Aalto University
School of Science

# Mat-2.4177 Seminar on Case Studies in Operations Research 2015

## Estimation of consumer repurchase behavior

## Final Report

## Client: Microsoft Mobile Oy

06/05/2015

***Authors:***

*Niko Lietzén (Project Manager)*

*Alessandro Mancuso*

*Antti Melén*

*Pekka Alli*

# Contents

# 1 Introduction

This project was made during spring 2015 in Aalto University School of Science for the course Mat-2.4177 Seminar on Case Studies in Operations Research. The project's client was Microsoft Mobile Oy and the goal of the project was to evaluate, whether we can create a forecasting model based on the customer database they have collected. Additionally, the project's more far-reaching goal was to create a functioning model that the marketing division of Microsoft can use in the future. The model should forecast customers' future purchasing behavior. Our client asked us to take a mathematical approach to the problem, rather than explaining the drivers that explain the buying behavior.

Customer-base analysis models have been the subject of much research over the past few decades. These models are used for a variety of purposes such as customer targeting, customer valuation or pricing segmentation. One of the most influential models so far is the Pareto/NBD model (Schmittlein et al. [1987]). This model applies to a non-contractual setting in which the end of the customer relationship is not observed.

In recent years, there has been progress in the methodologies in the field of customer-base analysis. However, in practice simple heuristics are still commonly applied, see Verhoef et al. [2003]. Some authors have stated that in the contractual setting, these complex models do not offer a substantial improvement, see Donkers et al. [2007]. However, it is not clear whether there is a clear difference in the non-contractual setting. We have decided to review and implement one of these more sophisticated models. The result of this project gives Microsoft Mobile relevant insight into whether it is feasible to further develop and implement a more complex forecasting model. The project is particularly interesting, because applications and empirical analyses of these models are rare.

## 1.1 Marketing

Structural models, which rely on economic and/or marketing theories of consumer or company behavior, have recently become more frequent among marketing studies. They make it possible to test the behavioral theories, such as consumer demand or choices, from which they are derived and to obtain behavioral predictions of consumers: i.e. they aim both to explain and to predict. When less emphasis is put on the actual theory and data fitting is prioritized, we can talk about reduced-form models that represent the consumer's historical decision rules as derived from marketing data. The resulting estimates can furthermore be used to predict future behavior of the consumers, and the models can be validated by using for example time series, i.e. hold-out data (Chintagunta et al. [2006]). In recent years, advances in information technology have resulted in the increased availability of customer transaction data, largely due to the reduced costs of collecting and storing customer records and the availability of distribution channels that provide direct assess to the customer. This trend is closely linked to an ever-growing desire on the part of the marketing manager to use the customer's transaction databases to learn as much as possible about the customer base. In this framework, nowadays predicting the customer behavior is an important element for success in any direct marketing activity, above all in a very competitive market like the mobile industry.

## 1.2 Research Problems and Objectives

We are faced with a large customer transaction database, without the monetary values of the purchases. The working title of our project is the following: *Estimation of consumer repurchase behavior*. We wish to develop a model that would help the marketing department of Microsoft Mobile. The research problem is divided into the following questions:

- What are the models for replacement forecasting proposed by academic literature?
- Can we implement the models using R?

Furthermore, the chosen model should answer to the following questions:

- Which customers are most likely to be active in the future?
- What is the level of transactions we can expect in future periods from our customers, both individually and collectively?
- How can we measure the quality of our forecast?

## 1.3 The Probability Modeling Approach

We approach the problem by assuming that a hidden underlying stochastic process determines the behavior of a single customer. We only have a "foggy window" as we attempt to see the customers' true behavioral tendencies. We cannot assume that the future behavior is going to be identical with the past. For instance if a customer purchased two phones last year, there is a chance that he might purchase zero or two or even four phones next year. Our goal is to find suitable estimates for these unobservable and hidden processes.

We start by specifying a mathematical model where the observed behavior is a function of an individual's hidden behavioral characteristics. Here we denote the characteristics parameter by $(\theta)$ and the function of the behavior by $f(\theta)$. We start by trying to characterize the observed behavior with a basic probability distribution. The next step is to make assumptions about a mixing distribution that captures the heterogeneity of $\theta$. The distribution of $\theta$ tells us how much the characteristics of different customers vary. Combining the distribution of $\theta$ with the distribution for individual level behavior gives us a mixture model. By using the mixture model we can estimate the behavior of a randomly chosen customer. Hereby, we can make future predictions related to the behavior of a new customer.

## 1.4 Focus of the Study

First, we performed a literature review to get a general perspective on the field of customer-base analysis. Our client gave us the freedom to choose the model. However, they requested that we make the implementation using R-programming language. The size of the data we have is challenging by size, so we had to focus on keeping the R-code as efficient as possible. Furthermore, the information above is divided into different files. After trying few alternatives, we implemented an algorithm that combines these files relatively fast.

We reviewed several different models suggested by literature. We decided to focus on the Pareto/NBD model. Literature suggest that the Pareto/NBD model performs relatively well, when compared to other customer-base analysis models (von Wangenheim and Wübben [2009]). Furthermore, the Pareto/NBD model has been implemented in the R-package BTYD (Dziurzynski et al. [2014]). Due to the mathematical complexity of these models, building a model from scratch would have required more time than we have available.

## 1.5 Structure of the Report

The report is organized in six parts: introduction, theoretical study, model implementation, results and conclusions. The theoretical study reviews some basic probability theory required to understand the later parts of the section. Later in that section, we explain the theory behind our model and give the appropriate references to find the full formulations. In the model implementation section we introduce our dataset and the implementation process and difficulties related to it. Finally, the results are summarized in the fourth section and the last part of the report contains the conclusions we made related to the findings.

# 2 Theoretical Study

The purpose of this section is to introduce academic literature on the field of customer base analysis. We begin by reviewing some elementary probability theory required to quantify the behavior of the customers. Following that, we summarize the Pareto/NBD model and discuss the estimation of the model parameters. We have chosen to implement the Pareto/NBD model after discussions with our Microsoft Mobile contact, Lauri Salminen. Furthermore, the Pareto/NBD model is currently discussed extensively in the field of customer base analysis, see the work of Peter S. Fader and Bruce G. S. Hardie.

## 2.1 Survival Functions and Hazard Rates

Suppose that $T$ is a random variable, which represents the lifetime of a single customer. We define the lifetime of a customer to be the the time interval between the first and the last Microsoft Mobile product purchase. The customers go through two stages in their lifetime. They are alive for some period of time and then become permanently inactive. The customers who are expected to make a repurchase purchase are denoted as alive customers. Whereas, the customers who are no longer purchasing Microsoft Mobile products are denoted as dead customers.

In this report, we denote random variables with capital letters: $T$, $X$, $Y$, etc. Additionally, we use lower case letters to denote the possible values of random variables. Hence, "$T = t$" says that the random variable $T$ takes the value $t$. Let the cumulative distribution function (cdf) of $T$ be

$$F_T(t) = P\left(T \leq t\right)$$

which specifies the probability that the random variable $T$ is less than some given value $t$. If we assume that $F_T$ is differentiable, the derivative is called the density function of $T$. The corresponding density function is then

$$f_T(t) = \frac{dF_T(t)}{dt}.$$

The distribution of a random variable that describes lifetime is often given in terms of the survival function,

$$S_T(t) = P\left(T > t\right) = 1 - F_T(t).$$

The survival function gives a probability of an individual surviving beyond time $t$. The force of mortality or more commonly the hazard rate is then defined as

$$h_T(t) = \frac{f_T(t)}{F_T(t)},$$

when we assume that the density $f_T(t)$ exists. The hazard rate tells us the probability of a customer death just after time $t$, given that the customer was alive at time $t$. The survival function can be written in terms of the hazard rate, see Bedford and Cooke [2001]. Thus, the hazard rate and the probability distribution function are equivalent ways of specifying the probabilistic behavior of a lifetime variable. Often in practical situations we try to guess the form of the hazard rate and recognize an appropriate family of models. In this report, we use the recommendations of literature to choose the hazard rates. The chosen hazard rates are discussed later on this report. In the context of customer-base analysis, the hazard rate is often denoted as dropout rate. We will be using the term dropout rate in the later sections.

## 2.2 Classifying Business Settings

In the Microsoft Mobile buyer-seller relationships neither the length nor the usage or monetary volume is contractually fixed. In this non-contractual setting it is especially challenging to forecast whether a customer will make a repurchase and if so how many transactions will he conduct in the future. The opposite is the contractual relationship, e.g. a mobile phone service provider knows exactly when the buyer-seller relationship has ended. In our case, Microsoft Mobile is not notified when a customer decides to purchase a phone from a different company. It is relevant to correctly classify the business setting before the modeling process, since the models suggested by literature are notably different in the contractual and non-contractual case.

## 2.3 Heuristics

According to the study by Verhoef et al. [2003], many companies base their marketing strategies on "gut feeling" and simple mathematics. Given the time and money cost associated with implementing complex stochastic models, a majority of the companies feel that they do not have to implement advanced methods. To convince the companies, they would require a study where the superiority of these academic models is clearly demonstrated.

One of the most common heuristics for the given problem is the RFM method. The method analyses the value of a customer based on the recency of the latest purchase, frequency of the purchases and monetary value of the purchases. We omit the monetary value, because we are not considering models with monetary value in this report. Common methodology for RFM is to scale the recency, frequency ( and monetary value ) of every customer from 1 to 10. The customers with recency value 10 are those that have made the latest purchases. Likewise, the customers with frequency value 10 are the most frequent buyers. After selecting the values, we segment the data from the intersection of the attribute values. We then have 1000 (10 x 10 x10 for every attribute) different combinations. The results of the segmentation are ordered in descending order, thus identifying the most valuable customers which have the most recent and frequent purchases. The decision maker then applies direct marketing according to some rule based on the order of the customers.

## 2.4 Gamma Distribution

Gamma distribution is a two-parameter probability distribution function family. The two parameters are shape parameter $r$ and scale parameter $\alpha$. gamma distribution is frequently used to model waiting times. Furthermore, in Bayesian statistics gamma distribution is commonly used as a prior distribution for various types of rate parameters, such as the $\lambda$ of an exponential distribution.

The probability density function for gamma function is

$$f(x|r,\alpha) = \frac{\alpha^r x^{r-1} e^{-\alpha x}}{\Gamma(r)}. \tag{1}$$

Probability density functions for altering the shape parameter are shown in figure 1 and for different scale parameters in figure 2. By choosing shape parameter $r = 1$, we get an exponential distribution with scale parameter $\alpha$. With positive integer as a shape parameter $r$, gamma distribution is also known as Erlang distribution. In our report, we use gamma distribution as the Bayesian prior distribution for the customers latent parameters $\lambda$ and $\mu$, that are introduced later in the report.

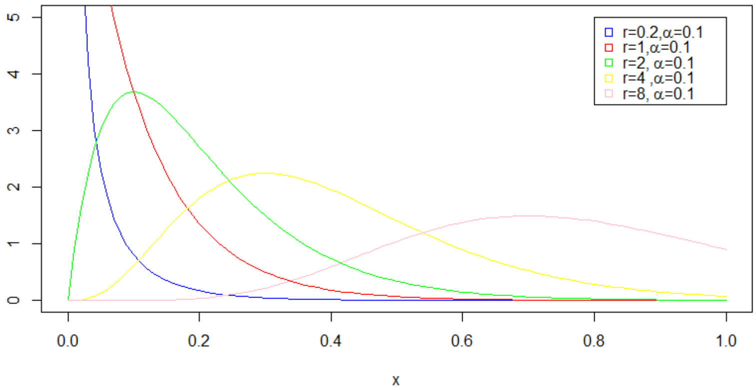


Figure 1: Probablity density of gamma function with different shape parameters

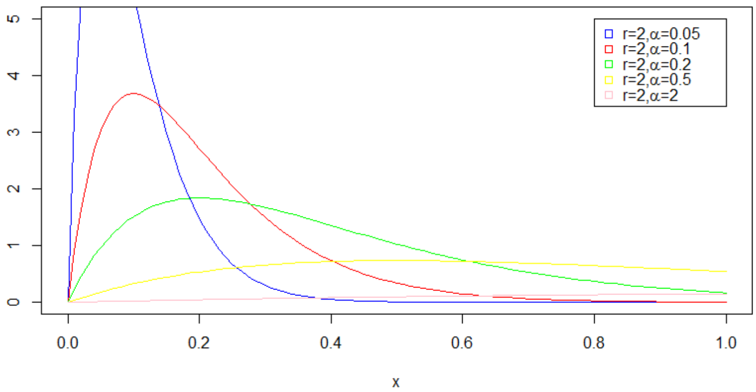


Figure 2: Probablity density of gamma function with different scale parameters

## 2.5 Negative Binomial Distribution (NBD)

Negative binomial distribution gives the probability of $r-1$ successes and $x$ failures in $x+r-1$ trials, and success on the $(x+r)$th trial. The negative binomial (NBD) distribution arises as a continuous mixture of Poisson distributions, where the Poisson rate $\lambda$ follows the gamma distribution. This is the less common way to define the NBD. In this formulation, the NBD has a shape $r$ and a scale $\alpha$ parameter, that come from the mixing gamma distribution. We can write the probability density function of the negative binomial distribution as:

$$\begin{aligned} f(X = x|r, \alpha) = & \int_0^\infty f_{poisson(\lambda)}(x) f_{gamma(r, \frac{1-\alpha}{\alpha})}(\lambda) d\lambda \\ = & \frac{\Gamma(r+x)}{\Gamma(r)x!} \left(\frac{\alpha}{\alpha+1}\right)^r \left(\frac{1}{\alpha+1}\right)^x, \end{aligned}$$

where $f_.$ are the corresponding probability density functions. Note that usually NBD is defined so that the distribution is discrete. Some examples of the NBD can be seen in figures 3 and 4.

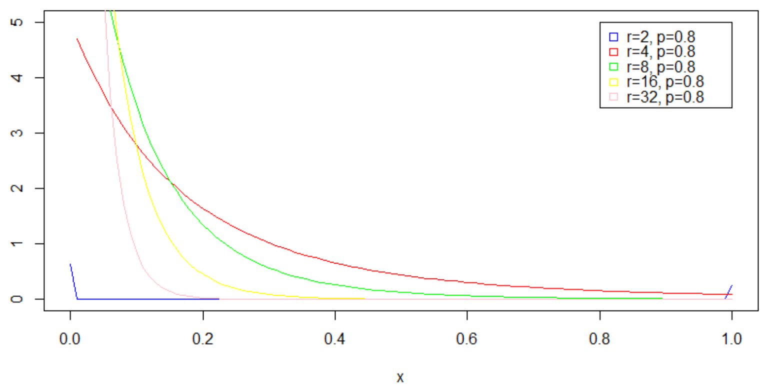


Figure 3: Probability density of NBD function with different number of trials

## 2.6 Pareto Type II Distribution

Pareto II distribution is a distribution specified by two parameters that are scale $s$ and shape $\beta$. The Pareto II is a standard Pareto distribution that has been shifted along the $x$-axis such that it starts at $x = 0$. The Pareto distribution arises as a continuous mixture of exponential distributions, where the rate parameter $\mu$ follows the gamma distribution. We can write the probability density function of the Pareto distribution as:

$$\begin{aligned} f(t|s, \beta) = & \int_0^\infty f_{exp(\mu)}(t) f_{gamma(s, \frac{1-\beta}{\beta})}(\mu) d\mu \\ = & \frac{s}{\beta} \left(\frac{\beta}{\beta+t}\right)^{s+1}, \end{aligned}$$

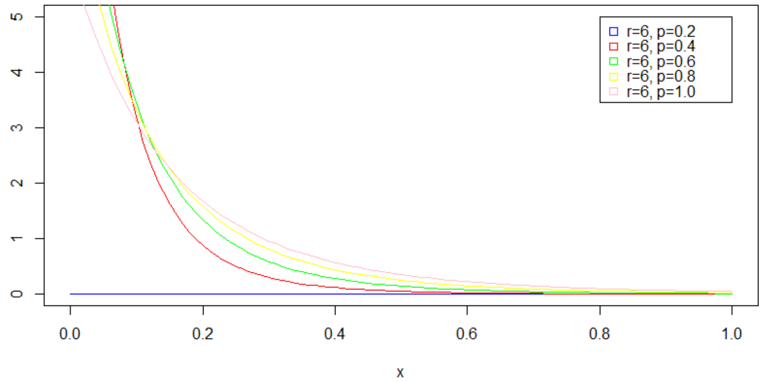


Figure 4: Probability density of NBD function with different probability parameters

where $f_.$ are the corresponding probability density functions. Even though, the exponential is distribution is considered to be a light tailed distribution, the mixing of exponentials produces a heavy tailed function. Examples of Pareto distribution with different parameters can be seen in figures 5 and 6.

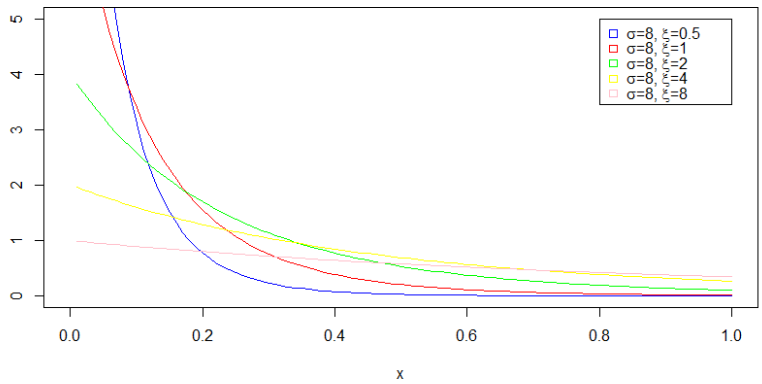


Figure 5: Probability density of Pareto II function with different shape parameters

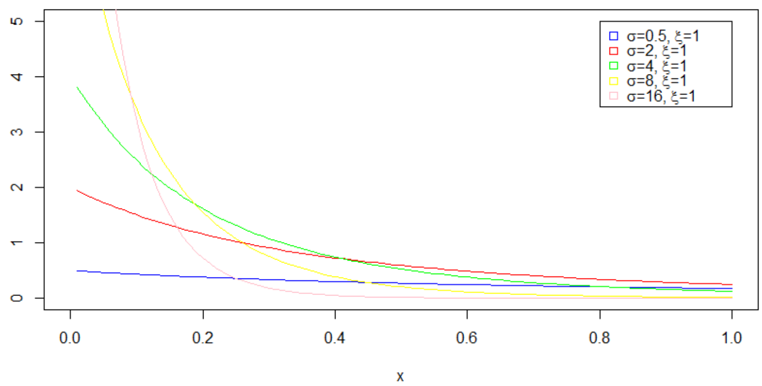


Figure 6: Probability density of Pareto II function with different scale parameters

## 2.7 Pareto/NBD model

The Pareto/NBD model was developed by Schmittlein et al. [1987], to describe repurchase behavior in a non-contractual setting. We have chosen the model based on the recommendations in academic marketing literature (von Wangenheim and Wübben [2009]). The model is attractive for several reasons:

1. The theoretical foundation is well established.
2. Limited information requirements, only purchase dates are required.
3. The model generates probabilistic outputs about future activity.

For the theoretical foundation, see Ehrenberg [1972]. The Pareto/NBD model operates on three values ($X = x, t_x, t$), based on customers' past purchase behavior. Here $x$ is the number of purchases made by a single customer, $t_x$ is the time when the latest purchase has occurred and $t$ is the time of observation (time currently). The Pareto/NBD model builds upon the assumption that purchases follow Ehrenberg's NBD model and the dropout events follow a Pareto distribution of the second kind. The Pareto/NBD model makes the following assumptions:

(i) Customers go through two stages in their lifetime: they are first alive for some period of time, then become permanently inactive.

(ii) The number of transactions made by a customer ($X$) follows a Poisson process with transaction rate $\lambda$. Therefore, the probability of observing $x$ purchases in the time interval $(0, t]$ is

$$P(X(t) = x|\lambda) = \frac{(\lambda t)^x e^{-\lambda t}}{x!}, \quad x = 0, 1, 2, ...$$

(iii) A customer's unobserved lifetime of length $\tau$ (after he is viewed as being inactive) is exponentially distributed with dropout rate $\mu$,

$$f(\tau|\mu) = \mu e^{-\mu\tau}.$$

(iv) Heterogeneity in dropout (hazard) rates among customers follow a gamma distribution with shape parameter $s$ and scale parameter $\beta$,

$$g(\mu|s, \beta) = \frac{\beta^s \mu^{s-1} e^{-\mu\beta}}{\Gamma(s)}.$$

(v) Heterogeneity in transaction rates among customers follow a gamma distribution with shape parameter $r$ and scale parameter $\alpha$.

(vi) The transaction rate $\lambda$ and the dropout rate $\mu$ are independent and vary between customers.

A Poisson process is a common and simple stochastic process for modeling the times at which customers make the purchases. Thus, it is a natural choice for our model. Assumption (ii) is equivalent to assuming that the time between purchases follows the exponential distribution with transaction rate $\lambda$,

$$f(t_j - t_{j-1}|\lambda) = \lambda e^{-\lambda(t_j - t_{j-1})} \quad t_j > t_{j-1} > 0,$$

where $t_j$ is the time of the $j$th purchase. The third assumption follows directly from this. Assumptions (iv) and (v) are purely based on recommendations of marketing literature. By heterogeneity, we mean that the personal preferences vary according to the gamma distribution in the population. The estimation of the model parameters is discussed in section 2.9.

The most relevant information that our models yields are:

- $E[X(\tau)]$ is the total expected number of transaction in a time period of length $\tau$.
- P(Active$|X = x, t_x, t)$ is the probability that a customer with parameters $(x, t_x, t)$ is still active at some time $t$.

- $\mathrm{E}(\hat{X}|X = x, t_x, t, \hat{t})$ is the expected number of transactions $\hat{X}$ of a customer with parameters $(x, t_x, t)$ in a time frame $(t, t + \hat{t}]$.

Using this information, we can identify the customers that are the most suitable targets for direct marketing. The customers with a high $\mathrm{E}(\hat{X}|X = x, t_x, t, \hat{t})$ value are most likely going to purchase a phone in the given time interval. Thus, these customers should get extra attention in the form of direct marketing.

## 2.8 Model Parameters

We begin the estimation process by dividing the database into two parts, the calibration period and the holdout period. The length of both these periods should be sufficiently long. We have decided to use two periods with equal lengths, which is suggested by Dziurzynski et al. [2014]. The holdout period is used to validate the model and the calibration period is used to estimate the model parameters. The model has four parameters $(r, \alpha, s, \beta)$, where $(r, \alpha)$ are the shape and scale parameters of the gamma distribution that determines the distribution of the variation of customer purchase rates across individuals. Furthermore, $(s, \beta)$ represent the scale and shape parameters of the gamma distribution that determines the variation of dropout rates across individuals.

We made the assumptions that the customer transaction rate is a Poisson process and that the parameter $\lambda$ varies according to the gamma distribution. We can then use the formulation from the previous sections to define the probability density function (pdf) of the mixture distribution:

$$f(X = x|r, \alpha) = \frac{\Gamma(r + x)}{\Gamma(r)x!} \left(\frac{\alpha}{\alpha + 1}\right)^r \left(\frac{1}{\alpha + 1}\right)^x,$$

which is the pdf of the negative binomial distribution (NBD). Likewise, we assumed that the dropout rate is distributed exponentially and that the parameter $\mu$ varies according to the gamma distribution. Now the probability density function of the mixture distribution is:

$$f(t|s, \beta) = \frac{s}{\beta} \left(\frac{\beta}{\beta + t}\right)^{s+1},$$

which is the pdf of the Pareto II distribution. Furthermore, we defined $\lambda$ to be the latent transaction rate of single customer. Similarly, we defined $\mu$ to be the latent dropout rate of single customer. Within the Pareto/NBD model we get the following relations:

- $r/\alpha$ represents the number of purchases an average customer has made in one time unit
- $s/\beta$ represents the dropout rate of an average customer in one time unit
- Lifetime of an average customer is exponentially distributed with expected value $1/(s/\beta)$.

When the shape parameter $s$ approaches 0, the mean of $\mu$ tends to zero. This means that the lifetime of the customers become infinite, since there is no dropout. Hence, the Pareto/NBD reduces to the simple NBD model. When $s$ approaches infinity the gamma distribution shrinks towards a mass point at its mean $s/\beta$. The Pareto distribution gets close to an exponential distribution with parameter $\mu = s/\beta$. Thus, the customers become more homogeneous in their mean lifetimes. The effects of different parameters is illustrated in figures 1 - 6. Overall, depending on the value of $s$, the Pareto/NBD can approach either the NBD model or the Exponential/NBD model. We expect our model to lie between these two extremes.

## 2.9 Parameter Estimation

The Pareto/NBD model is currently recommended by many researchers in the field of marketing. However, only a few researcher have successfully implemented the model. This is a result of the computational complexity of the model. The complexity issues arise from the difficulty of estimating the model parameters. The marketing literature lacks a consensus which of the proposed parameter estimation methods we should use. The methods can be divided into two branches: the method-of-moments and the maximum likelihood estimation. Both of these branches contain several different formulations. However, in a recent paper Fader et al. [2005] stated that method-of-moments approach is statistically more unstable and does not have the desirable statistical properties commonly associated with maximum likelihood estimation.

Maximum likelihood estimation (MLE) has desirable statistical estimator properties and is the estimator of choice for many statistical models. The important statistical properties can be found in Stuard and Ord [1987]. We have decided to use the maximum likelihood formulation by Fader et al. [2005].

Next we will provide the formulation for the model output estimation. Consider a customer who had $x$ transactions in the period $(0, t]$ with transaction times $t_1, t_2, ..., t_x$. There are two possible ways this pattern of transaction could arise:

i. The customer is still active at the end of the observation period ($\tau > T$). Now the individual-level likelihood function is the product of the exponential density functions and the associated survival function is:

$$\begin{aligned} L(\lambda | t_1, t_2, \ldots, t_x, t, \tau > t) &= \lambda e^{-\lambda t_1} \lambda e^{-\lambda (t_2 - t_1)} \ldots \lambda e^{-\lambda (t - t_x)} \\ &= \lambda^x e^{-\lambda t} \end{aligned}$$

ii. The customer became inactive at some time $\tau$ in the interval $(t_x, t]$. Now the individual-level likelihood function is

$$L(\lambda | t_1, t_2, \ldots, t, \text{inactive at } \tau \in (t_x, t]) = \lambda^x e^{-\lambda \tau}.$$

From here we can see that the information on the timing of the $x$ transactions is not relevant in the model. The sufficient information we require related to a single customer is $(X = x, t_x, t)$, as pointed out earlier. We can remove the conditioning on $\tau$, $\mu$ and $\lambda$ by integrating from $t_x$ to $t$ so that we get the likelihood function for a randomly chosen customer. The steps can be found in Fader et al. [2005]. We take the logarithm from the likelihood function to make the calculations more simple. The likelihood function and the log-likelihood function have the same maximizing parameters, hereby it does not change the results. The sample log-likelihood function is

$$LL(r, \alpha, s, \beta) = \sum_{i=1}^{N} ln \left( L(r, \alpha, s, \beta) | X_i = x_i, t_{x_i}, t_i) \right).$$

This can be maximized by standard numerical optimization routines. We use an optimization routine provided in the R-package BTYD. The difficulties of finding these parameters is discussed further in the report.

# 3 Model Implementation

## 3.1 Data Overview

The basis for modeling was two datasets provided by Microsoft Mobile: customer database and device database. The customer database consists of consumer id, date of last active day and the market of which the customer have activated their phone. The device database consists of consumer id, purchase date and platform id which tells us which operating system runs on the purchased phone. The size of the customer database is close to 3 million records and the size of device database is 3.8 million records. The size of the dataset is large which has quite an impact to the data manipulation as it brings hardware requirements to the manipulation of the full dataset.

Noteworthy factor with the given datasets is that the they do not necessarily cover the full purchase history of given customers, but the dataset is only a sample.

## 3.2 Data Preparation

Before we could use our dataset with our way of modeling, we had to prepare the datasets. The first task in the preparation was to remove outliers: 1) every activation time of a phone which is done within 50 days of previous one for the same customer is believed to be an outlier. 2) Every customer which has more than 10 purchases are expected to be a reseller or other individual which has other use for the phone than a regular customer.

After removing outliers from the datasets, the datasets had to be merged into a one dataset which has the format required for the modeling. The format of the final dataset consists of individual rows which have: customers id, number of purchases with given time, time from last purchases and length of customers time interval. The customers time interval is determined by the first purchase and the last active date.

## 3.3 Case: Microsoft Mobile

By applying the Pareto/NBD model we wish to determine:

- Active and inactive customers

- Individual and collective transaction forecasts

Using this information, the Microsoft Mobile marketing division should be able to target customers that are most likely going to make a purchase in the near future. Furthermore, the marketing personnel can try to reactivate some identified inactive customers. By focusing the direct marketing, the customers get less annoyed and the direct marketing does not impact the brand negatively. We hope that the results we present are motivation enough, to look for and analyze alternative methodologies and approaches to find an optimal marketing strategy.

We decided not to utilize the information regarding the date of latest activity since we were unable to find a suitable extension to the Pareto/NBD model that utilizes this information. Furthermore, the true meaning of this latest activity date was unclear to us. It was not clear, if the date was an indicator of latest use of a device or if the date was the latest response to a direct marketing campaign. However, in the case of Microsoft Mobile the use of this information would be a natural extension to the model.

We also had some modeling difficulties. The Pareto/NBD maximum likelihood estimation revealed some abnormal behavior on the data set. Especially, the parameter $s$ tended to converge to its estimation upper bounds, reaching unrealistically high values. This means that the Pareto distribution gets close to an exponential distribution. However, this kind of behavior is not uncommon, as stated by von Wangenheim and Wübben [2009]. We have chosen examples to the next section where the parameters are realistic and do not present this kind of behavior.

# 4 Results

In this chapter we test our model with the given data and show some examples of the results. We have seven different platforms and the customer behavior might be different for each of them. We test some of them separately and also together. Furthermore, we focus on the platform that Microsoft Mobile finds the most relevant for marketing. By platform we mean the operating system of the mobile phones. To run the test we first choose a sample size of 10000 purchases to ensure we have enough of those customers who make repurchases. To make the data more realistic we leave out those purchases that have occurred within 50 days. We assume that this kind of activity is a result of the same phone being activated multiple times. This should make sure that the customers are actually buying the devices for themselves. We also limit the maximum number of purchases per customer to 10 purchases to get rid of the possible phone dealers.

## 4.1 Case: All Platforms

First we see how the model fits a situation where we use all six device platforms together and use the sample size of 10000.

In figure 7 the purchase intervals are plotted as a histogram. The function is decreasing exponentially with most intervals within 500 days but there are some intervals even longer than 1000 days (3 years). The purchase intervals less than 50 days are omitted here as explained earlier. This mean that we have more active customers who are likely to purchase a new phone within 500 days than customers who make repurchases less frequently.

The customers are divided by the number of their repurchases and their numbers are shown in figure 8. Our model overestimates the number of customers who make 0, 2 and 3 or more purchases compared to the data. With our model the number of customers decreases as the number of repurchases increases. In this situation there are actually more customers who make one repurchase compared to those who do not make a repurchase.

In figure 9 cumulative purchases are plotted against the time. The model fits quite well to the actual data but it overestimates the number of purchases with long time period.

The estimated number of repurchases a new customer makes during one year (52 weeks) is 0.74 purchases. Parameters for the model in this case are

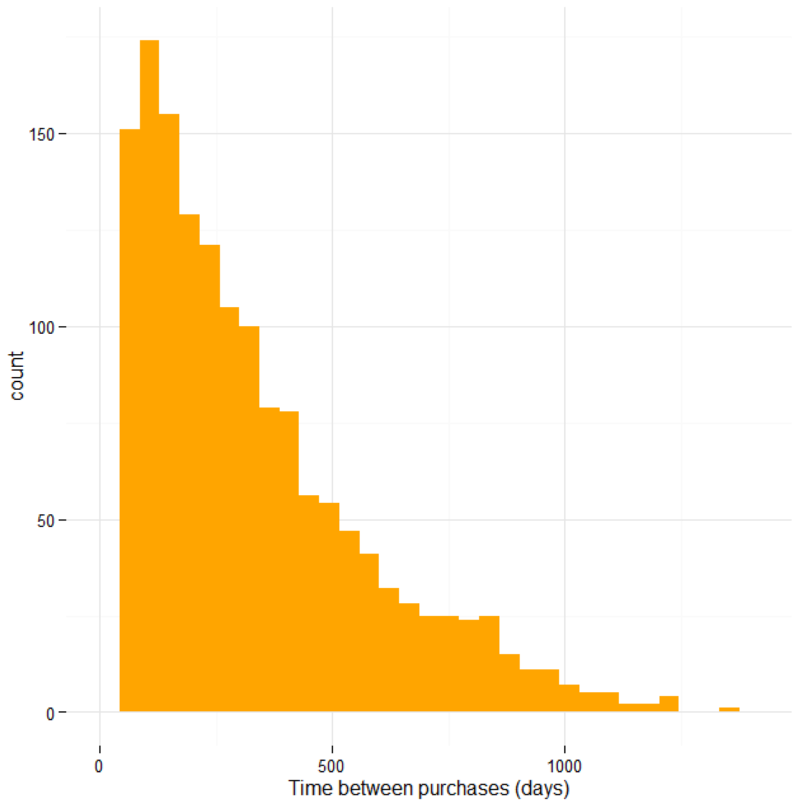


Figure 7: Histogram of purchase intervals for all the customers for all platforms

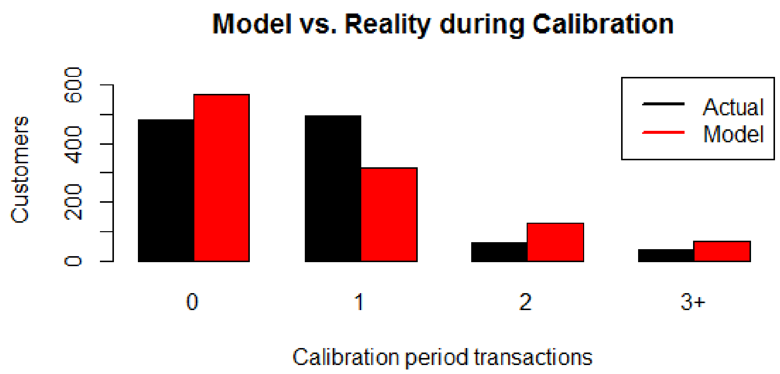


Figure 8: The actual number of repurchases plotted against the purchases predicted by the model for all platforms

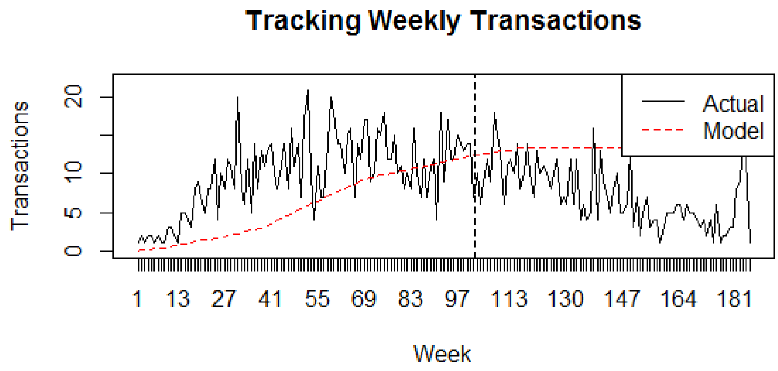


Figure 9: The actual transactions and the transaction predicted by the model by the week of occurrence for all platforms

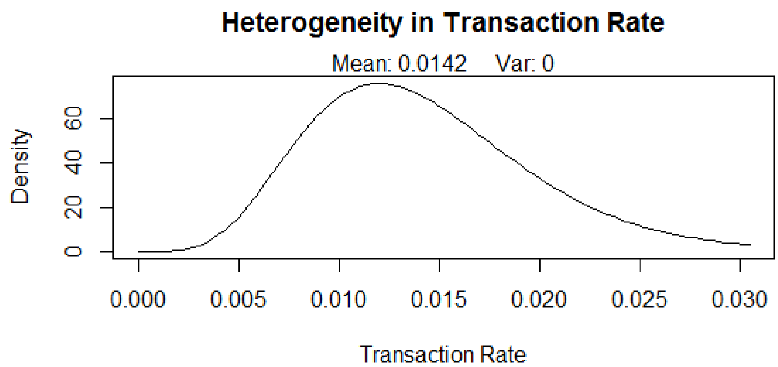


Figure 10: The distribution of parameter $\lambda$

$\beta = 6.41$, $s = 4.50 \times 10^2$, $r = 6.38 \times 10^{-6}$ and $\alpha = 1.61 \times 10^2$.

In figure 10 we see the estimated distribution of $\lambda$, which describes the heterogeneity in this dataset. Furthermore, you can clearly see a peak in the figure, that represents the behavior of an average customer.

## 4.2 Case: Single Platform

Now we try the model with only one platform that we consider the most relevant and use sample size of 10000. The estimated number of repuchases a new customer makes during one year (52 weeks) is 0.44 purchases in this case.

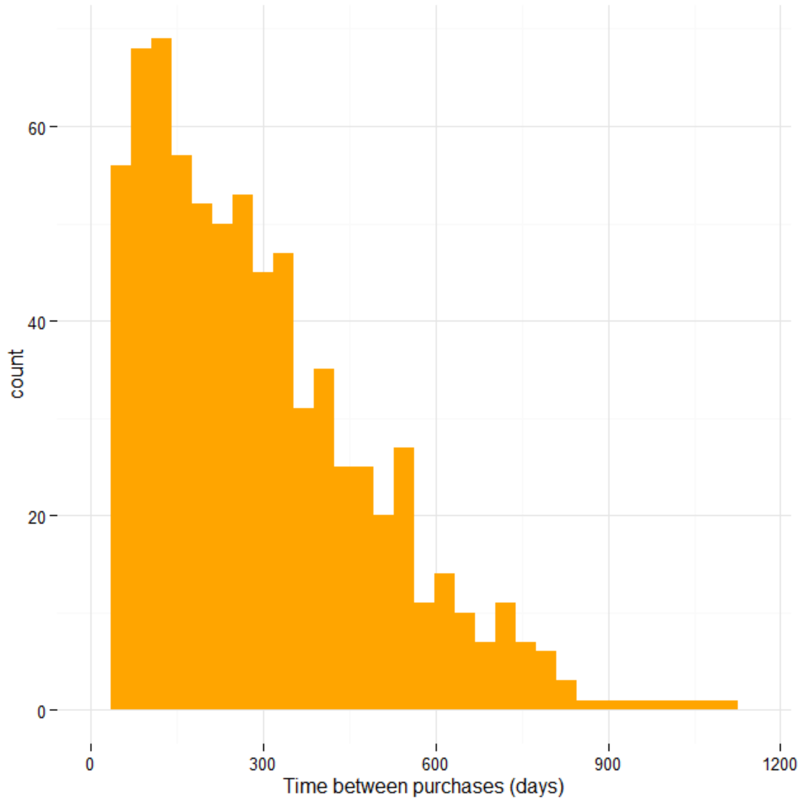


Figure 11: Histogram of purchase intervals for all the customers for a single platform

Now the estimated parameters are $\beta = 2.05$, $s = 2.45 \times 10^2$, $r = 2.10 \times 10^{-6}$ and $\alpha = 3.84 \times 10^2$.

The purchase intervals are plotted in figure 11 and the function is quite similar to the case with all platforms. When comparing the customers by the number of their repurchases the model seems to be very close to the actual data as seen in figure 12. However the model fit in figure 13 is not perfect for weekly transactions. One difference with this platform is that observated time period is shorter than for some other platforms. This might effect the results as customers are likely to have less repurchases in shorter time period. Especially several repurchases are rare.

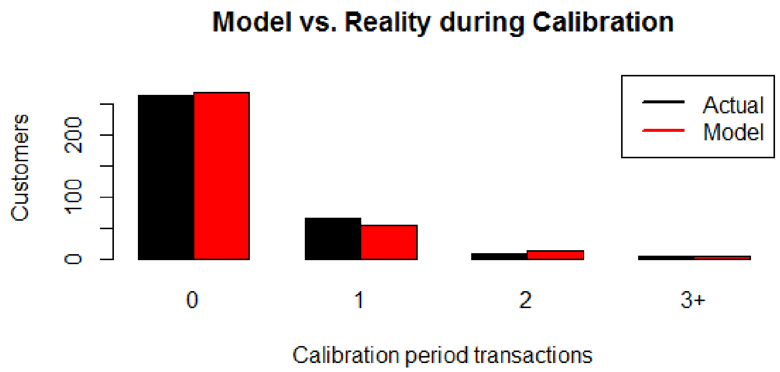


Figure 12: The actual number of repurchases plotted against the purchases predicted by the model for a single platform

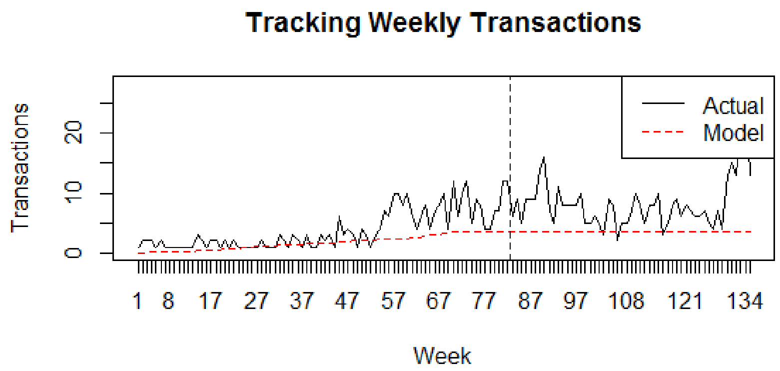


Figure 13: The actual transactions and the transaction predicted by the model by the week of occurrence for a single platform

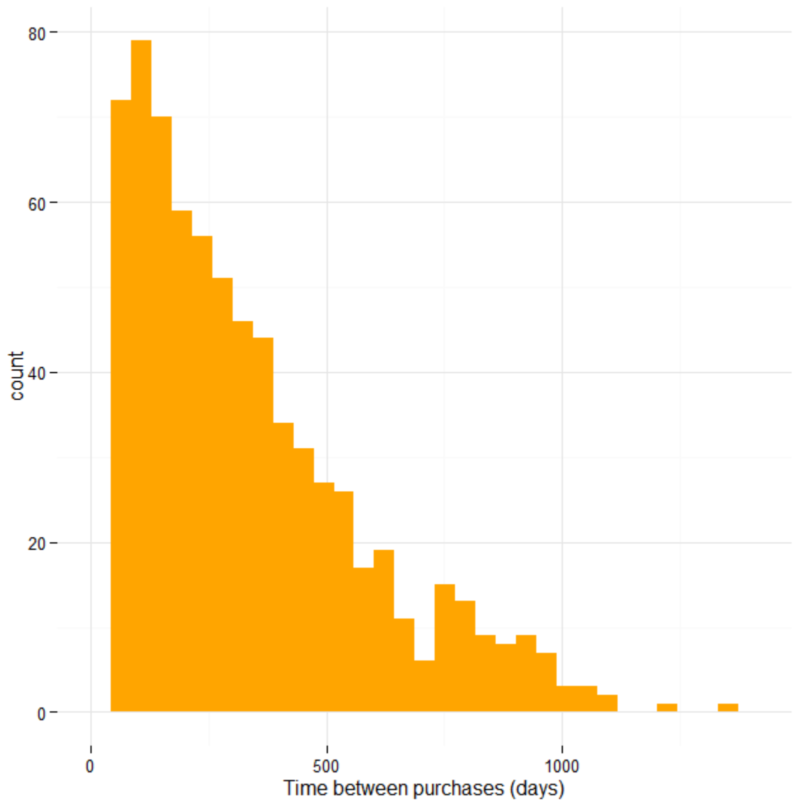


Figure 14: Histogram of purchase intervals for all the customers for all platforms with 5000 samples

### 4.3 Effect of Sample Size

To test what effect sample size has for our results we estimate the model for all platforms again using first only 5000 samples and then 50000 samples. The estimated parameters are $\beta = 2.14 \times 10$, $s = 1.58 \times 10^3$, $r = 1.27 \times 10^{-4}$ and $\alpha = 2.62 \times 10^1$. Number of estimated purchases is now 0.70 in one year for new customer, 0.04 slower than when using sample size of 10000.

The histogram of purchase intervals is plotted in figure 14 and again it is quite similar to previous results. In figure 15 the number of customers per repurchases is shown and the distribution is really similar with the case with 10000 samples. Our model predicts the number of customers to decrease as the repurchases increase which is not the case with real data for the first repurchase.

For the weekly transactions the outcome is again similar to the case with

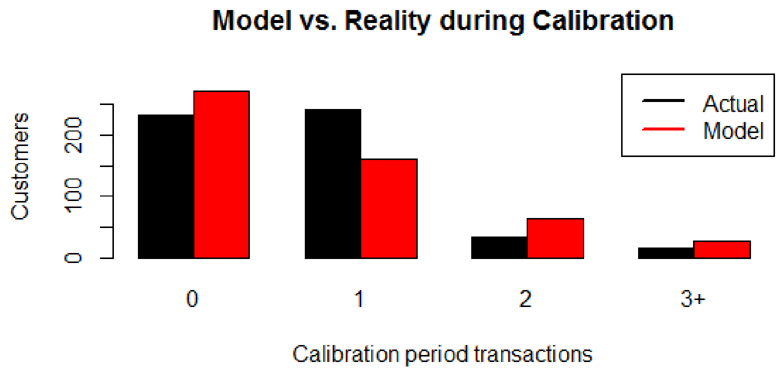


Figure 15: The actual number of repurchases plotted against the purchases predicted by the model for a single platform

10000 samples as seen in figure 16 but now the transactions are more evenly distributed over time. It seems that decreasing the sample size from 10000 to 5000 does not have strong effect on the outcome of our estimations.

Now we estimate the model for all platforms using sample size of 50000 which is 5 times the original sample size. The distribution of intervals is shaped as before but now it remains more of a logarithmic function as the sample size increases as seen in figure 17. The estimated parameters are now $\beta = 2.08$, $s = 1.28 \times 10^2$, $r = 5.10 \times 10^{-6}$ and $\alpha = 3.07 \times 10^2$. Now the expected repurchases for a new customer in one year is 0.84.

The number customers per repurchases is shown in figure18 and the weekly purchases in figure 19. The results are really similar to the case with sample size of 10000. It seems that increasing the sample size this much does not bring much of a benefit but requires more computing time.

The parameters from previous results are also shown in table 1.

Table 1: Parameters using different samples.

| | $r$ | $\alpha$ | $s$ | $\beta$ | -LL | $r/\alpha$ | s/$\beta$ |
|---|---|---|---|---|---|---|---|
| $n_{10k}$ | 6.41 | $4.50 \times 10^2$ | $6.38 \times 10^{-6}$ | $1.61 \times 10^2$ | -4147.9 | $1.42 \times 10^{-2}$ | $3.96 \times 10^{-8}$ |
| $n^*_{10k}$ | 2.05 | $2.45 \times 10^2$ | $2.10 \times 10^{-6}$ | $3.84 \times 10^2$ | -544.7 | $8.37 \times 10^{-3}$ | $5.47 \times 10^{-8}$ |
| $n_{5k}$ | 21.4 | $1.58 \times 10^3$ | $1.27 \times 10^{-4}$ | $2.62 \times 10^1$ | -1982.9 | $1.36 \times 10^{-2}$ | $4.83 \times 10^{-6}$ |
| $n_{50k}$ | 2.08 | $1.28 \times 10^2$ | $5.10 \times 10^{-6}$ | $3.07 \times 10^2$ | -22369.8 | $1.62 \times 10^{-2}$ | $1.66 \times 10^{-8}$ |

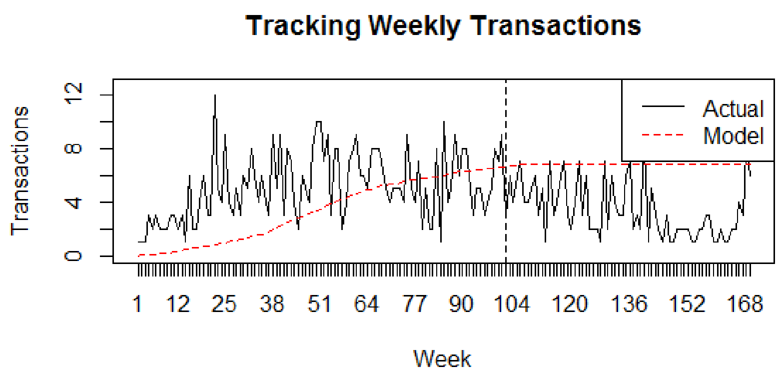


Figure 16: The actual transactions and the transaction predicted by the model by the week of occurrence for a single platform

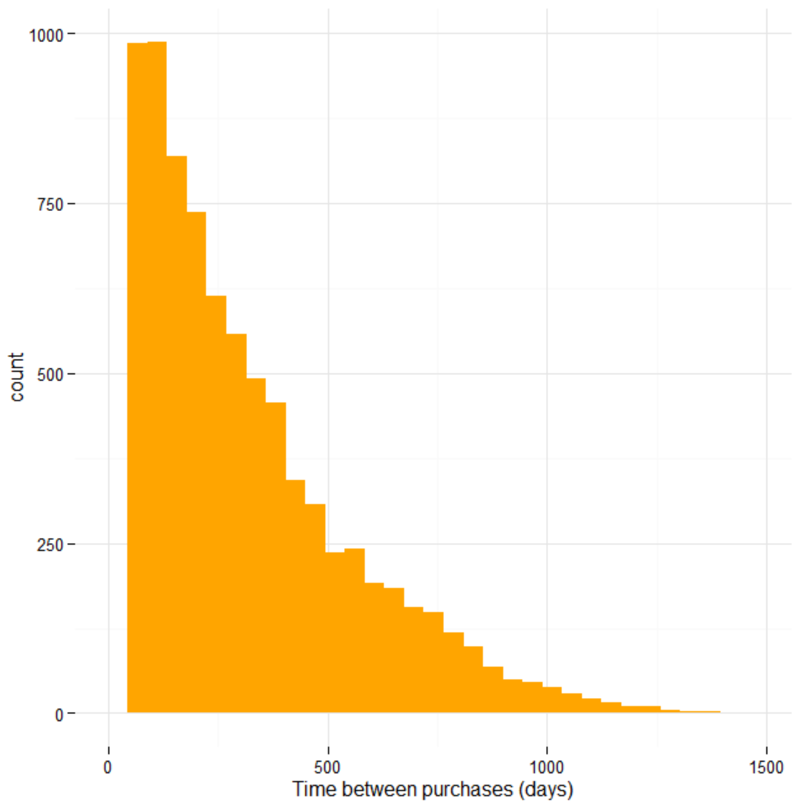


Figure 17: Histogram of purchase intervals for all the customers for all platforms with 50000 samples

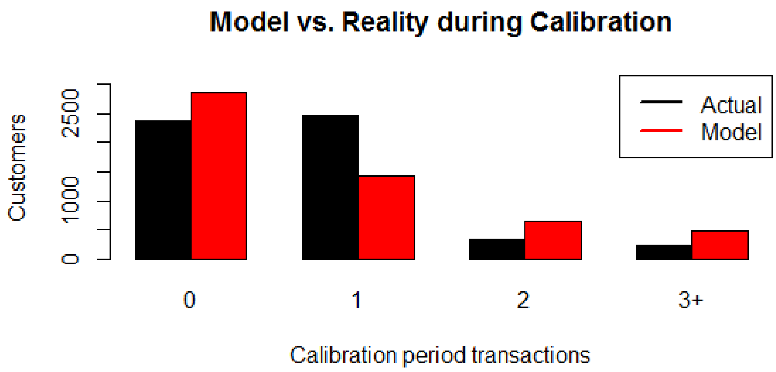


Figure 18: The actual number of repurchases plotted against the purchases predicted by the model for a single platform with 50000 samples

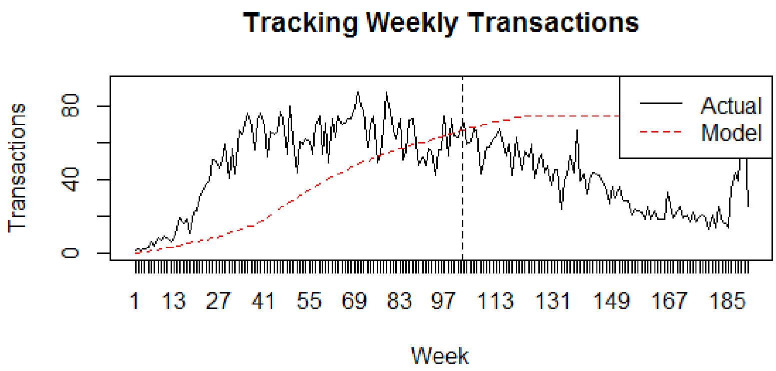


Figure 19: The actual transactions and the transaction predicted by the model by the week of occurrence for a single platform with 50000 samples

# 5 Conclusions

Our project showed that it is possible to model customer repurchase behavior to some point. However, the estimations require a large quantity of data from a long period. In the fast changing world of mobile phone markets it is not easy to gather data of repurchases as customers tend to change their mobile phone company when buying a new phone. Also the fact that the data is only from last few years and the customers are not buying phones often, makes it difficult to estimate repurchases. This model might work better with a customer database, where repurchases happen more often than once in a year on average. The privacy laws are also limiting the data that a company can gather from customers. Our Microsoft contact Lauri Salminen made it clear that Microsoft Mobile is following all these regulations, so that it compares favorably with its competitors.

The model we chose was far from perfect as it was often inaccurate and slow to compute with large datasets. It was difficult to estimate parameters that would make the model fit the data. This was due to the fact that the model was mathematically really complicated. However we were able to estimate the likelihood that a single customer will buy new phone in given time period which is a good result as this knowledge can be used for marketing purposes and especially for the timing of marketing emails.

The model also lacks the possibility to include the seasonality of the mobile phone markets. We did not see any seasonality in our data so we decided not to model it. With our model we can only estimate the time intervals of purchases but we should also consider when the purchases take place. Also, due to time constraints, we did not fully test if the model behaved differently in separate markets.

# 6 Appendix: Self-Evaluation

## 6.1 Summary of the Project

The target of our project was to find a model from academic literature that can give us information about customer repurchase behavior and then implement it to our client's data. We achieved this goal and found a suitable model that we implemented with R-language. The results are seen in this project and even if they are not perfect some of the results can be useful for marketing purposes.

## 6.2 Project Schedule

Our team was following the planned schedule well for most of the project. However, problems with certain critical tasks led to delays on later tasks. It was surprisingly difficult to find the correct academic literature. The very last part of our project was finished behind our own schedule but we still managed to finish our whole project before the deadline. Even tough the course had a really long time span of nearly four months the most important parts of our project were done during the final weeks. The first and interim reports took a lot of time to finish and it resulted our team to spend too much time on those instead of concentrating on the actual modeling. The amount of work was more than we expected at the beginning.

## 6.3 Lessons Learned

Our project was a real eye-opener for us as it made us realize how versatile applications mathematical models like Pareto/NBD can have and how they can yield business benefits. Working with large data had its own challenges but we managed to overcome them.

We also learned about project management. We learned how important the planning part is and also to analyze the risks beforehand to know what to to in case of failing something. Dividing tasks based on every team member's individual skills is important to utilize every member's skills so that we maximize the team performance.

### 6.4 Comments about the Course

The concept of the course was interesting as it enables us to get a catch on industry projects based on operation research. Every student was also able to choose a topic suiting his or her own interests and work on this project. However opposing other team's work did not bring much additional learning to the course as it was difficult to understand what was really going on with other projects. This was emphasized with interim reporting. Furthermore, it would be nice if the course followed the schedule of the periods. Many of the older students do not have any courses in the fifth period and hereby have started their summer jobs, so it is difficult to attend mandatory meetings.

# References


Tim Bedford and Roger Cooke. *Probabilistic Risk Analysis: Foundations and Methods.* Cambridge University Press, 2001.

Pradeep Chintagunta, Tülin Erdem, Peter E Rossi, and Michel Wedel. Structural modeling in marketing: Review and assessment. *Marketing Science*, 25(6):604–616, 2006.

Bas Donkers, Peter C Verhoef, and Martijn G de Jong. Modeling CLV: A test of competing models in the insurance industry. *Quantitative Marketing and Economics*, 5(2):163–190, 2007.

Lukasz Dziurzynski, Edward Wadsworth, Peter Fader, Elea McDonnell Feit, Daniel McCarthy, Bruce Hardie, Arun Gopalakrishnan, Eric Schwartz, and Yao Zhang. Package btyd. 2014.

Andrew SC Ehrenberg. *Repeat-buying: Theory and Applications.* North-Holland Amsterdam, 1972.

Peter S Fader, Bruce GS Hardie, and Ka Lok Lee. Counting your customers the easy way: An alternative to the Pareto/NBD model. *Marketing Science*, 24(2):275–284, 2005.

David C Schmittlein, Donald G Morrison, and Richard Colombo. Counting your customers: Who-are they and what will they do next? *Management Science*, 33(1):1–24, 1987.

Alan Stuard and Keith Ord. *Kendall's Advanced Theory of Statistics.* Griffin London, 1987.

Peter C Verhoef, Penny N Spring, Janny C Hoekstra, and Peter SH Leeflang. The commercial use of segmentation and predictive modeling techniques for database marketing in the netherlands. *Decision Support Systems*, 34 (4):471–481, 2003.

Florian von Wangenheim and Markus Wübben. *Analytical CRM: Developing and Maintaining Profitable Customer Relationships in Non-Contractual Settings.* Springer Science & Business Media, 2009.