beta1... constants of the mean demand function
#trendpatt... output of the function ‘‘trendpattern’’
#dayfact... different factors of the weekdays
#munoise... mean value of random noise
#sdnoise... standard deviation of random noise

demand <- rep(0, n*m)
#creates a vector of length n*m

day <- 1+(0:(n*m-1))%%m
#modular of n*m many samples between 1 to n*m

pv <- as.vector(t(matrix(rep(p,n*m/length(p)),ncol=n*m/length(p))))
#replicates p vector, so that we obtain a price vector o length n*m, with 4 different
#prices

logmu_demand <- beta0_prime+beta1*log(pv)
#deterministic mean demand function (exponential)

    demand<- rpois(n*m,exp(logmu_demand + log(trendpatt) + log(dayfact[day])
    +  epsilon))
#generates n*m many Poisson data under price, trend, weekday and random influences
data.frame(demand,pv,t=1:length(pv), day)}

m=6
n=52
beta0_prime=log(30)
beta1=-1
p=c(1,0.9, 0.8, 0.7)
ps = c(0.7, 0.8 ,0.9, 1)
p_united =cbind(p,ps)
#use of sequence b as price timing sequence
trendpatt <- trendpattern(n*m, 0.2)
#yearly trend function
dayfact = c(0.7, 0.6, 0.9, 0.8, 1.3, 1.7)
munoise=0
```

```
sdnoise=0.2
epsilon<- abs(rnorm(n*m, munoise, sdnoise))

res <- generate.demand(beta0_prime=log(30),beta1=-1,p_united=c(1,0.9, 0.8, 0.7, 0.7,
 0.8 ,0.9, 1), n=52, m=6, munoise=0, sdnoise=0.2, trendpatt=trendpattern(n*m, 0.2),
dayfact=c(0.7, 0.6, 0.9, 0.8, 1.3, 1.7), epsilon=abs(rnorm(n*m, munoise, sdnoise)))
```

In the case where lognormally distributed noise is directly added to the log mean demand without any exponentiation, the important point is to find the correct mean and standard deviation value in order to get the lognormal random noise with expectation 1. Therefore, the following equations are used:

$$\mu_{LN} = e^{\mu+\sigma^2/2}\sigma_{LN} = (e^{\sigma^2} - 1)e^{2\mu+\sigma^2} \tag{5.7}$$

In order to make the expectation of the lognormal distribution equal to 1 and $\sigma_{LN} = 0.1$, the parameters in the above equations have to be $\mu = -0.048$ and $\sigma = 0.31$. If $\sigma_{LN} = 0.2$, then it can be calculated that $\mu = -0.091$ and $\sigma = 0.43$.

5.4. Simulation of Demand Under Other Price Influences

5.4.1. Simulation of Demand Under Competition

As we discussed in Section 4.3.1, when we add the competition effect to the log mean demand function, we obtain the formulation:

$$log(\mu_{demand}) = \beta_0' + \beta_1 log(p) + \beta_2 log(p_c) \tag{5.8}$$

Here β_2 is a constant like β_1, but the important thing to notice is that β_1 and β_2 should have opposite signs due to the nature of competition. If the prices of the competitor product decrease, the demand of our product decrease too ($\beta_2 > 0$). This functional relationship is considered during the generation of this mean function.

Here only one competitor existence is assumed. As expected more than one competitor can be added to the model with their price parameters.

$$log(\mu_{demand}) = \beta_0' + \beta_1 log(p) + \beta_2 log(p_c) \tag{5.9}$$

In the case where competition is taken into consideration, the only difference in the demand data simulation code is the mean demand function. Since, a new variable, namely the price vector of the competitor, is added to the model, it should be defined in the code as well. However, the main logic of the simulation code remains unchanged.

In real world data it is not likely that the sales prices of competitors are available. However, it is possible to make an assumption about the effect of price actions of the competitor. Here, we define a variable "(1-rs*promday_comp)", which considers the impact of the sales price of the competitor and this variable is added to the mean demand function. Here, "rs" is the indicator of the assumption of the demand amount percentage reduction of the product demand when the competitor makes promotion. The factor "promday_comp" indicates if there is a promotion on the competitor side or not, so its values are 0 or 1. Moreover it is considered that, the number of weeks that one of the promotion prices is applied, is limited and set as a variable in the simulation function. This number can be defined by the user when running the simulation code. In the experiment this value is set to 3, so in total when there are 3 promotion price levels, we have 3X3, namely 9 weeks of promotion in the whole year. The choice of promotion weeks are determined randomly. After that, different price levels of promotions are assigned to these randomly chosen weeks, so at the end a price timing sequence is obtained. Again at this point only the price impact on the mean demand function is taken into consideration. In order to explore the effect of trend, various day factors and random effects, random noise; trend factor and day factors are added to the mean demand function one by one and then the total effect is examined.

```
generate.demand<- function(beta0_prime,beta1, rs, p_vec, n, m,d=3,munoise, sdnoise,
trendpatt , dayfact,epsilon){
```

```
# demand function: demand(p,) = exp(beta0+ beta1*log(p1)) * (1-rs*promday_comp),
#where p1 is the price of "our" product, prom2 the promotion indicator of the
#competitior
# c,b ... parameters of demand function
# rs ... assumption of the reduction (percentage) of "our" sales when the
#competitor make promotion
# p_vec ... different prices (in the moment length four)
# n ... number of weeks for simulation
# m ... number of days per week
# d ... number of week we have one of the promotion prices
#       in the moment we have a total of 3*d weeks with promotion
demand <- rep(0, n*m)
day <- 1+(0:(n*m-1))%%m
promweek <- rep(1, n)
index1 <- sample.int(n, 3*d) # index1 now holds numbers of all the
#"promotion weeks"
promweek[index1[1:d]] <- 2  # we have set 2 to the promweek values for
#the first d promotion week numbers
promweek[index1[(d+1):(2*d)]] <- 3  # we have set 3 to the promweek values for
#the second d promotion week numbers
promweek[index1[(2*d+1):(3*d)]] <- 4  # we have set 4 to the promweek values for
#the third d promotion week numbers

makedayvec <- function(wv, m) {
# returns the day vector for a given week vector
# wv ... week vector
# m ... number of days per week

as.vector(t(matrix(rep(wv,m),nrow=length(wv))))
}

# now the promweek vector holds the pricelevel for each week
promday <- makedayvec(promweek,m)

p <- rep(1, n*m)
p[promday == 2] <- p_vec[2]    # 4 price levels are assigned to the promday
#in order to get the price vector
```

```
p[promday == 3] <- p_vec[3]
p[promday == 4] <- p_vec[4]

prom1 <- rep(1, n*m)
prom1[promday == 1] <- 0   #sets 1 to all promotion days,
#and 0 to non-promotion days in "our" price vector

promweek_comp <- rep(0, n)
index2 <- sample.int(n, 3*d) # index2 now holds numbers of all the
#"promotion weeks" of the competitor
promweek_comp[index2[1:(3*d)]] <- 1  # we have set 1 to the promweek
#values for the competitor's promotion week numbers
promday_comp <- makedayvec(promweek_comp,m)   #contains only zeros and ones.
logmu_demand <- beta0+ beta1*log(p)

demand<- rpois(n*m, exp(logmu_demand + log(1 - rs*promday_comp)
+ log(trendpatt) + log(dayfact[day]) +  epsilon))
#let's assume promotion of the competitor reduces of our sales by 10%.
data.frame(demand,p,promday, prom1, promday_comp, t=1:length(p), day)
}

res <- generate.demand(beta0_prime=log(30),beta1=-1, rs= 0.1,
 p_vec=c(1, 0.9, 0.8, 0.7), n=52, m=6,
 d=3,munoise=0, sdnoise=0.2, trendpatt=trendpattern(n*m, 0.2),
 dayfact=c(0.7, 0.6, 0.9, 0.8, 1.3, 1.7), epsilon=abs(rnorm(n*m, munoise, sdnoise)))
```

5.4.2. Simulation of Demand Under Substitution

In Section 4.3.2, we have discussed the log mean of the demand function under substitution effect and got the equation below:

$$log(Q) = \beta_0' + \beta_1 log(p) + \beta_2 log(p_c) + \beta_3 log(p_s) \tag{5.10}$$

There are some important points to mention here about this equation. In this function β_3 is another constant which is given as default in the code. Although β_3 has the same sign as β_2, it is important to note that β_2 should be greater than β_3, because the competition effect is more drastic than the substitution effect. The rest of the demand generation is the same in the previous cases. So, we can use the same code in Section 5.4.1, with only changing the mean of the demand function and adding the corresponding parameters to the code.

5.4.3. Simulation of Demand Under Complementary Goods Effect

Referring to the Section 4.3.3, we use the log mean of the demand function below, in order to add the complement effect in the demand model, which in fact does not bring a big difference:

$$log(Q) = \beta_0' + \beta_1 log(p) + \beta_2 log(p_c) + \beta_3 log(p_s) + \beta_4 log(p_{cg}) \tag{5.11}$$

The important thing to notice here is the sign of the coefficient of the complementary product. It should be the same as β_1 due to negative cross elasticity of demand. As a result, the signs of the coefficients of competitor and substitute products are opposite to the coefficient of the complement (β_4). As in the previous subsection, by just changing the mean demand function in the code and adding the related parameters, we can use the same simulation as in Section 5.4.1.

6. DATA ANALYSIS

Due to the fact that data on hand are not useful for our experiments, demand data are generated, as it is explained in Chapter 5. The available data can differ in number of observations or mean of the demand. Therefore, we generated and stored two different data sets and used them in order to understand the impacts on the demand, estimate the demand model parameters the data, forecast the mean of the demand and examine the differences in the results of statistical analysis' of different data sets. The commands used in the generation process of two data sets are given below. We called them data set A and B, respectively. With the help of the "set.seed" command in R we can generate and used the same data again.

Data Set A:

```
set.seed(334217)

trendpattern <- function(n,maxChange){
# outputs a trend factor vector of length n with a linear increasing trend.
# the average of the result vector is 1
# the first value in the result vector is 1-maxChange
# the last value in the result vector is 1+maxChange

1+maxChange*(1:n-(n+1)/2)/((n-1)/2)
}

generate.demand<- function(beta0_prime,beta1, rs, p_vec, n, m,d=3,munoise, sdnoise,
trendpatt , dayfact,epsilon){
# demand function: demand(p,) = exp(beta0+ beta1*log(p1)) * (1-rs*promday_comp),
#where p1 is the price of "our" product, prom2 the promotion indicator of the
#competitior
# c,b ... parameters of demand function
# rs ... assumption of the reduction (percentage) of "our" sales when the
#competitor make promotion
# p_vec ... different prices (in the moment length four)
```

```
# n ... number of weeks for simulation
# m ... number of days per week
# d ... number of week we have one of the promotion prices
#       in the moment we have a total of 3*d weeks with promotion
demand <- rep(0, n*m)
day <- 1+(0:(n*m-1))%%m
promweek <- rep(1, n)
index1 <- sample.int(n, 3*d) # index1 now holds numbers of all the
#"promotion weeks"
promweek[index1[1:d]] <- 2  # we have set 2 to the promweek values for
#the first d promotion week numbers
promweek[index1[(d+1):(2*d)]] <- 3  # we have set 3 to the promweek values for
#the second d promotion week numbers
promweek[index1[(2*d+1):(3*d)]] <- 4  # we have set 4 to the promweek values for
#the third d promotion week numbers

makedayvec <- function(wv, m) {
# returns the day vector for a given week vector
# wv ... week vector
# m ... number of days per week

as.vector(t(matrix(rep(wv,m),nrow=length(wv))))
}

# now the promweek vector holds the pricelevel for each week
promday <- makedayvec(promweek,m)

p <- rep(1, n*m)
p[promday == 2] <- p_vec[2]    # 4 price levels are assigned to the promday
#in order to get the price vector
p[promday == 3] <- p_vec[3]
p[promday == 4] <- p_vec[4]

prom1 <- rep(1, n*m)
prom1[promday == 1] <- 0   #sets 1 to all promotion days,
#and 0 to non-promotion days in "our" price vector
```

```
promweek_comp <- rep(0, n)
index2 <- sample.int(n, 3*d) # index2 now holds numbers of all the
#"promotion weeks" of the competitor
promweek_comp[index2[1:(3*d)]] <- 1  # we have set 1 to the promweek
#values for the competitor's promotion week numbers
promday_comp <- makedayvec(promweek_comp,m)   #contains only zeros and ones.
logmu_demand <- beta0_prime + beta1*log(p)

demand<- rpois(n*m, exp(logmu_demand + log(1 - rs*promday_comp)
+ log(trendpatt) + log(dayfact[day]) +  epsilon))
#let's assume promotion of the competitor reduces of our sales by 10%.
data.frame(demand,p,promday, prom1, promday_comp, t=1:length(p), day)
}
m=6
n=52
beta0_prime=log(30)
beta1=-1
p_vec=c(1, 0.9, 0.8, 0.7)
trendpatt <- trendpattern(n*m, 0.2)
dayfact = c(0.7, 0.6, 0.9, 0.8, 1.3, 1.7)
munoise=0
sdnoise=0.2
epsilon<- abs(rnorm(n*m, munoise, sdnoise))

res <- generate.demand(beta0_prime=log(30),beta1=-1, rs= 0.1,
p_vec=c(1, 0.9, 0.8, 0.7), n=52, m=6, d=3,munoise=0, sdnoise=0.2,
trendpatt=trendpattern(n*m, 0.2),
dayfact=c(0.7, 0.6, 0.9, 0.8, 1.3, 1.7),
epsilon=abs(rnorm(n*m, munoise, sdnoise)))
```

Data Set B:

```
set.seed(334219)

trendpattern <- function(n,maxChange){
# outputs a trend factor vector of length n with a linear increasing trend.
# the average of the result vector is 1
```

```
# the first value in the result vector is 1-maxChange
# the last value in the result vector is 1+maxChange
1+maxChange*(1:n-(n+1)/2)/((n-1)/2)
}
generate.demand<- function(beta0_prime,beta1, rs, p_vec, n, m,d=1,munoise, sdnoise,
trendpatt , dayfact,epsilon){
# demand function: demand(p,) = exp(beta0+ beta1*log(p1)) * (1-rs*promday_comp),
#where p1 is the price of "our" product, prom2 the promotion indicator of the
#competitior
# c,b ... parameters of demand function
# rs ... assumption of the reduction (percentage) of "our" sales when the
#competitor make promotion
# p_vec ... different prices (in the moment length four)
# n ... number of weeks for simulation
# m ... number of days per week
# d ... number of week we have one of the promotion prices
#       in the moment we have a total of 3*d weeks with promotion
demand <- rep(0, n*m)
day <- 1+(0:(n*m-1))%%m
promweek <- rep(1, n)
index1 <- sample.int(n, 3*d) # index1 now holds numbers of all the
#"promotion weeks"
promweek[index1[1:d]] <- 2  # we have set 2 to the promweek values for
#the first d promotion week numbers
promweek[index1[(d+1):(2*d)]] <- 3  # we have set 3 to the promweek values for
#the second d promotion week numbers
promweek[index1[(2*d+1):(3*d)]] <- 4  # we have set 4 to the promweek values for
#the third d promotion week numbers

makedayvec <- function(wv, m) {
# returns the day vector for a given week vector
# wv ... week vector
# m ... number of days per week

as.vector(t(matrix(rep(wv,m),nrow=length(wv))))
}
```

```
# now the promweek vector holds the pricelevel for each week
promday <- makedayvec(promweek,m)

p <- rep(1, n*m)
p[promday == 2] <- p_vec[2]    # 4 price levels are assigned to the promday
#in order to get the price vector
p[promday == 3] <- p_vec[3]
p[promday == 4] <- p_vec[4]

prom1 <- rep(1, n*m)
prom1[promday == 1] <- 0   #sets 1 to all promotion days,
#and 0 to non-promotion days in "our" price vector

promweek_comp <- rep(0, n)
index2 <- sample.int(n, 3*d) # index2 now holds numbers of all the
#"promotion weeks" of the competitor
promweek_comp[index2[1:(3*d)]] <- 1  # we have set 1 to the promweek
#values for the competitor's promotion week numbers
promday_comp <- makedayvec(promweek_comp,m)   #contains only zeros and ones.
logmu_demand <- beta0_prime + beta1*log(p)

demand<- rpois(n*m, exp(logmu_demand + log(1 - rs*promday_comp)
+ log(trendpatt) + log(dayfact[day]) +  epsilon))
#let's assume promotion of the competitor reduces of our sales by 10%.
data.frame(demand,p,promday, prom1, promday_comp, t=1:length(p), day)
}
m=6
n=13
beta0_prime=log(30)
beta1=-1
p_vec=c(1, 0.9, 0.8, 0.7)
trendpatt <- trendpattern(n*m, 0.2)
dayfact = c(0.7, 0.6, 0.9, 0.8, 1.3, 1.7)
munoise=0
sdnoise=0.2
epsilon<- abs(rnorm(n*m, munoise, sdnoise))
```

```
res <- generate.demand(beta0_prime=log(30),beta1=-1, rs= 0.1,
p_vec=c(1, 0.9, 0.8, 0.7), n=13, m=6, d=1,munoise=0, sdnoise=0.2,
trendpatt=trendpattern(n*m, 0.2), dayfact=c(0.7, 0.6, 0.9, 0.8, 1.3, 1.7),
epsilon=abs(rnorm(n*m, munoise, sdnoise)))
```

In the table below, the characteristics of these data sets are summarized.

Table 6.1. Characteristics of Data Set A and B.

	No.of Weeks	**No.of Weeks w. prom.**	**Daily mean demand**
Data Set A	52	9	30
Data Set B	13	3	30

In this data analysis and understanding process, firstly we plot the data and try to display a graphical representation, in order to understand and interpret the demand data better. Plots play an important role in statistics. The right graphs, plots or visual displays of a dataset can uncover anomalies or provide insights that go beyond what most quantitative techniques are capable of discovering. Therefore, we find it important to start with plotting graphs of the available data sets before we estimate the model parameters and make tests.

After the graphical representation of the demand data is done, we are ready to start to estimate the parameters of the mean demand model, in order to quantify the effects of promotions on demand. Different statistical techniques are used in this process, and according to the used technique, the interpretations of the estimated parameters change. In this chapter, the use of the various statistical techniques in our study is explained. In order to understand the impacts of price promotions on the demand and estimate the parameters of the mean demand model, Analysis of Variance (ANOVA), regression and Analysis of Covariance (ANCOVA) with Poisson response are used.

In the constant mean model, which is explained in details in Section 4.2.1, the mean of the demand is constant and promotions occur. So, the data include daily demand and an indicator of promotions. In order to test the significance of promotions the first and easiest technique to apply is ANOVA. The generated data are grouped into two, namely demand with and without promotions, and accordingly the ANOVA-test compares the means of the two groups and shows the significance level of promotions. For ANOVA-test, the price information is not necessary.

Although, ANOVA results give some insights into the importance of promotions, this method is not enough to examine and quantify the effects of price promotions and other covariates on demand. The observations, namely the daily demands are sampled from independent random variables, but it does not follow normal distribution. It is count data, and we assume that demand is Poisson distributed in our study. Moreover, the variance of the response variable is not constant; in this case it is a function of the mean. In addition, the normality assumption of the error term is violated. Generalized linear models can fit any non-linear relationship that has a linearizing transformation, with the help of link functions. Due to these reasons, GLM is the right tool for forecasting the demand, estimating the model parameters and quantifying the effects of promotions on the demand. Since, our dependent variable in GLM model, namely mean of the demand, is Poisson distributed, we select a GLM of the Poisson family, in other words we use Poisson regression. Detailed explanations about GLM and Poisson regression were given in Section 3.4.

In order to use Poisson regression, we have to decide which link function to use. Due to the nature of the mean of the demand function mentioned in Chapter 4, the log link function is used.

With the help of the GLM, different influences on demand, which are discussed in Chapter 4 can be examined and the parameters of the mean demand model can be estimated. The dependent variable is always demand, as it is in the ANOVA case. The independent variables, namely covariates, used in the GLM model can differ according to the nature of the data on hand. In our case, demand and time influences are the covariates. These covariates. When the price, trend factor and the week day factor are added to the model, we are able to estimate their parameters. However, since week day is a nominal variable, this covariate should be added to the model as a factor. In this full model, while price and trend are interval scale covariate, different week days variable is a nominal variable. Consequently, the new model includes covariates both of nominal and interval scale; hence ANCOVA with Poisson response could be the name of the statistical technique used for this model.

6.1. Statistical Analysis of Data Set A

6.1.1. Plotting

In Figure 6.1, the daily demand data of data set A and its corresponding price levels are illustrated. The data contain 312 days and 4 different price levels. In the full year data, we can see that there are 9 weeks with promotions in total and the mean of the demand value is about 30. The relationship between demand and promotions is obvious; in the days with low price, higher demand values are observed. This relationship is shown in a strip chart in Figure 6.2, which illustrates the daily demand values at different price levels. It is shown that the most of the demand values at price level 1 is around 30. In Figure 6.3, we show the mean of the demand A values for different price levels. According to the price values the mean of the demand varies. For lower price levels we observe higher mean demand values and at high price levels lower mean demand values. Looking at the histogram at Figure 6.4, we can see that the distribution of demand is similar to the Poisson distribution, because the distribution of all daily demands is a mixture of Poisson distribution with different mean values.

Box plots can help us to visualize the difference between mean and variance of the demand values at different price levels or in different week days. In Figure 6.5, a box plot of the daily demand A at different price levels is shown. From this plot we can directly realize that the mean and the variance of the demand with low price are high

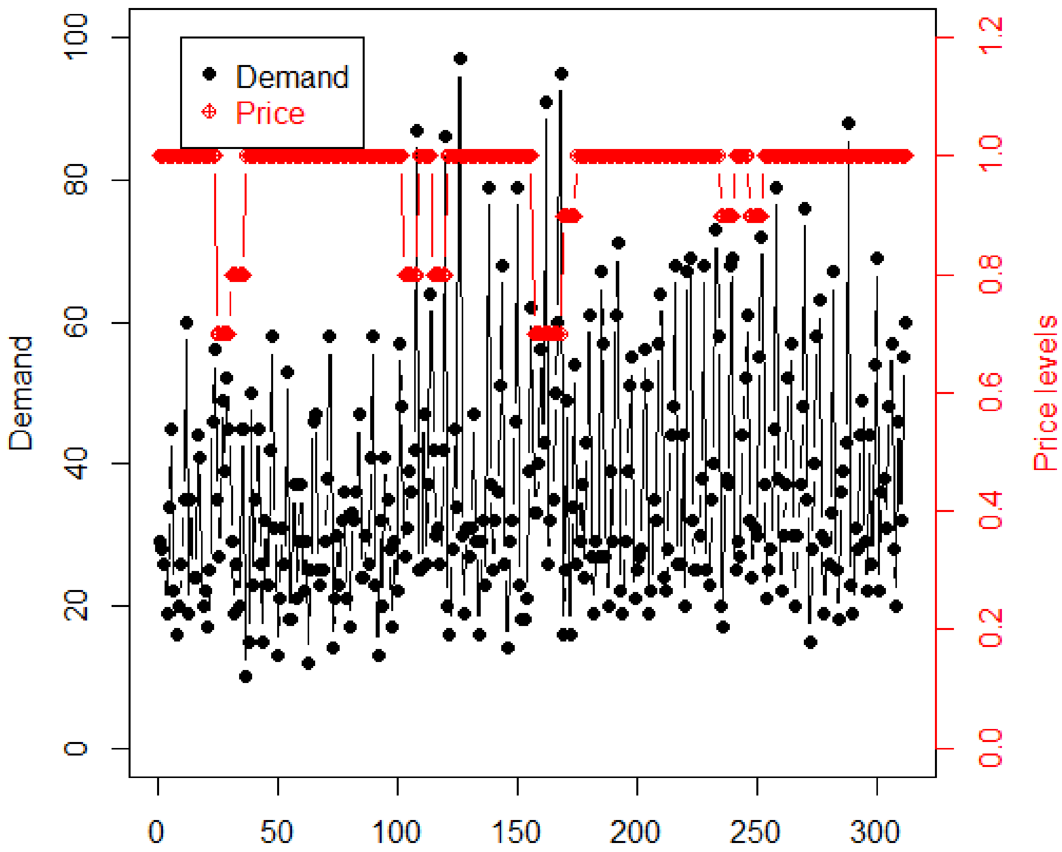

Figure 6.1. Data Set A: Demand vs. Price Levels.

compared to the situation at high price level. In Figure 6.6, we display the differences between means and variances of the demand on various week days. We assume that the highest demand is observed on Saturday, and the highest mean of the demand is on Day 6, which is Saturday.

Since we have different week days and different price levels, we can plot an interaction plot in order to see the influences of the week days to the mean of the demand. This graph helps us to understand the impacts of different week days on demand and makes the interpretation during the parameter estimation process easier. In the Figure 6.7, the interaction plot of the mean of demand A according to different days of the week is shown. Means of the demands in different week days develop almost parallel lines and every line has a decreasing pattern when the price increases. Only the lines of Monday, Wednesday and Thursday are very close to each other, but this does not

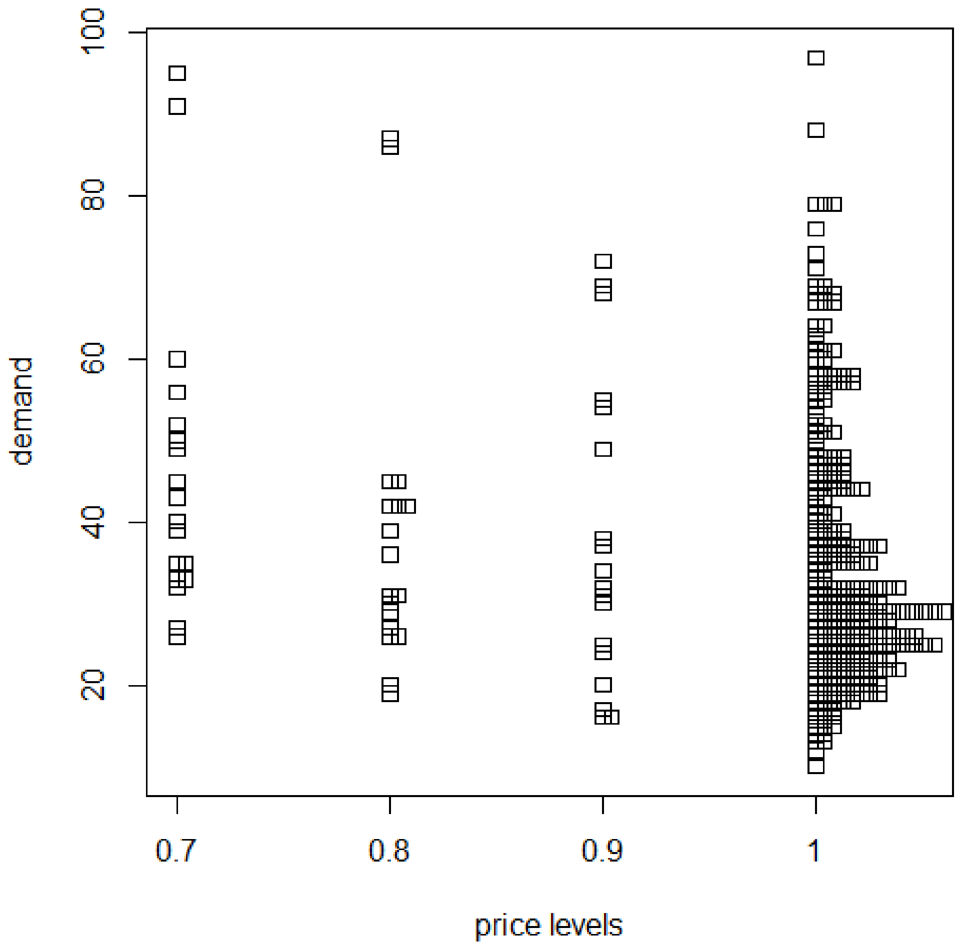

Figure 6.2. Data Set A: Daily Demand vs. Price Levels.

show that there is interaction. Due to the parallel nature of lines we can distinguish that our no interaction assumption between week days is not incorrect.

6.1.2. ANOVA-Test

As we stated in Section 4.2.2, demand can be modelled in a more sophisticated way with the help of the deterministic mean demand models. ANOVA is performed in this type of demand model as well, in order to assess the significance of price promotions. therefore in this case, the data matrix contains only demand data and indicator of promotions. In other words a vector that holds only “0”s and “1”s, can explain if there is a price promotion on that day. 0 means there is no price promotion, while 1 means there is price promotion. As it is known, ANOVA compares the groups, namely

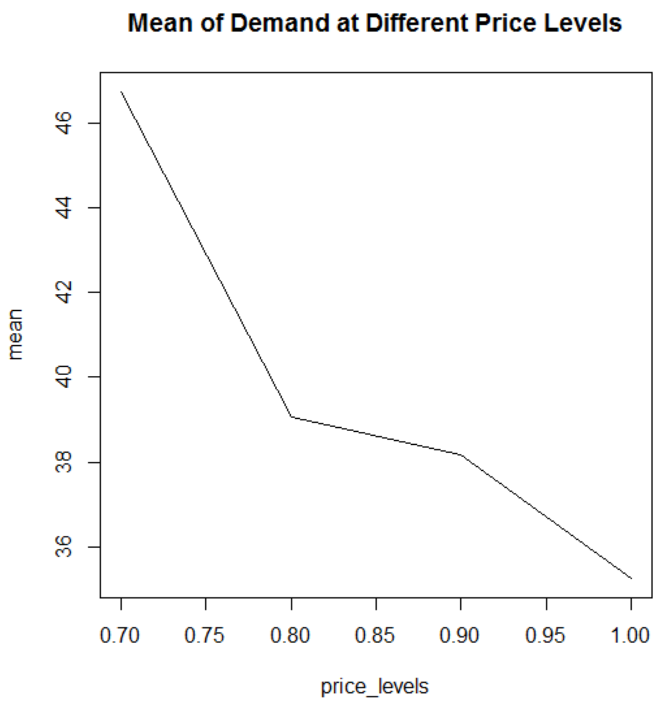

Figure 6.3. Mean of Demand A at Different Price Level.s

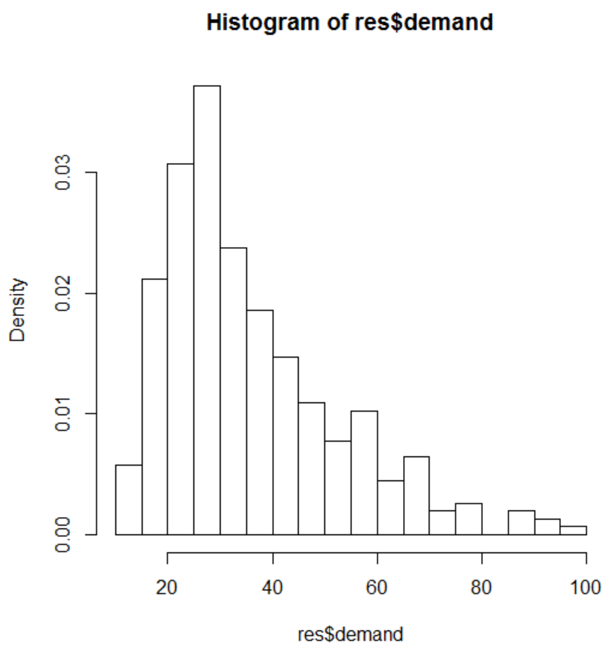

Figure 6.4. Histogram of Poisson Distributed Demand A.

nominal variables, so if there is a price promotion, corresponding demand data belong to the second group, if there is no price promotion, in other words, if the sales price is equal to the standard price, then the corresponding demand data are from the first group. By doing this, the nominal covariates of the ANOVA model have been created.

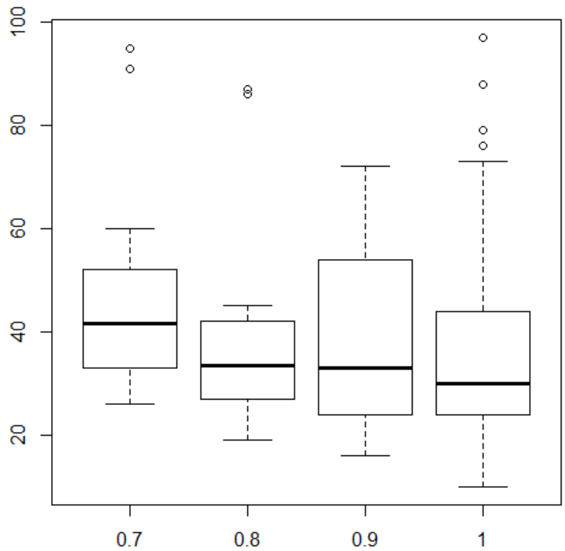

Figure 6.5. Box plot of daily demands A at various price levels.

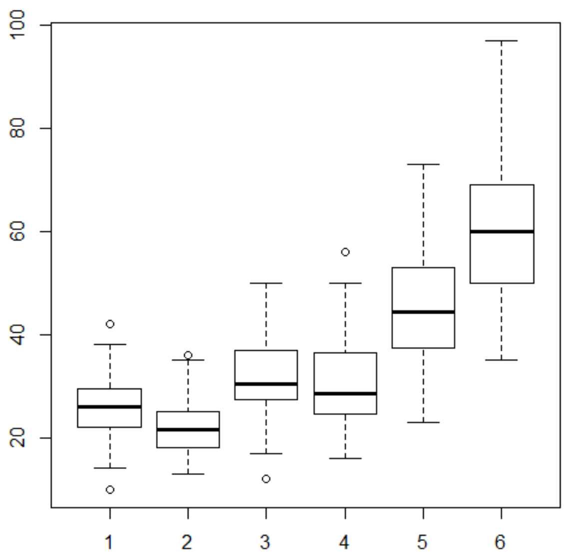

Figure 6.6. Box plot of daily demands A at various week days.

It is important to note here that, using t-test in order to compare them is enough, because we have only two groups. Since ANOVA is a general name of this method, in the thesis we mention ANOVA instead of t-test. Here, the data described and simulated in Section 5.4.1, are used, which are generated under the assumption that the price of the competitor product cannot be known exactly but the effect of price actions of the competitor on our product can be assumed. In the data matrix, there is a column called "prom1", which shows the promotion situation. Using the data set A, we apply ANOVA in order to test the significance of price promotions.

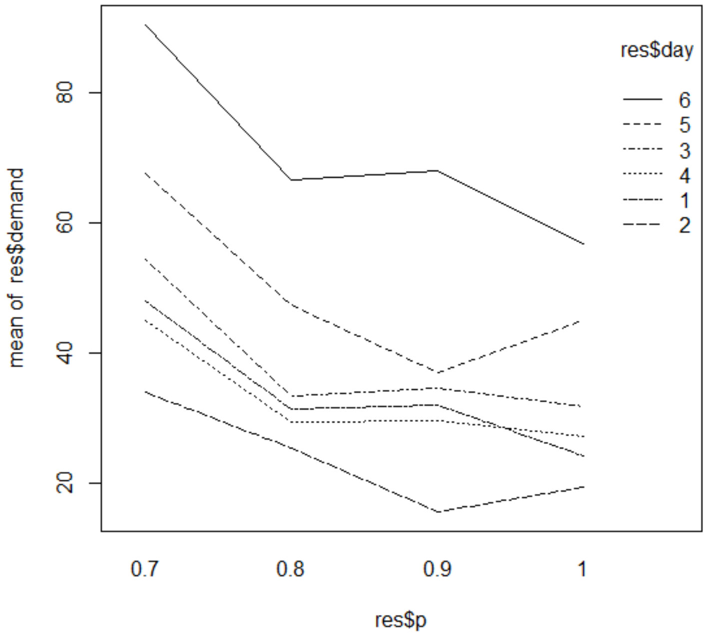

Figure 6.7. Mean of the Demand A in Different Week Days.

This model is fixed when promotion (yes - no) effect is investigated in the experiments, no matter which other effects are considered in simulating the mean of the Poisson distributed demand data. To analyse the data we use ANOVA with Poisson responses.

Here are the commands to make ANOVA-test:

```
myobj<- glm(res$demand ~ as.factor(res$prom1), family=poisson(link="log"))
summary(myobj)
```

Here are the results of ANOVA with Poisson response:

```
Call:
glm(formula = res$demand ~ as.factor(res$prom1), family = poisson(link = "log"))

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-5.0332  -2.0154  -0.9112   1.4146   8.5323
```

```
Coefficients:
                      Estimate Std. Error z value Pr(>|z|)
(Intercept)            3.56307    0.01048 339.897  < 2e-16 ***
as.factor(res$prom1)1  0.15815    0.02362   6.694 2.17e-11 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 2173.3  on 311  degrees of freedom
Residual deviance: 2129.9  on 310  degrees of freedom
AIC: 3800.2

Number of Fisher Scoring iterations: 4
```

It is obvious from the results that, the difference between the mean values of the two groups, namely observed demand with and without promotion is significant.

The same analysis is performed with “identity” link function as well and here are the results:

```
Call:
glm(formula = res$demand ~ as.factor(res$prom1), family = poisson(link = "identity"))

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-5.0332  -2.0154  -0.9112   1.4146   8.5323

Coefficients:
                      Estimate Std. Error z value Pr(>|z|)
(Intercept)            35.2713     0.3697  95.394  < 2e-16 ***
as.factor(res$prom1)1   6.0435     0.9496   6.364 1.97e-10 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 2173.3  on 311  degrees of freedom
Residual deviance: 2129.9  on 310  degrees of freedom
AIC: 3800.2

Number of Fisher Scoring iterations: 3
```

As we used "log" as link function first, the expectation for the default group is $exp(3.5631)$ $= 35.2713$. The result for the demand with promotions shows that the mean response of this group is $exp(0.1582) = 1.1713$ times larger than the default group. Thus we get for that group as mean response: $exp(3.5162) * exp(0.3798) = 41.3180$. This is equal to $35.2713 + 6.0435$, which is obtained from the model with identity link function. So, we can state that, the choice of the link function influences only the way we have to interpret the results, it does not change the estimates of the mean demands. The same AIC values in both of the results are another proof of this argument.

6.1.3. Poisson Regression

Poisson regression helps us to forecast the demand with and without promotions and estimates the demand model parameters. The full Poisson regression equation with all of the covariates is as follows:

```
glm(res$demand ~ res$p + res$t + as.factor(res$day) ,
family=poisson(link=log))
```

We can use this model for all kind of demand data. First we want to demonstrate the results of Poisson regression for the data set A.

```
Call:
glm(formula = res$demand ~ res$p + res$t + as.factor(res$day),
    family = poisson(link = log))
```

```
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.3704  -0.9415  -0.1677   0.7709   4.7871

Coefficients:
                     Estimate Std. Error z value Pr(>|z|)
(Intercept)          4.065062   0.104410  38.934  < 2e-16 ***
res$p               -1.013661   0.106975  -9.476  < 2e-16 ***
res$t                0.001074   0.000107  10.031  < 2e-16 ***
as.factor(res$day)2 -0.165377   0.040127  -4.121 3.77e-05 ***
as.factor(res$day)3  0.203550   0.036617   5.559 2.72e-08 ***
as.factor(res$day)4  0.151247   0.037047   4.083 4.46e-05 ***
as.factor(res$day)5  0.559221   0.034058  16.419  < 2e-16 ***
as.factor(res$day)6  0.859201   0.032415  26.506  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 2173.33  on 311  degrees of freedom
Residual deviance:  543.21  on 304  degrees of freedom
AIC: 2224.3

Number of Fisher Scoring iterations: 4
```

The Poisson regression results show that β_0' is estimated as 4.065062 and β_1 is estimated as -1.013661. The standard errors of both estimates are around 0.1. Since the coefficient of price covariate, β_1, has a very small p-value, we can state that it is significant. So we observe that price has an influence on demand. With the help of the coefficient estimates of the mean demand model, we can estimate the mean of the demand on a certain week day and at a certain price level. For example, we can calculate the mean response for a Wednesday in the middle of the year and for price

0.8 using

$$exp(4.065062 - 1.013661 * 0.8 + 0.001074 * 156 + 0.203550) = 37.53 \tag{6.1}$$

This means that for Wednesday, p=0.8 and t=156, the response variable follows a Poisson distribution with estimated mean value $\mu_{hat} = 37.53$.

If we want to estimate the mean demand without promotions on Wednesday in the middle of the year, we have to take p=1. In this case our estimated mean value is $\mu_{hat} = 30.65$.

It is shown that, the mean of the demand value differs due to the existence of promotions. We calculate a higher mean demand when the price is low. It is obvious that, when the price decreases by 20% the demand increases by 18%.

In order to test if the link function correctly describes the relationship in our data, we can also try to use the identity link instead of the log-link. In this case we obtain the results below:

```
Call:
glm(formula = res$demand ~ res$p + res$t + as.factor(res$day),
    family = poisson(link = identity))

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.4530  -0.9391  -0.1601   0.7144   4.4825

Coefficients:
                      Estimate Std. Error z value Pr(>|z|)
(Intercept)          54.927482   4.154926  13.220  < 2e-16 ***
res$p               -34.708854   4.239660  -8.187 2.68e-16 ***
res$t                 0.030786   0.003556   8.656  < 2e-16 ***
as.factor(res$day)2  -3.809289   0.955253  -3.988 6.67e-05 ***
as.factor(res$day)3   5.770435   1.049245   5.500 3.81e-08 ***
```

```
as.factor(res$day)4   4.016247   1.033134   3.887 0.000101 ***
as.factor(res$day)5  19.314654   1.168631  16.528  < 2e-16 ***
as.factor(res$day)6  35.182880   1.293668  27.196  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 2173.33  on 311  degrees of freedom
Residual deviance:  571.53  on 304  degrees of freedom
AIC: 2253.8

Number of Fisher Scoring iterations: 5
```

We can calculate again the mean response for a Wednesday in the middle of the year and for price 0.8 using

$$54.927482 - 34.708854 * 0.8 + 0.030786 * 156 + 5.770435) = 37.73 \tag{6.2}$$

We obtain almost the same result as for the log-link function. However, since we use "identity" link-function we do not exponentiate the last calculation.

In order to compare the two link-functions, we need to check their AIC-Values. The AIC-Value of the model with "log-link" is smaller than the model with "identity-link". So we can conclude that "log-link" presents the models better. This can be also considered as a test for the demand function, Therefore, we continue to use "log-link" in the rest of our study.

6.1.4. Checking No Interaction Assumption Between Covariates

An interaction in the context of regression analysis may arise when considering the relationship among three or more variables, and describes a situation in which the simultaneous influence of two covariates on the response variable is not additive. In the models, which are used to estimate the mean of demand, it is assumed that there is no interaction between the covariates, however testing this model assumption is necessary. Therefore, different models with different kinds of interactions are created and with the help of Akaike's Information Criterion (AIC) the best generalized linear model is tried to be found. The best model is the model with the smallest AIC-value. In this context, models with and without interactions are created and their AIC-values are compared with each other. In total we obtained 5 different models. Model 1, is the model without any interaction between covariates. There is an interaction between trend and weekday covariates in Model 2. In Model 3 we created interaction between price and trend. Price and week day are interacting with each other in Model 4. In the last model all covariates of the model are interacting with each other. These models and their AIC-values are given in the table below.

Table 6.2. AIC-Values of different models.

		AIC-Value
MODEL 1	$demand \sim p + t + as.factor(day)$	608.61
MODEL 2	$demand \sim p + t * as.factor(day)$	616.42
MODEL 3	$demand \sim p * t + as.factor(day)$	609.45
MODEL 4	$demand \sim p * as.factor(day) + t$	614.12
MODEL 5	$demand \sim p * t * as.factor(day)$	618.16

As it is seen, AIC-values are not really different from each other, so no interaction assumption is not wrong. In addition, the model without any interaction, Model 1, has the smallest AIC- value, which indicates that it is the best model. Moreover, the estimates for the interaction parameters are all small and have large p-values, which supports the correctness of our no interaction assumption. The results of the described five models are given below. They are obtained by using Data Set A.

```
Call:
glm(formula = res$demand ~ res$p + res$t + as.factor(res$day),
    family = poisson(link = log))

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.3704  -0.9415  -0.1677   0.7709   4.7871

Coefficients:
                     Estimate Std. Error z value Pr(>|z|)
(Intercept)          4.065062   0.104410  38.934  < 2e-16 ***
res$p               -1.013661   0.106975  -9.476  < 2e-16 ***
res$t                0.001074   0.000107  10.031  < 2e-16 ***
as.factor(res$day)2 -0.165377   0.040127  -4.121 3.77e-05 ***
as.factor(res$day)3  0.203550   0.036617   5.559 2.72e-08 ***
as.factor(res$day)4  0.151247   0.037047   4.083 4.46e-05 ***
as.factor(res$day)5  0.559221   0.034058  16.419  < 2e-16 ***
as.factor(res$day)6  0.859201   0.032415  26.506  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 2173.33  on 311  degrees of freedom
Residual deviance:  543.21  on 304  degrees of freedom
AIC: 2224.3

Number of Fisher Scoring iterations: 4

Call:
glm(formula = res$demand ~ res$p + res$t * as.factor(res$day),
    family = poisson(link = log))

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.4869  -0.9073  -0.1135   0.7411   4.8464
```

```
Coefficients:
                            Estimate Std. Error z value Pr(>|z|)
(Intercept)                4.107e+00  1.139e-01  36.062  < 2e-16 ***
res$p                     -1.014e+00  1.070e-01  -9.482  < 2e-16 ***
res$t                      8.188e-04  3.046e-04   2.688  0.00718 **
as.factor(res$day)2       -8.144e-02  8.123e-02  -1.003  0.31608
as.factor(res$day)3        2.034e-01  7.510e-02   2.708  0.00678 **
as.factor(res$day)4        8.211e-02  7.679e-02   1.069  0.28489
as.factor(res$day)5        4.594e-01  7.070e-02   6.498 8.15e-11 ***
as.factor(res$day)6        7.895e-01  6.702e-02  11.779  < 2e-16 ***
res$t:as.factor(res$day)2 -5.291e-04  4.483e-04  -1.180  0.23792
res$t:as.factor(res$day)3  4.336e-06  4.093e-04   0.011  0.99155
res$t:as.factor(res$day)4  4.262e-04  4.145e-04   1.028  0.30376
res$t:as.factor(res$day)5  6.088e-04  3.811e-04   1.597  0.11018
res$t:as.factor(res$day)6  4.278e-04  3.625e-04   1.180  0.23790
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 2173.33  on 311  degrees of freedom
Residual deviance:  532.51  on 299  degrees of freedom
AIC: 2224.8

Number of Fisher Scoring iterations: 4

Call:
glm(formula = res$demand ~ res$p * res$t + as.factor(res$day),
    family = poisson(link = log))

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.3833  -0.9317  -0.1718   0.7787   4.7803

Coefficients:
                     Estimate Std. Error z value Pr(>|z|)
(Intercept)          3.906082   0.219575  17.789  < 2e-16 ***
```

```
res$p               -0.850007   0.225701  -3.766 0.000166 ***
res$t                0.002373   0.001576   1.506 0.132019
as.factor(res$day)2 -0.165401   0.040127  -4.122 3.76e-05 ***
as.factor(res$day)3  0.203501   0.036617   5.558 2.74e-08 ***
as.factor(res$day)4  0.151173   0.037048   4.081 4.49e-05 ***
as.factor(res$day)5  0.559122   0.034059  16.416  < 2e-16 ***
as.factor(res$day)6  0.859079   0.032415  26.502  < 2e-16 ***
res$p:res$t         -0.001329   0.001608  -0.827 0.408379
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 2173.33  on 311  degrees of freedom
Residual deviance:  542.52  on 303  degrees of freedom
AIC: 2226.8

Number of Fisher Scoring iterations: 4

Call:
glm(formula = res$demand ~ res$p * as.factor(res$day) + res$t,
    family = poisson(link = log))

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.4414  -0.9089  -0.1333   0.8012   4.8077

Coefficients:
                           Estimate Std. Error z value Pr(>|z|)
(Intercept)                4.018891   0.294348  13.654  < 2e-16 ***
res$p                     -0.965565   0.306068  -3.155  0.00161 **
as.factor(res$day)2        0.227377   0.425217   0.535  0.59284
as.factor(res$day)3        0.285246   0.394946   0.722  0.47015
as.factor(res$day)4        0.483568   0.394913   1.224  0.22077
as.factor(res$day)5        0.138131   0.374817   0.369  0.71248
as.factor(res$day)6        0.962951   0.349793   2.753  0.00591 **
res$t                      0.001074   0.000107  10.032  < 2e-16 ***
```

```
res$p:as.factor(res$day)2 -0.409982   0.441932  -0.928  0.35356
res$p:as.factor(res$day)3 -0.085161   0.409910  -0.208  0.83542
res$p:as.factor(res$day)4 -0.346804   0.410245  -0.845  0.39791
res$p:as.factor(res$day)5  0.438025   0.388447   1.128  0.25948
res$p:as.factor(res$day)6 -0.108162   0.363030  -0.298  0.76575
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 2173.33  on 311  degrees of freedom
Residual deviance:  536.42  on 299  degrees of freedom
AIC: 2228.7

Number of Fisher Scoring iterations: 4

Call:
glm(formula = res$demand ~ res$p * as.factor(res$day) * res$t,
    family = poisson(link = log))

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.4439  -0.8967  -0.1192   0.7559   4.8269

Coefficients:
                                 Estimate Std. Error z value Pr(>|z|)
(Intercept)                      4.976083   0.567012   8.776  < 2e-16 ***
res$p                           -1.906309   0.589765  -3.232 0.001228 **
as.factor(res$day)2             -1.079566   0.858308  -1.258 0.208470
as.factor(res$day)3             -0.127370   0.776602  -0.164 0.869724
as.factor(res$day)4             -0.899298   0.814057  -1.105 0.269285
as.factor(res$day)5             -0.586523   0.751905  -0.780 0.435362
as.factor(res$day)6             -1.245410   0.723379  -1.722 0.085132 .
res$t                           -0.007635   0.004406  -1.733 0.083131 .
res$p:as.factor(res$day)2        1.014921   0.890869   1.139 0.254600
res$p:as.factor(res$day)3        0.337306   0.807102   0.418 0.676004
res$p:as.factor(res$day)4        0.991302   0.844846   1.173 0.240654
```

```
res$p:as.factor(res$day)5        1.092447   0.779706   1.401 0.161184
res$p:as.factor(res$day)6        2.084904   0.749647   2.781 0.005416 **
res$p:res$t                      0.008630   0.004498   1.919 0.055035 .
as.factor(res$day)2:res$t        0.010833   0.006492   1.669 0.095186 .
as.factor(res$day)3:res$t        0.003798   0.005944   0.639 0.522829
as.factor(res$day)4:res$t        0.012409   0.006060   2.048 0.040578 *
as.factor(res$day)5:res$t        0.007126   0.005671   1.257 0.208899
as.factor(res$day)6:res$t        0.018663   0.005380   3.469 0.000523 ***
res$p:as.factor(res$day)2:res$t -0.011568   0.006631  -1.744 0.081077 .
res$p:as.factor(res$day)3:res$t -0.003864   0.006069  -0.637 0.524336
res$p:as.factor(res$day)4:res$t -0.012175   0.006188  -1.968 0.049110 *
res$p:as.factor(res$day)5:res$t -0.006735   0.005784  -1.164 0.244299
res$p:as.factor(res$day)6:res$t -0.018612   0.005490  -3.390 0.000699 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 2173.33  on 311  degrees of freedom
Residual deviance:  510.23  on 288  degrees of freedom
AIC: 2224.5

Number of Fisher Scoring iterations: 4
```

6.2. Statistical Analysis of Data Set B

6.2.1. Plotting

The graphical representation of data procedure is applied for our second data set as well. In Data Set B we have smaller sample size. Therefore, distinguishing the patterns from the graphs is easier. In Figure 6.8, demand data B and corresponding price levels are illustrated. The data contain 78 days and 4 different price levels. In the 3 months of demand data, we assumed that there are 3 weeks with promotions in total and the mean of the demand value is about 30. The relationship between demand and promotions is clear. This relationship is shown in a strip chart in Figure 6.9. We observe 6 different demand values at each of the promotion price.

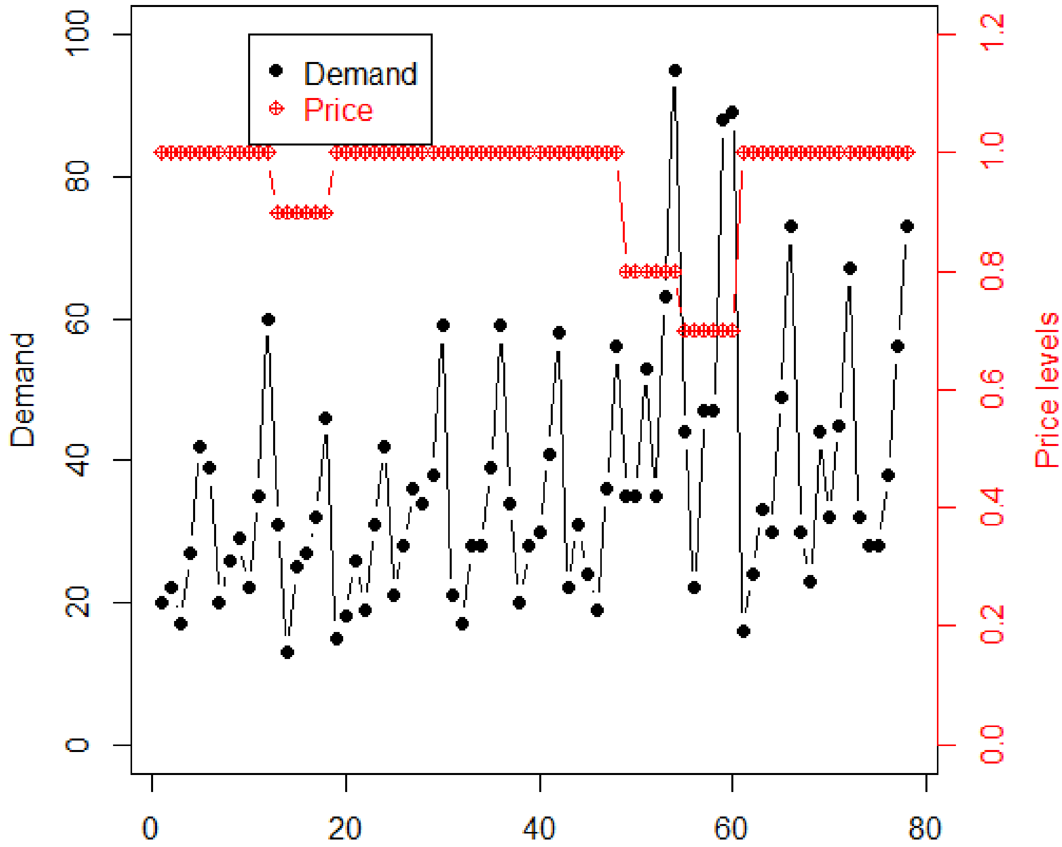

Figure 6.8. Data Set B: Demand vs. Price Levels.

In Figure 6.10, due to the small sample size we see anomalies, such that demand without promotions is higher that the demand when there is 10 % price discount. This is related with the randomness of the Poisson distribution and also. with the different demands on the different week days. The histogram at Figure 6.11 is more skewed than the histogram in Figure 6.4, due to the small sample size, and the densities of the demand differ more. Box plots of the daily demand vs. price levels and vs. weekdays are shown in Figure 6.12 and 6.13 respectively. It is obvious that in these cases the variances of the demands are higher when compared to the Data Set A.

In Figure 6.14, we see the interaction plot of demand at different week days. Like in Figure 6.10, we do not get exponentially decreasing lines. The reasons behind are the small sample size and Poisson randomness. The demand lines of Friday and

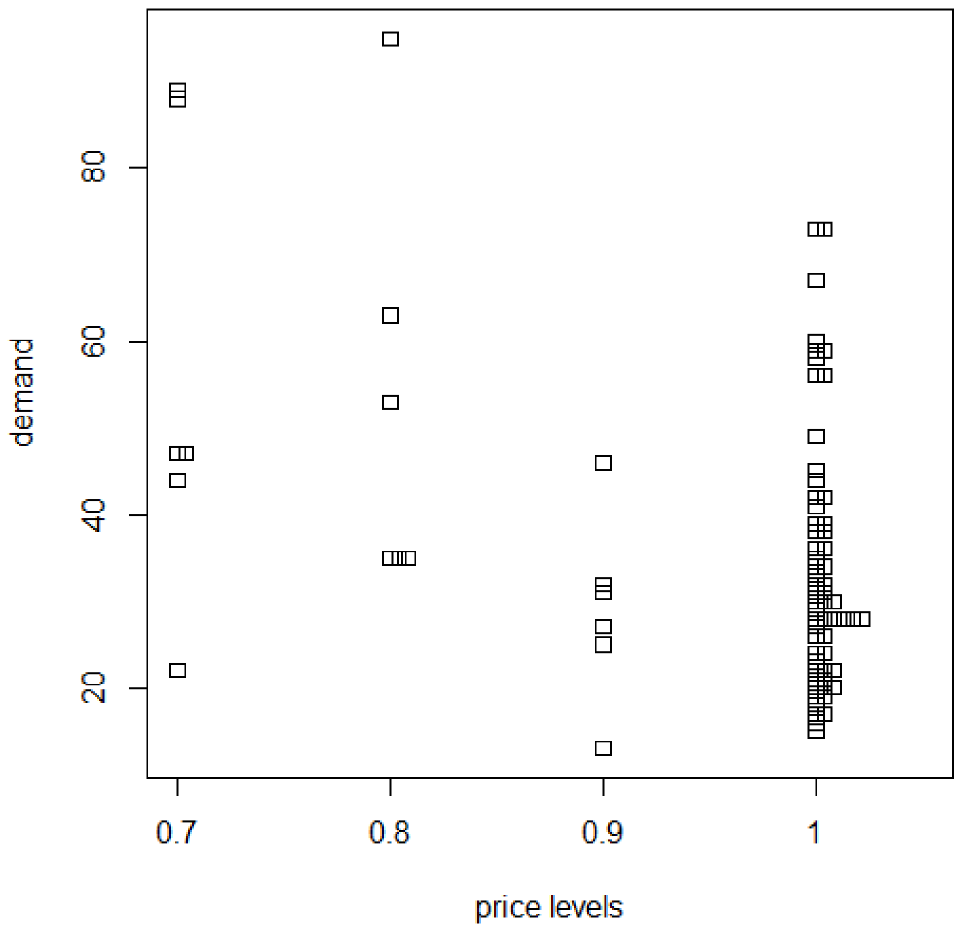

Figure 6.9. Data Set B: Daily Demand vs. Price Levels.

Saturday do not intersect each other due to relatively high demand values on these days. However, on other week days the demand lines intersect, but again this cannot be a proof of interaction, because we do not have enough observations to detect interaction.

6.2.2. ANOVA-Test

Making ANOVA in order to test the significance of price promotions in the demand data with less observations demonstrates that the mean values of the demand of these two groups are significantly different form each other. It is important that, the standard errors of the estimates in this model are higher than the standard errors of the results obtained from Data Set A.

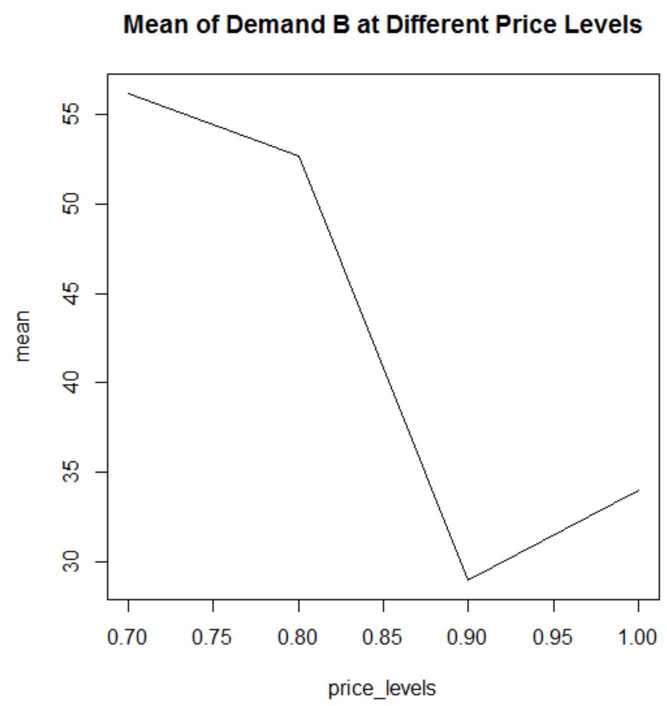

Figure 6.10. Mean of Demand B at Different Price Levels.

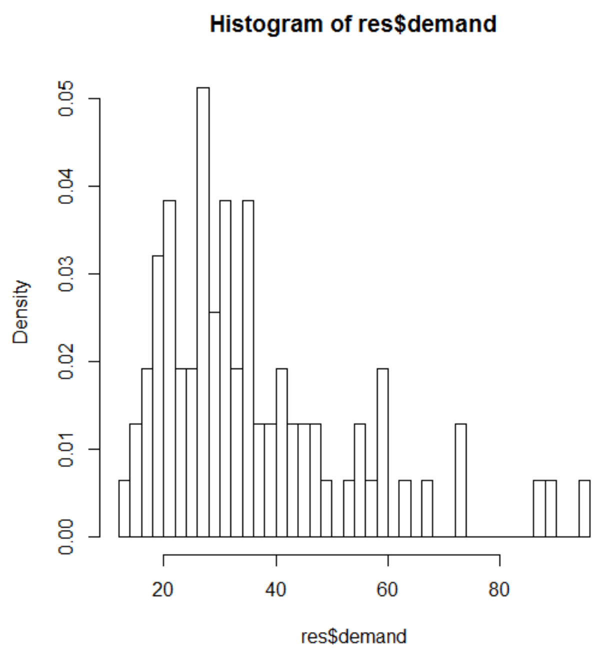

Figure 6.11. Histogram of Poisson Distributed Demand B.

```
Call:
glm(formula = res$demand ~ as.factor(res$prom1), family = poisson(link = "log"))
```

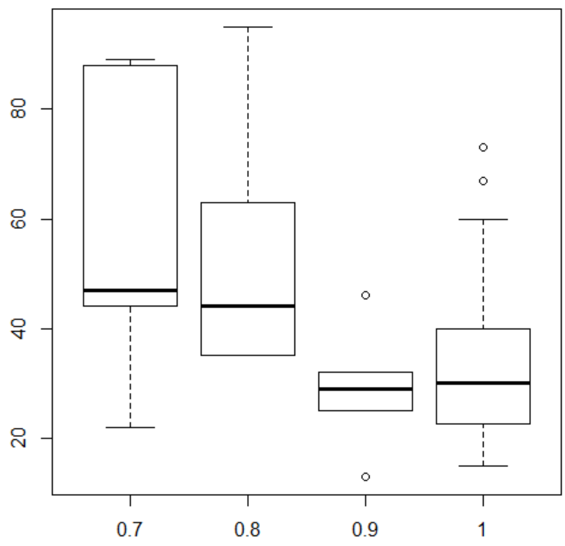

Figure 6.12. Box plot of the daily demands B at various price levels.

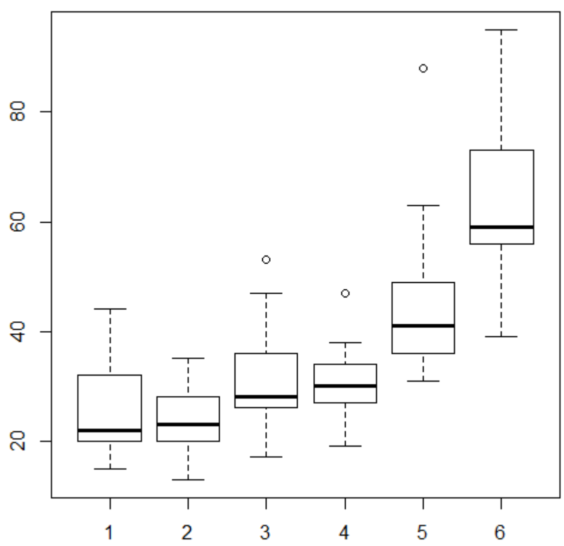

Figure 6.13. Box plot of the daily demands B at various week days.

```
Deviance Residuals:
   Min       1Q   Median       3Q      Max
-5.044   -2.220   -1.103    1.633    7.652

Coefficients:
                      Estimate Std. Error z value Pr(>|z|)
(Intercept)            3.56530    0.02171 164.199  < 2e-16 ***
as.factor(res$prom1)1  0.14827    0.04274   3.469 0.000521 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

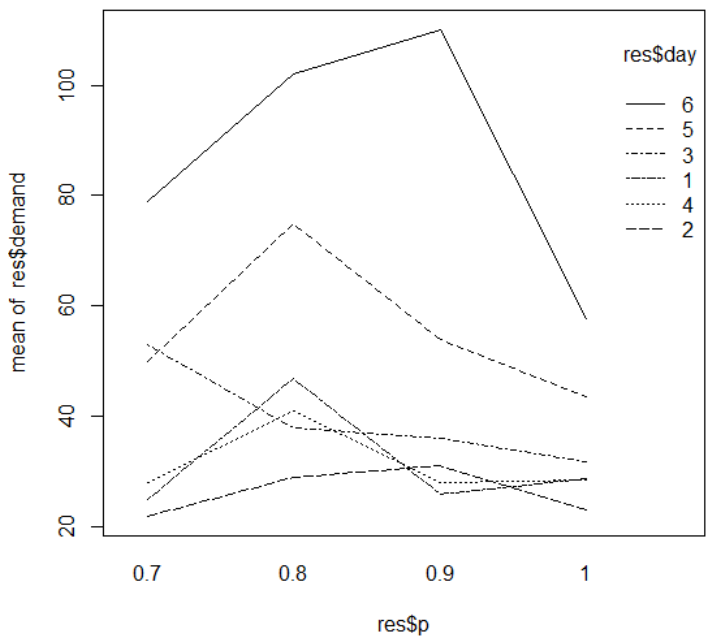

Figure 6.14. Mean of the Demand B in Different Week Days.

```
(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 705.77  on 77  degrees of freedom
Residual deviance: 694.03  on 76  degrees of freedom
AIC: 1113.1

Number of Fisher Scoring iterations: 4
```

Using "identity" link function does not change the results. It can be applied as it has been explained in Section 6.1.2.

6.2.3. Poisson Regression

Forecasting the demand and estimating the model parameters are done by Poisson regression again. Here are the results obtained from Data Set B:

```
Call:
```

```
glm(formula = res$demand ~ res$p + res$t + as.factor(res$day),
    family = poisson(link = log))

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.2590  -0.6047   0.0365   0.5169   2.5258

Coefficients:
                      Estimate Std. Error z value Pr(>|z|)
(Intercept)          3.9515065  0.1933987  20.432  < 2e-16 ***
res$p               -0.9516510  0.2039028  -4.667 3.05e-06 ***
res$t                0.0054793  0.0008843   6.196 5.78e-10 ***
as.factor(res$day)2 -0.3121335  0.0836684  -3.731 0.000191 ***
as.factor(res$day)3  0.1351052  0.0743948   1.816 0.069362 .
as.factor(res$day)4  0.0769822  0.0753465   1.022 0.306919
as.factor(res$day)5  0.5957515  0.0676747   8.803  < 2e-16 ***
as.factor(res$day)6  0.9426178  0.0641231  14.700  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 705.77  on 77  degrees of freedom
Residual deviance: 110.64  on 70  degrees of freedom
AIC: 541.71

Number of Fisher Scoring iterations: 4
```

The Poisson regression results with 3 months of observations show that β_0' is estimated as 3.9515065 and β_1 is estimated as -0.9516510. The standard errors of both estimates are around 0.2 in this case. Since the coefficient of price covariate, β_1, has a very small p-value, we can state that it is significant. So we observe that price has an influence on demand even the sample size is small. Using the estimated parameters, we can calculate the mean response for a Wednesday in the middle of the year and for price

0.8 using

$$exp(3.9515065 - 0.9516510 * 0.8 + 0.0054793 * 156 + 0.1351052) = 65.37 \qquad (6.3)$$

This means that for Wednesday, p=0.8 and t=156, the response variable follows a Poisson distribution with estimated mean value $\mu_{hat} = 65.37$. In Figure 6.9 we see how the demand values as price 0.8 differs.

Estimating the mean of the demand under the same conditions but p=1 leads us to get a mean demand value of $\mu_{hat} = 54.21$.

The mean of the demand value differs due to the existence of promotions. We calculate a higher mean demand when the price is low. It is obvious that, when the price decreases by 20% the demand increases by 17%.

If we compare the standard errors of the model parameter estimates of two different data sets, we observe that the standard error increases with decreasing sample size. Accordingly we obtain less precise estimates.

6.3. Other Possible Models

Since we want to explore how demand changes as a function of the covariates explained in details in Chapter 4, linear regression is the first and easiest method to use, however because of the nature of demand, linear regression is useless in our case. Due the non-negativity and discreteness of demand and non-constant variance function, where standard deviation changes with the mean, Poisson regression is a natural choice in order to estimate the parameters of the mean demand model.

In fact Poisson regression is not the only possibility to estimate the parameters of the mean demand model. There are other possible models that can be applied to quantify the effects on demand. These models are standard regression for log(demand)

and gamma regression.

6.3.1. Ordinary Linear Regression for log(demand)

Many people may prefer to use ordinary linear regression rather than the GLM, because it can be applied easily and it is well known. The ordinary linear regression models have normal distributed mean responses with constant variance. The shape of the distribution is therefore fixed. The change in mean responses are provided only by changing the mean parameter, namely μ. The variance does not change with the mean response, which means the variance of the response variable is the same for all possible mean values. However, it assigns in principle positive probability to all values on the real line, negative and positive, while demand is non-negative.

Different from Poisson regression, we want to try to estimate the parameters of our demand model with the help of the ordinary linear regression, which is easier to apply for many people. However, due to the non-negative nature of our response variable, we have to formulate the ordinary linear regression differently. In order to fulfill the non-negativity constraint of our response variable, we use "log-level" standard regression, where we take the logarithm of the response variable. In fact these two techniques, namely Poisson regression and "log-level" linear regression are similar, if μ_{demand} is not too small. We can use the same mean demand functions, explained in Chapter 4. The main difference is that, "log-level" standard regression produces continuous responses, while Poisson regression provides discrete values.

We can formulate our "log-level" regression equation as:

$$log((Demand)) = \beta_0' + \beta_1 log(p) + \epsilon \tag{6.4}$$

which is similar to our mean demand function given in Section 4.5, and can be written as:

$$Demand = exp(\beta_0')exp(\beta_1(p))exp(\epsilon) \tag{6.5}$$

The other important difference between these models is their variance functions and consequently their error structures. In "log-level" standard regression model $log(Demand)$ and error term ϵ ($\epsilon \sim N(0, \sigma^2)$) follow a normal distribution, which implies that the demand follows a lognormal distribution.

Since we have log responses in this model, when we exponentiate the equation $log(y) = log(\mu) + z\sigma$, we obtain a multiplicative error term in the equation: $y = \mu exp(z\sigma)$. So, standard deviation of the response variable of the "log-level" linear regression increases proportional to the mean, while in Poisson regression error term is proportional to the $\sqrt{\mu_{demand}}$. This means that, compared to Poisson regression, the standard error of parameters estimated by "log-level" linear regression increases faster.

However, due to the fact that for large μ values, the Poisson distribution can be approximated by the Normal distribution, the error structures of these regressions should approximate each other well.

We have already calculated the results of Poisson regression in Section 6.1.3. We used Data Set A again in "log-level" linear regression. Below, the results of both methods are demonstrated in order to make a clear comparison. The first one is the results of Poisson regression and the second one belongs to standard regression with log responses.

Results of the Poisson regression:

```
Call:
glm(formula = res$demand ~ res$p + res$t + as.factor(res$day),
    family = poisson(link = log))
```

```
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.3704  -0.9415  -0.1677   0.7709   4.7871

Coefficients:
                     Estimate Std. Error z value Pr(>|z|)
(Intercept)          4.065062   0.104410  38.934  < 2e-16 ***
res$p               -1.013661   0.106975  -9.476  < 2e-16 ***
res$t                0.001074   0.000107  10.031  < 2e-16 ***
as.factor(res$day)2 -0.165377   0.040127  -4.121 3.77e-05 ***
as.factor(res$day)3  0.203550   0.036617   5.559 2.72e-08 ***
as.factor(res$day)4  0.151247   0.037047   4.083 4.46e-05 ***
as.factor(res$day)5  0.559221   0.034058  16.419  < 2e-16 ***
as.factor(res$day)6  0.859201   0.032415  26.506  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 2173.33  on 311  degrees of freedom
Residual deviance:  543.21  on 304  degrees of freedom
AIC: 2224.3

Number of Fisher Scoring iterations: 4
```

Results of the "log-level" linear regression:

```
Call:
lm(formula = log(res$demand) ~ res$p + res$t + as.factor(res$day))

Residuals:
     Min       1Q   Median       3Q      Max
-0.81060 -0.13519 -0.00379  0.15287  0.67793

Coefficients:
                      Estimate Std. Error t value Pr(>|t|)
(Intercept)          4.0643439  0.1557704  26.092  < 2e-16 ***
```

```
res$p               -1.0342889  0.1597373  -6.475 3.80e-10 ***
res$t                0.0010553  0.0001471   7.173 5.61e-12 ***
as.factor(res$day)2 -0.1674069  0.0453646  -3.690 0.000265 ***
as.factor(res$day)3  0.1989614  0.0453653   4.386 1.60e-05 ***
as.factor(res$day)4  0.1460651  0.0453665   3.220 0.001422 **
as.factor(res$day)5  0.5597547  0.0453681  12.338  < 2e-16 ***
as.factor(res$day)6  0.8601436  0.0453703  18.958  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.2313 on 304 degrees of freedom
Multiple R-squared: 0.7207,     Adjusted R-squared: 0.7143
F-statistic: 112.1 on 7 and 304 DF,  p-value: < 2.2e-16
```

The standard error of the coefficient of price is equal to 0.104410, while the same value is 0.1557704 in "log-level" linear regression. Comparing the two models, as we stated before in this section, standard error values of Poisson regression estimates are smaller than the standard errors of "log-level" linear regression. The estimates in both results are almost the same and in their confidence intervals. The "log-level" linear regression is able to give the correct sign of the price coefficient as well.

The way of interpreting the results does not change in this model too. Again the response variable is the mean of the demand. Since we estimate the logarithm of the mean of the demand with this model, we have to exponentiate the fitted value, in order to find the real value of the mean of the demand. As an example we want to find the mean demand value for a Wednesday in the middle of the year and for price 0.8 using:

$$exp(4.06434398 - 1.0342889 * 0.8 + 0.0010553 * 156 + 0.1989614) = 36.62 \tag{6.6}$$

This means that for Wednesday, p=0.8 and t=156, the response variable follows a Poisson distribution with estimated mean value $\mu\_hat = 36.62$.

The estimated mean value of the Poisson regression is equal to 37.53. So, the two models are similar when used in this context.

6.3.2. Gamma Regression

Gamma regression is the third possibility for estimating parameters of the demand model. It is the most suitable one, for the case of continuous but nonnormal data, where the response variable can only be positive and the variance increases with the mean. In some conditions such as demand of oil, demand is a continuous positive random variate, so using gamma regression can be applied under these conditions. Since gamma regression in also a GLM, it is appropriate to be used for the responses taken at different covariate settings that do not have the same variance.

The density function of the gamma distribution is demonstrated in Equation 6.6:

$$f(x;k;\theta) = (\Gamma(k)\theta^k)^{-1}x^{k-1}e^{-x/\theta} \tag{6.7}$$

The gamma distribution has 2 parameters, shape parameter k and scale parameter θ. Its expectation of Y can be written in the form of these parameters as

$$E(Y) = k\theta. \tag{6.8}$$

Since $var(Y) = k\theta^2$, we can write it in terms of μ;

$$var(Y) = \mu\theta \tag{6.9}$$

which is also equal to

$$var(Y) = \mu^2/k \tag{6.10}$$

Note that μ must be positive and the variance of Y is nonconstant; it depends on the value of μ. For the Gamma distribution

$$CV = (\mu^2/k)^{1/2}/\mu = 1/k^{1/2} = k^{-1/2} = \psi^{1/2} = constant \tag{6.11}$$

so that, regardless of the value of μ, the ratio of noise to signal is the same. Here ψ stands for dispersion parameter and is equal to $1/k$. So $var(Y)$ can be written as $\mu^2\psi$. We can denote the dispersion parameter as $\sigma^2 = \psi$. Thus, rather than having constant variance, the gamma distribution imposes constant coefficient of variation. This is often a realistic model for demand and other data taking only positive values.

It was shown in Equation 6.9 that, the variance function of the gamma distribution depends on the shape k. If the value of k is large, then the shape of the distribution is similar to normal, if k is small, then the distribution is more skewed. Due to the fact that $var(Y) = \mu^2/k$, the standard error term of the gamma regression is proportional to the mean μ, like "log-level" linear regression.

Regression with the gamma model is going to use input variables and related coefficients to make a prediction about the mean of response variable, but actually we are focused on the scale parameter. This is because for gamma regression we have to assume that shape parameter k is the same for all observations, and so variation from case to case in $E(Y_i) = k\theta_i$ is due simply to variation of θ_i. Hence, in order to change the mean response in gamma generalized linear models, the scale parameter is changed.

Since the mean demand is in logarithmic form, we use the "log" link function. Applying Gamma regression model directly to the demand data A gives the results below:

```
Call:
glm(formula = res$demand ~ res$p + res$t + as.factor(res$day),
    family = Gamma(link = log))

Deviance Residuals:
     Min         1Q     Median         3Q         Max
```

```
-0.74229  -0.15901  -0.02983   0.13429   0.71529

Coefficients:
                      Estimate Std. Error t value Pr(>|t|)
(Intercept)          4.0899088  0.1550160  26.384  < 2e-16 ***
res$p               -1.0238618  0.1589637  -6.441 4.64e-10 ***
res$t                0.0009907  0.0001464   6.767 6.78e-11 ***
as.factor(res$day)2 -0.1638332  0.0451449  -3.629 0.000334 ***
as.factor(res$day)3  0.2030874  0.0451456   4.499 9.75e-06 ***
as.factor(res$day)4  0.1457633  0.0451468   3.229 0.001380 **
as.factor(res$day)5  0.5580191  0.0451484  12.360  < 2e-16 ***
as.factor(res$day)6  0.8556283  0.0451506  18.951  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Gamma family taken to be 0.05298898)

    Null deviance: 59.294  on 311  degrees of freedom
Residual deviance: 16.040  on 304  degrees of freedom
AIC: 2162.4

Number of Fisher Scoring iterations: 4
```

Standard error values of the estimates are similar to the results of "log-level" linear regression given in Section 6.4.1, which indicates that Gamma and "log-level" linear regression models are similar and have higher standard error values of the estimates due to their variance function $\sim$ mean relationships.

Pursuant to the results $\beta_0 = 4.09$ and $\beta_1 = -1.02$. Hence, for reference price level $E(Y) = exp(4.09 - 1.02 * 1) = 21.54$. Dispersion parameter ψ is calculated as 0.053. Therefore, we can calculate $k = 1/\psi = 1/0.053 = 18.87$ and $\theta = E(Y)\psi = 1.14$. So we have used the results of the GLM to find that for standard price level $= 1$ the response follows a Gamma distribution with $k = 18.87$ and $\theta = 1.14$.

I conclude this chapter with a check list that describes the standard steps that should be done in order to model daily demand data with known prices.

- Check yearly seasonality.
- Consider the week day effects on demand.
- Make interaction plots and detect interaction between your covariates.
- Estimate your model parameters. If the daily demand is smaller than 30 use Poisson regression in order to estimate the model parameters. Otherwise, you may prefer to use ordinary linear regression.
- Create interactions between your covariates in different models and check their AIC-values.
- Continue with the best model.

7. ASSESSING THE INFLUENCES ON DEMAND

In Chapter 6, we fixed the data we simulated and made the data analysis for two different data sets as if they were the real data. In Section 6.1.3 and 6.2.3 we calculated how much the demand are increased by promotions depending on the "real data analysis". However, with our demand model we can perform a sensitivity analysis in order to understand and formulate the influence of promotions on the demand. As we have already shown in Chapter 4, we model the mean of the demand with exponential mean demand model as follows:

$$log(\mu_{demand}) = log(\beta_0) + \beta_1 log(p) \tag{7.1}$$

If we increase the price by 10 %, we obtain:

$$log(\mu_{demand}) = log(\beta_0) + \beta_1 log(p * 1.1), \tag{7.2}$$

which can be rewritten as:

$$log(\mu_{demand}) = log(\beta_0) + \beta_1 log(p) + \beta_1 log(1.1) \tag{7.3}$$

which is equal to

$$log(\mu'_{demand}) = log(\mu_{demand}) + \beta_1 log(1.1) \tag{7.4}$$

$$log(\mu'_{demand}) = log(\mu_{demand}) + log(1.1)^{\beta_1} \tag{7.5}$$

which can be expressed as:

$$log(\mu'_{demand}) = log(\mu_{demand}) + log(1 + \delta)^{\beta_1} \tag{7.6}$$

where δ is the change in product price.

Hence, as we discussed in Section 4.1, the value of β_1 gives the relationship between price and mean demand. If it is below -1, then we have an elastic demand, which decreases with price increase. In the case of standard price, namely 1, β_0 gives the standard quantity.

Since we treated to the data sets as they were real data in the previous chapter, we assumed that we only know the "observed" values and do not know the theoretical values. However, since in fact we simulated our demand data, we have theoretical values of the model parameters. Therefore, we have the chance to check how accurate are our estimates by calculating their confidence intervals and checking whether the estimated values are in these intervals.

Let's look at again to the Poisson regression results of Data Set A in Section 6.1.3. Since we know all the theoretical values of our parameters from the data simulation chapter (i.e. Chapter 5), we can directly say that most of our estimates are close to the theoretical values. For example, our model estimates the coefficient of price, namely β_1 correctly, which is equal to -1. This value is in its confidence interval $-1.013661 \pm 2 * 0.106975$. Moreover, the estimated intercept, $\beta_0' = 4$ in our case, has a standard error of 0.104410, which means that, it is not very close to its theoretical value which is equal to $log(30) = 3.4$. But this depends on the data. Using another data set with same parameters may give more accurate β_0 estimate.

Moreover, the coefficients of different week days are significant and have small standard errors as well, which supports the idea that different week days have different mean demands.

We can make the same analysis for the Data Set B as well, whose Poisson regression results are given in Section 6.2.3. All our estimates in the result of the Poisson regression are close to the theoretical values. The intercept is estimated as 3,95. Its theoretical value is equal to $log(30) = 3.4$. The CI of the intercept can be expressed as $3.9515065 \pm 2 * 0.1933987$. The coefficient of the price is estimated precisely, since it is confidence interval $-0.9517 \pm 2 * 0.2039$ and its theoretical value -1 is in this range.

In the rest of this chapter, we want mainly to answer the question: "When can we assess the influence on the demand?". For this reason we performed some experiments. We tried to find out how good are we in reproducing the results whether we have the correct parameters.

Please note that, in the experiments, the time value of money is not considered. That is the inflation is ignored.

7.1. How Many Days of Data are Required?

The first experiment tries to answer the question: How many days of data are required to see the influence of price changes with high probability, in other words in order to get an acceptable power of the test? We have already discussed the parameter estimates of the Poisson regression models of Data Set A and Data Set B in Section 6.1 and 6.2. these data sets have the same mean of the daily demand, but different sample sizes. Both models are able to estimate significant and precise estimates, but the standard error values of the estimates are higher when we use the data with smaller sample size.

We can make the same experiment with a data set of more smaller observations, for instance 1 month of data. The data set used in this experiment, Data Set C, is given in the appendix. We demonstrated below the Poisson regression results with this new data set.

```
Call:
glm(formula = res$demand ~ res$p + res$t + as.factor(res$day),
    family = poisson(link = log))
```

```
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.3314  -0.5783   0.1851   0.6089   1.3454

Coefficients:
                     Estimate Std. Error z value Pr(>|z|)
(Intercept)          4.003426   0.546504   7.326 2.38e-13 ***
res$p               -0.981675   0.631647  -1.554  0.12015
res$t                0.013623   0.008063   1.690  0.09112 .
as.factor(res$day)2 -0.497869   0.162971  -3.055  0.00225 **
as.factor(res$day)3  0.275473   0.133506   2.063  0.03908 *
as.factor(res$day)4  0.143134   0.138165   1.036  0.30022
as.factor(res$day)5  0.543344   0.129216   4.205 2.61e-05 ***
as.factor(res$day)6  0.938883   0.124163   7.562 3.98e-14 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 205.297  on 23  degrees of freedom
Residual deviance:  22.896  on 16  degrees of freedom
AIC: 166.07

Number of Fisher Scoring iterations: 4
```

Although we calculate the value of β_1 correctly and in its CI, its p-value does not permit us to reject H0, which states that β_1 is equal to 0, because it is not smaller than 0.01. However, it is important to notice that the standard error values of the estimates are higher than the standard error values in the previous two examples. So, we see as expected that, when the sample size reduces, the standard error values of the estimated coefficients of covariates increases, which leads to less precise estimates.

7.2. Influence of the Values of β_0' and β_1?

In quantifying the effects of promotions another important thing to notice is the mean of the demand. Since β_0' and β_1 are given as default in the data simulation process, decision about the values of these parameters is very critical, because at that step, mainly the mean of demand is determined according to the used deterministic demand model. Assigning small values to the parameter β_0' means that the mean of the daily demand will be small. On the other hand, if large values are assigned, then the mean will be large. It is necessary to explain here again that, in the data sets we use β_0' which is equal to $log(\beta_0)$. Hence, if we want to generate demand data with a mean value around 30, than the value of β_0 should be equal to $log(30)$. Moreover, in the data sets shown in the introduction of Chapter 6, we have β_1 equal to -1.

In order to test the influence β_0' and β_1 we have different data sets with different parameter values. An example is shown below with a new data set, namely Data Set D. Its sample size is the same with the sample size of Data Set B (13 weeks), but Data Set D has a larger mean of demand, i.e. 450. As it is seen, although the sample size is small, the estimates are all significant in the model, however they are not precise, because they are not in their confidence interval. Different from the Poisson regression results of Data Set B shown in Section 6.2.3, the estimate of the coefficient of the price covariate is not in its CI: $-1.1299 \pm 2 * 0.0265$.

```
Call:
glm(formula = res$demand ~ res$p + res$t + as.factor(res$day),
    family = poisson(link = log))

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-7.9249  -2.3673  -0.4083   1.6986  16.4014

Coefficients:
                     Estimate Std. Error z value Pr(>|z|)
(Intercept)          6.801145   0.026301  258.59   <2e-16 ***
```

```
res$p               -1.129857   0.026544  -42.57   <2e-16 ***
res$t                0.001409   0.000027   52.17   <2e-16 ***
as.factor(res$day)2 -0.159554   0.010491  -15.21   <2e-16 ***
as.factor(res$day)3  0.284003   0.009420   30.15   <2e-16 ***
as.factor(res$day)4  0.144203   0.009715   14.84   <2e-16 ***
as.factor(res$day)5  0.610414   0.008836   69.08   <2e-16 ***
as.factor(res$day)6  0.914162   0.008419  108.58   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 32326.4  on 311  degrees of freedom
Residual deviance:  3549.4  on 304  degrees of freedom
AIC: 6079.4

Number of Fisher Scoring iterations: 4
```

If we increase β_1, which means its value approaches 0, then the significance of price impact on demand disappears. On the other hand, when β_1 is smaller than -1, for instance -2, then we see very clear effects of price on demand, which is an indication of the elastic demand.

7.3. What Happens If the Demand Is Even Smaller?

Another question to be answered in this topic would be "What happens if demand is even smaller, such as when 3 or 4 per day?". In this case, even for large sample sizes, in other words for instance whole year of data is available, it becomes impossible for the model to estimate the coefficients correctly, especially the effects of different week days. Due to the low demand, we are not able to observe significant difference between the days. We have generated Data Set D, which has an average mean demand of 3 per day and 312 observations in total and we used this data set for estimating the model parameters when the demand is very small. The model results show that the differences between mean responses on Monday and the mean responses on other week days (i.e. Tuesday, Wednesday and Thursday) are not significant. Since the p-values of the estimates of Tuesday, Wednesday and Thursday are not below 0.01, we cannot reject H_0. This means that when the mean of the daily demand is very small, the differences of the mean demands on some weekdays are not distinguishable.

```
Call:
glm(formula = res$demand ~ res$p + res$t + as.factor(res$day),
    family = poisson(link = log))

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.2202  -0.8360  -0.0791   0.6744   2.3351

Coefficients:
                      Estimate Std. Error z value Pr(>|z|)
(Intercept)          1.8533143  0.3276590   5.656 1.55e-08 ***
res$p               -1.3583996  0.3304498  -4.111 3.94e-05 ***
res$t                0.0017945  0.0003362   5.337 9.45e-08 ***
as.factor(res$day)2  0.0390275  0.1278024   0.305  0.76008
as.factor(res$day)3  0.3148648  0.1199766   2.624  0.00868 **
as.factor(res$day)4  0.2947212  0.1204458   2.447  0.01441 *
as.factor(res$day)5  0.6817940  0.1118894   6.093 1.10e-09 ***
as.factor(res$day)6  0.9139630  0.1079232   8.469  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 498.43  on 311  degrees of freedom
Residual deviance: 328.20  on 304  degrees of freedom
AIC: 1245.7

Number of Fisher Scoring iterations: 5
```

Although daily demand is very little, we can distinguish the price impact on demand correctly, because the actual value β_1 is in its confidence interval; $-1.3584 \pm 2 * 0.3304$.

So, in this case aggregation of the daily data into the weekly is a possibility in order to obtain data which can be interpreted more easily. We have done this aggregation to obtain weekly demand. Therefore, at the end we have 52 weekly observations instead of 312 daily observations. Since we aggregated demand data, we calculate the mean value of the price in each week. As a result, week day factors disappear and are not included in the GLM model. Below, you can find the codes for weekly analysis. We used again Data Set D.

```
mydemand<- res$demand
aggregatedays <- function(mydemand,ndays=6){
# retunrs the aggregated vector
# xv .. vector of daily observations
# ndays.. number of days to aggregate
# assumes that xv holds "full weeks"
if(length(mydemand)%%ndays!=0) {
  print("ERROR in aggregatedays() length of vector not a multiple of ndays");
  return(NULL)
}
hv<-(0:(length(mydemand)-1))%/%ndays
tapply(mydemand,hv,sum)
}

myd<- aggregatedays(mydemand,ndays=6)

myprice<- res$p
aggregatedays <- function(mydemand,ndays=6){
# retunrs the aggregated vector
# xv .. vector of daily observations
# ndays.. number of days to aggregate
# assumes that xv holds "full weeks"
if(length(myprice)%%ndays!=0) {
  print("ERROR in aggregatedays() length of vector not a multiple of ndays");
  return(NULL)
}
hv<-(0:(length(myprice)-1))%/%ndays
tapply(myprice,hv,mean)
```

```
}

myp<- aggregatedays(myprice,ndays=6)

agg_demand<- data.frame(myd, myp, tagg=1:length(myp))
```

With the help of this method, we eliminate the effects of week days and we are able to see the price influence, but the estimates for price remain the same, because the used information does not change, so it cannot give different results. In the experiment below, the aggregated data contains 52 observations, in other words 52 weeks, with a weekly average demand of 20 sales. Here, "agg_demand" is a matrix which contains demand, price and trend data in its columns.

```
Call:
glm(formula = agg_demand$myd ~ agg_demand$myp + agg_demand$tagg,
    family = poisson(link = log))

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-2.29479  -0.62470   0.06685   0.60410   1.95106

Coefficients:
                 Estimate Std. Error z value Pr(>|z|)
(Intercept)      4.071665   0.316223  12.876  < 2e-16 ***
agg_demand$myp  -1.358400   0.330450  -4.111 3.94e-05 ***
agg_demand$tagg  0.010767   0.002017   5.337 9.45e-08 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 84.132  on 51  degrees of freedom
Residual deviance: 43.647  on 49  degrees of freedom
AIC: 302.82
```

```
Number of Fisher Scoring iterations: 4
```

The estimates and the confidence intervals for β_1 in daily and weekly results are the same. But the value of the intercept, β_0 changes, because daily average demand and weekly average demand values are not the same.

We can illustrate the difference between two demand data with the help of box plots. In Figure 7.1, the box plot of non-aggregated data is seen and the difference in mean of the demand at different price levels cannot be distinguished very well. However, in Figure 7.2, although the information remains the same but the demand data are aggregated into weekly, we can distinguish different mean demand values due to different price levels obviously. It is important to notice again that, by aggregation, we do not change the data, we only change the way of analysis.

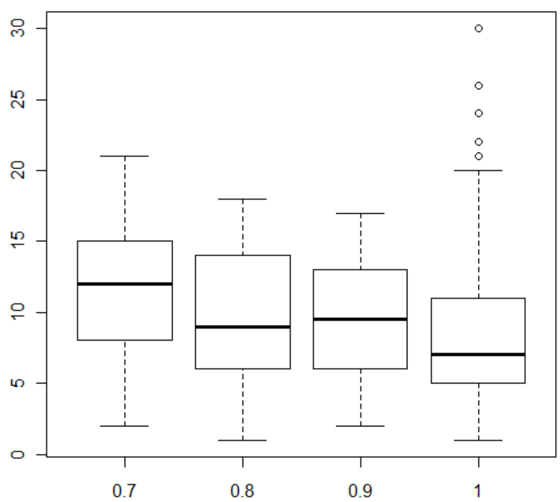

Figure 7.1. Box plot diagram of mean of the demand = 3.

7.4. What Happens If Strange Demand Is Observed?

Despite the fact that demand increases when the price goes down, there can be some days with strange demand, independent from the price of the product. Days before holidays, Mothers' Day, St. Valentines Day are some examples for this situation. While on these days exaggerated demand can be observed, on other days there can be very

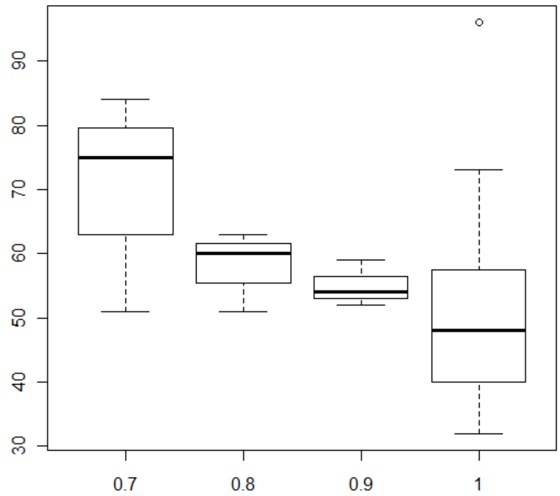

Figure 7.2. Box plot diagram of the aggregated mean of the demand.

little demand, due to different reasons. In this case we used again Poisson regression in order to estimate the model parameters and forecast the demand. We have another data set, namely Data Set E, which is shown in the appendix, and we used this data set in the Poisson regression analysis. Demand Data E contain high demands on some days without any price promotion. We have 312 observations with a mean demand of 30 per day. 8 days out of the data have the demand between 150 and 250. Results show that although the p-values of the estimates are small, the coefficient of price is not estimated correctly. The actual value of β_1 ($\beta_1 = -1$) is out of its confidence interval $-0.7692 \pm 2 * 0.09877$. So we can conclude that, when the exaggerated demand is observed in the demand data, our model cannot estimate the model parameters precisely.

```
Call:
glm(formula = res$demand ~ res$p + res$t + as.factor(res$day),
    family = poisson(link = log))

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-4.4285  -1.9314  -1.0756   0.0631  25.1335

Coefficients:
```

```
                     Estimate Std. Error z value Pr(>|z|)
(Intercept)          3.890e+00  9.936e-02  39.152  < 2e-16 ***
res$p               -7.692e-01  9.877e-02  -7.788 6.82e-15 ***
res$t                1.814e-03  9.831e-05  18.454  < 2e-16 ***
as.factor(res$day)2 -9.320e-02  3.590e-02  -2.596  0.00943 **
as.factor(res$day)3  4.002e-01  3.203e-02  12.496  < 2e-16 ***
as.factor(res$day)4  1.727e-01  3.361e-02   5.137 2.79e-07 ***
as.factor(res$day)5  3.670e-01  3.222e-02  11.389  < 2e-16 ***
as.factor(res$day)6  6.658e-01  3.047e-02  21.850  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 5876.5  on 311  degrees of freedom
Residual deviance: 4585.6  on 304  degrees of freedom
AIC: 6284.6

Number of Fisher Scoring iterations: 5
```

7.5. For How Many of the Experiments Significant Results Are Obtained?

In all experiments, simulated demand data are used, which contains not only Poisson randomness by nature, but also added randomness, that defines unpredictable uncertainties. Consequently, every run of the same experiment gives different results, although the common findings remains unchanged. Due to randomness the experimenter can have bad luck and get wrong results. It is clear that, in statistics the power of the test (i.e. the probability that the test will find a statistically significant result) is never 100%. So we estimate the power of the test with $H_0 : \beta_1 = 0$ by simulation.

In order to give an answer to the question: “For how many of the experiments significant results are obtained?”, the same experiment is repeated 1000 times. For example, let’s consider the Data Set B; the sample size is 3 months, namely 78 days, there are promotions in 3 weeks of these observations and mean of demand is about

30. Running this experiment 1000 times and calculating the probability of getting p-values of β_1 smaller than 0.05 result in 0.98, which means that for 1000 runs just 20 give non-significant estimates of the model covariate price. The probability to reject is close to 1 for Data Set A (52 weeks of data); however, when the sample size is smaller, (e.g. 4 weeks - Data Set C), power of the test drops to 0.60. The other important thing to mention here is that, the power of the test not only increases with the sample size, but also with increasing β_0.

A histogram is presented in Figure 7.3, which gives the frequencies of the p-values related to β_1 for 1000 experiments with 4 weeks of observation, 1 week with promotion and daily mean demand of 3. It is evident, that the power of this test is very small, namely 0.15, which means that only 15 % of the experiments give very small p-values for β_1.

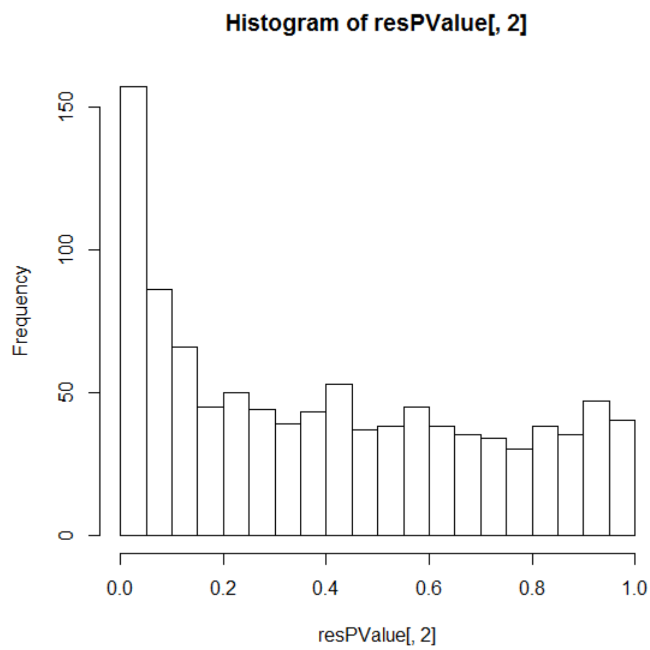

Figure 7.3. Histogram of p-values of β_1 for mean demand 4.

8. CONCLUSION

Price promotions are used by almost all companies in the retail sector. Since they are widely used tools of marketing, the dynamics behind them should be understood completely. In this study we intended to expand our knowledge about the impacts of price promotions on the demand of a product, to investigate in detail the models representing the demand and to quantify the effects of price promotions on the demand.

We started our work by modelling the mean of the demand using the exponential demand function well known in microeconomics. We tried to build the best and easiest demand model which is able to describe the mean of the demand. We used the logarithm of the generated mean demand model in order to avoid interactions and to obtain an additive model. Although we assumed that price and week day are not interacting with each other, price of the last week with promotions may have an influence on demand. This means that, demand can decrease in the week after promotion. In this situation an indicator can be added to the mean demand function, which indicates the week after promotion. This may be considered as the negative influence promotions may have on demand.

Considering the non-negativity and discreteness of demand we assumed that the observed demand is Poisson distributed. If real data are available we can estimate the parameters of the exponential demand function. Since we could not obtain reliable real data, we decided to simulate our own daily demand data by considering different characteristics of real data. In fact, this part of our work helped us a lot in understanding the underlying dynamics.

After modelling the demand and simulating demand data according to our assumptions, we used statistics in order to observe the influence of promotions on demand. We used Poisson regression to quantify these effects and forecast the demand.

We can summarize our main findings according to these experiments as follows:

GLM can be used to answer our main questions. Poisson regression is necessary if the mean of the daily demand is small. Poisson regression can help us to observe the influences of promotions on demand. Using Poisson regression in order to forecast the demand allows to estimate how much is the demand increased by promotions. For instance, we observed in our simulated data that when the price decreased by 20%, demand increased by about 18 %.

Since Poisson regression is a type of generalized linear models, it needs some extra knowledge in order to apply and interpret it. Many people may prefer to use ordinary linear regression due to its ease of use and interpretation. Therefore, we investigated ordinary linear regression as well, whose response variable is the logarithm of the mean of the demand. Since we can obtain positive responses by this type of regression too and its results are similar to Poisson regression results, we can use ordinary linear regression for log(demand) instead. However, for daily demand with small mean values, it cannot be used in place of Poisson regression. Moreover, note that the standard errors calculated by ordinary linear regression method are larger than the standard errors of Poisson regression results. The underlying reason of this situation is the difference in their variance functions.

If the sample size of the daily demand data is large, for example data of a whole year, we can see the influence of price changes with an acceptable power of the test. For small sample size, for instance one month of daily data, it is hard to show that the price has an influence. The values of β_0' and β_1 play an important role on the results. β_0' determines the mean value of the demand. If it is large, we can observe the effect of price on demand more easily. The value of β_1 is related to the price elasticity of demand. If it is close to 0, the impact of price on the demand disappears. However, when β_1 takes values smaller that -1, we see very expressive effects of price on the demand. If β_0' is small, analysing data becomes more difficult. In this case we can aggregate daily demand data into weekly, in order to be able to distinguish the price effect more easily. So results depend very much on the data and number of observations.

We tested our no interaction assumption between covariates in the Poisson regression model by allowing different kinds of interactions. When we compare the AIC criterion values of these models, we observe that our "no interaction" model has the best fit.

On the other hand if the price of a product remains in a very small range in the time of promotions, then we cannot distinguish clear effects with our model. It is important to mention here that in a competitive environment, if all competitors make promotions, then distinguishing the effects of promotions on the demand and estimating correct values for the model parameters become impossible. These facts highlight once again the importance of the available data quality. However, having such a model can helps us to plan promotions, forecast the effects of promotions and make revenue maximization.

To conclude, the aim of this thesis is opening a path and expanding the readers' horizons in building a reasonable model for demand and quantifying the impacts of price promotions on the demand data. The models we suggested in this work are easy to implement, but the interpretations of the results are crucial. It does not claim to be able to solve the entire puzzle but we hope, it will contribute to the existing knowledge and stimulate new research.

APPENDIX A: APPENDIX

Data Set C:

```
set.seed(334597)

trendpattern <- function(n,maxChange){
# outputs a trend factor vector of length n with a linear increasing trend.
# the average of the result vector is 1
# the first value in the result vector is 1-maxChange
# the last value in the result vector is 1+maxChange
1+maxChange*(1:n-(n+1)/2)/((n-1)/2)
}
generate.demand<- function(beta0_prime,beta1, rs, p_vec, n, m,d=3,munoise, sdnoise,
trendpatt , dayfact,epsilon){
# demand function: demand(p,) = exp(beta0+ beta1*log(p1)) * (1-rs*prom2),
#where p1 is the price of "our" product, prom2 the promotion indicator of the
#competitior
# c,b ... parameters of demand function
# rs ... assumption of the reduction (percentage) of "our" sales when the
#competitor make promotion
# p_vec ... different prices (in the moment length four)
# n ... number of weeks for simulation
# m ... number of days per week
# d ... number of week we have one of the promotion prices
#       in the moment we have a total of 3*d weeks with promotion
demand <- rep(0, n*m)
day <- 1+(0:(n*m-1))%%m
promweek <- rep(1, n)
index1 <- sample.int(n, d) # index1 now holds numbers of all the
#"promotion weeks"
promweek[index1[1:d]] <- 2  # we have set 2 to the promweek values for
#the first d promotion week numbers

makedayvec <- function(wv, m) {
```

```
# returns the day vector for a given week vector
# wv ... week vector
# m ... number of days per week

as.vector(t(matrix(rep(wv,m),nrow=length(wv))))
}
# now the promweek vector holds the pricelevel for each week
promday <- makedayvec(promweek,m)

p <- rep(1, n*m)
p[promday == 2] <- p_vec[3]    # 4 price levels are assigned to the promday
#in order to get the price vector

prom1 <- rep(1, n*m)
prom1[promday == 1] <- 0   #sets 1 to all promotion days,
#and 0 to non-promotion days in "our" price vector

promweek_comp <- rep(0, n)
index2 <- sample.int(n, 3*d) # index2 now holds numbers of all the
#"promotion weeks" of the competitor
promweek_comp[index2[1:(3*d)]] <- 1  # we have set 1 to the promweek
#values for the competitor's promotion week numbers
promday_comp <- makedayvec(promweek_comp,m)   #contains only zeros and ones.
logmu_demand <- beta0_prime+ beta1*log(p)

demand<- rpois(n*m, exp(logmu_demand + log(1 - rs*promday_comp)
+ log(trendpatt) + log(dayfact[day]) +  epsilon))
#let's assume promotion of the competitor reduces of our sales by 10%.
data.frame(demand,p,promday, prom1, promday_comp, t=1:length(p), day)
}
n=4
m=6
munoise=0
sdnoise=0.2
res <- generate.demand(beta0_prime=log(30),beta1=-1, rs= 0.1, p_vec=c(1, 0.9, 0.8, 0.7),
n=4, m=6,  d=1,munoise=0, sdnoise=0.2, trendpatt=trendpattern(n*m, 0.2),
 dayfact=c(0.7, 0.6, 0.9, 0.8, 1.3, 1.7), epsilon=abs(rnorm(n*m, munoise, sdnoise)))
```

Data Set D:

```
set.seed(334218)

trendpattern <- function(n,maxChange){
# outputs a trend factor vector of length n with a linear increasing trend.
# the average of the result vector is 1
# the first value in the result vector is 1-maxChange
# the last value in the result vector is 1+maxChange
1+maxChange*(1:n-(n+1)/2)/((n-1)/2)
}
generate.demand<- function(beta0_prime,beta1, rs, p_vec, n, m,d=3,munoise,
sdnoise, trendpatt , dayfact,epsilon){
# demand function: demand(p,) = exp(beta0+ beta1*log(p1)) * (1-rs*promday_comp),
#where p1 is the price of "our" product, prom2 the promotion indicator of the
#competitior
# c,b ... parameters of demand function
# rs ... assumption of the reduction (percentage) of "our" sales when the
#competitor make promotion
# p_vec ... different prices (in the moment length four)
# n ... number of weeks for simulation
# m ... number of days per week
# d ... number of week we have one of the promotion prices
#       in the moment we have a total of 3*d weeks with promotion
demand <- rep(0, n*m)
day <- 1+(0:(n*m-1))%%m
promweek <- rep(1, n)
index1 <- sample.int(n, 3*d) # index1 now holds numbers of all the
#"promotion weeks"
promweek[index1[1:d]] <- 2  # we have set 2 to the promweek values for
#the first d promotion week numbers
promweek[index1[(d+1):(2*d)]] <- 3  # we have set 3 to the promweek
#values for the second d promotion week numbers
promweek[index1[(2*d+1):(3*d)]] <- 4  # we have set 4 to the promweek
#values for the third d promotion week numbers
```

```
makedayvec <- function(wv, m) {
# returns the day vector for a given week vector
# wv ... week vector
# m ... number of days per week

as.vector(t(matrix(rep(wv,m),nrow=length(wv))))
}

# now the promweek vector holds the pricelevel for each week
promday <- makedayvec(promweek,m)

p <- rep(1, n*m)
p[promday == 2] <- p_vec[2]    # 4 price levels are assigned to the promday
#in order to get the price vector
p[promday == 3] <- p_vec[3]
p[promday == 4] <- p_vec[4]

prom1 <- rep(1, n*m)
prom1[promday == 1] <- 0   #sets 1 to all promotion days,
#and 0 to non-promotion days in "our" price vector

promweek_comp <- rep(0, n)
index2 <- sample.int(n, 3*d) # index2 now holds numbers of all the
#"promotion weeks" of the competitor
promweek_comp[index2[1:(3*d)]] <- 1  # we have set 1 to the promweek
#values for the competitor's promotion week numbers
promday_comp <- makedayvec(promweek_comp,m)   #contains only zeros and ones.
logmu_demand <- beta0_prime + beta1*log(p)

demand<- rpois(n*m, exp(logmu_demand + log(1 - rs*promday_comp)
+ log(trendpatt) + log(dayfact[day]) +  epsilon))
#let's assume promotion of the competitor reduces of our sales by 10%.
data.frame(demand,p,promday, prom1, promday_comp, t=1:length(p), day)
}
m=6
n=52
beta0_prime=log(3)
```

```
beta1=-1
p_vec=c(1, 0.9, 0.8, 0.7)
trendpatt <- trendpattern(n*m, 0.2)
dayfact = c(0.7, 0.6, 0.9, 0.8, 1.3, 1.7)
munoise=0
sdnoise=0.2
epsilon<- abs(rnorm(n*m, munoise, sdnoise))

res <- generate.demand(beta0_prime=log(3),beta1=-1, rs= 0.1,
p_vec=c(1, 0.9, 0.8, 0.7), n=52, m=6,
d=3,munoise=0, sdnoise=0.2, trendpatt=trendpattern(n*m, 0.2),
dayfact=c(0.7, 0.6, 0.9, 0.8, 1.3, 1.7),
epsilon=abs(rnorm(n*m, munoise, sdnoise)))
```

Data Set E:

```
set.seed(332255)

trendpattern <- function(n,maxChange){
# outputs a trend factor vector of length n with a linear increasing trend.
# the average of the result vector is 1
# the first value in the result vector is 1-maxChange
# the last value in the result vector is 1+maxChange
1+maxChange*(1:n-(n+1)/2)/((n-1)/2)
}
generate.demand<- function(beta0_prime,beta1, rs, p_vec, n, m,d=3,munoise,
sdnoise, trendpatt , dayfact,epsilon, strg, min, max){
# demand function: demand(p,) = exp(beta0+ beta1*log(p1)) * (1-rs*promday_comp),
#where p1 is the price of "our" product, prom2 the promotion indicator of the
#competitior
# c,b ... parameters of demand function
# rs ... assumption of the reduction (percentage) of "our" sales when the
#competitor make promotion
# p_vec ... different prices (in the moment length four)
# n ... number of weeks for simulation
# m ... number of days per week
# d ... number of week we have one of the promotion prices
#       in the moment we have a total of 3*d weeks with promotion
demand <- rep(0, n*m)
strg_demand<- sample.int(n*m, strg)
day <- 1+(0:(n*m-1))%%m
promweek <- rep(1, n)
index1 <- sample.int(n, 3*d) # index1 now holds numbers of all the
#"promotion weeks"
promweek[index1[1:d]] <- 2  # we have set 2 to the promweek values for
#the first d promotion week numbers
promweek[index1[(d+1):(2*d)]] <- 3  # we have set 3 to the promweek
#values for the second d promotion week numbers
promweek[index1[(2*d+1):(3*d)]] <- 4  # we have set 4 to the promweek
#values for the third d promotion week numbers
```

```
makedayvec <- function(wv, m) {
# returns the day vector for a given week vector
# wv ... week vector
# m ... number of days per week

as.vector(t(matrix(rep(wv,m),nrow=length(wv))))
}

# now the promweek vector holds the pricelevel for each week
promday <- makedayvec(promweek,m)

p <- rep(1, n*m)
p[promday == 2] <- p_vec[2]    # 4 price levels are assigned to the promday
#in order to get the price vector
p[promday == 3] <- p_vec[3]
p[promday == 4] <- p_vec[4]

prom1 <- rep(1, n*m)
prom1[promday == 1] <- 0   #sets 1 to all promotion days,
#and 0 to non-promotion days in "our" price vector

promweek_comp <- rep(0, n)
index2 <- sample.int(n, 3*d) # index2 now holds numbers of all the
#"promotion weeks" of the competitor
promweek_comp[index2[1:(3*d)]] <- 1  # we have set 1 to the promweek
#values for the competitor's promotion week numbers
promday_comp <- makedayvec(promweek_comp,m)   #contains only zeros and ones.
logmu_demand <- beta0_prime + beta1*log(p)

demand<- rpois(n*m, exp(logmu_demand + log(1 - rs*promday_comp)
+ log(trendpatt) + log(dayfact[day]) +  epsilon))
#let's assume promotion of the competitor reduces of our sales by 10%.
demand[strg_demand]<- as.integer(runif(strg, min, max))

data.frame(demand,p,promday, prom1, promday_comp, t=1:length(p), day)
}
```

```
m=6
n=52
beta0_prime=log(30)
beta1=-1
p_vec=c(1, 0.9, 0.8, 0.7)
trendpatt <- trendpattern(n*m, 0.2)
dayfact = c(0.7, 0.6, 0.9, 0.8, 1.3, 1.7)
munoise=0
sdnoise=0.2
epsilon<- abs(rnorm(n*m, munoise, sdnoise))

res <- generate.demand(beta0_prime=log(30),beta1=-1, rs= 0.1,
p_vec=c(1, 0.9, 0.8, 0.7), n=52, m=6,
 d=3,munoise=0, sdnoise=0.2, trendpatt=trendpattern(n*m, 0.2),
 dayfact=c(0.7, 0.6, 0.9, 0.8, 1.3, 1.7), epsilon=abs(rnorm(n*m, munoise, sdnoise)),
 strg=10, min=150, max=250)
```

REFERENCES

1. Agresti A., *An introduction to categorical data analysis* , 2nd Edition, John Wiley & Sons, Inc., New Jersey, 2007.

2. Dobson, A., *An introduction to generalized linear models* , 3rd Edition, CRC Press, New York, 2008.

3. Berk, R. and J. MacDonald, "Over dispersion and Poisson regression", *Journal of Quantitative Criminology*, Vol. 24, No. 3, pp. 269-284, 2008.

4. Blattberg, C. R. and S. A. Neslin, *Sales Promotion: Concepts, methods, and strategies*, Prentice Hall, New Jersey, 1990.

5. Faraway, J. J., *Extending the linear model with R : generalized linear, mixed effects and nonparametric regression models*, Chapman & Hall/CRC, Florida, 2006.

6. Faraway, J. J., *Linear models with R*, Chapman & Hall/CRC, Florida, 2005.

7. Gilbert, D. C. and N. Jackaria, "The efficacy of sales promotions in UK supermarkets: a consumer view", *International Journal of Retail & Distribution Management*, Vol. 30, No. 6, pp. 315-322, 2002.

8. Hormann, W. and M. G. Guler, "Statistical Inference - Lecture Notes", İstanbul, 2011.

9. Mittelhammer, R., *Mathematical statistics for economics and business*, Springer, New York, 1996.

10. Mulhern, F. J. and D. T.Padgett, "The Relationship Between Retail Price Promotions and Regular Price Purchases", *Journal of Marketing*, Vol. 59, pp. 83-90, 1995.

11. Pauwels, K., et. al., "New Products, Sales Promotions and Firm Value: The Case of the Automobile Industry," *Journal of Marketing*, Vol. 68, pp. 142-156, 2004.

12. Pauwels, K., et. al., "The Long-Term Effects of Price Promotions on Category Incidence, Brand Choice, and Purchase Quantity," *Journal of Marketing Research*, Vol. 39, No. 4, pp. 421-439, 2002.

13. R Development Core Team, "R: A language and environment for statistical computing", *R Foundation for Statistical Computing, Vienna, Austria*, ISBN 3-900051-07-0, URL http://www.R-project.org, 2008.

14. Walters, R. G., "Assessing the Impact of Retail Price Promotions on Product Substitution, Complementary Purchase, and Interstore Sales Displacement," *Journal of Marketing*, Vol. 55, pp. 17-28, 1991.