Master's thesis

NTNU
Norwegian University of Science and Technology

Rune Bach
Jakob Rief Holst

Exploring the Duration of Client-Bank Relationships in Mortgage Lending

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

NTNU
Norwegian University of Science and Technology

Rune Bach
Jakob Rief Holst

Exploring the Duration of Client-Bank Relationships in Mortgage Lending

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

Norwegian University of Science and Technology

NTNU
Norwegian University of
Science and Technology

Forord

Denne masteroppgaven ble utarbeidet i forbindelse med avslutningen på vår mastergrad i økonomi og administrasjon ved NTNU Handelshøyskole i Trondheim våren 2022. Denne prosessen har gitt oss økt kompetanse og forståelse for økonomi i praksis, hvor fokuset hovedsakelig har vært innen finans –og bankvirksomhet. Oppgaven, som er på 30 studiepoeng, har gjort klart hvor viktig det er å ha god forståelse av flere ulike temaer innenfor økonomifaget. Skrivingen har vært spennende og utfordrende, men takket være kompetansen vi har opparbeidet oss gjennom fem kunnskapsrike år føler vi oss godt fornøyde med resultatet og håper du som leser finner oppgaven nyttig og interessant.

Vi ønsker også å takke våre veiledere Are Oust og Endre Jo Reite for gode innspill og råd som har gitt oss verdifull innsikt i ferdigstillingen av oppgaven. Avslutningsvis ønsker vi å takke BN Bank for tilgangen til deres unike datasett.

Innholdet i denne oppgaven står for forfatternes regning.

Trondheim, Mai 2022

Rune Bach

Jakob Rief Holst

Abstract

Our research aims to improve banks' profitability by analyzing customer signals available before a relationship. We believe our research will improve banks' ability to identify long-lasting and profitable customers by analyzing real customer data. We tested four separate hypotheses about the bank-customer relationship using a robust multiple regression with a cross-sectional setup. The customer's duration, i.e. the full length of the bank-customer relationship, is used as a proxy for the bank's profitability. As for the independent variables, they have been separated into creditworthiness, demographic characteristics, and loan conditions. The dataset consists of 2579 observations from a local bank's mortgage portfolio, spanning from 2011 to 2022. We are the first people to receive access to this dataset, which grants our thesis unique data compared to existing research. During the dataset's timespan, two financial regulations were implemented. Therefore, the dataset was segregated into three sets, a table representing all observations, a table before 2017 and a table after 2017. The results indicated that poor creditworthiness led to shorter expected customer duration, but whether their effect is due to active risk management or due to regulations are hard to tell. Larger households have a longer expected duration. Younger borrowers have a shorter duration compared to older borrowers, this relationship is nonlinear but significant when segregated into age brackets. Finally, a customer's loan condition showed conflicting results. Some stricter conditions prolong the relationship while others seem to shorten it. To the best of our knowledge, our research paper is unique in analyzing factors impacting the longevity of a relationship between households and banks.

Sammendrag

Formålet med vår forskning er å forbedre bankers lønnsomhet ved å analysere kundesignaler som er tilgjengelige før innvilgelsen av et lån. Ved å analysere reelle kundedata, tror vi at vår forskning vil forbedre bankenes evne til å identifisere langvarige og lønnsomme kundeforhold. Ved å bruke en robust multippel regresjon med et tverrsnitt oppsett testet vi fire separate hypoteser om bank-kundeforholdet. Kundens varighet hos banken brukes som en proxy for bankens lønnsomhet, mens de uavhengige variablene er delt inn i tre kategorier; kredittverdighet, demografiske særtrekk og låneforhold. Datasettet består av 2579 observasjoner fra en lokal banks boliglånsportefølje, som strekker seg fra 2011 til 2022. Vi er de første som har blitt gitt mulighet til å jobbe med dette datasettet, noe som gir oss et unikt datagrunnlag sammenlignet med eksisterende studier. I løpet av datainnsamlingsperioden ble to finansforskrifter implementert. Datasettet ble derfor delt i tre sett; en tabell som representerer alle observasjoner, en tabell med observasjoner før 2017 og en tabell etter 2017. Resultatene indikerte at dårlig kredittverdighet førte til kortere forventet kundevarighet, men om deres effekt skyldes aktiv risikostyring eller er grunnet forskriftene er vanskelig å si. Større husholdninger har lengre forventet varighet. Yngre låntakere har kortere varighet sammenlignet med eldre låntakere. Alderen på kunden er ikke-lineært, men signifikant når alder deles inn i grupper. Kundens låneforhold viste motstridende resultater, noen strengere betingelser forlenger forholdet mens andre ser ut til å forkorte det. Så vidt vi vet er forskningsartikkelen vår unik når det gjelder å analysere faktorer som påvirker levetiden i forholdet mellom husholdninger og banker.

Table of Contents

List of Figures

List of Tables

1. Introduction

For banks, attracting and retaining long-lasting customer relationships is key to profitability (Storbacka *et al.*,1994). The motivation behind this paper is to provide useful insight into the bank-customer relationship. The research is conducted to optimize banks' decision-making in the private mortgage market by screening new customers who are more likely to be long-lasting. By analyzing informational signals given by customers, this paper will provide insight into the expected duration and profitability of a customer's relationship with the bank, limiting the banks' risk of attracting unprofitable customers. To the best of our knowledge, we are the first researchers to be given access to this set of customer data, granting our research paper a unique dataset.

This paper will investigate what factors affect the duration of the relationship between a bank and their customers**,** where the customers duration will serve as a proxy for the bank's profitability. We will analyze the informational signals given by the customers at the initiation. The information has been separated into three categories: creditworthiness, demographic characteristics, and loan conditions. This has further inspired the following research question: ***Can initial information about the customer give reliable signals regarding the bank's potential customer profitability?***

Previous research focuses on the relationship between mortgage pricing and customers switching mortgage banks, often from the customer's viewpoint. Ongena *et al.* (2021) found that the price difference on mortgages between current and new clients is 26 basis points, indicating a clear incentive for clients to switch mortgage banks. In a similar study, Lukas & Nöth (2019) found that interest rates are as high as 40 basis points between borrowers who switch to a new lender in contrast to non-switching customers when refinancing. In the same study, Lukas & Nöth (2019) also found that borrowers tend to switch banks in periods when interest rates increase compared to periods with declining interest rates. Furthermore, Brunetti *et al.* (2020) found that switching cost is a significant factor in explaining the switching rate of customers. By analyzing the switching rate before and after the Bersani reform, they found that the reform reduced the switching costs. Higher switching insinuates that the duration in which the customer stays at the bank stagnates.

While switching costs and mortgage interest rate differences have been found to determine the switching rate of customers, the duration of the customer relationship is less explored. Therefore, this study will analyze factors that have significant explanatory power on the duration of the customer relationship and the associated profitability that the customer longevity entails. The duration is used as a proxy for profitability because banks have costs associated with new loans given to the customer that only become profitable over time. Also, banks will often compete with several other financial institutions in attracting new customers. This often results in banks offering favorable terms to new customers (Carbo-Valverde *et al.*, 2011; Brunetti *et al*., 2020). Additionally, if one combines this with the winner's curse concept resulting from asymmetrical information (Thadden, 2004) and the extra guidance and attention expected to be given to new customers, the result is meager profitability from new customers. This lack of profit needs to be compensated, and the solution banks often take is to gradually increase interest rates for customers during the lifespan of their loan engagements (Brunetti *et al*., 2020; Sharpe, 1997; and Sharpe, 1990).

To analyze the duration of customers, robust multiple regression analysis with a cross-sectional setup was used. The dataset includes a total of 2579 observations from a local medium-sized bank, between June 2011 and January 2022. A key aspect is that all observations within the dataset have already left the bank within the aforementioned timeframe. However, their entry date may vary prior to the specified period. The dataset also includes 14 variables separated into three categories: creditworthiness, demographic characteristics, and loan conditions. Furthermore, to accommodate for the effect of regulatory changes during the period of data collection, the dataset was divided into three; the first dataset contains all 2579 observations, the second dataset only includes the 1356 observations predating 2017, and the third dataset contains the 1223 observations after 2017.

In order to answer our research question, we developed four separate hypotheses to test the impact of important aspects concerning a customer relationship. For the most part the findings support our initial hypotheses. The analysis indicates that lower creditworthiness also led to shorter expected customer duration, older generations have longer expected customer relationships compared to younger generations, large families are expected to have longer duration than small families. However, an interesting finding is that borrowers offered the best interest rates at initiation is in fact the borrowers with the lowest expected customer duration. We have also explored the impact of newly implemented government regulations on the interaction between the dependent and the independent variables, hence the 2017 divide. After the Norwegian government implemented regulations restricting bank's ability to issue risky mortgage loans, one can see that explanatory factors such as loan-to-value (LTV) and debt-to-income (DTI) gain more importance.

The rest of this paper is organized first through an analysis of previous research and underlying theory. In section 3, it provides insight to the background of the Norwegian banking and lending market, its new regulations, and its characteristics. Section 4 describes the dataset, followed by section 5 which introduces the constructed hypothesis based on previous research and existing theory. Section 6 proceeds to the paper's methodology, including the setup, choice of model and the empirical process. In section 7 the result from the analysis is presented, while finally section 8 and 9 finishes the paper with a thorough discussion and a conclusion.

2. Related Literature

Researchers have tried to analyze the profitability of customer relationships within private retail banking. Early studies of service marketing have influenced the method to analyze a bank's profitability, called the bank's "service-profit chain" (Garland, 2001). The main components of this chain are customer retention, satisfaction, and the quality of service. Storbacka *et al.* (1994) viewed profitability in light of service quality and customer satisfaction, where customer service profitability is obtained through a strong relationship that generates steady income during the relationship's longevity. The paper also highlights that banks have three types of customers: those who are profitable, those who are unprofitable and those hovering around break-even. Storbacka *et al.* (1994) acknowledge that there is no point in keeping customers who cannot turn profitable and there is evidence of measures taken by banks to weed out such customers (Stern, 2012). Furthermore, Garland (2001) used activity-based cost accounting to measure the non-financial forces of customer profitability. Moreover, Reichheld (1992) wrote that a non-specific company gained a 20% increase in annual earnings by increasing the retention rate by only one percent. Among the research on a bank's profitability, there is a consensus that the customers that stay with their bank over a longer period are most profitable to the bank.

The duration of a bank-customer relationship is scarcely explored. To the best of our knowledge, no study has contained data and an analysis of the duration between a retail bank and their household customers. However, this study has similarities to that of Ongena & Smith (2001) in the sense that a similar method to that of a duration analysis is used to analyze the customer duration. Ongena & Smith (2001) analyzed the duration of a firm-bank relationship while differentiating firms by leverage, size and potential growth. Their results show that it is more likely to end as the relationship matures. Furthermore, the study indicates differences in expected duration based on firm characteristics, where small, highly leveraged firms will have shorter relationships than bigger and less leveraged firms. In contrast to their study, this study has obtained a unique dataset about private household customers and their characteristics. Additionally, more interest was given to the duration's profitability and how it benefits the bank. However, both papers are interested in the duration of a customer relationship and the characteristics of households could be seen as similar to those of a firm. The household's age, size and leverage ratios are somewhat comparable to the different characteristics between firms.

Existing literature seems to agree that banks entice new customers by offering them favorable terms at the beginning of the relationship, only to worsen the terms later as the loan matures and the customer becomes locked in (Carbo-Valverde *et al*, 2011; Brunetti *et al.*, 2020). This locked-in concept can take several forms, often resulting from actions taken by the bank, actions or lack of actions from the customer, or a combination of the two. By "locked-in", it is referred to situations where the customer is either unaware of or unable to obtain refinancing opportunities to improve their cost of financing. Such locked-in situations result in the loanee remaining a loyal customer of the bank, despite paying significantly more in interest payments than they otherwise would had they decided to refinance. Locked-in cases can purely be due to a customer's inattentiveness, inertia or general lack of financial sophistication (Cox *et al*., 2015; Andersen *et al.*, 2018). However, they can also be the product of actions taken by the bank to increase switching costs by decreasing transparency and increasing the complexity surrounding the loan obligation (Ongena *et al*., 2021).

Furthermore, this study has taken additional inspiration from studies that focus on the differences in interest rates. To understand the client-bank relationship in the mortgage market, it is critical to establish a foundation regarding the interest rates that the customer receives when initiating the relationship and how it evolves. Previous literature shows a significant difference in mortgage interest rates between new and old customers (e.g. see Lukas & Nöth, 2019; Ongena *et al*., 2021), which intuitively impacts the duration of the customer relationship. Studies show that new customers receive lower interest rates than older customers, indicating that customers will receive worse mortgage terms as the duration matures. As mentioned previously, banks do this to compensate for the lack of profit from new customers (Brunetti *et al.*, 2020; Sharpe, 1997; Sharpe, 1990). To analyze this context, this study includes the interest rate at initiation and whether the customer has a higher or lower interest rate than the bank's average lending rate.

3. Background

3.1 The Norwegian Mortgage Market

Capturing the underlying factors explaining the duration of a customer relationship requires an understanding of both the Norwegian mortgage market as a whole and the actors competing in it. According to Statista Research Department (2021), over 80 percent of the Norwegian population own their own home. Compared to the 28 European countries, this represents over 11 percent more the average of 69.3 percent (Statista, 2021). The high level of homeownership creates a particularly coveted mortgage market for banks. Because of the potential market gains, the Norwegian lending market is very competitive. According to Ongena *et al.* (2021), DNB has the highest market share of 25.6%, while remaining banks all have market shares below 12%. The Norwegian banking market is then composed of several small and medium-sized banks that make up the rest of the mortgage market. Additionally, the competitive market is developing, and digitalization has made document signings fully digital, and switching mortgages between banks can now be concluded within days and at a low cost (Ongena *et al.*, 2021).

3.2 The Mortgage Regulations

The Ministry of Finance, which regulates the Norwegian mortgage market, has made significant adjustments to the regulation in the lending market during the period that the data was collected. The Ministry of Finance made these regulations as a response to the high level of debt in Norwegian households that could cause risks in households meeting their future obligations, and to reduce the growth of the housing prices (Regjeringen, 2016). In 2020, Norway had in excess of 16% higher debt-to-income compared to the Nordic countries' average of 207.55 percent of income (Trading Economics, 2020).

Within a two-year period, the Norwegian Ministry of Finance implemented two important lending regulations on the bank and finance industry to reduce the debt growth of households. The first regulation came into effect on July 15. 2015, which stated that a mortgage loan should not exceed 85 percent of the justified value of the home (Regjeringen, 2015). The regulation is captured in the loan-to-value (LTV) variable and should limit individuals' ability to obtain loans without equity and assets. The second regulation came into effect on January

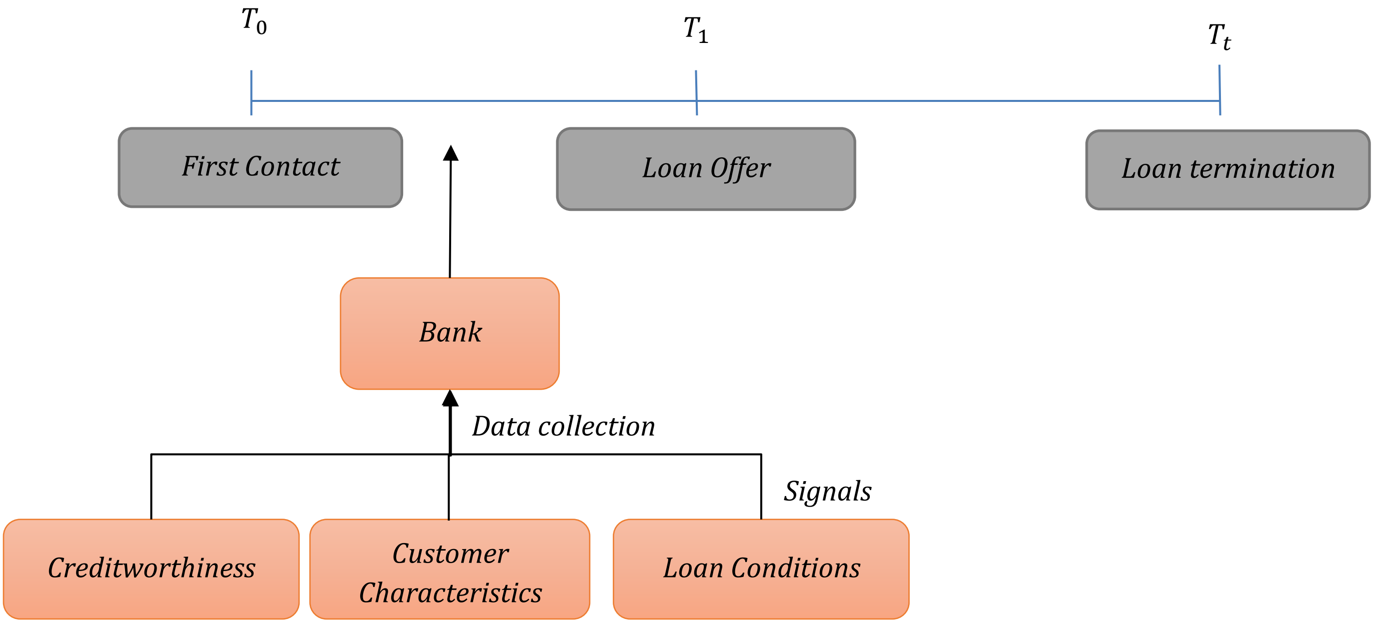
1. 2017 and stated that the customer's total loan should not exceed five times gross income (Regjeringen, 2016 (2)). This regulation is captured by the debt-to-income (DTI) variable. The regulations have put breaks on the Norwegian lending market, which is expected to impact the lending variables at the time of initiation. This difference should be particularly profound between customers that joined the bank before and after the beginning of 2017.

The government's mandated limit on debt compared to equity is 15%. The government has seemingly concluded that an LTV of more than 85% constitutes a critical level of debt with an increased risk of default. The regulations have a high chance of impacting the longevity of the customer duration, as it gives the banks greater insight and control of risk management. High-risk customers may see themselves switching banks because of their bank's stubbornness not to disregard the regulations, thus ending their relationship with said bank. According to the regulations imposed in 2015, banks may deviate from one or more of the strict terms for up to ten percent of the quarterly reported granted loans (Regjeringen, 2015). This deviation allows banks to pick which customers will be given an extra helping hand, which may vary between competing banks in each quarter.

3.3 Information Gathering Process

Figure 1 illustrates the information-gathering process between a bank and its customers. When customers apply for a bank loan, they give away a set of signals, either from their initial credit rating or through additional background checks conducted by the financial institution. The bank also receives information about the customers in demographic characteristics such as age and household size. This information is considered externally available, as all competing banks will have the same opportunity to register these signals about the customer. The signals, like good or bad credit history, can offer valuable information for the bank concerning the customers' ability and willingness to repay their future obligations. However, the amount of information the bank can garner from these externally available signals does not paint a complete picture of the customers' creditworthiness. The lack of a complete customer profiles results in asymmetrical information in favor of the borrowers, as they know more about their creditworthiness than the bank, and they have incentives to refrain from sharing personal information that could increase their cost of financing.

Figure 1. Overview of the information gathering process between banks and their borrowers



This paper wishes to explore these signals. The thesis aims to make a statement about whether any of these signals has a statistically significant predictive power on the potential profitability of the customer relationship. In order to achieve this aim, the paper is reliant on suitable variables to serve as proxies for externally gathered signals as well as a proxy for what constitutes a profitable bank-customer relationship from the bank's perspective.

4. Data

4.1 The Dataset

The dataset consists of observations from a medium-sized bank's mortgage portfolio. The observations span from June 2011 to January 2022. The dataset comprises of 14 variables, including the dependent variable, and contains 2579 observations. All the observations in the dataset come from real customers in our cooperating bank's loan portfolio. The customer information has been anonymized, and the dataset has been approved by the cooperating bank's compliance officers so that it should not be possible to uncover the identity of the customers behind the observations. The anonymization process involves granting each customer a random and non-traceable customer number and refraining from using any variable containing personal information that can be traced to the customer.

The collected data is centered around the information that said bank was able to procure during the issuance of the loan in question. The data is then stored on the bank's databases, some of which we were granted access to. The dataset can be described as random as it was decided to include all customers redeeming a mortgage loan during the selected period. The scope of the final dataset was naturally limited to only include mortgage loans that had an exit from the bank during the dataset collection. The timespan selected includes multiple periods with macroeconomic incidents that have the potential to influence the longevity of the bank-customer relationship. Such incidents include the rise of digital banking and the introduction of new market strategies such as low-cost online banking and exclusive union partnerships. However, even though all of these events can potentially affect the duration of the customer relationship, it seems likely that every customer will be affected by these events similarly and thus still comparable.

Due to the dataset being influenced by some macroeconomic factors, particularly the financial regulations imposed on the lending market between 2015 and 2017 (Regjeringen, 2015), it was deemed appropriate to analyze separate datasets reflecting the most important timeframes. Thus, the dataset was segregated into three sets, a table representing all observations, a table before 2017 and a set starting from 2017. The full dataset contains 2579 observations, where 1356 of the observations predate 2017 and 1223 of the observations are collected after 2017. The separation was useful to see changes in behaviors from both the bank and the customers, and how it affects the expected duration. The division is mainly based on the debt-to-income regulations that came into effect on January 1. 2017, in order to observe the significance of this regulation on the customers (Regjeringen, 2015). Even though the 2015 LTV-regulation could serve as a division basis, it was decided that it would serve a greater purpose to divide the dataset by the regulation that came into effect later, hence capturing both regulations in the analysis.

4.2 The Variables and Descriptive Data

This chapter will introduce the variables chosen for the regression analysis. The dependent variable and each independent variable will be given a short description followed by some selected comments regarding their descriptive statistics.

4.2.1 Customer Duration

The dependent variable `CustomerDuration` (in *table 1* below, and *table 5* and *6* in appendix) was created by calculating the difference between the start of the initial loan engagement and the redemption of the final loan engagement. Consequently, every observation in the dataset has a final exit date from the bank. Customer duration is measured in years and includes the total duration of a customer's cumulative engagements with the bank, i.e. internal refinancing, increased borrowings, and down payments. Descriptive statistics (*Table 1*) show that the observation with the shortest duration was only a customer with the bank for 11 days (0,0301*365.25), and the longest relationship with a final exit date spanned more than 33 years. The median duration for a borrower is around four years.

4.2.2 Creditworthiness

4.2.2.1 Loan-to-Value (LTV)

The LTV ratio is a measurement of the total outstanding loan compared to the underlying asset's value. High LTV ratios increase the risk of lending money to the borrower due to the underlying asset's role as collateral. Higher LTV will require less fluctuations before a proportion of the loan becomes unsecured. Most banks operate therefore with a strict policy governing the acceptable LTV ratio for any issued loan, depending on the risk level of the borrower (Akin, 2020).

$$LTV = \frac{Loan\ amount}{Market\ value\ of\ asset} * 100$$

The LTV ratio is a combination of the total loan amount and the market value of the underlying asset, measured in percentage. This means that higher LTV ratios represent either a high level of debt or a low market value of the underlying asset. According to *table 1* the lowest LTV value is 0, i.e. the dataset contains debt-free borrowers. The highest LTV ratio observed is 104%, with a median of 60%. The most common LTV value is 75%. Due to government regulations, it is expected that most LTV values will be below 85%.

4.2.2.2 Probability of Default (PD)

Unlike other risk metrics, the PD cannot be calculated by using a one-size-fits-all formula. Each loan is dependent on a number of customer-specific factors, which combined contribute to the calculated PD value. The PD variable in this paper is provided by the cooperative bank, and most banks operate with a maximum value applicable to a customer, which is set to 20 by said bank. This limits the internal risk the bank is willing to take regarding each individual customer. According to *table 1* the riskiest customer in this dataset has a PD of 12.358%, whilst the least risky customer has a PD of 0.04%. The median PD in the dataset is 0.53% and the mode lies at 0.083%. In other words, most borrowers have a PD below 1%.

4.2.2.3 Debt-to-Income (DTI)

The DTI ratio is a measurement of the total debt divided by the gross annual income of the customer. The DTI ratio highlights the income of a borrower, hence indicating the possible risk of lending to said borrower. Customers with higher DTI will be more likely to default on their obligations because a small increase in the costs will have a higher impact on their ability to service the debt. The DTI is given by the following formula:

$$DTI = \frac{Total\ household\ debt}{Gross\ annual\ income}$$

According to *table 1,* the dataset contains everything from debt-free borrowers (i.e. DTI = 0) to borrowers with 21 times DTI. The median DTI is 3.479 and the mode is 3.333, demonstrating that most DTI values are below the regulatory limit of 5.

4.2.3 Demographic Characteristics

4.2.3.1 Age

The age of the customer is originally given in a continuous variable measured in years but has later gone through a dummy-transformation measured in age groups. The age of the customer could influence the expected duration and was therefore implemented into the regression. *table 1* shows the bank contains customers between the age of 23 and 95, with a median age of 53.

4.2.3.2 Size of Household

The relationship between bank-customer duration and the size of the borrower's household is another relevant context we want to study. The size of the household regards the number of people living in the same home as the borrower and is measured by individuals. Household size could affect the expected duration of the relationship. A Hungarian study found that the probability of default increases with the number of dependents in a household (Papp, 2007). Therefore, it is interesting to find the relationship between the expected duration and the size of the household (`NrInHousehold` in *table 1*). The dataset contains households ranging from singles to families of eight, where the most common observation is singles, and the median observation is a household of 2.

4.2.4 Loan Conditions

4.2.4.1 Total Obligations

Total obligations constitute the total debt of the household, measured in NOK. It is an interesting metric in both its absolute terms and relative terms. Because total obligations are a result of the aforementioned risk metrics and regulations, borrowers who achieve large loans in absolute terms are more likely to be characterized by high income and personal wealth. At the same time, observations with large loans compared to equity could also represent individuals on the brink of financial insolvency. As a result, borrowers with large loans could represent both a valuable asset and a risky liability for the bank. According to *table 1,* the largest registered loan is 26 million NOK, while the median loan is 2.475 million NOK.

4.2.4.2 Interest Rate at Initiation

The interest rate at initiation is the rate offered to the customer at their first loan engagement with the bank. This interest rate will always be affected by the sum of risk metrics concerning the customer (Shea, 2017), combined with competition from rival banks. An important aspect of the interest rate is how it differs from those belonging to other borrowers. Comparing a customer's interest rate with that of the bank's average interest rate is therefore highly relevant. Two dummy variables capture this comparison between customers (`Highest20%` and `Bottom20%` in *table 1*). `Highest20%` reflects the 20% of customers offered the highest interest rate compared to the bank's average, while the opposite is true for `Bottom20%`.

Table 1. Descriptive statistics for the full set of observations

	Obs	Min	Max	Mean	Median	Std.Dev
CustomerDuration	2579	0,0301	33,3413	5,7403	3,9590	5,4930
Creditworthiness						
LTV	2579	0	104	56,4967	60	21,0882
PD	2559	0,04	12,358	0,8011	0,53	0,8664
DTI	2579	0	21,4286	3,4792	3,1746	2,0972
DTI2	2579	0	459,1837	16,5010	10,0781	26,2506
DTIAbove5	2579	0	1	0,1404	0	0,3474
Characteristics						
NrInHousehold	2579	1	8	2,2094	2	1,2448
AGE	2579	23	95	53,2276	53	13,2928
AGE2	2579	529	9025	3009,807	2809	1467,889
Youngsters	2579	0	1	0,0857	0	0,2799
Retirees	2579	0	1	0,1539	0	0,3610
Loan Conditions						
Highest20%	2579	0	1	0,2009	0	0,4007
Bottom20%	2579	0	1	0,2020	0	0,4016
TotalObligation	2579	50802,52	2,60e^(7)	3031362	2475000	2288645

Note:
See Table 9 for a short description of the variables

5. Hypothesis

5.1 Creditworthiness

High risk entails an increase in the possibility of defaulting on a given obligation, causing friction in the customer relationship. Due to banks being restricted by internal and governmental policies, a high-risk customer could be forced to search for external refinancing. This naturally reduces the expected duration of their relationship with the bank. Higher PD, LTV and DTI values are generally undesirable for a bank as it increases the risk associated with the banks' loan portfolio. As a result, banks will require compensation for the increased risk, such as higher interest rates (Kenton, 2021). According to an article written by the news outlet Reuters (Stern, 2012), when faced with undesirable customers, banks often implement measures such as credit card fees, service fees, and checking fees. The aim of these actions is either to change a customer's behavior or an attempt to get the customer relationship off their books. This could force the customer to seek better terms and conditions at other banks, reducing the expected duration. The combined effect of risks metrics leads to the following hypothesis:

$$\mathrm{H_1}: \text{Poor creditworthiness will reduce expected customer duration}$$

5.2 Demographic Characteristics

An article from the Financial Brand shows that younger generations are more prone to use online/mobile banking and are more inclined to adopt new financial technologies than older generations (McDade, 2016). These technologies reduce the transaction costs associated with changing banks, as elaborated in Ongena *et al.* (2021). This lowers the switching barriers, particularly for customers adopting the technology. Older generations meanwhile, often require support in their financial dealings, which increases the borrower’s tendency to focus on their relationship with the bank rather than interest savings (McDade, 2016). This results in an expectation that younger customers are more frequent switchers.

Furthermore, young adults are new to the labor market and will likely not have the same income and saved equity level as older generations. Therefore, younger generations may have more difficulty satisfying the government-mandated restrictions on LTV and DTI. Younger customers also tend to make significant life decisions such as moving for a new job or starting a family. Such decisions will often require the customers to apply for a loan, representing a natural situation for the borrower to compare their bank's offer with the competition. The combined effects result in the following hypothesis:

$$\mathrm{H_1}: \text{Younger borrowers will have lower expected customer duration}$$

Regarding the household size, it is believed that larger households will have a longer relationship with the bank. Larger households often indicate that the customer has more dependents that rely on the customer's financial solidity to get by. This could limit the time and energy they have available to spend on personal finances. As previous literature (Ongena *et al*., 2021; Lukas & Nöth, 2019) has highlighted, time and effort are one of the primary explanations as to why households choose not to search for refinancing, despite the considerable savings potential. Another potential explanation is that larger households are often associated with more risk averse borrowers due to their responsibility to provide for the family. A study conducted by Cox *et al.* (2015) demonstrated that households regard refinancing as a risky decision, and risk averse individuals are therefore less likely to search for external refinancing. This results in the following hypothesis:

$$\mathrm{H_1}: \text{Larger households will have longer customer duration compared to smaller households}$$

5.3 Loan Conditions

Larger customers are more profitable for banks than smaller ones (Stern, 2012). After all, the total size of the obligation combined with interest rate reflects the revenue from the mortgage. Large loans also reduce the proportional costs associated with customer interactions and generate opportunities for cross-sales through additional products and services. For these reasons, customers with large loans are very attractive to banks, giving them more incentives to retain them. If the banks succeed in appeasing these high-income borrowers, they would be expected to have longer customer relationships than smaller borrowers.

Previous literature, such as Lukas & Nöth (2019) and Andersen *et al.* (2018), have found social status and personal income significantly affect how active borrowers are in searching for alternative financing. The studies found that borrowers with higher personal income are also more active in searching for improved terms through refinancing.

Because the interest rate offered at initiation is partially the product of the customer's creditworthiness, it is expected that high initial interest rates will affect customer duration similarly to low creditworthiness. However, there are several contradicting arguments against this point of view. First of all, the savings potential of external refinancing could attract borrowers more prone to search for alternative financing (Ongena *et al.*, 2021). Secondly, the interest rate at initiation serves as a benchmark that borrowers use to compare their current interest levels. This anchoring effect (Lukas & Nöth, 2019) has shown that higher interest rates today than the offered rate at initiation often increase a customer's tendency to search for external refinancing. Finally, at the initiation of a mortgage loan, the customer will likely be very attentive to their offered interest rate and more prone to discuss the cost of borrowing with acquaintances. Thus increasing comparability and reducing the switching barriers.

The sum of potential influences loan conditions may have on the longevity of a customer relationship has led to the following hypothesis:

H_1: $Generous\ loan\ conditions\ is\ expected\ to\ increase\ the\ duration\ of\ the\ customer\ relationship$

6. Methodology

6.1 Modeling Customer Duration

The regression analysis aims to answer our research question: ***Can initial information about the customer give reliable signals regarding the bank's potential customer profitability?*** It was decided that a multiple regression with cross-sectional data, which specializes in gathering data for a specific point in time, was the most suitable method to answer our research question. The quick and easy benefit of using a cross-sectional framework enabled us to conduct multiple separate tests to identify and understand the variables' predictive properties and interactions.

A Cox regression (survival analysis) was considered for this task, to analyze the potential duration of a customer relationship. However, a Cox regression requires a particular event of "survival", i.e., the end of the relationship, deemed 1 if the event occurs and 0 otherwise. All observations within our dataset did end their relationship with the bank, thus our dataset is unable to meet the requirements of the Cox regression model. Furthermore, a Cox regression model requires the data to be censored. However, the limitations of the data did not allow for such censoring to take place.

The main objective is to analyze the customers' duration, determined by several explanatory factors. The duration variable (`CustomerDuration`) is the proxy for the regression. The impact of `CustomerDuration` is highly affected by the fact that the variable contains both an entry date and an exit date. Since this research relies on an exit date to measure the customer's total duration, excluding all current bank customers that would have impacted the results is necessary.

As shown in *table 2*, *3* and *4*, six different variations of a linear cross-sectional regression were created per dataset. The aim was to analyze the different effects of each variable on the duration. The process started therefore with a baseline regression. More variables and extensions were gradually added to the model to capture each variable's effect on the overall regression and the relationship between the independent and dependent variables. This process was repeated on all three datasets.

As for the independent variables, the customer's creditworthiness is accounted for by the probability of default (`PD`), debt-to-income (`DTI`) and loan-to-value ratio (`LTV`). DTI has been extended by the variable DTI squared (DTI2 in *table 1*) to check for any nonlinearity between customers' relative debt burden and the longevity of their relationship with the bank. Due to new regulations, we created a dummy variable of DTI (DTIAbove5 in *table 1)* to account for any significant differences between borrowers with DTI above 5 and those below 5. The dummy is 1 for customers with DTI above 5, and 0 for DTI below 5. The relationship between duration and customer's creditworthiness was further examined by adding the interaction variable `PD*LTV`. The aim of this variable is to highlight how changes in one of the variables may change the impact of the other.

Meanwhile, the customer's demographic characteristics consist of the variables `AGE` and `NrInHousehold`. It was decided to examine the relationship between customer duration and age through the additional variable age squared (`AGE2` in *table 1).* However, the regression analysis showed that the continuous variable `AGE` was a bad fit for the model. Instead, two dummy-variables were created to better capture the effect of the customer demographic *(*see *figure 3* in appendix*)*. By dividing age into segments (`Youngsters` and `Retirees` in *table 1)*, the model will provide information regarding significant differences in expected customer duration between different age groups. `Youngsters` are assigned the value 1 for customers below the age of 35 (0 otherwise) and `Retirees` is assigned the value 1 for customers above the age of 66 (0 otherwise). Customers aged between 35 and 66 constitute the model's baseline.

Finally, the customer's loan condition is made up of the interest offered by the bank at initiation and the total size of the customer's debt (`TotalObligation`). The variable for interest rate offered at initiation was constructed by calculating the difference between the rate paid by a customer during their first three months at the bank and the bank's average rate during the same time span. However, it became immediately apparent that the variable for the interest rate at initiation would struggle with heteroscedasticity. Therefore, the interest rate variable was standardized in order to reduce this problem. This transformation resulted in the variable `ScaledRentDiff`. Further additions to the regression were made to improve the model's interpretability by only including two dummy variables of `ScaledRentDiff`, namely `Highest20%` and `Bottom20%`. `Highest20%` has been given the value 1 for the 20% of borrowers offered the highest interest rate compared to the bank's average during the same period; otherwise, the variable's value is 0. The opposite is true for `Bottom20`%. Most initial interest rates offered by banks are closely tied to the a, thereby limiting the spread between most customers. By transforming the initial interest rate into two dummy variables, the model disregards incremental differences in the initial interest rate in favor of the major interest spreads that may impact the customer relationship.

The resulting cross-sectional expression for calculating the duration of the bank's customer relationship can therefore be given by the following equation:

$$
\begin{aligned}
&CustomerDuration \\
&\quad = \beta_1 * Highest20\%_i + \beta_2 * Bottom20\%_i + \beta_3 * TotalObligation_i + \beta_4 \\
&\quad * LTV_i + \beta_5 * PD_i + \beta_6 * NrInHousehold_i + \beta_7 * DTIAbove5_i + \beta_8 \\
&\quad * Youngsters_i + \beta_9 * Retirees_i + \beta_{10} * PD_i * LTV_i + \epsilon_i
\end{aligned}
$$

Where i represents a given customer

6.2 Validity

As a robustness check for the model, six different regressions were created (as seen in *table 2, 3* and *4* below). This was done to check for any positive or negative changes in the coefficients, and if one variable changed the outcome of others. As seen from the tables, all coefficients are consistent across the regressions, indicating that the regression is robust in terms of the variables.

Furthermore, we conducted a series of post-estimation regression tests to check for heteroscedasticity, multicollinearity, and normality (See *Figure 4* in appendix). The tests include Breusch-Pagan test for heteroscedasticity, Skewness and Kurtosis test for normality and a VIF test for multicollinearity. These tests make it possible to detect if the prerequisites for using OLS estimation are violated. The tests indicated that the dataset is subject to both heteroscedasticity and non-normality, but not subject to multicollinearity (see appendix: *table 7* and *table 8*).

To remove the heteroscedasticity problem, it was decided to conduct a robust regression. The drawback of this method is an increase in standard error amongst the regression coefficients, this has the potential of turning some variables from statistically significant to not statistically significant. However, this was not the case for this model, as the variable had the same significance after the robust regression. The benefit is that the estimated coefficients remain the same and the estimation method is robust to outliers.

Lastly, we conducted a Variance Inflation Factor (VIF) test on the predictors to check for possible multicollinearity. As a rule of thumb, a VIF of more than 5 suggests some problems with multicollinearity, however this did not seem to be an issue for the regression model (*Table 8* in appendix). A high VIF score for interaction-predictors are to be expected.

7. Results

This chapter is divided into two parts. Chapter 7.1 regards the results of the analysis based on the complete dataset, while chapter 7.2 discusses how the results differ between the two time-separated datasets and the complete dataset.

7.1. Main Results

As seen in *table 2*, *3* and *4* below, the model result with the highest explanatory power belongs to the simpler regressions, where the highest R^2 is 0,2595. However, the results are promising as the explanatory power decreases as more variables are introduced. The most complex regression (4b) has an R^2 of 0,2391 as seen from *table 2*. The aim of this research paper is not to find the model with the highest explanatory power but to find which factor has a significant impact on customer duration and how these factors impact customer duration. Therefore, the most important results are the variable's significance and the direction their impact has on the dependent variable.

From *table 2*, one can observe that the coefficients are steady across the regressions. The `LTV` has a low coefficient but is measured in integers, hence having a high overall effect on the duration. `LTV` shows a significant negative relationship at a 1-percent confidence level across the regressions. The regression coefficient indicates that the higher the level of LTV, the shorter the duration is expected to be. An increase in the LTV ratio is expected to reduce the expected duration by around 0.08 years. Making a rough prediction, a borrower with an LTV of 75 is expected to have a 4-year longer relationship with the bank compared to a borrower with an LTV of 25, all other variables being equal. According to *table 1* the median customer duration is 4 years, indicating that LTV has the potential to significantly impact the dependent variable.

The `PD` variable is significant across the regression models on a 1-percent confidence level. The results show that the higher the probability of default, the shorter the duration is expected to be. According to the regression coefficient to the full dataset's model 4b, an increase in the probability of default by 1% will reduce the expected duration by two and a half years. Furthermore, the interaction variable `PD*LTV` was added to the later regression models.

However, the variable is non-significant for regression 3b and 4b according to *table 2*. The results also contradict with the standalone PD and LTV variables, as the interaction variable shows that the duration increases as their values increase.

Results of the `DTI` variable shows that its relationship with the dependent variable is significant at a 1-percent confidence level. Also, it demonstrates that `DTI` has a negative impact on the expected duration. `DTI2` has similar results to that of the `DTI` variable, however the significant confidence level varies between 5-percent and 1-percent. Interestingly, the positive coefficient to `DTI2` suggests that there is nonlinearity within the relationship between DTI and customer duration. According to the model the effect of an increase in DTI will shorten the customer relationship until the DTI ratio reaches 10.5 (calculated by taking the derivatives of the variables from *table 2,* and setting the sum equal to zero), afterwards DTI will have a prolonging effect on the customer relationship. For `DTIAbove5`, the results show that customers who have a ratio above 5 will indeed have a shorter expected duration. The significant levels reflect those of the original `DTI`. According to the regression coefficient in *table 2*'s regression 4b, borrowers with a DTI ratio above 5 are expected to have on average 0.69 years shorter duration compared to borrowers with a DTI ratio below 5.

`Youngsters` and `Retirees` attempt to capture the effect of age on the duration more correctly. The results were not surprising as they complement the existing theory (McDade, 2016). The results from the full table show that the younger customers' duration is on average 1.34 years shorter than the base age, and the older customers' duration is on average 2.06 years longer than the base age. The variables are significant at a 1-percent confidence level for every regression model. For `NrInHousehold`, the size of the household has a positive impact on the duration. As the size of the household increases, the expected duration increases. The variable is significant at a 1-percent confidence level. Compared to the other variables, `NrInHousehold` has a smaller effect on the overall duration as the maximum observed size is 8. According to Regression 4b in *table 2,* one additional household member is expected to increase the borrower's stay with the bank by 0.33 years.

The results of the analysis found that a customer's total obligation has a significant effect on the duration for the full table. The relationship between customer duration and total obligation is positive, meaning the duration is expected to increase as the total obligation increases. However, the significant confidence level varies between 1-percent and 5-percent across the regressions. Finally, for `Highest20%` and `Bottom20%` the results show that the variables are significant at a 1-percent confidence level. The estimated regression coefficients show that customers who receive the lowest interest rates at initiation have a shorter duration compared to the majority base. Those who receive the highest interest rates have a longer duration compared to the majority base. *Table 2's* regression 4b predicts the top 20% to have on average 0.77 years longer duration than baseline, while the bottom 20% is expected to have on average 0.70 years shorter duration compared to baseline.

Table 2. Regression results for the full set of observations

	Baseline	Baseline w/best age	Regression 3a	Regression 4a	Regression 3b	Regression 4b
Creditworthiness						
LTV	-0,0422 *** (0,0061)	-0,0555 *** (0,0062)	-0,0603 *** (0,0066)	-0,0666 *** (0,0066)	-0,0692 *** (0,0098)	-0,0761 *** (0,0098)
PD	-1,2767 *** (0.1891)	-1,5124 *** (0,2058)	-1,5657 *** (0,2036)	-1,6315 *** (0,2061)	-2,3636 *** (0,7745)	-2,4994 *** (0,8092)
DTI	-0,5025 *** (0,1343)	-0,5330 *** (0,1375)	-0,4677 *** (0,1375)	N/A	-0,4485 *** (0,1396)	N/A
DTI2	0,0240 *** (0,0082)	0,0237 *** (0,0084)	0,0194 ** (0,0084)	N/A	0,0185 ** (0,0084)	N/A
DTIAbove5	N/A	N/A	N/A	-0,7123 ** (0,3491)	N/A	-0,6856 * (0,3509)
Characteristics						
NrInHousehold	0.2875 *** (0.0802)	0,2184 *** (0,0822)	0,2583 *** (0,0824)	0,3301 *** (0,0785)	0,2604 *** (0,0823)	0,3294 *** (0,0782)
AGE	0.0765 (0,0502)	N/A	N/A	N/A	N/A	N/A
AGE2	0.0003 (0,0005)	N/A	N/A	N/A	N/A	N/A
Youngsters	N/A	-1,3395*** (0,1986)	-1,3238 *** (0,1985)	-1,2897 *** (0,1999)	-1,3717 *** (0,2032)	-1,3439 *** (0,2064)
Retirees	N/A	2,0582*** (0,3822)	1,9816 *** (0,3806)	1,9465 *** (0,3805)	1,9347 *** (0,3830)	1,8967 *** (0,3827)
Loan Conditions						
Highest_20%	N/A	N/A	0.7735*** 0,2573	0,8487 *** (0,2584)	0,7008 *** (0,2676)	0,7658 *** (0,2708)
Bottom_20%	N/A	N/A	-0,6863*** (0,2342)	-0,7144 *** (0,2354)	-0,6719 *** (0,2344)	-0,6975 *** (0,2357)
TotalObligation	2,01e^(-7) *** (6,40e^(-8))	2,19e^(-7) *** (6,58^(-8))	2,36 e^(-7) *** (6,61e^(-8))	1,41e^(-7) ** (5,70e^(-8))	2,35e^(-7) *** (6,65e^(-8))	1,44e^(-7) ** (5,7e^(-8))
Interaction Variable						
PD*LTV	N/A	N/A	N/A	N/A	0,0133 (0,0114)	0,0145 (0,0120)
Constant	3,5971 (0,7483)	10,2104 (0,4948)	10,2198 (0,4901)	9,5429 (0,4324)	10,6564 (0,6283)	10,0490 (0,6175)
R^2	**0,2595**	**0,2342**	**0,2408**	**0,2373**	**0,2423**	**0,2391**
Observation N	2559	2559	2559	2559	2559	2559

Note:
** - significant at 10-percent confidence level*
*** - significant at 5-percent confidence level*
****- significant at 1-percent confidence level*
Red color – Non-significant
( )- the standard error of the coefficient
N/A – Not-applicable

7.2. Robustness Checks

Table 3 and *table 4* reflect the regression results of the observations before and after 2017. The major take-away from the robustness test is that the signs stay consistent across the regressions and the different datasets. However, the regressions show some differences in the significance levels between the full dataset and the two subsets. The most profound difference regards the DTI ratio, where the `DTI` variable is no longer significant in *table 3*. A change in the relationship between `DTI` and the dependent variable was expected since the two subsets were divided based on the implementation date of the new DTI regulations. This is also reflected in the results for `DTI2` and `DTIAbove5`. As for *table 4*, the significance varies between 10-percent and 1-percent confidence level across the regression models.

Regarding `PD` and `LTV`, the significance levels are similar across the tables, however the interaction variable `PD*LTV` turns significant at a 5-percent confidence level for *table 3*. Just like `DTI`, the variable `Youngsters` is statistically significant for both *table 2* and *table 4*, but not significant for *table 3*. Whereas the opposite is true for the variable `NrInHousehold`, here we find the variable to be non-significant for *table 4* but statistically significant according to *table 2* and *table 3*.

`TotalObligation` sees a reduction in significance between the observations before and after 2017. The variable changes from weak significant levels across the regression models in *table 3*, to a more profound significance at a 5-percent and 1-percent confidence level across *table 4*. Finally, the significance of `Bottom20%` and `Highest20%` differ between the tables. `Bottom20%` is non-significant in *table 4* but significant in *table 3*. However, `Highest20%` is only significant at a 10-percent confidence level for the first two regressions in *table 3*, but significant at a 1-percent level for all regressions in *table 4*.

Table 3. Regression result from the observation pre 2017

	Baseline	Baseline w/best age	Regression 3a	Regression 4a	Regression 3b	Regression 4b
Creditworthiness						
LTV	-0,0360*** (0,0098)	-0,0472 *** (0,0105)	-0,0503 *** (0,0109)	-0,0554 *** (0,0106)	-0,0665 *** (0,0144)	-0,0715 *** (0,1390)
PD	-2.2879*** (0.5590)	-2,5824 *** (0,5736)	-2,6302 *** (0,5701)	-2,6752 *** (0,5656)	-4,4426 *** (1,0727)	-4,5492 *** (1,0570)
DTI	-0,4429 * (0,2483)	-0,3765 (0,2542)	-0,3166 (0,2538)	N/A	-0,2725 (0,2560)	N/A
DTI2	0,0351 * (0,0182)	0,0282 (0,0189)	0,0221 (0,0189)	N/A	0,0182 (0,0190)	N/A
DTIAbove5	N/A	N/A	N/A	-0,4149 (0,5421)	N/A	-0,4193 (0,5403)
Characteristics						
NrInHousehold	0.3230*** (0.1151)	0,3132 *** (0,1164)	0,3470 *** (0,1170)	0,3647 *** (0,1154)	0,3476 *** (0,1164)	0,3623 *** (0,1144)
AGE	0.0898 (0,0793)	N/A	N/A	N/A	N/A	N/A
AGE2	0.0001 (0,0007)	N/A	N/A	N/A	N/A	N/A
Youngsters	N/A	-0,1695 (0,4503)	-0,3121 (0,4475)	-0,1779 (0,4354)	-0,4987 (0,4221)	-0,3922 (0,4104)
Retirees	N/A	1.3112 *** (0,6847)	1,2735 *** (0,4573)	1,2647 *** (0,4547)	1,2430 *** (0,4578)	1,2321 *** (0,4552)
Loan Conditions						
Highest_20%	N/A	N/A	0,7003 * (0,4110)	0,7751 * (0,4127)	0,4756 (0,4280)	0,5301 (0,4318)
Bottom_20%	N/A	N/A	-0.8156 ** (0,3190)	-0,8393 *** (0,3206)	-0,8039 ** (0,3194)	-0,8263 *** (0,3210)
TotalObligation	1,43e^(-7) (9,42e^(-8))	1,52e^(-7) (9,76e^(-8))	1,86e^(-7) * (9,74e^(-8))	1,71e^(-7) ** (7,92e^(-8))	1,86e^(-7) * (9,76e^(-8))	1,73e^(-7) ** (7,9e^(-8))
Interaction Variable						
PD*LTV	N/A	N/A	N/A	N/A	0,0297 ** (0,0144)	0,0308 ** (0,0143)
Constant	5,3030 (2,0821)	10,6153 (0,6847)	10,6267 (0,6810)	10,2753 (0,5998)	11,4446 (0,8451)	11,1752 (0,7977)
R^2	**0,2420**	**0,2242**	**0,2297**	**0,2290**	**0,2331**	**0,2326**
Observation N	1343	1343	1343	1343	1343	1343

Note:
** - significant at 10-percent confidence level*
*** - significant at 5-percent confidence level*
****- significant at 1-percent confidence level*
Red color – Non-significant
( )- the standard error of the coefficient
N/A – Not-applicape

Table 4. Regression results from the observation post 2017

	Baseline	Baseline w/best age	Regression 3a	Regression 4a	Regression 3b	Regression 4b
Creditworthiness						
LTV	-0,0456 *** (0,0082)	-0,0586 *** (0,0083)	-0,0690 *** (0,0090)	-0,0749 *** (0,0091)	-0,0769 *** (0,0122)	-0,0834 *** (0,0124)
PD	-0,8003 ** (0.1649)	-0,9576 ** (0,1762)	-1,0421 *** (0,1736)	-1,0979 *** (0,1764)	-1,6251 *** (0,6178)	-1,7403 *** (0,6628)
DTI	-0,5332 *** (0,1665)	-0,6001 *** (0,1684)	-0,4648 *** (0,1692)	N/A	-0,4463 *** (0,1700)	N/A
DTI2	0,0213 ** (0,0093)	0,0229 ** (0,0094)	0,0164 * (0,0096)	N/A	0,0156 * (0,0095)	N/A
DTIAbove5	N/A	N/A	N/A	-0,8770 ** (0,4169)	N/A	-0,8394 ** (0,4188)
Characteristics						
NrInHousehold	0.1182 (0.1056)	-0,0302 (0,1078)	0,0210 (0,1078)	0,1127 (0,1004)	0,0250 (0,1079)	0,1136 (0,1002)
AGE	-0,0058 (0,0752)	N/A	N/A	N/A	N/A	N/A
AGE2	0.0011 (0,0008)	N/A	N/A	N/A	N/A	N/A
Youngsters	N/A	-1,2968 *** (0,2178)	-1,1444 *** (0,2210)	-1,1273 *** (0,2237)	-1,1589 *** (0,2189)	-1,1447 *** (0,2219)
Retirees	N/A	2,7722 *** (0,7174)	2,4627 *** (0,7153)	2,3868 *** (0,7168)	2,3878 *** (0,7203)	2,3068 *** (0,7214)
Loan Conditions						
Highest_20%	N/A	N/A	1,4837 *** (0,3302)	1,5673 *** (0,3304)	1,4496 *** (0,3330)	1,5262 *** (0,3342)
Bottom_20%	N/A	N/A	-0,3151 (0,3281)	-0,3424 (0,3278)	-0,3020 (0,3281)	-0,3265 (0,3278)
TotalObligation	2,63e^(-7) *** (8,82e^(-8))	2,96e^(-7) *** (8,94e^(-8))	2,9e^(-7) *** (9,01e^(-8))	1,91e^(-7) ** (7,53e^(-8))	2,88e^(-7) *** (9,05e^(-8))	1,93e^(-7) ** (7,55e^(-8))
Interaction Variable						
PD*LTV	N/A	N/A	N/A	N/A	0,0098 (0,0091)	0,0108 (0,0098)
Constant	5,9056 (1,7965)	9,7129 (0,7523)	9,5730 (0,7377)	8,7696 (0,6366)	9,9495 (0,8670)	9,2197 (0,8100)
R^2	0,2582	0,2389	0,2567	0,2524	0,2581	0,2541
Observation N	1216	1216	1216	1216	1216	1216

Note:
** - significant at 10-percent confidence level*
*** - significant at 5-percent confidence level*
****- significant at 1-percent confidence level*
Red color – Non-significant
( )- the standard error of the coefficient
N/A – Not-applicable

8. Discussion

8.1. Creditworthiness

Regarding creditworthiness, we find support in our hypothesis that the customers expected duration with the bank is reduced with poor ratings. As the variable is a measure of risk, the results of the regression models indicate that the higher level of risk the bank accepts the shorter the expected duration. Because the Ministry of Finance addressed the issue of increased debt in Norwegian households (Regjeringen, 2016) and because the banks are operating with strict policies following the regulations, a customer with high-risk metrics will struggle to obtain refinancing at the bank. This struggle forces the customer to search for refinancing at competing banks, thus ending the relationship sooner.

The analysis found a significant negative relationship between the LTV and customer duration. A significant impact on the duration implies that the banks assess the LTV as a ratio connected to customer defaults. However, the interaction between PD and LTV is non-significant, suggesting that there are other explanations for the reduced duration. A possible explanation could be that the banks are actively trying to remove such customers, possibly due to the desire to maintain regulated policies on acceptable LTV and DTI ratios in their customer portfolio. At the same time, an external bank could be more inclined to deviate from the flexible quota (Regjeringen, 2015). This opens the possibility that competing banks attract customers with higher DTI and LTV ratios, thus decreasing the expected duration. An interesting takeaway from this is the opportunity for further research. Are banks who are more willing to offer refinancing to borrowers with higher LTV and DTI more likely to obtain borrowers who are more prone to switching banks and, by extension, less profitable?

Furthermore, the regression indicates a significant negative relationship between a customer's PD and their expected customer duration with the bank. The most obvious explanation for this is that defaults have a negative effect on the relationship. Thus, higher PD will also yield shorter expected customer relationships. Similar to LTV, banks could be actively trying to weed out risky customers. This could be because customers with a high PD value may yield a low expected return for the bank, or it could be a necessary action the bank takes to uphold the risk profile of the customer portfolio. Banks usually make weighted judgments on what types of customers they decide to initiate a relationship with and will demand additional risk compensation from customers with a higher probability of default. However, such an analysis is outside the scope of this paper.

From the regression, observe that higher DTI ratios will have a negative impact on the length of a customer relationship. However, the analysis shows a nonlinear relationship between customer duration and the customer's DTI level. The direction of the DTI and DTI squared coefficients contradicts the initial hypothesis that the combined effect should be exponential, not curved. A possible explanation for this is that DTI levels above 5 are significant, indicating that the mandated restriction on acceptable DTI is the main reason for the shorter customer relationships and not due to a high level of DTI.

It seems likely that the reason for LTV having a negative impact on customer duration are also the same reason for DTI having a negative impact on customer duration. Both variables have been the target of government regulations, and the regulations have in both instances been aimed at limiting the bank's opportunity to provide risky loans. Interestingly, the DTI ratio is non-significant in the observations before 2017. However, after the regulations were implemented, DTI has become a significant explanatory variable, indicating that the relationship between a bank and their customers has changed due to the new regulations. From *table 3* one can observe that the PD coefficient is more than double from *table 4*, which indicates that banks were more prone to using the PD as a risk metric prior to 2017. However, the significance of DTI in *table 4* indicates that DTI has become a more useful tool in assessing the expected duration.

Finally, during this regression analysis, all possible interactions between the predictors were tested, and the result yielded no significant interactions amongst any predictors belonging to the original regression model, except the interaction PD*LTV. These results indicate a significant interaction between PD and LTV in *table 3*, however in *tables 2* and *4*, the interaction is non-significant. The results indicate that the interaction between PD and LTV is predicted to have a positive impact on customer duration. Meaning that an increase in LTV will have an increasing impact on the customer duration for borrowers with a high probability of default and vice versa. However, this interaction is only moderately significant for observations before 2017 and yields a very low degree of explanatory power to our model. The variables' insignificance might be correlated with the relationship mentioned above between DTI and PD.

8.2. Demographic Characteristics

The results and analysis of a customer's demographic characteristics support our hypothesis. In terms of the household size, the results show that borrowers with larger families, i.e., more individuals in the household, have longer relationships with the bank. The household size has a similar expected duration to that of firms, where Ongena & Smith (2001) found that larger firms have longer relationships with the bank than smaller firms. Because switching banks require additional time and effort, bigger families are less prone to take the time to fulfill the process (Lukas & Nöth, 2019). The importance of other duties and priorities make up the time available to do the process. Another possible explanation might be that borrowers in larger families have more dependents relying on their financial stability, and therefore appreciate predictability and stability over the risk of switching banks.

From the results, borrowers younger than the age of 35 have a statistically shorter expected duration than customers aged between 36 and 66, and individuals older than the age of 66 have a statistically longer expected duration compared to individuals aged between 36 and 66. There are several potential reasons for this. Firstly, the digitalization of financial services has increased dramatically in the last five years. These technological innovations are usually adopted by younger generations before they are adopted by the older generation (McDade, 2016). As discussed previously, these new technologies make transaction costs of switching banks lower and increase transparency and comparability amongst different financial products. The result is reduced switching costs, which the younger generations take advantage of, thus making them more prone to seek refinancing. Older generations are typically the last demographic to adopt new digital innovations.

Furthermore, older generations are usually well-established and not subject to as many major life decisions as younger generations. At the same time, they enjoy a lifetime's worth of built-up equity, limiting the impact of government regulations on LTV and DTI. As a result, older generations are more inclined to seek advisory services and appreciate good customer service and long-lasting relationships (McDade, 2016). On the other hand, younger generations are less likely to stay within the regulatory limits and must therefore seek refinancing from banks that are willing to deviate from the flexible quota. Also, the younger generation is more transaction-oriented and looking for low-cost borrowings. The combined generational effect reflects the outcome of the results.

8.3. Loan Conditions

Regarding loan conditions, the results seem to have conflicting support for our initial hypothesis. As hypothesized, the total obligation has a significant positive relationship with the expected duration of the customer. Because the total obligation is a direct consequence of other calculated metrics, such as DTI and LTV, customers with higher total obligations must also have top-tier salaries and wealth. In other words, these customers are not considered very risky by the bank. Therefore, the banks could be keener on retaining these customers, offering improved conditions and additional products and services to boost retention rates and profits. On the other hand, it is expected that these customers are well economically educated and therefore are expected to be more active in searching for better terms and conditions. The results do not seem to support this. The more likely case is that these customers need clear finances, keeping their obligations with one steady financial lender, hence overshadowing the need for refinancing.

Regarding interest rates, the analysis shows that the 20% of borrowers offered the highest interest rates compared to the bank's average are expected to have longer customer relationships than the remaining 80% of customers. Conversely, the opposite tendency is found amongst borrowers offered the lowest interest rates compared to the company's average. However, the result is not significant for the 20% of bottom interest-paying borrowers after 2017.

The customers offered the highest interest rates are likely to experience fewer increases in their current interest rates, as they are already amongst the top 20% compared to the average. The anchoring effect would suggest that borrowers who receive a low initial interest rate would be more prone to search for refinancing. Once you already are amongst the 20% of borrowers paying the least interest rate, it will most likely develop in an increasing direction. Simultaneously, individuals who are offered the lowest initial interest rate could also represent the most price-sensitive customers, i.e., those who are the most prone to search for the cheapest source of credit. This would explain why lenders who are offered the lowest initial interest rate also have a tendency towards shorter customer relationships. This has some important implications concerning the marketing strategies banks deploy today. By offering the most beneficial conditions to new customers, banks could indirectly be attracting the least desirable borrowers.

Whether the amount of bank switches is significantly different before and after 2017 is outside the scope of this research paper. Although, if the anchor effect from the initially offered interest rate can explain the duration of the customer relationship, then it would be expected to see significant changes in the models pre-and post-2017. However, this is not the case. There are some variables that change from significant to not significant, but there are no major changes in the regression coefficient, nor in the explanatory power of the model. The results seem to suggest that the interest rate development has no significant impact on the factors affecting the duration of customer relationships.

9. Conclusion

We have studied the impact informational signals have on future profitability and longevity in bank-customer relationships. To answer our research question, we constructed and tested four hypotheses based on three key aspects of a bank-customer relationship: the customer's creditworthiness, their demographic characteristics, and the conditions of their loan engagements. The duration of a customer relationship was used as a proxy for the banks' profitability. By dividing our data into three, we have accounted for recently issued government regulations and enhanced our research paper by incorporating the effects government intervention has had on customer relationships in the mortgage market. After the regulations, variables such as debt-to-income has become more prominent factors in determining the duration of a customer relationship. We can also see a nonlinear divide in expected duration between the borrowers above and the borrowers below the government-mandated limits.

Through a robust multiple regression analysis with a cross-sectional setup, we confirmed that customers with poor creditworthiness are expected to have a shorter relationship with the bank and customers with larger households and older generations are expected to have longer relationships. The variables describing a customer's loan conditions do not seem to impact the customer duration in the same direction. A customer with a large amount of loan is expected to have a longer customer relationship compared with a customer with a smaller loan size. At the same time, customers offered the highest initial interest rate compared to the bank's average is expected to have a longer duration compared to the average borrower.

Our research contributes to the existing literature by uncovering which and how signals available to a bank before a loan agreement impact the upcoming relationship's profitability. We have enriched existing literature by offering a rare perspective from the bank's point of view, based on actual bank-customer data. To the best of our knowledge, no other researchers have been given access to our set of bank-customer data. We recommend implementing other variables for further research, such as customer gender, education, and location. Simultaneously, an interesting scope would be to find if competing banks are more or less willing to offer refinancing to borrowers with higher LTV and DTI and if those customers are less profitable.

10. Appendix:

Figures:

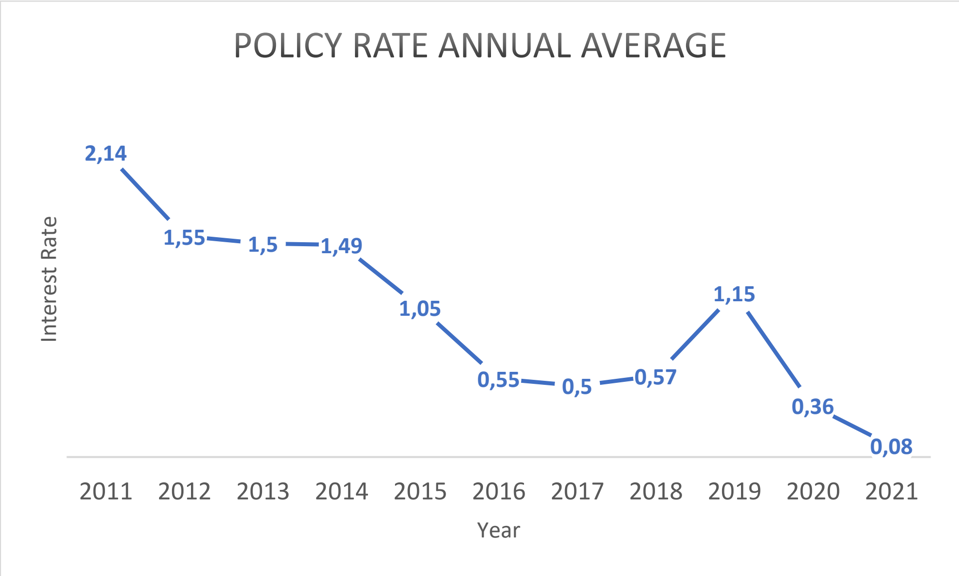


Figure 2. Norwegian key policy rate development spanning from 2011 to 2022. Source: Norges Bank, 2021

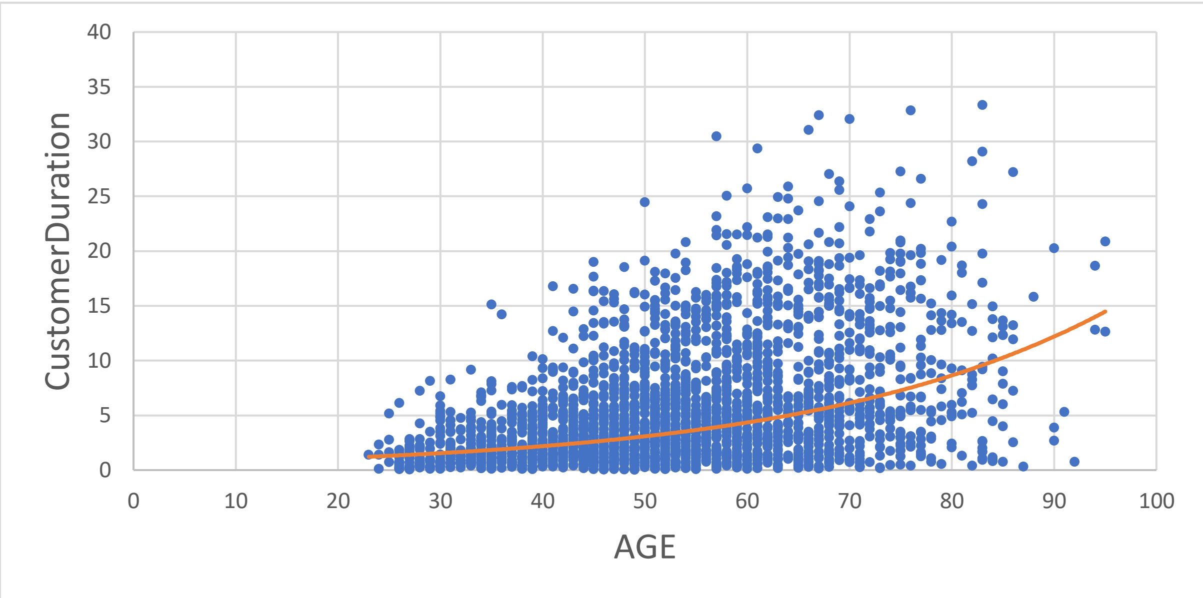


Figure 3. Scatterplot of age and the expected duration

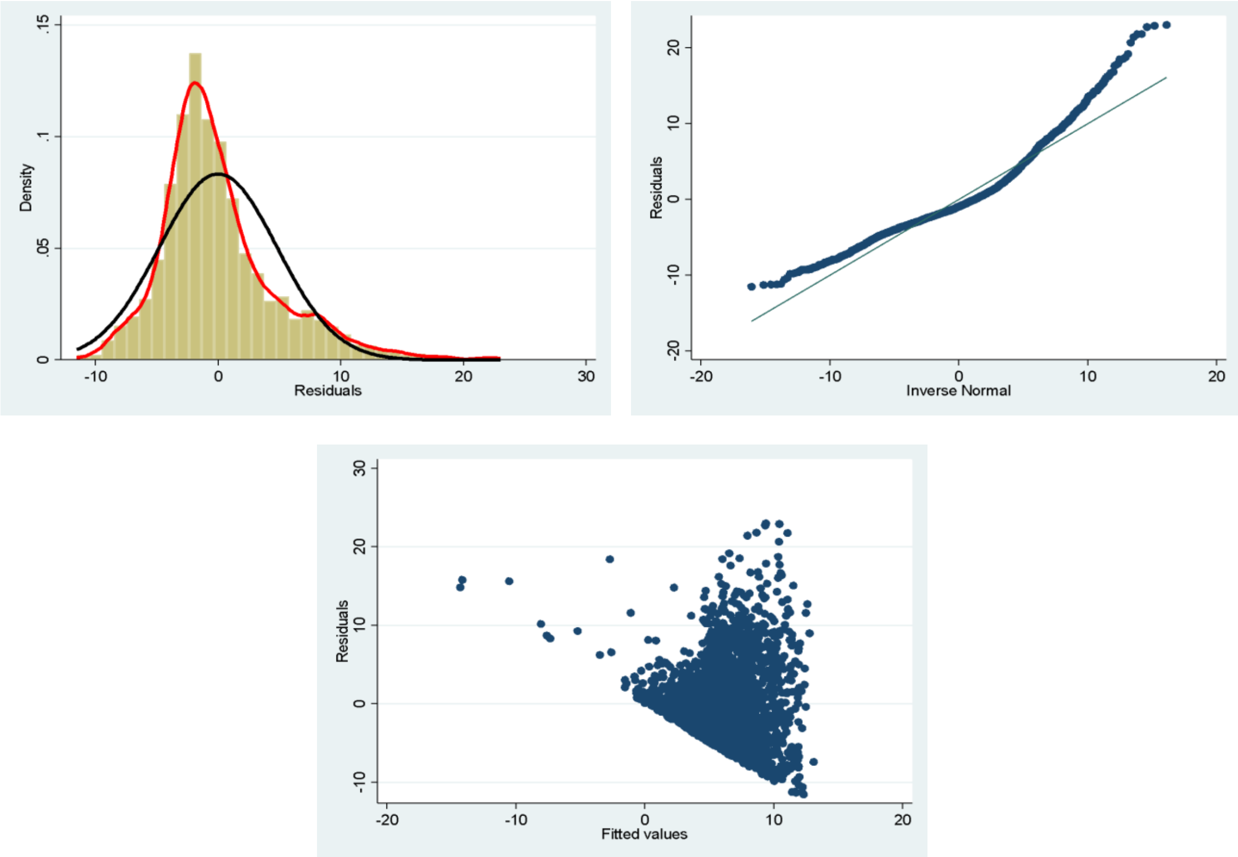


Figure 4. Diagnostics tests for the regression

Tables:

Table 5. Descriptive statistics of the observations pre 2017

	Obs	Min	Max	Mean	Median	Std.Dev
CustomerDuration	1356	0,15027	33,3412	6,8110	5,1566	5,5922
Creditworthiness						
LTV	1356	0	85	54,3783	60	21,1229
PD	1343	0,044	12,358	0,7089	0,476	0,6864
DTI	1356	0	15,246	3,1099	2,8251	1,8378
DTI2	1356	0	232,4406	13,0468	7,9813	19,1328
DTIAbove5	1356	0	1	0,1128	0	0,3165
Characteristics						
NrInHousehold	1356	1	8	2,3178	2	1,2425
AGE	1356	25	95	56,3238	55	12,5922
AGE2	1356	625	9025	3330,812	3025	1467,17
Youngsters	1356	0	1	0,0361	0	0,1867
Retirees	1356	0	1	0,2006	0	0,4006
Loan Conditions						
Highest20%	1356	0	1	0,1253	0	0,3313
Bottom20%	1356	0	1	0,1984	0	0,3989
TotalObligation	1356	80000	2,60e^(7)	2762085	2200000	2142087

Note:
See Table 9 for a short description of the variables

Table 6. Descriptive statistics of the observations post 2017

	Obs	Min	Max	Mean	Median	Std.Dev
CustomerDuration	1223	0,3014	32,8334	4,5513	2,6646	5,1279
Creditworthiness						
LTV	1223	3	104	58,8455	61	20,8076
PD	1216	0,04	11,013	0,9030	0,607	1,0198
DTI	1223	0	21,4285	3,8885	3,6198	2,2835
DTI2	1223	0	459,1837	20,3309	13,1029	31,9362
DTIAbove5	1223	0	1	0,1709	0	0,3766
Characteristics						
NrInHousehold	1223	1	7	2,0891	2	1,2368
AGE	1223	23	92	49,7948	49	13,2105
AGE2	1223	529	8464	2653,893	2401	1384,802
Youngsters	1223	0	1	0,1406	0	0,3478
Retirees	1223	0	1	0,1022	0	0,3030
Loan Conditions						
Highest20%	1223	0	1	0,2845	0	0,4514
Bottom20%	1223	0	1	0,2061	0	0,4046
TotalObligation	1223	50802,52	2,20e^(7)	3329923	2702360	2406770

Note:

See Table 9 for a short description of the variables

Table 7. Heteroscedasticity and normality tests

	Chi2	Prob > Chi2
Breusch–Pagan / Cook-Weisberg	406,16	0.000
Skewness and kurtosis for normality	397,49	0.000

Table 8: VIF-test for multicollinearity

	VIF	1 / VIF
PD*LTV	14.16	0.070627
PD	11.71	0.085376
LTV	2.20	0.454185
TotalObligation	1.73	0.576890
DTIAbove5	1.51	0.661641
Retirees	1.23	0.813349
Highest20%	1.18	0.848441
NrInHousehold	1.16	0.863250
Youngsters	1.14	0.876514
Bottom20%	1.11	0.899769
Mean VIF	3.71	

Note:
See Table 9 for a short description of the variables

Table 9. Variable descriptions

	Description
CustomerDuration	The duration of the customer
Creditworthiness	
LTV	Loan-to-Value of the customer
PD	Probability of default of the customer
DTI	Debt-to-Income of the customer
DTI2	Debt-to-income squared for nonlinearity
DTIAbove5	Dummy-variable of DTI for values above 5
Characteristics	
NrInHousehold	Number of people in the customer's household
AGE	The age of the customer
AGE2	The age of the customer squared for nonlinearity
Youngsters	Customers below the age of 35
Retirees	Customers above the age of 66
Loan Conditions	
Highest20%	Dummy-variable for the 20% highest interest rates
Bottom20%	Dummy variable for the 20% lowest interest rates
TotalObligation	The total amount of loan of the customer

References:

Akin, J. (2020). "Understanding Loan-to-Value Ratio". Experian. URL: https://www.experian.com/blogs/ask-experian/what-is-loan-to-value-ratio-and-why-is-it-important/ (Accessed: March 2022)

Andersen, S., J. Y. Campbell, K. Meisner-Nielsen, and T. Ramadorai (2018). "Sources of inaction in household finance: evidence from the Danish mortgage market". *Working paper*.

Brunetti, M., R. Ciciretti, and L. Djordjevic (2020). "Till mortgage do us part: Mortgage switching costs and household's bank switching". *Journal of Banking and Finance, 119*, 105904.

Carbo-Valverde, S., T. H. Hannan, and F. Rodriguez-Fernandez Francisco (2011). "Exploiting old customers and attracting new ones: The case of bank deposit pricing". *European Economic Review, 55.7*, 903–915.

Cox, R., D. Brounen, and P. Neuteboom (2015). "Financial Literacy, Risk Aversion and Choice of Mortgage Type by Households". *Journal of Real Estate Finance and Economics*, 50(1), 74–112.

Garland, R. (2001) "What Influences Customer Profitability? Service-Profit Chain", *Journal of Targeting, Measurement and Analysis for Marketing*, 10(3), 233-248.

Kenton, W. (2021). "Default Probability". *Investopedia*. URL: https://www.investopedia.com/terms/d/defaultprobability.asp (Accessed: March 2022)

Lukas, M. and M. Nöth (2019). "Interest rate changes and borrower search behavior". *Journal of Economic Behavior & Organization, 163*, 172–189.

McDade, S. (2016). "How age impact consumer behavior in retail banking". *The Financial Brand*. URL: https://thefinancialbrand.com/61246/age-consumer-behavior-patterns-banking/ (Accessed: April 2022)

Norges Bank (2022, 29. April). *Styringsrenten*. URL: https://www.norges-bank.no/tema/pengepolitikk/Styringsrenten/ (Accessed: April 2022)

Ongena, S. and D. C. Smith (2001). "The duration of bank relationships". *Journal of Financial Economics*, 61.3, 449–475.

Ongena, S., F. Paraschiv and E. J. Reite (2021). "Determinants of price discrimination and switching mortgage provider in times of regulation and digitalization". *Swiss Finance Institute Research Paper*, 21-67.

Papp, D. H. (2007). «Assessing household credit risk: evidence from a household survey".

Magyar Nemzeti Bank (MNB) Occasional papers, 70.

Regjeringen (2015). *Details on the new regulations from 2015*. URL: https://www.regjeringen.no/no/dokumenter/forskrift-om-krav-til-nye-utlan-med-pant-i-bolig/id2417408/

Regjeringen (2016) (2). *Details on the new mortgage regulations from 2016.* URL: https://www.regjeringen.no/no/tema/okonomi-og-budsjett/finansmarkedene/boliglansforskriften-1.-januar-201730.-juni-2018/id2523977/

Regjeringen (2016), *New mortgage regulations*. URL: https://www.regjeringen.no/no/dokumentarkiv/regjeringen-solberg/aktuelt-regjeringen-solberg/fin/pressemeldinger/2016/fastsetter-ny-boliglansforskrift/id2523967/

Reichheld, F. F. (1992) "The truth of customer retention". *Journal of Retail Banking*, 13(4), 21–26.

Sharpe, S. A. (1990). "Asymmetric Information, Bank Lending, and Implicit Contracts: A Stylized Model of Customer Relationships". The Journal of Finance, 45.4, 1069–1087.

Sharpe, S. A. (1997). "The Effect of Consumer Switching Costs on Prices: A Theory and its Application to the Bank Deposit Market". *Review of Industrial Organization, 12*, 79–94.

Shea, N. (2017) "Seven factors that determine your mortgage interest rate". *Consumer Financial Protection Bureau*. URL: https://www.consumerfinance.gov/about-us/blog/7-factors-determine-your-mortgage-interest-rate/ (Accessed: February 2022)

Statista (2021). *Homeownership rate in Norway, from 2011 to 2020*. URL: https://www.statista.com/statistics/543453/house-owners-among-population-norway/ (Accessed: March 2022)

Statista (2021). *House ownership rate in the European Union, from 2010 to 2018*. URL: https://www.statista.com/statistics/503864/house-owners-among-european-union-eu-28-population/ (Accessed: March 2022)

Stern, L. (2012, 02. March) *When your bank doesn't want you.* URL: https://www.reuters.com/article/us-column-personal-finance-idUSTRE82026K20120302 (Accessed: May 2022)

Storbacka, K. (1994) "The nature of customer relationship profitability". *Swedish School of Economics and Business Administration, Research Report 55,* Helsingfors

Thadden, E. L. von (2004). "Asymmetric Information, Bank Lending and Implicit Contracts: The Winner's Curse". *Finance Research Letters 1.1*, 11–23.

Trading Economics (2020). *Statistics on debt-to-income for Nordic countries.*

NTNU

Norwegian University of
Science and Technology