Floating-rate mortgages: Why do they prepay?

October 7

2010

Master Thesis
C.T.J. Hamstra,
University of Twente, Industrial Engineering & Management.
Corporate Market Risk Management, ING.

A Logit model for prepayment drivers

Supervised by:
Dr. R.A.M.G. Joosten - University of Twente
Dr. B. Roorda - University of Twente
MSc. J.S. Holtman - ING CMRM

Preface

Two years ago I made the decision to advance from a Bachelor at *Saxion Universities of Applied Sciences* to a Master at the *University of Twente,* because I felt an academic Master would give me valuable skills and insights for future work as an Industrial Engineer.

The first weeks were tough, getting used to a new system and level of education and having to attend lectures in 4 different classes to be able to follow the *Pre-Master* as well as the Master curriculum. Fortunately, the curriculum of *Financial Engineering & Management* fit my interests in finance perfectly and, together with a talk with Track coordinator B. Roorda, inspired me to work even harder to finish the Master. I would like to thank Mr. Roorda for this. In March this year, the final hurdle remained, which I knew was going to be the toughest, the Master thesis…

This research was conducted for the department *Corporate Market Risk Management - Retail Netherlands (CMRM Retail NL)* of ING and in the context of a master thesis assignment for the studies *Industrial Engineering & Management* of the University of Twente.

Mrs. J.C.M. van de Rijt, manager of the department *CMRM-Retail NL,* formulated the research topic and it was supervised by her and Mr. J.S. Holtman from ING and Mr. R.A.M.G. Joosten and Mr. B. Roorda from the University of Twente. I would like to thank all of them for their guidance, and in particular Mr. Holtman for his patience and the pleasant day-to-day collaboration. I thank Mrs. Van de Rijt for adhering to my request for a more scientific research topic than the one initially planned. Moveover, I would like to thank Mr. Joosten and Mr. Roorda for their critical assessment of the thesis and the helpful suggestions they gave.

I am grateful for the help throughout the project of my direct colleagues and in particular Msr. C. Lubbers-Benning who provided me with insight into the statistical model applied and Mr. R. Stakenburg and R. Fanciully for data collection and statistical analysis respectively, which has been essential in getting the results.

My internship, research, and the development of a mortgage hedge-tool have given me insight into the application of *Market Risk management* theories in practice, and I have had a more than pleasant time working for CMRM and ING.

Floating-rate mortgages: *Why would they prepay?*

Summary

Researchers have given little attention to prepayment analysis on floating-rate mortgages, because of the supposedly small interest rate risk. The credit crisis, however, has led to unexpected increases in the absolute height and volatility of liquidity costs, i.e. the cost of obtaining funds. These developments introduced the risk of gains and losses, whenever the actual mortgage prepayment rates (early repayments) deviate from the ones projected. This was exactly what happened when central banks' injections of cheap short-term funds lowered short-term interest rates and lowered the incentives for clients to prepay. We have specifically identified the risks and drivers of these prepayments.

A portfolio of floating-rate mortgages that charge a fixed spread over Euribor has been analyzed for the period December 2006 until April 2010.

At first we conducted a literature research to find suggestions for explanatory variables (drivers). Suggestions were *relative contract age, loan notional, portfolio burnout, seasonality* and *interest rate volatility*. In contrast to *fixed-interest-rate* mortgages, the prepayments are not driven by the developments of solely one particular interest rate, but by an entire *term structure*, i.e. the yield curve. For this purpose we conducted research to characterize yield curves. *Level, slope,* and *curvature* were suggested as characterizing variables.

Secondly, the functional forms of prepayment functions were addressed to model prepayments at a contract level. Proportional hazard models [Cox, 1972] are often applied for prepayment analysis. They require the formulation of a *log-likelihood function* to conduct *Partial Likelihood Estimation*. Furthermore, we investigated *Probit* and *Logit* models. They are designed to perform regression analysis on binary response variables, which is the case for the data (prepay or not).

Because of its simpler functional form, the bounded output values, and the intuitive measure of a driver's strength offered by the *Odds Ratio,* we preferred the *Logit* model over the *Probit* model and proportional hazard models.

Functional form of the Logit model:

$$\mathrm{v}_{\mathrm{i,t}} = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k + \varepsilon_{\mathrm{i,t}}$$

$$\hat{p}_{i,t} = \frac{1}{1 + e^{-v(i,t)}}$$

Partial prepayments can be seen as a risk-free investment, yielding the mortgage rate. As such the prepayment drivers were expected to differ from full prepayments and both prepayment types were, therefore, modelled separately.

Four samples were drawn for both types of prepayments. We adopted an iterative approach to obtain the best-performing model specifications with the suggested explanatory variables. Multi-collinearity was taken into account and led to the omission of the explanatory variables *level*, *curvature* and *interest rate volatility*.

The samples for the so-called *credit mortgages* resulted into inconsistent models and require more data and further analysis. The results for the other mortgages show that full prepayments are mainly driven by the yield curve *slope* (negative relationship) and the *contract age* (positive relationship). Partial prepayments are mainly driven by *contract age* (positive relationship) and the second strongest driver is *slope* (negative relationship). As was hypothesized partial and full prepayments require significantly different models, which was confirmed by a *likelihood-ratio* test. The fraction of correct (balanced sample) predictions for full and partial prepayments is 66.8% and 63.5% respectively. These figures (compared to 50%) show the performance of the model and its included prepayment drivers. The best performing models are:

		Full prepayments			Partial prepayments		
Best-fit-models	**(a;b)**	Coefficient	St.error	Odds ratio	Coefficient	St.error	Odds ratio
C	(1;1)	1.76	(0.079)*	1.00	-0.19	(0.052)*	1.00
Contract age	(0.50;0.25)	2.43	(0.141)*	1.84	3.67	(0.0207)*	2.50
Slope(-1M)	(0.01;0.04)	-57.73	(1.053)*	5.65	-15.12	(1.544)*	1.57
Burnout	(10;30)	-0.038	(0.002)*	2.14	0	N/A	N/A
Notional	(2;1)	0.12	(0.016)*	1.27	-0.17	(0.030)*	1.40
dummy (winter/JanMar)	(0;1)	-0.20	(0.027)*	1.22	-0.32	(0.054)*	1.38
dummy (lownot)	(1;0)	0	N/A	N/A	2.86	(0.328)*	17.46

The odds ratio indicates how many times more likely prepayment is for value ‘a’ than for ‘b’. * = significant at level 2.5%

According to the *odds ratios* a client faced with a slope of 1% is nearly 6 times more likely to *fully* prepay than when faced with a 4% slope. A contract that has served 50% of its lifetime is nearly twice as likely to *fully* prepay as a contract that served 25%. This shows that if the *slope* could be modelled to project future values the Logit model could give great insight into expected future full prepayments.

The *Hosmer Lemeshow test* indicates there are still significant drivers not included in the model (omitted variables). This is economically intuitive, because whether a client prepays or not is in part driven by idiosyncratic (client-specific) factors. The *LR statistic* confirms the overall significance of both prepayment models. The *specificity* (identifying non-prepayments) is far lower than the *sensitivity* (identifying prepayments), due to the fact that predictions are made for specific months and not for, e.g., 6-month buckets.

To conclude, one should keep in mind that all these models and their drivers are based on historical data analysis. The results consequently tell us which variables were significant prepayment drivers in the past. Bank capital- and liquidity requirements as well as the collective memory and financial sophistication of clients may not return to the old situation. One should consequently question whether relationships derived from turbulent times, such as the period of 2006 until 2010, are a reliable predictor for the future or whether a regime shift has taken place. Conducting this research again with additional data in one or two years is, therefore, recommended to perform out-of-sample validation.

1. Introduction

1.1. Market developments

During the credit crisis liquidity became scarce and liquidity costs increased significantly. This has led to increased potential losses and gains in case of funding adjustments for specific floating-rate mortgages (called roll-overs)[1]. Lower (or higher) than expected non-contractual repayments, called *prepayments*, are causing lent out funds to return to the bank later (earlier) than expected, causing a funding (hedge) mismatch. These inaccurate prepayment projections in combination with changing liquidity costs pose re-investment uncertainty and potential liquidity shortage.

As opposed to the conventional prepayment analysis in fixed-rate mortgages, these developments led to this research into prepayment drivers for floating-rate mortgages with fixed spreads (interest add-ons) over Euribor.

1.2. Current situation

The department *Corporate Market Risk Management Retail Netherlands* (CMRM Retail NL) of ING is, among others, responsible for the assessment of market risks in retail products in the Netherlands and the modelling of the risks and periodically hedging of them. Hedges focus on liquidity and interest rate risk[2]. To hedge these risks one needs to project (forecast) prepayments which requires insight into what drives prepayments.

The floating-rate mortgage with fixed spreads is no longer sold by ING, nevertheless, a portfolio exists for which funding, based on prepayment projections, is locked in and risks need to be managed. The changing liquidity spreads (costs) have made prepayment analysis and prepayment modelling relevant, because the liquidity spread in these mortgages and in their funding is locked in for the entire contract tenor. As opposed to the prepayment risk in fixed-rate mortgages, this risk is, therefore, not limited (truncated) to the fixed-interest-rate period (typically 5-10 years), after which re-pricing of costs can occur, but instead it involves the entire liquidity typical maturity (typically 30 years).

Moreover, significant product feature differences between fixed-rate and floating-rate mortgages lead to the idea that prepayments may be driven by other factors. One of the significant differences is the option that roll-over (floating-rate mortgage) clients have to switch from a roll-over mortgage to a fixed-rate mortgage without penalty. The incentive is expected to depend, among others, on the yield curve shape and on the increase in mortgage interest rate that clients are willing to suffer in order to obtain the certainty that a fixed-rate mortgage offers.

[1] These mortgages have interest rates (*client rates*) that fluctuate according to the Euribor rate but are offset with a fixed additional spread, which covers the bank's costs and the profit margin.

[2] See Appendix II for definitions of the 4 sources of interest rate risk.

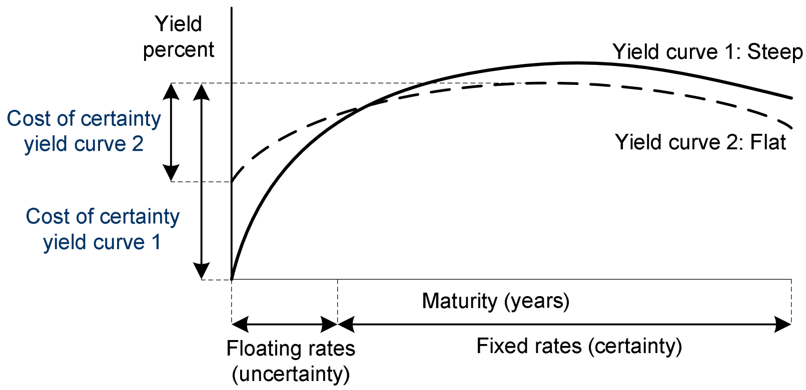


Figure 1: The cost of obtaining certainty by switching to a fixed-rate mortgage depends on the yield curve shape.

Because previous research projects were mainly focused on the refinance incentive (caused by decreasing interest rates) for fixed-rate mortgages the conclusions and drivers cannot be copied on a one-to-one basis, to floating-rate mortgages. Moreover, the absence of prepayment penalties in these floating-rate mortgages affects the clients' incentives to refinance.

This paper is intended to give insight into clients' prepayment behaviour in a specific portfolio of roll-over (floating-rate) mortgages to identify the variables that have driven the past prepayments (i.e. *prepayment drivers* or *explanatory variables*).

"Why would someone prepay a floating-rate mortgage?"

Before prepayment analysis is initiated, preliminary research has been conducted to clarify the context of the problem and the *core problem.* This preliminary research can be found in Appendix I and helps to give insight to those readers not familiar with the risks of prepayments in general (a.o. fixed-rate mortgages). Furthermore it explains explicitly which risks are posed by a prepayment for floating-rate mortgages with fixed spreads.

In the next subsection *"Problem analysis"* the conclusions from the preliminary research are summarized by answering the following two questions:

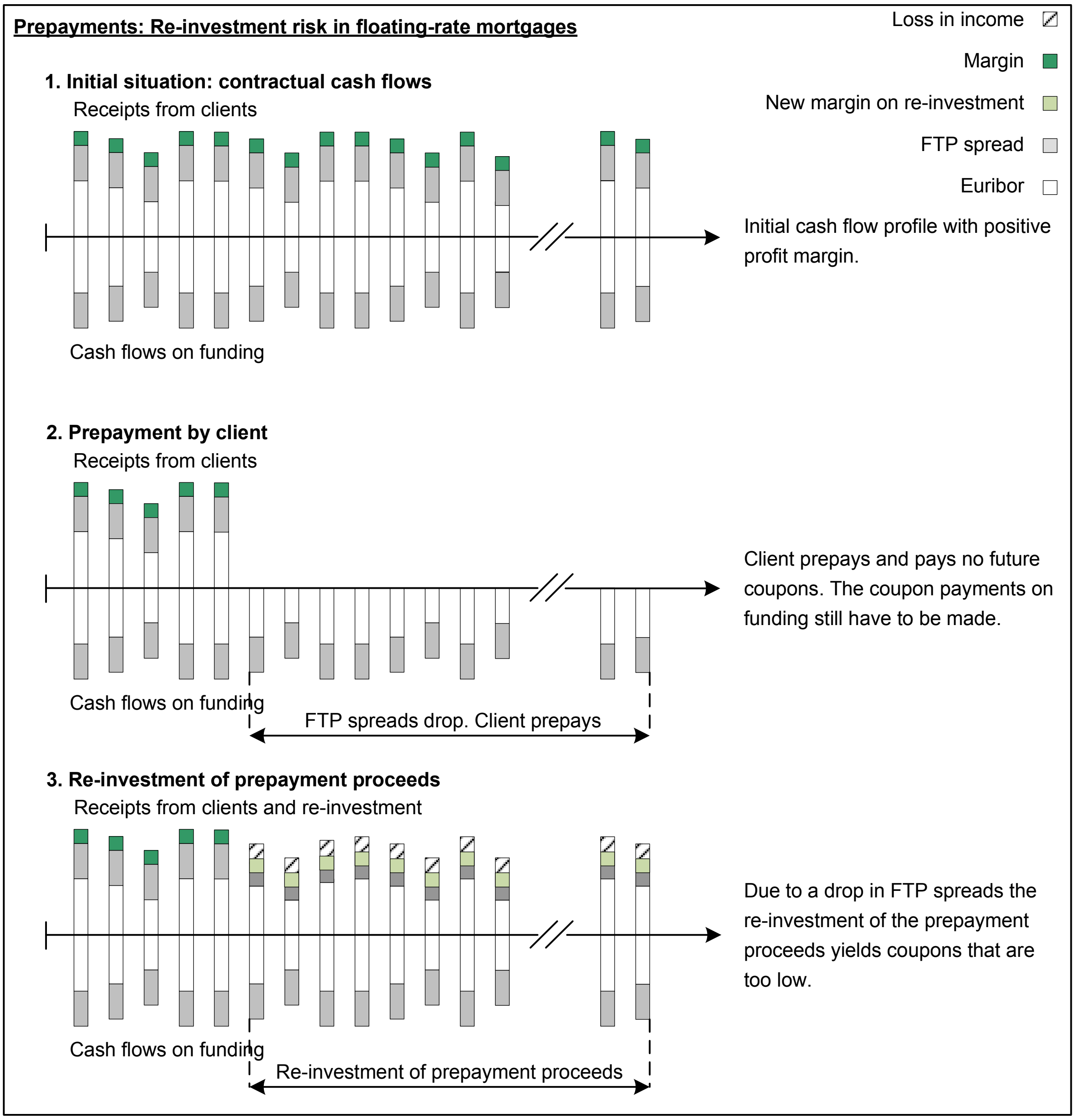
I. *What are the main risks involved in a portfolio of floating-rate mortgages with a fixed spread over Euribor?*

II. *How are these risks hedged and how could a prepayment function facilitate this, if forecasts of explanatory variables were available?*

1.3. Problem analysis

The key issues in prepayment risk, that result from the preliminary research in Appendix I, are:

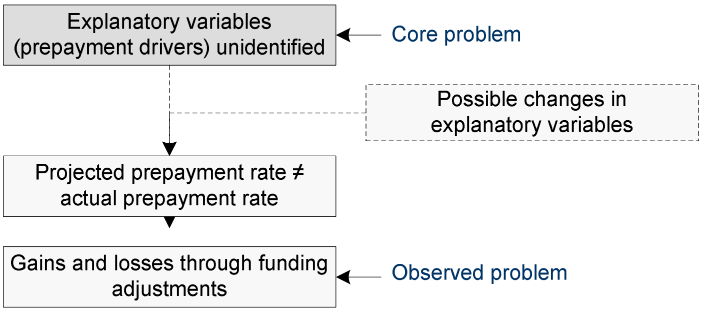
1. Funding spread volatility in combination with the long funding tenor (fixed spread is locked in) which, unlike the Euribor interest rate, cannot be transferred onto the client after the fixed-interest-rate period.
2. Reinvestment of prepayment proceeds occurs at a shorter tenor than the initial funding tenor.
3. Projections are not altered, because prepayment drivers are unknown.



It can be seen that when funding has high FTP spreads locked-in and FTP spreads drop, prepayment by a client poses a risk. The proceeds from prepayment cannot be re-invested at sufficient interest rates (sufficiently high spread over Euribor) to offset locked-in funding rates completely. A loss is consequently incurred. It should be noted that a prepayment directly after an increase in FTP spreads (liquidity costs) would constitute an economic gain, through the same reasoning.

1.3.1. Core problem identification

The final core problem identified is the issue that should be dealt with in order to tackle the observed problem. The observed problem is the gains and losses arising from funding adjustments. The link between the *core problem* and the observed problem is depicted in the causal chain below.[3]



1.3.2. Problem statement

Changing costs of liquidity, as defined by the FTP spread, can trigger economic losses when higher or lower prepayments rates occur than were expected. As opposed to fixed-rate mortgages re-investment risk applies to the entire remaining contract tenor, instead of only to the remaining time until re-pricing at the moment the fixed-interest-rate period ends. To accurately project prepayment rates CMRM first requires insight into the factors that drive these prepayments and the strength and nature of their impact.

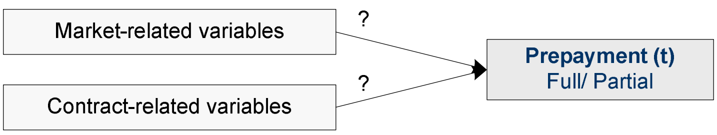


Figure 2: In the above model for the problem statement the nature, strength and direction of relationships between explanatory variables and the independent variable are unknown.

[3] This causal chain has been constructed according to the "Managerial Problem Solving Method" of J.M.G. Heerkens.

1.3.3. Problem owner and stakeholders

Problem owner

Transferring the liquidity- and interest rate risk to FM ALM is the task of CMRM. CMRM is the *problem owner* when it comes to modelling these risks accurately.

Stakeholders

CMRM and business units: hedge ineffectiveness may lead to *Profit & Loss* volatility at the business units and possibly to losses.

The entire bank: a wrong assessment of risk can lead to i) *Economic Capital* levels which are insufficient to cover shocks, ii) pricing being too low leading to clients not being charged for the risk they impose on the bank, iii) charging clients insufficiently for the prepayment option (and corresponding potential gains) they receive.

1.3.4. Conclusions preliminary research

I. *What are the main risks involved in a portfolio of floating-rate mortgages with a fixed spread over Euribor without prepayment penalties?*

Although the ever-changing Euribor component from the client rate in a floating-rate mortgage is exactly charged onto clients every month by adjusting the client interest rate. The other components, however, cannot be adjusted whenever they change, leading to the risk that some of these costs are not compensated for in case of unexpected high or low prepayment rates.

For a portfolio, with relatively low liquidity (FTP) spreads locked into client rates, the main risk lies in lower-than-expected prepayment rates requiring extra funding to be attracted at prevailing (high) liquidity costs (higher than clients are paying), constituting an economic loss.

For a portfolio with high liquidity (FTP) spreads locked into client rates, the main risk lies in higher-than-expected prepayment rates requiring funding to be unwound at prevailing (low) liquidity costs (lower than clients are paying), constituting an economic loss. A constant (high or low) FTP spread does not pose a risk.

II. *How are these risks hedged and how could a prepayment function facilitate this, if forecasts of explanatory variables were available?*

If insight in prepayment drivers led to accurate prepayment projections and accurate cash-flow-replicating portfolios, than no funding adjustments would be required and the exact liquidity costs paid on funding could be charged (transferred) onto clients.

1.3.5. Research questions

The preliminary research has led to the formulation of the following research questions:

1. Which explanatory variables can be identified as drivers for full and partial prepayments during the period Dec. 2006 - Apr. 2010?

> *1A. Which explanatory variables are suggested by the literature?*
>
> *1B. Which explanatory variables are found to be significant for the fixed-spread floating-rate mortgages?*

2. Which models and drivers explain full and partial prepayments between Dec. 2006 and Apr. 2010 most accurately?

> *2A. Which functional form is most suitable to model prepayments and identify drivers?*
>
> *2B. Should partial prepayments be modelled separately? Are the two models for full- and partial prepayments significantly different?*

The suitability and fit of the model will be determined by appropriate measures for the type of model chosen (see subsection 5.2.3)

The next section addresses the research design. Section 3:*"Literature research"* addresses research question 1A by analyzing previous research on explanatory variables and functional forms of prepayment models, which are, however, mainly focused on fixed-rate mortgages. Descriptive analysis of dependent and independent variables is conducted in section 4 in order to formulate hypothesized relationships between variables and prepayments. The explanatory research conducted in section 5 will answer the research questions 1B, 2A and 2B. Section 6 contains the conclusions and recommendations for further research.

2. Research design

This section elaborates on the scope of analysis in subsection 2.1, it covers the research questions in subsection 2.2, and subsection 2.3 and 2.4 cover the research design and the data requirements respectively.

2.1. Scope

2.1.1. Prepayment types and level of analysis

A portfolio of roll-over mortgages with a fixed spread over Euribor is analyzed. The portfolio contains contracts with the same fixed spread but with different start dates and contract characteristics. The contracts and prepayments will, therefore, be analyzed at an individual level in order to eventually explain prepayment rates at an aggregated level, i.e. the portfolio. This prevents the loss of information. Prepayment rate changes could in that case be explained by portfolio composition changes as well[4].

Prepayment types

Full prepayments: The entire outstanding mortgage notional is prepaid and the contract is ended before the end of the contract period has been reached. If the previous month's outstanding notional, however, was ≤25% of the initial notional, this will be defined as the *final partial prepayment*, instead of a *full prepayment*.

Partial prepayments: A part of the mortgage notional is prepaid, the contract is not ended.

Trigger prepayments: A trigger prepayment for roll-over mortgages occurs when a client switches from a floating-rate to a fixed-rate mortgage. This is seen as a special case of a full prepayment.

Analysis

Analyzing at contract level and aggregated level will include:
- Types of prepayments: *full* and *partial*.
- Time frame prepayment analysis: Dec. 2006-Apr. 2010: the portfolio is older, but no more data can be obtained for analysis.
- Time frame of contract origination dates: Contracts originated before June 2010 (98.3% is originated after January 1999)
- Variables:
    - Market-related variables such as the yield curve (height and shape) and the liquidity spread over time will be the main focus of analysis. Besides that the explanatory strength of contract-related variables will be tested.
    - Variables such as price inflation rate and housing market behaviour are outside the scope of analysis because data is not available.

Due to data restrictions trigger prepayments cannot be distinguished as a subset of the full prepayments. Full and partial prepayments are identified separately at a contract level and, together (expressed as a percentage of the aggregated notional), they form the dependent variables which are to be explained and for which drivers are to be found. The data filters applied to obtain the prepayment data can be found in Appendix III.

[4] This is closely related to the concept of *burnout*.

2.1.2. Empirical vs. optimal-call approach

Empirical vs. optimal-call approach (Pliska [2006])
While there is widespread agreement that the value of a mortgage contract subject to prepayment but not default risk should be given by an expectation of the present value of the cash flow, the devil is in the details. A wide variety of approaches have been considered, most of which are commonly classified into one of two categories.

One kind of approach has been variously called (…) an empirical approach and an econometric approach. The basic idea is to build a stochastic model for interest rates and possibly other economic factors, and then add a statistical model describing how the mortgagor's prepayment behavior depends on the factors.

The other main kind of approach for the valuation of mortgage contracts is called an option-based or a structural approach. The basic idea is to incorporate some kind of optimal behavior with respect to the mortgagor's decision about when to refinance. Moreover, the way to do this is to appeal to some intuition based upon the theory of the optimal early exercise decision for American options, usually leading to a recursive valuation procedure that resembles the one used for the binomial option pricing model.

The above box shows two approaches to prepayment modelling explained by Pliska (2006). We have adopted the *empirical approach* in this research because the goal of this research is to determine drivers for prepayments. Eventually ING intends to use this to attract mortgage funding with a suitable duration that mirrors the expected cash flows, subject to the *expected* prepayments instead of the *optimal prepayments.* The model should consequently fit *observed* prepayment behaviour which might be *non-optimal,* because prepayments can occur for *non-financial* reasons as well.

2.1.3. Credit mortgages vs. other mortgages

Credit mortgages are analyzed separately. These mortgages are offered to clients so that they have credit available when needed, which can be prepaid at all times. Prepayment behaviour for these mortgages is, consequently, expected to have more idiosyncratic (client-specific) explanatory variables.

If not analyzed separately these mortgages might add to the unexplained variance in the dependent variable because the omitted explanatory variables are unsuitable. The reason they are important is because their credit needs (notional increase) might at times offset prepayments (notional decrease) by other clients. On the other hand prepayments could be boosted when clients with a credit mortgage prepay as well.

2.3. Prepayment analysis

2.3.1. Longitudinal design

Longitudinal design means the prepayment behaviour of a group of clients (portfolio of mortgages) is monitored over time to identify drivers (explanatory variables). In this case the analysis is done *retrospectively* with logged data.

	Pre-test	**Event(1)**	**Post-test(1)**		**Post-test(n-1)**	**Event(n)**	**Post-test(n)**
Experimental group (**fixed spread**)	O	X	O	…	O	X	O
Control group (variable spread)	O		O	…	O		O

Time span: 2006 - 2010

Figure 3: The research design does not include a control Group. Retrospective data analysis is applied.

In Figure 3 the *"O"* signifies a measurement on the outcome variable, which in this case depicts whether a client has prepaid or not (binary). The *"X"* signifies an event (the prevailing values of the explanatory variables). Keeping the research goal in mind explains why a method such as *time series analysis* is not evaluated as an option; it does not result in drivers, other than time and a lagged dependent variable. Because the portfolio involves a great amount of contracts, and data is several years old a questionnaire or interviews are too cumbersome to execute with the resources available. The methods of regression analysis is thus adopted to answer the research questions.

Characteristics of research

- Time span: Dec. 2006 Apr. 2010
- Time interval between measurements: 1 month
- Dynamic panel: Full and trigger prepayments lead to panel attrition, because clients no longer have the initial roll-over mortgage. Besides this new contract inflow occurred over the time period spanned (mortgage production).

Purpose:

- Track and explain aggregate level changes over time by analyzing contract level data and prepayments.
- Distinguish *developmental* effects from *historical* effects: A change in prepayment behaviour may be triggered by the ageing of a contract, instead of specific events (changes in explanatory variables).

Internal validity issues:

- *Absence of randomized control group:* In order to prove that the supposed relationship is causal, it has to be shown that the randomized control group has significantly different values for the outcome variable than the experimental group. This is especially relevant for the *liquidity spread* as *explanatory* variable. See explanation in section "control group".
- *History*: Multiple events, which are not tracked in the research, may have impacted the outcome variable. Not all change in variability is then due to the explanatory variables. A significant part of the decision to prepay might be idiosyncratic (i.e. client specific) that cannot be explained by generic market-related variables, nor contract-related variables.

External validity issues:

- *Attrition bias:* The loss of cases over time (prepayment) is assumed to not be random. Because this is exactly what is to be explained, this does not hurt generalizability.

Potential future control group

In June 2009 new roll-over mortgages were introduced and sold with a variable spread over Euribor. The two portfolios (fixed-spread and variable-spread roll-overs) have approximately a one-year time overlap[5]. In the future, with more data this portfolio could serve as a control group. This is especially relevant for determining the sole impact of changes in the *liquidity spread*.

[5] Ageing effect can hardly be identified and distinguished from other effects.

2.4. Prepayment data

The type of data determines, among others, the type of analysis and conclusions that are feasible. This subsection addresses the origin of the data, the definition of a *prepayment* and the unit and level of analysis.

2.4.1. Sources

The data are collected from the database HP. Prepayment data at a contract level of these mortgages can be deduced from the databases as of December 2006.[6]

2.4.2. Data quality requirements

The data should be such that *reliability* and *internal validity*[7] are sufficient and maximized. Several factors are thus important with respect to the data:

- Time span covered: minimum of three years of prepayment data to cover variability in explanatory variables.
- Contract age range: The portfolio consists of several contract ages (main part up to twelve years.)
- Yield curve changes (shifts, twists) observed during time span of prepayment data.
- Significant FTP spread changes observed during time span of prepayment data.
- Prepayment can be identified according to explicit definitions.

Variability in explanatory variables in the time span covered is crucial in regression analysis for drawing reliable conclusions and maximizing internal validity of the research. The actual data that are required and used and the prepayment definitions used are found in Appendix III. The amount of contracts and amount of prepayment data is enormous and should not pose an issue for determining whether explanatory variables are statistically significant or not.

2.4.3. Data quality assessment

The data obtained span a sufficiently long time period (41 months) and ages are well distributed over the portfolio to meet the *data quality requirements*. See Appendix IV for information regarding the distribution of contract *start dates* and *reset periods*.

Information regarding actual liquidity spreads, prepayment rates and portfolio sizes, is confidential and has not been reported. Developments of these factors are shown, but in graphs without numbers on the axes.

[6] See Appendix III for the definitions of the different types of prepayments and the filter applied to obtain the data.

[7] See subsection: “Definitions” for an explanation of the concepts.

3. Prepayment drivers and functional forms

In this section literature research regarding prepayment analysis and corresponding explanatory variables (prepayment drivers) is conducted in order to answer research sub question 1A. Potential drivers following from this research will be analyzed in section 4 *"Descriptive research"* and tested for their explanatory strength in section 5 *"Explanatory research"*. Literature about fixed-rate mortgages or prepayments in countries other than the Netherlands will be assessed on relevance before it is analyzed further.

Research question 1A:

Which explanatory variables are suggested by the literature?

Definition: Seasoning vs. burnout

These concepts are addressed in the literature and, therefore, explained upfront. Seasoning refers to the ageing of an individual client's *contract*. Directly after taking on a new mortgage clients tend to show low prepayment rates, but these increase gradually over time as the contract ages.

Burnout, on the other hand, refers to an ageing mortgage *pool* that exhibits decreasing rates of prepayments, supposedly, because interest-rate-sensitive clients tend to prepay earlier and are therefore less and less represented in the ageing pool. The older mortgage pool is therefore biased towards interest-insensitive clients.

As is seen the *contract age* tends to dominate the prepayment rates initially, but over time the factor *burnout* becomes more prevalent. Especially for a portfolio that has no new production inflow of mortgages, as is the case for the fixed-spread floating-rate mortgages, the burnout effect is important.

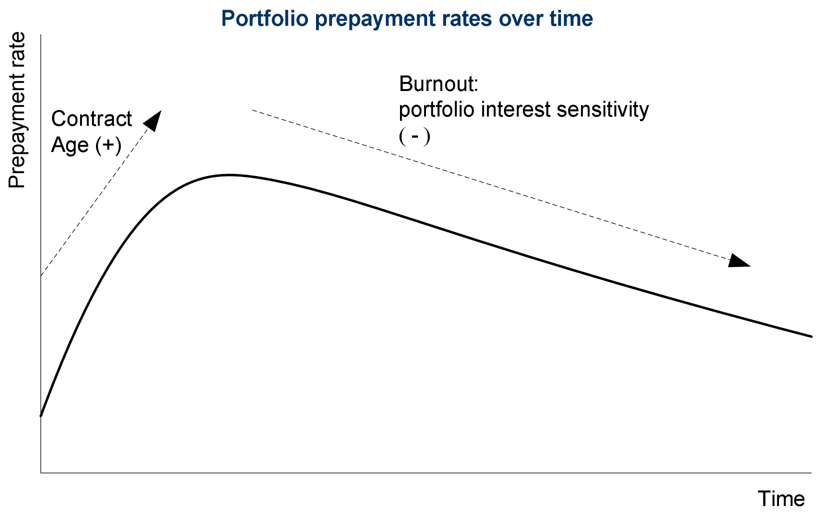


Figure 4: The effects of contract age (contract level) on prepayments rates vs. the effects of burnout on the portfolio level.

3.1. Research into explanatory variables

3.1.1. Previous ING research:

The conventional explanatory variable that is used for fixed-rate mortgages is the interest spread (as a proxy for the *refinance incentive*) between the client rate locked-in in the mortgage and the five-year SWAP-FTP rate.[8] ING introduces a three-month-lagged SWAP-FTP rate to compensate for the client's reaction time to interest rate changes. Previous research of Fanciulli (2009) has resulted in several additional significant explanatory variables:

"The data present in at least two instances a very compelling case for (…):
1. Relative age of fixed interest rate period.
2. Initial notional"

The parameters of the prepayment function are not necessarily the same for different business units, due to a potential difference in *"financial sophistication"* of clients, which is actually observed in prepayment data of ING. The hypothesis is that financially sophisticated (e.g. educated) clients are more aware of prepayment benefits when they occur than others.

Application to roll-over mortgages

The *relative age* might be a relevant explanatory variable in case of roll-overs.[9] The interest spread for roll-overs could be defined as the spread between a long-term fixed rate and the client rate (short-term floating). A positive spread would constitute an incentive to switch from a roll-over mortgage to a fixed-rate mortgage.

[8] The SWAP-rate is the rate at which interest swaps are settled, but where the loan notional never flows to a counter party. The SWAP-FTP rate is based on the SWAP-rate but includes the costs of liquidity, because the notional is transferred as well.

[9] Point of attention may be that *age*, defined in this way, is closely related to the *burnout* phenomenon which is addressed in the next subsection. A fundamental difference is the level of measurement of both variables. Whereas the *absolute age* applies to an individual contract, burnout applies to a portfolio and cannot apply to individual contracts.

3.1.2. Mortgage prepayments in the Netherlands

Alink (2002) analyzes prepayments for fixed-rate mortgages (Appendix V). He suggests *refinance incentive*, *burnout*, *seasoning*, and *seasonality* to be investigated. Refinance incentive is defined as the difference between the client rate locked in and the prevailing market rate, dampened by potential penalties. Seasonality refers to seasonal differences in prepayment rates.

The part of *burnout* that takes into account the heterogeneity of clients in a mortgage pool could be a significant explanatory variable for this portfolio of roll-over mortgages.

Burnout: *pool factor*

Alink discusses the operationalization of burnout and suggests using a pool factor as a proxy. The pool factor, for a pool of *n* mortgages, is defined as:

$$PF_t = \frac{\sum_{i=1}^{n} Notional_{\text{outstanding, i, t}}}{\sum_{i=1}^{n} Notional_{\text{initial, i}}} = \frac{Pool_{balance,t}}{Pool_{initial}}$$, where, $0 \le PF_t \le 1$ for $t \ge 0$.

The *initial pool* is without prepayment. In case the *pool balance* at time *t*, is close to the initial balance (PF ≈ 1), there have not been a lot of prepayments. Otherwise, if *PF* approaches zero it means a lot of prepayments have already taken place.

This method, however, has significant weaknesses because analysis occurs at a pool (aggregated) level. It does not group individual clients, based on identified differences, *fast* prepayers and *slow* prepayers (heterogeneity) to monitor the portfolio composition. On the contrary, using the pool factor *assumes* that the fraction of *fast* prepayers decreases over time. If this is, however, not the case the variable will simply turn out to be an insignificant driver after analysis. An alternative *burnout* proxy is constructed and listed in Appendix VIII.

3.1.3. Prepayment risk in adjustable-rate mortgages subject to initial discounts

Ambrose & LaCour-Little (2001) state the following:

Explanatory variables for floating-rate mortgages

"In the case of ARM (adjustable-rate mortgages), some argue that there may be a third motivation, namely a desire to switch from an ARM to FRM (fixed-rate mortgages) contract, based on expectations that rates have reached a trough. (…) As an indicator of borrower expectations concerning future interest rates, we include a measure of the term structure (TERMSTRU) defined as the 10-year Treasury bond rate minus the 1-year Treasure bond rate. Kau et al. (1993) argue that interest rate volatility has a significant impact on the prepayment option value (…) accordingly we use interest rate volatility, ARMVOL, defined as the standard deviation of the Freddy Mac 1-year ARM rate measured over the previous 24 months."

These comments concern yield curve characteristics, which are elaborated on in research summarized in the next subsection. The authors go on to say that they incorporate the *age* of a contract to account for seasoning and they include the quadratic term *age-squared* to *"control for nonlinearity in the shape of the hazard function"*, which, however, risks introducing multi-collinearity with the variable *age*.

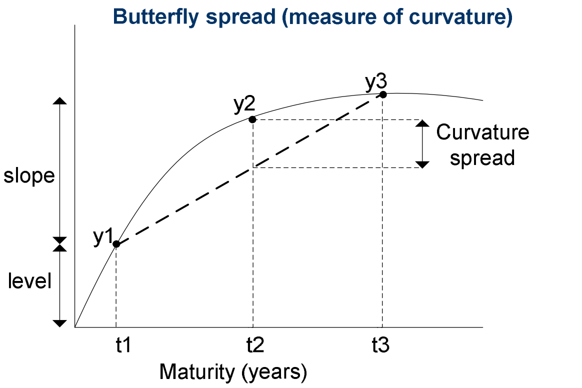
3.2. Research into yield curve characteristics

One of the options to a client with a roll-over mortgage is to switch to a fixed-rate mortgage with a fixed interest rate period of, for example, five or ten years. An interest rate corresponding to a single tenor is in that case no longer sufficient to quantify *refinance incentive*. An entire range of tenors and corresponding interest rates (a yield curve) may in that case impact clients' prepayment behaviour.

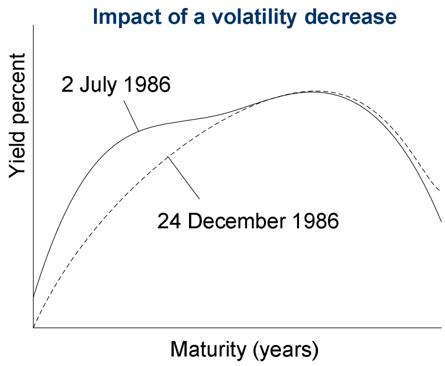
Using the yield curve in explaining prepayments requires identifying explanatory variables[10]. This subsection discusses the characterization of yield curves in order to use the characteristics as explanatory variables instead of all the individual interest rates.

3.2.1. Curvature, level and steepness

In *'Common factors affecting bond returns'* Litterman & Scheinkman (1991) identify common explanatory variables that affect the returns on U.S. government bonds and related securities. It is concluded that: *"most of the variation in returns on all fixed-income securities can be explained in terms of three factors, or attributes of the yield curve, which we will call level, steepness, and curvature."*



A butterfly spread can be used as a measure of curvature (concavity). Important are the choices of t1, t2, and t3. Source: "Revisiting the Shape of the Yield Curve: The Effect of Interest Rate Volatility" (Christiansen, Lund, 2005).



The volatility level decreased over the period of 2 July 1986 to 24 December 1986. This was accompanied by a decrease in curvature of the yield curve. Source: "Volatility and the yield curve" (Litterman, Scheinkman, 1991).

[10] Using individual interest rates corresponding to a range of tenors is undesirable since it would result in a large prepayment (regression) model, and could show significant *multi-collinearity*[10].

Curvature

"The third factor, which we call curvature, increases the curvature of the yield curve in the range of maturities below twenty years; the effect on yields tails off above twenty years." (Litterman, Scheinkman, 1991). Litterman, Scheinkman, and Weiss (1991) found that changes in curvature of the yield curve have a relationship to the changes in rate volatility. Prices of fixed-income securities with embedded options, such as callable bonds, are sensitive to this volatility.

A roll-over mortgage has such an embedded option as well, namely the option to prepay. Low long-term interest rates in combination with a flat or inverted yield curve might induce clients to switch their mortgage from floating-rate to fixed-rate in order to lock in these low rates and reduce uncertainty (due to interest rate volatility) in their monthly instalments.

Christiansen & Lund (2005) suggest a specific butterfly spread (Appendix VI) as a measure of curvature. An extra remark is made that w should be chosen such that the duration of the entire butterfly spread is zero. The butterfly spread is given by: $c_t = y_{2t} - \left(wy_{1t} + (1-w)y_{3t}\right)$. The bond tenors y1, y2, and y3 have been chosen by the authors such that the final curvature measure (historically) has a strong relationship (correlation) to interest rate volatility. The concluded tenors are three months, three years and ten years respectively.

Level

"(…) the first factor represents essentially a parallel change in yields (…)" (Litterman, Scheinkman, 1991). As a measure of *level*, Christiansen & Lund (2005) recommend using the three-month yield, stating: *"The 3 month yield is applied as a proxy for the instantaneous short term interest rate, and in our analysis it corresponds to the level of the term structure; $l_t = y_{1t}$ "*. The variable y_{1t} corresponds to their variable used in the measure for the *curvature*.

Steepness (slope)

"(…) a shock from the steepness factor (as defined here) lowers the yields of zeroes up to five years, and raises the yields for zeroes of longer maturities." (Litterman, Scheinkman, 1991). As stated, changes in steepness (slope) can be observed by a changing spread and a decreasing slope may pose an incentive for clients to switch from a floating- to a fixed-rate mortgage.

Moreover, Litterman, Scheinkman, and Weiss (1991) suggest that a high slope is an indicator of expectations for interest rates to rise. To determine a measure of the *slope* Christiansen & Lund (2005) state the following: *"In the empirical analysis, the slope of the yield curve is defined by $s_t = y_{3t} - y_{1t}$ (…)"*[11]. The variables y_{1t} and y_{3t} correspond to the ones used in measuring the *curvature*.

[11] The reason the authors state is: *"(…) in terms of keeping the correlation between the slope and the butterfly spread at a reasonable level, the definition (…) is the most appropriate choice."* This could create multi-collinearity in regression analysis.

3.3. Research into prepayment models: functional forms

This subsection addresses three papers that address *proportional hazard models* to explain prepayment rates. These models have the form $\lambda(t;z)=\lambda_0(t)e^{z\beta}$ introduced by Cox (1972)[12]. The hazard rate, variable $\lambda(t;z)$, depicts the expected fraction of clients that will prepay in a given time period. All three articles apply *baseline hazards,* $\lambda_0(t)$, which are boosted or dampened by proportionality factors, $e^{z\beta}$, which depend on the values of explanatory variables *z,* and the coefficients β. The entire model, therefore, accounts for the dependence of the hazard rate, $\lambda(t;z)$, on explanatory variables.

Parametric assumptions about the distribution of the survival time and consequently the hazard rate have thus been made before estimation of the parameters starts. Besides the *proportional hazard models*, this subsection covers *Logit, Probit and Extreme Value* models as well.

Proportional hazard models
"The proportional hazard model is an appropriate method when one is willing to assume that the conditional termination rate or hazard function (defined as the ratio of the density to the survivorship proportion, $\lambda(t)=f(t)/S(t)$ *can be written as the product of an exponential function of covariates and a baseline hazard that captures the underlying pattern of the hazard function over time which is identical across all subjects in the sample.(…) Estimation of the covariate parameter vector β can be achieved using a variety of canned routines for computing Cox (1975) partial-likelihood estimates that do not require specification and estimation of the baseline hazard component given"* [Calhoun, Deng, 2000].

3.3.1. Prepayment Behaviour of Dutch Mortgagors

Charlier and Van Bussel (2001) performed a study on fixed-rate mortgages in the Netherlands. They apply an S-shaped seasoning curve, and use the following proportional hazard model:

Proportional hazard model: [Charlier and Van Bussel (2001)**]**
"Let h_{it} *denote the hazard rate for mortgage i in month t:*

$$h_{it}=h_0(age_t;\upsilon_1,\upsilon_2).\pi(x_{it};\tilde{\upsilon})$$

The first parameter indicates the baseline-hazard for a mortgage that has just been originated, whereas the second parameter indicates how the baseline-hazard changes with an increase in the age of the mortgage. The function $\pi(x_{it};\tilde{\upsilon})$ *is the proportionality factor that depends on explanatory variables* x_{it} *and the parameters* $\tilde{\upsilon}$*. Given the explanatory variables, which will be defined later on, the parameters* $\tilde{\upsilon}$ *indicate the effect on the proportional hazard by a change in the related explanatory variable. The baseline hazard contains the seasoning effect, which refers to the gradual increase in prepayment speeds until a reasonably steady-state speed is reached (…). Therefore the S-shaped seasoning curve should hold for both purchase loans and refinancing loans and as such the following specification serves as the baseline hazard:"*

$$h_0(age_t;\upsilon_1,\upsilon_2)=\frac{1}{1+e^{-\upsilon_1-\upsilon_2.age_t}}$$

$$\pi(x_{it};\tilde{\upsilon})=e^z,\ z=-e^{-\tilde{\upsilon}'x_{it}}$$

[12] z_t is a vector of explanatory variables, β the corresponding vector of coefficients.

The functional form of $\pi(x_{it};\tilde{\upsilon})$ reflects the extreme-value distribution which can be used to estimate binary models. The functionial form of $h_0(age_t;\upsilon_1,\upsilon_2)$ reflects a logistic function (Logit model). The authors analyzed four explanatory variables: refinance incentive, burnout, seasoning and seasonality. For savings and interest-only mortgages it is among others concluded that *"the likelihood that a (…) mortgage will be prepaid increases with the age of the mortgage. Models excluding burnout also lead to a positive relationship between prepayments and the refinance incentive. However, when burnout is included the direct effect of the refinance incentive disappears and is taken over by burnout."*

3.3.2. Mortgage Prepayment Behavior in a Market with ARMs only

He & Liu (1998) have performed a study on ARMs (Adjustable-rate-mortgages) in Hong Kong.

Proportional hazard model: [He, Liu (1998)]

The authors introduce a log-logistic proportional hazard function with the proportional hazard model:

$\pi(t;\underline{v};\underline{\theta}) = \pi_0(t,\gamma,\rho).e^{\underline{\beta v}}$, and a baseline hazard of: $\pi_0(t;\gamma;p) = \dfrac{\gamma p(\gamma t)^{p-1}}{1+(\gamma t)^p}$ $\quad t^* = \dfrac{(p-1)^{1/p}}{\gamma}$

They state: *"The baseline hazard function measures the probability of prepayment under homogenous conditions,* $\underline{v} = \underline{0}$ *. The log-logistic hazard function admits a variety of relationships between the probability of prepayment and the age of the mortgage. In particular for p > 1, the probability of prepayment increases from zero to a maximum at* t^* *and decreases to zero thereafter. This kind of prepayment behaviour is consistent with the seasoning effects. On the one hand, people tend not to prepay too soon after paying hefty setup costs; on the other hand, as time goes by, most prepayments would already have been made and the prepayment would slow down."*

The log-likelihood function corresponding to the probability density function of the above model is:

$$\ln L(\theta) = \sum_{i=1}^{I}\left[\ln\gamma + \ln\rho + (p-1)\ln(\gamma t_i) - \ln\left(1+(\gamma t_i)^p\right) + \sum_{h=1}^{s}\beta_h v_h(t_i) - \exp\left(\sum_{h=1}^{s}\beta_h v_h(t_i)\right)\ln\left(1+(\gamma t_i)^p\right)\right]$$
$$-\sum_{j=1}^{J}\exp\left(\sum_{h=1}^{s}\beta_h v_h(t_i)\right)\ln\left(1+(\gamma t_i)^p\right)$$

Where I = 1,2,…, I denotes the set of prepaid mortgages, and the set of surviving mortgages is denoted by J=1,2,…,J

This baseline hazard function shows the prepayment rates in the absence of explanatory variables' influences other than *age* (*"homogenous conditions"*). The above description of the baseline hazard seems to indicate that the *seasoning* effect that is accounted for, implicitly takes *burnout* into account as well. The complexity of the log-likelihood function is a boundary to applying this model.

3.3.3. Mortgage Prepayment and Default Decision: a Poisson Regression Approach

Schwartz & Torous (1993) apply the proportional hazard function as well, but suggest Poisson distributed events. This justifies Poisson regression. Their study involves geographically dispersed fixed-rate mortgages in the United States.

Proportional hazard model: [Schwartz, Torous (1993)]
Here the age-dependency of prepayment rates is modelled by estimating multiple *age-dependent* parameters for the baseline hazard function.

$$P(X = x) = f(x) = \frac{\pi^x e^{-\pi}}{x!}, \quad x = 0,1,2,...$$

where $\pi = (v, \theta)$ and thus depends on explanatory variables. The final model for the prepayment rate becomes:

$$\pi[\tau, v(t), \alpha_\pi, \beta_\pi] = \pi_0(\tau, \alpha_\pi) e^{\beta_\pi v(t)}$$

The difference with the log-logistic baseline hazard of He & Liu (1998) is thus the way time dependency is modelled. Schwartz & Torous (1993) divide mortgage contracts over age-buckets and estimate constant baseline hazards for each bucket, whereas He and Liu use a log-logistic function to account for the entire range of ages in the portfolio.

3.3.4. Probit, Logit and Gumbel regression

Hosmer & Lemeshow (2000) and Cramer (2003) discuss the application of the Logit and Probit model and the interpretation of the model results (see User's guide Eviews 6 as well). This subsection summarizes some of their conclusions and discusses Gumbel regression as well.

"Whereas the hazard function in continuous time is defined as the instantaneous rate of failure conditional on survival to a given point in time, the discrete-time hazard is the probability that an event occurs (…) in the interval from t to t +1 , given that an event has not already occurred prior to t" [Calhoun, Deng, 2000]. A prepayment either occurs or does not in a specified month.

Probit, Logit and Gumbel models are meant for these kinds of binary dependent variables, because they address the issue of the bounded outcome values, whereas inputs are unbounded. The predicted values are bounded between zero and one by transforming a linear equation $\beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_k X_k$ with range $(-\infty, \infty)$ to a range $[0,1]$. This transformation is done through the use of a cumulative distribution function.

These binary models do not require the assumption of normality with respect to the residuals of the regression analysis. Statistical tests on coefficients are thus valid even if errors are not normally distributed. Coefficients are determined by *Maximum-Likelihood-Estimates*.

Probit regression

A *Probit model* uses the cumulative standard normal distribution, Φ , as transformation function. This leads to a probability of prepayment of:

$$P\left(y_{i,t}=1 \mid x_{1,i,t}, x_{2,i,t}, \ldots, x_{k,i,t}, \boldsymbol{\beta}\right)=\Phi\left(\beta_0+\beta_1 x_{1,i,t}+\beta_2 x_{2,i,t}+\ldots+\beta_k x_{k,i,t}\right)$$
$$P\left(y_{i,t}=1 \mid \mathbf{x}=0, \beta_0=0\right)=\Phi(0)=0.5$$

Logistic regression

Logistic regression is often referred to as the *Logit model*. The cumulative distribution function used as transformation function is the *logitistic equation*, as opposed to the cumulative standard normal distribution. This leads to a probability of prepayment:

$$P\left(y_{i,t}=1 \mid x_{1,i,t}, x_{2,i,t}, \ldots, x_{k,i,t}, \boldsymbol{\beta}\right)=\frac{1}{1+e^{-v(i,t)}}$$
$$v_{i,t}=\beta_0+\beta_1 x_{1,i,t}+\beta_2 x_{2,i,t}+\ldots+\beta_k x_{k,i,t}+\varepsilon_{i,t}$$
$$P\left(y_{i,t}=1 \mid \mathbf{v}_{i,t}=0\right)=\frac{1}{1+e^{-0}}=0.5$$

$v_{i,t}$ can be interpreted more intuitively as the log of the odds $p_{i,t}/\left(1-p_{i,t}\right)$ where $\left(1-p_{i,t}\right)$ is the probability of non-prepayment.

$$Logit\left(p_{i,t}\right)=v_{i,t}=\ln\left(\frac{p_{i,t}}{1-p_{i,t}}\right)=\ln(odds)$$

Gumbel regression

Gumbel regression is also referred to as the Extreme Value model and *"is based upon the CDF for the Type-I extreme value distribution. Note that this distribution is skewed."* [Eviews 6 Users guide]. It is often used to model extreme (rare) events.

$$P\left(y_{i,t}=1 \mid x_{1,i,t}, x_{2,i,t}, \ldots, x_{k,i,t}, \boldsymbol{\beta}\right)=\exp\left(-\exp\left(-v_{i,t}\right)\right)$$
$$v_{i,t}=\left(\beta_0+\beta_1 x_{1,i,t}+\beta_2 x_{2,i,t}+\ldots+\beta_k x_{k,i,t}\right)$$
$$P\left(y_{i,t}=1 \mid \mathbf{x}=0, \beta_0=0\right)=\exp(-\exp(0))=e^{-1} \approx 0.37$$

The asymmetry of the distribution and the fact that it is intended to model extreme events make the *Extreme-Value model* less suitable than *Probit* and *Logit* models. For logistic regression the *odds ratio* helps interpreting final model results and its drivers. It is a measure of association of *odds* instead of *probabilities.* The *odds ratio* is explained in the following example:

Odds ratio: a measure of association for Logistic regression

Another way to look at coeffiients is to calculate the *Odds-ratio.* Odds represent the relative frequency with which different outcomes occur. The difference of prepayment odds when x =1 compared to when x = 0 is:

$$Logit\left(p_{i,t}\right)=c+\beta x$$

$$\left.\begin{array}{l} Odds(x=a)=\dfrac{p_{x=a}}{1-p_{x=a}} \\[2ex] Odds(x=b)=\dfrac{p_{x=b}}{1-p_{x=b}} \end{array}\right\} Odds\,ratio=\frac{\left(\dfrac{e^{c+a\beta}}{1+e^{c+a\beta}}\right)\Big/\left(\dfrac{1}{1+e^{c+a\beta}}\right)}{\left(\dfrac{e^{c+b\beta}}{1+e^{c+b\beta}}\right)\Big/\left(\dfrac{1}{1+e^{c+b\beta}}\right)}=\frac{e^{c+a\beta}}{e^{c+b\beta}}=e^{\beta(a-b)}$$

$$Odds\,ratio=\frac{p_{x=1}/\left(1-p_{x=1}\right)}{p_{x=0}/\left(1-p_{x=0}\right)}=e^{0.70} \approx 2.01$$

A client is, therefore, approximately twice as likely to prepay when x=1 compared to when x=0, all other things equal.

The Odds-ratio is a measure of association (…) as it approximates how much more likely (or unlikely) it is for the outcome to be present among those with x=1 than among those with x = 0. [Hosmer, D.W., Lemeshow, S., 2000]

The confidence interval for beta and consequently for the *Odds-ratio* is: $C.I._{95\%}\ \beta_{slope}:\{\hat{\beta}_{slope} \pm 1.96\sigma_{\hat{\beta}}\}$

3.4. Distinction among types of prepayments

Research question 2B

Should partial prepayments be modelled separately? Are the two models for full- and partial prepayments significantly different?

No literature was found that aims to model partial prepayments separately from full prepayments. This might be because of practical considerations resulting from the typical relatively small part of the portfolio that partial prepayments make up, as measured in currency.

The distinction is relevant in case these prepayments have different drivers. This may be likely if one considers a partial prepayment as an investment with a risk-free rate of return equal to the mortgage client rate. If the mortgage client rate exceeds alternative risk-free investments in the market then clients in this portfolio have the incentive to prepay partially[13]. The benefit is reduced by the tax rate, because of deductibility of mortgage interest payments. Lowering of tax deductibility in combination with low short-term interest rates, might consequently induce partial prepayments.

Extracting the partial prepayment data from the full prepayment data prevents these data from clouding the regression analysis on full prepayments. Eventually this may lower the unexplained regression variance, which is our goal.

Although there might be a practical consideration not to estimate two separate models (partial prepayments rates are relatively low), the academic question *"Are the two models for full- and partial prepayments significantly different?"* remains, and will be further addressed in section 4 and 5.

[13] It should be noted that the risk-free rate of return obtained is floating and is, therefore, not a certain fixed rate of return.

3.5. Conclusion

Partial prepayments in particular may have other drivers than full- and trigger prepayments. For this reason partial prepayments are analyzed separately in the descriptive- as well as the explanatory research. Trigger prepayments are not distinguished from other full prepayments, because of restrictions in data structure.

Two separate models are, therefore, constructed in order to reduce the unexplained variance in the regression results. A hypothesis test will be performed to test the statistical significance of the difference between the model for partial- and the model for full prepayments which will answer research question 2B.

Research question 1A: *Which explanatory variables are suggested by the literature?*
To characterize yield curves the characteristics *level, slope,* and *curvature*, as defined by Christiansen & Lund (2005), are used in subsequent sections. Other suggested prepayment drivers are *relative contract age (seasoning), seasonality, level, slope, curvature, burnout* and *interest rate volatility*. Moreover, it is hypothesized that clients with high incomes and a high initial loan notional are more likely to prepay than others. Besides the definition of *burnout* according to Alink (2002) a second definition will be used to investigate the burnout's explanatory strength (Appendix VIII).

Burnout's importance in this particular portfolio comes from the fact that no inflow of new mortgages will occur, meaning that the portfolio size will slowly decrease and any burnout effect in prepayment rates would not be compensated for by new mortgages entering the pool.

Unique for floating-rate mortgages are *Interest rate volatility* and *slope*. Both variables are expected to be related to *switches* from floating-rate to fixed-rate mortgages, assuming a fixed-rate mortgage is perceived as a haven of *certainty* as opposed to a highly volatile floating-rate mortgage.

Market related	Contract related	Client related
Seasonality dummy variables Alink (2002)	(Initial) loan notional Fanciulli (2009)	Age of client
Interest rate volatility Ambrose & LaCour-Little (2001)	(Relative) contract age Fanciulli (2009)	Income of client
Level of yield curve Litterman & Scheinkman (1991)	Burnout Alink (2002)	
Slope of yield curve Litterman & Scheinkman (1991)		
Curvature of yield curve Litterman & Scheinkman (1991)		
Adjusted refinance incentive/ TermStr Ambrose & LaCour-Little (2001)		

Figure 5: The variables that will be investigated further in section 4 and 5.

Proportional hazard models and binary response models such as *Logit, Probit* and *Gumbel* were found to model prepayments. To make a final model choice, the empirical distribution of prepayments is first observed in section 4. Section 5 contains a further analysis and overview of the beneficial aspects of these models as well as potential unwanted aspects.

4. Descriptive research

In this section the historical values and developments of potentially relevant explanatory variables are investigated and the dependent variables. This serves several goals:

1. The empirical distributions of the dependent variables influence the choice (suitability) of the prepayment's functional form.
2. Historical developments are analyzed to check whether sufficient variability in explanatory variables has been present in the time period covered.
3. A pair-wise correlation matrix is constructed to identify potential sources of multi-collinearity.
4. Preliminary insight is obtained into potential relationships between explanatory variables and the dependent variables.

Research question 1B:

Which explanatory variables are found to be significant for the fixed-spread floating-rate mortgages?

The insight this section gives will help formulate hypothesized relationships to be tested in section 5, which will consequently answer research sub question 1B.

Subsection 4.1 addresses the historical prepayments and their empirical distribution, which form the basis for the choice of a suitable prepayment function in section 5 *"Explanatory research"*. Subsections 4.2 and 4.3 cover the historical developments of explanatory variables and their pair-wise correlations. Subsection 4.4 summarizes the conclusions drawn from the literature research and the descriptive research by formulating hypotheses graphically.

4.1. Historical prepayments

4.1.1. Empirical distributions

The prepayment rates will be analyzed as a fraction of contracts in the portfolio, which coincides with the research goal to model and explain client behaviour. To investigate the portfolio impact, each client's probability of prepayment can be calculated with the corresponding outstanding notional.

Full prepayments

The overall portfolio prepayment rate for a certain month follows from individual contracts either prepaying fully, partially, or not at all. If the partial prepayments are left out and analyzed separately, an individual contract exhibits binary behaviour.

Distribution of full prepayments

$$x_{i,t} \sim bernoulli\, r.v., \quad for\, i = 1,2,\ldots,n_t$$

where $x_{i,t}$ is a *Bernoulli* trial that has value one when a client prepays in month *t* and zero otherwise and $p_{i,t}$ is the unknown probability of full prepayment for contract *i* at time *t*, which depends on explanatory variables. For this reason the $p_{i,t}$, and consequently $x_{i,t}$, are expected to show dependency, which is what we wish to model. Unfortunately this dependency means the dependent variable is not binomially distributed, the *Central Limit Theorem* cannot be applied and therefore nothing can be said about the distribution of X_t / n_t .

$$\sum_{i=1}^{n} x_{i,t} = \frac{X_t}{n_t}, \quad for\, t = 0,1,\ldots,T$$

The expected value $E[X_t / n_t]$, however, could be projected when the expected value of explanatory variables is known as well as the portfolio composition and each contract's characteristics. The expected impact on the portfolio notional is denoted by:

$$E[L_{prepaid}] = \sum_{i=1}^{n} [p_{i,t} * L_{\text{outstanding},i,t}]$$

$L_{\text{outstanding},i,t}$ is the outstanding notional of contract *i* at time *t.*

Partial prepayments

Partial prepayments do not lead a client to drop out of the portfolio, so *n* is unaffected, however, the outstanding notional does decrease. Again the dependent variable is binary, since a client either prepays (partially) or does not. Individual partial prepayments may again be dependent on one another through common explanatory variables (drivers) although the impact of idiosyncratic factors (surplus cash that is invested by prepaying partially) is expected to be greater. Whether or not a client prepays is driven by drivers, whereas the amount that is prepaid is assumed to be more client-specific. For this reason the fraction of clients prepaying is, just as for full prepayments, used as dependent variable.

Distribution of partial prepayments

$$E[L_{Partial}] = \sum_{i=1}^{n} [p_{i,t} * E[F] L_{\text{outstanding},i,t}]$$

where $E[F]$ denotes the expected fraction of $L_{\text{outstanding},i,t}$ that will be prepaid, in case one prepays partially. $E[F]$ can be estimated by:

$$E[F] = \frac{\sum_{t=1}^{41} \sum_{i=1}^{n} M_{i,t}}{41k}$$

where *k* is the amount of contracts in the 41 months under analysis that have prepaid and $M_{i,t}$ is the fraction of the outstanding notional of contract *i* in month *t* that was prepaid. $E[F]$, therefore, is the average of past partial prepayment fractions.

4.1.2. Historical prepayments

In Appendix VII the full and partial prepayment rates (as a fraction of the portfolio) for credit mortgages and *other mortgages* are displayed in graphs. The actual prepayment rates are not depicted, but the developments over time are visible. This subsection will summarize the preliminary findings obtained from analysing these graphs.

Full prepayments

While credit mortgages are expected to have different prepayment drivers the graphs displaying the prepayment rates move seem to show the same movements. This suggests that they are strongly impacted by a common driver, i.e. a common explanatory variable.

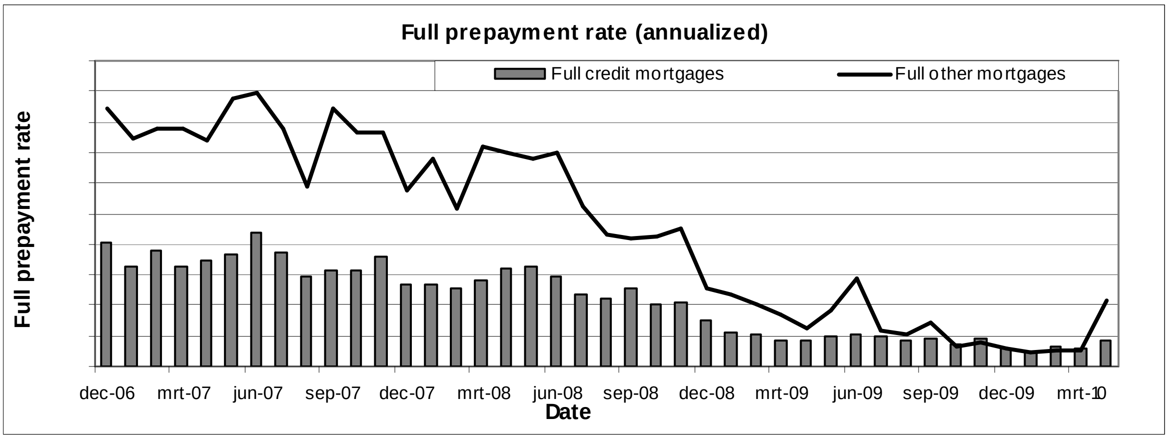


Figure 6: The full prepayment rates of source system HP are displayed. A distinction is made between credit mortgages and other mortgages, because of their potentially different prepayment drivers.

Partial prepayments

A negative partial prepayment for credit mortgages means extra credit was drawn by the client from the credit facility. The correlation (partial prepayments) for credit mortgages and *other mortgages* is remarkably high with a correlation coefficient equal to 0.64.

Low partial prepayments intuitively are correlated to low interest rates, which is also a beneficial time for clients with a credit mortgage to draw extra on their credit facility. Because of this positive correlation hardly any diversification is obtained which is unfortunate for the overall portfolio variance. Because the partial prepayments have a significant correlation with the full prepayments as well, with a correlation coefficient of 0.57, the variability of the aggregated prepayments is hardly reduced.

Although the variance may not benefit significantly from diversification effects, the impact of prepayments on the hedge portfolio is reduced by the negative partial prepayments of the credit mortgages because they compensate for part of the partial and full prepayments of the other mortgages, which are by definition positive. In short, when part of the prepayments is compensated for by other clients drawing extra credit the portfolio impact is reduced.

4.2. Historical values of explanatory variables

This subsection will discuss the historical developments of contract- and market-related variables, to get a first indication of potential relationships that may exist with the prepayment rates in the same period. Multi-collinearity will be addressed by analyzing pair-wise correlations.

4.2.1. Operationalization

For fixed-rate mortgages the main driver for prepayments is said to be *refinance incentive* (for household relocation, which prevents a fine to be paid), which arises from decreasing market rates. The equivalent for roll-overs is the refinance incentive that arises due to changes in the client spread that is paid over Euribor.

Besides this, the incentive to switch from floating-rate to fixed-rate mortgage could arise from the difference between a long-term fixed rate and the client rate (short-term floating). Clients may even prefer a fixed rate (certainty) over a floating rate (uncertainty) whenever the ten-year-fixed rate is slightly above the short-term floating rate. The resulting variable that measures the incentive to switch to a fixed-rate mortgage is referred to as the *adjusted refinance incentive*. See Appendix VIII for the operationalized explanatory variables.

Lagged explanatory variables

To account for the client's reaction time to interest rate changes a lag in the explanatory variables is introduced. A big interest rate change is quickly noticed by a client with a floating-rate mortgage, because of the consequent change in monthly instalments. Lags up to 5 months are investigated.

4.2.2. Contract-related variables

In Appendix IV it can be seen that the portfolio contains mortgages of all ages between 0 and 10 years (1.7% is older). Appendix IX gives some insight into the developments of the average initial contract notional over time.

4.2.3. Market-related variables

This subsection will give an overview of historical developments in market-related variables, namely yield curve shape[14] and Euribor short-term rates (height and volatility).

Liquidity spread (FTP): 2006 - 2010

The variability in the liquidity spread is the trigger that led to this research. A quick scan of the FTP spread over time has indicated that liquidity costs have increased sufficiently during the credit crisis, for them to be used as explanatory variable.

---

[14] In subsection 3.2.: "Research into yield curve characteristics", the method of measurement of *level, slope,* and *curvature* has been researched and explained in more detail.

Yield curve shape: Jan. 2006 – Jun. 2010

Client rates are based on the SWAP-FTP curve but contain risk spreads and the profit margin as well[15]. Even though clients observe client rates offered in the market, the use of the SWAP-FTP curve (and its characteristics) as explanatory variable is accurate as long as the spread between both is relatively constant.

During and after the credit crisis, however, the discrepancy of the five-year SWAP-FTP rate and the five-year mortgage client rate has grown. Projecting prepayment rates based on the SWAP-FTP projections might in that case lead to overestimating prepayment rates, whereas a future decrease in the discrepancy will most likely result in underestimates of the prepayment rates.

Over the period 2000-2010 several yield curves shapes are observed which are summarized by using the characteristics *level, slope, and curvature* in Figure 7*.* As can be seen the actual level of interest rates has varied significantly as well as the shape. The variability is even greater when quarterly or monthly data is analyzed. It should be noted that *level* and *slope* have a high inverse relationship, which follows from their definition[16].

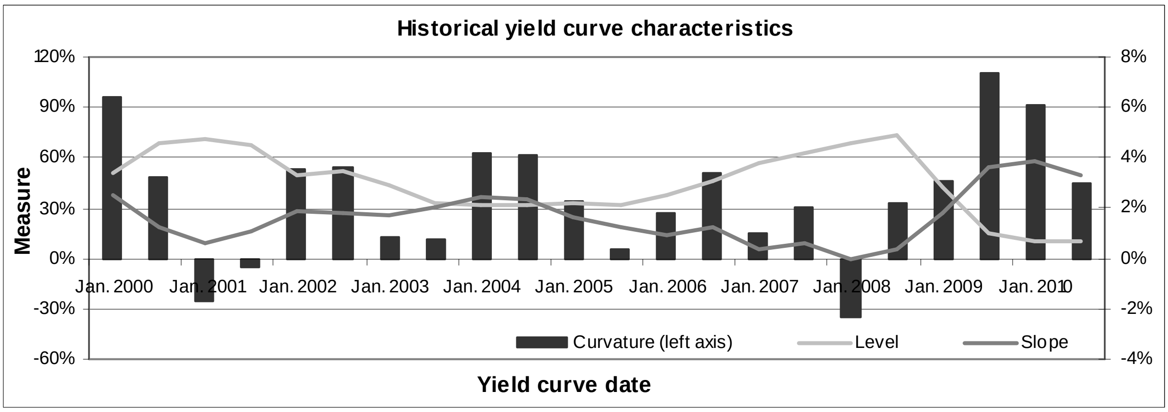


Figure 7: Observed yield curve (SWAP-FTP) characteristics between 2000 and 2010. Characteristics are 'level', 'slope', and 'curvature'. Measurement moments are the first available curve of January and June of each year.

[15] The reason SWAP-FTP is practical is its easy availability and because projections of it can be modelled by *Hull-White* interest rate models, which consequently serve as inputs to project future prepayment rates. Modelling market client rates would require projecting future risk spreads and profit margins as well.

[16] They both depend on the three-month rate. Although the slope depends on the ten-year rate as well, this rate is far less volatile. The changes in slope and level therefore follow mainly from changes in the slope.

Interest rate volatility vs. Curvature

The relationship between curvature of the yield curve and the interest rate volatilities of different reset periods is shown in Appendix X. From the correlation coefficients in Figure 8 it can be seen that the supposed relationship by Litterman, Scheinkman, and Weiss (1991) between the curvature and the interest rate (SWAP-FTP) volatility has been quite strong over the period Jan. 2008 - Jun. 2010 and weak over the other periods. The highest correlation is found with the short rates, the three-month, six-month and one-year SWAP-FTP interest rates.

Correlation coefficient: Vol (SWAP - FTP) vs. curvature						
Period	**1-month**	**3-month**	**6-month**	**1-yr**	**2-yr**	**10-yr**
2001-2010	0.55	0.54	0.54	0.51	0.44	-0.02
2006-2008	0,02	-0.01	-0.03	-0.06	-0.04	-0.16
2008-2010	0,77	0.76	0.75	0.73	0.69	0.16

Figure 8: The table shows the correlation coefficient between the volatility of different SWAP-FTP rates with the curvature measure for different time periods.

4.3. Pair wise correlations

Multi-collinearity clouds the individual explanatory variables' impact on the dependent variable and is caused by highly correlated explanatory variables. For this reason pair wise correlations have been investigated for the data samples. Appendix XI shows the correlation matrix.

4.4. Conclusion

The full and partial prepayment rates have varied substantially over the period analyzed. As expected prepayment rates have shown a declining trend due to the decreasing short term rates and the increasing yield curve slope. The value of full prepayments is far higher than that of partial prepayments, making them the most important ones to explain by drivers.

The yield curve shape, as defined by *level, slope,* and *curvature,* as well as the other explanatory variables have shown sufficient variability over the time period 2006-2010, in order for them to be used in regression analysis. Due to high correlations among explanatory variables, they cannot all be used in the same regression model.

Hypotheses:

By analyzing the results from the descriptive research in combination with the articles in the literature research the following relationships seem to be present:

1. High FTP spreads decrease the *refinance incentive* and consequently prepayments
2. Steepening of the yield curve reduces the *adjusted refinance incentive*, and lowers prepayments.

Because these changes in explanatory variables can have different effects on different types of prepayments (trigger, partial, full) the relationships will be tested for partial and full prepayments separately. See Figure 9 for the hypotheses[17].

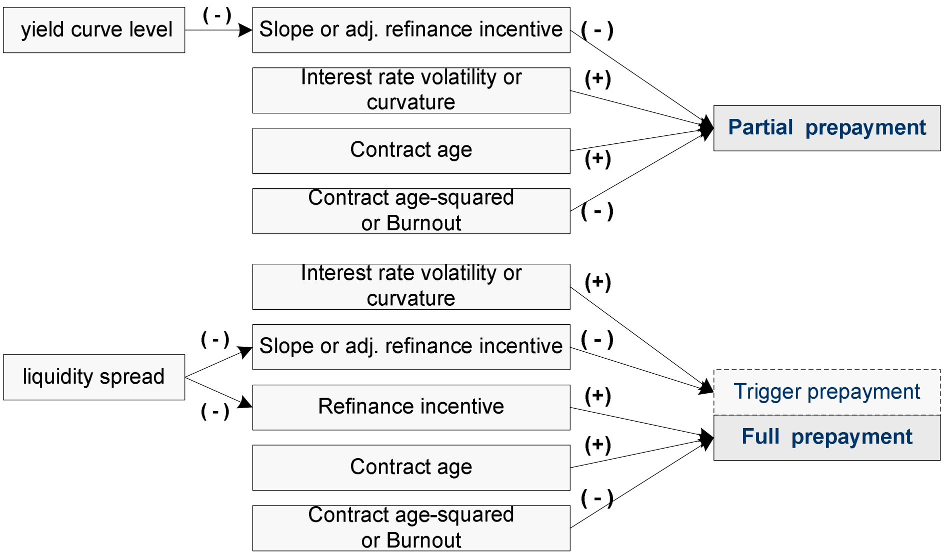
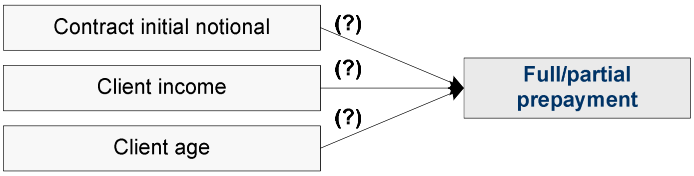


Figure 9: Hypotheses; relationships among explanatory variables and the types of prepayments (the dependent variables.)

The following relationships have no hypotheses but will be tested in the next section for their strength and direction if applicable.



[17] Although high interest rate volatility accompanied by an upward trend in the yield curve level is expected to pose an incentive to switch to a fixed-rate mortgage, the combination of these events has not occurred during the time period covered in analysis. For this reason this supposed interaction effect cannot be tested and consequently no hypothesis has been formulated.

5. Explanatory research

In this section the data analysis is conducted in order to answer research questions 1 and 2 by first answering their sub questions 1B, 2A and 2B.

Research questions

Which models and drivers explain full and partial prepayments between Dec. 2006 and Apr. 2010 most accurately?

> ***1B.*** *Which explanatory variables are found to be significant for the fixed-spread floating-rate mortgages?*
>
> ***2A.*** *Which functional form is most suitable to model prepayments and identify drivers?*
>
> ***2B.*** *Should partial prepayments be modelled separately? Are the two models for full- and partial prepayments significantly different?*

First, the most suitable functional form to model prepayments for roll-overs is chosen and substantiated in subsection 5.1. which answers sub question 2A. In subsection 5.2 the process of data analysis and the method of sampling data and obtaining the best model parameters are explained. Finally, in subsection 5.3 sub question 2B and the main research question 2 are answered. The models' results are summarized and an explanation is given on the interpretation of the model parameters and its coefficients.

5.1. Choice functional form

This subsection evaluates the suitability and applicability of the models from the literature research.

	Model	Output	Positive aspects	Negative aspects
Proportional hazard models	Poisson regression model with multiple age-dependent constant baseline hazards [Schwartz, Torous, 1993]	Survival analysis (duration model)	Clear distinction of prepayment rates between 'contract age' groups.	Multiple models and parameters are necessary to model age-dependency.
	A log-logistic baseline hazard function [He, Liu, 1998]	Survival analysis (duration model)	Intuitive interpretable modelling of time-dependency.	Multiple parameters are to be estimated through a complex probability density function and MLE procedure.
	Constant baseline hazard [Cox, 1972]	Survival analysis (duration model)	All dependency on explanatory variables is modelled in the exponent.	Outcome values not bounded between zero and one. Complex partial likelihood estimation.
Binary response models	Logistic regression (*Logit model)*	Discrete events. Prepayment probability of an individual contract (becomes a prepayment rate on a portfolio level)	All: Bounded outcome variable. Built-in pdf in Eviews and consequently Maximum Likelihood Estimation. No normality assumptions with respect to the residuals. Logit: Odds ratio makes output interpretable.	No disadvantages.
	Probit regression			Both: More complex cdf than Logit model. No odds ratio or derivative calculations to interprete coefficients intuitively.
	Gumbel regression			Gumbel: Meant to model extreme events.

Proportional hazard with log-logistic or Poisson baseline hazard

The reasoning behind the log-logistic baseline hazard of He & Liu (1998) with an age-dependent baseline hazard fits the economic intuition for prepayments. The application of the model, however, would require the formulation of a complex Log-likelihood-function to obtain point estimates through MLE and bootstrapping to obtain standard errors for the parameters in order to test for statistical significance. The same applies to the Poisson baseline hazard model of Schwartz & Torous (1993).

Proportional hazard with constant baseline hazard

The age-dependency of the baseline hazard could be approximated by a polynomial equation (in the exponent) instead of the log-logistic function. This would simplify the baseline hazard. In order to use OLS-regression, however, and conduct statistical tests, the residuals would have to be normally distributed which is not the case since the dependent variable is binary. An alternative is Partial- or Maximum-Likelihood-Estimation of the parameters, which requires the *probability density function*.

Logit, Probit and Gumbel regression

The binary character of the dependent variable makes the use of *Logistic*, *Probit*, or *Gumbel regression* suitable. The Extreme-Value distribution (Gumbel) is disregarded because it is intended to model extreme events.

Because of the functional form of the *Logit* and *Probit* model the projected dependent variable is bounded between zero and one. The coefficients of the regression equation are estimated by Maximum-Likelihood-Estimation, which is performed by EViews. No normality assumptions have to be made with respect to the residuals' empirical distribution. Because of its simplicity and the intuitive measure of association that is offered by the *odds ratio* (see subsection 3.3.4. for a definition), we prefer the *Logit* model over the *Probit* model.

5.1.1. Conclusion

Research question 2A:

Which functional form is most suitable to model prepayments and identify drivers?

The *Logit* model is most suitable and will be applied in the next subsection for the following reasons:

1. Explanatory variables can be tested for statistical significance without having to make normality assumptions with respect to the regression residuals.

2. EViews contains a built-in function to evaluate *Logit* models which does not require the formulation of a probability density function in order to determine maximum-likelihood estimates of the explanatory variables' coefficients and perform bootstrapping.

3. Results (the strength of drivers) can be interpreted intuitively by the *odds ratio*.

5.2. Data analysis

This subsection shows the process applied for the data analysis, explains the use of sample data and the method of stepwise regression analysis.

5.2.1. Analysis process

Figure 10 shows the entire data analysis process. Appendix III shows the query applied in step 1, and the definitions used to determine prepayments in step 4. Step 9 is explained in more detail in subsection 5.2.3.

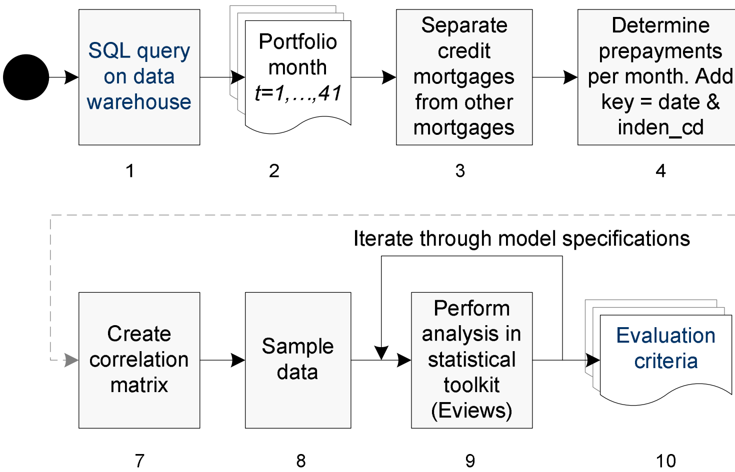


Figure 10: The process of data gathering, selection, and analysis.

5.2.2. Sample data

Because the data are analyzed on a monthly basis the overall fraction of the data that represents a prepayment is relatively small. Conducting the analysis on a random sample might, therefore, easily lead to bias or loss of information because part of the relatively few prepayments is not selected. Therefore, the analysis is performed on *balanced sample* data that contains all of the prepayments and a random sample of non-prepayments of the same size.

Without a model one would be able to predict 50% of the *balance sample* data correctly by just guessing. The final model should thus be able to predict a fraction well above 50% correctly indicating that significant prepayment drivers have been found.

Preferably the model were estimated on a sub sample up to a certain time point, and tested for predictive power on (available) future data outside of that sample (i.e. in-sample out-sample predictive power analysis). However, the limited time period spanned by the data requires the model estimation to be based on sample data spanning the full time period of December 2006 until April 2010. The performance is thus measured on data whose dependent variable realizations were known and were used to estimate the model.

5.2.3. Stepwise regression analysis

Different model specifications are iterated through according to the methodology described in the article *procedure for stepwise regression analysis* by Johnsson (1992) in order to find the models with the combination of explanatory variables that fits the best. The steps applied within point 9 of Figure 10 (previous page) are:

1. Each individual explanatory variable's correlation with the dependent variable is tested. Variables that show non-significant (significance level 2.5%) relationships are omitted from the model and from further analysis.
2. Pair-wise correlations of remaining significant explanatory variables are tested. The less significant of two highly correlated[18] variables is omitted from further analysis.
3. The evaluation is performed by iterating through all combinations of the significant explanatory variables left after step 2 and evaluating each model's performance and significance.

The evaluation in step 3 is based on its fit with the data by applying the criteria in the box below.

Model evaluation criteria

1. Akaike-Information-Criterion (AIC): measures the model fit (log likelihood) corrected for the amount of explanatory variables.
2. Schwarz criterion: measures the model fit with a stricter correction for the amount of explanatory variables than AIC.
3. McFadden R-squared: measures the relative model log likelihood improvement compared to the restricted log likelihood (all variables are omitted).
4. LR statistic: tests the overall significance of the model. This has to be satisfied for the model to be valid. A P-value of zero means the model is significant.
5. Hosmer-Lemeshow: This is a Goodness-of-fit test criterion that concludes whether there are still omitted variables that should be included in the model or not. A p-value of zero means there are still omitted variables.
6. Prediction accuracy: This measures the overall number fraction of correct predictions for a sample that contains 50% prepayments and 50% non-prepayments. As explained in the previous section this value should lie significantly above 50%.
7. Sensitivity: Measures the proportion of actual prepayments which are (correctly) predicted as prepayment.
8. Specificity: Sensitivity measures the proportion of non-prepayments which are (correctly) predicted as non-prepayment.

In Appendix XII the criteria are further explained. The prioritization of these criteria is as follows: *LR statistic* has to be satisfied and after that the prediction accuracy is maximized because this indicates the best drivers have been found which answers *research question 4*. The AIC and Schwarz criterion do not conflict with these conclusions.

Hosmer & Lemeshow (2000) remark the following regarding the use of *McFadden r-squared*: *"Unfortunately, low R^2 values in logistic regression are the norm and this presents a problem when reporting their values to an audience accustomed to seeing linear regression values…".*

[18] There is no formal definition for "highly correlated" with respect to multi-collinearity; there are just rules of thumb. The conservative boundary of 0.30 is used in the stepwise regression analysis here.

5.3. Final model

In this section the data analysis is conducted. The results are displayed in subsection 5.3.1. Subsection 5.3.2. gives special attention to the interpretation of the *sensitivity* and the *specificity* of the models. In subsection 5.3.3. the significance of the model differences between partial and full prepayments is tested, which consequently answers the question whether these prepayments should be modelled separately (research question 2B.) Subsection 5.3.4. and 5.3.5. help interpreting the model specifications and the method of sampling.

5.3.1. Model results

The *LR statistic* in all cases leads to rejection of the joint null-hypothesis[19] which shows the overall significance of the models evaluated. The *Hosmer-Lemeshow test* leads to a rejection of the null-hypothesis for each model, indicating that there are still omitted explanatory variables that should be included into the model to improve the goodness-of-fit. The detailed results of the other evaluation criteria per model can be found in Appendix XIV.

The model results for the credit mortgages varied among different samples[20]. One of the three samples drawn resulted in a significant model, whereas the other two did not. The seemingly high sensitivity to the sample drawn, led to the conclusion that more data and research are needed to identify consistent prepayment drivers for credit mortgages. Four different samples have been drawn and used for the portfolio *excluding* credit mortgages. This portfolio does provide consistent and stable model compositions and parameters, for partial as well as full prepayments. Lags up to 5 months were investigated for *slope*. The results of this can be found in Figure 11.

No significant relationship was found between prepayments and the client-related variables *client income* and *client age*. Furthermore, *Slope(-1M)* performed better than the *adjusted refinance incentive*, whereas *Volatility_1M* is preferred over *curvature* because it is more intuitive. The *FTP spread* is included in the *slope* and, therefore, not analyzed separately. The second burnout proxy performed better then the *pool factor* (Appendix VIII). The one-month lagged dependent variable for partial prepayments, *Preps(-1M)*, showed correlation, but was left out of the model because it cannot be used to forecast more than one month ahead. Moreover, these prepayments typically involve amounts below € 200 which limits their impact on the hedge position and consequently their risk.

The different signs for *notional* for partial and full prepayments are perhaps a consequence of the definition of a partial prepayment and the introduction of the dummy variable *'dlownot'*. Full prepayments that concern less than 25% of the initial loan notional are labelled a partial prepayment. This potentially creates the bias in the prepayment data that results in the negative sign for notional in the model for partial prepayments (Appendix III.)

---

[19] Appendix XII shows explanations of the evaluation criteria.

[20] The samples are drawn by the "=rand()" function of excel, which generates random numbers.

Summary statistics

The below statistics and model specifications give the preliminary answer to research question 2. Whether the model specifications for partial prepayments are significantly different from the ones for full prepayments is tested in subsection 5.3.3. *"Significance of model differences"*.

Research question 2:

Which models and drivers explain full and partial prepayments between Dec. 2006 and Apr. 2010 most accurately?

		Full prepayments			Partial prepayments		
Best-fit-models	{a;b}	Coefficient	St.error	Odds ratio	Coefficient	St.error	Odds ratio
C	{1;1}	1.76	(0.079)*	1.00	-0.19	(0.052)*	1.00
Age	{0.50;0.25}	2.43	(0.141)*	1.84	3.67	(0.0207)*	2.50
Slope(-1M)	{0.01;0.04}	-57.73	(1.053)*	5.65	-15.12	(1.544)*	1.57
Burnout	{10;30}	-0.038	(0.002)*	2.14	0	N/A	N/A
Notional	{2;1}	0.12	(0.016)*	1.27	-0.17	(0.030)*	1.40
dummy (winter/JanMar)	{0;1}	-0.20	(0.027)*	1.22	-0.32	(0.054)*	1.38
dummy (LowNotional)	{1;0}	0	N/A	N/A	2.86	(0.328)*	17.46

The odds ratio indicates how many times more likely prepayment is for value 'a' than for 'b'. * = significant at level 2.5%

Omitted explanatory variables:	**Alternative included:**
Curvature →	Volatility_1M
Level →	Slope
Adjusted refinance incentive →	Slope
Client age	N/A
Client income	N/A
Partial prepayment (-1M)	N/A

Evaluation criteria	**Full prepayments**	**Partial prepayments**	
Akaike-Information-Criterion:	1.224	1.275	
Schwarz criterion:	1.226	1.281	
McFadden R-squared:	0.117	0.081	
P[LR statistic]:	0.000	0.000	<- The model is significant
P[Hosmer-Lemeshow]:	0.000	0.000	<- There are still omitted variables
Overall fraction of correct predictions:	66.8%	63.5%	
- **Sensitivity**	87.3%	68.5%	
- **Specificity**	65.8%	78.7%	

Figure 11: Model specifications, omitted variables, and evaluation criteria (model performance.)'LowNotional' depicts a dummy variable that has value 'one' when the outstanding notional is below 15% of the initial notional and has value zero others.

The *Odds ratio* for the dummy LowNotional is strikingly high. One should, however, keep in mind that this is due to the definition of a partial prepayment. Since *full* prepayments that involve less than 25% of the initial loan notional are labelled a *partial* prepayment, the dummy *LowNotional* in fact accounts for the bias in the data. Because of its relatively low occurrence it does not have high predictive power.

5.3.2. Model Sensitivity vs. Specificity

The evaluation criteria show that the *sensitivity* of the model for full prepayments is far higher than the *specificity*, which means that there is a bias towards predicting too many prepayments. The overall fraction of correct predictions is consequently mainly determined by the model predicting prepayments well, and not by predicting non-prepayments well. See Figure 12 for a graphical representation of *sensitivity* and *specificity*.

The low specificity becomes clear when one considers the following;

1. Whenever market- and contract conditions are such that a prepayment might be expected, the model will predict this prepayment immediately.
2. However, if this prepayment occurs after, for example, 6 months the model will have predicted 5 *non-prepayments* incorrectly, whereas it predicted the *prepayment* correct.
3. The result is a high sensitivity and a low specificity.

This issue could be addressed by aggregating the data in time buckets of, for example, 6 months. The model, in that case, would have predicted a prepayment to occur *in the next 6 months*, which in fact it also did.

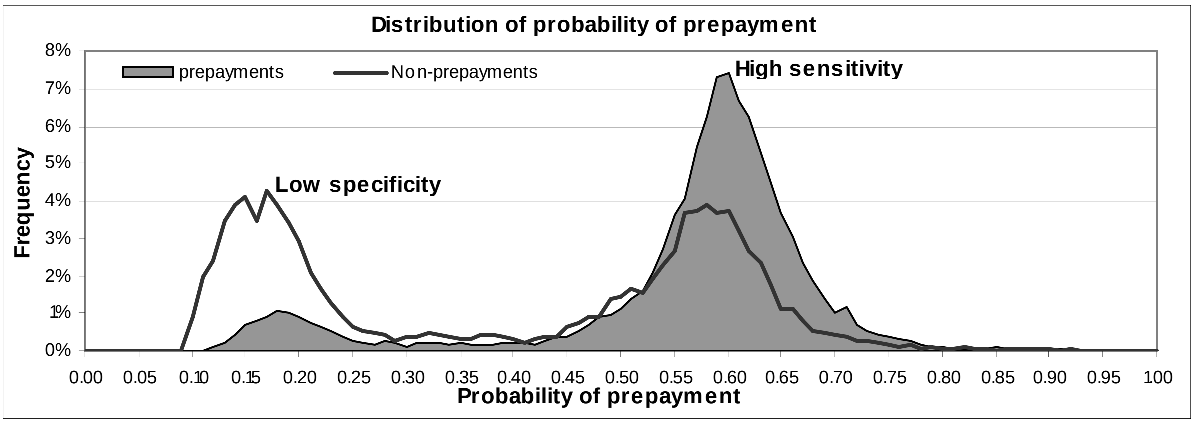


Figure 12: A graphical representation for full prepayments of the evaluation criteria 'sensitivity' and 'specificity'. A perfect model would predict a probability of 1 for all prepayments, and a probability of zero for all non-prepayments.

5.3.3. Significance of model differences

Research question 2B:

Should partial prepayments be modelled separately? Are the two models for full- and partial prepayments significantly different?

To optimize model performance it was decided to model partial prepayments separately from full prepayments. Now that this has been done research question 2B will be answered through a *Likelihood Ratio* test.

The *Likelihood Ratio* test is composed of three steps:

1. Calculate the log-likelihood of the estimated best model for partial prepayments, which is referred to as the *unrestricted model*.
2. Calculate the log-likelihood of the model for full prepayments when applied to data of partial prepayments. This model is referred to as the *restricted model*. Because this model is per definition suboptimal its log-likelihood will be lower.
3. Test whether the decrease in log-likelihood when the *restricted model* is applied, compared to when the *unrestricted model* is applied, is statistically significant.

The *log-likelihood* for the unrestricted model is known (see Appendix XIII), $l_u = -4931.34$, whereas the *log-likelihood* for the restricted model has to be calculated. The log likelihood functions (EViews 6 User's Guide II, 2007) are defined as:

Log - likelihood functions

$F()$ depicts the cumulative distribution function, y_i are the realizations of the dependent variable and x_i, β are the vectors of explanatory variables and their coefficients:

$$P\left[y_i = 1 \mid x_i, \beta\right] = F\left(-x_i'\beta\right)$$
$$P\left[y_i = 0 \mid x_i, \beta\right] = 1 - F\left(-x_i'\beta\right)$$
$$l(\beta) = \sum_{i=1}^{7744} y_i \ln\left[F\left(-x_i'\beta\right)\right] + \left(1 - y_i\right)\ln\left[1 - F\left(-x_i'\beta\right)\right]$$

Cumulative distribution function: logistic equation

Since the cumulative distribution function in this case is the logistic equation the Log-likelihood, $l(\beta)$, becomes:

$$F\left(-x_i'\beta\right) = \frac{1}{1 + e^{-x_i\beta}}$$
$$1 - F\left(-x_i'\beta\right) = 1 - \frac{1}{1 + e^{-x_i\beta}} = \frac{1}{1 + e^{x_i\beta}}$$
$$l(\beta) = \sum_{i=1}^{7744} y_i \ln\left[\frac{1}{1 + e^{-x_i\beta}}\right] + \left(1 - y_i\right)\ln\left[\frac{1}{1 + e^{x_i\beta}}\right]$$

The next page shows the results of the hypothesis test and also shows the programming code used for obtaining the log likelihood of the *restricted model.*

Hypothesis test on statistical significance of model differences

Significance of Model differences: Log likelihood ratio test

If using the unrestricted model for partial prepayments yields a statistically significant improvement in log likelihood (higher value) then H_0 will be rejected in favor of H_1.

Hypotheses

$$H_0 : l_u - l_r = 0$$
$$H_1 : l_u - l_r > 0$$

Test statistic/ Distribution

The test statistic is defined as: $LR = -2\left(l_r - l_u\right)$ and has a χ_5^2 distribution, because 5 parameters are restricted.

Critical region

$$\chi^2_{5,\alpha=0.01} = 15.09$$
$$C.R.: LR > 15.09$$

Value test statistic

The log likelihood of the restricted model is obtained through the following Eviews code:

Generate series:

```
Betax = 1.7610+2.433*age-57.73*slopelag1-0.0383*burnout+0.1215*notional-0.2004*dwinter
likelihood=preps*log(1/(1+exp(-Betax)))+(1-preps)*log(1/(1+exp(Betax)))
```

Run program

```
scalar loglikelihood=@sum(likelihood)
```

$$l_r\left(\beta\right) = -5612.81 \qquad l_u\left(\beta\right) = -4931.34$$
$$LR = -2\left(l_r - l_u\right) = 1362.94$$

Conclusion

$$1362.94 > 15.09$$
$$LR \in C.R.$$
$$\text{Reject } \mathrm{H}_0$$

The nul hypothesis is rejected indicating that the difference in log likelihood between the unrestricted and the restricted model for full prepayments is significant. This in turn means that the model for partial prepayments is significantly different from the model for full prepayments.

The above statistical test concludes that the model for partial prepayments is significantly different from the model for full prepayments. This indicates that one benefits from modelling partial prepayments separately since the performance will be higher. The models will be dealt with separately in the following subsections.

5.3.4. Drivers: strength, direction and *Odds Ratio*

The impact of the coefficients in the regression equation on the probability of prepayment eventually follows from the logistic equation, which is illustrated in the following text box. The importance of an explanatory variable is not seen from the height of its coefficient, because this depends on the unit of measurement of the explanatory variable.

The *odds ratio* shows that a client is nearly 6 times more likely to prepay when the slope is 1% compared to when it is 4%. The negative sign of the *constant* for partial prepayments is counterintuitive. It should, however, be noted that it is only negative in the presence of the other explanatory variables. Moreover, through the logistic equation, the eventual probability of prepayment is positive at all times[21].

Driver strenght and direction

By taking the derivative of the logistic equation with respect to $v(i,t)$ the impact of a small increase in $v(i,t)$ on the probability of prepayment can be seen if we assume a sample that contains 50% prepayments.

$$\hat{p}_{i,t} = \frac{1}{1+e^{-v(i,t)}} \qquad \frac{\delta(\hat{p}_{i,t})}{\delta v_{i,t}} = \frac{e^{-v(i,t)}}{\left(1+e^{-v(i,t)}\right)^2} > 0$$

Because this derivative is positive, a higher coefficient increases the probability of prepayment. The impact of an increase in the *slope* will be illustrated by an example for the probability of full prepayment. The explanatory variables are, e.g., set to: Age = 0.25, burnout = 30M, notional = 1 (* 100,000), dWinter = 0, leading to: $v_{i,t} = 1.036 - 57.73x$ and $x = slope_{t\text{-}1}$

$$\frac{\delta(\hat{p}_{i,t})}{\delta x} = \frac{\delta(\hat{p}_{i,t})}{\delta v_{i,t}} \frac{\delta v_{i,t}}{\delta x} = -57.73 \frac{e^{-(1.036-57.73x)}}{\left(1+e^{-(1.036-57.73x)}\right)^2}$$

Evaluated at $x = 3\%$ this becomes:

$$\left.\frac{\delta(\hat{p}_{i,t})}{\delta x}\right|_{x=3\%} = -1282\%$$

The unit of measurement of the *slope* is 100%. The impact of a slope-increase of 1% on the probability of full prepayment is, therefore: -12.82%.

Odds ratio

Another way of looking at a coeffiient's influence is to calculate the Odds-ratio. The difference of prepayment odds for a slope of 1% and 4% is:

$$Odds\,ratio = e^{-(-57.73*[0.04-0.01])} = 5.65$$

The confidence interval for beta and consequently for the Odds-ratio is:

$$C.I._{95\%}\ \beta_{slope} : \{\hat{\beta}_{slope} \pm 1.96\sigma_{\hat{\beta}}\} = \{\hat{\beta}_{slope} \pm 1.96\sigma_{\hat{\beta}}\} = \{-57.73 \pm 1.96 * 1.05\} = \{-59.80; -55.70\}$$

$$C.I._{95\%}\ OR : \{e^{55.70*(0.04-0.01)}; e^{59.80*(0.04-0.01)}\} = \{5.32; 6.01\}$$

A client is, therefore, nearly 6 times more likely to prepay when the slope of the yield curve is 1% than when it is 4%, all other things equal.

[21] Slope(-1M) indicates a one-month lag in the slope variable. This means the independent variable (slope) of last month is used to explain the dependent variable of this month.

5.3.5. Correction for biased sample

In the original full data set a far lower fraction (say: 2%) of the dependent variables' realizations[22] involve full prepayments, whereas for the sample used for analysis this is 50% (see subsection 5.2.2). The final model will, consequently, have to be corrected for this bias if one wishes to apply the model to contracts randomly drawn from the full data set.

Drawing randomnly from the full data set

The predicted value from the Logit model has to be corrected by the correction factor which is calculated by: $c.f. = 0.0158/0.5 = 0.0316$. This follows from the fact that a contract is drawn randomnly from a subset of the full data set:

$$P[y_i \in sample] = 0.04$$
$$P[y_i = 1 \mid y_i \in sample] = 0.50$$
$$P[y_i = 1] = P[y_i = 1 \mid y_i \in sample] * P[y_i \in sample] = 0.50 * 0.04 = 0.02$$

For example, if one again wishes to identify the impact of a change in the *slope*, the equation becomes:

$$\frac{\delta(\hat{p}_{i,t})}{\delta x} = 0.04 * \frac{-57.73e^{-(1.036-57.73x)}}{\left[1+e^{-(1.036-57.73x)}\right]^2} = -51.3\%$$

An increase in the slope from 3% to 4% would, thus, lower the probability of prepayment by 0.41% for a randomly chosen contract from the full data set. Because of the low fraction of prepayments among the realizations of the dependent variable the model will hardly ever predict a probability of prepayment of 50% or higher for a single, randomly drawn, contract. A decrease in the *slope* from 3% to 2% will only increase the probability of prepayment by 0.41% which would not change the expectation of the single contract, because it will still be below 50% and thus is expected not to prepay. Consequently one might conclude that the model has low predictive ability. On average, however, the *portfolio* is expected to show 0.41% more prepayments among all the contracts and on a portfolio basis, therefore, the 0.41% has a significant impact on the fraction of portfolio notional that is expected to be prepaid,, through the following equation:

$$\hat{P}_t = \frac{\sum_{i=1}^{I} \hat{p}_{i,t} N_{i,t}}{\sum_{i=1}^{I} N_{i,t}}$$

With respect to the predictive ability of the model one should realize two things:

1. It will remain hardly possible to predict correctly that a contract will not prepay in month *t*, but will prepay in month *t+1*, because of the marginal impact this single month has on the *(relative) age* of the contract (±1/360). Allocating contracts to time buckets of, for example, 6 months, would improve the predictive ability. This explains the low *specificity* of the model.

2. As can be seen from the *Hosmer-Lemeshow* test values there are still omitted explanatory variables that should be included in the model to get a *goodness-of-fit* close(r) to 1. A significant part of these omitted variables is expected to be *idiosyncratic*, meaning that not only *common* factors but unknown *client-specific* factors exist as well that influence their propensity to prepay.

[22] Each month that a contract is seen in the data is a realization. One contract, therefore, has a maximum of 41 realizations corresponding to the 41 months under analysis.

6. Conclusions

In subsection 6.1 the research questions are answered. The potential use of the results is explained in subsection 6.2 and recommendations for further research are given in subsection 6.3.

6.1. Answer to research questions

1. *Which explanatory variables can be identified as drivers for full and partial prepayments during the period Dec. 2006 - Apr. 2010?*

Contract age, yield curve slope, burnout, loan notional, Winter/JanMar (dummies) and LowNotional (dummy) were found significant drivers for prepayments for floating-rate mortgages with a fixed spread over Euribor as is depicted in the summary statistics in Figure 10 of subsection 5.3.1.

2. *Which model and drivers explain full and partial prepayments between Dec. 2006 and Apr. 2010 most accurately?*

2A. A Logit model was concluded to be the most suitable functional form, because of its bounded outcome variable and intuitive measure of association offered by the *Odds ratio*.

2B. A *likelihood ratio* test concluded that partial prepayments should in fact be modelled separately from full prepayments. Their model differences were found significant.

From an academic perspective, the best-fit-models are displayed in Figure 10 in subsection 5.3.1. The more practical simplified models, however, have around 88% of the explanatory strength and predictive power of the complex models and are attractive practical alternatives because of their simplicity.

		Full prepayments			Partial prepayments		
Simplified models:	**(a;b)**	Coefficient	St.error	Odds ratio	Coefficient	St.error	Odds ratio
C	(1;1)	0.352	(0.022)*	1.00	-0.403	(0.041)*	1.00
Contract age	(0.50;0.25)	1.747	(0.134)*	1.55	3.995	(0.201)*	2.71
Slope(-1M)	(0.01;0.04)	-59.34	(1.041)*	5.93	-17.145	(1.500)*	1.67

The odds ratio indicates how many times more likely prepayment is for value ‘a’ than for ‘b’. * = significant at level 2.5%

The primary drivers in these simplified models are *slope* and *contract age*. According to the *Odds ratio* for the slope, a client is nearly 6 times more likely to prepay fully when the yield curve slope is 1% compared to when it is 4%. A contract that has served 50% of its lifetime is nearly twice as likely to *fully* prepay as a contract that served 25%. The other variables displayed in Figure 10 are statistically significant, although their addition to the explanatory strength is limited.

Eventually prepayment projections can only be made when explanatory variables are projected first. When the inaccuracy of the projections of explanatory variables exceeds the explanatory strength they have, they do not add to forecast performance. For this reason the simplified models are justified. They contain the most influential drivers.

6.2. Application of results

6.2.1. Reliability of the explanatory variables

Although the variable *burnout* improves the fit on historical data, its coefficient seems to overstate its actual effect when projections of full prepayment rates for future years are simulated with a *constant slope*. This may be due to the distorting effect the increased spread between SWAP-FTP rates and the client rate has had (see subsection 4.2.3. Market-related variables) that occurred simultaneously.

The simplified models in the previous subsection, therefore, pose an attractive alternative, when prepayment projections are to be made. Appendix XV shows the sensitivity of the projected prepayment rates to the slope for the coming ten years.

6.2.2. Projection of future yield curve slopes

Long term fixed rates (five-year tenor) are typically forecast by so-called no-arbitrage models such as the *Hull-White* model[23]. These models use the prevailing forward rate curve as input as well as interest rate volatilities implied from market derivative products. Such a model constructs an interest rate tree for a certain long-term fixed rate (e.g. 5-year) which is used as input for the fixed-rate prepayment models.

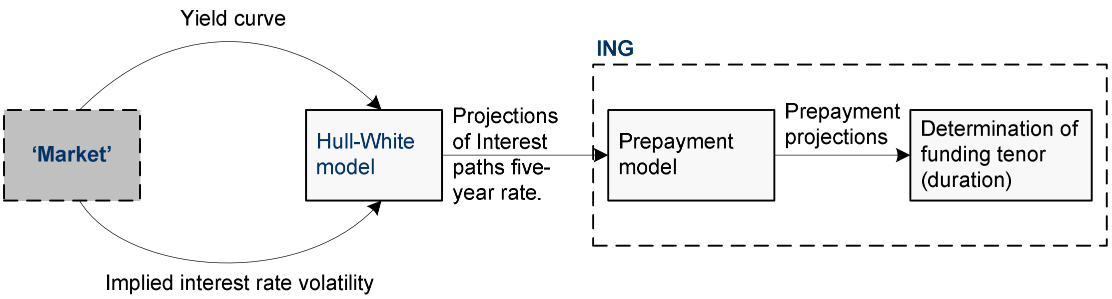


Figure 13: The process that leads to prepayment projections for fixed-rate mortgages starts at the forward rate curve and implied interest rate volatility from the 'market'.

<u>*The market's interest rate expectations*</u>

Prices and consequently implied volatilities are set by everybody in the market and at the same time by nobody in particular. The final result depicts a weighted average expectation of multiple scenarios and the market, therefore, symbolizes a black box from which implied volatilities and interest rate expectations are extracted. According to the *Efficient Market Theory* (for an overview, see Malkiel [2003]) the market is rational and as a collective has a correct view on inflation expectations which are consequently embedded in the yield curve. The following quote, however, also shows that the average market view may cloud the developments in the underlying scenarios and their likelihood.

[23] Alternatives to no-arbitrage models, are the 'equilibrium models' such as Vasicek and 'Cox, Ingersoll, and Ross', however, these models may result in output that is inconsistent with the prevailing structure of interest rates. No-arbitrage-models on the contrary take the initial term structure as given and define how it can evolve (Hull, 2008).

Stated by the Federal Reserve System (2002):
"long-term bond rates remained as high as 5 percent right up until the start of 1995. The failure of economists and financial markets to forecast Japan's deflationary slump in the early 1990s poses a cautionary note for other policymakers in similar circumstances: deflation can be very difficult to predict in advance."

The Bank of International Settlements states the following (Working Paper 188, [2005])
"The possibility remains, however, that changes in the composition of the BOJ's balance sheet caused by its market operations have had some effects on the term structure of interest rates."

The previous quote shows that even if the collective view of investors were true central bank interventions cloud this view by manipulating interest rates through monetary policy or *market operations* (e.g. purchasing of long-term government bonds.) The use of the market data without critical assessment is, therefore, questionable.

Projection of future roll-over prepayment rates
In order to project future prepayments in roll-over mortgages the variable *Slope(-1M)* would have to be forecast in a similar manner. A method for this does not exist yet, because the models such as *Hull-White* are intended to model *level* changes (parallel shifts) and do not capture dynamics such as *yield curve twists*. Another complicating factor is the fact that slopes are highly impacted by central bank interventions. As stated by the ECB herself *"Central bank liquidity management means supplying to the market the amount of liquidity consistent with a desired level of short-term interest rates"*[24]. It is this intervention that is hard or even impossible to embed in a model.

Factors not included in the model
1. Economic conditions (housing market turnover)
2. Tax regulation changes (deductibility of mortgage interest payments)
3. Big spread changes between SWAP-FTP and mortgage client rates (see subsection 4.2.3.)

To conclude

One should keep in mind that all the models and their drivers are based on historical data. The results consequently tell us which variables were significant prepayment drivers in the past. Bank capital- and liquidity requirements as well as the collective memory and financial sophistication of clients may not return to the old situation. One should question whether relationships derived from turbulent times, such as the period of 2006 until 2010, are a reliable predictor for the future or whether a regime shift has taken place. Conducting this research again with additional data in one or two years is, therefore, recommended.

[24] http://www.ecb.int/mopo/liq/html/index.en.html

6.3. Recommendations further research

As stated before, generalizing the specific results to other countries should not be done without care, because of the impact that tax regulation and penalties could pose on the clients' perceived refinance incentives.

With respect to *other portfolios* and *mortgages* it is recommended to conduct prepayment analysis on the *variable-spread* roll-overs when sufficient data (≥3 years) are available for analysis. Comparing this portfolio to the *fixed-spread* roll-overs makes the analysis of the sole impact of the FTP spread possible.

Furthermore, the explanatory strength of the yield curve slope on fixed-rate mortgage prepayments could be analyzed, because decreasing short-term floating rates pose an incentive for clients to *switch* to a floating-rate mortgage. Regarding other explanatory variables it is recommended to especially investigate the impact of client-related explanatory variables (*level of education, income, household size* etc.). A *"sophistication factor"* might have an impact on the client's propensity to identify and profit from beneficial prepayment opportunities[25].

Finally, analysis of the prepayment option value could be interesting, because for roll-over mortgage contracts the value is expected to depend on the volatility of long-term rates as well as short-term rates such as in callable bonds (Litterman, Scheinkman, Weiss [1991]). The absence of a penalty potentially makes the option value significant and relevant because of the high risks it poses.

Because it is eventually the goal to project future prepayment rates, research into yield curve slope modelling is advised. The slope is the main driver of full prepayments and slope expectations could consequently offer great insight in expected future full prepayment rates. Research into yield curve (shape) forecasting from a time-series perspective conducted by Yu and Salyards (2009) based on the model of Nelson and Siegel (1987) shows that this topic is on the agenda.

[25] In this paper the age of the oldest family member and a proxy for the family income were tested for explanatory strength, but were concluded insignificant. The income proxy data showed high data pollution (missing values) which made its use questionable even if the variable were found significant.

7. Reflection on personal learning goals

In subsection 7.1. I introduce the goals I wished to achieve during my Master thesis research assignment and the corresponding internship at the department *Corporate Market Risk Management Retail Netherlands* of ING. In subsection 7.2. I evaluate the goals and I discuss points of improvement.

7.1. Personal learning goals

1. I chose a quantitative assignment at a Risk Management department because I wanted to learn more about financial/ statistical analysis during my internship and thesis assignment. This as opposed to more *qualitative* oriented analyses.
2. Learning programming skills in order to perform financial analysis flexibly.
3. To conduct a broadly oriented research assignment methodologically.

1. Financial analysis

The Master thesis assignment for the study *Industrial Engineering & Management* could go several ways from my perspective. One could analyse a company's strategy and/or organization as was taught in courses such as *Organization & Strategy, Introduction to Industrial Engineering & Management,* and *Management of Technology*. On the other hand one could conduct an analysis within the field of the *track specialization*, in my case *Financial Engineering & Management*. Because I thought I would learn more from a quantitative financial analysis assignment and that this experience would be more appreciated by future employers I decided to pursue the latter.

2. Programming skills

Although initially it was not part of my goals, it became one when I discussed the assignment with the department CMRM from ING. The assignment required programming skills and I figured this would become important as a financial analyst, so I started some courses at home during the weeks prior to the internship.

3. Methodological research assignment

Because I have not done an academic Bachelor I realized I was behind on this aspect compared to class mates. My idea was to immediately at the start of the internship construct the research structure and thesis structure. I wanted to prevent doing an analysis which I eventually wrote a thesis around.

I enrolled in the Master because I was not satisfied with the things I had learned during my Bachelor study. I wanted to compensate for the lack of challenge and drive I had felt during the final years of my Bachelor study. When I made the decision to attend an academic Master programme I, therefore, committed and challenged myself to obtain the highest possible results. To obtain these results I knew one should not only perform data analysis well, but this analysis has to be put in context and has to be arrived at not by chance but by a methodological and scientific approach.

7.2. Reflection on personal learning goals

1. Financial analysis

I am satisfied with the extra insight I gained into statistical analysis and statistical hypothesis testing. This is not the actual financial knowledge or insight into financial products, but they are helpful tools with which I had little experience outside of textbook examples.

Besides this, I learned more about which (interest rate and liquidity) risks are hedged and in what manner in a bank such as ING. Moreover, I saw how funds flow through a bank, and learned the terminology used in the world of *Risk Management* and banking in general. Looking back I knew less details than I thought when I started the internship. But the general concepts and tools taught in the studies did give me the ability to quickly understand risks, hedge methods, calculation methods etc.

2. Programming skills

The first week I was given an existing *Visual Basic* programme to analyze and learn more. Because I had already completed some internet courses on *Visual Basic* I could read some of it. However, the programme used Object-oriented programming, which I had no experience with. During the first weeks I gained experience through asking questions and copying the methods applied in the existing programme. With this I made an automated hedge tool for a specific bonus/savings mortgage. Extra features of the tool were calculations of *Economic Capital* levels and pricing of the *bonus-feature*.

All-in-all I learned basic object-oriented programming skills which facilitates the database gathering and financial analysis of large quantities of data. Since I do not wish to become an expert in this field I am satisfied with my final level of understanding of programming right now.

3. Methodological research assignment

Instead of being pragmatic and starting the data analysis immediately I choose to structure my research. At first I drafted an (in my opinion) ideal approach by making the thesis structure to guide me accordingly. The reason I did this, is because I am normally pragmatic, and have the tendency to go right to the final analysis. The risk is that, in that case, I would not have weighted all the alternatives against each other. Moreover, one could have tunnel vision and pursue an initial idea, whereas literature research could have opened up the mind for alternatives.

Although I of course improved earlier analyses and writings because new insights emerged, I am pleased with the fact that, overall, I did follow my initially drafted structure. My *preliminary* research led to the correct formulation of specific research questions, whereas it also gave me insight into the nature of the problem at hand. Secondly, I performed the *literature* research to obtain the already existing models and variables, and the hypotheses (relationships among variables) drafted by other researchers. With these alternatives I started the descriptive analysis, where I described the variables, their distribution and where I formulated my own hypotheses based on the *literature* and *descriptive* research. And only then did I start the actual *explanatory* research and correspondingly the data analysis.

I feel this has protected me from observing relationships in data for which I afterwards formulated a hypothesis. Instead I formulated hypotheses based on the literature- and descriptive research which were either confirmed or rejected.

Points for improvement

I believe I could have planned my data collection better. Unfortunately, due to some struggling, I choose to finish an earlier phase first, before I started the meetings with the IT department to formulate the data requirements and prepayment definitions. Planning this better could have potentially saved me a month.

Besides that I wish I could have pressed deadlines by colleagues more strictly. The reason I did not do this was because I felt I did not have the authority as an intern. Moreover there were multiple urgent simultaneous projects being executed. That having said, I do wish to improve this, particularly because of the job (*Operations & IT Banking*) that I will start this year which will consist of change projects and improvement projects. To perform these activities well, getting other people's commitment and keeping deadlines is essential.

All-in-all I really feel I have gained much new financial insight and insight about how to scientifically and practically conduct a research project in a professional environment.

8. Definitions

Balanced sample	A sample containing an equal amount of prepayments and non-prepayments.
Causal relationship:	A change in the explanatory variable leads to a change in the dependent variable.
Correlation relationship:	Variables are correlated, no clear causal relationship necessarily exists.
Client rate	The mortgage interest rate offered to a client
External contracts:	Mortgages
External validity:	Refers to the generalizability of a research design.
FIRP	Fixed interest rate period (also referred to as interest rate typical maturity)
FTP spread	Funds transfer price. A liquidity spread placed upon the yield curve which together result in the internally charged prices of funds.
Household relocations:	Clients who move from one house to another are said to relocate.
Internal contracts:	Contracts used internally to aggregate and hedge risks arising from external contracts.
Internal validity:	The validity of assumed relationships between explanatory variables and the dependent variable. A relationship's nature (direct causal, intervening, non-causal, spurious), its direction and strength should be substantiated.
M.L.E.	Maximum Likelihood Estimation refers to an iterative algorithm that determines the coefficients that are most likely to have produced the data.
Loan notional:	The amount of money borrowed, i.e. the height of the mortgage.
Logit model:	A model that is based on the logistic equation and logistic regression. Dependent variable is bounded between zero and one (binary).
Log-likelihood:	A measure of the likelihood of obtaining the data when the model under the null-hypothesis is correct. The goal is, thus, to maximize the log-likelihood.
Prepayment rate:	The annualized rate in month *t* is defined by $p^{rate}_{annualized,t} = 1 - \left(1 - \frac{L_{prepaid,t}}{L_{outstanding,t}}\right)^{12}$
Probit model:	A model that is based on the standard normal cumulative distribution function. Dependent variable is bounded between zero and one (binary).
Reliability (of a model):	Measures the extent to which research will yield the same results and conclusions when performed multiple times and is, therefore, a measure for internal validity.
Reset period:	The amount of months or years between consecutive interest rate reset moments.
Refinance incentive:	The incentive caused by market interest rates, to refinance or prepay a mortgage.
SWAP-FTP:	The yield curve that business of ING can fund themselves against, and to which they add their cost spreads and profit margins.
Tenor:	Liquidity typical maturity (i.e. contract tenor).

9. Reference list

Alink, B.J., 2002, “Mortgage prepayments in the Netherlands”, *Twente University Press.*

Ambrose, B.W., LaCour-Little, M., 2001, Prepayment risk in adjustable-rate mortgages subject to initial discounts: Some new evidence, *Real Estate Economics*, **29, 2, 305-327.**

BIS Working Papers, 2005, Japan’s deflation problems in the financial system and monetary policy, *Working Paper*, **188.**

Box-Steffensmeier, J. M., Zorn, C. J. W., 1998, Duration Models and Proportional Hazards in Political Science, Working paper Annual Meeting of the Midwest Political Science Association 1998.

Calhoun, C.A., Deng, Y., 2000, A Dynamic Analysis of Fixed- and Adjustable-Rate Mortgage Terminations*, The Journal of Real Estate Finance and Economics,* **24, 1 & 2, 9-33.**

Charlier, E., Bussel, A. van, 2001, Prepayment behaviour of Dutch Mortgagors: An Empirical Analysis, *Real Estate Economics,* **31, 2.**

Christiansen, C., Lund, J., 2005, Revisiting the Shape of the Yield Curve: The Effect of Interest Rate Volatility, *Working Paper Aarhus School of Business.*

Cox, D.R., 1972, Regression Models and Life-Tables, *Journal of the Royal Statistical Society. Series B (Methodological)***, 34**, **2, 187-220.**

Cramer, J.S., 2003, “Logit Models from Economics and other Fields”*,* Cambridge University Press.

Daniel, J., 2008, A Variable-Rate Loan-Prepayment Model for Australian Mortgages, *Australian Journal of Management*, **33, 2, 277-305.**

EViews 6 User’s Guide II, 2007, *Quantitative Micro Software.*

Fanciulli, R., 2009, QUICK STUDY OF PREPAYMENT INCENTIVES (QUICK-SCAN), *CMRM research ING*, **version 5.**

Federal Reserve System, 2002, Preventing Deflation: Lessons from Japan’s Experience in the 1990s, *International Finance Discussion Papers*, **279.**

He, J., Liu, M., 1998, Mortgage Prepayment Behavior in a Market with ARMs only, *Journal of the Asian Real Estate Society*, **1, 1, 64-80.**

Hosmer, D. W., Lemeshow, S., 2000, "Applied Logistic Regression (Second edition)", *Wiley-Interscience Publication.*

Johnsson, T., 1992, A procedure for stepwise regression analysis, *SpringerLink Statistical Papers*, **33, 1, 21-29.**

Kroot, J., Slowinska, E., 2009, Prepayment Model Description, *CMRM research ING*, **version 1.1.**

Malkiel, B.G., 2003, The Efficient Market Hypothesis and Its Critics, *Journal of Economic Perspectives*, **17, 1, 59-82.**

Lindstrom, J., 2007, Liquidity typical prepayment for rollovers and savings, *CMRM Retail Research.*

Litterman, R., Scheinkman, J., 1991, Common factors affecting bond returns, *The Journal of fixed income,* **1, 1, 54-61.**

Litterman, R., Scheinkman, J., Weiss, L., 1991, Volatility and the yield curve, *The Journal of Fixed Income*, **1, 1, 39-53.**

Mitra, D., Golder, P.N., 2002, Whose Culture Matters? Near-Market Knowledge and Its Impact on Foreign Market Entry Timing, *Journal of Marketing Research*, **39, 3, 350-365.**

Nelson, C.R., Siegel, A.F., 1987, Parsimonious Modeling of Yield Curves, *The Journal of Business,* **60, 4, 473-489.**

Pliska, S.R., 2006, Mortgage Valuation and Optimal Refinancing, *Stochastic Finance,* **1, 183-196**

Schwartz, E.S., Torous, W.N., 1993, Mortgage Prepayments and Default Decisions: A Poisson Regression Approach, *Journal of American Real Estate and Urban Economics Association*, **21, 4, 431-449.**

Thoolen, C., 2009, The prepayment function, quick wins of work stream 2, *CMRM presentation.*

Yu, W.C., Salyards, D.M., 2009, Parsimonious Modeling and Forecasting of Corporate Yield Curve, *Journal of Forecasting*, **28, 1, 73-88.**

10. Appendix

Appendix I: Preliminary research; floating-rate vs. fixed-rate mortgages

1.1. Corporate Market Risk Management - Retail Netherlands

CMRM Retail NL transfers the interest rate- and liquidity risk to the department *Financial Markets Assets & Liabilities Management* (FM ALM), by the use of roll-overs (floating-rate coupon-bearing contracts) and internal deposit contracts (fixed-rate zero-coupon contracts). This way risks are aggregated at FM ALM. FM ALM nets these positions, which leads to risk mitigation and economies of scale. Whenever the remaining risk (after netting) is too great, external parties (the capital markets) are addressed in order to hedge the remaining exposure. This is done by the *Trading desk*.

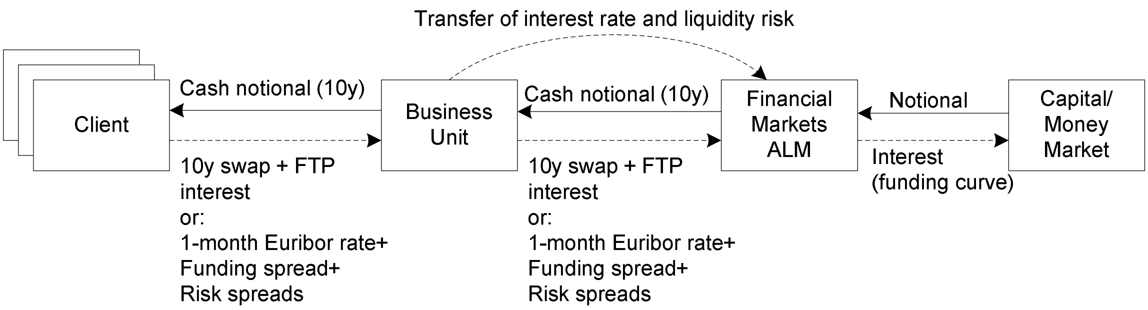


Figure 14: Aggregation of interest rate risk and liquidity risk from business units and transfer to Financial Markets. FTP = liquidity.

<u>*Hedge by CMRM Retail NL*</u>

CMRM hedges the exposure monthly. If the clients' repayment schedules were fixed, it would be possible to exactly replicate the client payments with the internal contracts (funding) up front, at least for the Fixed Interest Rate Period (FIRP).

However, clients usually have some flexibility in repaying the loan. They have the option, subject to certain conditions, to pay a higher amount than the contractual payment in a certain month, and consequently decrease the outstanding notional of the loan more than was planned (partial prepayment.) On the other hand, they may decide to fully prepay their loan notional and refinance their mortgage. For fixed-rate mortgages this is typically done at times of low interest rates to reduce monthly instalments and/or because a larger debt (and consequently house) could be serviced by a mortgagor's income against these lower interest rates, increasing the likelihood of household relocation. Clients, thus, have a prepayment *option*, which they will tend to exercise when it is in their own interest (exceptions are interest independent cases such as mortality, divorces, and some household relocations).

This option introduces uncertainty in the cash flows received from clients and must, therefore, be taken into account when a hedge is created (funding is attracted), that is intended to offset the *'expected'* future (interest-rate) exposure of the portfolio. For fixed-rate mortgages a penalty is often charged for full prepayment to compensate the bank (fully or partially) in case an economic loss is incurred. Mortgage prepayments of individual clients can also lead to gains for the bank, but on a portfolio basis the *option* clients have typically leads them to act (prepay) to their economic benefit and against the economic benefit of the bank.

1.2. Fixed-rate mortgage features and hedge

Standard fixed-rate mortgages are assumed to have interest rate prepayments made up off two components:

- *Interest rate independent:* Prepayments due to mortality, divorces and people who move to other houses.
- *Interest rate dependent prepayments*: lowering market rates form incentives to clients to refinance their mortgage potentially in combination with household relocation to avoid a penalty.

A lower boundary in prepayment rates (portfolio level) exists that indicates the (relatively stable) interest rate independent prepayments. An upper boundary indicates that some clients are *sticky*. Even if it were beneficial to refinance their mortgage they do not, i.e. they show interest rate insensitive behaviour.

<u>*Penalties*</u>

Penalties exist for prepayment and pose a hurdle to clients, although they are sometimes insufficient to compensate banks for their losses. In case no prepayment occurs, a new proposal will be offered to the client at the interest rate typical maturity to extend their contract into. This effectively limits certain risks, such as the risk of changing liquidity spreads, since all costs in renewed funding can be charged onto the client by offering an increased client rate.

The prepayment function projects future prepayment behaviour, based on historical data and interest rate projections. Funding is attracted based on the estimated prepayment behaviour in the future. The prepayment behaviour of an individual client is binary; a client either prepays or does not. On a large portfolio this smoothes out, which justifies the application of a continuous prepayment function and the consequent smoothed funding profile (see next page for a graphical representation.)

The bank funds fixed-rate mortgage loans with a maturity equal to the fixed-interest-rate period. After the fixed-interest-rate period the bank can refund and offer the client a new client rate. The interest-rate risk is therefore transferred onto the client. If a client, however, pays back the loan early (full or partial prepayment) the bank is faced with a liability with maturity *m* (i.e. its own funding) and with cash. This cash will have to be reinvested until time *m*, at the prevailing interest rate, to cover the remaining liability. This creates uncertainty and poses an interest rate risk.

Even if the client were to take on a new mortgage after prepayment, the new mortgage will most likely have a different maturity and FIRP, which means funding for the previous mortgage will have to be unwound by re-investing the prepayment proceeds.

Prepayment replication: funding

The exact prepayment behaviour of an individual client cannot be forecast, because it exhibits binary behaviour (either prepay or not) which depends on idiosyncratic (client-specific) factors as well. On a portfolio basis, however, the decisions of individual clients smooth out and result in a prepayment *rate*, which can be modelled more easily. Funding is, therefore, attracted to replicate the behaviour observed and modelled on a portfolio level, which is depicted in Figure 15. A lower prepayment rate leads to a longer profile and consequently requires a longer funding profile. An unexpected drop in the prepayment rate consequently requires a funding adjustment (lengthening).

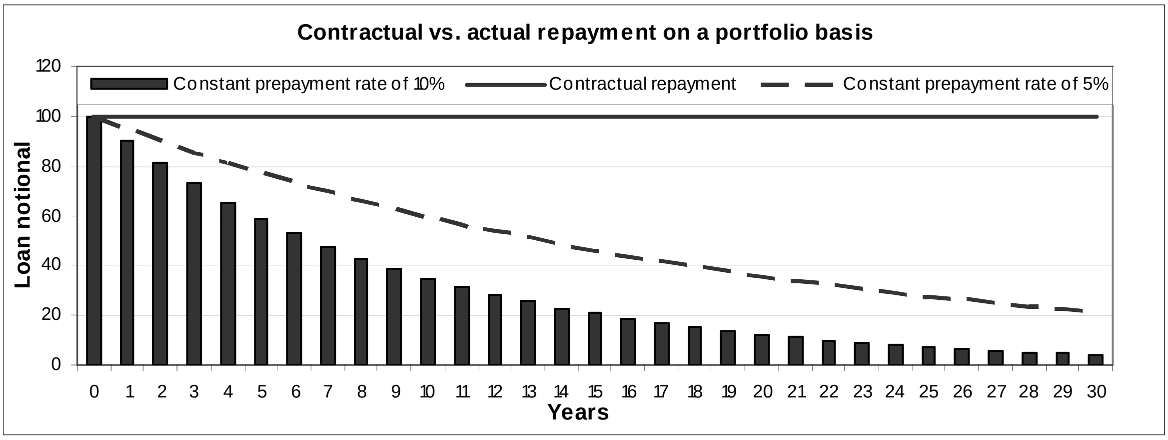


Figure 15: The binary prepayment behaviour of an individual client can be hedged by modelling the portfolio-level behaviour. This figure shows that only 20% of the portfolio loan notional is expected to still be present after 15 years in case of a constant annual prepayment rate of 10%.

If non-contractual prepayments were underestimated or overestimated the actual amount of prepayments would lead the funding (attracted and locked in) to become either too long or too short (in duration) and un-hedged cash flows and consequently interest-rate risk and liquidity risk would arise. Adjusting the hedge portfolio to offset the un-hedged cash flows may constitute an economic gain or loss for the bank.

Clients' prepayment behaviour should be analyzed in order to project prepayment rates accurately, such that the expected prepayments are accounted for in advance by adequate funding.

1.3. Floating-rate mortgage features and hedge

A mortgage is considered a floating-rate mortgage when the fixed interest rate period (FIRP) is equal to or shorter than one year. The reset period (equal to FIRP) can be 1 month, 3 months, 6 months, 9 months or 12 months. Unlike with fixed-rate mortgages, a client is not offered a new client rate after each FIRP. Instead the client rate is automatically reset, based on the contract and Euribor rate in the market, after each FIRP. This eliminates the possibility of charging higher costs to the client, such as risk or liquidity costs.

Because the tenor of funding is based on the contract tenor, no refunding occurs as long as a client serves his contract up until maturity. Higher funding costs are in such cases irrelevant. Floating-rate mortgages are contractually (excluding unforeseen prepayments) repaid by a single repayment at maturity ('bullet'). The prepayment option is embedded (optionality risk[26]), and no penalty is charged.

Product features

When a client prepays his loan notional fully the interest rate risk is far lower for a floating-rate mortgage than for a fixed-rate mortgage. The risk of a market interest rate change is now restricted to one interest reset period, and *m* monthly mortgage payments, where *m* is the amount of months in a reset period. The roll-overs that charge a fixed spread over Euribor have, however, an additional risk.

The client rate for these mortgages consists of two components, namely:

- Euribor
- Fixed spread composed of
    - Risk spreads (credit risk, market risk, etc.)
    - Production costs
    - Liquidity spread (FTP spread paid on funding)
    - Profit margin

The fixed spread a client is charged with should cover multiple components, such as production costs, funding spreads and risk spreads. Furthermore, a client is allowed to fully prepay without penalty which eliminates a big part of the hurdle that a prepayment might otherwise pose.

Hedge method

A *'cash flow'*-hedge is performed for floating-rate mortgages. This means the 'expected' cash flow profile is offset by internal contracts. The actual cash flows are uncertain due to the uncertainty that prepayments cause[27]. The most important difference with the hedge (funding) of fixed-rate mortgages is the fact that funding is attracted with a tenor based on the contract tenor instead of the fixed-interest-rate period.

[26] See Appendix I: "Sources of interest rate risk" for definitions of the 4 sources of interest rate risk.

[27] The floating Euribor rate itself creates uncertainty in cash flows but this is offset/ hedged by another floating internal contract.

1.5. Market developments and model risk

"Generally: in times of crisis, model risk is higher as models in that case do not predict very well" (Thoolen, 2009). This did apply to roll-over mortgages as well. Before 2007 liquidity was abundant and cheap. Euribor and the margin, consequently, made up the main part of the client rate. When the costs of liquidity in the market increased significantly, due to the credit crisis, its part in the total client rate increased. The prepayment risk arising from these developments were not modelled.

FTP spread height and volatility

The FTP spread is determined based on the liquidity typical tenor[28] and gives the mark-up for non-amortizing (i.e. bullet) contracts. The internally charged interest rates (SWAP curve) are, thus, increased to reflect market conditions, however, not on a one-to-one basis per se. The result is the SWAP-FTP curve. Nevertheless, the deviation of the costs charged to the business units (SWAP-FTP) from *Financial Market's* actual funding costs is restricted, because it would otherwise pose an internal arbitrage opportunity[29] and would form an incentive for longer or shorter funding and investments.

For this reason the FTP spread will be interpreted as a proxy for liquidity costs, and it is consequently assumed that clients are offered client rates with comparable liquidity spreads by competing mortgage lenders. This is relevant since the incentive to prepay depends among others on offered client rates in the market.

[28] 'Liquidity typical' maturity can be defined as the moment in time ING can get the funds back, or the next moment in time, ING will have the opportunity to re-charge the client with the new price of liquidity.

[29] Too high FTP spreads would give business units the opportunity to offer relatively high interest rates on savings, which are not in line with the market and would cause high inflows of expensive funding. Too low FTP spreads would make it possible for business units to get cheap funds from de department *Financial Markets*, and consequently to offer relatively low interest rates on mortgages, which would result in low-yielding assets. Both situations are undesirable.

1.6. Prepayment, reinvestment and funding costs

For fixed-rate mortgages funding is attracted with a tenor based on the fixed-interest-rate period (FIRP), which means, that funding is rolled over after each FIRP. When this happens a client is offered a new client rate and in case necessary higher funding costs can be transferred onto the client. Prepayment before the end of the FIRP is (partially) covered by a penalty payment.

For floating-rate mortgages, funding is attracted with a tenor based on the contract tenor (projected prepayments shorten the funding tenor). Unexpectedly high full or partial prepayment rates would require the bank to re-invest the proceeds from prepayment at prevailing liquidity spreads and with a tenor equal to the remaining time to the funding contract end date.

If spreads at the time of prepayment are lower than when funding was initially attracted, these proceeds from prepayment will be insufficient to unwind/offset existing funding completely, which would constitute an economic loss whereas the opposite would constitute a gain. Since the tenor of the re-investment is shorter than the tenor of initial funding (at the time of contract origination), a lower liquidity spread is likely to apply even when spreads are unchanged, because the liquidity costs (FTP spreads) are typically an increasing function of the tenor. Therefore the focus, from a market risk perspective, lies on the cost of liquidity in the market, i.e. liquidity spreads.

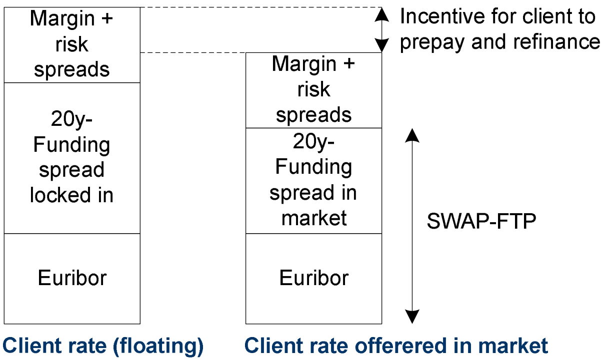


Figure 16: A decreasing funding spread (which includes FTP) poses clients who have a (higher) fixed spread embedded in their client rate an incentive to refinance (prepay).

Fixing the spread charged to clients introduced a new risk that depends on the height of the FTP spread and that arises from its volatility. The economic gain or loss from adjusting existing funding becomes potentially larger as the funding spread increases, but is only triggered when funding spreads decrease. A constant high or low liquidity spread, alone thus poses no risk because it is priced into client rates, but together with funding spread volatility it does. A more detailed explanation of how FTP spread volatility can lead to gains and losses can be found in section 1.3. “Problem analysis”.

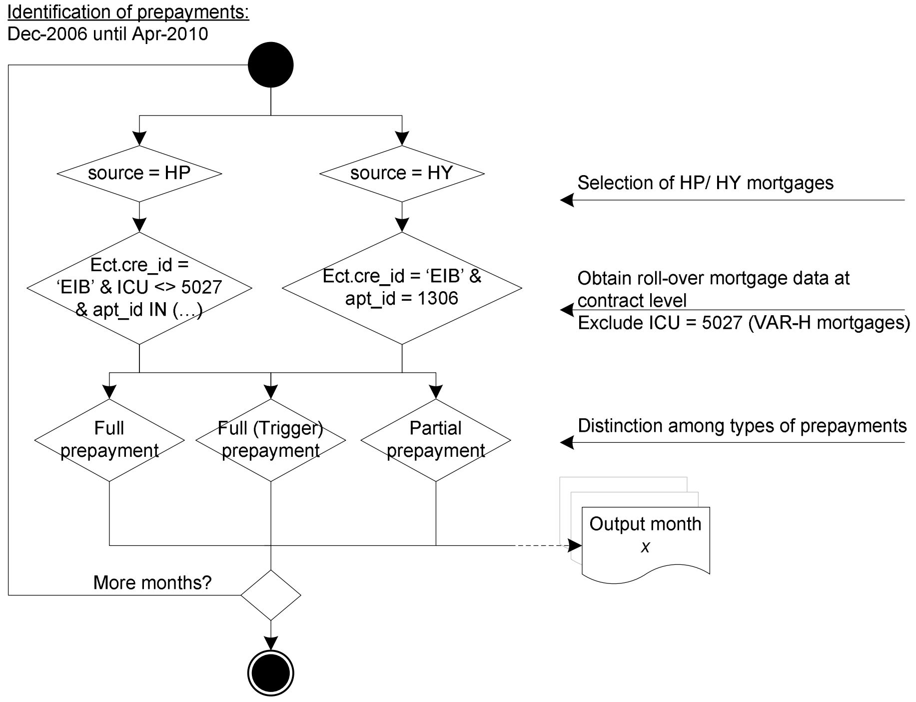
Appendix II: Sources of interest rate risk

Interest rate risk arises from the exposure of a bank's financial condition to adverse movements in interest rates.

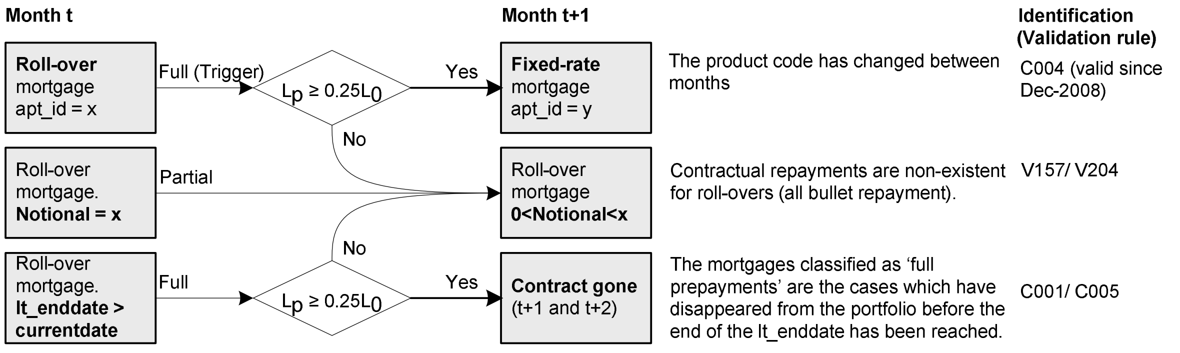
Sources of interest rate risk:

- *Repricing risk:* Repricing risk arises from timing differences in the maturity (for fixed-rate) and repricing (for floating-rate) of bank assets, liabilities and off-balance sheet positions.
- *Yield curve risk:* Yield curve risk is the risk that the bank is subjected to both changes in the slope and shape of the yield curve. Yield curve risk arises when unanticipated shifts of the yield curve have adverse effects on a bank's income or underlying economic value.
- *Basis risk:* Basis risk arises from imperfect correlation in the adjustment of the rates earned and paid on different financial instruments with otherwise similar re-pricing characteristics.
- *Optionality:* Optionality risk arises from the options embedded in many banking positions. These options provide the holder the right, but not the obligation, to buy, sell, or in some manner alter the cash flow of an instrument or financial contract. Interest rate risk resulting from optionality could arise because the bank has not fully hedged this risk or due to the fact that it is impossible to perfectly model client behaviour.

Appendix III: Prepayment data identification



The data cannot be selected by gathering all mortgages of a certain ICU, because this ICU was not available before November 2008. All 'apt_id's that existed under the ICUs 4910 en 4909 are tracked. The roll-over mortgages are selected by adding the criterion ect.cre_id = 'EIB' and VAR-H mortgages are excluded by excluding all mortgages with ICU=5027.



The different types of prepayments can be identified by observing specific changes in contract-features or by contracts that disappear completely under certain conditions. Some checks have been programmed in REALM through validation rules and checks. These validation rules are operational for roll-overs as of Dec. 2008. The prepaid amount and the initial notional are denoted by L_p and L_0 respectively.

Because the validation rules for roll-overs were not operational before December 2008, the prepayments have been deduced manually by comparing the portfolio from month to month by using the *INDEN_CD* field (identification code) to track a contract over time. If a contract dropped out of the portfolio for 2 consecutive months it has been labelled as a *full prepayment*. If a contract has not disappeared, but the outstanding notional has decreased, it has been labelled a *partial prepayment*. All contracts that were labelled *full prepayment* that had a relative age < 4% were excluded from analysis, because they are assumed to be data errors. Partial prepayments smaller than € 200 have been excluded from analysis because of their minor impact on the portfolio outstanding notional. They may be data errors and their low values create a distorting effect to the results.

Appendix IV: Roll-over portfolio information

Portfolio composition

Reset period	Fraction
1 month	90.7%
3 months	1.3%
6 months	1.9%
12 months	6.1%

Figure 17: 90.5 percent of the total amount of the fixed-spread roll-over contracts has an interest reset period of one month.

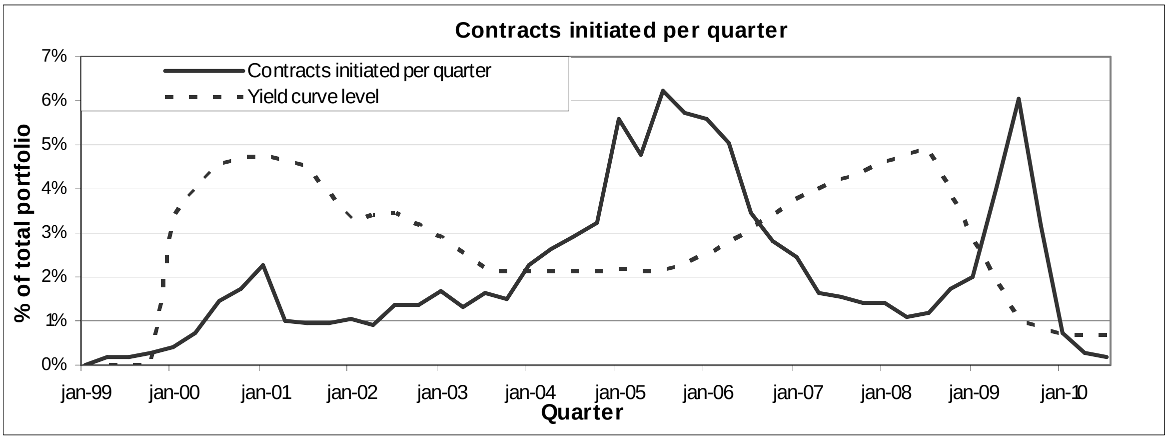


Figure 18: The fraction of contracts in the portfolio with a certain origination date (quarter). Only contracts originated after 1999 are included. Times of low yield curve levels (three-month Euribor) as expected, tend to show high roll-over mortgage production.

Appendix V: Mortgage prepayments in the Netherlands

Optimal call vs. Empirical observations

> *"U.S. literature on mortgage prepayment makes a distinction between 'optimal call' and empirical models. Optimal call prepayment models assume that borrowers with similar loans will react in a similar fashion to changes in market interest rates. These models treat the prepayment option in a mortgage as a call-option and assume that the option-holder (borrower) will exercise this option optimally. However, such behaviour is not consistent with empirical evidence, as Archer, Ling and McGill (1995) noted."* [Alink, 2002]

Quote 1: The optimal-call-model vs. empirical observations approach.

An alternative is working with empirical observations. This means the historical behaviour of clients is used as a proxy for future behaviour. This approach is adopted in this research.

> 1. ***Seasonality***: Captures seasonal differences in house sales over the course of a year.
> 2. ***Refinancing incentive***: Low market rates form an incentive to refinance loans.
> 3. ***Seasoning***: Refers to the age of the contract. Prepayment rates are typically low shortly after origination and then rise in the early years of a contract.
> 4. ***Burnout***: Prepayment activity tends to decrease for older mortgages. This may be due to:
>   - Heterogeneity in clients in a mortgage pool.
>   - Decreasing interest rate payments over time, in the case of annuity (amortizing) contracts, lower the benefit from refinancing.
>
> Burnout:
> *"The phenomenon they observed was a gradual slowing down of the rate of prepayment for pools of mortgages from which a substantial proportion of loans had already prepaid. The decline in prepayment rate was not steady - it could be overwhelmed by acceleration caused by a decline in interest rates. However, even when prepayment rates reacted to lower interest rates, burned-out pools were seen to prepay more slowly than they had in the same circumstances earlier in their lives. This phenomenon derives from the heterogeneity of borrowers, and a simple example will illustrate. Suppose that borrowers have fixed constant probability of prepayment in a particular month, given that they have not already prepaid. Suppose further that all borrowers prepay at one of two rates: they are either fast prepayers (10% per month) or slow prepayers (2% per month). If a pool of borrowers is initially 50% fast prepayers and 50% slow prepayers, how will the overall prepayment rate for the pool evolve? Because it is a 50%–50% mix, the initial prepayment rate will be 6%. But many more of those who prepay will be fast prepayers, so the proportion of fast prepayers will decline, and the proportion of slow prepayers will rise. As a result, the prepayment rate for the pool will decline over time and eventually approach 2%. The pool will become burned out."* [Arden Hall, 2000]

Quote 2: The classification of prepayments into four components and explanation of the phenomenon "Burnout" by Arden Hall.

Appendix VI: Volatility and the yield curve

Volatility and the yield curve

Additional issues addressed by in the article "Volatility and the yield curve" Litterman, R., Scheinkman, J., Weiss, L., 1991] are summarized in this appendix. The authors conclude that higher interest-rate volatility tends to have two opposing effects, (1) through a multiplicative process it increases expected future short rates, and (2) by increasing the convexity of the discount factor function[30], it reduces yields. The first effect dominates at the short end of the curve, whereas it is dominated at the long end of the curve. An increase in volatility thus moves the peak of the yield curve to the left and lowers the long-term rates, i.e. increases the *curvature* of the yield curve.

Measure of volatility and curvature

The authors go on to say that a measure of volatility can be obtained from the yield curve by considering the yield spread on a *butterfly*, which is a portfolio that is long two bonds of different maturities and short a bond of intermediate maturity. A regression of implied volatility on the levels of the one-month, three-year, and ten-year zero yields led to the following model: $V(t) = 0.497\text{-}0.065x_1(t) + 0.157x_2(t) - 0.126x_3(t)$, where x_1,x_2, and x_3 are the one-month, three-year, and ten-year zero yields, respectively[31]. This regression model could, thus, help quantify the implied volatility and indirectly the curvature of the yield curve, to eventually test the significance of the *curvature* as an explanatory variable for prepayments. However, it was found that this measure of implied volatility does not seem to be a good indicator of convexity for the yield curves observed over the time period spanning 2000 through 2010.

[30] This is not straightforward, but is explained in more detail in the article "Volatility and the yield curve" by Litterman, Scheinkman and Weiss.

[31] *"for weekly data from January 1, 1984, through June 23, 1998 (…) this regression explains 70% of the variation in V, with a standard error of estimate equal to 0.03"*

Appendix VII: Prepayments rates Dec. 2006 - May 2010

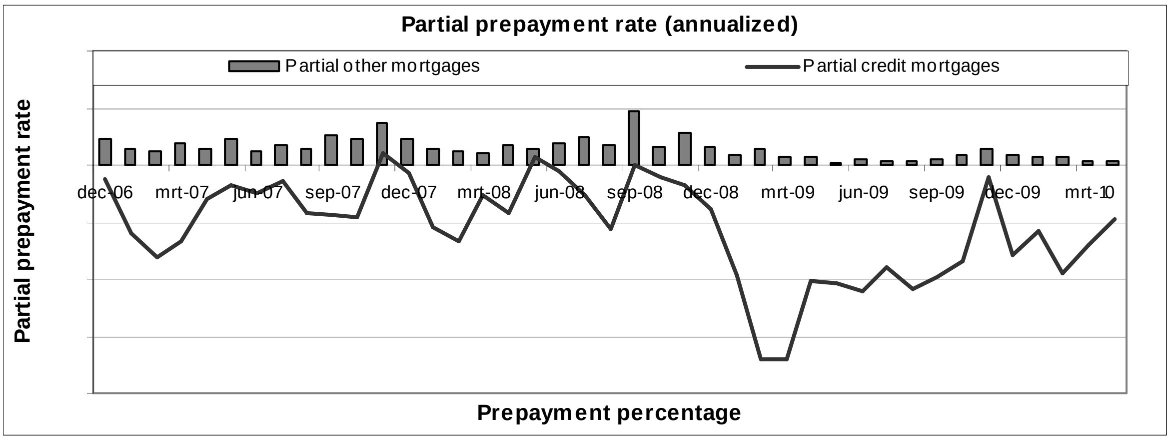


Figure 19: The partial prepayment rates of source HP are displayed. A distinction is made between credit mortgages and other mortgages, because of their potentially different prepayment drivers.

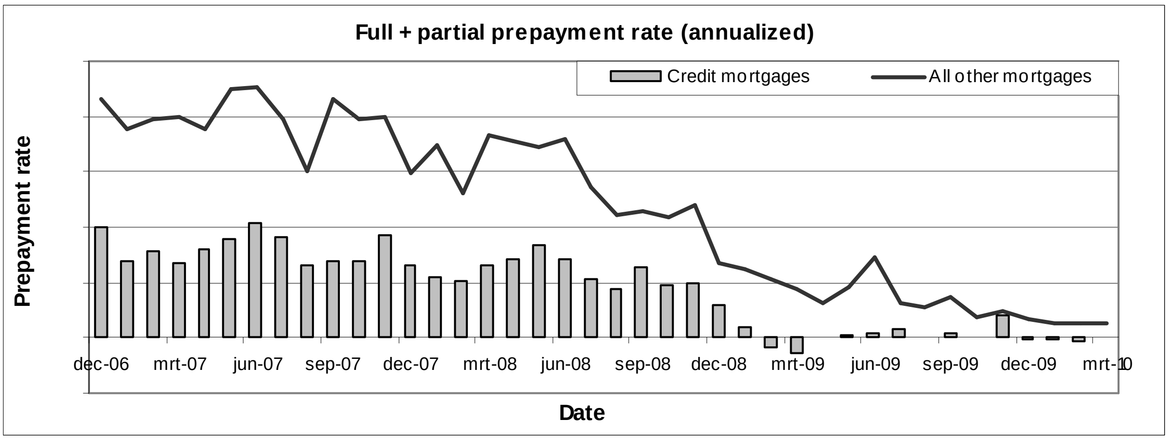


Figure 20: The full prepayments are added to the partial prepayments. For the credit mortgages this leads to lower overall prepayment rates, because of the typically negative partial prepayments.

Appendix VIII: Operationalization of explanatory variables

The standard refinance incentive compares the client spread (over Euribor) locked in with the client spread offered in the market. Due to the increased FTP spreads this incentive has typically been negative over the passed years.	$R.I._t = \frac{Eur_{1M,t} + Clientspread_{locked-in}}{Eur_{1M,t} + Clientspread_{market}} 100\% - 1$
The variable that measures the incentive to switch to a fixed-rate mortgage is referred to as the *adjusted refinance incentive*. $X_{m,t}$ denotes the SWAP-FTP rate of tenor *m*, at time *t*.	$A.R.I._t = \max\left[\frac{Eur_{1M,t} - X_{10Y,t}}{Eur_{1M,t} + Clientspread} 100\%; 0\right]$
Interest rate volatility	$\sigma_{3M,t} = stdev(X_{3M,t-k}, X_{3M,t-k+1}, \ldots, X_{3M,t}), k = 12$
Level of the yield curve:	$l_t = X_{3M,t}$
Slope of yield curve:	$s_t = X_{10Y,t} - X_{3M,t}$
Curvature of yield curve:	$c_t = X_{3Y,t} - (wX_{3M,t} + (1-w)X_{10Y,t})$
(Initial) loan notional	$N_{0,i} / 100{,}000$
(Relative) age	$age_{t,i} = 1 - \frac{remaining\,months\,lt\_enddate_{t,i}}{initial\,maturity}$
Burnout (pool factor):	$PF_t = \frac{\sum_{i=1}^{n} Notional_{outstanding,i,t}}{\sum_{i=1}^{n} Notional_{initial,i}} = \frac{Pool_{balance,t}}{Pool_{initial}}$
Burnout (weighted-average portfolio age in months):	$Burnout_{portfolio,t} = \frac{\sum_{i=1}^{n} Age_{i,t} * Notional_{initial,i,t}}{Notional_{initial,Portfolio,t}}$
Dummy variable Winter "dWinter":	$dWinter = \begin{cases} 1 & \text{if month} = Nov/Dec/Jan/Febr \\ 0 & \text{otherwise} \end{cases}$
Dummy variable JanMar "dJanMar"	$dWinter = \begin{cases} 1 & \text{if month} = Jan/Febr/Mar \\ 0 & \text{otherwise} \end{cases}$
Dummy variable "dlownot"	$dWinter = \begin{cases} 1 & \text{if } outst\_notional < 0.15 * \text{initial notional} \\ 0 & \text{otherwise} \end{cases}$

Appendix IX: Historical values of contract-related variables

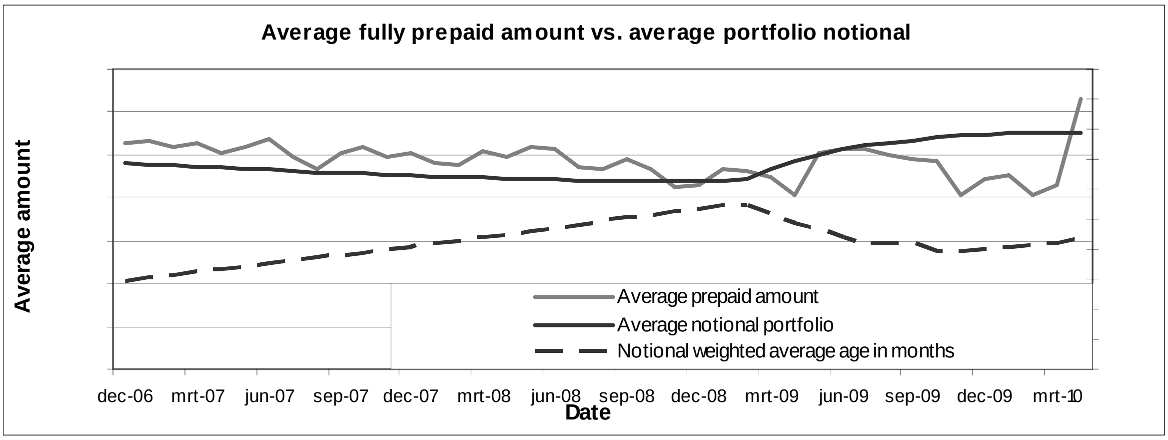


Figure 21: The large inflow of new mortgages (production) in 2009 has increased the average notional of the portfolio as well as decreased the notional-weighted-average age of the portfolio.

Because of the low short-term interest rates in 2009 the average initial notional of the new production (new inflow of mortgages) was higher than the average notional of the existing portfolio. This phenomenon reduced the overall average portfolio notional. Because the older mortgages are overrepresented in the full prepayments this leads to the average full prepaid notional to being lower than the average notional of the overall portfolio. One should, however, be careful drawing conclusions with respect to causal relationships.

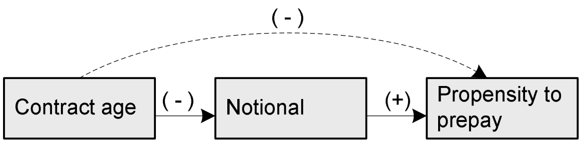


Figure 22: In this portfolio and in this situation the older contracts on average have lower notionals, and have a higher propensity to be prepaid because the new mortgage production is so recent (young) that in the overall prepayments the low notionals (of older contracts) are overrepresented.

Appendix X: Historical values of explanatory variables

Figure 23 shows the interest rate volatility clients have been exposed to. This is compared to the curvature measure of the SWAP FTP yield curve.

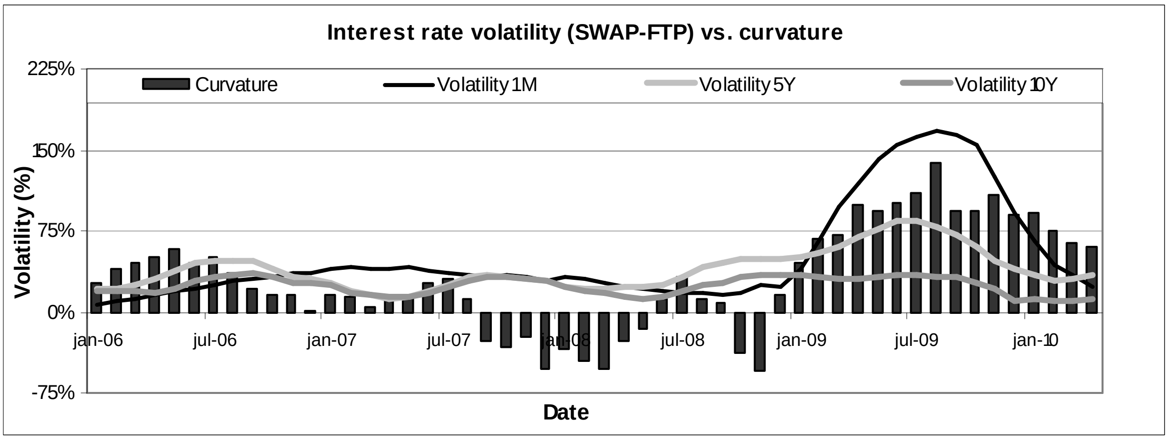


Figure 23: The volatilities at each date are calculated from the 12 previous months. The curvature is calculated according to the measure suggested by Christiansen and Lund Christiansen, C., Lund, J., 2005]. Volatility of short term rates is higher than that of long-term rates.

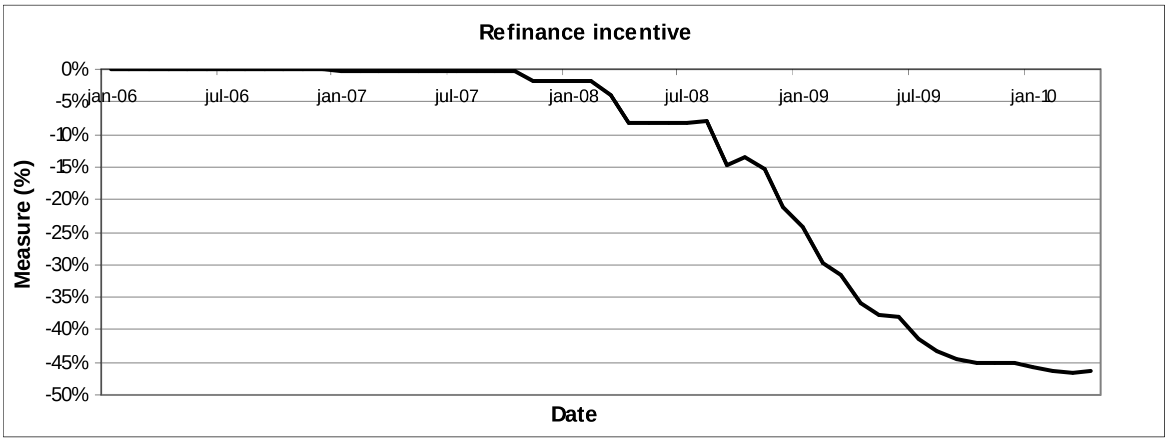


Figure 24: The incentive to refinance a roll-over depends on the locked-in client spread over Euribor and the prevailing client spread offered in the market. The client spread offered in the market is based on the 6-month instead of the 3-month reset period, which causes an inaccuracy.

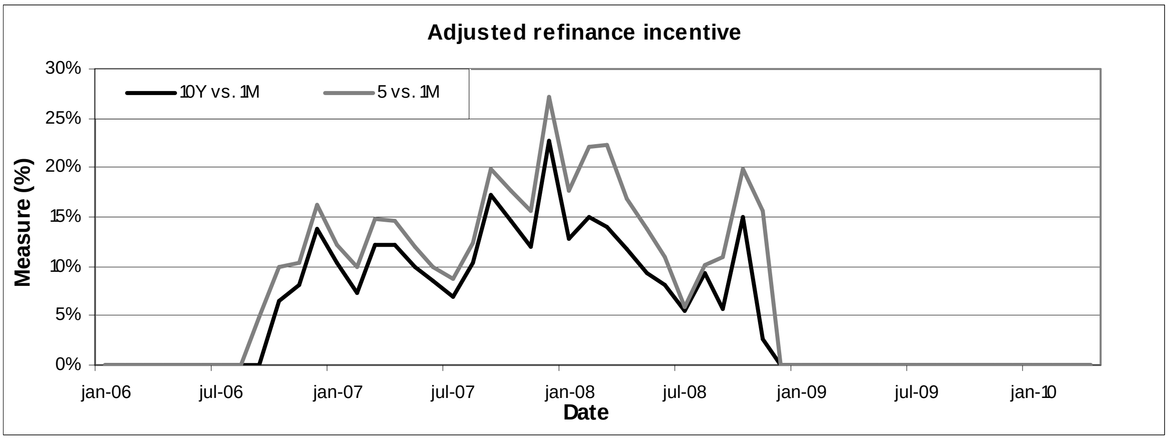


Figure 25: The adjusted refinance incentive adds a factor a (= 20%) to the potential decrease in client rate to account for the fact that clients prefer certainty over uncertainty whenever the ten-year-fixed rate is equal to or close to the three-month-floating rate.

Appendix XI: Pair wise correlation matrices

Full prepayments excluding credit mortgages

	Age	Age^2	Burnout	Curvature	Dlownot	DJanMar	DWinter	Notional	Slope(-1M)	Vol_1M
Age-squared	**0.91**									
Burnout	0.21	0.11								
Curvature	0.07	0.0.6	0.02							
Dlownot	-0.01	0.00	0.03	0.01						
DJanMar	-0.03	-0.02	-0.08	-0.18	0.00					
Dwinter	-0.02	0.00	-0.07	-0.19	0.00	**0.37**				
Notional	-0.10	-0.09	-0.04	0.01	-0.13	-0.01	-0.01			
Slope(-1M)	0.12	0.09	0.17	**0.89**	0.01	-0.17	-0.11	0.01		
Volatility_1M	0.12	0.09	0.19	**0.83**	0.01	-0.16	-0.12	0.01	**0.94**	
Level	-0.10	-0.09	-0.12	**-0.87**	-0.01	0.03	-0.03	-0.01	**-0.97**	**-0.93**

Figure 26: IF the variables 'curvature', 'volatility1M', and 'age-squared' are omitted, multi-collinearity will not be an issue.

Partial prepayments excluding credit mortgages

	Age	Burnout	DJanMar	DWinter	Notional	Vol_1M
Age						
Burnout	0.11					
DJanMar	-0.03	-0.07				
DWinter	-0.01	0.04	**0.41**			
Notional	-0.19	-0.01	0.00	0.01		
Slope(-1M)	0.04	0.07	0.01	-0.02	0.08	
Volatility_1M	0.04	0.14	-0.23	-0.14	0.03	0.70

Figure 27: The pair wise correlation coefficients are important to prevent multi-collinearity.

Appendix XII: Model evaluation criteria

Akaike information criterion

"It is grounded in the concept of entropy, in effect offering a relative measure of the information lost when a given model is used to describe reality and can be said to describe the tradeoff between bias and variance in model construction, or loosely speaking that of accuracy and complexity of the model." The lower the value the better the model.

Schwarz criterion

This criterion measures the same as Akaike but corrects stricter for the amount of explanatory variables.

McFadden R-squared

McFadden R-squared is the likelihood ratio index Akaike(…) as the name suggests, this is an analog to the R^2 reported in linear regression models. It has the property that it always lies between zero and one. [EViews 5 Users Guide]. A value of one indicates a perfect model fit, whereas a value of zero indicates that it has no explanatory strenght.

LR statistic

LR statistic tests the joint null hypothesis that all slope coefficients except the constant are zero (…). This statistic (…) is used to test the overall significance of the model. [EViews 5 Users Guide].

Hosmer-Lemeshow *Goodness-of-fit test*

The H-L test (Hosmer and Lemeshow) allows one to perform Pearson chi-squared tests of goodness-of-fit. "*The idea underlying these tests is to compare the fitted expected values to the actual values by group. If these differences are "large", we reject the model as providing an insufficient fit to the data". [EViews 5 Users Guide].*

Predictive power

This tests which fraction of a sample of contracts the model predicts well. 50% of the sample contracts have involve a prepayment and the other 50% does not. If the model predicts a probability of prepayment bigger than 0.5, and the contract indeed involves a prepayment the prediction is considered a succes. If the model predicts a probability smaller than 0.5 and the contract indeed does not involve a prepayment the prediction is again considered a succes. The final overall fraction of succesful predictions is a proxy for the predictive power and the identification of significant drivers.

a) Sensitivity

Sensitivity measures the proportion of actual prepayments which are (correctly) predicted as prepayment.

$$Sensitivity = \frac{\#true\ \ positives}{\#true\ \ positives + \#false\ \ negatives}$$

b) Specificity

Sensitivity measures the proportion of non-prepayments which are (correctly) predicted as non-prepayment.

$$Specificity = \frac{\#true\ \ positives}{\#true\ \ positives + \#false\ \ negatives}$$

Appendix XIII: Model specification and statistics

Full prepayments - portfolio excluding credit mortgages

Method: ML - Binary Logit (Quadratic hill climbing)
Included observations: 28536
Convergence achieved after 4 iterations
Covariance matrix computed using second derivatives

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	1.761018	0.079806	22.06624	0.0000
AGE	2.433088	0.141164	17.23584	0.0000
SLOPELAG1	-57.73463	1.053036	-54.82686	0.0000
BURNOUT	-0.038251	0.001910	-20.02327	0.0000
NOTIONAL	0.121471	0.016193	7.501599	0.0000
DWINTER	-0.200368	0.026881	-7.453959	0.0000

Mean dependent var	0.500000	S.D. dependent var	0.500009
S.E. of regression	0.460384	Akaike info criterion	1.224021
Sum squared resid	6047.033	Schwarz criterion	1.225757
Log likelihood	-17458.33	Hannan-Quinn criter.	1.224579
Restr. log likelihood	-19779.65	Avg. log likelihood	-0.611800
LR statistic (5 df)	4642.641	McFadden R-squared	0.117359
Probability(LR stat)	0.000000		

Obs with Dep=0	14268	Total obs	28536
Obs with Dep=1	14268		

Method: ML - Binary Logit (Quadratic hill climbing)
Included observations: 28536
Prediction Evaluation (success cutoff C = 0.5)

	Estimated Equation			Constant Probability		
	Dep=0	Dep=1	Total	Dep=0	Dep=1	Total
P(Dep=1)<=C	6868	2078	8946	14268	14268	28536
P(Dep=1)>C	7400	12190	19590	0	0	0
Total	14268	14268	28536	14268	14268	28536
Correct	6868	12190	19058	14268	0	14268
% Correct	48.14	85.44	66.79	100.00	0.00	50.00
% Incorrect	51.86	14.56	33.21	0.00	100.00	50.00
Total Gain*	-51.86	85.44	16.79			
Percent Ga...	NA	85.44	33.57			

	Estimated Equation			Constant Probability		
	Dep=0	Dep=1	Total	Dep=0	Dep=1	Total
E(# of Dep=0)	8220.64	6047.36	14268.00	7134.00	7134.00	14268.00
E(# of Dep=1)	6047.36	8220.64	14268.00	7134.00	7134.00	14268.00
Total	14268.00	14268.00	28536.00	14268.00	14268.00	28536.00
Correct	8220.64	8220.64	16441.28	7134.00	7134.00	14268.00
% Correct	57.62	57.62	57.62	50.00	50.00	50.00
% Incorrect	42.38	42.38	42.38	50.00	50.00	50.00
Total Gain*	7.62	7.62	7.62			
Percent Ga...	15.23	15.23	15.23			

*Change in "% Correct" from default (constant probability) specification
**Percent of incorrect (default) prediction corrected by equation

Partial prepayments - portfolio excluding credit mortgages

Method: ML - Binary Logit (Quadratic hill climbing)
Included observations: 7744
Convergence achieved after 5 iterations
Covariance matrix computed using second derivatives

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	-0.194651	0.051793	-3.758264	0.0002
AGE	3.673921	0.207041	17.74491	0.0000
SLOPELAG1	-15.12276	1.544283	-9.792745	0.0000
NOTIONAL	-0.167621	0.030463	-5.502427	0.0000
DLOWNOT	2.859170	0.327547	8.729032	0.0000
DJANMAR	-0.318826	0.053628	-5.945126	0.0000

Mean dependent var	0.500000	S.D. dependent var	0.500032
S.E. of regression	0.472692	Akaike info criterion	1.275141
Sum squared resid	1728.962	Schwarz criterion	1.280529
Log likelihood	-4931.344	Hannan-Quinn criter.	1.276988
Restr. log likelihood	-5367.732	Avg. log likelihood	-0.636796
LR statistic (5 df)	872.7748	McFadden R-squared	0.081298
Probability(LR stat)	0.000000		
Obs with Dep=0	3872	Total obs	7744
Obs with Dep=1	3872		

Method: ML - Binary Logit (Quadratic hill climbing)
Included observations: 7744
Prediction Evaluation (success cutoff C = 0.5)

	Estimated Equation			Constant Probability		
	Dep=0	Dep=1	Total	Dep=0	Dep=1	Total
P(Dep=1)<=C	2825	1779	4604	3872	3872	7744
P(Dep=1)>C	1047	2093	3140	0	0	0
Total	3872	3872	7744	3872	3872	7744
Correct	2825	2093	4918	3872	0	3872
% Correct	72.96	54.05	63.51	100.00	0.00	50.00
% Incorrect	27.04	45.95	36.49	0.00	100.00	50.00
Total Gain*	-27.04	54.05	13.51			
Percent Ga...	NA	54.05	27.01			

	Estimated Equation			Constant Probability		
	Dep=0	Dep=1	Total	Dep=0	Dep=1	Total
E(# of Dep=0)	2136.94	1735.06	3872.00	1936.00	1936.00	3872.00
E(# of Dep=1)	1735.06	2136.94	3872.00	1936.00	1936.00	3872.00
Total	3872.00	3872.00	7744.00	3872.00	3872.00	7744.00
Correct	2136.94	2136.94	4273.87	1936.00	1936.00	3872.00
% Correct	55.19	55.19	55.19	50.00	50.00	50.00
% Incorrect	44.81	44.81	44.81	50.00	50.00	50.00
Total Gain*	5.19	5.19	5.19			
Percent Ga...	10.38	10.38	10.38			

*Change in "% Correct" from default (constant probability) specification
**Percent of incorrect (default) prediction corrected by equation

Appendix XIV: Model evaluation results

Full prepayments - portfolio excluding credit mortgages

Model specifications	AIC	Schwarz	McFad. R^2	Correct predictions	Remark
c age	1.386	1.386	0.001	49.2%	
c slopelag1	1.248	1.249	0.100	65.6%	
c switchincentive	1.273	1.274	0.082	64.7%	
c dWinter	1.386	1.387	0.000	50.8%	p-value =0.0268
c dJanMar	1.385	1.386	0.001	51.4%	
c dAprJun	1.384	1.385	0.001	51.9%	
c dJulSep	-	-	-	-	insignificant
c dOctDec	-	-	-	-	insignificant
c age slopelag1	1.242	1.243	0.104	65.7%	
c age slopelag1 notional	1.240	1.241	0.106	65.9%	
c age slopelag1 burnout	1.228	1.229	0.115	66.6%	
c age slopelag1 notional burnout	1.226	1.227	0.116	66.9%	
c age slopelag1 notional burnout dJanMar	1.225	1.227	0.117	66.8%	
c age slopelag1 notional burnout dWinter	**1.224**	**1.226**	**0.117**	**66.8%**	**Best model**

Significance level 2.5% applied.

Partial prepayments – Portfolio excluding credit mortgages

Model specifications	AIC	Schwarz	McFad. R^2	Correct predictions	Remark
c prepslag1*	1.152	1.154	0.170	65.7%	
c prepslag1 contract age slopelag1 notional djanmar*	1.109	1.114	0.201	71.1%	Best performance
c age	1.323	1.325	0.046	60.7%	
c slopelag1	1.375	1.377	0.008	55.2%	
c dlownot	1.352	1.354	0.025	53.0%	
c dwinter	1.383	1.384	0.003	52.2%	
c djanmar	-	-	-	-	inconsistent sign
c age slopelag1	1.370	1.371	0.012	55.7%	
c age slopelag1 notional	1.306	1.305	0.062	61.0%	
c age slopelag1 notional dlownot	1.301	1.287	0.075	62.7%	
c contract age slopelag1 notional dlownot djanmar	1.283	1.281	0.081	63.5%	Best valid model

* These models are invalid because prepslag1 only gives information for one month ahead.

Full prepayments - portfolio excluding credit mortgages

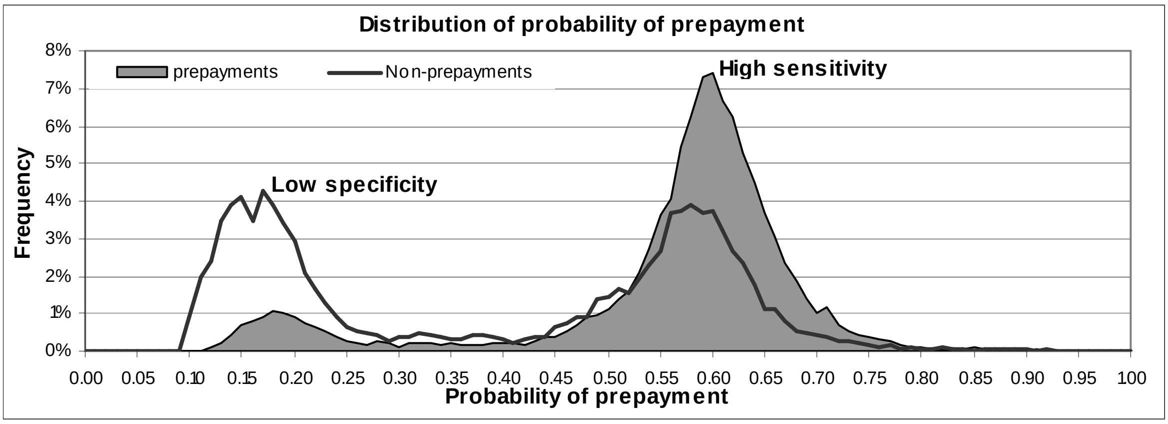


Figure 28: A graphical representation for full prepayments of the evaluation criteria 'sensitivity' and 'specificity'. A perfect model would predict a probability of 1 for all prepayments, and a probability of zero for all contracts that turn out to be non-prepayments.

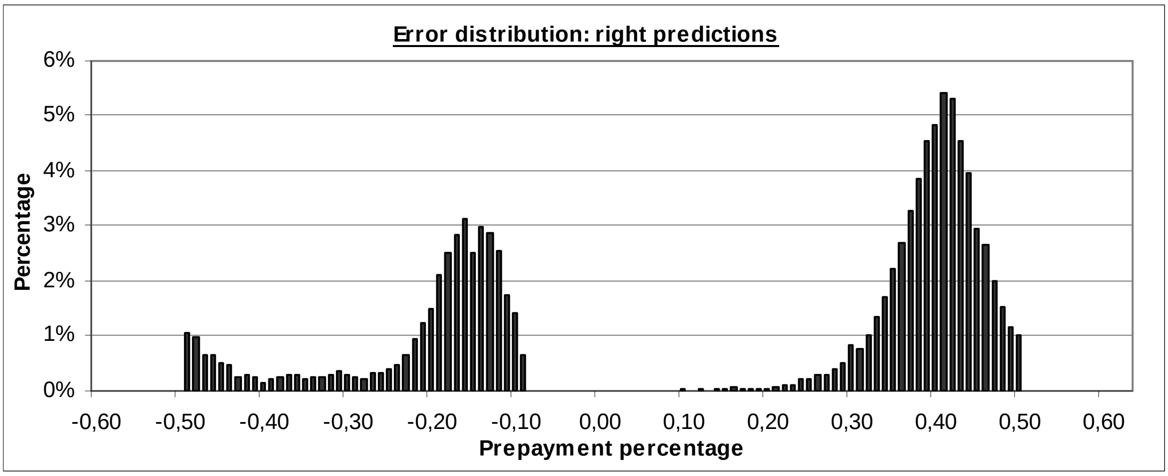


Figure 29: The errors of the correctly predicted contracts. Left: the non-prepaid contracts. Right the fully prepaid contracts.

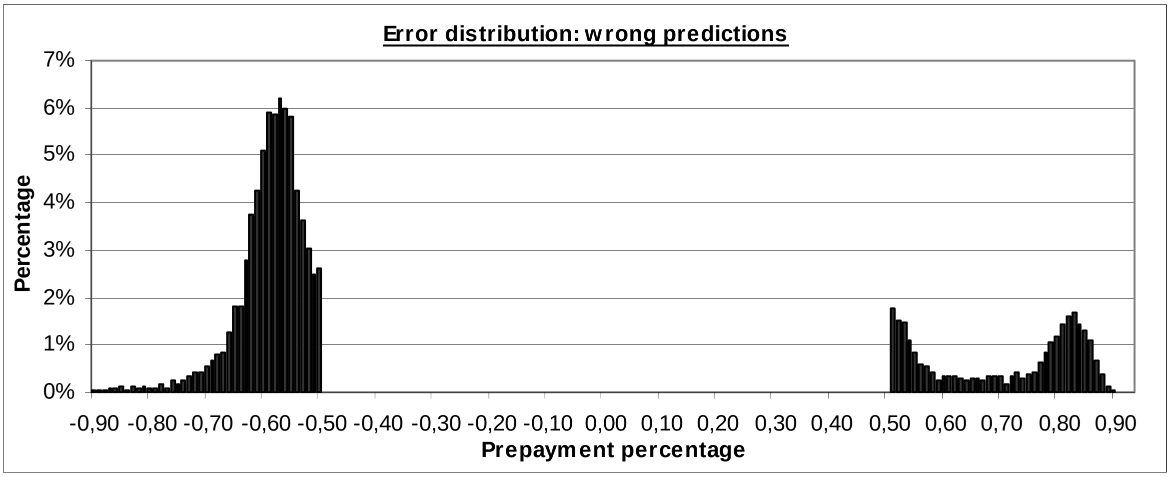


Figure 30: The errors of the wrongly predicted contracts. Left: the non-prepaid contracts. Right: the fully prepaid contracts.

Appendix XV: Prepayment rate sensitivity to the yield curve slope

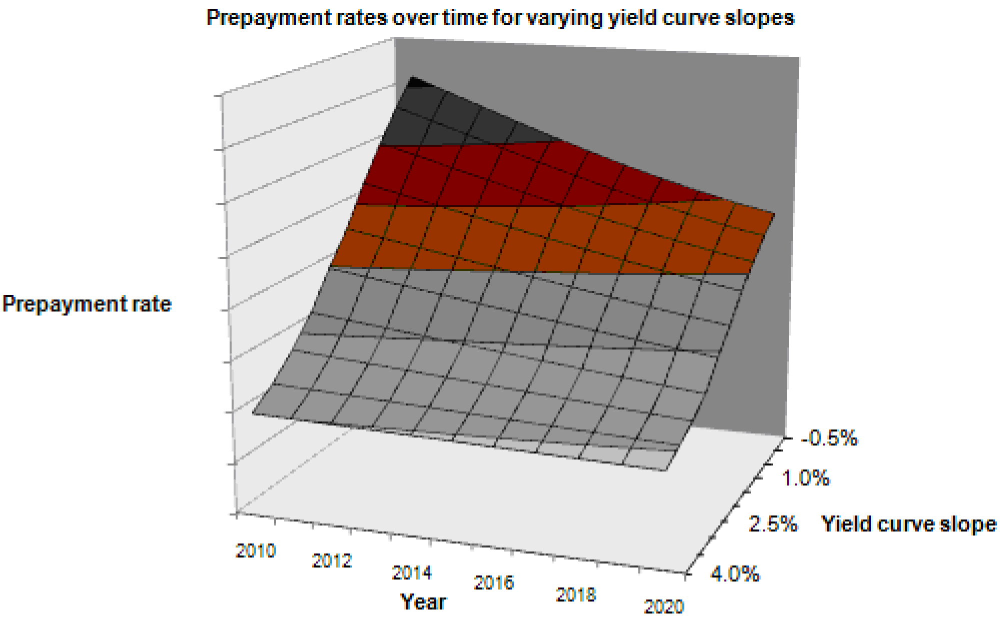


Figure 31: The projected prepayment rates for the coming 10 years for the current portfolio for varying slopes.