ERASMUS UNIVERSITY ROTTERDAM

Erasmus School of Economics

Econometrics and Management Science

Business Analytics and Quantitative Marketing Master Thesis

Predicting Customer Lifetime Value using a Hierarchical Bayes Model combined with a Time-Varying Monetary Value on an Online Fashion Platform

Abstract

Every company wants to increase their profit and a well-known way to do so is by focusing more on the most profitable customers. These customers can be found by predicting the customer lifetime value (CLV), a value which reports the future revenue a customer brings to a company. This value can be predicted using a probability model to forecast the number of future transactions a customer will perceive and a prediction about the monetary value of these transactions. This research uses a hierarchical Bayes framework with a Markov chain Monte Carlo procedure to predict the transactions and combines it with the Pareto/Dependent model to predict a time-varying monetary value. This newly proposed combination is tested against combinations with the 'classical' Buy-Till-You-Die models, namely the Pareto/NBD and the gamma-gamma model. The data are from a Dutch online fashion platform, Winkelstraat.nl. The results showed that the newly proposed combination does not outperform the combination with the 'classical' models using this data set.

January 2021

Tobias van Emst

456546

Supervisor: Prof. dr. ir. R. Dekker

Second assessor: Prof. dr. D. Fok

Contents

1 Introduction

In the recent years, customers are preferring more and more to purchase clothing online over a purchase at the physical store. Some reasons for this behavior are that online shops have more product variety, are more convenient to visit and one can easily compare the same products at different stores (Jiang et al., 2013). This has led to an enormous increase in online shopping and as a result, an increase in the number of online shops. As more and more online retailers are present on the market, the competition has grown as well. Therefore, every company has to make the most out of its business strategy to compete with others on the e-commerce market.

In order to improve the business strategy, the main focus should be to increase the profits with as little as possible costs. By this matter, companies can keep growing and be a step ahead of their competitors. The best way to do so is by focusing more on the customer's needs, such that the customer satisfaction increases. Lots of firms make a distinction between customer acquisition and customer retention to focus on the right customer group. Customer acquisition is about obtaining new customers and customer retention's main goal is to keep customers at the company. Pfeifer (2005) concluded that firms should devote more attention and money to customer retention as the costs to retain a customer are much lower than the costs of bringing in new customers. This is often done by customer loyalty and brand loyalty initiatives. Other ways for online shops to satisfy customer demands is by giving free shipment on their products, giving lots of discounts or invest in customer support. Either way, every company is determined to allure customers to buy their products, such that the sales and the revenue increase.

As was mentioned before, online shops should search for the most profitable customers to gain the most revenue. A common way to increase a company's revenue is by improving their customer relationship management (CRM). CRM is an approach to manage a company's interaction with potential and existing customers. By focusing on data analyses of customer's purchase history, it can help improve business relationships with these customers and therefore resulting in a growth of the sales. Furthermore, CRM can learn a company about the customer needs and how to fulfill them by making use of, e.g., social media. Buttle and Maklan (2019) concluded that it can increase the company's profit by creating a long-term relationship with the most profitable customers.

An e-commerce company who also wants to improve their customer satisfaction is Winkelstraat.nl. Winkelstraat.nl is a Dutch online fashion platform with lots of items from the premium and luxury segment, as designer clothing, shoes, bags and other accessories. They give physical stores the opportunity to present their collection via a quick and efficient platform. By this matter, they expand the range of these boutiques and broaden their market. Over 300 unique boutiques are working together with Winkelstraat.nl, located over the Netherlands and West of Europe. One can shop over 600 popular brands. They started in October 2012 and have had an enormous growth in brand awareness since then[1].

As mentioned before, retaining customers is a top priority for online shops, but it can be quite difficult to measure. As Winkelstraat.nl is a platform that sells premium and luxury items, the prices of these products can be seen as quite high. This can make a customer who has bought multiple items during some time very valuable. On the other hand, a customer who buys less products, but at a higher price, can also be pointed out as a valuable customer. To focus on the right group of customers, this research wants to answer the following question: **Which customers will be the most profitable ones in the future for Winkelstraat.nl?** This is the main topic in this research. Once they can answer this question, they can allocate more of their marketing resources to the profitable customers in order to increase their revenue. Further, knowing who the less profitable customers are, Winkelstraat.nl can rethink their current marketing strategy for these customers in order to make them more satisfied, which would also boost their profits.

To answer the question above, the problem should be described in more detail, which can be seen in Section 2. After that, Section 3 gives a selection of the literature written about this subject. The models used in this research are given in Section 4. Section 5 reports the data that will be used and Section 6 describes the software implementation of the models and the data. Finally, the obtained results are given in 7, a discussion of these results in Section 8 and the conclusion of this research is given in Section 9.

2 Problem Description

As was mentioned before, CRM can increase a company's profit by creating a long-term relationship with the most profitable customers. In order to measure the impact and outcomes of the firm's CRM, the customer lifetime value (CLV) can be used. Berger and Nasr (1998) formulated the definition of CLV as follows: *a prediction of the net profit or loss attributed over the entire lifetime of transactions of a customer with the firm, i.e., the revenue the customer brings to a company subtracting the costs of attracting, selling and servicing that customer.* CLV can be insightful as it

[1]Follows from `https://www.winkelstraat.nl/over-ons`

can predict which customers are profitable and which ones are not, such that firms can focus on the most profitable ones.

Getting a prediction about the customer lifetime value (CLV) can be very useful in helping firms make strategic decisions, because it can help predicting and identifying who its best and worse customers will be in the future. Also, one can get to know their characteristics and retrieve which customers to go after in the long run. This can help quantify the relationship of the firm with its customers and subsequently allow the firm to make more informed decisions in a structured framework. Predicting CLV can also be helpful for companies by making tactical decisions, in a sense that it can help allocate short-term resources among marketing variables and focus on marketing activities. Marketing efforts are best directed at the most profitable customers. For instance, a company could direct an email with discounts on items to customers who are highly profitable to try and retain them and lower the marketing costs by not giving this discount to the lower profitable ones. Knowledge of CLV enables firms to develop customer-specific marketing programs leading to an increase in efficiency and effectiveness of such programs. Winkelstraat.nl focuses on the age group of 16-36 years old, so they are already using strategic decisions, but can improve this using the CLV of these particular groups.

To calculate CLV, one needs to predict the number of transactions a customer will purchase and what the expected monetary value of each transaction will be. A lot of literature has been written about how to make a prediction of these two measures. The most well-known, and therefore the 'classical' models, are the Pareto/NBD and gamma-gamma model to predict the number of transactions and the monetary value of these transactions, respectively. This research will make use of an extension of the Pareto/NBD model, namely the hierarchical Bayes framework, to predict the future transactions and the probability whether a customer will ever buy another item in the future. This framework relaxes some assumptions which are used in the classical model, and could therefore lead to an increases in the prediction performance. This framework will be combined with an extension on the gamma-gamma model, namely the Pareto/Dependent model, which can predict a time-varying monetary value of the future transactions. By making use of a time-varying monetary value instead of a static one, one could predict more flexible. This combination will eventually make a prediction about the customer lifetime value. This will be further explained in Section 3, where all models present in the literature will be discussed. In the end, this research wants to answer the following question: **Does a combination of a hierarchical Bayes framework with a time-varying monetary value better predict the customer lifetime value than**

the classical predictors?

By answering this question, the customer lifetime value can be predicted by the better predictors. By knowing this, the marketing department of Winkelstraat.nl can efficiently redirect the marketing costs per marketing channel on these customers and lay less focus on the less profitable ones. Take, for example, the group with the highest CLV level. If this particular group can be extended or their profits can be improved or the costs for this group can be redirected more efficiently, Winkelstraat.nl's revenue would improve. The combination of the hierarchical Bayes framework with the time-varying monetary value from the Pareto/Dependent model is new to the literature, so this paper will contribute to the current marketing research literature.

3 Literature

Nowadays a lot of papers have been written about customer lifetime value and its formulations. Variation of these formulations can be enormous, depending on the specific task, the availability of the data at hand and how the user wants to use them. Dwyer (1997) firstly explained the different CLV formulations in words, where after Berger and Nasr (1998) made these words more abstract by using formulas to express the customer lifetime value. Eventually, these formulas are expressed more clearly by Reinartz and Kumar (2000), Thomas (2001) and Jain and Singh (2002) who reported the following formula to predict the customer lifetime value (CLV).

$$\text{CLV}_i = \sum_{t=0}^{T} \frac{r_{it} - c_{it}}{(1+d)^t} \ , \tag{1}$$

where i denotes the index which can be a customer or a segment of customers, r_{it} the revenue obtained from customer i at time t, c_{it} the direct costs generating r_{it}, d the pre-determined discount rate and T the time period used for estimating the CLV.

This basic formula neglects some important understandings of the relationship between the customer and the company. This is important, as the CLV formula differ critically between differences in this relationship. It is important to distinguish a contractual and a non-contractual customer-company relationship, i.e., whether a customer has a legal relationship with the company, such that one can observe whether the customer becomes inactive. Further, CLV formulas can differ on whether purchases made by a customer are made in discrete or continuous points in time, so whether purchases can be made on strict points or at any point in time.

To obtain the CLV formula that fits the best for Winkelstraat.nl, one needs to decide how the customer-company relationship looks like. Firstly, they have a non-contractual relationship, i.e., no legal relationship between the customer and the company, such that one cannot observe when a customer becomes inactive. A customer can stop buying products from Winkelstraat.nl at any time without them knowing it, which makes it non-contractual. Secondly, Winkelstraat.nl is continuous in time, i.e., orders can take place at any point in time. This is certainly the case for online retailers, where the customer can buy at any point in time he or she wants. Reinartz and Kumar (2003) and Gupta et al. (2004) implemented these customer-company relationship restrictions, which resulted in the following CLV formula

$$E(\text{CLV}_i) = \sum_{t=1}^{T} \frac{(r_{it} - c_{it})p_{it}}{(1+d)^t} \; - \; \text{AC}_i, \tag{2}$$

where p_{it} indicates the probability that customer i is active at time t (retention rate) and AC_i shows the acquisition costs for customer i. The summation has an upper bound, namely the time horizon T. When we say the 'lifetime' of a customer, we must incorporate the fact that customers can buy at any point in time, and therefore T should be set to infinity. In order to make predictions about the customer lifetime value, the T can be a finite time, as one wants to predict over a specific time frame. Further, this formula results in a prediction of the CLV, not the true value. That is why the term CLV_i is changed into $E(CLV_i)$. Finally, the beginning of the summation is changed to $t = 1$, as the first prediction is made in the future. The term $(r_{it} - c_{it})$ is also known as the customer margin. This margin, the customer retention (p_{it}) and customer acquisition (AC_i) can be modelled separately and in the end combined to compute CLV.

The customer margin can be seen as the revenue a customer brings to the company. This amount can be retrieved by multiplying the number of transactions a customer purchases with the monetary value of these bought items (Jasek et al., 2018). So, to make a prediction about the revenue a customer would bring to the company in the future, a prediction about the future transactions and the monetary value of those transactions should be made. The models used to retrieve a prediction of the number of transactions and to get the probability of a customer to still be active in the future (retention rate) are explained in Section 3.1. The expected monetary value of these transactions can also be modelled and will be discussed in Section 3.2. The acquisition costs are neglected here, as many firms (including Winkelstraat.nl) do not have these costs specifically for each customer.

3.1 Number of Transactions

As was said before, to calculate the customer lifetime value one needs to use the customer margin, which can be split up in the number of transactions and the monetary value of these transactions. This Section focuses on the calculation of the number of transactions and the probability of being active (retention rate). These measures can be modeled by probability models. Ehrenberg (1959) was one of the first to model repeat buying behavior. His probability model assumed that the purchases made by a customer are randomly around their individual-specific, time-invariant purchase rate. Further, he assumed that this purchase rate differs per customer. These assumptions can be modeled using a Poisson process with a gamma distributed purchase rate, which results in a negative binomial distribution (NBD) model.

The assumption about time-invariant purchase rates usually does not hold, as customers may purchase more products when they earn more money, e.g., when a student gets his/her first job. To allow for time-variant purchase rates, Schmittlein et al. (1987) suggested to use a Pareto/NBD model. This model assumes that customers are 'alive' (making purchases) when they purchased their first product and after an unobserved period of time they 'die' (become inactive). During the time a customer is alive, his/her purchase behavior is characterized by the NBD model (customer transaction is modeled by a Poisson distribution). The unobserved period of time a customer stays alive is treated as random and is modelled using an exponential distribution. The heterogeneity in the drop-out rates across customers can be computed using a gamma distribution. A gamma mixture of exponentials is better known as the Pareto distribution, that is why the complete model is known as the Pareto/NBD model. It turns out that the only information required to make a prediction about a customer's future number of purchases are the recency (when the customer's last transaction occurred) and frequency (how many purchases this customer made during the time he/she was alive). This model is applicable in contexts where the time when a customer becomes inactive is unknown to the analyst (non-contractual) and the customer can make any number of purchases at any time and can die at any time (continuous). This is also known as a non-contractual continuous customer-company relationship.

Despite the fact that this Pareto/NBD model originates from 1987, it is still commonly used by researchers (e.g., Avinash et al. (2019), Xie and Huang (2020)). This model has been extended in a number of directions. A well-known extension is the one suggested by Reinartz and Kumar (2000, 2003). They changed the probability that a customer is alive into a dichotomous 'alive' or 'death' measure. When a customer has left the relationship is approximated using the time the customer first entered the relationship and a threshold model. The time between this cut-off value and the first time he or she was 'alive' then shows the lifetime of this customer. By using this cut-off value, the lifetime is finite for each customer, which makes the analysis much easier. However, the authors concluded that the data set they used (3 years of transactional data) was not enough to get valuable insights into the phenomenon. Moreover, they concluded that each customer has a different buying frequency, such that by taking a short transactions time frame, the profitable customers can be pointed out as unprofitable ones.

Although the Pareto/NBD model has been used in a lot of researches, it has rarely been used in 'action'. The main reason for this is that the parameter estimation is computational challenging as the likelihood function to estimate these parameters is seen as quite daunting (Fader and Hardie, 2005). To overcome this problem, Fader et al. (2005b) proposed a variant of the Pareto/NBD model, namely the beta-geometric/NBD (BG/NBD) model. Changing the fact that a customer can 'die' at any time, the BG/NBD model assumes that a customer can become inactive after any transaction with a certain probability and the heterogeneity in dropout probabilities across customers is captured by a beta distribution. This results in a model that is easier to implement, e.g., with Microsoft Excel one can easily obtain the desired parameters. The BG/NBD model offers therefore a much straightforward parameter estimation process with almost no loss in the model's fit and forecasting performance. By this matter, this model can be seen as an attractive alternative. Although this model can be much easier implemented, it makes heavy buyers more likely to 'die', as they have more transactions and therefore more opportunities to become inactive. Furthermore, a drawback of this model is that it assumes that customers with no repeat purchase still remain active. This leads to performing worse at estimating inactive customers compared with other models which predict CLV (Jerath et al., 2011). The BG/NBD model itself has a few extensions, e.g., Mzoughia and Limam (2014), who implemented a COM-Poisson instead of a Poisson distribution to model purchase behavior. To better explain the differences between customers and therefore better predict the parameters, Fader et al. (2007) implemented time-invariant covariates into the Pareto/NBD and the BG/NBD models. Unfortunately, this requires very difficult computations, so this will not be used in this research.

Another extension on the basic Pareto/NBD model (and on the BG/NBD model) was made by Jerath et al. (2011), who presented a variant on the assumption that a customer's unobserved 'death' could occur at any point in time. They stated that this 'death' can only occur at discrete points in calendar time, which they call periodic death opportunity (PDO) model. When the time period after which the customer makes his/her dropout decision is very small, the PDO model converges to the Pareto/NBD model and when it is longer, the dropout process is 'shut off' and the PDO model converges to the basic NBD model. When the manager's main focus is the customer dropout, the PDO framework deserves some attention (Korkmaz et al., 2013). On the other hand, Jerath et al. (2011) encourage using the Pareto/NBD model when the manager's primary goal is forecasting purchases, as the data set they used is rather specific. As this paper wants to predict the customer lifetime value, a huge focus should be to forecast the number of purchases as accurate as possible. Therefore, this model would not apply as a useful extension.

Abe (2009), Singh et al. (2009) and Ma and Liu (2007) showed that the use of a hierarchical Bayes framework extends the Pareto/NBD model for customer-base analysis. The model presented in Abe (2009) uses the same assumptions as the classical Pareto/NBD model, but relaxes the independence assumption of the purchase and death rate. It is assumed that the two processes follow a bivariate log-normal distribution. Korkmaz et al. (2013) showed that when there is a non-zero correlation between the processes, and thus when the assumption is violated, the hierarchical Bayes model outperforms the other models in unbiased prediction performance. Furthermore, this model can also incorporate customer characteristics as covariates. When the other assumptions given in the hierarchical Bayes model assumptions hold, the lifetime predictions accurately predict the actual values (Korkmaz et al., 2014). This model can give a prediction about the lifetime of a customer and about the number of future purchases a customer would make. A drawback of this approach is that it is computationally expensive. This model makes use of a Markov chain Monte Carlo simulation, and in order to converge to the right parameter values, a lot of draws should be taken (Korkmaz et al. (2014) used 200.000 iterations).

3.2 Monetary Value

As stated before, in order to calculate the customer margin, the predicted number of future purchases can be used. However, this prediction is not the same as the customer margin. In the beginning of this Section it was stated that the customer margin also depends on the monetary value of the items purchased, as one does not know the exact amount a customer would spend per item. This problem is addressed by Schmittlein and Peterson (1994), who created a normally distributed submodel which predicts the average future monetary value per customer. Using the assumption that the number of future purchases and the average purchase value per customer are independent, they stated that these values can be multiplied to obtain an approximation of the total future spending of customers, and therefore an approximation of the customer margin. The assumption about a normally distributed submodel has been replaced by Fader et al. (2005a). They claim that this distribution is not bounded from below by zero and it is a symmetric spend distribution. As spend data tend to be right skewed, Fader et al. (2005a) suggested a gamma-gamma model.

Fader et al. (2005a) also concluded, using their data set, that the correlation between the number of future transactions and the average spend per transaction has an average value of 0.06 with a p-value of 0.08. However, they assumed that the number of transactions and their monetary value are independent. This contradicts each other and therefore this assumption is violated. This can be logically explained; when a customer buys more products, the monetary value per product has to drop, to maintain the same spending behavior. To deal with this violation, Glady et al. (2009) suggested to drop this assumption to further increase the accuracy of the prediction of the CLV and came up with the Pareto/Dependent model. They claim that it does not require complex adjustments and when the correlation is significantly present, their model works better than the gamma-gamma model. This model uses some features from the Pareto/NBD model and the gamma-gamma model for the prediction of the monetary value. Furthermore, the monetary value obtained by this model is time-varying, which could increase the predictions even more.

3.3 Conclusion

In the end, lots of papers have been written about predictions of customer lifetime value (CLV) and its applications. By taking into account the customer-company relationship of Winkelstraat.nl, this paper will predict the CLV as stated in Equation 2. To measure this value, one has to look at the retention rate and customer margin, as these measures should be modelled using probability models. This research will evaluate the retention rate by making use of the Pareto/NBD model and an extension of this model, namely the hierarchical Bayes framework of Abe (2009) and will include customer characteristics as covariates. We expect that the latter will outperform the basic Pareto/NBD model, as it relaxes the independence assumption between the transaction rate and the death rate.

To calculate the customer margin, this research needs to know the future average monetary value per item for each customer. This will be modelled using the gamma-gamma model suggested by Fader et al. (2005a) and by the Pareto/Dependent model given by Glady et al. (2009), which extends the first model by relaxing the assumption of a customer's spending per transaction to be independent of his/her transaction flow. Further, this model results in a monetary value which varies over time.

Finally, the future number of purchases obtained by the Pareto/NBD model and the hierarchical Bayes model will be combined with the predicted monetary value of these purchases obtained by the gamma-gamma model and the Pareto/Dependent model. A combination of the hierarchical Bayes framework with the Pareto/Dependent model has not yet been done in the marketing literature, so this research will be the first to perform this.

4 Methodology

This Section will explain the models which will be used in this research. First the models to predict the number of transactions will be given, the Pareto/NBD model in Section 4.1 and the hierarchical Bayes framework in Section 4.2. After that, the models to predict the monetary value of a transaction will be given, the gamma-gamma model in Section 4.3 and the Pareto/Dependent model in Section 4.4. The prediction performance of these models will be evaluated and this is explained in Section 4.5. Finally, the usefulness of these models based on the literature is given in Section 4.6.

4.1 Pareto/NBD model

To obtain the input measures needed to calculate the customer lifetime value (CLV), namely retention rate and customer margin, this research will make use of the Pareto/negative binomial distribution (NBD) model. This subsection will explain this model by following the research performed by Schmittlein et al. (1987). This model has a convenient feature, namely that the only input needed are the recency, frequency and the observation period. Recency is a measure of how recently a customer made a purchase, frequency is a measure of how often the customer purchases during the observation period, excluding the first purchase. For each customer, these measures are written as $(X,\ t,\ T)$, where X is the frequency, t stands for the recency and T is the complete observation period. By making use of these symbols, the Pareto/NBD model can be explained by the following six assumptions:

Assumption 1. *Customers go through two stages in their 'lifetime'. 1) They are alive for some period of time and 2) then become permanently inactive.*

Assumption 2. *While the customer is alive, the number of transactions made by a customer follows a Poisson process with transaction rate λ; therefore, the probability of observing x transactions in the time interval (0, t] is given by*

$$P(X(t) = x \,|\, \lambda,\, t,\, T) = \frac{(\lambda\, t)^x\, e^{-\lambda t}}{x!}. \tag{3}$$

This is equivalent to assuming that the time between transactions is distributed exponential with transaction rate λ,

$$f(t_j - t_{j-1} \,|\, \lambda) = \lambda\, e^{-\lambda(t_j - t_{j-1})}, \tag{4}$$

where t_j is the time of the jth purchase.

Assumption 3. *A customer's unobserved 'lifetime' of length τ (after which he or she is inactive) is exponentially distributed with dropout rate μ:*

$$f(\tau \,|\, \mu) = \mu\, e^{-\mu\tau}. \tag{5}$$

Assumption 4. *Heterogeneity in transaction rates across customers follows a gamma distribution:*

$$g(\lambda \,|\, r;\, \alpha) = \frac{\alpha^r\, \lambda^{r-1}\, e^{-\lambda\alpha}}{\Gamma\,(r)}. \tag{6}$$

Assumption 5. *Heterogeneity in dropout rates across customers follows a gamma distribution:*

$$g(\mu \,|\, s;\, \beta) = \frac{\beta^s\, \mu^{s-1}\, e^{-\mu\beta}}{\Gamma\,(s)}. \tag{7}$$

Assumption 6. *The transaction rate λ and the dropout rate μ vary independently across customers.*

Assumptions 2 and 4 give us the NBD model for the distribution of the number of transactions while the customer is alive:

$$\begin{aligned} P(X(t) = x \,|\, r,\, \alpha) &= \int_0^\infty P(X(t) = x \,|\, \lambda)\, g(\lambda \,|\, r,\, \alpha)\, d\lambda \\ &= \frac{\Gamma\,(r + x)}{\Gamma\,(r)\; x!} \left(\frac{\alpha}{\alpha + t}\right)^r \left(\frac{t}{\alpha + t}\right)^x. \end{aligned} \tag{8}$$

Assumptions 3 and 5 give us the Pareto distribution of the second kind, which report the prediction whether a customer is still alive:

$$\begin{aligned} f(\tau \mid s, \beta) &= \int_0^\infty f(\tau \mid \mu)\, g(\mu \mid s,\, \beta)\, d\mu \\ &= \frac{s}{\beta}\Big(\frac{\beta}{\beta+\tau}\Big)^{s+1}, \text{ and} \\ F(\tau \mid s,\, \beta) &= \int_0^\infty F(\tau \mid \mu)\, g(\mu \mid s,\, \beta)\, d\mu \\ &= 1 - \Big(\frac{\beta}{\beta+\tau}\Big)^{s}. \end{aligned} \tag{9}$$

The NBD and Pareto labels for each of the submodels naturally lead to the name of the integrated model, namely the Pareto/NBD model.

As one can see, the two models need the parameters r, α, s and β. Schmittlein et al. (1987) suggested that these parameters can be obtained by maximizing the likelihood for the observed transaction data of the customers. This likelihood function is given as

$$\begin{aligned} L(r,\, \alpha,\, s,\, \beta) &= \prod_{i=1}^{N} P(X_i = x_i,\, t_i,\, T_i \mid r,\, \alpha,\, s,\, \beta) \\ &= \frac{\Gamma(r+x)\alpha^r\beta^s}{\Gamma(r)}\Big(\frac{1}{(\alpha+T)^{r+x}(\beta+T)^s} + (\frac{s}{r+s+x})A_0\Big), \end{aligned} \tag{10}$$

where

$$A_0 = \frac{F(a_i, b; c_i; z(t_i))}{(\alpha+t_i)^{r+s+x_i}} - \frac{F(a_i, b; c_i; z(T_i))}{(\alpha+T_i)^{r+s+x_i}}, \tag{11}$$

with

$$\begin{aligned} a_i &= r + x_i + s; \qquad & b &= s+1; \\ c_i &= r + x_i + s + 1; \qquad & z(y) &= \frac{\alpha-\beta}{\alpha+y}; \\ F(a; b; c; z) &= \sum_{j=0}^{\infty} \frac{(a)_j (b)_j}{(c)_j}\frac{z^j}{j!}, \end{aligned}$$

where F is the Gaussian hypergeometric function (Abramowitz and Stegun, 1970) and $(a)_j$ is the Pochhammer's symbol which is a rising factorial and is further explained in Srivastava et al. (2014). The derivation of equation 10 can be seen in Appendix A.1.1 and follows Fader and Hardie (2005). Maximizing this equation give us the parameter values which can be used to calculate the number of future transactions for a customer and whether this customer is still alive.

The probability that customer i is alive, conditional on the data, is given as

$$P(alive|x_i, t_i, T_i) = 1\Big/\Big(1 + \frac{\hat{s}}{\hat{r}+\hat{s}+x}(\hat{\alpha}+T_i)^{\hat{r}+x_i}(\hat{\beta}+T_i)^{\hat{s}}A_0\Big), \tag{12}$$

where the derivation of this equation can be found in Appendix A.1.2. The value of A_0 is given in Equation 11.

The Pareto/NBD model also provides a formula for the (unconditional) expected number of transactions customer i will make up to time k.

$$E(x_{ik}) = \frac{\hat{r}\hat{\beta}}{\hat{\alpha}(\hat{s}-1)}\Big[1 - \Big(\frac{\hat{\beta}}{\hat{\beta}+k}\Big)^{\hat{s}-1}\Big]. \tag{13}$$

This derivation is stated in Appendix A.1.3. However, to get a prediction of the CLV value of customer i, one needs to estimate the conditional expectation of the number of transactions a customer will make up to time $T_i + k$, given his/her transaction data. This is formulated as follows

$$\begin{aligned} E(x_{i,T_i+k}|x_i, T_i, t_i) = \hat{x}_{i,T_i+k} = x_i + &\frac{\Gamma(\hat{r}+x_i)\hat{\alpha}^{\hat{r}}\hat{\beta}^{\hat{s}}}{\Gamma(\hat{r})(\hat{\alpha}+T_i)^{\hat{r}+x_i}(\hat{\beta}+T_i)^{\hat{s}}\hat{L}_i} \\ &\times \frac{(\hat{r}+x_i)(\hat{\beta}+T_i)}{(\hat{\alpha}+T_i)(\hat{s}-1)}\Big[1 - \Big(\frac{\hat{\beta}+T_i}{\hat{\beta}+T_i+k}\Big)^{\hat{s}-1}\Big], \end{aligned} \tag{14}$$

where $\hat{L}_i = L(\hat{r}, \hat{\alpha}, \hat{s}, \hat{\beta}|x_i, t_i, T_i)$ is the estimated likelihood function from Equation 10. Therefore, the expected number of future transactions in the k-th time period is given as $\hat{x}_{i,T_i+k} - \hat{x}_{i,T_i+k-1}$. The derivation of this equation can be found in Appendix A.1.3.

4.2 Hierarchical Bayes framework

This research will continue with an extended version of the Pareto/NBD model. Abe (2009) suggested to use a hierarchical Bayes (HB) framework which alters the heterogeneity assumptions on the purchase and dropout rate (Assumptions 4 and 5 from Section 4.1) and drops the independence assumption between them (Assumption 6 in Section 4.1). They do so by introducing a new assumption and keeping the first 3 assumptions from Section 4.1. The new assumption is as follows.

Assumption 7. *Individuals' purchase rate λ and dropout rate μ follow a multivariate log-normal (MVN) distribution. This makes sure that a correlation between the purchase and dropout processes is permitted. It can be mathematically expressed as*

$$\begin{bmatrix} log(\lambda) \\ log(\mu) \end{bmatrix} \sim MVN\left(\theta_0 = \begin{bmatrix} \theta_\lambda \\ \theta_\mu \end{bmatrix}, \Gamma_0 = \begin{bmatrix} \sigma_\lambda^2 & \sigma_{\lambda\mu} \\ \sigma_{\mu\lambda} & \sigma_\mu^2 \end{bmatrix}\right). \tag{15}$$

Further, the HB framework will use a Markov chain Monte Carlo (MCMC) simulation to estimate the model parameters instead of estimating them analytically. A convenient advantage of performing no analytical estimation is that the model can now obtain time-invariant covariates. To model these covariates, this research will perform a linear regression as follows, where index i indicates customer i:

$$\begin{bmatrix} log(\lambda_i) \\ log(\mu_i) \end{bmatrix} = \theta_i = \beta' d_i + e_i \, , \text{ where } e_i \sim MVN(0, \Gamma_0), \tag{16}$$

where d_i is a vector that contains K characteristics of customer i and β is a parameter vector. If d_i equals 1 (only an intercept), the formula becomes Equation 15.

The parameters that this model will estimate are the purchase rate λ and dropout rate μ. In order to estimate these parameters, Abe (2009) suggested to use two unobserved latent variables. z is a dummy parameter which equals 1 if a customer is active at time T and 0 otherwise. The other latent variable is τ, which denotes the dropout time when $z = 0$. If both these variables are known, one could estimate the likelihood function for the transaction data. One should make a distinction in whether the customer is active at time T $(z = 1)$ or not.

$$\begin{aligned} &P(\text{xth purchase at } t \text{ and active until } T \text{ and no purchase between } [t, T]) = \\ &P(X = x, t, T | \, \lambda, \mu, z = 1) \times P(z = 1 | \, \mu, \lambda) = \frac{\lambda^x \, t^{x-1}}{\Gamma(x)} e^{-(\lambda+\mu)\, T}, \end{aligned} \tag{17}$$

$$\begin{aligned} &P(\text{xth purchase at } t \text{ and active until } \tau \text{ and no purchase between } [t, \tau]) = \\ &P(X = x, t, T | \, \lambda, \mu, z = 0) \times P(z = 0 | \, \mu, \lambda) = \frac{\lambda^x \, t^{x-1}}{\Gamma(x)} \mu e^{-(\lambda+\mu)\tau} (t \leq \tau \leq T). \end{aligned} \tag{18}$$

These formulas are based on Assumption 3, which states that the customer's lifetime is characterized by an exponential distribution. The above two cases can be combined to a more compact notation for the likelihood function.

$$L(x, t, T | \, \lambda, \mu, z, \tau) = \frac{\lambda^x \, t^{x-1}}{\Gamma(x)} \mu^{1-z} e^{-(\lambda+\mu)\, (zT + (1-z)\tau}. \tag{19}$$

Unfortunately, neither z nor τ is known, so we have to treat them as missing data and apply a data augmentation technique. In order to simulate z in the MCMC procedure, a new expression for the probability that a customer is active at time T has to be evaluated.

$$P(\tau > T | \, \lambda, \mu, T, t) = P(z = 1 | \, \lambda, \mu, T, t) = \frac{1}{1 + (\frac{\mu}{\lambda\mu}) \, [e^{(\lambda+\mu)\, (T-t)} - 1]} \tag{20}$$

Using the MCMC procedure requires to have priors. Equation 16 shows that the prior for λ and μ is chosen to be lognormal. The parameters for this lognormal specification are β and Γ_0.

These parameters are also estimated in a Bayesian manner with two priors. β follows a multivariate normal distribution ($\beta \sim MVN(\beta_0, \Sigma_0)$) and Γ_0 an inverted Wishart ($\Gamma_0 \sim IW(v_{00}, \Gamma_{00})$). Many Bayesian models choose the priors to be uninformative, such that these priors would not affect the posterior results. However, as Korkmaz et al. (2014) showed, setting a very diffuse prior on Γ_0 (i.e., having a large variance) has a major impact on the posterior distribution of behavioral parameters and leads to unstable estimates. In order to incorporate this information, we will use the following conjugate prior specification: β_0 is a null matrix of K by 2, Σ_0 a K by K diagonal matrix with 0.01 on the diagonal, the degrees of freedom v_{00} = 3 + K and $\Gamma_{00} = v_{00}\ I_2$.

Now we have all the information needed to obtain estimates for the parameters $(\theta_i, \tau_i, z_i \forall i; \beta, \Gamma_0)$ by making use of the MCMC procedure. In order to estimate the joint density, each parameter is sequentially generated given the remaining parameters from its conditional distribution until there appears convergence. The procedure used in this research is slightly different from the one used in Abe (2009) [2]. This research uses the procedure which is presented in Korkmaz et al. (2013):

(0) Set initial value for θ_i^0 $\forall$ i using priors β and Γ_0.

(1a) Generate $[z_i|\, \theta_i]$ according to Equation 20.

(1b) If $z_i = 0$, generate $[\tau_i|\, z_i,\, \theta_i]$ using a truncated exponential distribution with parameter $\lambda_i + \mu_i$ and truncation such that $t_i < \tau_i < T_i$.

(2) Generate $[\beta,\, \Gamma_0|\, \theta_i,\, \forall i]$ using a standard multivariate normal regression update (see Rossi and Allenby (2003)).

(3) Sample $[\theta_i \,|\, z_i,\, \tau_i,\, x_i,\, T_i \,\forall i,\, \beta,\, \Gamma_0]$ using a random-walk Metropolis-Hastings sampler. The likelihood function used in this sampler is given in Equation 19.

(4) Iterate (1) $\sim$ (3) until convergence is achieved.

After convergence, simulation draws for (λ, μ) can be obtained for every customer. These values can be implemented in Assumptions 2 and 3 to obtain the remaining lifetime and the number of future purchases.

[2] In the paper from Abe (2009) step 4 in Section 4.5 on page 544 must be altered.

4.3 Gamma-Gamma model

By using one of the two aforementioned models, the researcher can obtain the predicted number of future purchases. As was mentioned in Section 3.2, the customer margin does not equal the number of future purchases, as it is not known how much a customer will spend per purchase. Fader et al. (2005a) suggested to use a gamma-gamma model, which can predict the average future purchase value of the transactions per customer. If this value is multiplied with the predicted number of future purchases, the customer margin is obtained.

Every customer has different number of transactions, which we call frequency, and is denoted as x (subscript i, indicating customer i, is removed for simplicity). Let z_1 ... z_x denote the profit of each observed transaction for a specific customer. An assumption should be made about the fact that the value of each transaction varies randomly around the customer's unobserved mean transaction value. Further, we denote the customer's average observed transaction value to be $\bar{z}$, where $\bar{z} = \frac{1}{x}\sum_{j=1}^{x} z_j$. This value is an imperfect estimate of the customer's (unobserved) mean transaction value, which we denote by $E(\bar{Z})$. Assume all Z_j, where j denotes the jth transaction, to be independent and identically distributed (i.i.d.) gamma variables with shape parameter p and scale parameter v, i.e., all $Z_j \overset{iid}{\sim} Gamma(p, v)$.

To account for heterogeneity in underlying mean transaction value across customers, the scale parameter v is also assumed to be gamma distributed, with shape parameter q and scale parameter γ, i.e., $v \sim Gamma(q, \gamma)$. Further, assume that the shape parameter of Z_j (i.e., p) is constant across customers. For clarity, we have $Z_j|v \sim Gamma(p, v)$ and $v \sim Gamma(q, \gamma)$ (hence the name gamma-gamma model). Using this, we can arrive at the (unconditional) expected transaction value of a specific customer for purchase j. $E(Z_j)$ is given by

$$E(Z_j) = E(\,E(\,Z_j|v\,)\,) = E\Big(\,\frac{p}{v}\,\Big) = p\,E\Big(\,\frac{1}{v}\,\Big) = \frac{p\,\gamma}{q-1}, \tag{21}$$

where the last step is performed because $\frac{1}{v}$ can be seen as an inverted-gamma distribution ($Inv\text{-}Gamma(q, \gamma)$) which has mean $\frac{\gamma}{q-1}$.

In the end, this research wants to know the average expected transaction value $E(\bar{Z})$. Using the assumption that all $Z_j \overset{iid}{\sim} Gamma(p, v)$ one can make two statements:

1) sum of x i.i.d $Gamma(p, v)$ random variables are $Gamma(px, v)$ distributed.

2) $Gamma(px, v)$ multiplied by $\frac{1}{x}$ give a $Gamma(px, vx)$ distribution.

This results in $\bar{Z} \sim Gamma(px, vx)$. Using this result, the desired expected mean transaction value is given by

$$\begin{aligned} E(\bar{Z}|\, p,\, q,\, \gamma,\, \bar{z},\, x) &= \frac{(\gamma + \bar{z}x)\, p}{px + q - 1} \\ &= \Big(\frac{q-1}{px+q-1}\Big)\frac{\gamma p}{q-1} + \Big(\frac{px}{px+q-1}\Big)\,\bar{z}. \end{aligned} \tag{22}$$

To obtain results from this equation, the model parameters (p, q, γ) have to be evaluated. Fader et al. (2005a) suggested that this can be done by maximizing the following likelihood function

$$L(p,\, q,\, \gamma|\, data) = \prod_{i=1}^{N} f(\bar{z}_i|\, p,\, q,\, \gamma,\, x_i). \tag{23}$$

As $\bar{z}$ is assumed to be gamma distributed, the probability function becomes the following

$$f(\bar{z}|\, p,\, q,\, \gamma,\, x) = \frac{\Gamma(px+q)}{\Gamma(px)\,\Gamma(q)}\frac{\gamma^q\, \bar{z}^{px-1}\, x^{px}}{(\gamma + \bar{z}\, x)^{px+q}}. \tag{24}$$

After obtaining these parameter values, the expectation for the mean transaction value can be conducted.

4.4 Pareto/Dependent model

The gamma-gamma model assumes that the number of future transactions is independent of the average spend per future transaction. Nevertheless, Fader et al. (2005a), who presented this model, concluded that there does exist a small correlation between these values within their data set, but they argued that this value was small enough to drop. In general, this correlation does exist, such that dropping this assumption can increase the performance of the predictions. Glady et al. (2009) had the same reasoning and came up with the Pareto/Dependent model. This new approach takes into account the possibility of dependency between the number of purchases and the average value per purchase. This dependency will be designed at customer level, such that the heterogeneity in the population is accounted for. Furthermore, this new model does not require constant expected profit per transaction over time.

First, assume that the number of transactions and the average profit per transaction of a customer i are dependent and relate to each other by the following model:

$$log\Big(\frac{Z_{i,k}}{E(Z_{i,k})}\Big) = r_i\, log\Big(\frac{x_{i,k}}{E(x_{i,k})}\Big) + \epsilon_i\,, \tag{25}$$

where r_i is a coefficient of dependence, $Z_{i,k}$ is the observed average profit per transaction on time k for customer i and $x_{i,k}$ is the observed number of transactions on time k for customer i. $E(Z_{i,k})$

and $E(x_{i,k})$ are the (unobserved) expected values of $Z_{i,k}$ and $x_{i,k}$. The first expected value can be obtained using the gamma-gamma model and can be found in Equation 21 in Section 4.3. The latter expectation can be found using one of the two models which predict the number of future transactions, the Pareto/NBD model or the hierarchical Bayes model. Glady et al. (2009) used the Pareto/NBD model in his research and therefore used Equation 13 from Section 4.1. This research will use the same procedure, but will also use the hierarchical Bayes model to forecast the number of transactions. Therefore, the (unconditional) expectation of the number of transactions is the expected value of the Poisson process from Assumption 2, which is λt.

The dependency between the monetary value of a customer and the number of transactions he/she makes can differ across customers. Customers having a high number of future transactions may have a high monetary value, but the opposite can also be true. That is why the dependency coefficient r_i will be modeled as a function of explicative variables, for which we take the 'lifetime' (T_i), recency (t_i) and the probability that a customer is active (p_i). p_i differs for the two frequency models. In the Pareto/NBD model, this probability can be calculated using Equation 12. Using the hierarchical Bayes model, the probability of being alive is the mean over the draws of the latent dummy variable z, which indicates whether a customer is active at the end of the training period or not. This results in the following formula:

$$r_i = \alpha_1\,\hat{p}_i + \alpha_2\,T_i + \alpha_3\,t_i + \alpha_0\,. \tag{26}$$

Taking Equation 25 with $k = T_i$ and implementing Equation 26, one ends up with:

$$\begin{aligned} log\Big(\frac{Z_{i,k}}{E(Z_{i,k})}\Big) = \alpha_1\,\hat{p}_i\,log\Big(\frac{x_i}{E(x_{i,k})}\Big) + \alpha_2\,T_i\,log\Big(\frac{x_i}{E(x_{i,k})}\Big) \quad &+ \alpha_3\,t_i\,log\Big(\frac{x_i}{E(x_{i,k})}\Big) \\ &+ \alpha_0\ log\Big(\frac{x_i}{E(x_{i,k})}\Big) + \epsilon_i\,. \end{aligned} \tag{27}$$

Estimating this formula gives estimates for $\alpha_{0,1,2,3}$ and therefore an estimate for r_i.

Now the future average profit per transaction can also be estimated. Using the period $[0,\,T_i+k]$, the formula becomes

$$\hat{Z}_{i,T_i+k} = \bar{Z}_i\Big(\frac{\hat{x}_{i,T_i+k}}{E(x_{i,T_i+k})}\Big/\frac{x_i}{E(x_i,T_i)}\Big)^{\hat{r}_i} \tag{28}$$

where $\bar{Z}_i$ is the observed customer's average transaction value from Equation 22 from Section 4.3, x_i (frequency) is the observed number of transactions during the observation period T_i and $\hat{x}_{i,T_i+k}$

is the prediction of x_{i,T_i+k} using the Pareto/NBD model or the hierarchical Bayes framework. Note that $\hat{Z}_{i,T_i+k}$ can now vary over time.

Eventually, with a time-varying monetary value of the transactions, the customer lifetime formulation from Section 3 does no longer hold. A new CLV formula is given by Glady et al. (2009). It is approximately the same as the previous one, but now the time variation of the monetary value is taken into account.

$$E(\mathrm{CLV}_{ih}) = \sum_{k=1}^{h} \frac{(\hat{x}_{i,T_i+k}\hat{Z}_{i,T_i+k} - \hat{x}_{i,T_i+k-1}\hat{Z}_{i,T_i+k-1})p_{it}}{(1+d)^t}, \tag{29}$$

4.5 Prediction Performance

To evaluate and compare the previous mentioned models on their prediction, two performance measures are conducted. Glady et al. (2009) suggested to make use of the *root mean square error* (RMSE) and the *mean absolute error* (MAE) between the prediction and the actual value. The actual values can be obtained from the data and the predictions are obtained using the models. These measures can be applied to the performance of predicting the future number of transactions, the future average monetary value of these transactions and the customer lifetime value. To account for possible outliers, Glady et al. (2009) also applied a trimming of 1% to both measures, which result in discarding the largest 1% of the prediction errors. The RMSE can be formulated as follows

$$RMSE = \sqrt{\frac{1}{0.99 \times N} \Sigma_{i \in RP} (\widehat{CLV}_{i,k} - CLV_{i,k})^2} \tag{30}$$

with N as the number of customers, k the number of future days ahead for the predictions and RP is the set of the remaining predictions. Here, the CLV is given as the observed customer lifetime value and $\widehat{CLV}$ for the prediction. In order to measure the performance of the models individually, these two values should be changed to the prediction and actual values of the number of transactions or the monetary value of these transactions. Further, the MAE is defined as

$$MAE = \frac{1}{0.99 \times N} \Sigma_{i \in RP} |\widehat{CLV}_{i,k} - CLV_{i,k}|. \tag{31}$$

The measures are quite alike when one looks at the average model prediction error, as it is in units of the variable of interest. Further, they both report a better score for the prediction model when the measure values are lower. On the contrary, there also exist some sustainable differences between these metrics. For instance, when the RMSE is used, the average is the best predictor, where the median is the best predictor when the MAE is used. Further, as the RMSE firstly square the errors before they are averaged, this metric gives high weights to large errors in contrast to the MAE metric, where the absolute is taken and thereafter the errors are averaged. This should be kept in mind when the results from these metrics are given.

4.6 Usefulness of the Models

It is useful to compare the models to each other based on the literature written about them. By this matter, one knows beforehand what can be expected from the different models while trying to understand/program them. Some small criteria that a practitioner would like to know are how well the models predict, whether the formulation in the paper is understandable, how flexible the models are based on the number of assumptions and how computational expensive the models are.

The most well-known Buy-Till-You-Die model is the Pareto/NBD model, firstly mentioned by Schmittlein et al. (1987). This commonly used model predicts the number of transactions in the future and can give a probability about a customer being alive in the future. The formulation of this model and the derivations of the formulas are quite hard, as Fader et al. (2005b) reported that the likelihood function is quite complex, involving evaluations of the Gaussian hypergeometric function. Multiple evaluations of this function are very demanding from a computational standpoint. A convenient feature of this model is that it only needs the recency and frequency of a customer in order to make predictions. Once the assumptions hold and the model is correctly implemented, the prediction performance is impressive (Fader et al., 2005a). Abe (2009) extended this model with a hierarchical Bayes (HB) framework and relaxed an independence assumption between the transaction rate and death rate. Borle et al. (2008) concluded that this assumption is not suitable in many situations, which makes the HB model more flexible. Once there exists a strong correlation between these two rates, the HB model outperforms the Pareto/NBD model (Korkmaz et al., 2013). A drawback of the HB model is that it uses a Markov chain Monte Carlo simulation, where the number of iterations should be high to make sure convergence appears (Korkmaz et al. (2014) used 200.000 iterations).

Moving on, the gamma-gamma model is easy to understand and implement (Fader and Hardie, 2013). However, this model assumes an independence between the number of transactions per customer and the monetary value of these transactions. Glady et al. (2009) question this assumption and propose the Pareto/Dependent model, which drops this assumption. They concluded that when there does exist a dependence, their model performs better on parameter estimation. Nevertheless, when this correlation is neglectable and the independence assumption hold, the two models perform approximately the same. A drawback of this model is that it must use a model which predicts the number of transactions (e.g., Pareto/NBD model or the hierarchical Bayes model). Therefore, the Pareto/Dependent model depends on the results and computation time of these other models, which can lead to wrong predictions (e.g., when the expected number of transactions is wrongly predicted) and long computation times (e.g., when the HB model is used with a lot of iterations). Table 1 below shows the summary of what is said before and gives the criteria and how the different models perform on them. The index '-' indicates a poor evaluating, '0' a mediate performance and '+' reports a positive performance on the criteria.

Table 1: Criteria valuation for the models

Criteria	Pareto/NBD	Hierarchical Bayes	Gamma-Gamma	Pareto/Dependent
Understandable formulation	-	0	+	+
Prediction accuracy	+	+	+	+
Number of assumptions	-	+	0	+
Computation time	+	-	+	0

5 Data

This Section discusses the data that is given by Winkelstraat.nl. First, in Section 5.1, the complete data set and the variables hidden within are given. To gain more knowledge of the data, Section 5.2 dives deeper into the behavior of this data set. Then, an explanation of the splitting of the data set is given in Section 5.3. Finally, the data that will be used for the different models will be given in Section 5.4.

5.1 Data Available

This research makes use of data that has been made available by Winkelstraat.nl, a Dutch online retailer which specializes in designer and luxury clothing. They have organized their data based on the orders that are placed by their customers. They started their online webshop on the second of October. The start date of this research is where we have drawn the line as end date for the data set, which is the first of April 2020. The data set Winkelstraat.nl has obtained over the years contains a good amount of information. The orders placed by the customers can be seen as the rows of the dataset and the columns are the different attributes of this order and the customer who made this order. Therefore, the variables can be split up in two classes, the ones that describe the order and the variables which explain some information about the customer who made the order.

Firstly, we will discuss the specification of the order. During the time between the start of Winkelstraat.nl (2nd of October 2012) and the start of this research (1st of April 2020) 670,167 orders have been placed. Every order has a unique ID, which we call *Order ID*. Within each order, multiple items can be present, which is given as variable *Number of Items*. As Winkelstraat.nl sells designer and luxury clothing, the price of these items is relatively high compared to 'normal' clothing. That is why customers tend to buy less items per order, which is also seen in the data set where the average number of items per order is around 1.29. The variable *Purchase Date* shows when an order has been placed on the website. Over the years, the total number of purchases per month has increased. This will be further explained in the next Section. The price of the orders is given as *Price*, where it ranges from a minimum of 9.95 to a maximum of 4696.23 euro (this maximum contained 21 items in one order) and the average price per order is around 185.77 euro. The minimum of the variable *Price* maybe lower than expected. This is feasible, as the items can be bought in sale. The variable *In Sale* reports how much items within the order are bought in sale. The total number of orders which contained an item in sale is 231.268, which is 34.51% of the total number of orders. Further, the data set shows whether an order contains an item which is returned, reported as deficiency or had to be cancelled, with variables *Is Return*, *Is Deficiency* and *Is Cancellation*, respectively. The first variable is an interesting feature, as a company wishes to have a small number of returns, as returns can be a metric of customer dissatisfaction. When customers return their item, they get (a part of) the price of the item back, which lowers the gross revenue of the company. The latter two variables are present by a failure of Winkelstraat.nl, where a deficiency is an item which is no longer in stock and a cancellation of an order is done when Winkelstraat.nl could not deliver the desired item for any reason. The three variables are present with 25.79%, 3.49% and 2.80%, respectively, in the data set. Finally, one knows the *Brand*, *Category* and *Size* of the clothing within the order.

Other variables present in this data set are the ones that give information about the customers. Each customer has a unique *Customer ID*. This ID is based on the email the customer uses to make an order at Winkelstraat.nl. As they value the privacy of the customers, they hash the email addresses, such that each email has a unique hashed ID. So, every time a specific email is being used for a repeat order, the data set assumes that this specific *Customer ID* has ordered again. There are 331,162 unique *Customer ID*'s present in the data set over the time frame used in this research. As every order contains a *Customer ID*, all variables given in the previous paragraph can also be informative about the customer. For example, the average value of *Number of Items* per customer is 2.61 and the average price spend per customer per order is 144.04 euro. This paragraph focuses on the variables which give more information about the customer.

The items can be split up in the gender the clothing is meant for, namely male and female, or it shows whether it is made for children. By counting all different genders per item per customer and taking the most frequent one, the gender of each customer is decided and given as variable *Gender*. 117,252 customers bought items which are labeled as male clothing, 65,160 as female clothing and 17,353 as clothing for children. Further, the first time a customer bought an item at Winkelstraat.nl is reported as *First Purchase* and the payment type used to indicate how a customer have paid for the orders is given as *Payment Type*. The most frequent types are paying using Ideal or using an Open Invoice (49.08% and 29.49% respectively). Open Invoice is a relatively new approach to pay where customers can try the goods and either return them or pay within a time period, usually 14 or 28 days. Buyers like it because they are not required to pay up front. The variable *Account* reports whether a customer created an account on Winkelstraat.nl. When customers make an account, more information is present about them. It is known what their *Date of Birth* is, such that we created the variable *Age*, which reports the age of the customer on the moment they placed an order. This age ranges from 7 till 70 years old with an average value of approximately 33. Finally, the data set shows whether a customer is subscribed for their newsletter and what the start and end date of the subscriptions are, denoted as *Subscriber Start Date* and *Subscriber End Date*, respectively. Unfortunately, variables which are present because of customer's own reporting, such as *Date of Birth*, always have some uncertainty as customers could report incorrectly for any reason. This should always be kept in mind.

Companies are only allowed to use personal data for marketing purposes if the user gives informed opt-in consent. Winkelstraat.nl follows this rule as they meet the GDPR and the Cookie Law. The Cookie Law is a piece of privacy legislation that requires websites to get consent from visitors to store or retrieve any information on a computer, smartphone or tablet. The reason it was designed was to protect online privacy by making customers aware of how information about them is collected and used online and gives them a choice to allow it or not. As Winkelstraat.nl meets this law, they use legal personal data by means of an informed opt-in consent. Furthermore, the data set they presented to this research is anonymized. As said earlier, the email addresses this research uses to identify customers are hashed.

5.2 Data Analysis

In order to get a better understanding of the purchase behavior of the customers at Winkelstraat.nl, Figure 2 shows the customer share for the number of orders made. One can see that the majority of customers, namely 63.75%, have only made one purchase and that approximately 12.13% made more than 3 purchases. Winkelstraat.nl wants to increase the group with repeat buyers, as that can be a metric for customer satisfaction. The one-time buyers are difficult for this research as their purchase history is one observation and predicting their CLV can be hard. Another interesting result based on the data is the share that the top customers have on Winkelstraat.nl's net revenue (so excluding returns, deficiencies and cancellations), as can be seen in Figure 3. The 20% most profitable customers have a net revenue share of 63.52%. So, when Winkelstraat.nl can identify who their top customers will be in the future, they can increase their net revenue.

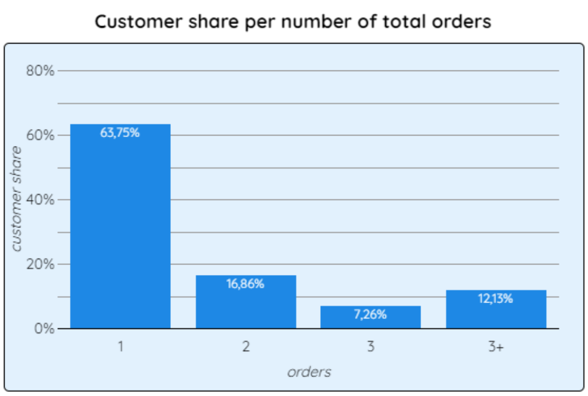

Figure 2: Number of orders per customer

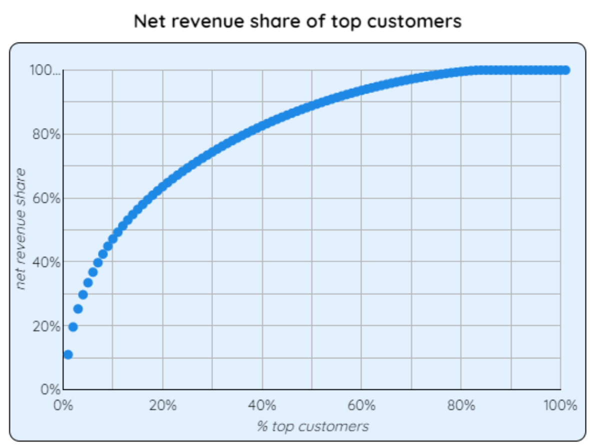

Figure 3: Net revenue share of top customers

As was mentioned before, the number of orders per month is increasing over the years, which can be seen in Figure 4. This figure suggests that there is a clear trend present. This is of utmost importance for this research, as the main subject is forecasting, which will be harder once there is a trend present. This should be taken into account when there will be made a prediction about the future number of purchases. In the last two years of the figure, one can see that there are two clear peaks. These peaks are based in the week of Black Friday, the week of the last Friday in November. Within this week a lot of orders are made as there is a week full of discounts on a lot of items. Another interesting figure is Figure 5, which shows the average amount of money spend per customer per month over the years. One can see that it fluctuates around a mean value, which indicates that there is no trend present and therefore one can suggest that predictions will not suffer from this. Further, as the values fluctuate around a mean, the monetary value of each customer based on their transaction history can be calculated by taking the mean over all prices paid per order. This is done in the marketing literature and will be done in this research. As was said before, the figures show fluctuations. This suggest that there are seasonal patterns present. This can be perfectly seen in the week of the last Friday of November, where Black Friday is in. To incorporate this kind of seasonal patterns, one should predict over whole years, such that the actual and predicted time period incorporates the same seasonal appearance.

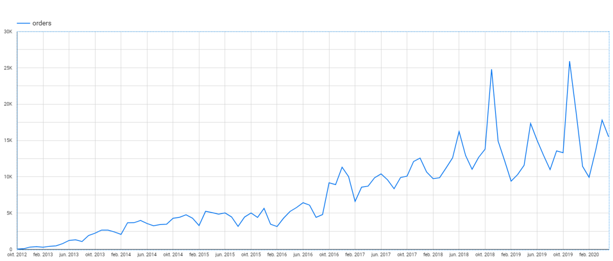

Figure 4: Number of orders per month over the years

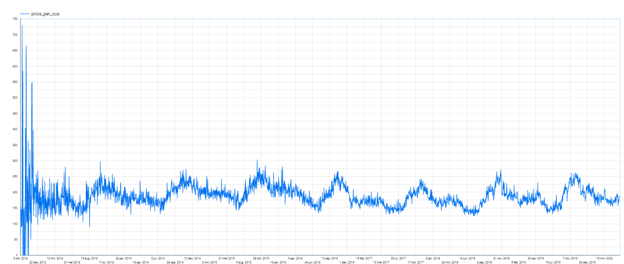

Figure 5: Amount spend by customers per month over the years

5.3 Data Splitting

The previous Section implied that there are seasonal patterns present in the data. In order to take this into account, the data set should consist of whole years. As the data presented by Winkelstraat.nl consist of 7.5 years of data, the first six months are neglected. This means that the dates range between the first of April 2013 and the 31th of March 2020, which results in a time span of 7 complete years. Dropping the first six months is a loss of 0.22% of the total observations. As these are the first months of the existence of Winkelstraat.nl, their data does not represent the buying behavior of their customers, so dropping these observations has little effect on the results.

To measure the prediction performance of the different models as discussed in Section 4.5, a training and test set should be made. A training set makes sure the model is properly trained, such that it can make good predictions. In order to evaluate these predictions, a test set will be made.

This research can make use of seven years of transaction data. The data will be split up in five years of training data and two years of test data, which implies that the predictions are 730 days ahead (2 years with 365 days). This cannot be done randomly, as the prediction for the training set can only be done for later years. That is why the training set consists of the data from 01-04-2013 till 31-03-2018 and the test set from 01-04-2018 till 31-03-2020. By this matter, the training set has a lot of data to train the model for prediction, which would result in very precise predictions. Important to note is that the customers who purchased their first order in the test set are discarded from the research, as they cannot be trained because they have no transaction history. Further, the customers who have bought in the training set, but are not present in the test set, are given a value of zero in the test set for the number of transactions and the monetary value of these transactions. This is done as customers who bought items in the training set, but are not returning in the test set, are labelled inactive after the training set. By using this reasoning, we see that 60.45% of the customers in the training set have a zero in the test set and therefore have not repeated buying items at Winkelstraat.nl. The Pareto/NBD and hierarchical Bayes model make a prediction about the probability of a customer to be alive, such that it can predict whether a customer is not buying anything in the test set. In the end, the number of customers in the training and test set is 49,892.

5.4 Data Used

The previous Sections described the available data, but did not specify which parts of the data will be used in the remainder of this research. This Section explains what data each model requires in order to properly perform. The Pareto/NBD model has a nice feature, namely that it only requires a customer's recency and frequency (RF data). Recency can be obtained using the variable *Purchase Date* in order to retrieve the number of days between the first transaction and the latest one. Frequency is equal to the number of purchases a customer made, after their first purchase, during the time he/she was alive and can be calculated by summing the number of times the *Customer ID* of a specific customer is present in the data. In order for this model to work, the one-time buyers are neglected from the dataset, as these customers have no purchase history.

For the gamma-gamma model and the Pareto/Dependent model, the RF data should also be used. The purpose of these models is to predict the future monetary value of all customers. In order to retrieve such a prediction, they use the present monetary value. This value can be computed using the *Price* variable, which shows the price of the orders placed by customers. By taking the average over the prices of all transactions for each customer, a monetary value is developed. The

RFM (recency, frequency and monetary value) data in the training set which are needed for these models is given in Table 2, where the customers are ranked based on their frequency. The value T reports the number of days since the customer made his/her first purchase.

Table 2: Top and bottom frequent buyers

(a) Top 10

ID	Frequency	Recency	T	Monetary value
4250	147	1747	1778	148.37
48707	144	1173	1623	236.78
33259	128	1730	1761	302.12
37543	128	1231	1428	247.62
45797	114	1789	1808	205.59
30970	107	1513	1561	239.50
48332	106	1343	1365	128.71
22872	103	1466	1736	241.90
5581	98	394	1025	222.69
43592	97	1803	1823	255.77

(b) Bottom 10

ID	Frequency	Recency	T	Monetary value
45675	1	30	248	11.95
25354	1	20	178	12.50
26533	1	274	838	14.00
43121	1	701	1067	14.94
4376	1	327	347	14.99
8256	1	394	422	15.98
46023	1	1567	1601	15.98
195	1	386	662	17.98
30681	1	810	1102	17.98
9460	1	433	597	19.46

This Table shows that there are profitable customers who buy lots of items from Winkelstraat.nl and have a high average purchase value. On the other hand, there are also customers who have a low monetary value and are one-time repeat buyers. This information is needed for the three aforementioned models to get predictions about the future number of transactions and the future average monetary value.

The last model performed in this research is the hierarchical Bayes (HB) framework developed by Abe (2009). This model extends the classical Pareto/NBD model not only by dropping the assumption between the transaction and death rates, but it can also include covariates. This research will implement the HB model with and without covariates in order to see the effect on the prediction performance. The HB model will need the RF data as mentioned before and will include covariates. The variables that will be used as covariates are a subset of the ones mentioned in Section 5.1, namely *Is Return*, *In Sale*, *Age*, *Number of Items*, *Is Deficiency*, *Is Cancellation*, *Payment Type* and *Gender*. In the end an intercept is merged to the dataset.

The variable *Payment Type* is split up in three dummy variables, namely whether a customer paid using *Ideal* or not, paid using an *Open Invoice* or not and whether an *Other Payment Type* was used. *Gender* is also split up in dummies, namely whether a customer orders clothing indicated as

Male clothing or not, *Female* clothing or not, whether the clothing is meant for *Children* or not and whether the gender of the ordered items within the order is *Unknown* or not. Further, the variables *Is Deficiency* and *Is Cancellation* are merged together as one dummy variable, indicating whether a customer ever had an order being a deficiency and/or being cancelled or not. As the data should be customer-specific, the variables which are order-specific must be transformed. Therefore, each variable takes the average value over the orders for all customers. For example, when a customer made four repeat transactions with three of them reported as returned, the variable *Is Return* will show a value of 0.75 (3 of the 4 orders were returned). Finally, the variable *Age* is used. The hierarchical Bayes model can only incorporate time-invariant covariates, but the age of a customer differs over the years. This research has decided that the time frame used as training period (five years) has little influence on the person's age, as one can take the average over the five years and still have a good prediction about the age of the customer. Further, for some customers there is no information about their age. This research used the mean over all customers' age in order to fill the missing values. The descriptive statistics of this data set can be found in Table 3.

Table 3: Descriptive statistics of the data

	Mean	Std. dev.	Min	Max
Frequency	3.146	5.250	1.000	147.000
Recency	399.276	405.353	1.000	1814.000
T	804.351	486.515	1.000	1825.000
Monetary Value	189.347	141.800	11.950	3159.813
Is Return	0.273	0.289	0.000	1.000
In Sale	0.312	0.421	0.000	8.667
Age (years)	33.849	8.309	7.000	70.000
Number of Items	1.294	0.540	1.000	12.500
Is Deficiency/Cancellation	0.102	0.177	0.000	1.000
Ideal	0.438	0.423	0.000	1.000
Open Invoice	0.276	0.385	0.000	1.000
Other Payment Types	0.285	0.375	0.000	1.000
Male	0.106	0.194	0.000	1.000
Female	0.058	0.143	0.000	1.000
Children	0.007	0.046	0.000	1.000
Gender Missing	0.830	0.231	0.000	1.000

6 Software Implementation

To run the data on these model specifications, the programming language *Python* (Langa, 2020) is used. This is due to the fact that Winkelstraat.nl is using this software for their other programming implementations. The data is stored in Google BigQuery, where it can be automatically updated once an order is placed by a customer. All data that is made available is given in Section 5.1.

To implement the Pareto/NBD model and the gamma-gamma model, a *Python* package can be used. This package is called the *Lifetimes* package [3]. These models only require transaction data, where the customer ID, the datetime a customer made a purchase and the price of these transactions should be given. The package transforms this information into Recency, Frequency and Monetary value (RFM) data together with the time since the customer made his/her first purchase (T). The parameter functions are obtained using the *Lifetimes* package, which maximizes the likelihood functions as given in Section 4. Further, this package uses these parameters to implement them in the desired functions to obtain the predictions for the number of transactions and the monetary value of these transactions for each customer. Finally, this package has a convenient feature, namely it can calculate the customer lifetime value using a programmed function.

Unfortunately, not all models can be implemented using this package. As this package is from February 2017 and the latest version is from October 2019, it is still in progress and development. The Pareto/Dependent model is an extension of the gamma-gamma model and uses some results from this model and from the Pareto/NBD and hierarchical Bayes model, which is explained in Section 4.4. Therefore, some functions from the package can be used to get the desired parameters, which can thereafter be implemented in the functions for this model to obtain useful results. Glady et al. (2009) described the procedure very well, such that the models could be quickly implemented. Further, the CLV has to be evaluated. The previous models could make use of the *Lifetimes* package, but this cannot be done for this model, such that the CLV formula had to be self-programmed to get the desired results. As the CLV runs over a lot of iterations and every iteration the monetary value has to be calculated using this model, the computation time is somewhat higher. Eventually, this model runs smoothly and can be used for other practitioners/data sets.

The last model that had to be programmed is the hierarchical Bayes model as described in Section 4.2. As this model is implemented completely different than the other models, the *Lifetimes* package is of no use. This results in the fact that the complete implementation of the model is new

[3] Documentation can be found at https://lifetimes.readthedocs.io/en/latest/

to *Python* and should be programmed during this research. This model was described in Abe (2009), where the formulas were clearly reported, but difficult to program. Fortunately, the programming language *R* (R Core Team, 2020) was helpful. This language has a package which completely programs this model, namely the *BTYDplus* (Platzer, 2016) package. By diving into this package, the complete programming code was given. This research has used this code and transformed it into *Python* code. This was a difficult and time-consuming job as this code consists of over 350 lines of code. In the end, we managed to get the entire code, with some adjustments, into *Python*. We checked our implementation by performing the hierarchical Bayes model in *R* using the same data set. The results were approximately the same, thus we can conclude that our implementation performs well. It can be used by other programmers and other data sets can be implemented. Drawbacks of this approach is that it takes a lot of time to run the entire simulation and the output is very extensive, as every parameter (4) is given for every draw (200) for all customers (49.892). The numbers in parentheses are the numbers used in this research.

7 Results

As described in Section 5, the data that has been used in this research is from Winkelstraat.nl. The transaction history has been used to calculate the different parameters for each model. The specification of all models can be found in Section 4. The obtained parameters are than used to get a prediction for the future number of purchases and the future monetary value of these purchases. Section 5.3 explained that the training set consists of five years and the test set of two years, which implies that the predictions from the training set are 730 days ahead (2 years of 365 days). The test set is used to compare the true number of purchases and the true monetary value of these purchases with the predicted ones. In the end, by combining the two predictions and actual values, the predicted and actual customer lifetime value is expressed. All predictions are than compared to the actual values using metrics which were described in Section 4.5. This Section reports all results obtained in this research.

7.1 Pareto/NBD model

Firstly, the results from the Pareto/NBD model are given. The specification of this model can be found in Section 4.1. First of all, the model parameters are computed using the Maximum Likelihood Estimator and thereafter used to compute a prediction for the number of transactions and the probability that a customer is alive. The results are $r = 1.686$, $\alpha = 260.756$, $s = 3.221$ and $\beta = 4077.277$. These parameter values can be used to predict the future number of transactions for each customer and his or her probability of being alive at the end of the training period by implementing them in Equations 14 and 12 respectively. To get an insight in the results from this model, Table 4 is presented with two subtables. The first one reports the results from a random subset of the total results, as it is based on the ranking of the customer ID's, where the middle ten ID's are taken. One should note that, as previously mentioned, approximately 60% did not make a repeat purchase in the test set, such that a lot of the actual frequency reports zero. Also note that the average, min and max values are presenting information over all results. The second subtable reports the top ten customers in the test set, which are the ones who have the highest actual buying frequency. Both tables also report the prediction about the buying frequency and the probability of these customers to be alive after the training period.

Table 4: Results Pareto/NBD model

(a) Subsample based on customer ID's

ID	P(alive)	Predicted frequency	Actual frequency
24940	0.826	1.334	2
24941	0.158	0.092	0
24942	0.149	0.021	0
24943	0.267	0.095	0
24944	0.127	0.023	0
24945	0.027	0.001	0
24946	0.856	2.916	1
24947	0.830	1.346	0
24948	0.062	0.004	0
24949	0.078	0.006	0
Average	0.589	1.680	1.287
Min	0.000	0.000	0
Max	1.000	45.435	93

(b) Top 10 based on the actual buying frequency

ID	P(alive)	Predicted Frequency	Actual Frequency
10635	0.989	19.828	93
635	0.999	38.494	76
3530	0.119	0.331	75
31657	0.890	10.924	74
18251	0.999	24.195	69
8990	0.850	4.659	59
3739	0.985	5.415	52
15588	0.997	7.373	51
42277	0.997	8.651	51
43888	0.075	0.091	51

The first subtable shows that the predictions are quite accurate based on this subsample. Even though most of the actual buying frequency consist of zero, the predictions measure that pretty well. Further, the average of the predictions and the actual values are in close range. A rather insightful table is the second subtable. Here, the top ten most frequent actual buyers are presented next to their prediction. One can see that the top buyers are hard to predict, where some even got a prediction of about zero, but had an actual buying frequency above 50. The prediction performance of this model based on some performance measures can be found in Section 7.5.

7.2 Hierarchical Bayes Model

Another model which can make a prediction about the number of future transactions is the hierarchical Bayes model proposed by Abe (2009). This model extends the Pareto/NBD model by relaxing the assumption of independence between the transaction and death rates. Another positive side of this model is that it can incorporate time-invariant covariates. This research performed the model two times, without covariates (model 1) and with covariates (model 2), such that we can conclude whether the covariates show a positive grow in the prediction performance. To retrieve the results from this model, we have to initialize the model parameters. The Markov chain Monte Carlo (MCMC) steps are repeated 100.000 times, where only the last 50.000 were used with a thinning value of 500 (this results in 200 draws per customer per parameter). This resulted in twelve hours of computation time to run this model once.

As was mentioned in Section 3.1, the hierarchical Bayes model in literature mostly outperforms the classical Buy-Till-You-Die models, such as the Pareto/NBD model, when there is a correlation between the transaction and death rates. The correlation between the two can be calculated by using the variance of both rates and the covariance between them. Every draw has other values for these (co)variances such that we obtained 200 correlations. Figure 6 shows a histogram where the correlations between the values of the rates (λ and μ) are given. One can see that it does not differ much from zero, with an average value of 0.0065.

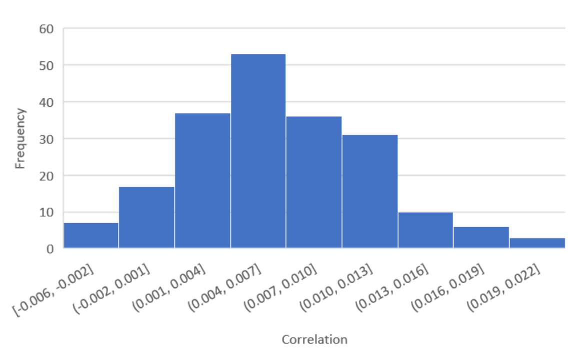

Figure 6: Correlation between $\log\lambda$ and $\log\mu$

The hierarchical Bayes model is implemented with and without covariates. Section 4.2 suggested that the rates λ and μ are bivariate normally distributed, as can be seen in Equation 16. This equation also shows that both rates are regressed against each covariate. The obtained results are given in Table 5, where the first subtable shows the results when the model runs without covariates (model 1) and the second subtable shows the results with them (model 2). The numbers presented between parentheses are the standard errors obtained in this regression. The asterisk above some values indicate that this variable has a significant influence on the rates. Note that the regression of the covariates on the rates is done on the logarithm of λ and μ, such that the magnitude of the coefficients must be interpreted accordingly.

Table 5: Hierarchical Bayes models

* Indicates significance

(a) Without covariates (model 1)

	log λ	log μ
Intercept	-5.395 (0.038)	-7.281 (0.200)
Variance	0.658 (0.010)	0.052 (0.007)

(b) With covariates (model 2)

	log lambda	log mu
Intercept	-5.806 (3.770)	-8.176 (1.742)
Is Return	0.183 (0.019)	0.243* (0.030)
In Sale	-0.030 (0.025)	-3.613* (0.450)
Age	0.004 (0.001)	-0.006 (0.001)
Number of Items	0.139 (0.011)	-0.053 (0.026)
Is Deficiency/Cancellation	0.797* (0.037)	1.662* (0.063)
Ideal	-0.312* (3.728)	-1.607* (1.556)
OpenInvoice	0.209 (3.726)	-0.707 (1.551)
Other Payment Type	-0.391* (3.728)	-0.563 (1.551)
Male	0.346* (0.300)	0.260 (1.259)
Female	-0.251* (0.303)	-3.273* (1.166)
Children	-0.131 (0.312)	-0.084 (1.277)
Gender Missing	0.132 (0.306)	2.406* (1.204)
Variance	0.557 (0.007)	0.088 (0.007)

One can see that the variance of the purchase rate at model 2 is lower than the one at model 1. This is the opposite for the death rate. Further, looking at the model with covariates, one can see that the *In Sale* variable has a negative influence on the death rate, which implies that when a customer buys more items in sale, the death rate drops, which makes this customer less likely to drop out in the future. We can also see that some other covariates are significantly related to the purchase and/or death rate. When an order has an item, which is a deficiency or a cancellation, the logarithm of the purchase rate will increase, as well as the death rate. This would imply that a customer who has had more items which were reported as a deficiency or cancellation tend to buy more, but have a shorter lifetime. The opposite is true for the variable *Ideal*. When a customer pays more often using Ideal, both rates would drop, which would suggest that this customer tend to shop less at Winkelstraat.nl, but has a longer lifetime. This variable has high standard errors, making it a volatile result. This should be kept in mind. Other payments than paying with Ideal or OpenInvoice also seems to have a negative effect on logλ and thus customers who pay with a different payment type tend to purchase less. Again, the standard errors are relatively high. Then, the 'gender' the bought clothes are meant for have some effect on the rates. First of all, when a customer buys more clothing labeled as men clothing, he or she will purchase more often as the purchase rate is positively influenced by this variable. On the other hand, Female clothing seems to have a significant negative effect on both rates, implying that customers who buy these clothes will longer wait until they place a new order and have a longer lifetime. Finally, when the gender of the clothing is not reported, the death rate is positive, indicating that customers who bought these clothes are more likely to 'die' sooner.

In the end, one wants to get predictions about the future number of transactions and the probability of being alive. This model provides us with parameter values for the transaction and death rates. The average values over the draws for these rates, namely logλ and logμ, are -2.272 and -3.151 without covariates (model 1) and -2.297 and -3.691 with covariates (model 2), respectively. The convergence of these parameters is checked by the Geweke test (Geweke et al., 1991). The results from the two specifications from this model are next used to get predictions, as was reported in Section 4.2. Firstly, for every draw, each value of logλ is used in the Poisson distribution, given in Equation 3, to get a prediction about the number of future transactions. After that, the draws are aggregated to get one value for each customer. The same procedure is followed for each value of logμ to get a prediction of the expected lifetime for each customer, using Equation 5. The results are given in Table 6, where a subsample is chosen when the results are ranked on the customer ID, such that it can give a representation of the entire results database. These are the same customer ID's as in Table 4 (Pareto/NBD model). There are two tables given; the results from the model with and without covariates.

One can see that the results from both model specification report relatively the same, with some small differences. Both models seem to predict the actual value well, where an average value of the expected frequency from the model without covariates is closer to the average value of the actual buying frequency, indicating that this model performs somewhat better. A rather strange result is seen in the results from the second model, where the expected lifetime of a customers is high, with an average value of approximately 2×10^{13}. This would suggest that most customers would remain active for an infinite time. Korkmaz et al. (2014) encountered the same problem and mentioned that this probably occurs when one of the assumptions is violated. The prediction performance and

Table 6: Hierarchical Bayes models

(a) Without covariates (model 1)

ID	Lifetime	P(alive)	Predicted frequency	Actual frequency
24940	1636.95	0.86	1.75	2
24941	1336.08	0.19	0.30	0
24942	1680.66	0.11	0.00	0
24943	1390.20	0.30	0.45	0
24944	1452.42	0.14	0.00	0
24945	1481.66	0.03	0.45	0
24946	1527.19	0.88	4.00	1
24947	1404.54	0.84	1.45	0
24948	1350.58	0.08	0.00	0
24949	1599.68	0.09	0.00	0
Average	1491.17	0.61	2.03	1.29
Min	1037.19	0	0	0
Max	2004.73	1	68.85	93

(b) With covariates (model 2)

ID	Lifetime	P(alive)	Predicted frequency	Actual frequency
24940	73649.82	1.00	2.80	2
24941	2547.25	0.31	0.20	0
24942	824.61	0.07	0.00	0
24943	2076.88	0.41	0.95	0
24944	290.72	0.00	0.00	0
24945	684.88	0.01	0.00	0
24946	20630.14	1.00	3.80	1
24947	43730.68	0.99	2.70	0
24948	719.18	0.03	0.00	0
24949	759.19	0.04	0.00	0
Average	2.24E+13	0.68	2.61	1.29
Min	1.19E+02	0	0	0
Max	1.10E+18	1	92.20	93

the comparison with the Pareto/NBD model can be found in Section 7.5.

7.3 Gamma-Gamma Model

After having a prediction about the number of future transactions, we continued with obtaining a prediction about the monetary value of these transactions for each customer. This can be evaluated using the gamma-gamma model as presented in Section 4.3. This model uses a maximum likelihood estimation (MLE) to get a prediction about the model parameters. This model was implemented using the programming language Python and the results for the parameters (p, q and v) are given as (3.955, 5.161 and 203.293), respectively. These parameters are used to get a prediction about the expected monetary value of each customer's buying behavior. These values can be implemented in Equation 24 from Section 4.3, which is a gamma distribution in order to get this prediction. This model also reports the unconditional expected monetary value which can be obtained using the same parameter values, with a value of 193.23 euro. This unconditional expectation will eventually be used in the Pareto/Dependent model.

As was mentioned in Section 5.3, more than 60% of the customers in the test set did not purchase any items during this time frame. This model predicts the monetary value of the future transactions and thus cannot be equal to zero. In order to get an idea about the prediction performance of the models which predict the monetary value of the future purchases (gamma-gamma model and the Pareto/Dependent model), one must only incorporate the customers who have bought items in the test set and therefore have an actual monetary value of the transactions above zero. The obtained predictions are reported in Table 7. This Table is again divided in two subtables, where the first one is based on the customer ID's and the latter one shows the top ten based on the actual monetary value of the future transactions. Also, the average, minimum and maximum values of the results, excluding the zero values, is given.

Table 7: Results gamma-gamma model without the actual zeros values

(a) Subsample based on customer ID's

ID	Predicted monetary value	Actual monetary value
24803	186.49	249.85
24810	126.34	92.03
24811	122.29	213.38
24819	293.96	407.50
24822	199.56	349.86
24823	268.43	216.30
24826	127.20	70.45
24827	81.19	77.46
24828	101.95	115.93
24829	317.65	350.74
Average	189.09	179.08
Min	55.95	13.98
Max	1423.52	489.79

(b) Top 10 based on the actual monetary value

ID	Predicted monetary value	Actual monetary value
48450	181.88	489.79
22087	521.39	489.68
7353	91.36	489.50
36329	181.16	489.49
13143	156.35	489.48
39211	223.53	489.38
35317	407.53	489.34
7925	296.77	489.12
40613	200.91	489.00
32820	382.42	488.85

By first looking at the subtable which reports a subset of the results, one can see that some predictions are quite well, where others are nowhere near the actual values. This model seems to predict the mediate monetary values well, where the high actual monetary values are poorly predicted. This can also be seen in the second subtable, where the top ten customers based on the actual values are given. These high actual values are hard to predict by this model. The average of the predictions and the actual values are quite alike, where it differs by ten euro. The maximum of the predictions seems to be enormous compared to the actual maximum value. The evaluation of the prediction performance will be further discussed in Section 7.5

7.4 Pareto Dependent model

The previous model makes use of the independence assumption between the purchase frequency and the monetary value of these purchases for each customer. This assumption is questionable as customers with a high average monetary value per order can be seen as satisfied customers and therefore these customers tend to buy more at Winkelstraat.nl, which means that there is a dependence present. Glady et al. (2009) suggested to drop this assumption and came up with the Pareto/Dependent model, which is described in Section 4.4. This assumption can be tested using the Pearson correlation coefficient between the frequency and the monetary value from the RFM data. This correlation resulted in a value of 0.097, which is significantly different from zero. This would suggest that this model can have a better fit than the gamma-gamma model.

This model uses some parameters and functions from a model which predicts the number of transactions (Pareto/NBD, hierarchical Bayes model 1 (without covariates) and hierarchical Bayes model 2 (with covariates)). Therefore, this model has been run three times in order to get these three models as input for the Pareto/Dependent model. This has been explained in Section 4.4. Using these models and the unconditional predicted monetary value from the gamma-gamma model, the parameter value r can be calculated for each customer, with an average value of 0.064 using the Pareto/NBD model, 0.0032 for the hierarchical Bayes model 1 and -0.039 with the hierarchical Bayes model 2. Making use of this parameter and the aforementioned values from the other models, one can make a prediction about the monetary value. This value can have a dependence with the buying frequency of a customer. Further, this monetary value can change over time, making it more flexible.

Again, the many zero values in the test set cannot be incorporated in the results when the monetary value of the transactions is discussed. Therefore, as was done with the gamma-gamma model results, the results for the Pareto/Dependent model with the different frequency models is given for all customers who have bought one or more items in the test set. The results are given in Table 8, where the columns indicate which frequency model is used (HB is hierarchical Bayes) for the Pareto/Dependent model to get a prediction about the monetary value and the last column shows the actual values. The first subtable reports a subsample of the customer ID's, the same ID's as with the gamma-gamma model. The second subtable shows the top ten customers based on the actual monetary value of the transactions (again, the same customer ID's as before). Finally, the average, minimum and maximum are given at the bottom of the first subtable.

Table 8: Results Pareto/Dependent model without the actual zeros values

(a) Subsample based on customer ID's

ID	Pareto/NBD	HB model 1	HB model 2	Actual
24803	181.76	48.32	191.62	249.85
24810	54.62	282.78	54.96	92.03
24811	81.75	131.11	83.35	213.38
24819	400.98	324.51	398.85	407.50
24822	197.25	222.93	200.29	349.86
24823	290.06	174.23	287.60	216.30
24826	92.80	189.01	92.49	70.45
24827	50.79	69.92	51.32	77.46
24828	69.10	188.67	70.04	115.93
24829	348.18	73.26	350.85	350.74
Average	181.47	189.32	183.78	179.08
Min	11.80	15.18	12.16	13.98
Max	1896.26	3105.96	1914.83	489.79

(b) Top 10 based on the actual monetary value

ID	Pareto/NBD	HB model 1	HB model 2	Actual
48450	174.05	134.62	162.26	489.79
22087	541.01	455.49	544.62	489.68
7353	53.93	252.29	55.53	489.50
36329	176.84	189.19	177.97	489.49
13143	136.05	253.11	160.13	489.48
39211	237.42	216.46	252.04	489.38
35317	449.00	130.00	452.19	489.34
7925	297.10	49.99	306.73	489.12
40613	207.11	119.63	208.62	489.00
32820	404.59	68.58	402.77	488.85

The first table shows some interesting results. One can see that the Pareto/Dependent in combination with the hierarchical Bayes model 1 predicts the poorest. This would suggest that the inclusion of covariates has a positive effect on the prediction results. The other two model combinations seem to predict approximately the same and these values are close to the actual values, apart from the high monetary value. This can also be seen in the second subtable, which shows that all three models poorly predict the highest monetary values. The predictions are mostly well below the actual value. Again, HB model 1 predicts the worst and the other two models predict approximately the same. The average predictions are somewhat higher than the average actual values, but are really close to each other, especially the combination with the Pareto/NBD model. As was seen in the gamma-gamma model, the maximum of the predictions are much higher than the actual maximum. The prediction performance and the comparison with the gamma-gamma prediction performance will be discussed in the next Section.

7.5 Overall Model Comparison

The previous subsections discussed the results obtained by the different models. We want to know which model has the best prediction performance on this data set, such that Winkelstraat.nl has the best tools to predict who the most profitable customers are. This can be done by reporting the model's prediction performance using the root mean squared error (RMSE) and the mean absolute error (MAE) as was discussed in Section 4.5. This Section gave the formulas to calculate these measures using the customer lifetime value, but the prediction of the frequency and monetary values can also be implemented with their actual values. The model performance of predicting the future number of transactions and the future monetary value of these transactions can be seen in Table 9. These two subtables show the prediction performance of each model discussed in this research and used two benchmarks in order to compare the results of the models with understandable benchmarks. The first benchmark is the *Only Zeros*, which contains only zeros as prediction for the frequency and monetary value. The second one is *Training Data*. In the first subtable this benchmark contains the frequency from the training data set as prediction for the future multiplied by 2/5, as the training set contains five years of purchase time and the test set two years. For the second subtable, the *Training Data* contains the monetary value from the RFM data from the training set.

Table 9: Prediction performance of the different models

(a) Frequency models

	RMSE	MAE
Only Zeros	3.193	1.286
Training Data	2.920	1.507
Pareto/NBD	2.798	1.584
Hierarchical Bayes (model 1)	3.428	2.049
Hierarchical Bayes (model 2)	3.080	1.810

(b) Monetary value models

	RMSE	MAE
Only Zeros	204.835	179.082
Training Data	175.542	108.507
Gamma-Gamma	107.600	79.933
Pareto/Dependent with Pareto/NBD	133.778	91.715
Pareto/Dependent with HB model 1	174.695	120.843
Pareto/Dependent with HB model 2	134.888	92.305

The first subtable in Table 9 shows that the RSME values are very close to each other, where the Pareto/NBD model has the best fit. Further, the purchase history of a customer also seems to be a good predictor for the future number of transactions. The hierarchical Bayes model shows some difficulties predicting the future buying frequency, especially with no presence of covariates. The MAE values show roughly the same results, where using only zeros performs the best predictions. This can be due to the fact that the MAE's best predictor is based on the median and the median of the actual frequency is zero.

The second subtable above shows the prediction performance of the models which predict the future monetary value of the transactions for each customer. The previous subsections, where the gamma-gamma model and Pareto/Dependent model results were given, used only the customers in the test set which purchased at least once in order to get a good impression of the model performance. The second subtable will also use the data with customers who has bought in the test set. One procedure is clearly the best predictor, namely the gamma-gamma model. This model has the lowest RMSE and MAE values. Further, the Pareto/Dependent model with the Pareto/NBD model and the Pareto/Dependent model with the hierarchical Bayes model 2 (with covariates) perform approximately the same and relatively well. Finally, the benchmarks and the Pareto/Dependent model with the HB model 1 perform the worst, where the *Only Zeros* has the highest value of all. This is clearly due to the fact that the zero values are discarded for the performance measures of the monetary value models.

The main focus of this research is knowing which model combination best predicts the customer lifetime value. Table 10 shows the prediction performance from the different combinations of the models and the benchmarks previously mentioned. The actual value of the customer lifetime values which has been used in the RMSE and MAE formulas is the multiplication of the frequency and monetary value of the test set. The benchmark *Only Zeros* again predicts only zeros and *Training Data* uses a multiplication of the frequency and the monetary value from the training set.

Table 10: Customer lifetime value performance

	RMSE	MAE
Only Zeros	847.676	278.794
Training Data	808.430	343.410
Pareto/NBD with gamma-gamma	736.046	283.927
Pareto/NBD with Pareto/Dependent	721.737	280.300
Hierarchical Bayes model 1 with gamma-gamma	743.608	312.164
Hierarchical Bayes model 1 with Pareto/Dependent	737.784	300.990
Hierarchical Bayes model 2 with gamma-gamma	737.781	300.111
Hierarchical Bayes model 2 with Pareto/Dependent	728.590	304.900

One can see that the models outperform the benchmarks on the RMSE values. Again, the MAE result of the *Only Zeros* benchmark is lower than every other predictor, but has the highest RMSE value. This has the same reasoning as before, namely the best predictor of the MAE is the median. The median of the CLV is zero, making the *Only Zeros* benchmark a good predictor. The model which performs the best on both prediction performance measures is the Pareto/NBD model with the Pareto/Dependent model. This model has the lowest RMSE value with respect to all other predictors and has the lowest MAE value of all models. Comparing the Pareto/NBD model combinations with the hierarchical Bayes model combinations, one can see that the Pareto/NBD combinations both score better on the MAE measure. Further, the hierarchical Bayes model with covariates (model 2) performs better than the model without covariates (model 1) on the RMSE values. At last, the most interesting combination for the marketing literature is the new one, namely the hierarchical Bayes model with covariates in combination with the Pareto/Dependent model. The RMSE measure show that this model is second best in predicting the customer lifetime value, but it does not work better than the best predictor, namely the combination between the Pareto/NBD and the Pareto/Dependent model.

8 Discussion

The main focus of this research is to predict the customer lifetime value for Winkelstraat.nl using the best predicting models. In order to get this prediction, some classical models can be used to predict the number of future transactions and the monetary value of these transactions per customer, as these two values could report the profitability of a customer to a company. A new combination of existing models was introduced in this research in order to try and get the best performing model combination which can predict the customer lifetime value. This combination consist of the hierarchical Bayes model by Abe (2009) and the Pareto/Dependent model by Glady et al. (2009). These models are based on two extensions of the 'classical' models (Pareto/NBD and gamma-gamma models) and has not been combined in the literature so far.

Following the marketing literature available about the different models gave us a valid reason to believe that this new combination could lead to better predictions. This is due to the fact that the 'classical' models have some questionable assumptions. The Pareto/NBD model assumes independence between the transaction and death rates and the gamma-gamma model assumes independence between the buying frequency and the monetary value of the items bought. First of all, a customer's transaction rate can be dependent of his/her death rate, as customers who have a low transaction rate (buy with longer inter-transaction times) are more likely to have a high defection rate (more likely to stop buying). Dropping this assumption can lead to improved predictions. Further, the independence assumption between a customer's number of purchases and the monetary value of these purchases can be questionable. Customers with a high average monetary value per order probably have a lower transaction rate, as the spend more per transaction than customers who have a low monetary value, indicating that there is a dependence present. The hierarchical Bayes model drops the first assumption and Pareto/Dependent model the second one, which could make the combination of them in order to predict the customer lifetime value more flexible.

The results showed that this combination did not outperform the 'classical' combination of the models where the Pareto/NBD and the gamma-gamma model are used. The new combination had less prediction accuracy than the others. A reason for this could be that the assumptions, which were dropped by the two extension models, have not been hardly violated by the data. This is the case when the data set contains no correlations between various measures, which was assumed by the models. Therefore, the assumptions the two classical models assumed are still valid and the hierarchical Bayes and Pareto/Dependent model would no longer outperform the other models.

Another explanation could be that a lot of customers which were present in the training set were not present in the test set. This would imply that their actual value of the frequency, monetary value and customer lifetime value is zero. This is difficult for the models to capture. They do take a probability of being alive into account, but predicting some many non-repeat buyers seems to be really hard. When predicting the number of future transactions this would not be a huge problem, as a prediction of someone's future purchase behavior are low (average of the prediction models was around 1.5). In contrast, when the customer lifetime value has to be predicted, the buying frequency is multiplied by the monetary value of the bought items, where the average of the monetary value without zeros is about 200 euro. If the actual frequency is zero, the CLV is zero, but when the predicted frequency is 1.5, the CLV becomes 300. The large difference between the predicted and actual values can lead to a drop of the accuracy of the overall model prediction.

Further, Winkelstraat.nl focusses on a specific age group. Morisada et al. (2019) showed that using their data set from an online fashion platform, customers under 30 years old are mostly having a low transaction rate and a long lifetime, indicating that they would not buy anytime soon, but will buy on a later moment. Winkelstraat.nl's main focus is on young buyers ranging between 16 and 36 years old. This would suggest that the customers of Winkelstraat.nl are not repeat buying any time soon. Therefore, with a test period of two years, it could be the case that a lot of customers are not buying in the test set, suggesting they have left the company, but are actually customers who have a very long lifetime. If this test set is extended with more years of data, the predictions could be more accurate.

A different customer buying frequency could also be due to a change in the supply of Winkelstraat.nl. As they have been growing ever since they first started in October 2012, their assortment has been growing as well. This could attract more customers and therefore increase the revenue, but it could also change a customer's buying behavior. Predicting a varying buying behavior is very hard and could lead to poor predictions performance.

Finally, Winkelstraat.nl is an online webshop for luxury and designer clothing. To our knowledge, only a few recent papers have discussed customer lifetime value for the online fashion industry and no empirical research has been developed about the online luxury fashion shops. This could be a dataset which is very hard to predict, as the items presented by Winkelstraat.nl can be seen as expensive ones, making the buying pattern more volatile compared to cheaper clothing. This could be a reason for customers to buy these items on variable moments of time, which makes a prediction about the future buying behavior very hard.

9 Conclusion

Customers are preferring more and more to purchase items online as it is more convenient, there is more product variety and products can be easily compared (Jiang et al., 2013). This has increased the number of e-commerce webshops over the past years. One of them is Winkelstraat.nl, a Dutch online fashion platform with items from the premium and luxury segment as designer clothing, shoes, bags and other accessories. They want, as any other firm, to increase their profit. This can be done by satisfying lots of customers, but especially the most profitable ones, as they bring the most revenue to a company. That is why this research wants to answer the research question: "Which customers will be the most profitable ones in the future for Winkelstraat.nl?" Having an answer to this question can result in increasing their revenue, as they can lay more focus on these customers with, e.g., allocating more of their marketing resources on them.

In order to retrieve who the most profitable customers will be in the future, the customer lifetime value can be predicted. This value reports the revenue the customer brings to a company over their lifetime with the firm. This value can be predicted using a prediction of the number of transactions a customer will make in the future multiplied by the average predicted monetary value of these transactions. These two values can be predicted using prediction models and by combining these models, the customer lifetime value can be calculated. This research proposes a new combination of two existing models, namely a hierarchical Bayes model to predict the future number of transactions and a model which predicts a time-varying monetary value, the Pareto/Dependent model. This combination is tested against 'classical' Buy-Till-You-Die models, such as the Pareto/NBD and gamma-gamma model, to report the accuracy of this combination. The second research question therefore is reported as: "Does a combination of a hierarchical Bayes framework with a time-varying monetary value better predict the customer lifetime value than the classical predictors?" This combination is new to the literature, such that answering this question is also relevant for the current marketing literature. By using the models as they were specified in the relevant papers and by using data available by Winkelstraat.nl, this research obtained the desired results and compared them using the root mean squared error (RMSE) and the mean absolute error (MAE) as prediction accuracy measures.

The data set used during this research did not have significant dependence between a customer's transaction and death rates, namely a correlation of 0.0065. This would imply that this assumption is valid and that the Pareto/NBD model, which assumes this assumption, and the hierarchical Bayes model, which drops this assumption, perform relatively the same. On the other hand, the data did consist of a dependence between the buying frequency and monetary value of the purchases, with a value of 0.097, which is significantly different from zero. This suggests that the model which drops this assumption (Pareto/Dependent model) could have a better fit than the gamma-gamma model. This could also be seen in Section 7, where the results are given. The Pareto/NBD model outperforms the hierarchical Bayes model on both prediction measures (RMSE and MAE). The gamma-gamma model has better predictions than the Pareto/Dependent when one only looks at the prediction of the monetary value, but the Pareto/Dependent model seems to better predict the customer lifetime value, independent of which combination with a frequency model is used. In order to answer the research question whether the new model combination performs better, we can conclude that when using this data set, it is better to use the Pareto/NBD model in combination with the Pareto/Dependent model, because this combination gives the best predictions. The most profitable customers can be obtained by using this combination to express the customer lifetime value and take the customers with the highest scores. One should keep in mind that despite the fact that the hierarchical Bayes model did not predict better than the Pareto/NBD model, it does give valuable insights in the customer characteristics due to the covariates it can incorporate. It is the choice of the practitioner to get the desired results; better predictions or insights in the customers characteristics.

Further research could be developed on customer churn. Our research showed that a lot of customers are 'death' in the future, which led to wrong predictions. We think that the luxury online webshops will encounter a lot of one/two time buyers, as the prices are high. That is why a model could be developed where more attention is given on the customer churn process. Lemmens and Gupta (2020) proposed a profit-based loss function in order to rank customers based on the incremental impact of the intervention of churn. By laying more focus on the customer churn, they concluded that their model leaded to significantly more profitable campaigns than competing models. Another direction of future research could be to perform the same research, using the same combination of the hierarchical Bayes model with a time-varying monetary value, but by using another data set. A data set which has more years to train and test this combination, such that the purchase history consist of more information and the predictions are made over a longer test period, leading to less customers to 'die'. And/or a data set which harder violates the assumptions the 'classical' models assume, making the models which drop these assumptions more flexible. This research could also be developed on the non-luxury e-commerce market to see whether the newly proposed combination does work in that scenario. Finally, this research can be developed on a data set which also focuses on the age group above 30 years old, because Winkelstraat.nl's main focus is the age group ranging between 16 and 36 years old. Morisada et al. (2019) concluded that older customers are easier to predict, as they are less occasional buyers.

10 Acknowledgement

We thank Winkelstraat.nl for the availability of their data set for this research and for the opportunity to perform this research while having supervision from their employees. Further, we want to thank prof. dr. ir. Rommert Dekker for his guidance and expertise involving this paper. His supervision helped this research to constantly grow.

References

Abe, M. (2009). "Counting Your Customers" One by One: A Hierarchical Bayes Extension to the Pareto/NBD Model. *Marketing Science*, 28(3):541–553.

Abramowitz, M. and Stegun, I. A. (1970). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, volume 55. US Government printing office.

Avinash, A., Sahu, P., and Pahari, A. (2019). Big Data Analytics for Customer Lifetime Value Prediction. *Telecom Business Review*, 12(1):46.

Berger, P. D. and Nasr, N. I. (1998). Customer Lifetime Value: Marketing Models and Applications. *Journal of interactive marketing*, 12(1):17–30.

Borle, S., Singh, S. S., and Jain, D. C. (2008). Customer Lifetime Value Measurement. *Management science*, 54(1):100–112.

Buttle, F. and Maklan, S. (2019). *Customer Relationship Management: Concepts and Technologies*. Routledge.

Dwyer, F. R. (1997). Customer Lifetime Valuation to Support Marketing Decision Making. *Journal of Direct Marketing*, 11(4):6–13.

Ehrenberg, A. S. (1959). The Pattern of Consumer Purchases. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 8(1):26–41.

Fader, P. S. and Hardie, B. G. (2005). A Note on deriving the Pareto/NBD model and related Expressions. *accessed July*, 31:2005.

Fader, P. S. and Hardie, B. G. (2013). The Gamma-Gamma model of Monetary Value.

Fader, P. S., Hardie, B. G., and Hardie, B. G. (2007). Incorporating Time-Invariant Covariates into the Pareto/NBD and BG/NBD models.

Fader, P. S., Hardie, B. G., and Lee, K. L. (2005a). RFM and CLV: Using Iso-Value Curves for Customer Base Analysis. *Journal of marketing research*, 42(4):415–430.

Fader, P. S., Hardie, B. G., and Lee, K. L. (2005b). "Counting Your Customers" The Easy Way: An Alternative to the Pareto/NBD model. *Marketing science*, 24(2):275–284.

Geweke, J. et al. (1991). *Evaluating the Accuracy of Sampling-based Approaches to the Calculation of Posterior Moments*, volume 196. Federal Reserve Bank of Minneapolis, Research Department Minneapolis, MN.

Glady, N., Baesens, B., and Croux, C. (2009). A Modified Pareto/NBD Approach for Predicting Customer Lifetime Value. *Expert Systems with Applications*, 36(2):2062–2071.

Gupta, S., Lehmann, D. R., and Stuart, J. A. (2004). Valuing Customers. *Journal of marketing research*, 41(1):7–18.

Jain, D. and Singh, S. S. (2002). Customer Lifetime Value Research in Marketing: A Review and Future Directions. *Journal of interactive marketing*, 16(2):34.

Jasek, P., Vrana, L., Sperkova, L., Smutny, Z., and Kobulsky, M. (2018). Modeling and Application of Customer Lifetime Value in Online Retail. In *Informatics*, volume 5, page 2. Multidisciplinary Digital Publishing Institute.

Jerath, K., Fader, P. S., and Hardie, B. G. (2011). New Perspectives on Customer "Death" using a Generalization of the Pareto/NBD Model. *Marketing Science*, 30(5):866–880.

Jiang, L. A., Yang, Z., and Jun, M. (2013). Measuring Consumer Perceptions of Online Shopping Convenience. *Journal of Service Management*, 24(2):191–214.

Korkmaz, E., Fok, D., and Kuik, R. (2014). The Need for Market Segmentation in Buy-Till-You-Defect Models. *ERIM Report Series Reference No. ERS-2014-006-LIS*.

Korkmaz, E., Kuik, R., and Fok, D. (2013). " Counting Your Customers": When will they buy next? An Empirical Validation of Probabilistic Customer Base Analysis models based on Purchase Timing. Technical report, Erasmus Research Institute of Management.

Langa, L. (2020). Python (version 3.9.1).

Lemmens, A. and Gupta, S. (2020). Managing Churn to Maximize Profits. *Marketing Science, Forthcoming*.

Ma, S.-H. and Liu, J.-L. (2007). The MCMC Approach for solving the Pareto/NBD model and possible Extensions. In *Third international conference on natural computation (ICNC 2007)*, volume 2, pages 505–512. IEEE.

Morisada, M., Miwa, Y., and Dahana, W. D. (2019). Identifying Valuable Customer Segments in Online Fashion Markets: An Implication for Customer Tier Programs. *Electronic Commerce Research and Applications*, 33:100822.

Mzoughia, M. B. and Limam, M. (2014). An Improved BG/NBD Approach for Modeling Purchasing Behavior using COM-Poisson Distribution. *International Journal of Modeling and Optimization*, 4(2):141.

Pfeifer, P. E. (2005). The Optimal Ratio of Acquisition and Retention Costs. *Journal of Targeting, Measurement and Analysis for Marketing*, 13(2):179–188.

Platzer, M. (2016). Customer Base Analysis with BTYDplus.

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Reinartz, W. J. and Kumar, V. (2000). On the Profitability of Long-Life Customers in a Noncontractual Setting: An Empirical Investigation and Implications for Marketing. *Journal of marketing*, 64(4):17–35.

Reinartz, W. J. and Kumar, V. (2003). The Impact of Customer Relationship Characteristics on Profitable Lifetime Duration. *Journal of marketing*, 67(1):77–99.

Rossi, P. E. and Allenby, G. M. (2003). Bayesian Statistics and Marketing. *Marketing Science*, 22(3):304–328.

Schmittlein, D. C., Morrison, D. G., and Colombo, R. (1987). Counting Your Customers: Who are they and what will they do next? *Management science*, 33(1):1–24.

Schmittlein, D. C. and Peterson, R. A. (1994). Customer Base Analysis: An Industrial Purchase Process Application. *Marketing Science*, 13(1):41–67.

Singh, S. S., Borle, S., and Jain, D. C. (2009). A Generalized Framework for Estimating Customer Lifetime Value when Customer Lifetimes are not observed. *QME*, 7(2):181–205.

Srivastava, H., Çetinkaya, A., and Kıymaz, I. O. (2014). A Certain Generalized Pochhammer symbol and its Applications to Hypergeometric Functions. *Applied Mathematics and Computation*, 226:484–491.

Thomas, J. S. (2001). A Methodology for Linking Customer Acquisition to Customer Retention. *Journal of marketing research*, 38(2):262–268.

Xie, S.-M. and Huang, C.-Y. (2020). Systematic Comparisons of Customer Base Prediction Accuracy: Pareto/NBD versus Neural Network. *Asia Pacific Journal of Marketing and Logistics*.

A Appendix

A.1 Estimation Pareto/NBD model

A.1.1 Model Parameters

As was described in Section 4.1, the Pareto/NBD model can be estimated by first estimating the model parameters (r, α, s, β). The estimation of these parameters is based on the derivation of Fader and Hardie (2005). The likelihood that has to be maximized is as followed

$$L(r,\, \alpha,\, s,\, \beta) = \prod_{i=1}^{N} P(X_i = x_i,\, t_i,\, T_i \,|\, r,\, \alpha,\, s,\, \beta). \tag{A.1}$$

To maximize this equation, one needs to compute the probability model first. This can be done by first determining whether customer i is alive at time T and if not, what value of τ did he/she become inactive. This results in the following expression

$$\begin{aligned} &P(X_i = x_i,\, t_i,\, T_i \,|\, r,\, \alpha,\, s,\, \beta) = \\ &\quad \int_0^\infty \int_0^\infty P(X = x,\, t,\, T|\, \lambda,\, \mu,\, \tau > T)\, P(\tau > T)\, g(\lambda|\, r,\, \alpha) h(\mu|\, s,\, \beta)\, d\lambda\, d\mu + \\ &\quad \int_0^\infty \int_0^\infty P(X = x,\, t,\, T|\, \lambda,\, \mu,\, t < \tau < T)\, P(t < \tau < T)\, g(\lambda|\, r,\, \alpha) h(\mu|\, s,\, \beta)\, d\lambda\, d\mu, \end{aligned} \tag{A.2}$$

where $g(\lambda|\, r,\, \alpha)$ and $h(\mu|\, s,\, \beta)$ are given by the assumptions given in Section 4.1 (A4 and A5 respectively).

This results in four new probabilities. Using Assumption 3 about the unobserved lifetime of a customer, which is exponentially distributed, one can give an expression for the probability that a customer is alive after time period T

$$P(\tau > T|\, \lambda,\, \mu) = e^{-\mu T}, \tag{A.3}$$

and an expression for the probability that a customer is inactive within the time period $t < \tau < T$

$$P(t < \tau < T|\, \lambda,\, \mu) = e^{-\mu t} - e^{-\mu T}. \tag{A.4}$$

The other two probabilities can be expressed by first looking at the time to get the x-th purchase. Assumption 2 gives the interpurchase time between two transactions by means of an exponential distribution. By summing all these interpurchase times until the x-th transaction, a sum of i.i.d exponential distributions has to be computed, which results in a gamma distribution with the same rate parameter. This gives $\frac{\lambda^x}{\Gamma(x)} t^{x-1} e^{\lambda t}$. Further, we need the probability that no purchase happens in the time period $(t, T]$. This can be obtained by again using Assumption 2, where the Poisson distribution can be used in the period $(t, T]$ with no purchases $(x = 0)$. This results in $e^{-\lambda(T-t)}$. Using this explanation, one ends up with

$$\begin{aligned} P(X = x,\, t,\, T|\, \lambda,\, \mu,\, \tau > T) &= \frac{\lambda^x}{\Gamma(x)} t^{x-1} e^{\lambda t} * e^{-\lambda(T-t)} \\ &= \frac{\lambda^x t^{x-1} e^{-\lambda T}}{\Gamma(x)}. \end{aligned} \tag{A.5}$$

For the last probability function the case when a customer became inactive within the time period $t < \tau < T$ has to be evaluated. By using the Law of Total Probability and the probability of τ, as stated in Assumption 3 ($\mu e^{-\mu\tau}$), we end up with

$$\begin{aligned} P(X = x,\, t,\, T|\, \lambda,\, \mu,\, t < \tau < T) &= \int_t^T \frac{\lambda^x}{\Gamma(x)} t^{x-1} e^{\lambda t} * e^{-\lambda(\tau-t)} * \mu e^{-\mu\tau} d\tau \\ &= \int_t^T \frac{\lambda^x t^{x-1} e^{-\lambda\tau}}{\Gamma(x)} \mu e^{-\mu\tau} d\tau. \end{aligned} \tag{A.6}$$

Now one can obtain the likelihood function in Equation A.1 by implementing the previous four equations in Equation A.2, and implementing this in the likelihood function. This difficult calculation has been done by Fader and Hardie (2005) and concluded that the likelihood looks as follows

$$L(r,\, \alpha,\, s,\, \beta) = \frac{\Gamma(r+x)\alpha^r \beta^s}{\Gamma(r)} \Big(\frac{1}{(\alpha+T)^{r+x}(\beta+T)^s} + (\frac{s}{r+s+x}) A_0 \Big) \tag{A.7}$$

with $\alpha \geq \beta$ [4] and

$$A_0 = \frac{F(a_i, b; c_i; z(t_i))}{(\alpha+t_i)^{r+s+x_i}} - \frac{F(a_i, b; c_i; z(T_i))}{(\alpha+T_i)^{r+s+x_i}} \tag{A.8}$$

where,

$$\begin{aligned} a_i &= r + x_i + s; & b &= s + 1; \\ c_i &= r + x_i + s + 1; & z(y) &= \frac{\alpha-\beta}{\alpha+y}; \end{aligned}$$

$$F(a, b; c; z) = \sum_{j=0}^{\infty} \frac{(a)_j (b)_j}{(c)_j} \frac{z^j}{j!},$$

where F is the Gaussian hypergeometric function (Abramowitz and Stegun, 1970) and $(a)_j$ is the Pochhammer's symbol (Srivastava et al., 2014).

In the end, a maximization algortihm (MLE) has to be evaluated on the likelihood function in Equation A.7 to get estimates for the model parameters.

[4] For the case where $\alpha < \beta$ the function differs slightly. See Fader and Hardie (2005) for details.

A.1.2 Probability being Alive

Once the model parameters are estimated, the probability for a customer to be alive can be calculated. Given the information about the customers, this probability is expressed as follows

$$P(alive|x,t,T) = \int_0^\infty \int_0^\infty P(\tau > T|\lambda,\mu,x,t,T)g(\lambda,\mu|r,\alpha,s,\beta,x,t,T)d\lambda d\mu \tag{A.9}$$

Using the Bayes Theorem, the second term on the right hand side can be calculated as follows

$$g(\lambda,\mu|r,\alpha,s,\beta,x,t,T) = \frac{L(\lambda,\mu|x,t,T)g(\lambda|r,\alpha)g(\mu|s,\beta)}{L(r,\alpha,s,\beta|x,t,T)} \tag{A.10}$$

Finally, using the fact that the denominator of this equation is the likelihood function stated in Equation A.7, we end up with

$$P(alive|x_i,t_i,T_i) = 1\Big/\Big(1 + \frac{\hat{s}}{\hat{r}+\hat{s}+x}(\hat{\alpha}+T_i)^{\hat{r}+x_i}(\hat{\beta}+T_i)^{\hat{s}}A_0\Big) \tag{A.11}$$

where A_0 is defined in Equation A.8.

A.1.3 Number of future transactions

Using the same parameter values as before, the Pareto/NBD model also provides an expectation for the number of transactions of a customer. The (unconditional) expectation is given as

$$E(X(t)) = \int_0^\infty \int_0^\infty E(X(t)|\lambda,\mu)g(\lambda|r,\alpha)g(\mu|s,\beta)d\lambda d\mu \tag{A.12}$$

where the two density functions are produced by the assumptions. The expectation can be split up in two parts. The first scenario would be that $\tau > t$, in which the expected number of transactions simply becomes λt. The other scenario would be that $\tau \leq t$, which would suggest that the number of transactions in the interval $(0,t]$ is $\lambda\tau$. This results in the following

$$\begin{aligned} E(X(t)|\lambda,\mu) &= \lambda t P(\tau > t|\mu) + \int_0^t \lambda\tau f(\tau|\mu)d\tau \\ &= \lambda t e^{-\mu t} + \frac{\lambda}{\mu}\int_0^t \mu^2\tau e^{-\mu\tau}d\tau \end{aligned} \tag{A.13}$$

Now one can end up with the (unconditional) expectation for the number of transactions.

$$E(X(t)) = \frac{\hat{r}\hat{\beta}}{\hat{\alpha}(\hat{s}-1)}\Big[1 - \Big(\frac{\hat{\beta}}{\hat{\beta}+t}\Big)^{\hat{s}-1}\Big] \tag{A.14}$$

For the CLV calculation, the expectation conditional with the data should be evaluated. This is done as

$$\begin{aligned} E(Y(t)|r,\alpha,s,\beta,x,t,T) = \int_0^\infty \int_0^\infty \Big[\quad & E(Y(t)|\lambda,\mu,aliveatT)P(\tau > T|\lambda,\mu,x,t,T) \\ & g(\lambda,\mu|r,\alpha,s,\beta,x,t,T)\Big]d\lambda d\mu \end{aligned} \tag{A.15}$$

Now implementing the different equations stated above to end up with

$$\begin{aligned} E(Y(t)|r,\alpha,s,\beta,x,t,T) = x + & \frac{\Gamma(\hat{r}+x)\hat{\alpha}^{\hat{r}}\hat{\beta}^{\hat{s}}}{\Gamma(\hat{r})(\hat{\alpha}+T)^{\hat{r}+x}(\hat{\beta}+T)^{\hat{s}}\hat{L}} \\ & * \frac{(\hat{r}+x)(\hat{\beta}+T)}{(\hat{\alpha}+T)(\hat{s}-1)}\Big[1-\Big(\frac{\hat{\beta}+T}{\hat{\beta}+T+t}\Big)^{\hat{s}-1}\Big] \end{aligned} \quad \text{(A.16)}$$

One can see that the bracket term is the expression for the probability of being alive as stated in Equation A.11. The rest of the formula is the unconditional expectation of the Pareto/NBD model with 'updated' parameters that for the individual behavior up to time T (assuming no death in $(0, T])$.