Modelling Prepayment Risk: Multinomial Logit Model Approach For Assessing Conditional Prepayment Rate

Non-Confidential Version

Submitted to the University of Twente in conformity with the requirements for the degree:

Master of Science in Financial Engineering and Management

Master Thesis - Pedro Vasconcelos
S0211702

UNIVERSITY OF TWENTE.

Supervisor: Dr. B. Roorda
Supervisor: Dr. H. Kroon

NIBC

Supervisor: Jurgen Peters
Supervisor: Dr. Petra Danisevska
Supervisor: MSc Bastiaan Stücken

Modelling Prepayment Risk: Multinomial Logit Model Approach For Assessing Conditional Prepayment Rate

Abstract

In this Thesis, a Multinomial Logit Model is used to explain prepayment risk on a loan by loan level. A new way of defining the refinance incentive of mortgagors is implemented. An out-of-sample back test is performed and a Mark-to-Market model is applied to calculate the P&L impact caused by prepayment risk on the Dutch mortgage portfolio of NIBC.

Acknowledgements

"Attitude is a little thing that makes a big difference."
Winston Churchill

After two years in the Netherlands, this Thesis is the final step of my Master Degree in Financial Engineering and Management at the University of Twente. Completing a master abroad was a rather challenging task. It required a personal effort on a daily basis to adapt to a new culture, allied to the difficulties underlying a Master Degree. I am surprised how easily I was able to adjust to this new environment and how much I developed in the past two years as a student and even more as a person.

This research was carried out during my internship at NIBC in The Hague. The realization of this master Thesis would not have been possible without the contribution of numerous people. Firstly, I would like to thank my supervisor at the University of Twente Dr. B. Roorda. It was an honor to work with him during the past six months. His insights, assistance, comprehension and general knowledge were of key importance and gave a clear direction to this research. Secondly, I would like to thank my supervisors at NIBC, Jurgen Peters, Petra Danisevska and Bastiaan Stücken. Without a doubt, their experience, support, skills and guidance were a main driver in the completion of this project. I would also like to thank Frits van der Scheer , Thijs Poorthuis and especially Peter Kuijpers for their insights and collaboration during every major step of my research.

Last but not least, I would like to express my gratitude towards my family. Although it was difficult to be apart from them, I felt their full support in every step of my experience in the Netherlands. Without my parents, my accomplishments here would have never been possible. They always encouraged me in my goals and ambitions and were always there for me during my frustrations, achievements and failures. For this and much more, I am proud to be their son. I also want to show my gratefulness to my beloved brother, as he has always set a great example for me. His guidance and advice inspire me daily to try to become a better person.

Pedro Vasconcelos
September 2010

Table of Contents

List of Figures

List of Tables

Chapter 1: Introduction

1.1 NIBC

NIBC is a bank focused on the mid-cap segment in the Benelux and Germany. NIBC offers answers to complex financial challenges to corporate clients, financial institutions, institutional investors and family offices. The bank has offices in The Hague, London, Brussels, Frankfurt, New York and Singapore.

In the heart of NIBC operations is the Risk Management business unit. This strategic business unit is responsible for managing all the risks on a bank-wide basis. In order to be able to cover all types of risks across the bank, the Risk Management business unit has the following structure:

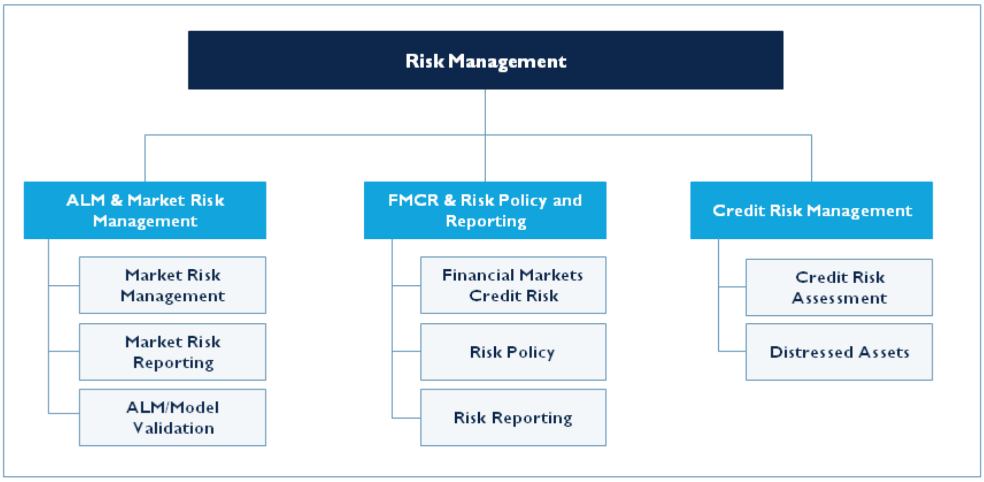

Figure 1 - Structure Risk Management BU

This Thesis is part of a research project of the Asset Liability Management / Model Validation (ALM/MV) department. This department has the following responsibilities within NIBC:

- Risk Control.

ALM/MV performs the daily risk control on the NIBC Retail Markets portfolio, consisting of Dutch and German consumer mortgages. Portfolio size is around € 10 billion (Year End 2009). ALM/MV monitors and forecasts sensitivities to interest rates, market mortgage rates and consumer prepayment.

- Liquidity Risk Management

ALM/MV monitors and reports on the types and amounts of funding available to NIBC, as well as the maturity characteristics of the bank's assets and liabilities. Liquidity forecasts of different lengths are made; these analyses may include stress event and expected funding actions.

- Balance Sheet Reporting and Research

NIBC is exposed to many types of financial and non-financial risk. ALM/MV brings these risks together at a bank level by researching and reporting on bank-wide risk measures (such as Economic Capital) and asset quality. ALM/MV is also responsible for the communication with the Dutch National Bank (DNB) and with external rating agencies and stress testing.

- Model Validation

The business critical models used within NIBC require validation by an independent party. ALM/MV judges the models on their theoretical and numerical correctness, and is also responsible for their documentation, regulatory compliance and comparison to external benchmarks.

This Thesis will focus on mortgage portfolio risk control, and in particular, on the prepayment behaviour of the mortgagors. Prepayment behaviour of the mortgagor is a crucial variable in the valuation of mortgages. When managing a mortgage portfolio, it is important to model the mortgagors' prepayment behaviour. Knowing how prepayment will change, as economic conditions change, is valuable information that enables mortgagees to value mortgages in an accurate way, enabling them to hedge their risks properly.

1.2 The Research Framework

This research is conducted with the goal to improve NIBC prepayment assumptions within their interest rate risk framework for mortgages. For valuation and hedging purposes, NIBC currently assumes a constant prepayment rate. The first objective of this Thesis is to improve the prepayment risk model that already exists within NIBC, based on Sterk (2005), in order to come up with a prepayment rate that depends on a set of explanatory variables determined on a loan by loan basis. Therefore, an introduction to the Dutch mortgage market is presented in the next Section, followed by the literature research on prepayment models in Section 1.4. Then Chapter 2 describes the adopted prepayment model and Chapter 3 looks into the empirical results.

The second objective of this Thesis is to back-test the prepayment model, proposed in Chapter 2. Out-of-sample forecast performance is usually seen as the acid test of an econometric model and it will examine the validity of the model, proposed in this Thesis. Chapter 4 presents this back-test.

Chapter 5 determines the historical marked-to-market (MtM) losses when using the prepayment model, described in Chapter 2, on the existent valuation model of NIBC. Then this result is compared with the current method, in which the conditional prepayment rate (CPR) is assumed to be constant per different categories (buckets) of time until mortgage interest rate reset. Since this model is supposed to predict prepayment rates that are more accurate than the current CPR assumptions, lower MtM results are expected to be observed. Therefore, a model to estimate the mark to market value of NIBC mortgage portfolio will be discussed in Chapter 5. Consequent P&L impacts using NIBC current assumptions will be compared with the P&L impact when using the prepayment model.

Chapter 6 summarizes the conclusions of the previous chapters and describes the recommendations that can be drawn from this research project.

1.3 The Dutch Mortgage Market

Since 1991, the Dutch Mortgage market has developed at a higher pace than the Gross Domestic Product (GDP). This is easily deducted by looking at the substantial growth of the market and the wide range of mortgage types a mortgagor has access to nowadays. Figure 2 shows the GDP of the Netherlands and the respective outstanding mortgage debt in billion euros (CBS Netherlands, 2008). The increasing percentage of the GDP that mortgage debt represents, over the time period considered, reveals the importance of mortgage debt in the Dutch economy.

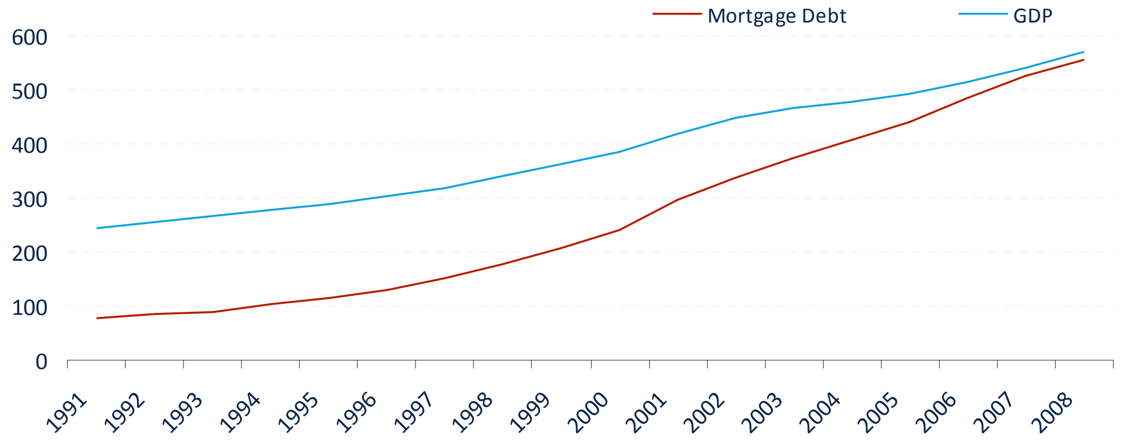

Figure 2 - Mortgage Debt Vs GDP

There are various types of mortgage contracts available to mortgagors Mortgage types can be divided on the basis of their redemption procedure. The most popular mortgage types in the Netherlands are:

- Linear and Annuity Mortgages

These are the simplest contracts a mortgagor can enter into. At every payment date during the mortgage life the mortgagor repays an interest component and a principal component. The difference between linear and annuity mortgages comes from the way the principal component is redeemed. For a linear mortgage, the principal component is fixed (value of the principal divided by the number of payment dates until maturity), while for an annuity mortgage, the total mortgage payments are fixed. Hence, for annuity mortgages the interest payments are initially high whereas the principal payments are low (this relation reverses towards maturity).

It is important to notice that the interest payment for both types of mortgages decreases over time until maturity. Therefore, these types of mortgages do not fully exploit the tax deductibility feature offered by Dutch Law. Linear Mortgages and Annuity Mortgages were very popular during the 80s.

- Life-Insurance and Savings Mortgages

In the early 90s, these two mortgages types appeared to better exploit the tax deductibility feature in the Dutch market, as no principal is repaid during the life to the mortgages. The mortgagor, on a regular basis, puts money in a savings account, held by the mortgagee. On maturity date, the principal is repaid using the money that was put into the savings account. The difference between life-insurance and savings mortgages comes from the savings account. In a savings mortgage, the savings account offers a fixed return on the invested money, while in a life-insurance mortgage this return may vary. Thus, for life-insurance mortgages, there is the risk that the amount saved, in the savings account, may not be enough to repay the principal amount at maturity.

Since no principal is repaid during the mortgage life, the outstanding mortgage principal remains intact. Thus, interest payments do not decrease during the life of the mortgage. This special feature of these mortgages provides takes full advantage of the tax system in the Netherlands.

- Interest-only and Investment Mortgages

The difference between these two and the mortgage types discussed above is that no principal payments or periodic savings are required. The idea behind this is that the principal can be repaid at maturity by entering a new mortgage contract or by selling the house. The Dutch Government imposes that the loan underlying these mortgages can not exceed 75% of the foreclosure value of the property. This is done in order to mitigate the additional credit risk that such mortgages carry, and to prevent the over-credit of Dutch mortgagors.

Usually mortgages have an interest rate fixed period. At the interest reset date, the contractual mortgage rate is set equal to the prevailing mortgage rate (current mortgage rate offer to the same contract at that point in time). In the Netherlands, usually this period is between one and fifteen years.

During the period when the contractual rate is fixed, the value of the mortgage varies, implying risk for both the mortgagor and the mortgagee. For example, if mortgage rates increase during that period, the mortgage is more valuable for the mortgagor, since the mortgage rate that is being paid out is less than the market mortgage rate. This feature shows how important it is for banks the mortgagor's prepayment option, during the life of a mortgage contract. Prepayment risk is, therefore, the risk that a mortgagor prepays between interest rate fixed periods, when the contractual rate differs from the prevailing one.

In order to take into account prepayment risk, banks include some surcharge in their mortgage pricing. Nevertheless, prepayment risk is still exists because a mortgagor can prepay at any time, while the spread to account for this risk is received on a monthly basis across the life of a mortgage.

Due to this, banks usually charge a prepayment penalty. This penalty is meant to discourage mortgagors to refinance their mortgages, when mortgage rates decline, to compensate the bank for losses that the bank will incur, if the mortgagor prepays. The penalty mount is the difference between future contractual interest payments of the mortgage contract and those of a newly originated mortgage with the same characteristics.

The Dutch Government imposes a set of prepayment possibilities for the mortgagors where the mortgagee can not charge any penalty to the mortgagor. Thus, in the Netherlands at least 10% of the initial principal can be prepaid within any full calendar year. Prepayment is also free in the following situations:

- House sale;
- Demise of the Mortgagor;
- Bankruptcy;
- At any interest reset date;
- Reception of a fire-insurance benefit.

Another important characteristic in the Dutch mortgage market is that some mortgages have *a Nationale Hypotheek Garantie (NHG)*. This guarantee was created by the Dutch Government to permit low-income earners to be able to own their own houses. In case of mortgage default and if the house foreclosure value is below the outstanding principal, the government pays the difference. The effect of mortgages with a NHG is straightforward for mortgagors. This guarantee lowers the default risk for the mortgagee. As a consequence, the mortgagor is able to get a mortgage with lower rate, making it easier for low income mortgagors to own a house.

1.4 Literature on Prepayment Modelling for Residential Mortgages

A wide range of prepayment models can be found in academic prepayment literature. These prepayment models can be split into two groups shown in the following Figure:

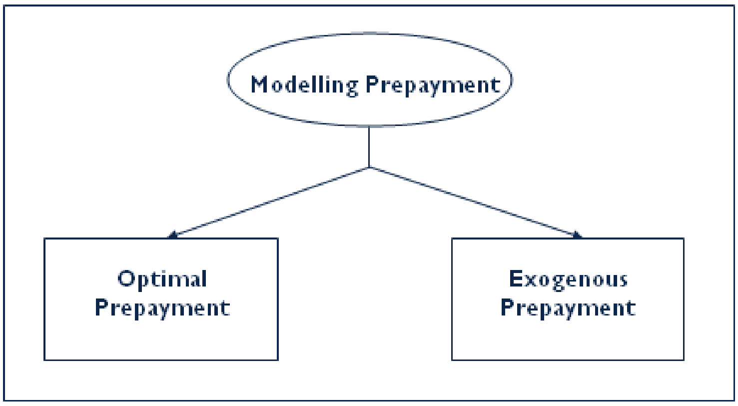

Figure 3 - Modelling Prepayment

On the one hand, there are models that work under the assumption that prepayment is exercised in an optimal way. The mortgagor would prepay when the value of the mortgage is greater than the outstanding debt plus transaction costs. Van Bussel (1998) assumes optimal prepayments and values mortgages as callable bonds.

On the other hand, there are models that assume an exogenous prepayment rule. This makes more sense than the above models, because actual prepayments often appear to be non-optimal or irrational from a risk-value perspective. In other words, mortgagors may prepay when the prevailing mortgage rate is above their contractual rate.

In the literature, exogenous models can be divided into two groups (see Figure 4). The first group contains models that are built upon endogenous models and the second contains strictly empirical models. The former simply add exogenous calls (that are not related to the interest rate) to the optimal prepayment model explained above, in order to take into account the non-optimal prepayment behaviour of mortgagors. The latter looks into the observed prepayments to determine which explanatory variables can explain such behaviour.

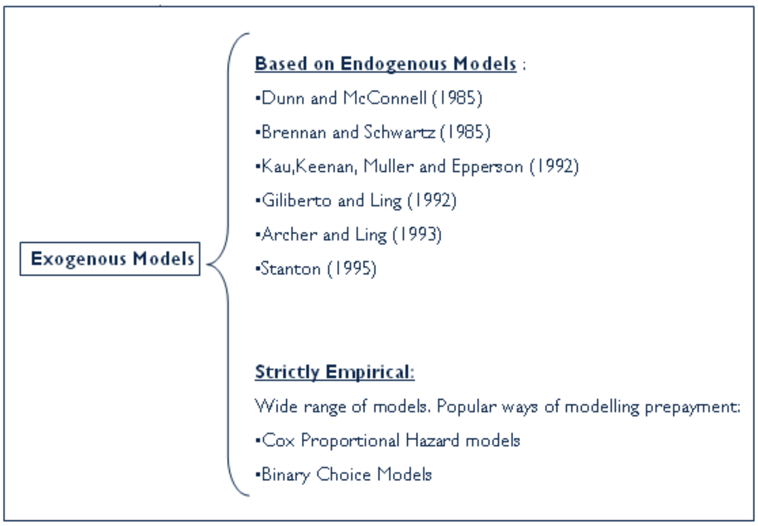

Figure 4 - Exogenous Models

Dunn and McConnell (1985) use a Poisson process to explain the non-optimal prepayments. Brennan and Schwartz (1985) and Kau, Keenan, Muller and Epperson (1992) build upon the Dunn and McConnell prepayment valuation model. It is important to notice that the above authors did not relate prepayment behaviour to transaction costs. This was already included in the models proposed by Gilberto and Ling (1992), Archer and Ling (1992) and Staton (1995).

As explained above, strictly empirical models use observed prepayments to predict prepayments in the future. The literature on these models is extensive and more recent. By looking at the literature on Strictly Empirical models two popular ways for modelling prepayment can be found as it can be seen in Figure 4.

The Cox Proportional Hazard models (CPH) were introduced by Cox (1972). They are mainly used to model the duration of dependent variables. Therefore, in prepayment models the CPH explain the duration of a mortgage until it is prepaid or redeemed.

The Binary Choice Models use a dependent variable that takes a value one if a determined event occurs and zero otherwise. Thus, for prepayment models it would take the value one if the mortgage is prepaid. The most known Binary Choice models are the Logit Model and the Probit Model.

Both CPH and Binary Choice Models have many variations. There are various reasons for a mortgage to end like refinance, house sale, default and others. Therefore, the above mentioned models can be extended to incorporate competing risks.

It is important to notice that the majority of the strictly empirical models are based on observations from aggregated pool-level data. Examples of these are the articles such as, Brennan and Swartz (1985), Clapp et al. (2000), Collin-Dufresne and Harding (1999), Hayre (1994), Hayre and Rajan (1195), Hayre et al. (2000), Huang et al. (1999), Jegadeesh and Ju (2000), Patruno (1994), Singh and McConnell (1996).

The evident problem that arises from using aggregate pool-level data is that an important part of information is lost. The characteristics of individual loans are hidden behind the averages. And variables like refinance incentive are likely to be underestimated or overestimated.

In the past, models on individual loan data were already proposed by Abrahams (1997), Green and Shoven (1986), Van Bussel (1998) and by Sterk (2005). The model used by Sterk (2005) is used in the remaining of this Thesis and it is explained in the next Chapter.

Chapter 2: Prepayment Risk Model

2.1 Introduction

The objective of this Chapter is to describe the current prepayment risk model used within NIBC and to propose improvements on it. Therefore, Section 2.2 describes the current model while Section 2.3 explains the adjustments that will be incorporated. Section 2.4 describes all the explanatory variables that were chosen to model prepayment behaviour of mortgagors while Section 2.5 describes how predictions about prepayment rates can be done. Finally, Section 2.6 draws some conclusions and discusses some expectations about the development of this research.

2.2 Description of the Current Model Prepayment Risk Model

Sterk (2005) proposed a Multinomial Logit Model to model prepayment in the Netherlands during his internship at NIBC. He argues that a competitive model "is likely to produce better predictions than a standard model", because the reasons that leads a mortgagor to prepay due to refinancing or due to moving from a house to another are very different.

Multinomial Logit Models are just an extension of the popular Logit Model used when the dependent variable has more than two possible outcomes. In this case the dependent variable (prepaid) can have three distinguished outcomes. At each point in time, t, a mortgage i can continue or be prepaid due to moving (house sale) or be prepaid due to refinancing. The dependent variable prepaid should be seen as a polytomous response variable that does not have an ordered structure.

Let,

$$y_{it} = \begin{cases} 1, & \textit{if Mortgage Continues} \\ 2, & \textit{if Mortgage Move} \\ 3, & \textit{if Mortgage Refinance} \end{cases} \tag{1}$$

The Multinomial Logit Model focuses on the mortgage as a unit and uses mortgage characteristics as explanatory variables. It should be stated that since the explanatory variables are the characteristics of a mortgage, it follows that they are constant for the three possible outcomes of the dependent variable.

The probability distribution of y_{it} is defined in Equation 2:

$$\tau_{ij} \equiv \Pr(y_{it} = j \mid X_{it}) = \frac{\exp(X_{it}\beta_j)}{\sum_{l=1}^{3}\exp(X_{it}\beta_l)} \qquad \textit{for} \quad j = 1,2,3 \tag{2}$$

Where, X_{it} is the vector of the explanatory variables and $\beta_{l,j}$ is the vector of the regression coefficients.

The multinomial Logit Model consists of several binary Logits, which are estimated simultaneously. In this case two binary Logits are considered by setting continue as the base category. One binary Logit is considered for refinance versus continue and other for move versus continue.

Since, the probabilities in Equation number 2 have to sum to one and a base category as been selected. The above expression is rewritten:

$$\tau_{i1} \equiv \Pr(y_{it} = 1 \mid X_{it}) = \frac{1}{1 + \sum_{l=2}^{3}\exp(X_{it}\beta_l)} \ , \tag{3}$$

$$\tau_{ij} \equiv \Pr(y_{it} = j \mid X_{it}) = \frac{\exp(X_{it}\beta_j)}{\sum_{l=2}^{3}\exp(X_{it}\beta_l)} \qquad \textit{for} \quad j = 2,3 \tag{4}$$

In a binary Logit the regression coefficients individually represent a change in the Logit of the probability associated with a unit change in the predictor (associated with that coefficient) holding all other predictors constant. It is important to remember that the regression coefficients can not be interpreted in the same way as linear models since on the left side is a Logit rather than a mean. A Logit is defined as the natural log of the odds ratio.

In this case, since we are using a Multinomial Logit Model, the odds ratio is defined as the ratio of the probabilities between two alternatives:

$$\Omega_{j\,|\,l} = \frac{\Pr(y_{it} = j \mid X_{it})}{\Pr(y_{it} = l \mid X_{it})} = \frac{\exp(X_{it}\beta_j)}{\exp(X_{it}\beta_l)} \quad for \quad j,l = 1,2,3 \tag{5}$$

The Multinomial Logit model is used to model choices. In this case, it models the choices continue, refinance or move. The model relies on the assumption of Independence of Irrelevant Alternatives (IIA). When considering the two binary Logit models that are estimated simultaneously, the first considers refinance versus continue and it is irrelevant that the mortgage may have or not the possibility of moving. The same question is assumed in the other binary logit where it is irrelevant if a mortgage may or not have the possibility of refinancing. The problem is that in reality this may not always hold, because if a mortgagor can not refinance the mortgage the probability of him exercising the option of moving is lower than if he could refinance.

The regression coefficients are calculated using the maximum likelihood method. The likelihood function of this model is defined as:

$$L(\beta) = \prod_{i=1}^{n} \prod_{t=1}^{T_i} \tau_{ij}^{\,I[y_{it}=j]} \tag{6}$$

Where $I[.]$ is an indicator function, taking the value one if $y_{it} = j$ and zero otherwise. For further information, see Sterk (2005).

2.3 Adjustments to the Current Model

- Data

- Refinance incentive

In the current model the refinance incentive is defined as the absolute difference between the prevailing swap rate and the swap rate at the date the mortgage rate was fixed. Sterk stated that "Absolute changes in mortgage rates are highly correlated to absolute changes in long rates". However, in line with recent events one could argue whether this correlation exists, as it can be seen in Figure 5, the swap rates declined more than mortgage rates, while the retail spread increased significantly.

Figure 5 - Mortgage Rate and components since January 2008

NIBC can derive internally mortgage rates for itself, intermediaries or Dutch Banks taking into account any combination of the following factors:

- Risk Class;
- Tenor;
- Product.

Mortgage rates are usually built by adding spreads to a long interest rate (with relevant maturity). This retail spread varies for each mortgage taking into account the above mentioned factors. Therefore, the mortgagor can have different refinance incentives since if the mortgagor refinances the new mortgage can have a completely different set of characteristics. Since absolute changes in mortgage rates are not highly correlated to absolute changes in long rates anymore, a benchmark for the retail spread was computed (Blended Market Retail Spread), using an internal model of NIBC. By doing this, a Blended Market Mortgage Rate is estimated by adding a long rate to this spread.

In this Thesis the refinance incentive is defined as the difference between the Blended Market Mortgage rate at the mortgage origination and the prevailing Blended Market Mortgage rate. An internal model is used to blend the data to the NIBC mortgage portfolio. This is done in order to be in a position to have a benchmark throughout time that can be compared.

One could argue that the difference between the contractual rate and the benchmark can be used instead. The reason not to do so is that it does not compare the same thing. Each individual mortgage is a combination of different characteristics, and at prepayment date or reset date, the mortgagor can change the mortgage to a completely different combination. Hence, numerous contractual rates are available and as a consequence, the set of refinance incentives becomes nearly impossible to calculate. This justifies the use of the benchmark.

In the remainder of this Thesis, the refinance incentive is defined as:

$$RefinanceIncentive(t) = Blended\ Market\ Mortgage\ Rate_0 - Blended\ Market\ Mortgage\ Rate_t \quad (7)$$

Where,

$$Blended\ Market\ Mortgage\ Rate_t = 5yr\ Swap_t + Blended\ Market\ Retail\ Spread_t \quad (8)$$

Note that there are several ways to define the refinance incentive. The most obvious is simply to calculate the difference between the contractual mortgage rate and the prevailing rate on the mortgage market. However, applying the refinance incentive as defined in equations 7 and 8 should lead to more accurate results since a benchmark for the retail spread is created.

For different definitions of the refinance incentive, see Alink (2002) and Charlier and Van Bussel (2001).

2.4 Explanatory Variables

The following explanatory variables for prepayment were included in the Multinomial Logit Model, taking into consideration the above explained adjustments and in line with Sterk (2005) reasoning:

- <u>Redemption type</u>

There are many types of mortgage contracts that mortgagors can enter. The most popular types, in the Netherlands, are the following:

- Linear Mortgages;
- Level Payment Mortgages;
- Savings Mortgages;
- Life Insurance Mortgages;
- Investment Mortgages;
- Switch Mortgages;
- Interest-only Mortgages.

One could argue that the prepayment behaviour varies across different mortgage types. In the past, Charlier and Van Bussel (2002) and Alink (2001) estimated a single model for each mortgage type. A dummy variable for each mortgage type will be used, in order to account for the effect of each redemption type on the prepayment rates for moving and refinancing.

- <u>Seasonality</u>

Prepayment rates have the particular characteristic to spike in certain months. This phenomenon is known as seasonality. In his studies Van Bussel (1998) notices that prepayment apparently has a seasonal pattern. He discovered that prepayment rates are relatively low in the early spring and after that they increasing, reaching a peak in July. His study also found that another peak occurs in December related to the "*annual limited prepayment option embodies in most Dutch contracts due to tax effects*". This Seasonal pattern was also confirmed empirically by Richard and Roll (1989) and by Kang and Zenios (1992).

Therefore, the Multinomial Logit model will include twelve dummy variables, one for each month, to take into account this phenomenon.

- Seasoning

This variable relates the age of the mortgage to the prepayment. It is believed that shortly after a borrower has entered a mortgage contract, it is unlikely for him/her to prepay it, since this will incur extra costs to the borrower, namely moving costs if the borrower moves from a house to another or a penalty, if he/she chooses to refinance. Thus, at origination the probability of prepayment is zero, after this it gradually increases until, according to the findings of Alink (2002), a seasoned level is reached more or less about three years after the origination date.

A variable will be included in the prepayment model to account for this behaviour. This variable is defined ,as described in Alink (2002):

$$seasoning = \begin{cases} \ln(\frac{age}{36}) & if\ age < 36\,months \\ 0 & if\ age \geq 36\,months \end{cases} \tag{9}$$

- Refinance Incentive

It is one of the crucial variables in the prepayment model. When interest rates decrease, the borrower has an incentive to prepay, if he/she behaves in a rational way. As explained in the previous Section, the refinance incentive in this Thesis is defined as in Equations 7 and 8.

- Age of the Mortgagor

Some people argue that elder mortgagors have different prepayment behaviours than younger ones. Young people move more often than elder mortgagors. Moreover, elder mortgagors are believed to be more conservative. Therefore, as any conservative person, they are less open to change, which may implicate a more passive behaviour towards prepayment of their mortgages. Moreover, Alink (2002) proved that the aging of mortgagors is negatively related to the probability of prepayment.

In line with Vicent Sterk prepayment model, the following dummy variables for different age categories were included:

- agecat1, if the age of the mortgagor is between 18 and 30 years;
- agecat2, if the age of the mortgagor is between 30 and 40 years;
- agecat3, if the age of the mortgagor is between 40 and 50 years;
- agecat4, if the age of the mortgagor is between 50 and 60 years;
- agecat5, if the age of the mortgagor is older than 60 years.

- Interest rate fixed period

It makes sense to think that a longer interest rate fixed period will imply a higher prepayment rate. The reasoning behind that is the same as in the valuation of American options. If the time until maturity is higher, then the value of the option is also higher, since the holder of the option has more time, and therefore more opportunities to exercise the option. Hence, in theory a mortgagor that has a thirty-year fixed rate period is more likely to prepay than a mortgagor that has a five-year fixed rate period.

In order to take this into account a dummy variable is included for each interest fixed rate period that have a significant percentage of NIBC portfolio.

- Property Type

Usually, people first live in a flat and later on, when they are wealthier, they move to a house. This should reflect a higher prepayment rate on mortgages, when the underlying is a flat than when it is a house. Since mortgages on flats are expected to be prepaid more frequently, a dummy variable on the property type will be included in the Multinomial Logit model, taking the value one if the underlying is an apartment and value zero if it is a house.

- National Mortgage Guarantee (NHG)

Some academics like De Jong (1998) argue that guaranteed mortgages are prepaid less often. The reason for this is that the prepayment on these mortgages may lead to a loss of the guarantee in a future contract, since the new mortgage may exceed the maximum loan size allowed by the Dutch Government. While others, like Alink (2002), did not find any significant difference in the prepayment behaviour of mortgagors that have a guarantee and those who do not have it.

The significance of the above will be studied. A dummy variable is included in the Multinomial Logit model. The variable is called NHG and has value one if the mortgage is guaranteed and zero otherwise.

- Time until Reset

Mortgages may differ in the way they prepay relative to the time left until next reset. For instance, if a mortgage is 2 months away from its reset date and it has a huge refinance incentive, the mortgagor may wait until the reset date to refinance it, in order to not pay a penalty fee. Thus, to see how prepayment rates are explained based on the time until next reset, the following dummy variables are included in the prepayment model:

- Remaincat 1=1 if the time until reset is less or equal to 1 month;
- Remaincat 2=1 if it is between 2 and 4 months before reset;
- Remaincat 3=1 if it is between 4 and 12 months before reset;
- Remaincat 4=1 if it is between 1 and 5 years before reset;
- Remaincat 5=1 if it is between 5 and 10 years before reset.

- Region

In order to see how prepayment varies across geographic regions/areas in the Netherlands five dummy variables are added to the model. STATER provides the postal codes, linked to the mortgage and based on the code, it is possible to split the data in the following categories:

- Big cities, postal codes 1000-1099 (Amsterdam), 2500-2599 (The Hague), 3000-3099 (Rotterdam), and 350-3599 (Utrecht);
- South, postal codes 4000-6499 (Limburg, Noord-Brabant and Zeeland);
- East, postal codes 6500-8299 (Gelderland, Overijssel, Drenthe, Northern Flevoland);
- North, postal codes 8300-9999 (Groningen and Friesland);
- West, postal codes 1000-3999 (Noord-Holland, Zuid-Holland, Utrecht, Southern Flevoland), except for the above mention cities.

2.5 Prediction of Single Month Mortalities for Moving and Refinancing

In the academic mortgage literature prepayment is defined as the redemption of a mortgage loan before its maturity. Following Fabozzi (2001), the *Conditional Prepayment Rate* (CPR) is the annualized *Single Month Mortality* (SMM) of a mortgage portfolio. These definitions simply are:

$$SMM_t = \frac{Debt\ prepaid\ at\ Month\ t}{Total\ Outstanding\ debt\ at\ month\ t} \tag{10}$$

$$CPR = 1-(1-SMM)^{12} \tag{11}$$

The objective behind the application of the Multinomial Logit Model is to use the resulting estimates of the explanatory variables to predict prepayment rates. Following from Equations 10 and 11 and the choice to predict prepayments related to moving and refinancing, the monthly prepayment rates, also known as SMM, are defined bellow:

$$SMM_{refinance}(t) = \frac{\sum_{i=1}^{N} B_{it} I(y_{it} = 2)}{\sum_{i=1}^{N} B_{it}} \tag{12}$$

$$SMM_{Move}(t) = \frac{\sum_{i=1}^{N} B_{it} I(y_{it} = 3)}{\sum_{i=1}^{N} B_{it}} \tag{13}$$

Where, B_{it} is the outstanding notional balance at time t for mortgage i. In order to get the debt prepaid at month t, $I(\cdot)$ is an indicator function. From Equation 1, y_{it} is a random variable with three possible outcomes. Hence, it follows a Multivariate Bernoulli Distribution, implying that $I(y_{it=j})$ follows a Univariate Bernoulli distribution with mean τ_{itj}. Then, the respective expected prepayment rate and variances expressions are:

$$E[SMM_{refinance}(t)] = \frac{\sum_{i=1}^{N} B_{it}\tau_{it2}}{\sum_{i=1}^{N} B_{it}} \qquad Var[SMM_{refinance}(t)] = \frac{\sum_{i=1}^{N} B_{it}^{\ 2}\ \tau_{it2}\,(1-\tau_{it2})}{(\sum_{i=1}^{N} B_{it}\ )^2} \tag{14), (15}$$

$$E[SMM_{Move}(t)] = \frac{\sum_{i=1}^{N} B_{it}\tau_{it3}}{\sum_{i=1}^{N} B_{it}} \qquad Var[SMM_{Move}(t)] = \frac{\sum_{i=1}^{N} B_{it}^{\ 2}\ \tau_{it3}\,(1-\tau_{it3})}{(\sum_{i=1}^{N} B_{it}\ )^2} \tag{16), (17}$$

From the above Equations, one can predict prepayment rates by replacing the probabilities τ_{itj} by their predictions. These predictions are just the result of using the $\hat{\beta}$ estimates provided by the Multinomial Logit Model.

Therefore, the following three Equations can predict prepayment rates:

$$S\hat{M}M_{refinance}(t) = \frac{\sum_{i=1}^{N} B_{it}\hat{\tau}_{it2}}{\sum_{i=1}^{N} B_{it}} \qquad S\hat{M}M_{Move}(t) = \frac{\sum_{i=1}^{N} B_{it}\hat{\tau}_{it3}}{\sum_{i=1}^{N} B_{it}} \tag{18), (19}$$

$$Where, \quad \hat{\tau}_{ij} \equiv \hat{\Pr}(y_{it} = j \mid X_{it}) = \frac{\exp(X_{it}\hat{\beta}_j)}{1+\sum_{l=2}^{3}\exp(X_{it}\hat{\beta}_l)}, \qquad for \;\; j = 2,3 \tag{20}$$

2.6 Conclusions

In this Chapter the model for prepayment that used in this Thesis was explained. It was chosen to use the model of Sterk (2005) since we are interested in a model on individual loan data and we believe that the reasons that lead mortgagors to prepay due to moving differ from the reasons that lead to refinance. Therefore, the Multinomial Logit model is the appropriate model since it allows taking into account the two, previously mentioned, competitive risks.

Two main adjustments were made to Sterk's model. Firstly, the data-set used by him at NIBC was extended and mortgages with an interest fixed period less than five years were included since there is no reason to believe that prepayment risk is negligible for these types of mortgages. Secondly the way the refinance incentive was defined by Sterk was changed since in Section 2.3 it was proven that absolutes change in swap rates were not correlated with absolute changes in mortgage rates in the recent past. Therefore a new way of defining refinance incentive was proposed in Section 2.3. Thus in the remainder of this Thesis refinance incentive is defined as the difference between the Blended Market mortgage rate at the origination of the mortgage and the prevailing Blended Market mortgage rate. An internal model was used to blend the data to the NIBC mortgage portfolio. This is done in order to be in a position to have a benchmark throughout time that can be compared.

Because the input data-set of the Multinomial Logit model was extended and the definition of the refinance incentive was improved, we expect that the predictive power of the model will be improved.

Section 2.4 describes all the explanatory variables that will be incorporated in the model. The selection of redemption type, seasonality, seasoning, refinance incentive, age of mortgagor, interest fixed period, property type, national mortgage guarantee, time until reset and region as explanatory variables explains the prepayment behaviour of the mortgagors.

Finally, Section 2.5 presents the way the parameter estimates produced by the Multinomial Logit model can be used for predicting prepayment rates. SMM and CPR were defined and they are concepts that will be often used during this Thesis.

Chapter 3: Empirical Results

3.1 Introduction

In this Chapter the previously discussed model will be applied. Section 3.2 describes the data-set that will be used as input and Section 3.3 gives provides insight into the characteristics, underlying the data-set. Section 3.4 discusses the realized prepayment rates due to moving and refinance. The estimation results of the Multinomial Logit model are shown in Section 3.5 and Section 3.6 compares the full model with three restricted models that leave out some of the explanatory variables. Finally, Section 3.7 draws the conclusions of this Chapter.

3.2 Description of the Data

3.3 Characteristics of the Data-set

The used data-set contains of 87,351 mortgages and 202,160 mortgage parts originated from January 1961 until October 2009, as it was said in Section 2.2.

Figure 6 - Yearly Production of Mortgages Parts

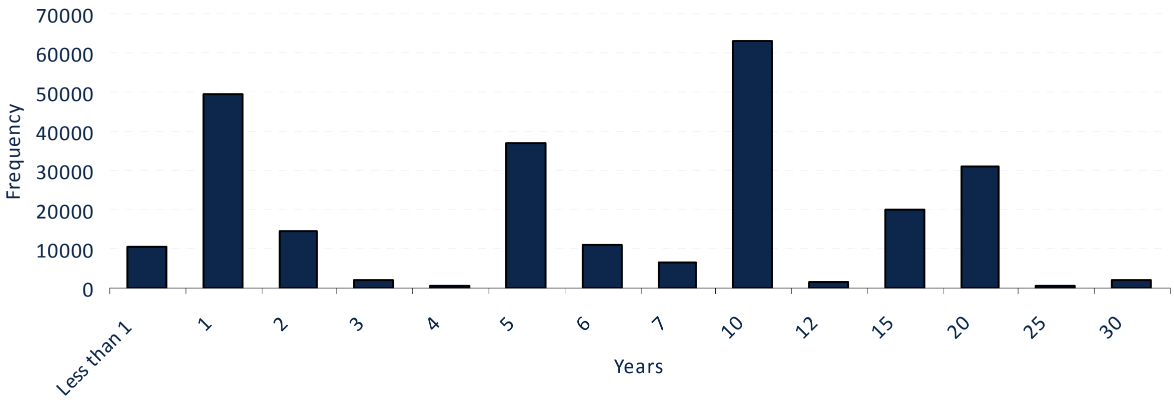

Figure 7 - Fixed Rate Periods

In Figure 7, the fixed interest rate period of the mortgages is shown. Mortgages with a fixed interest rate period of 10 years are the most common among the data, representing 25% of the portfolio. Moreover one, five, and twenty years fixed rate period are also popular.

In line with Section 2.2, mortgages with an interest fixed period less than 1 year will be deleted from the data-set, since they have a limited prepayment risk and most of them are variable interest rate mortgages that almost do not have any prepayment risk. Also, mortgages with an interest fixed period greater than 20 years will be erased from the Data-set, because they do not have a significant presence in the portfolio of NIBC.

Comparing to Sterk (2005) data-set, significant differences are observed. Mortgages with one year interest rate period are not dropped out and mortgages with five and twenty years, that had a more substantial presence in the portfolio, 20,4% and 20,8% respectively, now have substantially lower percentages of the portfolio.

After the above mentioned restrictions to the data-set, the number of mortgages decreased to 73,349 and the number of mortgages parts decreased to 136,769. This gives, on average, approximately 1.9 mortgage parts per mortgage.

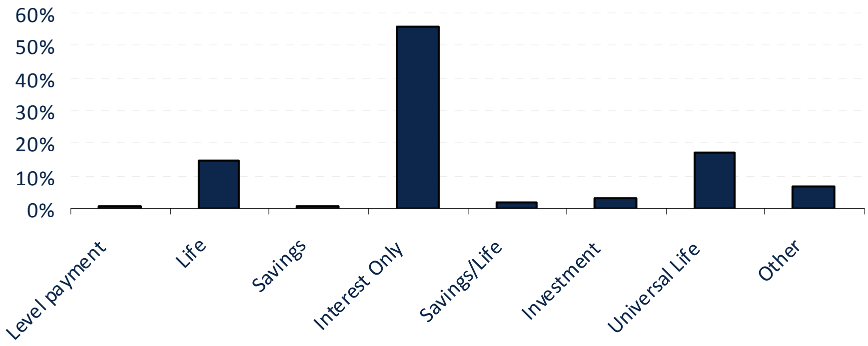

Figure 8 - Size Mortgage Types in the Portfolio

As it can be seen in Figure 8, the most popular type of mortgage is interest only, which represents 71% of the portfolio. The explanation behind that is the fiscal attractiveness of mortgage debt in the Netherlands, where interest payments are tax deductible. Also popular, for fiscal reasons, are Life and Universal Life mortgages, which represent more than 10% of the portfolio.

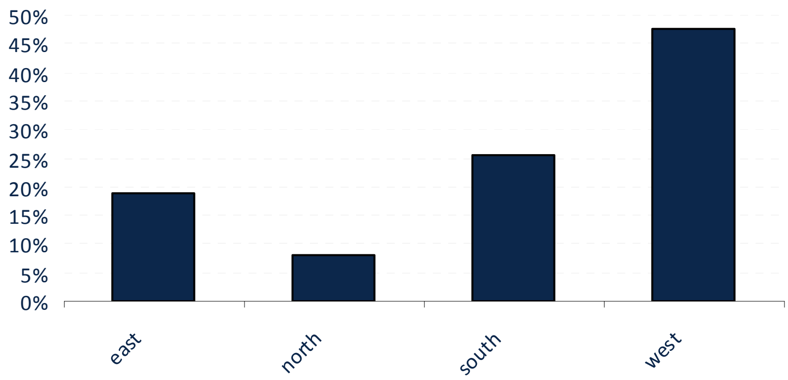

Figure 9 - Portfolio Geographic Distribution

It is also interesting to notice that almost half of the mortgages are originated in the west part of the Netherlands (Figure 9). These results are in line with the data-set used by Sterk (2005).

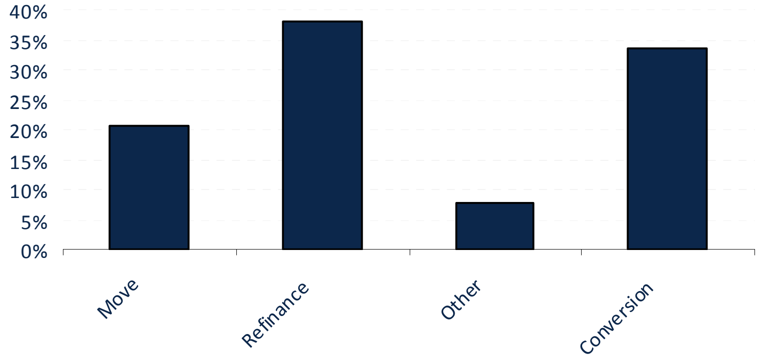

Figure 10 - Causes of Prepayment

Only prepayment due to house sale or refinancing are going to be considered, since they are the main reasons for prepayment, as shown in Figure 10. The high percentage of conversion is not surprising because too often, in the STATER data, the reason of repayment is set equal to conversion when nothing actually changed. The percentage for refinancing is slightly lower when comparing to Sterk (2005). However, for moving the decrease was around 8 percent. It may seem that people prepaid less after 2005. However, the real reason for this is that mortgages with a one year interest fixed rate were included in the model and these mortgages are expected to have more conversions, in line with what was explained in Section 3.1.

3.4 Observed Prepayments

The observed prepayment rates for moving and refinancing from September 1998 until October 2009 can be seen in Figure 11 together with the total Prepayment Rate (PR):

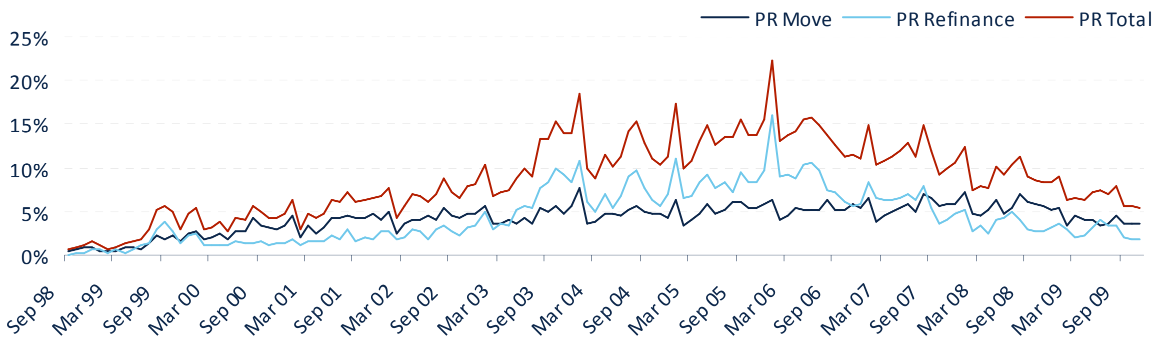

Figure 11 - Observed Prepayment Rates since September 1998

It is surprising how volatile the total prepayment rate is. Nevertheless, the PR for moving usually takes values between 4 and 6 percent. The PR for refinance is more volatile and it has sharply increased since the end of 2002 until January 2006. Looking at the period after 2006, prepayment rates decreased. Although interest rates decreased sharply after 2008, few mortgages were refinanced due to the increase in retail spread and risk aversion in the market.

The PR's seem to have strong seasonality effects. Thus, to study this, more closely, in the Figure 12 the PR's for each month are plotted.

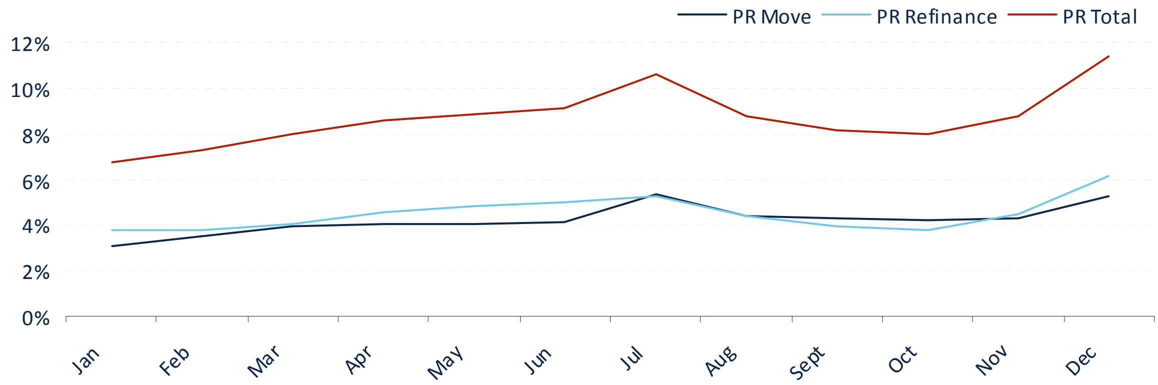

Figure 12 - Seasonality

In agreement with the findings of Van Bussel (1998) and Alink (2002), prepayment rates peak in July and December. It is easy to understand why the PR for moving peaks in these months, since moving from one house to another takes time and from a borrower point of view, it makes more sense to move during holidays. For the refinance incentive, the peak in December can be explained by fiscal reasons. As expected, these results are in line with the observations made on Sterk (2005) data-set.

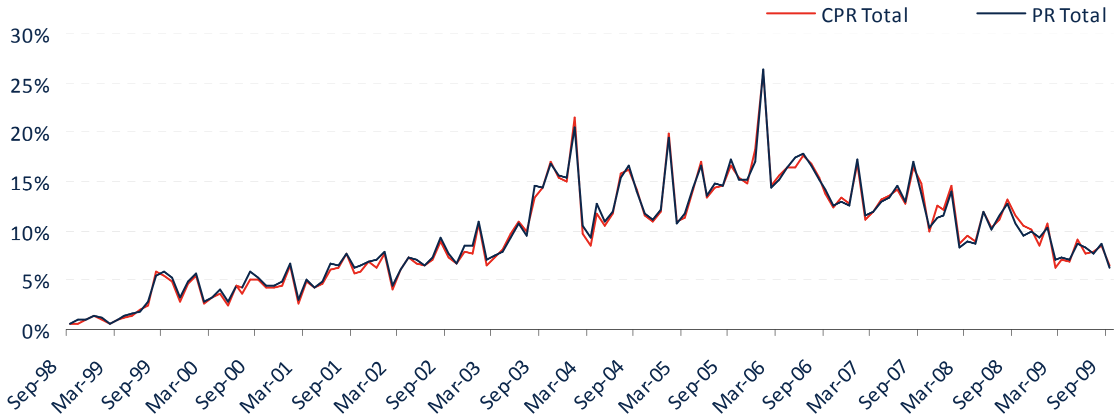

Figure 13 - Observed PR and CPR since September 1998

Until now, we are looking at the prepayment rates as the number of prepayments divided by the total number of mortgages at each period. Figure 13 compares the Total Prepayment Rate with the total CPR to see if the size of the loan has a significant impact. One could argue that a mortgagor, prepaying with a notional of one hundred thousand euros, has a completely different impact on the bank than a mortgagor that prepays a mortgage with a notional of ten million euros. Although, the size of a notional does not seem to be feasible as an explanatory variable for prepayment, no explanations are stated in the literature to say why prepayment behaviour varies across different principals. The differences are not substantial and it seems that they can be neglected. This simply tells us that the heterogeneity between mortgagors, regarding the value of the mortgage principal, is not sufficient to have a considerable impact, at least on the portfolio data provided by STATER.

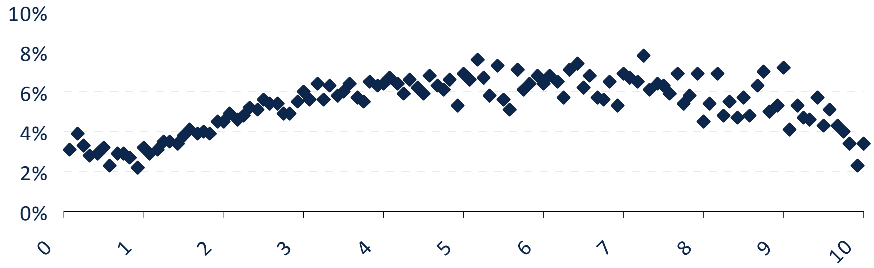

Figure 14 - Seasoning for prepayment due to moving

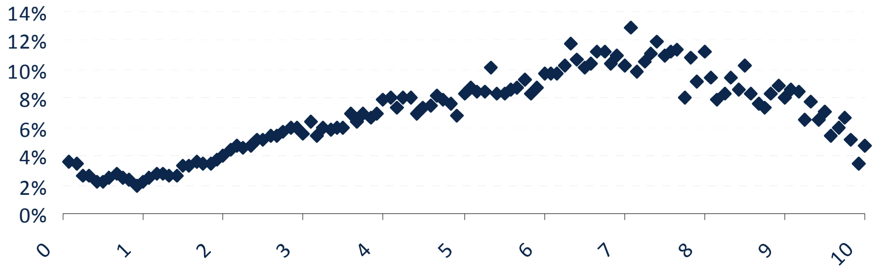

Figure 15 - Seasoning for prepayment due to refinance

Figures 14, and 15, show the observed seasoning in the data-set for PR move and PR refinancing. By looking at the scatter plots, it can be concluded that for moving the portfolio becomes seasoned after three years, while for refinancing it takes around five and a half years for a mortgage to reach its seasoned level. As previously explained, this phenomenon is linked to the fact that both situations, refinancing and moving, imply transaction costs. The former has a penalty fee and the latter has moving costs. Considerable differences are seen when comparing to Sterk's results, who observed a period of 3.5 years for reaching a seasoned level due to refinancing, while our data-set seasoned after 5.5 years. This difference is partially explained by the inclusion of the mortgages with one year interest fixed rate period.

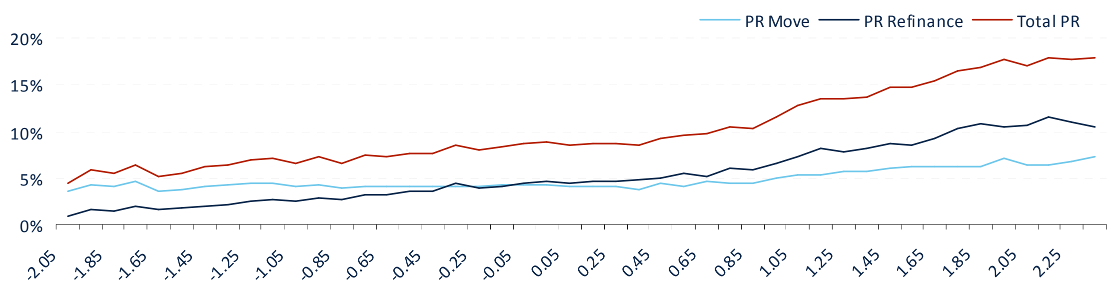

Figure 16 - Prepayment as a Function of the Refinance Incentive

Figure 16 shows the PR move and PR refinance as a function of the refinance incentive. As expected, a substantial increase in the PR refinancing is observed, when the refinance incentive increases. It can be easily seen how sensitive the rate of refinancing is to changes in the refinance incentive (when the refinance incentive increases from 0.00% to 1.45%, the PR refinance doubles). The rate of moving is not that sensitive to a refinance incentive change. Nevertheless, when the refinance incentive increases, the PR Move tends to smoothly increase.

It is important to notice that the new way of defining the refinance incentive leads to the same trends as in Sterk (2005). However, increases in prepayment, when the refinance incentive is higher, are much more significant. Therefore, the data shows that mortgagors are more sensitive to refinance incentive, as defined in this Thesis, compared to Sterk (2005) results. As an example, when the refinance incentive is 2%, the PR for refinancing is around 12%, while in Sterk (2005) is around 7.7%.

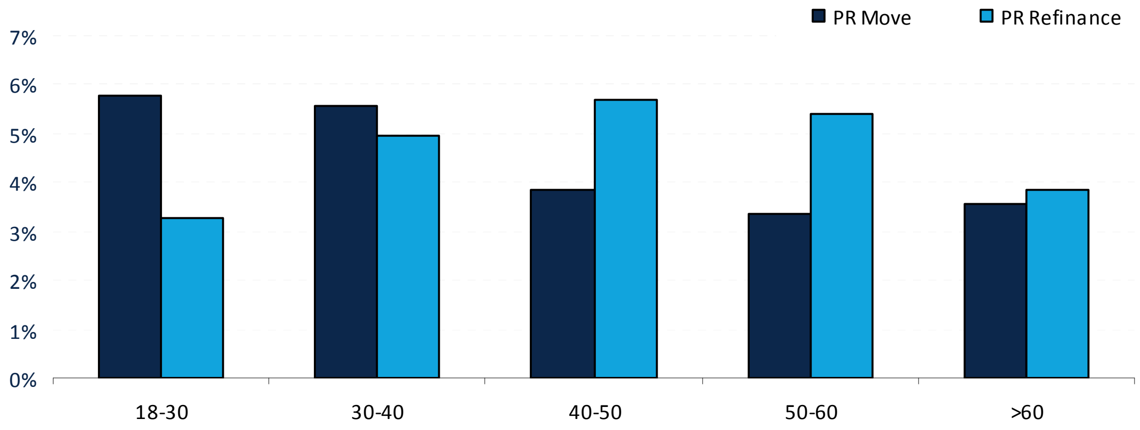

Figure 17 - Prepayment by Age Cohort of the Mortgagor

By looking at the observed prepayment rates as a function of different age cohorts of the Mortgagors, see Figure 17, we can conclude that the prepayment rate for moving decreases with the age of the mortgagor, except for the last cohort. This is logical, because elderly people tend to settle down and are less willing to move. On the other hand, the prepayment rate for refinance increases with the age of the mortgagors, except for the last cohort (>60) where people seem to be less sensitive to mortgage rate moves in market. One could wonder why the youngest cohort has such a low percentage for prepayment rate due to refinancing. However, since younger people move more frequently they have to refinance less because the two options can not be exercise at the same time. The same trend was also found by Sterk (2005).

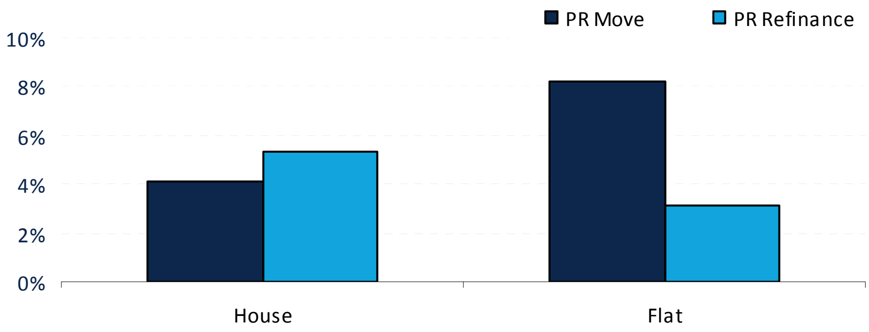

Figure 18 - Prepayment by Property Type

As expected, and in line with Sterk (2005) results, when the underlying of a mortgage is a house, mortgagors move less frequently. From the refinancing point of view, people that own a house are more likely to refinance their mortgages than if they were living in an apartment as it can be seen in the Figure 18. Although the results have the same trend as in Sterk (2005) the PR Move when the underlying is a Flat is substantially higher, representing an increase from 5% to 8.9%. Overall, the results show that people are much more likely to move, when they live in the Flat.

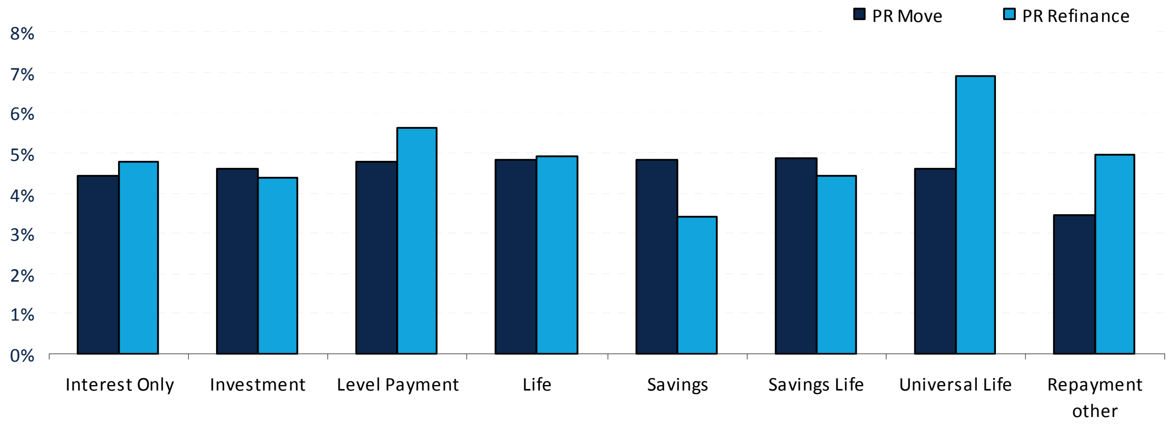

Figure 19 - Prepayment by Mortgage Type

Prepayment rates for moving do not seem to vary significantly across different mortgages types, as shown in Figure 19. The prepayment rate for refinance is high for universal life mortgages and level payment mortgages, when comparing to the other types of mortgages.

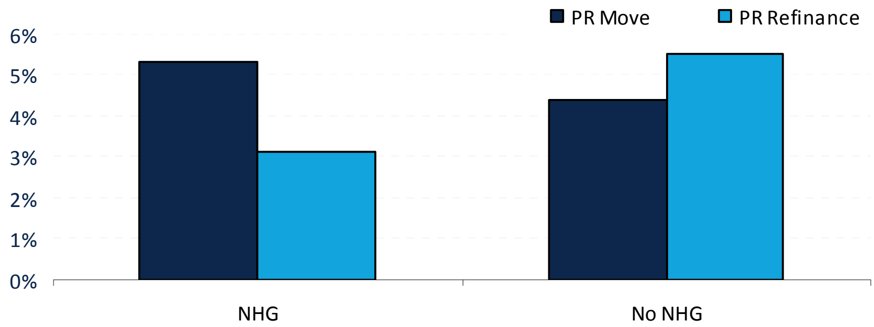

Figure 20 - Prepayment by NHG

In line with De Jong (1998), prepayment rates due to refinancing are higher for mortgages that are not guaranteed by the NHG program (Figure 20). However, the same is not observed for the PR Move. Moreover, the data show a significant difference in the prepayment behaviour of mortgagors that have a guarantee and those that do not have it, going against the findings of Alink (2002). This further substantiates the inclusion of this dummy variable in our model. When comparing to Sterk (2005), the observed results show the same patterns for both PR Move and PR refinance.

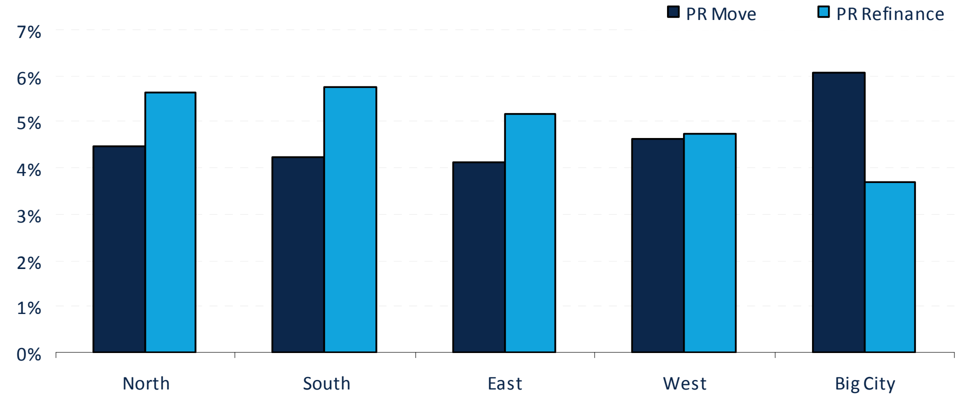

Figure 21 - Prepayment per Region

In Figure 21 the prepayment rates are distributed across regions in The Netherlands. By looking at the graph, it may be surprising that the PR move is higher in big cities than in the rest of the regions. However, cities usually have more apartments and this may help to explain both, the high value for PR move and low value for PR refinance. As expected, the extended data-set has the same characteristics with respect to these dummy variables, as in Sterk (2005).

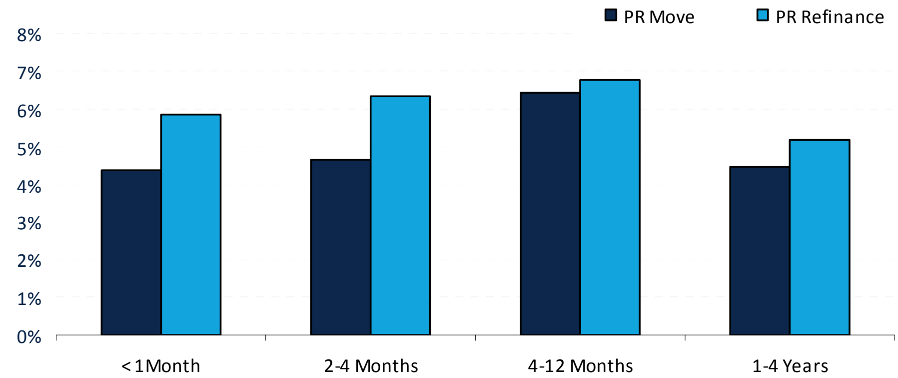

Figure 22 - Prepayment by Categories of Time until Reset

Looking at Figure 22 it can be striking that prepayment rates are higher for mortgages that have two to twelve months until reset, when comparing to those that have one to four years to go. However, this may be explained if we take into account that mortgagors only are aware that they have a mortgage closer to the interest reset period.

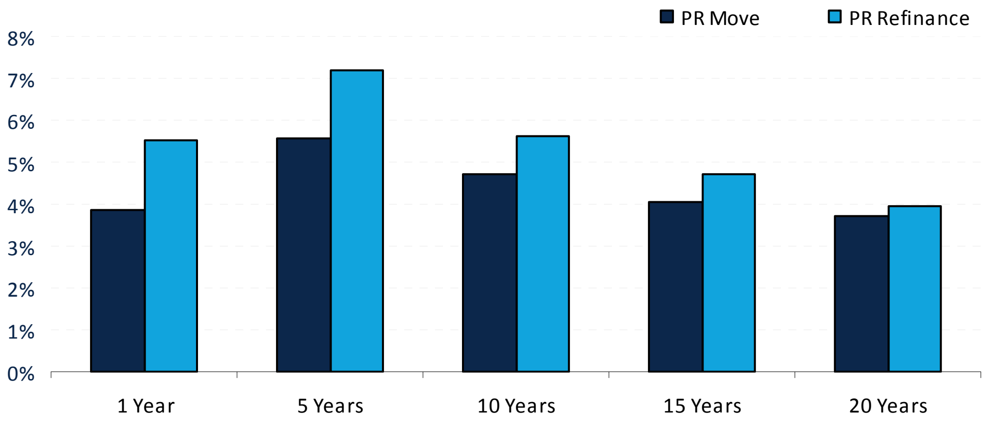

Figure 23 - Prepayment per Fixed Rate Period

Surprisingly, prepayment rates decrease while the interest rate fixed period increases, exception made for the first two categories in Figure 23. As explained before, the opposite was expected. One could argue that mortgagors do not behave in a rational way or that mortgagors that take a longer fixed interest rate period are not aware of interest rate moves. They may prefer a long fixed period to not worry about any changes in their payments. Sterk (2005) draw the same conclusion. However, Sterk (2005) did not include mortgages with interest fixed rate period of one year in the data-set. It is important to reflect why these mortgages do not follow the same trend as others. One possible explanation is that mortgages that have an interest fixed rate period of one year have less prepayment risk, compared to the rest, because the refinance incentive is set equal to zero every year when the mortgage rate is set equal to the prevailing one at the reset date.

3.5 Estimation Results Multinomial Logit

In order to obtain the estimation results, the CATMOD procedure was used in SAS. This procedure is used to estimate the coefficients for the predictors. It is important to note that the interpretation of the estimations is not as straightforward as in linear regressions, since on the left hand side of the regression equation there is a Logit instead of the mean.

The dependent variable is called *prepaid* and it can take one of the following values:

- **Zcontinuing**: if a part of a mortgage in a certain month continues to the next one without any prepayment occurring;
- **Refinance**: if a part of a mortgage in a certain month is prepaid due to refinance;
- **Move**: if a part of a mortgage in a certain month is prepaid due to the mortgagor moving to other house/flat.

The SAS program merges and computes the data-set and checks if a prepayment occurred. If not *prepaid* equals “Continue”, if it did occurred, it checks if a penalty fee existed. If penalty fee existed, then *prepaid* is equal to “Refinancing”, because if the principal is paid due to a house sale, no penalty can be charged to the mortgagor under Dutch Law. Obviously, if the mortgage was prepaid and penalty was not charged, *prepaid* is set equal to “Move”.

In the following table the observed frequencies of the dependent variable, *prepaid*, can be seen:

Prepaid	Frequency
move	32787
refinance	36303
zcontinue	8368864

Table 1 - Frequencies

As expected, "Continue" is the most common output of the variable *prepaid* in the data-set and it will be considered as the reference category in the Multinomial Logit Model. Therefore, as explained in Chapter 2, one binary Logit is considered for refinance versus continue and other for move versus continue. Both are estimated simultaneously.

The majority of the explanatory variables are composed by a set of dummy variables. Thus, one category will be set as the reference category and it does not need its own dummy to be included in the regression, since it is uniquely identified by the case when all other variables are equal to zero. For the month variables, "January" was set as the reference, while the base category for the fixed rate period was set equal to the variable "Fixed1y". For the region, the variable "North" is the reference and for the age of the mortgagor, "agecat1" is the reference. Finally, remaincat1 is the reference for the time until reset and for the mortgage type "repayment other" is set as reference.

In table 5 on the next page, one can see the estimated results, produced by the Multinomial Logit Model.

Full Model						
Parameter	Move			Refinance		
	Estimate	Standard Error	Pr > ChiSq	Estimate	Standard Error	Pr > ChiSq
Intercept	-5.7549	0.0709	<.0001	-5.516	0.0659	<.0001
refinance_1	0.1185	0.00526	<.0001	0.3618	0.00496	<.0001
level_payment	0.2831	0.0853	0.0009	-0.0879	0.078	0.2597
life	0.2162	0.0586	0.0002	-0.2006	0.0532	0.0002
savings	0.1541	0.067	0.0215	-0.4798	0.0663	<.0001
interest_only	0.2091	0.0579	0.0003	-0.1549	0.0525	0.0031
savings_life	0.1467	0.0632	0.0202	-0.2812	0.0589	<.0001
investment	0.1004	0.0684	0.1422	-0.3721	0.0645	<.0001
universal_life	0.1485	0.0596	0.0127	0.067	0.0536	0.2119
seasoning	0.3452	0.0082	<.0001	0.4508	0.00878	<.0001
south	-0.1173	0.0239	<.0001	-0.0519	0.0217	0.017
east	-0.1608	0.0249	<.0001	-0.1159	0.0228	<.0001
west	-0.1122	0.0227	<.0001	-0.2545	0.0212	<.0001
bigcity	-0.2006	0.0285	<.0001	-0.3989	0.0304	<.0001
flat	0.6667	0.0156	<.0001	-0.3524	0.0219	<.0001
agecat2	0.0672	0.0175	0.0001	0.2361	0.0209	<.0001
agecat3	-0.2837	0.0191	<.0001	0.2896	0.0211	<.0001
agecat4	-0.4771	0.0229	<.0001	0.172	0.0231	<.0001
agecat5	-0.4869	0.0289	<.0001	-0.1821	0.0301	<.0001
NHG	-0.011	0.0143	0.4424	-0.4125	0.017	<.0001
fixed5y	-0.1322	0.02	<.0001	0.1355	0.0202	<.0001
fixed10y	-0.2023	0.0193	<.0001	-0.1859	0.0204	<.0001
fixed15y	-0.2198	0.0259	<.0001	-0.2927	0.0256	<.0001
fixed20y	-0.2082	0.0276	<.0001	-0.2385	0.0279	<.0001
feb	0.1288	0.0312	<.0001	-0.0209	0.0284	0.4626
mar	0.2356	0.0305	<.0001	0.069	0.0279	0.0133
apr	0.2925	0.0301	<.0001	0.2033	0.0271	<.0001
may	0.2819	0.0302	<.0001	0.2679	0.0268	<.0001
jun	0.299	0.0301	<.0001	0.3036	0.0266	<.0001
jul	0.5345	0.0287	<.0001	0.3416	0.0264	<.0001
aug	0.4053	0.0292	<.0001	0.2039	0.027	<.0001
sep	0.3527	0.0294	<.0001	0.0557	0.0277	0.0444
oct	0.3493	0.0294	<.0001	0.0421	0.0279	0.1309
nov	0.3639	0.0299	<.0001	0.178	0.0274	<.0001
dec	0.5747	0.0287	<.0001	0.5337	0.0255	<.0001
remaincat2	0.1798	0.0431	<.0001	0.4856	0.0385	<.0001
remaincat3	0.3349	0.0308	<.0001	0.6713	0.0282	<.0001
remaincat4	0.4124	0.0252	<.0001	0.4088	0.0244	<.0001
remaincat5	0.1832	0.0229	<.0001	0.4136	0.0217	<.0001

Table 2 - Parameter Estimates Full Model

The intercepts are Multinomial Logit estimates for move relative to continue and for refinance relative to continue, when all the predictor variables in the model are evaluated at zero. They show that move or refinance are less common than continue.

Both estimates for the refinance incentive (refinance_1) are positive and significant at 1% level. By looking at the estimates, one could say that an increase in the refinance incentive by one point would make the multinomial log-odds for preferring move to continue increase by 0.1185 and the multinomial log-odds for preferring refinance to continue increase by 0.3618. As expected, a higher value of the refinance incentive increases more the probability of a termination of a mortgage due to refinance than due to move. It should be stated that Sterk (2005) found a negative relation for the refinance estimate for the model that studies the termination of a mortgage due to moving. Our results seem to be closer to reality, since it is feasible to acknowledge that mortgagors would not move less if the refinance incentive becomes higher.

Most of the parameter estimates for the dummy variables for the type of the mortgage are not significant at 1% level, when taking into account the model that compares move to continue. This is understandable since the literature does not provide any explanation why the mortgage type could be related to a house sale. On the other hand, for the model that compares refinance to continue only *interest_only* is not significant at 1% level.

The majority of the month variables estimates are significant at 1% level in both models. It is interesting to note that the estimates are particularly high for *December* in both models and for *July* for the model that compares move to continue. This is in line with our expectations, since the literature suggests these two when considering seasonality phenomenon in prepayment rates.

In the model that compares move to continue, the parameter estimate for mortgagors with age betweens 30 and 40 years old (*agecat 2*) relative to *agecat 1* (age of the mortgagor is between 18 and 30 years old) is 0.0399 for preferring move to continue. Moreover, for the other categories, relative to *agecat1* the multinomial estimates have a negative relation. These negative relations were expected, since it is believed that older mortgagors move less often that young people. When looking at the other model, one realizes that the probability of refinance relative to continue increases as the age of the mortgagor increases with an exception of the last category *agecat5*. In both models all the estimates are significant at 1% level.

The parameter estimates for the dummy variable *flat* is significant at 1% level for both binary Logit Models. If the underlying is a flat the probability of move increases relative to continue by 0.6667 while the probability of refinance decreases relative to continue by 0.3524. This is in line with the expectations since people that live in a flat would move more often than when the underlying is a house.

All the parameter estimates of dummy variables concerning the region of mortgage origination are significant at 2% level. The multinomial estimates are relative to the region *north* and these estimates have a negative influence on the probability of move relative to continue and refinance relative to continue.

The parameters estimates for Seasoning are significant at 1% level for both binary Logits.

The Multinomial Logit estimates that compare the interest rate fixed period of 5, 10, 15, and 20 years to one year interest rate reset, given that all other variables are held constant, are negative for both models (except *fixed 5y* for refinance comparing to continue). It should be noted that all dummy variables, that enclose the explanatory variable, for the fixed interest rate period are significant at 1% level. This is an improvement since these variables were not significant in Sterk (2005).

The Multinomial Logit estimates for the variable time until reset are all significant at 1% level for both binary Logits. Moreover, all the estimates are positive, implying that an increase in the time until reset has a positive effect on the probability of move relative to continue or refinance relative to move.

The parameter estimate for the variable *NHG* is not significant for the model that compares move to continue. Nevertheless, the estimate is negative in both models, implying that an increase of mortgages with a guarantee from the Dutch Government will have a negative impact on the probabilities of prepaying due to move and refinancing.

As explained above, the majority of the parameter estimates for mortgage types are not significant. Moreover, the NHG parameter for move relative to continue is also not significant. Therefore, a set of restricted models will be analysed, where these explanatory variable are excluded. The following restricted models were produced:

	Variables Not Included	Results
Restricted Model 1	Mortgage Type	Appendix 3.A
Restricted Model 2	Guarantee	Appendix 3.B
Restricted Model 3	Mortgage Type and Guarantee	Appendix 3.C

Table 3 - Restricted Models Characteristics

In the following Section, the full model and the restricted ones will be evaluated, to check which is the most appropriate to use in this Thesis.

3.6 Comparison between Full and Restricted Models

In 1974, Hirotsugu Akaike developed an information criterion (AIC) useful for model selection. AIC rewards the goodness-of-fit and it also includes a penalty as increasing function of the number of explanatory variables used. By doing this, the AIC gives a better ranking for models that use fewer parameters, discouraging, in this way, over-fitting. The AIC is equal to:

$$AIC = \frac{1}{N}(-2l + 2n) \tag{21}$$

Where, N is the number of observations, l is the log likelihood and n is the number of explanatory variables.

In 1978, Gideon E. Schwarz introduced a new criterion for model selection. It was named the Bayesian Information Criterion (BIC). In econometric models, when the estimation is done by using maximum likelihood estimation, the likelihood can be increased by simply adding parameters. This may lead to over-fitting. The BIC is very close to the AIC criterion, with the main difference being that BIC has a stronger penalty for additional parameters than the AIC. As with AIC, a model with a lower BIC is preferred. The BIC is equal to:

$$BIC = \frac{1}{N}[-2l + n\log(N)] \tag{22}$$

The AIC and BIC values were computed for the full model and the three restricted models. The results can be seen in the following table:

	Full Model	Restricted Model 1	Restricted Model 2	Restricted Model 3
N	8437954			
n	39	32	38	31
-2LogL	870795.52	871281.42	871433.84	872042.34
AIC	0.103209086	0.103265012	0.103284498	0.103354953
BIC	0.103231855	0.103283694	0.103306683	0.103373051

Table 4 - AIC and BIC Criterions

In our case, the full model has the lowest value for the AIC and the BIC. However, since the differences in the AIC and BIC are small across the different models, it was chosen to use restricted model 1 for the remaining of this Thesis. The reason behind this is that it simplifies the coordination with assumptions from other models within NIBC that use the prepayment forecasting model.

3.7 Conclusions

This Chapter started by describing the data-set used as input for the Multinomial Logit Model. The data-set has around 8.4 million observations where each mortgage part has a monthly observation during the time it existed.

Section 3.3 presented the basic characteristics of the data-set. The portfolio can be considered recent, since 85% of the mortgages were originated after 1999. The most common fixed interest reset periods are one, five, ten, and twenty years, which differs from Sterk (2005) observations. Sterk (2005) excluded mortgages with a fixed interest reset period lower than 5 years. From our point of view, excluding those mortgages from the data-set is not optimal since they carry substantial prepayment risk. Therefore, only variable interest reset mortgages were excluded from the data-set.

In Section 3.4, the observed prepayment rates due to moving and refinance were shown. On an aggregate level, they are highly volatile. This is mainly due to prepayment caused by refinancing. Prepayment rates, derived from moving, usually took values between 4 and 6 percent since January 2000 until October 2009. Total prepayment rates show peaks in July and December. These strong seasonality effects are mainly due to holidays and fiscal reasons. It was also found that the heterogeneity between mortgages regarding the principal value is not sufficient to have a considerable impact. Hence, prepayment rates do not vary across mortgages with different principals. Moreover, prepayment rate due to refinance is highly sensitive to the refinance incentive. This sensitivity was higher than in Sterk's (2005) results due to the new way of calculating the refinance incentive. This seems to be closer to reality and therefore, it shows an improvement from Sterk (2005) results. This can be considered a key result of this Thesis, as defining refinance incentive is rather difficult, because at each reset date the mortgagor can refinance to a new mortgage with a completely different set of characteristics. Capturing in a better way the sensitivity of mortgagors to the refinance incentive is a clear sign of improvement.

Section 3.5 presented the estimates results of the Multinomial Logit Model. Both estimates for the refinance incentive (*refinance_1*) are positive and significant at 1% level. Sterk (2005) found a negative relation for the refinance incentive estimate in the model that studies the termination of a mortgage due to moving. Our results seem to be closer to reality, since it is feasible to acknowledge that mortgagors would not move less if the refinance incentive gets higher. The other main discoveries were that the parameter estimates for seasonality seem to catch the peaks in July and December and the majority of the months estimates were significant at 1% level for both models. Moreover, the Multinomial Logit estimates that compare the interest rate fixed period of 5 , 10, 15, and 20 years to one year interest rate reset given that all other variables are held constant are negative for both models. It should be noticed that all the dummy variables that enclose the explanatory variable for the fixed interest rate period are significant at 1% level. This is an improvement since these variables were not significant in Sterk (2005). It was also shown that majority of the parameters estimates for mortgage types are not significant. The parameter estimate for the variable NHG is not significant for the model that compares move to continue. Nevertheless, both estimates are negative for both models, implying that an increase of mortgages with a guarantee from the Dutch Government will have a negative impact on the probabilities of prepaying due to move and refinancing. All the other parameter estimates regarding the other explanatory variables were in line with expectations as explained in Section 3.5.

Section 3.6 compares the full model with the restricted models by using the Akaike information and the Bayesian information criterions. The full model has the lowest value for the AIC and the BIC. It was chosen to use restricted model 1 for the remaining of this Thesis. This was done because it simplifies the coordination with assumptions from other models within NIBC that use the prepayment forecasting model.

Appendix 3.A - Parameter Estimates Restricted Model 1

Restricted Model 1						
Parameter	Move			Refinance		
	Estimate	Standard Error	Pr > ChiSq	Estimate	Standard Error	Pr > ChiSq
Intercept	-5.5532	0.0429	<.0001	-5.6664	0.0424	<.0001
refinance_1	0.1179	0.00524	<.0001	0.3614	0.00493	<.0001
seasoning	0.3447	0.00819	<.0001	0.4545	0.00877	<.0001
south	-0.1192	0.0239	<.0001	-0.0482	0.0217	0.0266
east	-0.1615	0.0249	<.0001	-0.1123	0.0228	<.0001
west	-0.1141	0.0227	<.0001	-0.2584	0.0211	<.0001
bigcity	-0.2055	0.0285	<.0001	-0.3935	0.0304	<.0001
flat	0.6633	0.0156	<.0001	-0.3477	0.0218	<.0001
agecat2	0.0634	0.0175	0.0003	0.2427	0.0209	<.0001
agecat3	-0.2866	0.0191	<.0001	0.2965	0.021	<.0001
agecat4	-0.4783	0.0228	<.0001	0.1825	0.0231	<.0001
agecat5	-0.4781	0.0287	<.0001	-0.1864	0.0299	<.0001
NHG	-0.0124	0.0142	0.3792	-0.4444	0.0169	<.0001
fixed5y	-0.1285	0.02	<.0001	0.1441	0.0201	<.0001
fixed10y	-0.2052	0.0192	<.0001	-0.1698	0.0202	<.0001
fixed15y	-0.2116	0.0258	<.0001	-0.307	0.0254	<.0001
fixed20y	-0.2121	0.0274	<.0001	-0.2328	0.0276	<.0001
feb	0.1289	0.0312	<.0001	-0.0208	0.0284	0.4634
mar	0.2356	0.0305	<.0001	0.0689	0.0279	0.0135
apr	0.2926	0.0301	<.0001	0.2031	0.0271	<.0001
may	0.2819	0.0302	<.0001	0.2679	0.0268	<.0001
jun	0.299	0.0301	<.0001	0.3036	0.0266	<.0001
jul	0.5345	0.0287	<.0001	0.3409	0.0264	<.0001
aug	0.4054	0.0292	<.0001	0.203	0.027	<.0001
sep	0.3528	0.0294	<.0001	0.0551	0.0277	0.0466
oct	0.3494	0.0294	<.0001	0.0415	0.0279	0.1364
nov	0.3638	0.0299	<.0001	0.1779	0.0274	<.0001
dec	0.5747	0.0287	<.0001	0.5337	0.0255	<.0001
remaincat2	0.1777	0.0431	<.0001	0.4784	0.0385	<.0001
remaincat3	0.3318	0.0307	<.0001	0.6657	0.0281	<.0001
remaincat4	0.407	0.0251	<.0001	0.4086	0.0243	<.0001
remaincat5	0.1808	0.0229	<.0001	0.4134	0.0216	<.0001

Table 5 - Parameter Estimates Restricted Model 1

Appendix 3.B - Parameter Estimates Restricted Model 2

Restricted Model 2						
Parameter	Move			Refinance		
	Estimate	Standard Error	Pr > ChiSq	Estimate	Standard Error	Pr > ChiSq
Intercept	-5.7622	0.0703	<.0001	-5.7254	0.0654	<.0001
refinance_I	0.1188	0.00524	<.0001	0.3699	0.00494	<.0001
level_payment	0.2834	0.0853	0.0009	-0.0798	0.0779	0.3058
life	0.2165	0.0585	0.0002	-0.1969	0.0532	0.0002
savings	0.1529	0.067	0.0225	-0.5194	0.0663	<.0001
interest_only	0.2101	0.0579	0.0003	-0.1322	0.0524	0.0116
savings_life	0.1455	0.0632	0.0213	-0.3336	0.0588	<.0001
investment	0.1013	0.0684	0.1385	-0.35	0.0644	<.0001
universal_life	0.15	0.0596	0.0118	0.1017	0.0536	0.0577
seasoning	0.3448	0.00818	<.0001	0.4386	0.00874	<.0001
south	-0.1164	0.0239	<.0001	-0.0231	0.0217	0.2865
east	-0.1604	0.0249	<.0001	-0.1019	0.0228	<.0001
west	-0.1111	0.0227	<.0001	-0.2225	0.0211	<.0001
bigcity	-0.1993	0.0284	<.0001	-0.3629	0.0304	<.0001
flat	0.6649	0.0154	<.0001	-0.4094	0.0218	<.0001
agecat2	0.0682	0.0174	<.0001	0.2687	0.0209	<.0001
agecat3	-0.2814	0.0189	<.0001	0.3621	0.0209	<.0001
agecat4	-0.4742	0.0226	<.0001	0.2598	0.0229	<.0001
agecat5	-0.4834	0.0286	<.0001	-0.0807	0.0299	0.0068
fixed5y	-0.1308	0.0199	<.0001	0.1739	0.0201	<.0001
fixed10y	-0.2011	0.0193	<.0001	-0.1561	0.0203	<.0001
fixed15y	-0.2181	0.0258	<.0001	-0.2482	0.0255	<.0001
fixed20y	-0.2066	0.0276	<.0001	-0.1946	0.0279	<.0001
feb	0.1288	0.0312	<.0001	-0.0214	0.0284	0.4511
mar	0.2356	0.0305	<.0001	0.0695	0.0279	0.0128
apr	0.2925	0.0301	<.0001	0.2034	0.0271	<.0001
may	0.2819	0.0302	<.0001	0.2684	0.0268	<.0001
jun	0.2991	0.0301	<.0001	0.3041	0.0266	<.0001
jul	0.5345	0.0287	<.0001	0.3417	0.0264	<.0001
aug	0.4053	0.0292	<.0001	0.2036	0.027	<.0001
sep	0.3526	0.0294	<.0001	0.0551	0.0277	0.047
oct	0.3492	0.0294	<.0001	0.0418	0.0279	0.1341
nov	0.3639	0.0299	<.0001	0.1781	0.0274	<.0001
dec	0.5747	0.0287	<.0001	0.5339	0.0255	<.0001
remaincat2	0.1817	0.043	<.0001	0.5362	0.0385	<.0001
remaincat3	0.3367	0.0307	<.0001	0.7173	0.0282	<.0001
remaincat4	0.4133	0.0252	<.0001	0.4342	0.0245	<.0001
remaincat5	0.1835	0.0229	<.0001	0.422	0.0218	<.0001

Table 6 - Parameter Estimates Restricted Model 2

Appendix 3.C - Parameter Estimates Restricted Model 3

Restricted Model 3						
Parameter	Move			Refinance		
	Estimate	Standard Error	Pr > ChiSq	Estimate	Standard Error	Pr > ChiSq
Intercept	-5.5607	0.0421	<.0001	-5.8772	0.0418	<.0001
refinance_I	0.1182	0.00523	<.0001	0.3698	0.00491	<.0001
seasoning	0.3441	0.00817	<.0001	0.4402	0.00873	<.0001
south	-0.1181	0.0239	<.0001	-0.0163	0.0217	0.4519
east	-0.161	0.0249	<.0001	-0.0963	0.0228	<.0001
west	-0.113	0.0227	<.0001	-0.2242	0.0211	<.0001
bigcity	-0.2041	0.0284	<.0001	-0.3539	0.0304	<.0001
flat	0.6613	0.0154	<.0001	-0.4102	0.0217	<.0001
agecat2	0.0646	0.0174	0.0002	0.2784	0.0208	<.0001
agecat3	-0.2839	0.0189	<.0001	0.3767	0.0209	<.0001
agecat4	-0.4748	0.0225	<.0001	0.2812	0.0229	<.0001
agecat5	-0.4739	0.0283	<.0001	-0.0712	0.0297	0.0164
fixed5y	-0.1267	0.0198	<.0001	0.1898	0.02	<.0001
fixed10y	-0.2038	0.0191	<.0001	-0.1357	0.0201	<.0001
fixed15y	-0.2097	0.0257	<.0001	-0.2608	0.0254	<.0001
fixed20y	-0.2103	0.0274	<.0001	-0.1877	0.0276	<.0001
feb	0.1289	0.0312	<.0001	-0.0213	0.0284	0.4523
mar	0.2357	0.0305	<.0001	0.0694	0.0279	0.0128
apr	0.2926	0.0301	<.0001	0.2033	0.0271	<.0001
may	0.2819	0.0302	<.0001	0.2685	0.0268	<.0001
jun	0.299	0.0301	<.0001	0.3042	0.0266	<.0001
jul	0.5345	0.0287	<.0001	0.3411	0.0264	<.0001
aug	0.4054	0.0292	<.0001	0.2025	0.027	<.0001
sep	0.3527	0.0294	<.0001	0.0544	0.0277	0.0495
oct	0.3493	0.0294	<.0001	0.0411	0.0279	0.1399
nov	0.3638	0.0299	<.0001	0.178	0.0274	<.0001
dec	0.5747	0.0287	<.0001	0.5339	0.0255	<.0001
remaincat2	0.1799	0.043	<.0001	0.531	0.0385	<.0001
remaincat3	0.3337	0.0306	<.0001	0.7128	0.0281	<.0001
remaincat4	0.408	0.0251	<.0001	0.4328	0.0244	<.0001
remaincat5	0.1811	0.0229	<.0001	0.4204	0.0217	<.0001

Table 7 - Parameter Estimates Restricted Model 3

Chapter 4: Back-testing

4.1 Introduction

The aim of this Chapter is to back-test the model proposed in this Thesis. The Chapter starts by describing the types of prediction errors that can be found, when using the Multinomial Logit Model in Section 4.2.

Section 4.3 describes the back-testing framework. Section 4.4 presents the results of out-of-sample ex-post back-testing results while Section 4.5 shows the results of two out-of-sample ex-ante back-tests. Section 4.6 back-tests the model using fewer observations as in-put to see if the model still produces acceptable predictions. This is done from an operational point of view, since the programs used to obtain predictions cost a lot of time. Finally, Section 4.7 discusses the results and draws the conclusions.

4.2 Types of Prediction Errors

There are three main sources of prediction errors when applying the Multinomial Logit Model to predict prepayment rates:

- The observed prepayment rate is a random variable

Using Equations 12 and 13, in Chapter 2 Section 5, one could arrive to the observed SMM for moving and refinance in a certain month. These values will always contain an error because we are looking at an aggregate level (monthly) rather than at an individual level (observation). For example, if we consider a portfolio of two mortgages with the same outstanding notional balance and with the same probability of prepayment, equal to:

$$SMM_{Total} = SMM_{Move} + SMM_{refinance} = 15\% \tag{23}$$

It is easy to understand that three realizations are possible:

- Both do not prepay;
  $$SMM_{total} = (1-0.15)*(1-0.15) = 72{,}25\% \tag{24}$$
- One prepays and the other not;
  $$SMM_{Total} = 2*[0.15*(1-0.15)] = 25{,}5\% \tag{25}$$
- Both prepay;
  $$SMM_{total} = 0.15*0.15 = 2{,}25\% \tag{26}$$

Since it was calculated that the prepayment rate in that month was 15% a prediction will always exist error. The reason behind this is in the way the SMM is calculated. As it can be seen in Equation 12 and 13, the numerator is a sum of independent random variables.

If we assume that the variances provided by the Equations 12 and 13 are bounded by the Central Limit Theorem (CLT) then the prediction error follows a normal distribution with zero mean and a variance given by Equations 15 and 17. It should be highlighted that assuming that prediction errors follow a normal distribution is a rather strong assumption. Nevertheless, it is a common practice in prepayment literature.

Sterk (2005) proves that this source of prediction error tends to be negligible if the number of included mortgages is large enough. If we considered that all mortgages have an equal outstanding balance Equations 15 and 17 transform into:

$$Var[SMM_{refinance}(t)] = \frac{\sum_{i=1}^{N} \tau_{it2}(1-\tau_{it2})}{N^2} \tag{27}$$

$$Var[SMM_{move}(t)] = \frac{\sum_{i=1}^{N} \tau_{it3}(1-\tau_{it3})}{N^2} \tag{28}$$

Therefore if N (number of mortgages), goes to infinity, the Variance of prepayment rate for moving and refinance converges to zero. Since the data-set used in this work includes more than 8 million observations this source of error is considered irrelevant.

- Errors induced by the parameter estimates

The Nobel Laureate in Physics Niels Bohr once said that "Prediction is very difficult, especially if it is about the future", thus even the parameter estimates that best maximize the likelihood of the data-set will fail to predict 100% correctly the prepayment rates in the future. This source of error arises from the variability in Equations 18, 19, and 20 expressed by the terms $\exp(X_{it}\widehat{\beta}_j)$ and τ_{ij}. It should be kept in mind that the model was estimated using Maximum Likelihood. Considering this the parameters follow a Multivariate Normal distribution. The arguments of this distribution are the means of the parameter estimates and the covariance matrix, obtained by the Multinomial Logit model.

It is important to notice that since we are using a Multinomial Logit Model the predicted probabilities of moving and refinance are drawn by applying a nonlinear function to the parameter estimates. Since, $E[g(x) \neq g(E[x])$, except for particularly cases (e.g. $g(x)$ is linear), substituting the expected values of the parameters, which are random variables, into a nonlinear function to obtain a prediction will imply that the prediction will not yield the expected value of the function forming some prediction bias. HSU and Wilcox (2000) provide a Monte Carlo approach for approximating the exact stochastic prediction. Nevertheless, it is a standard practice to ignore the estimation error explained above since it implies a great deal of computational time. Thus, it is not feasible. Moreover, the prediction bias associate with large portfolios such as the one we are using tend to be small or even negligible implying that the prediction error in our case will steam almost completely from the randomness of the parameter estimates.

- Omitted relevant variables

As in any model, there will always be an error associated with non-inclusion of variables that help to explain the prepayment rates behaviour. Some of these variables are even not observable.

- Model class error

As in any empirical model, we have always the error associated with the model class. There is always a risk that prepayment behaviour can not be explained by a Multinomial Logit Model. This is a risk inherent in modelling people behaviour since people do not behave optimally.

4.3 Back-testing Framework

Out-of-sample forecast performance is usually seen as the acid test of an econometric model. The reasoning behind that is that a good out-of-sample performance provides a strong support to the model underlying theory. Usually this can be done through a comparison to rival forecasts or relative to in-sample performance. Since it is not possible to access data from a different mortgage portfolio, the out-of-sample forecast performance will be done relative to the in-sample performance.

Out-of-sample forecast performance versus in-sample performance can be done in two ways. If the predictions are made before the outcomes have occurred and evaluated at a later stage, when the outcomes are known, it is denominated as ex-ante. On the other hand, if they are evaluated against a sub-set of the original data-set, it is known as ex-post.

Since our observations can be sorted by date, we are able to apply both approaches to access the out-of-sample prediction performance of our Multinomial Logit Model. The back-testing framework is presented in the following table:

Back-Testing	Column A	Column B
I	All Mortgages with an Odd ID	All Mortgages with an Even ID
II	Mortgages that were alive or existed until Dec 2005	Mortgages that were alive or originated after Dec 2005
III	Mortgages that were alive or existed until Dec 2007	Mortgages that were alive or originated after Dec 2007
IV	Mortgages that were alive or existed between Jan 2006 and Dec 2008	Mortgages that were alive or originated after Dec 2008

Table 8 - Back-testing Framework

As shown in the above Table, in all back-testing frameworks the initial data-set is split into two data-sets. After that, the predicted CPR for moving and refinance are calculated based on the parameters estimated by the Multinomial Logit Model using data-set in the first column as input for the out-of-sample data-set observations (second column). Moreover, in-sample predictions are made for the out-of-sample data-set (from now on referred to as in-sample predictions). The reason for this is that the in-sample predictions are made with the parameters estimates that better maximize the likelihood for the out-of-sample data-set and therefore can be seen as reference since they are the best predictions our model could ever produce for that data-set. This information is summarized in the following Figure:

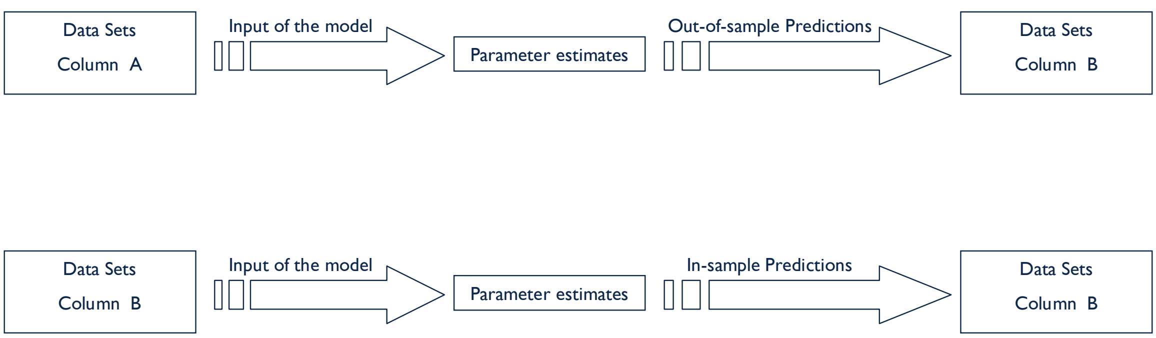

Figure 24 - Explanation of production of in-sample and out-of-sample predictions

It is important to note that Back-testing framework I incorporates all mortgages that have an odd I.D. number as input to the Multinomial Logit Model. Thus, the data-set used for estimation incorporates the full date range of the initial data-set. Therefore, since the entire regime shifts are captured by the Multinomial Logit Model better results are expected compared to the other frameworks where an Ex-Ante approach is taken.

Back-testing Framework IV is done from an operational point of view, since the Multinomial Logit Model takes a great deal of time to run with the full data-set. Moreover, since we are looking at the aggregate level of prepayment rates (SMM) and since the output of this model will be used in other models that are built in

MATLAB within NIBC, the predictions have to be constructed in MATLAB. This implies a long process of exporting and importing data. Hence, it is relevant to know whether a model that uses as input only three years of historical data presents satisfactory results, from an operational point of view.

4.4 Back-testing I

Figure 25 - Back-testing I, CPR Total Observed Vs CPR Total Predicted

Figure 26 - Back-testing I, CPR Move Observed Vs CPR Move Predicted

Figure 27 - Back-testing I, CPR Refinance Observed Vs CPR Refinance Predicted

Table 9 - Back-testing I, Statistics on the Residuals

Figure 28 - Back-testing I, 95% C.I for CPR Total Predicted

Figure 29 - Back-testing I, In-Sample Vs Out-of-sample

4.5 Back-testing II and III

Figure 30 - Back-testing II, 95% C.I for CPR Total Predicted

Figure 31 - Back-testing II, In-Sample Vs Out-of-sample

Table 10 - Back-testing II, Statistics on the Residuals

Figure 32 - Back-testing III, 95% C.I for CPR Total Predicted

Figure 33 - Back-testing III, In-Sample Vs Out-of-sample

Table 11 - Back-testing III, Statistics on the Residuals

4.6 Back-testing IV

Figure 34 - Back-testing IV, 95% C.I for CPR Total Predicted

Figure 35 - Back-testing IV, In-Sample Vs Out-of-sample

Table 12 - Back-testing IV, Statistics on the Residuals

4.7 Conclusions

Appendix 4.A - Parameter using Odd Mortgages

Table 13 - Parameter using Odd Mortgages

Appendix 4.B - Histogram Errors

Figure 36 - Back-testing I, Histogram Errors CPR Total

Figure 37 - Back-testing I, Histogram Errors CPR Refinance

Figure 38 - Back-testing I, Histogram Errors CPR Total

Appendix 4.C - Back-testing I, C.I for CPR Move and CPR Refinance

Figure 39 - Back-testing I, 95% C.I for CPR Move Predicted

Figure 40 - Back-testing I, 95% C.I for CPR Refinance Predicted

Appendix 4.D - Parameter Estimates using Even Mortgages

Table 14 - Parameter Estimates using Even Mortgages

Appendix 4.E - Back-testing I, In-Sample Vs Out-of-sample

Figure 41 - Back-testing I, In-Sample Vs Out-of-sample (CPR Move)

Figure 42 - Back-testing I, In-Sample Vs Out-of-sample (CPR Refinance)

Table 15 - Back-testing I, Statistics on the Residuals (CPR Move and CPR Refinance)

Appendix 4.F - Parameter Estimates Using Mortgages until December 2005

Table 16 - Parameter Estimates Using Mortgages until December 2005

Appendix 4.G - Back-testing II, C.I for CPR Move and CPR Refinance

Figure 43 - Back-testing II, 95% C.I for CPR Move Predicted

Figure 44 - Back-testing II, 95% C.I for CPR Refinance Predicted

Appendix 4.H - Parameter Estimates using Mortgages after December 2005

Table 17 - Parameter Estimates using Mortgages after December 2005

Appendix 4.I - Back-testing II, In-Sample Vs Out-of-sample

Figure 53 - Back-testing IV, In-Sample Vs Out-of-sample (CPR Move)

Table 26 - Back-testing IV, Statistics on the Residuals (CPR Move)

Figure 54 - Back-testing IV, In-Sample Vs Out-of-sample (CPR Refinance)

Table 27 - Back-testing IV, Statistics on the Residuals (CPR Refinance)

Chapter 5: P&L Impact due to Prepayment Risk

5.1 Introduction

The aim of this Chapter is to calculate the P&L impact caused by prepayment risk. Moreover, the P&L prepayment impact is compared when calculated with current CPR assumptions and when calculated with CPR assumptions, produced by the Multinomial Logit Model. Section 5.2 explains the method of valuation and how the cash flows of a mortgage can be calculated. Section 5.3 presents the current NIBC assumptions regarding CPR while Section 5.4 will discuss the model used to Mark to Market NIBC portfolio while Section 5.5 discusses the MtM results produced by both assumptions. Finally, Section 5.6 presents the main conclusions of this Chapter.

5.2 Cash Flows Calculation

Up to a certain degree, mortgages are still considered safe investments. Nevertheless, there is always the risk associated with the change of interest rates because the present value of the cash flows, which characterize a mortgage, will always be dependent on the future development of interest rates. The existence of prepayments simply makes it more complicated, as the payoff function of a mortgage becomes quite difficult to estimate.

It is important to understand why mortgagees should assign a market value to their mortgage portfolios. Firstly, any financial institution has the option to price their assets at a market value under IFRS (International Financial Reporting Standards). Secondly, the valuation techniques used to value a mortgage portfolio can be used as a risk management tool to hedge the underlying risk.

In this Thesis the valuation model of a mortgage portfolio will be based on what is known as the deterministic valuation method. This method assumes a constant prepayment rate. The present value of the cash flows are calculated using spot rates, also known as zero rates, plus some spread to account for costs and risks. The deterministic valuation, Equation 32, gives the price of a mortgage at time $t = 0$ (in months):

$$\pi(0) = \sum_{t=1}^{Tr} \frac{C_t}{(1 + R_t(0) + Spread)^{t/12}}, \tag{32}$$

Where, C_t is the mortgage cash flow at time t, $R_t(0)$ is the spot rate at time t and T_r is the time until maturity measured in months.

The negative aspect of this valuation model is that it does not take into account the stochasticity in interest rates and prepayment rates. Moreover, the spread that is added in the discount factor, to take into account prepayment risk, credit risk, fair value risk, servicing fees and other costs are set by NIBC models and expert judgement. Hence, it is intrinsically linked to the experience of the mortgagee.

To overcome the above negative points, of the deterministic valuation method, the dynamic asset pricing theory could be applied. This method uses two models as input, one for prepayment risk, and another for the term structure. The deterministic valuation method is used in the remaining of this Thesis.

In this Section three different ways of valuing the cash flows of a mortgage throughout time are explained. The first method addresses the way in which a mortgage cash flow should be calculated, if the mortgagor had no possibility of prepayment. The second method explains how a mortgage cash flow should be calculated from a theoretical point of view, if prepayments are allowed. Finally, the third method explains how NIBC currently calculates the cash flows of the mortgages included in their portfolio.

It should be stated that the calculation of the payoff function of a mortgage portfolio is complex since contractual mortgage cash flows usually consists of interest payments and principal repayments which vary across mortgage types. For instance, for annuity mortgages the interest payments and principal repayments are combined in a manner that the total cash flow is constant over time, while for linear mortgages the principal repayments are the same every month, meaning that the total monthly cash flows decreases over the life of the mortgage.

- **Mortgage cash flows when there is no possibility of prepayment**

Contractual cash flows, C_t^C, have two components. The first one is the money a mortgagee receives due to interest payments, IP_t^C, while the other is the money received due to principal repayments PP_t^C as it can been seen in the following Equation:

$$C_t^C = IP_t^C + PP_t^C \tag{33}$$

If we now define I^C as the contractual mortgage rate, the calculation of the interest payments is straightforward:

$$IP_t^C = B_{t-1}^C * \frac{I^C}{12} \tag{34}$$

Where, B_t^C represents the outstanding principal at the end of month t:

$$B_t^C = B_{t-1}^C - PP_t^C \tag{35}$$

In this method of calculating cash flows, as well in the remaining two methods, every cash flow is assumed to take place at the start of every month. As stated before, the principal repayments vary across mortgage types. Therefore, for interest-only mortgages, they are equal to zero, exception made for the month that corresponds to the maturity of the mortgage. For linear mortgages, principal repayments are equal every month, with the value of the principal at origination divided by the number of months until maturity:

$$PP_t^C = \frac{B_{t=0}^C}{T_M} \tag{36}$$

Finally, for a level payment mortgage, the principal prepayments are equal to:

$$PP_t^C = B_{t=0}^C * \left[\frac{i*(1+i)^n}{(1+i)^n - 1}\right] \tag{37}$$

Where, i is equal to $I^C/12$ and n is the number of months of the mortgage.

- **Mortgage cash flows adjusted for prepayment**

The main difference, when a mortgagor has the option to prepay is that the principal payments, PP_t, should include both repayments and prepayments. Moreover, the cash flow associated with prepayment penalties, PE_t, should be calculated. Thus, the total cash flow at each month is equal to:

$$C_t = IP_t + PP_t + PE_t \tag{38}$$

By adjusting the contractual cash flows for prepayment, one arrives to the correct way of accessing the cash flows of a mortgage, where the mortgagor has the option of prepayment. In order to do this, the concept of single month mortality, SMM, should be recalled. SMM is the conditional probability that a mortgage is prepaid in a given month given that it has not been prepaid before. Hences, we are able to define the unconditional probability that a mortgage is prepaid in a given month by using the following Equation:

$$q_t \equiv (1 - SMM_{t=1}) * (1 - SMM_{t=1}) * \ldots * (1 - SMM_{t=t-1}) * SMM_{t=t} \tag{39}$$

From the above Equation one can easy arrive to the unconditional probability that a mortgage is not prepaid before month t because it is simply equal to $q_{t+1} + \ldots + q_{T_r}$, where T_r is the time until the next reset. Having this in mind and following a result derived by Jamshidian in an internal NIBC document, which can be seen in appendix 5.A, the following relations with the contractual cash flows hold:

$$B_t = (q_{t+1} + \ldots + q_{T_r}) * B_t^C \tag{40}$$

$$IP_t = (q_t + \ldots + q_{T_r}) * IP_t^C \tag{41}$$

$$PP_t = (q_t + \ldots + q_{T_r}) * PP_t^C + q_t * B_t^C \tag{42}$$

Bearing the above Equations in mind, the only component in Equation 38 that is missing is the cash flow associated with prepayment penalties. As explained before, penalties exist in order to protect the bank from losses, when mortgage rates decline because although the mortgagor has a refinance incentive, the extra fee may discourage her/him from prepaying. The penalty is the difference between the future contractual interest payments of the mortgage contract and those of a newly originated mortgage with the same characteristics until the next interest reset date. It should be highlighted that in the Netherlands mortgagors are usually allowed to prepay part of their initial principal for free (generally 10 to 20 percent), in order words without a penalty. Therefore, if the mortgage market rate, I_t^{market}, for a mortgage with the same characteristics is lower than the contractual mortgage rate, I^C, and the outstanding principal balance is higher than the percentage of the balance that can be prepaid without penalty, B_{PF}, then the penalty is set equal to:

$$PE_t = (B_{t-1} - B_{PF}) * [e^{(I^C - I_{t-1}^{market})(T_r - t)/12)} - 1] \tag{43}$$

Our prepayment model can be combined with a term structure model such as short rate models, Libor market model or even by Monte-Carlo simulation of PCA (Principal Component Analysis) to forecast prepayment scenarios $SMM, \ldots, SMM_{T_r}$. This fact combined with the above given Equations would permit NIBC to value the portfolio on the monthly basis using this cash flow framework. The problem with this approach is that we are interested of having cash flows streams that might be thirty years ahead. Even if we believe the model predictive power on the short-term it seems quite unreasonable that we can predict the SMM thirty years from now in an accurate way. This fact allied with the monthly change of the prepayment path dramatically increases the volatility in the valuation of the portfolio implying also an increase on the hedging turnover, which can be costly. Moreover, to implement this is not straightforward, because the database where we can access loan part data level does not have all the explanatory variables needed as input in our model making it difficult to forecast prepayment paths for each loan part. Nevertheless, it is certainly a good idea to implement this way in the future for valuing mortgages, at least as a comparison tool with the current framework explained below.

- **NIBC mortgage cash flows**

The cash flows of a mortgage, within NIBC, are adjusted for a constant annualized prepayment rate (CCPR) during the interest rate term of a mortgage. Basically, NIBC assumes that each loan part prepays every month at a constant prepayment rate. Thus, the cash flow Equations are equal to the ones, presented before, in the case where the mortgagor has no option to prepay with a slight change in Equation number 35 where the idea that mortgagors prepay at a constant prepayment rate is incorporated:

$$B_t^C = B_{t-1}^C - PP_t^C - [(B_{t-1}^C - PP_t^C) * (1 - (1 - CCPR_{NIBC})^{\frac{1}{12}})] \tag{44}$$

In the above Equation, the CCPR can be substituted by a constant SMM, since a CPR is simply the annualized version of a single mortality rate. Thus, Equation 44 transforms into:

$$B_t^C = B_{t-1}^C - PP_t^C - [(B_{t-1}^C - PP_t^C) * CSMM_{NIBC}] \tag{45}$$

The advantage of the current framework is that it creates less volatility in the valuation and in the hedging turnover of the mortgage portfolio on the long run if the assumed CCPR is valid.

NIBC does not apply the same CCPR to all mortgages. Mortgages are divided in four different buckets that reflect the remaining time to the next interest reset, where each bucket has a different CCPR, as explained in the next Section.

5.3 Current NIBC Assumptions

Table 28 - NIBC CCPR Assumptions per Bucket

Figure 55 - NIBC CPR Assumption Vs Predicted CPR for Bucket I

5.4 Mark to Market Model

For this and the next Section of this Thesis a SAS program was used as a base for the calculation of the prepayment P&L impact. The SAS program determines the P&L effects of realized mortgage prepayments and it is used mainly for result measurement and risk monitoring of prepayments under NIBC mortgage portfolio.

The idea behind the mark to market model is described in the figure below:

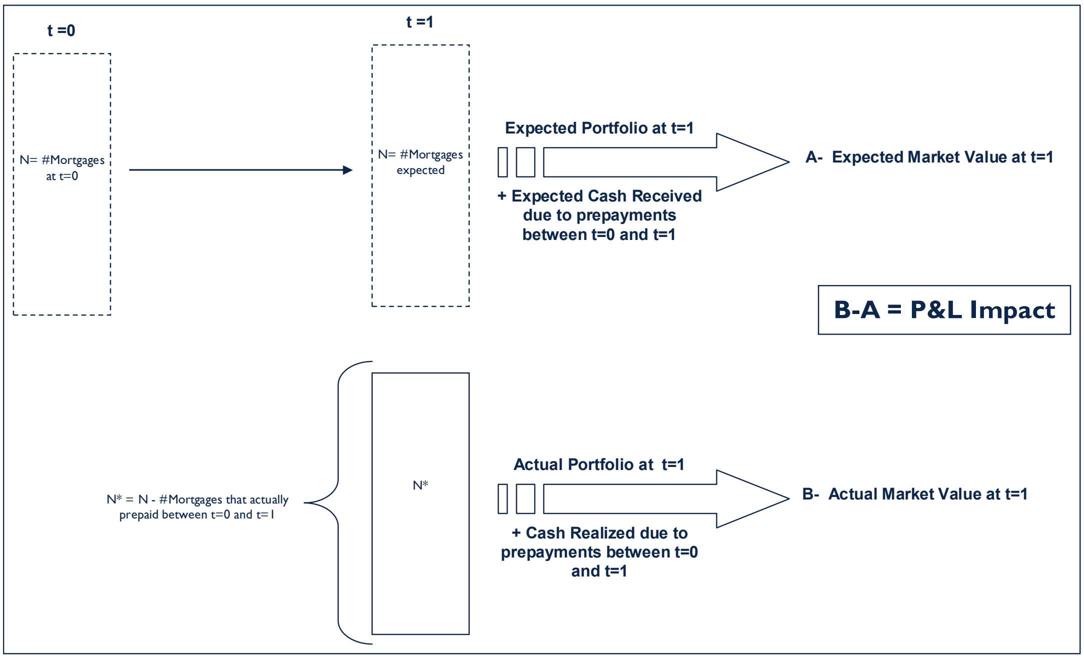

Figure 56 - MtM Model Framework

The model calculates the expected market value of the mortgage portfolio of NIBC at t=1 and adds to that the value of the expected cash received due to prepayments between t=0 and t=1 (A, in Figure 55) by using the following Equation:

$$A = Expected\ Market\ Value\ portfolio = \sum_{i=0}^{N}\left(\sum_{t=1}^{\min\{T_r,360\}}\left[\frac{C_{t,i}^{NIBC}}{(1+R_{t-1}(0)+Spread_{t-1})^{(t-1)/12}}\right]+\right)+(B_0^C-PP_1^C)*CSMM_{NIBC} \tag{50}$$

Where N is equal to the number of mortgage parts included in the portfolio at t=0, T_r is the number of months until next reset of each individual mortgage, R_t are the zero rates derived by bootstrap method using the EURIBOR curve (source: Bloomberg) at t=1.

After this, the program compares the actual portfolio at t=1 and the portfolio at t=0 in order to check the exact number of mortgages that were actually prepaid and it calculates the cash realized due to these prepayments. Having this value it calculates actual market value of the portfolio at t=1 (B, in Figure 55) by adding the present value of the mortgage portfolio at t=1 to the cash realized between t=0 and t=1, by using the following Equation:

$$B = Actual\ Market\ Value\ portfolio = \sum_{i=0}^{N^*}\left(\sum_{t=0}^{\min\{T_r,360\}}\left[\frac{C_{t,i}^{NIBC}}{(1+R_t(0)+Spread_t)^{t/12}}\right]\right)+Cash\ Realized_1 \tag{51}$$

The difference between the Market Value of the actual portfolio and the Market Value of the expected portfolio is the P&L impact due to the difference between prepayment assumptions and realizations in months t=1.

In order for this to work some assumptions were made regarding the portfolio configuration that is included as input. The assumptions were the following:

- New production and mortgage part principal increases between t=0 and t=1 are excluded;
- Bridge mortgages are excluded since the mortgagor has the option to repay the loan at any moment without any penalty;

- Mortgages with interest rate reset date at t=1 are excluded since at each interest reset date the mortgagor has the option of repay or prepay without any penalty;
- In an event of conversion the mortgagor receives a new mortgage loan part and the old loan part is considered to be prepaid due to refinance and a penalty is paid out;
- The cash flows are calculated for each mortgage until the minimum of months until the upcoming reset date and 360 months. This means that we are assuming that the mortgagor redeems the mortgage on the next reset date if no prepayment happens meanwhile or in 30 years from now if the next interest reset date is further away;
- Every mortgage cash flow is assumed to occur at the beginning of each month.

5.5 Historical Mark to Market Results

Table 29 - P&L Impact due to prepayment Risk

Figure 57 - P&L impact NIBC Assumptions Vs P&L impact prepayment model

5.6 Conclusions

Appendix 5.A - Proof Jamshidian

The proof presented below was discovered by Jamshidian in an internal NIBC document (2002). This appendix is a copy of his document using the same notation presented in this Thesis.

1. Mortgage cash flow scheme

A mortgage cash flow scheme describes the contractual cash flows of a cluster or pool of mortgages. It consists of a triple (PP_t^C, IP_t^C, t) where:

$$PP_t^C = principal\ paid\ in\ month\ t \tag{52}$$

$$IP_t^C = Interest\ paid\ in\ month\ t \tag{53}$$

$$t = 1, \ldots, T \tag{54}$$

Hence the total cash flow at month t is $IP_t^C + PP_t^C$. Let B_t^C denote the outstanding principal at month t:

$$B_t^C := PP_{t+1}^C +, \ldots, + PP_t^C . \ (1 \le t \le T) \tag{55}$$

The total principal is $B_0^C := PP_1^C +, \ldots, + PP_T^C$. By convention, $B_T^C = 0$. Note,

$$PP_t^C = B_{t-1}^C - B_t^C . (1 \le t \le T) \tag{56}$$

Let I_t^C denote the mortgage interest rate (coupon) at month t:

$$I_t := \frac{P_t^C}{PP_{t-1}^C} \tag{57}$$

A single mortgage with a fixed coupon I_t and maturity T can be thought as a mortgage cash flow scheme where $PP_t^C = 0$ for $t < T$, and $I_t^C = I_t$ is constant.

2. Conditional Prepayment Rate (CPR)

A mortgage cash flow scheme can be prepaid. If it is prepaid at month $t \ge 1$, then the borrower pays $IP_t^C + PP_t^C +, \ldots, + PP_T^C = IP_t^C + B_{t-1}^C$ at month t and makes no more payment. The conditional prepayment rate SMM_t is defined as:

$$SMM_t := \text{Probability that scheme prepays at month } t \text{ given it has not prepaid before} \tag{58}$$

Equivalently,

$$SMM_t := \frac{probability\ that\ scheme\ prepays\ at\ month\ t}{probability\ that\ scheme\ has\ not\ prepaid\ before} = \frac{q_t}{q_t+,\ldots,+q_T}, \tag{59}$$

Where,

$$q_t := probability\ that\ scheme\ prepays\ at\ month\ t\ (1 \le t \le T). \tag{60}$$

By assumption, $q_1 + \ldots + q_T = 1$.

In general, the SMM depends on the mortgage cash flow scheme. For example, the higher the mortgage interest (coupon) the higher the SMM . Note, $SMML_1 = q_1$, as we assume that the cash flow scheme has not been prepaid at the present time $t = 0$ (i.e., that $q_0 = SMM_0 = 0$). From the above definition, one easily obtains (say by induction) that

$$q_t = (1 - SMM_1)\ldots(1 - SMM_{t-1})\, SMM_t \tag{61}$$

and that the probability that there is no prepayment by time t is

$$q_{t+1} + \ldots + q_T = 1 - (q_1 + \ldots + q_t) = (1 - SMM_1)\ldots(1 - SMM_t). \tag{62}$$

3. (SMM) - Transformation of a mortgage cash flow scheme

Given a cash flow scheme (PP_t^C, IP_t^C, T) and a known prepayment rate SMM_t for the scheme, we can, for valuation purposes, assume that the scheme is a portfolio of T cash flow schemes (PP_t^m, IP_t^m, T), $m = 1,\ldots,T$ where

$$PP_t^m = IP_t^m = 0\ for\ t > m \tag{63}$$

$$IP_t^m = q_m\ IP_t^C\ \ for\ t \le m \tag{64}$$

$$PP_t^m = q_m PP_t^C\ \ for\ t < m \tag{65}$$

$$PP_m^m = q_m (PP_m^C + \ldots + PP_T^C) \tag{66}$$

Define a new cash flow scheme (PP_t^C, IP_t^C, T), where

$$PP_t := PP_t^1 + \ldots + PP_t^T = (q_t + \ldots + q_T)PP_t^C + q_t(PP_{t+1}^C + .. + PP_t^C) \tag{67}$$

$$= (q_t + \ldots + q_T)PP_{t+1}^C + q_t B_t^C \tag{68}$$

$$IP_t := IP_t^1 + IP_t^T = (q_t + \ldots + q_T)IP_t^C\ \ . \tag{69}$$

We note that the associated outstanding principal and coupon rates are given by:

$$B_t := PP_{t+1} + \ldots + PP_T = (q_{t+1} + \ldots + q_T)B_t^C \tag{70}$$

$$I_t := \frac{IP_t}{B_{t-1}} = I_t \tag{71}$$

Indeed, the second follows from the first, and to verify the first, note the following consistency check:

$$B_{t-1} - B_t = (q_t + \ldots + q_T)(B_t^C + PP_t^C) - (q_{t+1} + \ldots + q_T)B_t^C \tag{72}$$

$$= q_t(B_t^C + PP_t^C) + (q_{t+1} + \ldots + q_T)PP_t^C = PP_t \tag{73}$$

In terms of SMM_t, we can also write these various equations as

$$B_t = (1 - SMM_1)\ldots(1 - SMM_t)B_t^C \tag{74}$$

$$IP_t = (1 - SMM_1)\ldots(1 - SMM_t)IP_t^C \tag{75}$$

Appendix 5.B - NIBC CPR Assumption Vs Predicted CPR for Bucket II, III, and IV

Figure 58 - NIBC CPR Assumption Vs Predicted CPR for Bucket II

Figure 59 - NIBC CPR Assumption Vs Predicted CPR for Bucket III

Figure 60 - NIBC CPR Assumption Vs Predicted CPR for Bucket IV

Chapter 6: Conclusions and Recommendations

The development of the Dutch mortgage market has been of great importance for the Dutch economy, historically and more than ever in the past two decades. The mortgage debt market represents a very high percentage of GDP, which further intensifies the importance of mortgages in the Netherlands. The prepayment behaviour of the mortgagor is a crucial variable in the valuation of mortgages. When managing a mortgage portfolio, knowing how prepayment will change as economic conditions change can enable mortgagees to value mortgages in a more accurate way, thus, permitting them to hedge their risk correctly.

Much research has been dedicated to this topic. In prepayment literature, a wide range of prepayment models can be found. On the one hand, there are models that work under the assumption that prepayment is always exercised in an optimal way. On the other hand, there are models that assume an exogenous prepayment rule. For a bank, determining an optimal prepayment strategy is quite a difficult task, mainly due to transaction costs. Moreover, mortgagors do not behave optimally with respect to their prepayment strategies. Thus, empirical exogenous models are preferred as when looking at the observed prepayments, these models consider a set of explanatory variables that justify the mortgagor behaviour towards prepayment. The academic literature on exogenous models entails two popular ways for modelling prepayment, namely the Cox Proportional Hazard Models and the Binary Choice Models. In this Thesis, an extension of the Logit model (a binary choice model) was used to model prepayment. Therefore, a thorough description of the applied Multinomial Logit Model can be found in Chapter 2. The main reason for choosing this model is that it allows banks to take into account competitive risks, i.e. it permits banks to model the reason for prepayment due to swapping houses (move) or refinance on an individual loan level. This becomes even more essential, when we consider that prepayment penalties can only be charged when the mortgagor refinances.

In Chapter 2 it was shown that absolute changes in mortgages rates were not correlated anymore to the changes in long rates, if we examine the recent past. Therefore, a new way of classifying the refinance incentive of the mortgagors was proposed. Refinance incentive was defined as the difference between the blended market mortgage rate at origination of the mortgage and the prevailing blended market mortgage rate. This new definition permits NIBC to have a benchmark throughout time that can be compared across different types of mortgages. The chapter continued with listing all explanatory variables of the model together with explaining the way, in which the parameter estimates produced by the Multinomial Logit model can be used for predicting prepayment rates.

Chapter 3 described in detail the data-set, used as input in the Multinomial Logit Model. After that, the empirical study was produced. On an aggregate level, the observed prepayment rates are highly volatile. This volatility is mainly due to prepayment caused by refinancing, as prepayment rates derived from moving are quite stable (around 6%). Prepayment rates exhibit peaks in July and December. These strong seasonality effects can be explained by holiday and fiscal reasons. It was also found that the heterogeneity between mortgages regarding the principal value is not sufficient to have a considerable impact. . i.e. prepayment rates do not vary across mortgages with different principals. Moreover, the prepayment rate due to refinance is highly sensitive to the refinance incentive.

The estimate results resulting from the model were derived and presented in Chapter 3. The mortgage type variable was not considered significant. One of the main improvements, when comparing to Sterk (2005), was that both estimates for the refinance incentive are positive and significant at 1% level. A higher value of the refinance incentive increases the probability of termination of a mortgage more due to refinance than due to move. For comparison, Sterk (2005) found a negative relation for the refinance incentive estimate when studying the termination of a mortgage due to moving. Our results seem to be closer to reality, since it is feasible to acknowledge that mortgagors would not move less if the refinance incentive gets higher but the other way around.

Chapter 4 presented three main sources of prediction error that should be considered when applying the prepayment model. The first one is related to the observed prepayment rate being a random variable itself. It was proven that when the number of mortgages in the data-set is large enough, the variance of the observed prepayment rate for moving and refinance converges to zero. Since the used data-set contained more than 8 million observations, this source of error was considered irrelevant. The second source of error is induced by the parameter estimates and the back-testing was performed to address this error. The third source of error addresses the issue of omitting relevant variables. As with any empirical model, the prediction power of the model will always fall below 100%, as some explanatory variables can not be observed or included.

Out-of-sample forecast performance is usually seen as the acid test of an econometric model as a good out-of-sample performance strongly supports the model underlying theory. Usually, this is achieved by making a comparison to rival forecasts or to in-sample performance. Since it is not possible to access data from a different mortgage portfolio, a comparison between the out-of-sample forecast performance was made relative to the in-sample performance. As the observations can be sorted by date, it was possible to implement ex-ante and ex-post approaches to access the out-of-sample prediction performance of the Multinomial Logit Model. Overall, the model manages to predict prepayment rates quite accurately and it captures seasonality patterns of prepayment rates, a particularly difficult pattern to model. The main conclusions of the back testing are that the model's predictive power is satisfactory for the three proposed frameworks and that the inclusion of more data in the model improves its predictive power.

In Chapter 5, three different ways for calculating mortgage cash flows were presented. In the first one no prepayment is allowed, the second allows for prepayment using the multinomial Logit model jointly with a term structure model to predict prepayment paths and the last one concern the NIBC mortgage cash flow calculation model in place, where mortgages are assumed to prepay at a constant rate per bucket. This method creates less volatility in the valuation and in the hedging turnover of the mortgage portfolio. It is believed that applying the second method for calculating cash flows is not optimal because it is rather overoptimistic to expect that our prepayment model, jointly with a term structure model, will be able to accurately predict prepayment paths for every mortgage for the next thirty years. Nevertheless, better results are expected if the prepayment model to forecast prepayment paths for the next year is used and after it the current method (using a constant CPR) is applied for calculating cash flows from year one until the next interest reset date. This should be true because the prepayment forecasting model should accurately predict prepayment rates in the short-term. Therefore, this approach is recommended for further research projects within NIBC.

NIBC CCPR assumptions were compared to the predictions, produced by the Multinomial Logit model. The model predictions outperform NIBC current assumptions in every bucket, indicating that at least NIBC assumption for the first bucket should be changed in the near future since it is abnormally far away from the observed mean CPR for bucket I for the considered period.

In the end, a mark-to-market model was proposed to calculate the P&L impact due to the difference between prepayment assumptions and realizations in a given month. This model was used under NIBC assumptions and under assumptions produced by the Multinomial Logit model. The results showed that the use of the prepayment model for producing CCPR assumptions reduce the MtM losses on mortgages due to prepayment. However, it was proven that an expected CPR closer to the observed CPR does not implicitly imply a decrease on P&L volatility. The portfolio of NIBC per bucket is not homogeneous. Therefore, it is recommended to further research the possibility to bucket mortgages by refinance incentive, as this should permit to have the portfolio divided in a more homogenous manner (in terms of sub-par and above-par value). Hence, if the portfolio is homogeneous, the bank can be confident that an expected CPR closer to the observed one will imply a decrease in the P&L volatility due to prepayment risk.

After all data bases are align, the bank will be able to use this research to come up with a prepayment path in the future for every observation (mortgage). Hence, in order to optimize the use of the prepayment model, the following scenarios should be studied:

- Calculate the historical MtM losses, when using the forecasted prepayment path of each observation for the next 30 years on the valuation of the portfolio. (No bucketing)

- Calculate the historical MtM losses, when using the forecasted prepayment paths in the short term and current NIBC CCPR assumptions in the long term. (Existing bucketing)

- Calculate the historical MtM losses, when using new CCPR with bucketing by refinance incentive. One should create three buckets, namely a bucket where the refinance incentive is high, another where it is low and a third one where it is neutral. (More buckets are not advisable as due to mortgage rate shifts, one bucket could contain few observations, making it difficult to assess the realized CPR).

- Calculate the historical MtM losses, when using the forecasted prepayment path of each observation in the short run and the new CCPR assumptions for the refinance incentive buckets in the long run.

- Calculate the historical MtM losses, when using the forecasted prepayment path of each observation in the short run and an average of the forecasted prepayment path of each observation as a CCPR assumption in the long run.

Every scenario should be compared and the homogeneity of the portfolio per bucket should be tested, when bucketing is used. Hence, if any CCPR assumption is used per bucked, the portfolio should be homogeneous in every bucket.

If MtM losses volatility can be decreased from one of the above scenarios, this will imply a positive impact on the P&L. After all, the money from the prepayment penalties and prepayment spread will still be received.

References

Abrahams, S.W., 1997, *"The New View in Mortgage Prepayments: Insight from Analysis at the Loan-by-Loan level"*, The Journal of Fixed Income, 7:1, 8-21.

Alink, B.J., 2000. "Mortgage Prepayments in the Netherlands", Enschede University, PhD Thesis.

Allan H. Murphy, 1998, *"Skill Scores Based on the Mean Square Error and Their Relationships to the Correlation Coefficient"*, Oregon State University.

Archer, W.R. and D.C. Ling, 1993, *"Pricing Mortgage-Backed Securities: Integrating Optimal Call and Empirical Models of Prepayment,* Journal of the American Real Estate and Urban Economics Association, 21:4, 373-404.

Archer, W.R., D.C. Ling, and A.G. McGill, 2003, *"Household Income, Termination Risk and Mortgage Pricing"*, Journal of Real Estate Finance and Economics, 27:1, 111-138.

Boris I. Ederov, 2009, *"Mortgage Prepayments and Economic Capital Estimation"*, University Amsterdam, Master Thesis.

Brennan, M.J. and E.S. Schwartz, 1985, *"Determinants of GNMA Mortgage Prices,* Journal of the American Real Estate and Urban Economics Association, 13:3, 209-228.

Bussel, A. van, 1998, *"Valuation and Interest Rate Risk of Mortgages in the Netherlands"*, Maastricht University, PhD Thesis.

Bussel, A. van, and E. Charlier, 2001, *"Prepayment Behaviour of Dutch Mortgagors: an Empirical Analysis"*, CentER discussion paper, Tilburg University, 2001:64.

Cox, J.C., J.E. Ingersoll, and S.A. Ross, 1985a, *"A Theory of the Term Structure of Interest Rates,* Econometrica, 53:2, 385-407.

Cox, J.C., J.E. Ingersoll, and S.A. Ross, 1985a, *"An Intertemporal General Equilibrium Model of Asset Prices,* Econometrica, 53:2, 385-407.

Cramer, J.S., 2003, *"Logit Models from Economics and Other Fields"*, Cambridge University Press.

Dunn, K.B, and J.J. McConnell, 1981a, *"A Comparison of Alternative Models for Pricing GNMA Mortgage-back Securities,* The Journal of Finance, 36:2, 471-487.

Dunn, K.B, and J.J. McConnell, 1981b, *"Valuation of GNMA Mortgage-Backed Securities,* Journal of Finance, 36:3, 599-616.

Frank J. Fabozzi, 2006, *"The Handboof of Mortgage-Backed Securities"*, 6th Edition, McGraw-Hill, New York.

Gilberto, S.M., and D.C.Ling, 1992, *"An Empirical investigation of the Contingent Claims Approach to Pricing Residential Mortgage Debt",* Journal of the American Real Estate and Urban Economics Association, 20:3, 393-426.

Goncharov Y., 2005, *"An internal-based Approach to the Valuation of Mortgage Contracts and Computation of the Endogeneous Mortgage Rate"*, International Journal of Theoretical and Applied Finance, vol.09, 889-914.

Green, J. and J.B. Shoven, 1986, *"The Effects of interest rates on Mortgage Prepayments",* Journal of Money Credit and Banking, 18:1, 41-59.

Hendeshott, P.H. and R. van Order, 1987, *"Pricing Mortgages: An Interpretation of the Models and Results"*, Journal of Finance, 41:5, 1011-1029.

Hsu, A., and R.T. Wilcox, 2000, *"Stochastic Prediction in Multinomial Logit Models"*, Management Science, 8, 1137-1144.

Hull, J., 2008, "Options, Futures, and Other Derivatives", 6th Edition, Prentice Hall International Editions, Englewood Cliffs, New Jersey.

Hull, J., and A. White, 1990, *"Pricing Interest-rate Derivative Securities"*, The Review of Financial Studies, 3:4, 573-592.

Kang and Zenios, 1992,*"Predicting Prepayment Rates for Mortgages Using the Cascade-Correlation Learning Algorithm"*, the Journal of Fixed Income, 2:4, 86-96

Kau J., Keenan D., and Smurov A., 2004, *"Reduced-form Mortgage Valuation"*, Journal of Housing Research, 6, 217-244.

Kau, J.B. and C.V Slawson, *"Frictions, Heterogeneity and Optimality in Mortgage Modeling"*, Journal of Real Estate Finance and Economics, 24:3, 239-260.

Kau, J.B. and D.C. Keenan, 1995, *"An overview of Option-theoretic Pricing of Mortgage Modeling"*, Journal of Real Estate Finance and Economics, 24:3, 239-265.

Kau, J.B., D.C.Keenan, W.J. Muller, and J.F. Epperson, 1992, *"A Generalized Valuation Model for Fixed-Rate Residential Mortgages"*, Journal of Money Credit and Banking, 24:3, 279-299.

Kevin Dowd, Andrew J.G. Cairns, David Blake, Guy D. Coughlan, David Epstein, and Marwa Khalaf-Allah, 2008, *"Back testing Stochastic Mortality Models: An Ex-Post Evaluation of Multi-Period-Ahead Density Forecasts"*, Pensions Institute, Discussion Paper PI-0803.

McConnel, J.J and M. Singh, 1994, *"Rational Prepayment and the Valuation of Collaterized Mortgage Obligations,* The Journal of Finance, 49:3, 891-921.

Michael P. Clements and David F. Hendry, 2003, *"Evaluating a Model by Forecast Performance"*, Journal of Economic Literature, 41, 788-829.

Richard, S.F and Roll, R., 1989, *"Prepayments on Fixed-rate Mortgage-Backed Securities"*, Journal of Portfolio Management, 73-82.

Sharp N., 2006, *"Advances in Mortgage Valuation: An Option-Theoretic Approach"*, University Manchester, Master Thesis.

Staton R., 1995, *"Rational Prepayment and the Valuation of Mortgage Backed Securities"*, Review of Financial Studies 8, 677-708.

Vincent Sterk, 2005, *"Prepayment Risk Modeling and the Valuation of Dutch Mortgages"*, Tilburg University, Master Thesis.