Predicting Customer Lifetime Value
Understanding its accuracy and drivers from a frequent flyer program perspective

Kevin Do Ruibin
Tobias Vintilescu Borglöv

Supervisor: Christina Grundström
Examinator: Thomas Rosenfall
Copyright

The publishers will keep this document online on the Internet – or its possible replacement – for a period of 25 years starting from the date of publication barring exceptional circumstances.

The online availability of the document implies permanent permission for anyone to read, to download, or to print out single copies for his/hers own use and to use it unchanged for non-commercial research and educational purpose. Subsequent transfers of copyright cannot revoke this permission. All other uses of the document are conditional upon the consent of the copyright owner. The publisher has taken technical and administrative measures to assure authenticity, security and accessibility.

According to intellectual property law the author has the right to be mentioned when his/her work is accessed as described above and to be protected against infringement.

For additional information about the Linköping University Electronic Press and its procedures for publication and for assurance of document integrity, please refer to its www homepage: http://www.ep.liu.se/.

© Kevin Do Ruibin and Tobias Vintilescu Borglöv
Abstract

Each individual customer relationship represents a valuable asset to the firm. Loyalty programs serve as one of the key activities in managing these relationships and the well-developed frequent flyer programs in the airline industry is a prime example of this. Both marketing scholars and practitioners, though, have shown that the linkage between loyalty and profit is not always clear. In marketing literature, customer lifetime value is proposed as a suitable forward-looking metric that can be used to quantify the monetary value that customers bring back to the firm and can thus serve as a performance metric for loyalty programs. To consider the usefulness of these academic findings, this study has evaluated the predicted airline customer lifetime value as a loyalty program performance metric and evaluated the drivers of customer lifetime value from a frequent flyer program perspective.

In this study, the accuracy of the Pareto/NBD Gamma-Gamma customer lifetime value has been evaluated on a large dataset supplied by a full-service carrier belonging to a major airline alliance. By comparing the accuracy to a managerial heuristic used by the studied airline, the suitability as a managerial tool was determined. Furthermore, based on existing literature, the drivers of customer lifetime value from a frequent flyer perspective were identified and analyzed through a regression analysis of behavioral data supplied by the studied airline.

The analysis of the results of this study shows that the Pareto/NBD customer lifetime value model outperforms the managerial heuristic in predicting customer lifetime value in regard to almost all error metrics that have been calculated. At an aggregate-level, the errors are considered small in relation to average customer lifetime value, whereas at an individual-level, the errors are large. When evaluating the drivers of customer lifetime value, points-pressure, rewarded-behavior, and cross-buying have a positive association with customer lifetime value.

This study concludes that the Pareto/NBD customer lifetime value predictions are only suitable as a managerial tool on an aggregate-level. Furthermore, the loyalty program mechanisms studied have a positive effect on the airline customer lifetime value. The implications of these conclusions are that customer lifetime value can be used as a key performance indicator of behavioral loyalty, but the individual-level predictions should not be used to allocate marketing resources for individual customers. To leverage the drivers of customer lifetime value in frequent flyer programs, cross-buying and the exchange of points for free flights should be facilitated and encouraged.
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>BTYD</td>
<td>Buy ‘Till You Die</td>
</tr>
<tr>
<td>CLV</td>
<td>Customer Lifetime Value</td>
</tr>
<tr>
<td>CE</td>
<td>Customer Equity</td>
</tr>
<tr>
<td>CRM</td>
<td>Customer Relationship Management</td>
</tr>
<tr>
<td>FFP</td>
<td>Frequent Flyer Program</td>
</tr>
<tr>
<td>FSC</td>
<td>Full-Service Carrier</td>
</tr>
<tr>
<td>FvA</td>
<td>Forecast vs Actual</td>
</tr>
<tr>
<td>LCC</td>
<td>Low-Cost Carrier</td>
</tr>
<tr>
<td>LP</td>
<td>Loyalty Program</td>
</tr>
<tr>
<td>MAE</td>
<td>Mean Absolute Error</td>
</tr>
<tr>
<td>MAPE</td>
<td>Mean Absolute Percentage Error</td>
</tr>
<tr>
<td>MSLE</td>
<td>Mean Squared Logarithmic Error</td>
</tr>
<tr>
<td>RMSE</td>
<td>Root Mean Squared Error</td>
</tr>
<tr>
<td>SRQ</td>
<td>Specified Research Question</td>
</tr>
<tr>
<td>VIF</td>
<td>Variation Inflation Factor</td>
</tr>
</tbody>
</table>
Table of Figures and Tables

Figure 1. Antecedents and effects of loyalty based on Watson et al. (2015) ........................................ 3
Figure 2. Structure of chapter 2. ............................................................................................................. 5
Figure 3. Classification of customer-bases (Fader and Hardie, 2009) ...................................................... 7
Figure 4. How reward program mechanisms influence retention and purchase volume (Blattberg, Kim and Neslin, 2008, p.550). ........................................................................................................ 16
Figure 5. The effect of points-pressure and rewarded-behavior around the redemption event (Taylor and Neslin, 2005, p.294) ........................................................................................................... 17
Figure 6. Conceptual framework of CLV drivers according to Blattberg et al. (2009, p.158) ................. 18
Figure 7. Analytical model including the calculation of Customer Lifetime Value and its drivers .21
Figure 8. The research process split into seven phases ......................................................................... 24
Figure 9. An overview of the methodology concerning the calculations of Customer Lifetime Value, ........................................................................................................................................... 24
Figure 10. Data split for CLV calculations .............................................................................................. 27
Figure 11. Scatterplot of residuals. ......................................................................................................... 33
Figure 12. Histogram of the residuals. ..................................................................................................... 34

Table 1. An overview of customer-base analysis models (Fader and Hardie, 2009) .................. 7
Table 2. An overview of aggregate-level error measures used by the literature. ....................... 9
Table 3. An overview of individual-level error measures used by the literature ......................... 10
Table 4. An overview of performance metrics from previous studies .............................................. 11
Table 5. A typology of LP types adopted with Swedish examples (Berman, 2006) .................. 12
Table 6. Synthesis of the different views on program types and their corresponding reward type (Berman, 2006; Blattberg, Kim and Neslin, 2008) ................................................................................................. 12
Table 7. Overview of frequent flyer program types (de Boer and Gudmundsson, 2012, p.23) ...... 13
Table 8. Description of the dataset used for Customer Lifetime Value calculations .................. 27
Table 9. Engineered features with definitions used for the regression analysis ................................ 31
Table 10. Descriptive statistics of the independent variables used in the regression analysis .. 31
Table 11. Model summary of the regression model .............................................................................. 32
Table 12. Correlations of the dependent and independent variables .................................................. 35
Table 13: Collinearity statistics of the regression model ...................................................................... 35
Table 14. Individual-level error metric in percent of average holdout CLV .................................. 38
Table 15. Forecast vs Actual error metric as a percentage of the average holdout CLV ............... 38
Table 16. Standardized coefficients from the regression analysis ...................................................... 39
1 Introduction

This chapter provides the reader with an introduction consisting of the background to the problem, the problem itself and the purpose of this study.

1.1 The Importance of Profitable Customers and Loyalty Programs

There has been an increasing realization over the past decades that existing customers represent a valuable asset to the firm and that some customers are more valuable than others (Lindgreen and Wynstra, 2005). Much of today’s marketing efforts concerns how to best deliver value to the firms’ customers (Kumar and Reinartz, 2016). Of equal importance, though, is the value that the customers bring back to the firm (Ramsay, 2005). Along these lines, balancing the efforts of marketing initiatives with the return given by the customers is an important strategic prioritization for the successful companies of today (Kumar and Reinartz, 2016).

A firm’s loyalty program (LP) is a Customer Relationship Management (CRM) tool that can identify, reward and successfully retain profitable customers (Kumar and Reinartz, 2012). In industries such as the airline industry, the value derived from customers and the cost of serving them are largely heterogeneous, thus calling for a focus on value alignment (Kumar and Reinartz, 2012). A prime example of LPs attempting to match value to and from customers is the Frequent Flyer Programs (FFP) that came into existence over 30 years ago when American Airlines launched their AAdvantage program (de Boer and Gudmundsson, 2012). These early programs lead to the expansions of LPs in other industries such as hotels, rental cars, and financial services (Vinod, 2011). The FFPs often offer a wide range of benefits and products to frequent flyers, including miles from flying and partners, preferential treatment, lounge access, upgrades, credit cards, and more (Scandinavian Airlines System, 2018; American Airlines, 2018; Lufthansa, 2018).

At the same time, the benefits of such programs from a corporate perspective are not always clear as they require major investments and the link between loyalty and profitability is often weaker than expected (Kumar and Reinartz, 2012; Reinartz and Kumar, 2002; McKinsey, 2013). Furthermore, few airline executives say that they have fully understood the relationship between financial results and loyalty (Deloitte, 2013). Presumably, the lack of an appropriate measure to follow up on could be the cause for this. There has been research conducted on effective LP design, but there are a limited number of empirical studies on the resulting effects on the performance of the LPs (Kumar and Reinartz, 2012; McCall and Voorhees, 2010; Watson et al., 2015). The forward-looking metric Customer Lifetime Value (CLV) is proposed as a good measure and tool for managing value from customers (Kumar and Reinartz, 2016). As such, CLV can be the link between loyalty and financial results.

1.2 Problem Analysis

According to Vinod (2011), LPs can be the key to new revenue growth if data intelligence and marketing programs are designed and leveraged to its full potential. The use of data intelligence to enhance marketing efficiency can be defined as database marketing and one of the cornerstones of database marketing is, in turn, CLV which can be used to both to diagnose the health of a business and to assist in making tactical decisions (Blattberg, Kim and Neslin, 2008). Due to the abundance of customer data accessible to firms, forward-looking metrics such as CLV and the associated customer analysis can contribute to sustainable competitive advantage (Erevelles, Fukawa and Swayne, 2016; Kumar and Reinartz, 2016).

Since the advent of FFPs, significant development has been made leading up to the programs of today. Most notably, FFPs has transitioned from pure cost-centers to value-adding profit-centers for airline companies.
carriers (de Boer and Gudmundsson, 2012). Due to the high degree of competition, consolidation, and volatility in the airline industry, FFPs play an increasingly important role for carriers to gain a competitive advantage through revenue growth and profit (Liu and Yang, 2009). Since firms can build huge databases of customer data through their LPs, they serve as an important source of valuable customer information that can be used for future marketing activities, such as activities aimed at increasing a customer’s total purchases from the firm (Basso, Clements and Ross, 2009; Berman, 2006).

1.2.1 Capturing Customer Lifetime Value

Kumar and Reinartz (2012, p.4) define the value from customers as the economic value of the customer relationship to the firm – expressed on the basis of contribution margin or net profit. The forward-looking measure CLV is proposed as a measure that captures both the nature of the customers' behaviors and their contributions and is thus a suitable metric for measuring value from customers (Kumar and Reinartz, 2016). Pfeifer and Bang (2005, p.49) define CLV as the present value of all the future cash flows attributed to a customer relationship, whereas a different wording along the same lines is given by Kumar and Reinartz (2016, p.42) who define CLV as the present value of future profits generated from a customer over his or her lifetime. Both CLV definitions make it clear that CLV in addition to being a behavioral metric, also deals with the monetary value of a customer. With the assistance of probability models used to predict individual customers’ CLV, it is today possible to achieve excellent results with commonly available software packages such as Microsoft Excel (Fader, Hardie and Shang, 2010; Fader and Hardie, 2009). With an individual-level metric such as CLV, it is then possible to facilitate decisions on costs, increases in revenues and profits, return on investment, customer acquisition and retention, and realigning marketing resources (Kumar and Reinartz, 2016). Furthermore, Malthouse (2013) argues that CLV provides the best rationale for allocating marketing resources.

One research field of CLV analysis is investigating the effects of LPs on CLV and firm’s profitability (Jain and Singh, 2002). LPs can benefit from a forward-looking metric such as CLV as it can guide decisions on allocation of marketing activities to align value from and to customers (Blattberg, Kim and Neslin, 2008; Kumar and Reinartz, 2016). Research has been conducted by Blattberg, Malthouse and Neslin (2009) in regards to how mechanisms in LPs affect profits. The authors conclude that there is empirical support for these mechanisms, but that more evidence is needed to verify and quantify their impact on CLV. Furthermore, Kumar and Reinartz (2012) suggest that LPs can be a viable tool to use in order to align value to and from customers.

1.2.2 The Drivers of Customer Lifetime Value in the Context of Loyalty Programs

A firm’s strategic opportunities might be best viewed in terms of the firm’s opportunity to improve the drivers of its customer equity (Rust, Lemon and Zeithaml, 2004). The linkage between CLV and Customer Equity (CE) is that CE is the sum of lifetime value of all customers of a firm (Kumar and Reinartz, 2016). Thus, understanding the drivers of CE also means to understand the drivers of CLV. This enables firms to be truly customer-centered and CLV provides the tools for making strategic marketing decisions inherently information driven (Rust, Lemon and Zeithaml, 2004). Furthermore, Kumar and Reinartz (2016) note that understanding the drivers of customer value is needed to translate the results of CLV models to managerial decision making. Multiple studies have tried to find drivers of CLV and as more studies are conducted, the list of drivers can be expected to change (Kumar and Reinartz, 2016).

As stated earlier in 1.1, the linkage between profit and loyalty is not always clear. Kumar and Reinartz (2002) for instance, found a weak correlation between loyalty and profitability in an empirical study. Several authors state that true loyalty might only be achieved by a combination of attitudinal and
behavioral loyalty (Bijmolt, Dorotic and Verhoef, 2011; Kumar, 2008), but it is important to note that loyalty does not equal profitability. Loyalty and the actual performance of an LP are therefore two different subjects (Blattberg, Kim and Neslin, 2008). As an extreme standpoint to this, Blattberg, Kim and Neslin (2008) are even apprehensive on the usage of the term “loyalty program”, instead they opt to use the terms reward programs as a way to discriminate between loyalty and performance.

According to Watson et al. (2015), the conflicting results in the present literature regarding the effects of loyalty on performance (sales, share of wallet, profit performance, etc.) is due to the common research practice of using single-element measures on either attitudinal or behavioral loyalty as a metric of overall program performance. Attitudinal loyalty relates to factors such as the level of commitment, favorable attitudes, and positive affect while behavioral loyalty relates to the customer's purchase behaviors (Bijmolt, Dorotic and Verhoef, 2011; Blattberg, Kim and Neslin, 2008; Watson et al., 2015). Therefore, it is of interest to determine how to measure the effectiveness of LPs and FFPs. An interpretation of the antecedents and effects of loyalty by Watson et al. (2015) is illustrated in Figure 1.

![Figure 1. Antecedents and effects of loyalty based on Watson et al. (2015).](image)

Watson et al. (2015) have identified four antecedents to attitudinal and behavioral loyalty: (1) commitment, (2) trust, (3) satisfaction, and (4) loyalty incentives. These differentially affect attitudinal and behavioral loyalty. For instance, satisfaction has little effect on behavioral loyalty but a strong effect on attitudinal loyalty. The different design components of an LP can be seen to influence these four antecedents through different mechanisms; for instance, points-pressure might build commitment, relationship duration might build trust, preferential treatment might build satisfaction, and rewards are typical loyalty incentives. Furthermore, the authors found that attitudinal and behavioral loyalty affects word of mouth and performance differentially where attitudinal loyalty has a stronger connection to word of mouth and behavioral loyalty has a stronger connection to program performance. Thus, the task to measure how LPs affect profit performance points towards a focus on behavioral loyalty. This is in line with research by Qi et al. (2012), who conclude that loyalty is a driver of CLV while customer satisfaction is not.

The majority of existing LPs reward customers based on their behavioral loyalty with the simple tenet that the more you spend, the greater the reward (Kumar, 2008). The effects of an LP on customer behavior and attitudes may proceed through three mechanisms, of which points-pressure and rewarded-behavior mechanisms are directly related to reward redemption. For the rewarded-behavior effects, an LP member must redeem points for a reward; for points-pressure effects, he or she must value the reward (Bijmolt, Dorotic and Verhoef, 2011). The effects of these behavioral mechanisms lead to increased customer retention and customer expenditures (Bijmolt, Dorotic and Verhoef, 2011; Blattberg, Kim and Neslin, 2008). In a quantitative study, Gupta, Lehmann and Stuart (2004) have found that a 1% increase in customer retention is five times greater than the effects of a similar increase in margin. Leenheer et al. (2007) have in turn found a positive effect of LP membership on share-of-wallet. The linkage between behavioral loyalty and profit can thus be observed to be high, a
statement supported by many authors (Kumar, 2008; Reinartz and Kumar, 2003; Rust, Lemon and Zeithaml, 2004; Watson et al., 2015).

1.3 Synthesis
At the beginning of this introduction, the concept of customer value is defined as a dichotomy consisting of value to customers and value from customers (Kumar and Reinartz, 2016). In recent time, there has been an increasing realization that the value of existing customers is an important and variable asset dependent on each individual customer (Lindgreen and Wynstra, 2005). As such, it is critical for firms to target the right customers in terms of the value that they contribute back to the firm. One of the main CRM tools utilized by firms of today is the LP and the FFPs of the aviation industry is a prime example of this (Kumar and Reinartz, 2012; Terblanche, 2015; Vinod, 2011). As the main corporate goal of any LP is to retain customers and derive more revenue from them in the future, which ultimately leads to increased profits to the firm, it becomes of strategic value to determine on what basis this should be done (Noone and Mount, 2008). Historically, there has been an assumption that loyalty leads to higher profits, but this might not necessarily be completely true (Reinartz and Kumar, 2002; Watson et al., 2015). Instead, program performance differs from loyalty and measuring program performance is a different task than measuring loyalty (Blattberg, Kim and Neslin, 2008). Regarding loyalty, though, there is a stronger connection between behavioral loyalty and actual performance whereas attitudinal loyalty might have an impact on long-term loyalty.

With the advent of abundant customer data CLV is proposed by several authors as a suitable forward-looking metric in allocating marketing resources to balance value to and from customers (Blattberg, Kim and Neslin, 2008; Kumar and Reinartz, 2016; Malthouse, 2013). CLV is a monetary metric (Pfeifer and Bang, 2005; Kumar and Reinartz, 2016), that in the context of LPS might capture the main mechanisms driven by customer behavior (Blattberg, Malthouse and Neslin, 2009). It is thus a metric that can bridge the gap between loyalty and performance. Given the forward-looking nature of CLV, it is also necessary to understand the accuracy of the predictions in order to determine its suitability as a metric. Also, to make CLV translatable to managerial decisions, the drivers of CLV must be understood (Blattberg, Malthouse and Neslin, 2009; Kumar and Reinartz, 2016). This is crucial as LPS can be the key to new revenue growth if data intelligence and marketing programs are designed and leveraged to its full potential (Basso, Clements and Ross, 2009; Berman, 2006; Vinod, 2011). As such, the accuracy and drivers of CLV should be assessed in the context of LPS.

1.4 Purpose
The purpose of this thesis is therefore to:

Evaluate predicted airline CLV as a loyalty program performance metric and the drivers of CLV within frequent flyer programs

1.5 Disposition
The disposition of this report is as follows. Chapter 2 describes the frame of reference which further examines literature regarding CLV and LPS. Chapter 3 describes how the frame of reference leads to an analytical model and two specified research questions. Chapter 4 describes the general set-up of the study and gives motivation on how and why certain methodological choices were made. Chapter 5 describes the data and analysis corresponding to the two specified research questions. Finally, chapter 6 includes the conclusions that can be drawn from the analysis, contributions by this study, and suggestion on future studies.
2 Frame of Reference

This chapter presents a literature review based on the purpose of this study. Relevant theories and models from the academics are covered and serve as input to the analytical model of this study.

Since the purpose of this report is to evaluate predicted CLV as a loyalty program performance metric and to understand the drivers of CLV within FFPs, we must first examine suitable CLV models for this study. After that, it is necessary develop an understanding of FFPs and their main components, mechanisms, and effects to program performance to find drivers of CLV. As stated earlier, understanding the drivers of CLV will be crucial in developing successful strategies that can maximize CLV (Kumar and Reinartz, 2016; Rust, Lemon and Zeithaml, 2004). The overall structure of the frame of reference is illustrated in Figure 2.

![Figure 2. Structure of chapter 2.](image)

This chapter consists of four different sections where section one discusses how CLV predictions can be modeled and two to four are connected as they relate to how LPs and FFPs, in the end, affect CLV. In the end, the chapter will be synthesized which leads to the analytical model that will be applied and the specified research questions (SRQ) related to the model.

2.1 Understanding Customer Lifetime Value

Not all customers are created equal. The notion of CLV has long been used in industries such as financial services and magazine subscriptions where the retention of customers has been important (Blattberg, Kim and Neslin, 2008). At a fundamental level, a simple CLV model that takes revenues, costs, discount rate, repeat purchase probability and acquisition costs during a specified time horizon at a customer level can be defined as (Gupta et al., 2006, p.141):

\[
CLV = \sum_{t=0}^{T} \frac{(p_t - c_t)r_t}{(1 + i)^t} - AC
\]

where

- \(p_t\) = price paid by a customer at time \(t\),
- \(c_t\) = direct cost of servicing the customer at time \(t\),
- \(i\) = discount rate or cost of capital for the firm,
- \(r_t\) = probability of customer repeat buying or being alive at time \(t\),
- \(AC\) = acquisition cost of the customer, and
- \(T\) = time horizon for estimating CLV.

The issue with simple CLV calculations is that they are often calculated with averages across the entire customer base, resulting in a CLV for “the customer”, completely ignoring the heterogeneity of customers and thus ignoring that customers are not created equal (Fader, 2012). Arguably,
predicting the value of an individual customer require more advanced models to model purchase behavior successfully.

2.1.1 Modeling Approaches

A variety of models have been developed to capture CLV better. Gupta et al. (2006) have identified five modeling approaches: (1) Recency-Frequency-Monetary Value (RFM), (2) Probability models, (3) Econometric Models, (4) Persistence Models, (5) Computer Science Models, and (6) diffusion/growth models.

Firstly, RFM models are considered unsatisfactory as they perform worse than other CLV methods and only try to estimate the behavior in the next period (Gupta et al., 2006; Reinartz and Kumar, 2003; Venkatesan and Kumar, 2004). Secondly, probability models are defined by Gupta et al. (2006, p.142) as “a representation of the world in which observed behavior is viewed as the realization of an underlying stochastic process governed by latent (unobserved) behavioral characteristics, which in turn vary across individuals.”. Despite the unsatisfactory result of the RFM model, it can serve as an input for several probability models (Fader, Hardie and Lee, 2005b).

Thirdly, econometric models are similar to probability models as they assume similar underlying stochastic behavior but are more general as they use hazard functions and often incorporate covariates (Gupta et al., 2006). Customer acquisition, retention, and cross-selling or margin are combined to estimate CLV. One example is the Markov Chain Model with Decision Tree Learning used by Jasek et al. (2018) which uses RFM metrics and probability drives such as age, demographics type and intensity of product ownership and activity level as approaches for defining states.

Technically advanced models such as neural networks, classification and regression trees et cetera that fall under the category computer science models can be used to calculate CLV, but they have received limited attention in the marketing literature. Some authors have used these type of models, but there does not seem to be one model that fits all (Jasek et al., 2018; Malthouse, 2009).

Persistence models (Villanueva, Yoo and Hanssens, 2008) and diffusion/growth models (Gupta et al., 2006) are more aimed at calculating Customer Equity (CE) which is the sum of a firm’s CLV at an aggregate level and there is limited research on computer science models for CLV calculation (Jasek et al., 2018), making them unsuitable for the purpose of this study.

To summarize, both probability models, computer science models, and econometric models can be used for calculating individual-level CLV. Only probability models are considered a marketing research field with a variety of out-of-the-box models called customer-base analysis (Jain and Singh, 2002). Thus, probability models will be investigated further.

2.1.2 The Need for Customer-Base Analysis

As LPs facilitate the process of keeping track of individual customers and their transactions, large transaction-level databases are created. These datasets can be used to calculate simple analyses of descriptive character such as the average number of orders and the average order size, but it can also be used to create forward-looking predictions such as estimates of CLV (Fader and Hardie, 2009). As customer bases of firms can differ significantly from each other, Fader and Hardie (2009) describe two dimensions: opportunities for transactions and type of relationship with customers as shown in Figure 3. Opportunities for transactions can be either continuous where a purchase can occur at any given time such as the purchase of flights or discrete where the transaction can only occur at a specific period of time such as attending a conference taking place on a specific date. The relationship with customers can be either noncontractual where the firm is unaware of whether a customer has churned
or not between purchases, or *contractual* where the firm observes that a customer churns such as when the customer fails to renew its mobile plan.

<table>
<thead>
<tr>
<th>Continuous Opportunities for transactions</th>
<th>Discrete Opportunities for transactions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grocery purchases</td>
<td>Credit card</td>
</tr>
<tr>
<td>Doctor visits</td>
<td>Student meal plan</td>
</tr>
<tr>
<td>Hotel stays</td>
<td>Mobile phone usage</td>
</tr>
<tr>
<td>Event attendance</td>
<td>Magazine subs</td>
</tr>
<tr>
<td>Prescription refills</td>
<td>Insurance policy</td>
</tr>
<tr>
<td>Charity fund drives</td>
<td>Health club m.ship</td>
</tr>
<tr>
<td>Noncontractual Type of Relationships with Customers</td>
<td>Contractual Type of Relationships with Customers</td>
</tr>
</tbody>
</table>

**Figure 3. Classification of customer-bases (Fader and Hardie, 2009).**

Given the different characteristics of each type of customer base, different models are needed. In Table 1, a brief overview of the customer-base analysis models is presented.

**Table 1. An overview of customer-base analysis models (Fader and Hardie, 2009).**

<table>
<thead>
<tr>
<th>Setting</th>
<th>Example Model</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Continuous time non-contractual</td>
<td>Pareto/NBD (Pareto, negative binomial distribution)</td>
<td>(Schmittlein, Morrison and Colombo, 1987)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>BG/NBD (beta-geometric, negative binomial distribution)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MCMC frameworks (Markov Chain Monte Carlo)</td>
</tr>
<tr>
<td>Continuous time contractual</td>
<td>Various: exponential-gamma distribution, Weibull-gamma distribution</td>
<td>(Fader and Hardie, 2009)</td>
</tr>
<tr>
<td>Discrete time non-contractual</td>
<td>BG/BB (beta-geometric, beta-binomial distribution)</td>
<td>(Fader, Hardie and Shang, 2010)</td>
</tr>
<tr>
<td>Discrete time contractual</td>
<td>sBG (shifted beta-geometric)</td>
<td>(Fader and Hardie, 2007)</td>
</tr>
</tbody>
</table>

As flight tickets can be bought at any given time and without the firm knowing if a customer has churned between purchases, the purchase of flight tickets can be considered a continuous time, noncontractual setting.

2.1.3 Analysis of a Continuous-Time Noncontractual Customer Base

Over the years, multiple models for analysis of continuous-time noncontractual customer bases have been proposed. In 1987, Schmittlein, Morrison and Colombo proposed the Pareto/Negative Binomial Distribution (Pareto/NBD) with the purpose of identifying and counting the number of active members and predict their future activity in a non-contractual setting. These authors described the model as a generic Buy 'Till You Die (BTYD) model that theoretically can be used in any industry making it suitable for this study. Since then, several models have been proposed in the academic
literature. Fader, Hardie and Lee (2005b) propose the Beta-Geometric/Negative Binomial Distribution (BG/NBD) model which is based on the same assumptions as the Pareto/NBD model, but the difference lies in that the BG/NBD model assumes that dropout can occur immediately after a purchase, whereas Pareto/NBD assumes that the dropout can occur at any point in time. More recent research on CLV has focused on two areas: extensions of BTYD models such as the Pareto/Gamma-Gamma-Gamma (Pareto/GGG) that incorporates the regularity of interpurchase timing (Platzer and Reutterer, 2016) and Markov Chain Monte Carlo simulations (Mzoughia and Limam, 2015). Other contributions are extensions on the topic of incorporating time-invariant and/or time-variant covariates, Bayesian estimations of the Pareto/NBD model and incorporation of non-stationary repurchase behavior in a broader class of hidden Markov Models (Reutterer, 2015). The extensions of the BTYD models and Markov Chain Monte Carlo simulations have received limited attention in the marketing literature. Whereas the BG/NBD model has been praised for its easier implementation (Wübben and Wangenheim, 2008; Fader, Hardie and Lee, 2005a), the Pareto/NBD model has been shown as a good benchmark model in continuous-time noncontractual settings by multiple studies as it performs well across a variety of datasets from different industries when compared to other BTYD and econometric models (Gupta et al., 2006; Jasek et al., 2018).

An assumption of the Pareto/NBD model is that each customer makes purchases according to a Poisson process with rate \( \lambda \) and that each customer remains alive for a lifetime which has an exponential distribution with death rate \( \mu \). The model also assumes heterogeneity across customers so that the purchasing rate \( \lambda \) is gamma distributed across all customers and that the death rate \( \mu \) is distributed according to a different gamma distribution. The rates \( \mu \) and \( \lambda \) are independent of each other (Schmittlein, Morrison and Colombo, 1987).

Because of the behavioral differences across industries, it is of interest to validate that the Pareto/Model can be used in the aviation industry. As noted by Fader, Hardie and Lee (2005a), the performance of the Pareto/NBD model has been empirically validated on holdout sets whereas the performance of more recent models has not been thoroughly studied. Using a well-known model that is easily accessible through statistical software such as MATLAB, Python and R increases the reliability since it makes the calculations easier to repeat. Therefore, the Pareto/NBD model will be used in this study. A consequence of this choice can be that the predictive accuracy can be inferior to extended CLV-models such as the Pareto/GGG by Platzer and Reutterer (2016).

2.1.4 From Purchase Behavior to CLV

As the Pareto/NBD model only predicts the expected number of future purchases, it is necessary to multiply it with a monetary value of each purchase. Two ways of determining this monetary value are proposed, the mean transaction value of each customer or a model to estimate each customer’s underlying average transaction value denoted \( E(M) \) (Fader, Hardie and Lee, 2005b). In the case of the model, the more transactions each customer performs, the closer the average transaction value is to the predicted \( E(M) \). In other words, the model assumes there is a stationary average transaction value \( E(M) \) for each individual customer. This is considered superior to a mean transaction value as initial purchases is not necessarily representative (Fader, Hardie and Lee, 2005b). Three submodels have been proposed for this task: (1) standard normal (Schmittlein and Peterson, 1994), (2) log-normal (Borle, Singh and Jain, 2008) and (3) Gamma-Gamma (Colombo and Jiang, 1999). In the context of a noncontractual continuous time setting, the standard normal and Gamma-Gamma distribution are suitable (Fader, Hardie and Lee, 2005b). The standard normal submodel assumes that the overall distribution of transaction values follows a normal distribution, whereas the Gamma-Gamma model allows for a skewness in the underlying average transaction value (Schmittlein and Peterson, 1994; Fader, Hardie and Lee, 2005b).
2.1.5 Model Validation on Individual-level and Overall Customer-Base Level

As noted by Niels Bohr: "Prediction is very difficult, especially about the future.". Despite the challenge of accurately predicting the future, the managerial need for a forward-looking metric like CLV is high thus warranting a qualified guess (Kumar and Reinartz, 2016; Fader, 2012). In the last ten years, several studies have tested the performance of the Pareto/NBD model alongside both BTYD models and other models. Whereas certain authors (Jasek et al., 2018; Fader, Hardie and Lee, 2005a; b) have expressed confidence in the reliability of the Pareto/NBD model for the purposes of forecasting individual buying behavior, other authors have been more skeptical (Wübben and Wangenheim, 2008; Malthouse and Blattberg, 2005; Malthouse, 2009). Given the predictive nature of these models, it is necessary to validate their forecasting capabilities. As in the field of statistical learning such as machine learning, data is split up between a training set to fit the model and a test set to evaluate the model (Gareth et al., 2013). In the field of Customer-Base Analysis, the words calibration period set and holdout period set is often used instead (Fader, Hardie and Lee, 2005b; Platzer and Reutterer, 2016). To evaluate the managerial suitability of these models, it is important to evaluate both aggregate- and individual-level performance (Wübben and Wangenheim, 2008).

2.1.5.1 Aggregate-Level Performance

Two measures are used in the literature to evaluate the aggregate-level performance: Forecast vs Actual (FvsA) and Mean Absolute Percentage Error (MAPE) which can be seen in Table 2 (Wübben and Wangenheim, 2008; Jasek et al., 2018; Leeflang et al., 2015; Fader, Hardie and Lee, 2005b). The first metric, FvsA, evaluate the cumulative forecast profit versus the cumulative actual profit from the start of the calibration period to the end of the holdout period and is therefore a comprehensible analysis on the performance of the model (Jasek et al., 2018). The second metric, MAPE, evaluate the absolute error during the holdout period only and is relative to the corresponding actual value (Jasek et al., 2018; Leeflang et al., 2015; Wübben and Wangenheim, 2008). The MAPE metric is suitable when comparing forecast accuracy for different settings (Leeflang et al., 2015). The aggregate-level performance is of interest as multiple studies have shown that models do not perform equally well on both aggregate-level and individual-level.

Table 2. An overview of aggregate-level error measures used by the literature.

<table>
<thead>
<tr>
<th>Name</th>
<th>Authors’ evaluation</th>
<th>Examples of studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forecast vs Actual (FvsA)</td>
<td>Comprehensible relative metric that shows if the forecast over- or underestimates the actual CLV</td>
<td>(Jasek et al., 2018; Fader, Hardie and Lee, 2005b)</td>
</tr>
<tr>
<td>Mean Absolute Percentage Error (MAPE)</td>
<td>Relative absolute error during the holdout period used for comparing different settings</td>
<td>(Jasek et al., 2018; Leeflang et al., 2015; Wübben and Wangenheim, 2008)</td>
</tr>
</tbody>
</table>

2.1.5.2 Individual-Level Performance

A variety of error measures with different characteristics have been proposed in the literature to evaluate the individual-level performance of models. An overview of the six error measures can be seen in Table 3. Mean Absolute Error (MAE) is described as a simple measure that is used by multiple authors with a result that is easily comprehensible (Jasek et al., 2018; Platzer and Reutterer, 2016; Schwartz, Bradlow and Fader, 2014; Platzer, 2008). The Root Mean Squared Error (RMSE) is a common error measure but the result can be severely punished by outliers. To solve this issue the Root Median Squared Error can be used instead (Platzer, 2008). As CLV is proposed as a way to identify profitable customers to allocate marketing resources better, it is also important to evaluate how many of the top 10% most profitable customers are correctly identified by the model. Three sets of authors use some type of sensitivity analysis to evaluate this error metric. Fader, Hardie and Lee (2005a) use correlation as an error measure, which is questioned by Platzer (2008) as it only conveys
to what degree the variables change in unison. In the 2008 competition arranged by the *Journal of Interactive Marketing (JIM)*, the Direct Marketing Educational Foundation and the DMA Nonprofit Federation, the measure of accuracy of individual-level CLV was defined as the root mean squared error of the logged predicted values (MSLE) where lower values represent better predictions (Malthouse, 2009). As noted by Platzer (2008), MSLE puts more emphasis on the accurate estimate of the main group of customers, which usually is the low repeat transaction group.

Table 3. An overview of individual-level error measures used by the literature.

<table>
<thead>
<tr>
<th>Name</th>
<th>Authors’ evaluation</th>
<th>Examples of studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Absolute Error (MAE)</td>
<td>Simple measure used by multiple authors</td>
<td>(Jasek et al., 2018; Platzer and Reutterer, 2016; Schwartz, Bradlow and Fader, 2014; Platzer, 2008)</td>
</tr>
<tr>
<td>Mean Squared Logarithmic Error (MSLE)</td>
<td>Used in the JIM competition as the measurement of accuracy on individual-level CLV</td>
<td>(Platzer, 2008)</td>
</tr>
<tr>
<td>Root Mean Squared Error (Mean RMSE)</td>
<td>Sensitive to outliers</td>
<td>(Platzer, 2008; Wübben and Wangenheim, 2008)</td>
</tr>
<tr>
<td>Root Median Squared Error (Median RMSE)</td>
<td>Less sensitive to outliers compared to Mean RMSE</td>
<td>(Wübben and Wangenheim, 2008)</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>How many of the top 10% most profitable customers were correctly assigned to the top 10% class</td>
<td>(Jasek et al., 2018; Malthouse and Blattberg, 2005; Wübben and Wangenheim, 2008)</td>
</tr>
</tbody>
</table>

2.1.5.3 Determining the usefulness of the models

To determine the usefulness of the models, two perspectives can be used, a statistical perspective and a comparison with simple managerial heuristics used in the company. From a statistical perspective, the results of previous studies are mediocre. Using the combined Pareto/NBD, Gamma-Gamma CLV model, Jasek et al. (2018) calculated a MAE of 113.7% on the weighted average of six online retailers. In the 2008 JIM competition, Malthouse (2009) notes that the winner had an MSLE value of 1.69, which can be interpreted as that the best model for the competition dataset was off by a multiplicative factor of \(\exp(5.4) \approx 5.4\). From a simple managerial heuristics perspective, the results are slightly better. Wübben and Wangheim (2008) compare the individual- and aggregate-level performance of the Pareto/NBD model to what the authors call simple managerial heuristics in order to discuss their suitability as managerial tools. For estimating future purchase behavior, the authors conclude that the studied companies simply assume that every customer continues to buy at his or her past mean purchase frequency. For estimating sensitivity, the managerial heuristics used is that the previous top 10% customers will continue to be the top 10%. From a managerial perspective, these managerial heuristics serve as a benchmark when it comes to determining whether the CLV-calculations are useful in practice. An overview of the performance of the BTYD models from multiple datasets can be found in Table 4.
To summarize, the Pareto/NBD and Gamma-Gamma or Normal form can be used to calculate CLV as it is considered a good benchmark model in a continuous-time non-contractual setting. To validate the performance of the model, several error measures can be calculated at an aggregate- and individual-level. Furthermore, these metrics can be benchmarked against simple managerial heuristics used by managers to determine whether the predictions of the model are acceptable.

2.2 Types of Loyalty Programs
Since LPs have evolved over time and differ among firms and industries, it is beneficial to first get a macro-level understanding of the different types of LPs that exist. Authors such as Berman (2006) and Blattberg, Kim and Neslin (2008) have found different ways to divide LPs into categories and in relation to FFPs, authors such as de Boer and Gudmundsson (2012) have also identified distinct program types.

2.2.1 The Four Tiers of Loyalty Programs
According to Berman (2006), there are four tiers of LPs, type 1 to type 4, where the higher tiers build upon the lower tiers. While the benefits of type 3 and 4 programs are greater than those of type 1 and type 2 programs, the higher tiers are costlier to implement and maintain. Broadly speaking, type 1 programs are programs where no database is established for the firm’s customers and members receive discounts regardless of purchase history, type 2 programs are characterized by members receiving rewards after a certain number of purchases, type 3 programs are programs where there is an established customer database and members receive points based on cumulative purchases and type 4 programs are those who also segment customers based on their purchase history and where more sophisticated data mining capabilities are being utilized for tailored communications and promotions (Berman, 2006). The type 4 programs can be seen as the new approach to LPs, characterized by personalization and customization at the individual customer level (Kumar, 2008). In recent times, personalization and customization has been greatly enhanced by the rapid development in Internet technologies, mobile platforms, and social media (Breugelmans et al., 2015). As these technologies evolve, the capabilities for personalization and customization will only improve as time goes on.
Furthermore, the evolved LPs are customer-centric compared to program-centric. Identifying in which category an LP is situated in can help guide managers on the steps available to improve the program (Berman, 2006). The characteristics of each type of program are described in Table 5. The purpose of the lower tier programs is mainly to foster an increase in member purchases and purchase frequency, both being behavioral responses (Bijmolt, Dorotic and Verhoef, 2011; Leenheer et al., 2007). Type 4 programs in addition to what the lower tiers do, also incorporate personalized marketing to foster these behaviors.

Table 5. A typology of LP types adopted with Swedish examples (Berman, 2006).

<table>
<thead>
<tr>
<th>Program Type</th>
<th>Characteristics of Program</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Type 1:</strong></td>
<td>Members receive additional</td>
<td>Simple supermarket programs, IKEA.</td>
</tr>
<tr>
<td></td>
<td>discount at the register</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Membership is open to all</td>
<td></td>
</tr>
<tr>
<td></td>
<td>customers</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Clerk will swipe discount</td>
<td></td>
</tr>
<tr>
<td></td>
<td>card if member forgets or</td>
<td></td>
</tr>
<tr>
<td></td>
<td>does not have a card</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Each member receives the</td>
<td></td>
</tr>
<tr>
<td></td>
<td>same discount regardless</td>
<td></td>
</tr>
<tr>
<td></td>
<td>of purchase history</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• There are no targeted</td>
<td></td>
</tr>
<tr>
<td></td>
<td>communications directed</td>
<td></td>
</tr>
<tr>
<td></td>
<td>at members</td>
<td></td>
</tr>
<tr>
<td><strong>Type 2:</strong></td>
<td>Members receive 1 free</td>
<td>Pizzerias, Stamp cards.</td>
</tr>
<tr>
<td></td>
<td>when they purchase n units</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Membership is open to all</td>
<td></td>
</tr>
<tr>
<td></td>
<td>customers</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Firm does not maintain a</td>
<td></td>
</tr>
<tr>
<td></td>
<td>customer database linking</td>
<td></td>
</tr>
<tr>
<td></td>
<td>purchases to specific</td>
<td></td>
</tr>
<tr>
<td></td>
<td>customers</td>
<td></td>
</tr>
<tr>
<td><strong>Type 3:</strong></td>
<td>Members receive rebates or</td>
<td>Low-cost carriers, hotels, credit</td>
</tr>
<tr>
<td></td>
<td>points based on cumulative</td>
<td></td>
</tr>
<tr>
<td></td>
<td>purchases</td>
<td>card programs.</td>
</tr>
<tr>
<td></td>
<td>• Seeks to get members to</td>
<td></td>
</tr>
<tr>
<td></td>
<td>spend enough to receive</td>
<td></td>
</tr>
<tr>
<td></td>
<td>qualifying discount</td>
<td></td>
</tr>
<tr>
<td><strong>Type 4:</strong></td>
<td>Members receive targeted</td>
<td>Full-service carriers, ICA</td>
</tr>
<tr>
<td></td>
<td>offers and mailings</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Members are divided into</td>
<td></td>
</tr>
<tr>
<td></td>
<td>segments based on their</td>
<td></td>
</tr>
<tr>
<td></td>
<td>purchase history</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Requires a comprehensive</td>
<td></td>
</tr>
<tr>
<td></td>
<td>customer database of</td>
<td></td>
</tr>
<tr>
<td></td>
<td>customer demographics and</td>
<td></td>
</tr>
<tr>
<td></td>
<td>purchase history</td>
<td></td>
</tr>
</tbody>
</table>

2.2.2 Categorizing Program Types Based on Their Reward Structure

In addition to Berman’s (2006) program types, another way of categorizing program types is by their reward structure. According to Blattberg, Kim and Neslin (2008), two prominent types of LPs have emerged: frequency reward programs and customer tier programs. According to the authors, the main difference between the two program types lies in their structure; frequency reward programs reward customers with products, rebates, and points based on their purchase history, whereas customer tier programs segment customers into tiers and give different benefits depending on the tier level. Even though these two programs types are distinct in their structure, a mix of both structures is also common as in the case of FFPs (Kopalle, Neslin and Sun, 2009).

In comparison with the program types identified by Berman (2006), all program types below type 4 fall in the frequency reward program type as these types rewards customers based on their past purchases. Type 4 programs, on the other hand, does segment customers based on their purchase history and are thus aligned with the customer tier programs as described by Blattberg, Kim and Neslin (2008). A synthesis of the two types of categorizations can be seen in Table 6. As such, both categorization types can be used in conjunction to analyze where an LP belongs.

Table 6. Synthesis of the different views on program types and their corresponding reward type (Berman, 2006; Blattberg, Kim and Neslin, 2008).

<table>
<thead>
<tr>
<th>Reward Structure</th>
<th>Program Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency Reward</td>
<td>Type 1, Type 2, Type 3</td>
</tr>
<tr>
<td>Customer Tier</td>
<td>Type 4</td>
</tr>
</tbody>
</table>
2.2.3 Frequent Flyer Programs – the Loyalty Programs of the Aviation Industry

As stated in 1.1, FFPs are one of the prime examples of customer LPs. A comprehensive effort has been done by de Boer (2017) to summarize existing literature and knowledge about FFPs. According to Berman (2006), FFPs normally belong to type 3 programs, but FFPs are in fact quite heterogeneous and can in turn be divided into three general types according to their key characteristics and whether the airline carrier is a full-service carrier (FSC) or a low-cost carrier (LCC) (de Boer, 2017; de Boer and Gudmundsson, 2012). FFPs are part of the standard product offered by established FSC airlines, whereas the offering of FFPs is a newer phenomenon among LCCs (de Boer, 2017). It must be noted, though, that FFPs are now commonplace even among LCCs. The difference between FFPs in LCCs and FSCs is mainly that they are simpler in LCCs and mainly offer frequency awards such as a free flight after a certain number of flights have been purchased by the customer (de Boer, 2017). FFPs around the world tend to vary across ten different key dimensions which can be used to assess what type of program that a specific FFP belongs to. According to de Boer (2017), it is possible to categorize FFPs into three different program types: (1) legacy programs, (2) advanced programs, and (3) autonomous programs. An overview and comparison of the three types are given in Table 7.

Table 7. Overview of frequent flyer program types (de Boer and Gudmundsson, 2012, p.23).

<table>
<thead>
<tr>
<th></th>
<th>Legacy</th>
<th>Advanced</th>
<th>Autonomous</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Strategic focus</strong></td>
<td>Frequent flyers</td>
<td>Frequent flyers and high credit card spenders</td>
<td>Frequent flyers and everyday spenders</td>
</tr>
<tr>
<td><strong>Structure</strong></td>
<td>FFP department (part of Marketing / Sales)</td>
<td>Separate strategic business unit</td>
<td>Separate company</td>
</tr>
<tr>
<td><strong>Ownership</strong></td>
<td>100% owned by the airline</td>
<td>100% owned by the airline</td>
<td>Owned by airline and/or outside investors</td>
</tr>
<tr>
<td><strong>Suitable for third-party investment</strong></td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Reporting</strong></td>
<td>At aggregate level</td>
<td>May do segmental reporting</td>
<td>Income and balance sheet</td>
</tr>
<tr>
<td><strong>Non-air partner accrual as percentage</strong></td>
<td>Small (&lt;20%)</td>
<td>Medium (&gt;20%)</td>
<td>Large (&gt;50%)</td>
</tr>
<tr>
<td><strong>Partner range</strong></td>
<td>Travel related (hotel, car)</td>
<td>Travel related and financial services</td>
<td>Travel, financial and everyday spend</td>
</tr>
<tr>
<td><strong>Awards</strong></td>
<td>Award tickets and upgrades</td>
<td>Air travel and limited merchandise</td>
<td>Air travel, other travel, merchandise, experiential awards</td>
</tr>
<tr>
<td><strong>Staff profile</strong></td>
<td>Airline background</td>
<td>Airline and marketing background</td>
<td>Other backgrounds including retail and finance</td>
</tr>
<tr>
<td><strong>Award allocation policy</strong></td>
<td>Fixed – supplemented with distressed inventory</td>
<td>Combination of fixed and dynamic allocation</td>
<td>Combination of fixed and dynamic – any seat is available</td>
</tr>
</tbody>
</table>

Similar to Berman’s (2006) typology for LPs, the three types of FFPs can be seen as tiers where the autonomous programs are the most advanced form, whereas the legacy program type stems from the original program that American Airlines and other airlines of that era launched in the 1980s. One key difference between the two classifications, though, is that the FFP types by de Boer (2017) also concerns the ownership and suitability for investing in the programs. According to de Boer (2017) each of the program types have distinct features across ten key dimensions: (1) strategic focus, (2) structure, (3) ownership, (4) suitability for third-party investment, (5) type and level of reporting, (6) percentage of miles earned outside the airline, (7) partner range, (8) scope and width of awards, (9) staff profile, and (10) award allocation policy. The main difference between the advanced FFPs and the autonomous FFPs is the structure and relationship between the LP and the airline where the autonomous program operates as a standalone unit outside the airline (de Boer, 2017). No distinction
is made by de Boer and Gudmundsson (2012) regarding the reward structure of the different FFP types. They can therefore all be said to have a varying degree and mix of frequency reward structure and customer tier structure. For instance, most FSC FFPs offer both flight redemptions that can be redeemed through earned miles, and rewards based on the tier that the customer has qualified for (Scandinavian Airlines System, 2018; American Airlines, 2018; Lufthansa, 2018).

Given that LPs and FFPs can be categorized in different types according to their characteristics, it becomes interesting to see on what dimensions that LP characteristics can be differentiated on. Researchers agree that depending on the design of a program, different customer responses are facilitated which in the end lead to differences in program performance (Bijmolt, Dorotic and Verhoef, 2011; Breugelmans et al., 2015; Watson et al., 2015; McCall and Voorhees, 2010). Furthermore, if autonomous FFPs are the future of airline FFPs, it becomes increasingly important to assess the value of such programs and the results that they bring to the airlines (de Boer and Gudmundsson, 2012).

2.3 The Design Components of a Loyalty Program Affects Program Performance

According to Breugelmans et al. (2015) the characteristics of any LP can be divided into five key design components that are relevant for all types of LPs: (1) membership requirements, (2) program structure, (3) point structure, (4) reward structure, and (5) program communication. Several authors point out that the design components of an LP, in the end, affect how the program performs (McCall and Voorhees, 2010; Liu and Yang, 2009; Watson et al., 2015).

Membership requirements affect the convenience, effort, and costs associated with joining an LP (Breugelmans et al., 2015). According to the authors, a customer’s decision to join and adopt an LP is dependent on the perceived benefits relative to the perceived costs and risks of enrollment. Furthermore, the decisions on specific membership requirements is thus a trade-off between attracting a broader customer base by lowering the perceived costs and risks of joining a membership and increasing the perceived benefits, versus targeting a narrower and more profitable customer segment by doing the opposite.

There are two primary program structures: frequency reward programs that award all LP members who reach a required threshold with discounts and gifts, and customer tier programs that segment customers according to their value and actual profitability to the firm (Breugelmans et al., 2015). These are based on and are essentially the same as the two types of reward programs given by Blattberg, Kim and Neslin (2008) which are also mentioned as a key design component by Bijmolt, Dorotic and Verhoef (2011).

Point structure refers to the rules of how points are issued and expired, what the point thresholds are for redeeming rewards and whether tiered structures are used which may affect how points are earned (Breugelmans et al., 2015). A similar view on point structure is shared by Liu and Yang (2009) who categorize point structure as a program design component along with participation requirements, and rewards. According to McCall and Voorhees (2010), LPs are normally structured in tiers that are designed to reduce costs and provide firms with the flexibility to segment members within the LP. This builds on the Pareto principle that a small proportion of a firm’s customers contribute a large share of the firm’s revenue (McCall and Voorhees, 2010). Thus, by calculating CLV and CE, this phenomenon can be observed and a common usage of CLV is to segment a firm’s customers (Rust, Lemon and Zeithaml, 2004; Kumar, 2008; Kim et al., 2006). McCall and Voorhees (2010) state that research on program tiers have been focused on the impact of the number of tiers and the customer’s behavior as they approach and move between tiers.
Reward structure is the program component related to the type of rewards that can be earned by LP members. These can broadly be divided between monetary and non-monetary rewards, the aspirational value of luxury or necessity, the relation of the reward to the firm’s brand, and the reward timing which can be immediate or delayed (Breugelmans et al., 2015). LPs offer multiple forms of rewards and research on reward types have tended to focus on the utility associated with a particular reward and whether the reward is direct or indirect (McCall and Voorhees, 2010).

Lastly, program communication refers to all forms of contact between the LP and its members. According to Breugelmans et al. (2015), research has shown that small nuances on how a member’s progress is communicated to the member might influence the consumer’s behavior. In recent years, advances in internet technology and the general adoption of social media have enhanced the tools of how LPs and their members interact (de Boer, 2017; Breugelmans et al., 2015).

McCall and Voorhees (2010) have identified the following factors as output measures of how the program design affects program effectiveness: (1) increased purchase frequency, (2) decreased customer price sensitivity, (3) customer advocacy, (4) extended relationship length, (5) share of wallet, (6) consumer community and connectedness, and (7) increased firm performance. These are comparable to similar findings by other authors (Blattberg, Malthouse and Neslin, 2009; Liu and Yang, 2009; Bijmolt, Dorotic and Verhoef, 2011). Customer advocacy and community connectedness stand out as not being directly related to program performance as defined by Watson et al. (2015). Instead, these two factors are more related to word of mouth while the rest of the output factors are more related to program performance. The reasoning for this is that customer advocacy and community connectedness fall within indirect value, instead of direct value (Kumar and Reinartz, 2016). All the other factors affect financial performance and are thus components that lead to direct value for the firm and factors which might be relevant to measure in the context of program performance.

To summarize, the design components of an LP affect the output of the program and its performance. While categorizing LPs into different types and tiers might give a macro view of the differences among programs, a look at the LP design components presents a more detailed view on the mechanics of the programs. Regarding program structure, Blattberg, Kim and Neslin (2008) choose to make a clear distinction between frequency reward programs and customer tier programs, whereas Breugelmans et al. (2015) regards this as one of the five design components that differs among LPs. As stated in 2.2.3, FFPs can incorporate elements of both program structures. Thus, the view presented by Breugelmans et al. (2015), is more fit in the context of FFPs.

2.4 The Drivers of Loyalty Program Performance
From 2.3 it is motivated that the design of an LP affects how the program will perform. To further investigate how the design of an LP in the end leads to performance, it is necessary to identify the underlying mechanisms that are at play. By understanding these, the true drivers of program performance and thus CLV can be identified. Furthermore, another approach is to look at generalizable drivers of CLV and fit them within the context of LPs and program performance. The two following sub-chapters will deal with these two approaches. First, the mechanisms affected by program structure, reward structure, and customer factors will be examined. Second, the generalizable drivers of CLV will be examined.

2.4.1 The Main Mechanisms in Loyalty Programs that Drive Program Performance
Blattberg, Kim and Neslin (2008) have identified three main mechanisms that influence the typical LP member process of earning points, climbing membership tiers, and exchanging points for discounts or rewards: (1) points-pressure mechanism, (2) rewarded-behavior mechanism, and (3)
personalized marketing mechanism. These mechanisms all serve to increase the customer value of a firm (Blattberg, Kim and Neslin, 2008). Earned rewards such as free flights redeemed with earned miles are common among a variety of firms such as airlines, hotels, and supermarkets. Instead of a price discount, firms incentivize customers to focus their purchases on the firm to earn the reward. As a result, these reward programs create points-pressure and rewarded-behavior (Blattberg, Malthouse and Neslin, 2009). The mechanisms, and how they relate to behavioral and attitudinal responses are illustrated in Figure 4.

Figure 4. How reward program mechanisms influence retention and purchase volume (Blattberg, Kim and Neslin, 2008, p.550).

Points-pressure refers to the mechanism that a member is more likely to make additional purchases when the member perceives to be close to obtaining a reward (Bijmolt, Dorotic and Verhoef, 2011). The effect is said to be short-termed as customers are expected to increase their purchase frequency when they are in the vicinity of reaching a reward (Kopalle, Neslin and Sun, 2009; Taylor and Neslin, 2005). The mechanism occurs due to a combination of customer switching costs as customers accumulate points towards a goal and the future orientation of customers as they care about a future reward that can be gained by accumulating points (Taylor and Neslin, 2005). Additionally, the attractiveness of the reward can further increase points-pressure (Blattberg, Kim and Neslin, 2008). In the context of FFPs, the points-pressure mechanism has been observed both when members approach a cash-in reward as well as when members are getting close to a tier upgrade (Kopalle, Neslin and Sun, 2009).

The rewarded-behavior mechanism affects a members’ behavioral and attitudinal responses after they obtain a reward and leads to an increase in purchase rate after a member receives a reward (Bijmolt, Dorotic and Verhoef, 2011; Blattberg, Kim and Neslin, 2008). This may be due to either behavioral learning or increased effect and the implications of these underlying factors give implications on whether the results of rewarded-behavior truly increase loyalty or just purchase inertia (Blattberg, Kim and Neslin, 2008). While some authors argue that the mechanism have long-term effects (Taylor and Neslin, 2005), some conflicting results have been found regarding the long-term effects of the rewarded-behavior mechanism. For instance, Kopalle et al. (2009) could not observe any long-term effects in association with the rewarded-behavior mechanism.

The personalized marketing mechanism enhances the behavioral and attitudinal responses of a member. In particular, regarding program performance, personalized marketing can boost customer retention and purchase volume (Blattberg, Kim and Neslin, 2008). This can be achieved through individually targeted promotions, cross-selling, or personalized customer service. Especially cross-selling is seen as a major opportunity when companies have data on customer purchase history, which
can be collected through LPs (Blattberg, Kim and Neslin, 2008). The number of marketing contacts associated response rates are critical for managing CLV, but more contacts might result in customers being “worn-out” if they happen too close to each other (Blattberg, Malthouse and Neslin, 2009). Rewards programs facilitate the use of personalized marketing efforts, for instance, to increase cross-selling and could, therefore, be a driver of CLV (Blattberg, Kim and Neslin, 2008).

In conclusion, these three mechanisms can be identified as potential drivers of CLV in the context of LP performance. Points-pressure and rewarded-behavior both acts as mechanisms in LPs to increase the customers’ purchasing frequency. These mechanisms are centered around the event of a redemption in which points-pressure is present before the redemption and rewarded-behavior after the redemption (Taylor and Neslin, 2005). The interplay of these mechanisms around the redemption event is illustrated in Figure 5.

![Figure 5](image-url)

Figure 5. The effect of points-pressure and rewarded-behavior around the redemption event (Taylor and Neslin, 2005, p.294).

In addition to points-pressure, Dorotic et al. (2014) also propose an additional mechanism which they call redemption momentum. This mechanism is, according to the authors, independent of points-pressure and enhances purchase behavior before and after the redemption event. The authors mainly distinguish between points-pressure and redemption momentum as they define points-pressure as occurring when a member has an insufficient amount of points prior to a redemption, whereas they do have enough points to redeem the reward prior to doing so when exhibiting redemption momentum. Understanding how each of the mechanisms affects program performance can give implications on how to optimize the LP design for increased CLV.

2.4.2 Generalizable CLV Drivers in the Context of Program Performance

Through a review of extant literature, Blattberg, Malthouse and Neslin (2009) have identified four of these antecedents as empirically generalizable drivers of CLV: (1) customer satisfaction, (2) marketing efforts, (3) cross-buying, and (4) multichannel purchasing. These are according to Blattberg, Malthouse and Neslin (2009) well-defined, consistent effects found by at least three different set of authors that all have a positive relationship with CLV. A positive relationship means
that the drivers positively affect one or more of the following components of CLV: (1) relationship duration, (2) revenue, and (3) costs.

In Figure 6, a conceptual framework for the antecedents of CLV is illustrated. As discussed by Watson et al. (2015), program performance is mainly affected by behavioral factors which rules out customer satisfaction as a driver in the context of program performance. It must be noted, though, that effective customer responses along with marketing actions lead to behavioral customer responses and vice versa (Blattberg, Kim and Neslin, 2008).

Marketing efforts have a strong association with relationship duration which in turn has a positive effect on CLV (Blattberg, Malthouse and Neslin, 2009). This fits in the context of LPs as marketing efforts are related to the program communication design component mentioned by Breugelmans et al. (2015). Modeling and calculating CLV is crucial to correctly balance marketing efforts for profitable customer engagement (Kumar and Reinartz, 2016). Furthermore, the authors emphasize that CLV can serve both as a tool on how to strategize marketing efforts related to LPs as well as a performance metric of such activities.

Cross-buying or cross-selling is associated with higher CLV, mainly through increased lifetime duration and revenue (Blattberg, Malthouse and Neslin, 2009). Cross-buying is promoted by FFPs since they give airlines the opportunity to capture ancillary revenues, principally through the sale of miles to banks offering credit cards (Reales and O’Connell, 2017). According to the authors, cross-buying might be a result of effective customer retention. In the case of LPs, it is common to offer additional products such as credit cards in exchange for discounts or points in the LP. In one study on the causal relationship between behavioral loyalty and cross-buying, the results suggest that cross-buying is a consequence of behavioral loyalty rather than the reverse (Reinartz, Thomas and Bascoul, 2008).

Multichannel purchasing and pricing strategies for acquisition, retention, and reacquisition are two potential drivers of CLV which can be meaningful to evaluate in the context of an airline, but there
are no major mechanisms within FFPs that are associated with them since channel and pricing strategies lie outside the scope of FFPs.

To summarize, the generalizable drivers of CLV by Blattberg, Malthouse and Neslin (2009) that fit within the context of LPs, can be identified together with the three LP mechanisms as the drivers of program performance. Marketing efforts and cross-buying are of interest in this study since these drivers are commonly present in FFPs.
3 Specification of the Task with Specified Research Questions

The SRQs are derived from the analytical model. The answers to these will lead to the conclusion of this study.

3.1 From Frame of Reference to an Analytical Model

As stated in chapter 2.1 Understanding Customer Lifetime Value, CLV is a relevant measurement to ensure that the value from customers matches the value to customers. An issue, though, is that it has been challenging to measure CLV accurately at an individual level. This leads to questioning whether CLV is a suitable managerial tool and metric. In particular, Wübben and Wangenheim (2008) have made comparisons between the individual- and aggregate-level performance of the Pareto/NBD model to what the authors call simple managerial heuristics in order to discuss their suitability as managerial tools.

With the assumption that CLV is a suitable tool, multiple independent factors have been identified in LPs that, in the end, should affect one dependent factor which in our case is CLV. In this regard, CLV serves as the measurement used to determine how the LP performs. From chapter 1.2.2 it is understood that CLV encompasses many of the traditional metrics used to measure program performance, such as profits, revenue, share of wallet, etc. From the theory presented in chapter 2.4, the following rationale has been applied to reach the dependent variables that ultimately are part of the analytical model. The relationships in the analytical model are simple which means that only multiple parallel analyses together serve to answer the main research question.

In chapter 2.2, it is understood that different types of LPs have a different set of tools which are used to ultimately drive program performance. The program types are characterized by their differences in design components which cover factors such as the program structure, the point structure, the reward structure, and so on. These design components, in turn, affect customers that are participating in LPs through three main mechanisms, namely: (1) point-pressure, (2) rewarded-behavior, and (3) personalized marketing. These three mechanisms are part of all LPs with the exception of personalized marketing which is only apparent in more advanced program types defined as type 4 programs by Berman (2006). According to Berman (2006) FFPs are normally not in this category, which has lead us to omit personalized marketing among the independent variables in the analytical model.

Beyond the mechanisms present in LPs, another approach is to also look at the drivers of CLV itself and assess what drivers that fit in the context of LPs. From chapter 2.4.2 we have identified: (1) cross-buying, (2) marketing efforts, and (3) promotions as drivers of CLV that fit within this context. As such, these drivers have also been included as independent variables in our analytical model.

To summarize, we have identified five behavioral drivers of CLV in the context of LPs: (1) point-pressure, (2) rewarded-behavior, (3) cross-buying, (4) marketing efforts, and (5) promotions. As attitudinal loyalty aspects are considered to have limited impact on LP performance according to the meta-analysis by Watson et al. (2015), we have chosen to not include attitudinal aspects. This results in an analysis model presented in Figure 7.
Consequently, we have defined two major SRQs that will capture the suitability of CLV as a managerial tool and the effects of the identified drivers on CLV.

3.2 Specified Research Questions

1. How suitable are the Pareto/NBD CLV predictions as a managerial tool in the aviation industry compared to simple managerial heuristics:
   a. At an individual-level?
   b. At an aggregate-level?

2. How and to what extent do the following drivers affect the (airline) CLV from an FFP perspective:
   o Cross-buying
   o Points-pressure
   o Rewarded-behavior
   o Promotions
   o Marketing efforts
4 How the Study was Conducted

This chapter describes and motivates the methodology of the study. Support is drawn from relevant academic methodology theory.

4.1 Scientific View

The scientific view of the authors aligns with the post-positivistic philosophical worldview. The following description by Creswell (2013) applies. Postpositivists hold a deterministic philosophy in which causes determine effects or outcomes. Thus, the problems studied by postpositivists reflect the need to identify and assess the causes that influence outcomes. It is also reductionistic in that the intent is to reduce the ideas into a small, discrete set to test, such as the variables that comprise hypotheses and research questions. The knowledge that develops through a postpositivist lens is based on careful observations and measurement of the objective reality that exists in the world. Thus, developing numeric measures of observations and studying the behavior of individuals becomes paramount for a postpositivist.

Consequently, this study has been conducted by reviewing existing literature and theory and thereafter to collect data that either supports or refutes the theory. The study has aimed to develop relevant and true statements that explain the identified problem. As objectivism is an essential aspect of research, the standard of validity and reliability are important.

4.2 Overall Methodology Approach

According to Lekvall and Wahlbin (2001), the approach of a study is determined depending on two dimensions: (1) the depth or width of the study, and (2) if the study is of quantitative or qualitative nature. Furthermore, the authors list a third dimension which is time. As the purpose of this study is to evaluate predicted airline CLV as a loyalty program performance metric and the drivers of CLV within frequent flyer programs. The study takes on a quantitative approach as CLV is a quantitative metric. Furthermore, as CLV is calculated at both an individual-level and aggregate-level which requires transactional data from multiple customers, the study takes on a cross-sectional approach. The motivations for the design of this study will be given below.

4.2.1 The Suitability of a Quantitative Study

From the literature review performed in the problem analysis, it was observed that most of the previous studies regarding CLV have been of quantitative nature. Multiple studies in the past have also specifically been conducted in the airline industry setting. To advance the research further in terms of generalizable empirical contributions, it was therefore favorable to apply quantitative methods for this study as well. Even though the aim of this study is to draw qualitative and managerial implications of the analysis, we have chosen to utilize quantitative research methods to evaluate the SRQs for this study. As such the study can be classified as a quantitative study. This implies that the study will focus on testing objective theories by examining the relationships among variables (Creswell, 2013). The variables, in this case, are variables that describe customer behavior at an individual level. Quantitative research entails that concepts have to be measured, which means that numerical data has to be collected (Bryman and Bell, 2011). Furthermore, the reason for measurement in quantitative research can be summarized into three main reasons: (1) measurement allows the delineation of fine differences between people in terms of the characteristics in question, (2) measurement gives us a consistent device for making such distinctions, (3) measurement provides the basis for more precise estimates of the degree of relationship between concepts (Bryman and Bell, 2011). In the case of this study, we have examined concepts within LPs such as point pressure, in relation to program performance. The last point to be made regarding the suitability of a quantitative
study regards the availability of numerical data (Yin, 2018), which in this case have been readily available.

4.2.2 Deductive Approach
In quantitative studies, it is common to use theories deductively with the objective to test or verify the theory (Creswell, 2013). This is the case for this study as the initial problem analysis and frame of reference were conducted by reviewing literature and theories. Each variable to be analyzed have been identified and synthesized from existing theory and models. As such, comparisons can also be drawn between the results in this study and the results from previous studies. It must be noted, though, that the deductive process while seemingly linear, in reality might entail some elements of induction (Bryman and Bell, 2011). In particular, new theories might be generated out of the data that have been analyzed.

4.2.3 The Cross-Sectional Study Approach
The purpose of this study is to understand and describe how different factors drive CLV and to evaluate CLV as a performance metric. According to Churchill and Iacobucci (2005), this can be best described as descriptive research and one of the best known and most important types of descriptive design is the cross-sectional study. While cross-sectional studies traditionally use surveys as a tool to collect data (Bryman and Bell, 2011), the approach in this study has instead been to extract existing data from a customer database. According to Churchill and Iacobucci (2005), this is a more modern approach to marketing research. This is aligned with the database marketing approach highlighted by Blattberg, Kim and Neslin (2008), and has been possible due to technological advances and the availability of big data.

4.2.4 Time Perspective
Since CLV is used as a prediction of future behavior based on historical customer data, customer data covering multiple years have been extracted and analyzed. Thus, the study employs a time series analysis as described by Lekvall and Wahlbin (2001) and analyzes past behavior to predict future behavior.

4.3 An Overview of the Research Process
The research process consists of seven phases, which are illustrated in Figure 8. These steps correspond to the process of quantitative research outlined by Bryman and Bell (2011). As a first step, a literature review was conducted on the topic of CLV and LP performance to get an overview of the field. Secondly, the purpose and research questions of the study were concluded and thirdly the data for CLV calculations and regression analysis was extracted from the database of Northern Airlines. With the help of the Python Library Lifetimes (Davidson-Pilon, 2018), CLV was calculated subsequently evaluated to gather data to answer SRQ1. An operationalization of the identified CLV drivers in the context of an FFP were then evaluated with a regression analysis to answer SRQ2. Lastly, a discussion and conclusion of the findings are given to connect the findings of this study to the purpose.
4.4 Building the Theoretical Framework

To get an overview of the field of CLV and LP effectiveness, the EBSCO database in conjunction with Google Scholar and Microsoft Academic Search was used to search for relevant literature. Keywords such as “customer loyalty”, “loyalty” and “customer lifetime value” were used to build a catalog of literature which served as a basis for further study. After an initial overview of the obtained literature, relevant subfields for our study were identified. To further investigate issues regarding customer loyalty, search terms such as “loyalty program design”, “loyalty program effectiveness” were used while search terms such as “Pareto/NBD”, “BTYD”, and “non-contractual continuous time” were used to drill down on the issue regarding CLV.

To ensure that the literature was of high quality, sources to be included in this report have been vetted on the authors’ credibility and the number of citations for each source. Most of the articles cited are the works of recognized authors within the field of marketing. As the subject of LPs and CLV are expanding fields of research, recently published articles have had a higher priority than more dated articles. In exceptional cases, sources other than articles from academic journals have been used. In those cases, the authors have been considered trustworthy due to them being recognized academic authors within their field. Both the initial literature search according to selected keywords and the vetting of authors and credible sources have served to ensure that theories relevant to the study have been used. This ensures the construct validity of the study (Bryman and Bell, 2011).

4.5 Calculating Customer Lifetime Value

To answer SRQ1 and its sub-questions, CLV was calculated and evaluated using the open-source programming language Python. The objective of the calculations was to evaluate the suitability of the Pareto/NBD CLV predictions as a managerial tool in the aviation industry compared to simple managerial heuristics. The unit of analysis for SRQ1 was the airline CLV metric. LPs and FFPs can generate income through commission and the sale of points to partners, but as the purpose of this study was to evaluate the airline metric in order to understand how the FFP can contribute to the airline CLV these this type of income is not analyzed. An overview of the methodology concerning the calculation of CLV is presented in Figure 9.
4.5.1 Calculating Customer Lifetime Value
As discussed in chapter 2, there is a variety of ways to calculate CLV. The sale of flight tickets is categorized as a continuous time and non-contractual setting which limits the number of appropriate models. As the goal of the study is to evaluate drivers of CLV in the context of FFPs, it is necessary to use an individual-level CLV model to be able to connect the CLV to the customer’s interaction with the LP.

4.5.1.1 The Pareto/NBD Submodel to Predict Future Purchases
Despite being 30 years old, the Pareto/NBD model has been proposed as a good benchmark model in recent studies and is highly regarded among scholars (Fader, Hardie and Lee, 2005a; Jasek et al., 2018). As noted by Fader, Hardie and Lee (2005a), the performance of the Pareto/NBD model has been empirically validated on holdout sets whereas the performance of more recent models has not been thoroughly studied. Using a well-known model that is easily accessible through statistical software such as MATLAB, Python and R makes the study more replicable which increases the reliability. A consequence of this choice can be that the predictive accuracy can be inferior to extended CLV-models such as the Pareto/GGG by Platzer and Reutterer (2016). Fader, Hardie and Lee (2005b) calculate the discounted expected transactions (DET) for the Pareto/NBD model as:

\[
DET(\delta|\alpha, \beta, \gamma, x, t_x, T) = \frac{\alpha^\gamma \beta^s \delta^{s-1} - \Gamma(r + x + 1)\Psi[s, s; \delta(\beta + T)]}{\Gamma(r)(\alpha + T)^{r+x+1}L(r, \alpha, s, \beta|X = x, t_x, T)}
\]  

(2)

where
\(\delta\) is the rate of interest,
\(\Gamma\) is the gamma function,
\(r, \alpha, s, \beta\) are the Pareto/NBD parameters,
\(\Psi(\cdot)\) is the confluent hypergeometric function of the second kind,
\(L(\cdot)\) is the Pareto/NBD likelihood function, and
\((X = x, t_x, T)\) is the purchase history in frequency, recency and the current time

For this study, the Python package Lifetimes v0.9.0 was used to calculate number of future purchases through the Pareto/NBD model (Davidson-Pilon, 2018).

4.5.1.2 The Choice of Monetary Value Submodel
To calculate the CLV, we also need a submodel for the expected revenue per transaction (Gupta et al., 2006). Depending on the distribution of the monetary value, either the standard norm or the Gamma-Gamma submodel is suitable to use. A skewness analysis on the monetary value of the repeat customers shows a skewness of 5 indicating that the data does not follow a normal distribution. The Gamma-Gamma model allows for skewness and was thus a suitable submodel in our setting (Fader, Hardie and Lee, 2005b). Therefore, the Gamma-Gamma submodel was used instead of the norm submodel as the data cannot be considered to follow a normal distribution.

4.5.1.3 The Gamma-Gamma Submodel to Estimate the Mean Monetary Value per Transaction
Fader, Hardie and Lee (2005b, p.419) define the unobserved mean transaction value E(M) according to the Gamma-Gamma submodel as:

\[
E(M|p, q, \gamma, M_x, x) = \frac{(\gamma + m_x x)p}{px + q - 1}
\]

(3)

where
\(m_x\) is the average spend of the customer
\(x\) is the number of transaction observed by the customer
The Gamma-Gamma implementation in the Python package *Lifetimes v0.9.0* was used for calculating customers’ unobserved mean transaction value E(M).

4.5.1.4 Calculating CLV Using the Pareto/NBD and Gamma-Gamma Models

Using the submodels Pareto/NBD and Gamma-Gamma, the expected number of future transactions and the expected spend per transaction are given. Consequently, by multiply these two factors the expected total spend of the customer in the projected time period is given as shown by Fader, Hardie and Lee (2005b). Though, the combination of the two models relies on the assumption that the correlation between monetary value and frequency is close to zero (Fader, Hardie and Lee, 2005b). (Fader, Hardie and Lee, 2005b). Our dataset shows a correlation of 0.104, which can be compared to the correlation of 0.11 used by Fader, Hardie and Lee (2005b) and was therefore considered acceptable.

4.5.1.5 Simple Managerial Heuristics

To answer SRQ1, the Pareto/NBD CLV was benchmarked against what Wübren and Wangheim (2008, p.88) call simple managerial heuristics: the assumption that every customer continues to buy at his or her past mean purchase frequency. In this study, the same definition of simple managerial heuristic was adopted as this type of heuristic is also used by Northern Airlines, only adjusting the expected total spend of the customer with the discount rate.

4.5.2 Data Collection Methods

For this study, an airline that will be referred to as Northern Airlines has been studied. Northern Airlines is an FSC that is a member of a major airline alliance, mainly operating in Northern Europe. As previously mentioned in chapter 1.1, FFPs are one of the most mature examples of LPs and this is especially true for FFPs belonging to FSCs (de Boer and Gudmundsson, 2012). The structure and components of FFPs in FSCs belonging to major airline alliances such as Star Alliance, SkyTeam and OneWorld are heterogeneous as the alliances require certain benefits to be offered (Star Alliance, 2018; SkyTeam, 2018; OneWorld, 2018). Consequently, Northern Airlines was a suitable airline to study as the FFP is to be considered similar to FFPs belonging to FSCs in any major alliance. It also means that the results of this study can be generalizable to a high degree. In the following sections, the steps that have been taken to ensure that the data fits the purpose of this study will be described. The methodological choices made for the calculation of CLV will also be explained.

4.5.2.1 Revenues

As seen in the overview of FFP types in Table 7, FFPs can generate a substantial income from partners than from the airline itself. As we analyzed FFPs in the context of an LP, only revenues generated from the sale of the tickets were included. As a result, the regression analysis that will follow the CLV calculations will give information on how the LP affects the airline in selling tickets. Therefore, revenues from commissions and sale of LP points were not included. Furthermore, ancillary revenues such as upgrades, extra baggage, buy on board, seat selection etc. were not included as the studied airline did not offer most of these services at the start of the calibration period.

4.5.2.2 Marginal Cost versus Full Cost

As discussed by Blattberg, Kim and Neslin (2008), the choice between marginal cost and full cost approach to monetary value should be linked to the decision at hand for the firm. Revenue managers of airlines are tasked with balancing supply and demand by opening or closing booking classes. Given that the cost of operating a flight is largely fixed cost, it can be deemed better to sell a ticket below the fixed cost than flying with empty seats because additional revenue will still improve the profit of
the flight. The situation of airline tickets is therefore similar to the situation in which Blattberg, Malthouse and Neslin (2009) suggest the marginal cost approach to calculating CLV. As a result, a marginal cost approach was subtracted from the revenue in the monetary value found in the dataset, further on called monetary value. Two types of costs where considered: catering and on-ground related costs such as lounge and fast track which are major cost drivers. Other costs were not subtracted by the airline.

4.5.2.3 Calibration and Holdout Length

The literature proposes that CLV should be calculated for a period of three years because of the following four reasons presented by Kumar (2008): uncertainty in the prediction for longer time periods, future cash flows are discounted heavily, the managerial need is to be able to make marketing allocation based on the customers’ value in the near future. In the context of FFPs where high tier customers remain at a high tier for multiple years, a three-year horizon captures three qualification period cycles which is in line with the recommendations of Kumar (2008). For this study, the dataset supplied by Northern Airlines only contained a total of three years, which resulted in a split of a one-year calibration period and a two-year holdout period for the CLV prediction.

![Figure 10. Data split for CLV calculations](image)

4.5.2.4 Discount Rate

It is of interest to use a net present value calculation as the future profits can be considered less valuable due to the uncertainty involved and the cost of capital required to maintain the business until the profits are realized (Blattberg, Kim and Neslin, 2008). A higher discount rate means that future profits are less valuable. Blattberg, Kim and Neslin (2008) propose several ways of calculating discount rates, such as using the opportunity cost of capital, project weighted average cost of capital, beta and project-specific discount rates and more. Consequently, we have chosen to use a discount rate of 10% per the recommendation of Northern Airlines. This discount rate was applied to the holdout set in order to match the discounted value of the CLV calculations.

4.5.3 Analysis of the Data

The dataset received by Northern Airlines conforming to the methodological choices covered 13 139 049 flights flown by 369 332 LP members who all flew at least once in November 2014 for the time-period November 2014 to October 2017. A description of the dataset can be found in Table 8.

Table 8. Description of the dataset used for Customer Lifetime Value calculations.

<table>
<thead>
<tr>
<th>Number of transactions</th>
<th>Number of unique IDs</th>
<th>Mean frequency</th>
<th>Mean Monetary Value</th>
<th>Date range</th>
</tr>
</thead>
<tbody>
<tr>
<td>13 139 049</td>
<td>369 332</td>
<td>35.58</td>
<td>1 249</td>
<td>2014-11-01 - 2017-10-30</td>
</tr>
</tbody>
</table>
The data was retrieved from Northern Airlines’ internal databases, which means that the data used for analysis is of secondary nature. This implies that some aspects regarding the data have been out of control, such as how the data was initially collected. The company agreed to participate in this research on the condition of anonymity and with the prohibition of spreading the data to any third party. In the following subchapters, some methodological choices for the CLV calculations will be explained. It is not mandatory for customers to register their LP number. Consequently, there was a risk of undercoverage error as the dataset does not contain the flights of members and members where the LP number was not registered. It is likely that the lowest tier is most susceptible to this as the tendency to register your LP number is higher among customers in higher tiers as these customers are more inclined to take advantage of the benefits associated with higher tiers.

4.5.3.1 Description of the Calculation Process

The calculation process can be described as three phases: (1) preparing the data, (2) fitting the parameters of the submodels and (3) calculating the CLV.

Firstly, the dataset received by Northern Airlines was split into a calibration period and holdout period in order to be able to validate the predictions of the model (Fader, Hardie and Lee, 2005b). The data for each period was then transformed into a recency, frequency and monetary value for each customer to serve as input in the Pareto/NBD submodel (Fader, Hardie and Lee, 2005b). Secondly, the prepared dataset of the calibration period was used to fit the parameters of the Pareto/NBD and the Gamma-Gamma submodel to be able to make predictions based on the behavior of the cohort. Thirdly, the individual-level predictions on future number of flights and average monetary value were created with the fitted submodels and subsequently discounted. The managerial heuristic was also calculated by multiplying the spending in the calibration period by two and calculating the net present value with the discount factor as discussed in 4.5.1.5. As a result, three CLV values were available: the predicted CLV through the Pareto/NBD model, the predicted CLV through the managerial heuristic and the actual CLV in the holdout period.

To answer SRQ1, error metrics were calculated on the Pareto/NBD CLV model and the simple managerial heuristic (Jasek et al., 2018; Wübben and Wangenheim, 2008). Thereafter the error metrics of the Pareto/NBD CLV model was compared to the simple managerial heuristic to determine its suitability at an individual and aggregate level (Wübben and Wangenheim, 2008).

4.5.3.2 Individual-Level CLV Prediction Errors

To answer SRQ1a, the individual-level prediction of CLV were evaluated with multiple error metrics proposed by the literature. To evaluate the individual-level performance, MAE, MSLE, Mean and Median RMSE and sensibility was calculated and analyzed in this study. The metrics are defined below.

Mean Absolute Error (MAE) describes the size of the average error and is defined as (Jasek et al., 2018, p.12):

$$ MAE = \frac{1}{n} \sum_{i=1}^{n} |A_i - F_i| $$

where

$A_i$ is the actual profit of the customer during the holdout period,
$F_i$ is the predicted profit of the customer during the holdout period and
$n$ is the number of customers
Root Mean Squared Error (Mean RMSE) can be described as the standard deviation of the prediction errors and is defined as (Hyndman and Koehler, 2006, p.682):

\[
\text{Mean RMSE} = \sqrt{\text{mean}(A_i - F_i)^2}
\]  

(6)

To put the Mean RMSE into perspective, it was put in relation to the actual CLV in the holdout period, resulting in a percentage error.

Root Median Squared Error (Median RMSE) is similar to the Mean RMSE but uses the median squared error instead and is defined as (Hyndman and Koehler, 2006, p.682):

\[
\text{Median RMSE} = \sqrt{\text{median}(A_i - F_i)^2}
\]  

(7)

To put the Median RMSE into perspective, it was put in relation to the actual CLV in the holdout period, resulting in a percentage error.

Root Mean Squared Logged Error (MSLE) is similar to the Mean RMSE but penalizes underestimates more than overestimates due to the logarithmic function and is defined as (Malthouse, 2009, p.273):

\[
\text{MSLE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left[ \log(A_i + 1) - \log(F_i + 1) \right]^2},
\]

(5)

where

- \(A_i\) is the actual profit of the customer during the holdout period,
- \(F_i\) is the predicted profit of the customer during the holdout period and
- \(n\) is the number of customers

Given that the size of the MSLE error is determined by the ratio between \(A_i\) and \(F_i\), the unit used for the CLV calculations has an impact on the metric. If CLV is given in 10 000s compared to 10s, a customer who made zero repeat transactions (i.e., \(A_i = 0\)) in the holdout will have an MSLE error of 9.21 compared to 2.40 because the ratio is determined by the size of the CLV. Consequently, it can be difficult to compare MSLE scores across datasets from different industries and countries using different currencies.

Sensitivity states the percentage of customers that are correctly classified as the future top 10% or 20% and is defined as (Jasek et al., 2018):

\[
\text{Sensitivity} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}
\]  

(8)

Where True Positive is the number of customers that were correctly classified to belong to the top 10% and top 20% customers and False Negative the number of customers that were incorrectly classified to not belong in the top 10% and top 20% customers.

For the purpose of evaluating the suitability of the Pareto/NBD CLV model as a managerial tool, correlation between the predicted CLV and the actual CLV has not been studied as correlation describes the linear relationship between the two.

### 4.5.3.3 Aggregate-Level CLV Prediction Errors

To answer SRQ1b, the literature suggests two error measures that can be used to compare the predicted CLV values to the managerial heuristic: Forecast vs Actual (FvA) and Mean Absolute
Percentage Error (MAPE). Whereas other authors (Jasek et al., 2018; Wübben and Wangenheim, 2008) have used both when comparing the errors of different datasets, only one dataset has been provided in his study. Consequently, the FvA measure will be sufficient to understand the aggregate-level accuracy as the time period for calculating the Pareto/NBD CLV and managerial heuristic CLV (Leeflang et al., 2015). The Forecast vs Actual metric describes whether the forecast over- or underestimates the CLV and is defined as (Jasek et al., 2018, p.12):

\[
F_{\text{vs}A} = \frac{F_t}{A_t} * 100
\]

where

\[F_t\] is the sum of forecasted CLV for all customers in the holdout period
\[A_t\] is the sum of actual CLV for all customers in the holdout period

To further understand the aggregate-level accuracy, the accuracy across all customers as well as on the tier levels of the LP has been studied to understand the accuracy of the models on subsegments of the customers.

### 4.6 Performing the Regression Analysis

To answer SRQ2 and its sub-questions, a multiple regression analysis was conducted using the SPSS statistics program on the same sample of data spanning the calibration period of the CLV calculations. The objective of the regression analysis was to evaluate and explain CLV and its association to the following drivers: (1) points-pressure, (2) cross-buying, (3) rewarded-behavior, (4) promotions, and (5) marketing efforts. Regression analysis was chosen as a method since it is a common and versatile technique that is applicable to every facet of business decision making; It is used as a statistical technique to analyze the relationship between a single dependent variable and several independent variables with the following basic formulation (Hair, 2006, p.169):

\[
Y_1 = X_1 + X_2 + \cdots + X_n
\]

In our case, the dependent variable \(Y_1\) is the CLV value assigned to each customer at the end of the calibration period and the independent variables \(X_{1,n}\) are metrics of the drivers described above, which are the sub-questions of SRQ2. The unit of analysis for SRQ2 and its sub-questions is the LP and its members. This is because the drivers are derived from the customers’ purchasing behavior, which means that individual purchasing data is required.

### 4.6.1 Operationalization of Customer Lifetime Value Drivers

The process of creating variables from raw data suitable for prediction called feature engineering is used to create features that represent the above-mentioned drivers. Appropriate features are an important factor in making machine learning projects work and can thus also be important when constructing a regression analysis (Domingos, 2012). Domingos states that good feature engineering is complex as it requires both domain-specific knowledge, intuition and creativity and therefore the only way of determining if the feature engineering is good is through iterative trial-and-error testing.

To create engineered features, workshops were conducted with employees within the FFP department at Northern Airlines to capture the domain-specific knowledge and intuition of the employees and conduct trial-and-error tests as recommended by Domingos (2012). The workshops began with an introduction to the drivers and their definition, followed by an open discussion on each driver and how they could be engineered or operationalized. The current alternatives for engineered features are presented in Table 9.
Table 9. Engineered features with definitions used for the regression analysis.

<table>
<thead>
<tr>
<th>Engineered Feature</th>
<th>Operationalization</th>
<th>Abbreviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Points-pressure and Rewarded-behavior</td>
<td>Has the customer redeemed any flights with points</td>
<td>RDMP</td>
</tr>
<tr>
<td></td>
<td>Has the customer been upgraded to a higher customer tier</td>
<td>UPGT</td>
</tr>
<tr>
<td></td>
<td>Has the customer been downgraded to a lower customer tier</td>
<td>DWGT</td>
</tr>
<tr>
<td>Cross-Buying</td>
<td>Has the customer earned points from air/non-air partners except Northern Airlines</td>
<td>UNQP</td>
</tr>
<tr>
<td>Promotions</td>
<td>Has the customer purchased any flights through promotions</td>
<td>PRMT</td>
</tr>
<tr>
<td>Marketing Efforts</td>
<td>How many emails have the customer opened</td>
<td>-</td>
</tr>
</tbody>
</table>

After the workshops, it was deemed difficult to measure points-pressure and rewarded-behavior separately based on the data that was available. Instead, these two mechanisms will be measured as a combined driver which is motivated by the fact that these effects collectively act together in the case of redemptions (Bijmolt, Dorotic and Verhoef, 2011). Variable X...XN were set as categorical variables since only a small percentage of the customers would have any values greater than zero. Essentially this means that these variables can only take the values of 0 for cases equal to zero and 1 for cases with values greater than zero. It was also deemed difficult to assess the impact of marketing efforts as the data-set available lacked information on email interactions. Thus, marketing efforts had to be omitted in the analysis.

4.6.2 Data Collection Methods
The requirements for the data needed to perform the regression analysis were set by the operationalization of drivers into variables. Based on this, additional data corresponding to the operationalized variables was given by Northern Airlines from the same customers of the CLV-calculations during the calibration period. Descriptive statistics for the chosen variables are presented in Table 10.

Table 10. Descriptive statistics of the independent variables used in the regression analysis.

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Statistic</td>
<td>Statistic</td>
<td>Statistic</td>
<td>Statistic</td>
<td>Statistic</td>
<td>Statistic</td>
<td>Std. Error</td>
</tr>
<tr>
<td>DWGT</td>
<td>369332</td>
<td>0</td>
<td>1</td>
<td>-</td>
<td>0.224</td>
<td>3.989</td>
<td>0.004</td>
</tr>
<tr>
<td>UPGT</td>
<td>369332</td>
<td>0</td>
<td>1</td>
<td>-</td>
<td>0.418</td>
<td>1.316</td>
<td>0.004</td>
</tr>
<tr>
<td>FREQ</td>
<td>369332</td>
<td>0</td>
<td>1</td>
<td>-</td>
<td>6.582</td>
<td>2.041</td>
<td>0.004</td>
</tr>
<tr>
<td>REC</td>
<td>369332</td>
<td>0</td>
<td>52</td>
<td>-</td>
<td>18,360</td>
<td>0.775</td>
<td>0.004</td>
</tr>
<tr>
<td>CLV</td>
<td>369332</td>
<td>0</td>
<td>1</td>
<td>-</td>
<td>1,11040</td>
<td>0.208</td>
<td>0.004</td>
</tr>
<tr>
<td>RDMP</td>
<td>369332</td>
<td>0</td>
<td>1</td>
<td>-</td>
<td>0.454</td>
<td>0.926</td>
<td>0.004</td>
</tr>
<tr>
<td>PRMT</td>
<td>369332</td>
<td>0</td>
<td>1</td>
<td>-</td>
<td>0.212</td>
<td>4.262</td>
<td>0.004</td>
</tr>
<tr>
<td>UNQP</td>
<td>369332</td>
<td>0</td>
<td>1</td>
<td>-</td>
<td>0.480</td>
<td>-0.589</td>
<td>0.004</td>
</tr>
</tbody>
</table>

In addition to the engineered features and CLV, frequency (FREQ) and recency (REQ) are added to the regression model as control variables. These variables serve as input variables in the CLV-
calculations and therefore have a strong contribution to the dependent variable. Initially, monetary value was also considered as a control variable but was eventually excluded since the variable greatly increased the heteroscedasticity of the regression model. This might be due to the heterogeneous nature of the Pareto/NBD distributions of CLV and the Gamma-Gamma distribution of monetary value.

4.6.3 The Regression Model Used in This Study

After the initial steps of engineering the features that represent the drivers of CLV, the stepwise estimation technique was used to estimate the regression model. In this procedure, the independent variable with the greatest contribution is added first and independent variables are then added based on their incremental contribution over the variables already included in the model (Hair, 2006). This procedure was automated in SPSS. The final regression model that was generated can be described as:

\[
\ln CLV_i = \beta_0 + \beta_1 FREQ_i + \beta_2 REC_i + \beta_3 UNQP_i + \beta_4 RDMP_i + \beta_5 UPGT_i + \beta_6 DWGT_i + \beta_7 PRMT_i
\] (11)

where

- CLV<sub>i</sub> = the CLV value of each customer
- FREQ<sub>i</sub> = the number of purchases for each customer
- REC<sub>i</sub> = the recency for each customer
- UNQP<sub>i</sub> = if a customer has earned points from multiple partners
- RDMP<sub>i</sub> = if a customer has redeemed flights or upgrades with points
- UPGT<sub>i</sub> = if a customer has been upgraded to a higher tier
- DWGT<sub>i</sub> = if a customer has been downgraded to a lower tier
- PRMT<sub>i</sub> = if a customer has purchased a flight through a promotion

The dependent variable CLV has been log-transformed to treat non-linearity and heteroscedasticity. The overall fit of the model is given by the R square value of 0.84 as seen in Table 11, this is naturally very high due to the inclusion of frequency and recency as independent variables.

<table>
<thead>
<tr>
<th>Model Summary&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
</tr>
<tr>
<td>-------</td>
</tr>
<tr>
<td>1</td>
</tr>
</tbody>
</table>

a. Predictors: (Constant), FREQ, DWGT, PRMT, UPGT, UNQP, RDMP, REC
b. Dependent Variable: CLVpnbd_Ln1

Furthermore, the model has a significance level less than 0.01. To assess the relative magnitude of each driver, the standardized coefficient for each independent variable have been compared. All independent variables are statistically significant as all variables have a sig. value less than 0.01. This means that all predictors contribute significantly to the total variance of the regression model.

4.6.4 Evaluation of the Regression Analysis

To evaluate the regression analysis, four assumptions must be examined when conducting a linear regression analysis (Hair, 2006): (1) the linearity of the phenomenon measured, (2) constant variance of the error terms, (3) independence of the error terms, and (4) normality of the error term distribution. In addition, multicollinearity is also a concern in multiple regression since it may lead to skewed coefficients in the regression model (Hair, 2006). These assumptions are all in
consideration of the relationship between the dependent variable and the independent variables. In multiple regression, the variate of the independent variables acts collectively in predicting the dependent variable. This necessitates both testings for assumptions during the initial phases of the regression analysis, and when the model has been estimated (Hair, 2006). Checking the assumptions can be easiest done by observing the residual plots in Figure 11 and Figure 12.

Figure 11. Scatterplot of residuals.

Figure 11 shows a scatterplot of the standardized predicted value in comparison to the standardized residual. This shows how the residuals vary depending on the predicted value of the regression model. Ideally, this should show a random distribution across the range of values (Hair, 2006).
Figure 12. Histogram of the residuals.

Figure 12 shows the distribution of the standardized residuals of the regression model and their fit to the normal distribution. Ideally, the residuals should be normally distributed (Hair, 2006), which in this case can be seen as the residuals are distributed along the normal curve.

4.6.4.1 Linearity of the Phenomenon
To test the linearity of the phenomenon, the residual plot of the regression model was examined. This is according to Hair (2006) the easiest way to examine linearity as non-linearity would show a curvilinear shape of the residuals. Initial estimations showed signs of non-linearity. This was treated with a natural log-transformation of the dependent variable CLV. As a result, no major signs of non-linearity could be observed in Figure 11.

4.6.4.2 Constant Variance of the Error Terms
The residual plots were also used to assess any presence of unequal variance which is called heteroscedasticity (Hair, 2006). This was also treated with the same log-transformation of the dependent variable. As can be seen in Figure 11, the residuals do show small signs of heteroscedasticity as the variance is smaller for the higher predicted values. Ideally, the pattern should be randomly distributed across the plot (Hair, 2006). No corrective action was taken for this as the residuals in general have a homoscedastic pattern as most of the residuals are randomly distributed across the predicted values. Furthermore, the objective of the regression analysis is of explanatory nature rather than predictive, which means that our constraints for heteroscedasticity are more relaxed.

4.6.4.3 Independence of the Error Terms
In regression analysis, it is assumed that each predicted value is independent (Hair, 2006). As the regression analysis in this case is not in the form of time series data, there should be no carryover from one observation to another. Each observation represents a unique customer and its actions throughout the calibration period.
4.6.4.4 Normality of the Error Term Distribution

To check for the normality of the error term distribution, a histogram of the residuals was plotted which can be seen in Figure 12. This is one of the simplest diagnostic tools that are suitable for larger data sets (Hair, 2006). What can be observed is that the residuals are normally distributed with a sharp peak near the mean. In large samples with over 200 cases, though, the impact of normality effectively diminishes (Hair, 2006). Thus, we have taken a lenient stance on meeting this assumption.

4.6.4.5 Multicollinearity

An additional assumption that deals with the relation between the included independent variables is multicollinearity. High degrees of multicollinearity can prevent the estimation of coefficients and result in coefficients being incorrectly estimated with wrong signs (Hair, 2006). To assess this, the correlations of both the dependent variable and all the independent variables in Table 12 have been examined.

Table 12. Correlations of the dependent and independent variables.

<table>
<thead>
<tr>
<th>Correlations</th>
<th>CLVpnbd_Ln1</th>
<th>DWGT</th>
<th>UPGT</th>
<th>RDMP</th>
<th>PRMT</th>
<th>UNQP</th>
<th>REC</th>
<th>FREQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pearson Correlation</td>
<td>1.000</td>
<td>.020</td>
<td>.261</td>
<td>.391</td>
<td>.164</td>
<td>.399</td>
<td>-.793</td>
<td>.820</td>
</tr>
<tr>
<td>DWGT</td>
<td>.020</td>
<td>1.000</td>
<td>-.053</td>
<td>.072</td>
<td>.028</td>
<td>.064</td>
<td>-.022</td>
<td>-.014</td>
</tr>
<tr>
<td>UPGT</td>
<td>.261</td>
<td>-.053</td>
<td>1.000</td>
<td>.083</td>
<td>.034</td>
<td>.159</td>
<td>-.206</td>
<td>.235</td>
</tr>
<tr>
<td>RDMP</td>
<td>.391</td>
<td>.072</td>
<td>.083</td>
<td>1.000</td>
<td>.349</td>
<td>.318</td>
<td>-.278</td>
<td>.365</td>
</tr>
<tr>
<td>PRMT</td>
<td>.164</td>
<td>.028</td>
<td>.034</td>
<td>.349</td>
<td>1.000</td>
<td>.143</td>
<td>-.110</td>
<td>.176</td>
</tr>
<tr>
<td>UNQP</td>
<td>.399</td>
<td>.064</td>
<td>.159</td>
<td>.318</td>
<td>.143</td>
<td>1.000</td>
<td>-.314</td>
<td>.315</td>
</tr>
<tr>
<td>REC</td>
<td>-.793</td>
<td>-.022</td>
<td>-.206</td>
<td>-.278</td>
<td>-.110</td>
<td>-.314</td>
<td>1.000</td>
<td>-.568</td>
</tr>
<tr>
<td>FREQ</td>
<td>.820</td>
<td>-.014</td>
<td>.235</td>
<td>.365</td>
<td>.176</td>
<td>.315</td>
<td>-.568</td>
<td>1.000</td>
</tr>
</tbody>
</table>

According to Hair (2006), correlations of .90 and higher indicates substantial collinearity in the model. As can be observed, the highest correlation among the independent variables is between FREQ and REC (-.568) which is thus well below the threshold.

As collinearity might depend on the combined effect of multiple independent variables, the effects of multicollinearity must also be assessed based on the tolerance or variation inflation factor (VIF) (Hair, 2006). VIF is the inverse of tolerance, which means that the measures can be used interchangeably. The tolerance and VIF statistics of the regression model is given in Table 13.

Table 13: Collinearity statistics of the regression model.

<table>
<thead>
<tr>
<th>Independent variable</th>
<th>Tolerance</th>
<th>VIF</th>
</tr>
</thead>
<tbody>
<tr>
<td>DWGT</td>
<td>0.986</td>
<td>1.014</td>
</tr>
<tr>
<td>UPGT</td>
<td>0.927</td>
<td>1.078</td>
</tr>
<tr>
<td>RDMP</td>
<td>0.742</td>
<td>1.348</td>
</tr>
<tr>
<td>PRMT</td>
<td>0.875</td>
<td>1.143</td>
</tr>
<tr>
<td>UNQP</td>
<td>0.823</td>
<td>1.214</td>
</tr>
<tr>
<td>REC</td>
<td>0.651</td>
<td>1.537</td>
</tr>
<tr>
<td>FREQ</td>
<td>0.608</td>
<td>1.644</td>
</tr>
</tbody>
</table>
A common cutoff threshold is a VIF value of 10 (Hair, 2006). As can be observed, these measures are all within acceptable limits with the highest VIF measure of 1.644 for the FREQ variable, which is well below the cutoff threshold.

4.7 Ethical Considerations
This study has been conducted with the ethical considerations following the norms in the Swedish market research field (Lekvall, Wahlbin and Frankelius, 2001). As no primary data has been collected, no interaction has been held between the researchers and the studied customers. Even though this process has been out of reach for the researchers, we want to acknowledge the importance of consent and privacy. In this regard, all customer data have been collected at the consent of the customer and according to regulations. To access and use the customer data, the researchers have been clear on what the purpose of the study has been. In this regard, the researchers have involved Northern Airlines who holds the data throughout the study.

The anonymity of each customer has been protected as no information in the report can be traced back to a single customer through identifiers and personal information. To ensure that no distortion has been made to the data during the analysis, all steps of the analysis and the corresponding evaluation criteria have been shown as clearly as possible in the report. Furthermore, quality metrics have been used in the quantitative analysis and any results that impact the analysis negatively have been disclosed.

4.8 Evaluation of the Quality of the Study
According to Yin (2018), there are for tests common to all social science methods and these can be used to evaluate the quality of the study: (1) construct validity, (2) internal validity, (3) external validity, and (4) reliability. These quality factors are also advocated by Bryman and Bell (2011) who give the following description of each term. Construct validity concerns the identification of correct operational measures for the concepts being studied. Internal validity is mainly a concern for explanatory case studies that try to determine whether event x led to event y and is therefore not applicable to this study which is a descriptive cross-sectional study. External validity concerns the domain to which a study’s findings can be generalized. Reliability concerns the consistency of a measure of a concept and the replicability of a study, such that the data collection procedures can be repeated with the same results.

4.8.1 Construct Validity
To ensure the construct validity of our study, three tactics are proposed: the use of multiple sources of evidence that support each other, chain of evidence, and having the draft of the study reviewed by key informants (Yin, 2018). Firstly, multiple studies by different authors from peer-reviewed journals have been triangulated to form our understanding of the topic of this study. The literature used in the study have also been vetted on their relevancy to the problem of this study. Secondly, to support the validity of the third research question, the accuracy of the CLV metric has been evaluated. Thirdly, the study has been reviewed by opponents, an academic supervisor and Northern Airlines multiple times. To further improve the validity when addressing the purpose, the accuracy of the CLV metric has been evaluated.

4.8.2 External Validity
To increase the external validity of the study, we have used data from a full-service carrier with an advanced FFP and made general conclusions in response to the purpose. As a result, the findings of the study are relevant for other full-service carriers. This means that the results from this study may be applicable to other airlines with similar FFPs which ensures the generalizability of our findings.
To ensure external validity, theories must be generalizable to other settings than the setting of one study to ensure external validity (Bryman and Bell, 2011; Gibbert, Ruigrok and Wicki, 2008).

4.8.3 Reliability
To increase the replicability of the study, it is of importance to document the methodology used in the study (Yin, 2018). Therefore, an open-source Python package has been used for the advanced calculations and the associated code has been made available in the appendix. To check for internal reliability of the study, the error metrics were also calculated on a subset of 10% of the total number of customers. The result on the subset of customers was in line with the result on the entire dataset. Furthermore, regression analysis is a common analytical method and was conducted using the well-known SPSS statistics program.

4.8.4 Objectivity
The study has been performed on-site at the head office of Northern Airlines, but the authors have not had a close relation to the employees of Northern Airlines concerning the study. CLV has been a subject of interest, but Northern Airlines has had little influence on the direction of the study. The authors have taken the initiative when constructing the purpose and SRQs without the influence of the company. One of the authors was employed by Northern Airlines at the time of the study, which might have created bias. The negative impact of this bias was prevented by an objective mindset and by following acknowledged methodology seen in previous studies. The positive impact was that the domain-specific knowledge of the aviation and FFP industry helped the creation of the study.

4.8.5 Sources of Error
Given the quantitative approach used in this study, the sources of error lie mainly within the data used and the methodology used to analyze the data. The dataset used in this study was of secondary nature and therefore outside of the authors’ control. To decrease the risk of error related to the dataset a descriptive analysis was conducted and the results were verified with Northern Airlines. As seasonality trends exist in the aviation industry, there is a risk that the choice of cohort impacts the study. To minimize this impact, the airline was asked to avoid months they considered non-representative. Revenues from ancillary products such as upgrades, lounge and more were not included. This might affect the calculated CLV as ancillary revenues have become an increasingly important source of revenue for airlines in recent years as customers pay less for the tickets themselves and increasingly pay for ancillaries (Warnock-Smith, O’Connell and Maleki, 2017). Consequently, the calculated CLV will most likely overestimate income from the sale of tickets whereas the total income could potentially be higher. To decrease the risk of error related to the methodology used, open source tools have been used and methodology approaches similar to previous studies have been used for the CLV calculations. The error stemming from the operationalization of the CLV drivers has been minimized by holding workshops and discussions with Northern Airlines to capture the employee's domain-specific knowledge.
5 Data and Analysis

This chapter describes and analyses the result of the study by answering SRQs. Support is drawn from previous relevant academic literature.

5.1 The Pareto/NBD CLV Predictions as a Managerial Tool

To understand the suitability of the Pareto/NBD CLV model, we analyzed the error metrics at an individual- and aggregate-level and compare it to the managerial heuristic.

Firstly, the individual-level errors are large in relation to the average CLV as seen in Table 14. The MAE of the Pareto/NBD CLV model is equal to 66% of the average CLV. This can be compared to a weighted mean MAE of 113.7% in the six datasets analyzed by Jasek et al. (2018). Given the different industry studied and different currency used the MSLE seen in this study is difficult to analyze and compare the result of 3.54 and 3.53 to the JIM winner score of 1.69. The Mean RMSE is over 120% of average CLV for the Pareto/NBD CLV model and 145% for the managerial heuristic, indicating a large error. Unfortunately, previous studies have not described RMSEs in percentage of average CLV making them difficult to compare to the results of this study. The difference between the MAE and Mean RMSE can be explained by the fact that Mean RMSE is more sensitive to outliers due to the squared term. The managerial heuristic Median RMSE is almost half the size of the Pareto/NBD, at 34 versus 13.

Table 14. Individual-level error metric in percent of average holdout CLV.

<table>
<thead>
<tr>
<th>Model</th>
<th>MAE</th>
<th>MSLE</th>
<th>Mean RMSE</th>
<th>Median RMSE</th>
<th>Sensitivity 10%</th>
<th>Sensitivity 20%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pareto/NBD</td>
<td>66.46</td>
<td>3.54</td>
<td>119.62</td>
<td>33.81</td>
<td>62.80</td>
<td>69.67</td>
</tr>
<tr>
<td>Heuristic</td>
<td>78.38</td>
<td>3.53</td>
<td>144.78</td>
<td>13.00</td>
<td>62.26</td>
<td>68.78</td>
</tr>
</tbody>
</table>

Sensitivity is slightly better than in the study by Wübben and Wangheim (2008, p.91) who find that the Pareto/NBD model correctly classify about 67% of the top 20% customers and much better than the 45% achieved by Malthouse and Blattberg (2005). Similar results can be seen for the top 10% customers where both models correctly classify about 62-63% of actual top 10% customers. In line with Wübben and Wangheim (2008, p.91), we do not consider the Pareto/NBD CLV model to be a suitable managerial tool due to the large error in regards to MAE and Mean RMSE.

Secondly, the FvA aggregate-level error is small as seen in Table 15. The Pareto/NBD CLV model was 127.6% of the holdout period CLV, whereas the managerial heuristic resulted in 148.2% indicating that the Pareto/NBD is ~20 percentage points lower. This means that the Pareto/NBD CLV overestimates the CLV with 27.6%.

Table 15. Forecast vs Actual error metric as a percentage of the average holdout CLV.

<table>
<thead>
<tr>
<th>Tier</th>
<th>Pareto/NBD</th>
<th>Heuristic</th>
</tr>
</thead>
<tbody>
<tr>
<td>All tiers</td>
<td>127.6</td>
<td>148.2</td>
</tr>
<tr>
<td>Tier 1</td>
<td>140.8</td>
<td>143.6</td>
</tr>
<tr>
<td>Tier 2</td>
<td>128.6</td>
<td>146.9</td>
</tr>
<tr>
<td>Tier 3</td>
<td>124.7</td>
<td>149.4</td>
</tr>
<tr>
<td>Tier 4</td>
<td>122.8</td>
<td>151.8</td>
</tr>
</tbody>
</table>

To understand the FvA error better, the error was also calculated by LP tier level where tier 1 is the lowest tier. As Table 15 shows, the Pareto/NBD CLV model is most accurate among tier 2 and up whereas the managerial heuristic performs fairly equal across all tier levels. Our findings for all tiers are in line with the study by Wübben and Wangheim (2008, p.89), as they find the aggregate-level
CLV to be superior to the managerial heuristic when using the BG/NBD model which outperformed the Pareto/NBD model on their dataset. In line with Wübben and Wangheim (2008, p.91), we find the Pareto/NBD CLV model to be a suitable as a managerial tool due to the superior performance compared to the simple managerial heuristic.

On one hand, the Pareto/NBD CLV prediction is suitable as a managerial tool in the aviation industry at an aggregate-level because of the superior performance compared to the managerial heuristic. On the other hand, the errors at an individual-level are too high to consider it a suitable managerial tool.

5.2 The Drivers of CLV in the Context of Loyalty Programs
The final regression model included a sample size of 369332 cases which corresponds to unique customers and their transactions throughout the calibration period. The dataset is the same set that was used in the CLV calculations with the addition of measures on the operationalized variables. The results from the regression data and analysis show that the positive effect of cross-buying, points-pressure and rewarded-behavior on CLV can be empirically justified in the context of FFPs with cross-buying having the highest impact. This will be further explained and justified below.

5.2.1 Overall Description of the Regression Model Results
The overall model fit is $R^2 = 0.84$, which indicates that a high degree of the total variation of CLV is explained by the model. Furthermore, the confidence intervals for all regression parameters are excluded from zero. Table 16 provide the standardized coefficients for all parameters that were used as independent variables.

Table 16. Standardized coefficients from the regression analysis.

<table>
<thead>
<tr>
<th>Model</th>
<th>Standardized Coefficients</th>
<th>t</th>
<th>Sig.</th>
<th>95.0% Confidence Interval for B</th>
<th>Collinearity Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Beta</td>
<td></td>
<td></td>
<td>Lower Bound</td>
<td>Upper Bound</td>
</tr>
<tr>
<td>1</td>
<td>(Constant)</td>
<td>4788.838</td>
<td>0.000</td>
<td>9.364</td>
<td>9.372</td>
</tr>
<tr>
<td></td>
<td>DWGT</td>
<td>0.010</td>
<td>15.825</td>
<td>0.000</td>
<td>0.045</td>
</tr>
<tr>
<td></td>
<td>UPGT</td>
<td>0.031</td>
<td>45.951</td>
<td>0.000</td>
<td>0.080</td>
</tr>
<tr>
<td></td>
<td>RDMP</td>
<td>0.053</td>
<td>68.956</td>
<td>0.000</td>
<td>0.125</td>
</tr>
<tr>
<td></td>
<td>PRMT</td>
<td>-0.006</td>
<td>-8.865</td>
<td>0.000</td>
<td>-0.040</td>
</tr>
<tr>
<td></td>
<td>UNQP</td>
<td>0.073</td>
<td>100.104</td>
<td>0.000</td>
<td>0.165</td>
</tr>
<tr>
<td></td>
<td>REC</td>
<td>-0.460</td>
<td>-564.851</td>
<td>0.000</td>
<td>-0.028</td>
</tr>
<tr>
<td></td>
<td>FREQ</td>
<td>0.510</td>
<td>605.188</td>
<td>0.000</td>
<td>0.086</td>
</tr>
</tbody>
</table>

As expected the greatest contribution to CLV is given by the two control variables frequency (FREQ .510, sig. < .01) followed by recency (REC -.460, sig. < .01). This is intuitive since both variables are input variables for the Pareto/NBD-model used to calculate the individual-level CLV for each customer. Recency has a negative impact which is expected since a higher recency value lowers the survival probability of a customer and thus lowers the customers CLV. Following these two variables is unique partners (UNQP .073, sig. < .01) that expectedly has a positive impact on CLV. This gives
empirical evidence that cross-buying is a strong driver of CLV and supports previous theory (Blattberg, Malthouse and Neslin, 2009). All variables that are operationalizations of points-pressure and rewarded-behavior have a positive impact on CLV. The strongest being redemptions (RDMP .053, sig < .01) followed by upgraded customer tier (UPGT .031, sig. < .01), and downgraded customer tier (DWGT .010, sig. < .01). Surprisingly, promotions have a negative impact on CLV (PRMT -.006, sig. < .01). This deviates from Blattberg, Malthouse and Neslin (2009), who expects a positive relationship between CLV and promotions.

5.2.2 Cross-Buying has a Positive Impact on CLV
As the results show, cross-buying has, relative to the other drivers in the model, the highest impact on CLV. According to Blattberg, Malthouse and Neslin (2009), this is mainly due to an increase in relationship duration and customer profitability. In the setting of this study the positive impact of cross-buying cannot be attributed to an increase in profitability since the measure of cross-buying is if the customer has earned points from multiple partners. As such, the increase in CLV due to cross-buying can instead be attributed to an increase in relationship duration. A question posed by Blattberg, Malthouse and Neslin (2009) is regarding the causality of the effect of cross-buying, which could not be assessed in this study. Therefore, it must be noted that the evidence of cross-buying can be the result of customers who are already heavily invested in the product and company.

5.2.3 Points-Pressure and Rewarded-behavior have a Positive Impact on CLV
Evidently, points-pressure and rewarded-behavior have a positive effect on CLV which is in line with the assumption in the analytical model that CLV captures measures for LP performance. The impact of redemptions and customer tier upgrades are in line with findings by Kopalle et al. (2009), who attributed the effects of these actions to the points-pressure and rewarded-behavior mechanisms. Redemptions are the clearest evidence of these effects as it requires the active choice of a customer to use their points to redeem a flight, whereas customer tier upgrades should naturally contribute to higher CLV as the event is inherently engineered to award customers who fly more than their peers. These results also suggest that rewarded-behavior has a long-term effect due to CLV being a forward-looking and predictive metric of the future value of a customer. This can be linked with Taylor and Neslin (2005) who suggest that the effect of rewarded-behavior is long-term.

5.2.4 Promotions have a Negative Impact on CLV
The negative association of promotions and CLV is a surprising result as it deviates from previous literature (Blattberg, Malthouse and Neslin, 2009). A reasoning for this might be that customers who take advantage of promotions are more restrictive in their purchasing behavior and thus fly less frequently than customers not taking advantage of promotions. This would be in line with findings by Kopalle et al. (2009) who found evidence of two distinct customer segments in FFPs of which one is a deal prone segment that behaves differently from the other customers. Furthermore, the relative impact is small in comparison with the highest contributing variables in the regression model. Therefore, the practical significance of the results regarding promotions are weak.
6 Conclusions and Contributions

This chapter concludes the study by relating the answers of the SRQs to the purpose. Recommendations and suggestions for further studies are given and the contribution of the study is discussed.

6.1 General Conclusions from the Study

The purpose of this study is to evaluate predicted airline CLV as an LP performance metric and the drivers of CLV within FFPs. Stated below are general conclusions in bold that can be drawn from the analysis and their justifications.

The Pareto/NBD Customer Lifetime Value prediction is only suitable as an aggregate-level metric, meaning it is not suitable at an individual-level.

CLV is a suitable tool, but it is of importance to understand that it is forward-looking and therefore comes with an error. Understanding the implications of the error is crucial when determining the usage of the metric. On one hand, the Pareto/NBD CLV prediction show a strong performance with an overestimate of 28% at an aggregate-level making it suitable as a managerial tool as shown in 5.1. On the other hand, the prediction shows a weak performance that is slightly better than the managerial heuristic making it disputable as a managerial tool as shown in 5.1.

Loyalty program mechanisms have a positive effect on CLV.

The analysis in 5.2.3 shows that points-pressure and rewarded-behavior, in line with the analytical model, can be positively associated with an increase in CLV. These results provide a linkage between the main mechanisms of LPs, program performance, and CLV. With the analysis drawn from 5.1, measuring CLV at an aggregate-level can be a suitable tool to manage the impact of these mechanisms on different customer segments. Looking at the relative impact of each driver in 5.2.1 it can be concluded that cross-buying has the largest impact on CLV, followed by points-pressure and rewarded-behavior. This means that cross-buying is a generalizable CLV driver in the context of FFPs. Furthermore, it suggests that firms should utilize LPs to induce cross-buying. To conclude, the reward structure of LPs leads to higher CLV. Enhancing the design of the LP with mechanisms such as points-pressure and rewarded-behavior in mind should therefore lead to an increase in CLV and greater returns for the firm.

6.2 Discussion of the Analytical Model

In the analytical model it was assumed that points-pressure, rewarded-behavior, cross-buying, promotions, and marketing efforts would have a positive impact on CLV. Through the analysis, points-pressure, rewarded-behavior, and cross-buying have been confirmed to have this effect. Contrary to the analytical model, promotions had a negative association with CLV. This could warrant a revision of the analytical model but due to the small relative impact and the measurement of this variable, we have instead chosen to leave this for future studies. Furthermore, we have not been able to analyze the relationship between marketing efforts and CLV and therefore leave this for future studies.

6.3 Implications for Practitioners

From the two general conclusions in 6.1, four implications are suggested for practitioners.

The usage of Pareto/NBD Customer Lifetime Value should be limited to be used as an aggregate-level performance metric

Use the Pareto/NBD CLV as a Key Performance Indicator
As discussed in 1.2, CLV is a metric that should be of large interest for FFPs given the heterogeneity of the customer base where high tier customers can be exponentially more worth than low tier customers. With the Pareto/NBD CLV, it is possible to track behavioral changes for segments of the customer-base over time that indicate how strong or weak the relationship is between customer segments and the airline with an acceptable accuracy.

Do not use individual-level CLV for allocation of marketing resources at an individual-level

Given the large errors of the individual-level CLV predictions through the Pareto/NBD model seen in 5.1, we do not recommend the usage of individual-level CLV to allocate marketing resources despite what previous theoretical studies have suggested as seen in 1.2.1.

Loyalty program mechanisms have a positive effect on CLV.

Facilitate cross-buying for customer retention.

As stated in 6.1, the relative contribution of cross-buying is the highest among the drivers analyzed in 5.2, managers should look for opportunities to increase cross-buying among the firm’s customers such as through further incentivizing customers to earn miles through partners such as credit cards, hotels and more. Since cross-buying increases CLV by extending the relationship duration between a firm and its customers (see 5.2.2), it acts as a viable tool for customer retention.

Utilize redemptions to trigger points-pressure and rewarded-behavior.

From theory it is suggested that points-pressure and rewarded-behavior is induced through redemptions. The results from 5.2 supports this and suggest that redemptions are associated with an increase in CLV. As a main component of LPs such as FFPs, managers of these programs should therefore utilize redemptions to increase CLV.

6.4 Academic Contributions

Given the extraordinary research opportunity enabled by the accessibility of both data for CLV calculations and the LP data used for the regression analysis, this study has contributed both theoretically, methodologically and practically.

Given the generic nature of the Pareto/NBD CLV model which is applicable to any industry, we have theoretically contributed by validating its suitability at an aggregate-level in the aviation industry. Furthermore, the study has investigated to what degree the identified drivers contribute to CLV vis-à-vis each other whereas previous studies have examined the drivers independently (Blattberg, Malthouse and Neslin, 2009).

Methodologically, a link between CLV and LP performance has been made which is previously unseen in the studied literature. This study shows that CLV can be used as a measurement when evaluating the performance of an LP and the effects of design components in LPs. This is an addition to performance measures listed by Watson et al. (2015) such as sales, profit, and share of wallet

Practically, the results of the study can be of interest for airlines as it both proposes an easy way to quantitatively analyze behavioral loyalty and a way to evaluate how the FFP contributes to the airline CLV.

6.5 Suggestions for Future Studies

Given the methodology of this study, the causality between the operationalized drivers and CLV cannot be determined. This is similar to previous studies that have examined the relationship between the drivers and CLV without investigating causality (Blattberg, Malthouse and Neslin, 2009;
Blattberg, Kim and Neslin, 2008). To understand the relationship between the drivers and CLV a study on the causality should be conducted. Another limitation with the methodology regards the effects of points-pressure and rewarded-behavior. Future studies should evaluate how these mechanisms individually affect CLV to understand these mechanisms better. Due to the weak negative association of promotions on CLV, future studies should investigate the relationship between promotions and CLV further. Furthermore, since no analysis could be conducted on the relationship between marketing efforts and CLV, this is another area for future studies.

A limitation of using the well-known Pareto/NBD model to calculate CLV is that newer and less-known CLV models with potentially superior performance such as the Pareto/GGG (Platzer and Reutterer, 2016) were excluded. Further studies could compare these models to managerial heuristics to determine their suitability as managerial tools.

The purpose of the study has limited the study of drivers to the FFP. Considering that FFPs is only one type of marketing investments it can be of interest to put the FFP drivers of CLV in relation to other marketing drivers such as brand and pricing mentioned in the literature in order to better maximize marketing return on investment (Blattberg, Malthouse and Neslin, 2009).
References

The Harvard reference style by Anglia Ruskin University has been used for this thesis.


import numpy as np
from numpy import log
import matplotlib
import pandas
import math
from lifetimes import BetaGeoFitter, ParetoNBDFitter, GammaGammaFitter, ModifiedBetaGeoFitter
import matplotlib.pyplot
from sklearn.metrics import mean_squared_error
from lifetimes.utils import summary_data_from_transaction_data, calibration_and_holdout_data, _customer_lifetime_value
from lifetimes.plotting import plot_cumulative_transactions, plot_period_transactions, plot_calibration_purchases_vs_holdout_purchases
pandas.set_option('expand_frame_repr', False)

dataset = pandas.read_csv("C:\Users\TobiasZenBook\Documents\exjobb\transactions2014nov3year_wcost_agg_4week_lp_discounted.csv", sep=';')
dataset.columns = ['id', 'date', 'monetary_value']

# 1 year calibration 2 year holdout
cal_hold = calibration_and_holdout_data(dataset, 'id', 'date', monetary_value_col = 'monetary_value', calibration_period_end = '2015-10-31', observation_period_end = '2017-10-30', freq='W')
t = 52*2

## GammaGamma fitter
returning_cal_hold = cal_hold[cal_hold['frequency_cal']>0]
returning_cal_hold = returning_cal_hold[returning_cal_hold['monetary_value_cal']>0]
returning_cal_hold[['monetary_value_cal', 'frequency_cal']].corr()

ggf = GammaGammaFitter(penalizer_coef = 0)

ggf.fit(returning_cal_hold['frequency_cal'],
        returning_cal_hold['monetary_value_cal'])
cal_hold['monetary_value_GG'] = 
ggf.conditional_expected_average_profit(cal_hold['frequency_cal'],
   cal_hold['monetary_value_cal'])

heuristic = dataset.loc[(dataset['date'] <= '2015-10-31')]
heuristic.groupby(['id']).agg({'monetary_value': 'sum'})
cal_hold['CLV_heur'] = 
heuristic.groupby(['id']).agg({'monetary_value': 'sum'})/((1+0.13)**1) +
heuristic.groupby(['id']).agg({'monetary_value': 'sum'})/((1+0.13)**2)

cal_hold['CLV_hold'] =
cal_hold['monetary_value_holdout'] * cal_hold['frequency_holdout']

/*
# Pareto/NBD fitter
pnbd = ParetoNBDFitter(penalizer_coef = 0.025)
pnbd.fit(cal_hold['frequency_cal'], cal_hold['recency_cal'],
cal_hold['T_cal'], iterative_fitting=2)

# år 2 discountrated
year2_discounted =
(pnbd.conditional_expected_number_of_purchases_up_to_time(52*2,
cal_hold['frequency_cal'], cal_hold['recency_cal'], cal_hold['T_cal'])-
 pnbd.conditional_expected_number_of_purchases_up_to_time(52,
cal_hold['frequency_cal'], cal_hold['recency_cal'],
cal_hold['T_cal']))/((1+0.13)**2)

# år 1 discountrated
year1_discounted =
(pnbd.conditional_expected_number_of_purchases_up_to_time(52,
cal_hold['frequency_cal'], cal_hold['recency_cal'],
cal_hold['T_cal']))/((1+0.13))

# Add them together
   cal_hold['condexp_pnbd'] = year2_discounted+year1_discounted

   cal_hold['CLV_pnbd'] =
cal_hold['monetary_value_GG'] * cal_hold['condexp_pnbd']
## testa olika error measures

### MAE i percent av avg clv
MAE_pnbd = abs(cal_hold['CLV_hold']-cal_hold['CLV_pnbd']).mean()/cal_hold['CLV_hold'].mean()*100
MAE_heur = abs(cal_hold['CLV_hold']-cal_hold['CLV_heur']).mean()/cal_hold['CLV_hold'].mean()*100

### Mean RMSE
Mean_RMSE_pnbd = np.sqrt(mean_squared_error(cal_hold['CLV_hold'], cal_hold['CLV_pnbd']))/cal_hold['CLV_hold'].mean()*100
Mean_RMSE_heur = np.sqrt(mean_squared_error(cal_hold['CLV_hold'], cal_hold['CLV_heur']))/cal_hold['CLV_hold'].mean()*100

### Median RMSE
Median_RMSE_pnbd = np.sqrt(((cal_hold['CLV_pnbd'] - cal_hold['CLV_hold'])**2).median())/cal_hold['CLV_hold'].mean()*100
Median_RMSE_heur = np.sqrt(((cal_hold['CLV_heur'] - cal_hold['CLV_hold'])**2).median())/cal_hold['CLV_hold'].mean()*100

### MSLE
MSLE_pnbd = np.sqrt((np.log((cal_hold['CLV_hold']+1)/(cal_hold['CLV_pnbd']+1)))**2).sum() / cal_hold['CLV_pnbd'].count())
MSLE_heur = np.sqrt((np.log((cal_hold['CLV_hold']+1)/(cal_hold['CLV_heur']+1)))**2).mean())

### sum av clv
clv_sum_pnbd = cal_hold['CLV_pnbd'].sum()/cal_hold['CLV_hold'].sum()*100
clv_sum_heur = cal_hold['CLV_heur'].sum()/cal_hold['CLV_hold'].sum()*100

### Sensitivity
### top 10 %
# actuals
top_10_act = math.floor(cal_hold['CLV_hold'].count()/10)
top_10cust_act = cal_hold['CLV_hold'].sort_values(ascending = False).head(top_10_act)
top_10cust_act = top_10cust_act.to_frame()
# heuristic
top_10_heur = math.floor(cal_hold['CLV_heur'].count()/10)
top_10cust_heur = cal_hold['CLV_heur'].sort_values(ascending = False).head(top_10_heur)
top_10cust_heur = top_10cust_heur.to_frame()
# Pareto/NBD

```
top_10_pnbd = math.floor(cal_hold['CLV_pnbd'].count() / 10)
top_10cust_pnbd = cal_hold['CLV_pnbd'].sort_values(ascending=False).head(top_10_pnbd)
top_10cust_pnbd = top_10cust_pnbd.to_frame();
```

# heuristic result
```
test1 = pandas.merge(top_10cust_act, top_10cust_heur, left_index=True, right_index=True, how='left')
sensitivity10_heur =
test1['CLV_heur'].notnull().sum() / test1['CLV_hold'].count() * 100
```

# Pareto/NBD result
```
test2 = pandas.merge(top_10cust_act, top_10cust_pnbd, left_index=True, right_index=True, how='left')
sensitivity10_pnbd =
test2['CLV_pnbd'].notnull().sum() / test2['CLV_hold'].count() * 100
```

## top 20%

# actuals
```
top_20_act = math.floor(cal_hold['CLV_hold'].count() / 5)
top_20cust_act = cal_hold['CLV_hold'].sort_values(ascending=False).head(top_20_act)
top_20cust_act = top_20cust_act.to_frame()
```

# heuristic
```
top_20_heur = math.floor(cal_hold['CLV_heur'].count() / 5)
top_20cust_heur = cal_hold['CLV_heur'].sort_values(ascending=False).head(top_20_heur)
top_20cust_heur = top_20cust_heur.to_frame();
```

# Pareto/NBD
```
top_20_pnbd = math.floor(cal_hold['CLV_pnbd'].count() / 5)
top_20cust_pnbd = cal_hold['CLV_pnbd'].sort_values(ascending=False).head(top_20_pnbd)
top_20cust_pnbd = top_20cust_pnbd.to_frame();
```

# heuristic result
```
test3 = pandas.merge(top_20cust_act, top_20cust_heur, left_index=True, right_index=True, how='left')
sensitivity20_heur =
test3['CLV_heur'].notnull().sum() / test3['CLV_hold'].count() * 100
```

# Pareto/NBD result
```
test4 = pandas.merge(top_20cust_act, top_20cust_pnbd, left_index=True, right_index=True, how='left')
sensitivity20_pnbd =
test4['CLV_pnbd'].notnull().sum() / test4['CLV_hold'].count() * 100
```
```python
import pandas as pd

# Define d as a dictionary
d = {'Model': ['pnbd', 'heuristic'],
     'MAE': [MAE_pnbd, MAE_heur],
     'MSLE': [MSLE_pnbd, MSLE_heur],
     'Mean RMSE': [Mean_RMSE_pnbd, Mean_RMSE_heur],
     'Median RMSE': [Median_RMSE_pnbd, Median_RMSE_heur],
     'Sensitivity_10%': [sensitivity10_pnbd, sensitivity10_heur],
     'Sensitivity_20%': [sensitivity20_pnbd, sensitivity20_heur],
     'Sum_CLV': [clv_sum_pnbd, clv_sum_heur]}

df = pd.DataFrame(data=d)
df = df.set_index('Model')

data = pd.read_csv('C:\Users\TobiasZenBook\Documents\exjobb\regression2017apr2.csv', sep=';')
regression.columns = ['MBR_ID', 'Tierlevel', 'Redemption', 'Promotions',
                      'UniquePartners', 'Downgraded_TierLevel', 'Upgraded_Tier_level']
cal_hold['recency2'] = cal_hold['T'] - cal_hold['recency']
dataset = dataset.set_index('id')
cal_hold_LP = pd.merge(cal_hold, regression, left_index=True, right_index=True)
cal_hold_LP['Tierlevel'] = cal_hold_eb['Tierlevel'].fillna('1')

avgclv = cal_hold_eb.groupby(['Tierlevel']).agg({'CLV_pnbd': 'mean', 'Tierlevel': 'count'}).sort_values(by=['CLV_pnbd'])
```

Appendix 2. Regression Analysis

**Table 2-1. Descriptive statistics.**

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLVpnbd_Ln1</td>
<td>9.5267</td>
<td>1.11040</td>
<td>369332</td>
</tr>
<tr>
<td>DWGT</td>
<td>.05</td>
<td>.224</td>
<td>369332</td>
</tr>
<tr>
<td>UPGT</td>
<td>.23</td>
<td>.418</td>
<td>369332</td>
</tr>
<tr>
<td>RDMP</td>
<td>.2900</td>
<td>.45377</td>
<td>369332</td>
</tr>
<tr>
<td>PRMT</td>
<td>.0474</td>
<td>.21243</td>
<td>369332</td>
</tr>
<tr>
<td>UNQP</td>
<td>.6412</td>
<td>.47964</td>
<td>369332</td>
</tr>
<tr>
<td>REC</td>
<td>17.62</td>
<td>18.360</td>
<td>369332</td>
</tr>
<tr>
<td>FREQ</td>
<td>5.63</td>
<td>6.582</td>
<td>369332</td>
</tr>
</tbody>
</table>

**Table 2-2. Correlations.**

<table>
<thead>
<tr>
<th></th>
<th>CLVpnbd_Ln1</th>
<th>DWGT</th>
<th>UPGT</th>
<th>RDMP</th>
<th>PRMT</th>
<th>UNQP</th>
<th>REC</th>
<th>FREQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pearson Correlation</td>
<td>.1000</td>
<td>.020</td>
<td>.261</td>
<td>.391</td>
<td>.164</td>
<td>.399</td>
<td>-.793</td>
<td>.820</td>
</tr>
<tr>
<td>Sig. (1-tailed)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CLVpnbd_Ln1</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
</tr>
<tr>
<td>DWGT</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
</tr>
<tr>
<td>UPGT</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
</tr>
<tr>
<td>RDMP</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
</tr>
<tr>
<td>PRMT</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
</tr>
<tr>
<td>UNQP</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
</tr>
<tr>
<td>REC</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
</tr>
<tr>
<td>FREQ</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>N</th>
<th>CLVpnbd_Ln1</th>
<th>DWGT</th>
<th>UPGT</th>
<th>RDMP</th>
<th>PRMT</th>
<th>UNQP</th>
<th>REC</th>
<th>FREQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLVpnbd_Ln1</td>
<td>369332</td>
<td>369332</td>
<td>369332</td>
<td>369332</td>
<td>369332</td>
<td>369332</td>
<td>369332</td>
<td>369332</td>
</tr>
<tr>
<td>DWGT</td>
<td>369332</td>
<td>369332</td>
<td>369332</td>
<td>369332</td>
<td>369332</td>
<td>369332</td>
<td>369332</td>
<td>369332</td>
</tr>
<tr>
<td>UPGT</td>
<td>369332</td>
<td>369332</td>
<td>369332</td>
<td>369332</td>
<td>369332</td>
<td>369332</td>
<td>369332</td>
<td>369332</td>
</tr>
</tbody>
</table>
### Table 2-3. Model summary.

<table>
<thead>
<tr>
<th>Model</th>
<th>R</th>
<th>R Square</th>
<th>Adjusted R Square</th>
<th>Std. Error of the Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.917a</td>
<td>.840</td>
<td>.840</td>
<td>.44361</td>
</tr>
</tbody>
</table>

a. Predictors: (Constant), FREQ, DWGT, PRMT, UPGT, UNQP, RDMP, REC
b. Dependent Variable: CLVpnbd_Ln1

### Table 2-4. Coefficients.

<table>
<thead>
<tr>
<th>Model</th>
<th>Unstandardized Coefficients</th>
<th>Standardized Coefficients</th>
<th>95.0% Confidence Interval for B</th>
<th>Collinearity Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>Std. Error</td>
<td>Beta</td>
<td>Lower Bound</td>
</tr>
<tr>
<td>1</td>
<td>(Constant)</td>
<td>9.368</td>
<td>.002</td>
<td>4788.838</td>
</tr>
<tr>
<td></td>
<td>DWGT</td>
<td>.052</td>
<td>.003</td>
<td>.010</td>
</tr>
<tr>
<td></td>
<td>UPGT</td>
<td>.083</td>
<td>.002</td>
<td>.031</td>
</tr>
<tr>
<td></td>
<td>RDMP</td>
<td>.129</td>
<td>.002</td>
<td>.053</td>
</tr>
<tr>
<td></td>
<td>PRMT</td>
<td>-.033</td>
<td>.004</td>
<td>-.006</td>
</tr>
<tr>
<td></td>
<td>UNQP</td>
<td>.168</td>
<td>.002</td>
<td>.073</td>
</tr>
<tr>
<td></td>
<td>REC</td>
<td>-.028</td>
<td>.000</td>
<td>-.460</td>
</tr>
<tr>
<td></td>
<td>FREQ</td>
<td>.086</td>
<td>.000</td>
<td>.510</td>
</tr>
</tbody>
</table>

a. Dependent Variable: CLVpnbd_Ln1

### Table 2-5. Residual statistics.

<table>
<thead>
<tr>
<th>Residuals Statisticsa</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predicted Value</td>
<td>7.9203</td>
<td>13.9681</td>
<td>9.5267</td>
<td>1.01794</td>
<td>369332</td>
</tr>
<tr>
<td>Residual</td>
<td>-8.14469</td>
<td>2.63190</td>
<td>.00000</td>
<td>.44361</td>
<td>369332</td>
</tr>
<tr>
<td>Std. Predicted Value</td>
<td>-1.578</td>
<td>4.363</td>
<td>.00000</td>
<td>1.000</td>
<td>369332</td>
</tr>
<tr>
<td>Std. Residual</td>
<td>-18.360</td>
<td>5.933</td>
<td>.00000</td>
<td>1.000</td>
<td>369332</td>
</tr>
</tbody>
</table>

a. Dependent Variable: CLVpnbd_Ln1
Figure 2-1. Histogram of residuals.

Figure 2-2. Normal P-P Plot of regression standardized residual.
Figure 2-3. Scatterplot of residuals.