Alma Mater Studiorum – Università di Bologna

DOTTORATO DI RICERCA IN

GENERAL MANAGEMENT

Ciclo XXX

Settore Concorsuale: SECS-P/08

Settore Scientifico Disciplinare: Economia e Gestione delle Imprese

Exploring the Acquisition Process:

The Role of Marketing Activities and Searching Behavior in Driving New Users Toward Acquisition

Presentata da: Federica Vecchioni

Coordinatore Dottorato

Chiara Orsingher

Supervisori

Elisa Montaguti

Co-Supervisori

Sara Valentini

Fulvia Pennoni

Esame finale anno 2019

ABSTRACT

Firms are spending a lot of efforts and resources in acquiring new, and possibly valuable, customers. First purchase is often considered as the moment of acquisition and the most crucial point of the business relationship in the managerial practice. However, little is known about the process that leads to the acquisition of new customers, and when a customer can be considered fully acquired. The purpose of this research is twofold. First, we aim at analyzing the evolution of the acquisition process over time, investigating the role played by marketing activities and search behavior in turning prospects into valuable customers. Second, we intend to provide a better representation of the concept of customer acquisition, particularly in non-contractual settings. In order to achieve these goals, we estimate a Hidden Markov Model on a unique dataset, where both clickstream and transactional data of an e-tailer company are merged to provide a complete view on prospects and customers' behaviors. We find that: (1) potential new customers pass through up to five hidden states that characterize the acquisition process. While in these stages, individuals can be described as *directed searchers*, *deliberative searchers*, *first triers*, *acquired customers*, and *loyal customers*; (2) the road to acquisition is heterogeneous as different migration paths across states exist; (3) both marketing activities and pre-acquisition behavior have an effect in driving prospects towards acquisition, and this effect changes according to the state in which the individual is; (4) the first purchase does not necessarily entails the acquisition of the customer: it could be just a tentative approach to start a relationship with the firm.

TABLE OF CONTENTS:

1. INTRODUCTION

"You know that the beginning is the most important part of any work".

(Plato - The Republic, 380 B.C.)

As Plato recognized, the beginning is a fundamental step in order to build anything, including relationships.

Firms are very familiar with this problem, as they strive to find new valuable customers to acquire and with whom starting a relationship. Thus, it is clear that the acquisition of new customers is a fundamental phase of the Customer Relationship Management (CRM) of any company because, before maintaining a relationship with its customers, a firm first needs to identify customers who are not only willing to be acquired but who are also worth acquiring. In fact, a large proportion of prospects who are more likely to be acquired by companies turns out to be *bad customers*, as the additional value they bring to the firm is not worthy the acquisition costs the firm has to face in order to acquire them (Cao and Gruca, 2004).

Despite its importance, both academic and managerial literature on customer acquisition is scarce. We don't know much, for example, about the factors that are more likely to increase the probability to acquire new customers, driving them towards their first purchase, which kinds of marketing strategies are more effective, or how to target future customers on the basis of their pre-acquisition behaviors. This is mainly due to a chronic scarcity and difficulty of gathering data concerning the pre-acquisition stage (Min, Zhang, Kim, and Srivastava 2016; Tillmanns, Hofstede, Krafft, and Goetz, 2017), combined with firms' inability to identify the first customer's transaction, especially when such purchase takes place in a physical store. In fact, it is both hard – in most cases, almost impossible - and costly for firms to link a transaction occurring in a brick-and-mortar store to a customer record, and this is possible only when the customer owns a loyalty card. Nowadays, thanks to advances in technology, most firms seat on an extensive amount of data that were previously both unavailable and inaccessible. Most digital firms are trying to develop tools which allow them to unequivocally i) identify each customer's first purchase and track his/her pre-acquisition behavior; ii) monitor most of the marketing communications a customer receives and responds to; iii) track a customer's searching activities, or the device she uses. The unique identification of the user is one of the major issues for digital firm, and its accomplishment is not a trivial task for them.

A significant part of previous literature on customer acquisition focuses on issues like budget allocation between acquisition and retention efforts (e.g. Blattberg and Deighton, 1996; Thomas, 2001; Reinartz, Thomas, and Kumar, 2005; Natter, Ozimec, and Kim, 2015), or the relationship between acquisition and retention (e.g. Schweidel et al., 2008; Villanueva et al., 2008; Lewis, 2006b), but only a few studies attempt to analyze the antecedents of customer acquisition (e.g. Trusov et al. 2009; Wangenheim and Bayon 2007; De Vries, Gensler, and Leeflang, 2017). More importantly, a significant part of this work takes a product-centric perspective, according to which a customer is acquired when she purchases for the first time a specific product or service they focus on, regardless whether she had already bought from the company previously (e.g. Schweidel et al., 2008; Schwartz, Bradlow, Fader, 2017). By contrast, under a customer-centric perspective, firms' strategies are built around an in-depth knowledge of the customer. This means that firms ground their actions on each customer's history and preferences in order to develop a customized marketing activity (Fader, 2012). Thus, under a customer-centric perspective, a customer is acquired at the time s/he purchases a brand for the first time (e.g. Wangenheim and Bayon, 2007).

There is general consensus in both the academic and the trade press that, shifting the focus from products to customers, benefits companies by letting them move from mass-marketing to individual marketing (Sheth et al., 2000), enhancing the effectiveness of firms' cross-selling strategies (Blattberg et al., 2008), increasing customer satisfaction (Blattberg et al., 2008), and competitive advantage (Day, 2000).

Since customer-centric firms need to have a perfect knowledge of the history and preferences of their customer base (Fader, 2012), they are able to follow each of their customers through their entire *buying process*, which is the decisional process that each person undergoes before making a purchase, composed by several states, that helps her in making her final choice (e.g. Lavidge and Steiner, 1961; Rogers, 1962; Roberts and Lattin, 1991; Bettman, Luce, and Payne, 1998; Court et al., 2009; Kireyev, Pauwels, and Gupta, 2016; Dierks, 2017). Since we are interested in the first purchase of potential customers, in this research we will consider the buying process as the *Customer Acquisition Process*.

Even though the multistage decision process concept has gained an enormous consensus into the marketing field, in customer acquisition this concept is still largely unexplored, as most of the studies in acquisition literature consider customer acquisition as a single-step process (e.g. Hansotia and Wang, 1997; Lewis, 2006a; Wangenheim and Bayon, 2007; Nam, et al., 2010). Analyzing it as a series of multiple stages allows marketers and advertisers to better understand the customers' mental scheme in decision making (Van Lennep, 2014), providing some insights about the reasons underlying the changes in customers' behavior over time (Dierks, 2017), and helping in developing more customer-centric shopping experiences (Faulds, 2018) by sending the right information to customers at the right moment of their decisional process (Court, 2009), making their marketing communications more effective (Lovejoy, 2014). For these reasons, it would be worthy to explore the role of the marketing activities and of the touchpoints between the firm and the customers in each of the stages of the Customer Acquisition Process, in order to provide some insights about the development of the optimal acquisition marketing strategy.

Lastly, customer acquisition has been investigated mainly in contractual settings (e.g. Wangenheim and Bayon, 2007; Nam, et al., 2010), whereby a customer is bounded to the firm for a fixed period of time after having signed a contract (e.g. banking account, insurance, etc.).

However, in non-contractual settings, where customers do not subscribe any contract with firms, customers could be making just a single first transaction (allegedly the acquisition) and disappear thereafter. In such case, it could be argued that the customer is not really acquired, since her relationship with the firm started and ended with the first purchase occasion. For this reason, we contend that in non-contractual setting, acquisition can be treated as an unobservable state that is the outcome of a set of activities including search, purchases and spending. This means that we propose to overcome the traditional view of a first purchase as a unique indicator of acquisition.

Thus, the goal of this research is to analyze the customer acquisition funnel answering the following three research questions:

(i) How does the customer acquisition process evolve over time?

(ii) How does the pre-acquisition behavior influence the acquisition process? What is the role played by the interactions between the potential customers and the different marketing activities in each stage of the acquisition funnel in driving prospects toward acquisition?

(iii) How can we define customer acquisition in a non-contractual setting?

In order to answer these three questions, an empirical study is presented. It employs data from an Italian e-retailer who sells clothing and design objects worldwide. The intent of the study is to identify and describe the stages prospects go through in their acquisition process, analyzing the impact that marketing activities and pre-acquisition behavior have on the probability to move forward in the funnel and on the probability to be acquired in each stage of the process. The particular setting of the study allows us to take a customer-centric perspective since we are able to exactly identify the first customers' purchase and their previous behavior. Moreover, the availability of data concerning the post-acquisition transactional behavior allows us to treat acquisition as a latent state in order to test the need for a different definition of acquisition in a non-contractual setting.

Results show that prospects actually pass through five hidden states in their acquisition process: *directed searchers*, who do not purchase, and make short, goal-oriented searching sessions; *deliberative searchers*, who still do not purchase, but are the ones who search the most; *first triers*, when they purchase for the first time with a limited monetary value, *acquired customers*, the state in which they purchase for the first or second time, but with more consistent expenditure level, and *loyal customers*, with the most intense purchase activity. We also found that users do actually shift from one state to another in order to be acquired or to become loyal, but the path to acquisition is not the same for everyone: while some users need to have some searching sessions before passing to one of the purchase states, others have processes characterized by purchase states only. Accordingly, we found that more than half of our users never get into a purchase state. Results also highlight the fact that deliberative searchers are an interesting state that the firm should carefully monitor, as it contains prospects who are more likely to become first triers and acquired.

Our findings suggest that both marketing activities and pre-acquisition behavior have an effect in driving prospects towards acquisition, and this effect changes according to the state in which the individual is. For example, overall, customer-initiated touchpoints perform better than firm-initiated touchpoints in enhancing users' likelihood to search and to purchase, especially if they are in the later stages of the acquisition process. Results also show that the act of registering to the website has an impact on the acquisition process, as it increases people's likelihood to purchase at the very beginning, and to move on in the process, especially shifting from non-purchase to purchase states. Users registering in the website using a social account (e.g. Google+) have a higher probability to be deliberative searchers and to remain in that state, but only if they do not use a Facebook account, which increases their likelihood to move to a purchase state. The use of the Wish List for directed searchers makes them moving on in the process, as it helps users search in a more organized and efficient way (Close and Kukar-Kinney, 2010), but the effect is reversed for deliberative searchers, decreasing their likelihood to become first triers. Regarding the search-related activities purchase, only sorting the products by ascending price, clicking on suggested products, and use filters increase the transactional outcomes, but not for loyal customers, as they are not price sensitive. We found also the device choice to play a role in the process. In fact, the mobile usage has been found to be positively related to search activities, but negatively to purchase activities, especially for customers who are purchasing for the first time. Desktop users perform worse than tablet users both in terms of search, purchase, and monetary value, highlighting the fact that, even if it is not yet very widespread, nowadays the use of tablet is gaining importance in the digital retailing environment.

Regarding the new definition of customer acquisition in non-contractual settings, our findings corroborate the idea that the traditional definition of customer acquisition does not hold when customers are not bounded to the firm by a contract. In fact, our model suggests that the first purchase could be just a tentative approach to start a relationship with the firm, especially when it is not a high value purchase. However, customers who make the first purchase with a higher expenditure can already be considered as acquired. In the same way, a customer who make a second purchase, but still with a limited expenditure, should not necessarily be considered as acquired yet.

The contribution of this research is threefold. First, from a theoretical standpoint, this research increases the little literature on customer acquisition by focusing on the customer acquisition process *per se*, by investigating the process leading prospects to be acquired by the company and the antecedent factors that affect this process. It is also the first study to formally consider the prospects' pre-acquisition activity (e.g. registration and search activities, last touch marketing activities). Moreover, this is the first study investigating the customer acquisition process in order to discover which are the stages preceding acquisition, and which paths to acquisition are more likely to be observed, accounting for the impact of prospects' pre-acquisition activity (e.g. registration and search activities, last touch marketing activities). This research also provides a new way to define acquired customers when they are not bounded by a contract, looking at it as a hidden state in the process rather than an observed behavior. Secondly, from a managerial point of view, this work is relevant as it provides some hints to design more effective, and customer-centric acquisition programs by considering the state in which prospects are in the acquisition process, suggesting a way for developing more customized acquisition programs which better meet the prospects' needs according to their state. Moreover, our results are useful in order to better target prospects who are more likely to be acquired and to become loyal. Finally, this work provides an empirical contribution to the acquisition literature, since it is one of the first attempts to analyze the customer acquisition process in a completely digital setting, whereby both the pre and the post-acquisition behaviors are monitored.

The present dissertation is structured as follows: after this first introductive chapter, the second provides a brief overview on the customer acquisition topic and describes the state of the art in the literature about customer acquisition. Then, the third chapter is dedicated to the theoretical development of the three research questions and the related hypotheses. The fourth chapter describes the Hidden Markov Model, which is the methodology employed in the empirical study. Chapter 5 starts with the presentation of the setting and a brief description of the raw data provided by the firm. Then it explains the process of data cleaning and merging of the different sources of information, and describes the final database, followed by the model estimation and the presentation of the results. The document ends with a discussion of the main findings and the expected theoretical, managerial, and empirical contribution of the research to the field, together with the limitations and some suggestions for future research.

2. Theoretical Background

The intent of the current chapter is to introduce the reader to the customer acquisition concept, and its development in the marketing literature. The first part of the chapter narrates an example story with the purpose of introducing the customer acquisition issue. They will help with the explanation and – hopefully – the understanding of the concepts described in the second part of the chapter, which provides an overview about the concept of customer acquisition as it is in today's marketing academic and managerial environments, illustrating its definition, its importance, and the main acquisition-related managerial tasks.

The third and last section of this chapter deals with the state of the art of the customer acquisition literature. It starts with the challenges that researchers have to face in studying acquisition, mainly due to the data availability. Related to this issue, the discussion will proceed dealing with the perspective through which acquisition has been analyzed. It will review the results of some studies dealing with the factors influencing customer acquisition, and the targeting of new potential customers, then there will be a brief discussion about the acquisition and retention relationship.

The end of the chapter is dedicated to a summary of the reviewed works, highlighting the gaps that this thesis intends to fill.

2.1. Prologue

Kate's Shopping

Kate is a young woman who has just moved in London. Every morning, in her way to the office, she passes in front of Marks & Spencer's store. One day she decides to enter and have a look. Kate likes what she sees, but she does not buy anything.

As soon as the geo-tracker on her smartphone registers her visit, she receives a Marks' & Spencer sponsored post on her Facebook home page, suggesting to start following the company on social media, as she might be interested in it. She then gives her like to the M&S' Facebook page, and starts receiving the news about it.

For a few days, she continues following the M&S news on Facebook, until a post, advertising a discount on a summer dress, catches her attention. She enters in the store and tries to buy the dress on sale, but it is out of stock, so she decides to register to the website and buy it online. Kate feels a bit unsure about her choice, as she usually prefers to physically see and try the clothes she purchases, especially when she has no previous experience with the firm. However, when the receives the product, she loves it, and her trust toward the firm increases.

Soon, Kate discovers the existence of the M&S' app, "Cook with M&S", containing a lot of recipes to do at home. Since she loves cooking, Kate downloads the app on her smartphone and searches for an interesting recipe. When she realizes that some of the ingredients are missing, she goes to the M&S store to buy them. At the checkout, the shop assistant proposes her to join their loyalty program, and since she is sure to come back to shop there again, she accepts.

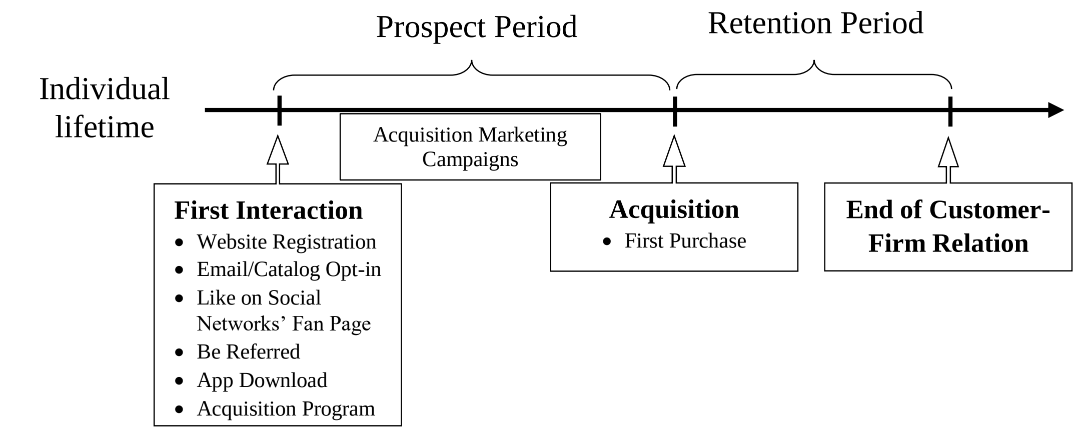
The Kate's shopping story above, tells us a story about the beginning of a relationship: at first, Kate becomes aware of the existences of M&S. Then, she collects more information, through social media, or directly interacting with it. When she gets an incentive to start a deeper relationship, she makes a first trial (e.g. using the discount for the first purchase). Finally, as the first trial increases her trust in M&S, she establishes a long-term relationship with it, by accepting to be part of M&S loyalty program (in other words, purchasing from it on a regular basis).

2.2. Overview

Kate's story highlights an obvious point, i.e., that there is a starting point for every kind of relationship. From a Customer Relationship Management (CRM) point of view, customer acquisition represents the starting point of the customer-firm relationship (Kumar and Petersen, 2012). Customer acquisition is generally defined in literature as the first-time a new customer purchases from a firm or subscribes for a service (e.g. Gupta and Zeithaml, 2006; Schweidel, Fader, and Bradlow, 2008). For example, Wangenheim and Bayon (2007) consider customers as acquired when they subscribe for an energy service for the first time from the provider. Following this definition, Kate is acquired by M&S when she places her order for the dress on the firm's website. However, the first purchase seldom represents the actual first interaction between the customer and the brand or the retailer.

Figure 1 depicts the classical customer lifecycle in a simplified way. At a certain point of their life, people start being interested in a brand. In the previous example, Kate starts to be interested in M&S by passing several times in front of the store during her way to the office. As soon as individuals demonstrate their interest through some concrete actions (e.g. sign-up to the firm's website, opt-in in the email program, receive a referral), firms start considering them as prospects. SendPulse, an American multichannel marketing platform providing marketing services to companies, defines a prospective customer as: "a customer who can buy your product if they want, i.e. that they have physical and financial resources. A prospect for the seller is a client who is in need of the proposed product, but she has some doubts. The task of the seller is to make a prospect buy this product. Marketers can consider as prospective customers all the people who have the characteristics of the target audience for which the product or service is intended. Any entrepreneur before the release of the products thinks about the image of a person, who can buy these products. Some researches of prospective customers are conducted in order to find out where and how to offer them goods" (SendPulse, 2017). In other words, prospects are individuals who are, or can be, interested in the firm's products, they do not have purchased yet, but they are likely to do it in the near future.

FIGURE 1: The simplified customer lifecycle



People can become prospects in several ways, besides being targeted *a priori* by the firm. In an online context, they can register to the website, opt-in in email programs (Kumar, Zhang, and Luo, 2014), or follow the firm fan pages on social networks (DeVries, Gensler, and Leeflang, 2017). Otherwise, they can allow the catalog reception (Anderson and Simester, 2004), be referred from other customers (Biyalogorsky, Gerstner, and Libai, 2001; Schmitt, Skiera, and Van den Bulte, 2011), or download the branded app. Kate, for instance, becomes an M&S' prospect by visiting the store and following its Facebook page.

A prospect's engagement with a firm can involve several interactions. Recalling the introductory examples, before making her first purchase from M&S, Kate visited the store,

gathered information through social media, and received promotional messages. As can be noted, the main feature of this "*prospect phase*" is that these interactions do not include any actual purchase. Usually, prospects engage in search activity, both through the website, the store, or the branded app. They can also receive marketing communications, and interact with the company through social media.
According to the definition of customer acquisition that has always been employed in literature, when customers make their first actual purchase, they turn into newly *acquired customers*. Thus, customer acquisition represents the boundary between the "prospect period" and the "retention period", which lasts until a customer decides to interrupt his/her relationship with the firm.

Managerial press usually describes the process leading prospects to actual customers as the "lead funnel" (Klipfolio, 2015), in which visitors of the store or website become prospects when they demonstrate some interest to the company by subscribing to the newsletter or registering to the website, then they engage with the firm, becoming qualified prospects until the actual purchase. IMSC (Internet Marketing Success Center) has a more complex view of the funnel, describing it as a six-stage process: inquiry, marketing qualified lead, sales accepted lead, sales qualified lead, and new customer. In the first stage, the firm chooses its target and sends them communications to generate awareness. People who get interested in the brand become marketing qualified lead and start getting information about it. Once they have enough information, some of them will consider buying from the firm, becoming sales accepted lead. Then, when they are ready to make the purchase, they move into the sales qualified lead phase, which ends when they actually buy and become new customers (Donahoe, 2016).

It is important to note that the process firms engage to acquire new customers should not be confused with the so-called "customer win-back" behavior, which is defined as "*the process of firms' revitalizing relationships with customers who have defected*" (Thomas, Blattberg, and Fox, 2004; p. 31). Strauss and Friege (1999) provide a detailed description of the main differences between prospects and expired customers: while the first have no experience with the company, and the company does not have much information about them, the latter have previous experience with the company (which is not likely to be positive), and the firm knows them, thus communication with expired customers is easier than communication with prospects. A study conducted by Marketing Metrics further highlights the difference between acquiring new and winning-back expired customers showing that companies are more likely to succeed in selling their products to lost customers (20-40%) than to prospects (5-20%) (Griffin and Lowenstein, 2002).

The marketing activities firms engage in, in order to acquire new customers, usually take the form of acquisition programs. The most widely used are targeted promotions (Fruchter and Zhang, 2004) and promotional discounts to encourage new customers to make their first trial (Anderson and Simester, 2004; Lewis, 2006b; Natter et al., 2015). This strategy rests on the idea that by lowering the price, firms can also decrease the customers' risk perception associated with their first trial (Lewis, 2006b). However, even if price promotions can increase the number of acquired customers, Lewis (2006b) demonstrates that customers acquired through acquisition discounts are less profitable than the others since they are likely to be single-time buyers, or deal-prone customers, who just make their purchases during the promotional period. In this case, one could question whether those kinds of customers can be considered by the firm as actually acquired, since their relationship with the company starts and ends with that single purchase. In addition, this kind of acquisition programs may have a negative effect on customer retention (Farquhar, 2005). About this, Dong, Yao, and Cui (2011) show that acquiring customers through discounts may lead to customer antagonism (Anderson and Simester, 2010) when existing customers realize that they are not the recipients of a promotion, thus firms should be careful in developing their acquisition promotions, since they can spoil the retention of their existing customers.

Referral programs are another widely used strategy to acquire new customers. They are "*a form of stimulated WOM that provide incentives to existing customers to bring in new customers*" (Schmitt, et al., 2011; p. 3). Biyalogorsky, et al. (2001) define two ways in which firms can encourage customers to refer others: low prices and rewards. However, both of them have some advantages and disadvantages. Low prices can increase the probability to purchase and to refer a friend, but at the same time, it can encourage also customers free riding, allowing them to purchase at a lower price without make any referral. On the other hand, rewards can assure the referral activity, but they might be given also to customers who would have referred anyway (Biyalogorsky et al., 2001). Moreover, customers can abuse of rewards by providing a lot of low-valuable referrals (Schmitt et al., 2011). Trusov et al. (2009) raised some concerns about the referral program effectiveness, saying that the firm-initiated WOM might be less valuable than customer-initiated WOM. In the more recent mobile marketing context, for example, the refer-a-friend program is catching on. Companies like Airbnb, American Express, and Uber include this feature in their mobile apps, rewarding users who refer their friends (Extole, 2015).

In the recent years, firms started being active also on social media (e.g. Facebook, Twitter, Pinterest) in order to communicate with their actual and potential customers. De Vries et al. (2017) analyzed the impact of traditional advertising, Facebook messages, and Twitter word of mouth between consumers on the firm's acquisition rate, showing that, even if traditional advertising is still the most effective tool to attract new customers, the firm's Facebook messages and the customers' word of mouth on Twitter play a role in enhancing the acquisition rate.

Finally, firms can also use apps themselves as customer acquisition tools, driving people in store and increasing their interest in the brand, but only if they are not designed as an advertising vehicle and provide some experiential value to their users, continuously engaging and helping them in their everyday lives (Luxury Daily, 2011).

The customer acquisition management allows firms to effectively shape the size and increase the profitability of the customer base (Blattberg, Kim, and Neslin, 2008). The importance of studying customer acquisition lies in the acquisition costs that firms have to face in their attempt to attract new customers (Blattberget al., 2008). Pfeifer (2005) defines the average acquisition cost as the ratio between the investment in acquisition and the percentage of prospects that has been actually acquired. According to eMarketer (2017), for instance, between September 2016 and August 2017 the cost of acquiring an app user who makes a purchase through the app was on average 64.46$, and 76.40$ for in-app purchases (e.g. boosting tools for gaming apps, or unlocking some pro features of the app). In order to help firms face the difficulties associated with the identification of the best trade-off between acquisition and retention investment, several studies try to identify the optimal budget allocation between acquisition and retention efforts (e.g. Blattberg and Deighton, 1996; Berger and Nasr-Bechwati, 2001; Thomas, 2001; Reinartz et al., 2005; Natter et al., 2015) to increase the customers' profitability. Blattberg et al. (2008) argue that in order to increase the number of acquired customers, one potential strategy for firms should be to change the shape of the acquisition curve[1]. This means that firms should be able to better target their prospects, by identifying in advance the ones who have the highest probability to respond, and sending them their marketing activities (Blattberg, et al., 2008; Tillmanns et al., 2017).

[1] The acquisition curve plots the relationship between the acquisition rate and the level of budget allocation in order to acquire new customers (Blattberg, Kim, and Neslin, 2008). Thus, it represents the response rate to different acquisition programs.

According to Monetate Ecommerce Quarterly (2018) in the first quarter of 2018 the overall e-commerce conversion rate[2] was about 2.59%. Specifically, it was 3.77% for traditional devices (e.g. desktop), 3.40% for tablets, and 1.53% for smartphones. These relatively low conversion rates highlight the importance of investigating the factors that are more likely to increase the individual's propensity to convert, in order to develop more effective customer acquisition strategies. As highlighted by Kumar and Petersen (2012): "Customer acquisition is an important step for companies in developing a successful and comprehensive CRM strategy. Identifying the right customers to acquire, predicting customer response to promotional activities, and understanding the long-term effects of marketing on customer acquisition are key components in the acquisition process" (p. 15).

2.3. State of the Art

In spite of the importance of customer acquisition for firms, literature about this topic is scarce. Blattberg, Kim, and Neslin (2008), for example, conclude their book chapter on customer acquisition saying:

> *"In general, there is very little research on acquisition marketing. The traditional marketing literature does not separate the issue of acquiring customers from retaining customers. […] Research in advertising studies the general impact of communications but does not separate newly acquired customers from retained customers. Therefore, it is necessary to develop a new literature on the theory of customer acquisition. […] as important as acquisition marketing is, researchers need to develop theories, principles and empirical generalizations that will help*

[2] The conversion rate is defined as the ratio between the number of prospects and the number of customers who actually make the purchase (Monetate Ecommerce Quarterly, 2018)

> *practitioners develop better acquisition marketing strategies and tactics."* (Blattberg, Kim, and Neslin, 2008; p.514).

Four years later, Kumar and Petersen (2012) still complain about the scarcity of research on customer acquisition:

> "*Without sustained customer acquisition there will eventually be no customers left for the company to try and retain. Despite the significance of customer acquisition, research on this topic is still not sufficient.*" (Kumar and Petersen, 2012; p. 22).

Since the publication of Kumar and Petersen's work in 2012, academic research moves on in the exploration of customer acquisition (e.g. Liu et al., 2015; De Vries et al., 2017; Tillmanns et al., 2017). However, empirical studies about this topic have always been thwarted by challenges, for both marketers and researchers, associated to the collection of information about the identity of perspective customers and their searching behavior within the firm before the first transaction.

2.3.1. Data Availability

The reason why customer acquisition is so much understudied, as anticipated before, lies in the difficulty for both firms and academics to get data about prospect customers. Firms are usually able to track their customers only after they have been acquired, but what they do before their first purchase is often a mystery (Thomas, 2001; Liu, Pancras, and Houtz, 2015; Min, Zhang, Kim, and Srivastava, 2016). Usually, to overcome this issue, firms buy data about their potential prospects from third parties, such as list vendors (Reinartz and Kumar, 2003; Tillmanns et al., 2017), or employ co-operative databases, defined as "*pooling of data across direct marketing firms by a third party vendor in order to provide a broader view of customer transactions and thus enable direct marketing firms that have access to the co-operative databases to refine their promotional strategies*" (Liu et al., 2015; p. 40). The first paper dealing with this data availability issue is the one of Lix, Berger, and Magliozzi (1995), in which the authors show the benefits of linking the geographic, socioeconomic, and lifestyle information contained in commercially available databases, with survey data to better target prospects for a direct mail campaign. In a similar vein, Liu et al. (2015) describe the benefits of using co-operative databases in predicting the risk of bad-debt in acquiring new customers for a direct mail campaign, showing that targeting prospective customers by using information contained in co-operative databases, such as the past purchase behavior with other firms, outperforms the use of more classical targeting strategies, which is based only on the addresses of potential prospects. The work of Tillmanns et al. (2017) deals with the issue of selecting most relevant information in predicting acquisition among the thousands of variables contained in the databases purchased from third parties. They develop a Bayesian variable selection method to optimize the selection of variable in order to target new customers to acquire.

Nowadays, this limitation on data availability is still particularly true in a physical setting, where keeping track of all the individuals' store visits and their searching activity is almost impossible. In fact, it is both hard and costly for firms to link a transaction occurring in a brick-and-mortar store to a customer record, and this is possible only when the customer owns a loyalty card. Moreover, it is impossible for firms to record all the physical store visits of their customers when they do not end with a purchase, or what happened during these visits (e.g. how many products they saw, which product they were looking for, etc.). However, thanks to the rise of the e-commerce and m-commerce, firms who operate in a purely digital environment might be able to overcome this limitation, recording not only most of the marketing communication they receive before their first purchase, but also all the individuals' log-ins on their websites, their searching activity within the web page and through the branded app, and, to some extent, their behavior on social media. According to Statista (2017), the worldwide number of digital shoppers, defined as customers who made at least one purchase through a digital channel from any device (e.g. desktop, tablet, smartphone), was 1.66 billion in 2017, estimated to grow up to 2.14 billion in 2021. This rapid growth of the digital environment in retailing is very promising to investigate issues relating to the searching behavior of customers before the first or between purchase occasions, such as customer acquisition, that were very challenging to analyze in the past years.

2.3.2. Product Centricity vs. Customer Centricity

Due to these challenges in data availability, a significant part of the studies on customer acquisition takes a product-centric perspective (e.g. Hansotia and Wang, 1997; Lewis, 2006b; Schweidel et al., 2008; Nam et al., 2010; Schwartz et al., 2017). Under a product-centric strategy, everything revolves around the development, the marketing, and the selling of the product itself to the mass market (Sheth, Sisodia, and Sharma, 2000), and the main goal of the firm is to sell the product to as many customers as possible (Shah et al., 2006). A key characteristic of product-centricity is that it does not account for customer heterogeneity, treating all the customers in the same way (Fader, 2012). Therefore, under a product-centric perspective, customers are considered as acquired when they purchase for the first time the product (or product category) of interest, no matter whether and what they had bought before. Hansotia and Wang (1997), for example, describe the acquisition likelihood as a function of the different packages of a motor club membership, Cao and Gruca (2005) develop a targeting acquisition tool for the direct mail campaign of a new financial service, while Lewis (2006b) analyzes the impact that acquisition promotional discounts have on the customer lifetime value for a newspaper's subscribers, considering the new subscribers as newly acquired customers, but not accounting for the fact that those people could already have purchased elsewhere that newspaper before the observed subscription. Another evident example of a product-centric study in acquisition literature is represented by the work of Shweidel et al. (2008). In their work, they consider the time needed to a customer to acquire a telecommunication service for the first time, but their starting sample were all the customers who had already subscribed another company's service. It is clear that under a customer-centric viewpoint, it cannot be considered as a real acquisition. In a similar vein, Nam et al. (2010) analyze the effect of WOM on customer acquisition, considering acquisition as the subscription to a video-on-demand (VOD) service. However, the entertainment company who tests the VOD service sells also other different services.

Customer centricity, on the other hand, is defined as "*a strategy that aligns a company's development and delivery of its products and services with the current and future needs of a select set of customers in order to maximize their long-term financial value to the firm*" (Fader, 2012). Thus, the organization is set up around the customer, rather than around the product (Blattberg et al., 2008), and the main question that managers should ask themselves is how many and which products they can sell to the right customer, rather than how many customers they can find for a specific product (Shah et al., 2006).

Customer-centric firms are expected to perfectly know who their customers are, they recognize the fact that customers are not all equal, but some are better than others and are worthy to be satisfied, thus firms should concentrate their efforts on retaining them and acquiring other customers like them, rather than wasting their resources in trying to satisfy and acquire everyone (Fader, 2012). The perfect knowledge of customers, which is pivotal for customer-centricity, requires a "single view of the customer", meaning that firms need to perfectly know all the purchase history and the preferences of everyone in their customer base (Fader, 2012). This is not a trivial task for organizations, especially for the biggest ones (Hart, 1999), since it often requires a complete reorganization of the way in which firms manage their data and evaluate their performances (Deighton, 1997; Shah et al., 2006). Thus, under a customer-centric perspective, customers are actually acquired at the time they make their very first purchase from the firm, regardless the product they buy.

A few studies in customer acquisition attempt to take a customer-centric perspective (e.g. Thomas, 2001; Reinartz et al., 2005; Wangenheim and Bayon, 2007; Chan, et al., 2011). However, most of them focus on the optimal budget allocation between acquisition and retention efforts, accounting for customer acquisition just as a sample selection tool in order to model the duration of the retention (Thomas, 2001; Reinartz et al., 2005). By contrast, Wangenheim and Bayon (2007) analyze the effect of WOM referrals on customer acquisition for a German energy service provider by taking a customer-centric perspective. However, these authors do not account for the fact that some of the acquired customers could have actually been won-back customers, thus, the observed adoption of the focal energy provider might not have been their real first purchase.

Recent marketing literature tends to emphasize the importance of customer-centricity, because it allows marketers to shift from a mass-marketing strategy, where one single product or a single message is supposed to fit to all the customers, to an individual marketing strategy, where every single customer is different from all the others and has to be treated in a customized way (Sheth et al., 2000). Customer-centricity increases customer satisfaction since firms can target their customers on the basis of their actual desires and needs, avoiding overwhelming them with too many communications and marketing actions (Blattberg et al., 2008). Moreover, taking a customer-centric approach enables firm to gain competitive advantage: first, thanks to the deep knowledge of clients that a customer-centric approach involves, firms are able to build special relationships with each of their best customers, and these relationships will be difficult for competitors to emulate (Day, 2000); second, it not only enables firms to better satisfy customers' needs but also to anticipate them, thus to fulfill future needs before other players (Ashley and Morrison, 1997). In addition, Homburg, Droll, and Totzek (2008) demonstrate that focusing the efforts to the most valuable customers enhances the customer profitability and the returns on sales, since customer-centricity allows firms to reduce marketing costs, and the focus on higher-level customers (customer prioritization) does not spoil the firms' relationship with lower-level ones.

As stated before, shifting from a product-centric to a customer-centric approach is not an easy task. Firms need to reorganize their databases around their customers rather than around their products (Fader, 2012). Today, thanks to the technological advancement, data is as available for firms as never before, and their data storage and management skills are dramatically enhanced. An excellent example of customer-centricity in the service industry is provided by Marriott International. Marriott International owns 19 brands internationally (e.g. Marriott Hotels & Resorts, Delta Hotels, BVLGARI Hotels & Resorts, Ritz-Carlton) and, instead of storing their guests' information in separated silos for each brand, the Marriott International managers made these databases "talk" with each other, allowing them to have a single-view of the customer (HVS, 2008), that is particularly relevant for cross-selling strategies[3] (Blattberg et al., 2008). In their book *Database Marketing*, Blattberg, Kim, and Neslin (2008) show clearly how taking a product-centric approach, rather than a customer-centric one, can lead to completely different outcomes. In fact, if marketing managers focus their attention on product X, they can try to sell that product to the customer who has the highest probability to buy it. However, by focusing the attention on the customer's side, they can realize that these particular customers are actually more likely to purchase other products. Thus, in trying to sell them these products rather than product X, the firm can not only maximize the probability to sell something to these customers, but it can also increase the customers' satisfaction toward their shopping experiences (Blattberg et al., 2008).

As can be seen, there are several differences between product-centricity and customer-centricity. Fader (2012) suggests that customer-centricity is important for firms also in

[3] Cross-selling is defined as the firms' strategy to encourage "*a company's customer who have already bought its Product A to also buy its Product B*" (Deighton, Peppers, and Rodgers, 1994; p.91)

developing their acquisition strategies because it allows companies to mainly focus their acquisition efforts on potential high-value customers.

2.3.3. Factors Influencing Customer Acquisition

In order to develop an effective customer-centric acquisition strategy, firms first need to know which factors influence more the prospects' likelihood to purchase from the brand for the first time. However, only a few studies have investigated the drivers leading prospects to make their first purchase.

Some of those studies explain acquisition by means of transactional variables (e.g. Hansotia and Wang, 1997; Thomas, 2001; Lewis, 2006a). Hansotia and Wang (1997), try to determine which prospects should be contacted basing on the customer lifetime value (CLV), Thomas (2001) considers acquisition as a sample selection tool to estimate her Tobit model for customer retention, but she adds the service that has been bought as a predictive variable for the acquisition, while Lewis (2006a) models the probability for a customer to be acquired by means of order size penalties (such as shipping fees based on the order size), showing that order size penalties have a significant detrimental effect on the probability of a customer to be acquired. Notably, these measures, which have been used in order to predict the acquisition, are actually available only after the customer has already been acquired.

Other studies dealing with the investigation of factors influencing customer acquisition are represented by the studies of Trusov, Bucklin, and Pauwels (2009) and Wangenheim and Bayon (2007), in which the authors analyze the effect of WOM referrals on customer acquisition for a social networking site, demonstrating that personal WOM significantly boosts users' sign-ins to the website, and, similarly, in Nam, Manchanda, and Chintagunta's (2010) study about the adoption of a video-on-demand (VOD) service, who demonstrate that the contiguous WOM has an effect on customer acquisition and, more specifically, the negative WOM effect has almost twice the effect of positive WOM. More recently, DeVries, Gensler, and Leeflang (2017) analyze the different impacts of traditional advertising, messages shared on social network (e.g. Facebook) by the company, and messages shared on Twitter by users on brand building metrics – such as awareness, consideration, and preference – and the acquisition rate, finding that traditional advertising, and both firm and customer-initiated messages on social networks have an impact on customer acquisition.

It is worthy to note that there is a lack of studies analyzing the relationship between the prospects' pre-acquisition behavior, that is, the interactions that prospects have with the firm before their first purchase, and their propensity to be acquired. The only exception can be found in Reinartz et al.'s (2005) work in a B2B context, in which they model customer acquisition as a function of the channels customers used to start their relationship with the firm (among other marketing and demographic variables), finding a significant impact of the customer initiated contact on acquisition.

Other studies in different branches of marketing, such as mobile marketing, demonstrate that the customers' behavior through different devices can predict future purchases. For example, De Haan et al. (2015) found that customers who search through different devices (e.g. smartphone and desktop) have a higher conversion rate than customers who use a single device. Moreover, when they are approaching the final purchase, they are more likely to use fixed devices. Thus, the device usage can be a predictor of customer acquisition, since prospects who use different devices for their searching activities might have a higher probability to make their first purchase. Moreover, the purchase intention can be enhanced after the adoption of the branded app (Bellman et al., 2011; Dinner, Van Heerde, and Neslin, 2015). Bellman et al. (2015) find that app usage increases brand awareness and subsequent purchase intention, Dinner et al. (2015) demonstrate that branded app users have higher propensity to buy in both online and offline stores, and this effect is higher for the purchases in online context. Finally, Kim, Wang, and Malthouse (2015) show that app usage increases customers' spending behavior. In their study, De Haan et al. (2015) find that the number of times that a customer enters in the website has a significant and negative impact on his/her probability to purchase. This is a first signal of the importance of accounting also for the number of interactions and the cross-device behavior that a prospect has before the acquisition time.

2.3.4. Customer Acquisition and Retention

The scarcity of studies investigating the factors leading to acquisition is also due to the fact that, as highlighted by Blattberg, Kim, and Neslin (2008), there is very little literature focusing on acquisition alone. The vast majority of previous studies accounts also for customer retention. These studies have several purposes. Most of them focused the attention on the optimization of the budget allocation between acquisition and retention efforts (Thomas, 2001; Reinartz et al., 2005; Ovchinnikov, Boulu-Reshef, and Pfeifer, 2014; Natter, Ozimec, and Kim, 2015; Schwartz, Bradlow, and Fader, 2017). For example, Natter, et al., (2015) and Schwartz et al. (2017) develop tools to optimize the budget allocation respectively for price comparison websites and display advertising. Among the studies dealing with the budget allocation issue, some of them investigate the relationship between market dynamics and acquisition and retention expenditures (Fruchter and Zhang, 2004; Voss, and Voss, 2008; Min, Zhang, and Kim, 2016). Fruchter and Zhang (2004) investigate the relationship between acquisition and retention costs, and the firm’s market share. They demonstrate that a firm with a small market share should focus more on customer acquisition, while bigger firms should focus on the retention of their existing customer base. Moreover, a firm A has a higher advantage of acquiring firm B’s customers when A has high and effective acquisition investments and the retention investments if firm B are limited and ineffective. Voss and Voss (2008) show that in markets with high competition it is better to focus on an acquisition-oriented strategy, and Min, et al. (2016) analyzed the way in which the market dynamics (such as market share leadership, competition, and market penetration) affect customer acquisition and retention expenditures, finding that the higher is the number of competitors in the market, the grater are the acquisition costs. Moreover, the acquisition costs are more sensitive than retention costs to the market position, and the stage of the life-cycle in which the product is. A study conducted by Arnold, Fang, and Palmatier (2009) shows that firms who invest more on their acquisition strategies are able to come up with more radical innovation rather than incremental innovation, which is in turn boosted by a more retention-oriented strategy. Finally, Musalem and Joshi (2009) investigate the trade-off between acquisition and retention efforts in the case in which two firms are competing for the same customers, showing that acquisition investments should be higher for customers who are marginally profitable for the competing firm.

Besides the budget allocation between acquisition and retention investments, other studies analyze different types of relationship between acquisition and retention (Schweidel et al., 2008; Villanueva et al., 2008; Schmitt et al., 2008; Chan, Wu, and Xie, 2011). For instance, Schweidel et al. (2008) investigate the relationship between the time that an individual takes to purchase a telecommunication service and the duration of his/her relationship with the firm, showing that people who wait longer to subscribe tend to have a longer retention period. Other researchers analyzed the relationship between the retention and the acquisition channel or acquisition program: Verhoef and Donkers (2005) use a financial service provider to investigate the effects on customers' retention and cross-buying behavior of the channel through which they have been acquired (i.e. direct mail, web site, TV or radio advertising). In an online context, Villanueva et al. (2008) analyze the difference in retention between users who register on a company's website coming from a marketing activity or from word of mouth (WOM), and demonstrate that, while marketing activities generate users with a higher short-term value, on his hand WOM brings to the website users with a higher long-term value. In a similar vein, Schmitt et al. (2011), investigate the effect of referral programs on customers' value and retention, finding that referred customers have both higher value and retention rate, and Chan, Wu, and Xie (2011) analyze the effect of being acquired through Google Search Advertising on the CLV, demonstrating that this channel brings more valuable customers to the firm compared to the other channels.

Other studies addressed the issue of the effects of acquisition promotions on customers' retention (Lewis, 2006a; Lewis, 2006b; Dong, Yao, and Cui, 2011), or the difference in the promotions effectiveness for new and established customers (Anderson and Simester, 2004). For instance, Lewis (2006a, 2006b) analyzes the impact of shipping fees and promotions on new customers' value, finding that new customers are attracted more by shipping fee discounts based on order size rather than by the shipping fee level itself, but customers acquired through promotions are less valuable than the others. Anderson and Simester (2004) show that a deeper price discount actually increases the purchase probability of new customers, but decreases the purchase probability of already established customers, while Fruchter and Zhang (2004), considering targeted promotions, demonstrate that the promotional activity for acquisition purposes of a firm should increase as its competitors' market share increases, and should decrease as the redemption rate of the promotion increases.

It is important to note that the most of the studies reviewed in this section do not analyze the antecedents of customer acquisition (e.g. Anderson and Simester, 2004; Lewis, 2006b; Schmitt at al., 2011; Natter et al., 2015; Min et al., 2016; Schwartz et al., 2017), but only the way in which the acquisition strategy of the firm influences the future retention of the new customers.

2.3.1. Targeting Prospects

It is evident that customer acquisition has a very close relationship with customer retention. Since acquiring new customers requires a lot of efforts, firms has at least to be quite sure that they are not going to waste their money with customers who are not valuable (Blattberg et al., 2008; Tillmanns et al., 2017). The customers they are going to acquire need to be likely to have a good expected lifetime value (Olensky, 2017). Of course, the more firms know about their prospective customers, the easier it is for them to target prospects properly and to create a long term relationship with them (Coler, 2018). For this reason, the level of risk and uncertainty associated to the acquisition exercise is significant, since the firm does not have any previous experience with prospects it is going to target, and the available information to predict the risk of acquiring a new customer is limited (Liu at al., 2015), as they usually have to rely on lists of addresses provided by a vendor. In fact, a large proportion of prospects who are more likely to be acquired by companies turns out to be *bad customers,* who are strongly deal-prone and buy only with deep promotions (Lewis, 2006b), or who require a significant amount of efforts to be retained, bringing little additional value to the company (Cao and Gruca, 2004), or turn out to be *bad debt* customers, as they are unwilling or unable to pay for the product or service they purchased (Liu at al., 2015), furthermore, they can be chronic returners (e.g. Anderson, Hansen, and Simester 2009; Ofek, Katona, and Sarvary 2011). In all these cases, the acquisition costs are much higher than the additional value that those new customers bring to the firm. The definition of "right" customers to target is "*customers who purchase a large amount with each order, buy frequently, and remain active over a long period of time*" (Hansotia and Wang, 1997; p. 8).

The issue of targeting prospects is one of the most explored in the customer acquisition field, right after the budget allocation for acquisition efforts discussed previously. One of the first works dealing with this issue is the one of Lix et al. (1996), in which the authors highlight the importance of considering individual survey data in addition to the classical information (e.g. geographic, sociodemographic, lifestyle) contained in the commercially available dataset in order to target their prospects in a more efficient fashion. One year later, Hansotia and Wang

(1997), advance the first analytical model to select prospects who are worthy to be acquire, and to find the most effective way to acquire them. Cao and Gruca (2005) develop a modeling framework to target prospects who are both valuable and likely to respond to the marketing activities of the firm, while Liu et al. (2015) model the risk of targeting bad debt customers and demonstrate that a targeting scheme accounting for this risk gives a better prediction of the future customer profitability. More recently, Tillmanns et al. (2017) develop a Bayesian variable selection tool to select the most effective variables in targeting prospects.

Firms can also employ a "two-step acquisition method" to target their prospects. This method is particularly effective in the online environment, as it requires people to "self-select" into the pool of recipient of the firm's marketing communication. In other words, companies can place a banner or a sponsored link within a general website or in a search engine, and then send their targeted communications only to people who clicked on that advertising, who have demonstrated some interest in the advertised product or service. This method is particularly cost effective, since the first step requires a cheaper mass-marketing communication, while the second step follows naturally, without any further investment in targeting-related activities (Blattberg et al., 2008).

2.4. Summary

This chapter has introduced the concept of acquisition as it is known in today's academic and managerial worlds, and has reviewed the current state of the art of customer acquisition literature. Table 2.4.1 summarizes the main empirical studies on customer acquisition.

TABLE 2.4.1: Empirical studies on customer acquisition

Studies	Antecedents of acquisition: Marketing actions	Antecedents of acquisition: Prospects' behavior	Retention	Perspective
Lix, Berger & Magliozzi (1995)	No	No	No	Customer-centric
Hansotia & Wang (1997)	No	No	Yes	Product-centric
Thomas (2001)	No	No	Yes	Product-centric
Reibstein (2002)	No	No	Yes	Customer-centric
Anderson & Simester (2004)	No	No	Yes	Customer-centric
Fruchter & Zhang (2004)	No	No	Yes	Not at individual level
Cao & Gruca (2005)	No	No	No	Product-centric
Reinartz, Thomas, Kumar (2005)	Yes	Yes	Yes	Customer-centric
Verhoef & Donkers (2005)	Yes	No	Yes	Customer-centric
Lewis (2006a)	Yes	No	Yes	Product-centric
Lewis (2006b)	No	No	Yes	Product-centric
Wangenheim & Bayon (2007)	Yes	No	No	Not at individual level
Schweidel, Fader, Bradlow (2008)	No	No	Yes	Product-centric
Villanueva, Yoo, and Hanssens (2008)	Yes	No	Yes	Customer-centric
Voss & Voss (2008)	No	No	Yes	Customer-centric
Trusov, Bucklin, and Pauwels (2009)	Yes	No	No	Customer-centric
Nam, Manchanda, and Chintagunta (2010)	No	No	No	Product-centric

Schmitt, Skiera, Van den Bulte (2011)	No	No	Yes	Customer-centric
Chan, Wu, Xie (2011)	Yes	No	Yes	Customer-centric
Liu, Pancras, Houtz (2015)	No	Yes	Yes	Product-centric
Natter, Ozimec, and Kim (2015)	No	No	Yes	Product-centric
Min, Zhang, Kim, and Srivastava (2016)	No	No	Yes	Not at individual level
De Vries, Gensler, and Leeflang (2017)	Yes	No	No	Not at individual level
Schwartz, Bradlow, and Fader (2017)	No	No	No	Product-centric
Tillmanns, Hofstede, Krafft, and Goetz (2017)	No	No	No	Customer-centric

The table highlights the fact that almost a half of the studies in customer acquisition takes a product-centric perspective, and the vast majority of research in acquisition deals also with retention. Very few papers analyzed customer acquisition as a stand-alone concept.

Predictive variables like marketing activity and pre-acquisition behavior has been seldom taken into consideration. The only studies investigating the role of prospects' behavior are the ones of Reinartz et al. (2005) and Liu et al. (2015). However, the latter consider the bad behavior that prospects of the focal firm had with other firms, and Reinartz et al. (2005) only account for the number of times a customer got in contact with the firm. None of them consider prospects' searching activity or their demonstrations of engagement with the focal brand.

The aim of this research is to contribute to this stream of literature by analyzing acquisition as a stand-alone concept, taking a customer-centric perspective, and accounting for factors influencing customer acquisition as the firm's marketing activities, and the pre-acquisition behavior of its prospects, in terms of interactions with the different touchpoints, and searching activity.

3. CONCEPTUAL FRAMEWORK

The current chapter deals with the theoretical development of the three research questions listed in the Introduction. The first section addresses the need for a new definition of customer acquisition in a non-contractual setting, explaining it through a recall of the adoption of innovation process, and highlighting the parallelism between the acquisition and adoption processes.

The second section is dedicated to the question: “How does the customer acquisition process evolve over time?”. The idea underlying this research question is that, before being acquired, as time passes by, prospects go through a process composed by different stages, defined by their knowledge of the firm and the intensity of their relationship with it. The section provides a brief overview of the notions of customer *buying process* and *funnel* in the marketing literature, in order to give credit to the idea of the Customer Acquisition Process.

In the third section the question “What is the role played by the different marketing activities in each stage of the acquisition process in driving prospects toward acquisition? How does the pre-acquisition behavior influence the acquisition process?”. This section recalls the most important steps made in the recent attribution literature, which examines the extent to which each touchpoint between the customer and the firm contributes to the final conversion.

The fourth and last section summarizes the main arguments of the previous in order to build the theoretical development of the research questions. This last part aims at introducing our concept of Customer Acquisition Process, by identifying different classes of prospects and the dynamics across these classes.

3.1. Alternative Definitions

The Kate's story depicted in the prologue of the previous chapter can be useful in order to better understand the issue that the current section intends to investigate, but first, we need to make a few changes, turning it into a non-happy ending.

Example – Kate's Shopping

[…] For a few days, she continues following the M&S news on Facebook, until a post, advertising a discount on a summer dress, catches her attention. She enters in the store and tries to buy the dress on sale, but it is out of stock, so she decides to register to the website and buy it online. Kate feels a bit unsure about her choice, as she usually prefers to physically see and try the clothes she purchases, especially when she has no previous experience with the firm. After a long wait, she receives the product, and she feels very disappointed. The dress quality is very poor, and it has a dark spot in the white skirt. Angry, she looks for a way to return it and get a refund, but then she discovers that she has to pay the returning shipping fees. Kate then decides to keep the dress and try to fix it by herself, but she swears she will never buy from M&S again.

As can be noted, this modified version tells the story of a very short relationship. In fact, Kate's buying relationship with M&S involves only one purchase. In light of this event, can M&S consider Kate as a newly acquired customer? Accordingly, one of the main objectives of this thesis is to find an alternative definition for customer acquisition in a non-contractual setting.

3.1.1. Customer Acquisition and Innovation Adoption Process

An interesting parallelism exists between the concept of acquisition and the concept of innovation adoption. In particular, the innovation adoption process, defined as a process consisting of "*a series of actions and choices over time through which an individual or an organization evaluates a new idea into ongoing practice*" (Rogers, 1962; p. 163), describes the stages individuals pass through before adopting a new product or service at a regular basis. According to Rogers (1962), the adoption process consists in a series of five stages: awareness, interest, evaluation, trial, and adoption[4].

It is important to note that both processes of acquisition and adoption share a common ground: they require the customer to engage in a new behavior. However, the difference between the concepts of acquisition and adoption is subtle.

Adoption focuses mainly on a product or service, as it refers to a consumer's decision to buy or use a new product, in other words it is a choice centering on the product and its features. (e.g. Manchanda et al., 2008; Risselada et al., 2014). Moreover, the adoption literature does not distinguish between a firm's existent customers and the customers whose first interaction with the firm occurs with the innovation ((e.g. Prins and Verhoef, 2007; Risselada et al., 2014; Bilgicer et al., 2015). Furthermore, it should be noted that previous studies on new product adoption focuses mainly on non-contractual settings (e.g. Coleman, Katz, and Menzel, 1966; Bilgicer et al., 2015). This issue will be deepened in the following section.

Acquisition, on the other hand, refers to the customer acquisition to the firm (e.g. Blattberg and Deighton, 1996), as it represents the beginning of a new relationship between a customer and a firm. This implies that, like the adoption, also the concept of acquisition entails an element of novelty, but this novelty is at the individual-level (in other words, it is a novelty only for the single potential customer), because the firm could be already established. However, a firm can be considered as an innovation for the customer as far as this customer has not purchased it yet (in other words, at the acquisition time), perceiving it as a novelty for her. In fact, Rogers (1983)

[4] This concept will be explained in detail in the next section, "The Customer Acquisition Process".

defines an innovation as "*an idea, practice or object that is perceived as new by an individual or the unit of adoption. It matters little [...] whether or not an idea is "objectively" new as measured by the lapse of time since its first use or discovery. The perceived newness of the idea for the individual determines his or her reaction to it. If the idea seems new to the individual, it is an innovation*" (Rogers, 1983; p.11).

Another important adoption of innovation process feature that is worth mentioning is the following Gatignon and Robertson (1985)' statement: "*The concept of adoption has been used in a rather limited way to refer to a single decision. Yet for consumer product diffusion, adoption should be conceptualized more multidimensionally. For many consumer products, repeat purchase is the key to adoption; for others, it is important to assess adoption as to both width and depth.*" (Gatignon and Robertson, 1985; p. 854). They claim that researchers should not only analyze adoption as the number of individuals using the innovation for at least one time (the width of adoption), but also the concept of depth, defined as the extent to which the product is purchased by a single individual, should be taken into account. However, the adoption depth has always been mainly overlooked. Manchanda et al. (2008) study represents a notable exception. In their work, they distinguish between physicians who only made their first trial and physicians who actually adopted the new drug through a measure of post-trial prescription intensity. In the current research, we contend that since a customer can see a firm as an innovation, as discussed before, the concepts of width and depth could also be applied to the acquisition process.

In this work, the unit of analysis is the single adopter. For this reason, the width of the acquisition, or the number of acquired customers, will not be considered. Instead, the attention will be focused on the concept of depth, defined as the new customers' purchase intensity from the focal firm. Thus, accounting for a different definition of acquisition is useful in order to distinguish acquisition from pure trial, as the acquisition concept entails not only the first purchase occasion, but this purchase occasion should represent the beginning of a relationship between the customer and the firm, while the concept of trial neglects this aspect.

3.1.2. Contractual and Non-Contractual Settings

The redefinition of the concept of acquisition can be particularly valuable in a non-contractual setting. Unlike contractual settings, where a customer is bounded to the firm for a fixed period of time after having signed a contract (e.g. banking account, insurance, etc.), and the firm can observe the end of this relationship (Min et al., 2016), in non-contractual settings, customers do not subscribe any contract, and they could be making just a single first transaction (commonly considered in literature as the acquisition moment) and disappear thereafter. In this context, when a customer makes her first purchase from the firm, she is not bounded to the firm for the future.

TABLE 3.1.1: Customer Acquisition studies in contractual and non-contractual settings

Contractual Setting		**Non-Contractual Setting**	
Study	**Product/Service**	**Study**	**Product/Service**
Hansotia and Wang (1997)	Club membership	Reibstein et al. (2002)	Online retailer
Thomas (2001)	Club membership	Reinartz et al. (2005)	High tech manufacturer
Cao and Grouca (2005)	Financial services	Lewis (2006a)	Online retailer
Wangenheim and Bayon (2007)	Energy provider	Chan et al. (2011)	Biochemical lab supplies
Schweidel et al. (2008)	Telecommunication service		
Villanueva et al. (2008)	Website registration		
Nam et al. (2010)	VOD service		
Schmitt et al. (2011)	Financial services		
DeVries et al. (2017)	Telecommunication service		
Schwartz et al. (2017)	Financial services		
Tillmanns et al. (2017)	Insurance		

Previous studies on customer acquisition are carried out mainly in contractual settings, making the new customers' identification easier (e.g. Wangenheim and Bayon, 2007; Nam, et al., 2010). Table 3.1.1 emphasizes the point by summarizing a selection of studies about customer acquisition in contractual and non-contractual settings. Thus, in a non-contractual setting, if a customer makes her first purchase and then she never comes back, as in the Kate's Shopping modified example, can she really be considered as an acquired customer by the firm?

According to Rogers (1962)'s adoption process model, after getting aware and interested in an innovation, in the trial stage the individual uses the innovation for the first time and decides whether or not to continue adopting. In this view, a product is not considered as adopted right after the first trial, but only when its usage becomes regular. Similarly, in the acquisition context in a non-contractual setting, after the first purchase from a retailer, customers can decide whether or not to further purchase from it, confirming their acquisition to the company, or to remain only triers. For this reason, in this work we propose that in non-contractual settings, acquisition is not represented by the first observed purchase, but it is an unobservable state that is the outcome of a set of factors revolving around search, purchase and the monetary amount spent.

3.2. The Customer Acquisition Process

The example at the beginning of the previous chapter shows clearly that, before making a decision, people usually have to pass through different steps. In the story, Kate became aware of the existence of M&S by passing every morning in front of it. This repeated exposure to the store triggered her interest, which in turn drove her in looking for more information through social media and by visiting the store. After processing the information gathered and evaluating them, Kate finally decided to make the first transaction.

The existence of this form of multistage decision-making process is well known in marketing literature, both academic and managerial. Researchers and marketers have already established that, in taking their buying decisions, customers often go through a series of steps, usually defined as hidden cognitive states, that help them make their final choices. The essence is that, as people move on in the stages of the process, they come closer to the purchase (e.g. Lavidge and Steiner, 1961; Rogers, 1962; Roberts and Lattin, 1991; Bettman, Luce, and Payne, 1998; Court et al., 2009; Dierks, 2017). This is because, as time passes by, people become more knowledgeable about the structure of the problem they are facing, and the way in which they approach the problem changes accordingly (Bettman et al., 1998). In fact, as explained by Bettman et al. (1998, p. 188) in their paper, "preferences for options of any complexity or novelty are often constructed, not merely revealed, in making a decision" (Bettman and Park, 1980; Tversky, Sattah, and Slovic, 1988; Payne, Bettman, and Johnson, 1992).

The advantages of looking at the consumers' decisions as a series of multiple stages rather than a one-shot process are several. First, it helps marketers and advertisers better understand the customers' mental scheme in decision making (Van Lennep, 2014), second it is useful in order to analyze how customers change their behaviors over time, according to their mindset evolution (Dierks, 2017). Third, understanding the customers' buying process scheme, allows marketers to develop more customized marketing programs, providing to customers the most relevant information in the most appropriate time, depending on the mental state in which the customer is in her buying process (Court, 2009), making their marketing communications more effective (Lovejoy, 2014). In this way, marketers can design their marketing activities around the customers, providing more customer-centric shopping experiences (Faulds, 2018). Fourth and last, empirically speaking, researchers show that multistage choice models have a better performance in predicting choice decisions than traditional simple choice models (Roberts and Lattin, 1991; Andrews and Srinivasan, 1995).

3.2.1. Multistage Processes in Advertising Studies

The idea of a multistage process traces back in the advertising field. At the very beginning of the XX century, St. Elmo Lewis developed the AIDA (Attention, Interest, Desire, Action) model (Strong, 1925), which states that, in order to be effective, and advertising has first to be able to catch the *Attention* of potential customers, then to trigger their *Interest* toward the advertised product, which in turn should turn into a *Desire* to buy it, leading them to the final purchase *Action*. This model has been widely cited in marketing research (e.g. Barry, 1987; Vakratsas, and Ambler, 1999; Ehrenberg, 2000; Lin and Huang, 2006; Kotler and Armstrong, 2011), even in the early days.

The AIDA model falls under the hat of the Hierarchy of Effects models (Lavidge and Steiner, 1961). In their paper, the authors build upon the AIDA model, developing their own advertising staged process: Awareness, Knowledge, Liking, Preference, Conviction, and Purchase. In their view, the role of an effective advertising is in the first stage to make potential customers *aware* of the existence of the product or brand, then it has to provide them some information in order to make people *knowledgeable* about it. Of course, this information should trigger a positive attitude (*liking*), that in turn drives *preference* over other similar products. At this point, potential customers have to be *convinced* to make the final *purchase*.

A lot of other versions of the hierarchy of effects model have been proposed in the subsequent years. Among them, it is worth mentioning the one of Starch (1923), as he made the first attempt to propose metrics to measure the advertising effectiveness. In his view, an advertising first has to be *seen*, *read*, *understood*, and *remembered*. At last, the potential customer should *act upon* it. Another mention has to be done for Colley (1961), who summarized Lewis, Lavidge and Steiner, and Starch's model into one: people shift from a stage of *unawareness* to a stage of *awareness*, then they have to *comprehend* the message of the advertising, which leads to the subsequent stages of *conviction*, *desire*, and, at last, *action*.

3.2.2. Multistage Processes in Buying Behavior Studies

In the 1969, Howard and Sheth extend the concept of hierarchy of effects by introducing the idea of a sequence of buying behavior (Barry, 1987). Their sequence of buying behavior spans across five stages: *attention*, followed by *brand comprehension*, *attitude, intention*, and, finally, *purchase*. The decision-making process concept is well established in the consumer research, especially in regards to the brand choice context, where people first search for alternatives to put into their consideration set, then evaluate these alternatives in order to finally make their final decision (e.g. Engel, Blackwell, and Miniard, 1968; Court et al., 2009). Researchers in consumer behavior studies, usually see the decision making process as a funnel, in which consumers start with a lot of possible brands in mind, and then they narrow their choice set (the set of considered brands to purchase) as they move down in the funnel, until they make their final choice (e.g. Engel et al., 1968; Engel, Miniard, and Blackwell, 2006; Jansen and Schuster, 2011; Kotler and Armstrong, 2011; Yadav and Pavlou, 2014). There is no convergence on the name of the funnel, neither on its stages (Dierks, 2017). It has been called in several ways, some examples are the buying funnel (Jansen and Schuster, 2011), shopping funnel (Yadav and Pavlou, 2014), purchase funnel (Barry, 1987; Wiesel, Pauwels, and Arts, 2011), and conversion funnel (Abhishek, Fader, and Hosanagar, 2016).

The first consumer decision process model was developed by Engel et al. in 1968. It is a seven-stage model: need recognition, information search, evaluation of the alternatives, purchase, consumption, post-consumption evaluation, and divestment. In the first stage, the individual perceives a *need*. This need can be elicited by internal or external sources (Kotler and Armstrong, 2011). In the former case fall the needs such as hunger, tiredness, which are perceived internally by the individual. The external sources, on the other hand, comprise the social interactions, marketing communications, advertising, friends, or the simple exposure to a product. In the second stage, the individual *looks for information* about the ways in which she can satisfy her need, forming her consideration set of alternatives. She can start paying more attention to advertising messages, talk with friends, browse the Internet, or in social media (Kotler and Armstrong, 2011; Dierks, 2017). When she gathered enough information and built her consideration set, the individual *evaluates the alternatives* she had identified in the previous stage, narrowing down the number of alternatives until she comes up with the best option to *purchase*. After the *consumption*, in the *post-consumption* phase, the customer evaluates the experience she had with the product or the service bought. The last phase, *divestment*, occurs when she stops using it. For example, if she had bought a car, divestment occurs when she scrapped it.

Building upon Engel et al.'s (1968) model, in the following years researchers have developed several other versions of the consumer decision process. The most famous is the buying decision process (Sirakaya and Woodside, 2005; Engel et al., 2006; Kotler and Armstrong, 2011; Faulds et al., 2018), which is a revised version of the consumer decision process. Unlike the former version, the buying decision process puts all the post-purchases stages together in one single *post-purchase* step. Accordingly, it has five stages: need recognition, information search, evaluation of the alternatives, purchase, post-purchase (e.g. Kotler and Armstrong, 2011). Other versions of the decision process in the brand choice field encompass stages such as: *Awareness, Familiarity, Consideration, Purchase, and Loyalty* (Court, 2009; Song et al., 2016), where an individual starts in the awareness phase with a lot of brands in mind, and then she progressively diminishes the number of alternatives as she moves further in the funnel. Jansen and Schuster (2011) consider the buying funnel as a four-step process: *Awareness, Research, Decision, and Purchase*, while Yadav et al. (2013) present a more synthetic shopping funnel (also called purchase decision process), by synthetize Engel et al.'s (1968) information search and evaluation of alternatives phases into a single *pre-purchase* phase. Thus, their model is formed by four stages: *Need recognition, Pre-purchase activities, Purchase, and Post-purchase activities*.

In the recent years, due to the development of Internet and mobile technologies which allows consumers to be always connected, and to interact with firms at any moment in time and from every place in space, some researchers start raising some concerns about the linearity of the customer decision funnel (e.g. Court et al., 2009; Edelman, 2010; Srinivasan, Vanhuele, and Pauwels, 2010; Pauwels and van Ewijk, 2013; Edelman and Singer, 2015; Lemon and Verhoef, 2016; Colicev et al., 2018). In their view, people can go back and forth among the states, adding, for example, new alternatives in the consideration set even from the later stage of the process. Court et al. (2009) call this new phenomenon the Customer Decision Journey, a nonlinear process in which customers can move from the *Initial Consideration* phase, in which the consumers have already in mind their consideration set, as they are continuously exposed to advertising, friends' experiences through social media, online reviews, etc. Then, in the *Active Evaluation* stage, they continue to adjust their choice set by adding new alternatives or dropping others on the basis of the incredible amount of information they get every day, until they *Buy* the product. The post-purchase phase of the Consumer Decision Journey is completely different from the one hypothesized in the funnel. In fact, an increasingly number of consumers do online research even after the purchase, and, when they *Enjoy* the product, they pass into the *Advocate* phase. In this phase, if they are satisfied with their purchase, they can provide good feedback, otherwise, they can be very harsh with the reviews. In the last stage, satisfied customers create a *Bond* with the product and, if it is strong enough, it can allow customers to enter in a loop which comprehends only the Enjoy, Advocate, and Buy stages, bypassing completely the Consideration and Evaluation activities (Edelman, 2010). To sum up, the Customer Decision Journey has six main stages: Initial Consideration, Active evaluation, Buy, Enjoy, Advocate, and Bond. The main characteristic of this model is that customers can move almost freely from one stage to another. Edelman and Singer (2015) propose a modified version of the Customer Decision Journey, the New Customer Decision Journey, in which they completely eliminate the first two phases, stating that companies performing well in delivering value to customers can easily compress, if not completely eliminate, these two initial stages, allowing customers to start their journey directly from the purchase stage.

3.2.3. Multistage Processes in Adoption of Innovation Studies

Besides the advertising and buying behavior contexts, Rogers (1962) expands the multistage decision-process concept into the adoption of innovation field, describing his *Adoption Process,* defined by Kotler and Armstrong (2013) in their book as "the mental process through which an individual passes from first hearing about an innovation to final adoption" (p. 186). The model encompasses a sequence of five stages: Awareness, Interest, Evaluation, Trial, and Adoption (AIETA model, Rogers, 1962), which lately, in a subsequent edition, he renamed as Knowledge, Persuasion, Decision, Implementation, and Confirmation (Rogers, 1983). According to his model, in the first *Awareness/Knowledge* phase, the potential adopter first becomes aware of the existence of an innovation and looks for more information (*Interest/Persuasion* phase) that can help her in forming an attitude toward the novelty, which could be positive or negative. Based on this attitude, she can decide to adopt or reject the innovation (*Evaluation/Decision* phase). If she adopts, she purchases and tries it (*Trial/Implementation* phase) and, if she is satisfied with the experience, she can decide to adopt it at a regular basis (*Adoption/Confirmation* phase).

Drawing on Rogers' model, other researchers propose their own versions of adoption process. Robertson (1971) defines the process of adoption of new process as a sequence of *Awareness, Comprehension, Attitude, Legitimation, Trial, and Adoption* stages (Barry, 1987), while Ehrenberg (1974) develop the *Awareness, Trial, Reinforcement* (ATR) model, stating that the advertising role should not be primarily to enhance the first trial, but rather it should trigger repeat buying (*Reinforcement* phase). Preston, in a series of papers published in 1982, 1983, and 1984, presents his Association Model (Preston, 1982), which in the final version includes three Action Stages: *Search, Trial,* and *Adoption* (Preston and Thorson, 1984).

Even if the multistage adoption process is theoretically recognized, empirically still very few studies in adoption of innovations analyze the adoption decision as a process. In fact, it has more often been treated as a one-shot decision (Tellis and Stremersch, 2003, Van den Bulte and Stremersch, 2004; Prins and Verhoef, 2007; Demoulin and Zidda, 2009). For example, Prins and Verhoef (2007) and Demoulin and Zidda (2009) analyze respectively the timing of adoption of a new telecommunication service and of a loyalty card of a retailer using a split hazard model, in which the adoption is modeled as a single, binary choice. Similarly, Manchanda, Xie, and Youn (2008) study the effect of marketing, personal communication, and of the temporal dimension on the adoption of a new pharmaceutical product, but still they treat adoption as a single, discrete choice.

The few exceptions of studies analyzing adoption as a multistage process are the ones of Beal, Rogers, and Bohlen (1957), Kalish (1985) and Van den Bulte and Lilien (2007). In their study, Van den Bulte and Lilien (2007) analyze the adoption of a medical innovation with a two-stage hazard model, accounting for awareness and positive evaluation. They model adoption as a three-stage Markov process, with the first state having people not aware, and hence non adopters, in the second stage people aware, but who did not evaluate the innovation positively, hence non adopters, and in the last stage the adopters, thus people who are aware of the innovation and evaluate it in a positive way. Kalish (1985) analyze the two-stage adoption process of durable goods, defined as awareness (generated by advertising activity and word of mouth) and adoption.

More recently, Lambrecht, Seim, and Tucker (2011) introduce the concept of Adoption Funnel. They analyze the adoption of a financial online service as a 4-stage funnel. In the first stage, users have to *sign up* to the website, then, the first time they log in to the service using the information provided at the registration, they *evaluate* it. In the third step, they *try* to make the first transaction, and in the last stage, they continue to *use* the service at a regular basis. In other words, Lambrecht et al (2011)'s Adoption Funnel comprises four stages: Sign up, Evaluation, Trial, and Regular Usage. It is worthy to note that their stages are observed, as people are required to do specific actions in order to move into the next.

3.2.4. Multistage Processes' Dynamics

The vast majority of the studies dealing with the consumer decision process analyze the differences in the impact of the firm's activities on the desired outcomes across the stages of the process (e.g. Jansen et al., 2011; Lambrecht and Tucker, 2013; Yadav et al., 2013; Colicev et al., 2018). For example, Colicev et al. (2018) consider awareness, purchase intention, and satisfaction metrics as the three stages of the funnel, analyze the effect that owned and earned media have on these three metrics, and, in turn, how these metrics influence the shareholder value. Lambrecht and Tucker (2013) study the effectiveness of retargeted display advertising on the purchase funnel, showing that more personalized messages work better in the late stages of the process, while Yadav et al. (2013) discuss the roles played by social commerce in the need recognition, pre-purchase, purchase, and post-purchase phases. In the acquisition literature, De Vries et al. (2017), analyze the effects that traditional advertising and social messages have on brand-building metrics, such as awareness, consideration and preference, and, ultimately, customer acquisition.

On the other hand, the factors influencing individuals to move from one stage to another have not been so much investigated. For example, Abhishek et al. (2016) find that the exposition to display advertising trigger people's propensity to move on in the funnel only if they are on the early stages, but an excessive exposition in the later stages can make people come back in the funnel. Schweidel, Bradlow, and Fader (2011) analyze the stages in the relationship between a multi-service provider and its customers and how the promotional activities of the firm affect customers' switching from one stage to another. They show that promotions decrease the probability for customers to move into the *end of the relationship* final stage.

3.2.5. Multistage Processes in Customer Acquisition Studies

Despite the enormous relevance of the multistage decision process concept into many branches of the marketing field, as discussed in previous sections, to the best of my knowledge, this concept still remains unexplored in the customer acquisition context. In fact, most of the studies in acquisition literature consider customer acquisition as a single-step process (e.g. Hansotia and Wang, 1997; Lewis, 2006a; Wangenheim and Bayon, 2007; Nam, et al., 2010), or as a first stage before the retention phase (e.g. Thomas, 2001; Schweidel et al., 2008). The only study that can be considered as closer to the concept of funnel is the one of De Vries et al. (2017), who use a VAR model to analyze the different effects of social messages and traditional advertising on branding metrics - such as awareness, consideration, and purchase - and customer acquisition. Nonetheless, in their paper, they never mention their intention to analyze customer acquisition as a multistage process, or a funnel.

In this work, we contend that before being acquired, people go through a multistage process, namely, the Customer Acquisition Process. Accordingly, one of the main purposes of the present research is to explore the number and features of the stages that make up the Customer Acquisition Process, and the factors influencing people to switch across those stages. This point will be discussed better in Section 3.4.

3.3. Attribution Modeling

3.3.1. Touchpoints

During the decisional process, customers and firms get in touch with each other several times and in several ways. Practitioners and academics call the occurrence of each one of such events a *touchpoint*. (e.g. Court et al, 2009; Edelman and Singer, 2015; Verhoef, Kannan, and Inman, 2015). Baxendale, Macdonald, and Wilson (2015) define a touchpoint as "an episode of direct or indirect contact with the brand" (p. 236).

Touchpoints can be distinguished between firm-initiated, and customer-initiated (e.g. Blattberg et al., 2008; Bowman and Narayandas, 2001; Shankar and Malthouse, 2007; Wiesel et al., 2011).

Firm-Initiated Touchpoints. In the past, touchpoints were mainly firm-initiated, meaning that firms took the initiative to contact their potential customers. For example, firm-initiated touchpoints can be traditionally found in the offline environment, such as print advertising, TV spots, radio, or advertising boards, which do not require customers to make any action in order to receive the communication, as they are passively reached by the firm. They are particularly useful in reaching prospects who are not knowledgeable about the firm yet (De Haan, Wiesel, and Pauwels, 2016). A backfire of firm-initiated contacts is that they are usually unwanted (Blattbergh et al., 2008), and can be increasingly seen as annoying by customers.

In the online environment there are a lot of different types of touchpoints between a firm and its customers, both firm and customer-initiated. Examples of online firm-initiated touchpoints include display advertising and retargeting, newsletter campaign, social media (Anderl et al., 2016), and co-marketing with other brands. Display advertisings are "images embedded in the contents of Web pages that link the user to the advertiser's website (Xu, Duan, and Whinston, 2014). Thanks to the retargeting, the content of the advertisement can be personalized on the basis of the consumers' browsing history (Lambrecht and Tucker, 2013). Research shows that personalized communications increase the effectiveness of the advertisement in terms of relevance, trust, and response (e.g. Hoffman and Novak, 1996; Komiak and Benbasat, 2006; Dias et al., 2008). Newsletters are a form of permission-based marketing in which the firm sends promotional or informational e-mails to its customer base almost on a daily basis (Cho and Khang, 2006; Ellis-Chadwick and Doherty, 2012), and social media advertising is "a set of advertising platforms belonging to the field of social media, such as networks (e.g., Facebook), micromedia (e.g., Twitter), or other (mobile) sharing platforms (e.g., Instagram)" (Anderl et al., 2016; p. 460; Saboo, Kumar, and Ramani, 2016). Co-marketing occurs when the marketing activities and the brands of two firms collaborate with each other. For example, years ago Nike started collaborating with Apple to create Nike+iPod, a new product designed for sportive people who want listening music during their workout (Bernazzani, 2018).

Customer-Initiated Touchpoints. In the recent years, thanks to the enormous development of the online environment, customer-initiated touchpoints, represented by any type of contact between a customer and a firm which is initiated by a prospect or an already established customer (Ghose and Yang, 2009; Wiesel et al., 2011; De Haan, et al., 2016), are catching on. Advantages of customer-initiated touchpoints are that they are seen as less intrusive by customers (Shankar and Malthouse, 2007), as customers are responsible for their initiatives, and are usually more profitable than firm-initiated touchpoints (e.g. Shankar and Malthouse, 2007; Wiesel et al., 2011; Li and Kannan, 2014), and receive higher response rates (Sarner and Herschel, 2008).

Examples of online customer-initiated touchpoints in the online environment are represented by direct visits to the firm's website, organic search, and paid search (Anderl et al., 2016). Direct visits occur when the user lands in the website by writing the address in the address bar, while organic search (SEO) are the list of web pages that browser search engines (e.g. Google, Yahoo…) provide to the user when she types a keyword to search (Jerath, Ma, and Park, 2014). Similar to SEO, paid search (SEM), are links to web pages that appear at the top of the list of the search engine results (Li et al., 2016). The difference is that SEM results are sponsored links that "are determined using online auctions in which advertisers bid to be placed in response to queries by consumers and therefore are more commercial" (Jerath et al., 2014; p. 480).

Hybrid Touchpoints. There are also other types of "hybrid" touchpoints, such as earned touchpoints, like word of mouth (Stephen and Galak, 2012; Baxendale et al., 2015; deVries et al., 2017), and customer/firm-initiated touchpoints (Anderl et al., 2016), like affiliation with third party websites, and referrals. Word of mouth (WOM) is defined as any peer-to-peer conversation regarding the firm (Stephen and Galak, 2012; Baxendale et al., 2015) and has been proved to be more effective than traditional advertising communications (e.g. Trusov et al., 2009; deVries et al., 2017). Affiliate websites are third party shopping websites, displaying the products of different retailers, through which users can land on the seller websites in order to get more information about the product, or make the purchase. Affiliate websites usually do not sell anything by themselves. Referrals are text links to the brand's webpage included in other websites, without necessarily being remunerated (Anderl et al., 2016).

3.3.2. Attribution Modeling

The challenge that marketers have to face with this multiplicity of touchpoints is how to give the right credit to each channel for the desired outcome, usually represented by a conversion (Moffett, Pilecki, and McAdams, 2014; Kumar et al., 2016), in order to optimally allocate their budget across channels and publishers[5] (Bettman, 2017; Abhishek et al., 2017). Finding the solution to this challenge is becoming increasingly relevant in the last years, and this task is known in literature with the name of *attribution modeling* (e.g. Shao and Li, 2011; Li and Kannan, 2014; Berman, 2017; Abhishek, Despotakis, and Ravi, 2017). Kumar et al. (2016) define attribution modeling as “the science of using advanced analytics to allocate appropriate credit for a desired customer action to each marketing touch point across all online and off-line channels” (p. 450).

The increasing growth of the online environment for shopping-related activities has been a godsend for attribution studies. In fact, thanks to the higher traceability of online touchpoints, it becomes easier for marketers to follow the path to purchase of their customers (Kannan et al., 2016). The availability of clickstream data paves the way for the development of attribution models with different degrees of complexity by practitioners and academics.

3.3.3. Attribution Methodologies

Practitioners often rely on services offered by publishers. Google, for example, provides marketers ad-hoc services to give the right credit to each marketing channel (Google, 2017a-b). Such services employ several heuristic rules, such as the “last-click” or “first-click” attribution rules. In practice, in the first case they attribute the credit for the desired outcome to the last touchpoint the customer interacted with before the conversion, and in the second case the credit is attributed to the first touchpoint, ignoring all the other touchpoints that occurs during the customer journey. Another adopted heuristic rule consists in giving the same credit to all the touchpoints occurred before the conversion (Kannan et al., 2016; Kireyev et al., 2016; Abhishek et al., 2017). On their hand, researchers approached attribution modeling in several

[5] A publisher is the subject who is paid by the advertiser firm and is in charge of linking the advertising message to the right individual (Abhishek et al., 2017), as it knows the past online behavior of users across websites (Bettman et al., 2017).

ways: Shao and Li (2011) are the first to deal with this topic. Using data about a multichannel advertising campaign of a brand selling software products and services, they employ bagged logistic regression and probabilistic models to first identify the model with the best predictive performance, and then to estimate the effectiveness of each touchpoint. Dalessandro et al. (2012), Berman (2017), and Abhishek et al. (2017) face the attribution issue by developing three analytical models according to the Shapley value rule (Shapley 1953), which is based on the concept of fairness (Rabin, 1993), and states that “the payoff of each player […] is a weighted sum of his marginal contributions to every subset of players” (Abhishek et al., 2017; p. 11). Thanks to is easiness in implementation, the Shapley value rule is also adopted in some Google’s attribution models (Google, 2017b). Li and Kannan (2014), in the multichannel online environment, model the consideration, visits, and associated purchases of online channels with a measurement model, Xu et al. (2014) use Bayesian estimation to estimate a multivariate point process model in the electronics industry, and Abhishek et al. (2016) analyze the impact of online display advertising in each stage of the purchase funnel by estimating a Hidden Markov Model (HMM). Ghose and Todri (2015) and Barjas et al. (2016), on their hand, employ experimental designs to investigate the effectiveness of display advertising. Anderl et al. (2016) demonstrate that their graph-based attribution model outperforms the last-click strategy in four different industries. Li et al. (2014) employ structural equation models to explore how the keywords bidding decisions of the advertisers, the publisher’s ranking rules for those keywords, and the click-through and conversion rates relate to each other. Saboo et al., (2016) use autoregressive models to analyze the relationship between consumers’ social media activity and their subsequent purchase behavior in the context of the music industry. Other studies employ a vector error correction model (VEC) to analyze the relationship between clicks on SEM and display advertising (Kireyev, et al., 2016), and a structural vector autoregressive model (SVAR) to analyze the effects of several types of online advertising on website performance metrics, such as visits, conversions, and revenues (De Haan, Wiesel and Pauwels, 2016). De Haan et al. (2016)'s work is the only paper yoking into account also offline advertising, together with the one of Joo, Wilbur and Zhu (2016), who study the relationship between TV advertising and online search behavior with a conditional choice model.

3.3.4. Attribution in Multistage Processes

Since the introduction of the funnel concept, academics and practitioners have been interested in exploring the role of marketing activities in the different stages of the funnel (e.g. Frambach, Roest, and Krishnan, 2007; Naik and Peters, 2009; Court et al., 2009). In their introductory paper to the IJRM special session about attribution modeling, Kannan et al. (2016) complain about the lack of studies accounting for the role of marketing activities in the different stages of the customer journey. In fact, only few studies deal with the impacts of different touchpoints across the stages of the purchase funnel, demonstrating that the effectiveness of the channels changes according to the stage in which the customer is in her decisional process. (DeHaan et al., 2016; Abhishek et al., 2016; Abhishek et al., 2017).

De Haan et al. (2016) investigate the role of customer-initiated marketing activities and firm-initiated marketing activities in each of the stages of their so-called *website funnel* (landing to the website's home page, examination of some product pages, creation of the shopping basket, and check-out) across five product categories. They demonstrate that customer-initiated marketing activities outperforms firm-initiated activities in each stage of the funnel. Moreover, the authors distinguish between customer-initiated contacts generated websites with product-related contents, and unrelated contents. They find that, as customers move on in the funnel, customer-initiated contacts generated by content-related websites perform better than contacts generated by unrelated websites. Abhishek et al. (2016), using data about the marketing campaign for the launch of a new car, analyze the impact of different types of online advertisings (e.g. display in generic or specific websites, paid search) in the stages of the conversion funnel (namely: disengaged, active, engaged, and conversion states). Their results indicate that display advertising impressions and clicks in both generic and specific websites increase the likelihood to move from the disengaged to the active state, but they do not have any effect in moving people from the active to the engaged state. Even worse, too many impressions on generic websites have a detrimental effect for the final conversion and increase the likelihood to come back from the engaged to the disengaged state, and clicks in the engaged state decrease the probability of moving out of it. They also demonstrate that display advertising in specific websites increases the conversion probability in the engaged state, and only in the later stage of the funnel people use search engines to search for the product. Abhishek et al. (2016)'s results are in line with Lambrecht and Tucker (2013)'s, who find that more personalized and less generic display advertising is more effective in the later stages of the funnel, when consumers are closer to the purchase. A different issue faced by Abhishek et al. (2017) is how to allocate the budget between two publishers in a two-stage funnel: awareness and conversion. They develop an analytical model to find the optimal budget allocation on the basis of Shapley value rule, demonstrating that in the awareness stage, advertisers should invest more in publishers who increase the probability to move in the next stage of the funnel, and not in publishers whose intent is to directly drive conversions.

3.4. Theoretical Development

3.4.1. The Customer Acquisition Process

In marketing research, it is already well established that customers are heterogeneous (e.g. Chintagunta, Jain, and Vilcassim, 1991; Wedel et al., 1999; Horsky, Misra, and Nelson, 2006; Blattberg et al., 2008; Zhang, Kumar, and Cosguner, 2017), both in terms of observed behavior and hidden preferences. For example, some people may prefer purchasing after a long search, others enjoy browsing several products without making any purchase, while others are more goal oriented: they visit the website with a clear idea about what they are looking for, and purchase the product with very little search.

Search activities are clearly related to purchase activities, since, before purchasing, people usually look for information about the product, the brand, or the retailer (e.g. Hoffman and Novak, 1996; Schlosser, White, Lloyd, 2006; Verhoef, Neslin, and Vroomen, 2007). Moreover, website visitors who undertake an intensive search activity also have a higher conversion rate (Moe, 2003; Johnson et al., 2004). As already seen in section 3.2, previous studies dealing with the multistage processes in marketing studies usually identify the stages in the processes through both search and purchase behaviors. For example, in the buying process, once the customer recognizes a need, she goes through a series of stages, such as *search for information* and *evaluation of alternatives* (e.g. Engel et al., 1968; Sirakaya and Woodside, 2005; Engel et al., 2006; Kotler and Armstrong, 2011; Faulds et al., 2018), or *research* and *decision* (Jansen and Schuster, 2011), which are characterized by a high amount of search activity. Thus, we assume prospects' search behavior intensity (which can be low, medium, or high) to play a pivotal role in the customer acquisition process.

Even more obvious than search intensity, we consider the purchase behavior as a second dimension of the acquisition process. As explained previously in section 3.1, in this work we are challenging the view that a customer is acquired as soon as she makes her first purchase. In this line, studies analyzing the adoption of innovation process (section 3.2.3 of this chapter) distinguish between the *trial/implementation* stage, in which the individual uses the innovation for the first time, and the actual *adoption* stage, when she adopts the innovation at a regular basis (e.g. Rogers, 1962; Robertson, 1971; Ehrenberg, 1974; Rogers, 1983; Preston and Thorson, 1984; Lambrecht et al., 2011). Thus, we assume prospects' purchase intensity (which can be none, low, or high) to be an important indicator of the stage in which the prospect is in her acquisition process.

Figure 3.4.1 represents a 3×3 matrix summarizing the way in which we expect prospects to be categorized in terms of search (low, medium, high) and purchase (no, low, high) intensities during the acquisition process.

FIGURE 3.4.1: Potential classes of customers in the Acquisition Process

		PURCHASE		
		NO	LOW	HIGH
SEARCH	LOW	(1)	(4)	(7)
	MEDIUM	(2)	(5)	(8)
	HIGH	(3)	(6)	(9)

Referring to the traditional definition of customer acquisition, according to which customers are acquired right after the first purchase (e.g. Gupta and Zeithaml, 2006; Schweidel, Fader, and Bradlow, 2008), in Figure 3.4.1 acquired customers in both contractual and non-contractual settings are the ones who populate cells from (4) to (9): even if they make very few purchases, if not only the first, the firm considers them as acquired. In this work, we contend that in non-contractual settings, acquisition is not represented just by the first purchase, but it is an unobservable state that new customers reach only after their shopping from the firm becomes regular. Thus, according to our work, acquired customers in non-contractual setting are expected to be the ones who populate cells from (7) to (9) in Figure 3.4.1.

As already discussed in section 3.2, the preferences and behaviors of customers change according to the stage in which they are in the process. Thus, in every cell of the matrix represented in Figure 3.4.1, marketing activities and prospects (or customers)' interactions with the firm can have different effects on their propensity to search or to buy. For example, in the online context, the device through which users log in to the firm's website can play a different role according to the cell in which the user is at the time she logs in. DeHaan et al. (2015), show that people move from more mobile devices to less mobile devices (e.g. from smartphones to desktops) as they get closer to the purchase phase. Moreover, the use of mobile devices is usually associated to search-related activities (Shankar and Balasubramanian, 2009; Okazaki and Hirose, 2009), and its use has been found to enhance brand metrics associated to the early stages of the customer journey, such as brand awareness and attitude (Bellman et al., 2011; Bart, Stephen, and Sarvary, 2014; Peters, Amato, and Hollenbeck, 2014; Wang, Kim, and Malthouse, 2016). Accordingly, the use of more mobile devices can be expected to have a positive effect on search activity in the no-purchase categories (cells from (1) to (3) in the matrix), and a negative effect on the purchase propensity in the high/low purchase categories, while the desktop use in these categories can enhance the probability to buy.

Previous studies demonstrate that customer-initiated contacts are more effective in increasing people's likelihood to convert (e.g. Wiesel et al., 2011; Jerath and Park, 2014; De Haan et al., 2016; Anderl et al., 2016). In fact, both De Haan et al. (2016) and Anderl et al. (2016)'s studies demonstrate that the effectiveness of customer-initiated touchpoints is higher than the one of firm-initiated touchpoints in the entire purchase funnel. The McKinsey research conducted by Court et al. (2009) shows that among the touchpoints that occur in the later stages of the customer decision journey, almost two-thirds are customer-initiated, as people use Internet to gather information and reviews about the product they are going to buy. Jerath and Park (2014) demonstrate that the use of paid search is associated to customers who are closer to purchase, and Abhishek et al. (2016) find that customers use to search for the products in search engines only at the end of the funnel. Thus, we can expect people who search and purchase more, and who take the initiative to get in contact with the firm, to be more likely to search and to buy.

Following Bruner and Kumar (2000), who show that the more people have search experience, the less they are influenced by advertisings, Court et al. (2009) state that brand advertising should be provided in the first part of the consumer decision journey, in order to create awareness in the potential customers and to help them in their information seeking. In fact, in their research, Court et al. find that 39% of touchpoints in the early stage of the customer decision journey are firm-initiated. This is supported by the findings of Xu et al. (2009), who show that display advertising do not affect purchases, but increase people's likelihood to use other types of touchpoints to get in contact with the firm in future. Similarly, Ghose and Todri (2015) find that display advertising triggers search behavior for the brand and the product. On the other hand, Li and Kannan (2014) demonstrate that retargeting and display advertising can have a detrimental effect on the purchase likelihood. Thus, we can expect people who search and purchase less, and who are contacted by the firm, to be more likely to search than people of the same class who are not contacted by the firm.

As discussed in chapter 2, the firm's promotional activity has a positive effect in acquiring new customers (e.g. Fruchter and Zhang, 2004; Anderson and Simester, 2004; Lewis, 2006b; Natter at al., 2015), because a price discount decreases the perceived risks associated with the trial of a new brand (Lewis, 2006b). However, Lewis (2006b) states that a significant part of customers acquired through a price discount can be single-buyers, deal-prone customers. Moreover, Farquar (2005) raises some concerns about the ability of promotions to retain customers. For these reasons, we can expect promotional activity on the website to enhance people's likelihood to convert in the low purchase state, but to decrease it in the high purchase state.

According to the state in which they are, people can search in different ways. Nowadays, all the main online retailers allow their users to refine their search by applying filters to the results (e.g. color, size, price, category…), or by ordering the search results on the basis of a specific characteristic, usually price or product novelty. The possibility to refine the search can enhance people's propensity to search, and can increase their likelihood to convert, as it makes easier for them to find easily what they are looking for. Moreover, users can use different tools according to their state. For example, people who are in the high search/no purchase cell can be more likely to use the novelty or price descendant rankings, as they are likely to be browsing for entertainment, while people who are more likely to purchase can look for items with a lower price. To the best of our knowledge, there is a significant lack of marketing studies dealing with the tools users use to refine their search. The only study explicitly account for filtering tools and ranking of the search results is the one of Chen and Yao (2016), who found that these refinement tools enhance people's likelihood to search, and increase their purchase utility by decreasing the search costs.

3.4.2. Customer Acquisition Process Dynamics

Since the idea of a linear funnel has been replaced by the more recent concept of customer journey (Court et al., 2009; Edelman, 2010; Edelman end Singer, 2015), during the acquisition process, prospects can not only move from one stage to the following one, but they can move freely from one stage to any other in the process.

Previous literature demonstrates that the switching probabilities from one state to another of the process are influenced by external factors (e.g. Netzer, Lattin, and Srinivasan, 2008; Montoya, Netzer, and Jedidi, 2010). For example, the registration to the firm's website or the subscription to the newsletter can determine an advancement in the acquisition process. In fact, once the user signs-in to the firm's website, the firm knows who she is, since at the moment of the registration she is asked to provide some personal information, and it is able to engage in direct marketing activities, which has been found to increase profits (Chiang, Cchajed, and Hess, 2003), the likelihood to adopt an innovation (Risselada et al., 2014), and purchases (Kim, and Kumar, 2018). Moreover, the provision of personal information to a company allows the firm to send more personalized communications, leading to a higher satisfaction (Peppers and Rogers, 2017). Furthermore, it indicates that the user trusts the company enough to share her information (Schumann, Wangenheim, and Groene, 2014; Martin, and Murphy, 2017). The subscription to the newsletter reception can be expected to play a similar role. Permission marketing analyzes the individuals' decision to opt-in or opt-out to email programs. "*Opt-in marketing refers to firms explicitly asking customers for permission, usually when an online account is created. Customers can opt-out any time after they opt in*" (Kumar et al., 2014; p. 404). Usually, people can decide to opt-in to email programs because they trust the brand and have a positive image of it (Barnes and Scornavacca, 2008), or because they perceive to have a high participation level in the program, since it provides them value (Krishnamurthy, 2001). Research finds that customers who opt-in to email programs are more responsive to marketing communications (Marinova, Murphy, and Massey, 2002), have a higher purchase intention (DuFrene et al., 2005), and spend more (Jolley et al., 2013). Factors like trust, higher satisfaction, and higher purchase intention can easily drive people in moving ahead in the acquisition process. However, e-mail marketing has also a dark side, as people can start feeling overwhelmed by the firm's communications and, consequently, they can decide to opt-out the newsletter program (Kumar et al., 2014), and to avoid interacting with the firm. A study conducted by Zhang, Kumar, and Cosguner (2017) shows that there is an optimal number of e-mails that firms should send to their customers, and sending too many communications has a detrimental effect on the firm's profits. This result is in line with Li and Kannan (2014)'s, who highlight the fact that e-mail marketing can annoy some customers and decreases their purchase likelihood.

In the online retailing setting, when users see an interesting product, but they need to further evaluate it, postponing their decision to buy it or not, they can save the item of interest in the Wish List, a virtual container in which users can save products without adding them into the final shopping cart. Popovic and Hamilton (2014) analyze the effect of the Wish List usage on the purchase likelihood, finding that delaying the buying decision to a subsequent visit, leads people to re-evaluate their choices, actually decreasing their desire to buy the items they put previously into the Wish List. On the other hand, Close and Kukar-Kinney (2010) confirm their hypothesis according to which users at the beginning of the session may add all the products of interest in the cart or in the Wish List in order to narrow their consideration set later during the visit before the actual checkout. To sum up, we can expect the usage of the Wish list to help people moving on in the process, but only when items are added or dropped during the current session. Otherwise, it may have a negative effect when it has been used in the previous visit. Table 3.4.1 summarizes the expectations presented above about the factors influencing the customer acquisition process and its dynamics.

TABLE 3.4.1: Expectations about the factors influencing the customer acquisition process

Factor	**Effect**	**Categories**
Mobile	Positive on search activity	No purchase
	Negative on purchase propensity	High purchase/Low purchase
Desktop	Positive on purchase propensity	High purchase/Low purchase
Customer-initiated touchpoints	Positive on search activity and purchase propensity	High search
Firm-initiated touchpoints	Positive on search activity	Low search/Low purchase Low search/No purchase
Promotional activity on the website	Positive on purchase propensity	Low purchase
	Negative on purchase propensity	High purchase

Factor	**Process dynamics expectations**
Website registration	Move on in the process
Newsletter subscription	Move on in the process
Too many newsletter	Do not move on in the process
Wish List usage in the current session	Move on in the process
Wish List usage in the previous session	Do not move on in the process

4. METHODOLOGY

The goal of this thesis is to analyze the customer acquisition process as a series of hidden states, defined by the prospects' searching and buying activities intensity. In each step of the process, we aim at investigate which factors (e.g. marketing activities, browsing behavior) drive users to move on in the funnel, and what kinds of activities affect people's likelihood to purchase and to search in each stage of the process. To do so, we need to:

1. Identify and describe the hidden stages of the process;
2. Investigate which factors affect the prospects' probability to switch between states;
3. In each stage, estimate the effects of marketing activities on both searching and buying behaviors.

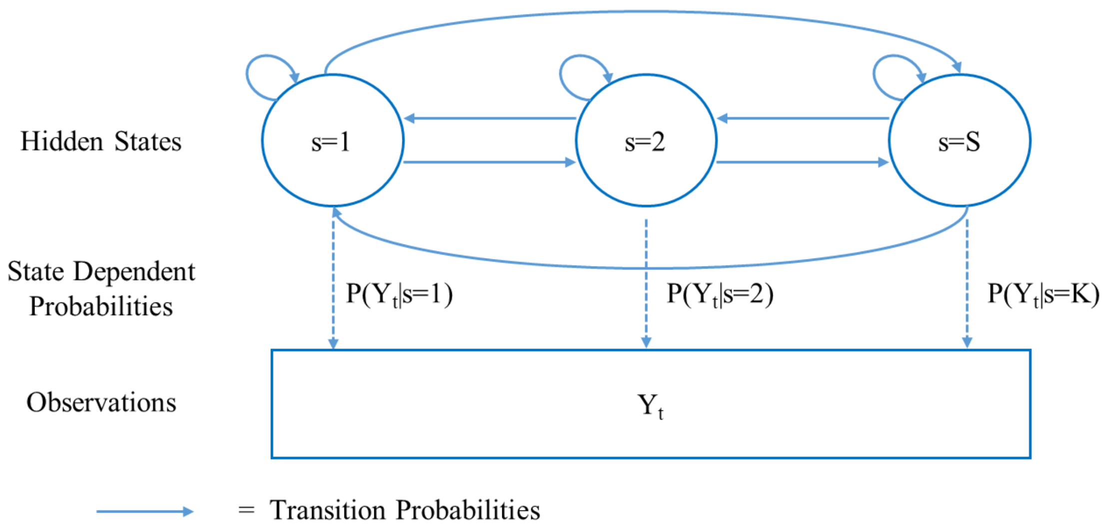
Hidden Markov Models (HMM) are particularly suited to accomplish these tasks.

In this chapter, we introduce the Hidden Markov Models methodology. The first part of the current chapter explains the general idea underlying HMM, and its applications in the different fields, in particular its employment within the marketing studies. The second part describes the main components of HMM: the initial state distribution, the transition probability matrix, and the state dependence probabilities. The third part of the chapter is dedicated to a brief discussion about the treatment of the unobserved customer heterogeneity in the proposed HMM. The final two parts of the chapter deal with the HMM estimation though the EM algorithm, introduced by the explanation of the concepts of forward and backward probabilities, and the model selection tools in order to select the number of latent states.

4.1. Introduction to Hidden Markov Models

HMM are defined as "models in which the distribution that generates an observation depends on the state of an underlying and unobserved Markov process" (Zucchini, MacDonald, and Langrock, 2016; p.3). In other words, HMM assume the existence of unobserved states in the process, and the individuals' transition among these states determines the realization of the observed behavior (Netzer, Ebbes, and Bijmolt, 2017). The latent states follow a Markov chain, and the realizations of the observed behavior are independent conditionally on the latent Markov process (Bartolucci, Farcomeni, and Pennoni, 2012). Since the states and the individuals' state membership cannot be observed directly, they are estimated from the observed behavior of the individual (Stamp, 2015). HMM can be seen as an extension of Latent Class modeling (Kamakura and Russell, 1989), as the latent states of HMM are actually latent classes in which individuals are grouped on the basis of their observed behaviors. The particularity of HMM which differentiate them from Latent Class Models lies in the fact that HMM's states are dynamic, in the sense that individuals can switch from one state to another at any point in time (Netzer et al., 2008; Bartolucci et al., 2012)

FIGURE 4.1: HMM structure



Adapted from Netzer et al., 2017

Figure 4.1 depicts the structure of a HMM: let's assume the existence of a process with a finite number of hidden states $s = \{1, \ldots, S\}$. At any point in time $t = \{1, \ldots, T\}$, an individual can be in any of the unobserved states. The researcher cannot observe the state in which the individual is, but he can observe the individual's behavior Y_t. The *state dependence probabilities* represent the probability that the individual behaves as Y_t given the state in which she is. Accordingly, the researcher can assess the state in which the individual is at time *t* by observing her behavior. In time *t+1*, the individual can stay in the same state as she was in time *t*, or she can shift to any of the other states. The probability to shift from a state to another is called *transition probability*.

The typical employment of HMM is to analyze long time series of observations of a single unit of analysis. The most famous example of HMM application to a time series situation is the Unfair Casino problem (Durbin et al., 1998). In this situation, a casino player – the unit of analysis – has to throw a dice. He can choose between two dices: one is fair, and has the same probability to produce any number from 1 to 6, the other is biased. The biased dice has a higher probability to produce a 6. The player does not know whether he is throwing the fair or the biased dice. Accordingly, the hidden states of the process are *fair* and *biased*. He can assess the state in which he is only through the observations of the dices outcomes.

To account for longitudinal data of a panel, as it usually happens in marketing studies, where a group of customers is monitored for a shorter period of time, Latent Markov Models (LMM) are employed (Bartolucci et al., 2012). Longitudinal data can be *univariate* or *multivariate*, depending on the number of observed variables considered: in the first case we consider only one response variable, and in the latter more than one (Bartolucci et al., 2012). To align with previous marketing studies, in which Latent Markov Models are usually known as Hidden Markov Models (Visser and Speekenbrink, 2010), in the present study we refer to Latent Markov Models as Hidden Markov Models.

HMM born in the late 60s (Baum and Petrie, 1966; Baum et al., 1970). Since then, HMM found applications in several fields, starting from speech and image recognition (respectively, Rabiner et al., 1989; Yamato, Ohya, and Ishii, 1992), to healthcare and genetics (Eddy, 1998; Rainer and Miller, 2000; Auranen et al., 2000; Cook, Ng, and Meade, 2000; Bartolucci, Lupparelli, and Montanari, 2009), criminology (Bijleveld and Mooijaart, 2003; Bartolucci, Pennoni, and Francis, 2007), passing from environmental studies (Hughes and Guttorp, 1994), organizational studies (Wang and Chan, 2011), and economics (Hamilton, 1989; Hamilton, 2008) and finance (Mamon and Elliott, 2007). In the recent years, marketing literature witnessed a growth of studies employing HMM (e.g. Montoya et al., 2010; Abhishek et al., 2016; Ascarza, Netzer, and Hardie, 2018). In fact, HMM are particularly useful in analyzing, for example, the stages of the purchase funnel (Park and Gupta, 2011), as well as the stages of the customer-firm relationships (Netzer et al., 2008; Ascarza and Hardie, 2013; Romero, Van der Lans, and Wierenga, 2013; Ma, Sun, and Kekre, 2015; Ascarza et al., 2018), or to analyze the searching activity on Internet (Montgomery et al., 2004; Abhishek et al., 2016). A lot of studies in marketing also found in HMM a useful tool to investigate the impact of the firms' marketing activities on the customers' likelihood to shift from one stage to another (e.g. Netzer et al., 2008; Montoya et al., 2010; Li, Sun, and Montgomery, 2011; Kumar et al., 2011; Luo and Kumar, 2013; Zhang, Netzer, and Ansari, 2014; Abhishek et al., 2016). Differently from the other fields, in marketing studies employing HMM, the focus is more on the hidden mechanisms that generate the observed behavior, rather than on the structure of the underlying latent process (Netzer et al., 2017). For this reason, many of the HMM applications in marketing employ non-homogeneous HMM. In other words, they allow the transition probabilities to be function of time-varying variables, such as marketing activities (e.g. Paas, Vermunt, and Bijmolt, 2007; Netzer et al., 2008; Montoya et al., 2010; Luo and Kumar, 2013; Abhishek et al., 2016; Ebbes and Netzer, 2017). In this way, researchers can explore the factors which play a major role in generating the transition scheme.

In this thesis, we are concerned about both the structure of the latent process, and the underlying mechanisms generating the observed behavior. In fact, our first purpose is to identify the hidden states of the acquisition process, as there are no previous studies in literature dealing with this issue. Specifically, our purpose is to find empirical evidence about the three-stage acquisition process hypothesized in Chapter 3, according to which prospects start in an explorative stage, followed by trial and acquisition stages. To accomplish this task, we rely on multivariate longitudinal data (Bartolucci et al., 2012; Ebbes and Netzer, 2017), as our stages are supposed to be the underlying mechanisms driving observed variables regarding both search and purchase behaviors.

In addition, the second and third goals of this research are to analyze which factors affect the transition probabilities and which factors affect the search and purchase behaviors (such as the click intensity of each session, and the presence of absence of a transaction) in each state of the customer acquisition process. To accomplish these goals, we employ a non-homogeneous HMM, since we include such factors in the model in the form of time varying covariates, which are employed in the estimation of the transition probabilities and of the state dependent probabilities (e.g. Zucchini et al., 2016; Netzer et al., 2017). The introduction of time-varying covariates allow us to relax the Markovian assumption that the observed behavior in time *t* is based only on the previous behavior in time *t-1* (Netzer et al., 2017). Chapter 5 will provide a detailed description of both the response variables, and the time varying covariates for the transition and state dependent probabilities. The key benefit of HMM is that it allows us to analyze the customer acquisition process in a dynamic way, as prospects can change state and, in turn, search and purchase observed behaviors, over time.

4.2. Main components

As explained before, HMM assume that individuals have a probability to move among a finite number (S) of hidden states, which follow a Markov process. Since the researcher is unable to observe the state in which the individual is at time *t*, she has to translate the hidden states to an observed behavior Y_t by means of state dependent probabilities, representing the probability to observe the behavior Y_t, conditional to the state *s* membership. Thus, with HMM, the researcher has to estimate the hidden states, their number S, the transition probabilities among the states, and the state dependent probabilities (Netzer et al., 2017).

Since we are using multivariate longitudinal data, let $\mathbf{Y}_i^{(t)}$ be a vector of observations $\{Y_{i1}^{(t)}, Y_{i2}^{(t)} \ldots, Y_{iJ}^{(t)}\}$ for a set of J observed behaviors j = (1, …, J) observed among *N* individuals *i* = (1, …, N) at any point in time *t* = (1, …, T). The basic assumption of HMM is that the vector of observed behaviors $\mathbf{Y}_i^{(t)}$ is a noisy measure of a set of a finite number {1, …, S} of latent states $S_i^{(t)}$, which can only be inferred through the observed outcomes $\mathbf{Y}_i^{(t)}$. To capture the dynamic component, HMM assume that the state in time *t* depends on the state in which the individual is in time *t-1*, so that:

$$P\left(S_i^{(t-1)} \middle| S_i^{(t)}, S_i^{(t-1)}, \ldots, S_i^{(1)}\right) = \left(S_i^{(t-1)} \middle| S_i^{(t)}\right)$$

Accordingly, the joint likelihood of HMM can be written as follows:

$$P\left(\boldsymbol{Y}_i^{(t)}\right) = \sum_{s^{(1)}=1}^{S} P(S_i^{(1)} = s^{(1)}) \prod_{t=2}^{T} P(S_i^{(t)} = s^{(t)} | S_i^{(t-1)} = s^{(t-1)}) \prod_{t=1}^{T} P(\boldsymbol{Y}_i^{(t)} | S_i^{(t)} = s^{(t)})$$

Where $P(\boldsymbol{Y}_i|S_i = s) = \sum_{j=1}^{J} P\left(Y_{ij}\middle|S_i = s\right)$ (Bartolucci et al., 2012; Netzer et al., 2017).

In the HMM likelihood formula described above, we can distinguish three main components, which are the three main building blocks of a Hidden Markov Model:

- Initial State Distribution: $\pi = P\left(S_i^{(1)} = s^{(1)}\right)$, which is a 1×S vector representing the probability distribution of an individual being in each one of the states $s^{(1)}$=(1, …, S) at time $t=1$;
- Transition Probability Matrix:$\boldsymbol{Q}_i^{(t)} = P(S_i^{(t)} = s^{(t)}|S_i^{(t-1)} = s^{(t-1)})$, which is a S×S matrix containing the individuals' probabilities to switch from one state to another;
- State Dependent Probabilities: $\boldsymbol{M}_i^{(t)} = P(\boldsymbol{Y}_i^{(t)}|S_i^{(t)} = s^{(t)})$, which is a S×S diagonal matrix containing the probabilities to observe $\boldsymbol{Y}_i^{(t)}$ given the state in which the individual i is at time t: $\boldsymbol{m}_{is}^{(t)} = \prod_{j=1}^{J} P(Y_{ij}^{(t)}|S_i^{(t)} = s^{(t)})$ (Netzer et al., 2017; Ebbes and Netzer, 2017).

Thus, following Zucchini et al. (2016), the likelihood function of a HMM can be written in the following matrix form:

$$L_i^{(T)} = P\left(\boldsymbol{Y}_i^{(t)}\right) = \pi \boldsymbol{M}_i^{(1)} \boldsymbol{Q}_i^{(2)} \boldsymbol{M}_i^{(2)} \ldots \boldsymbol{Q}_i^{(T)} \boldsymbol{M}_i^{(T)} \iota$$

Where ι is a S×1 vector of ones, and the matrixes $\boldsymbol{M}_i^{(t)}$ capture the individuals' behavior at any point in time t (Netzer et al., 2017; Ebbes and Netzer, 2017).

4.2.1. Initial State Distribution

The initial state distribution $\pi = \{\pi_1, \pi_2, \ldots, \pi_S\}$ is a 1×S vector representing the probability that an individual is in each one of the hidden states at the beginning of the observation period $t=1$. Putting it formally:

$$\pi_{is} = P\left(S_i^{(1)} = s^{(1)}\right), \qquad \text{with } s^{(1)} = 1, \ldots, S$$

The initial state probabilities can be function of a set of covariates affecting the individual's propensity to be in state *s* at time *t=1*. Following Ascarza et al. (2018) and Ebbes and Netzer (2017), we can estimate the S parameters of π_s by using a multinomial logit model:

$$\pi_{is} = P\left(S_i^{(1)} = s|\boldsymbol{X}_i^{(1)} = \boldsymbol{x}_i^{(1)}\right) = \begin{cases} \dfrac{1}{1 + \sum_{s=1}^{S} exp\left\{\boldsymbol{X}_i'^{(1)}\tau_s\right\}} & if\ s = 1 \\ \dfrac{exp\left\{\boldsymbol{X}_i'^{(1)}\tau_s\right\}}{1 + \sum_{s=1}^{S} exp\left\{\boldsymbol{X}_i'^{(1)}\tau_s\right\}} & if\ s = 2, \ldots, S \end{cases}$$

Where $\mathbf{X}_i^{(i)}$ is a vector of covariates $\{X_{i1}^{(1)}, X_{i2}^{(1)} \ldots, X_{iK}^{(1)}\}$ for a set of K covariates $k = (1, \ldots, K)$ which vary both across the *N* individuals $i = (1, \ldots, N)$ at time $t = 1$.

4.2.2. Transition Probability Matrix

The Transition Probability Matrix, $\boldsymbol{Q}_i^{(t)} = P(S_i^{(t)} = s^{(t)}|S_i^{(t-1)} = s^{(t-1)})$, is a S×S matrix containing the individual *i*'s conditional probabilities $q_{iss'}^{(t)}$ to move to state $s^{(t)}=s'$ at time *t* conditional on the individual *i*'s membership to state $s^{(t-1)}=s$ in the previous time *t-1*.

The matrix $\boldsymbol{Q}_i^{(t)}$ is represented as follows:

$$\boldsymbol{Q}_i^{(t)} = \begin{bmatrix} q_{i11}^{(t)} & q_{i12}^{(t)} & \cdots & q_{i1S}^{(t)} \\ q_{i21}^{(t)} & q_{i22}^{(t)} & \cdots & q_{i2S}^{(t)} \\ \vdots & \vdots & \ddots & \vdots \\ q_{iS1}^{(t)} & q_{iS2}^{(t)} & \cdots & q_{iSS}^{(t)} \end{bmatrix}$$

The matrix has on the diagonal the probabilities to be in time *t* in the same state *s* as in the previous time *t-1*. In other words, the diagonal represents the probabilities that an individual does not move to any other state in the time *t*. For example, the first element of the diagonal is the probability to be in state 1 at time *t* conditional to the state 1's membership in time *t-1*: $q_{i11}^{(t)} = P(S_i^{(t)} = 1|S_i^{(t-1)} = 1)$.

Outside the diagonal, the probability $q_{i12}^{(t)} = P(S_i^{(t)} = 2|S_i^{(t-1)} = 1)$ is the probability to switch from state *s=1* to state *s'=2*, $q_{i1S}^{(t)} = P(S_i^{(t)} = S|S_i^{(t-1)} = 1)$ represents the probability to move from state *s=1* to state *s'=S*.

Usually the states are considered as independent from the time, as well as their transition probabilities. However, in marketing studies, the states are often bounded to the customers' activities, and can change over time (Netzer et al., 2017). In such case, a non-homogeneous HMM has to be employed, which allows the transition probabilities to be a function of time-varying covariates, such as the firm's marketing actions, or the customers' activities (e.g. Paas et al., 2007; Netzer et al., 2008; Montoya et al., 2010; Abhishek et al., 2016). The time-varying covariates included in the specification of the transition probabilities are usually believed to have a long-term effect on the outcome variables (Netzer et al., 2017).

The specification of the transition probabilities $q_{iss'}^{(t)}$ takes the multinomial logit form:

$$q_{iss'}^{(t)} = P(S_i^{(t)} = s'|S_i^{(t-1)} = s, \boldsymbol{X}_i^{(t)} = \boldsymbol{x}_i^{(t)}) = \frac{exp\left\{\boldsymbol{x'}_i^{(t)}\beta_{1ss'}\right\}}{1 + \sum_{s'=1}^{S} exp\left\{\boldsymbol{x'}_i^{(t)}\beta_{1ss'}\right\}} \qquad if\ s \neq s'$$

$$q_{iss'}^{(t)} = P(S_i^{(t)} = s'|S_i^{(t-1)} = s, \boldsymbol{X}_i^{(t)} = \boldsymbol{x}_i^{(t)}) = \frac{1}{1 + \sum_{s'=1}^{S} exp\left\{\boldsymbol{x'}_i^{(t)}\beta_{1ss'}\right\}} \qquad if\ s = s'$$

Where $\mathbf{X}_i^{(t)}$ is a vector of covariates $\{X_{i1}^{(t)}, X_{i2}^{(t)} \ldots, X_{iK}^{(t)}\}$ for a set of K time-varying covariates *k* = (1, …, K) which vary both across the *N* individuals *i* = (1, …, N) and over time *t* = (1, …, T).

Since the number of the transition matrix parameters to be estimated increases dramatically in increasing the number of states (Bartolucci et al., 2012), in order to decrease the computational complexity, it is good practice to impose some restrictions to the transition matrix. For example, in the current study, we assume the last state s_S (acquisition) to be absorbing, so that people who enter in s_S are not allowed to move back to the other states (e.g. Abhishek et al., 2012; Schweidel et al., 2011; Schwartz et al., 2011). Given this assumption, the transition probabilities can be written as follows:

$$q_{iss'}^{(t)} = \begin{cases} \dfrac{exp\left\{\boldsymbol{x'}_i^{(t)}\beta_{1ss'}\right\}}{1+\sum_{s'=1}^{S} exp\left\{\boldsymbol{x'}_i^{(t)}\beta_{1ss'}\right\}} & if\ s \neq s' \\ \dfrac{1}{1+\sum_{s'=1}^{S} exp\left\{\boldsymbol{x'}_i^{(t)}\beta_{1ss'}\right\}} & if\ s = s' \\ 0 & if\ s = s_S\ and\ s \neq s' \\ 1 & if\ s = s' = s_S \end{cases}$$

Resulting in the following transition matrix:

$$\boldsymbol{Q}_i^{(t)} = \begin{bmatrix} q_{i11}^{(t)} & q_{i12}^{(t)} & \cdots & q_{i1S}^{(t)} \\ q_{i21}^{(t)} & q_{i22}^{(t)} & \cdots & q_{i2S}^{(t)} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

4.2.3. State Dependent Probabilities

The State Dependent Probability matrix, $\boldsymbol{M}_i^{(t)} = P(\boldsymbol{Y}_i^{(t)}|S_i^{(t)} = s^{(t)})$, is a S×S diagonal matrix containing the probabilities $\boldsymbol{m}_{is}^{(t)}$ to observe $\boldsymbol{Y}_i^{(t)}$ given the state in which the individual *i* is at time *t* (Netzer et al., 2017; Ebbes and Netzer, 2017). If the individual's state membership is known, conditionally on it, the probability $P\left(\boldsymbol{Y}_i^{(t)}\middle|S_i^{(t)} = s^{(t)}\right)$ is independent over time.

The matrix $\boldsymbol{M}_i^{(t)}$ is represented as follows:

$$\boldsymbol{M}_i^{(t)} = \begin{bmatrix} \boldsymbol{m}_{i1}^{(t)} & 0 & \cdots & 0 \\ 0 & \boldsymbol{m}_{i2}^{(t)} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \boldsymbol{m}_{iS}^{(t)} \end{bmatrix}$$

With the elements on the diagonal $\boldsymbol{m}_{is}^{(t)} = \prod_{j=1}^{J} P(Y_{ij}^{(t)}|S_i^{(t)} = s^{(t)})$, as we are in the multivariate case.

The state dependent probability matrix is the most flexible part of HMM, as it allows to choose the most suited distribution to fit the nature of the outcome variable, $\boldsymbol{Y}_i^{(t)}$. For example, one can use a normal distribution to account for continuous outcomes (Ebbes et al., 2010), or a logit or probit distribution for binary outcomes (e.g. Netzer et al., 2008; Schweidel et al., 2011).

Concerning this work, as we are dealing with a multivariate outcome, the state dependent distribution will be a combination of the single outcomes distributions (Ebbes and Netzer, 2017). This is in line with previous work, as, for instance, Abhishek et al. (2016) combine a probit and a logit model to analyze a multivariate outcome represented by the count variable, number of page views, and the binary outcome, conversion. Since we have both continuous and binary outcomes, we can split the J outcomes in two groups: the first containing the J_1 continuous outcomes, and the second containing the J_2 binary outcomes, so that: $J=J_1+J_2$. Accordingly, the state dependence probabilities will take the form:

$$\boldsymbol{m}_{is}^{(t)} = P(\boldsymbol{Y}_i^{(t)}|S_i^{(t)} = s^{(t)}) = (\prod_{j=1}^{J_1} f(Y_{ij}^{(t)}|S_i^{(t)} = s^{(t)}) \times \prod_{j=J_1+1}^{J} P(Y_{ij}^{(t)}|S_i^{(t)} = s^{(t)}))$$

With $s^{(t)} = \{1, 2, \ldots, S\}$, where the first component represents the continuous outcomes and the second component represents the binary outcomes.

The state dependent probabilities, as already seen for the transition probabilities, can be a function of time-varying covariates. The time-varying covariates included in the specification of the state dependent probabilities are usually believed to have a short-term effect on the outcome variables. This is due to the fact that, conditionally on the individual's state membership, the state dependent probability is independent over time, thus the covariates have an effect only on the current time (Netzer et al., 2017). Incorporating the covariates in the state dependence probabilities will result in:

$$\boldsymbol{m}_{is}^{(t)} = P(\boldsymbol{Y}_i^{(t)} | S_i^{(t)} = s^{(t)}, \boldsymbol{Z}_i^{(t)} = \boldsymbol{z}_i^{(t)})$$

$$= (\prod_{j=1}^{J_1} f\left(Y_{ij}^{(t)} \middle| S_i^{(t)} = s^{(t)}, \boldsymbol{Z}_i^{(t)} = \boldsymbol{z}_i^{(t)}\right)$$

$$\times \prod_{j=J_1+1}^{J} P(Y_{ij}^{(t)} | S_i^{(t)} = s^{(t)}, \boldsymbol{Z}_i^{(t)} = \boldsymbol{z}_i^{(t)}))$$

Where $\mathbf{Z}_i^{(t)}$ is a vector of covariates $\{Z_{i1}^{(t)}, Z_{i2}^{(t)} \ldots, Z_{iW}^{(t)}\}$ for a set of W time-varying covariates, $w = (1, \ldots, W)$, which vary both across the N individuals $i = (1, \ldots, N)$ and over time $t = (1, \ldots, T)$.

4.3. Unobserved Heterogeneity

In dealing with longitudinal data, researchers have to account for the fact that individuals in their data may behave differently from each other, and may have different preferences and hidden traits. This issue, namely the customer heterogeneity issue, is well known in marketing literature for a long time (e.g. Chintagunta, Jain, and Vilcassim, 1991; Wedel et al., 1999; Horsky, Misra, and Nelson, 2006; Blattberg et al., 2008; Zhang, Kumar, and Cosguner, 2017), especially in studies dealing with customer segmentation and targeting, where the heterogeneity between customers is at the very basis of these tasks (Netzer et al., 2017). Several studies demonstrate that not accounting for differences between individuals often leads to biased results (e.g. Heckman, 1981; Erdem and Sun, 2001; Kappe, Blank, and DeSarbo, 2018).

Heterogeneity can be both observed and unobserved. In the first case it is easy to take into account, as it can be simply treated by including observed covariates in the model. The unobserved heterogeneity is trickier to deal with, as researchers cannot rely on any observed information, thus they have to make inferences about the role of unobserved individual differences in the observed outcome.

HMM with longitudinal data can account for observed heterogeneity through the inclusion of observed individual covariates in the specification of the initial state, transition, and state dependence probability distributions. This allow the researcher to have a single initial state vector, transition matrix, and state dependence matrix specific to each individual (e.g. Paas et al., 2007; Netzer et al., 2008; Montoya et al., 2010; Netzer et al., 2012; Abhishek et al., 2016).

Since HMM belong to the same family of latent class models (Kamakura and Russell, 1989), HMM allow researchers to deal with unobserved heterogeneity through the use of discrete latent classes (Lazarsfeld, 1950; Wedel et al., 1999; Henry, 2004). In fact, the individual membership to the latent state reflects the effects of unobserved factors associated to the realization of the outcomes, $\boldsymbol{Y}_i^{(t)}$ (Bartolucci et al., 2012). The inclusion of observed covariates in the model is extremely useful in order to disentangle the effects of the observed heterogeneity from the effects of the unobserved heterogeneity on the outcomes of interest, conditional to the state membership (Bartolucci and Farcomeni, 2009). The key difference and advantage of using HMM over standard latent class models lies in the fact that HMM states are dynamic, meaning that individuals can switch from one state to another, while latent class states are static. This implies that, with HMM, researchers can account for unobserved heterogeneity in a dynamic way, allowing customers to change their preferences and hidden traits over time (Bartolucci and Farcomeni, 2009; Bartolucci et al., 2012).

4.4. Model Estimation

In order to estimate a HMM, one of the most commonly used methods is the implementation of the Expectation-Maximization (EM) Algorithm, or Baum-Welch algorithm (Baum et al., 1970; Baum, 1972; Dempster, Laird, and Rubin, 1977; Welch, 2003). This iterative method requires the calculation of forward and backward probabilities. Accordingly, the first part of this section describes the concepts of forward and backward probabilities, in order to introduce the second part of the section, which explains the EM algorithm underlying logic.

4.4.1. Forward and Backward Probabilities

Following Zucchini et al. (2016), let's consider *t*, a specific moment in time during the observation period, {1, …, *t*, …, T}. We then observe realizations of the observed outcomes $\boldsymbol{Y}_i^{(g)}$ both before (if *g* = 1, …, *t*) and after (if *g* = *t*+1, …, *T*) the considered time *t*.

The *forward probabilities* represent the 1×S vector of joint probabilities $\alpha_i^{(t)} = P(Y_i^{(1)}, \dots, Y_i^{(t)}, S_i^{(t)})$ up to time *t*. The vector of forward probabilities is computed as follows:

$$\alpha_i^{(t)} = \pi M_i^{(1)} \prod_{g=1}^{t} Q_i^{(g)} M_i^{(g)}$$

with *t* = (1, …, *T*). $\alpha_i^{(t)}$ considers the first (1, …, *t*) terms of the vector of the observed outcomes $\boldsymbol{Y}_i^{(t)}$. We can then derive the likelihood formulation up to time *T*:

$$L_i^{(t)} = \alpha_i^{(T)} \iota$$

where:

$$\alpha_i^{(t)} = \begin{cases} \alpha_i^{(t-1)} Q_i^{(t)} M_i^{(t)} & if\ t > 2 \\ \pi M_i^{(1)} & if\ t = 1 \end{cases}$$

The *backward probabilities* represent the 1×S vector of conditional probabilities $B_i^{(t)} = P(Y_i^{(t+1)}, \dots, Y_i^{(T)}, S_i^{(t)})$ after time *t*. The vector of backward probabilities is computed as follows:

$$B'^{(t)}_i = \left(\prod_{g=t+1}^{T} Q_i^{(g)} M_i^{(g)} \right) \iota$$

with $t = (1, \ldots, T)$. $\beta'^{(t)}_i$ considers the last $(t{+}1, \ldots, T)$ terms of the observed outcomes vector $\boldsymbol{Y}_i^{(t)}$. It is worthy to note that, if $t = T$, $B_i^{(T)} = \iota$.

The product of forward and backward probabilities gives the probability of observing the vector $\boldsymbol{Y}_i^{(t)}$:

$$\alpha_i^{(t)}(j) B_i^{(t)}(j) = P\left(\boldsymbol{Y}_i^{(t)}, S_i^{(t)} = j\right) = \sum_{j=1}^{S} \alpha_i^{(t)}(j) B_i^{(t)}(j) = \sum_{j=1}^{S} P\left(\boldsymbol{Y}_i^{(t)}, S_i^{(t)} = j\right) = P\left(\boldsymbol{Y}_i^{(t)}\right)$$

where $\alpha_i^{(t)}(j)$ and $B_i^{(t)}(j)$ are, respectively, the j^{th} elements of the vectors $\alpha_i^{(t)}$ and $B_i^{(t)}$. Accordingly, when $t = T$, the forward and backward probabilities can be used to compute the complete likelihood function, as seen in the section 4.2 (Netzer et al., 2012):

$$L_i^{(T)} = P\left(\boldsymbol{Y}_i^{(t)}\right) = \pi \boldsymbol{M}_i^{(1)} \prod_{g=1}^{T} Q_i^{(g)} M_i^{(g)} \iota = \alpha_i^{(T)} B_i^{(T)}$$

4.4.2. EM Algorithm

The "key idea" of the EM algorithm (Little and Rubin, 2002) resides in considering the hidden state membership as a missing data (Zucchini et al., 2012; Netzer et al., 2017). The EM algorithm consists in a series of iterations aiming at maximizing the HMM likelihood, following two main steps:

- The **E-step** computes the missing data's conditional expectations, given the observed outcomes. In this way, one can complete the data, without the missing values represented by the hidden state membership;

- The **M-step** maximizes the log-likelihood of the complete data obtained in the previous step

In this way, considering the outcome realizations $\boldsymbol{y}_i = (\boldsymbol{y}_i^{(1)}, \dots, \boldsymbol{y}_i^{(T)})$, and the estimated state membership $s_i = (s_i^{(1)}, \dots, s_i^{(T)})$, the log-likelihood function of the complete data (CDLL) at any point in time $t = (1, \dots, T)$ for all the individuals $i = (1, \dots, N)$ can be written as:

$$CDLL = \sum_{i=1}^{N} log\big(P(\boldsymbol{y}_i, s_i)\big) = \sum_{i=1}^{N}\left(\log\left(\pi_{s_i{}^{(1)}}\right) + \sum_{t=2}^{T} log\left(q^{(t)}_{is_i{}^{(t-1)}s_i{}^{(t)}}\right) + \sum_{t=1}^{T} log\left(m^{(t)}_{is_i{}^{(t)}}\right)\right)$$

Let's consider $v_i^{(t)} = \left(v_{i1}^{(t)}, \dots, v_{iS}^{(t)}\right)$, a S×1 vector of dummies indicating the state membership of the i^{th} individual at time t (for example, if the i^{th} individual at time t is in state 2, $v_i^{(t)}$ will take the form (0, 1, 0, …, 0)), and $W_i^{(t)}$, a S×S matrix of dummies indicating the transition among states of the i^{th} individual between time $t-1$ and time t. The CDLL can be re-written as follows:

$$CDLL = \sum_{i=1}^{N}\left(v'^{(1)}_i \log(\pi) + \sum_{t=2}^{T} \iota'\left(W_i^{(t)} \circ log\left(Q_i^{(t)}\right)\right)\iota + \sum_{t=1}^{T} v'^{(1)}_i log\left(M_i^{(t)}\right)\right)$$

$$= \sum_{i=1}^{N}\left(v'^{(1)}_i \tilde{\pi} + \sum_{t=2}^{T} \iota'\left(W_i^{(t)} \circ \tilde{Q}_i^{(t)}\right)\iota + \sum_{t=1}^{T} v'^{(1)}_i\, \widetilde{m}_i^{(t)}\right)$$

The CDLL is composed by the sum of three terms, that reflect the three main components of HMM: the first, $v'^{(1)}_i \tilde{\pi}$, regards the initial state distribution, the second, $\sum_{t=2}^{T} \iota'\left(W_i^{(t)} \circ \tilde{Q}_i^{(t)}\right)\iota$, is about the transition probability matrix, and the last, $\sum_{t=1}^{T} v'^{(1)}_i\, \widetilde{m}_i^{(t)}$, the state dependencies. In order to maximize the CDLL, one can maximize these three terms in a separate way. The EM algorithm works as follows:

1) **E-step**: through the forward and backward probabilities, we can estimate the conditional expectations, which, in turn, are used to estimate $v_i^{(t)}$ and $W_i^{(t)}$:

$$\hat{v}_i^{(t)}(j) = P\left(S_i^{(t)} = j|\boldsymbol{y}_i^{(1)}, \dots, \boldsymbol{y}_i^{(T)}\right) = \frac{\alpha_i^{(t)}(j)B_i^{(t)}(j)}{L_i^{(T)}}$$

$$\widehat{W}_i^{(t)} = P\left(S_i^{(t-1)} = j, S_i^{(t)} = k|\boldsymbol{y}_i^{(1)}, \dots, \boldsymbol{y}_i^{(T)}\right) = \frac{\alpha_i^{(t-1)}(j)Q_i^{(t)}(j,k)P\left(\boldsymbol{y}_i^{(t)}|k\right)B_i^{(t)}(k)}{L_i^{(T)}}$$

with $j, k = (1, \dots, S)$. $\hat{v}_i^{(t)}(j)$ and $\widehat{W}_i^{(t)}$ are, respectively, the individual i's likelihood of being in state j and of shifting from state j to k at time t;

2) **M-step**: We replace $\hat{v}_i^{(t)}(j)$ and $\widehat{W}_i^{(t)}$ to $v_i^{(t)}$ and $W_i^{(t)}$ in the CDLL, and maximize the CDLL, to obtain new updated parameters;

3) Iteration of E-step and M-step until the log-likelihood with the updated parameters does not significantly improve.

4.5. Model Selection

In HMM, by definition the states are hidden and cannot be observed. This means that, unless one has a very strong knowledge about the process and has a theoretical ground strong enough to know the number of states *a priori* (e.g. Ansari, Montoya, and Netzer, 2012), in the vast majority of applications, the number of states S in the latent process has to be estimated (Bartolucci et al., 2012; Netzer et al., 2017).

The selection of the number of states is one of the first steps in a HMM application, and consists in a series of estimations, increasing the number of states by 1 at any estimation. The process stops when the model fit does not improve in adding a further state. To evaluate the fit in order to select the model with the best number of states, the most used tools are the information criterions proposed by Akaike (1973) and Schwartz (1978), namely the Akaike Information Criterion (AIC), and the Bayesian Information Criterion (BIC) (Bartolucci et al., 2012; Netzer et al., 2017). In order to calculate these information criterions, we first need to know the number of parameters of a HMM, which are (S-1) in the initial state distribution vector, π; S×(S-1) in the transition matrix, $\boldsymbol{Q}_i^{(t)}$; and S in the state dependence matrix, $\boldsymbol{M}_i^{(t)}$. Putting it altogether, the number of parameters, p, is:

$$p = (S + 1)(S - 1) + S$$

Then, the AIC and BIC specifications are expressed as follows:

$$AIC = -2 \times ll + 2 \times p$$

$$BIC = -2 \times ll + \log(n) \times p$$

Where ll is the log-likelihood of the model, and n is the number of observations (N×T in the case of a balanced panel with N individuals observed for T time periods). The best model to select is the one with the lower AIC and BIC values.

The advantage of selecting the model by using AIC and BIC as selection criterions over the most common log-likelihood lies in the fact that the latter improves always in adding a hidden state, as the flexibility of the model increases. Information criterions, on the other hand, add a penalty to the model fit in relation to the number of parameters. In other words, AIC and BIC help in selecting the most parsimonious model which fits the data best.

Among the two criterions, BIC has been found to be more trustworthy, as its penalization for the parameters is more severe, while AIC tends to select models with more states than necessary (McLachlan and Peel, 2000; Dias, 2006; Bartolucci, Farcomeni, and Pennoni, 2014).

5. EMPIRICAL STUDY

In this chapter, we present the empirical study, which is the core of the present dissertation. It starts providing an overview of the empirical setting. The following section briefly describes the data and the different datasets provided by the firm, with an explanation of the process employed to join all the different sources of data and the creation of the final dataset.

The third section provides a detailed description of the final dataset, by distinguishing among different sets of variables: the outcomes of the model, the variables employed as covariates for the state dependence probabilities, and the covariates for the transition probabilities. The section ends showing the correlation matrixes among the different sets of variables.

The fourth section describes the HMM estimation, by presenting the LMest R package employed to estimate the model, and formally adding the covariates described previously in the methodological framework provided in Chapter 4.

The chapter ends with the presentation of the results. First, it presents the five identified states, the initial and transition probabilities and the effect of the covariates on the individual's likelihood to be in a state and to shift to any other state, and the effects of the covariates, such as the touchpoints and the search behavior, on the state dependent probabilities. The chapter ends with a description of the most frequent paths to acquisition.

5.1. Setting

Data come from an Italian multi-brand e-retailer who operates worldwide. It sells clothing, apparels, and design objects only digitally. The choice of this particular setting provides several advantages: first, since the environment is completely digital, it is possible to identify the very first transaction for each customer. The absence any physical store ensures that not only the first transaction, but also all the interactions occurring before and after the first purchase are recorded. Second, the company is a multi-brand retailer that sells several product categories. This ensures that the results will not be centered on a specific product category or a specific brand. Third, the relationship between the firm and its customers is non-contractual, thus customers are not bounded to the firm by a contract for a fixed period of time. As stated before, studies on customer acquisition in a non-contractual settings are scarce, and, as in this study, we actually question the concept and definition of acquisition we believe that these data enable us to test for different specifications of this construct. Fourth, the company sells mainly off-seasonal products, making it more difficult to find them in physical stores or in other e-retailers.

5.2. Dataset Creation

There are two different sources of data: Google Analytic Platform (GAP) for the clickstream data, and the firm's databases for the users, newsletter, and transactional data.

5.2.1. Firm's Databases

User Data. The User database contains demographic, registration, and first purchase information about 8,257,803 users who registered to the website, sign in the reception of the newsletter, or made a purchase between June 20th, 2000 and October 13th, 2017. Among them, 3,015,261 (36.51%) users registered to the website and never purchased. User Data does not contain any information about users who just visit the website without making any of the three activities listed above.

The two types of registration are independent one of another. In other words, people can register to the website without subscribing to the newsletter program, they can just register to receive the newsletter or they can sign up for both. Notably, they can unsubscribe to the newsletter reception at any time.

TABLE 5.1.1: Users by categories

User Type	Description	Prospect period	Newsletter	Wish List	Special Promotions	Proportion in the User Data
Direct purchase	Customers who purchase without any type of account. They can subsequently decide to register and/or subscribe to the newsletter.	No	No	No	No	18.57%
Newsletter	Users who opt-in to the newsletter. They can purchase and subsequently become full users, but not Rep customers.	Yes	Yes	No	Yes	14.09%
Full registration	Users who make a full registration to the website. They can also register as fast users. They cannot be Rep customers in the future.	Yes	Yes	Yes	Yes	67.34%

According to the so-called User Data of the company, 67.34% have a full registration, while 14.09% are users who only subscribes to the newsletter, with a 3.10% conversion rate. Registering to the website offers the prospects the opportunity to put their preferred items in the Wish List, and save them for a future order. It is important to remark that registering is not a necessary requirement to place an order. Prospects and Customers can make purchases without being registered to either the website or the newsletter, if this is the case their transaction will be hereafter defined as “direct purchase”. These types of customers account for 18.57% of the user database. Unlike newsletter subscribers and registered users, customers who make a purchase without being registered do not receive any special promotions, as promotional codes sent by e-mail to apply at the time of the checkout in order to receive a discount. Table 5.1.1 provides a summary of the different types of users.

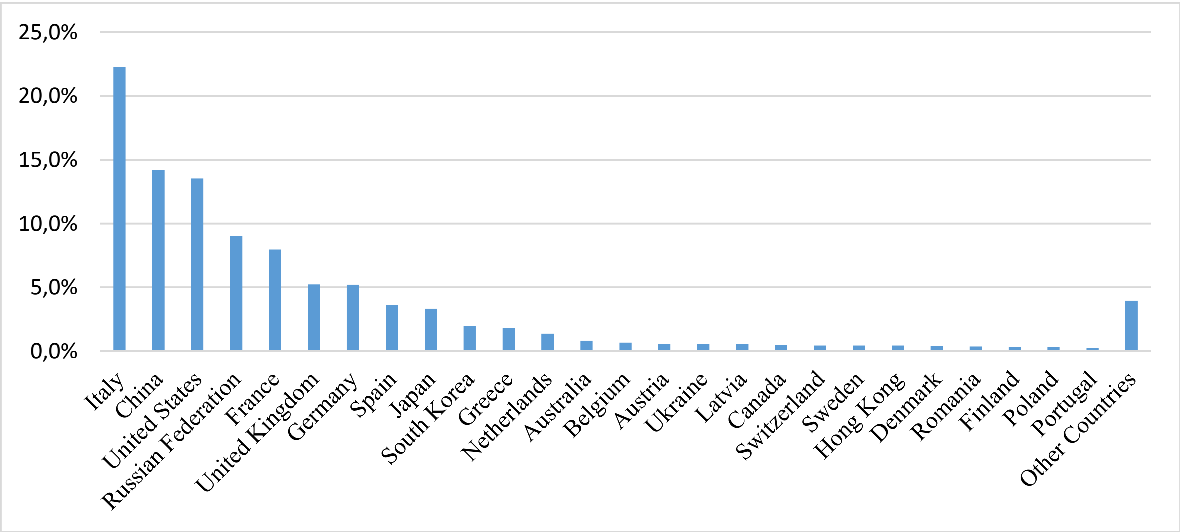
Almost half of the users in the database (49.40%) made at least a purchase. 37.59% of them (18.57% of all users) never registered to the website, 42.59% registered only after the first purchase, and the remaining 19.82% registered before or, at least, the same day of their first purchase (together, the last two conditions represent the 30.83% of the total users) [Table 5.2.1].

TABLE 5.2.1: Registration and purchase behavior

	Purchase	No Purchase	Total
Registered	2,545,474 (30.83%)	3,015,261 (36.51%)	5,560,735 (63.34%)
Not Registered	1,533,439 (18.57%)	1,163,629 (14.10%)	2,697,068 (32.66%)
Total	4,078,913 (49.40%)	4,178,890 (50.61%)	8,257,803 (100.00%)

The vast majority of users are women (70.16%), 28.90% are men, and the remaining 0.94% did not provide gender information. The average age is 38 years (36.18% of users did not provide age information).

FIGURE 5.2.1: Distribution of countries



They come from 109 countries, mainly from Italy (22.28%), China (14.18%), and USA (13.53%). Figure 5.2.1 represents the distribution of countries in the Users' database.

Newsletter Data. The Newsletter database contains information about 1,003,914,949 emails sent by the firm from December 13th, 2013 to May 5th, 2017. Information are about the number of times the email has been opened, the number of times the user clicked on it in order to follow the link contained, the delivery date, the first and last opening dates, and the first and last click dates. Unfortunately, it does not contain any information about the content of the email. Every email has been opened on average 0.26 times (SD = 30.48), ranging from a minimum of 0 times to a maximum of 99,374 times. After opening a newsletter, users click on the embedded link on average 0.26 times (SD = 2.09), from a minimum of 0 to a maximum of 3515 times.

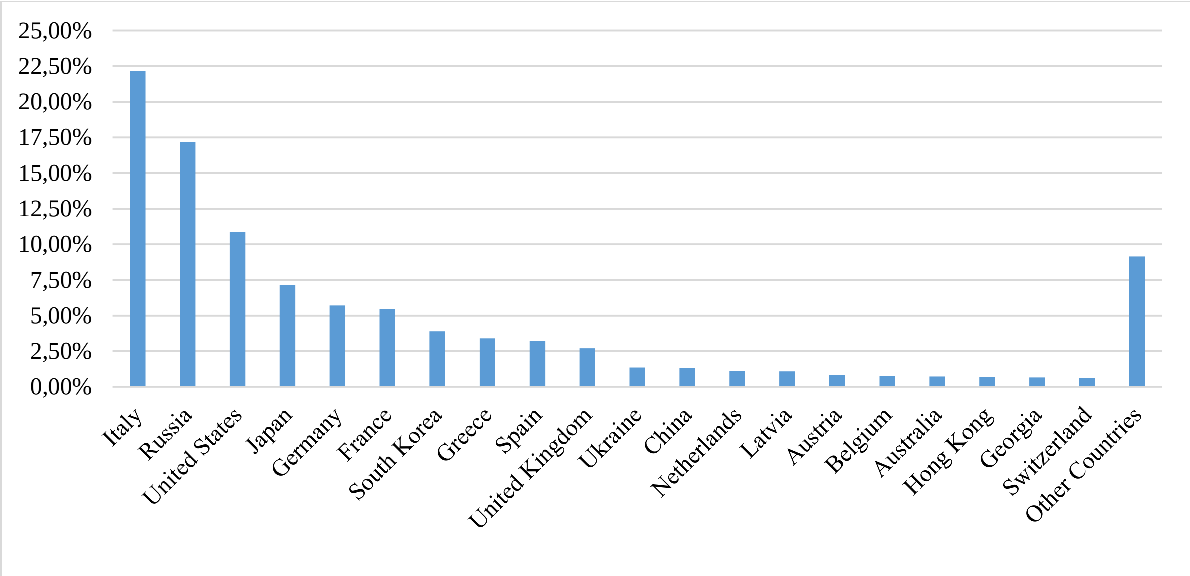
Transactional Data. The transactional database contains information (e.g. monetary value, discount, payment type, returns, etc.) about 24,035,515 orders made by 5,973,872 users between June 20th, 2000 and November 30th, 2017. The average monetary value of the transactions is 179.88€ (SD = 47495.26), ranging from a minimum of 0 to a maximum of 22452.34€.

5.2.2. Google Analytics Platform (GAP) 2017 Database

The firm initially provided us data coming from Google Analytics Platform (GAP), containing all the 1,340,108,261 clicks occurred in the website all over the world from January 1st to May 5th, 2017.

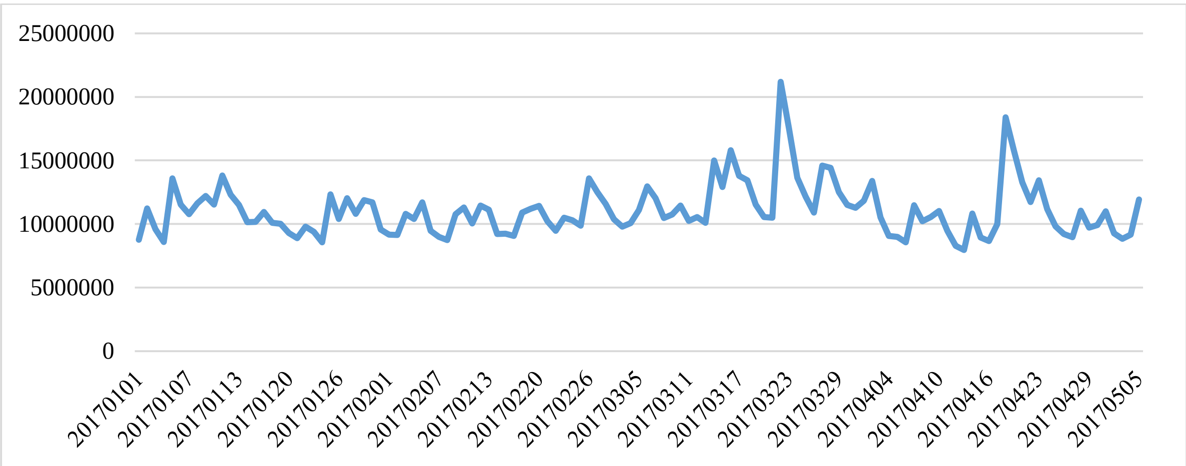
Figure 5.2.2 depicts the percentages of clicks distinct by the different countries. Clicks come from 241 countries, mainly from Italy (22.14%), Russia (17.17%), and USA (10.89%).

FIGURE 5.2.2: Clicks by countries



As shown in Figure 5.2.3, there is no time trend in the number of clicks, even though there are a couple of positive peaks in late March and April.

FIGURE 5.2.3: Clicks per day



In a more aggregate form, GAP data can be analyzed in terms of sessions. The database contains 72,603,038 sessions, generated by desktop (45.01%), mobile (42.92%), and tablet (12.07%) activity. The average session length is around 410.52 seconds, with 9.15 pages seen, and 18.46 clicks. For each Google ID, the average number of days that span between two subsequent sessions is 6.37 [Table 5.2.2]. The couples of subsequent sessions with more than 5 months between them are 0.2%.

TABLE 5.2.2: Sessions descriptive statistics

	Mean	**SD**	**Min**	**Max**
Length (ms)	410516.37	1036346.85	0	86386512
Pages	9.15	20.16	0	500
Clicks	18.46	38.50	1	500
Intersession days	6.37	18.22	0	269

User Identification. GAP database identifies users in two different ways:

- Google ID: Digital firms often rely on cookies in order to recognize their users, but this practice fails to work when the user decides to clean her cookies history (Kannan et al., 2016). Moreover, cookies are ineffective in the multi-device world, as they are device-specific. and cannot track users' cross-device behavior. Google IDs are based on this type of cookies. The initial data comprise 2,7180,693 Google IDs;
- Firm ID: Our firm tries to face the issue of the multi-device usage by creating its own cookie, which accounts not only for the classical cookies explained above, but also for the user's e-mail address. The initial data comprise 14,871,933 Firm IDs. It should be able to identify the same user across different devices as long as the user provided at least once an email address. The implementation of the Firm ID tool was a pilot: the firm intended to assess the viability of this approach to monitor cross-device search activities in a more precise fashion with respect to what could be done by using just traditional cookies, and to develop a unique ID able to link all the website visits and the transactions to the same user. Therefore, only a portion of the sessions could be uniquely related to a user, as 45.66% (33,150,728) of all the sessions do not have a Firm ID. The unique identification of the user is indeed not a trivial task. The two IDs concatenate

one with another, so that a Google ID can be associated to more than one Firm ID, and *vice-versa*.

Since the employment of the Firm ID to identify the same users across the different devices works only with registered users, as they use their email addresses in order to log-in the website from each device, the user identification follows the following steps, which are graphically represented in Table 5.2.3:

1) Step 1: Creation of a variable containing all the Firm IDs which are associated to at least one session with a login (the green squares in the table);
2) Step 2: Spreading of the Firm IDs identified in Step 1 to all the sessions with a Google ID associated to these Firm ID (the orange squares in the table);
3) Step 3: Spreading of the Firm IDs identified in Step 1 to all the sessions associated to a Firm ID which is concatenated to the Google IDs identified in Step 2 (the yellow squares in the table);
4) Step 4: Repeat steps 2-3 until a further step does not produce any changes.

TABLE 5.2.3: Steps for User Identification

Session	Google ID	Firm ID	Login	Step 1	Step 2	Step 3
1	A	1	No	0	0	2
2	B	1	No	0	2	2
3	B	2	Yes	2	2	2
4	C	2	No	2	2	2
5	D	2	No	2	2	2
6	E	2	No	2	2	2
7	E	3	No	0	2	2

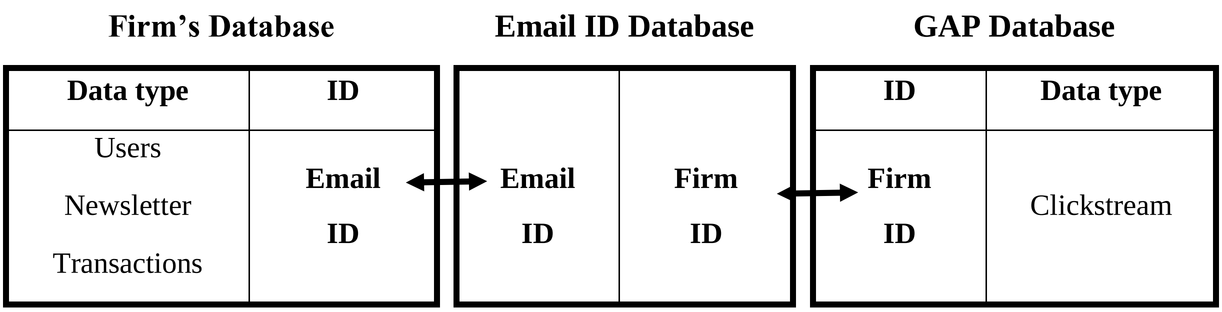
Some problems arise when a Google ID is associated with more than one login event having different Firm IDs, meaning that multiple users shared the same device. Since it is almost impossible to disentangle precisely the activity of each user, we drop these cases from the database, losing around 29.83% (4436239) of the initial Firm IDs.

After cleaning the database from these critical cases, and from people who never registered to the website (54% of the total clicks), as they did not provide any e-mail address, and do not log-in the website, making impossible their identification cross-device, the implementation of the steps 1-4 described above identified 610,848 Firm IDs (the final User IDs), who made 285,250,815 clicks in 7,085,525 sessions (40.26 clicks per session on average).

5.2.3. Merging the data

In order to join the GAP database with the other Firm's databases (Users, Newsletter, Transactions), we need to associate the Firm ID with the corresponding encoded email. The two sources of data can be matched through a third Email ID database, containing the encrypted email addresses of the users for the firm's data, and the correspondent Firm ID for the GAP data [Figure 5.2.4].

FIGURE 5.2.4: Matching of the databases

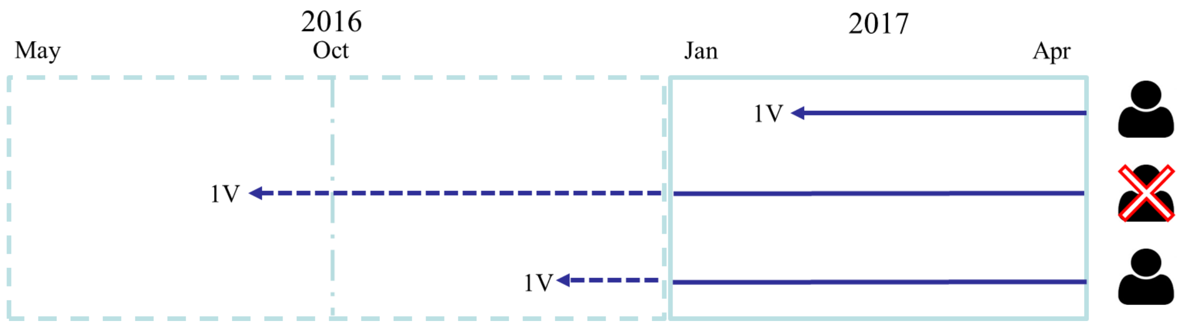


As mentioned before, the use of Firm IDs was a pilot with some fallacies, so that the firm decided to stop implementing it. Therefore, it did not apply to the entire database, resulting in only 15,103 out of 127,574 Email IDs of the Firm's Databases (11.84%) having a correspondence in the GAP database. The filters used to select the users in the final dataset are described in detail in the section 5.2.5.

5.2.4. GAP 2016 Database

After we created our database containing the 15,103 users mentioned above, the firm gave us the opportunity to integrate the 2017 clickstream data with another GAP database containing the clickstream data from May to December 2016. Since the database was too big to handle all in once, we split it in two: from May to September 30th, and from October 1st to December 31st. The reason underlying this decision to split the data before and after October is explained in Figure 5.2.5. As mentioned before, only 0.2% of sessions have an intersession time higher than 5 months. Accordingly, we assume users who do not make any session between May 1st and September 30th as new users, while people who visit the website in this time window cannot be assumed to be new and do not enter in the final database (the second user in Figure 5.2.5).

FIGURE 5.2.5: User selection timeline



1V = First Visit

To summarize, we select users in the following way:

1) We kept all the users who made at least one registered session in 2017 (the continuous-line square in Figure 5.2.5);
2) In GAP 2016 database (the dotted-line square in figure 5.2.5) containing the clickstream data after October 1st (GAP 2016_2), we kept all the users who appear in GAP 2017 database obtained in Step 1;

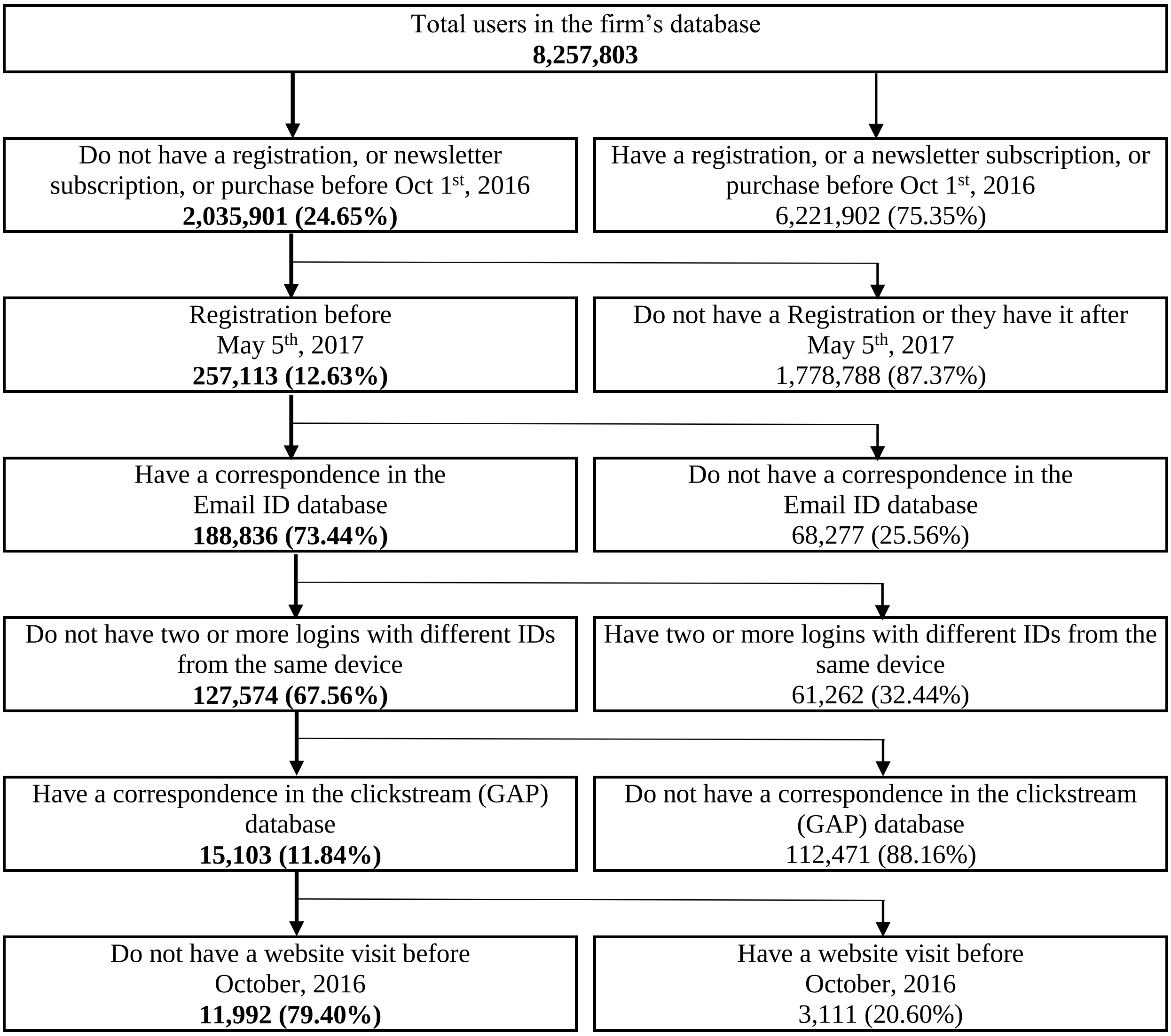
3) We checked whether GAP 2017 users identified in Step 1 had a session in GAP 2016 database containing the clickstream data before October 1st (GAP 2016_1). If so, they have been dropped from both GAP 2017 and GAP 2016_2.

Thus, the only use of GAP 2016_1 data is to identify users who should be dropped from the final dataset, since we cannot be sure they were new users of the website at the time we see them for the first time. On the other hand, we analyzed GAP 2016_2 dataset in order to extrapolate the sessions of interest, which contains all the 1,837,701,710 clicks (96,721,143 sessions) occurred in the website all over the world.

5.2.5. Filters Summary

The initial User data contains 8,257,803 users who registered to the website, sign in the reception of the newsletter, or made a purchase between June 20th, 2000 and October 13th, 2017. Since we want to identify users who start their interaction with the firm after October 2016, we first keep only those who do not have a registration, or newsletter subscription, or purchase before October 1st, 2016 (24.65% of the users in the initial User data). Then, since GAP data are available until May 5th, 2017 and we consider only users who were registered in that period, we drop all the users who registered to the website after May 5th, 2017 (87.37%). Of the remaining 257,113 users, 188,836 (73.44%) have their Email ID in the ID database. Due to identification issues already explained in section 5.2.2 (User identification), we drop users who have more than one Firm ID associated to the same device, keeping 127,574 users. In merging these users with the GAP database, we lost 88.16% of them. Only 15,103 Email IDs were successfully matched with a Firm ID in the GAP database. Finally, we checked whether these users had visited the website before October 1st, 2016, keeping the final pool of 11,992 users who did not have any [Figure 5.2.6].

FIGURE 5.2.6: Filters for the users' selection



5.3. Final Dataset Description

5.3.1. Overview

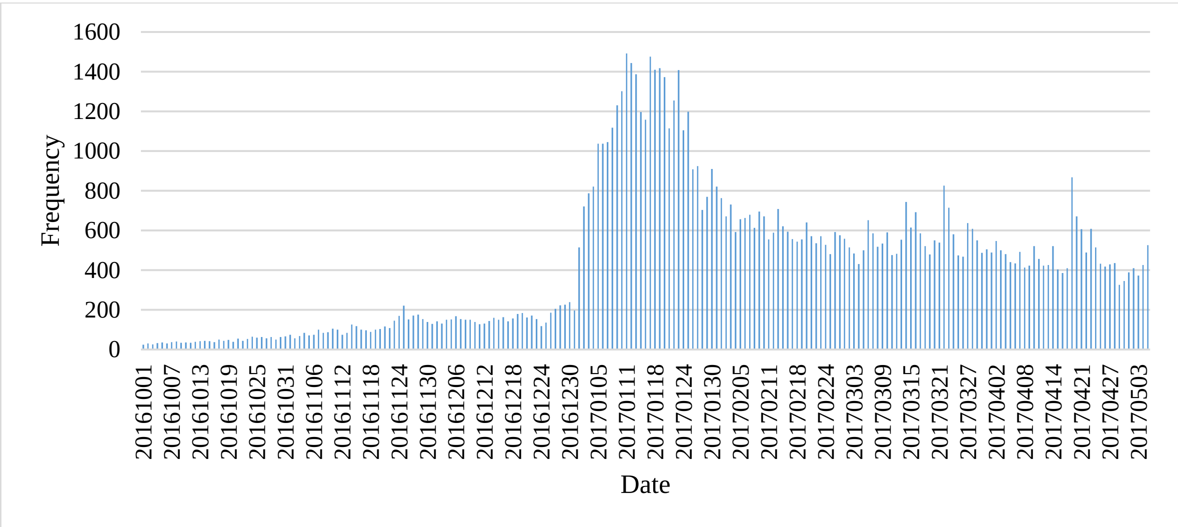
The final dataset contains 152,303 sessions made by 11,977 active users. Of them, 5,505 (45.96%) made at least one purchase. In total, we have information about 9,123 transactions [Table 5.3.1].

TABLE 5.3.1: Number of sessions, users, and transactions in the final dataset

	Number
Sessions (*t*):	152,306
Active users (*i*):	11,977
Transactions:	9,123
Users with transactions:	5,505

Active Users. We are following 11,977 users who visited the website between October 1st, 2016, and May 5th, 2017. They did not start browsing the website all in the same day, did not have sessions every day, and they can stop visit the website whenever they want. All the observed users are registered. Figure 5.3.1 shows the distribution of the number of active users in each day of the observation period. Due to the way in which the final database is built, 2016 is defined by a small number of users, the first part of 2017 is the one with the highest number of people, and in the last days, users diminish again, as some people stop engaging with the firm.

FIGURE 5.3.1: Frequency distribution of the number of active user per day



Number of Sessions. A session (*t*) is defined as a website visit, from the time the user lands on the website until she closes all the pages related to the website opened in the browser. In total, the 11,977 users made 152,306 sessions, ranging from a minimum of 1 to a maximum of 529. On average, they made 12.76 sessions. Figure 5.3.2 depicts the distribution of the number of sessions per user.

FIGURE 5.3.2: Frequency distribution of the number of sessions per user

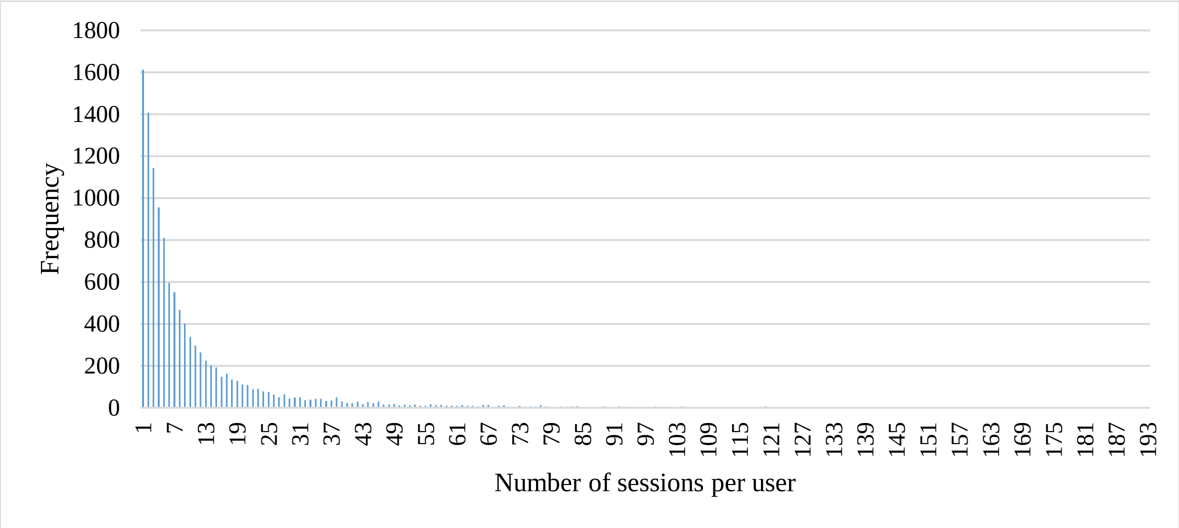
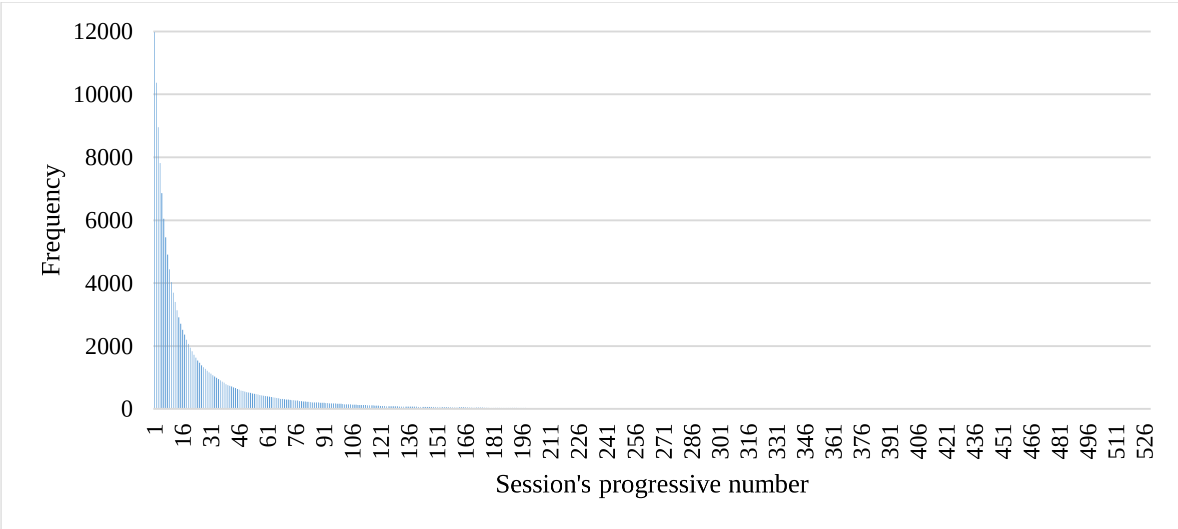


FIGURE 5.3.3: Frequency distribution of the sessions’ progressive numbers



Each session has a progressive number describing the chronological order of the times in which each user visits the website. For example, if user A has three sessions, she will have sessions 1 (the first), 2, and 3 (the last). Users with only one session will have only session 1. The distribution of the sessions’ progressive numbers, is represented in Figure 5.3.3. The

51.18% of the sessions is represented by all the sessions prior to the 12th, and only 5.03% are subsequent to the 114th.

5.3.2. Outcomes

Since search activity is highly related with purchase activity, as discussed in section 3.4, in this study, we consider more than one outcome variable, in order to account for both search and purchase actions. Table 5.3.2 describes the outcome variables considered.

TABLE 5.3.2: Outcome Variables

Macro category	Variable	Description
Search	n_click_{it}	Number of clicks of user *i* in session *t*
	n_pages_{it}	Number of pages seen by user *i* in session *t*
	$session\_length_{it}$	Time length of session *t* of user *i* (in seconds)
	$n\_products_{it}$	Number of products clicked by user *i* in session *t*
Transactional	$Purchase_{it}$	Dummy variable taking the value 1 if the user *i* made a purchase in session *t* and 0 otherwise
	$Monetary\_value_{it}$	Monetary value (in €) of the transaction. In other words, the final amount of € that the user *i* has to pay for the order in session *t*.

Table 5.3.3 shows the main statistics of the outcome variables.

TABLE 5.3.3: Main descriptives of the outcome variables

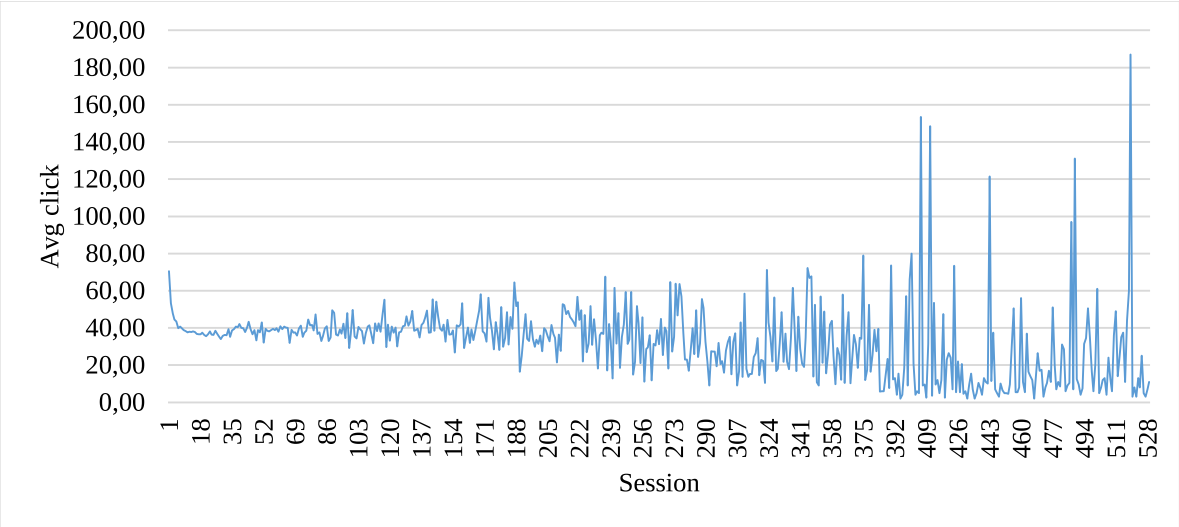
Variable	Obs	Mean	Std. Dev.	Min	Max
$session\_length_{it}$ (sec)	152306	933.762	1457.720	0	26270.7
n_click_{it}	152306	42.904	64.686	1	1000
n_pages_{it}	152306	19.304	33.112	1	960
$n\_products_{it}$	152306	6.503	13.173	0	410

People have on average 0.76 transactions each (SD=1.38), ranging from a minimum of 0 to a maximum of 37. They made in total 9123 orders in the considered period, with an average monetary value of 189.75€ (SD = 250.79; max = 8474.42 €).

Number of clicks. The average number of clicks of the 152,306 sessions is 42.90. Figure 5.3.4 depicts the distribution of the average number of clicks per session's progressive number.

As shown before, sessions over the 95th percentile (in other words, over the 113th session) have a very low frequency, less than 120 observations, in comparison with the first 15 sessions, which have more than 2700 observations. This can lead to a misleading interpretation of the distribution depicted in Figure 5.3.4, as the last sessions seem to have a very high average number of clicks, but it is coupled with a very few number of observations and, consequently, a low variance. As can be seen, the first two sessions are represented by the highest click intensity (respectively, 70.5 and 53.2 clicks on average), then it rapidly decreases and it stabilizes at around 35 clicks in the subsequent sessions.

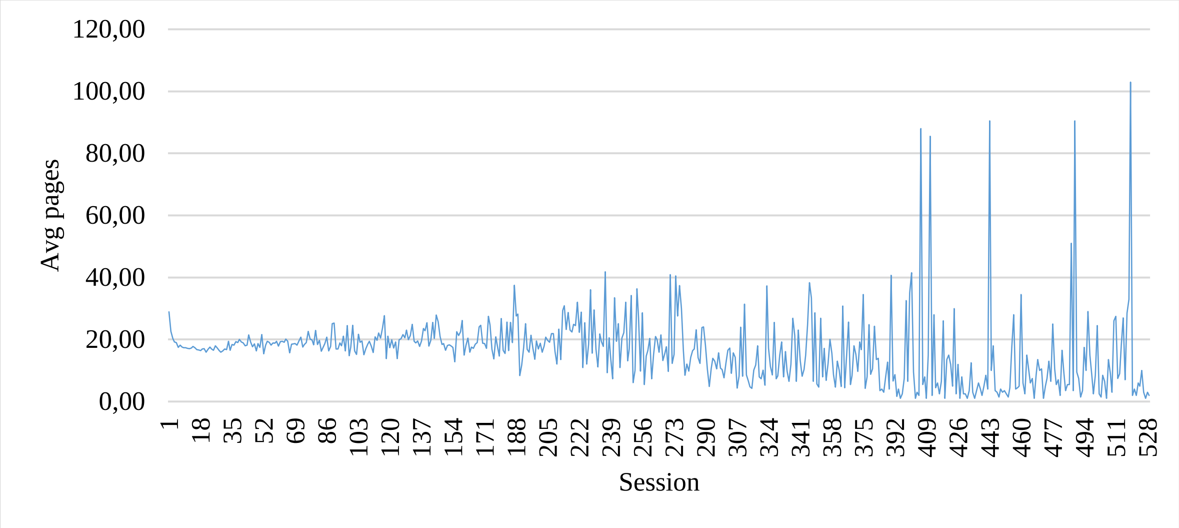
FIGURE 5.3.4: Distribution of the average number of clicks per session's progressive number



***Number of pages*.** The average number of pages seen in the 152306 sessions is 19.30, ranging from a minimum of 1 to a maximum of 960. Figure 5.3.5 depicts the distribution of the average number of pages per session's progressive number.

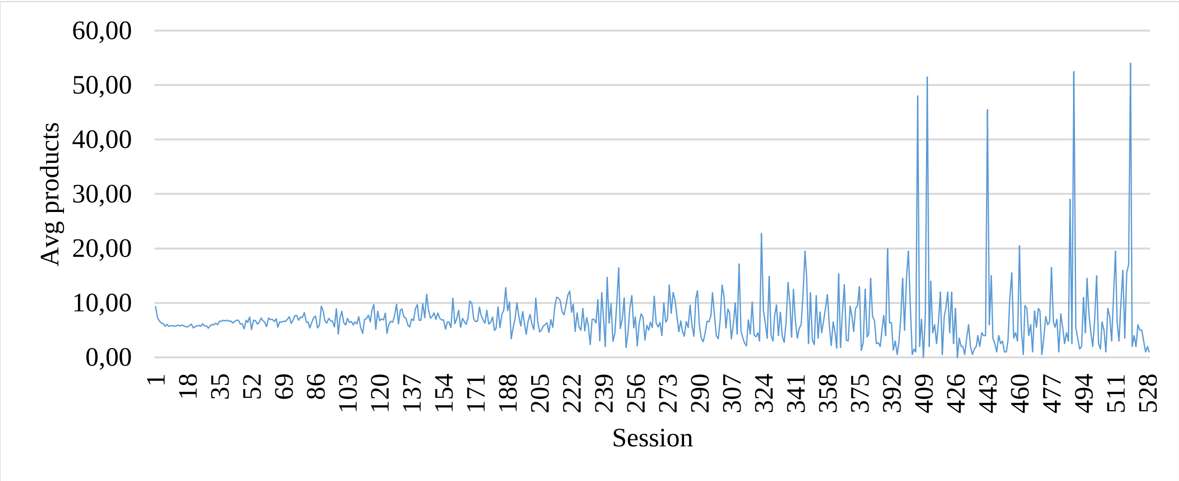
The shape is similar to the one of the average number of clicks: the first two sessions are represented by the highest number of pages seen (respectively, 28.9 and 22.6 pages on average), then it rapidly decreases and it stabilizes at around 17 pages in the subsequent sessions.

FIGURE 5.3.5: Distribution of the average number of pages per session's progressive number



Number of products. People click on average on 6.50 products per session. Figure 5.3.6 depicts the distribution of the average number of products per session's progressive number. The shape is similar to the one of the average number of clicks and pages: the first session is represented by the highest number of products clicked (9.33 products on average), then it rapidly decreases and it stabilizes at around 6 products in the subsequent sessions.

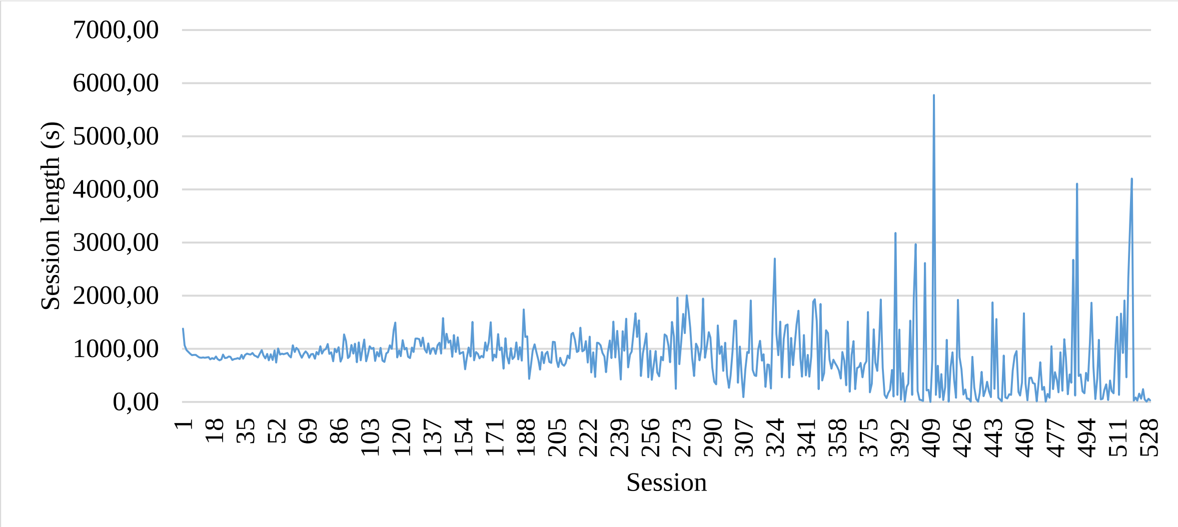
FIGURE 5.3.6: Distribution of the average number of products per session's progressive number



Session Length. The average length of a session is around 933.76 seconds. Figure 5.3.7 depicts the distribution of the average number of seconds per session's progressive number.

Like the previous distributions of clicks, pages, and products, the first two sessions have the longest duration (respectively, 1377.35 and 1065.77 seconds on average), then it rapidly decreases and it stabilizes at around 815 seconds in the subsequent sessions.

FIGURE 5.3.7: Distribution of the average session length per session's progressive number



Purchase. The percentage of sessions with at least one purchase is 5.61%. 0.256% of the sessions have more than one purchase, ranging from 2 to 16. Table 5.3.4 shows the frequency distribution of the number of purchases per session.

TABLE 5.3.4: Frequency distribution of the number of purchases per session

Purchases	Frequency	Percentage
0	143752	94,384%
1	8155	5,354%
2	314	0,206%
3	48	0,032%
4	22	0,014%
5	7	0,0046%
7	4	0,0026%
9	1	0,0007%
10	2	0,0013%
16	1	0,0007%
Total	152306	

The distribution of the percentage of sessions with purchases per session's progressive number is depicted in Figure 5.3.8. It immediately jumps to the eye that users are more likely to make a purchase in the first five sessions, which are the ones with the highest percentages of purchases (ranging from 12.36% of session 1, to 8% of session 5), while, as time passes by, their purchase propensity decreases.

FIGURE 5.3.8: Distribution of the percentage of sessions with purchases per session's progressive number

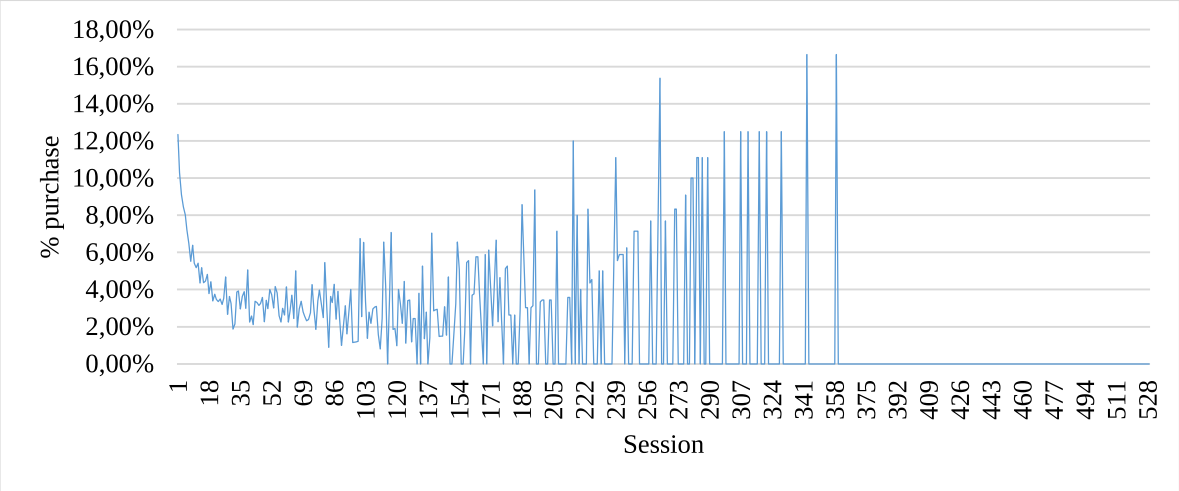


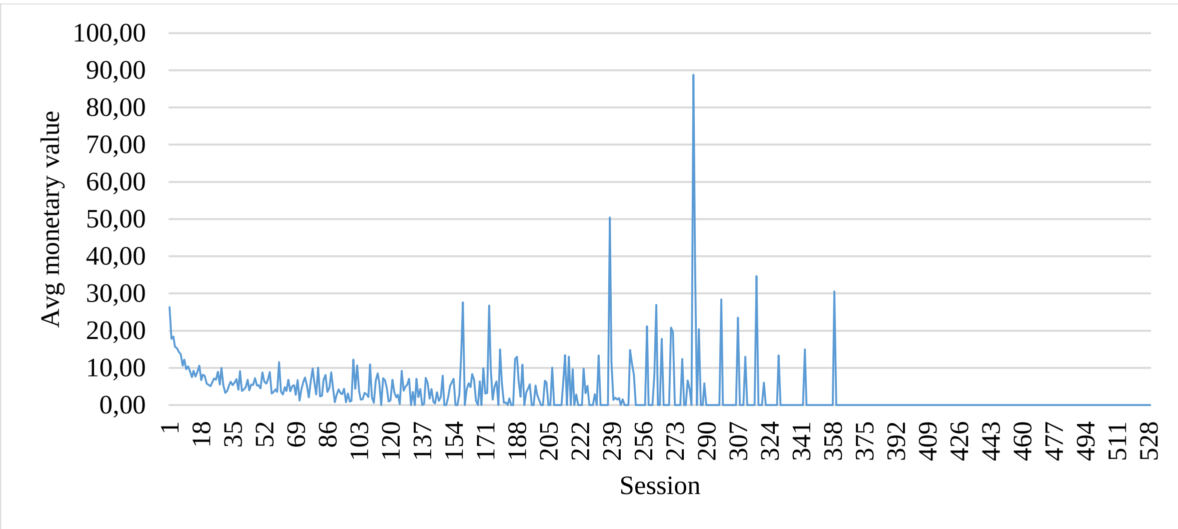
TABLE 5.3.5: Distribution of the number of purchases per user

Number of purchases	Freq	Percent	Cum.
0	6472	54.04	54.04
1	3725	31.1	85.14
2	1064	8.88	94.02
3	343	2.86	96.89
4	149	1.24	98.13
5	92	0.77	98.9
6	48	0.4	99.3
7	30	0.25	99.55
8	11	0.09	99.64
9	8	0.07	99.71
10	13	0.11	99.82
11	2	0.02	99.83
12	3	0.03	99.86
13	3	0.03	99.88
14	1	0.01	99.89
15	1	0.01	99.9
16	1	0.01	99.91
17	4	0.03	99.94
18	1	0.01	99.95
22	1	0.01	99.96
24	1	0.01	99.97
26	1	0.01	99.97
27	1	0.01	99.98
33	1	0.01	99.99
37	1	0.01	100
Total	11977	100	

People make more than one purchase during the observation period. On average, they make 0.76 purchases, ranging from 0 to 37. Table 5.3.5 shows the distribution of the number of purchases for each user. 94% of people have less than 3 purchase occasions.

Monetary Value. On average, customers spend 11.32€ per session[6]. Figure 5.3.9 depicts the distribution of the average monetary value per session's progressive number. In some way, the graph reflects the one of the percentage of purchases, with the higher-value orders concentrated in the first five sessions (ranging from 26.30€ of session 1, to 15.31€ of session 5), while, as time passes by, the average monetary value of the purchases decreases.

FIGURE 5.3.9: Distribution of the average monetary value per session's progressive number



5.3.3. Covariates

This section describes the time-varying covariates used in the model, divided in covariates for the state dependence probabilities and covariates for the transition probabilities. As already explained in chapter 4, the covariates included in the specification of the state dependent probabilities are believed to have a short-term effect on the outcome variables, while

[6] The calculation of the average monetary value per session comprises also sessions which do not involve any purchase.

the covariates included in the specification of the transition probabilities should have a long-term effect (Netzer et al., 2017).

Table 5.3.8 presents the two sets of variables. The first set regards the specification of the state dependence probabilities, and the second regards the transition probabilities. Following Ascarza and Hardie (2013) and Ascarza et al. (2018), we put marketing activities (e.g. promotional, last touch, and suggested products marketing variables) into the specification of the state dependent probabilities, together with the variables that are specific of the session, such as the device that has been used, the month in which it takes place, and the search-related activities that the user does during the session. The transition probabilities are defined by variables reflecting the characteristics of the state. Our set of variables comprises the number of days since the last session (Ascarza et al., 2018), the user's registration to the website or the subscription/unsubscription to the newsletter, the Wish list usage (since once the user adds some items in the wish list, they can be there also in the subsequent sessions), and the demographics.

TABLE 5.3.8: Covariates for the State Dependence and Transition Probabilities

State Dependence Probabilities		
Macro category	**Variable**	**Description**
Device	$Desktop_{it}$	Dummy variable taking the value 1 if the session *t* of user *i* is from computer and 0 otherwise
	$Smartphone_{it}$	Dummy variable taking the value 1 if the session *t* of user *i* is from smartphone and 0 otherwise
	$Tablet_{it}$	Dummy variable taking the value 1 if the session *t* of user *i* is from tablet and 0 otherwise
Search Presentation Order	$RankDate_{it}$	Number of times the user *i* ranked the search results by "latest arrivals" in session *t*
	$RankPriceHigh_{it}$	Number of times the user *i* ranked the search results from the highest price to the lowest in session *t*
	$RankPriceLow_{it}$	Number of times the user *i* ranked the search results from the lowest price to the highest in session *t*
	$Filters_{it}$	Number of filters (e.g. color, size, brand) the user *i* used during the searching activity in session *t*
Promotional	$Discount\_depth_{it}$	Depth of the discount that the user *i* sees in the website in session *t*. It can be 30%, 40%, 50%, 60%, 80%, or 0 (only in Africa).
	$Nl\_promotional_{it}$	Dummy variable taking the value 1 if there was a newsletter promotional campaign during the day of session *t* in user *i*'s

		country and user *i* subscribed to the newsletter, and 0 otherwise
	Nl_informational$_{it}$	Dummy variable taking the value 1 if there was a newsletter informational campaign during the day of session *t* in user *i*'s country and user *i* subscribed to the newsletter, and 0 otherwise
Marketing: Suggested Products	Suggested_Products$_{it}$	Number of products suggested by the firm (e.g. "You might also be interested in:…") on which the user *i* has clicked on in session *t*
Marketing: Last Touch	SEO$_{it}$	Dummy variable taking the value 1 if the user *i* landed on the website through SEO channel in session *t* and 0 otherwise
	Affiliation$_{it}$	Dummy variable taking the value 1 if the user *i* landed on the website through affiliation in session *t* and 0 otherwise
	Social$_{it}$	Dummy variable taking the value 1 if the user *i* landed on the website through social in session *t* and 0 otherwise
	Direct$_{it}$	Dummy variable taking the value 1 if the user *i* landed on the website through direct channel in session *t* and 0 otherwise
	Referral$_{it}$	Dummy variable taking the value 1 if the user *i* landed on the website through referral in session *t* and 0 otherwise
	Co-marketing$_{it}$	Dummy variable taking the value 1 if the user *i* landed on the website through co-marketing in session *t* and 0 otherwise
	Retargeting$_{it}$	Dummy variable taking the value 1 if the user *i* landed on the website through retargetig in session *t* and 0 otherwise
	Newsletter$_{it}$	Dummy variable taking the value 1 if the user *i* landed on the website through newsletter in session *t* and 0 otherwise
	SEM$_{it}$	Dummy variable taking the value 1 if the user *i* landed on the website through SEM channel in session *t* and 0 otherwise
	Other_channel$_{it}$	Dummy variable taking the value 1 if the user *i* landed on the website through other channels in session *t* and 0 otherwise
Temporal	October$_t$, November$_t$, December$_t$, January$_t$, February$_t$, March$_t$, April$_t$, May$_t$	Dummy variable taking the value 1 if the session *t* took place, respectively, in October, November, or December 2016, or January, February, March, April, or May 2017 and 0 otherwise
Transition Probabilities		
WishList	AddWishList$_{it}$	Number of items added in the Wish list in session *t* by customer *i*
	DropWishList$_{it}$	Number of items deleted from the Wish list in session *t* by customer *i*
	AddWishList_lag$_{it}$	Number of items added in the Wish list in session *t-1*
Registration	Registration_days$_{it}$	Days since the registration in the website
	Social_registration$_{it}$	Dummy variable taking the value 1 from the session in which the user *i* registered to the website through a social network account and 0 before
	Facebook_connected$_{it}$	Dummy variable taking the value 1 from the session in which the user *i* logged in the website through a Facebook account and 0 before
	Nl_registration_days$_{it}$	Days since the subscription to the newsletter

	$Nl\_unsubscription\_days_{it}$	Days since the unsubscription from the newsletter
	$Registration_{it}$	Dummy variable taking the value 1 if the user *i* registers to the website in session *t* and 0 otherwise.
	$Nl\_registration_{it}$	Dummy variable taking the value 1 if the user *i* subscribes to the newsletter in session *t* and 0 otherwise.
	$Nl\_unsubscription_{it}$	Dummy variable taking the value 1 if the user *i* unsubscribes to the newsletter in session *t* and 0 otherwise.
Search Activity	$Intersession\_days_{it}$	Days between session *t-1* and session *t*
Demographics	Age_i	Age of the user *i*
	$Gender_i$	Dummy variable taking the value 1 if the user *i* is female and 0 otherwise
	$Africa_i$, $America_i$, $Asia_i$, $Europe_i$, $Italy_i$, $Oceania_i$, $Russia_i$, USA_i	Dummy variables taking the value 1 if the user *i* logs in the website, respectively, from Africa, America, Asia, Europe (except Italy), Italy, Oceania, Russia, or USA, and 0 otherwise
	$Demo\_miss_i$	Dummy variable taking the value 0 if the user *i* provided her demographic information and 1 otherwise

5.3.4. State Dependence Covariates

Table 5.3.9 describes the State Dependence covariates. Since the incidence of co-marketing, other channel, and informational newsletter is lower than 5%, these three variables will not enter in the model.

TABLE 5.3.9: Main descriptives of the State Dependence covariates

Category	**Variable**	**Obs**	**Mean**	**Std. Dev.**	**Min**	**Max**
Search Presentation Order	$Rank\ Date_{it}$	152306	0.026	0.238	0	23
	$Rank\ Price\ High_{it}$	152306	0.051	0.359	0	26
	$Rank\ Price\ Low_{it}$	152306	0.367	1.321	0	71
	$Filters_{it}$	152306	2.341	6.689	0	304
Suggested Products	$Suggested\ Products_{it}$	152306	0.296	2.462	0	152
Promotional	$Discount\_depth_{it}$	152306	59.977	13.902	0	80
		Obs	**Freq**	**Percent**	**Cum**	
Marketing Last Touch	SEO_{it}	152306	31602	20.75	20.75	
	SEM_{it}	152306	27373	17.97	38.72	
	$Affiliazione_{it}$	152306	22846	15.00	53.72	
	$Newsletter_{it}$	152306	20793	13.65	67.37	
	$Retargeting_{it}$	152306	18476	12.13	79.50	
	$Direct_{it}$	152306	12585	8.26	87.77	
	$Referral_{it}$	152306	8942	5.87	93.64	
	$Social_{it}$	152306	8743	5.74	99.38	
	$Co\text{-}marketing_{it}$	152306	708	0.46	99.84	
	$Other\_channel_{it}$	152306	238	0.16	100.00	

Promotional	Nl promotional$_{it}$	152306	19647	12.90	
	Nl informational$_{it}$	152306	3198	2.10	
Device	Desktop$_{it}$	152306	105084	69.00	69.00
	Smartphone$_{it}$	152306	38123	25.03	94.03
	Tablet$_{it}$	152306	9099	5.97	100.00
Temporal	October$_{t}$	152306	2169	1.42	1.42
	November$_{t}$	152306	5291	3.47	4.89
	December$_{t}$	152306	7965	5.23	10.12
	January$_{t}$	152306	58039	38.11	48.23
	February$_{t}$	152306	25508	16.75	64.98
	March$_{t}$	152306	27940	18.34	83.32
	April$_{t}$	152306	21999	14.44	97.76
	May$_{t}$	152306	3395	2.23	99.99

Marketing Last Touch. Users can land on the website through 10 different channels:

- **SEO**: Organic search (e.g. Google);
- **Direct**: When the user lands in the website by writing the address in the address bar (e.g. www.companyname.com);
- **Retargeting**: Retargeting shows an advertising message embedded in another website to users who has previously demonstrated some interest in the firm;
- **Affiliation**: Affiliate websites are third party shopping websites, displaying the products of different retailers, through which users can land on the seller websites in order to get more information about the product, or make the purchase. Affiliate websites usually do not sell anything by themselves (e.g. www.vogue.com; www.skyscanner.com; www.bantoa.com);
- **Social**: Messages shared in a social media platform (e.g. Facebook, Twitter, Instagram, Google+);
- **Referral**: Link to the website in other websites, like blogs;
- **Co-marketing**: co-marketing occurs when the promotional activity of the firm is embedded in the marketing activities of another brand (e.g. when Vodafone includes

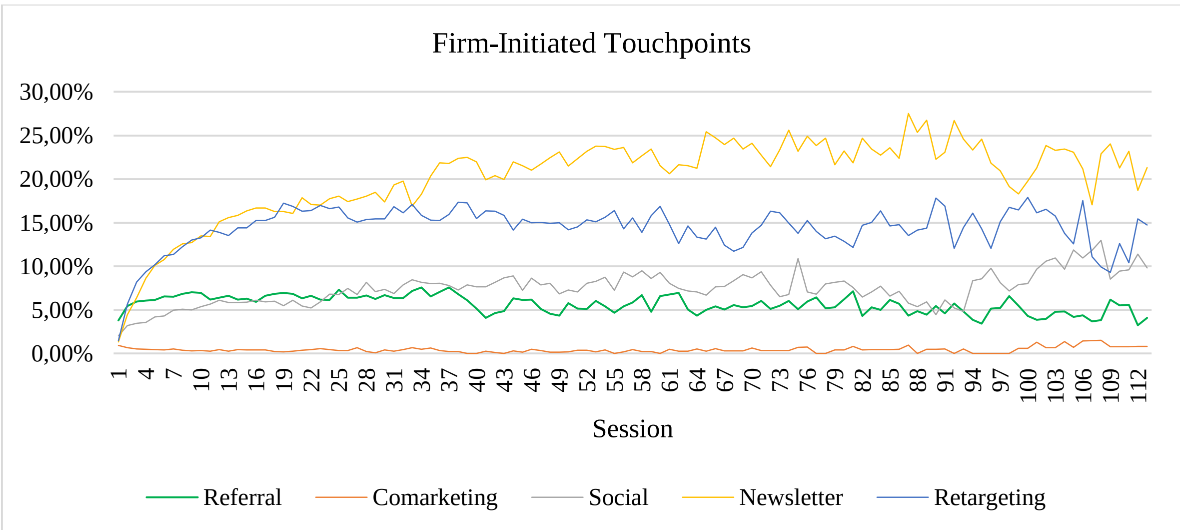
the Samsung smartphone in its monthly rate). In other words, the two brands make reciprocal discounts;

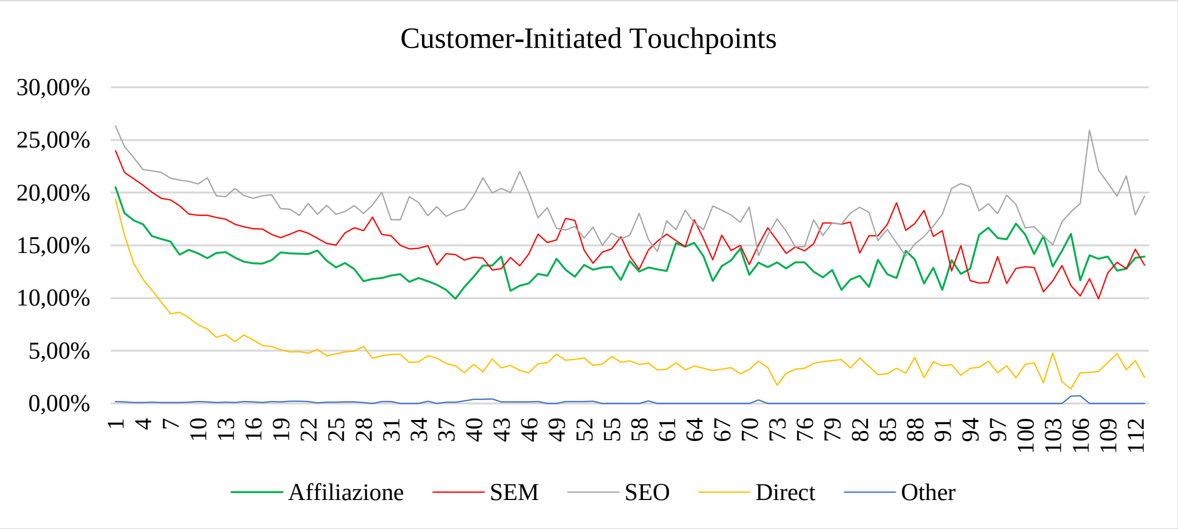
- **Newsletter**: Emails sent by the firm with informational or promotional purposes to people who opted in the reception of the newsletter;
- **SEM**: Paid search. In other words, sponsored links displayed as a result of search engines. They are different from organic search, since sponsored links are sold through auctions;
- **Other**: all the other ways in which users can land in the website (e.g. chat, navigation errors…).

Of the 152306 sessions, 20.75% and 17.97% are originated respectively by organic and paid search. Co-marketing and other channels (e.g. word of mouth via chat) represents only 0.46% and 0.16% [Table 5.3.9].

To make the interpretability easier, Figure 5.3.10 depicts the distribution of the number of times users landed in the website through each channel per session's progressive number up to the 95th percentile (113 sessions), in percentage.

FIGURE 5.3.10: Distribution of the channel percentages per sessions' progressive number for the first 95th percentile

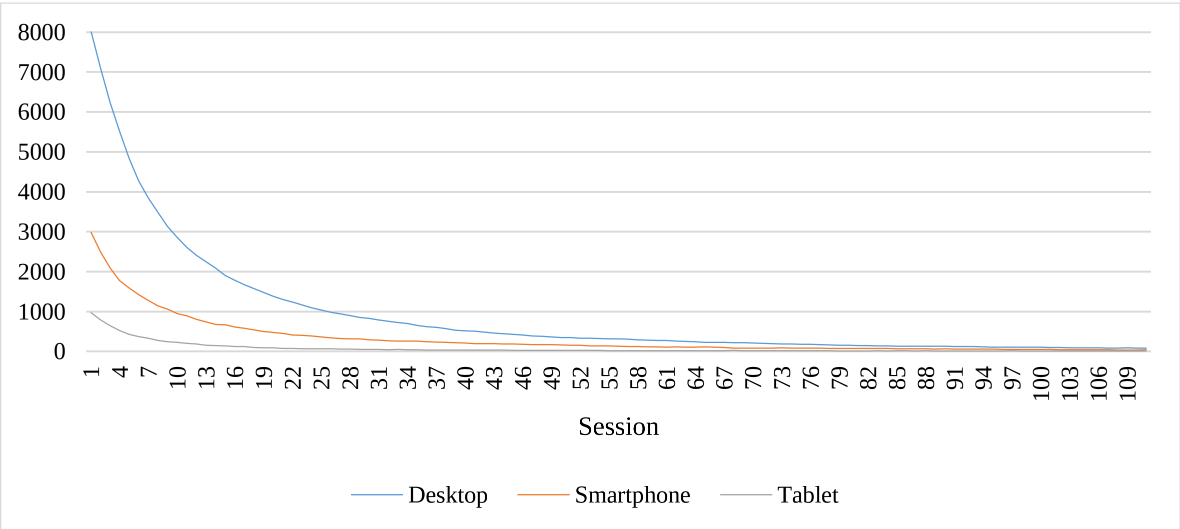




The graphs show that, among the firm-initiated touchpoints, the very first sessions come mainly from affiliated websites. Newsletter and retargeting catch on very quickly, and at around the 12th session, they surpassed the affiliation channel. Very few sessions are originated from co-marketing at any point in time. Regarding the customer-initiated touchpoints, SEO and SEM are always the most clicked, while the usage of the direct channel quickly decreases. Almost no sessions are originated from other types of channels.

Device.

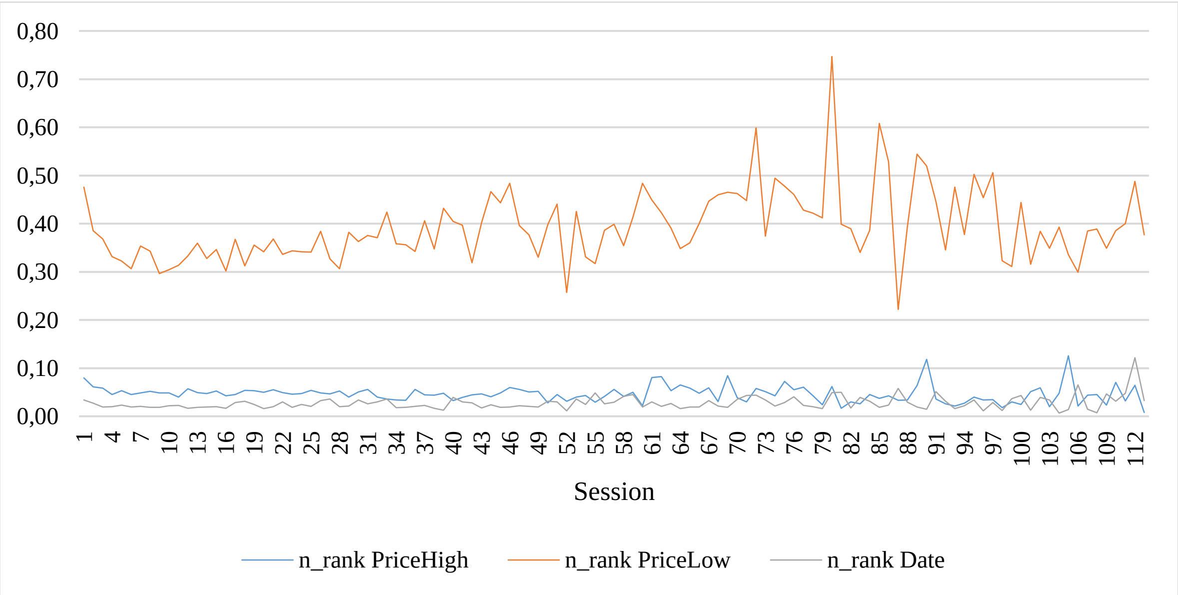
FIGURE 5.3.11: Distribution of the device usage per sessions' progressive number up to the 95th percentile



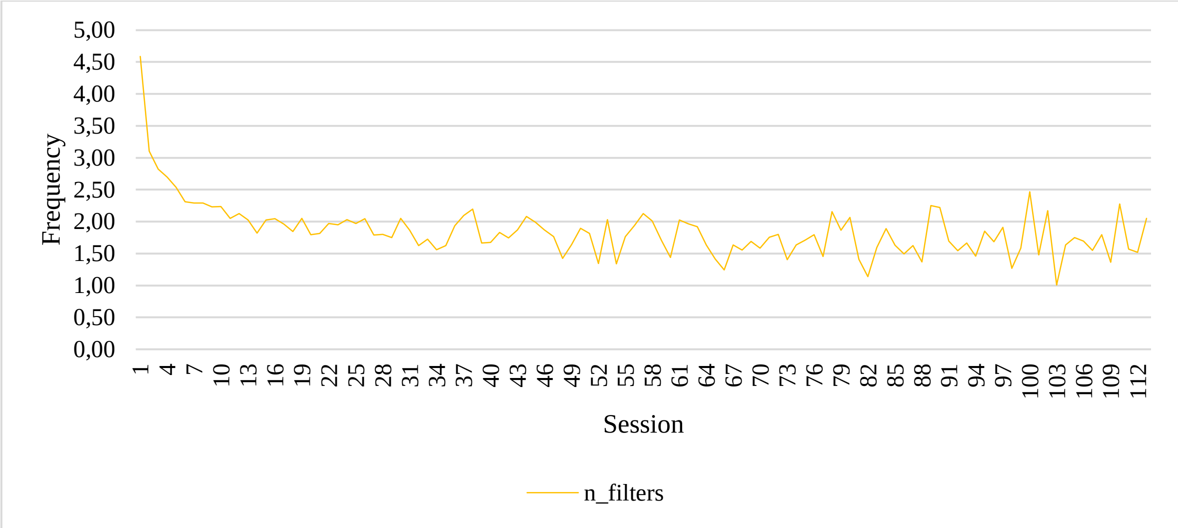
The vast majority of sessions are made from desktop (69%), 25% from smartphone, and the remaining 6% from tablet. Figure 5.3.11 depicts the distribution of the device usage across sessions up to the 95th percentile. Across all the sessions, the desktop is always the most used device, followed by the smartphone.

Search Presentation Order. When browsing for products in the website, users can decide to order the results of their search by price, both ascendant and descendant, or product novelty. They can also add filters to fine-tune their search. The average number of times people sort the results by price in the 152,306 sessions is 0.051 for the descendent sorting, 0.367 for the ascendant sorting, and 0.026 for the novelty sorting. The most used search tool is the possibility to apply filters. People apply on average 2.341 filters per session. Figure 5.3.12a depicts the distribution of the search presentation order tools usage (ranking tools) across sessions up to the 95th percentile, and Figure 5.3.12a depicts the distribution of the average number of filters across sessions up to the 95th percentile Across all the sessions, filters are always the most used tool, followed by the ranking from the lowest to the highest price.

FIGURE 5.3.12: a) Distribution of the ranking tools usage per sessions' progressive number up to the 95th percentile

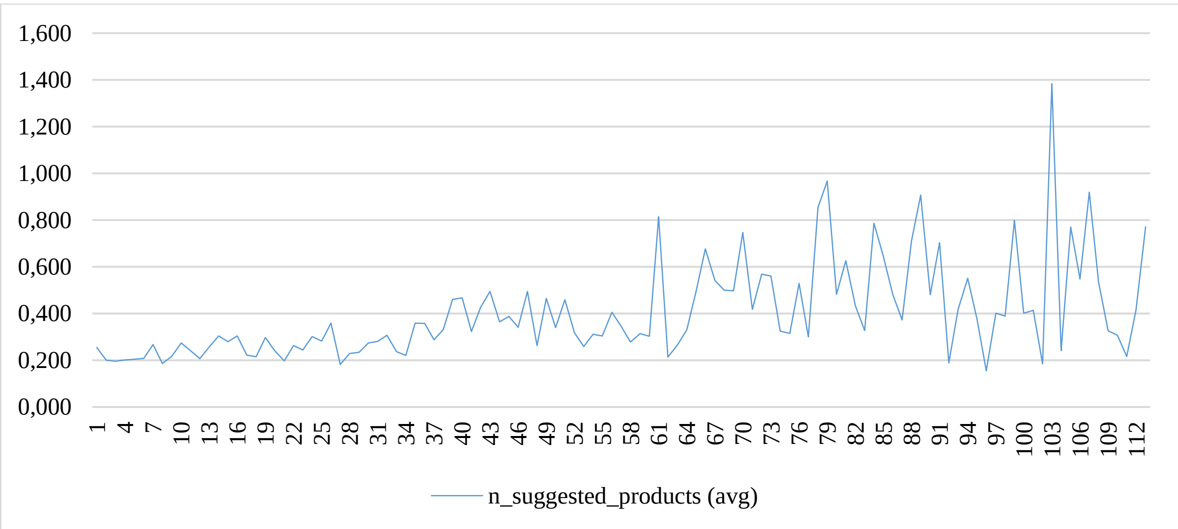


b) Distribution of the average number of filters used per sessions' progressive number up to the 95th percentile



Suggested Products. When a user clicks on a product, on the bottom of the product page the firm displays a list of suggested items that can be related to the main product of the page, with the intent of showing the user some items that can be of potential interest for her and in this way triggering her willingness to buy more. The average number of suggested products users click on is 0.296 per session. Figure 5.3.13 depicts the distribution of the average number of suggested products clicked per session's progressive number up the 95th percentile.

FIGURE 5.3.13: Distribution of the average number of suggested products clicked per session's progressive number up to the 95th percentile



Promotional Activities. The company engages in promotional activity through two main channels: the website and the newsletter. In the website, there are always discounts, available to any user who visits the website in a specific time window from a specific country. Discounts varies in terms of depth, both over time and across countries: 5.70% of the sessions have a 40% discount, 20.30% have a 50% discount, 31.10% a 60% discount, 15.30% an 80% discount, and the remaining 27.60% have a 30% discount. Discounts are the most effective activities, generally leading to a significantly higher purchase rate. This is true only for a discount depth higher than 50%. Reasonably, for 30% and 40% discounts, the purchase rate is significantly lower when the discount is present, since when it is absent, it means that a deeper discount is active [Table 5.3.10].

TABLE 5.3.10: Descriptives of the promotional activities

Event	**% Sessions**	**% Purchases if present**	**% Purchases if absent**	**Sig.**
Discount 30%	27.60	4.74	6.44	0.000
Discount 40%	5.70	3.09	6.14	0.000
Discount 50%	20.30	7.61	5.55	0.000
Discount 60%	31.10	6.23	5.85	0.003
Discount 80%	15.30	6.53	5.87	0.000
Promotional Newsletter	12.90	6.05	5.96	0.618
Informational Newsletter	2.10	4.95	5.99	0.013

Newsletters can be promotional or informational. Promotional emails contain discount codes available to anyone who receive the email. 12.90% of the sessions occur during a promotional email campaign and are made by users who subscribed to the reception of the newsletter. Surprisingly, promotional emails do not have any significant impact on the purchase rate (6.05% when present, 5.96% when absent, p-value: 0.618). Informational emails contain information about the new arrivals in the website. 2.11% of the sessions occur during an informational email campaign and are made by users who subscribed to the reception of the newsletter. Informational emails have a significant impact on the purchase rate, but they lead to a lower purchase rate (4.95% when present, 5.99% when absent, p-value: 0.013). This may be due to the fact that informational communications can trigger search activity more than the purchases.

5.3.5. Transition Probabilities Covariates

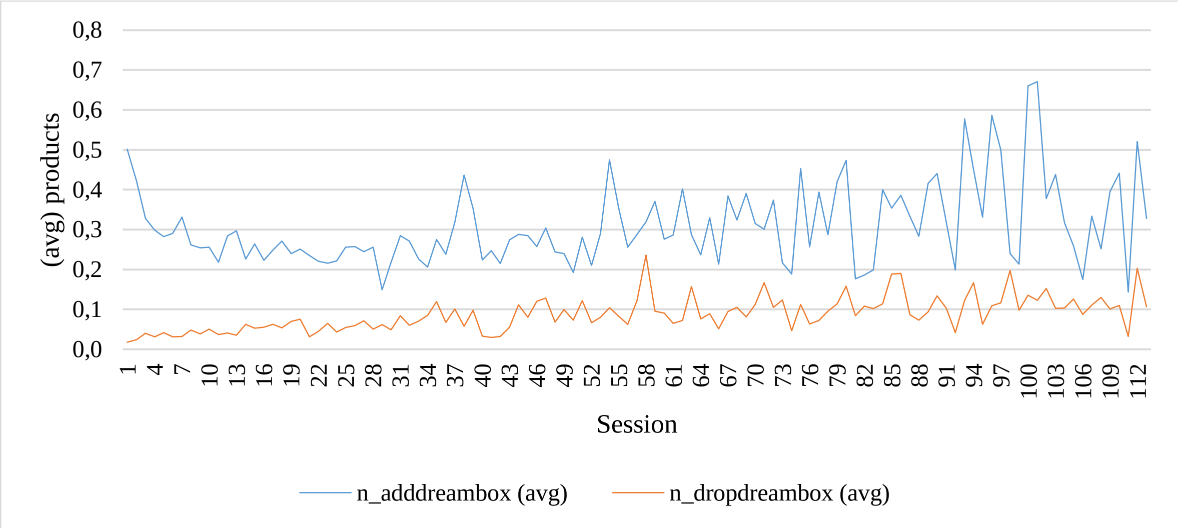
Table 5.3.11 describes the covariates for the Transition Probabilities. Since only 10% of the observed sessions are made from users who have a registration in another brand's website, the variable containing the number of days elapsed since the registration to a monobrand website will not enter in the model.

TABLE 5.3.11: Main descriptives of the Transition probabilities covariates

Main Category	Variable	Obs	Mean	Std. Dev.	Min	Max
Wish List	Add Wish List$_{it}$	152306	0.311	1.742	0	78
	Drop Wish List$_{it}$	152306	0.056	0.597	0	48
Search Activity	Intersession Days$_{it}$	140329	4.174	11.569	0	198
Registration	Registration Days$_{it}$	152306	25.162	33.914	0	207
	Nl Registration Days$_{it}$	152306	8.905	25.573	0	207
	Nl Unsubscription Days$_{it}$	152306	1.291	8.832	0	130
Demographics	Age$_i$	83105	37.81	11.645	13	98
		Obs	**Freq.**	**Perc.**	**Cum.**	
Registration	Facebook connected$_{it}$	152306	12797	8.40		
	Social registration$_{it}$	152306	20217	13.27		
Demographics	No country$_i$	152306	232	0.15	0.15	
	Africa$_i$	152306	322	0.21	0.36	
	America$_i$	152306	2156	1.42	1.78	
	Asia$_i$	152306	26475	17.38	19.16	
	Europe$_i$	152306	42660	28.01	47.17	
	Italy$_i$	152306	23628	15.51	62.68	
	Oceania$_i$	152306	1827	1.20	63.88	
	Russia$_i$	152306	36065	23.68	87.56	
	USA$_i$	152306	18941	12.44	100	
	Gender (Female$_i$)	152306	95890	62.96		
	Demo Missing$_i$	152306	69201	45.44		

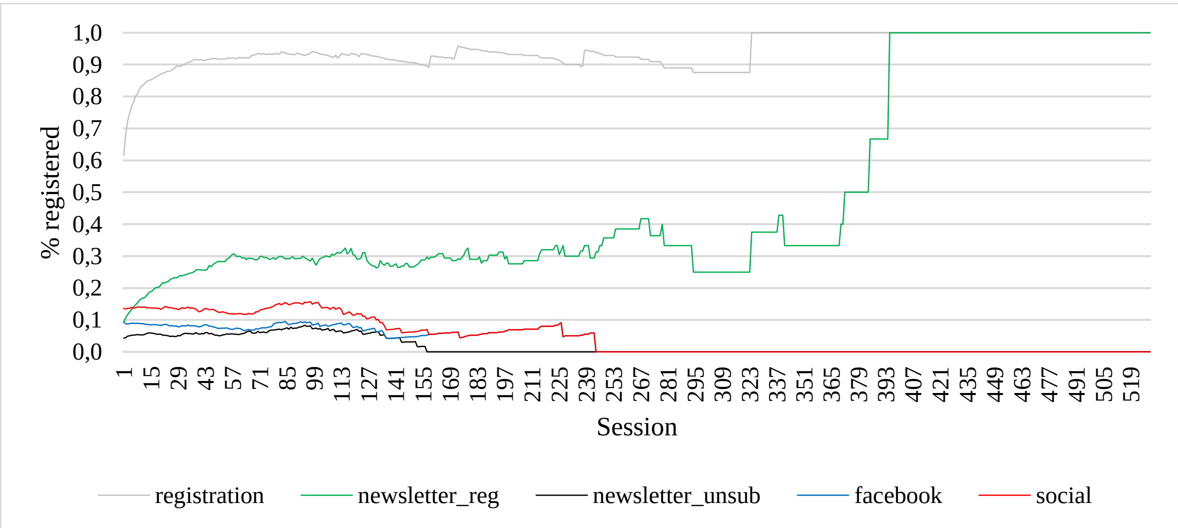
Wish List. Users can save (or drop) products of interest in the Wish List, in order to decide whether to buy them in any of the subsequent sessions. The average number of items added and dropped from the Wish List in each session are respectively 0.311 and 0.056 [Table 5.3.9]. Figure 5.3.14 depicts the distribution of the average number of items added and dropped from the Wish List per session's progressive number up to the 95th percentile.

FIGURE 5.3.14: Distribution of the average number of items added and dropped from the Wish List per session's progressive number up to the 95[th] percentile



Registration. Users can register to the website in two ways: they can opt-in to the newsletter, or they can make a full registration to the website. They can also decide to register through a social network account, like Google+, or Facebook.

FIGURE 5.3.15: Distribution of the percentage users for each type of registration per session's progressive number

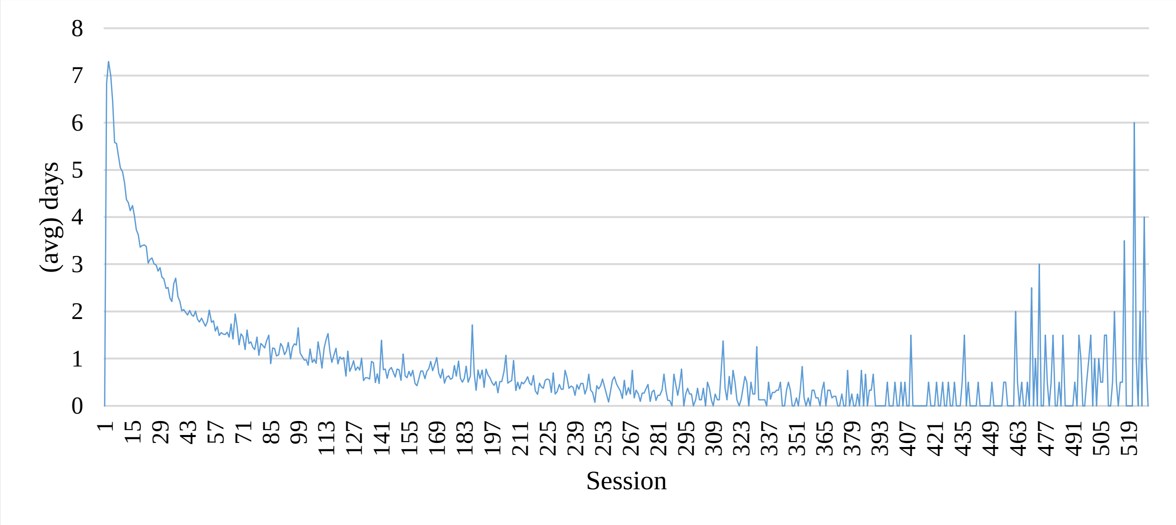


All the 11977 observed users (100%) have a full registration, but since they didn't register on the website at the same time, 1,3% of them registered after the end of the observation time (after May 5, 2017). 20,8% of the observed users made the registration with a social network account

(67,9% of them through Facebook). 24% subscribed to the newsletter, and 39,82% of them (9% of all the observed users) decided to unsubscribe. Figure 5.3.15 depicts the distribution of the percentage users for each type of registration per session's progressive number.

Intersession Days. The average number of days between session *t* and session *t-1* is 4.174 (SD = 11.569), ranging from a minimum of 0 to a maximum of 198. Figure 5.3.17 depicts the distribution of the average number of intersession days per session's progressive number. As anticipated previously, between the first sessions users need more days to visit the website again.

FIGURE 5.3.17: Distribution of the average number of intersession days per session's progressive number



5.4. Model Estimation

To estimate the HMM, we use the LMest R package (Bartolucci, Pandolfi, and Pennoni, 2017), which has been developed with the intent to estimate HMM with longitudinal data. The package has several features that make it particularly suited for our analysis over other alternative R packages: (i) it allows for time-varying covariates in the estimation of the transition and state dependent probabilities; (ii) it is able to handle multivariate outcomes; (iii) provides the estimation of standard errors; (iv) can deal with unbalanced panels.

This last characteristic is particularly useful in our context, as we are dealing with users with a high variation in the number of sessions. To deal with unbalanced panels, in order to create a sort of balancing, the package adds missing observations to users with a lower number of sessions, as the program allows for missing data under the missing-at-random assumption (Bartolucci et al, 2017). This solution raises some problems when few users have a very high number of sessions, while many others have few. To solve this issue, we estimate the model by dropping from the dataset all the users with a maximum number of sessions higher than 70 (2.51% of 11,977 users). The model will be then estimated using 11,676 users who made 115,589 sessions overall.

5.4.1. *Dependent Variables in the Model*

Since the LMest package is not able to handle multiple outcomes which are both categorical and continuous, we decided to discretize the continuous outcome variables, such as the search variables, and the monetary value of the transaction.

The correlations among the six outcome variables are described in Table 5.3.6, which shows a high correlation between the four search variables, especially between clicks and pages (0.906), clicks and products (0.805), and clicks and session length (0.781).

TABLE 5.3.6: Matrix correlation of the outcome variables

	Pages	Clicks	Products	Session Length	Monetary Value	Purchase
Pages	1.000					
Clicks	0.906	1.000				
Products	0.752	0.805	1.000			
Session Length	0.690	0.781	0.677	1.000		
Monetary Value	0.079	0.113	0.0630	0.093	1.000	
Purchase	0.107	0.148	0.0664	0.138	0.592	1.000

Search Outcomes. In order to reduce the number of parameters to be estimated with the HMM, we decided to join all the four search variables into one single factor, by running a factor analysis. Results confirm the existence of a single factor, which explains the 76.77% of the variance of all the four variables. This search factor ranges from a minimum of -0.627 to a maximum of 15.515, with an average of -0.015 (SD = 0.934). In order to discretize the variable, we split it into five classes, represented by the quintiles of the search factor distribution. Accordingly, in the first class fall all the sessions having a search factor value lower than the 20^{th} percentile (-0.556), in the second category the ones having a search factor between the 20^{th} and the 40^{th} (-0.443) percentiles, in the third between the 40^{th} and the 60^{th} (-0.232), in the fourth between the 60^{th} and the 80^{th} (0.285), and in the fifth sessions with a purchase factor higher than the 80^{th} percentile.

Transactional Outcomes. Regarding the transactional outcomes, we compute the cumulative variable of both purchase and monetary value, so that people who make only a first transaction will fall on the first purchase category, and move on the second category as soon as they make their second purchase. The last category comprises all the people who make more than two purchases from the time they buy from the firm for the third time. Category 0 comprises all the sessions before the first transaction. Similarly, the monetary value has been categorized by dividing all the sessions from the first transaction in quintiles. Category 0 contains all the sessions before the first transactions, category 1 has people who has a monetary value higher than 0 and lower than 70.97€, category 2 between 70.97 and 126.22€, category 3 between 126.22 and 210.48€, category 4 between 210.48 and 387.10€, and category 5 has a monetary value higher than 387.10€. Table 5.3.7 shows the dimensions of the purchase and monetary categories in the dataset.

TABLE 5.3.7: Frequencies of purchase and monetary categories

Purchase	Freq	Percent	Cum
0	83277	54.47	54.47
1	39539	25.86	80.34
2	13666	8.94	89.28
3	16393	10.72	100.00
Total	152875	100.00	

Monetary	Freq	Percent	Cum
0	83277	54.47	54.47
1	13914	9.10	63.58
2	13908	9.10	72.67
3	13933	9.11	81.79
4	13932	9.11	90.90
5	13911	9.10	100.00
Total	152875	100.00	

5.4.1. Correlations

Tables 5.3.12 show the correlations between the outcome variables and the state dependence covariates distinct by group. As can be seen, the highest correlations are the ones between the search factor and the number of filters applied (0.410), and the search factor and the number of times people rank the results by ascendant price (0.367).

TABLE 5.3.12: Correlation matrixes

a) Marketing last touch

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
(1) Search Factor	1.000												
(2) Purchase	0.137	1.000											
(3) Monetary Value	0.104	0.592	1.000										
(4) Affiliation	-0.031	0.020	0.009	1.000									
(5) SEO	0.063	0.004	0.000	-0.215	1.000								
(6) Other	-0.004	-0.007	-0.003	-0.017	-0.020	1.000							
(7) Social	-0.031	-0.026	-0.019	-0.104	-0.126	-0.010	1.000						
(8) Direct	0.046	0.049	0.052	-0.127	-0.154	-0.012	-0.074	1.000					
(9) Referral	-0.013	0.010	0.005	-0.105	-0.128	-0.010	-0.062	-0.075	1.000				
(10) Co marketing	0.016	0.005	0.003	-0.029	-0.035	-0.003	-0.017	-0.021	-0.017	1.000			
(11) Retargeting	-0.039	-0.039	-0.028	-0.156	-0.190	-0.015	-0.092	-0.112	-0.093	-0.025	1.000		
(12) Newsletter	-0.036	-0.033	-0.022	-0.167	-0.203	-0.016	-0.098	-0.120	-0.099	-0.027	-0.147	1.000	
(13) SEM	0.020	0.014	0.007	-0.197	-0.240	-0.019	-0.115	-0.141	-0.117	-0.032	-0.174	-0.186	1.000

b) Device and search presentation order tools

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
(1) Search Factor	1.000										
(2) Purchase	0.137	1.000									
(3) Monetary Value	0.104	0.592	1.000								
(4) Desktop	-0.054	-0.009	0.001	1.000							
(5) Mobile	0.056	-0.004	-0.009	-0.862	1.000						
(6) Tablet	0.003	0.025	0.015	-0.376	-0.146	1.000					
(7) Rank Date	0.109	-0.009	0.000	-0.007	0.014	-0.012	1.000				
(8) Rank Price High	0.189	0.011	0.021	0.018	-0.016	-0.006	0.089	1.000			
(9) Rank Price Low	0.367	0.011	-0.006	0.009	-0.002	-0.014	0.067	0.182	1.000		
(10) Filters	0.410	0.043	0.031	0.170	-0.200	0.034	0.069	0.190	0.310	1.000	
(11) Suggested Products	0.200	0.000	-0.003	-0.179	0.208	-0.030	0.018	0.003	0.040	-0.042	1.000

c) Promotional

	(1)	(2)	(3)	(4)	(5)
(1) Search Factor	1.000				
(2) Purchase	0.137	1.000			
(3) Monetary Value	0.104	0.592	1.000		
(4) Nl promotional	0.001	0.001	-0.002	1.000	
(5) Discount depth	-0.019	-0.0002	0.009	0.041	1.000

d) Demographics

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
(1) Search Factor	1.000														
(2) Purchase	0.137	1.000													
(3) Monetary Value	0.104	0.592	1.000												
(4) Age	0.000	-0.029	-0.018	1.000											
(5) Demo Missing	0.007	0.040	0.027	-0.910	1.000										
(6) Female	0.035	0.001	0.002	0.060	-0.011	1.000									
(7) Africa	0.000	-0.003	-0.004	-0.004	0.000	-0.009	1.000								
(8) America	-0.007	-0.011	-0.004	-0.029	0.035	0.002	-0.006	1.000							
(9) Asia	0.040	-0.008	0.000	-0.085	0.088	-0.016	-0.021	-0.055	1.000						
(10) Italy	-0.033	0.040	-0.002	0.052	-0.024	-0.001	-0.020	-0.051	-0.197	1.000					
(11) Europe	-0.011	0.026	0.012	0.045	-0.019	0.065	-0.029	-0.075	-0.286	-0.268	1.000				
(12) USA	-0.018	0.022	0.048	-0.074	0.084	-0.008	-0.017	-0.045	-0.173	-0.162	-0.235	1.000			
(13) Oceania	-0.003	0.013	0.016	-0.015	0.018	0.003	-0.005	-0.013	-0.051	-0.047	-0.069	-0.042	1.000		
(14) Russia	0.021	-0.071	-0.050	0.054	-0.117	-0.049	-0.026	-0.067	-0.255	-0.239	-0.347	-0.210	-0.061	1.000	
(15) No Country	-0.004	-0.002	0.001	0.003	-0.007	0.012	-0.002	-0.005	-0.018	-0.017	-0.024	-0.015	-0.004	-0.022	1.000

e) Transition Probabilities Covariates

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
(1) Add Wish List	1.000								
(2) Drop Wish List	0.245	1.000							
(3) Facebook connected	0.037	0.006	1.000						
(4) Social Registration	0.038	0.001	0.774	1.000					
(5) Registration Days	-0.004	0.032	0.022	0.032	1.000				
(6) Nl Registration Days	-0.011	0.006	-0.001	-0.019	0.605	1.000			
(7) Nl Unsubscription Days	0.004	0.012	0.168	0.261	0.179	0.050	1.000		
(8) Monobrand Reg Days	-0.009	-0.008	-0.017	-0.021	-0.056	-0.050	-0.014	1.000	
(9) Intersession Days	0.001	-0.010	0.017	0.017	0.138	0.062	0.049	0.032	1.000

5.4.2. Transition and Initial Probabilities

The transition probability $q_{iss'}^{(t)}$ of a user *i* to shift from state *s* to state *s'* in the acquisition funnel in each session *t* is a function of a set of individual time-varying covariates, categorized in four categories:

- Wishlist: $\boldsymbol{x}_{1i}^{'(t)}$
- Registration: $\boldsymbol{x}_{2i}^{'(t)}$
- Searching activity: $\boldsymbol{x}_{3i}^{'(t)}$
- Demographics: $\boldsymbol{x}_{4i}^{'(t)}$

The specification of the transition probabilities takes the multinomial logit form:

$$q_{iss'}^{(t)} = P(S_i^{(t)} = s' | S_i^{(t-1)} = s)$$

$$= \frac{exp\left\{\boldsymbol{x}_{1i}^{\prime(t)}\beta_{1ss'} + \boldsymbol{x}_{2i}^{\prime(t)}\beta_{2ss'} + \boldsymbol{x}_{3i}^{\prime(t)}\beta_{3ss'} + \boldsymbol{x}_{4i}^{\prime(t)}\beta_{4ss'}\right\}}{1 + \sum_{s'=1}^{S} exp\left\{\boldsymbol{x}_{1i}^{\prime(t)}\beta_{1ss'} + \boldsymbol{x}_{2i}^{\prime(t)}\beta_{2ss'} + \boldsymbol{x}_{3i}^{\prime(t)}\beta_{3ss'} + \boldsymbol{x}_{4i}^{\prime(t)}\beta_{4ss'}\right\}} \quad if\ s \neq s'$$

$$q_{iss'}^{(t)} = (S_i^{(t)} = s' | S_i^{(t-1)} = s)$$

$$= \frac{1}{1 + \sum_{s'=1}^{S} exp\left\{\boldsymbol{x}_{1i}^{\prime(t)}\beta_{1ss'} + \boldsymbol{x}_{2i}^{\prime(t)}\beta_{2ss'} + \boldsymbol{x}_{3i}^{\prime(t)}\beta_{3ss'} + \boldsymbol{x}_{4i}^{\prime(t)}\beta_{4ss'}\right\}} \quad if\ s = s'$$

Since the package does not allow us to put restrictions in the transition probability matrix, we employ the cumulative number of purchases and the cumulative number of spending described in section 5.3 as dependent variables. The structure of these variables does not allow people to move back in previous states once the customer makes one (or more) purchases, operating as an informal restriction.

We employ the same set of variables to estimate the initial state probabilities, π_{is}, representing the probability of each individual of being in state *s* in the first session *t=1*.

The specification of the initial probabilities takes the multinomial logit form:

$$\pi_{is} = P\left(S_i^{(1)} = s\right)$$

$$= \begin{cases} \dfrac{1}{1 + \sum_{s=1}^{S} exp\left\{\boldsymbol{x}_{1i}^{\prime(1)}\tau_{1s} + \boldsymbol{x}_{2i}^{\prime(1)}\tau_{2s} + \boldsymbol{x}_{3i}^{\prime(1)}\tau_{3s} + \boldsymbol{x}_{4i}^{\prime(1)}\tau_{4s}\right\}} & if\ s = 1 \\ \dfrac{exp\left\{\boldsymbol{x}_{1i}^{\prime(1)}\tau_{1s} + \boldsymbol{x}_{2i}^{\prime(1)}\tau_{2s} + \boldsymbol{x}_{3i}^{\prime(1)}\tau_{3s} + \boldsymbol{x}_{4i}^{\prime(1)}\tau_{4s}\right\}}{1 + \sum_{s=1}^{S} exp\left\{\boldsymbol{x}_{1i}^{\prime(1)}\tau_{1s} + \boldsymbol{x}_{2i}^{\prime(1)}\tau_{2s} + \boldsymbol{x}_{3i}^{\prime(1)}\tau_{3s} + \boldsymbol{x}_{4i}^{\prime(1)}\tau_{4s}\right\}} & if\ s = 2, \dots, S \end{cases}$$

5.4.3. State Dependent Probabilities

The outcome for the state dependent probabilities of the LMest package does not provide an estimation of the impact of the covariates in each state, but assumes that the impact of the covariates is homogeneous across states (Bartolucci et al., 2017). This assumption is incompatible with our research objective to explore the role of the marketing activities and search behavior in each stage of the process, thus we follow an alternative path.

After estimating the transition probabilities, $q_{iss'}^{(t)}$, we allocate every session *t* in the state *s* with the highest probability $q_{iss'}^{(t)}$. To estimate the state dependent probabilities, we took all the observations allocated to each state, and estimated the impact of the covariates by running a regression for each outcome in each state. It is evident that this is a suboptimal solution to the problem. A further development of the package is required to deal with this issue in a better way.

The outcomes of the model are a vector Yit of dimensions 3x1: Yit = [Y1it, Y2it, Y3it], where Y1it, Y2it, Y3it represent respectively the search factor, a binary variable indicating whether or not the user i placed an order during the session t, and the log-transformation of the monetary value of the transaction in session t (if any). Following Ebbes and Netzer (2017), to account for the long-tail distribution of the monetary value, we calculate the log-transformation as *ln*(monetary value+1), to manage the 0-values of the variable.

The state dependence probability is a function of a set of individual time-varying covariates, categorized in four categories:

- Marketing (last touch, promotional): $\boldsymbol{z}_{1i}'^{(t)}$
- Search Activity: $\boldsymbol{z}_{2i}'^{(t)}$
- Device: $\boldsymbol{z}_{3i}'^{(t)}$

- Temporal: $\boldsymbol{z}_{4i}^{\prime(t)}$

Linear regression model for the continuous outcomes. To model for the two continuous outcome variables $Y_{pi}^{(t)}$, with *p* = {1, 3}, we employ a simple linear regression model:

$$f\left(Y_{pi}^{(t)}\middle|S_i^{(t)}=s\right)=\varphi_{ps}+\boldsymbol{z}_{1i}^{\prime(t)}\gamma_{1p}+\boldsymbol{z}_{2i}^{\prime(t)}\gamma_{2p}+\boldsymbol{z}_{3i}^{\prime(t)}\gamma_{3p}+\boldsymbol{z}_{4i}^{\prime(t)}\gamma_{4p}$$

The component φ_{ps} represents the baseline value for each outcome variable *p* in state *s*.

Logit model for purchase probability. To model the binary outcome variable $Y_{2i}^{(t)}$ indicating whether the session *t* contains a purchase, we employ a binary logit model:

$$P\left(Y_{2i}^{(t)}=1\middle|S_i^{(t)}=s\right)=\frac{exp\left\{\varphi_{2s}+\boldsymbol{z}_{1i}^{\prime(t)}\gamma_{12}+\boldsymbol{z}_{2i}^{\prime(t)}\gamma_{22}+\boldsymbol{z}_{3i}^{\prime(t)}\gamma_{32}+\boldsymbol{z}_{4i}^{\prime(t)}\gamma_{42}\right\}}{1+exp\left\{\varphi_{2s}+\boldsymbol{z}_{1i}^{\prime(t)}\gamma_{12}+\boldsymbol{z}_{2i}^{\prime(t)}\gamma_{22}+\boldsymbol{z}_{3i}^{\prime(t)}\gamma_{32}+\boldsymbol{z}_{4i}^{\prime(t)}\gamma_{42}\right\}}$$

In which φ_{2s} represents the intrinsic likelihood to purchase in each state *s*.

5.5. Results

5.5.1. Latent States

We estimate the model by changing the number of states from 2 to 5. The model identifies the 5 states as the best solution, since it provides the lowest BIC value (530045.2) [Table 5.5.1].

TABLE 5.5.1: Models comparison

Number of states	2	3	4	5
BIC	626397.2	569322.7	566343.1	530045.2

The five states are defined by the level of search intensity in each session *t*, the cumulative number of purchases, and the cumulative monetary value up to session *t*. Table 5.5.2 describes the composition of the states, and table 5.5.3 provides some descriptives about the different types of search activity and purchase activity in each of the latent states.

TABLE 5.5.2: States description

Search	**State 1**	**State 2**	**State 3**	**State 4**	**State 5**
0	0.462	0.032	0.201	0.193	0.155
1	0.303	0.115	0.227	0.221	0.203
2	0.179	0.208	0.218	0.221	0.220
3	0.055	0.301	0.202	0.194	0.217
4	0.000	0.344	0.152	0.171	0.205
Purchase	**State 1**	**State 2**	**State 3**	**State 4**	**State 5**
0	1.000	1.000	0.000	0.000	0.000
1	0.000	0.000	0.902	0.680	0.344
2	0.000	0.000	0.092	0.232	0.265
3	0.000	0.000	0.006	0.088	0.390
Monetary value	**State 1**	**State 2**	**State 3**	**State 4**	**State 5**
0	1.000	1.000	0.000	0.000	0.000
1	0.000	0.000	0.500	0.000	0.000
2	0.000	0.000	0.500	0.000	0.000
3	0.000	0.000	0.000	0.500	0.000
4	0.000	0.000	0.000	0.500	0.000
5	0.000	0.000	0.000	0.000	1.000

TABLE 5.5.3: Search and purchase measures by state

State	**Clicks**	**Pages**	**Products**	**Length**	**Filters**	**Ranking by date**	**Ranking by price Low**	**Ranking by price high**	**Suggested Products**	**Pages per product**	**Number of purchases**	**Monetary value**
1	7.85	3.54	1.12	171.82	0.52	0.01	0.08	0.01	0.04	3.15	0.00	0.00
2	68.18	29.46	10.15	1429.26	4.23	0.04	0.57	0.08	0.34	2.90	0.00	0.00
3	35.12	15.86	4.43	783.59	1.83	0.02	0.35	0.03	0.24	3.58	1.10	73.12
4	39.15	16.86	5.24	857.12	2.06	0.02	0.25	0.04	0.20	3.22	1.44	223.07
5	47.39	20.65	6.64	974.41	2.17	0.02	0.18	0.08	0.20	3.11	2.58	946.86
Total	42.61	18.56	6.08	906.69	2.48	0.02	0.34	0.05	0.22	3.05	0.61	122.45

As can be seen, the first two states are defined exclusively by search activity. Following the work of Moe (2003), who categorized the different types of visits in an online store, State one can be defined as *directed searchers.* In her paper, Moe defined her *directed buyers* as visitors who "exhibit goal-directed search behavior: they have a specific product in mind when entering in the store and, as a result, are unlikely to exit the store without a purchase" (Moe, and Fader, 2004; p.327). They are defined by a little number of products, but a high number of pages with product-related information visited (Moe, 2003). Users in our group differ from

Moe's as they do not purchase in state one, however they seem to have a similar behavior: they do not browse for many products, and they see more pages per product (3.15) than users in the second group (2.90) [Table 5.5.3], i.e. they search for less products, while trying to acquire more information, as they presumably already know what they are looking for. The difference in pages per product seen is statistically significant (p-value = 0.004).

State two represents prospects with an intense search activity (search categories from 2 to 4 in Table 5.5.2) who, however, have not bought and, only later, may or may not convert to a different more purchase-oriented state. They are the ones with the highest search intensity, in terms of clicks, products, pages, length of the session, ranking tools, and filters [Table 5.3.3]. They spend a lot of time browsing among different products in the website. This can be a signal of the fact that they do not have a goal in mind, and still do not know what to search for. Gollwitzer and Bayer (1999) define the state in which people are still trying to understand what to do as *deliberative*. In this state, people search for information in order to learn more about the action they are going to undertake. Moreover, while in this state, prospects click on the highest number of products suggested by the firm (0.34), and undertake all the other search activities more intensively that prospects and customers in all the other states, another signal of the fact that they still do not have a clear idea in mind about what they are looking for. For this reason, we consider users in state two as *deliberative searchers*.

State three is defined mainly by the first purchase and by a low level of spending. Table 5.5.2 shows that while in this state, all customers have made at least a purchase, and 90.2% of them are first buyers. Their search activity is below the average in terms of clicks, pages, products, session length, and filters. Nonetheless, they dig into the pages they visit as they consult the highest number of pages per product (3.58). Both factors suggest the presence of a more goal oriented behavior. Furthermore, they make more intensive use of the ranking tool by ascendant price than prospects and customers in all the other states, signaling their willingness to look for cheaper prices. The average number of purchases of people in state three is 1.10, and the average spending is 73.12€, the lowest among the three purchase states. In summary, while state three buyers are moving their relationship with the firm to a new step, as they make their first purchase, however, they seem to be in an exploration mode as the amount spent is limited and they are mainly driven by a price search. This is in line with Kotler et al (2016)'s definition of trial stage in the new product adoption process: "The consumer tries the new product on a small scale to improve his or her estimate of its value" (Kotler et al., 2016; p. 186). For this reason, we define users in State three as *First Triers.*

Customers in state four have made at least a purchase with this retailer and 23.2% made also a second purchase. The portion of first time is still high (68.0%) as in State three, but unlike customers in state there they spend more. They make on average 1.4 purchases and their average spending is 223.07€ [Table 5.5.3]. Notably most of firms and business press (e.g. Gupta and Zeithaml, 2006; Schweidel, Fader, and Bradlow, 2008; Wangenheim and Bayon, 2007) consider first time buyers as acquired customers. Our findings indicate that first purchase could be just a tentative approach to the relationship with the firm (state three), but it can well be the start of a blossoming relationship (state four), whereby customers are not shy to spend more, to show their commitment by buying more items (on average 1.94 compared to 1.28 of the first triers. The difference is statistically significant, p-value = 0.000), and to start the repeat purchase process, we therefore consider them as *Acquired Customer*s only when they reach this stage.

State five is defined by a high level of expenditure, and does comprehend only some first buyers (34.3%) and second (26.5%) buyers, but also customers who have made more than two purchases (39.0%) [Table 5.5.2]. They make on average 2.58 purchases and their average spending is significantly higher (946.86€) than in state three and four [Table 5.5.3]. In terms of search, they are characterized by a relatively intense search activity. In fact, their number of clicks, pages, and products, and the length of their sessions are above the average. It is interesting to note that, while they are among the customers who rank their search by ascendant price the most, they are also the ones who rank the search results by descendant price the least, suggesting that they are less sensible to price (Suk, Lee, and Lichtenstein, 2002). This is in line with previous research, which demonstrated that the price sensitivity of loyal customers is lower than non-loyal ones (e.g. Krishnamurthy and Raj, 1991; Srinivasan, Anderson, and Ponnavolu, 2002; Guadagni and Little, 2008). We therefore define customers in this terminal as *Loyal*.

5.5.2. Initial State Probabilities

The initial state probabilities are described in Table 5.5.4. Users have a very high probability to start their process being in State two (0.786). This is reasonable, as people in state two are deliberative searchers, who do not have a clear idea about how and what to search for. They do not have any previous experience with the website, and are very likely to be in a deliberative mindset (Gollwitzer and Bayer,1999), eager to gather as much information as possible. Reasonably, the state with the lowest probability to be the initial is the last (0.019), which contains the loyal customers.

TABLE 5.5.4: Initial State Probabilities

State	1 Directed Searchers	2 Deliberative Searchers	3 First Triers	4 Acquired Customers	5 Loyal Customers
Initial Probability	0.088	0.786	0.063	0.044	0.019

In our modeling approach, we allow the initial state probabilities to be function of the same set of covariates affecting the transition probabilities. Table 5.5.5 presents the estimated coefficients. State one has been kept out as benchmark.

TABLE 5.5.5: Covariates affecting the initial state probabilities

	State 2			**State 3**			**State 4**			**State 5**		
	Coeff	**SE**	**t**	**Coeff**	**SE**	**t**	**Coeff**	**SE**	**t**	**Coeff**	**SE**	**t**
Intercept	3.41	0.60	5.67	-6.17	0.38	-16.44	-3.85	0.66	-5.79	-5.91	0.46	-12.90
Add Wish List	1.80	0.95	1.90	1.49	0.95	1.58	1.65	0.95	1.74	1.67	0.95	1.77
Drop Wish List	-1.05	0.88	-1.20	-0.87	0.88	-0.99	-0.96	0.88	-1.09	-0.96	0.88	-1.08
Add Wish List (lag)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Social Registration	0.89	0.44	2.01	-0.24	0.46	-0.51	-0.57	0.48	-1.20	-0.48	0.53	-0.89
Facebook Connected	-0.24	0.49	-0.49	0.55	0.52	1.07	0.16	0.55	0.29	-0.36	0.63	-0.57
Registration	1.63	0.15	10.69	3.42	0.19	18.21	3.51	0.21	16.64	3.18	0.22	14.20
Nl Registration	-0.56	0.23	-2.45	-0.76	0.23	-3.30	-0.86	0.24	-3.67	-1.03	0.25	-4.08
Nl Unsubscription	-0.19	0.65	-0.29	-0.39	0.65	-0.60	-0.11	0.66	-0.16	-0.05	0.67	-0.07
Intersession Days	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Africa	-2.84	0.80	-3.55	0.80	0.73	1.09	-3.04	1.20	-2.54	-3.48	0.02	-231.87
America	-1.94	0.70	-2.77	0.38	0.59	0.64	-1.34	0.72	-1.86	-3.98	0.01	-568.71
Asia	-1.97	0.58	-3.41	1.23	0.20	6.15	-1.28	0.53	-2.41	1.17	0.23	5.00
Europe	-1.23	0.58	-2.14	2.63	0.19	14.19	-0.30	0.53	-0.57	2.31	0.20	11.39
Italy	-2.03	0.58	-3.52	2.36	0.19	12.48	-1.17	0.54	-2.18	2.27	0.20	11.52
Oceania	-2.57	0.68	-3.80	0.36	0.50	0.72	-1.92	0.70	-2.75	1.55	0.48	3.22
Russia	-1.38	0.58	-2.38	1.27	0.24	5.34	-1.55	0.56	-2.79	0.70	0.36	1.94
USA	-1.74	0.58	-2.98	1.68	0.23	7.34	-0.35	0.54	-0.65	2.35	0.23	10.18
Age	-0.01	0.01	-1.57	0.00	0.01	-0.25	0.00	0.01	-0.33	-0.02	0.01	-1.91
Gender (female)	-0.15	0.12	-1.25	-0.31	0.15	-2.15	-0.29	0.16	-1.86	-0.32	0.19	-1.70
Demo missing	-0.31	0.29	-1.09	0.75	0.35	2.17	0.82	0.37	2.18	-0.24	0.44	-0.55

The Wish List usage in session 1 does not have any significant impact on the probability of being in one of the states at the beginning of the acquisition process. The main impact is represented by the type of registration people make in the first session. If they register to the website during the first visit, they are less likely to start from state one compared to all the other states. In particular, they are more likely to start from state three ($\tau_{reg,3}$ = 3.42, $t_{reg,3}$ = 18.21), state four ($\tau_{reg,4}$ = 3.31, $t_{reg,4}$ = 16.64), or state five ($\tau_{reg,5}$ = 3.18, $t_{reg,5}$ = 14.20), put it differently, they are more likely to start from a purchase state. This might be due to the fact that people who decide to make a registration already have a high level of interest and involvement with the company, which makes them more likely to buy. Another possible explanation is that they feel the need to register to the website before making a purchase, even if it is not formally required. On the other hand, users who decide to subscribe the reception of the newsletter during the first session are more likely to be in state one. In particular, they decrease their likelihood to start their acquisition process from one of the purchase states, especially being a loyal customer (state five: $\tau_{sub,5}$ = -1.03, $t_{sub,5}$ = -4.08), an acquired customer (state four: $\tau_{sub,4}$ = -0.86, $t_{sub,4}$ = -3.67), or a first trier (state three: $\tau_{sub,3}$ = -0.76, $t_{sub,3}$ = -3.30). A possible explanation for this result is that, by subscribing to the newsletter, people demonstrate some interest to the firm, but for some reasons they are not willing to interact more with it during the first visit, maybe for time constraints, or maybe because they simply want to postpone their information gathering in the future. Users registering in the website using a social account (e.g. Google+) have a higher probability to be deliberative searchers (state two: $\tau_{soc\_r,2} = 0.89$, $t_{soc\_r,2} = 2.01$). In fact, users who do not have previous experience with the website and are unsure about what to do can decide to use their social account as a way to register because it is easier and more convenient, as it does not require, for example, to create and remember a new password (Bauer et al., 2013), which is in line with one of the main characteristic of deliberative searchers.

5.5.3. Transition Probabilities

Table 5.5.6 represents the transition probabilities matrix. It contains the average probabilities to switch from one state to another in any session *t*. As can be seen, all the states are relatively sticky, as the highest probabilities are in non-transition cells (0.964 in state four, 0.963 in state three, 0.708 in state two, and 0.64 in the first state). Looking at the probabilities in the switching cells, it is more likely for users to move between the two purely search states (0.314 from state one to state two, and 0.230 from state two to state one), rather than moving to the purchase states.

TABLE 5.5.6: Transition Probabilities Matrix

		State *t+1*				
		1	**2**	**3**	**4**	**5**
State *t*	**1**	0.640	0.314	0.027	0.015	0.005
	2	0.230	0.708	0.031	0.024	0.006
	3	0.000	0.000	0.963	0.035	0.002
	4	0.000	0.000	0.000	0.964	0.036
	5	0.000	0.000	0.000	0.000	1.000

The three purchase states do not allow customers to move back in the process. For example, once a prospect becomes a trier, she cannot go back to a purely search state, but she will be a trier until she makes a subsequent purchase, which eventually makes her moving on and become an acquired, or a loyal customer. In the same way, an acquired customer cannot go back to the trial state, but she will only be allowed to move to the loyal stage.

TABLE 5.5.7: Covariates affecting the transition probabilities

	1 > 2			1>3			1>4			1>5		
	Coeff	**SE**	**t**	**Coeff**	**SE**	**t**	**Coeff**	**SE**	**t**	**Coeff**	**SE**	**t**
Intercept	-1.08	0.49	-2.18	-4.98	0.32	-15.50	-4.86	0.47	-10.33	-5.65	0.56	-10.04
Add Wish List	1.24	0.08	15.16	1.09	0.09	11.72	1.18	0.09	12.52	1.22	0.10	12.73
Drop Wish List	1.27	0.27	4.77	1.41	0.29	4.84	1.47	0.30	4.87	1.50	0.32	4.73
Add Wish List (lag)	-0.26	0.32	-0.81	-0.21	0.75	-0.28	-0.12	0.87	-0.14	-0.22	1.25	-0.17
Social Registration	0.09	0.13	0.70	0.08	0.24	0.34	-0.15	0.32	-0.47	-0.17	0.53	-0.31
Facebook Connected	-0.04	0.16	-0.24	0.06	0.29	0.20	0.15	0.41	0.37	0.28	0.66	0.42
Registration Days	0.00	0.00	-1.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.01	-0.67
Nl Registration Days	0.00	0.00	0.00	0.00	0.00	-0.50	0.00	0.01	-0.20	0.00	0.01	0.13
Nl Unsubscription Days	0.00	0.00	0.50	0.00	0.01	0.14	0.01	0.01	0.56	0.00	0.02	0.21
Registration	1.13	0.09	13.15	2.58	0.13	19.22	2.49	0.18	14.01	2.01	0.28	7.19
Nl Registration	-0.18	0.16	-1.13	-0.41	0.21	-1.96	-0.30	0.26	-1.13	-0.29	0.45	-0.66
Nl Unsubscription	-0.20	0.21	-0.96	-0.67	0.40	-1.67	-0.55	0.52	-1.06	-1.38	1.35	-1.02
Intersession Days	0.03	0.00	13.50	0.02	0.00	3.75	0.00	0.01	0.22	0.01	0.01	1.75
Africa	0.52	0.68	0.75	0.02	1.38	0.01	-1.36	2.38	-0.57	-0.78	2.48	-0.32
America	-0.85	0.53	-1.60	-0.63	0.72	-0.88	0.05	0.54	0.10	0.62	0.57	1.09
Asia	0.06	0.49	0.12	0.36	0.27	1.32	0.35	0.39	0.88	0.06	0.42	0.15
Europe	0.00	0.49	0.01	0.94	0.25	3.75	0.45	0.38	1.19	-0.25	0.40	-0.61
Italy	-0.07	0.49	-0.14	1.09	0.26	4.23	0.08	0.40	0.19	-0.89	0.47	-1.90
Oceania	-0.14	0.54	-0.26	0.69	0.44	1.58	0.58	0.58	1.01	-0.31	0.91	-0.35
Russia	0.13	0.49	0.26	0.51	0.26	1.96	-0.08	0.41	-0.19	-0.64	0.45	-1.44
USA	-0.33	0.49	-0.67	0.49	0.27	1.82	0.51	0.39	1.31	0.32	0.42	0.77
Age	0.00	0.00	-1.33	0.00	0.01	0.17	0.00	0.01	-0.50	0.01	0.01	0.75
Gender (female)	0.19	0.06	3.40	0.05	0.11	0.48	0.11	0.15	0.71	-0.05	0.23	-0.21
Demo missing	-0.19	0.12	-1.59	0.33	0.26	1.27	0.01	0.34	0.02	0.45	0.51	0.88

	2>1			2>3			2>4			2>5		
	Coeff	**SE**	**t**	**Coeff**	**SE**	**t**	**Coeff**	**SE**	**t**	**Coeff**	**SE**	**t**
Intercept	-0.91	0.43	-2.12	-4.84	0.78	-6.23	-4.61	0.74	-6.20	-6.36	0.51	-12.56
Add Wish List	-1.24	0.08	-16.08	-0.09	0.02	-3.95	-0.02	0.02	-1.50	-0.07	0.04	-1.70
Drop Wish List	-0.88	0.25	-3.46	-0.02	0.07	-0.32	0.05	0.04	1.30	0.12	0.04	2.75
Add Wish List (lag)	0.04	0.01	4.67	-0.03	0.02	-1.94	-0.03	0.02	-1.88	-0.01	0.03	-0.41
Social Registration	0.13	0.10	1.25	-0.33	0.15	-2.23	-0.59	0.17	-3.46	-1.30	0.44	-2.94
Facebook Connected	0.15	0.13	1.15	0.34	0.18	1.87	0.51	0.21	2.40	1.41	0.50	2.82
Registration Days	0.00	0.00	2.00	0.01	0.00	2.50	0.00	0.00	-1.00	0.00	0.01	0.40
Nl Registration Days	0.00	0.00	-0.50	0.00	0.00	0.33	0.00	0.00	0.75	-0.01	0.01	-1.00
Nl Unsubscription Days	0.01	0.00	2.33	0.01	0.01	1.00	-0.01	0.01	-0.62	0.01	0.02	0.44
Registration	-0.52	0.07	-7.80	1.86	0.08	24.14	1.71	0.08	21.94	1.78	0.15	11.68
Nl Registration	0.05	0.11	0.42	-0.16	0.10	-1.65	-0.56	0.12	-4.64	-0.26	0.20	-1.34
Nl Unsubscription	0.28	0.17	1.63	-0.18	0.20	-0.94	-0.30	0.23	-1.33	-0.32	0.47	-0.68
Intersession Days	-0.02	0.00	-16.00	-0.02	0.00	-4.00	-0.01	0.00	-3.33	-0.01	0.01	-2.00
Africa	-0.13	0.57	-0.23	-0.25	1.15	-0.22	-2.80	2.56	-1.09	-1.69	2.14	-0.79
America	-0.12	0.46	-0.26	0.38	0.82	0.47	-0.38	0.81	-0.46	-0.19	0.73	-0.26
Asia	-0.23	0.42	-0.55	0.29	0.77	0.37	0.35	0.73	0.47	0.14	0.43	0.33
Europe	0.02	0.42	0.04	1.05	0.76	1.37	0.55	0.73	0.76	0.31	0.42	0.73
Italy	0.02	0.42	0.06	1.51	0.77	1.97	0.29	0.73	0.39	-0.66	0.46	-1.42
Oceania	0.28	0.47	0.58	0.17	0.84	0.20	0.68	0.77	0.88	-1.09	1.03	-1.05
Russia	-0.26	0.42	-0.62	0.22	0.77	0.29	-0.25	0.73	-0.34	-0.50	0.44	-1.12
USA	0.16	0.43	0.37	0.53	0.77	0.69	0.87	0.73	1.20	1.25	0.43	2.93
Age	0.00	0.00	0.00	0.00	0.00	1.00	0.01	0.00	2.75	0.02	0.01	2.86
Gender (female)	0.02	0.05	0.36	0.04	0.07	0.61	-0.14	0.07	-1.97	-0.06	0.13	-0.48
Demo missing	-0.07	0.11	-0.64	0.55	0.16	3.48	0.79	0.17	4.55	1.32	0.33	4.05

	3>4			3>5			4>5		
	Coeff	**SE**	**t**	**Coeff**	**SE**	**t**	**Coeff**	**SE**	**t**
Intercept	-4.56	0.24	-18.78	-7.65	0.93	-8.21	-5.97	0.27	-21.79
Add Wish List	0.05	0.03	1.68	0.01	0.13	0.04	0.07	0.02	3.48
Drop Wish List	0.20	0.06	3.56	0.32	0.10	3.24	0.06	0.05	1.33
Add Wish List (lag)	0.03	0.03	0.82	-0.12	0.33	-0.37	0.06	0.02	2.59
Social Registration	0.14	0.21	0.68	0.03	1.05	0.02	-0.27	0.27	-1.00
Facebook Connected	-0.42	0.28	-1.49	-0.04	1.30	-0.03	0.00	0.33	0.01
Registration Days	0.00	0.00	0.50	0.01	0.01	1.14	0.00	0.00	-1.50
Nl Registration Days	0.00	0.00	-1.50	-0.03	0.02	-1.39	0.00	0.00	-0.67
Nl Unsubscription Days	-0.01	0.01	-0.86	-0.03	0.06	-0.45	0.01	0.01	1.00
Registration	0.79	0.16	4.82	1.42	0.58	2.44	1.41	0.16	8.96
Nl Registration	-0.09	0.29	-0.30	-0.32	1.10	-0.29	-0.26	0.35	-0.76
Nl Unsubscription	0.10	0.44	0.22	-2.97	0.00	-991.00	-0.35	0.60	-0.59
Intersession Days	0.01	0.00	1.75	-0.03	0.04	-0.86	0.01	0.00	2.50
Africa	1.73	0.55	3.16	-3.07	0.00	-1537.00	2.75	0.94	2.94
America	0.40	0.53	0.76	-3.04	0.01	-379.38	1.49	0.47	3.17
Asia	1.13	0.16	6.91	0.04	0.58	0.07	2.04	0.17	12.01
Europe	1.20	0.14	8.81	0.37	0.39	0.94	2.19	0.16	13.58
Italy	1.06	0.14	7.39	-0.93	0.58	-1.60	1.93	0.18	10.52
Oceania	1.41	0.49	2.87	1.69	1.02	1.66	2.06	0.33	6.18
Russia	0.68	0.17	4.12	-1.65	0.99	-1.68	1.67	0.20	8.54
USA	1.31	0.17	7.68	1.01	0.48	2.11	2.04	0.18	11.59
Age	0.00	0.01	0.00	0.02	0.02	0.95	0.01	0.01	1.67
Gender (female)	-0.22	0.10	-2.29	0.39	0.47	0.82	-0.10	0.10	-0.96
Demo missing	0.13	0.24	0.52	0.88	1.02	0.86	0.38	0.27	1.40

As explained previously, our model allows the transition probabilities to be function of time-varying covariates. Table 5.5.7 presents the estimated coefficients. Non-transition scenarios have been kept out as benchmark. It shows that directed searchers (users in state one) who add items in the Wish List are more likely to become deliberative searchers ($\beta_{aWL,12} = 1.24$, $t_{aWL,12} = 15.16$), or purchasers ($\beta_{aWL,15} = 1.22$, $t_{aWL,15} = 12.73$ to become loyal, $\beta_{aWL,14} = 1.18$, $t_{aWL,14} = 12.52$ to become acquired, and $\beta_{aWL,13} = 1.09$, $t_{aWL,13} = 11.72$ to become triers). Also dropping items from the Wish List increases significantly their propensity to move to a purchase state ($\beta_{dWL,14} = 1.47$, $t_{dWL,14} = 4.87$ to become acquired, $\beta_{dWL,13} = 1.41$, $t_{dWL,13} = 4.84$ to become triers, and $\beta_{dWL,15} = 1.50$, $t_{dWL,15} = 4.73$ to become loyal) or to become deliberative searchers ($\beta_{dWL,12} = 1.27$, $t_{dWL,12} = 4.77$). This may be due to the fact that the Wish List usage help users in make their search more organized and efficient (Close and Kukar-Kinney, 2010), and this feature is particularly useful for people who need to buy a particular product, but do not have a lot of time to spend on the website. Dropping items from the Wish List is a signal that they made a choice and move the product to the actual cart, discarding the items they are not willing to buy. The Wish List effect is reversed for deliberative searchers (state two): adding items into the Wish List has a detrimental effect on the probability to move to state three ($\beta_{aWL,23}$ = -0.09, $t_{aWL,23}$ = -3.95). This is because they dedicate more time in searching activity, and add items in Wish List in order to postpone their evaluation (Popovic and Hamilton, 2014). However, delaying the buying decision, leads people to re-evaluate their choices, decreasing their willingness to buy the items they put into the Wish List in previous sessions (Popovic and Hamilton, 2014). However, dropping items from the Wish List in state two increases the probability to move directly to state five ($\beta_{dWL,25}$ =0.12, $t_{dWL,25}$ = 2.75). For those who have already made a first purchase, adding products in the Wish List does not have any effect on the probability to move to a subsequent state, but dropping products makes triers more likely to move to state four ($\beta_{dWL,34}$ =0.20, $t_{dWL,34}$ = 3.56) or to be in state five ($\beta_{dWL,35}$ =0.32, $t_{dWL,35}$ = 3.24). A possible explanation for this result is that, as in state three people try to make a first purchase with a low expenditure level, they can have left in the Wish List other items they were willing to purchase, in order to buy them only after evaluating their first purchase experience within the firm. In this case, they drop items from the Wish List just to move them into the actual cart for the checkout. In shifting from state four to state five, adding products in the Wish List in both the previous and the current session has a positive effect ($\beta_{aWL,45}$ = 0.07, $t_{aWL,45}$ = 3.48; $\beta_{aWLl,45}$ = 0.06, $t_{aWLl,45}$ = 2.59), maybe because they are more knowledgeable about the firm and its products, and adding products in the Wish List means they have a stronger intention to buy them in future.

The action of registering to the website increases the probability to move on in the process from any state, especially from non-purchase states to states three and four ($\beta_{reg,13}$ = 2.58, $t_{reg,13}$ = 19.22; $\beta_{reg,14}$ = 2.50, $t_{reg,14}$ = 14.01; $\beta_{reg,23}$ = 1.86, $t_{reg,23}$ = 24.14; $\beta_{reg,24}$ = 1.71, $t_{reg,24}$ = 21.94). It also decreases the probability of moving from state two to state one ($\beta_{reg,21}$ = -0.52, $t_{reg,24}$ = -7.80). This is an unsurprising result, as the act of registration first is a signal of a high level of interest toward the firm, and it also indicates that the user trusts the company enough to share her information (Schumann, Wangenheim, and Groene, 2014; Martin, and Murphy, 2017), which are two factors that enhance the probability to purchase ant to create a long-term relationship with the firm (Spekman, 1988; Nooteboom, Berger, and Noorderhaven, 1997; Garbarino and Johnson 1999). Also the number of days passed from the registration has an effect: the higher the number of days, the higher is the probability for users in state two to make their first trial and move to state three ($\beta_{reg\_d,32} = 0.005$, $t_{reg\ g\_d,32} = 2.50$). It is interesting to note that users in state two who register with a social account (Google +) have a lower probability to move on in the process ($\beta_{soc,23} = -0.53$, $t_{soc,23} = -2.23$; $\beta_{soc,24} = -0.59$, $t_{soc,24} = -3.46$; $\beta_{soc,25} = -1.30$, $t_{soc,25} = -2.94$). As explained before, users who are unsure about what to do can use their social account to register because it does not require a lot of effort (Bauer et al., 2013) and knowledge, and, consequently, users who do not know much about the company and the functioning of its e-commerce may be less willing to make a purchase. However, for users who register through Facebook the effect is different: they are more likely to move from state two to higher states of purchase, as state four ($\beta_{Fb,34} = 0.51$, $t_{Fb,34} = 2.40$) and five ($\beta_{Fb,35} = 1.41$, $t_{Fb,35} = 2.82$). One possible explanation of this difference between Google+ and Facebook sign-ins can be the amount of personal information that people are sharing on the two social media platforms. Owyang (2012) states that, even if Google+ has more or less the same technical characteristics as Facebook, it suffers from a perception issue, and it is continuously outclassed by its competitor. This means that people share much more personal information on Facebook than on Google+, and the cost of privacy due to the Facebook registration are higher (Bauer et al., 2013). Thus, people who agree to share such an amount of personal information with the firm are more likely to establish a deeper relationship with it.

The subscription to the newsletter reception has a negative effect on the probability to move from state one to state three ($\beta_{sub\_d,13} = -0.41$, $t_{sub\_d,13} = -1.96$), and the number of days the user receives the newsletter has a detrimental effect on the probability to move from state two to four ($\beta_{sub,34}$ = -0.56, $t_{sub,34}$ = -4.66). Moreover, the number of days passing after the newsletter unsubscription increases people's likelihood to search less and to pass from state two to one ($\beta_{uns\_d,21}$ = 0.007, $t_{uns\_d,21}$ = 2.33). These results put some doubts about the efficacy of the firm's newsletter campaigns on prospects. While the negative effect of the subscription day can be explained by the delay effect discussed previously, the negative effect of the subscription duration seems to highlight the prospects' unwillingness to receive e-mail communications. This negative reaction to e-mails can be due to the reactance effect (Brehm, 1966), according to which when people feel forced to do something (in this case, they feel as the firm is trying to convince them to buy by overwhelming them with e-mail communications), they react engaging in the opposite way (in this case, they do not buy).

Finally, the number of days that elapse between session *t* and session *t+1* have a positive effect in moving on from state one to state two ($\beta_{int,12}$ = 0.02, $t_{int,12}$ = 13.50) and three ($\beta_{int,13}$ = 0.02, $t_{int,13}$ = 3.75), and from state four to five ($\beta_{int,45}$ = 0.10, $t_{int,45}$ = 2.50), and a negative effect in moving from state two to any other further state ($\beta_{int,23}$ = -0.02, $t_{int,23}$ = 4.00; $\beta_{int,24}$ = -0.01, $t_{int,24}$ = -3.53; $\beta_{int,25}$ = -0.01, $t_{int,25}$ = -2.00). This is because people who do not engage much in searching activity within the website need more days to search more, while people in state two, who have sessions in which they deeply explore the website and its products, can be closer to a conversion and need to make a subsequent session earlier in order to purchase, and people who have just purchased (the ones in state four) may need more time to purchase again and move to the loyal stage.

5.5.4. State Dependent Probabilities

In our modeling approach, we allow the state dependent probabilities to be function of a set of time-varying covariates supposed to have a short-term effect on the outcomes, thus concerning the distinctive characteristics of the session. Table 5.5.8 presents the estimated effects of those covariates on the search intensity, the probability to make a purchase, and the monetary value of the transaction.

TABLE 5.5.8: Covariates affecting the **s**tate dependent probabilities

	State 1: Search			**State 2: Search**			**State 3: Search**			**State 3: Purchase**		
	Coeff	**SE**	**t**	**Coeff**	**SE**	**t**	**Coeff**	**SE**	**t**	**Coeff**	**SE**	**t**
Intercept	-0.543	0.00	-131.48	0.019	0.04	0.53	-0.037	0.04	-0.96	0.248	0.18	1.39
Affiliation	0.004	0.00	2.11	-0.046	0.02	-2.92	-0.014	0.02	-0.78	0.728	0.08	9.47
Social	0.004	0.00	1.72	-0.086	0.03	-3.13	-0.090	0.03	-3.45	-0.545	0.14	-3.95
Direct	-0.001	0.00	-0.55	0.004	0.02	0.25	0.003	0.02	0.14	0.233	0.09	2.58
Referral	-0.002	0.00	-0.59	-0.037	0.02	-1.59	-0.128	0.02	-5.61	-0.205	0.11	-1.95
Retargeting	0.003	0.00	1.36	-0.065	0.02	-3.31	-0.110	0.02	-5.56	-0.555	0.10	-5.34
Newsletter	-0.004	0.00	-2.01	-0.162	0.02	-7.61	-0.133	0.02	-7.47	-1.330	0.11	-12.55
SEM	-0.004	0.00	-2.03	-0.042	0.01	-2.84	-0.040	0.02	-2.49	-0.013	0.07	-0.17
Mobile	0.003	0.00	1.17	0.259	0.02	10.61	0.007	0.03	0.22	-0.405	0.12	-3.52
Desktop	0.001	0.00	0.40	-0.021	0.02	-0.94	-0.080	0.03	-2.76	-0.321	0.11	-3.03
Ranking by date	0.051	0.01	8.07	0.192	0.02	12.72	0.397	0.03	12.93	-0.050	0.16	-0.31
Ranking by price high	0.053	0.01	10.15	0.187	0.01	17.48	0.242	0.02	13.15	-0.021	0.09	-0.24
Ranking by price low	0.045	0.00	26.69	0.148	0.00	50.51	0.203	0.00	44.81	0.088	0.02	4.33
Suggested Products	0.021	0.00	18.74	0.083	0.00	40.88	0.095	0.00	35.24	0.037	0.01	2.95
Filters	0.020	0.00	55.83	0.042	0.00	74.11	0.041	0.00	47.69	0.039	0.00	9.75
Nl promotional	0.005	0.00	2.77	0.032	0.02	1.93	0.054	0.02	3.41	0.040	0.07	0.54
Discount depth	0.000	0.00	0.05	0.001	0.00	1.64	-0.001	0.00	-2.85	-0.014	0.00	-5.35
February17	0.004	0.00	2.12	-0.106	0.02	-6.35	-0.099	0.01	-6.82	-1.225	0.08	-14.90
March17	0.002	0.00	0.93	-0.128	0.02	-7.03	-0.058	0.02	-3.45	-0.876	0.09	-9.85
April17	0.005	0.00	2.38	-0.115	0.02	-6.73	-0.103	0.02	-6.24	-1.564	0.09	-16.87
May17	-0.001	0.00	-0.19	-0.024	0.04	-0.59	-0.149	0.03	-4.45	-2.043	0.21	-9.87
October16	-0.016	0.00	-4.49	-0.046	0.03	-1.36	-0.330	0.60	-0.55	20.000	13.73	1.46
November16	-0.017	0.00	-7.90	-0.154	0.03	-5.95	0.655	0.62	1.06	20.450	14.64	1.40
December16	-0.002	0.00	-1.18	0.074	0.02	3.79	-0.285	0.26	-1.11	19.900	6.06	3.28
Sessions	26897			42711			18402			18402	18402	
Number of users	5123			10116			2767			2767	2767	
R-squared within	0.184			0.279			0.343					
R-squared overall	0.174			0.245			0.330					
R-squared between	0.132			0.241			0.327					
Log-likelihood										-7330	-7330	

	State 3: Monetary value			**State 4: Search**			**State 4: Purchase**			**State 4: Monetary value**		
	Coeff	**SE**	**t**	**Coeff**	**SE**	**t**	**Coeff**	**SE**	**t**	**Coeff**	**SE**	**t**
Intercept	2.191	0.10	22.15	-0.051	0.05	-1.12	-0.353	0.17	-2.03	2.299	0.12	19.48
Affiliation	0.501	0.04	11.49	-0.016	0.02	-0.75	0.455	0.08	5.52	0.407	0.06	7.33
Social	-0.203	0.06	-3.20	-0.108	0.03	-3.24	-0.390	0.15	-2.58	-0.170	0.09	-1.97
Direct	0.063	0.05	1.19	-0.021	0.02	-0.87	0.271	0.09	3.12	0.157	0.06	2.49
Referral	-0.138	0.06	-2.47	-0.067	0.03	-2.41	0.187	0.11	1.76	0.145	0.07	2.00
Retargeting	-0.184	0.05	-3.81	-0.120	0.02	-5.50	-0.747	0.10	-7.47	-0.324	0.06	-5.69
Newsletter	-0.494	0.04	-11.33	-0.135	0.02	-6.65	-1.112	0.10	-11.52	-0.469	0.05	-8.88
SEM	-0.016	0.04	-0.40	0.005	0.02	0.26	0.106	0.08	1.40	0.102	0.05	2.00
Mobile	-0.337	0.08	-4.17	0.082	0.04	2.22	-0.155	0.12	-1.31	-0.253	0.10	-2.60
Desktop	-0.292	0.08	-3.84	-0.054	0.03	-1.60	-0.373	0.11	-3.49	-0.377	0.09	-4.18
Ranking by date	0.001	0.07	0.02	0.234	0.03	7.65	-0.249	0.16	-1.58	-0.090	0.08	-1.14
Ranking by price high	-0.004	0.04	-0.10	0.130	0.02	6.70	-0.086	0.08	-1.11	-0.020	0.05	-0.40
Ranking by price low	0.062	0.01	5.71	0.198	0.00	40.74	0.119	0.02	5.67	0.093	0.01	7.37
Suggested Products	0.019	0.01	2.92	0.084	0.00	27.35	0.029	0.01	2.40	0.020	0.01	2.47
Filters	0.025	0.00	11.88	0.055	0.00	60.37	0.046	0.00	12.32	0.037	0.00	15.64
Nl promotional	-0.005	0.04	-0.13	0.026	0.02	1.35	0.041	0.08	0.52	-0.011	0.05	-0.21
Discount depth	-0.005	0.00	-4.59	0.000	0.00	-0.33	-0.003	0.00	-1.31	-0.001	0.00	-0.63
February17	-0.646	0.04	-18.30	-0.112	0.02	-6.79	-0.911	0.07	-12.31	-0.704	0.04	-16.41
March17	-0.635	0.04	-15.49	-0.117	0.02	-6.32	-0.892	0.08	-10.67	-0.807	0.05	-16.71
April17	-0.921	0.04	-22.52	-0.131	0.02	-7.01	-1.357	0.09	-15.93	-1.072	0.05	-21.92
May17	-1.218	0.08	-14.85	-0.143	0.03	-4.13	-1.451	0.17	-8.69	-1.261	0.09	-14.00

October16	3.179	1432.00	0.00	-0.196	0.48	-0.41	19.370	5.86	3.31	3.999	1243.00	0.00
November16	2.987	1484.00	0.00	-0.473	0.34	-1.40	19.500	4.20	4.64	3.761	0.88	4.28
December16	2.722	0.63	4.33	0.772	0.23	3.40	19.670	2.89	6.81	3.805	0.59	6.45
Sessions	18402			18382			18382			18382		
Number of users	2767			2434			2434			2434		
R-squared within	0.0934			0.327						0.0855		
R-squared overall	0.0824			0.318						0.0707		
R-squared between	0.0770			0.338						0.0758		
Log-likelihood							-7467					

	State 5: Search			State 5: Purchase			State 5: Monetary Value		
	Coeff	**SE**	**t**	**Coeff**	**SE**	**t**	**Coeff**	**SE**	**t**
Intercept	0.317	0.08	4.15	0.169	0.23	0.73	3.087	0.20	15.28
Affiliation	-0.050	0.04	-1.31	0.417	0.12	3.48	0.431	0.10	4.33
Social	-0.126	0.06	-2.13	-0.785	0.24	-3.31	-0.444	0.15	-2.88
Direct	0.067	0.04	1.77	0.186	0.11	1.63	0.113	0.10	1.14
Referral	-0.142	0.05	-2.83	0.294	0.16	1.86	0.218	0.13	1.66
Retargeting	-0.106	0.04	-2.70	-0.599	0.14	-4.22	-0.324	0.10	-3.15
Newsletter	-0.125	0.04	-3.46	-0.373	0.12	-3.06	-0.273	0.09	-2.90
SEM	-0.037	0.03	-1.06	0.197	0.11	1.82	0.115	0.09	1.28
Mobile	-0.088	0.06	-1.41	-0.036	0.17	-0.21	-0.048	0.17	-0.29
Desktop	-0.173	0.06	-3.03	-0.735	0.16	-4.71	-0.731	0.15	-4.78
Ranking by date	0.309	0.05	6.72	-0.085	0.18	-0.48	-0.010	0.12	-0.08
Ranking by price high	0.149	0.02	9.14	0.029	0.05	0.60	0.031	0.04	0.73
Ranking by price low	0.222	0.01	17.62	0.051	0.04	1.29	0.030	0.03	0.93
Suggested Products	0.081	0.01	14.38	-0.005	0.02	-0.25	-0.005	0.01	-0.34
Filters	0.060	0.00	38.84	0.043	0.00	9.08	0.043	0.00	10.70
Nl promotional	-0.010	0.03	-0.31	-0.147	0.11	-1.34	-0.088	0.09	-1.03
Discount depth	-0.002	0.00	-3.16	-0.001	0.00	-0.33	0.000	0.00	0.19
February17	-0.087	0.03	-2.79	-0.873	0.10	-8.64	-0.900	0.08	-11.10
March17	-0.110	0.03	-3.34	-0.990	0.11	-9.08	-1.080	0.09	-12.53
April17	-0.159	0.03	-4.89	-1.309	0.11	-12.23	-1.337	0.09	-15.64
May17	-0.258	0.06	-4.68	-1.899	0.21	-8.92	-1.710	0.14	-11.88
October16	0.000	0.00	0.00				0.000	0.00	0.00
November16	0.000	0.00	0.00				0.000	0.00	0.00
December16	-0.376	0.60	-0.62	20.230	8.86	2.28	4.730	1569.00	0.00
Sessions	9197			9197			9197		
Number of users	1041			1041			1041		
R-squared within	0.276						0.0665		
R-squared overall	0.260						0.0653		
R-squared between	0.234						0.0936		
Log-likelihood				-4175					

We can distinguish the last touch marketing variables in Table 5.5.8 into three groups: firm-initiated touchpoints, represented by Retargeting, Newsletter, and Social; customer-initiated touchpoints, comprising Direct, SEM, and SEO (which has been kept out as baseline), and hybrid-initiated touchpoints, such as Affiliation and Referral.

Results show that there is an overall negative effect of the firm-initiated touchpoints on all the three outcomes. Retargeting decreases the search activity, the purchase probability, and the monetary value of the transaction mainly in states three ($\gamma_{ret,search3}$ = -0.11, $t_{ret,search3}$ = -5.56; $\gamma_{ret,purch3}$ = -0.56, $t_{ret,purch3}$ = -5.34; $\gamma_{ret,mv3}$ = -0.18, $t_{ret,mv3}$ = -3.18), and four ($\gamma_{ret,search4}$ = -0.12,

$t_{ret,search4}$ = -5.50; $\gamma_{ret,purch4}$ = -0.75, $t_{ret,purch4}$ = -7.47; $\gamma_{ret,mv4}$ = -0.32, $t_{ret,mv4}$ = -5.69). It also has a negative impact on search in states two ($\gamma_{ret,search2}$ = -0.06, $t_{ret,search2}$ = -3.31) and five ($\gamma_{ret,search5}$ = -0.11, $t_{ret,search2}$ = -2.70), and a negative impact on the transactional outcomes in state five ($\gamma_{ret,purch5}$ = -0.60, $t_{ret,purch5}$ = -4.22; $\gamma_{ret,mv5}$ = -0.32, $t_{ret,mv5}$ = -3.15). The impact of the newsletter follows the same fashion, decreasing the search activity, the purchase probability, and the monetary value of the transaction in all the states, mainly in three ($\gamma_{nl,search3}$ = -0.13, $t_{nl,search3}$ = -7.47; $\gamma_{nl,purch3}$ = -1.33, $t_{nl,purch3}$ = -12.55; $\gamma_{nl,mv3}$ = -0.50, $t_{nl,mv3}$ = -11.33), and four ($\gamma_{nl,search4}$ = -0.14, $t_{nl,search4}$ = -6.65; $\gamma_{nl,purch4}$ = -1.11, $t_{nl,purch4}$ = -11.52; $\gamma_{nl,mv4}$ = -0.47, $t_{nl,mv4}$ = -8.88). Social activity does not perform differently, having a negative effect on search activity mainly in states three ($\gamma_{soc,search3}$ = -0.09, $t_{soc,search3}$ = -3.45) and four ($\gamma_{soc,search4}$ = -0.11, $t_{soc,search4}$ = -3.24), and a negative effect on the transactional outcomes in all the three purchase states, especially the third ($\gamma_{soc,purch3}$ = -0.55, $t_{soc,purch3}$ = -3.95; $\gamma_{soc,mv3}$ = -0.20, $t_{soc,mv3}$ = -3.20).

This means that firm-initiated activities work worse than SEO in triggering the search and purchase activities of prospects and customers. This is in line with the results of Anderl et al. (2016), who find that firm-initiated contacts have a negative effect on conversions than customer-initiated channels, and of Li and Kannan (2014), who demonstrate that retargeting and display advertising can have a detrimental effect on the purchase likelihood. Moreover, users increasingly perceive firm-initiated channels as intrusive (Edwards, Li, and Less, 2002; Goldfarb and Tucker, 2011) and unwanted (Blattberg et al., 2008). It is interesting to note that the worst performances are observed in the earlier stages of the purchase activity (first triers and newly acquired customers). This can be a matter of the degree of personalization of the communication. In fact, retargeting proposes advertising contents related to the customers' previous search and purchase activities (e.g. Lambrecht et al., 2013; Barjas et al., 2016), as well as the content of some newsletter, which are based on the previous purchases (e.g. Kumar, Morris and Pancras, 2008; Song et al., 2016). Our findings are in line with previous research, which demonstrates that personalized messages are more effective than general ones (e.g. Malthouse and Elsner, 2006; Hauser et al., 2009; Tucker, 2011), but only if the recipient is in the later stages of the purchase funnel (Lambrecht et al., 2013; Abhishek et al., 2016). However, this reasoning is valid only for the purchase states, not for the search ones. In fact, we found the negative effect to be lower or not significant in states one and two. This is in line with Court et al. (2009), who state that brand advertising should be provided in the first part of the consumer decision journey, in order to create awareness in the potential customers and to help them in their information seeking.

On the other hand, customer-initiated touchpoints show a positive effect. Directly typing the website address in the address bar of the browser increases both users' probability to purchase and the expenditure level in state three ($\gamma_{dir,purch3} = 0.23$, $t_{dir,purch3} = 2.58$; $\gamma_{dir,purch4} = 0.27$, $t_{dir,purch4} = 3.12$), and the monetary value of acquired customers ($\gamma_{dir,mv4} = 0.16$, $t_{dir,mv4} = 2.49$). Search levels are not significantly different from the SEO ones. The impact of SEM on search activity is lower than the SEO one in the first three states ($\gamma_{sem,search1} = -0.004$, $t_{sem,search1} = -2.03$; $\gamma_{sem,search2} = -0.04$, $t_{sem,search2} = -2.84$; $\gamma_{sem,search3} = -0.04$, $t_{sem,search3} = -2.49$), but it increases the monetary value of acquired customers ($\gamma_{sem,mv4} = 0.10$, $t_{sem,\ mv4} = -2.00$). Results are in line with previous studies, which demonstrate that customer-initiated touchpoints are more effective in increasing purchase probabilities (e.g. Wiesel et al., 2011; Jerath and Park, 2014; De Haan et al., 2016; Anderl et al., 2016). Jerath and Park (2014) demonstrate that the use of paid search is associated to customers who are closer to purchase. This can explain the lower impact on search of SEM over SEO in the first phases of the process, and its positive impact on the monetary value of acquired customers.

Affiliation and referrals do not have a clear definition in literature. De Haan et al. (2016) consider affiliation and referrals as customer-initiated contacts, as a user's landing on an affiliate website means that the brand offer is highly related to the user's need, thus the user might already be interested in the product category and looking for related information. Ghose and Todri (2015) associate affiliation to display advertising (firm-initiated contacts), and Anderl et al. (2016) group them into the "customer/firm-initiated touchpoints". Referrals are even more ambiguous, since they do not have any rule to be embedded in a website. Li and Kannan (2014) consider referrals as customer-initiated touchpoints, as they compare it to websites like TripAdvisor, where customers arrive if they are searching for that particular service or something related to it. Our results show that the effect of affiliation is very similar to the one of a customer-initiated touchpoint. In fact, it has a positive impact on search in state one ($\gamma_{aff,search1} = 0.004$, $t_{aff,search1} = 2.11$), and negative in state two ($\gamma_{aff,search2} = -0.05$, $t_{aff,search2} = -2.92$). This is not necessarily a bad result, as while in state one it seems to increase people's interest and propensity to look for information, for deliberative searchers affiliation help them in focusing their search activity. Coming from affiliate websites has also an overall positive effect on the purchase probability and monetary value in each of the purchase states, especially in states three ($\gamma_{aff,purch3} = 0.73$, $t_{aff,purch3} = 9.47$; $\gamma_{aff,mv3} = 0.50$, $t_{aff,mv3} = 11.49$) and four ($\gamma_{aff,purch4} = 0.46$, $t_{aff,purch4} = 5.52$; $\gamma_{aff,mv4} = 0.41$, $t_{aff,mv4} = 7.33$). Referrals, on the other hand, has a significantly lower impact on search than SEO in all the purchase states, especially the third ($\gamma_{ref,search3} = -0.13$, $t_{ref,search3} = -5.61$), it does not have any effect on the purchase probability, but it decreases the spending of first triers ($\gamma_{ref,mv3} = -0.14$, $t_{ref,mv3} = -2.47$) and increases the spending of acquired customers ($\gamma_{ref,mv4} = 0.15$, $t_{ref,mv4} = 2.00$). This behavior is similar to a firm-initiated touchpoint. In fact, even if they are not directly managed by the company, referrals can be found in websites not closely related to the firm. In this case, customers incur in them just by chance, when they are not explicitly looking for the firm or its products.

Regarding the mobile, we keep tablet as the benchmark, as it was the device with the lowest percentage of use. As expected, users who enter in the website through mobile (smartphone) are more prone to search, especially the deliberative searchers ($\gamma_{mob,search2}$ = 0.26, $t_{mob,search2}$ = 10.62), and the acquired customers ($\gamma_{mob,search4}$ = 0.08, $t_{mob,search4}$ = 2.22), but smartphone users who are in state three and four has a lower probability to purchase and to spend money ($\gamma_{mob,purch3}$ = -0.41, $t_{mob,purch3}$ = -3.52; $\gamma_{mob,mv3}$ = -0.34, $t_{mob,mv3}$ = -4.17; $\gamma_{mob,mv4}$ = 0.25, $t_{mob,mv4}$ = -2.60). This is in line with previous results, for example the study of DeHaan et al. (2015) shows that people shift from more mobile devices to less mobile devices as they get closer to the purchase. Moreover, the use of mobile devices is usually associated to search-related activities (Shankar and Balasubramanian, 2009; Okazaki and Hirose, 2009). Surprisingly, desktop users perform worse than tablet users both in terms of search ($\gamma_{desk,search3}$ = -0.08, $t_{desk,search3}$ = -2.76 for first triers, and $\gamma_{desk,search5}$ = -0.17, $t_{desk,search5}$ = -3.03 for loyal customers), purchase ($\gamma_{desk,purch3}$ = -0.32, $t_{desk,purch3}$ = -3.03; $\gamma_{desk,purch4}$ = 0.37, $t_{desk,purch4}$ = -3.49; $\gamma_{desk,purch5}$ = -0.74, $t_{desk,purch5}$ = -4.71), and monetary value ($\gamma_{desk,mv3}$ = 0.29, $t_{desk,mv3}$ = -3.84; $\gamma_{desk,mv4}$ = 0.38, $t_{desk,mv4}$ = -4.18; $\gamma_{desk,mv5}$ = -0.73, $t_{desk,mv5}$ = -4.78), highlighting the fact that nowadays the use of tablet is getting more and more important, even if it is still not so-widespread.

Regarding the search-related covariates, it is unsurprising to find out that they are all effective in increasing people's search activity in all the states. It is more interesting to look at their effect on the transactional outcomes. In fact, while ranking the products by novelty or by descending price does not affect the purchase probability or the expenditure level, sorting the products by ascending price, and clicking on suggested products, increase the purchase likelihood and the expenditure level of first triers ($\gamma_{rpl,purch3}$ = 0.09, $t_{rpl,purch3}$ = 4.33; $\gamma_{sp,purch3}$ = 0.04, $t_{sp,purch3}$ = 2.95; $\gamma_{rpl,mv3}$ = 0.06, $t_{rpl,mv3}$ = 5.71; $\gamma_{sp,mv3}$ = 0.02, $t_{sp,mv3}$ = 2.92), and newly acquired customers ($\gamma_{rpl,purch4}$ = 0.12, $t_{rpl,purch4}$ = 5.67; $\gamma_{sp,purch4}$ = 0.03, $t_{sp,purch4}$ = 2.40; $\gamma_{rpl,mv4}$ = 0.09, $t_{rpl,mv4}$ = 7.37; $\gamma_{sp,mv4}$ = 0.02, $t_{sp,mv4}$ = 2.47), but do not affect the behavior of loyal customers, reinforcing the evidence that loyal customers are not as price sensitive as new customers are. The positive use of filters to refine the search results on all the outcome variables encompasses all the states. These results are in line with the ones of Chen and Yao (2016), who found that search tools (e.g. rankings and filters) enhance people's likelihood to search, and increase their purchase utility.

The last set of variables is the promotional one. In periods with a promotional newsletter campaign, users who receive the e-mail search more if they are deliberative searchers ($\gamma_{nlp,search1}$ = 0.005, $t_{nlp,search1}$ = 2.77), or first triers ($\gamma_{nlp,search3}$ = 0.05, $t_{nlp,search3}$ = 3.41). This might be due to the fact that the reception of an e-mail containing a promotional communication can trigger the recipient's curiosity have a look to the discounted items, but they do not go further in their search activity. However, promotional newsletters do not affect transactional outcomes in any state. Even more surprising is the effect of the discount depth. As the depth of the discount increases, results suggests that the first triers' search and purchase activity decreases ($\gamma_{disc,search3}$ = -0.001, $t_{disc,search3}$ = -2.85; $\gamma_{disc,purch3}$ = -0.014, $t_{disc,purch3}$ = -5.35; $\gamma_{disc,mv3}$ = -0.005, $t_{disc,mv3}$ = -4.59). A possible explanation of this effect is that, as stated before, the presence of a discount limits the search activity to the items on sale. Moreover, the operationalization of the variable might provide some insights. The depth of the discount contained in the variable is the maximum level of discount available in the website, which is the one advertised in the banners (e.g. 60 for discounts up to 60%). This can generate some expectations in users who expect to buy the item they are interested with a deep discount. However, once they land to the website, they discover that the advertised 60% discount is actually a 30% for the product they were willing to buy, leading to a discrepancy between their expectation and the actual price, which leads to dissatisfaction and, consequently, a lower purchase intention (Oliver, 1980). This is especially true for users who are in the earlier stages of the purchase process, as the first triers, who have limited experience with the website and are more likely to have their own expectations.

5.5.5. *Paths to Acquisition*

The path leading prospects to be (or not to be) acquired can take different shapes. Table 5.5.9 shows the frequency distribution of all the observed paths of our 11,676 users.

TABLE 5.5.9: Frequencies of the possible paths

Path	Freq.	Percent.	Cum.
2	3058	26.19	26.19
21	2955	25.31	51.50
23	861	7.37	58.87
213	648	5.55	64.42
24	604	5.17	69.60
3	589	5.04	74.64
214	440	3.77	78.41
4	400	3.43	81.83
12	381	3.26	85.10
5	219	1.88	86.97
25	186	1.59	88.57
245	173	1.48	90.05
234	170	1.46	91.50
34	132	1.13	92.63
215	128	1.10	93.73
2134	114	0.98	94.71
45	111	0.95	95.66
123	105	0.90	96.56
2145	86	0.74	97.29
124	77	0.66	97.95
13	35	0.30	98.25
1234	26	0.22	98.48
125	23	0.20	98.67
1245	22	0.19	98.86
2345	22	0.19	99.05
1	21	0.18	99.23
21345	18	0.15	99.38
14	14	0.12	99.50
345	13	0.11	99.61
2135	9	0.08	99.69
35	9	0.08	99.77
15	7	0.06	99.83
235	6	0.05	99.88
134	5	0.04	99.92
145	4	0.03	99.96
12345	3	0.03	99.98
1235	1	0.01	99.99
135	1	0.01	100.00
Total	11676	100.00	

It is worth noting that 36.7% of the users “born and die” in the same state. They made on average 8.01 visits, with users who always stay in a non-purchase state having less visits than triers and acquired customers (directed searchers made on average 7.88 sessions, deliberative searchers 7.79, first triers 10.21, and acquired customers 10.68 sessions). Purely deliberative searchers are the most represented scenario (path 2: 26.19%), followed by the shifting from deliberative to directed searcher (path 2-1: 25.31%). Deliberative searchers are also highly likely to become first triers, both directly (path 2-3: 7.37%), or passing through the state one (path 2-1-3: 5.55%). This highlights the fact that deliberative searchers are an interesting state that the firm should carefully monitor, as it contains prospects who are more likely to become first triers and acquired, both directly (path 2-4: 5.17%), and passing through the directed searchers state (path 2-1-4: 3.77%).

FIGURE 5.5.1: Frequencies of paths in each transition

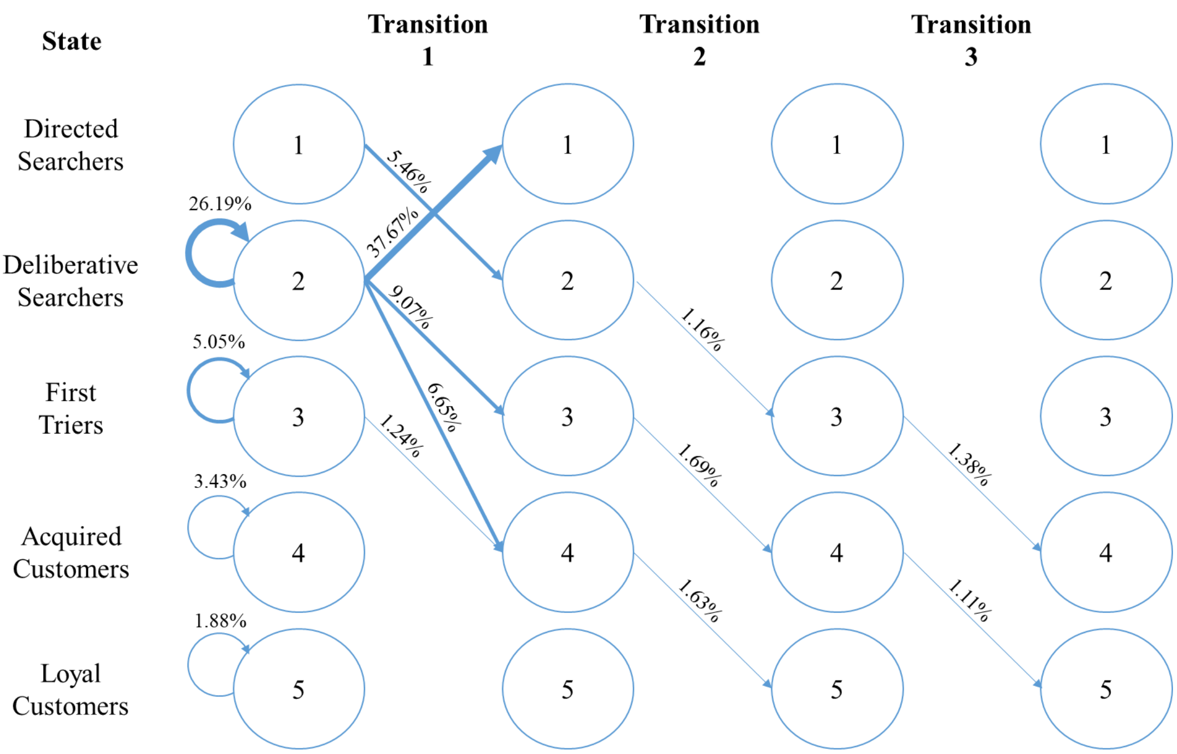


Figure 5.5.1 summarizes the most frequent paths that users follow in each transition. For sake of simplicity, it represents only the paths with a representativeness higher than 1%. The thickest arrows represent paths that have a percentage higher than 10%, and the thinnest a percentage below 5%. The graph shows that the most frequent paths in the first transition are all from state two, which leads to state one (37.67%), three (9.07%), and four (6.65%). The paths frequencies in the subsequent transitions are all around 1.60%. In the second transition it

is more likely to move from state three to four (1.69%) and from state four to five (1.63%), and in the third transition from state three to four (1.38%).

Since the number of possible paths is high, and many of them have a very low frequency, we join them into two types of groups. The first type distinguishes between users who are only in the purchase states, users who are only in purchase states, and users who are in both. Thus, the first group, *Only Search*, contains all the users who are only directed or deliberative searchers, corresponding to the first two states. It represents the 59.94% of the observed users. The second group, *Only Purchase*, contains users who are either first triers, acquired, or loyal customers, corresponding to states three, four, and five. It is the smallest group, representing the 12.62% of the observed users. The third and last group, Search and Purchase, represents all the remaining 32.44% of users who pass through both a search and a purchase state.

The second type distinguishes users on the basis of their arrival state. Thus, the three groups: *Up to Three, Up to Four,* and *Up To Five* contain all the users who arrive, respectively, at the first trial state (19.17% of the users), at the acquisition state (16.97%), and at the loyalty state (8.92%).

In order to analyze how long each user stays in each state, we compute the average number of sessions in which users were in each state for every group of paths described above [Table 5.5.11]. Results show that Only Searchers spend more sessions in state two (4.38) than in state one (2.93), and Only purchasers stay more in state four (2.34) and three (2.09) than in five (1.38). Users who are both searchers and purchasers spend more time in the acquisition state (4.04) and less in directed search (2.13) and loyalty (1.89) states. It is interesting to note that users in the *Up To Five* group stay more time in each state compared to people who are only triers or acquired customers. They spend also a significant amount of time in the acquisition state (5.73 sessions) before becoming loyal customers.

TABLE 5.5.11: Average number of sessions in each group of paths

State	Groups		
	Only Search	Only Purchase	Search and Purchase
1	2.934	0.000	2.131
2	4.357	0.000	3.866
3	0.000	2.094	4.044
4	0.000	2.236	3.983
5	0.000	1.382	1.891
State	**Groups**		
	Up to Three	**Up to Four**	**Up to Five**
1	1.635	1.641	3.165
2	2.751	2.984	3.466
3	6.623	1.546	2.750
4	0.000	7.754	5.729
5	0.000	0.000	10.495

At last, we want to provide a description of the first type of groups, in order to investigate which are the distinctive characteristics of users who only search, users who only purchase, and users who do both. Table 5.5.12 contains the description of registration, transactional, search, device, marketing, and demographic variables distinct by group of paths.

Users in the *Only Search* group have the highest proportion of social (27.3%) and Facebook (18.8%) registrations, and are also more likely to unsubscribe from the newsletter (11.0%). They are the ones who use more ranking tools (on average, 0.033 rankings by date, 0.065 by price high, and 0.400 by price low per session), and click on more products suggested by the firm (0.212). Only searchers have the lowest proportion of usage of SEM and newsletter touchpoints (respectively, 19.8% and 7.4%). In terms of demographics, they are youngest group (36.51 years).

The *Search and Purchase* group contains users who are registered for a longer time to the website (45.06 days) and to the newsletter (12.34 days). They do not distinguish themselves in terms of search activity, but they are the ones who use more the desktop to visit the website (70%). Among the three groups, searchers and purchasers have the highest proportion of usage of affiliate touchpoints (17.6%).

TABLE 5.5.12: Characteristics of the groups of paths

Variable type	Variable	Path group: Search and Purchase	Only Purchase	Only Search	Overall
Registration	Facebook connected	0,093	0,074	0,188	0,143
	Social Registration	0,142	0,108	0,273	0,210
	Newsletter Subscription	0,253	0,192	0,234	0,235
	Newsletter unsubscription	0,083	0,064	0,110	0,096
	Registration Days	45,061	25,890	27,461	32,973
	Nl subscription days	12,348	4,784	9,296	9,717
	Unsubscription days	2,763	1,559	2,256	2,333
Transactional	Purchases	1,521	1,536	0,000	0,687
	Monetary value	279,461	326,402	0,000	131,842
Search	Clicks	44,856	55,646	53,865	51,167
	Pages	18,917	23,542	22,430	21,431
	Products	6,053	5,805	6,806	6,436
	Session length	953,168	1094,685	1079,993	1040,701
	Filters	2,516	2,319	3,422	2,989
	Ranking by date	0,018	0,015	0,033	0,026
	Ranking by price high	0,043	0,034	0,065	0,054
	Ranking by price low	0,286	0,207	0,400	0,339
	Suggested products	0,169	0,129	0,212	0,187
	Add Wish List	0,252	0,153	0,634	0,449
	Drop Wish List	0,036	0,040	0,036	0,037
	Intersession days	5,039	4,725	5,782	5,408
	Number of sessions	15,915	5,711	7,310	9,900
Device	Mobile	0,222	0,281	0,249	0,245
	Tablet	0,074	0,086	0,079	0,078
	Desktop	0,704	0,634	0,671	0,677
Marketing	Affiliation	0,176	0,128	0,167	0,165
	SEO	0,225	0,242	0,235	0,233
	Other channel	0,001	0,001	0,001	0,001
	Social	0,034	0,013	0,034	0,031
	Direct	0,116	0,246	0,164	0,159
	Referral	0,061	0,038	0,052	0,053
	Co-marketing	0,004	0,004	0,009	0,007
	Retargeting	0,087	0,034	0,066	0,069
	SEM	0,206	0,203	0,198	0,201
	Newsletter	0,091	0,091	0,074	0,082
Demographics	Age	38,434	38,709	36,515	37,268
	Gender (Female)	0,639	0,620	0,649	0,642
	Africa	0,000	0,000	0,002	0,001
	America	0,006	0,005	0,009	0,007
	Asia	0,148	0,128	0,175	0,160
	Europe	0,338	0,327	0,256	0,291
	Italy	0,217	0,301	0,173	0,203
	Oceania	0,014	0,018	0,023	0,020
	Russia	0,114	0,043	0,185	0,144
	USA	0,163	0,179	0,155	0,161

The most interesting group is the *Only Purchase* one. They prefer not to use social channels to register (10.8% have a social registration, and 7.4% are registered through

Facebook), and not to subscribe to the newsletter (19%). The analysis of the search activity provides interesting insights: they are the group with more clicks (55.65), pages (23.54), and seconds spent on the website (1094.69). However, they have the lowest number of products seen (5.81), filters (2.32), and ranking tools (0.015 rankings by date, 0.034 by price high, and 0.21 by price low) used. Thus, they are goal-oriented customers, who come to the website with already a product in mind and go to buy it. The usage of the Wish List strengthens this idea, as they have the lowest number of products added (0.15) and the highest number of items dropped (0.04). Moreover, they visit the website less times (5.71 visits), but more frequently (on average, 4.73 days between two subsequent visits). Surprisingly, compared to the other groups, they do a massive use of mobile devices (28.1% of sessions from smartphone, and 8.6% from tablet). Among the three groups, only searchers have the highest proportion of customer-initiated touchpoints (SEO 24.2%, direct 24.6%), and the lowest of firm-initiated (social 1.3%, referral 3.8%, retargeting 3.4%). Regarding the transactional activity, their proportion of purchases do not differ significantly from the one of users in the *Search and Purchase* group (t test p-value = 0.653), but they have an average expenditure level that is significantly higher (*Only Purchase* = 326.40€, *Search and Purchase* = 279.46€, t test p-value = 0.003).

6. CONCLUSIONS

The present research aims at analyzing the evolution of the customer acquisition process over time, investigating the role played by the interactions between the potential customers and the different marketing activities in each stage of the acquisition funnel in driving prospects toward acquisition, and the influence of the pre-acquisition behavior on the acquisition process. It also proposes a new definition of customer acquisition in non-contractual settings, where customers do not subscribe any contract with the firm at the time of their first transaction.

In order to achieve these goals, this dissertation presented an empirical study, in collaboration with a big Italian e-retailer, that sells clothing and design objects worldwide. Building on a unique database storing information on individual-level transaction data, clickstream data and marketing activities (newsletter), we specified a Hidden Markov Model (HMM) with the intent to identify the stages that prospects go through before being acquired and, to analyze the impact that marketing activities and pre-acquisition behavior have on the probability to move forward in the process, and on the probability to be acquired and engage.

The particular setting of the study allows us to take a customer-centric perspective since we are able to exactly identify not only the first transaction a customer carries with the firm, but also his or her preceding behavior. Furthermore, the availability of data on customers' transactions allows us to represent acquisition as a latent state and to eventually provide a more thorough definition of acquisition in a non-contractual setting.

The results of our analysis provide an answer to the three main research questions of this dissertation:

(iv) How does the customer acquisition process evolve over time?

We have identified five states characterizing the acquisition process, defined by the level of search intensity, the cumulative number of purchases, and the cumulative monetary value, namely: *directed searchers*, *deliberative searchers*, *first triers*, *acquired customers*, and *loyal customers*.

The first two states are defined exclusively by search activity. *Directed searchers* exhibit a behavior characterized by a small number of products browsed, but a high number of pages per product visited, meaning that they search for less products, while trying to acquire more information, as they presumably already know what they are looking for.

Deliberative searchers are the ones with the most intense search activity, in terms of clicks, products, pages, length of the session, ranking tools, and filters, albeit they do not buy. This may be signaling that they do not have a goal in mind, and/or they still do not know what to search for. Our findings indicate that users have a very high probability to start their relationship with the firm being in this state. This is reasonable, as people do not have any previous experience with the website, and are very likely to be in a deliberative mindset (Gollwitzer and Bayer,1999), eager to gather as much information as possible.

The last three states entail an element of purchase, and are mainly defined by the cumulative expenditure level of customers. By construction, once people reach one of those states, they can only move on and are not allowed to go back. *First triers* are defined mainly by the first purchase and by a low level of spending. They are more goal oriented and price sensitive, as their search activity is below the average, but they dig more into the pages they visit, and make more intensive use of the ranking tool by ascendant price. In summary, while they are moving their relationship with the firm to the next step (i.e. they buy something), they seem to be still in an exploration mode as the amount spent is limited and they are mainly driven by a price search. *Acquired Customers* have made at least one or two purchases within the firm. The portion of first time buyers is still high, but, unlike first triers, they spend more. *Loyal customers* are defined by a high level of expenditure, have made more than two purchases. In terms of search, they are characterized by a relatively intense search activity, and they are less sensible to price, as they are among the customers who rank their search by ascendant price the most, and by descendant price the least. Of course, users are very unlikely to start their acquisition process being in this state.

All the states are relatively sticky, in fact, we found that 36.7% of users stay in their initial state for the entire process at least in the time window we observe.

The path leading prospects to be (or not to be) acquired can take different shapes. The most frequent movement occurs between directed and deliberative searchers. Accordingly, we found that more than half of our users never get into a purchase state. Results also highlight the fact that deliberative searchers are an interesting state that the firm should carefully monitor, as it contains prospects who are more likely to become first triers and acquired.

Another interesting group is the one composed by users who only move across purchase states. They are goal-oriented customers, who come to the website with already a product in mind and go to buy it. Their frequency of purchases do not differ from the one of who pass through purely search states, but they spend more. Almost 34% of loyal customers arrive at the loyalty stage by following an acquisition process which does not have any search state.

(v) What is the role played by the different marketing activities, and the prospects' pre-acquisition behavior, in each stage of the acquisition process?

Our findings suggest that both marketing activities and pre-acquisition behavior have an effect in driving prospects towards acquisition, and this effect changes according to the state in which the individual is.

Marketing. Results show that there is an overall negative effect of the firm-initiated touchpoints (retargeting, newsletters, social, and, to some extent, referrals) on the level of search activity, the purchase probability, and the monetary value of the transaction in all the three purchase states, meaning that firm-initiated activities work worse than SEO in triggering the search and purchase activities of prospects and customers. It is interesting to note that the worst performances are observed in the earlier stages of the purchase activity (first triers and newly acquired customers). This can be a matter of the degree of personalization of the communication, which is usually based on the customer's purchase history (e.g. Lambrecht et al., 2013; Barjas et al., 2016; Kumar, Morris and Pancras, 2008; Song et al., 2016), and has been found to be effective only if the recipient is in the later stages of the purchase funnel (Lambrecht et al., 2013; Abhishek et al., 2016). By subscribing to the newsletter, users demonstrate some interest to the firm, but for some reasons they are not willing to interact more with it during the first visit, maybe for time constraints, or maybe because they simply want to postpone their information gathering in the future. In fact, users who subscribe to the newsletter are more likely to start their acquisition process as deliberative searchers, and less likely to shift to a purchase state, especially if they have been subscribed for a long time. These results put some doubts on the efficacy of the firm's newsletter campaigns on prospects. This negative reaction to e-mails can be due to the reactance effect (Brehm, 1966). However, when deliberative searchers or first triers receive a promotional e-mail, they search more. This might be due to the fact that the reception of an e-mail containing a promotional communication can trigger the recipient's curiosity have a look to the discounted items, but they are still unwilling to go any further in their search activity.

On the other hand, customer-initiated touchpoints show a positive effect. Directly typing the website address in the address bar of the browser and using SEM increase both first triers and newly acquired customers' purchase activity. The use of paid search is associated to customers who are closer to purchase (Jerath and Park, 2014), in fact the impact of SEM on search activity is lower than the SEO one in the first three states, which are the ones characterized by none or low levels of purchase.

The discount depth decreases all the search and purchase activities of first triers. This should be interpreted as an alarm bell for managers who employ the "Up to …%" statement to signal their discount activity. In fact, this can generate some expectations in users who probably anticipate they will buy the item they are interested at a deep discount, leading to a discrepancy between their expectation and the actual price, which brings dissatisfaction and, consequently, a lower purchase intention (Oliver, 1980). This is especially true first triers, who have limited experience with the website and are more likely to have their own expectations.

Pre-purchase behavior. Results also show that the act of registering to the website has an impact on the acquisition process, as it increases people's likelihood to purchase at the very beginning, and to move on in the process, especially shifting from non-purchase to purchase states. This might be due to the fact that people feel the need to register to the website before making a purchase, even if it is not formally required, especially if they are at the beginning of their relationship with the firm, and do not have much experience with the website. The act of registration is also a signal of a high level of interest and trust (Schumann, Wangenheim, and Groene, 2014; Martin, and Murphy, 2017) toward the firm, which are two factors that enhance the probability to purchase and to become loyal customers (Spekman, 1988; Nooteboom, Berger, and Noorderhaven, 1997; Garbarino and Johnson 1999). Users registering in the website using a social account (e.g. Google+) have a higher probability to be deliberative searchers and to remain in that state, but only if they do not use a Facebook account, which increases their likelihood to move to a purchase state. In fact, deliberative searchers, who do not have a deep knowledge of the website, can use their social account as a way to register because it is easier and more convenient, as it does not require, for example, to create and remember a new password (Bauer et al., 2013). However, people use to share more personal information on Facebook than on Google+, and the cost of privacy due to the Facebook registration are higher (Bauer et al., 2013), requiring a higher level of trust towards the firm, leading to an increased likelihood to establish a deeper relationship with it.

The use of the Wish List for directed searchers makes them moving on in the process, as it helps users search in a more organized and efficient way (Close and Kukar-Kinney, 2010), a particularly useful feature for goal-oriented customers who do not have a lot of time to spend on the website. The effect is reversed for deliberative searchers, decreasing their likelihood to become first triers. This is because they dedicate more time in searching activity, and adding items in Wish List in order to postpone their evaluation decreases their likelihood to buy them in future (Popovic and Hamilton, 2014). When users in the later stages of the process drop items from the Wish List, it makes them more likely to move to the subsequent state. In fact, dropping items from the Wish List is a signal that they made a choice and move the checkout, discarding the items they are not willing to buy.

All the search-related activities (use of ranking and filtering tools, clicks on suggested products) are effective in increasing people's search activity in all the states. It terms of effect on the purchase activity, only sorting the products by ascending price, clicking on suggested products, and use filters have an effect on the purchase likelihood and the expenditure level. However, we did not find this effect for loyal customers, as they are not price sensitive.

Our results also highlight the fact that people who do not engage much in searching activity within the website need more days to search again, while deliberative searchers, who deeply explore the website and its products during their visits, can be closer to a conversion and need to make a subsequent session earlier in order to purchase. Customers who have just purchased, on the other hand, need more time to purchase again and move to the loyal state.

We found also the device choice to play a role in the process. In fact, the mobile usage has been found to be positively related to search activities, but negatively to purchase activities, especially for customers who are purchasing for the first time. This is in line with previous results, showing that the use of mobile devices is usually associated to search-related activities (Shankar and Balasubramanian, 2009; Okazaki and Hirose, 2009), and people shift from more mobile devices to less mobile devices as they get closer to the purchase (DeHaan et al., 2015). Desktop users perform worse than tablet users both in terms of search, purchase, and monetary value, highlighting the fact that, even if it is not yet very widespread, nowadays the use of tablet is gaining importance in the digital retailing environment.

(vi) How can we define customer acquisition in a non-contractual setting?

In chapter 3, we stated that in non-contractual setting, acquisition should be seen as a hidden state, that new customers reach only after their purchase activity within the company becomes regular, despite the fact that most of firms, business press, and previous literature (e.g. Gupta and Zeithaml, 2006; Schweidel, Fader, and Bradlow, 2008; Wangenheim and Bayon, 2007) consider first time buyers as acquired customers.

Our findings corroborate this idea. In fact, our model identified three states in which people who purchase can be: trial, acquisition, and loyalty. There is not a clear-cut in terms of number of purchases among these three states. In fact, while the first purchase could be just a tentative approach to the relationship with the firm, it can also represent the start of a blossoming relationship, whereby customers are not shy to spend more, and to purchase other times. Put in this way, customers who make the first purchase with a limited expenditure are categorized as triers, while those who make the first purchase with a higher level of monetary value of the transaction can already be considered as acquired. In the same way, a customer who make a second purchase, but continues limiting her expenditure should not necessarily be considered as an acquired customer yet.

6.2 Contributions

This dissertation was triggered by the desire to contribute to the customer acquisition literature in three ways: first, from a theoretical standpoint, we aim at increasing the little literature on customer acquisition and testing its traditional definition. From a managerial perspective, we believe this research may help managers in designing their acquisition strategies in a more personalized way, accounting for the prospect's state in the process. Finally, from an empirical perspective, it represents one of the first attempts to analyze the path to acquisition accounting for both pre and post-acquisition behaviors.

Theoretical. From a theoretical standpoint, this research increases the little literature on customer acquisition by focusing on the customer acquisition process *per se*. Since the vast majority of the studies in this field has not devoted much attention to acquisition alone, but they have mainly analyzed it jointly with other factors, such as the customer retention (Schweidel et al., 2008; Villanueva et al., 2008; Schmitt et al., 2011), and/or the optimal budget allocation (Blattberg and Deighton, 1996; Thomas, 2001; Berger and Nasr-Bechwati, 2001; Reinartz et al., 2005), this research extends the literature on acquisition by investigating the process leading prospects to be acquired by the company and the antecedent factors that affect this process.

It is also the first study to formally consider the pre-acquisition activity of prospects, such as their registration and search activities, and their former touchpoints with the firm. Moreover, analyzing acquisition under a customer-centric perspective, by taking into consideration all the prospect's interactions with the company, allows us to formally represent the customer acquisition as a multistage process. As far as we know, this is the first study investigating the customer acquisition process by identifying all the possible stages preceding acquisition, and also the path to acquisition. In this light, this work's contribution to the field consists the definition of our 5-stage acquisition process: directed search – deliberative search – first trial – acquisition – loyalty.

Finally, this research is in a non-contractual setting, where customers are not formally bounded to the firm for an observable period of time, and this setting has been highly understudied in acquisition literature, in part due to the fact that the definition of acquisition may be not as obvious as it is in a contractual setting. This research provides a new way to define acquired customers when they are not bounded by a contract, looking at it as a hidden state in the process rather than an observed behavior.

Managerial. From a managerial point of view, this work aims at helping marketing managers in designing more focused, effective, and customer-centric acquisition programs by considering the state in which prospects are in the acquisition process, suggesting a way for developing more customized acquisition programs which better meet the prospects' needs according to their state. This is relevant for managers, since the employment of a customer-oriented strategy has been found to provide several benefits to companies, because it allows them to shift from a mass-marketing to an individual marketing strategy (Sheth et al., 2000), it increases customer satisfaction (Blattberg et al., 2008), provides a certain degree of competitive advantage (Day, 2000), and enhances the effectiveness of firms' cross-selling strategies (Blattberg et al., 2008).

Moreover, our results are useful in order to better target prospects who are more likely to be acquired and to become loyal. The role of the search activity has been found to be particularly useful for firms in order to detect prospects in their earlier stages who are more likely to convert (for example, the deliberative searchers), since it employs measures which are directly observable by the company (e.g. the usage of ranking and filtering tools, of the Wish List, or the device), and hence to focus their acquisition efforts on them, without spending time and money for individuals who will likely never convert.

Empirical. Finally, we believe this work to have also an empirical contribution, since it is one of the first attempts to analyze customer acquisition process accounting for both prospects' pre and post-acquisition behavior, which is a novelty in acquisition studies. Moreover, the absence of physical stores allows the identification of the actual first transaction of each customer, and of all the visits that users made in the store (as far as technology allows), enabling us to gain the so much desired "full view of the customer" (Fader, 2012).

6.3 Limitations and Future Research

Several challenges have characterized the development of the current work. We did our best to cope with them, but sometimes they required us to take suboptimal decisions. The first regards the secondary data that the firm provided us. One of the major challenges that e-commerce retailers are facing nowadays is how to link all the sessions to the right user. This challenge is getting continuously harder to face, as the usage of multiple devices from the same user is growing fast. Digital firms often rely on cookies in order to recognize their users, but this practice fails to work when the user decides to clean her cookies history (Kannan et al., 2016). Moreover, cookies are ineffective in the multi-device world, as they are device-specific. Our firm faces this issue by creating its own cookie, which accounts not only for the classical cookies, but also for the user's e-mail address. This solution partially solves the problem arisen by the cookie deletion or the multi-device usage, but only as long as the user provides an e-mail account. However, it does not work perfectly, as it works only for users who provide an e-mail (in other word, register to the website), and log in every time they use a different device, or delete their cookies. For this reason, we decided to restrict our pool of users to registered ones only. This may lead to a selection bias, as they can be users who are already more likely to be acquired. An avenue for further investigation may be to overcome this limitation by considering also users who did not register to the website.

The firm did not provide us many exogenous information about marketing activities (e.g. number of impressions, marketing expenditures), or about the content of the e-mails, we only have the users' responses to the different activities (e.g. last-touch marketing, number of emails received), and the presence or absence of advertising and discount activity on the website, or of a promotional or informational e-mail campaign. It might be worthy for future research to incorporate exogenous information about marketing in the analysis.

The time window of the analysis spans from October, 2016 to May, 2017, which is less than one year. Due to computational issues, we consider 70 as the maximum number of sessions users can make, but some people may need a longer time to be acquired, or to make multiple transactions. Considering a longer time period might lead to more insightful results.

From a methodological standpoint, the LMest R package we have employed has some limitations. First of all, it does not account for unobserved individual heterogeneity. As discussed in section 4.3, HMM allows researchers to deal with unobserved heterogeneity through the use of discrete latent classes, which are dynamic over time (Lazarsfeld, 1950; Wedel et al., 1999; Henry, 2004; Bartolucci and Farcomeni, 2009; Bartolucci et al., 2012), but it accounts for heterogeneity between groups of similar users, not at the individual level. Not accounting for individual unobserved heterogeneity is dangerous, as it is likely to lead to biased results, capturing incorrect dynamics, and providing wrong insights (Heckman, 1981; Erdem and Sun, 2001; Netzer et al., 2016; Kappe et al., 2018). Future research might incorporate Bayesian estimation methods in order to deal with individual unobserved heterogeneity across users.

The package does not allow us to estimate the model jointly, by incorporating time varying covariates in both the transition and the state dependence probabilities. When including time varying covariates in the estimation of the state dependent probabilities, the package makes the assumption on the transition probabilities to be homogeneous over time (Bartolucci, Pandolfi, and Pennoni, 2017). This is made in order to reduce the computational complexity of the model. For this reason, we were forced to estimate them separately. Moreover, a second assumption on the state dependence probabilities with covariates is that the impact of the covariates is homogeneous across states (Bartolucci et al., 2017). This assumption is incompatible with our research objective to explore the role of the marketing activities in each stage of the process, thus we decided to follow an alternative path. We first estimated the transition probabilities, which provided us the probability of being in each state in every session *t*. Then, we allocated every session *t* in the state *s* which was more likely to be. To estimate the state dependent probabilities, we took all the observations allocated to each state, and estimated the impact of the covariates by running a regression for each outcome in each state. It is evident that this is a suboptimal solution to the problem. A further development of the package is required to deal with this issue in a better way.

The last limitation of the package is that it was initially built in order to deal with categorical outcomes. A recent update of the package allows the estimation of the model with continuous outcomes, but it is still unable to deal with mixed types of dependent variables. Since we had both continuous (search and monetary value) and categorical (purchase) outcomes, we opted for the more parsimonious way, by categorizing all the continuous dependent variables. This, of course, has the drawback to ignore part of the outcomes' variance. A possible development of this research should be to improve the LMest package code in order to cope with multivariate mixed outcomes.

Finally, our model does not account for customers who stop buying from the firm after the first or second transaction, but still go on with the search activity, considering them as "eternal triers". A last avenue for future research might be to relax this restriction and investigate for the existence of a "light quit" state, in which users still interact with the firm after they made at least one transaction, but do not purchase anymore.

References

Abhishek, V., Despotakis, S., & Ravi, R. (2017). "Multi-Channel Attribution: The Blind Spot of Online Advertising". Available at SSRN: https://papers.ssrn.com/soL3/papers.cfm?abstract_id=2959778

Abhishek, V., Fader, P., & Hosanagar, K. (2016). "Media exposure through the funnel: A model of multi-stage attribution". *Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2158421*

Akaike, H. (1973). "Information theory and an extension of the maximum likelihood principle". In: B. N. Petrov and F. Cs´aki, eds., *Second International Symposium of Information Theory*, 267–281.

Anderl, E., Becker, I., Von Wangenheim, F., & Schumann, J. H. (2016). "Mapping the customer journey: Lessons learned from graph-based online attribution modeling". *International Journal of Research in Marketing*, 33(3), 457-474.

Anderson, E. T., Hansen, K., & Simester, D. (2009). "The option value of returns: Theory and empirical evidence". *Marketing Science*, 28(3), 405-423.

Anderson, E. T., & Simester, D. I. (2004). "Long-run effects of promotion depth on new versus established customers: Three field studies". *Marketing Science, 23*(1), 4-20.

Anderson, E. T., & Simester, D. I. (2010). "Price Stickiness and Customer Antagonism". *The Quarterly Journal of Economics*, *125*(2), 729-765.

Andrews, R. L., & Srinivasan, T. C. (1995). "Studying consideration effects in empirical choice models using scanner panel data". *Journal of Marketing Research*, 30-41.

Ansari, A., Mela, C. F., & Neslin, S. A. (2008). "Customer Channel Migration". *Journal of Marketing Research*, *45*(1), 60-76.

Ansari, A., Montoya, R., & Netzer, O. (2012). "Dynamic learning in behavioral games: A hidden Markov mixture of experts approach". *Quantitative Marketing and Economics,* 10(4), 475-503.

Arnold, T. J., Fang, E. E., & Palmatier, R. W. (2011). "The effects of customer acquisition and retention orientations on a firm's radical and incremental innovation performance". *Journal of the Academy of Marketing Science*, 39(2), 234-251.

Ascarza, E., & Hardie, B. G. (2013). “A joint model of usage and churn in contractual settings”. *Marketing Science*, 32(4), 570-590.

Ascarza, E., Netzer, O., & Hardie, B. G. (2018). “Some customers would rather leave without saying goodbye”. *Marketing Science*, 37(1), 54-77.

Ashley, W. C., & Morrison, J. L. (1997). “Anticipatory management: Tools for better decision making”. *The Futurist*, *31*(5), 47.

Auranen, K., Arjas, E., Leino, T., & Takala, A. K. (2000). “Transmission of pneumococcal carriage in families: a latent Markov process model for binary longitudinal data”. *Journal of the American Statistical Association*, 95(452), 1044-1053.

Barajas, J., Akella, R., Holtan, M., & Flores, A. (2016). “Experimental designs and estimation for online display advertising attribution in marketplaces”. *Marketing Science*, 35(3), 465-483.

Barnes, S. J., & Scornavacca, E. (2008). “Uncovering patterns in mobile advertising opt-in behaviour: a decision hierarchy approach”. *International Journal of Mobile Communications*, *6*(4), 405-416.

Barry, T. E. (1987). “The development of the hierarchy of effects: An historical perspective”. *Current issues and Research in Advertising*, 10(1-2), 251-295.

Bart, Y., Stephen, A. T., & Sarvary, M. (2014). “Which Products Are Best Suited to Mobile Advertising? A Field Study of Mobile Display Advertising Effects on Consumer Attitudes and Intentions”. *Journal of Marketing Research*, 51(3), 270-285.

Bartolucci, F., & Farcomeni, A. (2009). “A multivariate extension of the dynamic logit model for longitudinal data based on a latent Markov heterogeneity structure”. *Journal of the American Statistical Association*, 104, 816–831.

Bartolucci, F., Farcomeni, A., & Pennoni, F. (2012). “*Latent Markov models for longitudinal data*”. Chapman and Hall/CRC.

Bartolucci, F., Farcomeni, A., & Pennoni, F. (2014). “Latent Markov models: a review of a general framework for the analysis of longitudinal data with covariates”. *Test*, 23(3), 433-465.

Bartolucci, F., Lupparelli, M., & Montanari, G. E. (2009). “Latent Markov model for longitudinal binary data: an application to the performance evaluation of nursing homes”. *The Annals of Applied Statistics*, 3(2), 611-636.

Bartolucci, F., Pandolfi, S., & Pennoni, F. (2017). “LMest: an R package for latent Markov models for longitudinal categorical data”. *Journal of Statistical Software*, 81(4), 1-38.

Bartolucci, F., Pennoni, F., & Francis, B. (2007). “A latent Markov model for detecting patterns of criminal activity”. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 170(1), 115-132.

Bauer, L., Bravo-Lillo, C., Fragkaki, E., & Melicher, W. (2013, November). “A comparison of users' perceptions of and willingness to use Google, Facebook, and Google+ single-sign-on functionality”. In *Proceedings of the 2013 ACM workshop on Digital identity management* (pp. 25-36). ACM.

Baum, L. (1972). “An inequality and associated maximization technique in statistical estimation of probabilistic functions of a Markov process”. *Inequalities*, 3, 1-8.

Baum, L. E., & Petrie, T. (1966). “Statistical inference for probabilistic functions of finite state Markov chains”. *The annals of mathematical statistics*, 37(6), 1554-1563.

Baum, L. E., Petrie, T., Soules, G., & Weiss, N. (1970). “A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains”. *The annals of mathematical statistics*, 41(1), 164-171.

Baxendale, S., Macdonald, E. K., & Wilson, H. N. (2015). “The impact of different touchpoints on brand consideration”. *Journal of Retailing*, 91(2), 235-253.

Beal, G. M., Rogers, E. M., & Bohlen, J. M. (1957). “Validity of the concept of stages in the adoption process”. *Rural Sociology*, *22*(2), 166-168.

Bellman, S., Potter, R. F., Treleaven-Hassard, S., Robinson, J. A., & Varan, D. (2011). “The effectiveness of branded mobile phone apps”. *Journal of Interactive Marketing*, 25(4), 191-200.

Berger, P. D., & Nasr-Bechwati, N. (2001). “The allocation of promotion budget to maximize customer equity”. *Omega*, *29*(1), 49-61.

Berman, R. (2017). “Beyond the last touch: Attribution in online advertising”. Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2384211

Bernazzani, S. (2018) “13 Examples of Successful Co-Branding Partnerships (And Why They're So Great)” [Available at: https://blog.hubspot.com/marketing/best-cobranding-partnerships]

Bettman, J. R., Luce, M. F., & Payne, J. W. (1998). “Constructive consumer choice processes”. *Journal of Consumer Research*, 25(3), 187-217.

Bettman, J. R., Johnson, E. J., & Payne, J. W. (1988). “A componential analysis of cognitive effort in choice”. *Organizational Behavior and Human Decision Processes*, 45(2), 111-139.

Bijleveld, C. C., & Mooijaart, A. (2003). “Latent Markov modelling of recidivism data”. *Statistica Neerlandica*, 57(3), 305-320.

Bilgicer, T., Jedidi, K., Lehmann, D. R., & Neslin, S. A. (2015). “Social Contagion and Customer Adoption of New Sales Channels”. *Journal of Retailing*,*91*(2), 254-271.

Biyalogorsky, E., Gerstner, E., & Libai, B. (2001). “Customer referral management: Optimal reward programs”. *Marketing Science*, *20*(1), 82-95.

Blattberg, R. C., & Deighton, J. (1996). “Manage marketing by the customer equity test”. *Harvard Business Review*, *74*(4), 136.

Blattberg, R.C., Getz, G., and Thomas, J.S. (2001). “*Customer Equity: Building and Managing Relationships as Valuable Assets”*. Boston, MA: Harvard Business School

Blattberg, R. C., Kim, B. D., & Neslin, S. A. (2008). "*Database Marketing: Analyzing and Managing Customers"* (Vol. 18). Springer Science & Business Media.

Bowman, D., & Narayandas, D. (2001). "Managing customer-initiated contacts with manufacturers: The impact on share of category requirements and word-of-mouth behavior". *Journal of Marketing Research*, 38(3), 281-297.

Brehm, J. W. (1966). "*A Theory of Psychological Reactance*". New York: Academic Press, Inc.

Bruner, G. C., & Kumar, A. (2000). "Web commercials and advertising hierarchy-of-effects". *Journal of Advertising Research*, 40(1-2), 35-42.

Burt, R. S. (1987). "Social contagion and innovation: Cohesion versus structural equivalence". *American Journal of Sociology*, 1287-1335.

Cao, Y., & Gruca, T. S. (2005). "Reducing adverse selection through customer relationship management". *Journal of Marketing*, 69(4), 219-229.

Chan, T. Y., Wu, C., & Xie, Y. (2011). "Measuring the lifetime value of customers acquired from google search advertising". *Marketing Science*, 30(5), 837-850.

Chen, Y., & Yao, S. (2016). "Sequential search with refinement: Model and application with click-stream data". *Management Science*, 63(12), 4345-4365.

Chevalier, J. A., & Mayzlin, D. (2006). "The effect of word of mouth on sales: Online book reviews". *Journal of Marketing Research*, *43*(3), 345-354.

Chiang, W. Y. K., Chhajed, D., & Hess, J. D. (2003). "Direct marketing, indirect profits: A strategic analysis of dual-channel supply-chain design". *Management Science*, 49(1), 1-20.

Chintagunta, P. K., Jain, D. C., & Vilcassim, N. J. (1991). "Investigating heterogeneity in brand preferences in logit models for panel data". *Journal of Marketing Research*, 417-428.

Cho, C.H. & Khang, H.K. (2006) "The State of Internet-Related Research in Communications, Marketing, and Advertising: 1994-2003", *Journal of Advertising*, 35 (3), 143 - 163

Choi, J., Hui, S. K., & Bell, D. R. (2010). "Spatiotemporal analysis of imitation behavior across new buyers at an online grocery retailer". *Journal of Marketing Research*, *47*(1), 75-89.

Close, A. G., & Kukar-Kinney, M. (2010). "Beyond buying: Motivations behind consumers' online shopping cart use". *Journal of Business Research*, 63(9-10), 986-992.

Coleman, J. S., Katz, E., & Menzel, H. (1966). "*Medical innovation: A diffusion study"* (Second Edition). Indianapolis: Bobbs-Merrill.

Coler, W. (2018). "Empowering Marketers: When Customer Acquisition Could Go Wrong Without Data Science". *Forbes*. [Available at: https://www.forbes.com/sites/forbestechcouncil/2018/07/23/empowering-marketers-when-customer-acquisition-could-go-wrong-without-data-science/#72f7e82f626e]

Colicev, A., Malshe, A., Pauwels, K., & O'Connor, P. (2018). "Improving Consumer Mindset Metrics and Shareholder Value Through Social Media: The Different Roles of Owned and Earned Media". *Journal of Marketing, 82*(1), 37-56.

Colley, R. H. (1961). "*Defining Advertising Goals for Measured Advertising Results*". Association of National Advertisers. Inc., New York, NY.

Conley, T. G., & Topa, G. (2007). "Estimating dynamic local interactions models". *Journal of Econometrics, 140*(1), 282-303.

Cook, R. J., Ng, E. T., & Meade, M. O. (2000). "Estimation of operating characteristics for dependent diagnostic tests based on latent Markov models". *Biometrics*, 56(4), 1109-1117.

Court, D., Elzinga, D., Mulder, S., & Vetvik, O. J. (2009). "The Consumer Decision Journey" [Available at: https://www.mckinsey.com/business-functions/marketing-and-sales/our-insights/the-consumer-decision-journey]

Dalessandro, B., Perlich, C., Stitelman, O., & Provost, F. (2012). "Causally motivated attribution for online advertising". *In Proceedings of the Sixth International Workshop on Data Mining for Online Advertising and Internet Economy* (p. 7). ACM.

Day, G. S. (2000). "Managing market relationships". *Journal of the Academy of Marketing Science, 28*(1), 24-30.

De Haan, E., Kannan, P. K., Verhoef, P. C., & Wiesel, T. (2015). "The Role of Mobile Devices in the Online Customer Journey". *Marketing Science Institute Working Paper Series*, Report No. 15-124.

De Haan, E., Wiesel, T., & Pauwels, K. (2016). "The effectiveness of different forms of online advertising for purchase conversion in a multiple-channel attribution framework". *International Journal of Research in Marketing*, 33(3), 491-507.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). "Maximum likelihood from incomplete data via the EM algorithm". *Journal of the royal statistical society. Series B (methodological),* 1-38.

Deighton, J. (1997). "Commentary on Exploring the implications of the Internet for consumer marketing". *Journal of the Academy of Marketing Science, 25*(4), 347-351.

Deighton, J., Peppers, D., & Rogers, M. (1994). "Consumer transaction databases: present status and prospects". Blattberg, Glazer, and Little (eds.), *The Marketing Information Revolution*. Boston, MA: Harvard Business School Press.

Demoulin, N. T., & Zidda, P. (2009). "Drivers of customers' adoption and adoption timing of a new loyalty card in the grocery retail market". *Journal of Retailing, 85*(3), 391-405.

De Vries, L., Gensler, S., & Leeflang, P. S. (2017). "Effects of traditional advertising and social messages on brand-building metrics and customer acquisition". *Journal of Marketing*, 81(5), 1-15.

Dias, J. G. (2006). “Model selection for the binary latent class model: a Monte Carlo simulation”. In: *A. Ferligoj, V. Batagelj, H.-H. Bock and A. Ziberna, eds., Data Science and Classification*, 91–100. Springer, Berlin.

Dias, M. B., Locher, D., Li, M., El-Deredy, W., & Lisboa, P. J. (2008). “The value of personalised recommender systems to e-business: a case study”. *In Proceedings of the 2008 ACM conference on Recommender Systems* (pp. 291-294). ACM.

Dierks, A. (2017). “*Re-Modeling the Brand Purchase Funnel”*. Springer.

Dinner, I. M., Van Heerde, H. J., & Neslin, S. (2015). “Creating Customer Engagement Via Mobile Apps: How App Usage Drives Purchase Behavior”. Available at SSRN: *https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2669817*.

Donahoe, S. (2016). “Lead Retention and Capture: How to Prevent a Leak in Your Sales Funnel”, *Internet Marketing Success Center* [Available at: http://imsuccesscenter.com/lead-retention-capture-prevent-leak-sales-funnel/].

Dong, Y., Yao, Y., & Cui, T. H. (2011). “When acquisition spoils retention: Direct selling vs. Delegation under CRM”. *Management Science*, *57*(7), 1288-1299.

DuFrene, D. D., Engelland, B. T., Lehman, C. M., & Pearson, R. A. (2005). “Changes in consumer attitudes resulting from participation in a permission e-mail campaign”. *Journal of Current Issues & Research in Advertising*, *27*(1), 65-77.

Durbin, R., Eddy, S. R., Krogh, A., & Mitchison, G. (1998). “*Biological sequence analysis: probabilistic models of proteins and nucleic acids*”. Cambridge University Press.

Ebbes, P., & Netzer, O. (2017). “Using Hidden Markov Models to identify and target job seekers for social network data”. *Working Paper*.

Eddy, S. R. (1998). “Profile hidden Markov models”. *Bioinformatics*, 14(9), 755-763.

Edelman, D. C. (2010). “Branding in the digital age”. *Harvard Business Review*, *88*(12), 62-69.

Edelman, D. C., & Singer, M. (2015). “The New Consumer Decision Journey”. [Available at https://www.mckinsey.com/business-functions/marketing-and-sales/our-insights/the-new-consumer-decision-journey]

Edwards, S. M., Li, H., & Lee, J. H. (2002). “Forced exposure and psychological reactance: Antecedents and consequences of the perceived intrusiveness of pop-up ads”. *Journal of Advertising*, 31(3), 83-95.

Ehrenberg, A. S. (2000). “Repetitive advertising and the consumer”. *Journal of Advertising Research*, *40*(6), 39-48.

Ehrenberg, A. S. (2000). “Repeat buying”. *Journal of Empirical Generalisations in Marketing Science*, *5*(2).

Ellis-Chadwick, F., & Doherty, N. F. (2012). “Web advertising: The role of e-mail marketing”. *Journal of Business Research*, 65(6), 843-848.

eMarketer (2017). “The Cost of Acquiring a Mobile App User - There’s always an app for that”. [Available at https://www.emarketer.com/Article/Cost-of-Acquiring-Mobile-App-User/1016688?ecid=NL1009]

Engel, J. F., Blackwell, R. D., & Miniard, P. W. (1968). “*Consumer Behavior*”. New York: Holt, Rinehart and Winston.

Engel, J. F., Miniard, P. W., & Blackwell, R. D. (2006). “*Consumer Behavior 10th Edition*”. Thomson South-Western. Mason. USA.

Erdem, T., & Sun, B. (2001). “Testing for choice dynamics in panel data”. *Journal of Business & Economic Statistics*, 19(2), 142-152.

Extole (2015). “Mobile Handbook. Optimize your customer referral programs for everyday moments”. [Available at http://go.extole.com/rs/377-QZQ-998/images/Extole_Mobile_Handbook.pdf].

Fader, P. (2012). “*Customer centricity: Focus on the right customers for strategic advantage*”. Wharton digital press.

Fader, P. S., Hardie, B. G., & Lee, K. L. (2005). “RFM and CLV: Using iso-value curves for customer base analysis”. *Journal of Marketing Research*, *42*(4), 415-430.

Farquhar, J. D. (2005). “Retaining customers in UK financial services: The retailers' tale”. *The Service Industries Journal*, *25*(8), 1029-1044.

Faulds, D. J., Mangold, W. G., Raju, P. S., & Valsalan, S. (2018). “The mobile shopping revolution: Redefining the consumer decision process”. *Business Horizons*, *61*(2), 323-338.

Frambach, R. T., Roest, H. C., & Krishnan, T. V. (2007). “The impact of consumer internet experience on channel preference and usage intentions across the different stages of the buying process”. *Journal of Interactive Marketing*, 21(2), 26-41.

Franses, P. H. (2005). “Diagnostics, expectations, and endogeneity”. *Journal of Marketing Research*, *42*(1), 27-29.

Fruchter, G. E., & Zhang, Z. J. (2004). “Dynamic Targeted Promotions a Customer Retention and Acquisition Perspective”. *Journal of Service Research*,*7*(1), 3-19.

Garbarino, E., & Johnson, M. S. (1999). “The different roles of satisfaction, trust, and commitment in customer relationships”. *the Journal of Marketing*, 70-87.

Gatignon, H., & Robertson, T. S. (1985). “A propositional inventory for new diffusion research”. *Journal of Consumer Research*, 849-867.

Ghose, A., & Todri, V. (2015). “Towards a digital attribution model: Measuring the impact of display advertising on online consumer behavior”. Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2672090.

Ghose, A., & Yang, S. (2009). “An empirical analysis of search engine advertising: Sponsored search in electronic markets”. *Management Science*, 55(10), 1605-1622.

Goldfarb, A., & Tucker, C. (2011). "Online display advertising: Targeting and obtrusiveness". *Marketing Science*, 30(3), 389-404.

Gollwitzer, P. M., & Bayer, U. (1999). "Deliberative versus implemental mindsets in the control of action". *Dual-Process Theories in Social Psychology*, 403-422.

Google (2017a). "About the default attribution models". [Available at: https://support.google.com/analytics/answer/1665189?hl=en]

Google (2017b). "Data-driven attribution methodology" [Available at: https://support.google.com/analytics/answer/3191594?hl=en]

Griffin, J., & Lowenstein, M. W. (2002). "*Customer Win back: How to Recapture Lost Customers--And Keep Them Loyal*". John Wiley & Sons.

Guadagni, P. M., & Little, J. D. (2008). "A logit model of brand choice calibrated on scanner data". *Marketing Science*, 27(1), 29-48.

Gupta, S., Hanssens, D., Hardie, B., Kahn, W., Kumar, V., Lin, N., & Sriram, S. (2006). "Modeling customer lifetime value". *Journal of Service Research*, *9*(2), 139-155.

Gupta, S., & Zeithaml, V. (2006). "Customer metrics and their impact on financial performance". *Marketing Science*, *25*(6), 718-739.

Hamilton, J. D. (1989). "A new approach to the economic analysis of nonstationary time series and the business cycle". *Econometrica: Journal of the Econometric Society*, 357-384.

Hamilton, J. D. (2010). "Regime switching models". In *Macroeconometrics and time series analysis* (pp. 202-209). Palgrave Macmillan, London.

Hansotia, B. J., & Wang, P. (1997). "Analytical challenges in customer acquisition". *Journal of Interactive Marketing*, *11*(2), 7-19.

Hart, C. W. (1999). "Customers are your business". *Marketing Management*,*8*(4), 6.

Hauser, J. R., Urban, G. L., Liberali, G., & Braun, M. (2009). "Website morphing". *Marketing Science*, 28(2), 202-223.

Heckman, J. J. (1981). "Heterogeneity and state dependence". In *Studies in labor markets*, University of Chicago Press, 91-140.

Henry, N. W. (2004). "Latent structure analysis". *Encyclopedia of Statistical Sciences*, 6.

Hoffman, D. L., & Novak, T. P. (1996). "Marketing in hypermedia computer-mediated environments: Conceptual foundations". *Journal of Marketing*, 50-68.

Homburg, C., Droll, M., & Totzek, D. (2008). "Customer prioritization: Does it pay off, and how should it be implemented?". *Journal of Marketing*, *72*(5), 110-130.

Horsky, D., Misra, S., & Nelson, P. (2006). "Observed and unobserved preference heterogeneity in brand-choice models". *Marketing Science*, 25(4), 322-335.

Howard, J. A., & Sheth, J. N. (1969). "*The Theory of Buyer Behavior*". New York: John Wiley & Sons, Inc.

Hughes, J. P., & Guttorp, P. (1994). "A class of stochastic models for relating synoptic atmospheric patterns to regional hydrologic phenomena". *Water resources research*, 30(5), 1535-1546.

HVS (2008). "Marriott's New Sales Force Strategy". [Available at http://www.hvs.com/article/3543/marriotts-new-sales-force-strategy/].

Jansen, B. J., & Schuster, S. (2011). "Bidding on the buying funnel for sponsored search and keyword advertising". *Journal of Electronic Commerce Research*, *12*(1), 1.

Jerath, K., Ma, L., & Park, Y. H. (2014). "Consumer click behavior at a search engine: The role of keyword popularity". *Journal of Marketing Research*, 51(4), 480-486.

Johnson, E. J., Moe, W. W., Fader, P. S., Bellman, S., & Lohse, G. L. (2004). "On the depth and dynamics of online search behavior". *Management Science*, 50(3), 299-308.

Jolley, W., Lee, A., Mizerski, R., & Sadeque, S. (2013). "Permission email messages significantly increase gambler retention". *Journal of Business Research*, *66*(9), 1617-1622.

Joo, M., Wilbur, K. C., & Zhu, Y. (2016). "Effects of TV advertising on keyword search". *International Journal of Research in Marketing*, 33(3), 508-523.

Kalish, S. (1985). "A new product adoption model with price, advertising, and uncertainty". *Management science, 31*(12), 1569-1585.

Kamakura, W. A., & Russell, G. J. (1989). "A probabilistic choice model for market segmentation and elasticity structure". *Journal of marketing research*, 379-390.

Kappe, E., Blank, A. S., & DeSarbo, W. S. (2018). "A random coefficients mixture hidden Markov model for marketing research". *International Journal of Research in Marketing*, 35(3), 415-431.

Katz, E., & Lazarsfeld, P. F. (1955). "*Personal Influence. The part played by people in the flow of mass communications"*. Transaction Publishers.

Katz, M. L., & Shapiro, C. (1994). "Systems competition and network effects". *The Journal of Economic Perspectives*, *8*(2), 93-115.

Kim, K. H., & Kumar, V. (2018). "The Relative Influence of Economic and Relational Direct Marketing Communications on Buying Behavior in Business-to-Business Markets". *Journal of Marketing Research*, 55(1), 48-68.

Kim, S. J., Wang, R. J. H., & Malthouse, E. C. (2015). "The Effects of Adopting and Using a Brand's Mobile Application on Customers' Subsequent Purchase Behavior". *Journal of Interactive Marketing*, *31*, 28-41.

Kireyev, P., Pauwels, K., & Gupta, S. (2016). "Do display ads influence search? Attribution and dynamics in online advertising". *International Journal of Research in Marketing*, 33(3), 475-490.

Klipfolio (2015). “The value of monitoring how prospects convert into paying customers”. [Available at http://www.klipfolio.com/blog/how-prospects-convert].

Komiak, S. Y., & Benbasat, I. (2006). “The effects of personalization and familiarity on trust and adoption of recommendation agents”. *MIS Quarterly*, 941-960.

Kotler, P., & Armstrong, G. (2011). “*Principles of marketing*”. Prentice–Hall of India, New Delhi, ed, 13, 61.

Kotler, P., & Armstrong, G. (2013). “*Principles of Marketing (16th Global Edition)*”. Pearson.

Krishnamurthy, S. (2001). “A comprehensive analysis of permission marketing”. *Journal of Computer-Mediated Communication*, *6*(2), 0-0.

Krishnamurthi, L., & Raj, S. P. (1991). “An empirical analysis of the relationship between brand loyalty and consumer price elasticity”. *Marketing Science*, 10(2), 172-183.

Kumar, V., George, M., & Pancras, J. (2008). “Cross-buying in retailing: Drivers and consequences”. *Journal of Retailing*, 84(1), 15-27.

Kumar, V., & Petersen, J. A. (2012). “*Statistical Methods in Customer Relationship Management”*. John Wiley & Sons.

Kumar, V., Sriram, S., Luo, A., & Chintagunta, P. K. (2011). “Assessing the effect of marketing investments in a business marketing context”. *Marketing Science*, 30(5), 924-940.

Kumar, V., Zhang, X., & Luo, A. (2014). “Modeling customer opt-in and opt-out in a permission-based marketing context”. *Journal of Marketing Research*, *51*(4), 403-419.

Lambrecht, A., Seim, K., & Tucker, C. (2011). “Stuck in the adoption funnel: The effect of interruptions in the adoption process on usage”. *Marketing Science*, *30*(2), 355-367.

Lambrecht, A., & Tucker, C. (2013). “When does retargeting work? Information specificity in online advertising”. *Journal of Marketing Research*, 50(5), 561-576.

Lavidge, R. J., & Steiner, G. A. (1961). “A model for predictive measurements of advertising effectiveness”. *The Journal of Marketing*, 59-62.

Lazarsfeld, P. F. (1950). “The logical and mathematical foundation of latent structure analysis”. *Studies in Social Psychology in World War II Vol. IV: Measurement and Prediction*, 362-412.

Lemon, K. N., & Verhoef, P. C. (2016). “Understanding customer experience throughout the customer journey”. *Journal of Marketing*, 80(6), 69-96.

Levitt, T. (1960). “Marketing myopia”. *Harvard Business Review*, 38(7/8), 26-44, 173-181.

Lewis, M. (2006a). “The effect of shipping fees on customer acquisition, customer retention, and purchase quantities”. *Journal of Retailing*, *82*(1), 13-23.

Lewis, M. (2006b). “Customer acquisition promotions and customer asset value”. *Journal of Marketing Research*, *43*(2), 195-203.

Li, H., & Kannan, P. K. (2014). “Attributing conversions in a multichannel online marketing environment: An empirical model and a field experiment”. *Journal of Marketing Research*, 51(1), 40-56.

Li, H., Kannan, P. K., Viswanathan, S., & Pani, A. (2016). “Attribution strategies and return on keyword investment in paid search advertising”. *Marketing Science*, 35(6), 831-848.

Li, S., Sun, B., & Montgomery, A. L. (2011). “Cross-selling the right product to the right customer at the right time”. *Journal of Marketing Research*, 48(4), 683-700.

Lin, Y. S., & Huang, J. Y. (2006). “Internet blogs as a tourism marketing medium: A case study”. *Journal of Business Research*, 59(10-11), 1201-1205.

Little, R. J., & Rubin, D. B. (2002). “*Statistical analysis with missing data (2nd edition)*”. John Wiley & Sons.

Liu, H., Pancras, J., & Houtz, M. (2015). “Managing customer acquisition risk using co-operative databases”. *Journal of Interactive Marketing*, 29, 39-56.

Lix, T. S., Berger, P. D., & Magliozzi, T. L. (1995). “New customer acquisition: Prospecting models and the use of commercially available external data”. *Journal of Direct Marketing*, *9*(4), 8-18.

Lovejoy, J. (2014). “The Consumer Journey: Explaining the Path to Purchase”. [Available at https://www.brandwatch.com/blog/path-to-purchase/]

Luo, A., & Kumar, V. (2013). “Recovering Hidden Buyer–Seller Relationship States to Measure the Return on Marketing Investment in Business-to-Business Markets”. *Journal of Marketing Research*, 50(1), 143-160.

Luxury Daily (2011). “Mobile apps can be customer acquisition tools: experts”. [Available at http://www.luxurydaily.com/mobile-apps-can-be-customer-acquisition-tools-experts/].

Ma, L., Sun, B., & Kekre, S. (2015). “The Squeaky Wheel Gets the Grease—An empirical analysis of customer voice and firm intervention on Twitter”. *Marketing Science*, 34(5), 627-645.

Malthouse, E. C., & Blattberg, R. C. (2005). “Can we predict customer lifetime value?”. *Journal of Interactive Marketing*, *19*(1), 2-16.

Malthouse, E. C., & Elsner, R. (2006). “Customisation with crossed-basis sub-segmentation”. *Journal of Database Marketing & Customer Strategy Management*, 14(1), 40-50.

Mamon, R. S., & Elliott, R. J. (Eds.). (2007). “*Hidden Markov models in finance*” (Vol. 460). New York: Springer.

Manchanda, P., Xie, Y., & Youn, N. (2008). “The Role of Targeted Communication and Contagion in Product Adoption”. *Marketing Science*, 27(6), 961-976.

Marinova, A., Murphy, J., & Massey, B. L. (2002). “Permission e-mail marketing as a means of targeted promotion”. *Cornell Hospitality Quarterly*, *43*(1), 61.

Martin, K. D., & Murphy, P. E. (2017). "The role of data privacy in marketing". *Journal of the Academy of Marketing Science*, 45(2), 135-155.

McLachlan, G. J., & Peel, D. (2000). "*Finite Mixture Models*". Wiley, New York.

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). "Birds of a feather: Homophily in social networks". *Annual Review of Sociology*, 415-444.

Min, S., Zhang, X., Kim, N., & Srivastava, R. K. (2016). "Customer acquisition and retention spending: An analytical model and empirical investigation in wireless telecommunications markets". *Journal of Marketing Research*, 53(5), 728-744.

Moe, W. W. (2003). "Buying, searching, or browsing: Differentiating between online shoppers using in-store navigational clickstream". *Journal of Consumer Psychology*, 13(1-2), 29-39.

Moe, W. W., & Fader, P. S. (2004). "Dynamic conversion behavior at e-commerce sites". *Management Science*, 50(3), 326-335.

Moffet, T., Pilecki, M., & McAdams, R. (2014) "The Forrester Wave™: Cross-Channel Attribution Providers, Q4 2014". [Available at https://www.forrester.com/report/The+Forrester+Wave+CrossChannel+Attribution+Providers+Q4+2014/-/E-RES115221]

Monetate (2018) "Ecommerce Quarterly Q1 2018". [Available at https://info.monetate.com/rs/092-TQN-434/images/EQ1-2018_First-Impressions.pdf].

Montgomery, A. L., Li, S., Srinivasan, K., & Liechty, J. C. (2004). "Modeling online browsing and path analysis using clickstream data". *Marketing Science*, 23(4), 579-595.

Montoya, R., Netzer, O., & Jedidi, K. (2010). "Dynamic allocation of pharmaceutical detailing and sampling for long-term profitability". *Marketing Science*, 29(5), 909-924.

Morrison, D. G., & Silva-Risso, J. (1995). "A latent look at empirical generalizations". *Marketing Science, 14*(3_supplement), G61-G70.

Musalem, A., & Joshi, Y. V. (2009). "Research Note—How Much Should You Invest in Each Customer Relationship? A Competitive Strategic Approach". *Marketing Science*, 28(3), 555-565.

Naik, P. A., & Peters, K. (2009). "A hierarchical marketing communications model of online and offline media synergies". *Journal of Interactive Marketing*, 23(4), 288-299.

Natter, M., Ozimec, A. M., & Kim, J. Y. (2015). "Practice Prize Winner—ECO: Entega's Profitable New Customer Acquisition on Online Price Comparison Sites". *Marketing Science*, 34(6), 789-803.

Nam, S., Manchanda, P., & Chintagunta, P. K. (2010). "The effect of signal quality and contiguous word of mouth on customer acquisition for a video-on-demand service". *Marketing Science*, *29*(4), 690-700.

Netzer, O., Ebbes, P., & Bijmolt, T. H. (2017). “Hidden markov models in marketing”. In *Advanced methods for modeling markets* (pp. 405-449). Springer, Cham.

Netzer, O., Lattin, J. M., & Srinivasan, V. (2008). “A Hidden Markov model of customer relationship dynamics”. *Marketing Science*, 27(2), 185-204.

Nitzan, I., & Libai, B. (2011). “Social effects on customer retention”. *Journal of Marketing*, *75*(6), 24-38.

Nooteboom, B., Berger, H., & Noorderhaven, N. G. (1997). “Effects of trust and governance on relational risk”. *Academy of Management Journal*, 40(2), 308-338.

Ofek, E., Katona, Z., & Sarvary, M. (2011). ““Bricks and clicks”: The impact of product returns on the strategies of multichannel retailers”. *Marketing Science*, 30(1), 42-60.

Okazaki, S., & Hirose, M. (2009). “Effects of displacement–reinforcement between traditional media, PC internet and mobile internet: A quasi-experiment in Japan”. *International Journal of Advertising*, 28(1), 77-104.

Olenski, S. (2017). “Marketing's Next Big Hurdle: The Rising Cost Of Customer Acquisition”. *Forbes*. [Available at: https://www.forbes.com/sites/steveolenski/2017/11/18/marketings-next-big-hurdle-the-rising-cost-of-customer-acquisition/#1a97d7d02298]

Oliver, R. L. (1980). “A Cognitive Model of the Antecedents and Consequences of Satisfaction Decisions”, *Journal of Marketing Research*, 17, pp. 460-469.

Ovchinnikov, A., Boulu-Reshef, B., & Pfeifer, P. E. (2014). “Balancing acquisition and retention spending for firms with limited capacity”. *Management Science*, 60(8), 2002-2019.

Owyang, J. (2012). “Google+ Has a Perception Problem”. [Available at: http://www.web-strategist.com/blog/2012/02/09/google-has-a-perception-problem/]

Paas, L. J., Vermunt, J. K., & Bijmolt, T. H. (2007). “Discrete time, discrete state latent Markov modelling for assessing and predicting household acquisitions of financial products”. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 170(4), 955-974.

Park, S., & Gupta, S. (2011). “A regime-switching model of cyclical category buying”. *Marketing Science*, 30(3), 469-480.

Pauwels, K., & van Ewijk, B. (2013). “Do online behavior tracking or attitude survey metrics drive brand sales? An integrative model of attitudes and actions on the consumer boulevard”. *Marketing Science Institute Working Paper Series*, *13*(118), 1-49.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1992). “Behavioral decision research: A constructive processing perspective”. *Annual review of psychology*, *43*(1), 87-131.

Peppers, D., & Rogers, M. (2016). “*Managing customer experience and relationships: A strategic framework*”. John Wiley & Sons.

Peters, C., Amato, C. H., & Hollenbeck, C. R. (2007). “An exploratory investigation of consumers' perceptions of wireless advertising”. *Journal of Advertising*, 36(4), 129-145.

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

Popovich, D., & Hamilton, R. (2014). “The Desire to Acquire Wish List Items: The Ironic Effect of Choosing to Delay Aspirational Purchases”. *ACR North American Advances*, 76-80.

Preston, I. L. (1982). “The Association Model of the advertising communication process”. *Journal of Advertising*, *11*(2), 3-15.

Preston, I. L., & Thorson, E. (1983). “Challenges to the use of hierarchy models in predicting advertising effectiveness”. In *Annual Convention of American Academy of Advertising*.

Preston, I. L., & Thorson, E. (1984). The expanded Association Model: Keeping the Hierarchy Concept alive.”. *Journal of Advertising Research*, *24*(1), 59-65.

Prins, R., & Verhoef, P. C. (2007). “Marketing Communication Drivers of Adoption Timing of a New E-Service Among Existing Customers”. *Journal of Marketing*, *71*, 169-183.

Rabin, M. (1993). “Incorporating fairness into game theory and economics”. *The American Economic Review*, 1281-1302.

Rabiner, L. R., Lee, C. H., Juang, B. H., & Wilpon, J. G. (1989). “HMM clustering for connected word recognition”. *Proceedings International Conference on Acoustics, Speech, and Signal Processing, IEEE Press*, 405-408.

Rainer, G., & Miller, E. K. (2000). “Neural ensemble states in prefrontal cortex identified using a hidden Markov model with a modified EM algorithm”. *Neurocomputing*, 32, 961-966.

Reibstein, D. J. (2002). “What attracts customers to online stores, and what keeps them coming back?”. *Journal of the Academy of Marketing Science*,*30*(4), 465-473.

Reinartz, W., & Kumar, V. (2003). “The Impact of Customer Relationship Characteristics on Profitable Lifetime Duration”. *Journal of Marketing*, 67(1), 77-99.

Reinartz, W., Thomas, J. S., & Kumar, V. (2005). “Balancing acquisition and retention resources to maximize customer profitability”. *Journal of Marketing*, *69*(1), 63-79.

Risselada, H., Verhoef, P. C., & Bijmolt, T. H. (2014). “Dynamic effects of social influence and direct marketing on the adoption of high-technology products”. *Journal of Marketing*, *78*(2), 52-68.

Roberts, J. H., & Lattin, J. M. (1991). “Development and testing of a model of consideration set composition”. *Journal of Marketing Research*, 429-440.

Robertson, T. S. (1971). “*Innovative behavior and communication”*. Holt McDougal.

Rogers, E. M. (1962). “*Diffusion of innovations”*. Simon and Schuster.

Rogers, E. M. (1983). “*Diffusion of innovations, 3rd edition”*. Simon and Schuster.

Romero, J., Van der Lans, R., & Wierenga, B. (2013). “A partially hidden Markov model of customer dynamics for CLV measurement”. *Journal of interactive Marketing*, 27(3), 185-208.

Saboo, A. R., Kumar, V., & Ramani, G. (2016). "Evaluating the impact of social media activities on human brand sales." *International Journal of Research in Marketing*, 33(3), 524-541.

Sarner, A., & Herschel, G. (2008). "A Checklist for Evaluating an Inbound and Outbound Multichannel Campaign Management Application". *Report G00160776, Gartner, Stamford, CT.*

Schlosser, A. E., White, T. B., & Lloyd, S. M. (2006). "Converting web site visitors into buyers: how web site investment increases consumer trusting beliefs and online purchase intentions". *Journal of Marketing*, 70(2), 133-148.

Schmitt, P., Skiera, B., & Van den Bulte, C. (2011). "Referral programs and customer value". *Journal of Marketing*, *75*(1), 46-59.

Schumann, J. H., von Wangenheim, F., & Groene, N. (2014). "Targeted online advertising: Using reciprocity appeals to increase acceptance among users of free web services". *Journal of Marketing*, 78(1), 59-75.

Schwarz, G. (1978). "Estimating the dimension of a model". *Annals of Statistics*, 6, 461–464.

Schwartz, E. M., Bradlow, E. T., & Fader, P. S. (2017). "Customer acquisition via display advertising using multi-armed bandit experiments". *Marketing Science*, 36(4), 500-522.

Schweidel, D. A., Bradlow, E. T., & Fader, P. S. (2011). "Portfolio dynamics for customers of a multiservice provider". *Management Science*, 57(3), 471-486.

Schweidel, D. A., Fader, P. S., & Bradlow, E. T. (2008). "A bivariate timing model of customer acquisition and retention". *Marketing Science*, *27*(5), 829-843.

SendPulse, 2017. "What is Prospective customer – Meaning". [Available at: https://sendpulse.com/support/glossary/prospective-customer]

Shah, D., Rust, R. T., Parasuraman, A., Staelin, R., & Day, G. S. (2006). "The path to customer centricity". *Journal of Service Research*, *9*(2), 113-124.

Shankar, V., & Balasubramanian, S. (2009). "Mobile Marketing: a Synthesis and Prognosis". *Journal of Interactive Marketing*, 23(2), 118-129.

Shankar, V., & Malthouse, E. C. (2007). "The growth of interactions and dialogs in interactive marketing". *Journal of Interactive Marketing*, 21(2), 2-4.

Shao, X., & Li, L. (2011). "Data-driven multi-touch attribution models". *In Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 258-264). ACM.

Shapley, L. S. (1988). "*A value for n-person games*". Princeton University Press, 1953. 11, 58.

Sheth, J. N., Sisodia, R. S., & Sharma, A. (2000). "The Antecedents and Consequences of Customer-Centric Marketing". *Journal of the Academy of Marketing Science*, *28*(1), 55-66.

Shugan, S. M. (2004). “Endogeneity in marketing decision models”. *Marketing Science, 23*(1), 1-3.

Sirakaya, E., & Woodside, A. G. (2005). “Building and testing theories of decision making by travelers”. *Tourism management, 26*(6), 815-832.

Song, Y., Sahoo, N., Srinivasan, S., & Dellarocas, C. (2016). “Uncovering Characteristic Paths to Purchase of Consumers”. *Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2619674*.

Spekman, R. E. (1988). “Strategic supplier selection: understanding long-term buyer relationships”. *Business Horizons*, 31(4), 75-81.

Srinivasan, S. S., Anderson, R., & Ponnavolu, K. (2002).” Customer loyalty in e-commerce: an exploration of its antecedents and consequences”. *Journal of Retailing*, 78(1), 41-50.

Srinivasan, S., Vanhuele, M., & Pauwels, K. (2010). “Mind-set metrics in market response models: An integrative approach”. *Journal of Marketing Research*, *47*(4), 672-684.

Stamp, M. (2015). “A revealing introduction to hidden Markov models”. *Department of Computer Science San Jose State University*, 26-56.

Starch, D. (1923). “*Principles of Advertising*”. New York, McGraw-Hill Book Company.

Statista (2017). “Number of digital buyers worldwide from 2014 to 2021 (in billions)”. [Available at https://www.statista.com/statistics/251666/number-of-digital-buyers-worldwide/]

Stauss, B., & Friege, C. (1999). “Regaining service customers. Costs and benefits of regain management”. *Journal of Service Research, 1*(4), 347-361.

Stephen, A. T., & Galak, J. (2012). “The effects of traditional and social earned media on sales: A study of a microlending marketplace”. *Journal of Marketing Research*, 49(5), 624-639.

Strong, E. K. (1925). “*The psychology of selling and advertising*”. McGraw-Hill.

Suk, K., Lee, J., & Lichtenstein, D. R. (2012). “The influence of price presentation order on consumer choice”. *Journal of Marketing Research*, 49(5), 708-717.

Tellis, G. J., Stremersch, S., & Yin, E. (2003). “The international takeoff of new products: The role of economics, culture, and country innovativeness”. *Marketing Science*, *22*(2), 188-208.

Thomas, J. S. (2001). “A methodology for linking customer acquisition to customer retention”. *Journal of Marketing Research, 38*(2), 262-268.

Thomas, J. S., Blattberg, R. C., & Fox, E. J. (2004). “Recapturing lost customers”. *Journal of Marketing Research*, *41*(1), 31-45.

Tillmanns, S., Ter Hofstede, F., Krafft, M., & Goetz, O. (2017). “How to separate the wheat from the chaff: Improved variable selection for new customer acquisition”. *Journal of Marketing*, 81(2), 99-113.

Trusov, M., Bucklin, R. E., & Pauwels, K. (2009). “Effects of word-of-mouth versus traditional marketing: findings from an internet social networking site”. *Journal of Marketing*, *73*(5), 90-102.

Tucker, C. E. (2014). “Social networks, personalized advertising, and privacy controls”. *Journal of Marketing Research*, 51(5), 546-562.

Tversky, A., Sattath, S., & Slovic, P. (1988). “Contingent weighting in judgment and choice”. *Psychological review*, *95*(3), 371.

Vakratsas, D., & Ambler, T. (1999). “How advertising works: what do we really know?”. *The Journal of Marketing*, 26-43.

Van den Bulte, C., & Lilien, G. L. (2001). “Medical innovation revisited: Social contagion versus marketing effort”. *American Journal of Sociology*, *106*(5), 1409-1435.

Van den Bulte, C., & Lilien, G. (1999). “A two-stage model of innovation adoption with partial observability: Model development and application”. *The Wharton School, University of Pennsylvania*.

Van den Bulte, C., & Stremersch, S. (2004). “Social contagion and income heterogeneity in new product diffusion: A meta-analytic test”. *Marketing Science*, *23*(4), 530-544.

Van Lennep, V. (2014). “The funnel is dead. Long live the consumer decision journey”. [Available at https://www.eiuperspectives.economist.com/marketing/funnel-dead-long-live-consumer-decision-journey]

Verbeek, M. (2008). *A guide to modern econometrics*. John Wiley & Sons.

Verhoef, P. C., & Donkers, B. (2005). “The effect of acquisition channels on customer loyalty and cross-buying”. *Journal of Interactive Marketing*, 19(2), 31-43.

Verhoef, P. C., Kannan, P. K., & Inman, J. J. (2015). “From multi-channel retailing to omni-channel retailing: introduction to the special issue on multi-channel retailing”. *Journal of Retailing*, 91(2), 174-181.

Verhoef, P. C., Neslin, S. A., & Vroomen, B. (2007). “Multichannel customer management: Understanding the research-shopper phenomenon”. *International Journal of Research in Marketing*, 24(2), 129-148.

Villanueva, J., Yoo, S., & Hanssens, D. M. (2008). “The impact of marketing-induced versus word-of-mouth customer acquisition on customer equity growth”. *Journal of Marketing Research*, *45*(1), 48-59.

Visser, I., & Speekenbrink, M. (2010). “depmixS4: an R package for hidden Markov models”. *Journal of Statistical Software*, 36(7), 1-21.

Voss, G. B., & Voss, Z. G. (2008). “Competitive density and the customer acquisition–retention trade-off”. *Journal of Marketing*, 72(6), 3-18.

Wang, M., & Chan, D. (2011). “Mixture latent Markov modeling: Identifying and predicting unobserved heterogeneity in longitudinal qualitative status change”. *Organizational Research Methods*, 14(3), 411-431.

Wang, B., Kim, S., & Malthouse, E. C. (2016). “Branded Apps and Mobile Platforms as New Tools for Advertising”. *In The New Advertising: Branding, Content, and Consumer Relationships in the Data-driven Social Media Era*. ABC-CLIO Santa Barbara, CA.

Wang, X., Yu, C., & Wei, Y. (2012). “Social media peer communication and impacts on purchase intentions: A consumer socialization framework”. *Journal of Interactive Marketing*, *26*(4), 198-208.

Wangenheim, F. V., & Bayón, T. (2007). “The chain from customer satisfaction via word-of-mouth referrals to new customer acquisition”. *Journal of the Academy of Marketing Science*, *35*(2), 233-249.

Wedel, M., Kamakura, W., Arora, N., Bemmaor, A., Chiang, J., Elrod, T., Johnson, R., Lenk, P., Neslin, S., & Poulsen, C. S. (1999). “Discrete and continuous representations of unobserved heterogeneity in choice modeling”. *Marketing Letters*, 10(3), 219-232.

Weiss, A. M., Lurie, N. H., & MacInnis, D. J. (2008). “Listening to strangers: whose responses are valuable, how valuable are they, and why?”. *Journal of Marketing Research*, *45*(4), 425-436.

Welch, L. R. (2003). “Hidden Markov models and the Baum-Welch algorithm”. *IEEE Information Theory Society Newsletter*, 53(4), 10-13.

Wiesel, T., Pauwels, K., & Arts, J. (2011). “Practice prize paper—Marketing's profit impact: quantifying online and off-line funnel progression”. *Marketing Science*, *30*(4), 604-611.

Xu, L., Duan, J. A., & Whinston, A. (2014). “Path to purchase: A mutually exciting point process model for online advertising and conversion”. *Management Science*, 60(6), 1392-1412.

Yadav, M. S., De Valck, K., Hennig-Thurau, T., Hoffman, D. L., & Spann, M. (2013). “Social commerce: a contingency framework for assessing marketing potential”. *Journal of Interactive Marketing*, *27*(4), 311-323.

Yadav, M. S., & Pavlou, P. A. (2014). “Marketing in computer-mediated environments: Research synthesis and new directions”. *Journal of Marketing*, *78*(1), 20-40.

Yamato, J., Ohya, J., & Ishii, K. (1992). “Recognizing human action in time-sequential images using hidden markov model”. *Computer Vision and Pattern Recognition, 1992. Proceedings CVPR'92., 1992 IEEE Computer Society Conference on*, 379-385. IEEE.

Zhang, X., Kumar, V., & Cosguner, K. (2017). “Dynamically managing a profitable email marketing program”. *Journal of Marketing Research*, 54(6), 851-866.

Zhang, J. Z., Netzer, O., & Ansari, A. (2014). “Dynamic targeted pricing in B2B relationships”. *Marketing Science*, 33(3), 317-337.

Zubcsek, P. P., Katona, Z., & Sarvary, M. (2015). "Social and Location Effects in Mobile Advertising". *Available at SSRN:* [*https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2616901*](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2616901).

Zucchini, W., MacDonald, I. L., & Langrock, R. (2016). "*Hidden Markov models for time series: an introduction using R*". Chapman and Hall/CRC.

APPENDIX A

R code for the Transition Probabilities estimation

```
library(haven)
data <- read_dta("181206 Final Dataset no70.dta")
#data <- data1[data1$out50==0,]

# convert id
uid <- unique(data$user_id)
head(uid); length(uid)

n <- length(uid)
id <- 0
for(i in 1:n){
  ind <- which(data$user_id==uid[i])
  id[ind]=i
}
tv = NULL
for(i in 1:n) tv = c(tv,1:sum(id==i))


YY <- cbind(
  data$search_cat2,
  data$purch_cum_cat, data$monetary_cum_cat)

# frequencies for the response categories
apply(YY,2,table)

X2 <- cbind(
  #wishlist
  data$n_addwishlist, data$n_dropwishlist, data$n_addwishlist_lag,
  #registration
  data$social_registration, data$facebook_connected,
   data$registration_days, data$nl_registration_days,
   data$nl_unsubscription_days, data$reg_day, data$nl_reg_day,
   data$nl_uns_day,
  #search
  data$intersession_days,
  #demographics
   data$africa, data$america, data$asia, data$europe, data$italy,
   data$oceania, data$russia, data$usa, data$age, data$female,
   data$demo_miss)

colnames(X2)<- cbind(
  #wishlist
  "n_addwishlist","n_dropwishlist", "n_addwishlist_lag",
  #registration
```

```
  "social_registration", "facebook_connected", "registration_days",
   "nl_registration_days", "nl_unsubscription_days", "reg_day",
   "nl_reg_day", "nl_uns_day",
  #search
  "intersession_days",
  #demographics
   "africa", "america", "asia", "europe", "italy", "oceania", "russia",
   "usa", "age", "female", "demo_miss")


require(LMest)

source("est_lm_long.R")
source("for_long.R")
source("lk_obs_long.R")
source("invglob.R")
system("R CMD shlib for_long.f --preclean")
dyn.load("for_long.so")
system("R CMD shlib back_long.f --preclean")
dyn.load("back_long.so")
system("R CMD shlib multi_Be1.f --preclean")
dyn.load("multi_Be1.so")
system("R CMD shlib multi_Be2.f --preclean")
dyn.load("multi_Be2.so")
system("R CMD shlib multi_Ga1.f --preclean")
dyn.load("multi_Ga1.so")
system("R CMD shlib multi_Ga2.f --preclean")
dyn.load("multi_Ga2.so")


# estimate

est = est_lm_long(id,tv,
                  YY,
                  5,
                  X1=X2[tv==1,],
                  X2=X2[-which(tv==1),],
                  out_se=TRUE)

est$bic
est$np

round(est$Psi,3)
round(est$Be,3)
round(est$seBe,3)
round(est$Ga,3)
round(est$seGa,3)

IP<-data.frame(round(est$Piv, 3))

TP <- data.frame(round(est$PI,4))
TP11 <- data.frame(round(est$PI[1, 1, ], 4))
TP12 <-round(est$PI[1, 2, ], 4)
TP11<- cbind(TP11, TP12)
TP13 <-round(est$PI[1, 3, ], 4)
TP11<- cbind(TP11, TP13)
TP14 <-round(est$PI[1, 4, ], 4)
TP11<- cbind(TP11, TP14)
TP15 <-round(est$PI[1, 5, ], 4)
TP11<- cbind(TP11, TP15)
```

```
TP21 <-round(est$PI[2, 1, ], 4)
TP11<- cbind(TP11, TP21)
TP22 <-round(est$PI[2, 2, ], 4)
TP11<- cbind(TP11, TP22)
TP23 <-round(est$PI[2, 3, ], 4)
TP11<- cbind(TP11, TP23)
TP24 <-round(est$PI[2, 4, ], 4)
TP11<- cbind(TP11, TP24)
TP25 <-round(est$PI[2, 5, ], 4)
TP11<- cbind(TP11, TP25)

TP31 <-round(est$PI[3, 1, ], 4)
TP11<- cbind(TP11, TP31)
TP32 <-round(est$PI[3, 2, ], 4)
TP11<- cbind(TP11, TP32)
TP33 <-round(est$PI[3, 3, ], 4)
TP11<- cbind(TP11, TP33)
TP34 <-round(est$PI[3, 4, ], 4)
TP11<- cbind(TP11, TP34)
TP35 <-round(est$PI[3, 5, ], 4)
TP11<- cbind(TP11, TP35)

TP41 <-round(est$PI[4, 1, ], 4)
TP11<- cbind(TP11, TP41)
TP42 <-round(est$PI[4, 2, ], 4)
TP11<- cbind(TP11, TP42)
TP43 <-round(est$PI[4, 3, ], 4)
TP11<- cbind(TP11, TP43)
TP44 <-round(est$PI[4, 4, ], 4)
TP11<- cbind(TP11, TP44)
TP45 <-round(est$PI[4, 5, ], 4)
TP11<- cbind(TP11, TP45)

TP51 <-round(est$PI[5, 1, ], 4)
TP11<- cbind(TP11, TP51)
TP52 <-round(est$PI[5, 2, ], 4)
TP11<- cbind(TP11, TP52)
TP53 <-round(est$PI[5, 3, ], 4)
TP11<- cbind(TP11, TP53)
TP54 <-round(est$PI[5, 4, ], 4)
TP11<- cbind(TP11, TP54)
TP55 <-round(est$PI[5, 5, ], 4)
TP11<- cbind(TP11, TP55)

state1 <- cbind(data, est$V[,1])
state2 <- cbind(state1, est$V[,2])
state3 <- cbind(state2, est$V[,3])
state4 <- cbind(state3, est$V[,4])
state5 <- cbind(state4, est$V[,5])

summary(IP)
summary(TP11)
```