Master's thesis

NTNU
Norwegian University of Science and Technology
Faculty of Economics and Management
NTNU Business School

Rebecca Margareta Bang & Stine Maurstad

Money Laundering

Assessing the relationship between money laundering and probability of default, customer size and change in customer characteristics

Master's thesis in Economics and Business Administration
Supervisor: Are Oust
Co-supervisor: Endre Jo Reite
May 2022

NTNU
Kunnskap for en bedre verden

Rebecca Margareta Bang & Stine Maurstad

Money Laundering

Assessing the relationship between money laundering and probability of default, customer size and change in customer characteristics

Master's thesis in Economics and Business Administration
Supervisor: Are Oust
Co-supervisor: Endre Jo Reite
May 2022

Norwegian University of Science and Technology
Faculty of Economics and Management
NTNU Business School

NTNU
Kunnskap for en bedre verden

Acknowledgements

This thesis was written as part of our Master of Science in Economics and Business Administration at NTNU Business School. The thesis counts for 30 credits of the specialization profile financing and investment.

First of all, we want to express our gratitude to our supervisors, Associate Professor Are Oust and PhD student Endre Jo Reite, who provided invaluable guidance during the whole journey, from finding a suitable topic and taking the time to discuss our thesis, to giving constructive feedback. We would also like to thank the Norwegian bank that has given us access to unique data in a field that has high data sensitivity where actors are reluctant to data sharing. Additionally, we would like to express our gratitude to family and friends, who have supported us through a challenging and exciting semester.

We hope this thesis provides new knowledge about factors that correlate with money laundering, which banks can use to improve existing anti-money laundering systems. We also hope that the thesis can provide input to other financial institutions and authorities as well, and motivate future research.

Please note that neither the institution nor the examiners are responsible - through the approval of this thesis - for the theories and methods used, or results and conclusions drawn in this work.

NTNU Business School

Trondheim, May 2022

Rebecca Bang

Rebecca Margareta Bang

Stine Maurstad

Stine Maurstad

Abstract

This article use a unique data set containing customer information from a Norwegian bank, to look for new factors that correlates with the bank's existing anti-money laundering (AML) system. To the authors' knowledge, these factors have not been researched in context of causality in the AML-field before. The factors are tested against both flagged and reported customers, making it possible to identify possible points of improvement in the bank's existing AML system. The main factors studied are a customer's probability of default (PD), the customer's size, and whether it has had any changes in characteristics during the period. The data set is in cross section form, containing administrative data on 8,538 corporate customers, of which 219 have been flagged and 64 have been reported. A binominal logit model has been used to look for causality.

The results show that there is no correlation between PD score and the probability of being reported. If a customer has had a change in PD score however, this increases the probability of both being flagged and reported. The results also show that the current AML system is size-dependent, meaning that the probability of being flagged depends on a customer's size. This is an important finding as it uncovers a weakness in the bank's current AML system. The greater the change in customer size, the greater the probability of being flagged or reported. Customer size, change in customer size, and change in PD score is therefore thought to increase efficiency if included in the bank's current AML system.

It is important to point out the originality and uniqueness of this article. The data set used is composed of customer information that is currently not being used in the bank's current AML system. To the authors' knowledge, this kind of data has never been used to study causality in the AML-field before. Due to the problems with data extraction in this field, similar previous research that can be directly connected to this research is practically non existent. This research is therefore considered a rewarding and exciting new contribution to the field of AML.

Sammendrag

Denne artikkelen bruker et unikt datasett som inneholder kundeinformasjon fra en norsk bank, for å se etter nye faktorer som korrelerer med bankens eksisterende anti-hvitvaskingssystem (AML). Dette har ikke blitt forsket på tidligere så langt forfatterne kjenner til. Denne korrelasjonen er testet mot både flaggede og rapporterte kunder, noe som gjør det mulig å identifisere mulige forbedringspunkter. Hovedfaktorene som er studert er en kundes sannsynlighet for mislighold og kundens størrelse, og om den har hatt endringer i kundeforholdet i løpet av perioden. Datasettet som blir brukt er i tverrsnittsform. Dataene inneholder informasjon fra 8,538 kunder hvor 219 av disse er flagget og 64 er blitt rapportert. Logistisk regresjon blir brukt for å foreta analysen.

Resultatene viser at det ikke er noen sammenheng mellom sannsynlighet for mislighold og sannsynligheten for å bli rapportert. Hvis en kunde har hatt en endring i sannsynlighet for mislighold derimot, har dette en signifikant positiv korrelasjon med sannsynligheten for både å bli flagget og rapportert. Dette antyder at en inkludering av endring i sannsynlighet for mislighold kan forbedre bankens AML-system. Videre viser resultatene at dagens AML-system er størrelsesavhengig, det vil si at sannsynligheten for å bli flagget avhenger av en kundes størrelse. Dette er et viktig funn da det avdekker en svakhet i bankens nåværende AML-system. En stor endring i kundestørrelse, gir større sannsynlighet for å bli flagget eller rapportert. Endringen i kundestørrelse er dermed en faktor som kan forbedre bankens nåværende AML-system.

Det er viktig å påpeke originaliteten og unikheten til denne artikkelen. Datasettet som benyttes er sammensatt av kundeinformasjon som foreløpig ikke brukes i bankens nåværende AML-system. Så vidt forfatterne vet er denne typen data ikke blitt studert i sammenheng med anti-hvitvasking før. På grunn av problemene med datautvinning på dette feltet, er lignende tidligere forskning som kan kobles direkte til denne forskningen praktisk talt ikke-eksisterende. Denne forskningen anses derfor som et givende og spennende nytt bidrag til feltet AML.

Contents

List of Tables

List of Figures

1 Introduction

Money laundering is the activity of hiding the origin of the funds by sending them through one or many transactions to make the funds seem legitimate (Iversen, 2021). It has been affecting the economy on both global and national levels for many years, and is increasing in scope due to globalisation and new technology. Anti-money laundering (AML) work is therefore becoming more and more important to combat money laundering itself and crime. Since money laundering is usually performed by the use of banks or other financial institutions (Økokrim, 2022), banks have a prominent role in the combat against money laundering through detecting and reporting suspicious activity. The amount of resources used on AML in banks increased in 2018 when Norway's Anti-Money Laundering Act was updated (Roaldsøy, 2018), increasing the need for efficiency improvement of current AML systems.

The majority of today's AML systems are rule-based (Chen et al., 2018), and both time and resource consuming. Maintaining the rules, keeping them relevant and deciding how to weigh them is difficult. The embedded rules are often too simplistic, causing missed cases and false alerts (Jullum et al., 2020). Such rule-based systems are therefore quite prone to weaknesses. Since only a small percentage of the flagged transactions are reported, most of the time and resources spent on such work are wasted. Improvements in current AML systems would increase the efficiency of money laundering detection in banks, which would also be benefiting on national and global levels.

Therefore, the aim of this research is to improve money-laundering detection by identifying new factors that correlates with a bank's existing AML system, which can further be used to improve detection accuracy in banks. The main factors investigated are a customers probability of default (PD) score, customer size and changes in customer characteristics. To the authors' knowledge, these factors are not researched in context of AML before. Suppose the link between these new factors and money laundering is proven to exist. In that case, the findings will shed light on which factors and information banks have to focus on to perform efficient AML-work.

Due to the high sensitivity of banks' data on anti-money laundering efforts, similar previous research is limited. However, due to the immense societal gain from increasing the efficiency of anti-money laundering measures, we have gained access to unique administrative data from a Norwegian bank and simultaneously worked on structuring the data and developing methodology, hypothesis, and theories in a field without access to proven methods and theories. To the best of the authors' knowledge, no similar research on actual transaction monitoring data in business banking exist. The data set is in cross section form, existing of 8,538 customers, of which 219 were flagged. Only 64 of the flagged cases were reported to Økokrim[1]. The flagged cases have been detected both manually and through the bank's rule-based system. Since banks are not informed whether the reported case is actually money laundering or not, this is not possible to study through the data set.

A binomial logit model will be used to examine the relationship between the new variables and the bank's AML system. Three different dependent variables will be used; One representing all flagged cases, one representing only cases flagged by the bank's automatic system and one representing only reported cases. This allows for identification of factors that correlates with actual reports, and to reveal possible weaknesses in the bank's automatic AML system. To assess the effect of change characteristics in AML systems, the factors were analysed through two main models; one with and one without change characteristics. Two additional models were also discussed for robustness.

The analysis presents three main findings. First of all, a model including change characteristics would perform better than a model without. Various variables for change support this. Secondly, there is no correlation between a customer's probability of default (PD) and the probability of being reported. However, if a customer has a change in PD score, this will increase the probability of being flagged and reported. The analysis also confirmed the assumption that the bank's AML system differentiate on customer size. This means that the probability of being flagged or reported is size dependent.

The article is divided into 8 chapters, including introduction. Chapter 2 introduce background on money laundering and regulations. Chapter 3 explains AML in Norway and the role of banks. Chapter 4 provides previous research on the topic. Chapter 5 presents

[1]National Authority for Investigation and Prosecution of Economic and Environmental Crime

a statistical description of the data. Chapter 6 presents the methodology, including a theoretical description of the models. In chapter 7, empirical results are presented and discussed which are concluded in chapter 8, summarizing the most important findings with a small review of proposals for further research.

2 Money laundering

2.1 Money laundering explained

Money laundering is necessary in order for criminals to profit from illegal activities (Iversen, 2021). Large sums of money are laundered every year, posing a threat to the global economy. The International Monetary Fund estimates that 2-5% of the global GDP[2] is laundered each year (Camdessus, 1998). To put it in perspective, the amount is equivalent to Spain's contribution to the global GDP in 2020 (The World Bank, 2021). Several attempts have been done to define money laundering through the use of models, mostly to obtain a standardized explanation of what it really is.

The most used model is the "Three-Stage Model" which defines money laundering through three stages; placement, layering and integration (Rui, 2012). The first phase, placement, is when the illegal funds are introduced into the financial system. The use of a bank account normally does this. In the next phase, the origin of the funds is concealed by moving and spreading the funds between more accounts or by marking them as payment for goods or services through fictitious invoicing. In the last phase the money is reintroduced to the financial system with a legitimate origin to be further used as legitimate funds. The model is however criticized for being unrealistic.

To define money laundering with a general explanation is in other words quite complicated. This is why organisations such as Financial Action Task Force (FATF) rather use so-called "typology studies" to describe money-laundering (Rui, 2012). The studies explain through practical examples instead of a rigid model. There is also a distinction to be made between low-level and high-level money laundering (Gibney, 2019). Low-level is considered laundering small amounts from low-criminal activities such as drug dealing, where the transaction is usually just small deposits of cash to one account spread out during the year. High-level is more complex, since the amount that has to be laundered is considerably larger. The process usually involves placing the illegitimate funds into the financial system and then distributing them to numerous accounts to create confusion and to avoid being detected. Finally the funds find its way back to the owner, to then appear as legitimate.

[2]Gross domestic product

The consequences of money laundering are plenty on both micro- and macro-economical level; it affects currencies and interest rates, a country's or institution's reputation, the integrity of financial markets, and maybe most important for governments - its revenue (McDowell and Novis, 2001). This is why organisations spend enormous amounts of both time and resources on detecting such activities. Money laundering is also both self-sufficient and self-strengthening. The more money that is laundered, the more can be used to hide and obscure funds and to fund the main criminal activity. In other words, it generates more crime. It is highly correlated with criminal acts such as corruption, threats/blackmailing, violence, and other types of crime such as tax and climate crime (Rui, 2012). It is also quite normal that money launderers, with time, try to infiltrate legitimate markets. If this is done properly, it will become harder to detect the laundering activities. As the illegal activities gets more mixed with legal activities, it will normally also get harder to pinpoint who's actually responsible for the criminal acts.

2.2 The importance of Anti-Money Laundering (AML)

The activity of preventing money laundering is called Anti-Money Laundering (AML). It has its origin back to 1988 when the FN convention against drugs was established due to increasing issues with international drug crime (United Nations: Convention Against Illicit Traffic in Narcotic Drugs and Psychotropic Substances, 1989). This shed light on the link between criminal acts and money laundering. The link was confirmed in 2009 by FATF's strategic surveillance study, which uncovered that the main source of money laundering is drug dealing and so called "white collar crimes" (Rui, 2012). This highlights one of the many important reasons to spend resources on AML-work; it doesn't just prevent money laundering as an activity, but it also makes it more difficult to profit on other criminal activities such as drug dealing and white-collar crimes. A more prominent problem with money laundering that needs to be emphasized is that such funds are often used to finance terror or other forms of criminal activities. Even legitimate funds are laundered, usually to avoid taxes, which results in significant costs for the society.

As mentioned, organized crime is directly linked to money laundering, since it isn't possible to profit from criminal activity without laundering the gains. AML complicates the money laundering process and makes it harder to profit on illegal activities. All combat against money laundering is therefore a combat against organized crime. Another major motivation behind AML is tax evasion. This was the main argument for the foundation of FATF back in 1989. Tax evasion in itself isn't money laundering, but to profit from the crime the funds have to be laundered. The Norwegian Tax Administration assume that the difference between what is actually taxed and what could be taxed in 2008 was approximately NOK 115 billion (Kristiansen, 2010). AML could therefore contribute to increasing a country's tax revenues considerably.

Money laundering is damaging to the global economy and trade as well. Fair competition implies that all actors compete under the same legal framework. Actors that gain profit through illegal activities, like tax evasion, are therefore damaging the normal market competition. This can potentially damage a country's economy and economic growth. If a country has trouble combating money laundering, or gains a reputation for having lax regulations, it will become a target for criminals that wish to profit from criminal acts. This can potentially evolve to high levels of corruption where criminals gain enough power to control the country through blackmailing and threats. E.g. the leader of UNDOC[3] stated in "The Guardian" in 2009, that criminals saved several banks from collapsing by providing capital during the financial crisis in 2008 (Rui, 2012). If this is true, it isn't unthinkable that at a certain point in time, these criminals might demand the favor returned.

2.3 Regulations

AML has been on the agenda since the end of the 1980s. The regulations regarding money laundering have therefore evolved extensively over the last 40 years both internationally and nationally. It is therefore relevant for this article to have knowledge about some of the most prominent actors in this field, such as FATF and the UN, and how it has evolved on a national level for Norway.

[3] United Nations Office on Drugs and Crime

2.3.1 The United Nations Office on Drugs and Crime (UNODC)

The combat against money laundering began in full with the UN convention on drugs in 1988, as a response to the continually growing drug trafficking problem (Rui, 2012). It was the convention that put money laundering on countries' radar. More specifically, art 3. letter b and c enforce the signature states to criminalize money laundering in context with drug trafficking. This was altered in later conventions, so that money laundering would be criminalized independent of primary offence. In 1997 UN established a sub-unit called United Nations on Drugs and Crime (UNODC), to assist member states in matters regarding drugs, organized crime, corruption and terrorism (UNODC, 2022a).

The UN has had several effects on Norwegian law. First of all in the form of enforcement. Signature states are enforced to initiate actions against money laundering, in terms of penalties, sanctions or regulations in accordance with the convention's demands. UNDOC, in collaboration with IMF and Commonwealth Secretariat(UNODC, 2022b), has also created model laws to assist member states in forming their own AML-framework. UN's resolutions also affects the member states. An example is the Resolution 1373 carried in 2001, which presumably increased the use of soft laws, from organisations like FATF, by the member states (Rui, 2012).

2.3.2 The Financial Action Task Force (FATF)

The Financial Action Task Force (FATF) is regarded as the world's most important organisation in the fight against global money laundering and terror financing (Rui, 2012). It was established in 1989 with the purpose of preventing the use of banks and other financial institutions to perform money laundering activities. As an inter-governmental body, they work to develop and promote policies to combat money laundering. Their recommendations are used worldwide as standards for global AML-regulations and counter-terrorist financing. Today they comprise 37 member countries and 2 regional organisations, which they are continually evaluating (FATF, 2022). As the work done by Johnson and Desmond Lim (2003) states, FATF's work is actually making a significant difference in the combat against money laundering.

The organisation has developed three main mechanisms to fight money laundering and terror financing; the development of guidelines, an evaluation system and a sanction system (Rui, 2012). The guidelines are undoubtedly the most important which in total consist of 125 pages with detailed demands, requests and advice, which has even affected the AML-directive of EU. This is a good example of the organisation's influence worldwide. The second mechanism, the evaluation system, is an evaluation of the member countries, and to what extent they follow the guidelines. The third mechanism, the sanction system, was originally a blacklist where both member and non-member countries could be listed, which proved to be highly effective.

Money laundering expert John Rui (2012) expresses that among all organisations combating money laundering, FATF has probably had the most influence worldwide. There are however some drawbacks. The FATF-regime is perceived to be undemocratic compared to other international organisations that create "hard law", probably due to the organisation being expert-driven. They are also solemnly focusing on fighting money laundering and terror financing, which introduces the risk of neglecting important factors such as GDPR[4]. Their legitimacy is also being questioned. They are being perceived as secretive with an unclear mandate and there have been attempts to use the organisation as political leverage. The organisation is however not able to create "hard law", which allows for alterations of their guidelines on a national level and lowers the risk of influence from the negative factors mentioned above.

2.3.3 The European Union (EU)

The European Union (EU) has had a central role in the AML-work in Europe, and has certainly had the most influence on Norwegian Law. Their instrument is the AML-Directive (AMLD) which they are continuously updating (Rui, 2012). They are currently on their sixth AML-Directive. In the previous AMLD (5AMLD), the directive introduced the demand of knowing who the "beneficial owners" of the bank accounts were to keep track of who actually owns the accounts (Hassoumi, 2021). This was the driving force behind the update of Norways AML Act in 2018.

[4]General Data Protection Regulation

The directive has initiated a lot of debate. Mostly due to the conflict between the obligation to report and confidentiality, which has even been taken to court (Rui, 2012). It is also criticised by banks and financial institutions since the obligation to report and register is extremely resource and time-consuming. The directive is a regulating force that introduces "hard law", and is therefore affecting nations in a more directly manner compared to FATF.

3 Anti-Money Laundering in Norway

The value of Norwegian hard cash has for the last 10 years been approximately NOK 50 billion. The central bank of Norway can only account for 40% (Nilsen, 2021). Circumstances such as this has initiated new regulations and increased focus on AML significantly the last years, especially in banks. Since 2015, Norway has used approximately 15 billion NOK on AML (Schultz, 2019).

Improvement of current AML systems in Norway is considered useful due to several reasons. First of all, Norway is a highly attractive country to place laundered funds since it has a financial system with high stability and credibility. It also has its own currency, which makes it possible to conceal tracks through currency trading (Trumpy and Lund, 2021). Transactions from Norway are also seldom questioned, due to Norway's good reputation internationally. Norway's bank sector is also considered small, and according to Gibney (2019) the risk of money laundering is higher for countries with a small bank sector. This is because such countries often use fewer resources on good AML-systems.

To further illustrate why focus on AML is important for banks in Norway, some recent cases of money laundering is worth presenting. In 2018, Danske Bank was accused of having laundered over $230 billion USD through its Estonian subsidiary (Gricius, 2018). This incident illustrate several important points. First of all, the importance of manual controlling. One of the reasons why the scheme was allowed to continue for over nine years was because the Danes blindly trusted the AML documents from the Estonian branch, which was written in Estonian and Russian. The incident also shows that money-laundering isn't necessarily linked to corruption and lax laws. Scandinavia is known for having good AML routines and low levels of corruption, and still managed to be involved in probably the largest (known) money-laundering scheme ever. Another example is the accusation of DNB, Norway's largest bank, for being involved in the laundering of almost $70 million through a tax haven in the Marshall Islands (Høgseth et al., 2019). This incident strengthens the argument for small countries with low corruption like Norway to continue improving their AML routines.

3.1 Regulations in Norway

The regulations on money laundering in Norway has been developed extensively in recent years. Money laundering was first criminalized in Norway's penal code in §337 through §341. As a member of the EØS, Norway is liable to introduce enactments from EU's directive. Norway therefore developed an independent law on money laundering called the Anti-Money Laundering Act in 2003. This was later updated in 2009, and again in 2018 to meet international standards and changes in EU's directive (Roaldsøy, 2018). The overall purpose of the act is to give guidance to the Norwegian banks and other actors in how to prevent and detect money laundering and terror financing (the Anti-Money Laundering Act, 2018). This aids in protecting Norway's financial and economic system by uncovering which banks and financial institutions are being used for such activities. The new act has included several recommendations from FATF and the EU's AML-Directive.

3.2 The Role of Banks

Both Norwegian banks and branches from foreign banks positioned in Norway are regulated through Norway's Anti-Money Laundering Act[5]. This means that banks have a particular responsibility when detecting and reporting suspicious activity. Banks neither work as detectives nor regulators, which means they have no responsibility to judge whether the case is actually money laundering or not (DNB, 2022). However, the Anti-Money Laundering Act enforces them to monitor their transactions and report to Økokrim[6] if something is considered suspicious. They are in other words, an obliged entity and has to act accordingly.

The Act has put pressure on the "Know Your Customer"-principle, and requires that the obliged entities performs risk-based customer due diligence (CDD) measures and ongoing monitoring of their customers[7]. This means that CDD and associated monitoring should cohere with the identified and assessed risk of money laundering (Finans Norge, 2017). The main motivation behind this is to identify deviating behaviour. This is considerably

[5] Anti-Money Laundering Act §4

[6] The Norwegian National Authority for Investigation and Prosecution of Economic and Environmental Crime

[7] Anti-Money Laundering Act §9

easier if the banks have knowledge about customers' expected capital flows or how they intend to use the bank. The law also requires that the obliged entities valuates their own activity to uncover possible internal involvement[8].

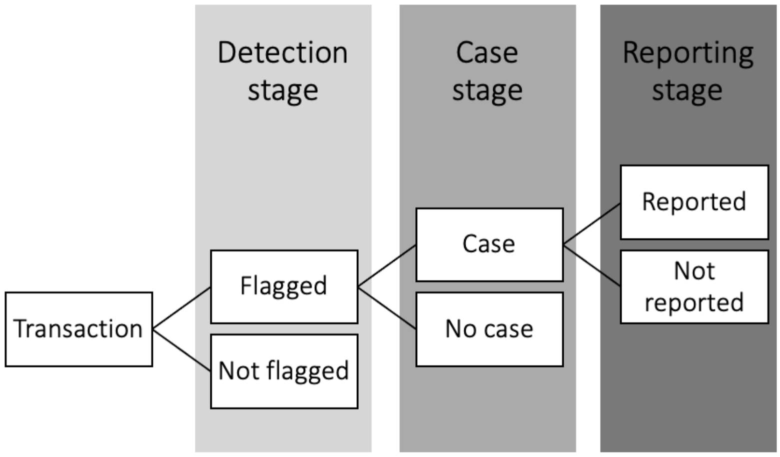
Obliged entities are as mentioned required to regularly perform CDD measures. This includes monitoring transactions to ensure they cohere with already registered information, such as main business activities and objectives. If circumstances that might be linked to money laundering is detected, the banks are obliged to investigate further[9] to either report or disclose the case. If the case is found suspicious, the banks are obliged to report to Økokrim[10], and provide whatever assistance and documentation they require. Due to confidentiality, the customer shall not under any circumstances be informed about the reporting[11]. Absent or deficient reporting to Økokrim or violation of the Act is punishable with penalties or fines, and up to 1 year imprisonment[12]. Obliged corporations can be sanctioned even if no one is held responsible for the act.

3.2.1 Detection process

The detection process in a typical Norwegian bank has three stages; the detection stage, the case stage and the reporting stage seen in figure 3.1 (Jullum et al., 2020). All transactions go through the initial detection stage, where they are evaluated based on a set of rules. This is the so-called automatic detection system. The flagged cases are further evaluated manually, where the cases that seem legitimate are left out of further investigation. The flagged cases that seem suspicious are then grouped into cases built around the main suspect party or possible related parties. Finally, these cases go through a thorough investigation where they either end up being reported to Økokrim or dismissed. The customer being investigated are as mentioned not informed by the bank nor Økokrim during the process. Banks are neither informed about the further investigation after reporting. This means that the reported activity is seldom confirmed, unless it becomes public knowledge that the customer is being accused of money laundering.

[8] Anti-Money Laundering Act §7
[9] Anti-Money Laundering Act §25
[10] Anti-Money Laundering Act §26
[11] Anti-Money Laundering Act §28
[12] Anti-Money Laundering Act §51

Figure 3.1: The three stages of the detection process in a typical Norwegian bank. Showing the handling of a transaction from start to finish, going through stages of automatic evaluation (detection stage), manual evaluation (case stage), then reported to Økokrim or dismissed (reporting stage).



In recent years, banks have mostly used rule-based systems to predict money laundering (Chen et al., 2018). One of the reasons why rule-based systems have been so popular is because they are easy to use and the simplest form of artificial intelligence (Grosan and Abraham, 2011). A rule-based system uses a set of rules that are coded in advance that tells what to do or conclude in various situations. The systems are typically amount-dependent with thresholds, e.g. transactions above 250 thousand or 20 transaction with the same amount. They are also normally sensitive to transactions to foreign countries or especially countries of risk. In recent years, due to new regulations from the EU, banks have also been obliged to initiate enhanced CDD if a customer is classified as risky according to the Anti-Money Laundering Act, as mentioned above. Currently used systems are therefore also sensitive to customers that have enhanced CDD. This is based on review of structure of rules and settings in the bank, and is assumed to be applicable to other Norwegian banks as well.

Since such threshold rules set the bank's automatic system, they are prone to have weaknesses. For example, the rules that are amount specific can lead to a small customer with low transactions falling under the radar. To avoid flagging too many transactions, the systems are typically programmed with upper limits on repeated flagging. Therefore, the system is most effective for customers with suspicious transactions within the threshold rules. It is plausible to assume that the system also neglects customers that fall outside the thresholds, such as small and very large customers.

This possible neglection of small customers is considered a weakness in context of the nature of money laundering. Since no one can deposit large sums of money without raising attention, the funds are usually divided in several small amounts and between several banks. This means that one big customer would be considered small by the various banks. It is therefore possible to assume that the probability of money laundering could be higher for customers that are considered small by the banks (Ringen and Gjesdahl, 2019).

4 Previous Research

Since the focus on money laundering has been relatively low until recent years, the research on the field is limited. It is mostly on how to measure the amount of money that is laundered each year, eg. studies of Walker (1999) or Claessens and Naude (1993), or qualitative studies on employee's or organisation's attitude towards money laundering, eg. studies of Nordvik and Reitan (2020) or Hafthun and Ulfsnes (2019). The use of machine learning in AML e.g. Jullum et al. (2020) and Chen et al. (2018) has also gained much interest in recent years, due to the weaknesses of rule-based systems and the immense societal gain of improving the systems efficiency.

Earlier empirical studies are dominated by overview articles e.g. Chen et al. (2018), and machine learning articles on limited to old data sets e.g. Deng et al. (2009) with only private customers, and Lopez-Rojas and Axelsson (2012) with synthetic data. Previous research that use raw administrative data extracted directly from a bank to look for causality, is however scarce. An article that has performed similar data extraction is Jullum et al. (2020), who also uses raw administrative data from a Norwegian bank. Their focus however is on prediction accuracy through machine learning and not causality. In other words, there is a research-gap in the AML-field on sufficient statistical analyses on suspicious transactions (Cotoc et al., 2021), that this article aims to supplement.

A major breakthrough in money laundering detection happened when the data perspective changed from transaction-oriented to subject-oriented (Kingdon, 2004). Instead of just looking at single transactions, the perspective shifted to single customers or groups and their behaviour pattern. This shift in focus introduced a good surrogate for the term "suspiciousness", which has been troublesome to define and a major obstacle in money laundering detection. The surrogate was "unsusualness" or unusual behavior, which is much easier to identify through artificial intelligence. This change in perspective improved the false positive rates significantly, highlighting dimensionality's importance.

The study of Gao and Ye (2007) points out that being too variable-oriented in money laundering detection can be damaging as well. It highlights that important patterns in customer behaviour might be missed when history is neglected. Their results might be improved if a time sequence analysis was made to obtain a more history-oriented perspective. Being too subject-oriented can be damaging as well (Jullum et al., 2020). Jullum et al. (2020) argues that since customers aren't warned when being reported they tend to continue their financial operations, which leads to the customer being reported several times. When modelling suspicious accounts or parties, this is a problem since the same account/party might present conflicting labels. This is avoided when detecting on transaction level.

Regardless of detection method or perspective, the previous research mentioned here agrees that research on the AML field is valuable for combating money laundering and crime on a superior level. The importance of AML work is supported by a comprehensive study done by Barone and Masciandaro (2010). The study is a direct contradiction of previous research on the controversial topic regulations and reduction in crime, such as the public choice theory on regulation from Coase (1937) and Stigler (1964). Barone and Masciandaro (2010) found that money laundering directly leads to an increase in crime, and that an increase in AML effectiveness, given its cost, lead to a decrease in the overall money laundering activity. All AML research is therefore considered as valuable contributions.

5 Data

One of the reasons why research on this field is limited is lack of data. The limited access to reliable data lies in the nature of money laundering, which makes the percentage of flagged cases in a bank relatively low and bank specific. Much could have been solved through data sharing between banks, but due to privacy and GDPR[13] this is non existent. This makes it hard to generalise findings because the variables and results might be bank specific, since the customer data is limited to that bank. The complete customer picture is in other words missing due to the problems with data sharing between banks.

In cooperation with a Norwegian bank, we have been given access to anonymized data on their corporate customers from August 2021 through January 2022. This kind of customer information, or data set, has to the authors' knowledge never been used in context of causality and money laundering before, possibly because data extraction is difficult. The data set itself is therefore a major contributor to the originality of this study. This gives reason to expect results that could be hard to justify or generalize, but regardless of results, the findings will anyways be a good contribution to AML research.

The data set has information on which customers that were reported, flagged and non-flagged. Whether a reported case actually turns out to be money laundering is not revealed to the bank, and is therefore not possible to research in this analysis. This section will explain any pre-processing of the received data set and provide a descriptive analysis of the final data.

5.1 Data description

The data set is in cross section form, making it possible to study causality between probability of detection and changes in customer characteristics over time in variable-form, like Gao and Ye (2007) and Kingdon (2004) motivates. From the total of 8,538 customers, 219 were flagged. Only 64 of the flagged cases were reported to Økokrim. The flagged cases have been detected both manually and through the bank's automatic system.

[13]General Data Protection Regulation

The customer base consist of mainly Norwegian customers where only 22 are foreign. The Norwegian customers are spread throughout Norway from 21 different sectors. 66% of the customers are based in Oslo or Viken county and 55% are private limited companies. 6.5% of the whole customer base has beneficial owners that is either a foreign citizen, has a foreign domicile, or has tax liability in a foreign country. The data set also contains information on any changes in the customer relationship, making it possible to investigate the causality between changes in customer characteristics and flagged cases.

The original data set consisted of 66 variables. From these, 4 new were created to assess the relationship between customer relationship and flagged and reported cases. The variable's extreme values have been examined and are actual values.

5.2 Data exploration

This section shows a detailed description of the variables used in the study, where the most important ones are described in detail. The software used for data analyses and processing is Excel and Stata.

5.2.1 Dependent variable

There are three dependent variables used in this study; flagged cases, flagged cases without manual investigation and reported cases. This is grounded in the bank's detection system represented in figure 3.1. The use of three dependent variables gives three different models, which makes it possible to analyse the variables in context of both the bank's automatic system and their manual evaluation, and to compare them. It also gives the opportunity to identify factors that leads to reporting and possible weaknesses in the bank's automatic system. The details of the variables are displayed in table 5.1.

Table 5.1: An overview of all dependent variables used in the study with a brief description of each variable.

Name	**Type**	**Obs.**	**Description**
Flagged[a] all	Binary	219	Flagged cases includes both flagged and reported
Flagged[b] automatic	Binary	169	Flagged cases without manual investigations
Reported	Binary	64	All reported cases

Flagged cases (*Flagged^a all*)

There are a total of 219 flagged cases in this data set, which is approximately 2.6% of all customers. The variable is binary where 1 indicates a flagged customer. A flagged customer in this case could either be reported or not reported, so the model with this as dependent variable is only investigating which factors that lead to a flagging. The variable is also a combination of both the bank's automatic system and manual investigations. Of all flagged cases 50 were from manual investigation.

Flagged cases without manual investigations (*Flagged^b automatic*)

Since the variable *Flagged^a all* includes manual cases, a new flagged variable excluding these was made to separate fully between the bank's automatic system and manual investigation. Therefore, the variable only includes 169 flagged cases, which is 2.0% of the data. Only 14 of the automatically flagged cases was reported. This variable does not separates between reported and not reported.

Reported cases

The flagged cases are manually evaluated and further reported to Økokrim if they're found suspicious, so the data set separates between reported and not reported cases. The reported variable is used as a dependent variable to investigate the difference between what is flagged and what is reported. This makes it possible to identify the factors that contribute to a case being reported. The amount of flagged cases that led to reporting is 64, or 0.8% of the whole data set. 50 of the reported cases are as mentioned discovered through manual investigation, 14 from the automatic system.

5.2.2 Independent variables

An overview of all independent variables is shown in table 5.2. Descriptive statistics for continuous variables are presented in table 5.3, and binary and categorical in table 5.4. *Customer size*, *PD*, and some variables that indicate change will be explained more in detail in this section since they are the main focus of this article.

Table 5.2: An overview of all independent variables used in the study with a brief description of each variable.

Name	Type	Description
Probability of default (PD)	Numeric	Values from 0% to 20%. A value close to 0% indicates a low probability of default, while a value close to 20% indicates a high probability of default.
Change in PD (1)	Binary	1 if there has been a change in PD the last year, 0 otherwise
Change in PD (2)	Binary	Same as Change PD 1 but without new customers
Customer size	Categorical	Customer size variable made categorical. Very Small: $<$ 0.1 mill Small: [0.1 mill, 0.5 mill) Medium: [0.5 mill, 2 mill) Large: [2 mill, 10 mill) Very Large: $\geq$ 10 mill
Change Customer size	Categorical	Change in customer size the last month 0: 0% 1: 100% 2: $<$ -100% 3: [-100%, 0%) 4: (0%, 100%) 5: $>$ 100%
Enhanced CDD	Binary	1 if the bank has implemented enhanced customer due dilligence (CDD), 0 otherwise
Change Enhanced CDD	Binary	1 if the bank has implemented enhanced CDD the last month, 0 otherwise
Inactive customer	Binary	1 if the customer has been inactive for 12 months, 0 otherwise
Number of deposit account	Numeric	Number of deposit accounts
Online banking	Binary	1 if the customer has an online bank account, 0 otherwise
Number of urgent payments	Numeric	Numeric of urgent payments the last year
Number of foreign payments	Numeric	Number of foreign payments last 2 years
Taxable abroad	Binary	1 if the beneficial owner (BO) is taxable abroad, 0 otherwise
Change Taxable abroad	Binary	1 if the BO has changed its tax location to abroad, 0 otherwise
Foreign domicile	Binary	1 if the BO has a foreign domicile, 0 otherwise
Change Foreign domicile	Binary	1 if the BO has changed its domicile location to abroad, 0 otherwise
Foreign citizen	Binary	1 if the BO has citizenship abroad, 0 otherwise
Change Foreign citizen	Binary	1 if the BO has changed its citizenship to abroad, 0 otherwise

Table 5.3: Descriptive statistics for the continuous variables used in the study. *Customer size* is used categorically in the analysis.

	Mean	**S.D**	**Min**	**Max**	**Median**	**Obs.**
Probability of default (PD)	0.026	0.031	0.0	0.2	0.018	8,358
Customer size	$2.8 \cdot 10^6$	$2.09 \cdot 10^7$	0	$5.78 \cdot 10^8$	16,333	8,538
Number of deposit acounts	1.29	1.15	0	17	1	8,538
Number of urgent payments	0.40	6.39	0	422	0	8,538
Number of foreign payments	1.19	19.19	0	1,387	0	8,538

Table 5.4: Descriptive statistics for binary and categorical variables used in the study. Shows the number of observations for each variable.

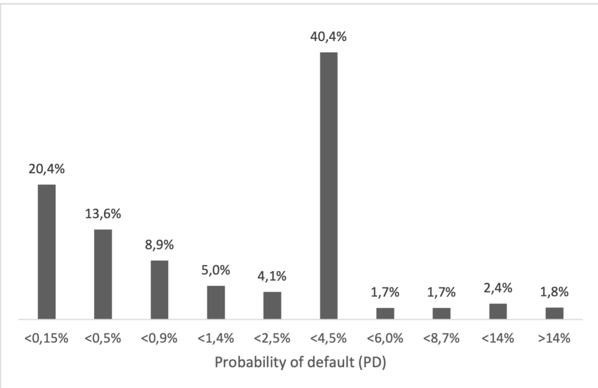
Variable name	**Obs.**	**Variable name**	**Obs.**	**Change Customer size**	**Obs.**
Change in PD (1)	104	Change in PD (2)	41	0%	5,074
Enhanced CDD	413	Change Enhanced CDD	127	100%	55
Inactive customer	3,598	Online banking	3,239	<-100%	433
Taxable abroad	119	Taxable abroad	5	[-100%, 0%)	1,525
Foreign domicile	150	Change Foreign domicile	5	(0%, 100%)	1,098
Foreign citizen	378	Change Foreign citizen	441	>100%	353

Probability of default (PD)

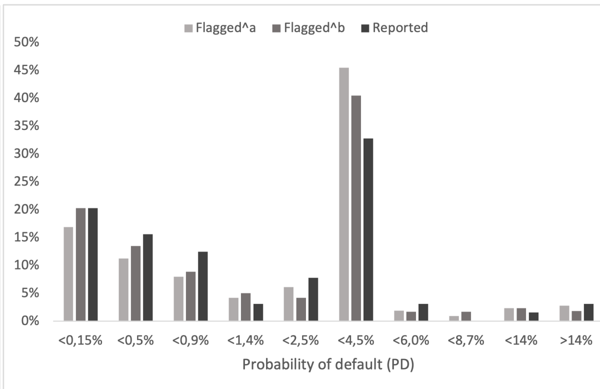
The variable *PD* shows the customers probability of default (PD) score for the last year. The number of observations is 8,358, meaning that 2% of the customer base has missing values. The missing values include 4 flagged cases but no reports. Since this reduces the sample with 180 observations, the models were tested without the variable as well to see if any major changes occurred. The exclusion led to no changes in the results. A customer can have any PD value between 0% and 20%. Table 5.3 shows that 50% of the customers has a score below 1.8%.

Figure 5.1 shows an overview of how the bank's customers are scored, separated into groups according to the bank's routine (table 5.1a), and how the flagged and reported cases are distributed across the different scores (table 5.1b). Figure 5.1a shows that most customers have a score between 2.5% and 4.5%. This group contains the bank's standard score of 4%, which is e.g. given to new customers if no other information is possible to obtain. A total of 3101 customers has 4%. Figure 5.1b shows how the flagged and reported cases are distributed across the different PD scores. The histogram show a trend similar to the distribution in figure 5.1a, which indicates that customers are flagged independent of PD score. This is further investigated in the analysis where PD is continuous and not categorical.

Figure 5.1: Distribution of the customers probability of default (PD) score, ranging between 0% and 20%, also in terms of flagged/reported cases. A value close to 0% indicates a low probability of default, while a value close to 20% indicates a high probability of default. 4% is the bank's standard score.



(a) Distribution PD



(b) Distribution of flagged/reported cases

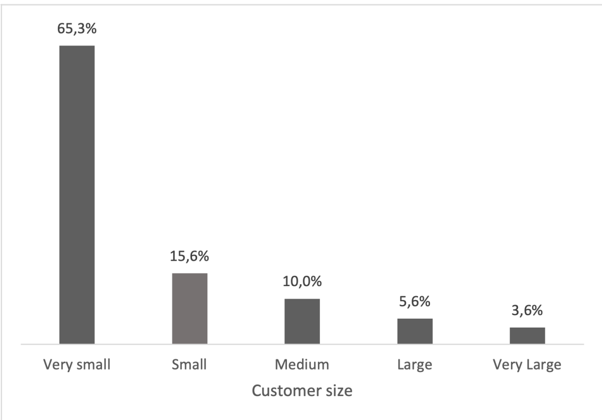
The variable *Change in PD (1)* tells if a customer has had a change in PD score the last year or not. Originally, new customers have been coded as a change during the past year. Since this is not considered as a real change in this analysis, this group was recoded to no change in a new variable *Change in PD (2)*. Therefore, the new variable, *Change in PD (2)*, makes a better representation of an actual change. 63 customers were considered new customers and recoded to no change.

Customer size

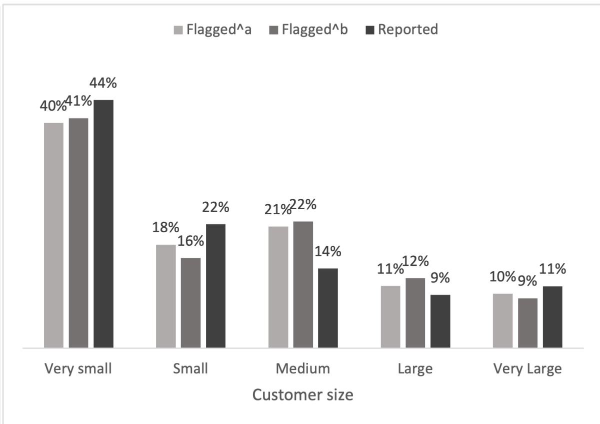
Customer size is defined as the total amount of cash a customer has either deposited or loaned from the bank. Figure 5.2 shows the distribution of the variable and how the flagged cases are distributed according to customer size. The definition of the labels are displayed in table 5.2. Figure 5.2a shows that the largest proportion of the customers are considered very small, which equals a customer size below NOK100,000. As seen from table 5.3, the mean customer size is NOK2.8 mill. At the same time, the median is only NOK16,333. However, a few *Very Large* customers that have a large effect on the data, which increases the average. These extreme values have been taken into account by making customer size categorical in the analysis.

Figure 5.2b shows how flagged and reported cases are distributed among the customer groups. *Very Small* has 44% of the reports, even though it accounts for 65% customers base. The other customer groups has amount of reports higher than their proportion of customers. The difference indicates that the bank's system might underperform on *Very Small* customers and overperform on the other groups.

Figure 5.2: Distribution of customer size in categorical form, and in terms of flagged/reported cases. Size definition is displayed in table 5.2.



(a) Distribution customer size



(b) Distribution of flagged/reported cases

The variable *Change Customer size* displayed in 5.4 shows whether a customer has increased, decreased or remained unchanged in size during the last month. It is defined as account activity the last month divided by size. 0% change is customers that has had no account activity during the last month, which is 5,074 customers or 59% of the whole customer base. 100% change is possible new customers the last month, while the four last categories represent small or large changes. Small changes are between 0% and 100%, while above 100% change is considered large changes in customer size. 31% of the data set had small changes while 9% had large changes. The remaining 1% are possible new customers the last month.

Remaining variables

The remaining variables are largely used to investigate other customer characteristics that can improve the bank's AML system, such as beneficial owner (BO) characteristics and payment patterns not included in the bank's system today. As displayed in table 5.2, there are also several variables indicating change in customer characteristics or how they use the bank.

The variables for change doesn't say what kind of change it is, just that there has been a change. The variables regarding BO characteristics, e.g. *Change foreign domicile*, doesn't say whether the customer has changed its domicile to Norway or abroad, just that there is a change. Since this article focus on change as a whole and doesn't separate between type of change, no change has been made to the variables.

5.2.3 Variable correlation

Appendix A2.1 shows an overview of the correlation between the independent variables. The categorical variables are excluded. The highest positive correlation is between *Foreign citizen* and *Change foreign citizen*, with a value of 0.79. Since the variables are binary, naturally connected, and doesn't impact results they were both included in the models. The value between *Change Enhanced CDD* and *Change in PD (1)* is 0.74. Both variables are binary, so this may indicate that several cases exist with both a change in PD and change in enhanced CDD. *Change in PD (2)* and *Change Enhanced CDD* have a correlation of 0.54.

Appendix A1.1 shows the VIF index for model 2. This model includes all variables listed in table 5.2. Table A1.1 shows no VIF values above 5, which indicates that there are no problems with multicollinearity among the explanatory variables, despite the fact that figure A2.1 display some high correlations. In model 4 (appendix A4.1) *Change Enhanced CDD* was dropped as a robustness test for the analysis, due to the high correlation between *Change Enhanced CDD* and *Change in PD (1).*

6 Methodology

A binomial logit model is used in the analysis. The choice of model comes mainly from the aim of this analysis, which is to examine the effect new factors has on the bank's AML system. More complex machine learning methods would be preferred if the goal was to improve prediction accuracy. However, regression models are a preferable choice to explain the relationship between a response variable and one or more explanatory variables (James et al., 2021). Our response variable is however binary with values 0 and 1. Therefore, the use of a logit model is a better choice than normal linear regression, since it will produce results within this range. A linear regression model will produce non-interpretive results outside this range.

6.1 The Binomial Logit Model

The binomial logit model is a preferred estimation technique when the response variable is binary. Using a variant of the cumulative logistic function avoids the problem with unboundedness that follows linear regression (Studenmund, 2017). The binomial logit model in linear form for each customer i is given by:

$$\ln\left(\frac{P_i}{1-P_i}\right) = \beta_0 + \sum_{k=1}^{n} \beta_k X_{ki} + \varepsilon_i \tag{6.1}$$

Where P_i is the probability that the customer is flagged or reported depending on the model, k is the number of explanatory variables, X_{ki} is the explanatory variable k for customer i, β_k is the regression coefficient for variable k, β_0 is the bias term and ε_i is the error therm.

Logit models do not meet the OLS[14] estimation requirements and instead use maximum likelihood estimation. This method gives values to the unknown parameters β that maximize the probability of obtaining the used data set (Hosmer et al., 2013). A so-called likelihood function (equation 6.1) is then estimated. In other words, the estimated parameters are the ones that best fits the data. The estimation technique does not require

[14]Ordinary Least Squares

constant variances like OLS, and heteroscedasticity is not an issue in this analysis.

How good the logit model fits the data is interpreted through goodness of fit (GOF) like the predictive power (*Pseudo* R^2) and Chi-square (χ^2) (Hammervold, 2020). The predictive power show how well the dependent variable is explained through the explanatory variables. Since our purpose is not to predict, *Pseudo* R^2 is not emphasized in the analysis (Tufte, 2000). The Chi-square value is the difference in the log-likelihood between a null model, where all the regression coefficients are zero, and our specified model. A good adjusted model has low Chi-square value with a p-value under 0.05.

What is considered a good *Pseudo* R^2 value is research field dependent. *Pseudo* R^2 can only be used to compare different models with the same dependent variable in the same sample. It is therefor not possible to compare the three versions of model 1 and 2 based on *Pseudo* R^2, since they have different dependent variables (Tufte, 2000).

6.2 Limitations

The research is naturally limited to Norway since the data used in this research is obtained from a Norwegian bank. The data is also limited to corporate customers and cannot be generalised to private customers. It can be argued that it is limited specifically to the Norwegian bank that is being used, though it is plausible to think that the coherence, if it exists, can be generalised to other Norwegian banks since most Norwegian banks use the same rulebased system for transaction flagging (Jonassen, 2022).

6.3 Variables included in model 1 and 2

This section presents an overview of the two models discussed in the analysis, in addition to the two models used to test robustness.

6.3.1 Variables used in model 1 (Without change characteristics)

Overview of the variables used in model 1 is displayed in table 6.1. This model includes no variables representing change, since one of the main purposes of this study is to analyze the effect of change characteristics which is done by comparing it to model 2.

Table 6.1: Variables included in model 1, where no variables representing change are being used.

Name	Type	Description
Probability of default (PD)	Numeric	Values from 0% to 20%. A value close to 0% indicates a low probability of default, while a value close to 20% indicates a high probability of default.
Customer size	Categorical	Customer size variable made categorical. Very Small: $<$ 0.1 mill (Used as reference group) Small: [0.1 mill, 0.5 mill) Medium: [0.5 mill, 2 mill) Large: [2 mill, 10 mill) Very Large: $\geq$ 10 mill
Enhanced CDD	Binary	1 if the bank has implemented enhanced customer due dilligence (CDD), 0 otherwise
Inactive customer	Binary	1 if the customer has been inactive for 12 months, 0 otherwise
Number of deposit account	Numeric	Number of deposit accounts
Online banking	Binary	1 if the customer has an online bank account, 0 otherwise
Number of urgent payments	Numeric	Numeric of urgent payments the last year
Number of foreign payments	Numeric	Number of foreign payments last 2 years
Taxable abroad	Binary	1 if the beneficial owner (BO) is taxable abroad, 0 otherwise
Foreign domicile	Binary	1 if the BO has a foreign domicile, 0 otherwise
Foreign citizen	Binary	1 if the BO has citizenship abroad, 0 otherwise

6.3.2 Variables used in model 2 (With change characteristics)

Model 2 is an extension of model 1, where variables representing change in customer characteristics were included. It includes the variables from model 1 displayed in table 6.1, in addition to those displayed in table 6.2.

Table 6.2: A list of the variables added to model 1 to form model 2. Makes it possible to study the effect of change characteristics.

Name	Type	Description
Change in PD (1)	Binary	1 if there has been a change in PD the last year, 0 otherwise
Change Customer size	Categorical	Change in customer size the last month 0: 0% (Used as reference group) 1: 100% 2: < -100% 3: [-100%, 0%) 4: (0%, 100%) 5: > 100%
Change Enhanced CDD	Binary	1 if the bank has implemented enhanced CDD the last month, 0 otherwise
Change Taxable abroad	Binary	1 if the BO has changed its tax location to abroad, 0 otherwise
Change Foreign domicile	Binary	1 if the BO has changed its domicile location to abroad, 0 otherwise
Change Foreign citizen	Binary	1 if the BO has changed its citizenship to abroad, 0 otherwise

6.3.3 Test for Robustness

To test the results, two robustness tests were performed. In the first one, *Change in PD (1)* is replaced with *Change in PD (2)*. This was done since new customers had been categorized as having a change in PD score. Since this is not considered as an actual change in this analysis, it was replaced with *Change in PD (2)*, where the new customers were coded to have no change in PD score. Then, real change could be examined in context with the probability of being flagged or reported.

The second test was done due to the high correlation between *Change in PD (1)/(2)* and *Change enhanced CDD*. *Change enhanced CDD* was chosen to be excluded because of the purpose of this study. This is elaborated more in detail in the data and analysis section.

7 Results and discussion

This section presents and discuss the outputs from the different logit models. The analysis resulted in two main models with and without variables for change characteristics. Two additional models are also discussed for robustness, which can be found in appendix A3.1 (model 3), and A4.1 (model 4). The main focus will be on model 1 and 2, while 3 and 4 are only commented. All four models have three versions, since three different dependent variables have been used in the analysis.

7.1 Model 1: Without change characteristics

Table 7.1 shows the results from model 1. All three models have a p-value close to 0, and are therefore significant at 1% level. The model explaining the bank's automatic system (*Flagged^b automatic*) has the highest pseudo R Square value (0.1897), while the model explaining reports (*Reported*) has the lowest (0.1056).

Probability of default (PD) is only significant, at 10% level, for the model with *Flagged^b automatic* as the dependent variable. This means that there is a weak significant link between a customer's PD score and the probability of being flagged by the bank's automatic system. The coefficient is positive, which means that the higher the PD, the higher the probability of being flagged. In model 1a and c (respectively where *Flagged^a all* and *Reported* are used as dependent variable) PD is not significant, indicating that an inclusion of PD in the bank's automatic system can lead to lower efficiency. It is plausible that a customer whom suffer economically might be tempted to attempt tax evasion or other "white-collar crimes", which could be enough grounds to assume there should be a connection. If this was true, the model explaining reports should have reflected significant values for *PD*. Since it isn't, it's more reasonable to assume that the link isn't as assumable as initially thought. *PD* is however not excluded in upcoming models due to comparison reasons, to study the effect of change compared to a stationary score, like previous research such as Gao and Ye (2007) motivates.

Table 7.1: Results Model 1. Where no variables for change is being used.

Dependent variable	**Flagged[a] all**	**Flagged[b] automatic**	**Reported**
	Model 1a	Model 1b	Model 1c
Number of obs.	8,358	8,358	8,358
χ^2	338.66	287.71	132.20
P-value	0.000	0.000	0.000
Pseudo R^2	0.1668	0.1897	0.1056
Probability of Default (PD)	2.449	3.919*	-4.024
Customer size:			
- Small	0.380*	0.135	0.709**
- Medium	1.015***	0.959***	0.717*
- Large	0.705***	0.719**	0.636
- Very Large	0.853***	0.599*	1.037**
Enhanced CDD	1.621***	1.571***	1.389***
Inactive customer	-0.921***	-1.278***	-0.186
Number of deposit account	0.148***	0.169***	0.047
Online banking	-0.670***	-0.802***	-0.558*
Number of urgent payments	0.008**	0.006*	0.011**
Number of foreign payments	0.009**	0.010**	-0.020
Taxable abroad	0.567	0.696	0.176
Foreign domicile	1.101***	1.200***	1.457***
Foreign citizen	1.066***	1.042***	1.455***

*** indicates significance at 1%, ** indicates significance at 5% and * indicates significance at 10%.

Customer Size represents several significant values for all three dependent variables. The basis for comparison is *very small.* The model with *Flagged[a] all* shows that *medium, large* and *very large* are significant at 1%, while *small* at 10%. In the model with *Flagged[b] automatic* as dependent variable, *small* is no longer significantly different from *very small.* The other groups are still significant, where *Medium* is the only group that is significant at 1% level. One aspect on customer size is that that money laundering is size independent. If this is true for the bank's system, *Customer size* should show no significant variables because no significance in context of customer size means no difference between groups in terms of flagging probability. This means that the bank's automatic system is size dependent in terms of flagging probability, where a medium customer (from 0.5 to 2 mill) has the highest probability of being flagged.

When comparing the model representing the bank's automatic system (*Flagged*[b] *automatic*) with the model representing reports (*Reported*), medium customers are no longer significant at 1% level, but at 10%. The group also has a lower coefficient, and is replaced with large as the group with the highest coefficient (1.037). This indicates that several of the flagged customers in this group are false cases, meaning they are not reported. A reason for this might be that the threshold values have been set in a way that makes the system react on transactions that more regularly come from this customer group (*medium*), leading to more false cases.

If the assumption of equal chance of money laundering between customer sizes is true, it should be reflected in the model representing *Reports* through no significant values for the different customer size groups. The model however shows the opposite. Customer size *large* is not significant, meaning no difference in terms of flagging between *large* and *very small* customers. *Small* and *very large* is significant at the 5% level, *medium* at 10% , indicating a difference. This doesn't necessary mean that the assumption is wrong. The differentiation occurs in the first stage of the detection process through automatic handling, meaning the differentiation is only regulated by dismissing the customers already detected by the bank's automatic system. It is in other words not possible to correct the differentiation entirely, since it requires that all transactions that are automatically handled is reviewed. The differentiation in the model representing reports (model 1c) could therefore be just a result of previous differentiation from the bank's automatic system.

There are arguments that could justify significant differences between customer sizes. As Ringen and Gjesdahl (2019) presents, customers who want to launder often use several banks to conceal their activity. From a banks' perspective, they would then be considered small in size, when in actual size they would be large. This gives reason to assume more money laundering from small customers and therefore reason to justify a significant difference between groups. However, it is possible that such bias is only detected since the data used is from one bank only. If the data contained information about the actual size of the customer (e.g. the sum of the customer size registered in all the banks used by the customer), the assumption about equal chance of money laundering between customer size is more plausible. It is therefore more reasonable to assume no differentiation when generalizing.

The models presents a few other points worth noticing. *Inactive customer*, *Number of deposit accounts* and *Online banking* show significant values at 1% level for the models using *Flagged*[a] *all* and *Flagged*[b] *automatic*, while for the model representing *Reported* only *Online banking* is significant at 10% level. This means that several of the false cases could possibly be explained through these variables. *Online banking* is negative for both models explaining flagged and reported cases, which means that customers with online bank accounts are less likely to get flagged and reported. The result could come from new customers who are trying to set up an account but are reported before they can start using the bank. Another reason might be that the customers without bank accounts are mostly large customers which accounts for more flagged cases. Lastly, two of the three beneficial owner (BO) characteristics are significant. *Foreign domicile* and *Foreign citizen* are significant at a 1% level for all three models, while *Taxable abroad* is not significant for either flagged or reported cases. This indicates that inclusion of such characteristics in an automatic system would be beneficial.

7.2 Model 2: With change characteristics

In model 2, several variables for change were introduced to study the effect of change in customer characteristics. Table 7.2 presents the results. The model is significant with a p-value close to 0 for all three dependent variables. The Psuedo R Square is above 21% for all models which is an improvement from model 1. The model using *Reported* had an Pseudo R Square of 0.1056 in model 1 which has increased to 0.2239. This shows that the variables for change strengthens the model significantly, confirming the hypothesis that change in customer characteristics could be a good predictor to detect money laundering. This underlines the point of Gao and Ye (2007), that emphasize the importance of customer history when analyzing transactions for suspiciousness.

Table 7.2: Results Model 2. Includes change variables in addition to those already presented for model 1.

Dependent variable	**Flagged[a] all**	**Flagged[b] automatic**	**Reported**
	Model 2a	Model 2b	Model 2c
Number of obs.	8,358	8,358	8,358
χ^2	393.83	317.44	256.50
P-value	0.000	0.000	0.000
Pseudo R^2	0.2196	0.2205	0.2239
Probability of Default (PD)	2.557	3.948*	-3.603
Change in PD (1)	2.508***	1.723***	3.452***
Customer size:			
- Small	0.284	-0.014	0.833**
- Medium	0.944***	0.854***	0.794*
- Large	0.721**	0.685**	0.711
- Very Large	0.876***	0.584*	1.220**
Change customer size:			
- 100%	-0.715	-1.846*	0.113
- <-100%	1.586***	1.493***	1.727***
- [-100%, 0%)	1.167***	1.087***	1.036**
- (0%, 100%)	1.168***	1.042***	1.151**
- >100%	1.860***	1.604***	2.324 ***
Enhanced CDD	1.583***	1.499***	1.422***
Change enhanced CDD	0.916**	1.028**	0.360
Inactive customer	-0.366	-0.707**	0.355
Number of deposit account	0.101**	0.128***	-0.019
Online banking	-0.569***	-0.756***	-0.249
Number of urgent payments	0.007*	0.004	0.010*
Number of foreign payments	0.008**	0.009***	-0.020*
Taxable abroad	0.640	0.657	0.179
Change taxable abroad	1.153	1.194	3.215*
Foreign domicile	1.184***	1.247***	1.498***
Change foreign domicile	1.786	(empty)	3.851***
Foreign citizen	1.255***	1.009***	1.947***
Change foreign citizen	-0.284	0.060	-1.255**

*** indicates significance at 1%, ** indicates significance at 5% and * indicates significance at 10%.

Table 7.2 shows that PD is still only significant at 10% level in the model using *Flagged[b] automatic* as dependent variable. The variable *Change in PD (1)* is significant at 1% for all three models. Which indicates that a change in PD is a better predictor than the customers PD score itself. The model using *Reported* as dependent variable has the highest coefficient in *Change in PD (1)*, which means that an automatic system that use *Change in PD (1)* would most likely improve its prediction accuracy. This however needs to be tested in a prediction model.

Change customer size shows significant values at 1% level for all categories except *100%* change in the models using *Flagged[a] all* and *Flagged[b] automatic*. For the model using *Reported*, *less than -100%* and *above 100%* is significant at 1%, while *between -100% and 0%* and *between 0% and 100%* is significant at 5%. *100%* change is only significant at 10% level in the model using *Flagged[b] automatic*. The category *100%* change was created to isolate new customers to study change in existing customer relationships. Basis for comparison is *0%*. Since the other categories shows significant values for all three dependent variables, these results reflects the benefits of including change characteristics in money laundering prediction. The categories *less than -100%* and *above 100%* change has the highest coefficients and is significant at 1% level in all three models, which indicates that customers with large changes in size has a higher chance of getting flagged and reported. The model using *Reported* also shows lower coefficients for changes *between -100% and 0%*, and *between 0% and 100%* compared to *less than -100%* and *above 100%*. This reflects that large changes results in higher probability of being reported compared to small changes.

The variables representing changes in BO characteristics, are only significant in the model explaining reports. *Change foreign domicile* is significant at 1% level, *Change foreign citizen* is significant at 5% level and *Change taxable abroad* is significant at 10% level. This shows that the bank could benefit from including such factors in their automatic system. *Change in foreign citizen* has a negative coefficient in the model, while the two others have a positive coefficient. This means that if a customer has a change in citizenship, the probability of getting reported is reduced. If it is taxable abroad or has a foreign domicile however, the probability increases.

7.3 Tests for robustness

Two additional models were made in order to test the robustness of the two main models. These are found in appendix A3.1 and A4.1.

Table 7.2 shows that *Change in PD (1)* has a significant effect at 1% for all three dependent variables. In *Change in PD (1)*, potential new customers are categorized to have a change in PD. Since this is not considered as an actual change in PD, a new variable was made to isolate the customers that have an actual change in PD, *Change in PD (2).* Model 3 (found in appendix A3.1) is therefore a reflection of model 2, but *Change in PD (1)* is replaced with *Change in PD (2).* In model 3, *Change in PD (2)* is no longer significant in the model using *Flagged^b automatic*, but positive and significant at 1% level in the model using *Reported.* This could indicate that *Change in PD (2)* is only linked to the manual reports, which means that the conclusion about including change in PD in the bank's automatic system still holds. Aside from this, the model alteration doesn't lead to any other remarkable changes in the other variables.

The variable *Change in PD (1)* has a strong correlation with *Change enhanced CDD*, which led to exclusion of *Change enhanced CDD* in model 4. The reason behind this choice lies in the focus of this article. Enhanced control is already used by the bank as a predictor, which makes it more interesting to study change in PD instead. Appendix A4.1 represents the new version of model 2, model 4, with *Change in PD (2)* and without *Change enhanced CDD.*

By removing *Change Enhanced CDD*, *Change in PD (2)* is again significant at 5% level in the model for *Flagged^b automatic.* This could indicate that *Change enhanced CDD* is the actual predictor of the flagged cases, and that the significance of *Change in PD (2)* is just caused by the correlation. The significance level of *Change in PD (2)* is however stable for the model using *Reported* from model 2 to 4, which means that the conclusion is still the same. There is a higher probability of being reported for customers with a change in PD score than for those without. The bank would therefore benefit from including change in PD as a predictor. Otherwise, no other changes in the model are worth mentioning.

8 Conclusion

The main purpose of this article isn't to form a prediction model for money laundering, but to research whether there exists causality between new variables and a Norwegian bank's AML system. This article is to the authors' knowledge the first to perform inference research on actual transaction monitoring data in business banking in a Norwegian context. The study by Jullum et al. (2020) is the closest resemblance to this article in terms of data, but like the majority of other previous research like Deng et al. (2009) and Lopez-Rojas and Axelsson (2012), the article focus on prediction accuracy. The scope of this article is in other words an area of AML that has little to no previous research, and is therefore considered a valuable contribution. The article's findings can be used as input on how AML measures should be elaborated to compensate for skewness in flagged and reported cases.

Three different dependent variables have been used in the analysis; one which represents the bank's detection system as a whole, one for just automatic flagged cases and one for manual reports. This has made it possible to evaluate which characteristics a reported customer has, and whether the bank's automatic system reflects the same or not. The analysis resulted in two main models, one with and one without variables for change, mainly to study the effect of change characteristics. The article's main findings will focus on a customer's PD score, the customer's size and whether it has had any changes in characteristics or not.

The model that represents the bank's automatic system shows a weak significant link between a flagged customer and its PD score. The model using reported cases however, shows that there is no connection between the customers PD score and the probability of being reported. This means that including PD score would not help the bank's automatic system, since it might just increase the number of false positive cases. This doesn't necessary mean that a customer's PD score is irrelevant. Whether a customer had a change in PD score or not turned out to be significant for both the bank's automatic system and for the reported cases. The relationship is positive, which means that a change in PD would lead to a higher probability of being reported. It is therefore possible to assume that a detection system that manage to capture a change in a customer's PD score would improve it's detection rate.

The general aspect on money laundering in terms of customer size is that it is size independent. The analysis presents significant values, which means the bank's system differentiate on customer size in terms of probability of being flagged and reported. With this perspective, the analysis uncovers a point of improvement in the bank's detection system. Since the data used in this article is limited to one singular bank, there is also reason to assume some differentiation (with possible higher probability for smaller customers) like Ringen and Gjesdahl (2019) pinpoints, due to customers spreading their activity between several banks. Some differentiation could therefore be justified to some degree. The conclusion about customer size would still be the same; the bank's automatic system would be significantly improved if customer size is taken into account.

Represented by both PD and customer size, including change in customer characteristics would likely have a positive effect on money laundering detection. The analysis shows that the probability of being flagged or reported increases if a customer has large changes in size. This is also confirmed by the Beneficial Owner (BO) characteristics, where the analysis shows a coherence between changes in such characteristics and the probability of being reported. This means that change in customer characteristics might be a good predictor of suspicious behavior, since a significant number of flagged and reported customers had changes when they were detected.

Comparison of the two models shows that change characteristics can possibly contribute to improve an already existing transaction-based prediction model. As both Gao and Ye (2007) and Kingdon (2004) points out in their study, being too transaction-oriented like traditional rule-based detection systems are today has several downsides. The findings in this article can therefore contribute by giving the systems a more subject- and history-oriented perspective while still maintaining the detection at transaction-level like Jullum et al. (2020) motivates, and hopefully increase the efficiency of AML measures in banks. Due to the immense societal gain from increasing the efficiency of AML-measures like Barone and Masciandaro (2010) states through their research, the findings in this article are considered as a valuable contribution in the field of AML.

8.1 Further Research

Since the research is solely an inference study and not a prediction study, possible further research would be to test the conclusions and see whether they actually would lead to better prediction models. It would also be interesting to perform the same analyses on data from other Norwegian or foreign banks, to see if the findings can be generalised or not.

References

Barone, R. & Masciandaro, D. (2010). Organized crime, money laundering and legal economy: theory and simulations. *European journal of law and economics*, *32*(1), 115–142.

Camdessus, M. (1998). *Money Laundering: the Importance of International Countermeasures*. Retrived 11. February 2022, from: https://www.imf.org/en/News/Articles/2015/09/28/04/53/sp021098.

Chen, Z., Van Khoa, L. D., Teoh, E. N., Nazir, A., Karuppiah, E. K., & Lam, K. S. (2018). Machine learning techniques for anti-money laundering (AML) solutions in suspicious transaction detection: a review.

Claessens, S. & Naude, D. (1993). *Recent Estimates of Capital Flight*. International Economics Department, The World Bank. https://www.researchgate.net/publication/23721660_Recent_Estimates_of_Capital_Flight.

Coase, R. H. (1937). The nature of the firm. *Economica*, *4*(16), 386–405. http://www.jstor.org/stable/2626876.

Cotoc, C.-N., Nițu, M., Șcheau, M. C., & Cozma, A.-C. (2021). Efficiency of money laundering countermeasures: Case studies from european union member states. *Risks (Basel)*, *9*(6), 120. https://doi.org/10.3390/risks9060120.

Deng, X., Joseph, V. R., Sudjianto, A., & Wu, C. F. J. (2009). Active learning through sequential design, with applications to detection of money laundering. *Journal of the American Statistical Association*, *104*(487), 969–981. https://doi.org/10.1198/jasa.2009.ap07625.

DNB (2022). *Anti-hvitvasking arbeid og bekjempelse av terrorfinansiering [Anti-money laundering work and the fight against terrorist financing]*. Retrived 24. January 2022, from: https://www.dnb.no/om-oss/samfunnsansvar/Okonomisk-kriminalitet/anti-hvitvasking-og-terrorfinansiering.html.

FATF (2022). *FATF Members and Observers*. Retrived 11. May 2022, from: https://www.fatf-gafi.org/about/membersandobservers/.

Finans Norge (2017). *Ofte stilte spørsmål - Hvitvaskingsregelverket [Frequently Asked Questions - Money Laundering Regulations]*. Retrived 7. Mars 2022, from: https://www.finansnorge.no/aktuelt/nyheter/2017/10/ofte-stilte-sporsmal-om-hvitvasking/.

Gao, Z. & Ye, M. (2007). A framework for data mining-based anti-money laundering research. *Journal of money laundering control*, *10*(2), 170–179.

Gibney, N. (2019). *Does Size Really Matter?* [Master's thesis, Norges Handelshøyskole (NHH)]. NHH Brage. https://openaccess.nhh.no/nhh-xmlui/handle/11250/2609822.

Grosan, C. & Abraham, A. (2011). *Rule-Based Expert Systems*, pages 149–185. Springer Berlin Heidelberg, Berlin, Heidelberg.

Hafthun, A. L. & Ulfsnes, H. (2019). *Rapportering av mistenkelige transaksjoner i banker : hva påvirker bankansattes beslutning om å rapportere? [Reporting suspicious transactions in banks: what influence bank employees' decision to report?]*. [Master's thesis, Norges

Handelshøyskole (NHH)]. NHH Brage. https://openaccess.nhh.no/nhh-xmlui/handle/11250/2612028.

Hammervold, R. (2020). *Multivariate analyser med STATA : en kort innføring [Multivariate analyzes with STATA: a brief introduction].* (1. ed.). Fagbokforlaget.

Hassoumi, N. (2021). *What the EU 6th AMLD means for your compliance programme.* Retrived 3. February 2022, from: https://www.refinitiv.com/perspectives/regulation-risk-compliance/what-the-eu-6amld-means-for-your-compliance-programme/.

Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression.* (3. ed.). John Wiley and Sons. https://doi.org/10.1002/9781118548387.

Høgseth, M. H., Haugen, A., & Nilsen, A. A. (2019). *Fiskerigigant skal ha brukt DNB for å overføre penger til skatteparadis [Fishing giant is said to have used DNB to transfer money to tax havens].* Retrived 03. May 2022, from: https://e24.no/boers-og-finans/i/LAXRd9/medier-fiskerigigant-skal-ha-brukt-dnb-for-aa-overfoere-penger-til-skatteparadis.

Iversen, M. L. (2021). *Hvitvasking [Money-laundering].* Retrived 21. January 2022, from: https://www.okokrim.no/hvitvasking.422268.no.html.

James, Witten, D., Hastie, T., & Tibshirani, R. (2021). *An Introduction to statistical learning : with applications in R.* (2. ed.). Springer.

Johnson, J. & Desmond Lim, Y. C. (2003). Money laundering: has the financial action task force made a difference? *Journal of financial crime*, *10*(1), 7–22.

Jonassen, O. (2022). *How to succeed with Anti-Money Laundering? Best practice from Norway.* Retrived 20. May 2022, from: https://www.tietoevry.com/en/newsroom/all-news-and-releases/articles/2021/how-to-succeed-with-anti-money-laundering-best-practice-from-norway/.

Jullum, M., Løland, A., Huseby, R. B., Ånonsen, G., & Lorentzen, J. (2020). Detecting money laundering transactions with machine learning. *Journal of money laundering control*, *23*(1), 173–186. https://doi.org/10.1108/JMLC-07-2019-0055.

Kingdon, J. (2004). Ai fights money laundering. *Intelligent Systems, IEEE*, *19*, 87 – 89.

Kristiansen, J.-E. (2010). *Regnskapstreff i sør 2010. Skattekriminalitet og skatteparadis [Accounting meeting in the south in 2010. Tax crime and tax havens].* Retrived 19. May 2022, from: https://www.yumpu.com/no/document/read/23060035/jan-egil-kristiansen-skatt-ast.

Lopez-Rojas, E. A. & Axelsson, S. (2012). Money laundering detection using synthetic data. *Annual workshop of the Swedish Artificial Intelligence Society (SAIS).* https://www.diva-portal.org/smash/get/diva2:834701/FULLTEXT01.pdf.

McDowell, J. & Novis, G. (2001). The consequences of money laundering and financial crime. *Economic Perspectives*, *6*(2), 6–8.

Nilsen, B. R. (2021). *Slik blir svarte penger hvite [How money laundering takes place].* Retrived 7. Mars 2022, from: https://www.finansforbundet.no/folk-og-fag/felles-ansvar/slik-blir-svarte-penger-hvite/.

Nordvik, E. M. A. & Reitan, K. A. (2020). *Anti-hvitvasking i norske sparebanker [Anti-money laundering in Norwegian savings banks]*. [Master's thesis, NTNU]. NTNU Open. https://hdl.handle.net/11250/2777212.

Ringen, G. H. & Gjesdahl, K. (2019). *Den nye hvitvaskingsloven (AML) – Dette må du vite [The new Anti Money Laundering Act (AML) - You need to know this]*. Retrived 24. January 2022, from: https://www.pwc.no/no/pwc-aktuelt/den-nye-hvitvaskingsloven-skaper-hodebry.html.

Roaldsøy, A. (2018). *Ny hvitvaskingslov: Større ansvar tildeles finansnæringen [New Anti Money Laundering Act: Greater responsibility is assigned to the financial industry]*. Retrived 24. January 2022, from: https://www.finansnorge.no/aktuelt/nyheter/2018/10/ny-hvitvaskingslov--storre-ansvar-tildeles-finansnaringen/.

Rui, J. P. (2012). *Hvitvasking : fenomenet, regelverket, nye strategier [Money laundering: the phenomenon, the regulations, new strategies]*. Universitetsforl.

Schultz, J. (2019). *De tre største bankene i Norge har brukt over 15 milliarder kroner på å bekjempe økonomisk kriminalitet [The three largest banks in Norway have spent more than NOK 15 billion on fighting economic crime]*. Retrived 24. January 2022, from: https://www.dn.no/marked/thomas-midteide/hans-thrane-nielsen/jan-petter-sissener/de-tre-storste-bankene-i-norge-har-brukt-over-15-milliarder-kroner-pa-a-bekjempe-okonomisk-kriminalitet/2-1-718618.

Stigler, G. J. (1964). Public regulation of the securities markets. *The Journal of Business*, *37*(2), 117–142. http://www.jstor.org/stable/2351027.

Studenmund, A. H. (2017). *A Practical guide to using econometrics*. (7th edition, global edition.). Pearson.

the Anti-Money Laundering Act (2018). *Act relating to Measures to Combat Money Laundering and Terrorist Financing (LOV-2018-06-01-23)*. Retrived 21. January 2022, from: https://lovdata.no/dokument/NLE/lov/2018-06-01-23.

The World Bank (2021). *GDP (current US$)*. Retrived 12. May 2022, from: https://data.worldbank.org/indicator/NY.GDP.MKTP.CD?most_recent_value_desc=true.

Trumpy, J. & Lund, K. (2021). *Tidligere Økokrim-direktør: To grunner til at norske banker er ekstra utsatt for hvitvasking [Former Økokrim director: Two reasons why Norwegian banks are particularly vulnerable to money laundering]*. Retrived 22. February 2022, from: https://www.dn.no/marked/hvitvasking/bill-browder/hermitage-capital-management/tidligere-okokrim-direktor-to-grunner-til-at-norske-banker-er-ekstra-utsatt-for-hvitvasking/2-1-453719.

Tufte, P. A. (2000). *En intuitiv innføring i logistisk regresjon [An intuitive introduction to logistic regression]*. Retrived 29. April 2022, from: https://oda.oslomet.no/oda-xmlui/bitstream/handle/20.500.12199/5950/file48351_arbeidsnotat08-2000web.pdf?sequence=1&isAllowed=y.

United Nations: Convention Against Illicit Traffic in Narcotic Drugs and Psychotropic Substances (1989). *International legal materials*, *28*(2), 493–526. https://doi.org/10.1017/S0020782900021756.

UNODC (2022a). *About UNODC*. Retrived 5. Mars 2022, from: https://www.unodc.org/unodc/en/about-unodc/index.html.

UNODC (2022b). *Model Treaties and Laws*. Retrived 5. Mars 2022, from: https://www.unodc.org/unodc/en/legal-tools/model-treaties-and-laws.html.

Walker, J. (1999). How big is global money laundering? *Journal of Money Laundering Control*, *3*(1), 25–37.

Økokrim (2022). *Statistikk MT-rapporter [Statistics MT reports]*. Retrived 11. May 2022, from: https://www.okokrim.no/statistikk-mt-rapporter.6156901-475337.html.

Appendix

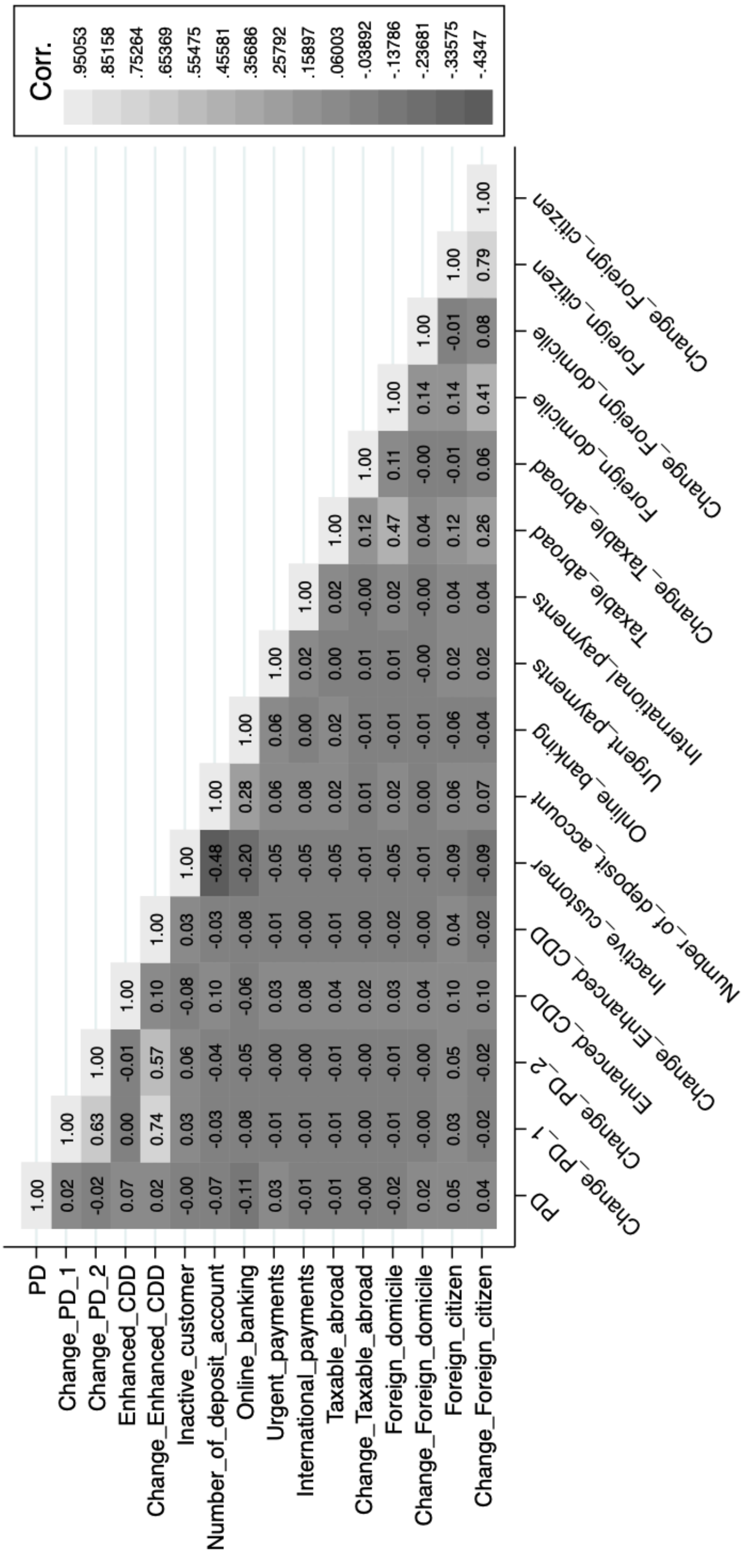
A1 VIF index for model 2

Table A1.1: Variance Inflation Factor (VIF) index for model 2. VIF values over 5 indicates possible multicollinearity problems.

Dependent variable	**VIF**
Probability of Default (PD)	1.06
Change in PD (1)	2.91
Customer size:	
- Small	1.33
- Medium	1.32
- Large	1.19
- Very Large	1.14
Change Customer size:	
- 100% change	1.22
- <-100% change	1.17
- [-100%, 0%)	1.65
- (0%, 100%)	1.56
- >100%	1.17
Enhanced CDD	1.08
Change Enhanced CDD	2.90
Inactive customer	2.00
Number of deposit account	1.49
Online banking	1.17
Urgent payments	1.01
International payments	1.02
Taxable abroad	1.30
Change Taxable abroad	1.02
Foreign domicile	1.61
Change Foreign domicile	1.04
Foreign citizen	3.02
Change Foreign citizen	3.54
Mean VIF	1.58

A2 Correlation matrix

Figure A2.1: Heatmap displaying correlation between independent variables. A number close to 1, or a white color, indicates high positive correlation, and a number close to -1, or a dark grey color, indicates high negative correlation.



A3 Result for model 3

Table A3.1: Results Model 3. The same variables included for model 2 are used, except that *Change PD (1)* is replaced with *Change PD (2)*.

Dependent variable	**Flagged[a] all**	**Flagged[b] automatic**	**Reported**
	Model 3a	Model 3b	Model 3c
Number of obs.	8,358	8,358	8,358
χ^2	411.11	338.46	283.64
P-value	0.000	0.000	0.000
Pseudo R^2	0.2164	0.2169	0.2303
Probability of Default (PD)	2.939	3.951*	-1.545
Change in PD (2)	2.507***	0.859	3.937***
Customer size:			
- Small	0.229	-0.061	0.763*
- Medium	0.906***	0.818***	0.770*
- Large	0.640**	0.633**	0.558
- Very Large	0.773***	0.510	1.069*
Change customer size:			
- 100%	0.461	-0.985	1.536
- <-100%	1.603***	1.508***	1.792***
- [-100%, 0%)	1.223***	1.123***	1.175***
- (0%, 100%)	1.190***	1.059***	1.218**
- >100%	1.867***	1.616***	2.399***
Enhanced CDD	1.533***	1.430***	1.463***
Change enhanced CDD	1.316**	1.564**	0.603
Inactive customer	-0.374	-0.637*	0.208
Number of deposit account	0.100**	0.129**	-0.033
Online banking	-0.612***	-0.787***	-0.302
Number of urgent payments	0.007*	0.004	0.001*
Number of foreign payments	0.008**	0.009**	-0.022
Taxable abroad	0.635	0.661	0.187
Change taxable abroad	1.152	1.214	3.170*
Foreign domicile	1.184***	1.239***	1.542***
Change foreign domicile	1.623	(empty)	3.347**
Foreign citizen	1.190***	1.025***	1.808***
Change foreign citizen	-0.256	0.053	-1.149**

*** indicates significance at 1%, ** indicates significance at 5% and * indicates significance at 10%.

A4 Result for model 4

Table A4.1: Results Model 4. The same variables included for model 2 are used, except that *Change PD (1)* is replaced with *Change PD (2)*, and *Change Enhanced CDD* is excluded.

Dependent variable	**Flagged[a] all**	**Flagged[b] automatic**	**Reported**
	Model 4a	Model 4b	Model 4c
Number of obs.	8,358	8,358	8,358
χ^2	376.94	311.90	261.33
P-value	0.000	0.000	0.000
Pseudo R^2	0.2130	0.2121	0.2297
Probability of Default (PD)	2.945	3.951*	-1.625
Change in PD (2)	3.672***	2.209**	4.480***
Customer size:			
- Small	0.261	-0.023	0.778*
- Medium	0.930***	0.846***	0.777*
- Large	0.661**	0.670**	0.560
- Very Large	0.818***	0.571*	1.085*
Change customer size:			
- 100%	1.084	-0.271	1.721
- <-100%	1.563***	1.467***	1.769***
- [-100%, 0%)	1.191***	1.091***	1.157***
- (0%, 100%)	1.166***	1.034***	1.210**
- >100%	1.858***	1.617***	2.395***
Enhanced CDD	1.647***	1.565***	1.518***
Inactive customer	-0.338	-0.604*	0.224
Number of deposit account	0.095*	0.119**	-0.035
Online banking	-0.613***	-0.787***	-0.306
Number of urgent payments	0.007*	0.004	0.010*
Number of foreign payments	0.008**	0.009**	-0.022
Taxable abroad	0.620	0.640	0.170
Change taxable abroad	1.126	1.181	3.172*
Foreign domicile	1.179***	1.230***	1.525***
Change foreign domicile	1.464	(empty)	3.285**
Foreign citizen	1.190***	1.005***	1.823***
Change foreign citizen	-0.283	0.035	-1.168**

*** indicates significance at 1%, ** indicates significance at 5% and * indicates significance at 10%.

NTNU

Kunnskap for en bedre verden