

Portfolio, Churn & Customer Value

Hugo Cornet, Pierre-Emmanuel Diot, Guillaume Le Halper, Djawed Mancer

2021-12-09

Contents

Abstract	5
1 Introduction	7
1.1 Context	7
1.2 A multi-purpose study	7
2 Definitions	9
2.1 How to define a customer portfolio?	9
2.2 What is attrition?	11
2.3 What does customer “value” mean?	11
3 Literature Review	13
3.1 On customer portfolio	13
3.2 On attrition	13
3.3 On customer value	13
4 Theoretical Framework	15
4.1 Duration models	15
4.2 Machine Learning techniques	21
5 Data	23
Appendix	25
Duration models	25

Abstract

This paper is being realized as part of our last year in master's degree in economics. It aims at studying a firm's most valuable asset namely its customers. To that end, we adopt a quantitative approach based on econometrics and data analysis with a threefold purpose to :

- model a *customer portfolio* as a set of customer segments;
- predict and analyse customer *churn*;
- estimate a customer portfolio's overall *value*.

After having defined the subject's key concepts, we apply duration models and machine learning techniques to a kaggle dataset related to customers of a fictional telecommunications service provider (TSP).

Keywords: *customer portfolio management (CPM), churn, customer value, duration models, machine learning, telecom.*

Chapter 1

Introduction

1.1 Context

1.2 A multi-purpose study

Chapter 2

Definitions

Before introducing the literature review on the topic and the quantitative methods implemented, it appears relevant to define each of the three key concepts of the study. Having a clear understanding of the notions of *portfolio*, *attrition* and *customer value* is essential to then choose proper methodologies and deliver insights from the data.

2.1 How to define a customer portfolio?

The notion of *portfolio* has greatly evolved before a firm's consumer base was considered as a set of portfolios. In chapter 3 a part of the literature review depicts an evolution of the *portfolio management* notion. A customer *portfolio* can be defined as a set of customers divided into several segments (or clusters) based on similar attributes. These discriminant features can be both economic - willingness to pay, budget constraint, etc. - and sociological - gender, age, socio-professional category, etc. . The underlying objective of this segmentation is to optimally allocate the company's resources. When dealing with *customer portfolio management* (CPM), two dimensions can be considered. On the one hand, it can be assumed that a customer stays in the same segment throughout their life in the firm's *portfolio*. On the other hand, a dynamic approach can be adopted as suggested by Homburg on dynamics in customer portfolios (Homburg et al., 2009). In their article, the authors question the static analysis by assuming that a customer can switch between segments. They explain that one of the firm's objectives is to convert less valuable customers into more valuable ones. In this context, customers can also be segmented depending on their *value* to the firm and their cost to serve (Thakur and Workmanb, 2016), as depicted by figure 2.1.

Value to the Company	High	Superior Service (Platinum Customer)	Best Service (Gold Customer)
	Low	Good Service (Silver Customer)	Better Service (Bronze Customer)
Cost to serve: Relative service level for optimal resource deployment			

Figure 2.1: CPM Matrix

2.2 What is attrition?

Customer *attrition* or churn occurs when a client discontinues using a service or a product offered by a firm. Churn analysis corresponds to measuring the *attrition* rate in the customer base of any company. However, evaluating *attrition* depends on the type of relationship between the firm and its clients. When it is defined by a contract, the customer has to inform the firm about their service termination. In the telecom industry, a consumer is required to notify their telecommunications service provider (TSP) before going to a competing company. In an opposite way, the firm/client relationship can be non-contractual. In that case, the service termination does not need to be notified. *Attrition* then becomes a latent variable and more advanced models are used to make forecasts.

2.3 What does customer “value” mean?

In *customer portfolio management*, one client’s value is represented by the **customer lifetime value** (CLV). CLV is the present *value* of all future purchases made by a customer over their lifetime in the firm’s portfolio, taking into account the *attrition* risk. CLV depends both on the purchase recency as well as on the purchasing rate and aims at identifying the most valuable customer groups. Formally, Gupta and Lehmann (Gupta and Lehmann, 2003) define CLV for customer i as follows:

$$CLV_i = \sum_{t=0}^{T_i} \frac{(p_t - c_t)r_{i,t}}{(1 + a)^t} - AC_i$$

with,

- p_t the price paid by customer i at time t
- c_t the cost at time t
- $r_{i,t}$ the **probability that customer i be active** at time t
- a the discount rate
- AC_i the acquisition cost of customer i
- T_i the duration of observation

An estimation of the portfolio’s overall value can be calculated through **customer equity** (CE) which amounts to the sum of all the CLVs. Since CE appears to be a good proxy of the firm’s value, the profit maximization pro-

gram can be rewritten as:

$$\begin{aligned} \max_p \quad & \text{CE} = \sum_{i=1}^N \sum_{t=0}^{T_i} \frac{(p_t - c_t)r_{i,t}}{(1+a)^t} - AC_i \\ \text{s.t.} \quad & r_{i,t} \in [0, 1] \\ & p_t > c_t \end{aligned}$$

Chapter 3

Literature Review

3.1 On customer portfolio

3.2 On attrition

3.3 On customer value

In recent years, *customer portfolio management* (CPM) has focused on optimizing clients' value to the firm. The company's interest lies in knowing how much net benefit it can expect from a customer *today*. These expectations are then used to implement efficient marketing strategies to get the highest return on investment. To that end, two key metrics are estimated by firms, namely customer lifetime value (CLV) and customer equity (CE) (see chapter 2 for definitions).

According to Blattberg and Deighton, CLV can be defined as the revenues derived from a customer minus the cost to the firm for maintaining the relationship with the customer over time (Blattberg and Deighton, 1996). As shown by Reinartz and Kumar, CLV modelling depends on the type of relationship a firm has with its clients (Reinartz and Kumar., 2003). In a contractual relationship, customer defections are observed which means that longer lifetime means higher customer value. Conversely, when the relationship is non-contractual, uncertainty arises between the customer's purchase behavior and lifetime.

With the development of data collection tools, companies have lots of customer-level data (or customer transaction data) at their disposal to measure CLV (Fader and al., 2005). Consequently, different modelling approaches can be adopted in order to estimate customers' value.

RFM (Recency Frequency Monetary) models are considered the simplest strategy to increase CLV and customer's loyalty (Gupta et al., 2006). It aims at targeting specific marketing campaigns at specific groups of customers to improve response rates. RFM models consists in creating clusters of clients based on three variables:

- *recency* which is the time that has elapsed since a customer's last activity with the firm;
- *frequency* that is the number of times a customer transacted with the brand in a particular period of time;
- *monetary* value of customer's prior purchases.

However, RFM models have a limited predictive power since they only predict clients' behavior for the next period.

In their article on CLV management, Borle and Singh draw the review of more advanced modelling techniques that can be implemented to estimate customers' value (Borle and Singh, 2008). A popular method to estimate customer lifetime value is the negative binomial distribution (NBD) - Pareto (Fader and al., 2005) which helps solving the lifetime uncertainty issue. The model takes past customer purchase behavior as input such as the number of purchases in a specific time window and the date of last transaction. Then the model outputs a repurchase probability as well as a transaction forecast for each individual. In Borle and Singh's research paper, a hierarchical bayesian model is implemented with a view to jointly predict customer's churn risk and spending pattern. Here, the main advantage of using a bayesian approach is to give priors on CLV's drivers. The study is base on data coming from a membership-based direct marketing company where firm/client relationships are non-contractual. In other words, the times of each customer joining the membership and terminating it are known once these events happen. Thus the implementation of a sophisticated estimation strategy is justified.

In our study, emphasize is placed on estimating the overall value of a customer portfolio. The methodology we will develop is based on a research paper written by our Econometrics teacher Alain Bousquet, whose goal is provide tools for an efficient management of a patent portfolio (Bousquet, 2021). The main idea is to consider each patent as an asset with a related value which can generate income if this very patent is exploited. This modelling approach can be transposed to customer portfolio analysis with the customer's value corresponding to the CLV and the probability of exploitation being the opposite of the risk of *attrition*. In this context, CLV can be estimated with techniques mentioned above. The client's risk of churn can be modelled with duration models or machine learning techniques as evoked in 3.2.

Chapter 4

Theoretical Framework

This chapter presents theoretical basis of the models and the algorithms that are used to model customer portfolios. As a customer's lifetime in a portfolio is usually represented by the time to churn, duration models are adapted to the data we have at our disposal. Thus, it is decided to introduce standard survival techniques as well as machine learning algorithms suited to time-to-event data. Methods to estimate customer's value are also introduced in this chapter.

4.1 Duration models

According to Cameron & Trivedi (Cameron and Trivedi, 2005), duration models (also called survival models) aims at measuring the time spent in a certain state before transitioning to another state. In Econometrics,

- a **state** corresponds to the class in which an individual i is at time t .
- a **transition** is movement from one state to another.
- a **duration** measures the time spent in a certain state and is also called a **spell** length.

Since measuring the time until the event is needed for multiple purposes, duration analysis is used in a variety of economic sectors as depicted by the following table.

Economic sector	Purpose
Macroeconomics	Length of an unemployment spell
Insurance	Risk analysis to offer a segmented pricing
Engineering	Time until a device breaks down
Epidemiology	Survival time of a virus

Economic sector	Purpose
Churn analysis	Time until a customer leave the portfolio

4.1.1 Censoring and Truncation

When dealing with survival data, some observations are usually **censored** meaning they are related to spells which are not completely observed. Duration data can also suffer from a selection bias which is called **truncation**.

Censoring mechanisms

Left-censoring occurs when the event of interest occurs before the beginning of the observation period. For example, an individual is included in a study of duration of unemployment at t_0 . At that time he has already been unemployed for a period but he cannot recall exactly the duration of this period. If we observe that he finds a job again at t_1 , we can only deduce that the duration of unemployment is bigger than $t_1 - t_0$, this individual is left-censored. Observation 2 on figure 4.1 is associated with a left-censored spell (Liu, 2019).

A spell is considered **right-censored** when it is observed from time t_0 until a censoring time t_c as illustrated by observation 1 on figure 4.1. For instance, the lifetime related to a customer who has not churned at the end of the observation period is right-censored. Let us note X_i the duration of a complete spell and C_i the duration of a right-censored spell. We also note T_i the duration actually observed and δ_i the censoring indicator such that $\delta_i = 0$ if the spell is censored. Then, $(t_1, \delta_1), \dots, (t_N, \delta_N)$ are the realizations of the following random variables:

$$\begin{aligned} T_i &= \min(X_i, C_i) \\ \delta_i &= \mathbb{1}_{X_i < C_i} \end{aligned} \tag{4.1}$$

Selection bias

Survival data suffers from a **selection bias** (or truncation) when only a sub-sample of the population of interest is studied. A customer entering the firm's portfolio after the end of the study is said to be **right-truncated**, whereas a client who has left the portfolio before the beginning of the study is considered **left-truncated**. Mathematically, a random variable X is truncated by a subset $A \in \mathbb{R}^+$ if instead of $\Omega(X)$, we solely observe $\Omega(X) \cap A$. On figure 4.1, the first and fifth observations suffers from a selection bias.

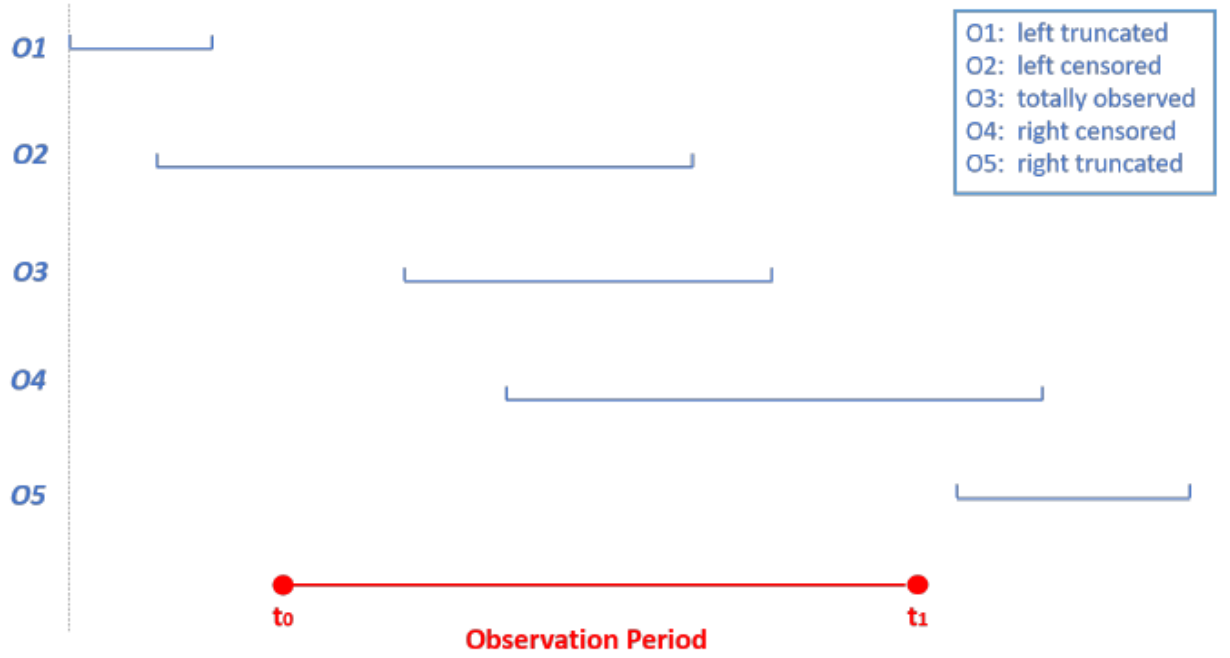


Figure 4.1: Censored and truncated data

4.1.2 Probabilistic concepts

In survival analysis, the response variable denoted T is a time-to-event variable. Instead of estimating the expected failure time, survival models estimate the *survival* and *hazard rate* functions which depend on the realization of T .

Survival function

The survival function $S(t)$ represents the probability that the considered event occurs after time t . For instance, $S(t)$ can measure the probability that a given customer survive in the portfolio at least until time t . Mathematically, the survival function is defined as:

$$S(t) = P(T > t) = 1 - F(t) \quad (4.2)$$

where $F(t)$ is the cumulative distribution function.

Hazard and Cumulative Hazard functions

Another key concept in duration analysis is the hazard function $\lambda(t)$ which approximates the probability that the event occurs at time t . For instance, $\lambda(t)$

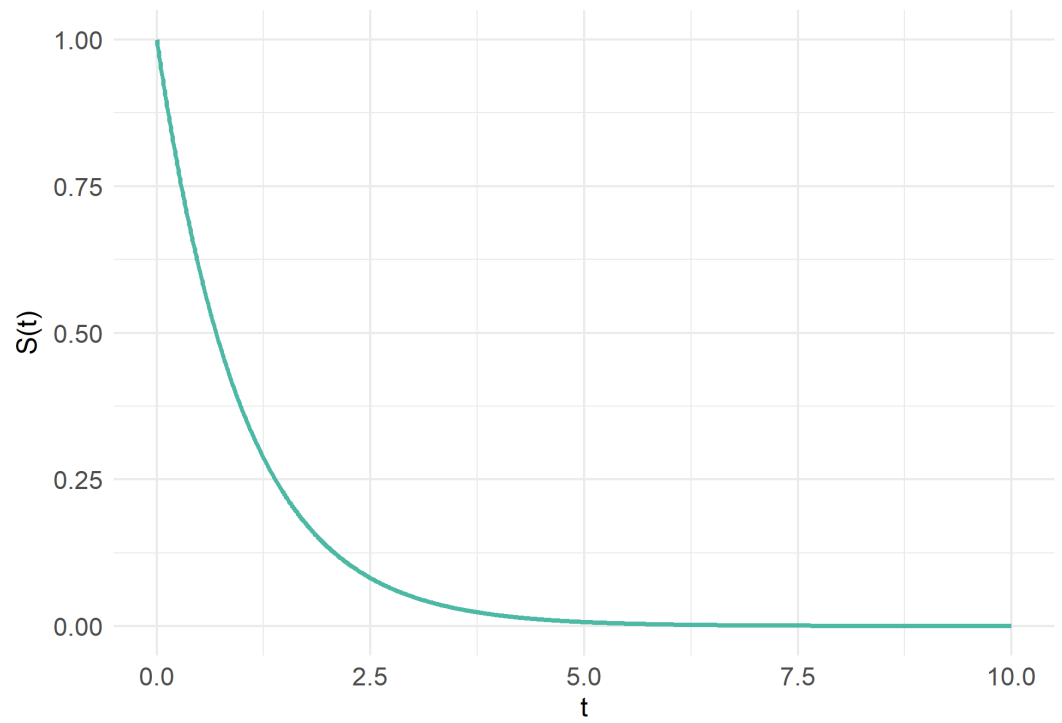


Figure 4.2: Survival function $S_T(t)$ with $T \sim \mathcal{E}(1)$

can measure the probability that a given individual leaves the firm portfolio at time t . Formally, it is expressed as follows:

$$\lambda(t) = \lim_{\Delta t \rightarrow 0} \frac{P[t \leq T < t + \Delta t | T \geq t]}{\Delta t} \quad (4.3)$$

Using the Bayes formula, equation (4.3) can also be written as (see equation (5.1) in appendix):

$$\lambda(t) = \frac{-d \ln(S(t))}{dt} \quad (4.4)$$

Finally, integrating the instantaneous hazard function gives the cumulative hazard function which can be more precisely estimated than the hazard function (Cameron and Trivedi, 2005) and is defined as:

$$\Lambda(t) = \int_0^t \lambda(s) ds = -\ln(S(t)) \quad (4.5)$$

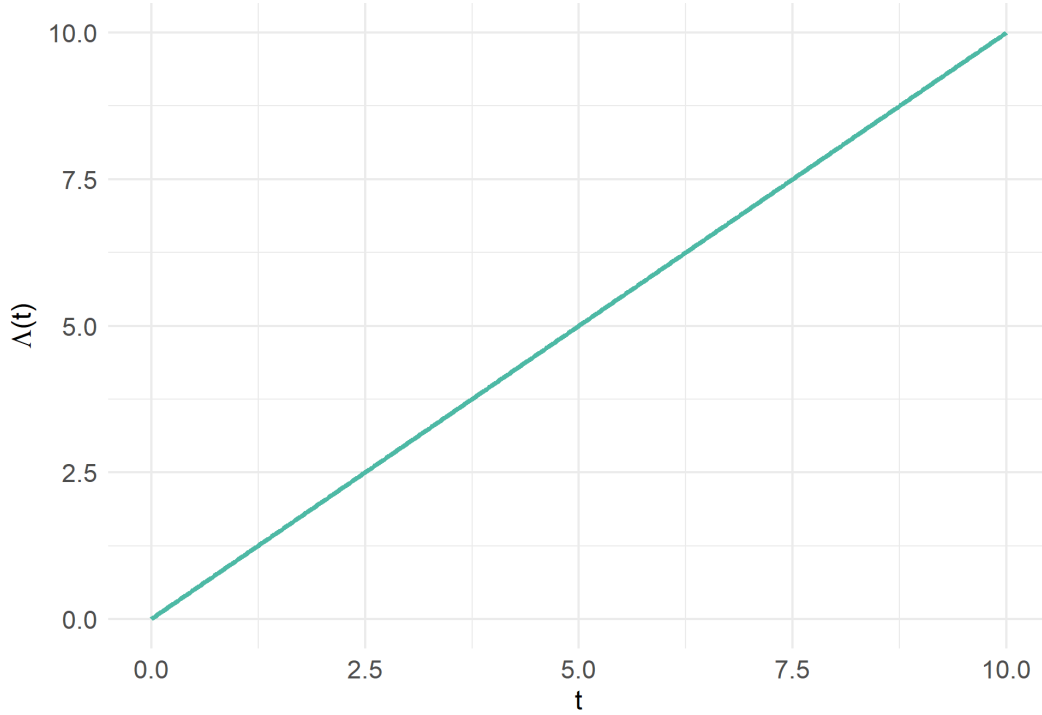


Figure 4.3: Cumulative Hazard function $\Lambda_T(t)$ with $T \sim \mathcal{E}(1)$

Thus, the hazard, survival and cumulative hazard functions are three mathematical functions which describe the same distribution.

4.1.3 Nonparametric models

When dealing with duration data, these methods are helpful to have a general overview of the raw (unconditional) hazard. Nonparametric models are rather used for data description rather than prediction. No explanatory variable is included in these models except for treatment variables such as the type of contract a customer has subscribe.

Notations

Let us consider a sample with N observations with k discrete failure time (e.g. a failure can be a churn event), such that $\forall j \in \llbracket 1; k \rrbracket$:

- t_j the j^{th} discrete failure time,
- d_j the number of spells terminating at t_j ,
- m_j the number of right-censored spells in the interval $[t_j, t_{j+1}]$,
- r_j the number of exposed durations right before time t_j such that:

$$r_j = (d_j + m_j) + \dots + (d_k + m_k) = \sum_{l|l \geq j} (d_l + m_l) \quad (4.6)$$

Hazard function estimator

As the instantaneous hazard at time t_j is defined as $\lambda_j = P[T = t_j | T \geq t_j]$, a trivial estimator of λ_j is obtained by dividing the number of durations for which the event is realized in t_j by the total number of exposed durations at time t_j^- . Formally, it is expressed as:

$$\hat{\lambda}_j = \frac{d_j}{r_j} \quad (4.7)$$

Kaplan-Meier estimator

Once the hazard function estimator computed, the discrete-time survivor function can be estimated using the Kaplan-Meier product-limit estimator. To estimate the survival at time t , this estimator computes the joint probability that a spell stays in the same state until t (e.g. remaining loyal to a firm until a certain time). This method is based on conditional probabilities and the survival function estimate is defined as:

$$\hat{S}(t) = \prod_{j|t_j \leq t} (1 - \hat{\lambda}_j) = \prod_{j|t_j \leq t} \frac{r_j - d_j}{r_j} \quad (4.8)$$

When plotting the survival curve after having performed the Kaplan-Meier estimation, confidence bands are also added to the plot in order to reflect sampling variability (Cameron and Trivedi, 2005). The confidence interval of the survival function $S(t)$ is derived from the estimate of the variance of $S(t)$ which is obtained by the Greenwood estimate (see equation (4.9)).

$$\hat{V}[\hat{S}(t)] = \hat{S}(t)^2 \sum_{j|t_j \leq t} \frac{d_j}{r_j(r_j - d_j)} \quad (4.9)$$

Nelson-Aalen estimator

The cumulative hazard function estimate is given by the Nelson-Aalen estimator which consists in summing up the hazard estimates for each failure time.

$$\hat{\Lambda}(t) = \sum_{j|t_j \leq t} \hat{\lambda}_j = \sum_{j|t_j \leq t} \frac{d_j}{r_j} \quad (4.10)$$

Exponentiating $\hat{\Lambda}(t)$, one can obtain a second estimate of the survival function (see equation (5.2) in the appendix):

$$\tilde{S}(t) = \exp(-\hat{\Lambda}(t)) \quad (4.11)$$

4.1.4 Parametric models

The nonparametric estimation is undoubtedly useful when it comes to have a general overview on the survival data. However, one may want to model the hazard and survivor functions with a functional form in which unknown parameters need to be optimized. In parametric estimation, there are two main groups of models to estimate the conditional hazard function $\lambda(t|\mathbf{x})$:

- the Parametric Proportional Hazards (AFT) models,
- the Accelerated Failure Time models (PH).

When implementing parametric models, λ , S and Λ are expressed based on the chosen parametric form. The instantaneous hazard function can either be constant or monotone.

4.2 Machine Learning techniques

Chapter 5

Data

Appendix

In this part, some proofs of the mathematical concepts used in the study are derived.

Duration models

Hazard function

$$\begin{aligned}\lambda(t) &= \lim_{\Delta t \rightarrow 0} \frac{P[t \leq T < t + \Delta t | T \geq t]}{\Delta t} \\ &= \lim_{\Delta t \rightarrow 0} \frac{P[t \leq T < t + \Delta t] / P[T \geq t]}{\Delta t} \\ &= \lim_{\Delta t \rightarrow 0} \frac{(F(t + \Delta t) - F(t)) / \Delta t}{S(t)} \\ &= \frac{dF(t)/dt}{S(t)} \\ &= \frac{f(t)}{S(t)} \\ \lambda(t) &= \frac{-d \ln(S(t))}{dt}\end{aligned}\tag{5.1}$$

Link between cumulative hazard and survivor functions

$$\Lambda(t) = \int_0^t \lambda(s) ds$$

$$\Leftrightarrow \Lambda(t) = \int_0^t \frac{f(s)}{S(s)} ds \quad (5.2)$$

$$\Leftrightarrow \Lambda(t) = -\ln(S(t))$$

$$\Leftrightarrow S(t) = \exp(-\Lambda(t))$$

Bibliography

- Blattberg and Deighton (1996). Manage marketing by the customer equity test. *Harvard Business Review*, pages 136–144.
- Borle, S. and Singh, S. S. (2008). Customer lifetime value measurement. *Management Science*, 54(1):110–112.
- Bousquet, A. (2021). Gestion optimale de portefeuilles de brevets.
- Cameron, A. C. and Trivedi, P. K. (2005). *Microeconometrics: Methods and Applications*.
- Fader and al. (2005). "counting your customers" the easy way: An alternative to the pareto/nbd. *Management Science*, 24(2):275–284.
- Gupta, Hanssens, and Hardie, D. (2006). Modeling customer lifetime value. *Journal of Service Research*, 9(2):139–155.
- Gupta and Lehmann, D. R. (2003). Customers as assets. *Journal of Interactive Marketing*, 17(1):9–24.
- Homburg, C., Steiner, V. V., and Totzek, D. (2009). Managing dynamics in a customer portfolio. *Journal of Marketing*, 73(5):70–89.
- Liu, W. (2019). *Inclusive Underwriting: the case of Cardiovascular Risk Calculator*. PhD thesis, ENSAE ParisTech.
- Reinartz and Kumar. (2003). The impact of customer relationship characteristics on profitable lifetime duration. *Journal of Marketing*, 67(1):77–99.
- Thakur, R. and Workmanb, L. (2016). Customer portfolio management (cpm) for improved customer relationship management (crm): Are your customers platinum, gold, silver, or bronze? *Journal of Business Research*, 69:4095–4102.