University of Twente

Financial Engineering and Management

Master Thesis

Determinants of curtailments in Dutch residential mortgages

Author

Wytze J.S.F. Driessens

University supervisors

B. Roorda
R.A.M.G. Joosten

Examination date

March 4, 2021

Company supervisors

P.B.N. Hendriks
G. Csapo

Non-Confidential Version

Abstract

Curtailments are payments on a mortgage outside the contractual redemption schedule that are smaller than the outstanding debt of the mortgage. Mortgages are subject to behavioral option risk because curtailments can alter the level and timing of the cash flows of the bank. To provide insight in curtailment behavior of mortgage customers, we study how different determinants affect curtailments of Dutch private residential mortgages.

We studied the impact of determinants on the probability of curtailment using a multivariate logistic regression model. Where previous literature studies different determinants of curtailments, activity-specific determinants were not part of the scope. Activity-specific determinants represent curtailment decisions or actions made by a borrower over the lifetime of his mortgage. With use of a multivariate logistic regression model, we find that activity-specific determinants significantly affect the probability of curtailment compared to other determinants. Borrowers with record of past curtailment, are more likely to curtail again. In addition, borrowers that made partial prepayments in higher frequency are also more likely to curtail again. However, the higher the percentage of the original principal amount that is prepaid within one year, the less likely a borrower is to curtail in the next period.

Furthermore, we used a beta regression model to study the effect of determinants on the conditional curtailment rate, which is the proportion of the outstanding debt that is paid off in case of curtailment. We find that activity-specific determinants significantly decrease the mean curtailment rate. Borrowers that prepaid in the past, prepay a smaller proportion of their outstanding debt. Also, for borrowers prepaying in higher frequency we see that the curtailment rate decreases.

Observing the dynamics of determinants for both the probability and curtailment rate, we deduce three curtailment groups. The first group is mortgages that never curtail, which is the case for the majority of observations. Within the group that does curtail; borrowers that curtail in high frequency curtail in lower rates whereas borrowers that curtail in low frequency curtail in a higher rate.

Not only activity-specific determinants, but also seasonal effects proved to affect curtailments. We confirmed findings from previous literature that the number of curtailments peaks in December. In addition, we found that the curtailment rate peaks in December as well. Also, we find a second peak in January for both the number of curtailments and curtailment rate. Another determinant affecting both the probability of curtailment and curtailment rate is the mortgage costs to principal rate. We find that higher mortgage costs result in a reduced probability of curtailment and also reduce the proportion of outstanding balance that is prepaid in case of curtailment.

We used multivariate logistic model to estimate the number of mortgages that curtail on portfolio level and to classify true curtailments. If we assess the performance from a classification perspective, the logistic model fails to identify high rates of true curtailments. In general, on monthly portfolios the logistic model classifies about 30 to 40 percent of true curtailments. If we assess the performance on estimating the percentage of mortgages that curtail on portfolio level, the average estimation error on yearly basis over out of sample months is 1.34 percent. However, the estimation errors do fluctuate over different out of sample months, where the standard deviation is 13.8 percent.

The beta regression model is used to predict the average conditional curtailment rate. The out of sample estimation error for the average conditional curtailment rate is 5.7 percent. The standard deviation of the out of sample months is 4.7 percent. However, the curtailment rate does not reflect differences in outstanding debts of mortgages but this does impact the estimation errors when estimating cash flows. We find that use of the beta regression may result in high estimation errors on cash flow level when estimating on a portfolio with high variety of outstanding debt.

Furthermore, with combined model use we estimated cash flows of curtailment. We used the logistic model to simulate the number of mortgages that curtail on out of sample data and subsequently estimated the proportion of outstanding debt that is curtailed. We simulated each out of sample month a 1000 times. Reviewing the results, only for two months the true total cash flow lied in within the 95 percent confidence interval. Other months all underestimated the true cash flow, where the estimation error peaked in December. Leaving December out, the average total cash flow was underestimated by 15.65 percent.

We find that borrowers that curtail show differences from borrowers that do not curtail. The beta regression model to assigns higher curtailment rates to mortgages without curtailment history and lower curtailment rates to mortgages with curtailment history. Next to that, the average outstanding debt is higher for curtailment observations over almost all years. Because of these differences and because the beta model that does not incorporate the outstanding debt, the low classification performance from the logistic model increases the estimation errors on cash flow level. We conclude that for the combined use of these models, the classification of true curtailments is not accurate enough to estimate well on cash flow level.

Key words: curtailments, determinants, multivariate logistic regression, beta regression

Preface

At this time of writing, time passed by quickly since I started my quest last September to close the chapter as a student Financial Engineering and Management at the University of Twente. When I started my journey on Industrial Engineering and Management about 5 years ago, I was the least expecting that I would write my thesis for 40 hours per week for 6 months behind the desk right next to my bed in my cosy room in Utrecht. These proved to be even more challenging times, asking for discipline and motivation. In the end, the result is what matters.

Since the first moments I have been around on Campus at the University of Twente, I always had the intention to make a picture standing on the roof of the Ravelijn building holding my certificate after graduation. After I graduated for my bachelor, I was not allowed due to reasons of safety. Right after, I set my mind on doing this photoshoot after I graduated for my masters. But this time, the Covid-virus had other plans. Right after starting my graduation internship at the Rabobank, I wanted to make a photo on the top of the Two Tower building from the Rabobank headquarters in Utrecht. Again, the universe seemed to be conspiring. Next time luck will be on my side, and I hope to celebrate the next milestone in life at good altitudes.

I am happy for having the opportunity getting to know the Rabobank and more specifically the Asset & Liability Modelling team. I want to thank Paul and Gergely, for being such involved along the way in these digital times and for all helpful discussions modelling wise but business wise as well.

Furthermore, I want to thank Berend and Reinoud not only for their support and guidance with their supervision on my master thesis, but for all entertaining and informative lectures during my time at the University of Twente.

At last, I want to thank my parents for always being approachable for unlimited support, guidance and coaching even at times I did not want or needed it. I want to thank my brother, on which I can always count on when it matters the most.

Table of contents

1. Introduction

In this chapter we introduce the research topic, the research goal and formulate the research proposal. Also, we introduce the company where we conduct our research.

1.1 The Rabobank

The Rabobank is the one of the largest Dutch private mortgage providers in the Netherlands. According to *De Nederlandsche Bank (2020)*, the market size of the outstanding debt of Dutch mortgages equalled 533.5 billion euros at the end of 2019. The balance sheet of the Rabobank noted about 188 billion euros for the Dutch mortgage portfolio. The Rabobank had a domestic market share of 21 percent on Dutch residential mortgages in 2019. The Rabobank provides data on Dutch mortgages holding a great diversity of variables.

1.2 Mortgage curtailments

Banks have to comply to the regulations imposed by the Basel Committee standards regarding Interest Rate Risk in the Banking Book (IRRBB). One of the three main sub-types of IRRBB is option risk that banks are exposed to. The Basel Committee defines the following: option risk arises from option derivative positions or from optional elements embedded in a bank assets, liabilities and/or off-balance sheet items, which can alter the level and timing of the cash flows. Option risk can be further characterized into automatic option risk and behavioral option risk. Mortgages are a common product with behavioral optionality subject to prepayment risk. The Basel Committee standards state that banks should understand the nature of prepayment risk and make reasonable and prudent estimates of the expected prepayments.

Dutch mortgages are offered in different product types with different amortization schedules which impacts the speed and timing in which the original principal amount is paid back. Outside this contractual amortization schedule the borrower can make a payment to reduce the outstanding debt, which is called a prepayment. A prepayment can either be a partial prepayment or a full prepayment, where both types are associated with different incentives to prepay.

A full prepayment is a prepayment that equals the full outstanding balance of the mortgage at that moment in time. *Hayre (2003)* suggested that full prepayments are due to a home sale by the borrower, a default by the borrower or the borrower willing to refinance their mortgage contract.

Partial prepayments, in literature often referred to as curtailments (which we will use throughout this research), are mortgage payments in excess of the amount due according to the amortization schedule of the specific mortgage, which do not equal the outstanding balance at that time. Curtailments can be done at any point in the lifetime of the mortgage. In general, mortgages have a contract optionality to curtail a certain percentage of the original principal amount each year penalty free. Rabobank provides basis and plus mortgages, where basis mortgages can curtail 10 percent of the original principal amount per year and plus mortgages 20 percent. One of the main

differences between full prepayments and curtailments is that full prepayments end the mortgage immediately, whereas after a curtailment the mortgage continues.

Both partial prepayments and full prepayments are examples of prepayment risk. Previous research on prepayment risk *(Stam (2015), Isbasoiu (2013), Vasconcelos (2010))* studies the effect of various drivers on all four classes of prepayment risk: move, refinance, default or curtailment. The articles related to prepayment risk often study generic drivers which are used to study all classes of prepayment. Less research can be found on solely on the prepayment class curtailments. Articles that do study solely curtailments often study the impact of different drivers on the occurrence of curtailments. In other words, we do not find many articles that study different determinants on curtailment that study probability and volume at the same time. Therefore, we are interested whether we can gain new insights in determinants affecting curtailment by studying determinants of curtailment for both the probability and volume.

The Rabobank data on Dutch mortgages hold a great diversity of variables on loan level, that provides opportunity to do research on determinants that influence both the occurrence and the volume of curtailments.

1.3 Problem statement

With an outstanding debt on private residential mortgages equaling around 188 billion euros, it is of interest for the Rabobank to quantify short term liquidity cash flows that are expected from curtailments. Curtailments change the expected cash flow schedule the bank planned, where cash that was expected in the future is received now which impacts the funding plan.

Therefore, it is important to understand what affects curtailments. To do this, we do research to determinants of curtailments. A determinant is a factor which decisively affects the nature or outcome of something.

1.4 Research proposal

Our goal is to study determinants that influence curtailments in Dutch private residential mortgages. Mortgage data of Dutch private residential mortgages are available from [start date] to [end date][1] on loan-level. To study the relation between determinants we model both the probability of curtailment as well as volume of curtailments. Next to that, we are interested in how both the determinants and models studied help in estimating the cash flows resulting from curtailment. Therefore, this research aims to answer the following questions:

Main research question:

What are determinants that affect curtailments of Dutch residential private mortgages?

Sub research questions:

(i) What is a suitable technique for modelling the probability and volume of curtailment?

(ii) To what extent do determinants that affect curtailments help to estimate cash flows?

The outline of the thesis is as follows. In the second chapter we are reviewing literature on curtailment. The goal of this chapter is to gain insight on determinants of curtailment that have been studied in other research and find determinants that have not been studied yet. In addition, we review literature on statistical models used to model the probability and volume of curtailment. Also, we analyse the Rabobank data. In addition, we propose an initial selection of determinants to include in this research.

In the third chapter we formally define curtailments and we discuss the selected models in more detail. Chapter Three also describes how the initial selected set of determinants is filtered into the final set of determinants before modelling. The fourth chapter provides information on the available dataset, the transformation of determinants before modelling, the correlation amongst determinants and the definition of all determinants.

In Chapter Five we discuss the outcome of the models. We will evaluate to what extent the determinants impact the probability of curtailment and the volume of curtailment. In addition, we discuss the results of the out of sample cash flow estimation. The last chapter, Chapter Six, concludes on important determinants affecting curtailments. In addition, we reflect on the performance of statistic models for estimating cash flows of curtailment. Lastly, we provide recommendations for further research.

[1] The range of the time period is left out due to confidentiality.

2. Literature review

In the first section we review literature on determinants of curtailment. Next, we review statistical models used to model curtailments. In the fourth section, we discuss how we analyse Rabobank data to make an initial selection of determinants. Next to this, for these determinants we formulate a hypothesis why and how these affect either probability of volume of curtailment.

2.1 Former research on curtailment

Curtailments are mortgage payments in excess of the amount due according to the amortization schedule of the specific mortgage, which do not equal the outstanding balance at that time. If a borrower elects to make a curtailment this reduces the outstanding debt of the mortgage, the life of the mortgage and future interest costs. From the mortgagee's perspective, making a curtailment is a voluntary action and may be done at any point in time in any amount less than his outstanding balance. Therefore, a curtailment can be seen as an expression of the mortgagee's desired debt level of the mortgage.

Although literature on curtailment is limited compared to literature on full prepayment, several articles have written about determinants affecting curtailment.

Fu (1997) finds that when a mortgagee has a record of previous curtailment, this increases the likeliness the mortgagee will make a curtailment in the future. This is backed by *Lin and Yang (2005)* who found that in Asian markets individuals who curtail increase the probability with 23 percent of future prepayment. *Fu (1997)* does not research how previous records of curtailment affect the volume of curtailments. *Fu (1997)* estimates mortgages with past curtailment to be 85 percent less likely to default during the remaining life.

In the context of jumbo loan[2] pools, *Budinger and Fan (1995)* find that curtailments seldom occur in the early life of the mortgage but curtailment rates significantly increase with the aging of the mortgage. In addition, they find that the loan to value (LTV) ratio helps to forecast curtailments. A low LTV ratio means the loan taken out on the mortgage is small compared to the collateral value at the origination of the mortgage. A low LTV ratio can be an indication that the borrower is in possession of more cash at the origination hence can afford a mortgage with low LTV. On the other hand, when the LTV ratio is high, this means that the loan taken out on the mortgage is close to the value of the collateral. This can be an indication that the borrower is not in excess of free capital. Mortgages with high LTV ratios are found to be negatively correlated with curtailment and low LTV ratios positively correlated with curtailment.

[2] A jumbo loan, also called a jumbo mortgage, is an American mortgage that with a type of finance that exceeds the limits set by the Federal Housing Finance Agency. A jumbo loan is not eligible to be purchased, guaranteed or securitized.

Budinger and Fan (1995) observed a trend of seasonality with curtailment being least likely in the fall months, but the frequency of curtailment beginning in December and reaching a peak in April. They observed this trend of seasonality for increased probability of curtailment, but did not study the effect on volume of curtailment. The curtailments peak in April because American citizens receive their tax refund. In addition, the peak in December is likely to be caused by two reasons. The first is that people often receive an additional month of salary in December hence they have more discretionary income. In addition, costs made in December can still be deduced from payroll tax of that fiscal year.

McCollum, Lee & Pace (2015) show that when the spread between the mortgage interest rate and the short term risk free rate increases, the probability of making a curtailment increases as well. This is particularly true for borrowers estimated to have positive equity in their property at the time of observation. A mortgage has positive equity if the amount amortized on the mortgage loan is greater than the change in the house price of the collateral between the origination of the mortgage and that moment in time. McCollum et al. (2015), call this the savings premium which is defined by the mortgage rate minus the short term risk free rate, for which they take the 12 month LIBOR. This saving premium can be used to measure the relative attractiveness of curtailment as an alternative investment opportunity. When this saving premium rate increases, the mortgagee can view a curtailment an investment in reducing the relatively high mortgage debt.

Adelman et al (2010) find that households with higher propensities to save are strongly positively correlated with curtailment, nearly doubling the probability of such payments. The reason that this could be measured in the study of Adelman et al, is because the authors made use of a survey from the Federal Reserve of Consumer Finances. Also *Lin et al. (2005)* linked curtailments to populations with high household saving rates.

Kuang et al. (2019) studied the effect of interest changes on mortgage curtailment for variable interest rate mortgages. They find that on average the cumulative curtailment over four months increase in response to an 1 percentage point increase in interest rate. In addition, Kuang et al. find that the degree of responsiveness is higher to interest rate increases as opposed to interest rate decreases.

One of the things we find worth noticing on *Fu (1997)* his finding on previous record of curtailment to increase the likeliness of curtailing again, is that this actually represents activity of the borrower on curtailment. In contrast, other determinants mentioned in the previous paragraphs are often related to the mortgage conditions, the borrower and the collateral. We will refer to *activity-specific* determinants, when we mention determinants that reflect the timing, size or frequency of curtailments by the borrower throughout the mortgage lifetime. When evaluating the articles from authors abovementioned, activity-specific determinants are not often included.

2.2 Former research on curtailment modelling

In this section we are reviewing statistical techniques that have been used to model curtailment.

Adelman et al. (2010) used a probit model to investigate the effect of determinants on curtailment. The probit model is a form of regression used to model the probability of an event occurring. In this case the event is a curtailment, which is also called the target variable, which can either have the value of 1 in case of curtailment and 0 otherwise. Because a probability is modelled, a link function is needed to make sure the predicted probability of the regression is between 0 and 1. The most common link function used for predicting dichotomous outcomes are the logistic and probit link function. *Adelman et al (2010),* does not implicitly mention why the probit link function is chosen.

Lin and Yang (2005) use a multinomial logistic regression model to estimate the quarterly conditional default and prepayment rates on loan level. This is done by using a multinomial logistic regression that estimates the probability of three different loan outcomes in a given period: the loan defaulted, the loan prepays or the loan is just active. All classified cases per event are aggregated to achieve default and prepayment rates of the studied portfolio. *Lin and Yang (2005)* study the effect of curtailment on default and prepayment probabilities. To achieve this, an additional curtailment term multiplied by the regression coefficient was added to each multinomial logistic regression equation for all three different events. The curtailment term represented the cumulative curtailment occurred up to the given period. Concluding, Lin and Yang (2015) did not model any curtailment metric directly but rather used curtailment data as tool to study the impact on default and prepayment rates.

McCollum et al. (2015) use the multinomial logistic regression model as well because of its ease of use. They argue, that when estimating multiple classes a multinomial logistic regression is the simplest way. The multinomial model predicts the probability of four events occurring: no event (the mortgage continues, base case), late scheduled payment, a curtailment or a full prepayment. Besides modelling the probability of these four events, McCollum et al. (2015) also try to capture the effect of chosen variables on the volume of curtailment. This is done by means of a two-limit tobit regression model. A tobit model is any class of regression models in which the observed range of the target variable is censored some way. The range of this tobit model is between curtailments with an amount between zero and thirteen thousand dollar.

Kuang et al. (2019) estimate the average four month curtailment amount change in response to changing interest rates using a distributed lag model. The distributed lag model is an regression equation used to predict current values of the curtailment amount (the target variable) based on both current values of an explanatory variable and the lagged (past period) values of this explanatory variable. The regression equations consists of an intercept to be estimated, the explanatory variable of the lagged periods (the curtailment amount) and corresponding lag weights. Since Kuang et al. (2019) are studying the relationship between changes in interest rates

on the amount curtailed for adjustable rate mortgages, the lag weights are the delta of the interest rate in previous periods. The number of lagged periods is three months.

Forms of logistic functions are often used and suitable to study curtailment. Because we want to study impact of different determinants on curtailment, we need a model that provides information about the impact on the probability of curtailment for each determinants. With impact, we mean the magnitude of change as well as the direction of change. The advantage of logistic regression is that it not only provides a measure of how appropriate a determinant is (coefficient size), but also its direction (positive or negative). This enhances understanding of different determinants on the probability of curtailment. Therefore, we choose to model the probability of curtailment using a multivariate logistic regression model. One disadvantage is that logistic regression requires average or no multi-collinearity between determinants. Therefore, we might have to make trade-offs between the variables that are used as final input to be studied.

Regarding the volume of prepayment, we want to review the relation between explanatory variables and the amount that is curtailed. One way to model the volume, is to model it as proportion of the outstanding debt that is curtailed. *Cribari-Neto & Ferrari (2010)* discuss the beta regression technique which is suitable for proportions and rates in the unit interval, with the extreme values excluded. Again, we would like to investigate how different determinants influence the rate of outstanding debt that is curtailed. The beta regression estimates regression coefficients for all determinants which tells us about the size as well as the direction of impact the predictors have on the proportion of outstanding debt that is curtailed. We will discuss the beta regression model and the multivariate logistic regression model in more detail in Chapter 3.

2.3 Analysis Rabobank data and initial determinant selection

In this section we discuss how we explore the Rabobank data and what approach we use to find interesting determinants in this data. Then, combining insights from the Rabobank data and literature review we form an initial set of determinants to research. In the subsections 2.4.1 and 2.4.2, we discuss determinants for the probability and curtailment rate and formulate a hypothesis why and how these affect the probability or curtailment rate.

To explore the Rabobank data, we proceed as follows. First of all, because of the size of the data available we make a quick-scan of variables that might be interesting to model the probability or curtailment rate. This results in a long list of variables that we review. For these variables, we review summary statistics such as average values, standard deviation and outliers to find differences. In addition, visualize this data by plotting variables against the number of curtailments or curtailment rates of curtailments in order to see whether stable trends are caused by certain variables. See for example the visualization of the borrowers age against the number of curtailments, in Figure 2.1.

Furthermore, there are variables that are simply available because they are collected by the Rabobank, for example the original principal amount, but we can construct variables ourselves using the available data as well. Examples of such variables are frequency of curtailment, recent curtailment activity or average curtailment amount.

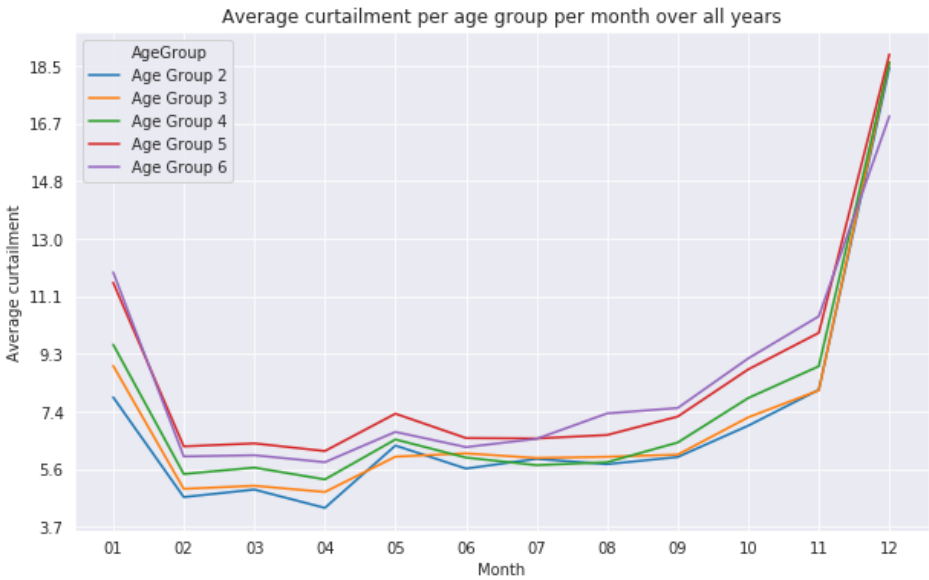

Figure 2.1 The probability of curtailments per age group.[3]

Based on the availability of variables in the data, the data analysis and literature review on determinants we made an initial selection of determinants to research. We go over the determinants selected for the probability and curtailment in separate sub sections below.

2.3.1 Determinants probability of curtailment

In this section we go over the determinants selected to study the impact on the probability of curtailment. We describe each determinant and form a hypothesis on the expected effect on the probability of curtailment. Due to the number of determinants, only a selection is discussed in this subsection. The remaining determinants can be found in Appendix A.1.

Percentage of months curtailed

This variable calculates the average months curtailed over all the months the mortgage is active. This variable can also be interpreted as to what frequency a borrower curtails. From the available data, we deduce there is a subset of borrowers that make a small prepayment on a more frequent basis. For these borrowers this average will be relatively high. Therefore, the expectation is that for high averages on this variable the probability of a curtailment in a next month increases.

[3] The y-axis values are displayed as percentage of the summed average curtailment percentages, due to confidentiality.

Mortgage costs to principal rate

Multiple factors determine the total monthly mortgage of a borrower. First of all, the amortization type determines whether the total monthly costs consist only of interest costs or redemption costs as well. Next to that, the yearly income of the borrower influences the maximum mortgage burden the borrower can afford month. Lastly, the borrower decides to what extent he wants to utilize this credit limit. We expect that high ratios of yearly mortgage costs to yearly income are negatively correlated with probability of curtailment.

Amortization type

There are five different amortization types in the dataset: the linear mortgage, annuity mortgage, bullet mortgages, saving mortgages and interest only mortgages. All of these types have different amortization schemes and interest rate components of the lifetime of the mortgage. Because these different types are expected to have different effects on the probability of curtailment, each type is used as a single variable.

First of all, we should distinguish the types bullet, interest only and saving from linear and annuity mortgages. This is because the latter two mortgage types include a redemption part, meaning that the mortgage debt is gradually decreased over time.

Interest only mortgages do not have an amortization schedule but only pay interest costs. The interest cost is calculated by multiplying the interest rate of mortgage by the complete outstanding balance amount (which is equal to the original principal amount in since there is no redemption). Therefore, the interest costs remain at the same level during the fixed interest period in case of no curtailment. In addition, there is no agreed end date for this mortgage type. The mortgage matures once the borrower repays the whole principal amount, for example when selling the house. Because there is no amortization schedule, the only way an interest only mortgage can reduce the outstanding balance is by curtailing (or full prepaying). When the borrower does not make any redemption over the life of the mortgage, it takes the risk that when house prices are decreasing the sale of the house will not be enough to cover the mortgage debt. In that case, the mortgagee is left with a residual debt.

Interest only mortgages that have a relatively high interest coupon rate (for example, in combination with a long fixed interest rate period where the interest coupon could be 4 or 5 percent) may choose to refinance (changing the mortgage contract terms) instead of making a curtailment. This will results in less curtailments. However, similar mortgages with a high coupon rate that originated before 2013 are allowed subtract interest costs from their payroll taxes. Especially for mortgages with high coupon rates, this will increase the tax benefits. Concluding on these arguments, we expect that for interest only mortgages we see a decreased probability of curtailment.

The bullet mortgage is a mortgage which does not have a redemption schedule but does have a pre-determined maturity time. This is how the mortgage type distinguishes itself from an interest only mortgage. The bullet mortgage can be seen as a lump-sum, where a principal amount is borrowed and the whole amount is repaid at the determined maturity. Because the bullet mortgage does not have a redemption schedule, we are interested how the bullet mortgage affects the probability of curtailment. We recognize that the bullet mortgage is quite similar to the interest only mortgage except it has no agreed end date, hence we should expect the bullet mortgage also to be correlated with a decrease in the probability of curtailment.

For saving mortgages the mortgagee opens a separate savings account where each month a scheduled savings amount is deposited. For this mortgage, a saving target amount is agreed on. This savings amount may be less or equal to the original principal amount. There is no redemption schedule for this mortgage. Therefore, the interests costs are calculated over the complete original principal amount if the mortgagee does not make a curtailment. By depositing to the savings amount the mortgagee builds up capital over time. Because the interest costs are calculated over the complete principal amount, the tax advantages for this mortgage are relatively high throughout the lifetime of the mortgage. In addition, for mortgages before 2013 the savings capital is exempted from Box 3 of the Dutch tax system. Amongst all amortization types the saving mortgage is therefore tax-wise an attractive option. Lastly, mortgage holders receive interest on the money on their savings account where the interest received is equal to the interest paid on the mortgage. Borrowers can either choose to deposit additional money to their mortgage savings account, which is an unscheduled savings deposit or specifically make an curtailment which reduced the outstanding debt. We do not take into account unscheduled extra savings, since these are formally not curtailments. We expect that the saving mortgage is negatively related with the probability of curtailment since making an extra deposit on the savings account is an attractive alternative seeing the tax advantages.

For the linear and annuity mortgage there is no obvious effect on the probability of curtailment over the lifetime of the mortgage. However, for linear mortgages the outstanding balance hence the interest costs decrease over time. This could imply that when the total monthly mortgage costs of the linear mortgage reduces over time, this provides opportunity for curtailment. Hence, for linear mortgages the probability of curtailment could be correlated with aging of the mortgage. For annuity mortgages, the monthly mortgage costs remain constant over time. The annuity mortgage is constructed in such way that early in the lifetime of the mortgage the interest costs component is higher than the redemption component. Over the lifetime, the interest costs component reduces and the redemption component increases. Therefore, the tax advantages of interest deduction decrease over time. Hence, for annuity mortgages the probability of curtailment could be negatively correlated with the loan age of the mortgage.

Interest rate incentive

McCollum et al.(2015) define the relative attractiveness of a curtailment, the savings premium, as the mortgage rate minus the short term risk free rate. When the mortgagee is in the position to contemplate on making a curtailment due to the gap between the rate on his mortgage and the interest rate for an alternative investment, this decision is bound until the end of the fixed interest rate of his mortgage. By the end of this fixed interest rate period, the mortgagee can negotiate a new mortgage rate. Therefore, the 12 months LIBOR is replaced here by the compared interest rate. The compared interest rate is the rate the mortgagee would get if he would currently take out the same mortgage he has, for the remaining time of his fixed interest rate period.

Interest rate type

The interest rate type for mortgages can either be fixed or variable. Variable rate mortgages can prepay without any limitation in any case. For fixed rate mortgages, this is only the case in between fixed interest rates periods. It would make sense if fixed rate mortgages therefore make smaller curtailments (below the penalty free prepayment threshold) whereas variable rate mortgages do not have to care about this and can make bigger curtailments. This would imply that the frequency of curtailment for fixed rate mortgages is higher, hence the fixed interest type mortgage is positively correlated with the probability of curtailment.

The complete list of determinants included in the initial selection can be found in Table A.3.1, in Appendix A.3

2.3.2 Determinants curtailment rate

In this section we go over the determinants selected to study the impact on the curtailment rate. We describe each determinant and form a hypothesis on the expected effect on the curtailment rate. Due to the number of determinants, only a selection is discussed in this subsection. The remaining determinants can be found in Appendix A.2.

Ever curtailed

Where *Fu (1997)* finds that previous record of curtailment increases the likeliness of curtailment, for the curtailment rate this does not necessarily hold. It would not make sense if a borrower is curtailing in a high curtailment rate on a frequent basis. Similarly, if borrowers curtail on a frequent basis it is more likely these curtailments are smaller amounts. We expect borrowers curtailing rarely to curtail in higher curtailment rates compared to borrowers that curtail on frequent basis. Therefore, we expect that once a borrower already curtailed this will be correlated with a decrease in the curtailment rate.

Percentage of months curtailed

First of all, this variable can be interpreted as the frequency to which the borrower curtails. We expect borrowers that frequently curtail to curtail in smaller curtailment rates. Therefore, we expect a high percentage of months curtailed to be negatively correlated with a high curtailment rates.

Amortization type

As mentioned earlier, there are five amortization types. The mortgage types bullet and interest only do not have amortization schedule. The idea behind these mortgages types is that the mortgage holders save up on their own responsibility, hence we expect that when these mortgages curtail it will be in higher proportion of outstanding debt than mortgages types such as linear or annuity mortgages. In short, we are expecting that interest only and bullet mortgages will curtail in higher curtailment rates compared to linear, annuity and saving mortgages.

Age

Depending on the age, different effects on the curtailment rate are expected. For example, younger borrowers are more likely to have recently bought their house. For starters, it is unlikely that large curtailments are done (since if the money was available, it would be more logical to take a lower mortgage loan). The middle aged are in the most unstable period of their lifetime, since switching jobs and therefore moving but also getting children and or getting divorced can impact curtailment decisions. Lastly, we expect older borrowers to be in the most stable phase concerning their private situation and older age could be associated with larger financial saving buffers thus the funds to curtail.

Franchise rate

Mortgages in the available dataset are either basis mortgages or plus mortgages, where basis mortgages can curtail 10 percent of the original principal amount yearly penalty free and plus mortgages can do this up to 20 percent. If they make a curtailment that exceeds this, the penalty is based on the amount that exceeds the franchise rate and the interest payments that the bank loses on this amount. We expect higher franchise rates are positively correlated with higher curtailment rates.

Percentage amortized

The percentage amortized is the percentage of the original principal amount that is paid off. For annuity and linear mortgages the initial debt is gradually decreased over the lifetime of the mortgage. Therefore the percentage that is already amortized automatically influences the proportion of outstanding debt that can be curtailed; when the percentage amortized is low the range of the possible curtailment volumes is high whereas when the percentage amortized is high, the range of the volume that can be curtailed gets smaller. In addition, when the percentage amortized is high, keeping the absolute amount constant, the curtailment rate will automatically

become higher during amortization. Hence, we think that higher percentages of amortization are correlated with higher curtailment rates.

Percentage franchise used

The amount the borrower is allowed to curtail yearly penalty free is called the franchise amount. Hence, the percentage franchise used is the amount curtailed as percentage of the penalty free amount. Since borrowers can determine the curtailment amount themselves, we expect them to want to avoid the penalty. Therefore, we expect a higher percentage of franchise used to be correlated with decreased curtailment rates.

Income to principal rate

High income to principal rates can indicate there is more financial capacity to curtail in higher volumes. Therefore, we are interested whether higher income to principal rates are correlated with higher curtailment rates.

Mortgage costs to principal rate

Higher mortgage costs to principals rates can indicate two things. First, high values show that the mortgage is amortized at faster pace than other mortgages. Second, it can indicate that because the mortgage costs are high there is less room left to additionally curtail. Therefore, we expect that borrowers having high mortgage costs to principal rates are correlated with decreased curtailment rates.

Percentage curtailed to principal

When the percentage curtailed to principal is high, this means that the borrower has already curtailed a significant share of his original principal in the past. Hence, we expect when the percentage curtailed to principal is high this is negatively correlated with high curtailment rates.

The complete list of determinants included in the initial selection can be found in Table A.3.2, in Appendix A.3

In Section 3.3 we explain how the initial set of determinants is filtered into the final set used for modelling.

3. Methodology

In the second chapter we chose the determinants to do research on and chose the modelling techniques. In this chapter, we will first describe how a curtailment is defined in this research. Subsequently, we will discuss the modelling techniques we use to model the probability of curtailment and curtailment rate in more detail. After the modelling techniques, we describe how we filter the set of variables to be considered for modelling to the final set of variables to use for modelling. In the last section of this chapter, we explain how we are estimating total portfolio cash flows using both models.

3.1 Definition curtailment and curtailment rate

A curtailment is recognized when a mortgage account receives more than the scheduled redemption amount but less than the complete outstanding balance on the mortgage in that period. The outstanding debt is equal to the original principal amount of the mortgage minus all scheduled redemptions up to that moment in time. The curtailment is given by (1):

$$C_{i,t} = \begin{cases} 1 & a \cap b \\ 0 & else \end{cases} \qquad (1)$$

where

a) $R_{i,t} - RS_{i,t} > 45.0$

b) $R_{i,t} - RS_{i,t} < 0.99 * OB_{i,t-1}$

where

$C_{i,t}$ the curtailment for mortgage *i* at time *t*

$R_{i,t}$ the redemption for mortgage *i* at time *t*

$RS_{i,t}$ the scheduled redemption for mortgage *i* at time *t*

$OB_{i,t}$ the outstanding balance of mortgage *i* at time *t*

The minimum amount for a curtailment for residential clients is set at 45 euros, hence we use this amount here as well. We are using the threshold of 99 percent of the outstanding balance to distinguish curtailments from full prepayments.

$C_{i,t}$ is the target variable of interest for modelling the probability of curtailment. We are interested how the determinants from Chapter Two affect the probability of a curtailment of mortgage *i* in period *t.*

The curtailment rate is the proportion of the outstanding debt that is curtailed in period *t-1,* minus the scheduled redemption in period *t*. The curtailment rate is formally expressed as:

$$CR_{i,t} = \frac{C_{i,t}}{OB_{i,t-1} - RS_{i,t}} \qquad (2)$$

We are interested how the selected determinants in Chapter Two affect the curtailment rate of mortgage *i* in the next period *t.*

3.2 Modelling techniques

In Chapter Two we chose to use a multivariate logistic regression model to model the probability of a curtailment and a beta regression model to model the curtailment rate. Both modelling techniques are now discussed in more detail.

3.2.1 The multivariate logistic regression

For each mortgage we model the probability a mortgage will curtail or not in the next period. We give the target variable $C_{i,t,}$ a value of 1 in case the mortgage curtails and 0 otherwise. The multivariate logistic model is a modelling technique for estimating the probability of a dichotomous outcome given a set of predictors. Hastie and Tibshirani (2017) generalize the multivariate logistic function as follows:

$$\log\left(\frac{p(C_{i,t})}{1 - p(C_{i,t})}\right) = \beta_0 + \beta_1 X_{1,t} + \cdots + \beta_p X_{p,t} \qquad (3)$$

Where the left side is called the log odds or the logit of curtailment and $X_{i,t} = (X_{1,t}, X_{2,t}, \ldots, X_{p,t})$ are p predictors for mortgage *i* in period *t*. The predictors p are all the determinants selected in Chapter Two. We can rewrite the logistic function as:

$$p(C_{i,t}) = \frac{E^{\beta_0 + \beta_1 X_{1,t} + \cdots + \beta_p X_{p,t}}}{1 + E^{\beta_0 + \beta_1 X_{1,t} + \cdots + \beta_p X_{p,t}}} \qquad (4)$$

Where $p(C_{i,t})$ is the probability that mortgage *i* makes a curtailment in period *t* given the set predictors.

We are interested in the regression coefficients $\beta_0, \beta_1, \ldots, \beta_p,$ since these explain how the log odds of the probability of curtailment increase or decrease when changing the predictor by one unit.

To estimate the regression coefficients $\beta_0, \beta_1, \ldots, \beta_p$, we use the maximum likelihood method, defined by the likelihood function (5):

$$L(\beta_0, \beta_1, \ldots, \beta_p) = \prod_{i:y_i=0} p(C_{i,t}) \prod_{i':y_i'=0} (1 - p(C_{i,t})') \qquad (5)$$

The estimates of $\beta_0, \beta_1, \ldots, \beta_p$ are chosen to maximize this likelihood function.

Hastie and Tibshirani (2017) explain the basic intuition behind using maximum likelihood to fit a logistic regression model is as follows: we seek estimates for $\beta_0, \beta_1, \ldots, \beta_p$ such that the predicted probability $p(C_{i,t})$ of curtailment for each individual, corresponds as closely as possible to the individual's observed actual curtailment. In other words, we try to find estimates $\beta_0, \beta_1, \ldots, \beta_p$ such that plugging these estimates into the model for $p(C_{i,t})$ given in (4), yields a number close to one for all individuals who curtailed, and a number close to zero for all individuals who did not.

The data is split into a sample and validation interval. We take a full year of data as validation interval, which equals 14 percent of the available data. We use the validation interval to make out-of-sample estimates for the probability of curtailment and estimating the conditional curtailment rate. The remaining data is used as sample interval. The sample interval is used to estimate the regression coefficients of all selected predictors.

All the determinants that are used for modelling have another scale. For example, the determinant borrowers age has another scale than the determinant percentage of months curtailed. After estimating the regression parameters we do want to compare the size of the regression parameters such that statements can be made to what extent the variables have different impact on the probability of curtailment. Therefore, we standardize the data first such that they are on the same scale. This is done by subtracting the mean determinant value from each individual determinant value and then divide by the standard deviation of this determinant. In this way, each value for all determinants will have a mean of zero and a standard deviation of one.

3.2.2 The beta regression

The beta regression model, formally defined later in this section, is commonly used for modelling continuous variables, such as proportions or rates, that have values on the standard unit interval (0,1). It is based on the assumption that the target variable, in our case the curtailment rate, is beta distributed. Also, it assumes that its mean (the mean curtailment rate) is related to a set of predictors through a linear predictor with unknown coefficient and a link function. The advantage of the beta distribution is that it can take on a lot of different densities, See Figure 3.1.

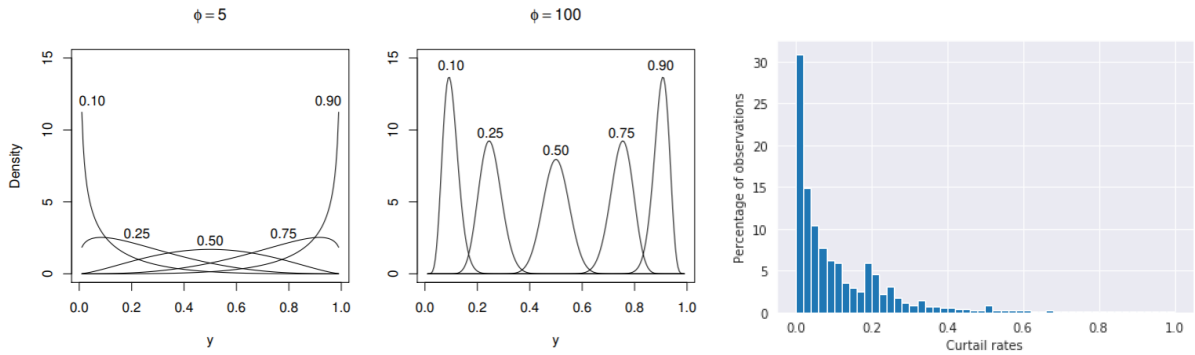

Figure 3.1: Possible densities of the beta distribution at different mean y and precision parameter φ in comparison to actual density in the available data. [4]

In the left two graphs of Figure 3.1 you can see different beta densities. The right graph of Figure 3.1 shows the distribution of curtailment rates from the available data. The majority of curtailment rates are around 0.0, meaning that most of curtailments are only a small proportion of the outstanding balance of the mortgage. If we compare the left and right graph of Figure 3.1, a beta distribution with a mean *u* of 0.10 in and with a φ parameter of 5 looks quite similar to the distribution of curtailment rates in the data.

However, the beta model does not support values that equal the extremes 0 or 1. In the case of curtailment rates, a value of 1 means the complete proportion of the outstanding balance of the mortgage is curtailed. Since this equals a full prepayment which is not examined in our research, it is not a problem that the beta model not include modelling values of 1. However, a value of 0 means the proportion of the outstanding balance curtailed is 0, thus there is no curtailment. This is true for the majority of observations in the dataset. Therefore, if we want to model also all observations without curtailment, the beta regression would not be suitable. An alternative is to look for a statistical technique that does incorporate the extreme value zero. For example, a fractional logistic regression model allows the target variable to be modelled to have value between and including 0 and 1. However, the dataset would include an enormous amount of curtailment rates of 0 compared to only a small proportion where the curtailment rate is greater than 0. When estimating the regression parameters on this dataset, it is likely the model will become biased. Therefore, we choose to model the curtailment rate *given the borrower curtails. This way, the model only needs to support instances where there was in fact a curtailment, excluding values of 0.* We refer to this as the conditional curtailment rate model.

[4] *Note:* "Probability density functions for beta distributions with varying parameters μ = 0.10, 0.25, 0.50, 0.75, 0.90 and φ = 5 (left) and φ = 100 (right)". Reprinted from *Beta Regression in R, Page 2. By* Cribari-Neto, F., Zeileis, A. 2012.

Ferrari and Cribari-Neto (2004) define the beta density with parameters u and φ by the following:

$$f(y;u,\phi) = \frac{\Gamma(\phi)}{\Gamma(u\phi)\Gamma\big((1-u)\phi\big)} y^{u\phi-1}(1-y)^{(1-u)\phi-1} \qquad (6)$$

with $0 < u < 1$ and $\varphi > 0$. In the application to modelling the curtailment rate, the parameter u is the *mean curtailment rate*. Let $y_1, \ldots, y_n$ be all the curtailment rates from the mortgages in the available dataset, such that $y_i \sim B(u_i, \varphi)$. The beta regression is defined as:

$$G(u_i) = b_0 + b_1 Z_{1,i} + \cdots + b_p Z_{p,i} \qquad (7)$$

G(•) is the link function making sure the estimations of the mean curtailment rate lie within the standard unit interval of (0,1). Many link functions are available. The logit link function is the default option. The advantage choosing the logit link function is that the estimated regression parameters can be interpreted as the log odds increase or decrease of the mean curtailment rate. Therefore, we choose the logit link function. $Z_1, \ldots, Z_1$ are the predictors which are the determinants in Chapter 2. Also, $b_1, \ldots, b_p$ are the corresponding regression coefficients. The regression coefficients are estimated by means of the maximum likelihood method.

3.3 Final determinant selection

In Chapter 2 we made an initial selection to model the probability of curtailment and curtailment rate. Before including these determinants into the final models, we want to review the relation between the determinant and target variable individually, we want to look at the correlation amongst determinants and look at transformation of determinants.

Statistical significance

In order to see whether the determinants are statistically significant, an univariate analysis is done. In this univariate analysis just one determinant serves as input for the logistic model or beta model to see the relation of an determinants to the target variable without interaction effect of other determinants. When an determinant is statistical significant to the target variable this means that changes of the determinants correlate with shifts in the target variable. The p-value of each determinant tests the null-hypothesis that the determinant has no correlation with the target variable. We will examine the p-value at a significance level of 5 percent. Therefore, when the p-value is less than the significance level we assume there is enough evidence to assume correlation of the determinant with the target variable hence we include this determinant in the final model.

Detecting correlation amongst variables

One disadvantage of the logistic regression model is that it requires average or no multi-collinearity between determinants. Multi-collinearity is the event of great inter-correlations among determinants in a multiple regression model. Multi-collinearity can prompt more extensive confidence intervals and less solid likelihood estimations for determinants. *Senaviratna & Cooray (2019)* describe two methods to diagnose multi-collinearity in a logistic regression model. These are reviewing the Pearson correlation coefficients among determinants and reviewing the Variance Inflation Factor (VIF).

A correlation matrix shows the value of the Pearson correlation coefficients for all pairs of variables. *Senaviratna & Cooray (2019)* state as a general rule of thumb that if correlation is greater than 0.8 or 0.9, multi-collinearity is a serious problem.

The VIF is a tool to quantify how much variance is inflated. The value of the VIF indicates how much times the variance is inflated as the predictor would be when uncorrelated to other predictors. The VIF is defined as:

$$VIF = \frac{1}{1 - R_i^2} \qquad (8)$$

Where R^2 is the coefficient of determination in linear regression. The VIF method is executed by picking each determinant and regress it against all other determinants. For each regression, the VIF is calculated as in (8). We calculate the VIF values for all determinants using a Python package (see Appendix D.1). VIF values equal to one mean no correlation, between 1 and 5 and moderately correlated and values higher than 5 are highly correlated. Therefore, VIF values above 5 are carefully reviewed. Concluding on this, for all determinants we check the Pearson correlation coefficients and VIF values and make adjustments once these exceeds the thresholds.

3.4 Cash flow estimation and evaluation

In this section we discuss how we estimate the curtailment cash flow using both the probability and rate model and the studied determinants.

The actual cash flow for an individual mortgage originates from three factors: the probability of curtailment, the conditional curtailment rate and the outstanding mortgage debt at that moment, adjusted for the scheduled redemption. In the data, the probability of curtailment is either 0 or 1. The curtailment rate can be any rate between 0 and 1 with exclusion of the extreme values 0 and 1. The outstanding debt can take on any value between observed values in the dataset. By multiplying these factors, we can derive the actual cash flow resulting from a curtailment on loan level of mortgage *i* in month *t*. See the following formula.

$$CF_{i,t} = [P(Curtail)] * [CR_{i,t} | \, P(Curtail) = 1] * [OB_{i,t} - RS_{i,t}] \qquad (10)$$

We want to estimate the total cash flow of the mortgage portfolio of each month from the validation sample. To do this, we estimate the cash flow for each mortgage using formula (10) and sum all individual cash flows to find the total cash flow of the portfolio.

The logistic model estimates for each individual the probability of curtailment, where this probability is between 0 and 1. Because the actual curtailments take values of 0 and 1, we want the estimations also to either be 0 or 1. To deduce whether the mortgage will curtail or not given its estimated probability, we draw a number from the uniform distribution on the unit interval. We say the mortgage curtails if the number drawn is below the estimated probability. This way, mortgages for which we estimate a low probability of curtailment often get a 0 assigned in the long run. In fact, we simulate whether mortgages curtail. To get more stable results, we simulate this a 1000 times for each out of sample months.

After we simulated curtailments all mortgages in the portfolio, we have a simulated group of mortgages that curtail. For these mortgages, we estimate the curtailment rate using the beta regression model and multiply this with the outstanding balance (adjusted for the scheduled redemption). Finally, we sum all estimated individual cash flows for the simulated mortgages to find the total cash flow. We compare this with the actual total cash flow, to find the estimation error using both models. We use the same sample and validation interval as mentioned in Section 3.2.1. For each month in the validation interval, we simulate the curtailing mortgages to estimate the total cash flow.

4. Data

The first section explains how the data is processed and what exclusions have been made or changes in construction of determinants. Section 4.2 shows the different steps of the determinant selection for the probability model. Section 4.3 follows up with the determinant selection for the curtailment rate model. Lastly, the definition of determinants are explained in Section 4.3 and the descriptive data of all determinants is shortly mentioned in Section 4.4.

4.1 Data processing

The initial available dataset contained an enormous amount of interesting determinants. During the data analysis we made some exclusions and changes in construction of determinants. Business loans have been excluded since we expect that determinants of curtailment are not comparable from those of private mortgages. Mortgage interest types were available in four categories: fixed, variable, stable and roll-over. Because the categories stable and roll-over only occurred a small number of times, these two have been excluded. Next to that, we observe seasonality for the months throughout the year. However, changes were sometimes quite small within a quarter and changes in the probability of curtailment showed the most significant changes in the first and last three months of the year. Hence, we chose to construct months April to September of the year into quarters, with the first and last three months as separate variables. Moreover, when visualizing probabilities of curtailment over ages, the probability of curtailment showed quite different patterns when looking at young, middle aged and older borrowers. Therefore, we split the predictor age up into three categories being dummy variables. The younger class applies when the age is below 35, the middle class applies when the age is between 35 and 65 and the older class applies when the borrower is older than 65. Finally, we started with all provinces included as individual determinants, but only found that the curtailment percentages were slightly higher in the provinces Noord-Holland, Zuid-Holland and Utrecht. Therefore, we constructed a dummy variable indicating whether the province of the mortgage holder belongs to the Randstad or not.

For the curtailment rate model, we saw that the average curtailment rates from the amortization types linear, annuity and interest only did not deviate from the average curtailment rate as much as the amortization type bullet and saving do. Therefore, we constructed the amortization type annuity, linear and interest only together as dummy variable and bullet and saving as separate determinants.

4.2 Determinant selection probability model

In this section we discuss the statistical significance of determinants for the probability of curtailment, we review the correlation matrix and the VIF values as discussed in Section 3.3. In addition, we briefly reflect upon possible transformations of variables.

Statistical significance

See the results for univariate analysis of determinants for the probability model in Table B.1.1., in Appendix B.1. From all determinants only the determinant "Second Mortgage Indicator" is statistically insignificant. Therefore, based on the p-value we choose to exclude this determinant.

Correlation amongst variables

First of all, we review the values of the Pearson correlation coefficients. Due to the high number of determinants and the size of the corresponding correlation matrix, we filtered the results based on determinants with a higher correlation coefficient of 0.5. See Table 4.1

Determinant	Determinant	Correlation Coefficient
Interest Only	Annuity	0.618
Fixed Interest Rate Period	Time to Interest rate reset	**0.898**
Age Group I	Age Group II	0.707
Age Group II	Age Group III	0.592
Unemployment rate	House price index	**0.968**
Curtailed Last 12 months	Ever Curtailed	0.721
Curtailed Last 12 months	Percentage Curtailed to Principal YTD	0.670

Table 4.1 Selection of the correlation matrix of determinants.

We see that two pairs of determinants exceed the threshold of 0.8 Therefore, we have to choose one of each of these pairs. We choose time to interest rate reset over fixed interest rate period since this determinant holds more information (the length of the period and the remaining time). We have no decision metric for the other pair, therefore we arbitrarily choose the unemployment rate.

Next to that, we review the VIF values. The complete table of VIF values for all determinants can be found in Table B.2.1, in Appendix B.2. Seeing the VIF values in Table B.2.1, we noticed that all determinants transformed to dummy variables exhibit high VIF values. Therefore, we will use the month February, annuity amortization type, Age Group 2 as base case scenario. This means they are removed from the final model, and therefore captured by the intercept of the model. After removing these determinants (first adjustment), the determinants interest type, unemployment rate, number of borrowers and curtailed last 12 months still show high VIF values. In the Pearson correlation coefficients we also see that curtailed last 12 months is quite highly correlated to two other determinants. Hence, we removed the unemployment rate and curtailed last 12 months. After this adjustment, all VIF scores are below 5.

Transformations of determinants

All determinants from Table B.1.1 are assumed to have a linear relation to the probability of curtailment. Seeing the average probability of curtailment over the range of the determinant loan age in Figure 4.1, it follows the trend of a logistic function where the average probability of curtailment slowly decreases when loan age increases. Therefore, we use a log function transformation on the loan age values.

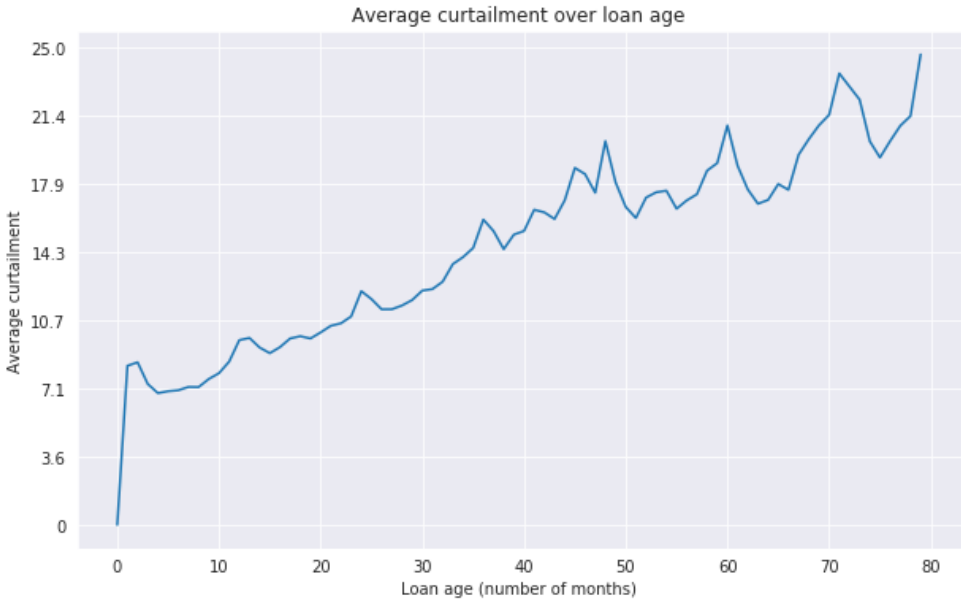

Figure 4.1 Average probability of curtailment over the value range of loan age.[5]

4.3 Determinant selection curtailment rate model

In this section we discuss the statistical significance of determinants for the curtailment rate, we review the correlation matrix and the VIF values as discussed in Section 3.3. In addition, we briefly reflect upon possible transformations of variables.

Statistical significance

See the results for univariate analysis for the determinants of the curtailment rate in Table B.3.1, Appendix 3.1. The determinant 'Rented House' is not statistically significant to the curtailment rate. Therefore, we exclude this determinant.

Correlation amongst variables

First of all, we review the values of the Pearson correlation coefficients. Due to the high number of determinants and the size of the corresponding correlation matrix, we filtered the results based on determinants with a higher correlation coefficient of 0.5. See Table 4.2

[5] The y-axis values are displayed as percentage of the summed average curtailment percentages, due to confidentiality.

Determinant	Determinant	Correlation
Age Group 2	Age Group 1	0.764
Age Group 2	Age Group 3	0.545
Last Curtailment	Percentage Amortized	0.574
Last Curtailment	Percentage Curtailed To Principal	0.588
Last Curtailment	Franchise Exceeded Indicator	0.624
Last Curtailment	Percentage Curtailed To Principal YTD	0.740
Last Curtailment	Average Curtailment	**0.944**
Last Curtailment	Highest Curtailment	**0.897**
Percentage Curtailed To Principal YTD	Percentage Curtailed To Principal	**0.987**
Percentage Curtailed To Principal YTD	Franchise Exceeded Indicator	0.718
Percentage Amortized	Percentage Curtailed To Principal	0.566
Average Curtailment	Percentage Amortized	0.566
Average Curtailment	Percentage Curtailed To Principal	0.587
Average Curtailment	Franchise Exceeded Indicator	0.629

Table 4.2 Selection of the correlation matrix of determinants.

Based on Table 4.2, we should make some decisions between determinants for the curtailment rate model. First of all, based on the Pearson correlation coefficients, one out of three determinants from last curtailment, average curtailment and highest curtailment should be included. We think the determinant average curtailment holds the most information, since it holds information on multiple curtailment observations instead of one (if this applies for the borrower). Therefore, we include the determinant average curtailment.

Next to that, we review the VIF values. The complete table of VIF values for all determinants can be found in Table B.4.1, in Appendix B.4.We filtered the determinants showing high VIF values in Table 4.3.

Determinant	VIF	VIF first adjustment	VIF second adjustment
Number of borrowers	1.028	14.028	-
Loan Age	1.965	26.457	-
Unemployment rate	1.293	20.652	-
Franchise Rate	1.217	56.113	-
Ever curtailed	2.202	4.980	4.013
Percentage Amortized	2.259	4.563	3.317
Percentage Curtailed to Principal	2.271	2.979	2.876
Percentage of Months Curtailed	1.878	2.973	2.117
Percentage Franchise Used	2.275	3.024	3.003
Franchise Exceeded Indicator	2.773	3.006	2.988
Average Curtailment	3.111	4.044	3.920

Table 4.3 Determinants of the curtailment rate model with high VIF values.

Also here, we notice that all determinants transformed to dummy variables exhibit high VIF values. Therefore, we omit one category of all categorical determinants in the final model. Furthermore, we note that after a first adjustment, some determinants exhibit high VIF values. Therefore,

Next to that, since the amortization types and age groups are transformed to dummy variables, these are creating high VIF values. Therefore, we will use reference encoding dropping the annuity determinant and the second age groups such that these are captured in the intercept of the

regression but do not cause high VIF values. Therefore, the determinants unemployment rate, franchise rate, number of borrowers and percentage curtailed to principal year-to-day are excluded. After excluding these determinants, all VIF values were below 5.

Transformations of determinants

All determinants from Table B.3.1 in Appendix B.3 are assumed to have a linear relation to the curtailment rate. We reviewed the average curtailment rate of the determinants percentage amortized and percentage franchise used over the range of the determinant loan age in Figure 4.1. The percentage amortized looks to follow the trend of an exponential function, therefore we use an exponential function transformation on all percentage amortized. In addition, we see that the average curtailment rates show a different pattern above the threshold 1.0. Therefore, we split this determinant into percentage franchise used and a franchise exceeded indicator.

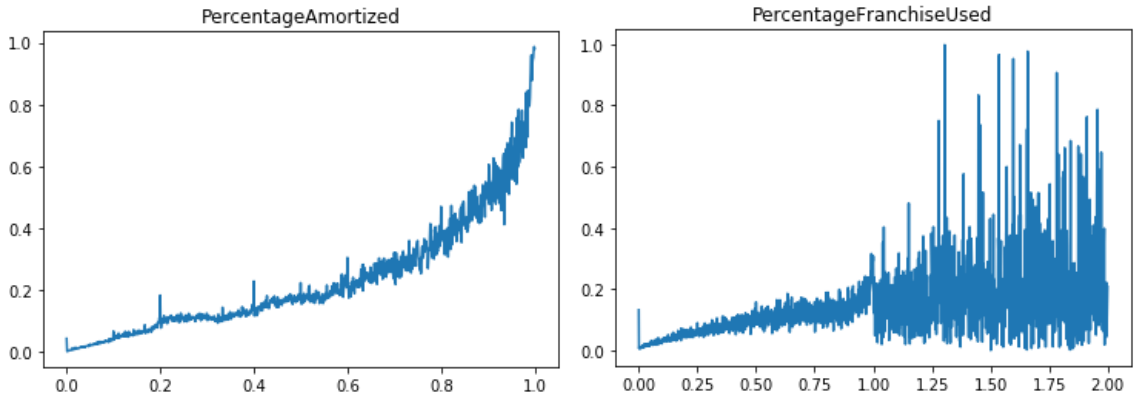

Figure 4.2 The average curtailment rate plotted over the range of determinants for percentage of months curtailed, the percentage franchise used and the percentage amortized.

4.4 Definition of determinants

In the probability model a lot of determinants are dummy variables, meaning they are either 1 when they apply and 0 when they do not apply. This is the case for all periods within the year, for the amortization types and for the age groups. The LTV ratio is also a dummy variable. When the LTV rate is higher than 0.9, then the LTV dummy variable is 1 and otherwise 0. For the Randstad indicator, the dummy variable is 1 if the mortgage holder lives in Noord-Holland, Zuid-Holland or Utrecht. The interest type is also a dummy variable, where the value of 1 means the mortgage is a fixed interest rate mortgage and the value of 0 means the mortgage is a determinant rate mortgage. Furthermore, the loan age is the number of months the mortgage is active. The interest rate incentive is the difference between the coupon rate of the mortgage and the compared interest rate. The determinant can have the value 0 to 6, where the value of 0 means the coupon rate is lower than the compared interest rate, the value of 1 means the difference between the coupon rate and compared interest rate is between 0 and 1 percent, a value of 2 means the difference is between 1 and 2 percent, et cetera. The number of borrowers is the number of borrowers on the mortgage which is either 1 or 2. The determinant in default is either 0 or 1, where the borrower is in default if payments on the mortgage have been missed for longer than 90 days. The mortgage costs to principal is the sum of yearly mortgage costs on the mortgage divided by the original principal amount.

Furthermore, for the conditional curtailment rate model the percentage amortized is the proportion of the principal amount that is paid down. The average curtailment is the average taken over the curtailments being done by the borrower over the life of the mortgage. The percentage of franchise used is the amount curtailed within a year divided by the total amount the borrower is allowed to curtail without a penalty. The franchise exceeded indicator is a dummy variable indicating whether the franchise amount is exceeded.

4.5 Descriptive data

According to the exclusions mentioned in this chapter and possibly changes in construction of determinants, find the descriptive data for the final determinant selection of the probability model and curtailment rate model in Table B.5.1 in and Table B.6.1 in Appendix B.5 and B.6.

5. Results

In this chapter we examine the results of the multivariate logistic regression and the beta regression. In addition, we show the results of the out of sample cash flow estimates.

5.1 Results multivariate logistic regression

All predictors listed in Table 5.1 are standardized before the regression, hence we can directly interpret different magnitudes of regression coefficients as relative importance. Since we use a multivariate logistic regression, we should interpret the regression coefficient as the log odds change in the probability of curtailment when changing the predictor by one unit. Also, several predictors are modelled as dummy variables. We are aware that dummy variables cannot increase with more than one unit, hence we cannot state directly that the highest regression coefficients in Table 5.1 are also the creating the most impact.

Variable	Coefficient [6]	Z-Value	P-Value
January	3.34	52.6	0.00
March	0.20	2.73	0.00
Second Quarter	0.99	10.3	0.00
Third Quarter	1.12	11.5	0.00
October	1.99	29.8	0.00
November	2.52	38.5	0.00
December	5.77	87.4	0.00
Bullet	-1.24	-16.01	0.00
Interest Only	-3.44	-50.4	0.00
Linear	3.35	81.3	0.00
Saving	-1.43	-28.2	0.00
Age Group 1	-0.75	-13.5	0.00
Age Group 3	-1.35	-25.2	0.00
LTV	-1.48	-22.6	0.00
Loan Age	2.95	44.6	0.00
Randstad	0.63	13.9	0.00
Interest Type	0.66	12.0	0.00
Interest Incentive	1.56	29.2	0.00
Time to Interest Reset	0.66	13.3	0.00
Number of Borrowers	0.62	12.9	0.00
In default	-0.64	-8.7	0.00
First Time Buyer	0.75	14.2	0.00
Move Indicator	0.28	5.2	0.00
Rented House	0.85	27.5	0.00
Free Capital	0.57	20.9	0.00
Mortgage Costs to Principal	-48.10	-120.1	0.00
Ever Curtailed	6.50	210.3	0.00
Curtailed Last Quarter	1.88	106.8	0.00
Percentage Principal Curtailed YTD	-2.66	-16.4	0.00
Percentage Months Curtailed	1.75	130.4	0.00

Table 5.1 Results of the multivariate logistic regression.

[6] Regression coefficients are depicted as percentage of the sum of regression coefficients due to confidentiality. The order and relative magnitude between different predictors is maintained.

As mentioned in Chapter 2, we are interested in the effect of activity-specific predictors because these are lacking in other literature. Predictors such as ever curtailed, curtailed last quarter, percentage of principal curtailed year to day and percentage of months curtailed are activity-specific predictors. First of all, we notice the magnitude of the regression coefficients of these predictors is high compared to other predictors. Mortgages that ever curtailed show high increased probability of curtailment. This is in line with findings from *Fu (1997).* Also, borrowers that recently curtailed (last quarter) are more likely to curtail. For higher percentages of months curtailed, in other words, for higher curtailment frequency, probability of curtailment is increasing as well. Curtailments are rare in the available data, but among these curtailments there are more borrowers that curtail on a more frequent basis. However, the higher the volume curtailed, the less likely a new curtailment is done within the same year.

Curtailments show a strong seasonal pattern throughout the year. We took February as base case, and compared the base case we see that from October curtailments increase and reach a peak in December. After the peak in December, we notice an increase in January as well. The peak in December confirms our hypothesis formulated in Chapter 2. Part of this hypothesis is that this peak occurs because borrowers receive a 13th month of salary in December. However, since often this 13th month is paid at the end of the month, this is short notice to prepay in the same month. The 13th month of salary is therefore more likely to explain the increase in curtailments in January. *McCollum et al. (2015)* find similar results for the peak in December. However, they do not find the same increase for January. The dataset used differs in time period and country, where McCollum et al. use American mortgage data. However, if the January peak is indeed caused by the 13th month, we should rather expect this effect to be universal.

The amortization type annuity is left out as predictor, and therefore functions as base case. Compared to annuity, we find linear type mortgages to increase the probability of curtailment. This is in line with expectation stated in Chapter 2, since we argue that tax-benefits are generally higher for annuity mortgages and interest costs (and therefore total mortgage costs) are gradually decreasing for linear mortgages. Furthermore, compared to annuity, the saving and bullet type decrease probability of curtailment. This confirms our hypothesis from Chapter 2. Finally, we see the strongest decrease in the probability of curtailment for the interest only mortgage, compared to annuity. This is according to our expectations from Chapter 2. However, the decrease in probability of curtailment is only due to one of the two proposed reasons, which is a high number of repricing prepayments. In the data we see that the number of repricing prepayments is the highest for interest only. To illustrate, it is about twice as high compared to annuity and linear mortgages, and two and a half times compared to saving mortgages. We state that this decrease is not related to mortgages originating before 2013 that have relatively high tax-benefits, as we do not see a structural difference in curtailment data for mortgages originated in before 2013 opposed to mortgages originated in or after 2013. Also when filtering on different interest incentive levels, no structural prepayment behaviour is noticeable.

The predictor Age Group 2 is left out due to perfect multi-collinearity, hence this functions as the base case. Compared to middle aged class, we see the younger class are less likely to curtail and the elderly class are the least likely to curtail. *Stam (2015)* finds that middle aged borrowers are the most likely to curtail as well. However, *Stam (2015)* reviews the predictor age separately for interest only and saving mortgages. Particularly for saving mortgages, *Stam (2015)* finds an significant increase of curtailment for the elderly class on saving mortgages. Looking at the combined effect in our results of Age Group 3 and saving mortgages, we confirm this finding.

The predictor LTV decreases the likeliness of curtailment. To be more precise, we find that mortgages with a LTV above 0.9 are less likely to curtail. This is in line with findings in other literature, *Budinger and Fan (1995).* Next to that, the predictor loan age is an important predictor of curtailment. Although the regression coefficient is of greatest magnitude compared to other predictors, for non-dummy predictors is it one of the highest meaning that for aged mortgages the likeliness of curtailment increases. *Stam (2015)* finds that specifically for interest only and saving mortgage are aging, the probability of curtailment decreases. Looking at the interaction of the predictors loan age, interest only and saving we find similar results. A possible explanation is that for aged mortgages, borrowers may prefer a full prepayment instead of partial prepayment. For higher interest incentives we see an increase in the probability of curtailment. This is in agreement with findings from *McCollum, Lee & Pace (2015) and Stam (2015)*

The positive regression coefficient of the determinant "interest type" indicates that fixed rate mortgages are more likely to curtail than variable rate mortgages. This is in contrast with results from *McCollum, Lee & Pace (2015),* who argue that adjustable rate mortgages are more likely to curtail, especially for larger curtailment amounts. *McCollum, Lee & Pace (2015)* argue lower future required monthly payments may make curtailment attractive to financially non-constrained variable rate mortgage borrowers. However, in the paper of *McCollum, Lee & Pace (2015)* there is an decreasing interest rate regime, hence variable rate mortgages already have lower monthly payments and less interest rate incentives to curtail. Hence, we cannot place the difference in findings on the effect of the interest type on curtailments.

The predictor "Randstad" represents houses that are located in Noord-Holland, Zuid-Holland or Utrecht. We find an increase of curtailment in these provinces compared to others. The predictor time until interest rate reset increases likeliness of curtailment for higher values. This is in contrast with our expectation and other literature. Looking at the individual regression of this predictor, the direction of the regression coefficient changed. We note that the interpretation of this predictor is biased, due to correlation effects with other predictors. Another predictor of which its regression coefficient changed in direction, is move indicator. We would expect probability of curtailment to increase the first year after a borrower moved places, but the coefficient is positive in Table 5.1. Again, we conclude a biased interpretation due to correlation amongst predictors.

Moreover, we notice a higher number of borrowers linked to the mortgage, higher capital at origination, first time buyers increase the likeliness of curtailment. Borrowers that are in default are less likely to curtail. The predictor that causes extreme decrease in the probability of curtailment are higher yearly mortgage costs compared to the original principal amount. This confirms our hypothesis from Chapter 2, arguing that mortgages amortizing at higher pace are less likely to curtail.

5.2 Results beta regression

Similarly for predictors of the conditional curtailment rate, all predictors listed in Table 5.2 are standardized before the regression, hence we can directly interpret different magnitudes of regression coefficients as relative importance. Since we use a beta regression, we should interpret the regression coefficient as the log odds change in the mean curtailment rate when changing the predictor by one unit. Also, few predictors are modelled as dummy variables. We are aware that dummy variables cannot increase with more than one unit, hence we cannot state directly that the highest regression coefficients in Table 5.1 are also the creating the most impact.

Variable	Estimate[7]	Z Value	P-Value
January	1.74	17.8	0.0
December	3.13	30.4	0.0
Bullet	0.64	7.2	0.0
Saving	1.88	16.3	0.0
Age Group 1	-0.69	-6.6	0.0
Age Group 3	0.45	4.5	0.0
Second Mortgage	0.15	-1.6	0.107
Income to Principal Rate	4.02	45.2	0.0
Mortgage Costs to Principal	-11.11	-86.0	0.0
Free Capital	1.29	14.3	0.0
Ever Curtailed	-18.25	-114.1	0.0
Percentage of Months Curtailed	-8.38	-58.4	0.0
Percentage Amortized	37.25	344.0	0.0
Percentage Curtailed to Principal	-3.27	-25.2	0.0
Percentage Franchise Used	-1.69	-13.4	0.0
Franchise Exceeded	1.49	12.5	0.0
Average Curtailment	4.56	35.16	0.0

Table 5.2 Results of the beta regression.

All predictors are statistically significant except for the predictor indicating whether the mortgage is a second mortgage.

For the conditional curtailment rate we notice peaks in January and December. This means that on average, a higher proportion of the outstanding debt of mortgages is curtailed. This is probably due to the same reasons for seeing higher peaks in the probability of curtailment, with December being the end of the fiscal year and January as aftereffect from the 13th month of salary received

[7] Regression coefficients are depicted as percentage of the sum of regression coefficients due to confidentiality. The order and relative magnitude between different predictors is maintained.

in December. Furthermore, compared to the middle aged class (this predictor is left out due to perfect multi-collinearity, and functions as base case) the younger aged class curtail in lower mean curtailment rates and older borrowers in higher mean curtailment rates. The differences between age groups is not extreme, because the regression coefficients are quite close.

We observe that the mortgage type bullet curtails in increased curtailment rates opposed to interest only, annuity and linear. Furthermore, whereas the saving mortgage is less likely to curtail, in case of curtailment it shows increased curtailment rates. One of the explanations we can think of for this, is saving mortgages that set their savings target amount lower than the original principal amount. Because the saving target amount is lower than original principal amount, using this to pay back debt is not full prepayment. Once the saved capital is used to pay back the mortgage debt, these are often higher volume redemptions. This might explain the high regression coefficient of the saving mortgage.

Moreover, we examine the activity-specific predictors such as ever curtailed, percentage of months curtailed, percentage franchise used, franchise exceeded and average curtailment. Interestingly, the predictor ever curtailed has the opposite effect on the conditional curtailment rate opposed to the probability of curtailment. Whereas borrowers that curtailed before show increased probability of curtailment, the mean conditional curtailment rate decreases. Next to that, higher percentage of months curtailed significantly decrease the mean curtailment rate. It makes sense that borrowers that curtail on more frequent basis do this in smaller proportions of their outstanding debt. From these dynamics, we can separate three type of borrowers. See Figure 5.1. The majority of borrowers never curtails. Amongst borrowers that curtail a smaller group of borrowers curtail in low frequency but high curtail rate and a group curtails in high frequency, but lower curtailment rates.

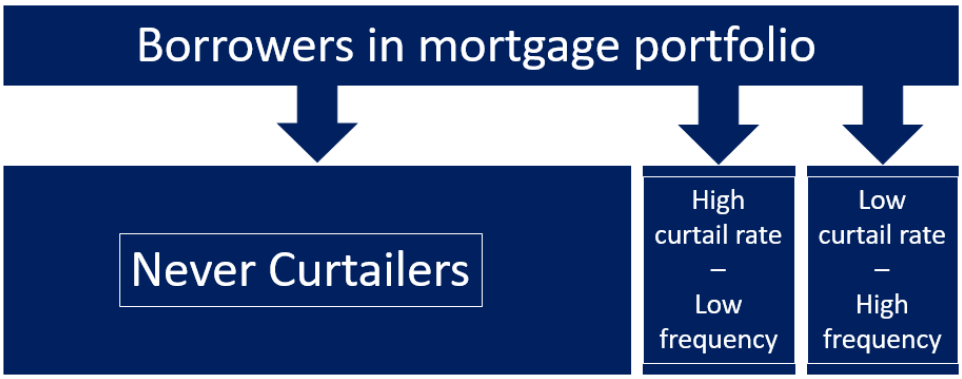

Figure 5.1 Three different borrower curtailment groups.

Furthermore, we assume that borrowers want to avoid prepayment penalties. Because of this, the negative regression coefficient of the predictor percentage franchise used is intuitive. It tells us the more volume is curtailed within a year, in case of another curtailment we should expect a small amount (hence a decreased mean curtailment rate). However, borrowers that exceeded the franchise also show also slightly increased curtailment rates. This is counter intuitive. It may be

the case that for some borrowers decreasing the mortgage debt in a short time frame is of more importance than the compensation that is charged for crossing the franchise limit. Finally, we observe that for higher average curtailment levels in the past we expect increased curtailment rates the next prepayment.

The predictor having the highest impact on the curtailment rate is the percentage amortized. The greater part of the mortgage debt that has been paid off, the higher we expect the curtailment rate to be. This makes sense because the when the percentage that is amortized is high, we assume the outstanding balance is low in comparison and keeping the curtailment amount equal, the curtailment rate will be higher. However, we should note this predictor is expressed as percentage and does not tell something about the absolute outstanding debt. For example, a mortgage with a principal amount of 10 million that has paid back 8 million and a mortgage with a principal amount of 200 thousand that paid back 160 thousand might both be amortized for 80 percent, while the outstanding debt differs significantly.

Higher incomes relative to the debt on the mortgage show increased curtailment rates. However, we notice that for higher yearly mortgages costs to principal the curtailment rate decreases more than it increases for higher income to principal. In addition, we also observe increased curtailment rates for borrowers with higher free capital when the mortgage originates. The above observations are in line with expectations.

5.3 Results cash flow estimation

In this section we examine the results of the cash flow estimations. See Table 5.3 and 5.4. Table 5.3 shows the estimated number of mortgages and actual number of mortgages that curtail, and the estimation error. Furthermore, for these groups of mortgages Table 5.3 shows the average curtailment rate and its estimation error. Next to that, Table 5.4 shows the average estimated curtailment amount per mortgage, for the simulated and actual group. Finally, Table 5.4 shows the total estimated cash flow, the total actual cash flow and the estimation error. For each out of sample month, we simulated the total expected cash flow a 1000 times. Table 5.3 and 5.4 show the average values of these 1000 runs. Using these 1000 runs, we made a 95 percent confidence interval for the total cash flow. In addition, we visualized the distribution of errors between the estimated cash flow and actual cash flow in using a histogram for the months January and March, see Figure 5.2.

	Simulated number of mortgages[8]	Actual number of Mortgages	Error (%)	Estimated average curtailment rate	Average curtailment rate	Error (%)
January	11.67	8.77	31.0	8.64	12.87	-28.6
February	5.97	5.48	7.2	8.36	7.84	13.1
March	5.98	5.56	5.9	8.43	7.54	19.1
April	5.94	5.69	2.8	8.57	7.39	24.4
May	6.14	7.19	-15.9	8.14	7.39	17.3
June	6.34	5.99	4.1	8.14	7.16	21.5
July	6.54	6.87	-6.3	8.07	7.62	12.8
August	6.42	6.23	1.6	8.14	7.62	14.5
September	6.50	7.37	-13.1	8.14	7.92	10.0
October	9.27	8.18	11.6	8.07	8.07	6.9
November	10.68	9.33	12.7	8.14	8.61	0.7
December	18.55	23.34	-21.7	9.14	9.98	-2.6

Table 5.3 Results of the out of sample cash flow estimation (1)

Seeing the cash flow estimation results, we notice a few things. First of all, for the number of mortgages that we estimate to curtail, we see quite high fluctuation over all months where the standard deviation equals 13.8 percent. However, if we review the average performance over the complete year, the estimate of the total number of mortgages that curtail only deviates from the actual results by 1.54 percent. Furthermore, we notice that the average curtailment rate is a overestimation compared to the actual result, except for January and December. Also, the estimated absolute curtailment amount per mortgage is remarkably low compared to the actual data. If we look at the dynamics between the different estimations errors (the error for the number of mortgages, average curtailment rate, average curtailment amount and the total cash flow) we notice that especially the combination of the number of mortgages and the average curtailment amount influences the total cash flow error.

[8] All column cells of Table 5.3 and 5.4 are displayed as percentage of the total sum of that column due to confidentiality. Order and relative magnitude between rows is maintained.

	Estimated average curtailment amount	Average curtailment amount	Error (%)	Estimated total cash flow	Total cash flow	Mean Error (%)	95 % CI - lower	95 % CI - upper
January	9.21	10.09	-24.8	12.65	9.90	-1.5	-4.1	1.0
February	8.05	7.23	-8.3	5.65	4.43	-1.7	-5.6	2.3
March	8.05	7.34	-9.7	5.67	4.57	-4.3	-8.1	-0.7
April	8.07	7.55	-11.9	5.64	4.80	-9.4	-12.9	-5.7
May	8.08	6.98	-4.6	5.84	5.61	-19.8	-22.7	-16.6
June	8.06	7.56	-12.2	6.01	5.07	-8.6	-11.7	-5.0
July	8.11	8.08	-17.3	6.25	6.21	-22.5	-25.6	-19.5
August	8.08	7.94	-16.1	6.11	5.73	-17.8	-20.8	-14.5
September	8.13	8.13	-17.6	6.22	6.70	-28.4	-31.4	-25.7
October	8.21	8.69	-22.3	8.96	7.96	-13.2	-15.7	-10.4
November	8.24	9.11	-25.5	10.35	9.51	-16.1	-18.3	-13.5
December	9.71	11.29	-29.2	20.64	29.50	-44.5	-45.7	-43.4

Table 5.4 Results of the out of sample cash flow estimation (2)

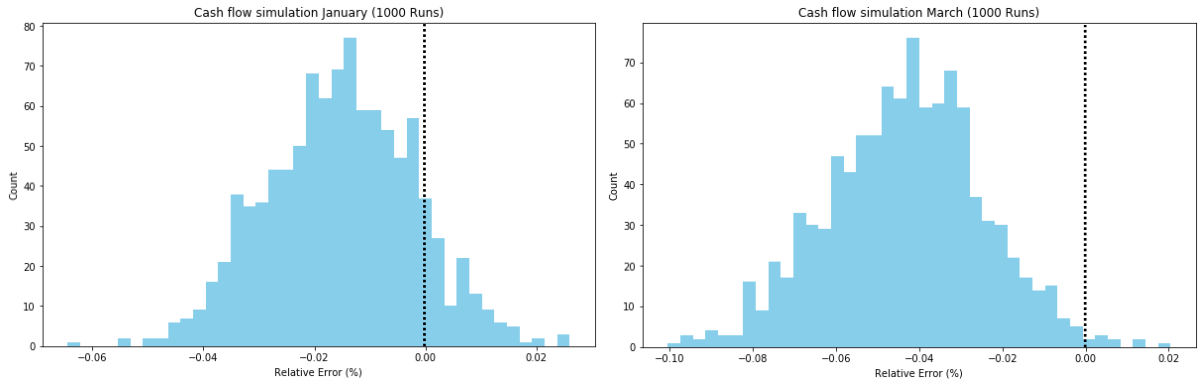

Figure 5.2 Histogram of the error between estimated and actual cash flow for January and March

In the end, we want to estimate the total cash flow correctly. We see that this error grows quite large, especially towards the end of the year. The total cash flow error is a result of the errors from both models.

To understand the total cash flow errors, we dig a little deeper into estimation errors of both models separately. We can assess the performance of the logistic regression model on two aspects: how well it estimates the absolute number of mortgages that curtail and how well the model classifies true mortgages that curtail.

Currently we assess the model on the absolute number of mortgages that curtail (see the results in Table 5.3.) If we look closer to the simulated group of mortgages, we notice that on average over all out-of-sample months, about 30 percent of the mortgages simulated are mortgages that in reality curtailed. Therefore, we can say that the logistic model in fact produces two errors: a deviation in the *number* of mortgages that we expect to curtail and a deviation from the group of mortgages that *truly curtail.*

To understand why the latter also influences the eventual cash flow estimation error, we look at the differences between mortgages that curtailed in 2019 and mortgages that did not. First of all, the average outstanding debt of mortgages that curtail is almost always higher than mortgages

that do not curtail. See Figure 5.3. Therefore, when the logistic model does not classify the mortgages that curtail correctly, the majority of mortgages simulated (about 70 percent) will have a lower outstanding debt on average. It is obvious that this deviation directly increases the cash flow estimation error. Furthermore, for all predictors selected for the beta regression model, the average values are different for the mortgages that do not curtail opposed to mortgages that do curtail. See Table C.1.1and C.1.2 in Appendix C1.

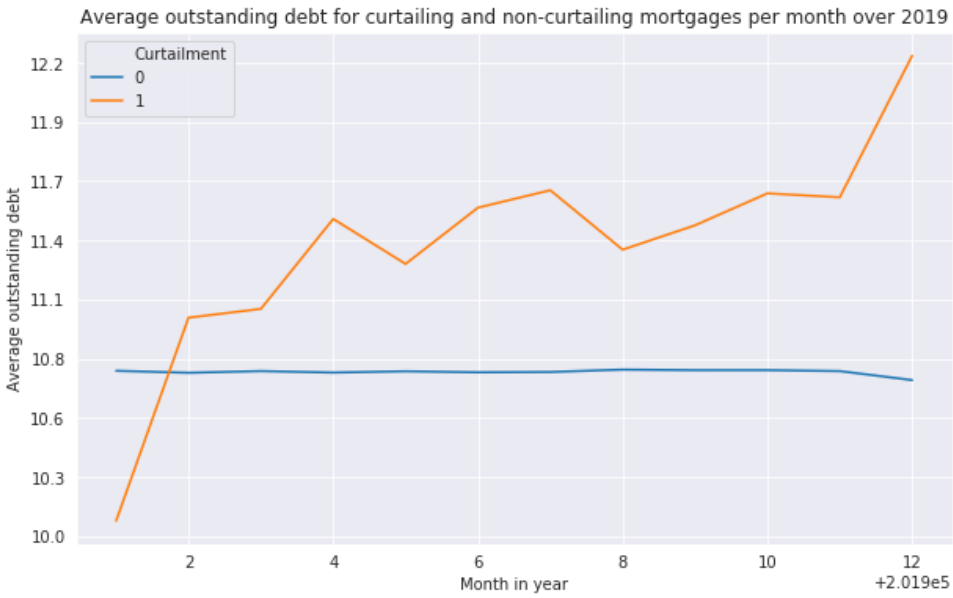

Figure 5.3 The average outstanding debt of curtailing and non-curtailing mortgages in 2019. [9]

Average values for the predictors such as ever curtailed, percentage of months curtailed, percentage franchise used, percentage curtailed to principal have a negative regression coefficient, meaning that the curtailment rate model generally assigns lower curtailment rates to mortgages that curtail in reality. Because the simulated group of mortgages contains far more mortgages that in reality do not curtail and the curtailment rate model assigns higher curtailment rates to these mortgages, the average conditional curtailment rate is higher in the simulation than the individual model performance.

The estimation errors on the number of mortgages and failing to classify the right mortgages, influence the estimation error of the beta regression model. To put it differently, the performance of the beta regression model is sensitive to mortgages with varying outstanding debts. To illustrate why this is the case, we show the cash flow estimations of the beta regression model when estimating on mortgages that actually did curtail.

[9] Values on the y-axis are displayed as percentage of the sum of the total axis, due to confidentiality. The order and relative magnitude between different predictors is maintained.

Period	Estimated average curtailment rate	Average curtailment rate	Error (%)	Estimated average curtailment amount	Average curtailment amount	Error (%)	Estimated total cash flow	Total cash flow	Error (%)
January	12.05	12.90	-1.85	9.78	10.09	-1.73	9.67	9.90	-1.73
February	8.21	7.85	9.93	7.60	7.23	6.59	4.70	4.43	6.59
March	7.94	7.53	10.84	7.59	7.34	4.85	4.76	4.57	4.85
April	7.86	7.38	12.03	7.81	7.55	5.02	5.01	4.80	5.02
May	7.67	7.42	8.61	7.45	6.98	8.26	6.04	5.61	8.26
June	7.50	7.13	10.56	7.82	7.56	4.90	5.29	5.07	4.90
July	7.68	7.63	5.73	8.03	8.08	0.74	6.22	6.21	0.74
August	7.74	7.59	7.13	7.95	7.94	1.55	5.78	5.73	1.55
September	7.86	7.90	4.45	8.09	8.13	0.93	6.72	6.70	0.93
October	7.75	8.09	0.67	8.24	8.69	-3.90	7.60	7.96	-3.90
November	8.10	8.60	-0.97	8.52	9.11	-5.22	8.96	9.51	-5.22
December	9.64	10.00	1.23	11.12	11.30	-0.18	29.26	29.50	-0.18

Table 5.5 Cash flow estimation errors of beta regression model on actual group of mortgages.[10]

Obviously, the estimation error of the total cash flow is improved significantly. It is interesting to see that that in this case, the total cash flow estimation error is equal to the average curtailment amount error. This stresses how sensitive the model is for varying outstanding debts. Although the model does not incorporate absolute outstanding debts, it does indirectly capture outstanding debts. This is because the model does include the predictor "percentage amortized", which assigns higher curtailment rates to mortgages that are amortized more. Since it is likely that mortgages that amortized for a high percentage also have smaller outstanding debts, the model does adjust its estimation indirectly based on the outstanding debt. However, it does not directly include the absolute outstanding debt when estimating the curtailment rate. Simply adding the absolute outstanding debt as predictor does not make sense, since it should on itself not have an assumed effect on the curtailment rate. Because the cash flow estimations improve significantly when classifying and prediction on the true mortgages, this illustrates the importance of good classification when using the combination of models that we do in this research.

[10] All column cells of Table 5.5 are displayed as percentage of the total sum of that column due to confidentiality. Order and relative magnitude between rows is maintained.

6. Conclusions

In this chapter we conclude on all findings in this research. The main research goal is to find determinants that affect curtailments of Dutch residential mortgages. In addition, the sub research goal is to investigate how selected determinants and selected models contribute to cash flow modelling.

Curtailments follow a seasonal effect, where the number of curtailments peaks in January and the highest peak in December. We find that this seasonal effect holds for the curtailment rate as well, where higher proportions of outstanding debt are curtailed in January and December.

Different amortization types show different effect on the probability of curtailment. Compared to annuity mortgages, linear mortgages show an increase in curtailment but saving, interest only and bullet mortgages a decrease. For saving mortgages, we see reduced probability of curtailments caused by high numbers of voluntary deposits to mortgage saving accounts.

Furthermore, the middle aged are most likely to curtail. The probability decreases for young borrowers, and even more for the elderly. However, the mean curtailment rates keeps increasing a little from younger, to middle aged and elderly borrowers. In addition, we find the predictor higher mortgage costs to principal rate to have the strongest impact on the probability of curtailment. Not only do higher mortgage costs decrease the likeliness of curtailment, also in case of curtailment the proportion of outstanding debt that is paid off is smaller.

In contrast to previous literature, we focussed on activity-specific variables. Studying activity-specific determinants both from the perspective of probability and the curtailment rate led to additional insight in curtailment behaviour. We conclude that higher frequency of curtailment in the past leads to higher probability of future curtailment. However, the higher the curtailment frequency the lower we should expect the proportion of outstanding balance that is paid off to be. Furthermore, borrowers that ever curtailed are far more likely to curtail, but show decreased curtailment rates if they prepay again. From these dynamics, we separated three type of borrowers (Figure 5.1). The first group of borrowers, which is the majority, does not curtail. Next to this group, we observe borrowers that curtail in high rate but low frequency and borrowers that curtail in low rate but high frequency.

We used a multivariate logistic regression to model the probability of curtailment. If we assess the performance from a classification perspective, the logistic model fails to identify high rates of true curtailments. In general, on monthly portfolios the logistic model classifies about 30 to 40 percent of true curtailments. If we assess the performance on estimating the percentage of mortgages that curtail on portfolio level, the average estimation error on total number of mortgages that curtail on yearly basis over out of sample months is 1.34 percent. However, the estimation errors do fluctuate over out of sample estimations, where the standard deviation is 13.8 percent.

In this paper we introduced the beta regression as an application to the curtailment rate. The beta regression does provide additional insight in the researched predictors. However, the curtailment rate does not reflect differences in outstanding debts of mortgages. Because of this, using the beta regression to estimate cash flows may result in high estimations on cash flow level particularly when estimating on a portfolio with high variety of outstanding debt.

In addition, borrowers that curtail show differences from borrowers that do not curtail. This is especially observable if we compare the summary statistics for predictors of the curtailment rate for mortgages that curtailed and did not. The beta regression model to assigns higher curtailment rates to mortgages without curtailment history and lower curtailment rates to mortgages with curtailment history. Another difference is that for curtailment observations, the average outstanding debt is higher. Because of these differences and because the beta model does not incorporate the outstanding debt when doing estimations, the low classification performance from the logistic model increases the estimation errors on cash flow level. We conclude that for the combined use of these models, the classification of true curtailments is not accurate enough to estimate well on cash flow level.

6.1 Limitations & recommendations for further research

In this section we reflect on shortcomings of this research, this that we did like to do differently and recommendations for further research.

First of all, the main research question focussed on broadening the scope of determinants to add to literature. Looking back on the research design at this point, we might have put more effort in broadening the selection criteria for choosing the statistical models but also focussing more on the advantages and disadvantages of using combined models for the purpose of cash flow modelling. The current selection criteria were mainly focussed on descriptive information of determinants, but less on predictive or classification qualities.

In this research we developed two models, one for the probability and one for the conditional rate. Both made use of a logit link function. The reason for this choice is enhanced interpretability of the regression parameters. Besides the logit link function, there are other link function such as the probit link, the complementary log-log link, the log-log link and the Cauchy link. These make interpretation of the regression coefficients difficult, but might result in a better fit with the data. For example, we noticed that for the beta model the log-log link resulted in an improved fit of data. Therefore, if interpretation is not of primary interest, it may be worth to see if other link functions that better fit the data can result in improved prediction accuracy.

Furthermore, we focused on determinants that influence the probability of curtailment. We observe that not only classifying the right number of mortgages that curtail but also the right mortgages that truly will improve the cash flow prediction. Therefore, it might be worth to look if advanced classification methods, which are primarily used to classify the right mortgages that will curtail in combination with a beta regression model can improve cash flow estimations.

We studied determinants and both models assumed that the determinants researched are linear predictors to the target variable. Therefore, a shortcoming is that we did not capture all non-linear effects of determinants. For example, the determinants percentage of months be related exponentially to the probability of curtailment. Also the percentage franchise used, might be negative exponentially related to the curtailment rate.

Lastly, this thesis describes three curtailment borrower types that show different behaviour in terms of frequency of curtailment and their conditional curtail rate when curtailing. These three types are never curtailing mortgages, mortgages that made one curtailment and mortgages that curtailed more than once. Segmenting mortgages into three segments before modelling might improve prediction accuracy for predicting the probability, the conditional rate and the total cash flow.

References

Adelman, S., Cross, M., & Shrider, D. (2010). Why do homeowners make mortgage curtailment payments?. *J. Hous. Res.*

Budinger, M., & Fan, Y. (1995). Rfc whole loan prepayment behavior. *The Handbook of Mortgage Backed Securities Chapter, vol. 13*, pp. 34–51.

De Nederlandsche Bank. (2020). *Mortgage lending to Dutch households, adjusted for securisations*. Retrieved from https://www.dnb.nl/statistieken/dashboards/woninghypotheken/

Ferrari, S., & Cribari-Neto, F. (2004). Beta regression for modelling rates and proportions. *Journal of Applied Statistics*, 31(7), 799-815.

Fu, Q. (1997). Retiring Early: An Empirical Analysis of the Mortgage Curtailment Decision (PhD thesis). *University of Wisconsin*.

Hastie, T., Tibshirani, R., Witten, D., & James, G. (2013). An Introduction To Statistical Learning .

Hayre, L. (2003). Prepayment Modeling and Valuation of Dutch Mortgages. *The Journal of Fixed Income*.

Isbasoiu, A. E. (2013). Drivers of mortgage prepayment: An application to Dutch data.

Kuang, W., Liu, C., & Wu, Q. (2019). How do interest changes affect mortgage curtailments? Evidence from China. *Real Estate Economic*.

Lin, C., & Yang, T. (2005). Curtailment as a mortgage performance indicator. *J. Hous. Econ. 14*.

Lin, C., Chu, T., Prather, L., & Wang, P. (2005). Mortgage curtailment and default. *Int. Real Estate Rev*.

McCollum, N., Lee, H., & Pace, R. (2015). Deleveraging and mortgage curtailment . *Journal of Banking & Finance* .

Senaviratna, N., & Cooray, T. (2019). Diagnosing Multicollinearity of Logistic Regression Model.

Stam, R. v. (2015). Application of the Multinomial Logit Method to Model Prepayment Risk.

Vasconcelos, P. (2010). Modelling prepayment risk: multinomial logit model approach for assessing conditional prepayment rate.

Appendix A.1

In this section, Appendix A.1, the remaining variables that are considered for modelling the probability of curtailment can be found with their description and expected impact on the probability of curtailment.

Ever Curtailed

This variable checks whether the borrower has ever prepaid since the origination of his mortgage. *Fu (1997)* found that past curtailment increases the probability of future curtailment. Additionally, *Lin and Yang (2005)* found that past curtailment increases future curtailment by 23 percent.

Borrower moving of residential place

Borrowers can decide to move to another residence while keeping the same mortgage. This variable indicates whether a borrower moved in the past year or not. Once a borrower moved, it can either redeem his old mortgage with the money received from the sale of his old house or he can take his old mortgage and use it (partly) for his new house. The latter case is the most likely if the old mortgage contains some benefits for the borrower such as a relative low interest rate or on the contrary relative high interest costs that can be tax-deducted. Transferring the mortgage contract to his new house may indicate relatively favorable terms. Curtailing on this mortgage therefore will only reduce those benefits, hence we expect this variable to be negatively correlated with the probability of curtailment.

Loan age

This variable keeps track of the number of months that have passed since the mortgage is originated, in other words the aging of the loan. *Stam(2015)* find a positive relationship between the aging of the loan and an increase in the probability of curtailment.

Current Loan to Value (categorical)

LTV is a recurring variable that literature (for example, *McCollum et al. (2015)*) points out to be correlated with the probability of curtailment. Collateral-specific variables are not available in the dataset, however the LTV that is updated throughout the lifetime of the mortgages is available as categorical variable. The LTV is available on loan-level and divided in three buckets: an LTV below 67.5, between 67.5 and 90 and above 90. A high LTV ratio is expected to reduce the probability of curtailment, whereas a low LTV is expected to increase the probability of curtailment.

Age

Depending on the borrowers age, different effect on the probability of curtailment is expected. For example, younger borrowers are more likely to have recently bought their house. For starters, it is unlikely that large curtailments are done (since if the money was available, it would be more logical to take a lower mortgage loan). The middle aged are in the most unstable period of their lifetime, since switching jobs and therefore moving but also getting children and or getting divorced can

impact curtailment decisions. Lastly, we expect older borrowers to be in the most stable phase concerning their private situation and older age could be associated with larger financial saving buffers thus the funds to curtail.

Location of collateral: province

Possible factors that influence curtailment are the change in value of the collateral and also whether the borrower has intentions to move. Provinces in the Netherlands show difference in responsiveness to fluctuations in the average housing price index in the Netherlands. According to the statistics of the Central Office for Statistics in the Netherlands, over the past 5 years the house price index rose above average in the provinces Noord-Holland, Zuid-Holland and Utrecht when the average house price index over all provinces rose. When the collateral is located in a province that show above average increase in the house price index when the overall house price index rises, a curtailment can be seen as a relatively attractive investment when putting equity in the home. Concluding, we expect provinces with above average increase in house price indexes to be positively correlated to the probability of curtailment.

Default status

When the borrower missed a monthly payment for longer than 90 days, the borrower enters default status. When a borrower is missing payments on the amortization of the mortgage, we assume that in any case there is no discretionary income left to make curtailments on the mortgage. Hence, the expectation is that this variable negatively impacts the probability a curtailment.

Percentage of principal curtailed year to date

We expect borrowers that curtail frequently to do this in smaller amounts. Especially within a range of one year, we do not expect borrowers to curtail multiple times in higher volumes. Therefore, we expect a high percentage of principal used to be negatively correlated with the probability of curtailment.

Curtailed last quarter

This variable tracks whether the borrower curtailed in the previous quarter, thus the previous three months. This variable attempts to capture recent curtailment activity of the client. We are interested whether borrowers that have been active in recent periods show higher probability of curtailment in the next month.

First time buyer

We expect that first time buyers are less experienced with mortgage structures and have less financial knowledge to know when to effectively curtail. Therefore, we expect this variable to be negatively correlated with the probability of curtailment.

Number of borrowers

Mortgages can be taken out by different number of people. Most likely within residential private mortgages, the number of borrowers linked to a mortgage is either a single borrower or a couple. Once the number of borrowers on the mortgage increases the aggregated income increased as well, hence this could increase financial room for curtailment. We are interested if higher number of borrowers attached to a mortgage results in a higher probability of curtailment.

End of fixed interest period

Between fixed interest rate periods a mortgage can be seen as risk free since there is no determined interest rate. Therefore, at the end of the fixed interest rate period the borrower is allowed to prepay the mortgage up to any amount. We are interested if the month in which the fixed interest period ends is positively correlated with the probability of curtailment.

Time to interest rate reset

The agreed interest rate is fixed for a certain period for fixed rate mortgages. Borrowers can curtail up to any amount in between fixed interest rate periods. Therefore, it is unlikely borrowers will curtail shortly after a new fixed interest period starts. Following this logic, we are interested if longer times to the interest rate reset increase the probability of curtailment.

Seasonality

According to previous literature, the month December is correlated with high numbers of curtailment. This is due to the fact that people often receive a 13th month of income which creates a financial opportunity to make an early redemption. In addition, seeing the tax system it can be beneficial to include a curtailment for the same tax year to reduce the tax burden of that year. To be more precise, Box 3 of the Dutch tax system taxes the income from savings and investments. If aggregated income in this box falls below a certain threshold, these assets are tax free. However, an early redemption can be subtracted from the income in Box 3. Hence, once an individual find himself slightly above the tax-free threshold, it can be beneficial to make an early redemption on the mortgage.

Rented House

Mortgages can be taken out with the purpose of renting it out. Rented houses often produce frequent recurring revenue hence this mortgage loan could be associated with more curtailments schedule compared to normal residential mortgages.

Unemployment rate

The unemployment rate serves as indicator for economic situations where average levels of income can decrease. With decreasing average level of income, there will be less room for financial opportunity to curtail on the mortgage. Additionally, dropping unemployment rates can lead to missed regularly redemption payments on the mortgage and get borrowers into default. In both

case, this will negatively impact the probability of a curtailment. The historical unemployment rate is available on quarterly basis.

House Price Index

Curtailment is a form of deleveraging behavior from the mortgagee. The leverage is influenced by the outstanding balance of the loan and the value of the collateral. For linear and annuity mortgages, the outstanding balance naturally decreases over time hence the leverage on the mortgage decreases as well. When the house price index rises, the value of the collateral most likely increases as well and the leverage decreases. Since we expect mortgage holders that pursue curtailments preference low leverage, the increase of the house price index can work as an incentive to curtail to decrease the leverage even more. Hence, the expectation is that an increase in house price index increases the probability of curtailment.

Appendix A.2

In this section, Appendix A.2, the remaining variables considered for modelling the volume of curtailment can be found with their description and expected effect on the volume of curtailment.

Free Capital at origination

When a mortgage is taken out for a client, the bank documents the free capital the client owns. The value of this free capital is not updated over the lifetime of the mortgage, therefore we call it free capital at origination. We realize that this information is relatively limited, since it only reflects the situation of the client at origination and because lifetimes of mortgages typically last between 20 and 30 years. However, we expect higher free capital at origination to be positively correlated with higher volumes of curtailment.

Second mortgage

When a homeowner takes on a second mortgage, this could give more burden to the borrowers monthly mortgage costs are higher compared to borrowers that own a single mortgage. Hence, we expect a loan that is taken out on a second mortgage to curtail in lower rates.

Last Curtailment

First, think about borrowers that never have curtailed. Here we do not have any information, thus de curtailment can either be high or low. However, once the client curtailed the first curtailment is either a relatively high or low amount. We assume that borrowers that make frequent curtailments do this in low amounts; hence if the last curtailment is low we expect the next one to be low as well. In addition, if the client made a high curtailment we expect that borrowers do not make high curtailments on frequent basis, therefore we expect that this also will result in a lower volume curtailment next time. Therefore, we expect that either low and high curtailments done in the past will be correlated with lower volumes.

Average curtailment

This variable has overlap of information with determinant "last curtailment". However, the advantage that is gives more information for borrowers that have a richer curtailment history. For borrowers that curtail more often, it gives a good indication about the average volume that should be expected once they make an additional curtailment. Following this reasoning, borrowers that on average curtail a high amount will likely do so next time. Hence, high average curtailments are expected to be correlated with higher volumes of curtailment.

Highest curtailment

This variable also contains overlapping information from both the average curtailment and last curtailment. If the highest curtailment is high compared to the average curtailment, this can be an indication that it was a one-time redemption. When the highest curtailment is low compared to average curtailment, this can be an indication that the borrower curtails more frequently in smaller amounts.

Percentage Curtailed to Principal YTD

This variable captures the same logic as the variable percentage curtailed to principal, but describes more recent activity of the client. Especially within the same year, we do not expect borrowers to curtail in high volumes in high frequency. Therefore, we expect a high percentage curtailed to principal year to day to be negatively correlated with a high volume curtailment in the next measurement period.

Appendix A.3

	Static continuous	Static categorical	Time-varying deterministic	Time-varying stochastic
Borrower-specific				
Age				
Number of borrowers				
Province: location of collateral				
Move indicator				
First Time Buyer				
Free Capital				
Mortgage costs to income				
Activity-specific				
Default status				
Percentage of months curtailed				
Curtailed last quarter				
Percentage of principal curtailed Year To Date				
Loan-specific				
Annuity mortgage				
Bullet mortgage				
Linear mortgage				
Interest Only mortgage				
Saving mortgage				
Interest rate type				
End of fixed interest rate period				
Time to interest rate reset				
Loan age				
Current LTV				
Interest rate incentive				
Macroeconomic				
Unemployment rate				
House price index				
General				
Seasonality				

Table A.3.1 Variables considered for modelling the probability of curtailment.

	Static continuous	Static categorical	Time-varying deterministic	Time-varying stochastic
Borrower-specific				
Age				
Free capital				
Income to principal rate				
Second Mortgage Ind				
Activity-specific				
Percentage amortized				
Percentage franchise used				
Franchise Exceeded Ind				
Income to principal rate				
Ever curtailed				
Percentage of months curtailed				
Last curtailment				
Average curtailment				
Highest curtailment				
Percentage curtailed to principal				
Percentage curtailed to principal YTD				
Loan-specific				
Annuity				
Bullet				
Linear				
Interest Only				
Saving				
Franchise rate				
(Yearly) Mortgage costs to principal rate				

Table A.3.2 Variables considered for modelling the volume of curtailment.

Appendix B.1

Variable	Coefficient	P-Value
January	0.9629	0.00
March	-0.9659	0.00
First Quarter	-1.5610	0.00
Second Quarter	-1.9787	0.00
Third Quarter	-0.6705	0.00
October	0.3434	0.00
November	0.6603	0.00
December	3.2463	0.00
Interest Type	0.4167	0.00
Loan Age	2.7674	0.00
Log Loan Age	3.6202	0.00
Annuity	-1.9777	0.00
Bullet	-0.7438	0.00
Interest Only	-0.2741	0.00
Linear	2.0480	0.00
Saving	0.8661	0.00
Fixed Interest Rate Period	0.2466	0.00
Time to Interest Reset	-0.8518	0.00
Randstad Ind	0.8528	0.00
Number of Borrowers	0.4983	0.00
Age Group 1	-0.3953	0.00
Age Group 2	1.0862	0.00
Age Group 3	-1.1463	0.00
First Time Buyer	0.7173	0.00
Move Indicator	-0.8263	0.00
Second Mortgage	0.0275	0.271
Rented House	0.9058	0.00
(Yearly) Mortgage Costs To Principal	-25.9845	0.00
Free Capital	0.5675	0.00
Interest Incentive	1.2594	0.00
LTV	-2.0195	0.00
In Default	-0.7509	0.00
Unemployment Rate	-4.4058	0.00
House Price	3.0792	0.00
Ever Curtailed	6.6077	0.00
Percentage Principal Curtailed YTD	3.0598	0.00
Curtailed Last 12 Months	11.6748	0.00
Curtailed Last Quarter	4.5994	0.00
Percentage of Months Curtailed	5.3341	0.00

Table B.1.1 Univariate analysis on variables for the probability model standardized

Appendix B.2

Variable	VIF	VIF first adjustment	VIF second adjustment
January	Infinite	1.442	1.342
February	Infinite	-	-
March	Infinite	1.115	1.319
First Quarter	Infinite	-	-
Second Quarter	Infinite	1.994	1.485
Third Quarter	Infinite	1.893	1.456
October	Infinite	1.257	1.179
November	Infinite	1.258	1.177
December	Infinite	1.313	1.211
Interest Type	1.115	32.018	1.087
Log Loan Age	4.738	11.694	1.326
Annuity	Infinite	-	-
Bullet	Infinite	1.018	1.012
Interest Only	Infinite	2.180	1.446
Linear	Infinite	1.227	1.119
Saving	Infinite	1.462	1.216
Fixed Interest Rate Period	8.755	-	-
Time to Interest Reset	8.942	4.211	1.156
Randstad Ind	1006	1.651	1.006
Number of Borrowers	1.049	13.260	1.042
Age Group 1	Infinite	1.592	1.350
Age Group 2	Infinite	-	-
Age Group 3	Infinite	1.246	1.112
First Time Buyer	1.141	1.610	1.372
Move Ind	1.185	1.387	1.177
Second Mortgage	1.001	-	-
Rented House	1.009	1.015	1.007
Mortgage Costs To Principal	1.078	1.768	1.067
Free Capital	1.009	1.168	1.009
Interest Incentive	1.415	4.557	1.181
LTV	1.078	1.200	1.038
In Default	1.002	1.006	1.001
Unemployment Rate	17.211	11.589	-
House Price	17.211	-	-
Ever Curtailed	2.197	2.254	1.601
Percentage Principal Curtailed YTD	1.837	1.879	1.304
Curtailed Last 12 Months	3.339	3.548	-
Curtailed Last Quarter	1.476	1.496	1.400
Percentage of Months Curtailed	1.685	1.590	1.510

Table B.2.1 VIF values variables from probability model

Appendix B.3

Variable	Coefficient	Z-Value	P-Value
January	2.49	38.69	0.00
December	1.93	29.57	0.00
Annuity	0.64	12.249	0.00
Bullet	0.46	92.492	0.00
Interest Only	0.55	10.6	0.00
Linear	0.44	8.517	0.00
Saving	-1.92	-36.311	0.00
Client Age I	-1.80	-33.992	0.00
Client Age II	0.66	12.488	0.00
Client Age III	1.48	28.583	0.00
Number of Borrowers	-0.67	-12.876	0.00
Second Mortgage Ind	0.31	61.939	0.00
Rent House Ind	0.08	14.626	0.144
Unemployment Rate	1.00	19.263	0.00
Franchise Rate	1.24	23.318	0.00
Income to Principal Rate	2.40	54.213	0.00
Mortgage Costs to Principal	-7.71	-146.31	0.00
Free Capital	1.30	26.103	0.00
Ever Curtailed	-3.45	-66.154	0.00
Percentage of Months Curtailed	-5.88	-109.97	0.00
Percentage Amortized	17.58	396.17	0.00
Percentage Curtailed to Principal	3.14	61.853	0.00
Percentage Franchise Used	1.11	21.314	0.00
Franchise Exceeded Ind	7.26	152.24	0.00
Percentage Curtailed YTD	5.35	109.8	0.00
Last Curtailment	10.43	221.89	0.00
Average Curtailment	10.23	215.27	0.00
Highest Curtailment	8.48	177.04	0.00

Table B.3.1 Univariate analysis variables volume model [11]

Appendix B.4

Variable	VIF	VIF first adjustment	VIF second adjustment
January	Infinite	1.193	1.381
December	Infinite	1.338	1.420
Other months	Infinite	-	-
Age Group I	Infinite	1.230	1.208
Age Group II	Infinite	-	-
Age Group III	infinite	1.147	1.095
Annuity	infinite	-	-
Bullet	infinite	1.004	1.001
Interest Only	infinite	-	-
Linear	infinite	-	-
Saving	infinite	1.731	1.233
Second Mortgage indicator	1.000	1.001	1.001
Income to principal rate	1.045	1.240	1.223
Mortgage costs to principal rate	1.750	4.629	2.674
Free Capital	1.007	1.206	1.181
Number of Borrowers	1.028	14.028	-
Loan Age	1.965	26.457	-
Unemployment rate	1.293	20.652	-
Franchise Rate	1.217	56.113	-
Ever curtailed	2.202	4.980	4.013
Percentage Amortized	2.259	4.563	3.317

[11] All column cells are displayed as percentage of the total sum of that column due to confidentiality. Order and relative magnitude between rows is maintained.

Percentage Curtailed to Principal	2.271	2.979	2.876
Percentage of Months Curtailed	1.878	2.973	2.117
Percentage Franchise Used	2.275	3.024	3.003
Franchise Exceeded Indicator	2.773	3.006	2.988
Average Curtailment	3.111	4.044	3.920

Table B.4.1; VIF values from variables of volume model

Appendix B.5

Variable	**No curtailment**		**Curtailment**		**All**	
	Mean	**Std. Dev**	**Mean**	**Std. Dev**	**Mean**	**Std. Dev**
January	0.694	0.879	0.082	0.981	0.615	2.451
March	0.067	0.898	0.045	0.727	0.657	2.517
First Quarter	2.282	4.299	1.560	3.840	2.261	4.148
Second Quarter	2.232	4.270	1.412	3.699	2.219	4.120
Third Quarter	1.990	4.106	1.560	3.840	1.970	3.961
October	0.560	2.415	0.565	2.500	0.557	2.330
November	0.568	2.435	0.654	2.670	0.565	2.358
December	0.711	2.686	1.664	3.935	0.715	2.610
Interest Type	8.201	1.324	7.326	1.123	8.145	1.277
(Log) Loan Age	24.393	9.390	23.865	7.841	24.270	9.050
Annuity	3.645	4.792	2.548	4.482	3.616	4.614
Bullet	0.033	0.618	0.007	0.340	0.033	0.597
Interest Only	2.792	4.550	2.385	4.406	2.776	4.390
Linear	0.769	2.782	1.248	3.529	0.765	2.694
Saving	1.112	3.285	1.219	3.491	1.114	3.169
Randstad Indicator	3.143	4.676	3.106	4.652	3.125	4.511
Number of Borrowers	14.052	4.492	12.661	4.303	13.972	4.334
Age Group 1	1.321	3.517	1.070	3.312	1.305	3.393
Age Group 2	6.094	4.289	5.743	3.954	6.059	4.138
Age Group 3	0.945	3.053	0.609	2.585	0.939	2.945
First Time Buyer	1.187	3.372	1.248	3.529	1.180	3.253
Move Indicator	1.346	3.546	0.988	3.208	1.338	3.421
Rented House	0.069	0.869	0.163	1.378	0.070	0.848
(Yearly) Mortgage Costs To Principal	0.368	0.193	0.267	0.179	0.366	0.186
Interest Incentive	19.887	13.564	18.969	10.766	19.790	13.095
LTV	1.028	3.178	0.528	2.434	1.022	3.057
In Default	0.033	0.628	0.007	0.359	0.033	0.606
Ever Curtailed	0.201	1.468	4.198	4.671	0.224	1.529
Percentage Principal Curtailed YTD	0.109	2.966	0.802	1.283	0.108	2.852
Curtailed Last Quarter	0.134	1.217	2.378	4.397	0.150	1.258
Percentage of Months Curtailed	0.033	0.242	1.122	1.585	0.042	0.289
Time to Interest Reset	20.85	12.95	19.81	12.59	20.84	12.95
Free Capital	8.41	22.43	11.83	26.44	8.44	22.46

[12]Table B.5.1 Summary statistics of determinants for probability model.

[12] All column cells are displayed as percentage of the total sum of that column due to confidentiality. Order and relative magnitude between rows is maintained.

Appendix B.6

Variable	All		Curtailment rate	
	Mean	Std. Dev	Mean	Std. Dev
January	2.099	5.221	6.720	6.242
December	2.723	5.876	4.822	4.453
Annuity	9.872	9.449	4.746	5.094
Bullet	0.034	0.715	8.694	8.165
Interest Only	8.851	9.191	4.708	4.926
Linear	4.851	7.464	4.632	5.094
Saving	4.709	7.385	2.999	4.420
Age Group 1	4.170	7.027	3.379	4.318
Age Group 2	21.929	8.298	4.518	5.027
Age Group 3	2.241	5.360	5.467	5.229
Second Mortgage Ind	0.025	0.586	6.302	5.466
Mortgage Costs to Principal	1.021	0.814	4.404	4.960
Ever Curtailed	15.858	9.846	4.138	5.128
Percentage of Months Curtailed	4.255	3.355	6.302	4.960
Percentage Amortized	7.347	4.625	4.404	4.960
Percentage Curtailed to Principal	2.496	3.077	4.404	4.960
Percentage Franchise Used	3.886	4.645	4.404	4.960
Franchise Exceeded	2.355	5.459	10.554	6.680
Average Curtailment	1.277	1.608	4.404	4.960
Free Capital	31.154	68.846	4.218	4.725
Income to Principal Rate	35.948	57.187	4.385	4.815

Table B.6.1 Summary statistics variables for volume model

Appendix C.1

Mortgages that made a curtailment in 2019										
Period	Outstanding Debt	Bullet Ind	Income To Principal Rate	Free Capital	Ever Curtailed	Percentage Months Curtailed	Percentage Curtailed To Principal	Percentage Franchise Used	Franchise Exceeded Ind	Average Curtailment
201901	7.39	18.42	9.47	4.63	9.12	6.94	8.65	6.70	16.46	15.15
201902	8.09	10.53	8.76	4.18	8.66	9.25	8.17	8.25	10.98	9.09
201903	8.12	5.26	8.70	4.19	8.78	9.83	8.65	8.25	9.15	7.58
201904	8.43	2.63	8.58	4.47	8.78	9.83	8.65	8.76	9.15	7.58
201905	8.27	7.89	8.23	4.10	8.66	8.67	8.17	8.76	7.93	7.58
201906	8.46	7.89	8.11	4.32	8.55	9.25	8.17	9.28	7.32	7.58
201907	8.52	13.16	7.76	4.21	8.55	9.25	8.17	8.76	7.32	7.58
201908	8.32	2.63	8.23	4.83	8.31	9.25	8.65	8.76	7.32	7.58
201909	8.40	6.58	7.99	4.78	8.20	8.67	9.13	8.76	6.71	7.58
201910	8.51	5.26	8.11	4.52	7.97	8.09	7.69	8.25	6.10	7.58
201911	8.50	7.89	8.23	4.75	7.62	6.94	8.65	8.25	6.10	7.58
201912	8.99	11.84	7.82	51.02	6.81	4.05	7.21	7.22	5.49	7.58

Table C.1.1 Average values of predictors of the beta regression model of curtailing mortgages [13]

Mortgages did not make a curtailment in 2019										
Period	Outstanding Debt	Bullet Ind	Income To Principal Rate	Free Capital	Ever Curtailed	Percentage Months Curtailed	Percentage Curtailed To Principal	Percentage Franchise Used	Franchise Exceeded Ind	Average Curtailment
201901	8.34	9.26	8.30	8.27	7.50	8.33	6.12	4.55	5.71	8.33
201902	8.33	8.89	8.30	8.30	8.33	8.33	7.14	9.09	8.57	8.33
201903	8.34	8.89	8.30	8.31	8.33	8.33	7.14	9.09	8.57	8.33
201904	8.33	8.89	8.30	8.30	8.33	8.33	7.14	9.09	8.57	8.33
201905	8.34	8.52	8.30	8.32	8.33	8.33	8.16	9.09	8.57	8.33
201906	8.33	8.52	8.30	8.35	8.33	8.33	8.16	9.09	8.57	8.33
201907	8.33	8.15	8.34	8.35	8.33	8.33	8.16	9.09	8.57	8.33
201908	8.34	8.15	8.34	8.36	8.33	8.33	9.18	9.09	8.57	8.33
201909	8.34	7.78	8.34	8.37	8.33	8.33	9.18	9.09	8.57	8.33
201910	8.34	7.78	8.34	8.39	8.33	8.33	9.18	9.09	8.57	8.33
201911	8.34	7.78	8.39	8.38	9.17	8.33	10.20	9.09	8.57	8.33
201912	8.30	7.41	8.44	8.29	8.33	8.33	10.20	4.55	8.57	8.33

Table C.1.2 Average values of predictors of the beta regression model non-curtailing mortgages

[13] All column cells are displayed as percentage of the total sum of that column due to confidentiality. Order and relative magnitude between rows is maintained.

Appendix D.1

In this section we briefly discuss the packages that we used for modelling.

The multivariate logistic regression model was executed using Python. More specifically, we used the following package:

Sklearn – linear model – Logistic Regression: version 0.24.1

Moreover, to calculate the VIF values we used the following package:

Statsmodels – stats outliers – influence variance inflation factor

From Statsmodels: v0.12.2

The beta regression package was not available in Python, hence we used R. We used the following package:

Package 'betareg': Version 3.1-4.

Repository CRAN