Exploring Merchants' Reluctance to Adopt e-Commerce Anti-fraud Tools

A Case Study in the Payments Industry

Maria I. Baltoglou

TU Delft Delft University of Technology

adyen
wherever people pay

Exploring Merchants' Reluctance to Adopt e-Commerce Anti-Fraud Tools:

A Case Study in the Payments Industry

Master thesis submitted to Delft University of Technology
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

in **Engineering and Policy Analysis**

Faculty of Technology, Policy and Management
by
Maria Ilektra Baltoglou
Student number: 4412842

To be defended in public on February 8th 2017

Graduation Committee

Chairman:	Prof. dr. M.J.G van Eeten,	TU Delft
First Supervisor:	Dr. ir. H. Asghari,	TU Delft
Second Supervisor:	Dr. ir. B. Enserink,	TU Delft
External Supervisor:	M. Apostolou,	Adyen

TU Delft

"Either write something worth reading
or do something worth writing."

- Benjamin Franklin

Preface

Writing my thesis has been a long journey, and surely not as easy as I would have imagined. When I started with this project, I left for the first time in my life the academic environment and I found myself in the payments industry, where everything was new to me. The process of conducting the research was often accompanied with feelings of frustration and mental fatigue; however, during this 9-month period I obtained new skills, evolved as a person and most importantly, I realized the value of my support network; people that without them, I would not have been able to successfully complete this challenge.

I want to begin by thanking my committee in TU Delft, since they have greatly contributed in the quality of this thesis. Specifically, I want to thank Michel van Eeten, the chairman, for guiding me from the very first moment and giving me concrete and sharp feedback. I would also like to extend my gratitude to my first supervisor, Hadi Asghari, who made me think in a critical way and set high research standards. I moreover thank Bert Enserink, my second supervisor, for his kind support throughout the research. I appreciated every single conversation with these three people, for they broadened my horizons, alleviated my stress and showed to me how it is to be passionate about your work.

At Adyen, I met several people that remarkably helped me this whole period. I am grateful towards my daily supervisor, Michail Apostolou, who took the initiative by himself to help me with this project. I feel that a large part of the thesis would not be feasible without his contribution and his willingness to share his knowledge with me. I also want to acknowledge Beau-Anne for her smart comments and guidance, as well as Mauricio for his constructive feedback. Furthermore, I feel privileged getting to know and having worked together with Daniel, Cristian, Traian, Sergios and Jelle. Many thanks go to my Greek colleague and friend Alexandros for the fun moments and for the daily cheering-up.

On a personal level, I would like to wholeheartedly thank my family and friends. Stefan for all the support, encouragement and belief in me; his advice to always keep a positive stance and to be faithful in my capabilities hugely contributed in accomplishing what I wanted. Thanks to my friend Hara who, no matter how far away, offered psychological support and moments of laughter. Finally, nothing of all these could ever be possible without the unconditional love and support of my family. I want to truly thank my parents for their understanding and inspiration throughout my whole journey. My gratitude also goes to my two elder brothers, who have always been there for me and act as an example to follow.

Maria

Delft, February 2017

Executive Summary

The current thesis is about risk management in the context of e-commerce fraud. E-commerce fraud poses a serious threat to business environments, as merchants' losses due to fraud amounted in $21.84 billion on a global scale during 2015. As such, several technical security solutions have been developed in order to detect fraudulent attempts in the payments industry and enable merchants protect their revenues.

From a merchant's perspective, while dealing with daily transactions, there are two costs that need to be balanced; costs related to chargebacks and costs related to refusals. In our case study, merchants have been offered a profit-maximizing anti-fraud tool, the *Risk Calculator*, which makes suggestions on how to balance chargebacks and refusals and thus reduce costs. However, from a practical perspective it has been noticed that merchants are reluctant to use the tool. The latter suggests they are not profit-maximizing, something that contradicts Rational Choice theory, where people act as self-maximizing individuals.

It should be noted at this point that the problem we are dealing with, is a socio-technical problem. By applying economic theories we can find plausible explanations in terms of people's *risk attitudes* and the implications of these attitudes. The main research question is defined as *"Why are merchants reluctant to adopt profit-maximizing risk management settings that are suggested by an anti-fraud tool that analyzes their transaction data and how can developers of such tools increase their acceptance?"*. To answer this question, the thesis entails three different research sections which altogether shed light on merchants' behavioral aspects; one explorative study regarding the usage of Risk Calculator through statistical analysis, one qualitative study through semi-structured interviews and finally one explanatory study where we look at broader factors that explain merchants' engagement with risk management. Through the empirical studies the main findings are the following.

Regarding the usage of the tool, we found that 153 companies visited the tool during a 3-month period. By comparing the visitors group with non-users of the tool, i.e. merchants that have never visited the page, we found that non-users are smaller companies than visitors in terms of transaction volume, and additionally have lower chargeback and refusal rates. Since the Risk Calculator is a statistical tool that needs certain volume of historical data in order to run the calculations, this finding suggest that the "good candidates" are indeed visiting the tool, or at least have seen it once, while there is a group of merchants for which the tool adds no value and hence they are not interested in using it. This is an expected observation, since according to Status Quo Bias in Behavioral Economics people prefer not to change behavior unless the incentive to do so

is strong; in this case, non-users seem to have already optimized transactions. Moreover, Prospect Theory captures this finding by suggesting that it is important to understand the general context when making a decision; in our case it is important whether a merchant is already doing well in terms of refused and fraudulent transactions and hence the tool should take this into account. The main problem related to the tool's usage was that practically none of the merchants is applying the suggestions, something that pinpoints to trust issues.

Furthermore, when merchants change risk scores through their settings they seem to increase the scores; on the contrary, the tool's suggestion most of the times is to decrease the risk scores. This means that merchants will be hesitant to accept suggestions which are not in line with their risk attitude; according to Prospect Theory merchants would prefer to forgo profit opportunities in order not to take the risk. Going a step further, we see that the suggested savings of the tool account for up to 0.5% of visitors' total revenues and up to 0.12% of non-users' total revenues; a quite small amount. This finding pinpoints to Security Economics, where people are not interested in using security tools when the time and effort they have to spend outweighs the benefits that the security advice brings.

Through the qualitative analysis, we were able to delve deeper into the problem. By talking to merchants and account managers we found out that the quality of output in terms of user interface and technical bugs, the insufficient documentation, the response time, interaction with it by the use of the slider, trust and the belief if it adds value to the business are the main challenges related to the tool's usage. Regarding the complexity of understanding how to use the tool, it was mentioned that people are not willing to spend time on it - a finding which is in line with Bounded Rationality Theory; decisions are subject to limitations in information and computational capacities and therefore might not be optimal, i.e. profit-maximizing. Apart from that, the indication of no trust in the underlying model is reflected firstly in the fact that all the interviewees stated they prefer a manual analysis since they consider the costs related to their transactions "too important to let a machine decide". This finding agrees with the concept of loss aversion: the psychological cost of loss is greater than the psychological benefit of gain, thus people prefer to put more energy (manual review vs. automated tool) in order to be sure they avoid losses. On the other hand, the non-trust is derived from the fact that the tool suggests extreme changes, i.e. putting a score from 100 to 0, to risk checks that are perceived by merchants very efficient in combating fraud. This finding might be a suggestion for the design of the tool, by giving the option to merchants to selectively apply the advices they consider most valuable.

It should be noted at this point that there was some contradictory feedback between developers in the company and merchants, as well as account managers regarding the reliability of the tool.

This might be reflected in the literature regarding alignment of incentives and actors' different perceptions; on one hand, developers are interested in company's infrastructure expansion, while account managers are more interested in potential loss of their reputation, since if the tool does not eventually work as expected, the merchants might no longer trust them; this is also reflected in their reluctance to promote the tool. Lastly, the merchants are the ones to be incurring the direct costs of chargebacks and refusals if the tool proves to be invalid. The above technical and non-technical issues influence engagement with the tool and can be regarded as the stepping stone for further improvements.

Moreover, we explored in more detail how active in terms of risk decisions is the group of merchants for which the tool is meant to be used from, and we saw that they are visiting their risk settings by making on average 1.35 changes per month. Additionally, most of the times merchants seem to ignore the tool's suggestions, but when they do make changes these are in line with the tool; a finding pointing again to the design of selectively applying the advices. Moreover, the majority of merchants does not make extreme changes to the risk scores, as the net score change equals to zero; under the finding that they do not trust the tool, they are faced with uncertainty when it comes to decide about scores. According to Behavioral Economics, when people are faced with uncertainty, it is more likely that they will go with the default, especially when it is presented as a recommended configuration.

Furthermore, we identified several drivers behind merchants' engagement with risk management. Firstly, the sector of merchants increases the probability of making changes to the risk settings. Specifically, it was found that merchants in information services and transportation industries are the ones that engage more in risk management. This can be explained by the fact that information services have zero marginal costs as they sell digital products and hence are not affected in terms of losses compared to physical goods industry. Regarding transportation industry, it is a sector heavily attacked by fraudsters and hence has developed awareness and maturity in trying to combat online fraud. Additionally, the higher transaction volume, reflecting merchant size, seems to be another factor increasing the risk activity. Large sized merchants have more resources - such as stuff and possibly financial tolerance towards losses - and therefore are more proactive in making risk choices. Moreover, a higher transaction value increases the probability of engaging more in risk changes. This is something expected, since the larger the transaction value, the more profit the merchant is losing. Lastly, the overall risk refusal rate is an indicator of increased risk activity; this is in line with the fact that merchants are highly interested in conversion. We also saw that chargeback rate seems not to have an effect on merchants risk activity, something that contradicts literature. However, it should be noted at this point that this is probably a result of

bias in the way chargebacks were stored in the database.

Finally, we identified the factors that influence the direction of the scores given to risk checks by merchants who do actively make changes to their settings. Specifically, sector seems to be the most significant factor determining the direction of scores. Merchants in transportation sector seem to decrease the scores, i.e. being more loose with their settings and thus willing to take more risk. This can possibly be explained by two facts; transportation industry, which in our case is mainly comprised by airlines, seems to be willing to absorb chargeback losses in an attempt to avoid declining transactions and lose out to the heavy competition. On the other hand, when buying a ticket it means that the person has to be physically present in order to travel, hence there is an increased probability of getting caught in case it is suspected they committed fraud. This might be a reassuring fact for the transportation industry, thus being more lenient towards risk. Apart from that, it was found that a decreasing transaction value leads to stricter scores. Although this might sound counter-intuitive, it was found that transactions with a value 0-20$ are more than twice as likely to be fraudulent than larger purchases.

Looking back at the research question, we can state that merchants' reluctance in adopting profit-maximizing advices suggested by anti-fraud tools can be explained both by behavioral as well as technical aspects. Firstly, the concepts in Behavioral Economics suggest that people do not always make the optimal decisions as noted in Rational Choice Theory; they do not change behavior when the incentives to do so are not strong, while their decisions are sub-optimal due to time and knowledge constraints. Moreover, they are not willing to follow advices if they are already satisfied with their status quo and if the effort they have to put on understanding the advice outweighs the economic benefit they gain. Apart from that, they tend to follow advices which are in line with their risk attitude. Secondly, technical aspects such as quality of output, ease of use, response time and accuracy in predictions are also important factors influencing the engagement with such tools. The research indicated that in order to increase the acceptance the anti-fraud tool examined in our case study, it is not sufficient to rely on analysis of transaction data, but rather provide different input to the tool. Particularly, the tool should take into account the type of merchants that use it, as well as the risk environment, and make different suggestions for optimization based on merchant classification. Apart from that, other improvements that can have a positive impact on its usage include (1) the embodiment of more information in the tool page, (2) the implementation of another algorithm of the alteration of the algorithm's constraints, (3) the implementation of A/B testing through the tool and (5) the redesign of the tool so that merchants can selectively apply the advices they find more valuable.

Contents

Chapter 1

Introduction

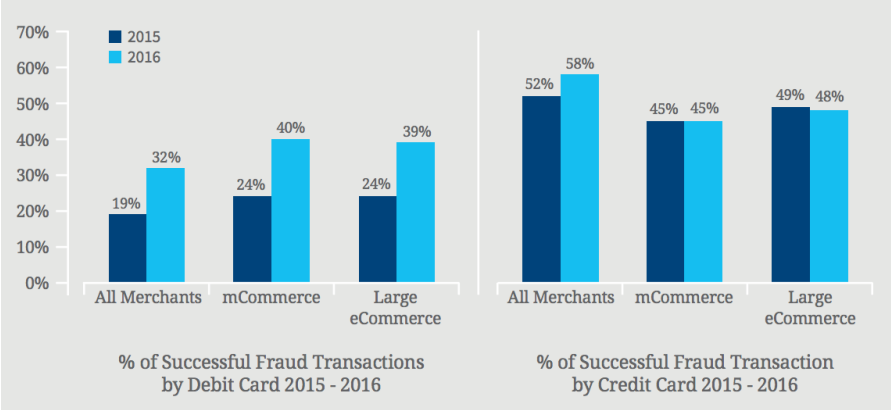
The rapid advancement of technology the last decades has made people's lives easier, especially in the field of transactions and online banking. The epitome of the progress in this field is represented by electronic payments, which benefit individuals with faster time of completion, ubiquity in terms of access, as well as reduced costs. The term "electronic payments" is used to indicate a transaction where credit cards, debit cards, internet banking, electronic funds transfers or direct credits/debits are used as payment systems.

Nowadays, e-commerce is taking over traditional commerce, with business-to-consumer (B2C) category being the most popular form, reaching the amount of 1.7 trillion US dollars in 2015 worldwide (Statista, 2016). Through e-commerce, firms have been able to easily establish a market presence or enhance their existing one. At the same time, shoppers have been able to benefit from lower prices and simplicity in the transaction process. As such, the growth of the payments industry was inevitable in order to facilitate this process.

However, e-commerce payment systems are used not only by legitimate shoppers, but as well by fraudsters who usually exploit cardholders' credit and debit cards without the victim's knowledge. According to LexisNexis annual report (LexisNexis, 2015), in 2016 merchants lost on average 1.47% of their revenues to fraud, a percentage which significantly increased over the past three years. Although fraudsters steal information of both credit and debit cards, credit card fraud is more widespread (Bhatla, Prabhu, & Dua, 2003); this fact is also confirmed in Figure 1.1. This Figure shows the percentage of successful fraudulent transactions by debit and credit card among eCommerce merchants, mCommerce merchants (i.e. mobile eCommerce merchants who accept payments through either a mobile browser or mobile application) and finally among all merchants. Credit cards represent a very popular payment method and are currently the second

most utilized mean for conducting transactions. The use of credit cards in daily transactions shows an exponential increase globally, with a yearly increment of 10-18% (Ferreira et al., 2015).

Figure 1.1: Percent of successful fraud transactions by debit / credit card (2015 – 2016) among mobile, eCommerce and all merchants (LexisNexis, 2015).



E-commerce fraud represents one of biggest threats to business establishments today (Bhatla et al., 2003) and apart from the revenue losses on a global scale, the cyber crimes committed through financial fraud support the growth of underground economy (Levchenko et al., 2011). Therefore, the need to secure against these risks in order to prevent losses is of critical importance. Crucial in achieving this safety is the notion of cyber security, i.e. the protection of Information Technology (IT) systems against fraud and theft (Moore, Friedman, & Procaccia, 2010).

1.1 Research Problem and Goal

In the context of e-commerce, fraud affects all the parties that are involved in the payment process. Although that according to common belief cardholders are the most affected stakeholder in card scams, several studies indicate that merchants are actually at higher risk (Bhatla et al., 2003; Wolters, 2012). The reason is that merchants carry not only the costs of the products sold and sent, but also the expenses related to chargeback fees - a term referred to reversing a payment after it has taken place. The complication that arises has mainly to do with the fact that merchants wish to combat fraudulent attempts, but at the same time allow legitimate shoppers come through.

The afore-described complexity has allowed dedicated businesses to develop risk-management solutions in order to solve the challenges related to fraudulent transactions. One of these

businesses is Adyen, a Dutch-based Payment Service Provider (PSP) that serves as a gateway to banks and card schemes, on behalf of the merchants. Adyen has developed a risk management system which is directly integrated in the payments platform and is designed to balance fraud defense and optimized conversion. Part of this risk system is an anti-fraud tool, namely the so-called *Risk Calculator*, which uses Machine Learning and based on historical data calculates the optimal configurations that merchants should use to combat fraud and maintain legitimate customers. At Adyen, we notice in practice a reluctance from merchants to use this anti-fraud tool. More specifically, out of Adyen's approximately 5,000 merchant customers, only around a hundred visit the tool, however without applying the suggestions it makes.

The assumption that merchants seem reluctant to adopt solutions for fighting fraudulent transactions is also verified by (LexisNexis, 2015). The research shows that even the merchants who invest resources in combating fraud, are not convinced that these solutions are effective. A reason contributing to this belief is that the majority of anti-fraud solutions relies on manual review, meaning that a large percent of the transactions flagged as potentially fraudulent are ultimately decided by humans. Moreover, according to the research, although merchants spend considerable amounts on these solutions, they still see fraudulent attempts increasing and conversion rates decreasing.

Industry and academia have focused a lot in creating solutions for combating fraud (Bahnsen, Aouada, Stojanovic, & Ottersten, 2016; Tan, Guo, Cahalane, & Cheng, 2016), without stretching the fact whether these solutions are being adopted by users or not. In this research we go a step further than adding solutions to the fraud problem or comparing which is the most effective way of dealing with fraudsters in e-commerce. Particularly, we explore the aspects that make merchants reluctant to adopt these solutions. Hence the main research goal can be summarized as: **Identifying the reluctance factors of adoption of anti-fraud tools and exploring whether analytics on transaction data is a good approach to indicate merchants what to do.**

From a societal perspective, this research can be beneficial in increasing trust in online payments by indicating what features anti-fraud solutions should enable in order to be used by merchants. As regards the scientific relevance, contemporary research focuses mainly on merely adding solutions to detect and combat financial fraud; to the best of our knowledge, little research is conducted on exploring through empirical research the factors that drive merchants to engage with risk management, especially in the payments industry context.

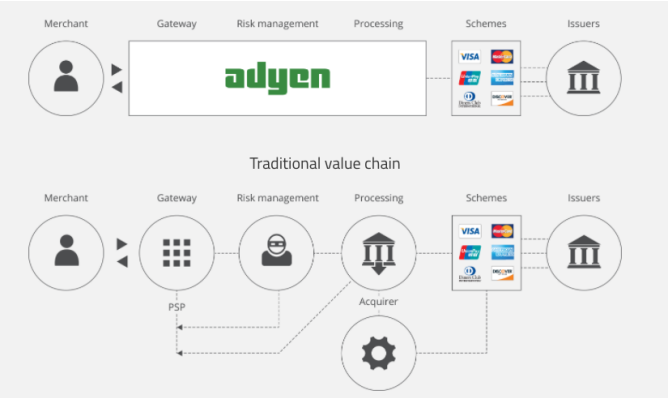
1.2 Research Opportunity

This research is embedded in the payment industry sector and its feasibility is highly dependent on the availability of appropriate data. Adyen, the leading PSP in the Netherlands, provided this unique opportunity for research. Adyen represents a player highly interested in the growth of trust in online transactions, and is thus influenced by the afore-described situation. However, quantitative research requires the availability of appropriate data, which are relevant to the research. Nevertheless, opportunities for quantitative research in the payment industry, and especially in the case where data related to fraud are involved, are uncommon.This is mainly due to legal and competition constraints. As such, most researches in this field are conducted with synthetic data. On the contrary, for our research real transaction data have been used.

1.2.1 Adyen

Adyen is a reliable and one of the world's largest PSPs headquartered in Amsterdam, The Netherlands. The company provides a single platform to accept payments anywhere in the world. Currently, this platform is processing payments for over 5,000 merchants, many of which having a transaction volume over 500,000 during a month. Adyen's platform is based on a robust technology which is maintained in-house, while there are frequent release cycles, namely on a monthly basis. The Figure below depicts the value chain of Adyen and how it facilitates customers by replacing multiple legacy providers used for payments processing.

Figure 1.2: The value chain of Adyen compared to the traditional value chain in payments.



The company is an omni-channel, which means that it is specialized in processing online, mobile

and Point of Sale (POS), accepting more than 250 different payment types. Payment types might include Visa Pay, Maestro, Apple Pay or iDeal (currently popular in The Netherlands).

At the same time Adyen offers solutions for mitigating risk and fraud in transactions. The risk management solutions offered by the company consist mainly of preventing chargebacks, i.e. the reversal of a payment by the bank, after the product or service has been delivered to the customer. Part of the risk management system is the scoring engine, where each transaction is assessed by assigning it a score. The logic behind this is as follows: There are several risk rules with a score associated to them. If a risk rule is triggered for a specific transaction, the respective score is added to the aggregate score for this transaction. In case the aggregate score for the transaction is 100 or higher, the payment is automatically rejected due to high risk reason and thus not sent to the bank. At the time this thesis is written, 663 of Adyen's customers use the score-based risk system. Accordingly, the company has access to a vast quantity of information related to transactions and risk. More details about the data used for the research can be found in chapter 3 and chapter 5.

1.3 Research Questions

The will for a successful solution to a problem, requires formulating the research question we aim to answer. According to Enserink et al. (2010), incorrect formulation of the problem will more likely lead to designing the wrong solution and thus failing to alleviate it. Following the description of the research problem, the main research question is formulated as follows:

RQ: *Why are merchants reluctant to adopt profit-maximizing risk management settings that are suggested by an anti-fraud tool that analyzes their transaction data and how can developers of such tools increase their acceptance?*

In order to answer the main research question, the following three sub-questions would also have to be investigated. It should be noted at this point that the research is based on a case study, by exploring a specific anti-fraud tool; depending on the results, generalizations might be made accordingly. The first two subquestions will be explored in order to answer the first part of the main research question, while the last subquestion will be used to answer the second part of the research question. Below, the three subquestions are formulated with a short explanation.

SQ1: What does the historical data indicate about the usage of the Risk Calculator and the risk settings adopted by merchants?

The answer to this question entails descriptive exploration of usage data, by means of descriptive statistics and individual test hypothesis. In this subquestion, we investigate what type of merchants use the Risk Calculator and how often. By the term "type" we mean in which industry they belong to, how large are they and what is the level of their chargebacks and refusals. Moreover, we are interested in knowing how these merchants interact with the tool's suggestions. Apart from that, we are interested in comparing them with non-users and check their common characteristics. This question will allow us to classify merchants in terms of risk attitude and plausibly find patterns indicating whether the tool is preferred by users with certain characteristics.

SQ2: How do merchants and account managers explain their engagement with the tool and how do they choose their risk profile?

In order to answer this question, we follow a qualitative research by means of semi-structured interviews. The interviews were conducted both with merchants who use and do not use the tool, as well as with account managers and developers within Adyen. The purpose of this subquestion is to help us delve deeper in the underlying opinions of users about the tool and about risk management in general. Since human behavior is unpredictable, exploring solely data might not be enough to come to robust conclusions. Thus, talking to merchants, account managers as well as developers can assist us to approach the problem from a more holistic view.

SQ3: *"By looking at broader patterns, which factors can we identify that are indicative of merchants' engagement with risk management?"*

Finally, the last subquestion is explanatory in nature and is being answered in terms of multivariate regression analysis. Through this empirical research, we look at factors that explain merchants' engagement with risk management, after identifying the subset of merchants that is of interest to us. Every tool designed to solve a problem, is meant to be used by a group of users that suffer from this problem. Hence, through descriptive analysis we first explore which is the subset of merchants that the tool adds more value, so that companies interested in pro-actively engaging with customers can benefit too. Afterwards, we determine the factors that drive the changes in this group's risk settings through the regression analysis. By determining this as the last step, we can then compare whether the findings are in line with the findings of the previous subquestions.

As noticed from the above description, the research of the thesis entails three distinct parts, since every subquestion is based on a completely different approach. Nonetheless, these approaches

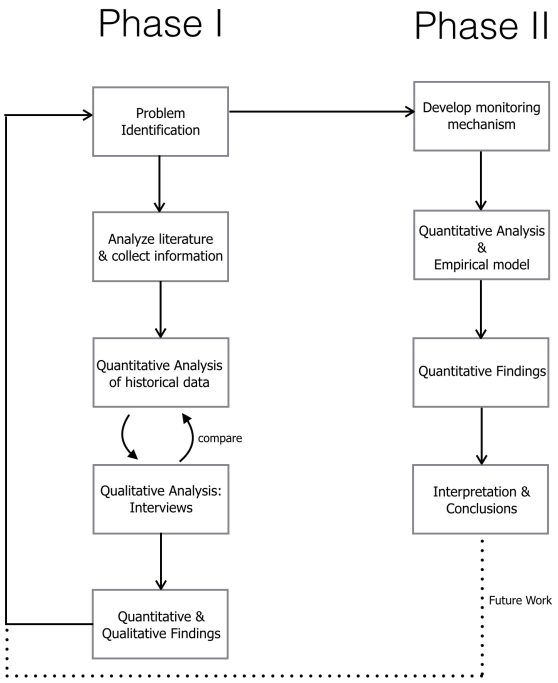
together help us approach the problem from a holistic view and infer conclusions about not only merchants' engagement with the specific tool, but also with risk management in general.

1.4 Research Methodology Overview

The research in this thesis is conducted by using a combination of *quantitative* and *qualitative* approach in order to provide an answer to the research questions. The use of both approaches is expected to give a more holistic answer to the identified problem. More specifically, the qualitative research helps to delve deeper into the problem by identifying motivations and underlying reasons and opinions. On the other hand, the quantitative research is used to quantify attitudes and generalize results.

Each of the research questions introduced above, follows a particular methodology which is being described in detail in the subsequent Chapters of the thesis. The overview of the overall methodology is illustrated in Figure 1.3.

Figure 1.3: The research flow followed in the thesis.



As can be seen from the Figure, there are two phases conducted. In the first phase, after identifying the initial problem we moved on analyzing the literature. The literature data, known as "secondary sources" include academic papers, proceedings of conferences, books and reports as main sources. To find these secondary sources, the Internet as well as scientific databases were used.

In relation to the quantitative data, access to datasets with transactions is granted by Adyen. The technique for working with these data is cross-sectional analysis, i.e. an observational analysis of data collected from a specific subset, at one specific point in time. More specifically, the first part of the quantitative analysis is exploratory and descriptive in nature; it aims to reveal usage patterns related to Risk Calculator through descriptive statistics and individual hypothesis testing. This is done by comparing the merchants that seem to be using the tool, versus those that have never used it.

Later on, we move to the qualitative analysis, where semi-structured interviews conducted with experts are used as a source of primary data. Particularly, we interviewed 11 merchants that are both users and non-users of the tool in order to obtain broad feedback about the Risk Calculator. Moreover, we talked to four account managers, one risk officer and two developers in Adyen in order to compare whether the opinions about the tool converge.

Moving on the second phase of the research, we developed a monitoring mechanism which actually stores the data calculated by the tool to a database and from this point on we continued with the quantitative analysis of the newly obtained data. This was deemed as a necessary step, since the available data in the first phase were insufficient to draw conclusions from.

The second part of the quantitative analysis is explanatory and is being answered in terms of statistics and multivariate regression analysis. As such, we first define the group of merchants for which the tool adds more value and explore their behavior in terms of frequency in risk changes. Afterwards, we build two regression models; one for predicting what drives merchants engagement with risk management, and one for predicting which factors influence the direction of scores for the merchants who make changes.

For the quantitative analysis, scripts in *SQL* were used in order to retrieve the appropriate data, while the main analysis was conducted by using the software tool *R Studio*. On the other hand, for the analysis of the qualitative data, we transcribed the interviews and used the software tool *Atlas.ti* in order to identify patterns through text coding. Lastly, the monitoring mechanism was developed in the programming language *Java*.

1.5 Document Structure

In this Section the overall structure of the thesis document is being presented. The document is partitioned in six Chapters, each of them including findings related to the main research question. More specifically, the rest of the document is organized as follows:

Problem Investigation (Chapter 2). The next Chapter focuses on understanding the theoretical background behind the problem, the involved stakeholders and the current literature review regarding the state-of-the-art solutions. The theories that are being used fall in within the following categories: Data Science, Machine Learning and Behavioral Economics.

Analysis of Historical Data (Chapter 3). Chapter 3 describes the hypothesis to be tested through the initial data exploration, by looking for patterns in historical data. The aim is to get a feeling on the common characteristics as well as the differences between the users of the Risk Calculator tool and the non-users.

Findings on Interviews (Chapter 4). Chapter 4 contains the results of the interviews conducted with merchants, account managers and developers. The results produced are based on text coding and aim at giving insights about people's incentives to engage with risk management, as well as at getting diverse feedback about the tool.

Findings on Target Group and Empirical Model (Chapter 5). Chapter 5 identifies the merchants for which the tool adds more value, and explores their risk activity by building an empirical model. This Chapter introduces the hypotheses that indicate which variables are expected to influence merchants' risk activity and drives conclusions on which are eventually the predictors.

Conclusions and Discussion (Chapter 6). The last part provides a summary of the research by reflecting on the results, concluding, discussing the limitations and suggesting future work.

Chapter 2

Background Information

In this Chapter we aim to introduce all the important concepts related to the research conducted in this thesis. The approach that is followed in order to provide an answer to the main research question is multidisciplinary and borrows notions of various theories. First, we begin with an introduction to the Payments Industry and we identify the main stakeholders involved in the problem under exploration. Then we move to E-commerce Fraud and to Machine Learning theories which can explain the contemporary solutions offered to the financial fraud problem. Finally, the theory on Economics of Cyber Security and Behavioral Economics presented in this Chapter will help us understand the basic concepts related to people's incentives in engaging with risk management and using anti-fraud solutions.

2.1 The Payments Industry

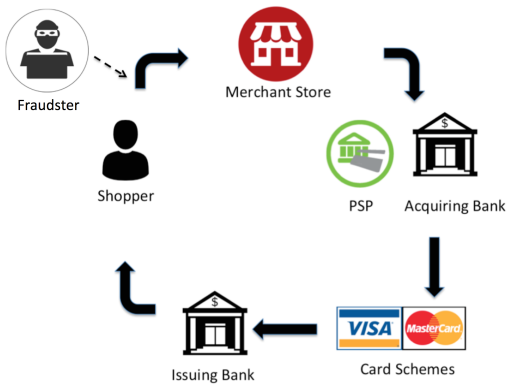
Crucial for understanding the topic of the thesis is the comprehension of how the payment industry works. From the point of view of a shopper, making an online purchase might seem a simple process. This is due to the fact that shoppers usually deal solely with the issuing bank and the merchant, hence they regard the whole underlying process of the transaction as a "black box" (De Gennaro, 2006), i.e. ignoring all the other participants. Below, we attempt to explain this black box that surrounds the payments industry by firstly introducing the involved parties and subsequently explaining the process, as well as the complications arising by the existence fraudsters.

2.1.1 Stakeholders in the Payment Industry

As in every real-life context, the Payments Industry represents a multi-actor setting where various actors with different interests are involved. According to Enserink et al. (2010), an *actor* can be defined as a person or organisation which is interested in a problem and can influence a decision that may be taken with regards to this problem. As reported by the authors, actors' behavior can usually be characterized by four distinct elements: networks, perceptions, values and resources. Networks reflect the interdependencies and social relations that are formed between different actors. This formation usually occurs when actors share the same goals and interests. Furthermore, perceptions indicate actors' general beliefs of how the world works and how other actors behave. Similar to this is the term of values, with the difference that values provide the directions towards actors would like to move, by indicating what they consider "good" or "bad". Finally, resources are the means that actors have to realize their goals. These means might be funds, knowledge or relations with other influential groups.

In the above context, when a simple transaction between a cardholder and a merchant takes place, players such as card acquirers, issuers and processors are participating in the process. Additionally, the card schemes, such as MasterCard and Visa, play an essential role in the completion of the transaction settlement. The most important actors in the payment industry are illustrated below in Figure 2.1.

Figure 2.1: The main stakeholders in the payments industry.



As can be seen in the Figure above, there are seven main actors. Initially, the *shopper* or the cardholder is the person who usually commences the transaction and provides the card details to the merchant in order to make a purchase. The *merchant* represents the retailer or service provider

that accepts card payments from the cardholders. Subsequently we have the *acquiring bank* which represents the merchant's bank, i.e. the financial institution that processes the payment details on merchant's behalf. Furthermore, a *PSP* is connected to multiple acquiring banks, card and payment networks and offers online services for accepting electronic payments. The *card associations* in their turn are a network of financial institutions that license payment cards. Finally, the *issuing bank* is the one that offers the branded payment cards directly to consumers. In order to ensure that the payment systems work efficiently for all the parties that use them, as well as to warrant that there is competition between PSPs and operators there is the *regulator*. Of course, another actor influencing the evenness of the payment process is represented by *fraudsters*. Especially in an era where technology continuously evolves, fraudulent attempts can be executed more easily than ever, posing threats not only to cardholders, but also to merchants. This is the point where the complication arises, making urgent the need for solutions to combat fraud.

Each of the aforementioned players has different interests when it comes to e-commerce fraud and the solutions that can be designed to mitigate this problem. Taking Adyen as the *problem* owner for finding a solution to e-commerce fraud, we can have a look at the other players' interests, interdependencies, goals and resources.

As regards the interdependencies, actors can have either supportive or conflicting interests with the problem owner. If an actor is affected by the problem situation by clear costs or benefits, then the actor is *"dedicated"*. On the contrary, if the actor does not experience any direct costs or benefits then is less likely to influence the problem analysis and thus is deemed as *"non-dedicated"*. Furthermore, actors are *"critical"* when they possess "realization power" or "blocking power", i.e. resources such as knowledge, expertise or funds that are willing to use either for contributing or blocking possible solutions; the critical actors are the ones that the problem owner cannot ignore when designing solutions (Enserink et al., 2010).

Table 2.1: Actors in the payment industry and their interdependencies.

	Dedicated Actors		**Non-dedicated Actors**	
	Critical actors	Non-critical actors	Critical actors	Non-critical actors
Similar/ supportive interests	Merchants Issuing & Acquiring Banks Card Associations	Cardholders	Regulator	
Conflicting interests	Fraudsters			

As can be seen from the above table, merchants, banks and card associations are dedicated actors,

as they incur costs related to fraudulent transactions (Thompson, 2014) and moreover have similar interests with Adyen, since they all wish to eliminate fraud. Crucial to consider however are also the fraudsters, who constantly try to create more sophisticated techniques in order to spread frauds. As regards cardholders, they are an important actor as they represent the victims that without them scams would not exist, nevertheless they are subject to "zero liability" (Consumer Action & Chase, 2009) and they do not possess the means for stopping frauds; they can just be more aware through getting informed. Finally, the Regulator possess a lot of authority and provides faster payments and refund rights to consumers (European Commission, 2016), however when credit card fraud is committed this actor does not experience any direct losses or benefits.

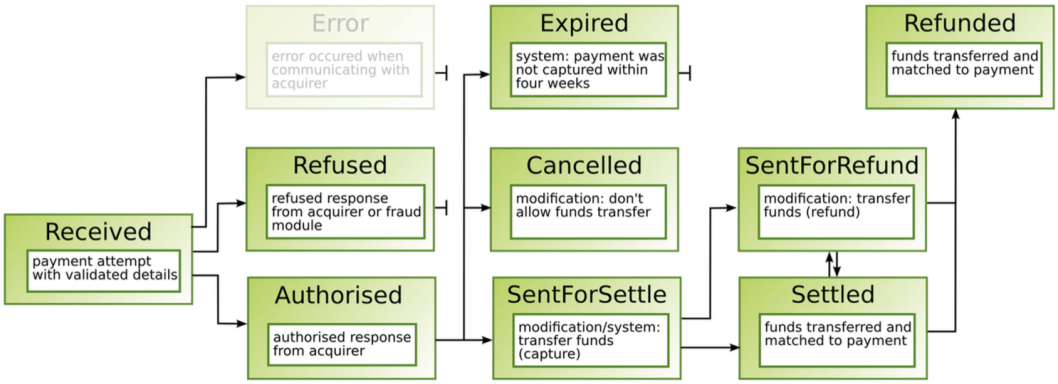
It derives from the preceding analysis that there is a continuous interaction between the different players, while each of them tries to exert influence on possible solutions designed for e-commerce fraud in order to achieve specific goals. This is a typical situation of a multi-actor system, encountered in complex situation settings.

2.1.2 The Payment Process

After introducing the stakeholders in the payments industry, we can now move on describing the actual process behind each transaction. Take a look again at Figure 2.1. The process is initiated when the shopper wishes to make a purchase and thus offers the card details to the merchant's online store. Subsequently, this information is processed by the acquiring bank or the PSP on behalf of the merchant. The next step is to send the information from the acquirer or the PSP to the card association and eventually to the issuing bank for authorization. Once the transaction is validated, the issuing bank deducts their charge and then pays the association network. Afterwards, the card association pays the acquirer processors on acquirers behalf, again after deducting their charge. In this way, the merchant's account is credited for the transaction amount by the processor. The issuer bills the shopper for the purchase and finally the shopper settles the bill.

However, the process might not always result in a successful payment. Depending on different factors, there are several states that a transaction may take. For instance, the acquirer may refuse to send the transaction information to the card associations due to fraud reasons. Regardless the payment method used, each payment goes through a basic sequence of statuses, as can be seen below in Figure 2.2.

Figure 2.2: The sequence of statuses in Adyen that each payment goes through (Adyen B.V, 2015).



The first status that can be seen in the picture is *"received"* and all payments get this status as their initial state; it actually denotes a validated payment attempt.

In the next step the payment will change status from "received" to either *"authorised"*, *"refused"*, or *"error"*. In case the payment is "authorised", this means that it has been approved by the financial institution and hence the delivery of goods or services may proceed. On the contrary, if the payment is "refused" this means that it has been declined by the financial institution or by the PSP due to fraud; the "refused" status is a final state, since the payment cannot proceed once declared refused. There is also the case where an error might occur in communicating with the financial institution, thus the "error" status is assigned to the payment; this is a final state too.

The payment can go to the next step only in the case it was "authorised". From this point there are again three states that it might reach: *"sent for settle"*, *"cancelled"* or *"expired"*. When the payment receives the "sent for settle" status, a request for transferring funds to the financial institution has been sent. This is usually known as a capture request. After the payment reaches this status it is not possible to cancel it. In case the shopper or the merchant wishes to cancel a payment that was initially authorised, then a request for blocking funds transfer for this authorised payment is sent. As such, the payment status is then "cancelled"; note at this point that it is only possible to cancel a payment when the "sent for settle" status has not been reached. On the other hand, when an authorised payment has not been cancelled or captured for 4 weeks, it reaches the "expired" status where funds transfer is no longer possible. The "cancelled" and "expired" states are final statuses.

Once again, when the payment reaches the status "sent for settle", it can proceed to the next

step. From this point on, it can be either *"settled"* or *"send for refund"*. In case it is "settled", the financial institution transfers the funds to the PSP. In the other case, a request to reimburse the shopper is sent to the financial institution; the refunding however can only occur for payments that have passed the "sent for settle" state. Lastly, the *"refunded"* state is reached when the financial institution has completed the reimbursement to the shopper.

2.1.3 The Notion of Chargebacks and False Positives

Delving deeper into the payments world, there are two terms that appear to be important especially from the merchants' point of view. These are the chargebacks and the false positives. As already mentioned, risk management solutions usually aim at achieving a balance between these two Key Performance Indicators (KPIs), since they both affect merchants revenues.

Chargebacks

So, what is a chargeback? According to Oxford University Press (2016) *chargeback* is *"a demand by a credit card provider for a merchant to compensate the loss on a fraudulent or disputed transaction"*. In other words, chargeback is a transaction that was initially labeled as authorized, however after the product or service was delivered, the customer disputed the transaction either due to dissatisfaction or due to fraud. Cases of dissatisfaction may include long delivery time or specifications of the product not matching with the online description. On the other hand, cardholders might find out that an unauthorized transaction was made on their account, hence they were victims of scammers. In this case, a chargeback helps the cardholder-victim to retrieve the money used from their account.

In order to file a chargeback, shoppers have to contact the issuing bank of the card they made the payment with. After providing the reasoning for the chargeback, the issuer examines the claim and if it is valid the money are directly deducted from the merchant's account, while a notification is sent to the merchant's acquirer. Beginning with the latest the day of delivery of the product or service, the cardholder has up to 4 months in order to claim a chargeback, depending on the card association (De Gennaro, 2006). Initially, the dispute is always in favor of the customer and the chargeback amount is deducted from the merchant's account; nevertheless, the merchant can recover the funds by providing evidence that the cardholder's claim is wrong. In most cases, the merchant is the party who finally bears the costs of chargebacks (Wolters, 2012); the acquirer

might be liable to pay the costs in case, for example, the merchant goes bankrupt between the time of the sale and the time the chargeback is debited in the account (De Gennaro, 2006).

It derives from the above that chargebacks have quite some implications for merchants and the costs that are incurred by them are significant. The costs related to chargebacks are summarized in Table 2.2. On one hand, the merchant carries not only the actual cost of the chargeback, but also the costs of buying and subsequently sending the product to the shopper, in case we are talking about physical goods (Bhatla et al., 2003). Moreover, if the shopper files the chargeback and additionally keeps the product, the merchant loses the revenue and any potential profit on it. It should also be mentioned that each time a chargeback is filed, the merchant must pay administrative costs related to the process that might range from $20 to $100 per transaction. Even in case the chargeback is later recalled by the shopper, the merchant does not recover this fee.

On the other hand, when the level of chargebacks exceeds a certain threshold, the card associations impose fees to the merchants. Let's take for example MasterCard and Visa programs. MasterCard with the so-called "Excessive Chargeback Program" (ECP) sets the threshold to 1.5% or more chargebacks-to-sales count ratio and additionally at least 100 chargebacks in 1 calendar month. Merchants that join the ECP are required to pay $25 per chargeback above the defined threshold. Similarly, Visa's Chargeback Monitoring Program (CMP) includes two categories: standard threshold program and excessive threshold program. For the first one, the chargeback-to-sale ratio is set to be 1% and 100 chargebacks. In case the merchant does not decrease this level till the fourth consecutive month, the fee from the fifth till the seventh month would be 45 euros, while for the eighth month onwards 85 euros per chargeback. In the latter case, the threshold is 2% and 500 chargebacks in a month, with a fee 85 euros per chargeback from first to sixth month and from the seventh month onwards the merchant would also have to pay 21.750 euros as a review fee (Adyen. B.V, 2016).

Table 2.2: Costs associated to chargebacks incurred by merchants.

Direct Costs	Chargeback value
	Administrative costs
	Fines to card schemes (in case of excessive chargebacks)
Indirect Costs	Product value
	Delivery costs
	Potential profit on product

False Positives

Although chargebacks pose a significant threat to merchants' revenues and hence is essential to keep their level low, revenues are also influenced by the number of legitimate transactions they allow. In some cases, in order to avoid fraudulent attempts by scammers, merchants might refuse to accept genuine transactions. This type of situation is known in the payments industry as *"false positives"*. More precisely a false positive refers to a transaction which is legitimate, but the risk management system that scans the payments perceives it as fraud; hence the transaction is blocked while it shouldn't have and the merchant loses potential revenues. Therefore, when trying to mitigate risks related to transactions merchants should consider both chargebacks and false positives as two variables with potential impact on their profit. Previous research (Wolters, 2012) in Adyen has proved that there should be a balance between false positives and fraud in order to optimize revenues.

As such, a risk management system which is set up too aggressively, might successfully block fraudulent attempts however it may result in many false positives and thus lost revenue. Similarly, a risk system which is loosely set up might reduce the number of false positives, but be unable to detect fraud and thus lead to lost revenue. This trade-off between the two KPIs arises complexity for risk managers who always try to reduce fraud without affecting legitimate shoppers.

2.2 E-commerce Fraud

The ease of misusing credit cards to commit fraud, as well as the anonymity that Internet offers has made it appealing to malevolent users to exploit unaware cardholders. Frauds related to e-commerce show an increasing trend, with fraudulent methods becoming more sophisticated and posing great threats to businesses (Bhatla et al., 2003). There are mainly two situations where e-commerce fraud can occur: card-present (CP) fraud and card-not-present (CNP) fraud. These two contexts are explained in more detail below. Additionally, we discuss some of the factors that are known to influence fraudsters decision in commiting fraud.

2.2.1 Card-Present Fraud

The CP fraud situation involves a fraudster physically stealing the credit or debit card of the cardholder/victim and then presenting it to the point-of-sale (POS) terminal. In order to complete

the transaction, the physical presence of the card to the POS is required. This fact makes it necessary for fraudsters to steal a card. The most obvious way to steal requires a direct contact with the cardholder, by for example stealing a wallet. Since this physical presence involves relatively high risk of getting caught, several other ways have been invented by scammers to make theft of card details easier and present them to merchants. The most popular methods are described below.

Counterfeit cards. A counterfeit is a fake card with real information details stolen from victims. The way the information is stolen is usually through copying it or "skimming" it. The term skimming refers to cloning the information through devices that scan and store the card's details from the magnetic stripe, during the transaction. This is usually done through electronic devices when the cardholder is waiting for the transaction to be validated in the terminal (Bhatla et al., 2003). In order to prevent this fraudulent technique, the schemes have introduced the "Chip-and-Pin" cards, known as EMV (Europay, MasterCard, Visa) which are explained in more detail later in this Chapter.

Mail Intercept Fraud. Another way to steal credit or debit card information is through intercepting either the mail containing the physical card from the bank before the cardholder receives it, or the email containing the passwords in electronic format. The precautions that the issuing banks have taken to prevent physical mail interception include sending the card, the PIN and the activation code in different mails.

White Plastic. This is a card-sized piece of plastic that a fraudster creates and looks like a hotel-room card. The white plastic contains legitimate stripe data that the fraudster has encoded. In this way it is easy to use them in (POS) terminals that do not require validation or verification (Bhatla et al., 2003).

2.2.2 Card-Not-Present Fraud

In the CNP environment, the card is not required to be present during the transaction, thus making it easier for fraudsters to collect the data. Moreover, criminals have to collect less information in order to commit CNP fraud, as only the credit card number and the security code are required for the transaction. The ways that the card information may be stolen can vary. On one hand, the conventional stealing or the skimming practices can still be used. On the other hand, there are some new approaches that are used by fraudsters, as described below.

Phising. This approach of "social engineering" is typically done by criminals who indicate to be trustworthy people or institutes, requesting the cardholders to provide their card details. The most common ways to perform phising is either through phone or through emails. Typically, the criminal asks the cardholder to "confirm" personal information usually for some made-up reason, thus trying to deceive the legitimate cardholder to provide as much details as possible.

Trial-and-error on the card number. Apart from directly stealing information, a fraudster can also use a "guessing" approach to generate credit cards. This is based on the fact that credit card numbers are not randomly generated, but they are based on some mathematical rules following Bank Identification Number (BIN) ranges. A BIN consists the first four to six digits that appear on a credit card and it uniquely identifies the institution that issued the card. As such, through the BIN the transactions can be matched to the issuer of the charge card. Beside the credit cards, the BIN system applies also to debit cards, prepaid cards and gift cards. Hence, it is possible for criminals to use the trial-and-error method in order to guess card numbers and then try to verify them. The verification process usually takes place through the *carding* method. In this method the fraudster verifies whether the card number is correct by attempting to do small amount transactions on web shops. In case the card number is accepted, then the fraudster uses it for large amount purchases in web shops.

2.2.3 Trends in e-Commerce Fraud

We can assume that fraudsters act as rational human beings and hence their actions are driven by profit-maximization incentives. According to recent studies, the trends regarding whether financial fraud is committed or not are the following.

On one hand, the characteristics of the transactions per se play an important role. As such, it was found that higher transaction values, with a peak at 100-150 euros, are more exposed to fraud (Ingenico Payment Services, 2015). Nonetheless, it has also been noticed that fraudsters prefer to test credit cards on small amounts and later on, after validating the card, move on making large amount purchases.

Moreover, studies have shown that in some countries, such as Mexico, Brazil, China, US, UK, Russia and France, the fraud rates are higher than the average global (Ingenico Payment Services, 2015). Additionally, the sector that the merchant belongs to has different exposure on fraud; merchants with high margins, such as Information Services, where the costs of chargebacks can be easily absorbed tend to have greater willingness to accept an increase in chargebacks in

exchanges of conversion (Ingenico Payment Services, 2015). Lastly, the size of the merchants can play a role in the engagement with risk management decisions. For example, larger merchants seem to prefer using anti-fraud solutions instead of opting for 3DS in order to avoid customer friction (Ingenico Payment Services, 2015).

Finally, the existence of appropriate anti-fraud tools offered by industry seems not to be greatly accepted by merchants. Accordingly, the acceptance of these solutions is determined by the awareness of merchants; from a merchant's perspective it seems to be a lack of understanding how fraud happens, whether there are any available solutions and ultimately how efficient they are (Cognizant, 2016).

2.3 Countermeasures for Preventing E-commerce Fraud

Since fraudsters are human beings, they adapt quickly and always try to find new methods for committing fraud. As such, the financial institutions have taken several countermeasures in order to combat some of the most common techniques that fraudsters use. These countermeasures have to do with the product as well as the service delivered to the merchants and they are described in detail below.

EMV Protocol. As already mentioned, EMV is a chip placed on the bank card that according to the card schemes is much more secure than the magnetic stripe. The idea of the EMV protocol is that the cardholder has to enter each time a PIN code during the transaction. This code is not on the card, thus stealing or skimming the card will not lead to loss of the PIN. The popularity of EMV is reflected in the fact that during 2015 35.8% of all transactions globally were EMV (EMVCo, 2016). Nevertheless, EMV is prone to some other weaknesses that make it still exploitable. Some examples are the *man-in-the-middle attack*, as well as the *relay attack*. In the first case, the fraudster intercepts the message for the PIN authentication during the EMV transaction process. More specifically, when the shopper enters the PIN code on the terminal, a message is sent to the card which replies with PIN OK/ NOT OK. By intercepting this message, the criminal can always reply PIN OK in any entered code, hence deceiving all the components that the transaction is valid. As regards the relay attack, it can be described as a coordinated effort between two criminals. Criminal A is selling a good and the shopper pays on the POS terminal, while criminal B is buying an expensive good on another shop. The POS terminal, however, of criminal A is actually proxing the card's retrieved data to the card of criminal B. In this way the cardholder-victim is actually paying for an expensive merchandise without being aware on the

fraud.

Card Security Code (CSC). The CSC is a code, usually 3 or 4 digits, printed on the bank card. This code is also known under the names "card verification value" (CVV), "card verification value code", "card verification code" (CVC), "verification code" (V-code or V code) or "card code verification", depending on the issuing bank. The strength of this code relies on the absence of electronic copies; it is only allowed for the issuing bank to have a copy of this code for verification issues.

3D Secure (3DS). In an attempt to mitigate CNP fraud, Visa introduced in 2001 the 3DS protocol. According to Bouch (2011), the aim of 3DS is to authenticate the cardholder during the transaction process. This is achieved when the cardholder is enrolled in an issuer-managed service, either during making a purchase or in advance, where he/she will be asked to choose a password. During the online purchase, the cardholder is asked to provide the 3DS password in order to prove that is in reality the legitimate cardholder. In order to understand deeper this protocol, it is essential to introduce the '3-Ds' that the acronym refers to:

1. Issuer Domain: This is the domain where the issuer manages the enrollment of the cardholder into the scheme, as well as the authentication during the purchase.

2. Acquirer Domain: This is the domain where the acquiring bank and the merchants are included. The acquiring bank ensures that the merchants are operating under the same scheme.

3. Interoperability Domain: This is the domain that describes the "connection" between the acquirers and the issuers, i.e. Visa and MasterCard schemes.

Opting for 3DS, the merchant can benefit from mitigating fraud and chargebacks, and accordingly reduce all the associated costs related to it. It should be noted, however, that 3DS cannot completely eliminate chargebacks and fraud, hence the merchant should continue to use anti-fraud systems. Another advantage of 3DS is the chargeback liability shift in which it is entitled. More specifically, the use of 3DS means that in case of disputed transactions the liability passes from the merchant to the card issuer, even if the issuer is not a participating member of the scheme. Following the above one could wonder why not all merchants use 3DS for their transactions? The answer here is quite simple. It seems that although 3DS prevents fraud and shifts liability to the issuer, it additionally decreases conversion (C. Nicholls, 2013).The reason is due to the fact that 3DS adds an extra step to the customer's checkout experience, asking for additional passwords and this is proven to be frustrating for the shopper who just wants to complete the payment as soon as possible. Thus, merchants when faced with this dilemma they prefer to take the risk

that some of the transactions might be fraudulent, rather than certainly lose some percentage of legitimate customers due to 3DS.

2.4 Machine Learning on Fraud Prevention

Since the problem of fraud was the new reality people had to face, a bunch of solutions aimed at detecting fraudulent attempts started being developed. Machine Learning (ML) has greatly contributed in this goal by offering a handful of applications. According to Han, Pei, and Kamber (2011), ML refers to making a computer learn and take decisions based on data. There are mainly four techniques used in ML, namely supervised learning, unsupervised learning, semi-supervised learning and active learning. the main difference between them is how the algorithm handles and interprets the data given as an input.

When the aim is to detect fraud in transactions, a very common solution includes developing heuristics around features that might be indicative of fraud (Guha, Manjunath, & Palepu, 2015). Two widespread ways for taking a decision about fraud is either having some rules that define whether a transaction needs to be send for investigation, or having a scoring system where an aggregation of the scores along with the claim value will determine whether the transaction needs to be investigated.

One of the most popular ways of combating fraud using ML is based on *Classification algorithms*. In essence, these algorithms focus on what is called "pattern recognition", i.e. identifying a pattern by analyzing raw data. This is usually done through the use of a training dataset, which enables the algorithm to determine in which pre-defined category a new observation belongs to. Some of the most popular classification algorithms include Decision Trees, Neural Networks, Naive Bayes, K-nearest Neighbor, Random Forests and many more.

In the case of financial fraud, the classification algorithm is based on raw transactional data. According to Bahnsen et al. (2016) most of the anti-fraud systems based on classification algorithms consider as features the place a transaction took place, the time, as well as the amount. Other approaches use more complex features, such as the consumer spending behavior through an aggregation strategy. The aggregation is actually a grouping of transactions by account, merchant, country etc. (Whitrow, Hand, Juszczak, Weston, & Adams, 2009).

In the study conducted by Dal Pozzolo, Caelen, Le Borgne, Waterschoot, and Bontempi (2014), the authors identify as a problem that fraud-detection tools are used reactively, meaning that they

send notifications for potentially risky transactions which later have to be assessed by human beings. This latency led the authors to the designing and testing of fraud-detection systems that use two training classifiers; one on user feedback and one on delayed samples, which are then aggregated.

Since the scientific community has explored in depth how we can benefit from applying ML classification in order to detect fraud in transactions, Table 2.3 below presents an overview of the most recent advances in this particular type of fraud. According to West and Bhattacharya (2016), early detection methods focused primarily on statistical analysis, such as logistic regression and neural networks, while recently the focus has been swifted more towards artificial intelligence methods.

Table 2.3: Summary of Machine Learning methods used for the problem of detecting credit card fraud.

Type of research	Method	Authors
Credit card fraud based on data provided by Europay International	Neural Networks	Maes et al. (2002)
	Bayesian learning	
Credit card fraud based on case study of retailer companies in Chile	Association rules	Sanchez, Vila, Cerda, & Serrano (2009)
Credit card fraud based on case study of bank	Hybrid methods	Duman and Ozcelik (2011)
Credit card fraud based on case study of bank	Decision trees	Sahin et al. (2013)
Credit card fraud of an international credit card operation	Logistic Regression	Bhattacharyya et al. (2011)
Building Adyen's fraud detection tool	3 Nearest Neighbours	Bert Wolters (2012)
	AdaBoost	
	CART	
	C4.5	
	Linear Regression	
	Logistic Regression	
	Naive Bayes	
	Neural Network	
	Random Forest	
	Support Vector Machines	
Credit card fraud based on case study of bank in Warsaw	Support vector machines	Olszewski (2014)
	Random forests	
	Self-organising map	
Credit card fraud based on case study of a Brazilian bank	Artificial immune system	Soltani Halvaiee and Akbari (2014)
Credit card fraud based on case study of a Turkish bank	Discriminant analysis	Mahmoudi & Duman (2015)
Credit card fraud by testing set of confirmed fraudulent transactions	Social network analysis	Van Vlasselaer et al. (2015)

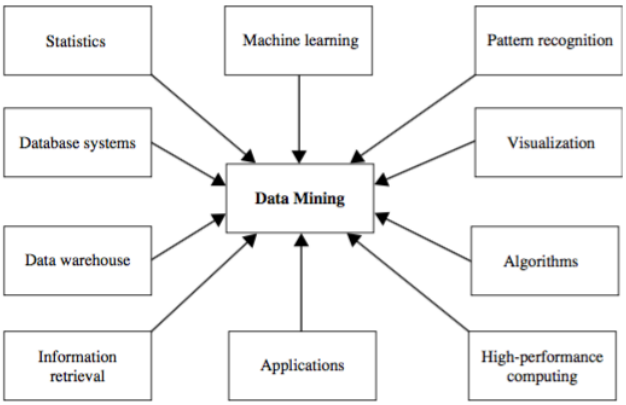
2.5 Data Mining

Since Information Systems became an integral part of everyday life, the amount of data received and stored in computer networks have inconceivably increased. Today, we are usually talking about petabytes of data, i.e. a compute storage unit which equals to one quadrillion bytes (Han

et al., 2011). Harnessing the power of data can lead to knowledge about several phenomena, one of them being detecting fraud in transactions. Literature refers to Data Mining as knowledge discovery from data (KDD), which emerged as part of IT's evolution.

In order to conduct data mining, several techniques can be used that will lead to retrieving knowledge from vast amounts of data. Machine Learning, as introduced above, represents a sub-category, or in other words a means for data mining. As Han et al. (2011) discusses, data mining borrows techniques from various different domains (Figure 2.3).

Figure 2.3: Domains contributing to data mining (Han, Pei, & Kamber, 2011)



Closely related to data mining is the notion of Big Data. The term "Big Data" emerged around a decade ago with the first research attempts being made in 2008, while in 2011 there was noticed a sudden increase in the attention payed to this term that has been growing ever since (Wamba, Akter, Edwards, Chopin, & Gnanzou, 2015; Gandomi & Haider, 2015). According to Kaisler, Armour, Espinosa, and Money (2013) *"big data is the amount of data just beyond technology's capability to store, manage and process efficiently"*. Similar studies offer definitions to big data through the lens of "Vs" (Uddin & Gupta, 2014; Gandomi & Haider, 2015; Jin, Wah, Cheng, & Wang, 2015). Most of the literature discusses big data in terms of 5Vs, namely volume, velocity, variety, veracity and value, however a recent study (Uddin & Gupta, 2014) has added two more dimensions, i.e. validity and volatility.

Volume is the dimension that refers to the size of big data, which as already mentioned currently is measured in petabytes or even exabytes. As the technology advances year by year, the volume of data tends to escalate, hence in the future the definition of big data in terms of size might change.

Variety is another important characteristic that refers to the heterogeneity of data, denoting that big data can be either unstructured, semi-structured or structured. Examples of unstructured data are usually images, audio and video whereas structured data refers to data retrieved from databases. Finally, a well-known example of semi-structured data is the XML language (Gandomi & Haider, 2015).

Velocity measures the generation rate of data, as well as the speed at which it should be analysed (Gandomi & Haider, 2015). The emergence of smartphones and of the Internet of Things, which is connected to the use of sensors have increased the velocity of big data. As regards *veracity*, it is an indicator of the quality of data and the trust of the sources from which the data are derived (Wamba et al., 2015). Indeed, knowing the degree of authenticity can help in the decision making process, since reliable data lead to correct analyses of opportunities.

Moreover, *value* is a dimension that correlates the exploitation of big data to economic benefits (Wamba et al., 2015); through harnessing its power, organizations can have increased revenues as a result from e.g. targeted advertisement.

The aforementioned five dimensions are the most cited in literature related to big data and hence the ones that most authors have agreed upon. Nevertheless, the study of Uddin and Gupta (2014) has contributed by adding two new dimensions. *Validity* refers to the degree at which the data is proper to use for a specific purpose. The very same dataset might be appropriate for one application but inappropriate for another ,thus validity should not be confused with the veracity dimension. Lastly, *volatility* refers to the flexibility of data storage; as long as the dataset is not anymore useful it can easily be destroyed.

After having introduced the concept of Big Data, we now understand why using techniques such as ML are inevitable, when the nature of data makes it impossible to manually investigate and look for patterns. Although logic dictates that data mining tools will probably be welcome by the interested users, contemporary research shows that there are some challenges related to user adoption.

Specifically, Huang, Liu, and Chang (2012) study the intentions of people to use data mining tools (DMTs). The findings showed that "perceived usefulness" and "perceived ease of use" are the two factors mostly influencing the adoption of a DMT. The perceived usefulness reflects whether the potential users think the tool creates benefit for their work, whereas the perceived ease of use indicates how easily users learn to use the tool. Regarding the latter it was found that the more time users spend to understand how the tool works, the less likely they will use it. Other

crucial factors were the quality of the output, the demonstrability of result and the response time.

Another study by Huang, Wu, and Chou (2013) investigates the factors that contribute to the continuance of using DMTs. The authors pinpoint to three characteristics that were found to influence people's willingness to continue the usage of the tools: task-technology fit, expectation–confirmation and habit. The task-technology fit includes notions regarding how individuals use technology to perform their tasks, as well as how they turn inputs to outputs. On the other hand, the expectation-confirmation model reflects customers' satisfaction and post-purchase behavior. Lastly, habit refers to a routinized behavior, targeted to obtain specific goals.

Furthermore, Chien, Kerh, Lin, and Yu (2016) highlight the importance of visual aesthetics and user experience as two features influencing users' perception for adoption. Accordingly, the authors develop a framework based on data-mining which identifies user's background information, such as gender, studies and personal aesthetics level, as influencing factors of user experience. The study concludes that by taking into account the factors influencing user experience, companies can target specific customer segments to proactively engage with, and thus gain a competitive advantage.

Although not ample, the existing literature has identified the challenge of user adoption regarding innovative tools. Moreover, an early study (Asare & Wright, 2004) questions the effectiveness of anti-fraud tools based on risk checklists in particular. This derives from the fact that one of the main findings was that auditors who used tools providing a standard checklist, made lower risk assessment than those without a checklist. The aforementioned observations lead us to the gap that this thesis aims to investigate, by focusing on a specific anti-fraud tool designed by Adyen and introduced in the following Section.

2.6 The case study of Adyen's Risk Calculator

After having presented all the relevant background information to the problem, we move towards exploring the actual risk management tool, developed and offered to merchants by Adyen. This section begins with an introduction to Adyen's risk management system and moves to describing the anti-fraud tool, aiming to help readers familiarize with how it works. This tool represents the stepping- stone of the research, since a case study is developed around it and the results are later generalized.

2.6.1 The Score-based Risk Engine System

Under Section 2.4 it was mentioned that there are two main approaches in risk systems provided by industry; either a score-based system or an attribute-based system. Adyen's risk management solution follows the first approach. More specifically the scoring system is fully embedded and integrated in the payments platform provided to merchants.

In order for the system to work effectively, merchants should provide as many data related to the transaction as possible. This might include name, phone number, IP address and shipping method. Each payment that comes into the system receives a score according to the following logic. There is a risk checklist comprised of 81 checks, each of them associated with a score. To get a better sense of of what the checklist looks like, we can have a look at Table 2.4. If a risk check is triggered for a specific transaction, its score is added to the aggregate score of the transaction. It should be noted at this point that the score might be either negative or positive, ranging from -1000 to +1000. If the fraud score for is 100 or higher, the transaction is refused by Adyen automatically. A score less than 100 is sent for authorization.

Table 2.4: Categories of risk checks in the Adyen System

Referral Checks	Block and trust lists of both good and bad attributes, affecting the risk score based on a known trend on many different shopper attributes.
Consistency Checks	Comparison of two or more transaction data points with each other (e.g. whether shopper's IP and and shipping address are consistent).
Velocity Checks	Setting thresholds in order to control how often a customer can attempt transactions. These rules are intended to identify high-speed frauds.
External Checks	These checks allow merchants to use scores and results from other third-party risk mitigation systems.
ShopperDNA Checks	ShopperDNA uses algorithms in order to track users across networks, devices and online identities by creating a single profile. ShopperDNA rules are risk checks that use this concept of a shopper to track behavior and velocities.

Depending on the merchant, these scores can be configured manually to form a risk strategy against fraud. Merchants might also establish some thresholds for manually reviewing transactions. In general, it is noticed that merchants differentiate their risk configurations by isolating low-risk from high-risk transactions.

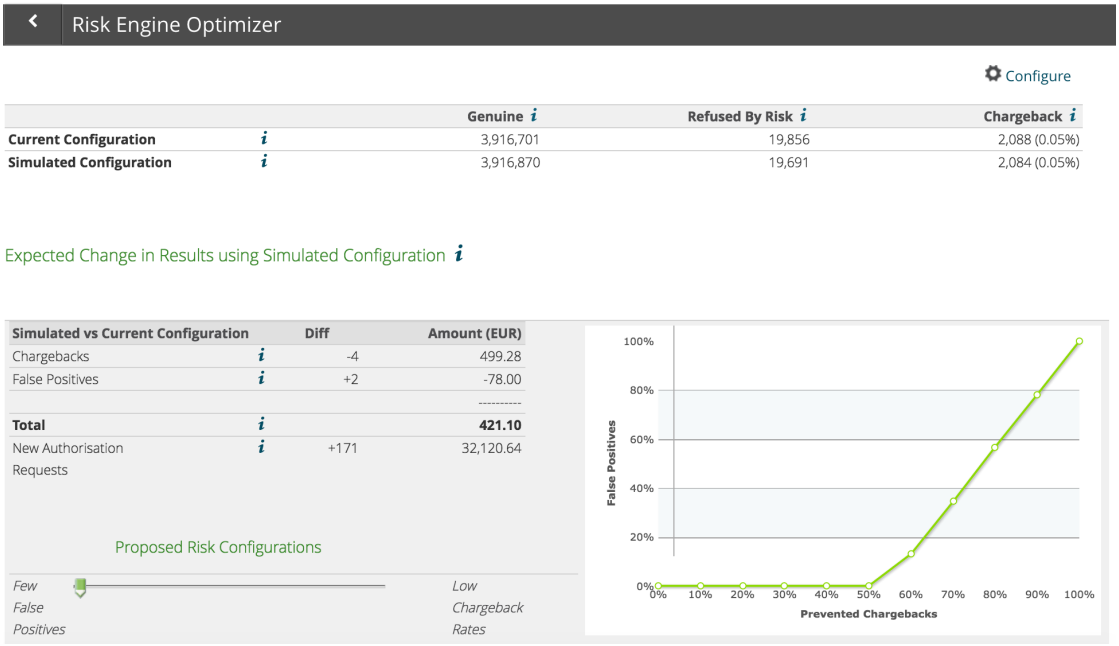
2.6.2 Introduction to Adyen's Risk Engine Optimizer

Part of the risk system described above is the Risk Engine Optimizer (REO), or simply the Risk Calculator tool. Risk managers within companies often take the responsibility of making changes to their risk or fraud settings, which sometimes comes at the cost of introducing several false positives or negatives into the merchants' profile. In order to make this process less stressful, the Risk Calculator, which based on a combination of the merchant's transaction data and statistical modelling, suggest changes in the risk settings by highlighting opportunities to stop fraud and on the same time to maintain conversion. In other words, the Risk Calculator makes suggestions on how to change the risk scores in order to maximize revenues, by either preventing chargebacks or false positives.

The strength of the tool is that it allows merchants to experiment with their risk settings in a "what-if" manner. In this way, they can understand the impact that their changes will have on their chargeback and refusal rates. In order to do so, the tool scans historical data of transactions up to the last 9 months and checks all the authorized transactions. From these are excluded the transactions that occurred during the last 3 months, since they are not yet considered to have a final state. From the authorized transactions it detects which of them resulted in chargebacks or refusals with the current risk configuration that the merchant uses. Then, by changing the risk configurations it predicts how many of these chargebacks or false positives would have been prevented, if using those different risk settings. The calculations to find the optimized configurations are based on Linear Optimization.

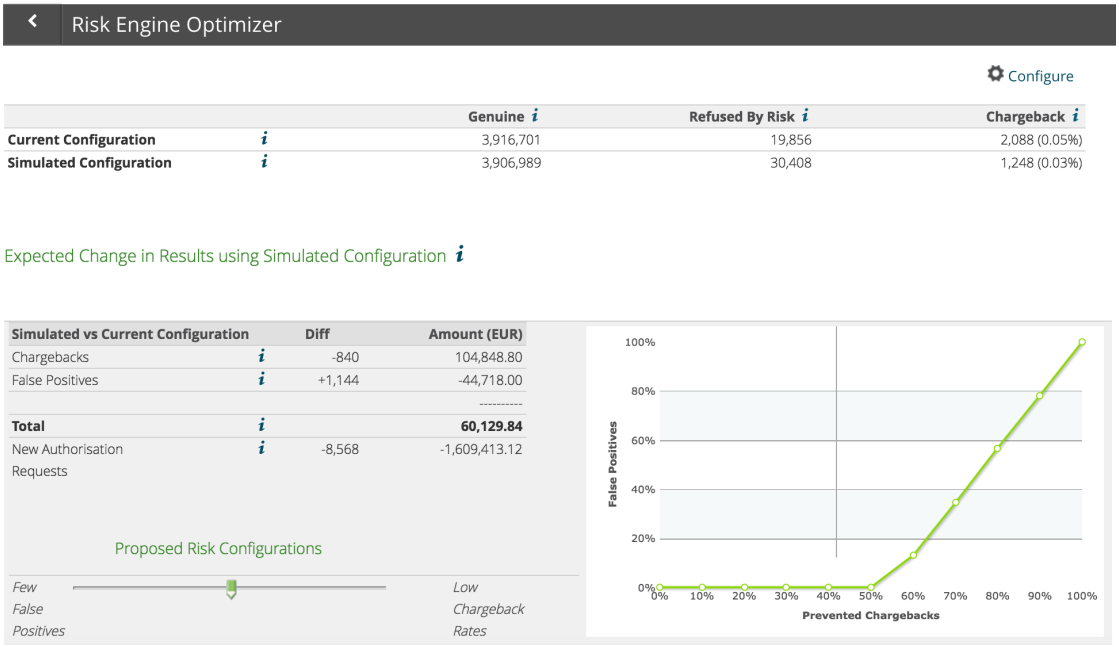
There are practically two ways to use the tool. The first one is more automated and suggests the optimal settings through the use of a slider button, which allows merchants to select whether they wish fewer false positives or fewer chargebacks. The second way is to manually change the risk scores and then calculate the expected effect that these changes would have on the revenues. A better illustration is shown through the following sample case in figures 2.4-2.6.

Figure 2.4: The REO set to minimize false positives.



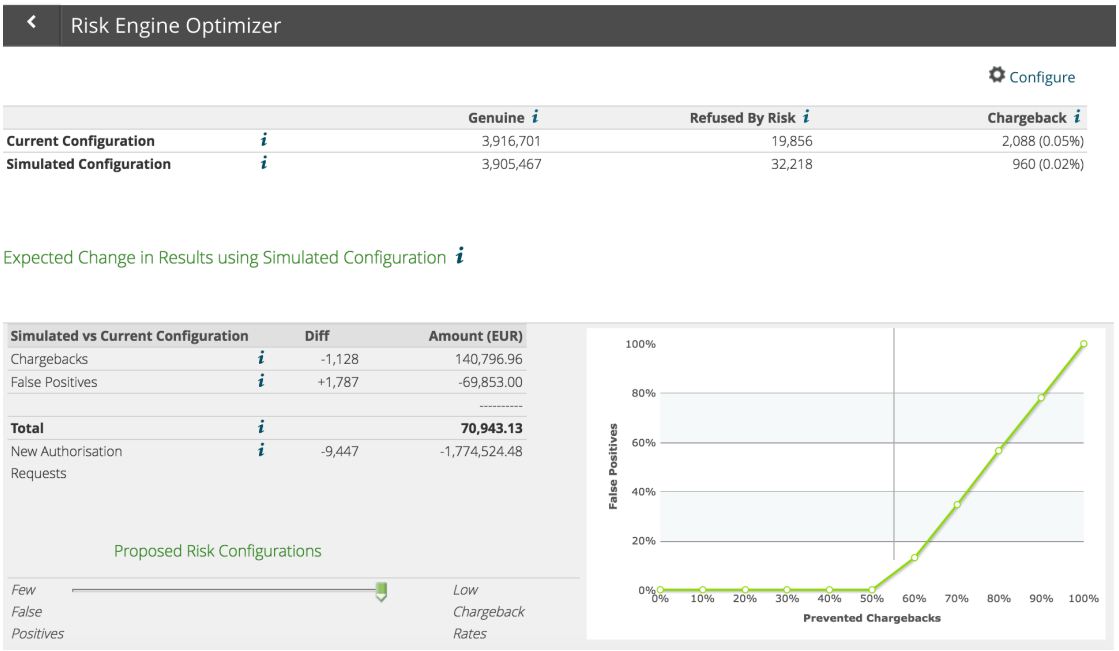
Beginning with explaining the slider button, in figure 2.4 can be seen that the merchant's preference is to minimize false positives, which means to accept as many transactions as possible. This can be the preferred choice if the merchant is more interested in conversion, i.e. turning the visitors to actual shoppers. The algorithm in this case identified that the minimum false positives that the merchant should accept is two more additional, leading to 78 euros loss. However, for this configuration the chargebacks are going to be reduced by 4, leading to a gain of 499.28 euros. Hence, there is actually a trade-off between the false positives and the chargebacks that a merchant can accept in order to minimize risks. Please note that in this case the number of chargebacks is high, thus even though the preferred configuration is "few false positives", the tool is still minimizing chargebacks in order to prevent the merchant from paying fines to the card schemes. The first row in the Figure, displays the genuine transactions, the refused by risk and the chargebacks that the merchant has with the current risk settings. The second row displays the "simulated configuration" i.e. how these three states (genuine, refused, chargebacks) will be modified if the merchant follows the tool's advice. The reader should also note that the system, as designed, cannot predict whether a false positive would have become a chargeback or not if it would have gone through.

Figure 2.5: The REO set to a middle solution between false positives and chargebacks.



Similarly, in figure 2.5, the merchant's preferred solution is somewhere in the middle of minimizing false positives or chargebacks. Thus the tool's suggested solution is to accept 1,144 more false positives and to decline 840 chargebacks, resulting in 60,129.84 euros gain. Finally, in figure 2.6 the preferred configuration is "low chargeback rate". Accordingly, the tool determines that the minimum number of chargebacks that can be achieved is -1,128 but at the same time the new false positives that would be introduced to the system are +1,787, resulting in a net increase of 70,943.13 euros.

Figure 2.6: The REO set to minimize chargebacks.

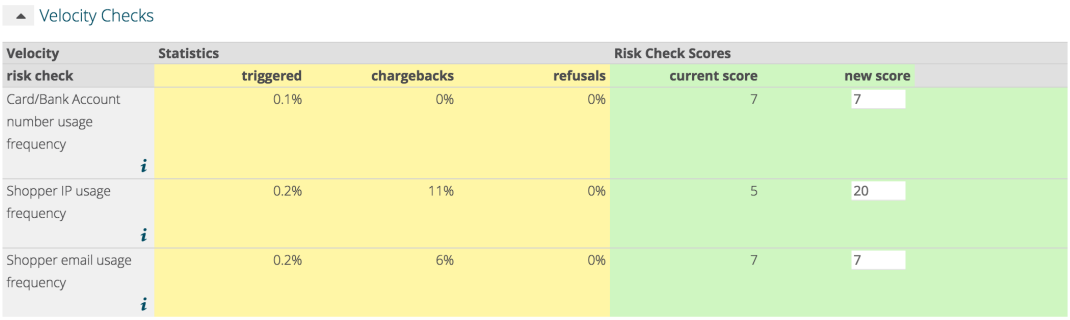


The next step after looking at the tool's predictions is to check the new scores that REO suggests for the risk settings in order to achieve the respective prediction in savings. For example, we can see in figure 2.7 that the tool suggests to change the score for the check "Shopper IP usage frequency" from 5 (that is currently set to) to 20. The user has to check for all the risk categories mentioned in Table 2.4 what are the new suggestions and then apply those in order to achieve the predicted savings.

Regarding the second way of using the tool we can have a look again at Figure 2.7. We can distinguish five columns. The "triggered" column refers to the perecentage of transactions for which the specific risk check was triggered. Next, the "chargebacks" column refers to the percentage of chargebacks out of the total number of transactions that triggered the specific risk check. Similarly, the "refusals" column refers to the percentage of chargebacks out of the total number of transactions that triggered the specific risk check. Subsequently, in the "current score" column we can see the current configurations that the merchant has chosen. In the "new score" column, merchants can manually modify the new scores of the risk checks and then check what the predicted savings would be by clicking on the "Calculate Expected Results". For example, if the user changes the score of "card/bank account usage frequency" from 7 to 50 and the "shopper email usage" from 7 to -10, the effect of these changes to the number of false positives

and chargebacks can be seen under the section "Expected Changes in Results Using Simulated Configuration" (as depicted in figures 2.4-2.6), and ultimately to the total amount being saved or lost.

Figure 2.7: Suggested scores for some of the velocity checks.



Technical Aspects

Adyen's Risk Calculator calculates the results based on the Simplex algorithm. The Simplex algorithm, invented by George Dantzig in 1947, solves linear optimization problems; when referring to linear optimization problems, we mean the minimization or maximization of a linear function, often called objective function, which is subject to *linear constraints* (Dantzig & Wolfe, 1960). Expressed in a mathematical way:

$$\max c^T * x$$
$$\text{subject to } Ax \leq b, x_i \geq 0$$

where $x = (x_1, ..., x_n)$ the variables of the problem, $c = (c_1, ..., c_n)$ the coefficients of the objective function, A is an m*n matrix and $b = (b_1, ..., b_m)$ constants with $b \geq 0$.

The solution of a linear problem can be achieved in the following way. Any $x = (x_1, ..., x_n)$ which satisfies the constraints of the linear problem is said to be a feasible solution of the problem. If this point represents the one where the objective function reaches its required maximum, then the point is said to be the optimal feasible solution (OFS). Therefore, in order to find the solution of the linear problem, the first step is to determine whether at least one feasible point exists or not. In the first case, we have to find the OFS and in the latter case we have to show that it does not exist. The Simplex algorithm thus accomplishes the result in two phases; in Phase 1 it starts with

an extreme point and the result is either a basic feasible solution or an empty feasible region - i.e. the linear problem is infeasible. In Phase 2 the algorithm is applied again using as starting point the feasible solution determined in Phase 1. The result of the second phase is either the OFS or an infinite edge where the objective function is unbounded below (Dantzig & Wolfe, 1960). Geometrically speaking, the aim of the Simplex is to find the furthest point in a desired direction on a convex polyhedron.

In the context of the Risk Calculator, the problem that is being solved is practically the maximization of revenues derived from authorized transactions minus chargebacks, with constraint the set of the risk rules. To delve deeper, the algorithm behind REO uses in essence 4 distinct steps (Wolters, 2012):

1. Aggregation of merchant's historical payment data.

2. Composition of the set of linear equations.

3. Solving of the linear problem.

4. Finding the preferable solution.

During the first step, the historical payment data of the merchant are aggregated in a way that for each triggered risk check the amount of authorized payments and chargebacks is known. Given the number of the authorized transactions, as well as the chargebacks and a decision parameter, a set of linear equations is composed leading to the action rule of *accept* or *reject*. Afterwards, the linear problem is being solved with the Simplex algorithm and in the end the preferable solution out of the feasible region is chosen.

In addition to the Simplex algorithm, a local search algorithm is used in order to find a solution in the feasible region that is closer to the merchant's current configurations. The reason for using the local algorithm is to find a solution which is "closer" to the merchant's current configurations, as according to Wolters (2012), it was found that merchants are only willing to make changes which are small and in line with their perceptions about risks.

2.7 Behavioral Economics

Delving into the problem of user adoption requires deep exploration and understanding of human behavior. If we reflect for a moment, we might realize that multiple times per day we have to take

decisions, either minor or more important. However, what kind of decision we take is influenced by several factors. As such, a whole sector has emerged that studies the reasoning behind people's decisions; the Behavioral Economics. In this Section we present some of the most essential notions of this field, according to their relevance for the problem under investigation.

2.7.1 Rational Choice Theory

The Rational Choice Theory (RCT) dominated economic concepts for years. The core of this theory is that people are expected to make decisions that result to the most optimal level or benefit for them. Hence, the "rational man" should always weight costs and benefits and choose the option that is going to bring the largest profit (Becker, 1976). In this theory, people have stable preferences. Many scientific branches, such as sociology, political science and criminology have extensively relied on RCT (Zey, 2015).

There are six main assumptions (Coleman & Fararo, 1992; Scott, 2000; Zey, 1997) that RCT is based on; (1) people act as self-maximizing individuals who are not influenced by others or by social norms, (2) people have complete knowledge regarding both theirs and others resources and preferences, (3) people evaluate the changing conditions and act based on rationality, (4) preferences are stable, (5) people are not influenced by emotions during the decision-making process, they remain rational, (6) interaction is defined as an exchange that occurs when benefits and costs are in equilibrium.

Based on the above, we can relate the concepts of RCT to both merchants' as well as fraudsters' behavior. Regarding fraud, we can infer that fraudsters follow a cost-benefit analysis in order to choose the most rewarding and safe type of fraud to commit, while taking into account all the risks (Gilmour, 2016). This happens in all types of crimes. According to RCT, fraud is not a random event but a careful weighing of rewards and losses. It should also be noted that according to Gilmour (2016), committing fraud reflects a rational choice from the perspective of the offender.

Regarding merchants' behavior, we know that according to RCT it is expected that they will always take a decision that maximizes their benefits. In the context of e-commerce, the largest benefit for a merchant is the maximization of revenues. Since tools like the Risk Calculator are designed as profit-maximizing auxiliaries, we expect that merchants would choose to use them as they have proven to be more effective than traditional alternatives (Wolters, 2012).

However, we frequently notice in everyday life that people are not always acting rational. As such, RCT has been criticized by several scholars. Common critics focus on the way RCT deals with psychology or with the failure of its empirical applications (Green, Shapiro, & Shapiro, 1994). Theories such as Bounded Rationality (Simon, 1982), as well as Prospect Theory (Kahneman & Tversky, 1979) explain people's decisions by focusing more on psychological aspects. In short, these theories support that people make choices based on heuristics, biases and overconfidence which often do not result in optimal benefits for them.

2.7.2 Bounded Rationality

One of the most dominant counter-theories of RTC is the concept of Bounded Rationality, which suggests that people are not always making the optimal decisions but are rather restricted by limits in knowledge and computational capacities (Simon, 1982). Since people in the modern era have various computational tools that aid them in their decision-making process, they are highly dependent on these machines which according to Simon (1972) *"these tools make more tractable the task of matching man's bounded capabilities with the difficulty of his problem"*.

Apart from that, the environment is a factor that can influence the rationality of the decisions taken by individuals (Gigerenzer & Goldstein, 1996). As such, people are considered "ecologically rational" when they make the best possible use of their limited processing capabilities, that results to near-optimal decisions.

Under the constraints of limited knowledge and processing capacity, scholars have highlighted the importance of feedback in order to make effective decision-making (Thaler & Sunstein, 2008). Specifically, experience, good information, as well as prompt feedback seem to be key factors that help people to make effective decisions. These factors are regarded as the best answer to the constraints posed by Bounded Rationality.

Lastly, Ariely (2008) indicates that Bounded Rationality affects the way people perceive prices, due to limits in thinking processes. Particularly, the concept of *zero price effect* illustrates that consumers give disproportionally higher value to products advertised as "free". As an example, a free chocolate is more attractive relative to a $0.19 chocolate than a $0.01 chocolate is compared to one priced at $0.20. In this example, according to RTC, a rational consumer should be indifferent between these two options since the price difference is 19 cents in both cases, however according to Bounded Rationality the consumer would always opt for the free product.

2.7.3 Prospect Theory

The mechanisms behind irrational behavior are further captured by Prospect Theory, developed by Daniel Kahneman and Amos Tversky in 1992. Prospect Theory aims at providing explanations on the decision making process when people are faced with risks and the outcome is unknown. In particular, this theory was developed as another alternative to RCT.

The core of Prospect Theory is that people, when faced with uncertainty, tend to overweight outcomes which are certain in comparison to outcomes which are merely probable (Kahneman & Tversky, 1979). This results in people being risk-averse in choices involving certain gains, and risk-seeking in choices involving certain losses. Additionally, people act inconsistently when the same choice is presented to them in a different way. The latter, known as *framing*, is proven according to the authors a crucial factor influencing decision making. Hence, risk attitudes as well as framing consist the basic concepts of the theory. Later, Prospect Theory was enhanced by introducing heuristics and biases.

Heuristics can be defined as mental shortcuts that people make in order to take quickly decisions (Kahneman, 2011; Plous, 1993). The *availability heuristic* influences decision makers by making them assess the probability of an event by the ease with which a concept can be brought to mind (Plous, 1993). The trap here is that improbable events capture people's attention and are subsequently considered to be the norm than the exception. For instance, merchants might believe that physical theft is more plausible to happen to them than electronic financial fraud, if the media is mostly focused on events of robberies and violence. Thus, they might underestimate the probability with which e-commerce fraud occurs and therefore might not find a reason for using anti-fraud tools.

In addition to heuristics, people tend to judge based on feedback and expertise, while they moreover make decisions based on memories and not on experiences (Kahneman, 2011). This practically means that if they did not have in the past any negative experiences regarding their transactions, they will probably not think about taking proactive measures to prevent fraud.

Furthermore, biases play an important role in shaping peoples' attitude towards risks. On one hand, *confirmation bias* leads people to only take into account events that support their beliefs (Plous, 1993). Therefore, people are more likely to rely on a single source that confirms their perceptions rather than on actual research that might prove the opposite beliefs. On the other hand, *hindsight bias* makes people regard what has already happened as obvious or inevitable (Plous, 1993).

The aforementioned notions are proven to systematically and unconsciously affect people while taking decisions. Dealing with risks related to transactions and hence to revenues, requires careful thinking. Nevertheless, psychological aspects act on the "backstage" thus decision makers are often influenced by several latent factors.

2.7.4 Dual System and Temporal Dimensions

According to Kahneman (2011), human decision making does not follow rationality because people's brain functions by using a *dual system*; System 1 is thinking in an intuitive, automatic and experience-based way, while System 2 provides deliberative and analytical way of thinking. The two systems often clash while we try to make a decision, and System 2 is the one to "fail" first.

System 1 is the home of heuristics, i.e. cognitive biases that we use when making decisions. One of these heuristics is the *Status Quo Bias*, which is defined as the preference for things to remain the same, such as a tendency not to change behavior unless the incentive to do so is strong (Samuelson & Zeckhauser, 1988). Inertia is one form of people's propensity to remain at the status quo (Madrian & Shea, 2001). A well-known example of this includes the low rates of pension plan enrolment when people have to make some effort to sign up ('opt-in'). It has been noted that in this situation, an effective way to increase enrolment rates is to change the default by making it 'opt-out', i.e. people do not have to make an active choice. Inertia, as well as a lack of self-control are problems that make changes in default options from opt-in to opt-out an effective strategy.

Another important branch of Behavioral Economics, includes the dimension of time as a factor influencing human perception. As such, the *time-discounting effect* suggests that present events are weighted more heavily than future ones (Frederick, Loewenstein, & O'donoghue, 2002). Generally speaking, the value of events that are further in the future are perceived as lower than the ones closer to the present. As an example, experiments show that people's preference to receive 100 euros and 110 euros in one month, will not be the same as receiving 100 euros now and 110 euros in one year and one month - although that the gap is one month in both cases. This acknowledges that people are biased towards the present and poor predictors of future experience.

2.8 Externalities and Stakeholder Incentives

In Economics, there are two other important notions that have been extensively used in order to assess the problem of users' behavior with regards to security measures and advices. These notions, closely interrelated, are the externalities and the stakeholder incentives. Below, we offer a brief explanation of the two terms.

On one hand, *externalities* refer to the costs or benefits which are borne by parties who did not choose to incur them. In the information security context, users that might follow advices to shield themselves against attacks, also burden themselves with indirect costs such as the time spent to understand those advices. Contemporary studies in this field (Herley, 2009) indicate that users' reluctance to follow security advices are in reality rational, since most of the times the time and effort they have to spend outweighs the benefits that the security advice brings. According to the author, there are cases where security advice causes more harm than the attack it aims to eliminate. Hence, policy makers should carefully reflect on externalities before designing security measures.

On the other hand, Bauer and Van Eeten (2009) emphasize that the overall level of security is the result of the decisions that the different stakeholders make by the factors, i.e. incentives, that influence them. These incentives, which are often in the form of revenue and reputation effects or social relations, are shaped by the direct costs and the externalities connected to the security advice.

2.9 Conclusions

In this Chapter we conducted a literature review in order to provide background information by introducing all the concepts related to the problem under exploration. We began with the basic notions related to the payments industry, such as payment statuses, chargebacks and false positives and identified the most important stakeholders into the field. Later on, we explained the two prevalent situations where e-commerce fraud occurs; the card-present and card-not-present environments. We also discussed the countermeasures developed by financial institutions to prevent fraud, and moreover referred to machine learning and data mining solutions developed and discussed by the academia. We additionally underlined that the problem is not on the solutions offered to combat fraud, but specifically on whether these solutions are being adopted by users; contemporary research indicates that there are several aspects that influence users

adoption of data mining tools. Subsequently, we explained Adyen's risk management system and in particular the Risk Calculator tool, which represents the case study on which the thesis is based on. Finally, we reviewed some of the most basic theories in Economics, such as Prospect Theory and Externalities, which can help us understand psychological and behavioral factors related to the users' decision-making, as well as incentives to adopt innovative tools.

Chapter 3

Analysis of Historical Data

After introducing the relevant background information, we move to the main body of the research. In this chapter, the first research sub-question is related to the usage of the Risk Calculator and is going to be answered by means of data analysis. To recall, this question was formulated as follows:

SQ: *What does the historical data indicate about the usage of the Risk Calculator and the risk settings adopted by merchants?*

The answer to this question is expected to reveal patterns regarding the usage of the tool, by means of comparing and contrasting users versus non-users through statistical analysis.

3.1 Methodology

As the research question implies, a quantitative empirical research with an exploratory objective is going to be performed.The research will be quantitative since it includes the analysis of quantitative transaction datasets and moreover is empirical, as it is a data-driven research with conclusions being verified by observations.

Relevant for this research, which aims to explore merchants' incentives for adopting profit-maximizing tools, is transaction data of users' of the Risk Calculator tool, as well as of non-users. Furthermore, we are interested in information regarding the configuration of the risk settings of the merchants, since this can shed light on their risk attitudes.

For this reason, the first step was to identify the group of users of the tool. This was done by accessing log files and getting a list of the names of companies that visited the page for a 3 month period, i.e from December 2015 to February 2016. The basic assumption made during this step was that every company that visited the tool is a user. This resulted in 153 companies, however a limitation that should be mentioned is that the Risk Calculator works on a merchant account level rather than on a company account level. This practically means that a single company might have multiple merchants, for example one different merchant for every country that the company is present, nevertheless there was no option to track which of the specific merchants visited the page. Thus, the second assumption made was that all the merchants of the companies that were identified in the user-list are using the tool.

After the identification of users, the next step was to identify the non-users in order to compare these two groups.As already mentioned, Adyen has thousands of customers, hence a random sample of those companies that did not visit the page was selected. Since the users were 153, we wanted a similar sample size to compare; this resulted in 154 companies. It should also be taken into account that the Risk Calculator is included in reports that merchants have to pay in order to have access to, and is generally observed that larger companies are those who pay for access to reports including this extra feature. Hence, we know beforehand that the non-users are generally smaller companies than the users.

The next step after the two groups were identified, was to extract data related to their transactions, risk attitude and configuration of tool. This is done through the use of SQL queries. Therefore, the datasets used for the research can be divided in the following categories.

3.1.1 Transaction Dataset of Users and Non-users of the tool

The first dataset consists of of 6,928 records containing information about the authorised, refused by risk and chargeback counts of the users of the tool for a 6-month period (December 2015-May 2016). As a reminder, the number of companies in this dataset is N=153. Moreover, the respective amount in euros for these transactions was included in the dataset. Based on these variables, the authorised, refusal and chargeback rates, as well as the average transaction value for these merchants were calculated. Although we have data of the visits to the tool for a 3-month period, in this dataset we expanded the time period to 6 months in order to capture effects in chargebacks that might take place later. An example of how the dataset looks like can be seen in table 3.1. For each of the merchants of one company, the respective transactions over the 6-month period were given. For the companies that were present in the dataset, the industry that they

belong to and the number of merchants they have was moreover found.

Table 3.1: Part of the dataset with merchants' transactions

company id	merchant id	date	authorised count	authorised eur	risk refused count	risk refused eur	chargeback count	chargeback eur
1	10	2015-12-01	4957	391676.67	171	-21057.71	2	-297.0
2	20	2015-12-01	28546	2097072.58	2979	-294735.77	1009	-138433.80
3	30	2015-12-01	94625	3924174.63	2145	-183873.84	106	-14020.4
4	40	2015-12-01	16910	701704.21	280	-11138.29	76	-4173.75

Similarly, the dataset of the non-users consists of 2,951 records with the same information for the same time span. The number of companies in the non-users is 154.

3.1.2 Dataset with Risk Parameter Changes of Users and Non-users

The second dataset obtained shows for each of the users and non-users what risk parameters did they change, when did they change them and what score they gave to these parameters. For the users group, this dataset consists of 10.786 records, while for the non-users of 2.847 records. These two datasets show the risk changes that the merchants made, for the desired time range, i.e from December to February. Table 3.2 shows how the dataset looks like.

Table 3.2: Part of the dataset with Risk Parameter Changes

merchant id	risk check id	score	begin date	end date
1	78	1	2015-12-01	9999-12-31
1	54	100	2015-02-01	9999-12-31
2	55	100	2016-01-01	2016-03-01
2	55	60	2016-03-01	9999-12-31

As can be noticed, there are two date columns. The first one, "begin date" denotes when did the merchant change the respective risk check, while "end date" denotes until when this change was active. Hence, the end date "9999-12-31" means that the change made to the risk check is valid till today.

3.1.3 Dataset with Tool's Suggestions for Users and Non-users

Finally, information regarding the use of the tool were essential in order to understand its functionality. Nevertheless, data related to the suggestions it makes for the users were not available in an exportable format, thus they had to be manually collected. This was done for 17 of the high visitors, 17 of the low visitors and 17 of the non-users, by entering the tool and writing down the suggested changes and savings for each merchant. The tool might suggest several optimizations for merchants, hence the name of the same merchant might appear for several times in the dataset as depicted under 3.3.

Table 3.3: Part of the dataset depicting the tool's suggestions

	Current genuine transactions	Current refused	Current chargebacks	Simulated false positives	Simulated chargebacks	Total amount (euros)	Risk suggestion	Proposed configurations
MerchantA	1500	85	23	-52	6	4200	low false positives	1
MerchantB	3000	112	50	-74	7	6500	low false positives	2
MerchantB	3000	112	50	15	-12	5000	low chargebacks	2

Analyzing the afore-described datasets in order to identify patterns and determine relations is the main objective of this phase. This step is executed by deploying statistical analysis, such as comparison of means for identifying similarities or differences between the group of users and non-users,through the use of R Studio software. Apart from that, a number of descriptive statistics are offered, providing information about the groups.

3.2 Findings on Usage Patterns

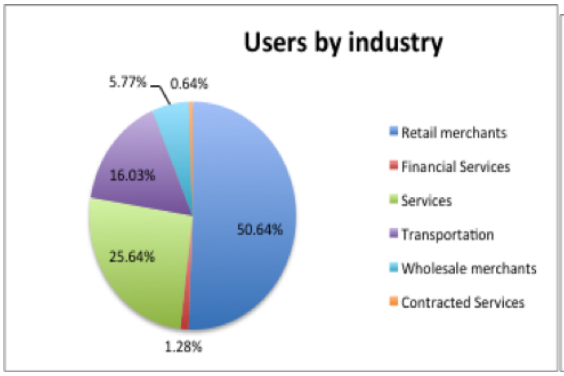
The first step of the statistical analysis aims at describing the properties of the two groups. The properties include *characteristics* such as the size of these companies and the industry they belong to. The second step aims at analyzing the *behavior* of the groups, i.e. how many times the users visited the tool and how often do merchants change their risk parameters. These insights on the characteristics and behavior of the merchants, are expected to shed light on their incentives for using the tool. Finally, we are interested in exploring the tool's suggestions regarding the savings and the risk attitudes it recommends to the merchants, in order to get a deeper understanding of its functionality. The results of each of these three steps are presented below.

3.2.1 Characteristics of Users and Non-users

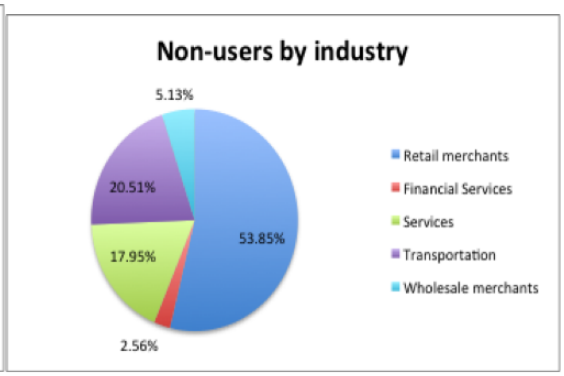
Beginning with the industry that the merchants belong to, the following figure depicts the sector for the users group.

Figure 3.1: Industry of merchants (users and non-users).

(a) Users classified by industry. N=153



(b) Non-users classified by industry. N=154



As can be seen in the pie chart, more than half of the users - i.e. 50,64% - belong to the Retailers industry, followed by Services and Transportation industries. From the Retailers industry, most of the merchants belong to the Commercial Cloth sector, while from the Services industry most of the merchants belong to Gaming services. Finally, regarding Transportation the majority of the merchants has to do with travel-related arrangements, such as Airlines.

The distribution of the non-users by industry seems to be quite similar to those of users, as illustrated by the pie chart below. The difference is that for the non-users, Transportation industry comes before Services industry.

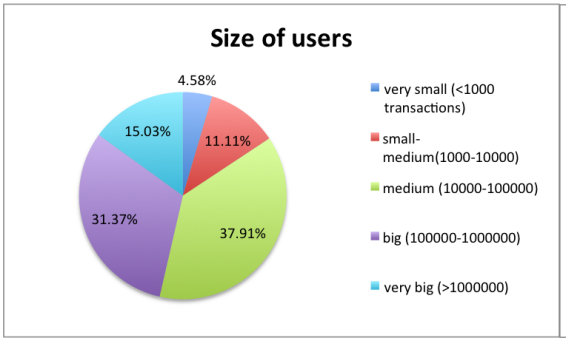
From the non-user Retailers, most seem to follow in the category of Digital Goods/Applications, whereas from the Transportation industry most fall in the category of Travel agencies and Tour Operators. Lastly, most non-users in the Services industry belong to Ticket Agencies.

As regards the size of the companies, most of the users can be classified as medium to big (Figure 3.2a). By "medium" we mean companies which have between 10.000-100.000 transactions for this 9-month period, while by "big" we mean companies with 100.000-1.000.000 transactions. For the non-users, we see that the chosen sample consists of very small companies (Figure 3.2b).This was an expected result, since the Risk Calculator is a statistical tool which means that

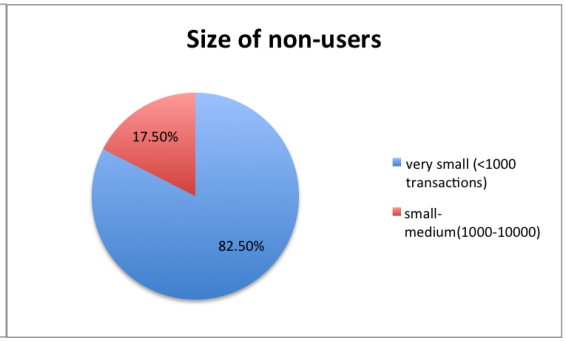
no optimal solution can be found for small transaction volume; hence the "good candidates" for the tool are the merchants with several hundred transactions.

Figure 3.2: Size of merchants (users and non-users).

(a) Users classified by size. N=153

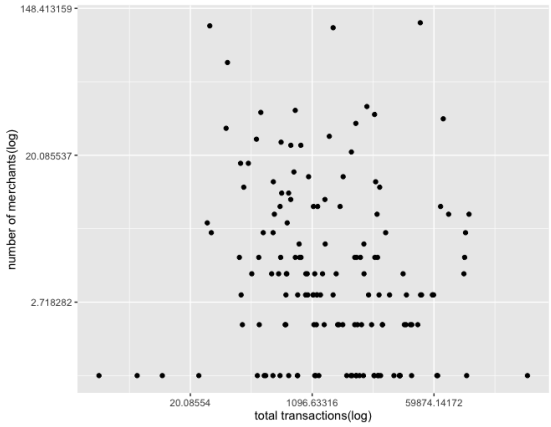


(b) Non-users classified by size. N=154

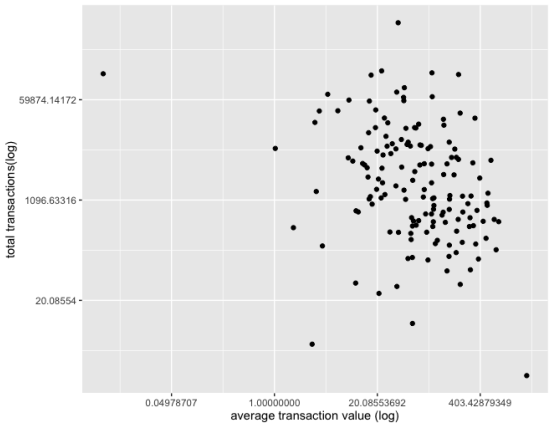


In order to explore more the properties of the users, the following hypotheses are made. The aim is to understand whether there is a correlation between the number of merchants of a company and the total transactions, i.e. if multiple merchants in one company equals having more transactions in total. By intuition, we would expect that the larger the company, the more merchant accounts it will have. Furthermore, we would like to test whether larger companies have also higher transaction values.

(a) Scatter plot of total transactions and average transaction value. N=152



(b) Scatter plot of total transactions and average transaction value. N=154



Since the number of total transactions is quite large compared to the number of merchants and the average transaction value, we use a logarithmic scale for the y axis. For the scatter plot of total

transactions-average transaction value, an outlier has been removed. The Spearman correlation coefficient can be seen for both graphs in table 3.4.

Table 3.4: Spearman correlation coefficients

Variables	Spearman coefficient
Total transactions - Number of merchants	-0.18
Total transactions - Average transaction value	-0.32

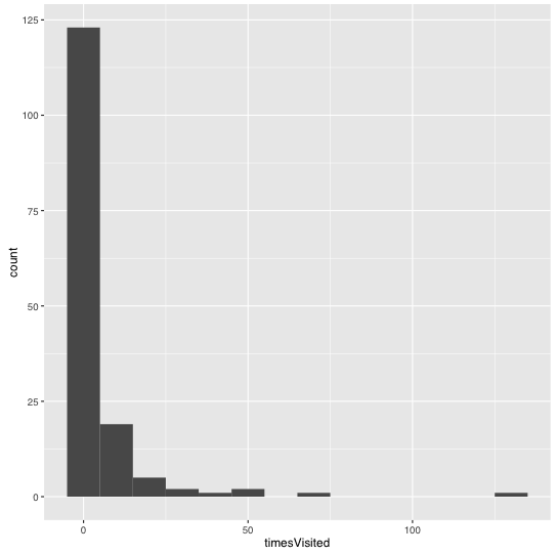
For the total transactions and number of merchants, the Spearman coefficient is -0.18 which indicates a very weak negative relation between the two variables. Similarly, for the total transactions and average transaction value the Spearman coefficient is -0.32 ; this indicates a moderate negative relation between the two variables. Therefore, we conclude that a higher transaction volume is *not* related to higher number of merchant accounts, nor higher average transaction value for the company.

The merchants visiting the Risk Calculator belong to retailers, services and transportation industries. Moreover, the majority of them processes 10,000 - 100,000 transactions during a 9-month period, indicating large-sized companies.

3.2.2 Behavior of Users and Non-users

Frequency of visiting the tool. Starting with the group of users, we first check the the number of times they have visited the tool. It appears that within a 3-month period the most active user visited the page 126 times, while there were also users who visited the page only once. This very low visiting frequency points to the fact that some of the companies identified in the user list, might practically not be active users of the tool but randomly visiting the page. For this reason, the median of the number of times the users visited the tool was calculated, as shown by the histogram below.

Figure 3.4: Histogram of the number of times that the users visited the tool.



The median is found to be 2, so the users group is split in two subgroups: high visitors and low visitors. High visitors group consists of all the companies that visited the page 3 or more times, while low Visitors group consists of the companies that visited the page once or twice. The aim of this splitting is to compare these two groups in terms of size and authorization/refusal/chargeback rates to see if they behave differently, i.e. if the frequent use of the tool has implications on these factors. The statistical analysis to be used for this comparison is known as "Comparison of Means" and the whole procedure is being described in the following subsection. Note that the low visitors group, consists of users who visited the page only once or twice in a 3 month period; since this is a very low frequency, we assume that the low visitors are probably a special category of non-users.

Comparison of Means: High Visitors and Low Visitors

In order to apply a statistical test, first we have to determine the sample size and formulate the hypotheses. After that, we check whether the data follows the normal distribution, and thus we choose either a parametric or a non-parametric test.

Hypotheses Building. Exploring the behavior of merchants requires describing the premises and thereafter checking whether they are verified or not. Below, we present the six hypotheses aiming to answer through the comparison of means.

Table 3.5: Hypotheses to be tested through comparison of means.

#	**Hypothesis**
1	H0: The average number of merchant accounts is the same in high and low visitors groups.
	H1: The average number of merchant accounts differs in high and low visitors groups.
2	H0: The average authorization rate is the same in high and low visitors groups.
	H1: The average authorization rate is higher in high visitors group rather than in low visitors group.
3	H0: The average risk refusal rate is the same in high and low visitors groups.
	H1: The average risk refusal rate differs in high and low visitors groups.
4	H0: The average chargeback rate is the same in high and low visitors groups.
	H1: The average chargeback rate differs in high and low visitors groups.
5	H0: The average total transaction volume is the same in high and low visitors groups.
	H1: The average total transaction volume differs in high and low visitors groups.
6	H0: The average transaction value is the same in high and low visitors groups.
	H1: The average transaction value is different in high and low visitors groups.

The intuition behind all these hypotheses is on the one hand to check whether merchants' behavior is dependent on their characteristics, and on the other hand to check the effect that the Risk Calculator has on the main KPIs, i.e. chargebacks and false positives. Hypothesis #1, is based on intuition regarding the fact that larger companies should have more merchants accounts, as it is more probable to be active in multiple countries or have the need to fragment their business in more accounts for better reporting. Furthermore, hypotheses #2 - #4 are related to the fact that the use of the tool aims to minimize chargebacks and false positives, i.e. chargeback and risk refusal rate (Wolters, 2012). Hence, we assume that merchants with high chargebacks and false positives would be the ones visiting the tool more frequently since they wish to minimize these transactions. Additionally, through hypotheses #5 we aim to check whether merchants with higher value products are more interested in preventing chargebacks and refusals (De Gennaro, 2006). Lastly, with hypothesis #6 we assume that the size of the merchant determines the use of automated anti-fraud tools or the investment of resources to prevent fraud (LexisNexis, 2015).

Hypotheses Testing. The sample size of the high visitors group is N=64, while for the low visitors N=88. After plotting the histograms for the authorisation, refusal and chargeback rates, as well as for the total transactions, average transaction value and number of merchants it was found that the data do not follow the normal distribution. Hence, the non-parametric Wilcoxon test was chosen to perform the comparison of means. The following two tables depict the statistical parameters for these two groups.

Table 3.6: Statistical parameters of High Visitors group (N=64).

	# of merchants	Authrate	RefusalRate	ChargebackRate	Total transactions	Average tr. value
Median	4	0.81	0.017	0.0004	5267	56.04
Mean	8.60	0.78	0.048	0.002	41723.26	118.06
St.dev.	11.94	0.123	0.079	0.005	164169.5	136.76
Range	1-71	0.45-0.97	0-0.38	0-0.02	65-1288862	0.0067-606.26

Table 3.7: Statistical parameters of Low Visitors group (N=88).

	# of merchants	Authrate	RefusalRate	ChargebackRate	Total transactions	Average tr. value
Median	3	0.83	0.006	0	1225.5	68.21
Mean	9.57	0.80	0.034	0.0016	8948.73	142.03
St.dev.	21.76	0.13	0.053	0.004	23161.61	212.43
Range	1-122	0.25-1	0-0.23	0-0.02	1-175015	3-1563.53

Table 3.8: Wilcoxon p-values.

#	**Wilcoxon p-value**
H1	0.2509
H2	0.1224
H3	0.1224
H4	0.2192
H5	0.01512
H6	0.5173

According to Table 3.8, the following conclusions are driven:

• High Visitors and Low Visitors have on average the same number of merchant accounts (p-value = 0.2509).

• High Visitors and Low Visitors have on average the same authorization (p-value = 0.1224), refusal (p-value = 0.1224) and chargeback rate (p-value = 0.2192).

• High Visitors have on average more transactions in quantity than Low Visitors (p-value = 0.01512).

• High Visitors and Low Visitors have on average the same transaction value in euros (p-value = 0.5173).

The finding that frequent visitors have the same refusal and chargeback rate as low-frequency visitors seems odd at first glance. This indicates that either the tool does not work properly or that

low-frequency visitors have other, more efficient mechanisms for monitoring their transactions. By looking at the log files, we eventually found out that no merchant is actually applying the tool's suggested configurations. This made us further explore whether merchants are changing the scores of the risk checks not through the tool, but through their settings page where they can manually select which scores they want to change. The results are later presented in this Section.

Comparison of Means: Users and Non-users

The same technique as described above is going to be applied between the users and non-users of the tool. Since the two subgroups of high and low visitors are similar, they are combined as one group and subsequently are compared to the non-users.

Hypotheses Building. Table 3.9 summarizes the respective hypotheses in order to compare the users versus non-users groups.

Table 3.9: Hypotheses for checking through comparison of means.

#	**Hypothesis**
1	H0: The average number of merchant accounts is the same in users and non-users groups.
	H1: The average number of merchant accounts differs in users and non-users groups.
2	H0: The average authorization rate is the same in users and non-users groups.
	H1: The average authorization rate is higher in users and non-users groups.
3	H0: The average risk refusal rate is the same in users and non-users groups.
	H1: The average risk refusal rate differs in users and non-users groups.
4	H0: The average chargeback rate is the same in in users and non-users groups.
	H1: The average chargeback rate differs in users and non-users groups.
5	H0: The average total transaction volume is the same in users and non-users groups.
	H1: The average total transaction volume differs in users and non-users groups
6	H0: The average transaction value is the same in users and non-users groups.
	H1: The average transaction value is the same in users and non-users groups.

The reason for not specifying the direction in H2-H4 is based on the fact that users might either be merchants with high chargebacks and refusals who wish to reduce those rates, or already have low chargebacks and refusals due to the systematic use of the tool.

Again, our data do not follow the normal distribution thus the Wilcoxon test is chosen to perform the analysis. As already mentioned, the sample size of the users is N=153, while the sample size of the non-users is N=154. The following tables show the statistical parameters for the groups of

users and non-users respectively.

Table 3.10: Statistical parameters of Users group

	# of merchants	Authrate	RefusalRate	ChargebackRate	Total transactions	Average tr. value
Median	3	0.82	0.013	0.00008	1674.5	64.68
Mean	9.17	0.79	0.04	0.0018	22748.54	131.94
St.dev.	18.23	0.128	0.077	0.00427	108707.8	184.24
Range	1-122	0.25-1	0-0.38	0-0.029	1-1288862	0.006-1563.53

Table 3.11: Statistical parameters of Non-users group

	# of merchants	Authrate	RefusalRate	ChargebackRate	Total transactions	Average tr. value
Median	1	0.93	0	0	41	72.49
Mean	7.9	0.88	0.008	0.008	1201.38	220.47
St.dev.	59.85	0.17	0.045	0.081	7597.038	477.53
Range	1 - 741	0 - 1	0-0.36	0 - 1	1 - 91215	0 - 4405.17

Table 3.12: Wilcoxon p-values for the comparison of means between users and non-users.

#	**Wilcoxon p-value**
H1	0.01103
H2	0.00000007365
H3	0.00000007365
H4	0.001762
H5	0.000002175
H6	0.2583

According to Table 3.12 it was found that:

• The non-users group appears to have on average less merchant accounts than the users group (p-value = 0.01103).

• The non-users group has on average less transactions in quantity than the users group (p-value = 0.0000002175).

• The non-users group has on average higher authrate (p-value =0.00000007365) and lower refusal (p-value =0.00000007365) and chargeback rates (p-value = 0.001762) than the users group.

• The non-users have the same average transaction value with the users group (p-value = 0.2583).

The results of the Wilcoxon test indicate that there are indeed some differences between the visitors and non-visitors group. Firstly, the non-visitors are smaller companies than the visitors; this might be related to the fact that the tool works better for larger amount of transactional information. Secondly, the hypothesis that merchants with higher chargeback and refusal rate will be the ones more interested to use the tool is verified. The fact that non-users have higher authorization rate than the users is most plausibly related to the fact that they have lower traffic, hence greater chance for legitimate transactions.

For the user group, we also checked the correlation between chargeback rate and times that the merchant visited the tool page, refusal rate and times of visiting the tool and finally between chargeback and refusal rate. The results are summarized in table 3.13. As can be seen from the table, no significant relation exists between the variables.

Table 3.13: Wilcoxon test values.

Variables	Spearman coefficient
Number of visits - Chargeback rate	0.22
Number of visits - Refusal rate	0.13
Chargeback rate - Refusal rate	0.23

Similarly for the non-users group, the correlation between the chargeback and refusal rate, as well as between the total transactions and number of merchants was checked. The results can be seen in table 3.14.

Table 3.14: Spearman correlation coefficient

Variables	Spearman coefficient
Refusal rate - Chargeback rate	0.47
Total transactions - Number of merchants	0.24

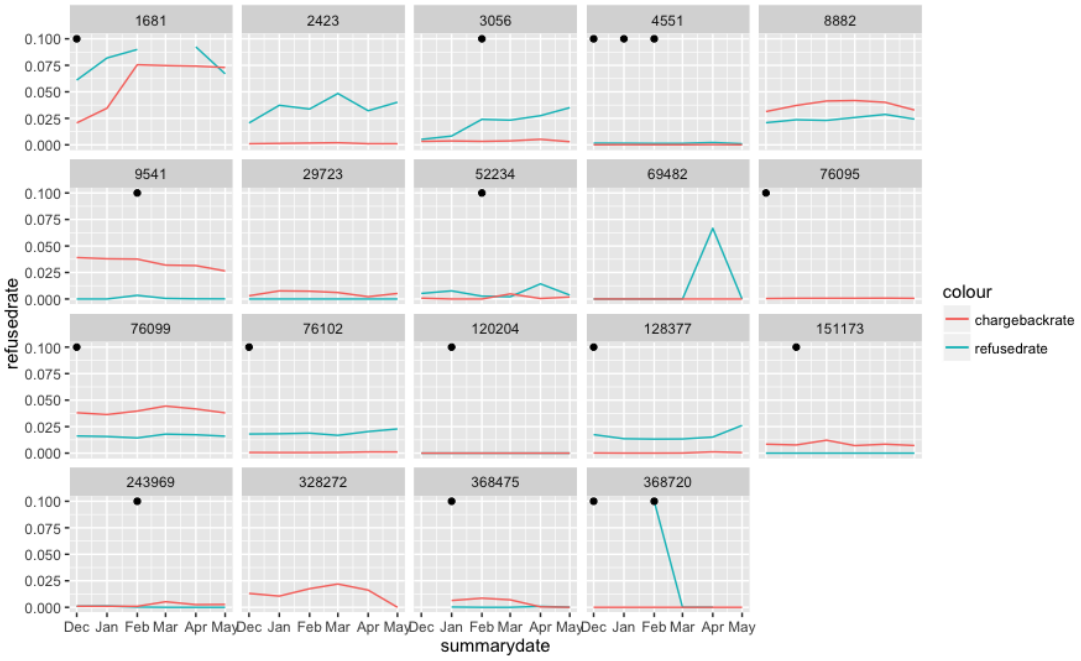
As we can see, the spearman coefficient between refusal and chargeback rate for the non-user group is moderate to strong, something that indicates that as the chargeback rate increases, so does the refusal rate. As regards the relation between total transactions and number of merchants, it can be characterised as a moderate positive relation.

The visitors of the tool seem to have on average higher chargeback and refusal rates than the non-visitors. Moreover visitors are larger companies in terms of transaction volume.

Frequency of Risk Activity

Users. Another aspect pointing to the behavior of the merchants, is how often they change the risk configurations and whether they generally engage with risk management. In order to measure the risk activity, we look at the dates when merchants make a change and the corresponding refusal and chargeback rates for this period. The expected finding would be that a correlation exists between changing risk settings and decreasing refusal or chargeback rate, either immediately or with a delay of a couple of months. For the sake of aesthetic simplicity, a random sample of merchants who use the tool was chosen and the respective plot showing the date when a risk change occurred, as well as the refusal and chargeback rate was drawn.

Figure 3.5: Risk changes versus risk refusal and chargeback rate of users (random sample out of n=1311 merchant accounts). Each dot represents when the change occurred, while the red line represents the chargeback rate and the light blue line the risk refusal rate.



First thing to be observed from Figure 3.5 is that not all merchants make changes to their settings. Merchants that do not make changes might have other mechanisms of protecting against risks, for example by using external risk tools instead of Adyen's risk system. Out of those who make changes, we notice in general that a change leads to a decrease in either risk refusal or chargeback rate. In order to quantify the visual correlation we observe in the Figure above, we created a

metric denoting the number of risk changes made per merchant per month, as well as the monthly increase/decrease in chargeback and refusal rates (delta values). The results of this correlation can be seen below.

Figure 3.6: Scatter plots between number of changes and deltas in refusal and chargebacks.

(a) Number of changes (monthly) Vs delta in chargebacks.

(b) Number of changes (monthly) Vs delta in refusals.

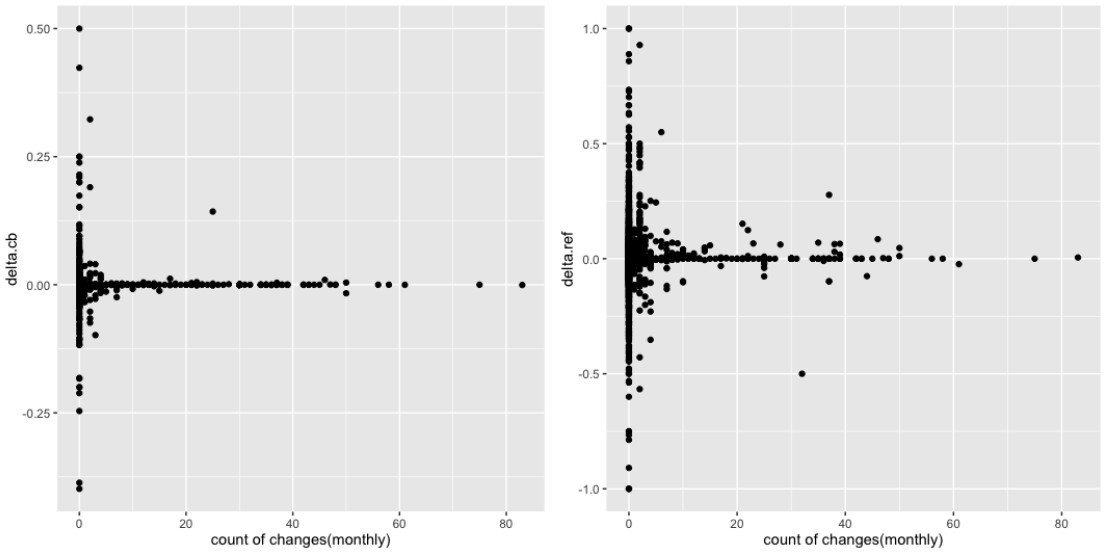


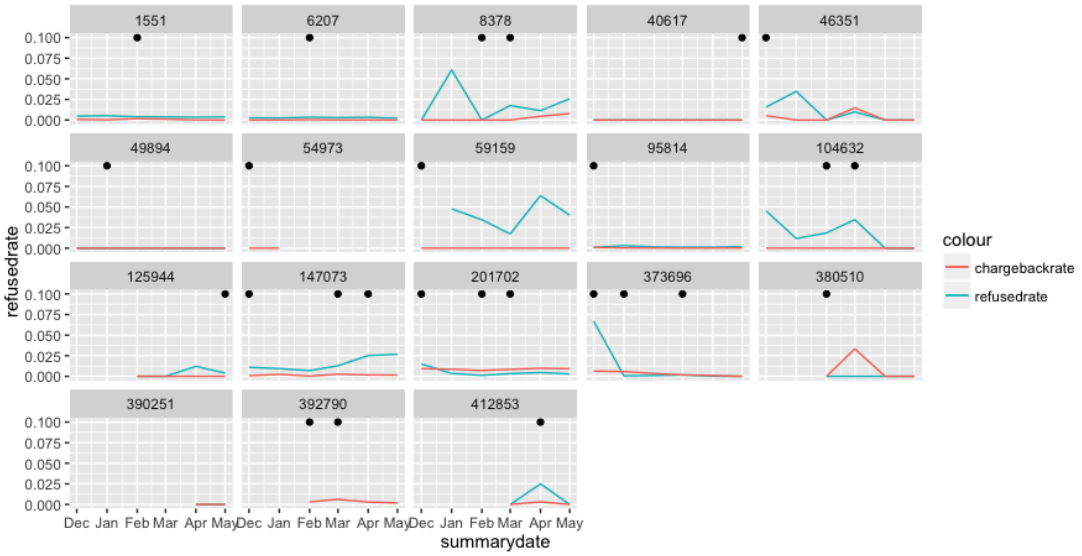
Table 3.15: Spearman coefficient between the number of changes and delta values.

Variables	Spearman coeff.
Counts - delta.cb	0.010
Counts - delta.ref	0.044

Interestingly enough, the Spearman rho does not indicate a strong relationship between the number of changes in the risk settings and the monthly increase/decrease in chargebacks and refusals. A possible explanation that this correlation is difficult to be captured is that the risk system is reactive; merchants first notice an increase in the two KPIs and later on they proceed to changes, hence the effect might not be as immediate as one would imagine.

Non-users. Obtaining a more holistic picture requires a comparison with merchants that have not visited the tool in the examined time period. Similarly, the risk activity for the non-users can be seen in Figure 3.7.

Figure 3.7: Risk changes versus risk refusal and chargeback rate of non-users (random sample out of n=629 merchant accounts). Each dot represents when the change occurred, while the red line represents the chargeback rate and the light blue line the risk refusal rate.



Again, we can distinguish visually some kind of correlation in terms of making a change and noticing an effect on the rates. In terms of Spearman correlation the results can be seen below:

Table 3.16: Spearman coefficient between the number of changes and delta values for non-users.

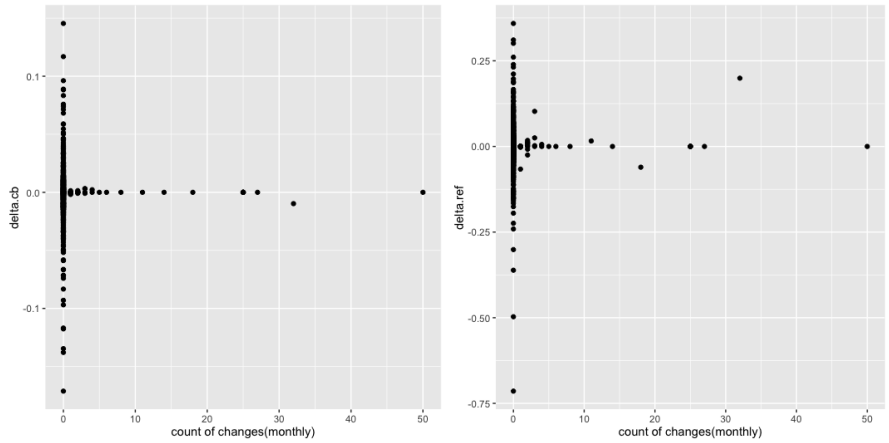
Variables	**Spearman coeff.**
Counts - delta.cb	0.0014
Counts - delta.ref	0.049

Table 3.16, as well as Figure 3.8 below, show no significant correlation between number of changes in the risk settings and increase/decrease in the KPIs. Generally speaking, the same observations mentioned under the section of users group hold for the non-users too.

Figure 3.8: Scatter plots between number of changes and deltas in refusal and chargebacks for non-users.

(a) Number of changes (monthly) Vs delta in chargebacks.

(b) Number of changes (monthly) Vs delta in refusals.



Users do not directly apply the tool's suggestions, however they are making changes to their risk settings in order to respond to chargbacks and refusals. Both merchants that visiti and do not visit the tool seem to be interested in making risk assessment through their settings.

3.2.3 Suggestions of the Tool

The last step of the analysis was to measure the suggestions of the tool in terms of projected savings and risk attitudes that it recommends to the users, as well as to the non-users.

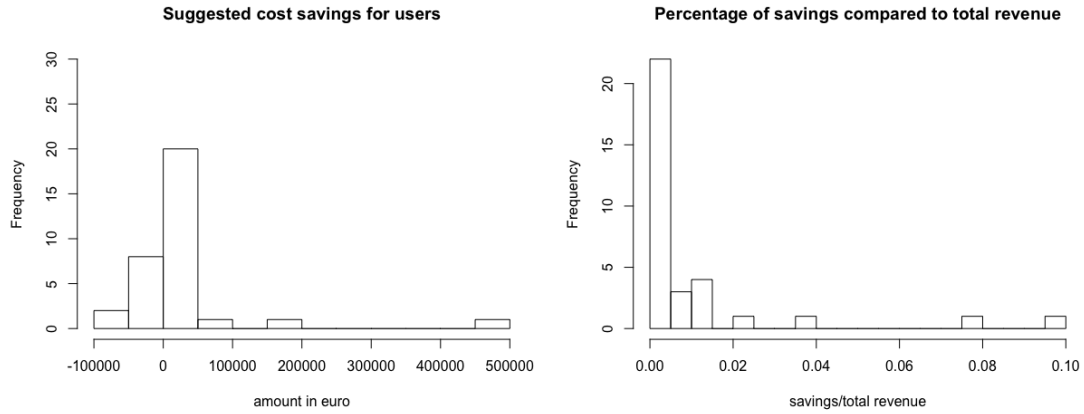
Projected Savings. The distribution of projected savings for the users can be seen in figure 3.10a. For most of the users, the projected savings account for up to 50.000 euros, while we can see also negative savings, i.e. losses. This is due to the fact that if merchants have a high number of chargebacks they are obliged to pay fines to the card schemes (e.g. Mastercard, Visa), therefore the tool shows that they have to decrease their chargebacks but on the same time accept more false positives thus resulting to a loss. Note that an outlier has been removed from the histogram, which was a company with projected savings 18 million euros. This might be due to the fact that this company is a financial institution with billions of revenues during this period, while most of the other companies belong to retailers, services or transportation industry.

For this sample of users, the projected savings were calculated as a percentage of their total revenue. The negative amounts, i.e. the losses, were turned into zero percentages since we are interested in capturing the savings. The distribution can be seen in figure 3.9b.

Figure 3.9: Suggestions of the tool for the user group.

(a) Projected savings of users as suggested by the tool.

(b) Users' projected savings as percentage of their total revenues.

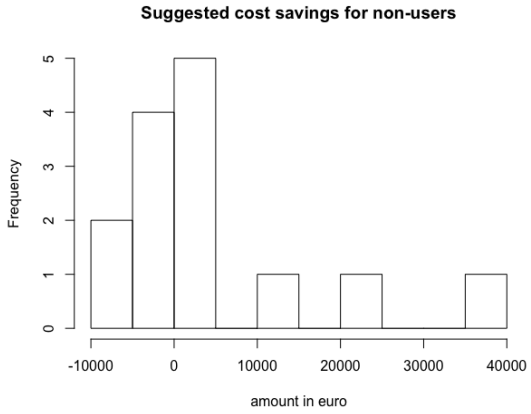


As can be seen from the figure, for most of the users the suggested savings account for 0.5% of their revenue. The highest percentage (10%) belongs to a retailer in the sector of online shopping mall.

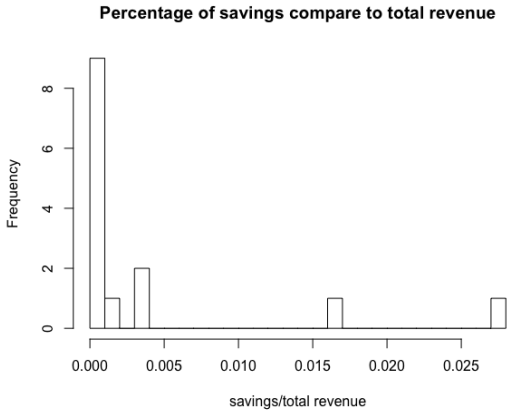
Regarding the non-users, the distribution of projected savings can be seen in Figure 3.10. For most of them, the calculated savings were up to 5.000 euros. Lastly, the savings expressed as a percentage of total revenues are depicted in Figure 3.10b. For most of the non-users the savings account for up to 0,12% of their total revenue.

Figure 3.10: Suggestions of the tool for the non-user group.

(a) Projected savings of non-users as suggested by the tool.



(b) Non-users' projected savings as percentage of their total revenues.



Suggested Risk Profile. As regards the risk attitude that the tool suggests, the results can be seen in table 3.17. For both the users and the non-users that the available configuration of the tool was one, i.e. when the tool provided only one feasible solution, the frequency of the suggested profile was calculated. As can be observed, for most of the merchants the tool suggested fewer false positives and no change in the number of chargebacks. This denotes a risk-taking attitude, since the false positives can be reduced when the settings are tweaked to be more loose rather than strict.

Table 3.17: Suggested risk profiles.

	Few False Positives	More False Positives	No change
Few Chargebacks	2	8	0
More Chargebacks	2	0	0
No change	11	6	0

In order to investigate whether the suggestions are in line with merchants' actual risk attitude, we created the histogram of the net score changes that visitors made to their settings for this specified 3-month period. As can be seen in Figure **??**, for the majority of the merchants the net score is zero or positive, denoting a risk-averse attitude. It should be noted that the histogram is created based on merchant account level; as explained earlier, a single company might have multiple merchant accounts and any changes to the settings occur on a merchant level rather than on company level, due to greater efficiency in reporting.

Figure 3.11: Histogram of net score given to risk checks by visitors.

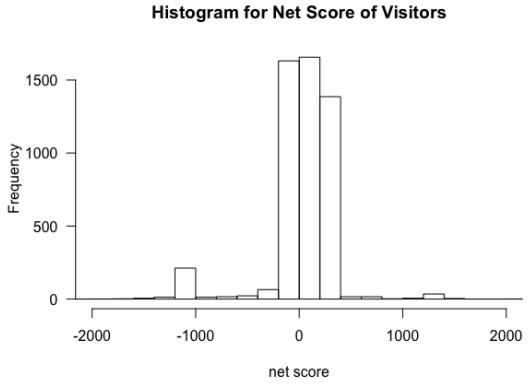


Table 3.18: Descriptive statistics of the histogram.

n	mean	median	st.dev
5106	38.99	99	325.5203

The tool's suggestions for the user group show slightly larger projected savings than for the non-users, however both amounts are quite low compared to total revenues. Moreover, for the majority of merchants the tool suggests a risk-taking attitude by decreasing false positives, however visitors of the tool seem to generally increase the scores and hence being risk-averse.

3.3 Threats to Validity

Before offering an overview of this Chapter, it is essential to reflect on the limitations that might have an impact on the analysis conducted. Generally, there are three types of validity, namely construct, internal and external validity. Since these types are also going to be used in the subsequent Chapters, we briefly introduce what they stand for. Firstly, *construct validity* refers to the degree to which a research is eventually capturing what it claimed to be capturing (Brown, 1996). Secondly, *internal validity* reflects whether the analysis was correctly executed, i.e. in terms of minimizing bias and systematic errors (Reis & Judd, 2000). Finally, *external validity* is the degree to which the results of the study can be generalized to other situations (Aronson,

Wilson, Akert, & Fehr, 2007). Since the findings can only be discussed as a whole after the different research parts are executed in each Chapter, we leave external validity for Chapter 6 and discuss only construct and internal validity.

3.3.1 Construct Validity

In this Chapter we were interested in exploring merchants' behavior through the historical data of the Risk Calculator. The aim was to make a comparison between users and non-users of the tool and identify whether they behave differently. Nevertheless, the identification of the "users" was not easy; for an extended period of time we looked at the log files and no merchant pressed the tool's button "apply changes to live". This led us to the simplification of assuming that every merchant who visited the tool is a user. Moreover, in Adyen's system there are two levels of accounts: company level and merchant level. The brand is actually reflected in the company level, and hence a company might have multiple merchant accounts (depending on the countries that it operates or for reporting reasons). The Risk Calculator works on a merchant level, however at this point we only had information on the company name that seemed to have visited the tool. Therefore, in the analysis all the merchant accounts of a single company were taken into account for exploring the behavior. These two simplifications might have had an impact on the accuracy of results, nevertheless since we had no evidence of merchant actually applying the tool's suggestions, this was the best proxy we could use.

On the other hand, we also had to choose a sample of non-users in order to make the comparison with the users. As already mentioned Adyen has around 5000 customers, nonetheless for statistical reasons the sample sizes should be of about equal size. Thus, a random sample of 154 companies was chosen.

Finally, the data for analyzing the tool's suggestions had to be manually collected since they were not stored in the database. The process was quite time-consuming, thus a limited number of records was gathered; a larger sample might have yield different results.

3.3.2 Internal Validity

The analysis in this Chapter was conducted by means of descriptive statistics as well as comparison of means. The main variables examined included the industry and the size of merchants, their authorisation, risk refused and chargeback rates, whether they visited the tool or not and with

what frequency. It is possible that other variables that were not taken into account might explain more about merchants' behavior. Such variables could for example be the country where a merchant is predominantly active and whether the merchant absorbs the liability for chargeback costs. However, these variable are more complex to capture than they seem; a merchant might be active in multiple countries, even continents while the liability depends on the type of transaction (method of payment) as well as whether 3D secure is provided for the specific transaction. Although these confounding variables might introduce bias, given the context of tool and the purpose of the study we included as many of the relevant variables as possible in the research.

3.4 Conclusions

In this Chapter we explored the historical data of the tool in order to identify usage patterns. The results obtained were related to merchants' characteristics, their risk behavior as well as the tool's suggestions regarding financial gains and risk profiles. A summary comparing the characteristics of visitors and non-visitors is presented in Table 5.14.

Table 3.19: Summary of comparison between visitors and non-visitors of the tool.

	High Visitors	**Low Visitors**	**Non-users**
Sector	Retailers	Retailers	Retailers
Size	Medium-Big	Medium-Big	Small
No of merchant accounts (median)	4	3	1
Total transactions (median)	5267	1225.5	41
Average transaction value (median)	56.04	68.21	72.49
Authorization rate (median)	81%	83%	93%
Chargeback rate (median)	0.04%	0	0
Risk refusal rate (median)	1.7%	0.6%	0

As regards merchants characteristics, we found that the users of the tool appear to be similar in terms of industry to the non-users. The largest categories in both groups are Retailer merchants, followed by Services and Transportation industries (Figures 3.1a and 3.1b). Moreover, the users

of the tool appear to be larger companies since they have on average more merchant accounts and more transactions in volume than the non-users group (Tables 3.10 and 3.11). This is due to the fact that merchants have to pay for access to the tool, and usually the larger companies are the ones who pay. Moreover, the tool in order to be able to calculate an optimal solution, needs a certain volume of several thousands transactions, thus larger companies represent better candidates for the tool. Apart from that, larger companies might have more dedicated staff to look at the tool, such as for example data science teams. However, it should also be noticed that Adyen's platform has in general more smaller accounts than big ones.

On the other hand, we expect non-users to have lower chargeback and refusal rates than the users, since the potential users would logically be the ones facing a problem with chargebacks or refusals.The analysis indicates that non-users have indeed both lower chargeback rate and refusal rate than the users (Table 3.11). Interestingly enough, the median values of these rates appear to be zero, so this practically means that for a group of merchants the tool cannot add any value, as they do not seem to have any problems to solve. Note that there might be some bias in this point, since as already mentioned non-users are quite small companies, which might be the reason that they do not experience no fraud at all.

The two aforementioned findings indicate that there is some type of engagement with the tool and moreover the "good candidates" are the ones who are visiting the tool, or at least have seen it once. This is based on the fact that the tool works better when large amount of transaction data is supplied to it, and also its aim is to help merchants that face the problem of increased chargebacks and refusals. For the non-users, we see that they have practically no problem to solve. Linking this to literature, we can mention two concepts capturing this finding. According to **Status Quo Bias** in Behavioral Economics, people prefer not to change behavior unless the incentive to do so is strong; in this case, non-users seem to have already optimized transactions. Moreover, **Prospect Theory** suggests that it is important to understand the general context when making a decision; in our case it is important whether a merchant is already doing well in terms of refused and fraudulent transactions and hence the tool should take this into account.

The finding that no merchant clicks on the "apply" button of the Risk Calculator, made us explore whether they are making any changes through their settings page. We saw that both the visitors and non-visitors groups tweak the scores not through the tool, but through their risk settings. We expect that there would be a correlation between the number of times a merchant changes the risk settings and the refusal or chargeback rate; users are interested in changing the scores since in this way they want to reduce risks related to transactions. The correlation coefficient however indicates a "very weak" relation between chargeback rate and risk changes (0.010) and between

refusal rate and risk changes (0.044) for the users of the tool. This might indicate that the effect of changing the scores of the risk checks is not immediately noticed on chargeback and refusal rate, but might be obvious after several months. Another possible explanation might be that there is some kind of bias in the data regarding chargebacks and refusals, which nevertheless at this point could not be identified.

As regards the tool's suggestions, we expect that the projected savings of the users would be on average larger than those of the non-users (i.e the tool is creating financial incentives to use it). This hypothesis is verified, since the suggested savings for most of the users account for 0,5% of their revenues (Figure 3.9b), whereas for the non-users they account for 0,12% (Figure 3.10b). Nonetheless, both of the amounts are quite small compared to merchants' total revenues. According to **Security Economics**, people are not interested in using security tools when the time and effort they have to spend outweighs the benefits that the security advice brings (Herley, 2009).

Going a step further, we saw that for most of the users as well as the non-users, the suggestion of the tool is to decrease the number of false positives, i.e. accept more transactions. This is a suggestion to become more risk-taking. Looking at how visitors of the Risk Calculator changed their scores during this period, we saw that the majority increased the scores and hence kept a more strict, or in other words risk-averse, attitude. According to Behavioral Economics, people are not willing to undertake advices wich are not in line with their risk attitude (Kahneman, 2011). As such, a risk-averse merchant would tend to reject risk-taking advices.

An important observation when analyzing the tool's suggestions was that the total amount of money can be both positive and negative. Logic dictates that the tool's suggestion would result in a positive amount of money being saved, however we also see negative amounts i.e. losses for the merchant. This is counter-intuitive since, according to the tool the negative amount is an indication that the users should reduce their chargebacks and accept more false positives, otherwise they will have to pay fines to the card schemes (Visa, MasterCard). However, this negative amount might "scare away" potential users if not explained explicitly: the merchant might think that the tool is wrong since it displays losses, hence this would imply less trust to the underlying model of the tool. Thus, the tool has to either show that the merchant gets a fine or offer the option to ignore this and calculate another optimal solution. Moreover, The absence of good communication of the result increases the complexity of the tool and thus lowers again the incentives to use it. This is an aspect related to the user interface that can be improved in order to increase usability. We aim to explore this in more detail during the interviews with the merchants.

Finally, during this first step of the research it became obvious that an internal mechanism for monitoring is essential, since we had to collect data manually. This would practically allow the calculations of the tool to be stored into the database and later on to be analyzed. Without this data, it proves to be particularly hard not only to identify patterns, but also to come to robust conclusions.

Looking back at the research question of this Chapter, which was framed as *"What does the historical data indicate about the usage of the Risk Calculator and the risk settings adopted by merchants?"*, we can state that firstly, the merchants who are visiting the tool are the ones who meet the requirements for triggering the tool's calculations. Moreover, they seem to have higher chargebacks and refusals than the non-users and they mainly belong to retailers and information services industry. Although they visit the tool on average 6 times during a 3-month period, no one is actually applying the suggested changes to the risk scores. Nevertheless, merchants do change their scores through the settings page and the majority of them is increasing the scores thus denoting a risk-averse attitude. This is not in line with the suggestions of the Risk Calculator, which most of the times suggests to decrease the scores in order to reduce false positives. Additionally, the financial savings that the tools suggests account for a small fragment of merchants' revenues.

Chapter 4

User Interaction With Risk Management Tools

In the previous chapter, we explored merchants' behavior in terms of data analysis. Particularly, we identified the properties of the merchants that both visit and do not visit the tool and we compared them. During this chapter we made several assumptions about the validity of the tool, the possible alternative mechanisms that merchants might deploy for monitoring risks, and the way people use the Risk Calculator. In this chapter, we try to give an answer to these assumptions by means of interviewing the merchants and thus exploring how they engage with risk management and what is their feedback about the tool. Furthermore, we compare those findings with internal people's opinion about the functionality and the added value of the tool, in order to contrast companies' to merchants' incentives. To recall, this chapter aims to provide an answer to the following research question:

SQ: *"How do merchants and account managers explain their engagement with the tool and how do they choose their risk profile?"*

The answer to this question is expected to lead to identification of the different stakeholders' needs, as well as shed light in the reasons that make merchants reluctant to use fraud tools. Below, the qualitative methodology followed in this chapter is described, while the findings are also depicted and explained. The results produced by a qualitative research are in principle descriptive than predictive, and the findings can often be generalized to populations with similar characteristics.

4.1 Methodology

This chapter is focused on qualitative analysis, in contrast to the previous chapter which followed a quantitative approach. Qualitative research is usually used in social sciences when there is a need to delve deeper into human behavior and identify motivations that explain a particular stance (Merriam, 1998). This information can be obtained by talking to relevant people and by determining patterns about how they make decisions.

4.1.1 Interview Design

The aforementioned research question is hence answered in a qualitative way, by means of interviews. More specifically, *semi-structured interviews* with merchants who are both users and non-users of the tool were conducted in order to explore their incentives. Apart from that, interviews with experts within Adyen were also conducted, in order to compare different perspectives on the problem. This group included account managers, risk officers, as well as developers who are familiar with the specific tool and its underlying concepts.
Semi-structured interviews consist of a set of questions that the researcher may or may not adhere to. This is mainly due to the reason that they serve as a free and open conversation that can vary between participants, as the boundaries to the topics that should be covered are relatively loose (Miles & Gilbert, 2005). The reason for using this type of interviewing is primarily because we want to explore *why* people behave in a certain way, as well as identify their perceptions and opinions about risk management in general and the Risk Calculator in particular. For the purpose of our research we chose the *purposive sampling* (Ritchie, Lewis, Nicholls, & Ormston, 2013), i.e. the selection of candidates according to preselected criteria relevant to the research question posed in the beginning of this chapter.

Each interview with the merchants lasted approximately 30 minutes and the questions aimed at revealing characteristics on one hand about the attitude of the merchants towards risk and on the other hand about the shortcomings of the tool. Thus, basic questions about risk management were asked to all the merchants who participated in the interviews; then, the questions were differentiated depending on whether the merchants seemed to be using the tool or not. The interview protocol with all the questions can be found in Appendix A. Please note that as the interviews were semi-structured, sometimes questions present at the protocol were omitted or new questions were added, depending on the interviewee's answers and stance. Similarly, the interviews with the developers and account managers lasted around 30 minutes too. Specifically,

two developers with in-depth knowledge on the functionality of the tool were interviewed, while four account managers and a risk officer further contributed in the qualitative part of the research. The approach followed here was more like an open discussion, with the leading question focusing around the respondent's opinion about the tool. The respective protocol can be found in Appendix B.

As a last step, the interviews were transcribed and inserted into the software *atlas.ti* in order to be analyzed in terms of open coding.The notion open coding refers to assigning labels to concepts by defining categories (Glaser & Strauss, 1967). Therefore, we read several times the transcripts and thought about interesting themes that can be applied to the interviewees' sayings. Subsequently, similarities and differences between the various transcripts were identified and collected under the main codes. The most important codes are described under Section 5.2.

4.1.2 Respondents Characteristics

In order to identify the most suitable candidates for the interviews, at first place the merchants were segmented into groups in order to be categorized according to their characteristics. This segmentation was done by creating a contingency table, based on four variables: the number of times that the merchants visited the tool, the number of times the merchants change their risk settings, their refusal rate and chargeback rate. The four aforementioned variables were turned into categorical by taking the median value for each of them. More specifically:

• The median value of the number of visits was 2, therefore merchants with visits>2 were given the value "high", those with visits<=2 were given the value "low" and finally non-users were given the value "zero".

• The median value of risk changes was 1.6, therefore merchants with number of changes>1.6 were given the value "high", while merchants with number of changes<=1.6 were given the value "low".

• The median value of the refusal rate was 0.0339, therefore merchants with refusal rate > 3.39% were given the value "high" and merchants with refusal rate<=3.39% were given the value "low".

• The chargeback rates were generally on a low level, as already mentioned in the Data Analysis chapter, thus the values that were given to this variable were "zero" and "non-zero".

The contingency table with the merchants can be seen below in Figure 4.1.

Table 4.1: Contingency table for selecting merchants to interview. The blue color corresponds to the groups of participants that were eventually interviewed, while the black color to the non-participants.

			High visits	**Low visits**	**No visits**
Risk Changes	**Refusal rate**	**Chargeback rate**			
high	high	non-zero	22	24	3
		zero	1	5	0
	low	non-zero	11	7	2
		zero	4	3	3
low	high	non-zero	13	16	0
		zero	1	2	0
	low	non-zero	12	22	7
		zero	0	10	24

The concept behind this segmentation was to pick a group of merchants as diverse as possible, that can provide with feedback on different aspects. As such, users and visitors of the tool can share their experience with the Risk Calculator and make comments about its validity. On the other hand, non-users might share the reasons they do not use the tool and which other mechanisms they have for doing their risk assessment. Furthermore, merchants with both high and low refusals or chargebacks can indicate whether the mechanisms they use are effective or not. Finally, merchants that change frequently their risk settings might provide reasons why they do not do that through the tool. Ideally, talking to at least two merchants of each category would enable us get very deep insights, however due to time and resource constraints this was not feasible. Hence, four categories were considered the most crucial:

- Users with frequent risk changes
- Non-users with frequent risk changes
- Users with infrequent changes
- Non-users with infrequent changes

The reasoning behind selecting these four categories is the following. The first group can provide us feedback about the interaction with the tool, while the second group can indicate the features missing from the tool that make them change the risk scores directly through the settings and not through the Risk Calculator. On the other hand, the third group can provide reasoning behind the frequency of their changes, whilst the last group might indicate if they use alternative mechanisms

to monitor their transactions and whether these mechanisms are more effective.

When sending the interview invitations, we made sure that at least one merchant from the aforementioned groups was contacted. Merchants from the remaining categories were also asked to participate in order to have a heterogeneous sample. This resulted in sending invitations to 40 different merchants, from which 11 responded positively and expressed their willingness to schedule the interview.

Out of the 11 merchants that were interviewed, three belong to the first group, i.e. users with chargebacks, high refusals and frequent changes. Moreover, two merchants belong to the group with chargebacks, high refusals and infrequent risk changes. From the remaining 6 groups, one merchant per group was interviewed. The procedure for selecting the merchants was the following; from the list with the names of the candidates, 40 that met the criteria mentioned in the contingency table 4.1 were picked randomly. Subsequently, the account managers responsible for the selected merchants were contacted and asked to reach out to the merchants.

The participants derived from the Retailers industry and differed in size, with the smallest one processing 47.087 transactions per year, and the largest one 42.200.522 per year. The specific industry sector of the respondents, as well as the position of the interviewees and the years of their experience, are depicted in table 4.2.

Table 4.2: Characteristics of the interviewees.

Participant	**Industry**	**Interviewee's Position**	**Experience**
P1	Shoe Store	Global Risk Manager	1.5 years
P2	Shoe Store	Customer Success Manager	2 years
P3	Ticket Agency	Global Expansion Coordinator	NA
P4	Ticket Agency	Customer Service	NA
P5	Automobile Parts	Head of Payments & Fraud Prevention	6 months
P6	E-commerce Marketplace	Regional Fraud Manager	5 years
P7	Audio Equipment	Global E-commerce Analytics Manager	3 years
P8	Property Rentals	Head of Payments	4 years
P9	Chauffeur Services	Head of Business Development	2 years
P10	Online Dating	Fraud Prevention Manager	1 year
P11	International Calls	Fraud Manager	3 years

4.2 Findings on Interviews

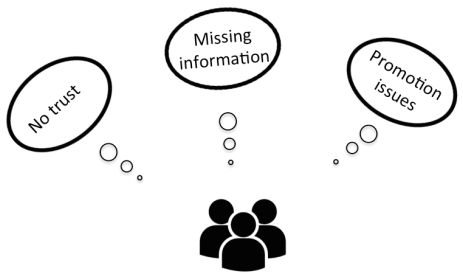
This section discusses the findings of the interviews with the merchants, as well as with the developers and account managers within Adyen. A comparison between them is subsequently performed in order to identify whether opinions converge or differ significantly. After the comparison, the most promising changes to be implemented to the tool are identified.

The insights obtained from the interviews with the merchants are divided in the following categories. The first category includes the findings that are related to the behavior and attitude of merchants towards risk management, while the second category includes all the challenges and shortcomings related to the Risk Engine Optimizer. Lastly, the third category identifies the suggestions for improvements to the tool that the merchants made.

4.2.1 Respondents' Feedback on the Tool

Regarding the shortcomings of the tool, we identified three broad challenges that merchants most frequently face when it comes to the usage of the Risk Engine Optimizer. The identification of those categories is related to the open coding process in Atlas.ti which resulted in 23 codes, aiming to answer questions related to merchants' behavior and the specific perceptions about the tool.Those challenges are illustrated below in Figure 4.1.

Figure 4.1: Main challenges regarding the Risk Calculator according to the interviewees.



There is no trust in the underlying model. This seems to be one of the most prevalent challenges, as 7 out of the 11 respondents mentioned several problems they experience or that they have noticed when trying to use the tool. More specifically, on one hand 3 merchants noticed that

there are some bugs in the tool such as (1) no change in the simulated results when a score in the risk rules is being changed, (2) the percentages in the statistics area sometimes exceed 100% and (3) the signs corresponding to the amount of money being saved are not accurate. Regarding the latter point, the tool shows that the merchant is going to lose money if the chargebacks are fewer, while it should be the other way around.

Apart from the bugs, some merchants do not seem to trust the underlying algorithm.As an argument, 3 merchants indicated that the tool is solely comparing the scores for the risk checks, while they would be interested in having a tool that would show the impact when they change the actual value of the risk check.

Furthermore one merchant perceived as drawback the fact that the calculations are made based on historical data, since the past does not always have to predict the future. Another merchant pinpointed to the fact the the tool suggests to decrease the score of rules which are proven to be very efficient in combating fraud, whereas respondent P4 argued that:

"[...] I do not think [it] is very reliable...The numbers it shows in the amount being saved seem extremely high to me."

Another issue that seems to lower the trust of merchants in the tool is the fact that the slider cannot always be moved. The explanation that the Simplex algorithm finds only one feasible solution is not sufficient, especially for merchants with non-technical background, resulting in them thinking that the tool is just not doing what it is supposed to do.

Going a step further, all of the respondents indicated that they wish to use the tool as an adviser and not as a risk-setter, since they want to have control of the risk decisions to be made within their company and not trust everything to a machine. This remark tells us a great deal about the context in which the tool should be embedded into.

Lastly, 3 merchants suggested to be able to perform A/B testing through the tool, i.e. sending part of the traffic to the suggested risk settings and directing another part to different settings. This suggestion also reflects a trust issue, since in this way merchants would like to double-check that the suggestions of the tool are indeed better than their current configurations.

There is need for additional information in order to understand how to properly use the tool. In 5 out of the 11 cases, merchants mentioned that they find difficult to understand how the tool works. The fact that there is no available documentation makes them spend considerable

amount of time in trying to figure out what the functionality of the tool is. Most of them however give up and prefer not to use it.

Moreover, six merchants indicated that they find difficult to interpret the results of the tool.Those merchants did not seem to grasp what the tool aims to do by just navigating into the page. According to participant P9:

"I'm actually not using it because it's not really clear, for me at least. [...] because what I understand is that you can basically put up a comparison between the current settings and then make predictions based on the past on how the - on how the new settings would change your current stats, but how exactly it works I have not a clear sense to be honest."

Additionally, two merchants completely ignored the "slider button" area and were just looking at the statistics of how many risk checks triggered for their transactions.This pinpoints to the fact that the flow of information does not guide the user through the page, making diffuse the purpose of the tool.

Another remark made by a merchant was that it is not clear whether the changes suggested by the tool are directly applied to the live account or not. Hence, the above comments indicate that there is need for more information to be integrated into the page in order to make it more self-explanatory.

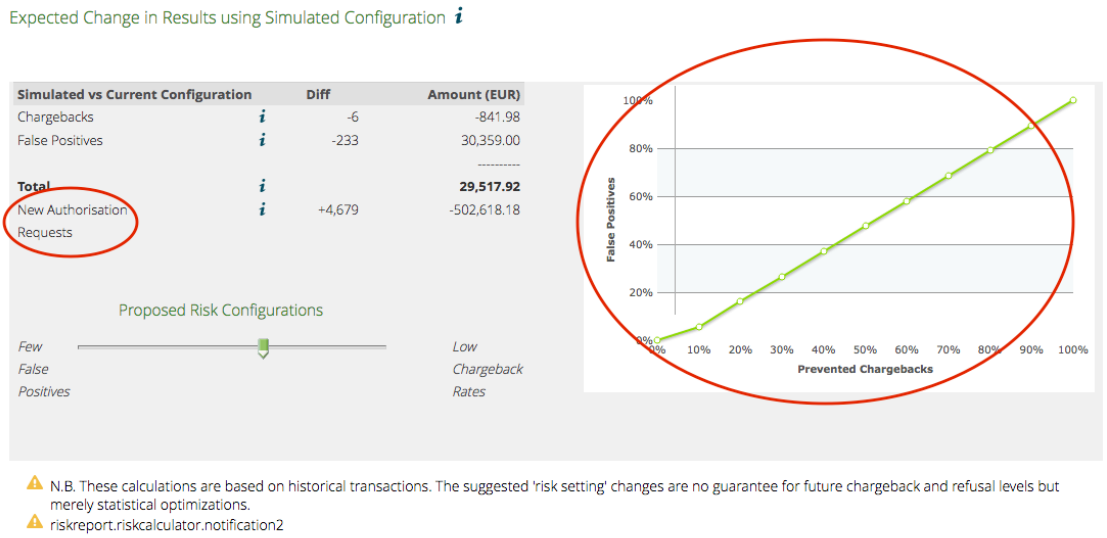
The promotion of the tool is not sufficient. Throughout the interviews it derived that one of the main problems related to the usage of the tool is the promotion from the appropriate people within the company. Especially in five cases, it was obvious that account managers deliberately urged merchants to not use the tool, although merchants were absolutely inclined to. This issue caused sometimes frustration to the merchants, as indicated by one of the interviewees

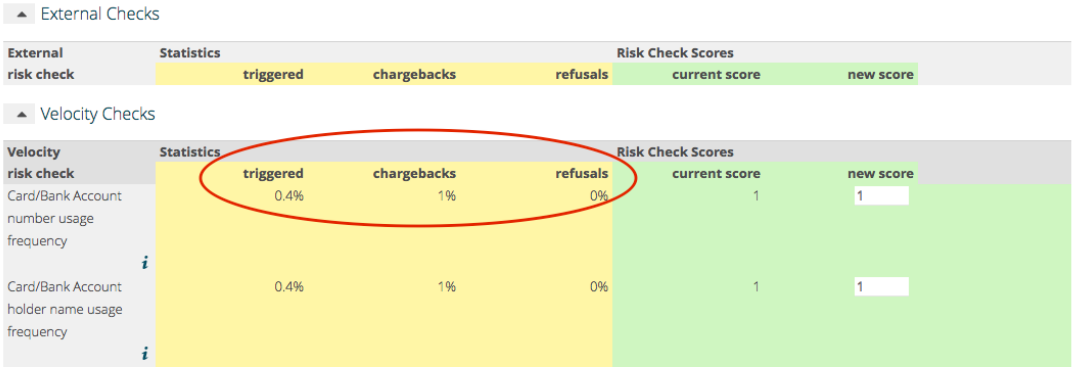
"The reason for not using it [...] is not knowing that it exists. So, for how long has this tool been in the reports page?"

Furthermore, talking to account managers was deemed crucial, since they facilitate merchants with risk management decisions and are responsible for promoting such tools. The four account managers that were interviewed separately, expressed the following concerns about the Risk Calculator.

The complexity of understanding how to use the tool is the main reason for not actively engaging with it. According to the respondents, there is no good introduction for the tool within the company and hence they try to figure out by themselves how to interpret the results. However, this proves to be time-consuming for them, since several terms are confusing and there is little explanation offered within the page of the tool. Below in Figure 4.2 the information that is not clear to the account managers can be seen.

Figure 4.2: Parts of the tool that account managers cannot interpret.





Regarding the term "New Authorization Requests" there was quite some confusion on how is it calculated and how is it different from the term "false positives". Apart from that, the respondents mentioned that they do not know how to interpret the graph and were confident that most of the merchants cannot either. Finally, two of the account managers indicated that they do not know what the statistics "triggered", "chargebacks" and " refusals" refer to.

Account managers do not trust the underlying model of the tool. This mainly derives from the fact that before being involved with the tool, the account managers ask Risk Officers' opinion who do not recommend its usage. Thus, influenced by this negative stance, none of the account managers is willing to take the risk of suggesting a tool to the merchants that does not seem to be working properly.

Account managers experienced different kinds of technical problems when trying to use the tool. More specifically, three out of the four account managers indicated that they perceive as a problem the fact that the slider is not moving for all merchant accounts. Especially for people with non-technical background, the explanation that the Simplex algorithm cannot find a "feasible solution" is not comprehensive enough, leading to the thought that the tool is just not working as it should. Furthermore, one respondent mentioned that the tool does not seem to take into account the changes in the risk settings per se. As an example, she mentioned that when she set the risk check regarding the email usage frequency to trigger when it reached 4 times per day and then the next month she changed the value to 3 in one hour, the tool did not take that into account.

Following the account managers' concerns about the reliability of the underlying model, a discussion with the Risk Officer within the company took place.

According to the risk officer the suggestions of the tool on the new risk scores seem too extreme. For example, the interviewee mentioned that the Risk Calculator suggests to change the score of a risk check from 100 to zero. Such dramatic changes, according to his experience, can never be implemented at once since it is more likely to have the exact opposite impact. Apart from that, the tool usually does not suggest to make any changes at all to sets of risk rules that are proven to be very effective in combating fraud. As such, the Risk Officer prefers to manually review merchants' accounts and give personalised advices to each of them. Nevertheless, the respondent expressed the eagerness to use actively this tool when it is improved in terms of the algorithm it uses, as it will make his everyday job easier.

The most prevalent challenges of the tool seem to fall both in the technical and non-technical sphere. Regarding the technical issues, need for additional variables in the tool as well as usability issues are the most crucial challenges. On the other hand, no trust in the underlying model, poor documentation and insufficient promotion pose further obstacles to the tool's usage.

4.2.2 Company's Feedback on the Tool

The next step was to compare internal opinions about the tool with those of merchants. As such, two developers were interviewed at first place. The discussion was mainly centered around the technical aspects of the tool. The respondents characteristics can be found in Table 4.3.

Table 4.3: Characteristics of the developers

Developer	**Position**	**Experience**
D1	Data Analyst	1 year
D2	Front End Developer	1.5 years

According to the developers the underlying model is mathematically accurate. Regarding the technical aspects, both of developers seemed confident about the underlying model, by highlighting that the suggestions are mathematically correct. Hence, according to the respondents, when a user follows the suggestions of the tool the expected results will be verified in a time horizons of around 3 to 4 months. As D1 mentioned:

"I think mathematically it makes sense... Based on historical data it just tries to optimize the settings, so the algorithm is accurate."

There are some minor bugs that need to be fixed. These bugs are mainly the opposite signs in the prevented chargeback amount and count, as well as the statistics displayed beside the risk checks. An illustration is shown in Figures 4.3 and 4.4. More specifically, when the amount of the chargebacks is negative this means that the merchant is avoiding those chargebacks, thus is saving money. Similarly, when the sign of the false positives is negative this means that the merchant is reducing the refusal of legitimate transactions, hence is gaining more money. On the other hand, the statistics displayed in the page denote the percentage of the transactions that triggered the specific risk scores, as well as the percentage of them that resulted in chargebacks or refusals. Hence, this amount cannot exceed 100 as it symbolizes percentages.

Figure 4.3: Bug in signs displaying count of chargebacks and amount in euros.

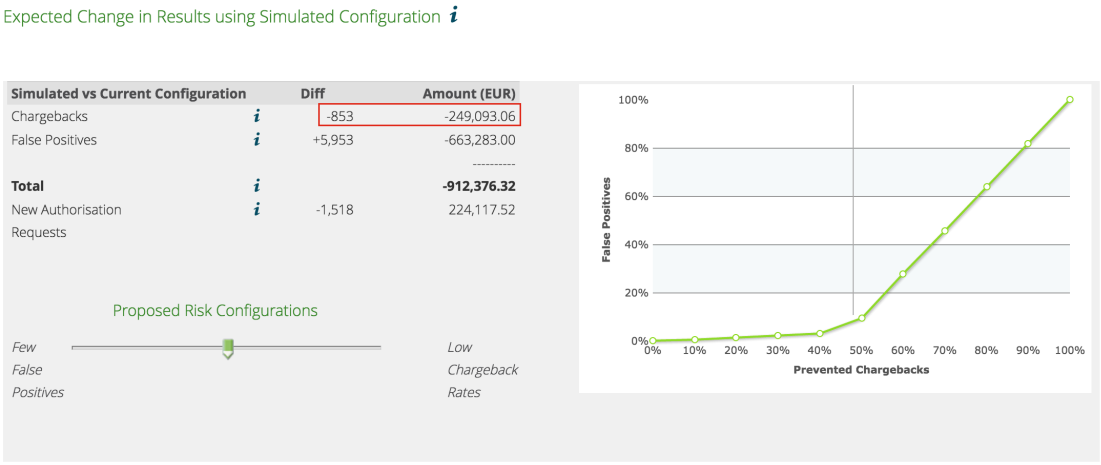


Figure 4.4: Bug in percentage of statistics.

ShopperDNA risk check	Statistics triggered	chargebacks	refusals	Risk Check Scores current score	new score
Different Countries used by Shopper	0.0%	127%	0%	30	30
Issuer Blocked (Previous) Card used by Shopper	0.0%	34%	0%	80	100

The tool displays the historical results without taking into account the settings of the risk parameters as set at a specific time. The fact that the tool is solely score-based is a limitation inherent to the algorithm itself. As we saw previously, this constraint was also mentioned by two merchants. For instance, the risk check "Card/Bank Account number usage frequency" verifies how many times a credit card or bank account has been used. It is obvious that in a situation where the same credit card has been used 8 times within 3 hours denotes a suspicious behavior; in this case it would be useful if the user could indicate through the tool what is the preferred action in such situation, e.g. reject the transaction directly. Currently the optimization occurs merely based on current scores given to the risk checks. According to the developer D1:

"It just lumps together historical risk results regardless of the way the checks were parametrized at the time. For example if in May a check triggered when more than 3 cards were used in one hour and in June this was changed to 5 cards in a day, the risk calculator won't know about this... It will just see that the check triggered, but not with what settings."

There is no possibility to know the final state of a previously blocked transaction. In the simulated result of the tool there is a variable that denotes the new authorisation requests, i.e. the transactions that were never sent to the issuer with the merchants' current risk configurations, but that are going to be send with the suggested configurations. However this is practically a "blind spot" since we can never know whether these transactions will eventually be authorised, refused or chargeback. The participant mentioned that he could not think of an alternative solution for resolving this challenge hence for the moment this can only be indicated as a limitation in the tool's suggestions.

"The tool can not know what would have happened if a transaction we blocked would have passed. It can suggest a configuration that would have allowed some of the blocked traffic to go through, but we don't know what would have happened with that traffic how much of it would have been bad. Yeah...And I don't know of a good way to solve this problem."

Usability most probably affects the willingness of using the tool. The participants expressed that the main problem with the tool is usability and therefore the most promising changes would be on the User Interface. One of the developers added the extra insight that one major problem is the fact that the Simplex algorithm cannot always find a feasible solution, resulting in no possibility for the slider to be moved. However, there is no message displayed in the screen that indicates that the algorithm could not find an optimal solution, thus resulting in users wrongly believe that the tool is not working as it should. Solving this problem could mean either adding a pop-up message, or changing the algorithm or alternatively not offering the tool to the merchants that do not have many opportunities for optimizations.

4.2.3 Merchants Attitude Towards Risk

In this section we aim to capture merchants' attitude towards risk. Crucial for exploring merchants' behavior are the notions of risk-aversion and risk-seeking. These concepts denote people's level of willingness to undertake risks, usually in exchange of some rewards, such as profits. We are also interested in knowing how involved with risk management the respondents are and what is their perception about it. Table 4.4 depicts the findings, with the first column indicating the percent of the interviewees that is related to the respective finding.

Table 4.4: Findings on merchants' attitude.

Percent	Findings
100%	Monitor risks related to transactions through manual review
100%	Prefer to always do some manual review/ use the tool as an adviser than as risk-setter
82%	Try to balance chargebacks and conversion when taking risk decisions
63%	Highly involved with risk management (4-6 hours daily)
27%	Trust more human than machine decisions
18%	Believe that risk mitigation is not in PSP's area of expertise
18%	Entered the Excessive Chargeback Program imposed by card schemes

All merchants praise the importance of reviewing their transactions manually. One of the most interesting findings while talking to the merchants was that they all perform some kind of manual analysis in order to monitor their chargebacks and refusals. Most of the respondents mentioned that they use Excel where they download their transactions and perform trend analysis by trying to identify some suspicious patterns. Although that the majority of the respondents mentioned that this is a time-consuming task, they also held the belief that it is inevitable to avoid it, since the risk decisions within a company are too important parameters. As participant P1 said:

"I would like to have some tool that works. But I will always have some manual analysis and control by myself. It is very tough to let everything on the tool. We can have some problems, so it's always good to have manual analysis. I think using both a tool and own analysis would be the best."

Other merchants also confirmed this statement by mentioning that they would not trust everything to a machine, as experienced humans can usually evaluate better a situation.

Apart from performing manual analyses, other mechanisms that merchants deploy for monitoring risks include getting advice from their account managers or using external anti-fraud tools. Regarding the first point, 4 merchants indicated that they rely to their account managers when it comes to handling chargeback and refusal issues. Participant P11 mentioned:

"We are also in good hands with our account manager checking our chargeback level [...] So, we rely on you telling us if there is anything wrong when it comes to chargebacks that we might do not notice."

Additionally, there were 2 interviewees who use an external anti-fraud tool, while one respondent also expressed her thought to deploy such a tool in the near future.Out of these 3 merchants, one mentioned that the reason for using an external tool is that they have different PSPs. However, the remaining 2 merchants implied that risk mitigation is according to their opinion not in PSPs' area of expertise. Accordingly, he merchants expressed the following beliefs:

"[...] because of course you are a PSP... This interface is to help us as a merchant, but maybe you don't have all the information or background of machine learning, of the malware or IP."

"and we still are convinced that Adyen is as a Payment Service Provider - I mean managing all of the aspects of the payments process - so it can't be specific only for fraud prevention and that's why we also decided to in parallel have another tool."

All the merchants that visit the tool use it as an adviser rather than as a risk-setter. During the interviews it emerged that the merchants who frequently visit the tool, do not actually apply the suggestions it makes but they solely look at the numbers it displays. This finding denotes that merchants are highly interested in having control of the risk settings for their transactions and moreover to understand how the risk rules work. This is reflected in participant's P5 quote:

"We use it as a recommendation, but apply our own risk settings."

The majority of merchants exhibits a risk-averse attitude. Regarding the risk bahavior merchants adopt, we see that the majority of the respondents holds a risk-averse attitude. This derives from the fact that on one hand most merchants mentioned that they always try to achieve a balance between chargebacks and conversion. When asked where would they give more weight if they had to choose between these two variables, the interviewees indicated that they would be willing to accept some risk only if chargeback rate is already low or if the payoff was guaranteed. For example, merchant P6 explained that:

"We would look at both but we would probably see that we could accept a few hundred more chargebacks if it means our revenue can increase by several thousand euros. So we kind of look to both together."

Only two merchants can be described as risk-seeking and this derives from the fact that they entered the Excessive Chargeback Program, imposed by the card schemes. The respective average chargeback rate of these merchants was 2.14% and 1.71% respectively, while the threshold defined by the card schemes is 1% or more chargebacks-to-sales count ratio and 100 or more chargebacks. Thus it can be deduced that these merchants were more accepting risk in order to increase their conversion rates. On the other hand, the majority of the merchants interviewed was highly involved with risk management within their company by spending on average 4-6 hours per day of doing risk assessment. This active monitoring of risks reflects a risk-averse attitude as well, since merchants seem to constantly try to mitigate hazards related to their transactions. On the contrary, 3 of the respondents engage in low levels of risk management within their company, with an average of 1 hour per week. We found that this is dependent on the size of the company, as the ones that engaged in lower levels of risk management are the smallest companies of the sample. Nevertheless, only one of this latter group of merchants has on average low refusals and almost zero chargebacks.

Limited resources and time constraints are the main drivers for using the tool. The merchants that seemed more keen to use the tool were the ones that reported to have limited resources for doing the manual analysis or the ones that mentioned that manual review is too time consuming for them and thus they would like to have an auxiliary tool in this process. This observation is helpful in defining a target group of users and later promote the tool to them. Specifically, participant P10 mentioned:

"Yes, 'cause the problem is that we don't have that many resources...I believe that this tool can be really helpful."

Most of the merchants claim that they do not change their risk settings frequently. The main reason is a "wait-and-see" attitude, i.e. making a change and waiting some months to evaluate whether the results were positive or not.The respondents indicated that the main reason for changing the risk settings is either an increase in chargebacks or a decrease in conversion. For instance, according to P7:

"[...] The way we do it is we make a change and then we wait - sometimes if something goes really bad we might consider to review earlier, but we usually wait about a month because it takes so long for fraud to show up, but we do not make a change each week cause we won't see the change."

Merchants seem to exhibit a risk-averse attitude by trying to balance chargebacks and conversion. This attitude implies that they would forgo profit opportunities in order not to take the risk and hence would reject configurations that advise them to be more risk-taking.

4.3 Discussion

After talking to different stakeholders about the Risk Calculator, we are now able to identify some further changes that can have a positive impact in the acceptance of this anti-fraud tool. The suggestions are focused on how the company might improve its internal processes in order to facilitate communication between the employees and better promote the tool.It should be noted at this point however that the list of suggestions is not exhaustive.

4.3.1 Improvements in Company's Processes

It derives from the interviews conducted internally in the company that the opinions about the tool among different specialists do not totally converge. Although developers pinpoint that the calculations are accurate, the Risk Officer and the account managers do not trust the tool's suggestions. This indicates a non-efficient coordination within the firm as there is no integration of knowledge between the members of the company (Grant, 1996). This means that the common knowledge, i.e. the intersection of the individual knowledge sets of the organizational members, needs to be enhanced. In such a way, account managers and risk officers should obtain basic information related to the developers' job in order to be able to fully understand how the tool functions and trust the model. As Grant (1996) mentions, *"production requires the coordinated efforts of individual specialists who possess many different types of knowledge"*, otherwise the integration of the same knowledge provides zero added value. It should be noted however that if the individuals have totally different knowledge bases, the integration can only happen in a primitive level. Hence, communicating this knowledge between the employees or investing in

training can positively influence the effectiveness of internal rules and directives, as well as align in a more concrete way the end goals of the employees.

Furthermore, by comparing the answers of different respondents we also notice a misalignment of incentives. On one hand, developers are interested in company's infrastructure expansion, thus the tool is an additional product that they can work on. On the other hand, account managers and risk officers are more interested in potential loss of their reputation, since if the tool does not eventually work as expected, the clients might no longer trust them. Adding into this the merchants and looking at the problem from a societal point of view, we can see that not only direct costs, but also indirect and implicit costs matter to the stakeholders (Bauer & Van Eeten, 2009). The following Table depicts the stakeholders as well as their respective considerations about costs that might influence their decision-making choices.

Table 4.5: Costs incurred by the different stakeholders that determine their alignment of incentives

Stakeholder	**Cost Incurring**	**Type of Cost**
Adyen	Cost of customer support	direct
	Brand damage	indirect
	Cost of product expansion	direct
	Legal provisions that shield PSPs	direct
	Cost of security measures	indirect
Merchants	Losses due to fraud	direct
	Costs of preventive measures	indirect
Developers	Cost of software development & testing	direct
Account Managers	Loss of reputation	indirect
Risk Officers	Loss of reputation	indirect

4.4 Threats to Validity

After presenting the results of the qualitative part of this research, it is essential to moreover discuss all the constraints that might have posed a threat to the validity of the findings and their interpretation. Constraints might be challenging for a research, especially in qualitative studies, since the measurements are not as objective as in quantitative research. For instance there are no formulas or quantifiable measures as a point of reference and the interpretation is done based on intuition. however we are confident that we tackled them as efficiently as possible.

4.4.1 Construct Validity

The construct validity in this Chapter is highly dependent on the way we chose our research instruments for interviewing people. As already described earlier, the selection of the candidates to be interviewed was made based on the four criteria depicted in table 4.1. It should again be noted at this point that the absence of an internal monitoring mechanism for the tool might have affected the extent of the results validity. Having directly the calculations of the tool stored in a database, we might be able to choose interviewees based on their projected savings or on the number of opportunities for optimization that the tool suggests; choosing such different criteria, might have produced different findings.

Nonetheless, these four criteria were deemed as the most crucial for the following reasons. Firstly, the number of visits enables us to get diverse feedback on the reasons for using or not the tool. Secondly, the context of the tool implies that risk decisions are taken by adjusting the level of chargebacks against the level of refusals (Wolters, 2012); hence we expect to see a difference in the behavior of these groups. Thirdly, high frequency in risk changes denotes that people are willing to be involved with risk assessment, but probably not through the tool. Understanding the reasons behind this can give us insights on what features the tool should enable. Thus, we are confident that to a high extent we captured valuable information from these different groups.

4.4.2 Internal Validity

Regarding the internal validity, we saw in this chapter that 11 merchants were interviewed. This sample size, if different, might have affected the results regarding merchants' attitude and feedback about the tool.Nevertheless, for the purpose of this research this number of interviews is considered to be adequate based on the following facts. On one hand, research as conducted by Guest, Bunce, and Johnson (2006) suggests that a number up to 6-12 interviews is enough to get in-depth and diverse insights, since according to the authors, data saturation occurs by the twelfth interview. The term saturation is used to denote the point where no new information is added to the research and hence the researcher reaches the stage of diminishing returns (Bonde, 2013).As can be also seen in the Findings section of this chapter, many recurring patterns where identified in the interviews with the merchants, thus it was considered that the saturation point had been reached. Additionally, Baker and Edwards (2013) advise that in a purely qualitative research, a number of interviewees falling within the range of 12-60 should be selected. Since our research follows both a quantitative and qualitative methodology, the number of 11 interviews is viewed

as satisfactory.

Secondly, the scope of the investigation is narrow, and specifically aims at understanding the aspects related to the appeal of an existing product. According to Bonde (2013), projects with narrow scope require less interviews than projects with broad scope in order to fully understand the underlying phenomena.

Thirdly, an important factor for the qualitative part of this project was the heterogeneity of the groups, in order to get feedback on different aspects. As such, the interviewees derived from 8 different groups as can be seen in table 4.1, where the groups are denoted with blue color. The diversity of the respondents thus in terms of frequency in risk changes, refusal rates, chargeback rates and number of visits to the tool was significant. Therefore, the participants were quite diversified and considered to represent a good sample.

Moreover, it is mentioned in Romney, Weller, and Batchelder (1986), that the need to interview a large sample of participants decreases with the level of expertise of the respondents in relation to the topic of inquiry. In our case, the interviewees held positions such as Risk Managers, Fraud Managers, Head of Payments, and Data Analysts, hence the level of their knowledge relative to the topic is considered to be high.

Lastly, the resources for this project are definite and not unlimited; particularly, the time constraint to be met is one academic semester, which equals to six months for gathering the information, analyzing it and writing the report. Furthermore, qualitative research is frequently entitled to high levels of non-response, which is reflected in the fact that approximately 25% of the participants that were reached out were willing to be interviewed.

As regards the concreteness of the analysis, the examination of the transcripts in terms of open coding was conducted only by one researcher. We can suppose that if more scientists worked on this analysis, might have resulted in slightly different codes being used for the interpretation of the respondents' answers. Due to confidentiality reasons, the transcripts cannot be made public, however they are available within Adyen if future researchers wish to re-evaluate them.

4.5 Conclusions

Before providing a summary of the results obtained through this qualitative research, we briefly compare the insights with the quantitative analysis. Under Chapter 3, several assumptions were

made with regards to the merchants' behavior and the usage of Adyen's tool. The qualitative part shed light in whether those assumptions hold true or not.

Firstly, the low refusal and chargeback rates of both users and non-users do not seem to be dependent on the usage of the tool; the assumptions that the low rates are the result of other mechanisms that the companies deploy, as well as that these companies are risk-aware and double check with the tool, are verified since all the respondents mentioned they are performing their own analysis (see Section 4.2.3 for more details).

Secondly, most of the merchants seem to be risk-averse by monitoring very closely the risks related to their transactions as well as by indicating that they always try to achieve a balance between chargebacks and conversion.

Thirdly, the fact that the chargeback and refusal rates of the high visitors do not differ from the non-visitors is mainly explained from the fact that most of the "high visitors" turned out to not actively use the tool, thus the level of these rates is dependent on other risk management mechanisms. All the participants perform a manual analysis of their transactions, with the majority using Excel, hence by identifying patterns they adjust the risk settings accordingly. However, the manual analysis is deemed as "not easy" by 63% of the respondents, since it is labor-intensive and time-consuming. They nevertheless prefer it, since they consider it as an integrated part of their daily job, while 27% of the merchants mention that to some extent manual review is required as machines cannot make better decisions than humans. This finding is aligned with the concept of **Loss Aversion**, which suggests that the psychological cost of loss is greater than the psychological benefit of gain, thus people prefer to put more energy (manual review vs. automated tool) in order to be sure they avoid losses. (Kahneman, 2011).

Additionally, the merchants that indicated capacity limits in doing risk assessment were the ones that expressed the highest willingness to use the tool in an effort to automate the analyses on their transactions. Many of the merchants explicitly said that they find the tool a good approach to automate tasks, however the time and effort it requires in order to understand how it works pose limits to using it. According to **Bounded Rationality Theory**, human decisions are often non-optimal due to constraints in time and knowledge capabilities (Simon, 1982).

One of the main findings of the interviews is that merchants are highly interested in conversion and they rate it equally important, or sometimes higher than fraud mitigation. Therefore the assumption that they change their risk parameters according to this two variables is verified. Furthermore, out of the 11 interviews it was verified only in one case that merchants are confused

by the losses in revenues displayed by the tool. However, no generalizations can be made out of this since most of the interviewees have not actively tried to use the tool.

Finally, the observation that the median value of zero chargebacks and refusals for non-users is mainly due to the fact that these companies recently started using the Adyen platform, thus the default value of zero was given for the period that the respective data was unknown. Hence, the assumption that there is some kind of bias in the sample is verified.

Concluding, the aim of this Chapter was to explore both merchants' and the company's opinion about the Risk Calculator tool. Regarding the feedback about the tool, the main findings are summarized below in table 4.6.

Table 4.6: Summary table of the main issues regarding the Risk Calculator, the stakeholders that expressed this opinion and the potential improvements to the tool.

Stakeholders	**Main Issues**	**Potential Improvements**
Merchants, Account managers	No trust	A/B testing through the tool, Implementation of another algorithm
Merchants, Developers, Account Managers	Slider not moving	Alteration of algorithm's constraints, Implementation of another algorithm
Merchants, Developers, Account Managers	Need for more complex information	Implementation of another algorithm
Merchants, Account managers	Poor documentation / information display in the tool	Embodiment of more information: User Interface / Documentation
Merchants, Developers	Usability	Changes in User Interface
Merchants, Developers	Bugs in the tool	Fixing
Merchants	Insufficient promotion	Pro-active engagement with users

We can see from the above Table that there is some contradictory feedback between developers in the company and merchants, as well as account managers regarding the reliability of the tool. This might be reflected in the literature regarding alignment of incentives and actors' different perceptions; on one hand, developers are interested in company's infrastructure expansion, thus the tool is an additional product that they can work on. On the other hand, account managers are more interested in potential loss of their reputation, since if the tool does not eventually work as expected, the merchants might no longer trust them; this is also reflected in their reluctance to promote the tool. Lastly, the merchants are the ones to be incurring the direct costs of chargebacks and refusals if the tool proves to be invalid.

In this Chapter the research question was formed as *"How do merchants and account managers engage with the tool and how do they choose their risk profile?"*. Through the interviews, we were able to find that both merchants and account managers do not totally trust the underlying model

of the tool and hence they are reluctant in applying the suggested changes. Moreover, it was found that that the quality of output in terms of user interface and technical bugs, the insufficient documentation, the response time, interaction with it by the use of the slider and the belief if it adds value to the business are the main challenges related to the tool's usage. Finally, the choice of the risk profile to be adopted, depends on manual review of the transactions with the plan to keep in balance chargebacks as well as refusals.

An interesting result that came out of the interviews is that out of the seven high visitors, the four claimed they have not used the tool and one of them has never seen it before. This pinpoints to the fact that there is no internal mechanism that accurately monitors who is using the tool by for example distinguishing account managers from actual merchants that can log in to the tool. Furthermore, from the seven interviewees listed as high visitors, only one reported to occasionally apply the suggestions of the tool, while other two use it as an auxiliary tool to double check with their own analysis. Nevertheless, all the participants expressed their willingness to use the tool primarily because manual reviewing is time-consuming, as well as because of capacity limits in their resources. We also saw that developers within Adyen stated the tool is mathematically accurate, while account managers mentioned they do not trust it. These contradictions pinpoint to the fact that the company should firstly make sure to monitor the usage of the tool and then proceed to the improvements.

Chapter 5

Analysis of Risk Behavior for Target Group

After the analysis of the historical data as well as the interviews with the merchants, we created a picture of the common characteristics between the users and non-users of the tool, and moreover discovered the benefits and challenges related to its usage. The next step is to explore the logic that drives merchants take risk decisions by changing their risk settings. However, before trying to find an answer to this research question, we found imperative the need to create the monitoring mechanism that would firstly allow us to identify the group of merchants for which the tool adds more value. Afterwards, the exploration of out last research question would be more meaningful, since we will be sure that we focus on the candidates that the tool is meant to be used by.

The analysis in this Chapter is on one hand expected to shed light on how the problem owner can pro-actively engage with the customers, and on the other hand what features are important when merchants are involved with risk decisions. Hence, the research question to be answered is formulated as follows:

SQ: *"By looking at broader patterns, which factors can we identify that are indicative of merchants' engagement with risk management?"*

In order to provide an answer, we first identify the group of merchants for which the tool adds more value, and we explore how they behave in terms of changing their risk settings. Furthermore, we ran a multivariate regression analysis in order to identify the drivers behind the engagement with risk management.

5.1 Methodology

The methodology in this Chapter follows a similar approach as in Chapter 3. More specifically, a quantitative analysis with empirical focus is being followed in order to answer the subquestion. Below, more details are given on how this is achieved.

5.1.1 Methodology for Target Group

On one hand, providing an answer to the sub-question required the identification of the "target group" for the Risk Calculator, since as discussed in earlier Chapters merchants with optimized transactions are not willing to use the tool and hence the company should pro-actively engage with the appropriate merchants. This included the development of a program (Java job) that automatically runs the tool's calculations for every merchant account in the company and stores the data to the database. More details on this are given in Chapter 6. Those aspects included whether merchants have access to the tool (recall that access to the tool is not free), how many years they are processing transactions and the volume of transactions they have. This resulted in a subset of merchants for whom it makes more sense to use such a tool.

Based on the result of the filtering, we analyzed a dataset with the changes made to the risk settings by this subset of merchants. The time range was chosen to be the year beginning on September 2015 and ending on September 2016, when this report is being written. As already noted, merchants are able to make changes to their risk settings without using the tool. Hence, we retrieved the data indicating when the merchant made a change, which risk check was changed and what score was given to it. An illustration of this dataset is shown below in Table 5.1. The first column identifies the merchant, while the second column identifies the risk check that was changed. Furthermore, "begin_date" denotes the time that the risk check was changed, whereas "end_date" denotes until when the change was valid. Hence, in the first row where the end date is 2015-10-06, we expect that there would be another record for this merchant with another change made to the same risk check, beginning at 2015-10-06. In case the risk check was not further changed by the merchant, the end date will have the format 9999-12-31. Lastly, score_1 indicates the score that the risk check was given initially by the merchant and score_2 the new value.

The analysis of this dataset included the creation of histograms and scatterplots and the calculatiion of descriptive statistics. More specifically, a metric counting the changes made to the risk checks on a *monthly* basis was created. Initially, we created also a metric on a weekly basis, however

Table 5.1: Part of the dataset with risk changes made by merchants.

account_id	riskcheck_id	begin_date	end_date	score_1	score_2	score_diff
5351	64	2015-09-30	2015-10-06	80	100	20
7296	4	2015-09-30	2015-09-30	50	50	0
125906	46	2015-12-30	2015-12-30	50	20	-30

since the results did not differ, we decided to keep only the monthly metric. Subsequently, frequency of changes was plotted.

Next, identifying how does this group behave, required looking at the proposals of the tool for the specific merchant accounts. In this way, we got a table with the tool's suggested scores for each of the risk checks and then we matched this to the actual changes merchants made to the particular risk checks. As a time frame, we took the week before running the Java job. Accordingly, we plotted the histogram in order to compare whether the tool's predictions were close to merchants' changes. Apart from that, histograms depicting which risk are being changed as well as the direction of changes were drawn.

5.1.2 Building the Empirical Model

Finally, based on the previous results, we tried to predict the correlation between the variables that drive the changes made to the risk settings. This is done by building two empirical models, following a *Linear Model*, as well as a *Generalized Linear Model (GLM) regression*, in order to predict the dependent variable risk changes by merchants. Firstly, we introduce all the concepts that will lead us to the development of a conceptual model and later on we introduce the actual measurement model.

As you might recall, under Section 2.1.3, we described that merchants have to face two types of costs while dealing with their transactions; the costs related to *chargebacks* and the costs related to *refused transactions*. We realise this as a fundamental relation influencing merchants engagement with risk management, since we assume that they aim for profit-maximization. Particularly, both the actual monthly level of chargebacks and refusals, as well as the deltas (change from one month to another) of these values are of importance. On one hand if the overall rates are high, merchants are expected to take more actions in comparison to merchants that have low rates and hence do not suffer from large losses. On the other hand, an increase in delta values from month to month, should imply more actions taken by merchants in order to adjust them. Nonetheless, the effect of this reaction should also be specified. According to account managers in Adyen,

chargebacks require a couple of months to be received after the shopper initiates the process.

We also saw under Chapter 2 that the *transaction characteristics*, such as the average transaction value and the offering of 3D-Secure, are factors influencing fraudsters' motivation to commit fraud. Moreover, *merchants characteristics*, including sector, country where they operate and size, are also contributing factors in how they respond to their risk environment. Regarding the size, as noticed during the interviews with merchants (Chapter 4), it determines the resources a merchant has to spend on risk assessment.

Furthermore, in Section 2.2 we described the different techniques for committing e-commerce fraud. As these techniques get more advanced and fraudsters can exploit cardholders' credentials more easily, merchants face greater risks since they become more exposed. However, as we know from our case study, merchants have the option to follow security advices that can mitigate fraudulent transactions. All the afore-described relations lead us to the development of the conceptual regression model, as presented in Figure 5.1. Table 5.2 additionally list the relations and variables presented in the model.

Figure 5.1: The conceptual regression model.

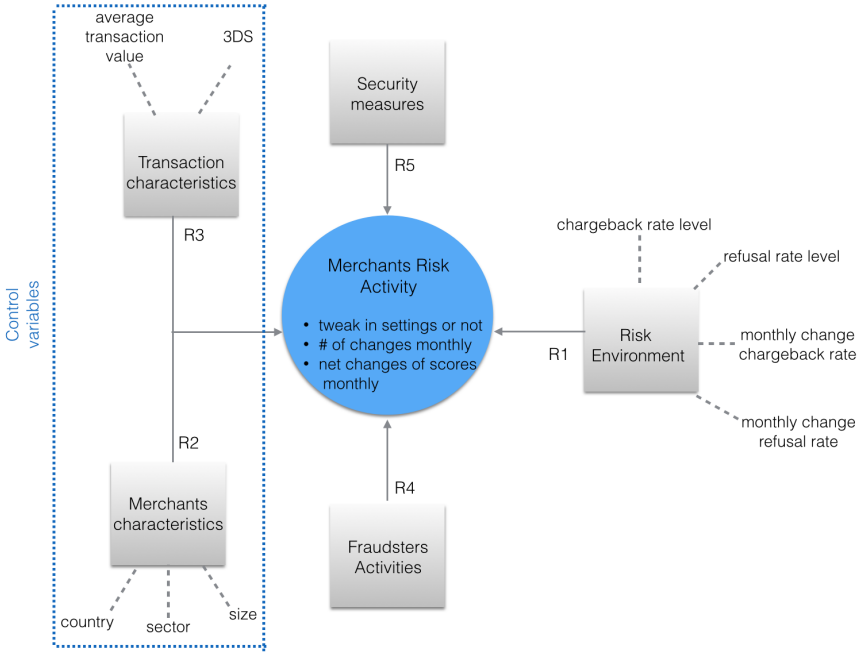


Table 5.2: The relations of the conceptual model linked to the literature.

Relation	Independent Variables	Source
Risk Environment	• monthly chargeback level	(De Gennaro, 2006)
	• monthly refusal level	(Wolters, 2012)
	• monthly change in chargeback	(LexisNexis, 2015)
	• monthly change in refusals	(Ingenico Payment Services, 2015)
Merchants Characteristics	• sector	(Ingenico Payment Services, 2015)
	• country	(Cognizant, 2016)
	• size	interviews
Transaction Characteristics	• average transaction value	(Ingenico Payment Services, 2015)
	• 3D-Secure	interviews
Fraudsters Activities	–	(Bhatla, Prabhu, & Dua, 2003) (Bahnsen, Aouada, Stojanovic, & Ottersten, 2016)
Security Measures	–	(Dal Pozzolo, Caelen, Le Borgne, Waterschoot, & Bontempi, 2014)

5.1.3 Building the Metrics

As outlined in the conceptual framework above, our dependent variable is an indicator of merchants risk engagement with risk management. Based on the available data, we propose the following three metrics in order to capture this activity:

• Whether merchants made any change in their risk settings for a specified period of time or not

• The number of the risk checks that merchants changed in a specific period of time

• The net score merchants gave to the risk checks they changed in a specific period of time

These different metrics can yield different results, by shedding light on various aspects that affect merchants' willingness to engage with risk decisions. As such, the 3 metrics are being described in detail below.

Metric 1: Binary. The first dependent variable denotes whether the merchant made changes to the risk settings per month during the year September 2015 to August 2016. This can be considered as a binary variable, where 1 indicates that the merchant changed the risk settings, while zero indicates that the merchant did not change the risk settings.

Metric 2: Counts. The second dependent variable is calculated as the number of changes the merchant made to the risk settings per month. As mentioned, the time frame is chosen to be one year, hence this metric captures how many risk checks each of the 2077 merchants changed each of the 12 months. Changes to the same risk check in the same month count as discrete.

Metric 3: Net score. The third and last metric is different from the previous two. On one hand,

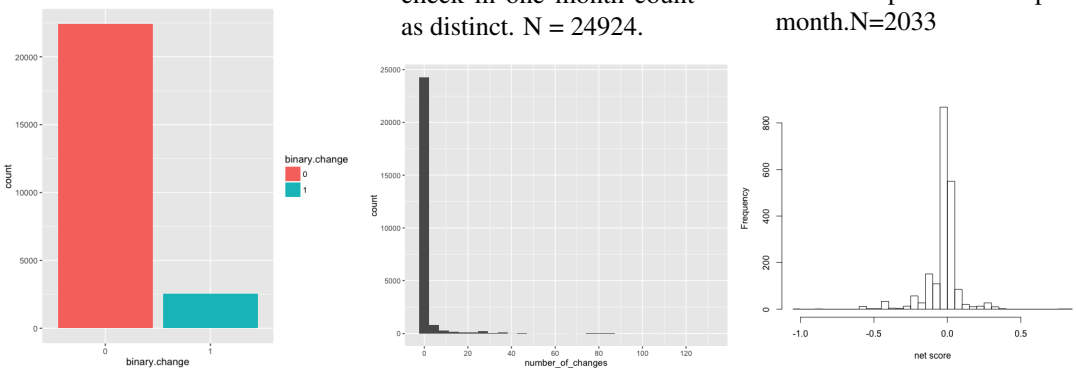
the major conceptual difference is that the two first metrics try to capture whether merchants make changes or not, while this third metric focuses on the population that *is making changes*, excluding those merchants that do not tweak the risk checks. Additionally, the third metric captures the direction of changes, i.e. the net increase or decrease of the scores that merchants gave to the risk checks, each month they tweaked the settings. In order to calculate this metric, the default scores given to the risk checks at zero time were taken into consideration in order to find the difference the first time a change occurred. Below in Figure **??** the distribution of the three metrics is depicted.

Figure 5.2: The three dependent variables.

(a) Metric 1. Bar plot of the binary dependent variable where 1 = changed risk settings, 0 = did not change. Sample size (n)= 24924.

(b) Metric 2. Number of changes to the risk settings per month, for 1 year period. Changes to the same risk check in one month count as distinct. N = 24924.

(c) Metric 3. Histogram of net increase/decrease in the scores per merchant per month.N=2033



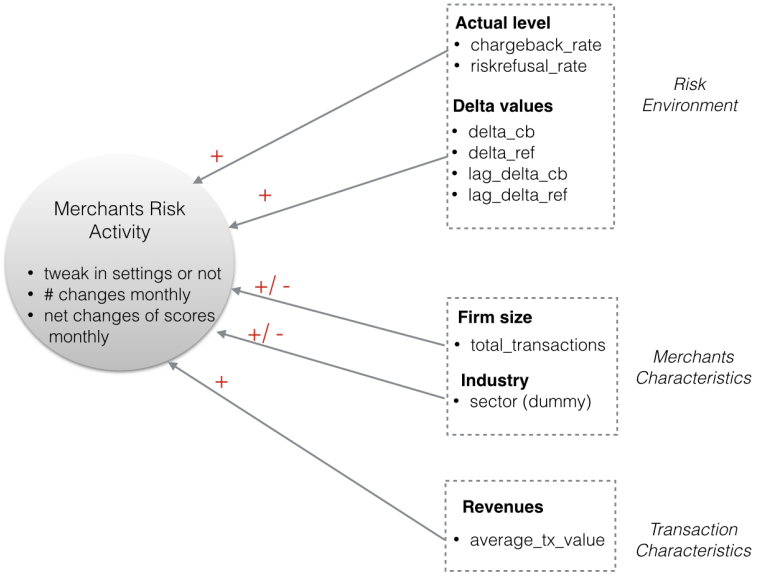
We see in Figure 5.2a that 22396 of the counts (that is, merchant per month) made no changes, while 2528 counts made a change. As we recall from previous findings, merchants do not seem to make changes every single month, but instead 2-3 months per year, hence the inflated zeros are somewhat expected. This dependent variable is binary, therefore a logistic regression is going to be used. The histogram in Figure 5.2b depicts a *negative binomial* distribution with mean = 1.35 and st.dev. = 6.59. Again, we notice that the majority of the observations is centered around zero. The distribution of the dependent variable dictates that a negative binomial GLM is going to be used in the regression analysis. Finally, Figure 5.2c shows the histogram of the third metric. Since this metric includes negative values, an *Ordinal Least Squares (OLS)* regression model is going to be used.

Based on these metrics, it would be possible to indicate the factors that drive the changes in

merchants risk settings, and ultimately have a picture of the determinants regarding engagement with risk management.

Ideally, we would like to test all the relations present in the conceptual framework under Figure 5.1. However, given the complexity of a single transaction and the time constraints, not all of the relations were possible to be tested. More details on the relations that were excluded are presented in the Threats to Validity Section. Figure 5.3 presents the measurement model, i.e. the parts of the conceptual framework that will eventually be included in the regression model, along with the names of the independent variables and the direction of the relationship.

Figure 5.3: The measurement model.



5.1.4 Building the Hypotheses

According to literature, we build a set of hypotheses which are later going to be tested by means of statistical analysis, as well as regression analysis. As noted in several studies (Wolters, 2012; LexisNexis, 2015; Ingenico Payment Services, 2015), but also from the findings of our interviews, chargebacks and false positives constitute the main risk environment that merchants have to deal with. Based on that, we assume that an increase in these two KPIs (H1), as well as their overall level throughout a specified period (H2) are concerning factors for merchants and will thus trigger

their reaction in order to mitigate them.

Furthermore, according to Ingenico Payment Services (2015), fraudsters tend to target products with an average value of at least 30 euros and the fraudulent attempts increase as the price reaches 150 euros, which according to the study represents the peak. However, another study mentions that transactions with value 0-20$ tend to be twice as fraudulent than higher amount purchases, due to the fact that they first want to test the stolen credentials. If merchants are aware about these facts, we expect that they will behave differently based on the average transaction value.

Moreover, we mentioned earlier in this Chapter that larger merchants tend to be more risk aware than smaller ones, in terms of willingness to adopt anti-fraud solutions (LexisNexis, 2015; Ingenico Payment Services, 2015). This depends on the resources they have, such as more stuff to make risk assessment and financial capability to pay for anti-fraud solutions.

Similarly, we know that merchants' sector might have an impact on merchants' engagement with risk decisions, depending on the margins and how easily the chargeback costs are being absorbed (Ingenico Payment Services, 2015).

Table 5.3: Hypotheses to be tested.

#	**Hypothesis**
H1	An increase in delta of chargeback/refusal rate increases merchants engagement with risk management
H2	An increase in overall chargeback/refusal rate increases merchants engagement with risk management
H3	Merchants significantly differ in terms of engagement with risk management, depending on the average transaction value
H4	Merchants significantly differ in terms of engagement with risk management, depending on their size.
H5	Merchants significantly differ in terms of engagement with risk management, depending on the sector they belong.

5.2 Findings About Target Groups

This Section describes the process followed in identifying the merchants for which the tool adds more value. Firstly, the filtering of merchants is presented, whereas the behavior of this identified target group is also analyzed by means of descriptive statistics.

5.2.1 Characteristics

In Chapter 4, during the interviews with various merchants, we noticed some contradictory observations; for instance merchants that were classified as users claiming not to have seen the tool before, merchants that did not seem very excited in using this kind of anti-fraud solutions, or others that expressed high eagerness in using the Risk Calculator. These findings led us to creating a split of the merchants and identifying the ones for which it makes more sense to use it. The main criteria for splitting the groups were the following:

• Identifying the merchants who have access to the tool

• Identifying the merchants who use Adyen as a PSP, and filter out those who use other PSPs since they typically provide little transactional information

• Identifying the merchants who are processing more than one year and thus have enough historical data

• Identifying the merchant accounts which are used for actual transaction processing, and not for example for testing reasons

• Identifying the merchants who process at least 1000 transactions during a 9-month period so that the tool is able to run the calculations

• Identifying the merchants for which the tool suggests multiple configurations, as during the interviews respondents with only one available proposal mentioned that the tool is not working for them

• Identifying the merchants for which the tool suggests the greatest optimizations, in terms of minimizing both chargebacks and refusals

The first six aspects could very easily be traced after we created the monitoring mechanism, by applying the appropriate filters. For the last one, however, we had to quantify the optimization that the tool suggests. Since the tool is trying to achieve a balance between chargebacks and false positives on order to maximize revenue, we captured the improvement of these two KPIs by using the following metrics:

$$\%\ \text{chargeback improvement} = (\text{current chargebacks} - \text{simulated chargebacks}) / \text{current chargebacks}$$

% refusal improvement = (current false positives - simulated false positives) / current false positives

In order to spot the merchants with the highest improvements as suggested by the Risk Calculator, we created the scatter plot of these two metrics (Figure 5.4). The best "candidates" are represented in the upper right quartile of the diagram (red colour). For these merchants the tool suggests an improvement of 25% in both chargebacks and false positives. However, we know from the preceding analysis as well as from previous research (Wolters, 2012; LexisNexis, 2015) that most of the times there is a trade-off between chargebacks and false positives; merchants also seemed to be aware of this fact during the interviews, hence we also defined the groups for which there is an improvement on one KPI and a slight loss on the other. As such, for the green colour group, merchants achieve a 10-24% improvement in chargebacks and a loss of maximum 15% in refusals. Similarly, the blue group can achieve a 10-24% improvement in refusals and a loss of maximum 15% in chargebacks. The next question for these two groups is whether the gain obtained for the one KPI compensates for the loss on the other KPI.

The application of the seven aforementioned filters resulted in the segmentation of merchants as depicted in Figure 5.5. The right part of the tree includes the merchants that the company should pro-actively engage with and hence the group on which the subsequent analysis will be performed. In order to have the largest sample possible, we chose the 2077 merchant accounts (used for transaction processing) in order to analyze their behavior.

We should remind you at this point that the tool works based on historical data, hence the transactional information per merchant change in time. Therefore, the identification of this target group should be re-evaluated by the company on frequent time intervals, as new merchants might start using Adyen's platform or as merchants are getting more or less volume of transactions.

Figure 5.4: Scatter plot of the improvement in chargebacks (%) versus the improvement in refusals (%) as suggested by the tool. The colour denotes the different groups of merchants

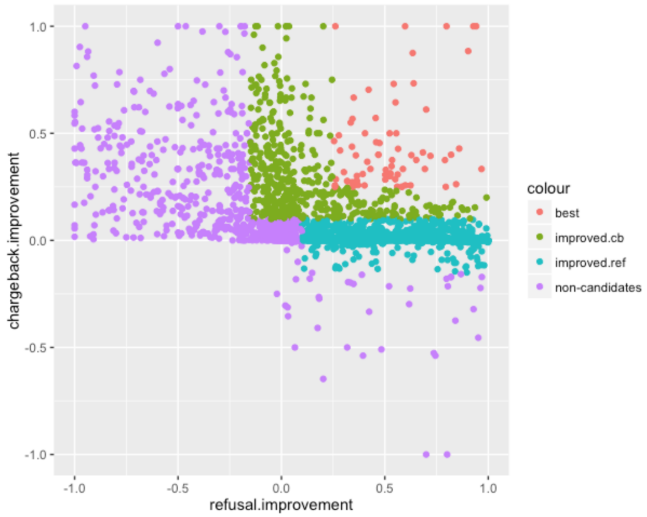
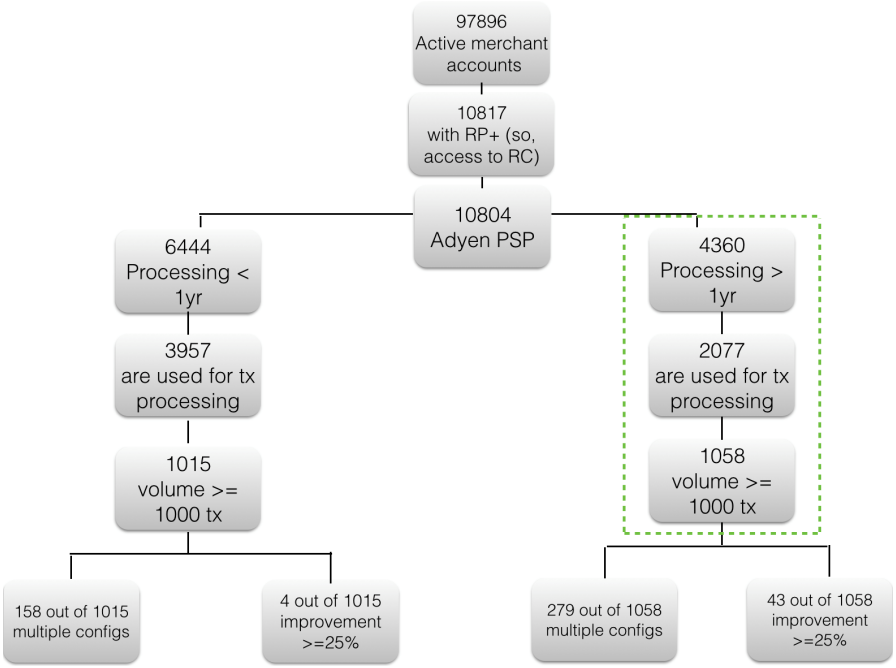


Figure 5.5: Identification of the merchant accounts that should use the Risk Calculator. The green-boxed area indicates the groups that represent the best candidates.



The group for which the tool can add more value includes merchants who meet the requirements for triggering the calculations. This basically means having enough transaction volume and actively processing transactions. Out of those, the merchants that practically benefit are the ones which can achieve a considerable improvement in both refusals and chargebacks, as well as the ones that have multiple opportunities for optimization (many slider positions).

5.2.2 Merchants' Risk Behavior and Frequency of Changes

In everyday life people are often required to take the role of a decision maker and thus decide about various issues that most of the time might be considered as risky, depending on the degree of knowledge they have about the results of their actions (Taghavifard, Damghani, & Moghaddam, 2009). This situation is especially predominant in business environments, where employees and managers have to take certain decisions that might affect the overall course of the business. Particularly, when it comes to payments it is evident, as already discussed earlier, that blocking too many legitimate transactions or accepting too many fraudulent comes at the cost of money. It is interesting hence to explore people's reaction in how they handle the risks regarding their payments and how they engage with risk management.

As such, in this case study, we analyze the data that are related to merchants' (target group) actions towards controlling the payments that can "pass through" and thus are considered safe, i.e. initiated by a legitimate shopper. Such actions can be taken through Adyen's backoffice system, where each merchant has an account and through the settings they can give a score to each of the 80 different risk checks related to transactions (see Section 2.6.1 for more details).

Frequency in changes. An indicator of the merchants' risk behavior is the frequency in which they make changes in their risk settings. For the specified sample of merchants, the histogram of the frequency in the risk changes can be seen in Figure **??**. As already discussed, the results might be different when considering changes made to the same risk check as one concrete change. Figure 5.6b below captures this different metric. In both histograms, the counts on y axis reflect the number of changes each merchant made per month.

Figure 5.6: Histograms of the two metrics.

(a) Metric 1. Number of changes to the risk settings per month, for 1 year period. Changes to the same risk check in one month count as distinct. N = 24924.

(b) Metric 2. Number of changes to the risk settings per month, for 1 year period. All changes to the same risk check in one month count as one.

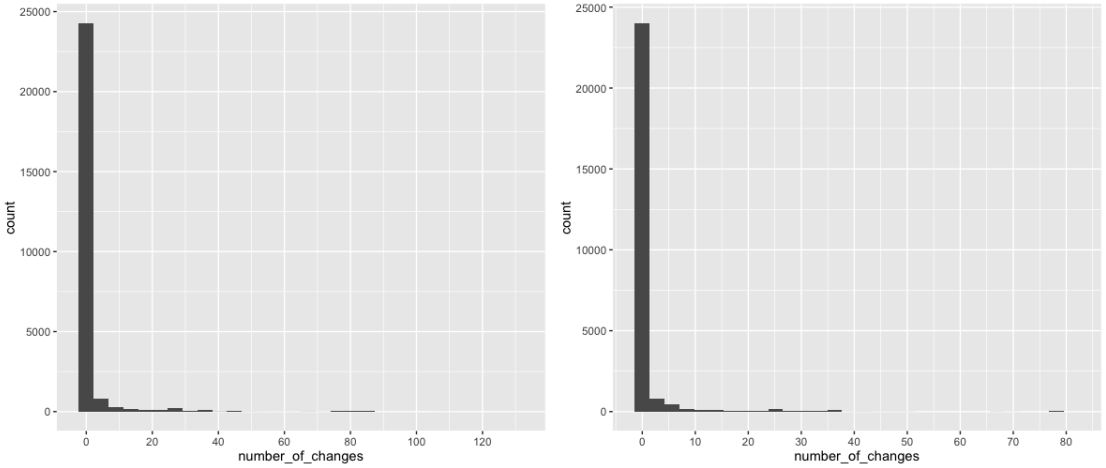


Table 5.4: Descriptive statistics of the two metrics. The n equals the number of merchants (2077) multiplied by 12 months.

	n	**mean**	**st.dev.**
Metric 1	24924	1.35	6.59
Metric 2	24924	1.18	5.69

We can see from the Figure above that the two metrics produce almost identical graphs, hence we can conclude that merchants change multiple risk checks when they adjust their risk settings. Moreover, we notice that the distribution of the number of changes is a *negative binomial* with a median of zero in both cases. The descriptive statistics of both metrics are presented in Table 5.4.

Next, we also looked at the mean and median number of changes that each merchant made during this period, as displayed in Figures 5.7a and 5.7b. Histogram *(a)* reflects the average number of changes each merchant made per month, while histogram *(b)* shows the median number of changes each merchant made per month. Table 5.5 shows the descriptive statistics for the two variables.

Figure 5.7: Histograms for number of changes per month.

(a) Histogram of the average number of changes per month per merchant.

(b) Histogram of the median number of changes per month per merchant.

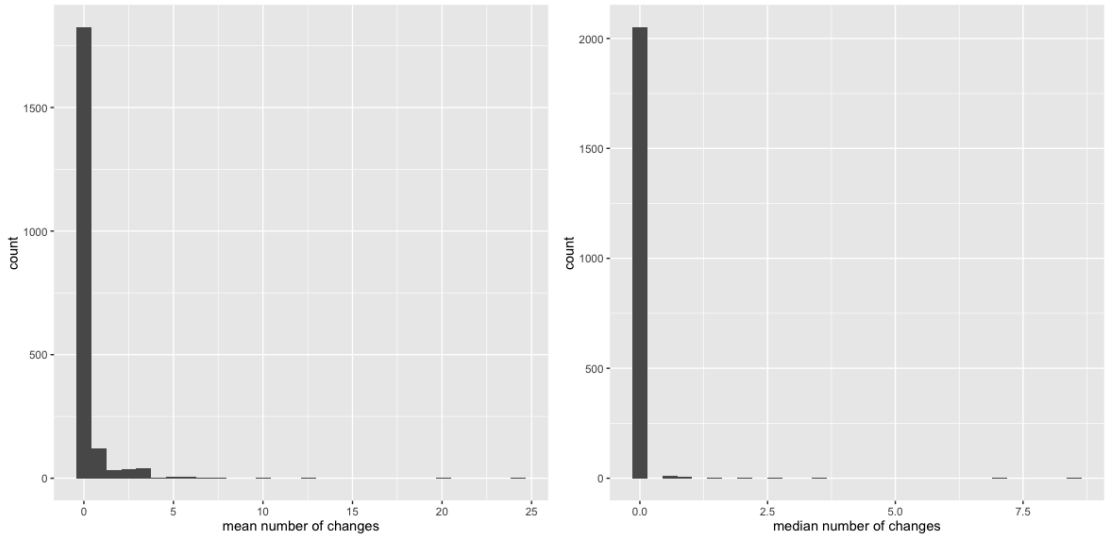
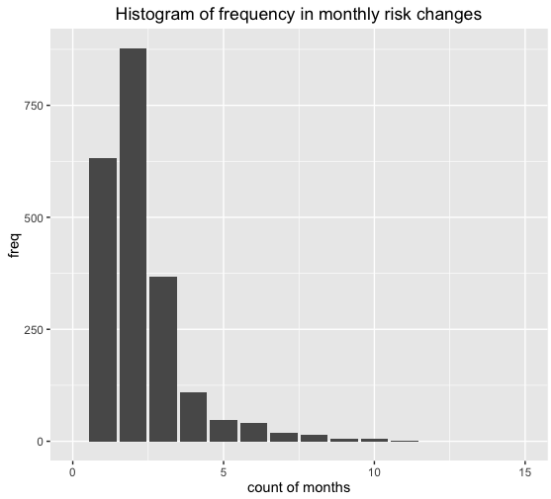


Table 5.5: Descriptive statistics

	n	**mean**	**st.dev.**
Metric 1	2077	0.28	1.11
Metric 2	2077	0.022	0.289

Going a step further, we also explored the total number of months during the year that merchants make changes to their settings. The reason for looking at this metric is due to the fact that we can assume, based on the interviews, that merchants might adopt a "wait-and-see" attitude when they change the score of the risk checks. As we see in Figure 5.8, most of the merchants visit their risk settings and adjust the scores two months per year. The y axis in the histogram shows the number of merchants that made any change in a particular month.

Figure 5.8: Count of months that merchants make changes throughout the year.



Number of risk checks being changed. Looking at how many risk checks merchants are changing at once can help us conclude whether some of them are considered more important and thus are being given more weight. Figure 5.9 shows the histogram of the distinct number of risk checks that merchants changed per month.

Figure 5.9: Histogram of the count of risk checks changed at once by merchant.

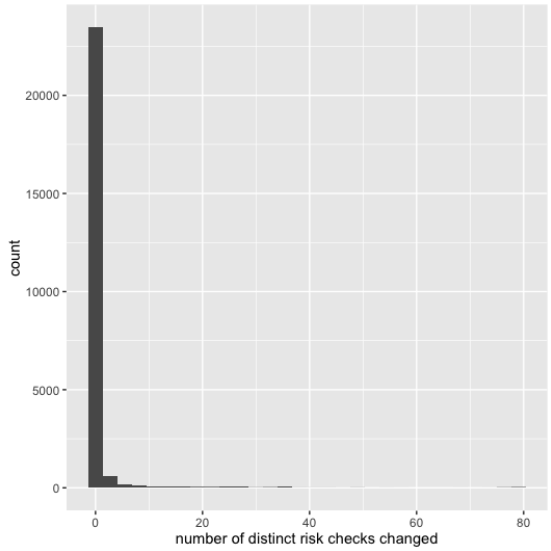
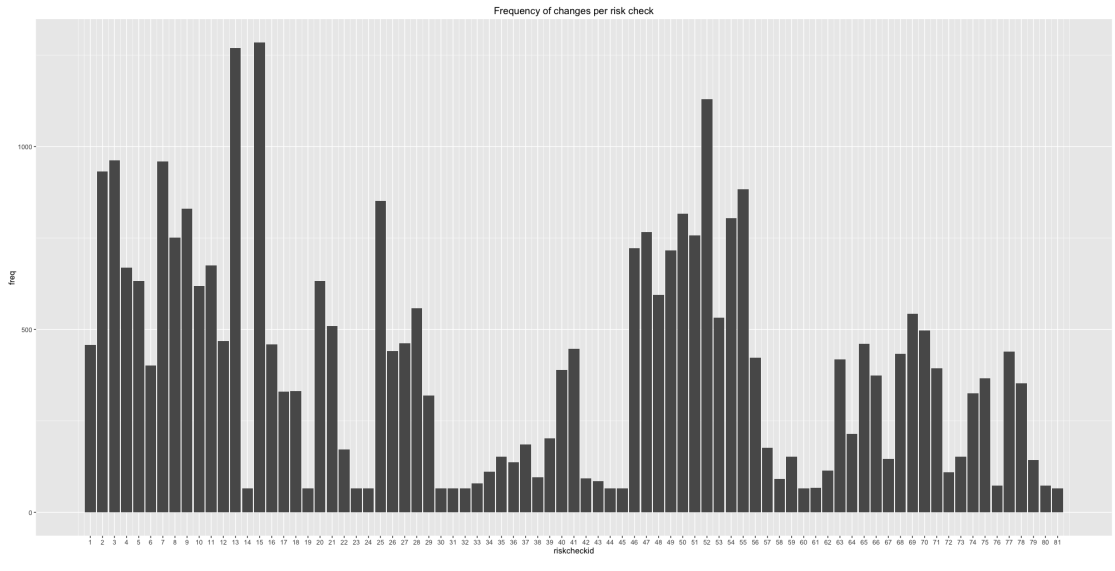


Table 5.6: Descriptive statistics

	n	mean	median	st.dev.
Metric	24924	0.76	0	4.74

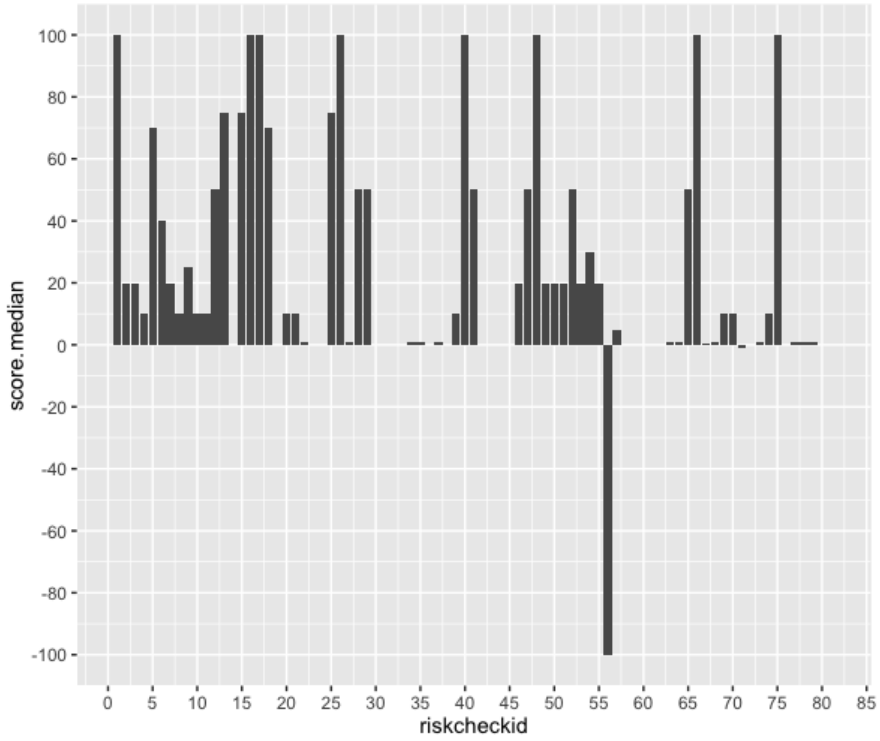
Subsequently, taking a look at the frequency that each of the risk checks was changed during this one year period can further indicate whether there is for example a specific subset of risk checks being changed more frequently. We notice in Figure 5.10 that all 81 risk checks have been changed during this period, with the 20 out of the 81 being changed 100 times, whereas the remaining were changed between 300 and 1500 times. Therefore, we see that there is no group of risk checks which is being ignored by merchants.

Figure 5.10: Frequency that each risk check was changed during one year.



Related to this aspect is also the actual values that merchants give to the risk checks, since negative scores denote in general a more risk-taking attitude by means that they make their settings more loose in order to accept more transactions. Figure 5.11 displays the median values given to each of the 80 risk checks during one year.

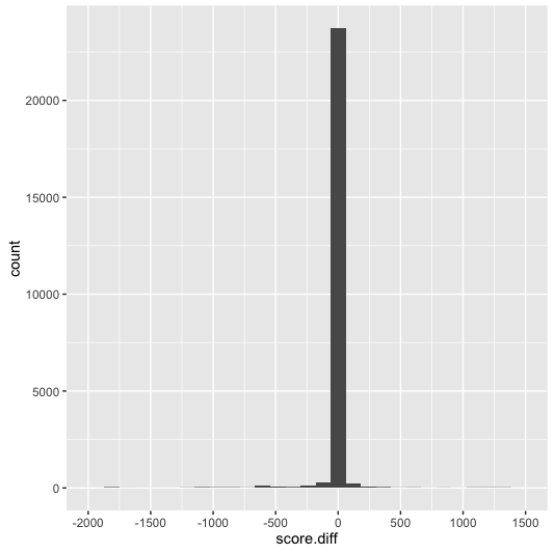
Figure 5.11: Median score given to each of the risk checks during one year.



As can be seen from Figure 5.11, only two risk checks are given a negative score. From the remaining, almost 20% of the risk checks is given a score of +100. We notice that the majority of the risk checks is being increased, hence we can infer that merchants tend to be risk-averse. Nevertheless, these values might make more sense when compared with the previous scores that the checks had, in order to identify the direction of changes.

Direction of changes. An aspect that can shed light on merchants' risk behavior is the direction of changes they make to their settings, in terms of increasing or decreasing the scores of the risk checks. Generally speaking, an increase in the scores denotes that the merchant is being more strict, i.e. trying to avoid risks, while a decrease denotes that the merchant is being more loose, i.e. risk-taking. In order to measure risk-averse and risk-taking attitudes, we created again two metrics. Metric 1 computes the overall net increase or decrease to the scores that merchants gave to the risk checks when changed their settings, by as an aggregation of the scores. Metric 2 calculates the net change as a ratio by dividing the sum of the score changes to the default values given when the merchant activated his account. The results can be seen below in Figure 5.12a and 5.12b.

(a) Histogram of the score differences per merchant per month. The score difference is aggregated per merchant for all the risk checks.



(b) Histogram of net increase/decrease of risk scores by merchant during 1 year period.

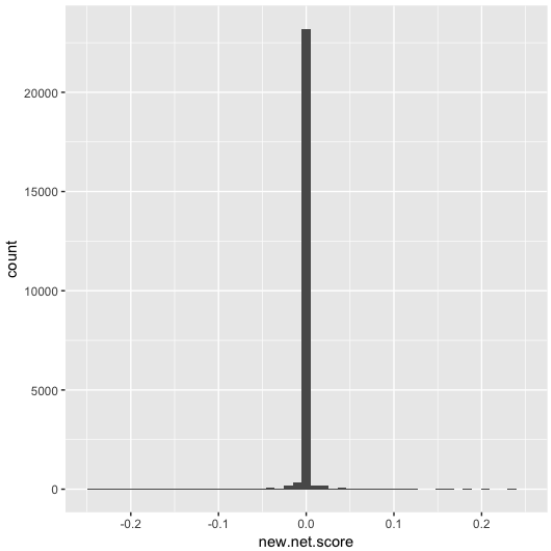


Table 5.7: The two metrics of net score change.

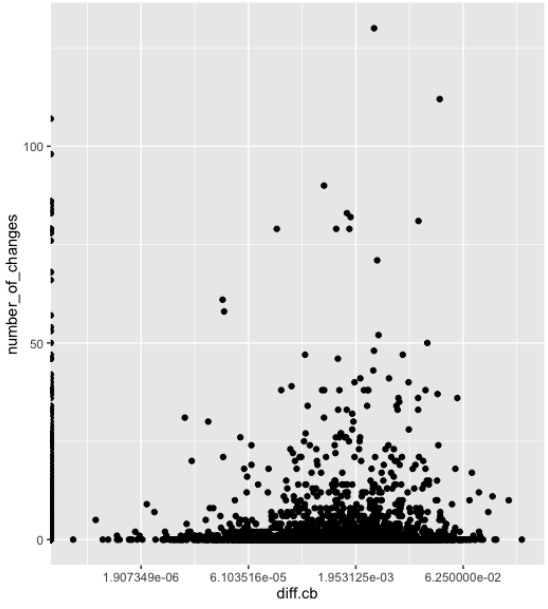
	n	**mean**	**median**	**st.dev.**
Metric 1	24924	-9.11	0	147
Metric 2	24924	-0.002	0	0.033

As we notice, the majority of merchants keeps a neutral stance when they make a change. This is reflecte in the fact that for the largest count of merchants, the net score change is zero. This implies a risk-averse attitude by trying to eliminate risks related to their transactions. This can be interpreted in two ways; on one hand, it may reflect that the merchant tried to achieve a balance by increasing some risk checks and on the same time by decreasing some others. On the other hand, it might indicate that the merchant activated or deactivated specific risk checks. This risk-averse attitude that can be distinguished from the two graphs indicates that merchants are probably faced with uncertainty when trying to decide about the scores. According to Behavioral Economics, when people are faced with uncertainty, it is more likely that they will go with the default, especially when it is presented as a recommended configuration (Kahneman, 2011).

Correlation of changes with chargebacks and refusals. After having looked into different factors that might influence the frequency in changing the scores of the risk parameters, we

also check the correlation between the number of changes and the chargeback, as well as the false positive rate. We expect as these two KPIs increase, so will the number of changes. This expectation derives from the fact that the context of making a risk score adjustment is to respond to threats that have to do with transactions, and these threats are practically fraud (i.e. chargebacks) and blocking of legitimate shoppers (i.e. false positives). Moreover, during the interviews with merchants (Chapter 4) the respondents claimed that changes to their risk settings are driven by trying to achieve a balance between chargebacks and false positives. The two following scatter plots depict the number of changes versus the delta values in chargebacks and in false positives respectively. In both of the Figures, no linear correlation can be identified. This comes at our surprise, since it implies that apart from chargebacks and false positives, there are other factors that drive merchants' changes in the settings. One possible explanation might be that this system is re-active instead of pro-active. This means that the merchant is notified when there is a chargeback, however the actual chargeback payment might be received several months later.

(a) Scatter plot of number of changes versus deltas in chargeback rate (logarithmic scale). Each point represents the delta value of chargebacks and the respective number of changes each merchant made this month.



(b) Scatter plot of number of changes versus deltas in refusal rate (logarithmic scale). Each point represents the delta value of refusals and the respective number of changes each merchant made this month.

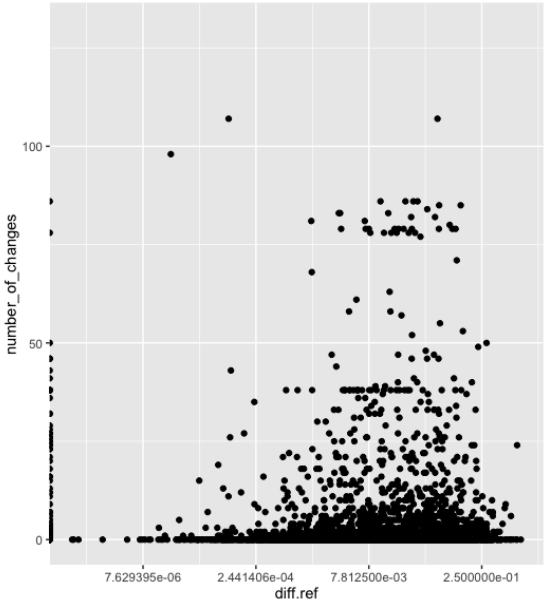


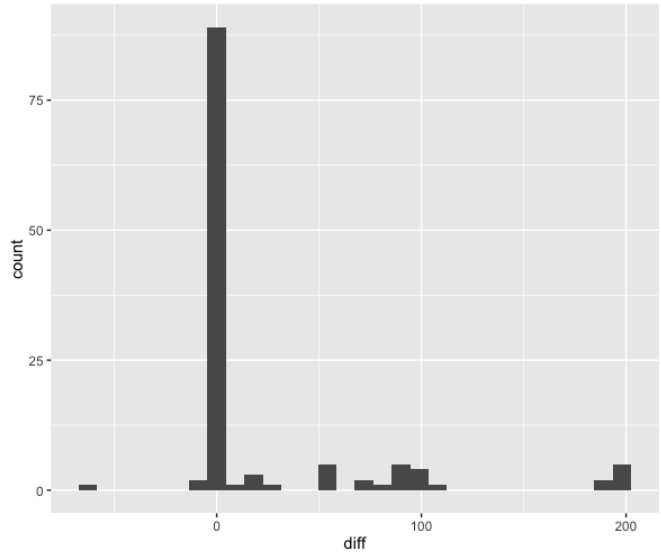
Table 5.8: Spearman correlation coefficient of the two variables.

	n	Spearman coeff.
changes Vs delta chargebacks	2077	0.0046
changes Vs delta refusals	2077	0.050

5.2.3 Merchants' Reaction Versus Tool's Suggestions

After getting a feeling about how merchants react to risks, we also wanted to compare the actual changes they make and the tool's proposals. Large difference between these two would indicate that the tool is failing to make the right suggestions. As such, we got the proposals of the tool on the 29th of September 2016 and we compared them with the changes the merchants made during the week 21-27 September. Out of the 2077 merchant accounts, 29 made changes in their risk settings during this week. It should be noted that not all of the 81 risk checks had been changed. The result is shown below in Figure 5.14.

Figure 5.14: Comparison of merchants' actual risk changes versus the tool's proposals. Each bar shows the difference between actual value and suggested value per risk check.



The x axis of the graph depicts the difference between the change that the merchant made and the suggested change as proposed by the Risk Calculator. As it can be seen, for the majority of the risk checks the difference is zero, indicating that the tool makes the "right" suggestions, by means that they are accepted by merchants. The next largest count is noticed on differences around 100

and 200. This practically means that for specific risk checks, the tool suggested to give a score of e.g. 0 while the merchant gave it a score of 100. Therefore, there is a group of risk checks that according to merchants are proven to be efficient in combating fraud and hence they will never accept to decrease them (see also Chapter 4). The problem spotted here is that the tool tries to solve the linear equation that contains as constraints the risk checks and mathematically it might make sense to decrease them. In order to prevent that, the linear equation should be modified in order to offer more gradual changes.

We see that merchants do engage with risk management, by visiting 2-3 months out of the year their settings and making changes. The majority is being neutral in the changes they make thus denoting a risk-averse attitude. Moreover, the changes merchants make are in line with the tool's suggestions.

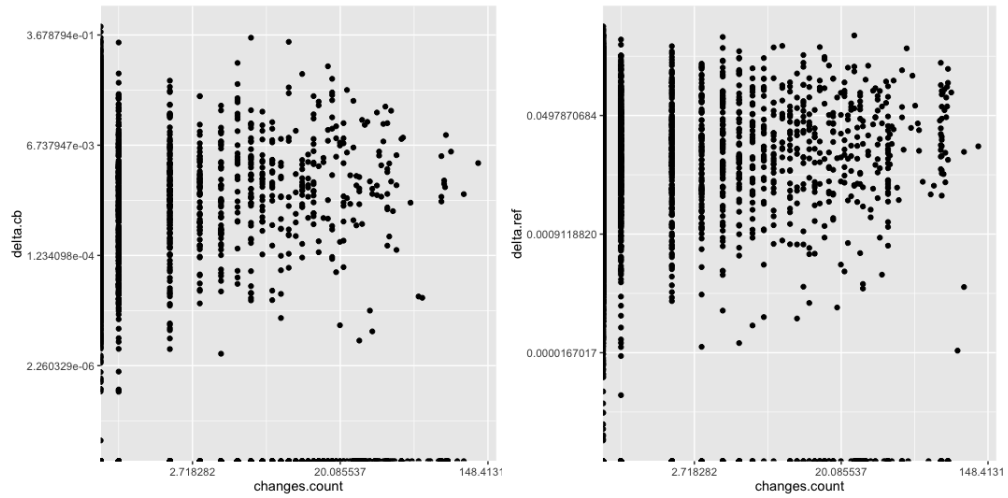
5.3 Hypothesis Testing

In this Section we test the hypotheses introduced under 5.1.2 through the use of descriptive statistics. After testing individually the hypotheses, the full regression model is going to be built. The reason for including the last part is to have a picture of how these factors simultaneously affect merchants risk activity. Please note that this Section includes a lot of figures and R outputs for the hypothesis testing. For aesthetic simplicity, only the scatter plots depicting the metric of counts (metric 2) versus the independent variables are presented. The Spearman coefficient for the dependent variable and the independent variables are presented at the end of this Section.

Testing Hypothesis 1. To recall, H1 was formulated as *"An increase in delta of charge-back/refusal rate increases merchants engagement with risk management"*. The scatter plots can be seen below; the natural logarithm was used for both x and y axis in order to make the graph more readable. By looking at Figures 5.20a and 5.20c, no strong correlation can be identified.

Figure 5.15: Metric 2 (count of changes) against the independent variables delta_cb and delta_ref (in log scale).

(a) Count_changes Vs delta_cb. (b) Count_changes Vs delta_ref.

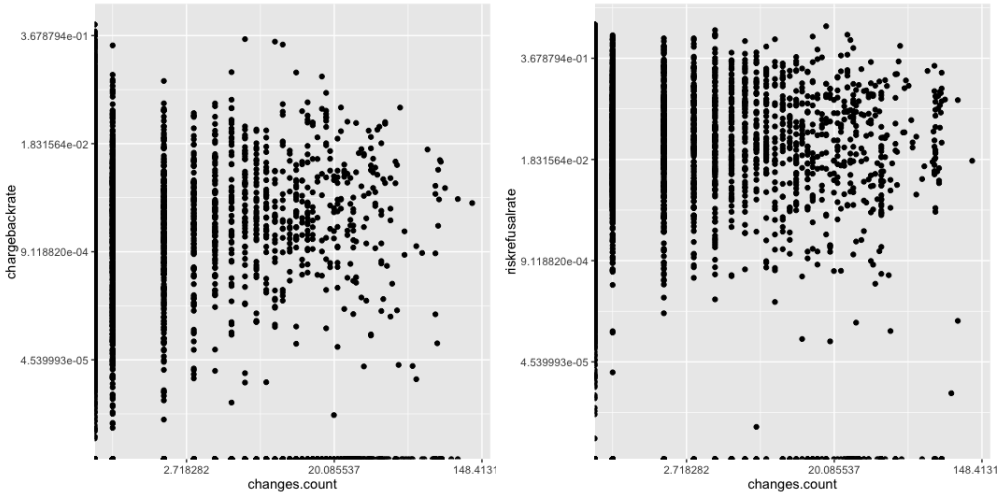


Testing Hypothesis 2. In order to test whether the higher overall chargeback/refusal rate leads to more frequent risk changes adopted by merchants, we created the scatter plots between each of the dependent variables and the independent variables "chargeback rate" and "risk refusal rate". The results together with the Spearman correlation can be seen below. Again, from Figure 5.20c no strong correlation between the two variables is visible.

Figure 5.16: Counts of changes against the independent chargeback rate and risk refusal rate (in log scale).

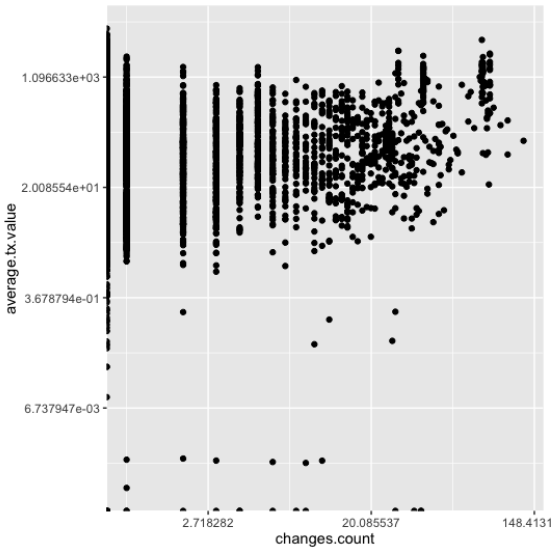
(a) Count_changes Vs chargeback rate.

(b) Count_changes Vs risk refusal rate.



Testing Hypothesis 3. Our third hypothesis was formulated as *"Merchants significantly differ in terms of engagement with risk management, depending on the average transaction value".* In order to test this relation, we created the following scatter plot. From Figure **??** no correlation can be identified between the two variables.

Figure 5.17: Count_changes Vs average_tx_value



Testing Hypothesis 4. In order to test whether the transaction volume, which reflects merchants' size, influences the risk activity adopted by merchants, we can have a look at the scatter plot below. A weak relationship seems to exist between count_changes and total transactions, when looking at Figure 5.18. The Spearman coefficient between all these variables is presented below in Table 5.9.

Figure 5.18: Count_changes Vs average_tx_value

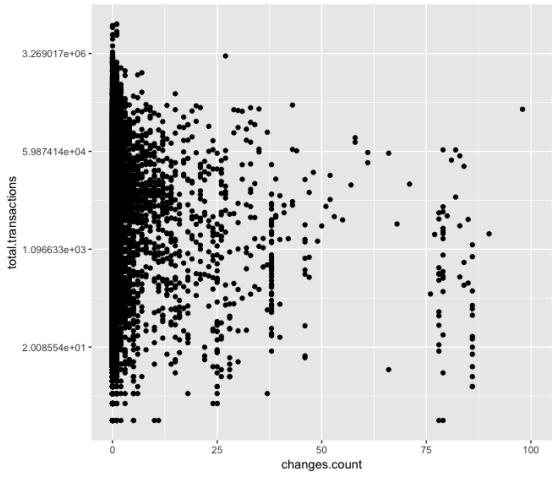


Table 5.9: The Spearman coefficient between all the dependent and independent variables.

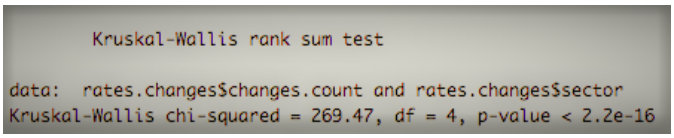
	Counts variable	**Net change variable**
delta_cb	0.0033	0.012
delta_ref	0.051	0.031
chareback rate	0.12	0.12
refusal rate	0.27	0.15
average_tx_value	-0.017	-0.19
total_transactions	0.16	0.027

Testing Hypothesis 5. In order to test whether the sector influences the risk activity adopted by merchants, we created a table with the descriptive statistics per sector and applied a comparison of means. Since the "sector" variable includes 5 levels and moreover the dependent variable (count of changes) is not normally distributed, we deploy the non-parametric Kruskal-Wallis test. The descriptive statistics of the "sector" variable along with the results of the statistical test can be seen below.

Table 5.10: Descriptive statistics for counts of changes based on merchants sector.

Sector	N	mean	st.dev.
Information services	10380	0.76	4.04
Physical goods	11244	0.56	4.18
Transportation services	2676	2.93	12.33
Hotel services	504	0.54	2.84
Unknown	120	0.22	0.69

Figure 5.19: Output of Kruskal-Wallis test.



The null and alternative hypotheses for the Kruskal-Wallis test are formulated as follows:

H0: The mean count_changes is equal in all groups of sector.

H1: The mean count_changes is not equal in all groups of sector.

We can see from the output presented in Figure 5.19, that the p-value is significantly low (2.2e-16), hence the null hypothesis is rejected; the mean count of changes is not the same for every identified sector. The Kruskal-Wallis test however does not indicate which group has the higher mean. This can be something that we can find out later from the regression model.

5.4 Findings on Multivariate Regression Analysis

As already explained earlier, in this Chapter we are expecting to predict what drives the changes in merchants' risk settings through a regression analysis. From the preceding analysis as well as the interviews with the merchants, we suspect that a change of the risk checks is triggered not only by one factor, but instead of many and hence we have a *multivariate* model.

Before proceeding to the regression analysis, it is essential to investigate whether these three different metrics capture indeed different things, as well as whether they are correlated with the

independent variables. In order to check how different are the three dependent variables, we can have a look at the scatter plots and the respective Spearman correlation.

Table 5.11: Spearman correlation coefficient for the three metrics.

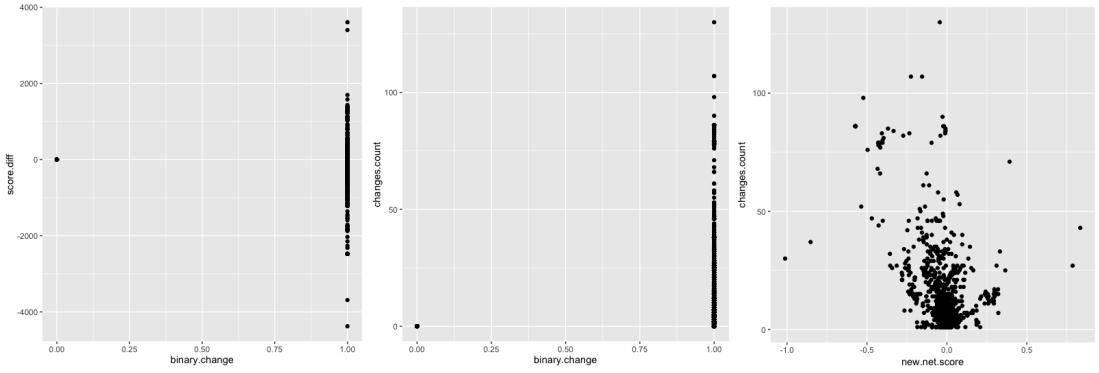
	Spearman coefficient
Binary change - Counts	0.99
Binary change - Net score	-0.24
Net score - Counts	-0.26

Figure 5.20: Scatter plots between the three dependent variables

(a) Scatter plot of binary change versus net score changes.

(b) Scatter plot of binary change versus number of changes per month.

(c) Scatter plot of number of changes versus net score.



We can see that the two first metrics (binary variable and counts variable) are highly correlated; this is of course expected as any count implies a change. After running the regression models for these two dependent variables, we saw that they produce identical results (as also expected from their high correlation), hence in the subsequent Section we present only the regression results for the binary variable.

Next step would be to check the correlation between all the independent variables to be used in the regression models. The scatter-matrix in Figure 5.21 depicts all the relations. Moreover, in Figure 5.22 we can see the partial correlations between the variables. We notice that the highest correlation we can observe is 0.37 and it is between chargeback rate and delta in chargeback rate. We also observe some correlation between total transactions and refusal and chargeback rates. The lowest correlations can be noticed between average transaction value and lagged chargebacks and refusals.

Figure 5.21: Scatter plots between all the independent variables for the regression models.

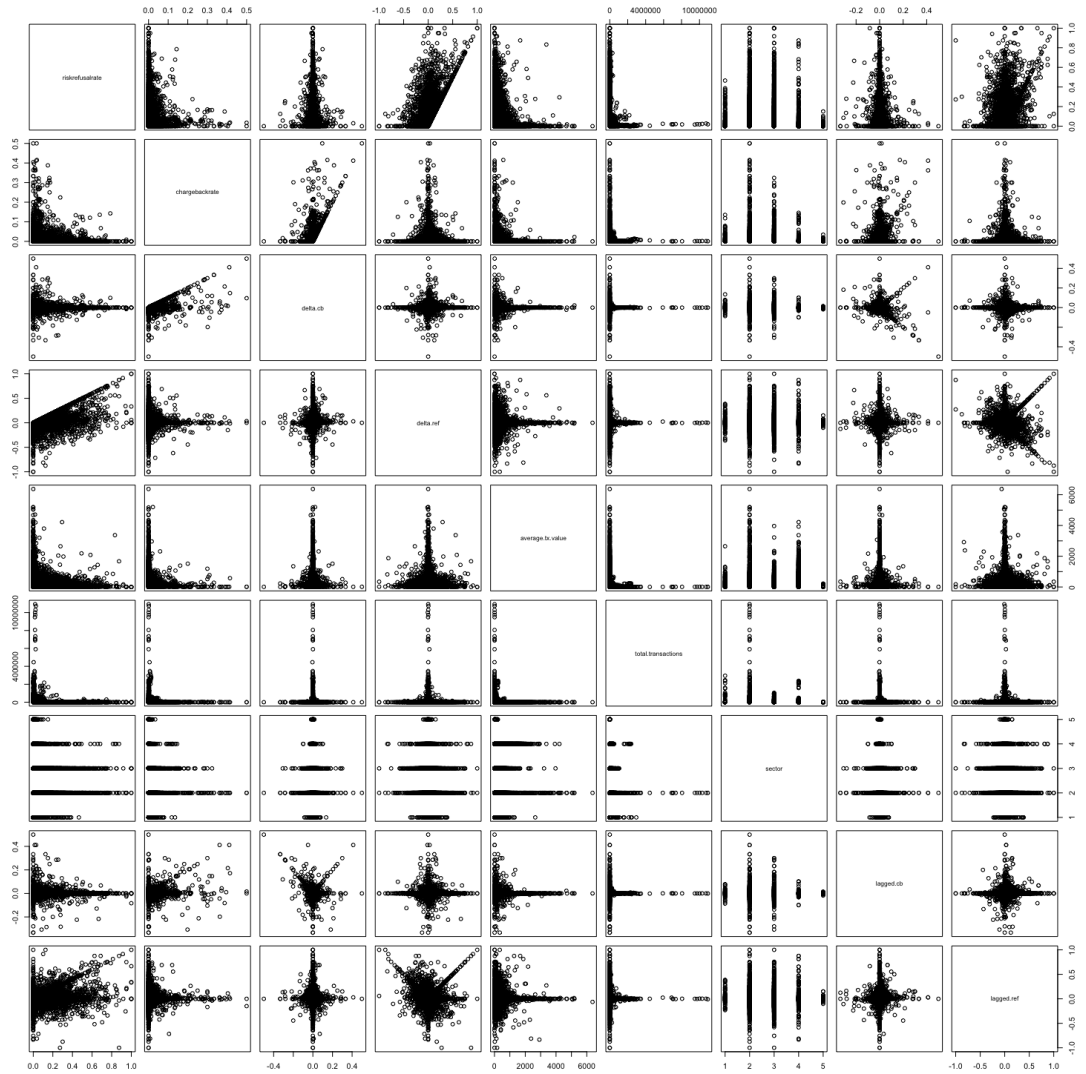


Figure 5.22: Partial correlation among all the independent variables.

$estimate

	riskrefusalrate	chargebackrate	delta.cb	delta.ref	average.tx.value	total.transactions	lagged.cb	lagged.ref
riskrefusalrate	1.00000000	0.24880895	-0.10724364	0.336428657	0.056547260	0.28049622	-0.067270294	0.220458143
chargebackrate	0.24880895	1.00000000	0.37193093	-0.090639419	-0.073510745	0.36768239	0.224138739	-0.053954319
delta.cb	-0.10724364	0.37193093	1.00000000	0.107365932	0.021309174	-0.10405054	-0.237314043	0.078706781
delta.ref	0.33642866	-0.09063942	0.10736593	1.000000000	-0.009434824	-0.06251137	0.113316043	-0.232459971
average.tx.value	0.05654726	-0.07351075	0.02130917	-0.009434824	1.000000000	-0.14735853	0.007205596	-0.001071035
total.transactions	0.28049622	0.36768239	-0.10405054	-0.062511373	-0.147358532	1.00000000	-0.024682014	-0.026196682
lagged.cb	-0.06727029	0.22413874	-0.23731404	0.113316043	0.007205596	-0.02468201	1.000000000	0.136456026
lagged.ref	0.22045814	-0.05395432	0.07870678	-0.232459971	-0.001071035	-0.02619668	0.136456026	1.000000000

5.4.1 Predicting the binary change variable

In the case where our dependent variable is the *binary_count*, i.e. "changed risk settings" / "did not change risk settings", we use a Logistic Regression. The choice of independent variables was introduced under Section 5.1.2. The result as produced by R can be seen in Figure 5.23. Important for the interpretation of the model are mainly the odds of coefficients, i.e. the beta coefficients to the power of *e*, as depicted in Figure 5.24.

Figure 5.23: The basic logistic regression output.

```
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.6432  -0.4976  -0.3862  -0.2987   2.8841

Coefficients:
                                Estimate Std. Error z value Pr(>|z|)
(Intercept)                    -4.201245   0.106722 -39.366  < 2e-16 ***
riskrefusalrate                 0.030640   0.002329  13.156  < 2e-16 ***
chargebackrate                 -0.014006   0.015221  -0.920   0.3575
delta.cb                       -0.010084   0.021618  -0.466   0.6409
delta.ref                      -0.006660   0.003186  -2.090   0.0366 *
log.avg.value                   0.030919   0.014449   2.140   0.0324 *
log.transactions                0.194739   0.008283  23.511  < 2e-16 ***
sectorhotel services           -0.094202   0.185613  -0.508   0.6118
sectorinformation services      0.598179   0.048962  12.217  < 2e-16 ***
sectortransportation services   0.573956   0.075579   7.594  3.1e-14 ***
sectorunknown                   0.559054   0.298619   1.872   0.0612 .
lagged.cb                      -0.008147   0.018527  -0.440   0.6601
lagged.ref                     -0.001094   0.002929  -0.373   0.7088
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 16361  on 24923  degrees of freedom
Residual deviance: 15316  on 24911  degrees of freedom
AIC: 15342

Number of Fisher Scoring iterations: 5
```

McFadden’s pseudo R squared: 0.0638

Figure 5.24: Exponent of coefficients.

(Intercept)	riskrefusalrate	chargebackrate	delta.cb
0.01497692	1.03111475	0.98609198	0.98996716
delta.ref	log.avg.value	log.transactions	sectorunknown
0.99336259	1.03140230	1.21499395	1.74901685
sectorhotel services	sectorinformation services	sectortransportation services	lagged.cb
0.91009887	1.81880291	1.77527634	0.44277764
lagged.ref			
0.89637252			

Figure 5.25: Confidence intervals.

	2.5 %	97.5 %
riskrefusalrate	0.026031879	0.0351677915
chargebackrate	-0.045951247	0.0138940676
delta.cb	-0.051521834	0.0330062593
delta.ref	-0.012891834	-0.0003949782
log.avg.value	0.003010236	0.0596571528
log.transactions	0.178559969	0.2110312015
sectorphysical goods	-4.412779426	-3.9943874771
sectorhotel services	-4.719797867	-3.8995694553
sectorinformation services	-3.808322159	-3.4017993286
sectortransportation services	-3.880360904	-3.3805567229
sectorunknown	-4.295171936	-3.0682728035
lagged.cb	-0.043718464	0.0285561414
lagged.ref	-0.006838083	0.0046475978

This is our basic regression model where we can see that 7 out of the 12 betas are significant (significance levels can be seen in Figure 5.2a). The model according to McFadden's R-square explains 0.063 of the variation. The signs of the betas are in the expected direction:

• positive for risk refusal rate: keeping the other variables at fixed value, a one unit increase in risk refusal rate increases the odds of making a change to the risk settings by 1.031, i.e. 3.1%.

• negative for the delta of risk refusal rate: keeping the other variables at fixed value, a one unit decrease in delta value of risk refusal rate increases the odds of making a change to the risk settings by 0.99, i.e. 1%. Intuitively the sign of this relation can be explained as following; when merchants notice a decrease in refusals, which means an increase in conversion, they decide to make changes in order to adjust chargebacks (as, theoretically, they already achieved the desired result regarding refusals).

• positive for logarithm of average transaction value: keeping the other variables at fixed value, a 2.7-fold increase in the average transaction value increases the odds of making a change to the risk settings 1.031, i.e. 3.1%.

• positive for logarithm of total transactions: keeping the other variables at fixed value, a 2.7-fold increase in total transactions increases the odds of making a change to the risk settings by 1.21,

i.e. 21%.

• positive for information services sector: keeping the other variables at fixed value, the odds of making a change to the risk settings for information services sector over the odds of making a change to the risk settings for physical goods sector is 1.81, i.e. 81%. Since information services have zero marginal costs, it is expected that they behave differently (Ingenico Payment Services, 2015).

• positive for transportation services sector: keeping the other variables at fixed value, the odds of making a change to the risk settings for transportation services sector over the odds of making a change to the risk settings for physical goods sector is 1.77, i.e. 77%. Since airlines and ticket agencies offer products that are very popular to fraudster since they can resell them easily (Ingenico Payment Services, 2015), it is expected that this sector behaves differently.

• positive for unknown sector: keeping the other variables at fixed value, the odds of making a change to the risk settings for unknown services sector over the odds of making a change to the risk settings for physical goods sector is 1.79, i.e. 79%.

It should be noted at this point that we use the *logarithmic scale* for the variables total transactions and average transaction value. The intuition behind this, is that the magnitude of these variables is more important than the actual values. Thus we are not interested in e.g. an actual increase of 1 euro, but in the order of magnitude.

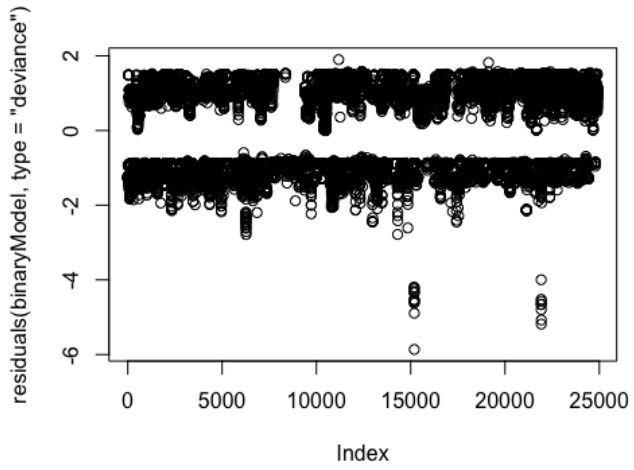
Fit of the model. One way to check the fitting of the model is through the analysis of deviance table which can be seen below. Noticing the difference between the null deviance and the residual deviance column indicates how our model compares to the so-called null model, i.e a model with only the intercept. The wider this difference, the better the fitting. Analyzing the table shows us the reduction in deviance when adding each variable, one at a time. We see that by adding risk refusal rate, the residual deviance is reduced by almost 200. The variables delta in refusal rate and total transactions only slightly decrease the deviance. Lastly, sector causes a reduction of 251 points. The other variables seem to improve the model even less. What should be noted here is that a large p-value in this table shows that the model without the variable explains more or less the same amount of variation.

Figure 5.26: The deviance table for the Logistic regression.

```
Analysis of Deviance Table

Model: binomial, link: logit

Response: binary.change

Terms added sequentially (first to last)

                   Df Deviance Resid. Df Resid. Dev             Pr(>Chi)
NULL                                24923      16361
riskrefusalrate     1  168.654     24922      16192 < 0.00000000000000022 ***
chargebackrate      1    0.004     24921      16192               0.95026
delta.cb            1    0.423     24920      16192               0.51568
delta.ref           1    4.462     24919      16187               0.03465 *
average.tx.value    1    1.841     24918      16185               0.17485
total.transactions  1   56.667     24917      16129   0.00000000000005161 ***
sector              4  250.260     24913      15878 < 0.00000000000000022 ***
lagged.cb           1    0.196     24912      15878               0.65803
lagged.ref          1    0.376     24911      15878               0.53978
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Often in the regression analysis we also see the so-called "regression diagnostics", i.e. some tests to check how well the model fits the data. In GLM models the ordinary assumptions of linear regression models, such as normality and homoskedasticity do not apply. An often used plot in GLM models is the deviance residual plot. A good fit would be indicated by a deviance residual plot where all the points are randomly dispersed around zero. In Figure 5.27, we see a not so random dispersion. This means that there is quite some unexplained variance.

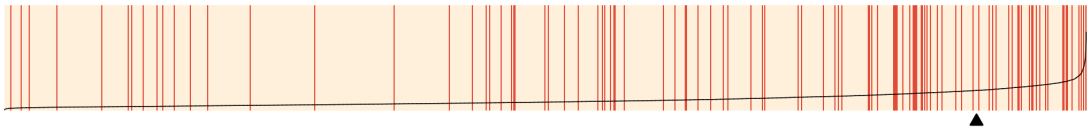
Figure 5.27: Deviance residual plot



Furthermore, for the logistic regression we also assess its predictive probability. In order to do so, we split the dataset into a "train" and a "test" dataset, based on the time frame. As such, the training dataset contains records for the first six months, while the test dataset contains records for the remaining six months. The accuracy obtained on the test set is 0.904 which is a pretty good result. However, bear in mind that this is dependent on the manual split made to the data earlier.

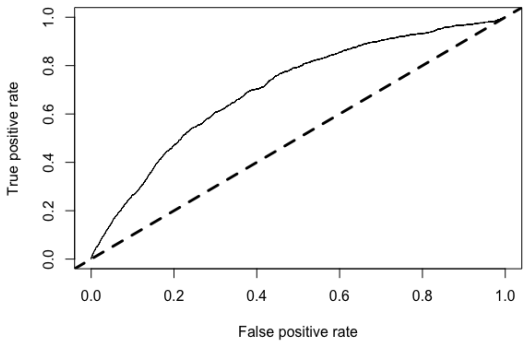
Performance Measurement. Finally, in order to assess the performance of the model, we plot the separation plot, as well as the ROC curve and calculate the area under the curve (AUC). Through the separation plot we can visually detect whether the model accurately matches high probability predictions to actual event occurrences, as well as low probability predictions to non-event occurrences. The separation plot as shown in Figure 5.28 shows a clustering of events on the right and a clustering of non-events on the left to middle; however there are event lines everywhere.

Figure 5.28: Diagnostic of logistic regression with separation plot.



On the other hand, the ROC is a graph that shows the true positive rate, i.e. positives that are correctly identified as positives, against the false positive rate, i.e. positives that are incorrectly identified as negatives (Figure 5.29) at various threshold settings. The AUC is the area under the ROC curve. A rule of thumb is that a model with a high predictive ability should have an AUC close to 1 (1 being the ideal) than to 0.5. The AUC obtained for our model equals 0.70, which indicates a rather weak accuracy.

Figure 5.29: The ROC curve.



Additionally, we mentioned earlier that McFadden's pseudo R-square equals 0.063, which can be assessed as a low value indicating much unexplained variance. This practically means that all the hypotheses together cannot explain sufficiently merchants' behavior. The reader should bear in mind however that R-square in GLM models are not a powerful measurement of goodness-of-fit (Hosmer & Lemeshow, 2000), and thus should be used in combination with AUC and AIC to derive conclusions.

Multicollinearity. The term multicollinearity denotes whether groups of more than two independent variables are high correlated. In order to check multicollinearity, we used the Variance Inflation Factor (VIF), where a low value - i.e. close to 1- indicates low multicollinearity. The results can be seen in Table 5.12.

Table 5.12: The VIF values for the independent variables.

Variable	**VIF**
chargeback_rate	1.488182
riskrefusal_rate	1.477796
delta_cb	1.485906
delta_ref	1.363859
lagged_cb	1.244210
lagged_ref	1.164612
average_tx_value	1.147958
total_transactions	1.022507

Altering the model using interaction terms

After running the initial model, we want to further experiment with it and increase its explanatory power, i.e. reduce AIC and increase R-squared. At first place we re-run the model by creating a second version of the dummy variable "sector", where we distinguish between "information services" and combine all the other categories to the level "other sectors". The intuition behind this is that information services can really behave differently from the other sectors since they have zero marginal costs.

Additionally, we can use *interaction terms* between the dummy variable (information services vs other sector) and all the other independent ones. The intuition behind this is that these terms, which all reflect aspects of the environment and characteristics of merchants, are influenced by the sector and are thus associated. Interaction terms, which are multiplication of such terms, are a common way to handle these parallel relations in regression analysis (the same model was ran by using transportation services vs other sectors and the results were the same). Finally, we excluded the chargeback rate, delta of chargeback rate and the lagged variables as they seem to not have a significant impact on the model according to the deviance table (Figure 5.26). The output of making the aforementioned modifications to the model is shown in Figure 5.30.

Figure 5.30: The modified logistic regression model with interaction terms.

```
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.4991  -0.5128  -0.3948  -0.2790   3.7454

Coefficients:
                                                 Estimate Std. Error z value Pr(>|z|)
(Intercept)                                     -5.309798   0.174176 -30.485  < 2e-16 ***
riskrefusalrate                                  0.034002   0.003218  10.568  < 2e-16 ***
sector.dummy2information services                2.555914   0.217615  11.745  < 2e-16 ***
delta.ref                                       -0.006984   0.004744  -1.472  0.14097
log.avg.value                                    0.198210   0.026422   7.502 6.30e-14 ***
log.transactions                                 0.253884   0.013347  19.022  < 2e-16 ***
riskrefusalrate:sector.dummy2information services -0.011817   0.004346  -2.719  0.00655 **
sector.dummy2information services:delta.ref      0.002579   0.006263   0.412  0.68051
sector.dummy2information services:log.avg.value -0.278880   0.032861  -8.487  < 2e-16 ***
sector.dummy2information services:log.transactions -0.109256   0.016922  -6.457 1.07e-10 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 16361  on 24923  degrees of freedom
Residual deviance: 15274  on 24914  degrees of freedom
AIC: 15294

Number of Fisher Scoring iterations: 6
```

McFadden's pseudo R-square: 0.066

Figure 5.31: The odds of coefficients for the altered Logistic regression.

```
                                (Intercept)                                 riskrefusalrate
                                  0.004942924                                     1.034587165
                sector.dummy2information services                                    delta.ref
                                 12.883065903                                     0.993040740
                                log.avg.value                                  log.transactions
                                  1.219218978                                     1.289022254
riskrefusalrate:sector.dummy2information services       sector.dummy2information services:delta.ref
                                  0.988252933                                     1.002582339
  sector.dummy2information services:log.avg.value sector.dummy2information services:log.transactions
                                  0.756631046                                     0.896501001
```

One thing to notice from the Figure is that AIC is reduced from 15342 to 15294 and R-squared is increased from 0.063 to 0.066, which shows a slightly better model. Apart from that, we notice that the delta of refusal rate is not significant anymore, while the following significant interactions appear:

- sector and risk refusal rate
- sector and average transaction value
- sector and total transactions

5.4.2 Predicting the net score variable

For our third dependent variable a linear regression model is being used. Remember that with this regression model, we are trying to capture the direction of score changes (i.e. increase versus decrease) only for the merchants that made changes to their risk settings. The results for this model as produced by R can be seen in Figure 5.32.

Figure 5.32: The OLS regression output for the net score variable.

```
Residuals:
     Min       1Q   Median       3Q      Max
-0.89534 -0.02085  0.01018  0.03374  0.95436

Coefficients:
                                Estimate Std. Error t value Pr(>|t|)
riskrefusalrate                0.0004833  0.0002679   1.804   0.0714 .
chargebackrate                 0.0042480  0.0018349   2.315   0.0207 *
log.avg.value                 -0.0119783  0.0013488  -8.880  < 2e-16 ***
delta.cb                      -0.0009773  0.0018913  -0.517   0.6054
delta.ref                     -0.0001312  0.0003311  -0.396   0.6921
lagged.cb                     -0.0017867  0.0018721  -0.954   0.3400
lagged.ref                    -0.0001726  0.0003125  -0.552   0.5809
log.transactions              -0.0007596  0.0008304  -0.915   0.3604
sectorhotel services           0.0579754  0.0227843   2.545   0.0110 *
sectorinformation services     0.0419235  0.0101672   4.123 3.89e-05 ***
sectorphysical goods           0.0374523  0.0114549   3.270   0.0011 **
sectortransportation services -0.0906122  0.0123342  -7.346 2.96e-13 ***
sectorunknown                  0.0647684  0.0482647   1.342   0.1798
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.105 on 1990 degrees of freedom
Multiple R-squared:  0.2403,    Adjusted R-squared:  0.2353
F-statistic: 48.41 on 13 and 1990 DF,  p-value: < 2.2e-16
```

In our last regression model we can see that seven out of the 12 coefficients are significant (significance levels can be seen in Figure 5.32). The model explains 0.24 of the variation. The sign of the seven betas is as follows:

• positive for risk refusal rate: keeping the other variables at fixed value, a 1% increase in overall risk refusal rate results in merchants increasing the net scores by 0.0004.

• positive for chargeback rate: keeping the other variables at fixed value, a 1% increase in overall chargeback rate results in merchants increasing the scores by 0.0042.

• negative for average transaction value: keeping the other variables at fixed value, one percent increase in the average transaction value results in merchants decreasing the net scores by 0.000119.

• positive for hotel services: keeping the other variables at fixed value, merchants that belong to hotel services industry increase the scores by 0.057.

• positive for information services: keeping the other variables at fixed value, merchants that

belong to information services industry increase the scores by 0.041.

• positive for physical goods sector: keeping the other variables at fixed value, merchants that belong to physical goods industry increase the scores by 0.037.

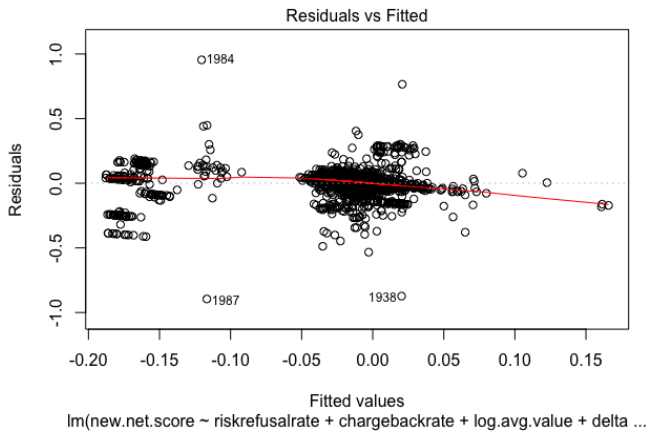
• negative for transportation services sector: keeping the other variables at fixed value, merchants that belong to transportation services decrease the net score by 0.09.

We notice that in this model, the overall chargeback rate becomes significant compared to the Logistic regression model, however the size of the coefficient is quite small and therefore the effect is not as substantial as one might expect. Apart from that, we see that the transaction volume in this model is not significant, and hence not influencing the direction of changes merchants make. This finding might be expected, since solely the transaction volume cannot be a determinant of increasing or decreasing scores when the other variables are kept constant.

Regression Diagnostics. In order to assess how well our data meet the conditions of the OLS regression, we run the regression diagnostics. These include assumptions about linearity of relations, normality of errors and homogeneity of variance. We also look at influential points, as well as multicollinearity issues.

• Linear relationships: The linearity between the dependent and independent variables in a multivariate model can be checked through plotting the predicted values against the standardized residuals. Figure 5.33 shows some random dispersion of the points, which is indicative of linear relations.

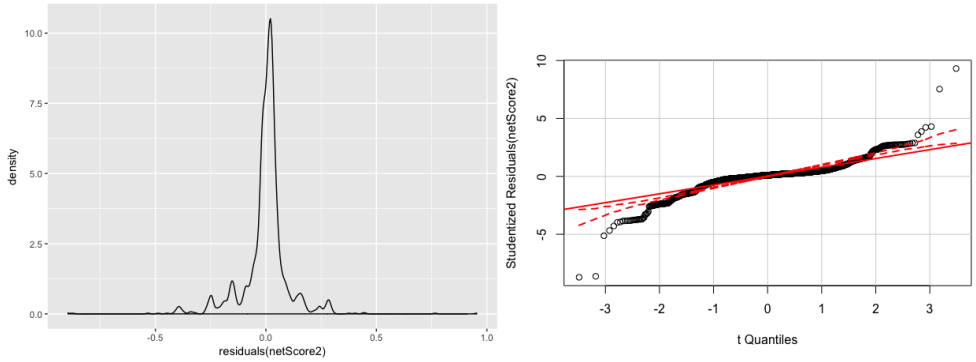
Figure 5.33: Predicted values against the standardized residuals.



• Normality of errors: We check the normality through the kernel density and QQ plots. both of the Figures 5.34a and 5.34b suggest some deviation of the normal distribution, denoted by the kinks in the density plot and the deviation from the red line in QQ plot.

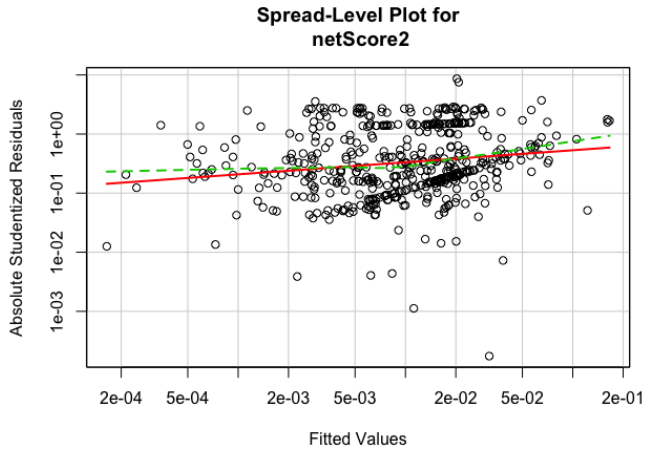
Figure 5.34: Plots checking the independence of errors.

(a) The density plot. (b) The qq plot.



• Homoskedasticity: In order to check whether the error variance is constant, we plot the fitted values against the residuals. Since no pattern can be identified in Figure 5.35, we conclude that there is no heteroskedasticity in our data.

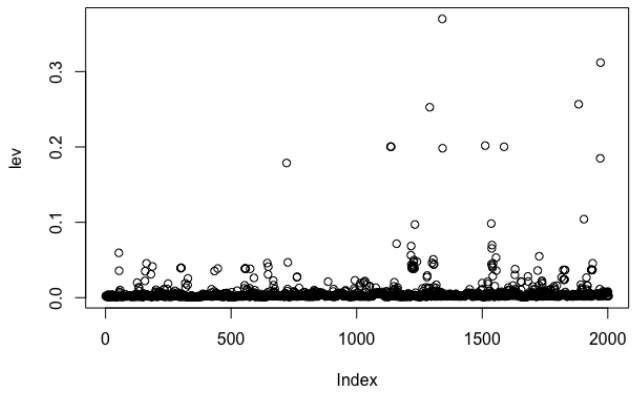
Figure 5.35: Fitted values against residuals.



• Influential observations: The leverage of an observation measures its ability to move the regression line all by itself. The leverage measures the amount by which the predicted value

would change if the observation was shifted one unit in the y direction. Observations can take leverage values between zero and one; all observations with zero leverage have no effect on the regression model. We can see from Figure 5.36 that there are some influential points with value greater than 0.2.

Figure 5.36: Leverage versus squared-residual plot.



• Multicollinearity: Lastly, we check whether there is correlation between the independent variables in the model. This is done by using the Variance Inflation Factor (VIF), as can be seen in Figure 5.37. We can see from the picture that there are no multicollinearity issues, since VIF is close to 1 for all the numeric independent variables.

Figure 5.37: The VIF of the independent variables for the model.

```
 riskrefusalrate   chargebackrate    log.avg.value         delta.cb        delta.ref        lagged.cb       lagged.ref
        1.447071         1.746711         1.144098         1.779206         1.334562         1.242261         1.164017
log.transactions
        1.135563
```

5.5 Threats to Validity

The limitations in this Chapter have mainly to do with the multivariate regression analysis. Regressions, as a statistical tool, have quite some uses such as for example providing functional relationships among variables. However, the limitations should also be carefully considered.

5.5.1 Construct Validity

The independent variables chosen for the models were considered as the ones with the highest relevance, given the scope of our research and the respective literature. Nevertheless, we cannot guarantee that the choice of more/other variables would have worse results than the ones currently obtained. Since we are trying to predict human behavior, plausibly we could be able to explain more variance by including variables such as the country where merchants are predominantly active, or the liability of costs related to fraudulent transactions. However the complexity of such variables is quite high and the time resources we have are not unlimited. For example, the majority of merchants are active in multiple countries, hence specifying one country per merchant would require checking where the largest part of transactions comes from. Apart from that, it could be the case that merchants might have an equally distributed country of transactions origin. On the other hand, liability is dependent on each individual transaction and hence specifying whether a merchant is held liable would require huge aggregations. Therefore, we are confident that given the time and data constraints, we chose the most relevant independent variables.

Moreover, it should be noted that there is a bias in the data regarding chargebacks. The information about chargebacks is being requested from the merchants by Adyen, who might provide it at any point in time. As such, merchants can "upload" chargebacks on the platform that were received several months ago, thus creating inconsistencies on the actual dates that these payments were received. This can plausibly be the reason that this variable seems to have no significant influence on merchants' risk decisions.

5.5.2 Internal Validity

The regression models were chosen based on the distribution of the dependent variables. However, for the logistic regression we selected our sample based on whether it represents a good candidate group to use the tool; taking a broader sample might have affected the results. Apart from that the number of merchants who did not make changes to their risk settings was disproportionally larger than those who did not. We also saw that AUC equals 0.70, indicating a weak accuracy.

Regarding the regression about the direction of changes, we see that R-square is 0.24. This might be due to the fact that, as discussed, more complex independent variables could have been included to the model. Moreover, the selection of the sample was based on the population of merchants who are making changes, leaving out those that do not. Nevertheless, this sample was again selected out of the merchants that represent the target group for the tool, as this would be

more valuable to Adyen. Selecting the sample throughout the whole merchant population might have produced different results. Lastly, we noticed some influential observations; due to time constraints, this was not further addressed.

For both of the models, as already mentioned, the chargeback variable was found to be biased.

It should also be noted that other analysis methods might also be appropriate. For instance, decision trees or random forests could also be used in order to explain the non-linear relationships between the dependent and independent variables.

5.6 Conclusions

In this Chapter we used empirical research in order to delve deeper into how merchants make risk assessment. As such, we firstly identified the group for which the tool adds more value by calculating the percent of improvement it suggests in both chargebacks and risk refusals, as well as by setting thresholds in the number of years processing and transaction volume so that the tool is given enough information to be able to run the calculations. This filtering resulted in 2077 merchant accounts representing the good candidates to use the tool. Next, we were interested in how these merchants engage with risk management. This was investigated initially through descriptive statistics and individual hypothesis testing. The main results are summarized in Table 5.13.

Table 5.13: Summary of results obtained through the descriptive statistics.

Method	**Findings**
Descriptive Statistics	Merchants make on average 1.35 changes per month in their risk settings
	Merchants visit their risk settings in order to make changes 2-3 times out of a year
	All the risk checks are being changed throughout the year
	A neutral net change in the scores of the risk checks is given by the majority of merchants, denoting a risk-averse attitude
	Merchants tend to ignore the suggestions of the Risk Calculator, but when they make changes in their risk settings, these are in line with the tool's suggestions

Apart from that, a complementary multivariate regression analysis was performed in order to test the effect of all the variables together. As explained, we used three different dependent variables in order to predict risk activity; (1) a binary variable indicating whether any change in the risk

settings occurred per month or not, (2) the count of risk checks changed per month, (3) the net score given to the risk checks that were changed per month. The first two metrics were highly correlated (0.99), hence only the results for the binary variable were included in this report. The third metric aims to capture the behavior only for the population that is making risk changes. The results of the hypotheses for the binary change variable are summarized in Table 5.14.

Table 5.14: Summary of the hypothesis results regarding whether merchants engage with risk management or not.

Hypothesis		**Logistic Regression**	**Logistic (Interaction Terms)**	**Individual Hypothesis**
H1	An increase in delta of chargeback rate increases merchants engagement with risk management	False	False	False
	An increase in delta of refusal rate increases merchants engagement with risk management	False	False	False
H2	An increase in overall chargeback rate increases merchants engagement with risk management	False	False	True
	An increase in overall refusal rate increases merchants engagement with risk management	True	True	True
H3	An increase in average transaction value increases merchants engagement with risk management	True	True	False
H4	Merchants significantly differ in terms of engagement with risk management, depending on their size	True	True	True
H5	Merchants significantly differ in terms of engagement with risk management, depending on the sector they belong	True	True	True

At first glance, for hypotheses H2 and H3 we see that there are some contradictory results between the regression and individual hypothesis methods. For H2, regarding the chargeback rate is was found to have no significant impact in the regression models, however the Spearman correlation indicated a weak positive relation of 0.16. Since the rho is weak, we can conclude that indeed the chargeback rate does not have a significant impact on whether merchants make changes to their settings or not. Regarding H3, we see that the average transaction value is found to be significant in the regression models, but not through the Spearman correlation. In this case, we can give more weight to the regression results and conclude that the average transaction value influences engagement with risk management but in a lower extent than the other significant variables, as also determined by the size of the coefficient.

Similarly, the results about the direction of changes, for the population that engages with risk assessment, can be seen in Table 5.15. Again, for H2 we notice that the Spearman correlation indicates a significant, however weak relation between the direction of changes and chargeback, as well as refusal rate. The respective rho equals to 0.12 for chargeback and 0.15 for refusal rate.

As the correlation is weak, we can conclude that eventually these two variables do not have a significant effect on the direction of scores.

Table 5.15: Summary of the hypothesis results regarding the direction of changes that merchants make.

Hypothesis		OLS	Individual Hypothesis
H1	An increase in delta of chargeback rate increases the net score given to risk checks by merchants	False	False
	An increase in delta of refusal rate increases the net score given to risk checks by merchants	False	False
H2	An increase in overall chargeback rate increases the net score given to risk checks by merchants	False	True
	An increase in overall refusal rate increases the net score given to risk checks by merchants	False	True
H3	An increase in average transaction value increases the net score given to risk checks by merchants	True	True
H4	The direction of score changes significantly differs depending on merchants' size	False	False
H5	The direction of score changes significantly differs depending on merchants' sector	True	True

From the above, we can conclude the following. On one hand, the **sector** and the **transaction volume** are the factors influencing most whether merchants make changes to their risk settings or not. More specifically, the Information Services and Transportation industries are the ones tweaking more their risk settings. This can be explained by the fact that information services have zero marginal costs as they sell digital products/services and hence are not affected in terms of losses compared to e.g. physical goods industry. Regarding transportation industry it is a sector that is heavily attacked by fraudsters (Ingenico Payment Services, 2015) and hence has developed awareness and maturity in trying to combat online fraud. In relation to transaction volume, larger merchants have more resources - such as stuff and possibly financial tolerance - and therefore are more proactive in making risk choices. In addition to sector and size, the overall **refusal rate** and the **average transaction value** seem to also influence whether merchants engage with risk management or not, however in a lower extent than the sector and size. Particularly, a higher refusal rate triggers more changes in merchants settings as merchants seems to be highly interested in conversion (Wolters, 2012; LexisNexis, 2015; Ingenico Payment Services, 2015), a fact that was also verified through the interviews. Lastly, a higher transaction value increases the probability of engaging more in risk changes. This is something expected, since the larger the

transaction value, the more profit the merchant is losing.

On the other hand, the **sector** seems to be the most significant factor in the direction of scores that merchants who engage with risk management give. Merchants in transportation sector seem to decrease the scores, i.e. being more loose with their settings and thus willing to take more risk. This can possibly be explained by two facts. On one hand, it has been noticed that approximately 27% of the transactions in the airline industry are filtered as potential fraud and hence need to be sent for manual review, which is a time-consuming and expensive task in this industry and nonetheless results in considerable false positives (Eaton-Cardone, 2016). Additionally, around 3.5% of all flight bookings are rejected by fraud filters, however many of those declined transactions are actually legitimate. Apart from that, nearly 40% cardholders abandon their transaction after being falsely declined. This results in airline merchants simply accepting suspicious transactions and absorbing the chargeback losses in an attempt to avoid declining transactions and losing out to the heavy competition. On the other hand, the tickets bought by fraudsters instead of being sold to the black market can also be used by the fraudster themselves; this means that the person has to be physically present, making it more probable of getting caught. This might be a reassuring fact for the transportation industry, thus being more lenient towards risk.

Moreover, we see that an increase in **average transaction value** leads to a decrease in scores. At first thought this is counter-intuitive; the higher the product/service value, the more revenue merchants lose. However, research (Sift Science, 2015) indicates that amounts of zero to 20 dollars have higher fraud rate; have the highest fraud rate. A purchase with average transaction value of $20 or less makes someone 2.16 times more likely to be a fraudster. In this way, fraudsters can actually test the stolen cards and then proceed to purchases with larger values. This might provide a justification of the fact that merchants seem to increase the scores when the average transaction value decreases.

One interesting finding is the fact that chargeback and refusal rate do not seem to have much of an impact on merchants' behavior. Contrary to our expectations, it seems that these variables are not the main drivers. Regarding chargebacks specifically, as it was mentioned under the Threats to Validity Section, there is some kind of systematic bias in the way they are stored in the database, hence probably affecting the results regarding its significance. On the other hand, refusals seem to have an impact on whether merchants make risk changes, but not on the direction of score. Logically speaking, an increased refusal rate denotes that the merchant is losing legitimate customers and thus profits; in this way the merchant would probably want to decrease the risk scores in order to allow more traffic. However, as noted during the interviews

merchants do not separately look at refusals and chargebacks, but they try to achieve a balance between the two and moreover they adopt a wait-ad-see attitude so that they can check the effect of the changes they made.

Taking a look back at our research question in this Chapter *"By looking at broader patterns, which factors can we identify that are indicative of merchants' engagement with risk management?"*, we can state that the type of merchant as well as the risk environment seem to be the most important drivers of risk engagement and choices. The sector, depending on the marginal costs and maturity level, can trigger the reaction of merchants by making them engage with risk decisions or not, while it can also denote the attitude towards risk. Furthermore, the size of the merchant, the average transaction value and the refusal rate are also indicators on whether merchants are willing to do risk assessment. Therefore, anti-fraud tools should take into account the type of merchants that use it, as well as the risk environment, and make different suggestions for optimization based on merchant classification.

Chapter 6

Discussion and Conclusions

This is the concluding chapter of the research followed in order to provide an answer to the research questions posed in the Introduction Chapter. Here, we will recall all the research questions and based on the results of the data analysis and interviews, or phrased differently on the empirical evidence, we will offer the answers. The main research question was formulated as follows:

Why are merchants reluctant to adopt profit-maximizing risk management settings that are suggested by an anti-fraud tool that analyzes their transaction data and how can developers of such tools increase their acceptance?

Our answer to the main research question consists of the following parts:

• Reviewing the context of e-commerce fraud, the solutions offered based on data mining and machine learning, as well as people's decisions as captured by Rational Choice Theory and Behavioral Economics.

• Conducting statistical analysis to datasets related to users, as well as non-users, of the anti-fraud tool Risk Calculator in Adyen.

• Interviewing merchants and account managers in order to capture their attitude towards engaging with risk-management. Interviews with developers within Adyen were also conducted.

• Developing a monitoring mechanism for extracting the data produced by the Risk Calculator.

• Conducting a statistical and multivariate regression analysis on the extracted data.

This Section is structured as follows. Firstly, we review the findings and offer a summary. Secondly, we provide recommendations and discuss the implications. Finally, we provide some insights about future work.

6.1 Findings on Analysis of Historical Data

Through the descriptive statistics of the initially obtained datasets, we compared the users of the Risk Calculator versus the non-users. Main limitation in this part was the lack of enough information in order to classify "users". More specifically, by intuition we would define a user as a merchant who is visiting the tool and then *applying* its suggestions. However, for a fairly long period of time we saw from the log files that no merchant was actually applying the suggestions. Therefore, we defined users as merchants who visited the tool, and moreover split them in two groups, i.e. high visitors and low visitors depending the number of times they visited the tool. For the non-visitors, a random sample was chosen. The main results from the descriptive analysis and the comparison of means are summarized in Table 6.1.

Table 6.1: Summary of the main findings of Chapter 3.

Category	**Findings**
Merchants Characteristics	Users and non-users appeared to be similar in terms of industry, with the largest category being Retailers followed by Services
	Users appear to be larger in size than non-users.
	Users (on a compny level) have more merchant accounts than the non-users.
Merchants Behavior	Users have larger refusal and chargeback rates than the non-users. Moreover, non-users have a median value of zero for both refusal and chargeback rates.
	The non-users group has on average less transactions in quantity than the users group.
	The non-users have the same average transaction value with the users.
	According to Spearman's rho, no significant correlation was found between the number of changes in risk settings and monthly increase/decrease in chargebacks and refusals.
Tool's Suggestions	The projected savings of users are slightly larger than the projected savings of non-users, however both amount for a very small fraction of their revenues.
	For the majority of the merchants the suggested risk profile is risk-taking, since the tools suggestions indicate decrease in false positives and no change in chargebacks.

Firstly, we saw that the merchants that visit more often the tool fall in the categories of Retailers and Services. The majority of them are large companies, i.e. ranging from 100000 to

1000000 transactions for a 9-month period. The most frequent visitor visited the tool 126 times during a 3-month period, while there were also merchants that visited the page once during this period.Therefore, at first glance, there is some type of engagement with the tool; nonetheless merchants' trust in it is under question since nobody is using the tool as a risk setter. By comparing the visitors group with non-users of the tool, i.e. merchants that have never visited the page, we found that non-users are smaller companies than visitors in terms of transaction volume, and additionally have lower chargeback and refusal rates. Since the Risk Calculator is a statistical tool that needs certain volume of historical data in order to run the calculations, the aforementioned findings suggest that the "good candidates" are indeed visiting the tool, or at least have seen it once, while there is a group of merchants for which the tool adds no value and hence they are not interested in using it.

Secondly, we saw that although merchants do not directly interact with the tool's suggestions, by means that they do not apply the suggested changes through the tool, they do change the risk scores through their settings page. Hence, we notice a willingness to engage with risk-management, nonetheless not by using the tool. For the visitors of the tool, we noticed that in this 3-month period the majority increased the scores they gave to the risk checks, however the suggestions that the tool made were more towards reducing false positives, i.e. advising merchants to reduce the scores and thus become more risk-taking.

Linking the above findings to the literature, according to Prospect Theory it is important to understand the general context that influences decision making; if merchants are already doing well in terms of refused and fraudulent transactions then they do not have particular interest in using such a tool. Moreover, according to Status Quo Bias people prefer not to change behavior unless the incentive to do so is strong; in this case, non-users seem to have already optimized transactions. Lastly, Behavioral Economics suggest that people tend to reject advices that are not in line with their risk attitude; as such, a risk-averse merchant would not accept applying a risk-taking advice.

By reflecting on these results, we are able to provide an answer to the respective sub-question which was formulated as follows:

What does the historical data indicate about the usage of Risk Calculator and the risk settings adopted by merchants?

Summarizing, the merchants who are visiting the tool are the ones who meet the requirements for

triggering the tool's calculations. Moreover, they seem to have higher chargebacks and refusals than the non-users and they mainly belong to retailers and information services industry. Although they visit the tool on average 6 times during a 3-month period, no one is actually applying the suggested changes to the risk scores. Nevertheless, merchants do change their scores through the settings page and the majority of them is increasing the scores thus denoting a risk-averse attitude. This is not in line with the suggestions of the Risk Calculator, which most of the times suggests to decrease the scores in order to reduce false positives. Additionally, the financial savings that the tools suggests account for a small fragment of merchants' revenues.

6.2 Findings on Interviews

The qualitative analysis followed in Chapter 4, helped us get deeper insights on people's behaviour regarding not only the specific anti-fraud tool, but also their risk decisions in general. During this Chapter we also compared merchants' opinions to those of account managers and developers in Adyen.The findings of this chapter are summarized below in Table 6.2.

The results of the interviews which were analyzed by means of open coding, gave us the opportunity to organize respondents' answers.

Firstly, the most prevalent challenges of the tool seem to fall both in the technical and non-technical sphere. Specifically, we found out that the quality of output in terms of user interface and technical bugs, the insufficient documentation, the response time, interaction with it by the use of the slider, trust and the belief if it adds value to the business are the main challenges related to the tool's usage. Regarding the complexity of understanding how to use the tool, it was mentioned that people are not willing to spend time on it since it costs considerable effort and time. Apart from that, the indication of no trust in the underlying model is reflected on one hand in the fact that all the interviewees stated they prefer a manual analysis since they consider the costs related to their transactions "too important to let a machine decide. On the other hand, the non-trust is derived from the fact that the tool suggests extreme changes, i.e. putting a score from 100 to 0, to risk checks that are perceived by merchants very efficient in combating fraud. This finding might be a suggestion for the design of the tool, by giving the option to merchants to selectively apply the advices they consider most valuable. Regarding merchants' risk profile, it was mentioned during the interviews that merchants try to achieve a balance between chargebacks and refusals, since if they have under control only the one variable, then they are losing a considerable amount on the other variable as there is always a trade-off.

Table 6.2: Summary of the main findings of Chapter 4.

Respondents	Findings
Merchants	All the interviewees mentioned they are manually reviewing their transactions in order to take risk decisions.
	All the interviewees mentioned they prefer to use the tool as an adviser than as risk-setter.
	82% of the interviewees mentioned they try to balance chargebacks and conversion when taking risk decisions, hence they are classified as risk-averse.
	63% of the interviewees was highly involved with risk management (4-6 hours per day).
	27% of the interviewees mentioned they trust more human than automated machine decisions.
	18% held the belief that risk mitigation is not in PSPs area of expertise.
	18% had entered the Excessive Chargeback Program imposed by schemes, hence were classified as risk-taking.
	The main feedback of the merchants who tried to use the tool was that they do not trust the suggestions, there is insufficient information on how to use the tool both in terms of interface and documentation and finally there is insufficient promotion.
Account Managers	Do not trust the underlying model of the tool.
	Believe that the tool is too complex to use it.
	Believe that the tool suggests too extreme changes in the risk scores.
Developers	Believe that the model is mathematically accurate.
	Find as a disadvantage that the tool is solely score-based and believe it should also take into accounts the features of the risk parameters per se.
	Mention as a challenge the fact that it is not possible to predict the final state of initially blocked transactions.
	Believe that the usability could be improved.

It should be noted at this point that there was some contradictory feedback between developers in the company and merchants, as well as account managers regarding the reliability of the tool. This might be reflected in the literature regarding alignment of incentives and different perceptions between various actors; on one hand, developers are interested in company's infrastructure expansion, thus the tool is an additional product that they can work on. On the other hand, account managers are more interested in potential loss of their reputation, since if the tool does not eventually work as expected, the merchants might no longer trust them; this is also reflected in their reluctance to promote the tool. Lastly, the merchants are the ones to be incurring the direct costs of chargebacks and refusals if the tool proves to be invalid.

A reflection on the above findings indicates that they are in line with the literature. On one hand, the main challenges with the usage of the Risk Calculator fall in the literature discussing challenges in user adoption of data-mining tools, which suggests that the intentions of users are

mainly related to its perceived usefulness and perceived ease of usage. Apart from that, we see the notion of Bounded rationality Theory which suggests that people's decisions are dependent on time and information constraints; this is reflected in the fact that merchants do not wish to spend a lot of time in understanding how the tool works. Furthermore, we see that the psychological cost of loss is greater than the psychological benefit of gain, thus people prefer to put more energy by doing manual review of the transactions against using the automated tool in order to be sure they avoid losses; this is underlined in the concept of loss aversion.

Overall, we can see from the qualitative analysis that merchants are interested in engaging with risk management, however the tool does not fully enable the features they would like to see. The above technical and non-technical issues influence engagement with the tool and can be regarded as the stepping stone for further improvements.

To recall, the research question in this Chapter was formulated as follows:

How do merchants and account managers engage with the tool and how do they choose their risk profile?

In summary, both merchants and account managers do not totally trust the underlying model of the tool and hence they are reluctant in applying the suggested changes. Moreover, it was found that that the quality of output in terms of user interface and technical bugs, the insufficient documentation, the response time, interaction with it by the use of the slider and the belief if it adds value to the business are the main challenges related to the tool's usage. Finally, the choice of the risk profile to be adopted, depends on manual review of the transactions with the plan to keep in balance chargebacks as well as refusals.

6.3 Findings on Risk Behavior of Target Group

In Chapter 5 we focused on identifying a target group for which the tool adds more value and furthermore predicting their engagement with risk management, as well as the direction of the score change by building an empirical model.

In order to identify the group for which the tool would be more beneficial to use, we created two metrics: the *percentage of improvement in chargebacks* and the *percentage of improvement in risk refusals*. Subsequently, we triggered the calculations of the tool for each merchant account in

Adyen and we spotted the group for which the tool suggests an improvement in both rates. Apart from that, we applied some filters based on whether merchants have access to the tool, on their transaction volume and on the number of possible optimizations that the Risk Calculator suggests for them. Please recall that the definition of the target group should be re-evaluated in some fixed time intervals by Adyen, as the tool works based on historical data.

Additionally, we ran two regression models; one for predicting what drives the changes to merchants risk settings and one for predicting the direction (net increase or decrease) of the scores for the merchants that make changes. The results of the first regression model are presented in Table 6.3, while the results of the second model are presented in Table 6.4.

Table 6.3: Summary of the hypothesis results regarding whether merchants engage with risk management or not.

Hypothesis		**Logistic Regression**	**Logistic (Interaction Terms)**	**Individual Hypothesis**
H1	An increase in delta of chargeback rate increases merchants engagement with risk management	False	False	False
	An increase in delta of refusal rate increases merchants engagement with risk management	False	False	False
H2	An increase in overall chargeback rate increases merchants engagement with risk management	False	False	True
	An increase in overall refusal rate increases merchants engagement with risk management	True	True	True
H3	An increase in average transaction value increases merchants engagement with risk management	True	True	False
H4	Merchants significantly differ in terms of engagement with risk management, depending on their size	True	True	True
H5	Merchants significantly differ in terms of engagement with risk management, depending on the sector they belong	True	True	True

Table 6.4: Summary of the hypothesis results regarding the direction of changes that merchants make.

Hypothesis		OLS	Individual Hypothesis
H1	An increase in delta of chargeback rate increases the net score given to risk checks by merchants	False	False
	An increase in delta of refusal rate increases the net score given to risk checks by merchants	False	False
H2	An increase in overall chargeback rate increases the net score given to risk checks by merchants	False	True
	An increase in overall refusal rate increases the net score given to risk checks by merchants	False	True
H3	An increase in average transaction value increases the net score given to risk checks by merchants	True	True
H4	The direction of score changes significantly differs depending on merchants' size	False	False
H5	The direction of score changes significantly differs depending on merchants' sector	True	True

From this Chapter the following findings can be inferred. Firstly, we saw that the merchant accounts that can theoretically use the tool in terms of access to the tool and time frame of processing transactions, are 2077. Narrowing it down, by taking into account that the tool needs enough data to run the calculations, we ended up with 1058 merchant accounts. Out of these, 279 are offered more than one opportunities for optimization (i.e. multiple slide configurations), while for 43 of them the tool suggests an improvement in both chargebacks and refusals. By exploring the behavior of the 2077 merchants, we saw that the majority of them makes changes in their risk settings 2-3 months throughout a year. Moreover, all the risk checks seem to have equal importance to those merchants, as all of them are being changed. The majority of the risk checks is given on average a positive score. However, when calculating the net score merchants give to their risk settings, it was found that the majority had a neutral change, i.e. merchants seem to prefer staying with the default choices. Finally, we saw that merchants tend to ignore the suggestions of the Risk Calculator, but when they change their risk scores the new values that they give are in most of the cases in line with the values suggested by the tool.

Linking the findings to the literature, we can see that according to Behavioral Economics, when people are faced with uncertainty, it is more likely that they will go with the default, especially when it is presented as a recommended configuration. This is reflected in the finding that for most of the merchants the net score change was equal to zero.

Secondly, from the regression analysis, the following can be derived. On one hand, the **sector** and the **transaction volume** are the factors influencing most whether merchants make changes to their risk settings or not. More specifically, the Information Services and Transportation industries are the ones tweaking more their risk settings. This can be explained by the fact that information services have zero marginal costs as they sell digital products/services and hence are not affected in terms of losses compared to e.g. physical goods industry. Regarding transportation industry it is a sector that is heavily attacked by fraudsters (Ingenico Payment Services, 2015) and hence has developed awareness and maturity in trying to combat online fraud. In relation to transaction volume, larger merchants have more resources - such as stuff and possibly financial tolerance - and therefore are more proactive in making risk choices. In addition to sector and size, the overall **refusal rate** and the **average transaction value** seem to also influence whether merchants engage with risk management or not, however in a lower extent than the sector and size. Particularly, a higher refusal rate triggers more changes in merchants settings as merchants seems to be highly interested in conversion (Wolters, 2012; LexisNexis, 2015; Ingenico Payment Services, 2015), a fact that was also verified through the interviews. Lastly, a higher transaction value increases the probability of engaging more in risk changes. This is something expected, since the larger the transaction value, the more profit the merchant is losing.

On the other hand, the **sector** seems to be the most significant factor in the direction of scores that merchants who engage with risk management give. Merchants in transportation sector seem to decrease the scores, i.e. being more loose with their settings and thus willing to take more risk. This can possibly be explained by two facts. On one hand, it has been noticed that approximately 27% of the transactions in the airline industry are filtered as potential fraud and hence need to be sent for manual review, which is a time-consuming and expensive task in this industry and nonetheless results in considerable false positives (Eaton-Cardone, 2016). Additionally, around 3.5% of all flight bookings are rejected by fraud filters, however many of those declined transactions are actually legitimate. Apart from that, nearly 40% cardholders abandon their transaction after being falsely declined. This results in airline merchants simply accepting suspicious transactions and absorbing the chargeback losses in an attempt to avoid declining transactions and losing out to the heavy competition. On the other hand, the tickets bought by fraudsters instead of being sold to the black market can also be used by the fraudster themselves; this means that the person has to be physically present, making it more probable of getting caught. This might be a reassuring fact for the transportation industry, thus being more lenient towards risk.

Moreover, we see that an increase in **average transaction value** leads to a decrease in scores.

At first thought this is counter-intuitive; the higher the product/service value, the more revenue merchants lose. However, research (Sift Science, 2015) indicates that amounts of zero to 20 dollars have higher fraud rate; have the highest fraud rate. A purchase with average transaction value of $20 or less makes someone 2.16 times more likely to be a fraudster. In this way, fraudsters can actually test the stolen cards and then proceed to purchases with larger values. This might provide a justification of the fact that merchants seem to increase the scores when the average transaction value decreases.

One interesting finding is the fact that chargeback and refusal rate do not seem to have much of an impact on merchants' behavior. Contrary to our expectations, it seems that these variables are not the main drivers. Regarding chargebacks specifically, as it was mentioned under the Threats to Validity Section, there is some kind of systematic bias in the way they are stored in the database, hence probably affecting the results regarding its significance. On the other hand, refusals seem to have an impact on whether merchants make risk changes, but not on the direction of score. Logically speaking, an increased refusal rate denotes that the merchant is losing legitimate customers and thus profits; in this way the merchant would probably want to decrease the risk scores in order to allow more traffic. However, as noted during the interviews merchants do not separately look at refusals and chargebacks, but they try to achieve a balance between the two and moreover they adopt a wait-ad-see attitude so that they can check the effect of the changes they made.

To recall, in this Chapter the analysis aimed at providing an answer to the following research sub-question:

By looking at broader patterns, which factors can we identify that are indicative of merchants' engagement with risk management?

In short, we can state that the type of merchant as well as the risk environment seem to be the most important drivers of risk engagement and choices. The sector, depending on the marginal costs and maturity level, can trigger the reaction of merchants by making them engage with risk decisions or not, while it can also denote the attitude towards risk. Furthermore, the size of the merchant, the average transaction value and the refusal rate are also indicators on whether merchants are willing to do risk assessment. Therefore, anti-fraud tools should take into account the type of merchants that use it, as well as the risk environment, and make different suggestions for optimization based on merchant classification.

6.4 Discussion

During this research we have argued that the problem of e-commerce fraud is mainly approached from a technical perspective, with academia and industry focusing on the creation of anti-fraud solutions based on machine learning, without paying much attention to the adoption of such solutions. However, as Anderson et al. (2013) states, most of the cybersecurity problems rely more often on economic and behavioral factors rather than on technical. As such, economic theories can offer explanations by pinpointing to actors' financial and behavioral drivers when technical solutions seem to fail.

Throughout the thesis, we have reviewed Behavioral Economics in order to study the factors that might provide explanations about how merchants make risk decisions regarding their transactions. We realized that theories in this field could provide insights related to merchants motivation in using profit-maximizing tools that aim to combat fraud. After reviewing the results, potentially two points of discussion could be raised based on the findings.

• The chargeback rate included in the regression models does not seem to be very important driver of merchants risk activity, based on the significance level and coefficients. Moreover, refusal rate was found to be significant only in logistic regression. Based on the interviews and the existing literature, we would expect that these two variables should have the largest impact on decisions regarding merchants' transactions. This does not mean that literature or our interviewees are incorrect, but it might suggest that predicting merchants behavior is a more complicated phenomenon which needs complementary theories or investigation of other information, that people might unconsciously think of when making decisions.

• The second part of the research question aims at suggesting improvements on increasing the acceptance of profit-maximizing anti-fraud tools. However, it should be critically reflected whether any efforts put into security countermeasures for financial fraud make sense in the context of Security Economics. As Herley (2009) mentions, the effort put by the end-user in terms of time and understanding of the advice, might cost more than the actual savings produced by the security tool. As such the author suggests that on one hand, the (potential) users of any security tools understand risks better than the party that provides the security tool. In our case study, for example, we do not have evidence that merchants who follow the suggestions of Risk Calculator fare better than the merchants who ignore it. On the contrary, we saw in Chapter 3 that non-users have actually lower chargeback and refusal rates. Moreover, it is usual to try to persuade users that if they do not follow the advice, the worst-case scenario might happen to them. Nonetheless, the

worst-case scenario is totally different than the actual harm they might get. If we consider for a moment the worst case scenario for not following the suggestions of the Risk Calculator, it is that merchants *might* end up with many chargebacks or refusals. However, we should bear in mind that the tool makes the calculations based on historical data and a critical question to be asked here is why does the past have to predict the future? Additionally, the amount of "lost" money shown in the tool is not actual revenue, but a lost opportunity to make profit. Finally, the time merchants spend in order to understand how the tool works should not be considered as a free resource. In the same study, Herley (2009) proves that time costs considerable amount of money, while during the interviews several merchants stretched the issue of the effort to understand how the tool works.

Given the above context, it should be well considered which of the suggested improvements might eventually be worth implementing. An essential thing to keep in mind is that as the complexity of the security advice increases, the willingness of the user to put effort in understanding it decreases.

6.5 Recommendations for Adyen

6.5.1 Identification of Improvements in the Tool

After the interviews, we are able to identify the improvements that can possibly be made to the tool. *"User stories" in agile software development* (Cockburn, 2006) can be very valuable in determining how the end product should look like, based on the end-users' needs. User stories are defined as short phrases denoting the functions that the system under development should provide, by capturing "who", "what" and "why" requirements with minimum level of detail (Cockburn, 2006). For the case of the Risk Calculator, the user stories are depicted in table 6.5.

Table 6.5: User Stories for the development of the Risk Calculator.

User	User Story
Merchants	As a potential user of the Risk Calculator, I want to have all the relevant information, so that I can understand by myself how to use it.
	As a potential user of the Risk Calculator, I want it to take into account apart from the scores the features of the risk parameters per se, so that it gives me broader advice about fraud.
	As a potential user of the Risk Calculator, I want to be able to interact with it by moving the slider, so that I can get feedback on my preferred variable.
	As a potential user of the Risk Calculator, I want to be able to perform A/B testing through it, so that I can check its validity.
	As a merchant, I want to use the Risk Calculator as an auxiliary tool rather than as a risk-setter, so that I can have control over my transactions
Risk Officer	As a risk officer, I want the Risk Calculator to suggest more gradual changes to the risk scores, so that I can trust its validity and promote it to merchants.

Based on the user stories presented above, the main improvements to be made to the tool include the following.

Embodiment of more information in the tool. A large percentage of the merchants that were interviewed indicated they need more explanation of how the tool works and how it can be used. Making the design of the Risk Engine Optimizer self-explanatory can reduce considerably the time needed to interpret its functionality and thus attract more users.

Alteration of algorithm's constraints. As already mentioned, several merchants perceived as a problem that the slider button does not move for them. Form a technical perspective, this means that the Simplex algorithm is able to find only one feasible solution for this group of merchants, hence offering only one optimization. One improvement that could be made here is to change the constraints of the algorithm so that for the same data it can find more than one feasible solutions.

Implementation of another algorithm. The Simplex algorithm is one of the most "basic"

algorithms, hence its power is limited. This is reflected in the fact that the changes it suggests to the risk scores are too extreme. Another machine learning algorithm could be more suitable in not only suggesting more gradual changes, but also taking into account more information apart from the scores of the risk checks.

Implementation of A/B testing through the tool. Since merchants seem to be highly concerned about the validity of the tool, applying A/B testing by sending one part of the shopper traffic to the suggested risk settings and keeping another part to the current settings could be a tangible way for the merchant to evaluate whether the suggestions are indeed optimal.

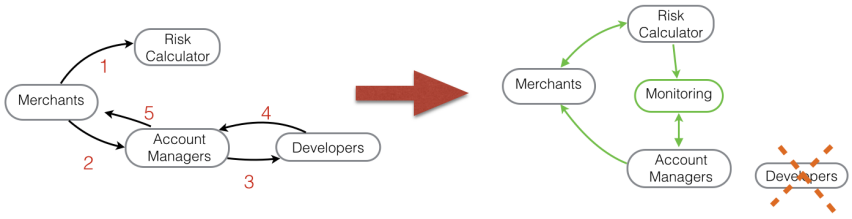
Redesign of the tool as an adviser. Since all the interviewees indicated that they are more interested in keeping the tool's suggestions as "a good piece of advice" rather than directly applying the actual changes, it might be wiser to change the context of the tool and re-model it as an adviser and not as a risk-setter. In this way, the tool can take into consideration broader information such as suspicious behavior versus normal behavior, or setting scores according to velocity rules and thus give recommendations on what actions need merchants to take.

6.5.2 Improvements in Internal Processes

Apart from that, at the initial stages of the research, we found some contradicting results related to the usage of the tool, as well as to merchants' feedback. If you recall, while selecting the groups of merchants to be interviewed, we identified the "high visitors" as a potentially interesting group to get feedback from. During the interviews, it emerged that merchants who were classified as high visitors supporting to have not visited the page of the tool. Apart from that, during Chapter 3 we saw that the calculations of the tool were not stored into the database, hence we had to manually collect data. Going a step further, merchants' feedback during the interviews seemed contradictory; for instance, merchants mentioned that they prefer manual review for their transactions but on the same time they would like to have an automated way to monitor risks. The aforementioned results highlighted the need for a mechanism that stores all the essential information related to the tool into the database. Since this mechanism was developed during this thesis, it is strongly recommended that Adyen uses it on e.g. monthly intervals in order to prevent unintended use and pro-actively engage with merchants. This will moreover simplify internal processes. As such, we are aware that at the moment this report is being written, when merchants wish to make changes to their risk scores, they contact their account managers who subsequently ask from the developers to provide them with solutions based on the appropriate data. Later on, developers inform account managers and account managers reach back to the merchants. As can

be seen from Figure 6.1, the loops in this process can be simplified by introducing the monitoring system:

Figure 6.1: Simplification of internal processes by the use of monitoring mechanism.



Concluding, the main finding of the research can be summarized in the fact that using solely analytics on transaction data in order to indicate merchants what to do with regards to risk assessment is not a good approach. The input to be given to the Risk Calculator should consider the type of merchant, including the sector, size in terms of transaction volume, average transaction value and the level of refusal, and possibly chargeback rate. In this way the tool can make suggestions for optimization based on user classification. The secondary changes to be made to the tool include the improvements listed under Section 6.4.1.

Summary. Taking a look back at the main research question, we can state that: Merchants' reluctance in adopting profit-maximizing advices suggested by anti-fraud tools can be explained both by behavioral as well as technical aspects. Firstly, the concepts in Behavioral Economics suggest that people tend to not change behavior when the incentives to do so are not strong, while their decisions are not always optimal due to time and knowledge constraints. Moreover, they are not willing to follow advices if they are already satisfied with their status quo and if the effort they have to put on understanding the advice outweighs the economic benefit they gain. Apart from that, they tend to follow advices which are in line with their risk attitude. Secondly, technical aspects such as quality of output, ease of use, response time and accuracy in predictions are also important factors influencing the engagement with such tools. The research indicated that in order to increase the acceptance the anti-fraud tool examined in our case study, it is not sufficient to rely on analysis of transaction data, but rather provide different input to the tool. Particularly, the tool should take into account the type of merchants that use it, as well as the risk environment, and make different suggestions for optimization based on merchant classification. Apart from that, other improvements that can have a positive impact on its usage include (1) the embodiment of more information in the tool page, (2) the implementation of another algorithm of the alteration

of the algorithm's constraints, (3) the implementation of A/B testing through the tool and (5) the redesign of the tool so that merchants can selectively apply the advices they find more valuable.

6.6 Contributions

6.6.1 Scientific Contributions

With regards to the scientific contributions of the thesis, after presenting the results of the three different research studies, we identify the following contributions. Firstly and most importantly, research in the field of user adoption regarding specifically anti-fraud tools is not ample. By looking at the existing literature, most of the research regarding human behavior related to security choices is centered around social media, mobile applications and the use of Internet in general. We were able to find only two papers discussing factors in user adoption of data-mining tools (Huang et al., 2012, 2013) and one paper focusing on victim's response to financial fraud and identity theft simulations (Rosoff, Cui, & John, 2014). Therefore, our research contributes to Security Economics and Usable Privacy by shedding light on psychological aspects that make merchants engage with security advices. Secondly, our research adds evidence to the body of Behavioral Economics by supporting theory with observations.

6.6.2 Practical Contributions

From a practical perspective, the contributions to Adyen can be divided in two categories. On one hand, we have listed several recommendations regarding how the Risk Calculator can be improved and hence accepted more widely by merchants, while we also made recommendations on the company's internal processes. On the other hand, we developed a program that stores the data and calculations of the Risk Calculator to the database.

Regarding this program, it was developed in Java language and is actually a *batch processing job*. According to Kannan (2013), a batch job is typically a long-running, bulk-oriented and data-intensive process which can be scheduled to run at a specific recurring time, or on demand. The Java job we developed included the creation of two SQL tables, where in the first one all the current configurations and transaction data of the merchants were stored, while in the second table the tool's calculations were being stored. For the storing and transport of the data, an XML file was moreover created. The result of this job was a dataset of 337029 records in the table with

the merchants' current configurations and 339572 records in the table with the tool's suggestions. The second table contained more records as for one merchant the tool might have more than one possible proposals suggested. By joining the two tables we obtained a dataset with all merchant accounts in Adyen and the results they can achieve by using the tool. In order to identify the target groups, we filtered this data based on different aspects. Therefore, another practical contribution is that we identified the group of merchants for which the tool adds more value and thus Adyen can pro-actively engage with this group. Moreover, this Java Job can be used as a monitoring mechanism by the company in order to re-evaluate its added value on merchants.

6.7 Limitations and Future Research

In this final section of the thesis we discuss the overall limitations of our work and we make some suggestions for future research. More specifically, we review construct, internal and external validity and based on the discussed findings, we make some recommendations for future work.

6.7.1 Construct and Internal Validity

We examined in detail the validity of our measurement instruments in the Section "Threats to Validity" under each Chapter. Here, we relist the main points in Table 6.6.

Table 6.6: Summary of research limitations.

Chapter	**Limitation Description**
Ch. 3: Analysis of Historical Data	Limitations in defining "users" group
	Limitations in data regarding tool's calculations (manual collection of data)
	Variables for the comparison of means could be extended
Ch.4: User Interaction with Risk Management Tools	Limitations regarding the number of interviewees
	Limitations regarding the correctness of open coding
Ch.5: Analysis of Risk Behavior for Target Group	The limited number of independent variables for explaining a complex phenomenon (merchants behavior), can be problematic
	Bias error in the independent variable chargebacks
	Selected population might not be representative

Apart from construct and internal validity, we should also mention the following. Our data for both the empirical and the qualitative part derived from a primary source. On the first case, we used the records of the company regarding merchants' transactions, whereas in the second case the audio recordings during the interviews. However since the data were provided by a third party, repeatability does not hold true and hence we will end up with the same dataset no matter how many times we run our scripts. Moreover, since the transaction data of merchants were collected at a certain point in the past, repeatability again does not hold true.

6.7.2 External Validity

Regarding the generalizability of the research, the following should be considered. The results obtained in this thesis are indicative of the characteristics and the behavior of merchants who use anti-fraud tools compared to those who do not. However, the results are applicable in Adyen since the case study is around the tool that the company has developed, and probably differs from other solutions offered by industry. Although the findings can be used as a proxy for other similar companies, it should be noted that the generalization of any conclusion must carefully be examined, plausibly by reproducing the study to other companies.

6.7.3 Future Research

The research conducted in this thesis can be extended in several ways. On one hand, further research to be conducted by Adyen, can be focused on implementing the improvements to the tool, listed under Section 6.4.1. In order to test the level of success of each suggestion, A/B testing can be deployed based on merchants' characteristics, such as their sector, size or risk attitude.

On the other hand, the multivariate regression model can be enriched by two ways. First, by creating another metric for the dependent variable "net score change". This is based on the fact that some of the risk checks might act as outliers, so maybe they should be left out of the calculation. Second, the regression can be extended by including more complex independent variables, related not only to merchant's attributes, but also to transactions characteristics such as the option of 3D-Secure,the country and the payment method. Alternatively, other data mining methods can be used such as cluster analysis, factor analysis or decision trees.

Additionally, post-qualitative interviews with the merchants can add more contextual information

and valuable insights to the findings. It would also be interesting to approach merchants representing the good candidates to use the tool (based on the results of the monitoring mechanism) and compare their behaviour and feedback to the merchants that represent non-good candidates to use the tool.

Appendix A

Interview Protocol

Interview Protocol Form

Institutions:

Interviewee (Title and Name):

Interviewer:

Survey Section Used:

A: Risk Awareness

B: Usage of the Risk Calculator

C: Risk Management within the Company

Other Topics Discussed:

Documents Obtained:

Post Interview Comments or Leads:

Adoption of Risk Management Tools Interviews

Introductory Protocol

To facilitate our note-taking, we would like to audio tape our conversations today. Please sign

the release form. For your information, only researchers on the project will be privy to the tapes which will be eventually destroyed after they are transcribed. In addition, this release form states that: (1) all information will be held confidential, (2) your participation is voluntary and you may stop at any time if you feel uncomfortable, and (3) we do not intend to inflict any harm. Thank you for your agreeing to participate. We have planned this interview to last no longer than one hour. During this time, we have several questions that we would like to cover. If time begins to run short, it may be necessary to interrupt you in order to push ahead and complete this line of questioning.

Introduction

You have been selected to speak with us today because you have been identified as someone who has a great deal to share about risk management practices and awareness of risks related to transactions. Our research project as a whole focuses on the improvement of Adyen's Risk Calculator tool, with particular interest in understanding how merchants engage with risk management tools, how they monitor their refused transactions and their chargebacks, and how they make decisions to minimize risks. Our study does not aim to evaluate your techniques or experiences. Rather, we are trying to learn more about how much companies value risk-management, and hopefully learn about practices that can make merchants' job in mitigating risks easier.

A. Risk Awareness

1. How important do you find monitoring refusal rates* and chargeback rates* within your company? Why?

2. Which are the actions you take for fraud and risk mitigation? How many Full Time Equivalent* you spent on it?

3. How satisfied are you with your average refusal and chargeback rates?

4. How risk averse would you consider yourself when changing your risk settings?

5. Which are the main reasons for changing your risk settings? How often you do that?

6. Do you use the same risk settings/person for all your merchant accounts? Why yes/no?

B. Usage of the Risk Calculator

1. How often do you use the Risk Engine Optimizer (daily/weekly/monthly basis)?

2. Which are your main motives for using the tool? What would be reasons for using it more?

3. Do you trust the tool's suggestions?

4. How often are you willing to apply the risk changes it suggests?

5. Did the predictions of the tool (i.e. the suggested savings it predicted in "total amount" when using the slider button) come true after you applied the changes?

6. Do you experience any problems with the tool?

7. Would you like to see any specific improvements/changes in the next version of the tool?

C. Risk Management Within the Company

1. Have you checked Adyen's Risk Calculator ("Risk Engine Optimizer") tool? Which are the main reasons for not using it actively?

2. How efficient and/or easy would you describe your mechanisms for monitoring refusals and chargebacks? What would make your job easier?

3. When would you consider using a risk management tool*?

GLOSSARY

Refusal rate: the amount of the transactions that were refused (not authorized) due to high risk level divided by the total amount of transactions.

Chargeback rate: the amount of the transactions that resulted in chargebacks divided by the total amount of transactions.

Full Time Equivalent: People working full time on this field.

Risk Management Tool: Software for automatically optimizing refusals and/or chargebacks by minimizing risks related to your transactions, such as fraud.

Appendix B

Interview Protocol within Company

Perceptions about the Risk Calculator

1. Which are your first thoughts on the Risk Calculator?

2. What are the most important benefits of the tool according to your opinion?

3. What are the most important shortcomings?

4. What explains, according to your opinion, merchants reluctance to use the tool?

5. What would you improve in the tool if we did not conduct the research?

Bibliography

Adyen B.V. (2015). *Adyen merchant manual*. Adyen B.V.

Adyen. B.V. (2016). *Chargeback & fraud manual eu region*. Adyen B.V.

Anderson, R., Barton, C., Böhme, R., Clayton, R., Van Eeten, M. J., Levi, M., ... Savage, S. (2013). Measuring the cost of cybercrime. In *The economics of information security and privacy* (pp. 265–300). Springer.

Ariely, D. (2008). *Predictably irrational*. HarperCollins New York.

Aronson, E., Wilson, T. D., Akert, R. M., & Fehr, B. (2007). *Social psychology* (fourth). Toronto, ON: Pearson Education.

Asare, S. K. & Wright, A. M. (2004). The effectiveness of alternative risk assessment and program planning tools in a fraud setting. *Contemporary Accounting Research*, *21*(2), 325–352. doi:10.1506/L20L-7FUM-FPCB-7BE2

Bahnsen, A. C., Aouada, D., Stojanovic, A., & Ottersten, B. (2016). Feature engineering strategies for credit card fraud detection. *Expert Systems with Applications*, *51*, 134–142.

Baker, S. E. & Edwards, R. (2013). How many qualitative interviews is enough? *Review Paper:National Centre for Research Methods*.

Bauer, J. M. & Van Eeten, M. J. (2009). Cybersecurity: stakeholder incentives, externalities, and policy options. *Telecommunications Policy*, *33*(10), 706–719.

Becker, G. S. (1976). *The economic approach to human behavior*. University of Chicago press.

Bhatla, T. P., Prabhu, V., & Dua, A. (2003, June). Understanding credit card frauds.

Bonde, D. (2013). Qualitative interviews: when enough is enough. *White Paper: Research by Design*.

Bouch, A. (2011). *3-d secure: a critical review of 3-d secure and its effectiveness in preventing card not present fraud*.

Brown, J. D. (1996). *Testing in language programs.* Upper Saddle River, NJ: Prentice Hall Regents.

Chien, C.-F., Kerh, R., Lin, K.-Y., & Yu, A. P.-I. (2016). Data-driven innovation to capture user-experience product design: an empirical study for notebook visual aesthetics design. *Computers & Industrial Engineering*, *99*, 162–173. doi:http://dx.doi.org/10.1016/j.cie.2016.07.006

Cockburn, A. (2006). *Agile software development: the cooperative game*. Pearson Education.

Cognizant. (2016). Secure payments: how card issuers and merchants can stay ahead of fraudsters.

Coleman, J. S. & Fararo, T. J. (1992). Rational choice theory. *Nueva York: Sage*.

Consumer Action & Chase. (2009). Questions and answers about credit card frauds. Press Release.

Dal Pozzolo, A., Caelen, O., Le Borgne, Y.-A., Waterschoot, S., & Bontempi, G. (2014). Learned lessons in credit card fraud detection from a practitioner perspective. *Expert systems with applications*, *41*(10), 4915–4928.

Dantzig, G. B. & Wolfe, P. (1960). Decomposition principle for linear programs. *Operations research*, *8*(1), 101–111.

De Gennaro, R. P. (2006). Credit card processing: a look inside the black box. *Economic Review*, *91*(1), 27–42.

Eaton-Cardone, M. (2016). Friendly fraud: an undetected risk in the airline industry. Retrieved January 20, 2017, from http://monicaec.com/friendly-fraud-undetected-risk-airline-industry/

EMVCo. (2016). Worldwide emv deployment statistics. Retrieved from https://www.emvco.com/about_emvco.aspx?id=202

Enserink, B., Hermans, L., Kwakkel, J., Thissen, W., Koppenjan, J., & Bots, P. (2010). *Policy analysis of multi-actor systems*. The Hague: Lemma.

European Commission. (2016, April). The payment services directive: what it means for consumers. Press Release.

Ferreira, G. A. L., Gonçalves, G. S., Otero, A. G. L., Lima, G. L. B., Tsoucamoto, P. T., Villaca, P. C. L., … Dias, L. A. V. (2015). Internet of things and the credit card market: how companies can deal with the exponential increase of transactions with connected devices and can also be efficient to prevent frauds. In *Information technology-new generations (itng), 2015 12th international conference on* (pp. 107–111). IEEE.

Frederick, S., Loewenstein, G., & O'donoghue, T. (2002). Time discounting and time preference: a critical review. *Journal of economic literature*, *40*(2), 351–401.

Gandomi, A. & Haider, M. (2015). Beyond the hype: big data concepts, methods, and analytics. *International Journal of Information Management*, *35*(2), 137–144.

Gigerenzer, G. & Goldstein, D. G. (1996). Reasoning the fast and frugal way: models of bounded rationality. *Psychological review*, *103*(4), 650.

Gilmour, N. (2016). Understanding the practices behind money laundering. a rational choice interpretation. *International Journal of Law, Crime and Justice*, *44*, 1–13. doi:http://dx.doi.org/10.1016/j.ijlcj.2015.03.002

Glaser, B. G. & Strauss, A. L. (1967). *The discovery of grounded theory: strategies for qualitative research*. New York, NY: Aldine de Gruyter.

Grant, R. M. (1996). Toward a knowledge-based theory of the firm. *Strategic Management Journal*, *17*.

Green, D. P., Shapiro, I., & Shapiro, I. (1994). *Pathologies of rational choice theory: a critique of applications in political science*. Cambridge Univ Press.

Guest, G., Bunce, A., & Johnson, L. (2006). How many interviews are enough? an experiment with data saturation and variability. *Field methods*, *18*(1), 59–82.

Guha, R., Manjunath, S., & Palepu, K. (2015). Comparative analysis of machine learning techniques for detecting insurance claims fraud. White Paper Wipro.

Han, J., Pei, J., & Kamber, M. (2011). *Data mining: concepts and techniques*. Elsevier.

Herley, C. (2009). So long, and no thanks for the externalities: the rational rejection of security advice by users. In *Proceedings of the 2009 workshop on new security paradigms workshop* (pp. 133–144). ACM.

Hosmer, D. W. & Lemeshow, S. (2000). Introduction to the logistic regression model. *Applied Logistic Regression, Second Edition*, 1–30.

Huang, T. C.-K., Liu, C.-C., & Chang, D.-C. (2012). An empirical investigation of factors influencing the adoption of data mining tools. *International Journal of Information Management*, *32*(3), 257–270.

Huang, T. C.-K., Wu, L., & Chou, C.-C. (2013). Investigating use continuance of data mining tools. *International Journal of Information Management*, *33*(5), 791–801.

Ingenico Payment Services. (2015). 2015 global online fraud panorama. Retrieved from https://payment-services.ingenico.com/~/media/files/2015-global-online-fraud-panorama_en.ashx?la=en

Jin, X., Wah, B. W., Cheng, X., & Wang, Y. (2015). Significance and challenges of big data research. *Big Data Research*, *2*(2), 59–64.

Kahneman, D. (2011). *Thinking, fast and slow*. Macmillan.

Kahneman, D. & Tversky, A. (1979). Prospect theory: an analysis of decision under risk. *Econometrica: Journal of the econometric society*, 263–291.

Kaisler, S., Armour, F., Espinosa, J. A., & Money, W. (2013). Big data: issues and challenges moving forward. In *System sciences (hicss), 2013 46th hawaii international conference on* (pp. 995–1004). IEEE.

Kannan, M. (2013). An overview of batch processing in java ee 7.0. Retrieved from http : //www.oracle.com/technetwork/articles/java/batch-1965499.html

Levchenko, K., Pitsillidis, A., Chachra, N., Enright, B., Félegyházi, M., Grier, C., ... Liu, H., et al. (2011). Click trajectories: end-to-end analysis of the spam value chain. In *2011 ieee symposium on security and privacy* (pp. 431–446). IEEE.

LexisNexis. (2015, September). True cost of fraud. Retrieved from http://chargebacktech.eu/study-shows-cnp-merchants-are-losing-more-revenue-to-fraud/

Madrian, B. C. & Shea, D. F. (2001). The power of suggestion: inertia in 401 (k) participation and savings behavior. *The Quarterly Journal of Economics*, *116*(4), 1149–1187.

Merriam, S. B. (1998). *Qualitative research and case study applications in education. revised and expanded from "case study research in education"*. ERIC.

Miles, J. & Gilbert, P. (2005). *A handbook of research methods for clinical and health psychology*. Oxford University Press.

Moore, T., Friedman, A., & Procaccia, A. D. (2010). Would a 'cyber warrior' protect us: exploring trade-offs between attack and defense of information systems. In *Proceedings of the 2010 workshop on new security paradigms* (pp. 85–94). ACM.

Nicholls, C. (2013, June). Are verified by visa and mastercard securecode conversion killers? Retrieved from http://www.practicalecommerce.com/articles/4059-Are-Verified-by-Visa-and-MasterCard-SecureCode-Conversion-Killers-

Oxford University Press. (2016). Oxford dictionaries language matters. In O. U. Press (Ed.). Retrieved from http://www.oxforddictionaries.com/definition/english/chargeback

Plous, S. (1993). *The psychology of judgment and decision making.* Mcgraw-Hill Book Company.

Reis, H. T. & Judd, C. M. (2000). *Handbook of research methods in social and personality psychology*. Cambridge University Press.

Ritchie, J., Lewis, J., Nicholls, C. M., Ormston, R., et al. (2013). *Qualitative research practice: a guide for social science students and researchers*. Sage.

Romney, A. K., Weller, S. C., & Batchelder, W. H. (1986). Culture as consensus: a theory of culture and informant accuracy. *American anthropologist*, *88*(2), 313–338.

Rosoff, H., Cui, J., & John, R. S. (2014). Behavioral experiments exploring victims' response to cyber-based financial fraud and identity theft scenario simulations. In *Soups* (pp. 175–186).

Samuelson, W. & Zeckhauser, R. (1988). Status quo bias in decision making. *Journal of risk and uncertainty*, *1*(1), 7–59.

Scott, J. (2000). Rational choice theory. *Understanding contemporary society: Theories of the present*, *129*.

Sift Science. (2015). United states of fraud. Retrieved from http://start.siftscience.com/hubfs/eBooks_and_White_Papers/United-States-Of-Fraud-Report.pdf?t=1459174607205&_ga=1.192285674.1326806301.1485167880

Simon, H. A. (1972). Theories of bounded rationality. *Decision and organization*, *1*(1), 161–176.

Simon, H. A. (1982). *Models of bounded rationality: empirically grounded economic reason*. MIT press.

Statista. (2016). B2c e-commerce sales worldwide from 2012 to 2018 (in billion u.s. dollars).

Taghavifard, M., Damghani, K. K., & Moghaddam, R. T. (2009). Decision making under uncertain and risky situations. Society of Actuaries.

Tan, F. T. C., Guo, Z., Cahalane, M., & Cheng, D. (2016). Developing business analytic capabilities for combating e-commerce identity fraud: a study of trustev's digital verification solution. *Information & Management*, *53*(7), 878–891. Special Issue on Papers Presented at Pacis 2015. doi:http://dx.doi.org/10.1016/j.im.2016.07.002

Thaler, R. & Sunstein, C. (2008). Nudge: improving decisions about health, wealth, and happiness. *Constitutional Political Economy*, *19*(4), 356–360.

Thompson, M. (2014, April). Credit card fraud jumps, so what's being done? Press Release.

Uddin, M. F., Gupta, N. et al. (2014). Seven v's of big data understanding big data to extract value. In *American society for engineering education (asee zone 1), 2014 zone 1 conference of the* (pp. 1–5). IEEE.

Wamba, S. F., Akter, S., Edwards, A., Chopin, G., & Gnanzou, D. (2015). How 'big data'can make big impact: findings from a systematic review and a longitudinal case study. *International Journal of Production Economics*, *165*, 234–246.

West, J. & Bhattacharya, M. (2016). Intelligent financial fraud detection: a comprehensive review. *Computers & Security*, *57*, 47–66.

Whitrow, C., Hand, D. J., Juszczak, P., Weston, D., & Adams, N. M. (2009). Transaction aggregation as a strategy for credit card fraud detection. *Data Mining and Knowledge Discovery*, *18*(1), 30–55.

Wolters, B. (2012). *Managing e-commerce credit card risk: an integral approach*.

Zey, M. (1997). *Rational choice theory and organizational theory: a critique*. Sage Publications.

Zey, M. (2015). Rational choice and organization theory. In J. D. Wright (Ed.), *International encyclopedia of the social & behavioral sciences (second edition)* (Second Edition, pp. 892–895). Oxford: Elsevier. doi:http://dx.doi.org/10.1016/B978-0-08-097086-8.73109-6