MODELLING PREPAYMENT RISK

Jan Jacobs
University of Groningen

Ruud H. Koning
University of Groningen

Elmer Sterken
University of Groningen

MODELLING PREPAYMENT RISK

J.P.A.M. Jacobs, R.H. Koning, E. Sterken
Dept. Economics
University of Groningen
The Netherlands

2nd August 2005

ABSTRACT

One of the most important financial decisions a household makes is the purchase of a home. In most European societies, households finance this purchase through a mortgage. The household promises to make regular interest payments, and repayment of the principal when the mortgage matures. Often, life insurance policies are attached to such mortgages so that the mortgage is repaid in case the mortgagor dies before the end of the contract. An important characteristic of such mortgage loans is the implicit option available to mortgagors: they can prepay their loan under certain conditions. This risk of prepayment makes the duration of a portfolio of mortgages stochastic which has implications for the re-finance policy by the mortgagee. We give an overview of recent literature on prepayment, and on empirical approaches to modelling the prepayment risk. Different models are explained using a data set of about ten thousand mortgage loan contracts covering the years 1998 to 2003 from the Netherlands. We find that prepayment behavior depends on the type of mortgage product, size of the loan, and the age of the mortgagor at the time the contract is signed.
KEYWORDS: prepayment, Cox proportional hazard, refinance incentive, mortgage.

1 INTRODUCTION

A mortgage loan is the largest loan contract for almost any individual. It is also an example of an incomplete, private contract. The lender (bank, insurance company, etc.), known as the mortgagee, and the borrower, known as the mortgagor, agree on various characteristics of the loan. Examples of such characteristics are the size of the loan, repayment schedule, life insurance policy attached to the loan, interest rate, collateral, and so on. The role of collateral is special in a mortgage loan, and distinguishes it to a large extend from other loans or private placements. Also, the tax treatment of mortgages differ from the treatment of other loans, at least in The Netherlands.

A mortgage loan is a private loan because only two (private) agents are involved in deciding upon the contract (although the interest rate may depend on a public institution that guarantees mortgage loans). The fact that only two parties are involved makes renegotiation of the loan simple and cheap.

A mortgage is also an incomplete contract. The contract does not contain clauses for all possible states of nature. A sudden change in (economic) conditions can change the incentives faced by both the mortgagee and the mortgagor.

One important unforeseen action by the mortgagor is prepayment. All mortgage contracts have a repayment scheme, but a change in, say, economic conditions may give the mortgagor incentives to deviate from that repayment schedule. An extreme deviation is full prepayment of the mortgage loan. Various arguments can be given to rationalize such a decision, these will be reviewed in section 2. Only the mortgagor has the option to prepay the loan, the mortgagee does not have such an option. It is known that the decision to exercise an option is not always taken rationally. One would expect that the option to prepay is exercised if this option is in the money, i.e., if the value of the option exceeds some (possibly contract specific) threshold. However, this is not always the case which make sit more difficult to model and predict prepayment behavior. It seems useful to model prepayment as the outcome of a random variable, an approach we pursue in this paper.

Usually, the mortgagee is not interested in prepayment of a single mortgage. He is, however, interested in prepayment on a portfolio level. A portfolio of mortgage loans has to be financed, for example through securitization or through borrowing money on the capital market. Even if the duration of the portfolio and the financing are matched perfectly, prepayment poses a risk to the mortgagee. Suppose a 30 years mortgage against 5% is financed through a 30 years bond paying 4.5%. If the interest rate on mortgages drops to 4%, the mortgagor may refinance his mortgage: he prepays the existing mortgage and takes on a mortgage for the remainder of the term against 4%. The mortgagee is left with the obligation to pay 4.5% interest on the bond used to finance the mortgage, and an amount of cash that will yield a lower return in these market conditions. In this example, the decision to prepay is driven by interest changes. In practice, there may be other determinants of prepayment as well (for example, marital situation, age of the mortgagor, moving, etc.). Clearly, a better understanding of prepayment behavior can be profitable knowledge for the mortgagee.

In this paper we focus on prepayment behavior in a mortgage portfolio of a Dutch bank. We discuss the most salient features of the Dutch market in section 2, together with a review of earlier findings. Section 3 covers our empirical approach to modelling prepayment behavior. It focuses on duration models as the tool of analysis. Empirical results are given in section 4, and section 5 concludes.

2 Review of Literature

Almost 7 million houses are occupied in the Netherlands. About 53 per cent of these house are owner occupied. Most Dutch house owners finance their house occupation by a mortgage loan. So the Dutch market for mortgage loans is of considerable size: about 389 billion euro mortgage loans is on the balance sheets of financial institutions. At the end of 2003 private banks owned about 311 billion euro, while institutional investors had about 36 billion euro mortgage loans on their balance sheets. Although 2003 was an extreme year (in a positive sense), it is insightful to acknowledge that 540 thousand new mortgage contracts were written, 290 thousand of those being renewals. It is estimated that more than 60 percent of the new contracts is originated by so-called intermediaries. But the final owner of the mortgage loan, as explained above, is in most of the cases a large financial institution. The concentration of the private bank mortgage market in terms of the market share of the four largest banks is about 80 per cent.

A mortgage loan contract is a typical incomplete financial contract: there is not a full set of contingent conditions in the contract that foresees all future actions. This implies that both lender (*initiator*) and client (*mortgagor*) need to think about handling unforeseen outcomes. Economic agents will use basic economic principles to take decisions in these unforeseen states. This allows us to model the likelihood of for instance two important contingent decisions: prepayment and default. We will focus on the former, and especially on the economics of prepayment decisions: if the value of future payments is higher than the loan balance plus refinancing costs the probability of prepayment will increase. In the literature this contingent action is seen as the execution of a call option embedded in the contract. It is well-known that some agents exercise this option, while it is not at all in the money, and some agents prepay while there are no economic incentives to do so. Apart from economic-financial arguments, there might be other factors driving prepayment decisions. Here one can think of migration due to job improvement or rotation. To the initiator it is valuable to estimate the value of the prepayment option and to forecast actual prepayment behavior in order to price the mortgage loans more competitively. We do not describe the deep underlying theoretical thoughts of the option model, these can be found elsewhere (see e.g. Schwartz and Torous, 1992), but focus on the main intuition.

In order to describe prepayment behavior in practice two basic approaches have been followed. First, one can use so-called aggregated data and use the law of the large numbers to forecast prepayment over a portfolio of mortgage loans. Examples of these studies are Brennan and Schwartz (1985), Kang and Zenios (1992) and Golub and Pohlman (1994). The main disadvantage of this class of models is that one cannot forecast prepayment behavior if the main characteristics of the population of the mortgagors changes. Therefore the second class of empirical studies of the prepayment option, the class that uses individual loan data, is by far more interesting. Examples of the second class of models are Archer, Ling, and McGill (1996) and Green and Shoven (1986). As we will construct an individual loan data model, the approach of the second class is more appealing. The theoretical prepayment motives though are identical across the two approaches. Hayre (2003) discusses pros and cons of the use of pool-level and individual loan data in more detail.

2.1 Mortgage Products in The Netherlands

The Dutch tax code is the main determinant of mortgage products that are available. Changes in the tax code has had profound implications for the popularity of different types of mortgages. Basically, interest payments on a mortgage loan are deductible from income tax. Considering the fact that the marginal rate of income taxation varies between 33.55% and 52% (June 2005), the net cost of interest payments on additional funds are between 0.48 and 0.66. Over time, certain restrictions have been imposed on this deductibility. Interest payments are tax deductible over a maximum period of 30 years, and only if the mortgage concerns the primary home.

Tax considerations also come into play in the redemption schedule. The mortgagor has to accumulate capital to redeem the loan at maturity. These savings could be taxable as property. The property (usually assets as cash, bonds, stocks, etc.) tax rate is 1.2% per year on the amount exceeding €19522. However, savings to redeem a mortgage loan can be in the form of an endowment insurance premium. The proceeds of this endowment insurance and the value at any moment in time are tax exempt under certain conditions.

The most important mortgage types are the following.

- Linear mortgage.
- Annuity mortgage.
- Savings mortgage. This type consists of a redemption free mortgage with an endowment insurance that is priced to guarantee enough capital to redeem the principal at maturity. Because no intermediate prepayments take place, the mortgagor enjoys the tax advantage discussed above fully throughout the term of the mortgage.
- Redemption free mortgage. This is also known as an interest-only mortgage. No capital accumulation takes place during the term of the mortgage. Usually the amount of the redemption free mortgage is small when compared to the value of the house.
- Investment mortgage. This is similar to a savings mortgage, except that capital is accumulated through investments (usually in a limited number of mutual funds offered by the mortgagee). There is no guarantee that capital at maturity suffices to redeem the principal.

More often than not, houses are financed by a combination of some of the products above.

A final important characteristic of the Dutch market is the choice of fixed interest rate period. Mortgagors can set the interest rate for a period ranging from one to 30 years. Usually, interest rates fixed for shorter periods of time are lower than interest rates fixed for a longer period. This variation is mainly determined by the yield curve on the capital market.

2.2 Key to the literature

In this review of the literature on prepayment modelling of mortgage loans we focus on the following items:

- What are the main determinants used? This comes back to the mortgage prepayment option model: what are the theoretical drivers of exercising the option?
- The modelling strategy: what kind of econometric or statistical model is used? Or in other words: how is theory implemented in describing the data?
- What are the main findings for the Dutch mortgage market up to now? Since we focus our study on the Dutch case it is good to have insight into the main other findings for Dutch data.

We do not give a full explicit institutional description of the Dutch mortgage market as well: this can be found in e.g. Hayre (2003) and other sources. Despite that we highlight some specific characteristics that are of interest to prepayment behavior:

- Dutch mortgage loans have coupons that are fixed during an extended period of time, which is mainly five or ten years. Mortgagors with a shorter horizon will opt for a shorter coupon period. At the reset date a new interest rate is set;
- Prepayments are free up to 10 percent per year of the original balance (in some cases this percentage is higher to 15 or 20 per cent), at the coupon reset date, in case the actual market interest rate exceeds the contractual interest rate, or any sudden or unexpected change in the personal settings of the mortgagor (like death). Also, the mortgage can be prepaid without additional charges if the mortgagor moves and sells the house. In other cases prepayments are discouraged by a *penalty*, mostly equal to the present value of the difference in monthly interest payments between the loan and an alternative until the next reset date.
- In case of selling the old house and buying a new one, a mortgage can be transferred (*portability*).
- Differences in mortgage loan types. Due to the Dutch tax law, deductibility of interest payments dominate the design of products. Amortizing loans are old-fashioned and replaced by interest-only mortgages, savings products, investment mortgages and insurance products. It might be so that certain types of mortgagors opt for more risky (say interest-only) products than others. This risk attitude might also have an impact on prepayment behavior.

2.3 The key determinants of prepayment

Prepayments are, without making a distinction with respect to any institution, caused by the following four classes of determinants (Hayre, 2003):

1. Refinancing components. These are mostly seen as the key determinants of prepayment and depend on tax considerations, the so-called burnout effect (that is that in a pool of mortgages, the slower adjusting contracts will stay longer in the pool), the media effect (reading lower mortgage rates in the news papers), prepayment penalties, etc.;
2. Housing turnover (about 4 percent in The Netherlands). This depends on the overall mobility rate of the labor force, the so-called loan-type effect (clients with shorter horizons typically buy shorter contracts), seasonal factors (the month of the year), the so-called seasoning curve (that is an S-shaped relation between the prepayment rate and the age of the loan), and a coupon effect (lock-in effects);
3. Defaults. Defaults depend on demographics and the phase of the business cycle. Defaults tend to increase with loan age, peaking between two and six years and then decline;
4. Curtailments (early partial prepayments) and full payoffs.

We discuss three elements in more detail: (1) refinancing, (2) seasoning, and (3) the coupon effect.

The refinancing incentive is the price mechanism of the prepayment decision. A sudden drop in the market rates may trigger the option to prepay, even though the penalty might neutralize the refinancing incentive. It is widely observed that the refinancing motive is not always handled rationally. Assuming that mortgagors are heterogeneous in gathering information and processing it in an appropriate way, it is likely that some clients will refinance quickly, leaving others in the pool of mortgagors. So for a constant refinancing incentive there will be so-called *burnout* of the pool. Refinancing may also be triggered by a so-called *media effect*. Hayre (2003) illustrates that for perceived all-time low values of the capital market interest rates refinancing peaks to irrational levels. Alink (2002) discusses the role of tax effects and argues that tax effects are minor in Dutch refinancing decisions.

Seasoning refers to the probability of a house sale as the age of the mortgage loan increases. The rate of seasoning also depends on the conditions on the housing market. It is likely that in a boom, more job changes will take place, and more houses will be sold. It will be likely that in those cases house prices will be high.

The coupon effect is simple: a low coupon will imply a lower speed of prepayment, since it is likely for instance that a home-owner will carry the mortgage loan to a new house in case of selling the old house.

Summarizing we can pin down the main micro and macro drivers of prepayment:

- refinance incentives: especially sudden changes in the capital market interest rate will trigger the prepayment option. In general a split between the coupon and market rate will be the key variable. This is a micro variable that we can use at the loan contract level;
- the business cycle: in a a booming housing market the prepayment rate will increase. On the other hand in a downturn we will observe more defaults, triggering prepayments. This is typical macro control variable;
- house sales: these are depending on personal income motives, seasoning effects, and maybe a coupon effect. This is a typical micro variable;
- the loan-type effect: mortgagors will self select into certain products. This is a typical micro argument.

Next we will describe how to model the prepayment decisions.

2.4 The Dutch case

There are not so many studies for the Netherlands. For a current review of the Dutch mortgage market see Rabobank (2003). The main studies focusing on prepayment are Van Bussel (1998), Charlier (2001), Alink (2002), Charlier and Van Bussel (2003), and Hayre (2003). We review these studies here in brief and focus on the key elements:

1. What kind of model is used?
2. What are the key determinants of mortgage prepayments?

3. What are the main findings?

2.4.1 Methods used

Van Bussel (1998) is the first empirical study for the Netherlands on mortgage prepayments. Van Bussel applies the methodology introduced by Kang and Zenios (1992) on their Wharton prepayment model. So it is an aggregate study that ignores heterogeneity and tries to predict prepayment at the mortgage pool level. Van Bussel's sample is much smaller than the sample used by Kang and Zenios and should therefore not be considered to be representative for the Netherlands. Charlier (2001) is not a full-fledged model of the prepayment decision but concentrates on the specification of the refinancing incentive for different loan types. Charlier studies three types of mortgage loans, annuities, savings products and redemption free mortgages, and finds evidence of different relations between refinancing incentives and refinancing rates. Alink (2002) estimates a logit model, using data provided by the Dutch SNS bank for the sample 1993-2001. Charlier and Van Bussel (2003) estimate a proportional hazard model for the Dutch mortgage market. Their econometric model resembles those of Green and Shoven (1986) and Schwartz and Torous (1992). The unit of observation is the loan level. The data are provided by the largest pension fund ABP and covers the sample 1989-1999. They consider two loan types: savings products and redemption free mortgage loans. Finally, Hayre (2003) uses partial correlations in a more descriptive study. At the end Hayre uses the projections from his four components, home sales, refinancing, defaults, and curtailments, in a total model. The precise details of the model remain unclear though. Hayre uses this model to forecast prepayment rates for Dutch Mortgage-Backed-Security portfolio's.

2.4.2 Determinants used to explain prepayment

Now we presented the general classes of determinants and their interpretation we can be brief about the description of the determinants used in empirical Dutch studies of mortgage prepayment. We summarize:

1. Van Bussel (1998): Seasonality, refinancing incentive, seasoning, and burnout;
2. Alink (2002): Loan to foreclosure value ratio, the distribution channel, mortgage rank, age of the client, property type, urbanization, geographic region, interest type, mortgage type, seasoning, refinancing incentives, momentum of interest rates, level of interest rates and steepness of the yield curve;
3. Charlier and Van Bussel (2003): Seasoning, refinancing incentives, burnout, seasonality, mortgage and property type;
4. Hayre (2003): See the classes mentioned above.

2.5 Main findings

What are the main findings? Alink concludes that his model tends to overestimate prepayment rates, when used in out-of-sample exercises. He also concluded that different mortgage loan types show different prepayment patterns. Savings mortgages for instance prepay slower. Charlier and Van Bussel (2003) conclude that the likelihood that a savings mortgage loan will be prepaid increases in the age of the loan. Models that exclude burnout show that the refinancing motive is significant. Prepayments in December are higher due to holidays and tax effects (prepayments in January and February are lower). For savings products they find that so-called upgrading from an apartment to a house is important. This effect also holds for interest only loans, but to a lesser extent. Finally, Hayre (2003) illustrates that the refinancing effect and seasonal house sales, among others, are relevant in simple partial correlations. He adds these findings in a linear form in total prepayment forecasting model. These results should be considered with care though. Other, mortgagor-specific determinants might be more important and collinear with seasonal and refinancing variables.

3 Empirical Approach to Modelling Prepayment

3.1 Modelling strategy

Two types of models to explain prepayment are used most frequently: We can distinguish so-called duration (or survival) models and binary choice (logistic) models. In the class of duration models we observe so-called semi-parametric and parametric specifications. The main difference is in the modelling of time. In our empirical part we use only survival models to model prepayment for reasons that will be explained at the end of this section. We start this section with a somewhat more technical discussion of survival models, followed by a short discussion of binary choice models.

3.2 Duration analysis: the proportional hazard approach

Without exception the proportional hazard approach has been the most popular tool in estimating prepayment risk of mortgage loans. The class of proportional hazard models is a special case of the so-called Mixed Proportional Hazard Model, the most popular tool in duration analysis. The mixed proportional hazard model is a reduced-form model: there are no attempts to estimate the deep structural parameters of the utility functions used by agents. Early references with respect to mortgage loan prepayments are Dunn and McConnell (1981), Brennan and Schwartz (1985), and Green and Shoven (1986). We introduce notation and recall the main line of modelling below.

The key variable in the model is the hazard rate λ_{it}: the probability that mortgage loan i will be prepaid at date t, given the fact that it has not been prepaid at date $t-1$ or before. This quantity is also known as the single month prepayment rate. Usually, such a hazard rate is modelled as:

$$\Pr(T = t | T \geq t-1) = \lambda_{it} = \lambda_0(t) P(x_{it}; \beta) \tag{1}$$

Here it is good to get a little intuition. The model expresses the hazard rate as a function of the so-called baseline hazard, which is not specific to each individual loan i, but evolves over time, and a proportionality factor. The baseline hazard reflects a natural prepayment rate, that varies over time (that is, with the age of the mortgage) and reflects the seasoning effect discussed earlier. The second term expresses the impact of loan- and time specific developments x_{it}. It is assumed in this model that the impact of a variable x_{it} on the hazard rate is constant through time (via β), although the value of the covariate may change over time. Contract heterogeneity is captured by the variables x_{it}. As such, model (1) is not statistically identified. We need to impose restriction such that estimation and unambiguous interpretation of the parameters becomes possible.

Charlier and van Bussel (2003) use the following specification for the hazard rate (see also Green and Shoven, 1986):

$$\lambda_{it} = \lambda_0(t, \mu_1, \mu_2) P(x_{it}; \beta) \tag{2}$$

where $\lambda_0(\cdot)$ is again the baseline hazard, which depends on the age of the loan t, and two parameters μ_1 and μ_2. Note that t (and the random duration of contracts T) is process time, not calendar time. μ_1 represents the baseline hazard for a mortgage loan that has just been originated, while μ_2 indicates how the baseline-hazard increases with the age of the mortgage loan. $P(\cdot)$ is the proportionality factor and x_{it} is again a set of determinants of $P(\cdot)$ (with parameters β).

The baseline hazard contains the seasoning effect. Normally this is an S-shaped relation between the prepayment rate and the mortgage age. One can use the popular exponential form:

$$\lambda_0(A_i, \mu_1, \mu_2) = \frac{1}{(1 + \exp(-\mu_1 - \mu_2 t))} \tag{3}$$

or use other specifications like the lognormal or the log-logistic functions (see Alink, 2002, p. 90). Again, this is no more than modelling the basic hazard rate using a simple functional form. There is no attention for loan-specific determinants. The function $P(\cdot)$ normally contains the other loan-specific factors explaining prepayment. One could use the functional form:

$$P(x_{it}, \beta) = \exp(-\exp(-\beta' x_{it})) \tag{4}$$

which ensures that λ_{it} is always in the interval $[0, 1]$. In the set x_{it} one normally finds the determinants like refinancing incentives, seasonality, and burnout. Schwartz and Torous (1992) use variations on the proportional-hazards approach together with a Poisson regression to integrate prepayment into an overall valuation framework. Another line of approach is to use nonparametric regression techniques (see e.g. Maxam and Lacour-Little, 2001).

The general approach therefore is quite simple. First a general shape of the hazard rate function is chosen. Next, cross-sectional heterogeneity is assumed to describe loan-specific variations of the general hazard function. Here we are able to introduce interactions between multiple variables. A problem is that in estimation time-varying regressors are difficult to handle.

3.3 Logistic regression

An alternative to the proportional hazard model is a binary choice (or more specifically a logistic) regression model. In a binary choice model one describes the probability that either state 0 (say no change) or state 1 (a change: prepayment) will occur. As opposed to duration models, it is possible to examine the implications of utility maximizing behavior in discrete choice models, see for example Koning and Ridder (2000, 2003).

However, we are not aware of any empirical application of such ideas in the context of prepayment analysis.

In general a binary choice model describes the probability that a discrete variable $y_i = 1$, in our example the case that a mortgage loan will be prepaid. In general one models:

$$P(y_i = 1|x_i) = G(x_i, \beta) \tag{5}$$

The probability that $y_i = 1$ depends on the vector x_i, containing individual characteristics. The function $G(\cdot)$ is allowed to take values in the interval $[0, 1]$. Two classes of binary choice models are widely used. The first is the probit-class, which uses a standard normal distribution for the distribution function $G(.)$. The logit model uses the following form:

$$P(y_i = 1|x_i) = \frac{\exp(\beta' x_i)}{1 + \exp(\beta' x_i)} \tag{6}$$

The logit model is close to the proportional hazard model. We can show this as follows. One can rewrite

$$\lambda_{it} = \lambda_0(t) \exp(x_{it}, \beta) = \exp(\log(\lambda_0(t)) + \beta' x_{it}) \tag{7}$$

and use

$$z_{it} = \log(\lambda_0(t)) + \beta' x_{it}) \tag{8}$$

in the specification

$$G(t, x_{it}) = \frac{\exp(z_{it})}{1 + \exp(z_{it})} \tag{9}$$

which leads to the logit model. For small hazard rates the proportional hazard model and the logit model are comparable (see Alink, 2002, for a comparison of both specifications).

3.4 Modelling strategy

In the next section we discuss an empirical example, where we use a survival model to model prepayment. Our choice to model prepayment using survival models (as opposed to binary choice models) is based on a number of considerations. The most common (and, one might add, the easiest) approach to modelling prepayment on a micro level is through binary choice models. Given a collection of mortgages in a portfolio on a particular date, the model estimates the probability that a mortgage with certain characteristics will be prepaid in the next month, or in the next year. These probabilities can be added to have an estimate for the aggregate prepayment rate in the portfolio. There are two major problems with this approach. First, as such models analyze the probability of prepayment in a certain time interval, it is not straightforward to extend the results over longer time intervals. Knowing the probability that a mortgage will not be prepaid in the next year does not translate to knowledge about the probability that it will not be prepaid during the next five years in a simple way. Often, the planning horizon of the mortgagee and the unit of time in the binary choice model differ. The second problem appears if data are available from multiple years. How does one account for dependence over time that occurs due to the fact that the same contract will be observed more than once? Prepayment behavior of a contract cannot be assumed to be independent over time.

Survival analysis is a far more suitable analytical tool. The data requirements for these models are only slightly different from the requirements of binary choice models. Mortgages need to be followed over time, and the researcher needs to know when the mortgage originated. This type of information is usually available in the bookkeeping data of mortgagees. Survival models specify the probability distribution of the duration of a contract, T. The parameters of this probability distribution may depend on explanatory variables x_i. Instead of the probability distribution of T, $F(T)$, one usually focuses on the survival function which is the complement of $F(T)$:

$$S(t) = \Pr(T > t) = 1 - F(t).$$

Clearly, the survival function and the distribution function contain the same information. Once we have estimated the survival function, it is simple to answer questions like ‘what is the probability that this new contract will not be prepaid during the first fixed coupon period of 5 years’. The answer is $S(60; \hat{\beta}, x_i)$. Also, the problem of dependence of observations between years has been taken care off by using the contract as the unit of observation, and not a contract-year. The price of these advantages is that we have to alow for censoring. Of some observations we only know that the survive for at least a certain number of months. This censoring point is determined by the observation window, and will vary between contracts. The other problem is that it is difficult to include time varying covariates. We elaborate on this problem later.

Survival models are slightly more complicated than normal linear regression models, or logit models. Harrell (2001) lists three reasons why they are better suited to model survival data:

1. Time to prepayment can have an unusual distribution. Certainly the distribution is skewed because T is non-negative, so the normal distribution is not applicable.

2. The probability of survival of a contract, which is the focus of survival models, is of more interest than expected survival, which is the focus of regression models.

3. The hazard function (see below) is often informative in a very intuitive sense, and helps us to understand the mechanism of prepayment.

4 Empirical Results

We illustrate the modelling approach discussed in the previous section by an empirical application. We use tools from survival analysis to determine prepayment at a portfolio level, and deviations from that survival for individual contracts. Our approach is based on the mixed proportional hazard model by Cox. In that model, the hazard rate is specified as:

$$\Pr(T = t | T \geq t - 1) = \lambda_{0t} \exp(\beta x_{it}). \tag{10}$$

In this model, the baseline hazard is not specified as some smooth function, known up to a few a parameters (as discussed earlier). Instead, the functional form of the baseline hazard is not specified and the baseline hazard is estimated nonparametrically. The main advantage of this type of model as opposed to fully parametric models (as (1)) is that possible misspecification of the baseline hazard does not bias the estimates of the (proportional) effects of the variables in x_{it}. More background information on survival models and the Mmixed proportional hazard model can be found in Harrell (2001), Smith (2002), Therneau and Grambsch (2000), and Venables and Ripley (2002).

Our application is based on a data set provided by a Dutch financial institution. In order to keep the analysis tractable and the computational effort involved limited, we use a subset of these data: five thousand redemption free mortgages and five thousand savings mortgages. These two types are chosen because they are predominant in this portfolio: 65% of the total number of contracts are of either type, with the remaining 35% distributed over four other types.

The data were originally organized by year, not by contract. In fact, usually data are organized by contract-year because they are derived from book-keeping systems. The data cover the period 1 January 1998 to 31 December 2003, six years (72 months) in total. We know of each contract when it was initiated, and whether or not it is prepaid at a certain date. The majority of contracts is not prepaid at the date when our observation window ends ultimo 2003.

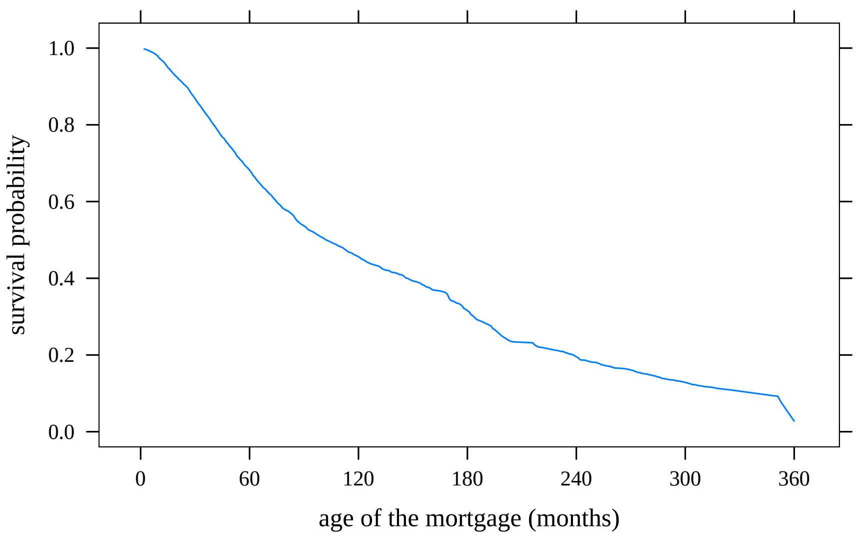
The following information is available:

- type of mortgage,
- date of initiation of the contract,
- status of the contract (prepaid or not),
- status of the contract at the end of the observation window (ie, an indicator whether the observed duration is censored or not),
- the interest rate paid during each year,
- age of the mortgagor at the moment of initiation,
- size of the principal,
- duration of the fixed interest period,
- socio-economic status indicator of the zip-code area in 1994, 1998, and 2002.

Other variables of interest as the value of the house at the initiation of the contract, income of the mortgagor, marital status and occupation of the mortgagor, etc. are unavailable, or missing for most cases. We assume that individual contract heterogeneity due to variation in these variables is captured appropriately by modelling prepayment as the outcome of a random variable.

We start by calculating the survival curve in our sample, and its associated hazard rate. Both are drawn in figure 1. The lower panel shows the hazard rate, and its shape is well interpretable: mortgages face a higher risk of prepayment after five, ten, of fifteen years of contract duration. These data often correspond to interest reset dates, when the mortgage can be prepaid fully without incurring any penalty.

Figure 2 shows the same data, but stratified by type of mortgage. Clearly, survival of the savings mortgage (as displayed by the black curve in the top panel) is markedly



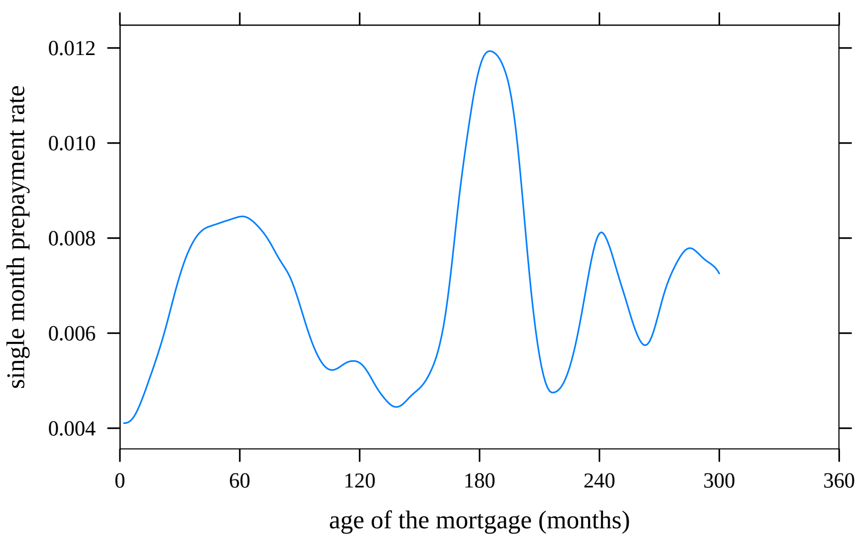


Figure 1: Estimated survival and hazard curve of all contracts in the sample.

Table 1: Median duration by type of mortgage.

	records	events	median	lower cl	upper cl
all contracts	10000	3073	103	98	109
spaarhypotheek	5000	1651	114	107	121
aflossingsvrije hypotheek	5000	1422	94	88	105

Table 2: Survival rates by type of mortgage.

age	all	saving	red.free
1	1.000	1.000	1.000
12	0.967	0.984	0.959
24	0.905	0.936	0.889
36	0.828	0.866	0.806
48	0.752	0.788	0.729
60	0.682	0.716	0.660
72	0.616	0.650	0.592
84	0.565	0.598	0.537
96	0.518	0.548	0.495
108	0.487	0.515	0.467

higher than survival of redemption free contracts. As savings mortgages were introduced around 1985, the survival curve of savings mortgages is not estimated after 225 months. The hazard rates in the lower panel of figure 2 show approximately the same pattern, but the hazard of redemption free mortgages is almost everywhere higher than the one of savings contracts.

Numerical data on survival confirm these qualitative observations, as listed in table 1. Median duration of contracts is 103 months, approximately 8.5 years. 3073 of all ten thousand contracts in our selected data set are prepaid during the observation window. It is also clear from the data in that table that median duration of redemption free mortgages is significantly shorter than median duration of savings mortgages. Survival rates of both types are given in table 2.

In figures 3 and 4 we examine whether survival varies by size of the loan and age of the mortgagor (at the initiation of the contract). Again, survival curves are drawn in the top panel, and hazard rates in the bottom panel. The classification used for these two continuous variables corresponds to the four quartiles of their distributions. From the figures, we see that mortgages with a higher principal have a higher survival rate, and that contracts initiated by younger mortgagors are more likely to be prepaid. The last finding is not surprising: younger people tend to move more frequently when the are starters on the labor market. Also, marital status and the presence of children determine housing needs. It would be interesting to examine whether these younger mortgagors who prepay their mortgage enter a new contract with the same or a different financial

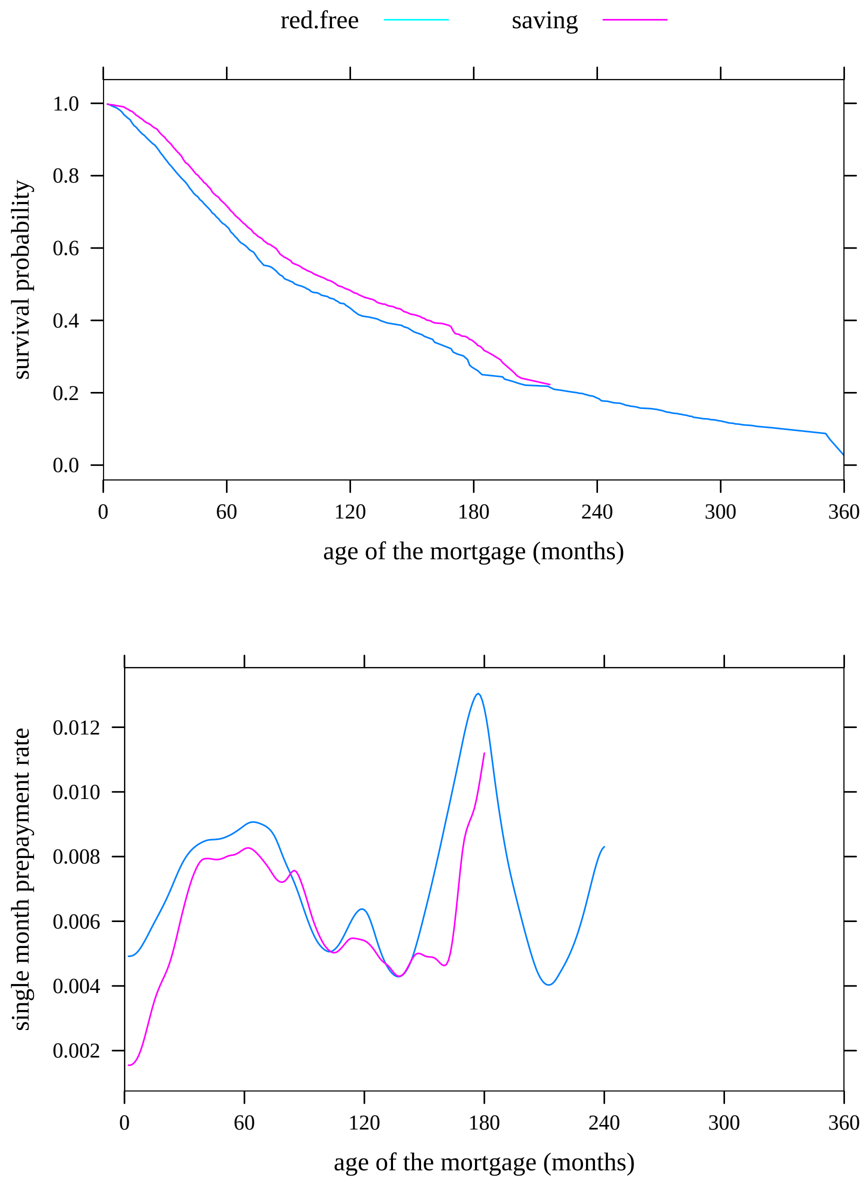


Figure 2: Estimated survival and hazard curve by product type.

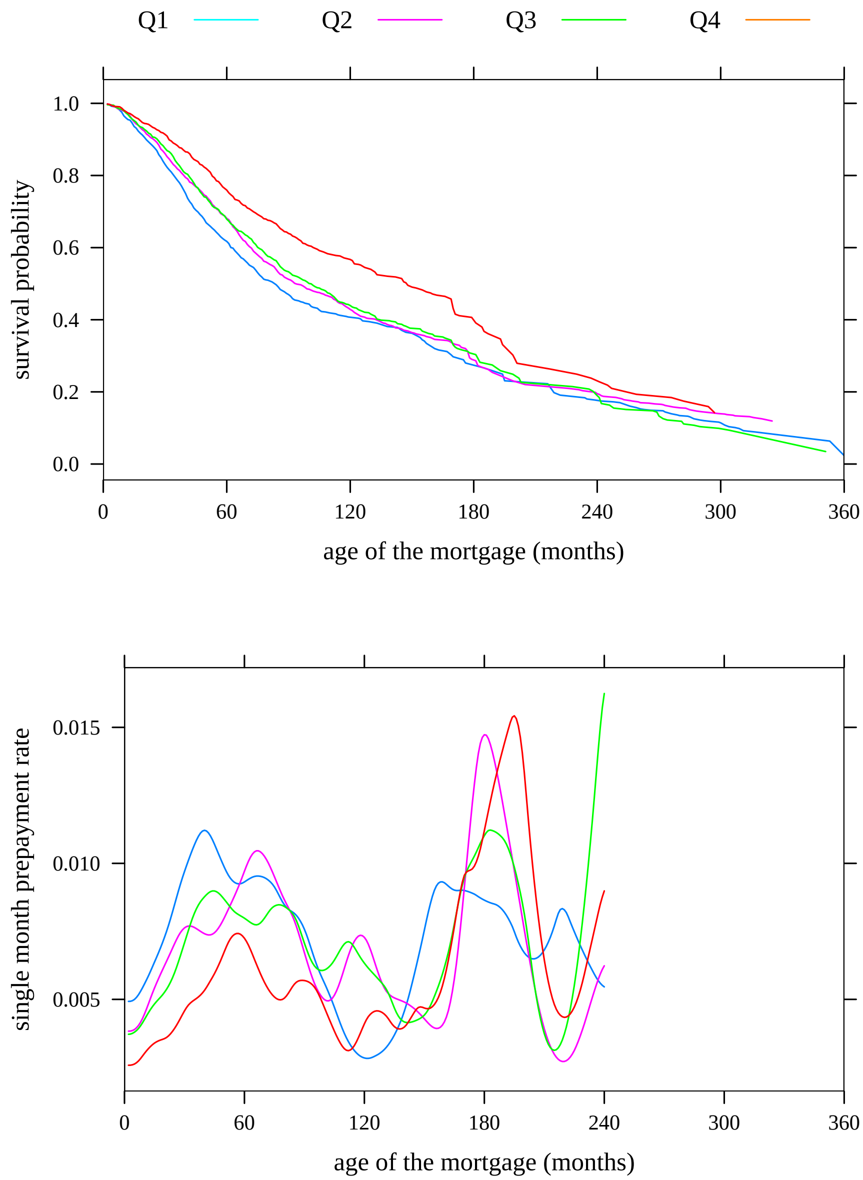


Figure 3: Estimated survival and hazard curve by principal.

institution. However, we do not pursue that issue here any further.

We proceed to look at the effects of socio-economic status of the zip-code area on the likelihood of prepayment. Measurement of this variable is described in Knol (2003). Even though we have measurements in three years (1994, 1998, and 2002), we graph survival curves for the status in 2002 only, assuming that the status is reasonably stable over time. Categorization of the socio-economic status score is such that each group contains approximately 1/3 of the observations. Mortgage contracts for houses in high status areas face lower prepayment than contracts in low status areas (figure 5). Of course, status in an zip-code area is correlated with the value of the house, and hence with the size of the loan. This finding is in agreement with the finding above that higher loans face lower prepayment rates.

Finally, we look at the effect of the market interest rates. We label the variable that measures the incentive of mortgagors to refinance their mortgage the 'refinance incentive' (*RFI*). Refinance incentive is a time varying covariate. It measures the (dis)advantage to the mortgagor if he were to prepay at time t, and refinance his mortgage until the next interest reset date against the current market rate at that moment. The refinance incentive changes over time for two reasons. First, market rates vary over time. Second, the time to the next interest reset date changes over time. We recall that at the interest rest date, a mortgagor can prepay his mortgage without incurring any interest penalty). These two aspects can be condensed in a single measure in different ways. We choose to define the refinance incentive as the net present value of interest payments saved (until the next interest reset date) if the mortgage could be refinanced against the prevailing interest rate, relative to the size of the loan:

$$RFI_{it} = \frac{\sum_{\tau=0}^{n} \frac{(cr_i - mr_{n,t})PP_i}{(1+d)^\tau}}{PP_i} \times 100\%, \tag{11}$$

where n is the number of months until the next interest reset date (which is contract specific and time-varying as well), cr_i the contract interest rate, $mr_{n,t}$ the market rate at time t for an n-months loan, and d the discount rate (taken to be 3% annually). The size of the loan, PP_i appears both in the numerator and the denominator of equation (11), and is in the end not relevant in our measure of refinance incentive.

From a computational point of view, introducing a time-varying covariate as the refinance incentive causes problems. It is possible to include time-varying covariates in a mixed proportional hazard model, but then one has to discretize the model in small observation windows during which the time-varying variables are constant. In our case, this means that we have one line per observation per month per contract. Our data set expands from 10000 contracts to 463075 observed contract months.

It is not possible to visualize the effect of time-varying covariate as in figures 2-5. Instead, we measure the effect of the refinance incentive through its impact on the hazard rate. We estimate an mixed proportional hazard model, with RFI_{it} as the only covariate:

$$\lambda_{it} = \lambda_{0t}^{PT} \exp(\beta RFI_{it}). \tag{12}$$

We stratify by product type, that is, each product type has its own baseline hazard (as indicated by the superscript PT for 'product type' in equation (12)). Deviations

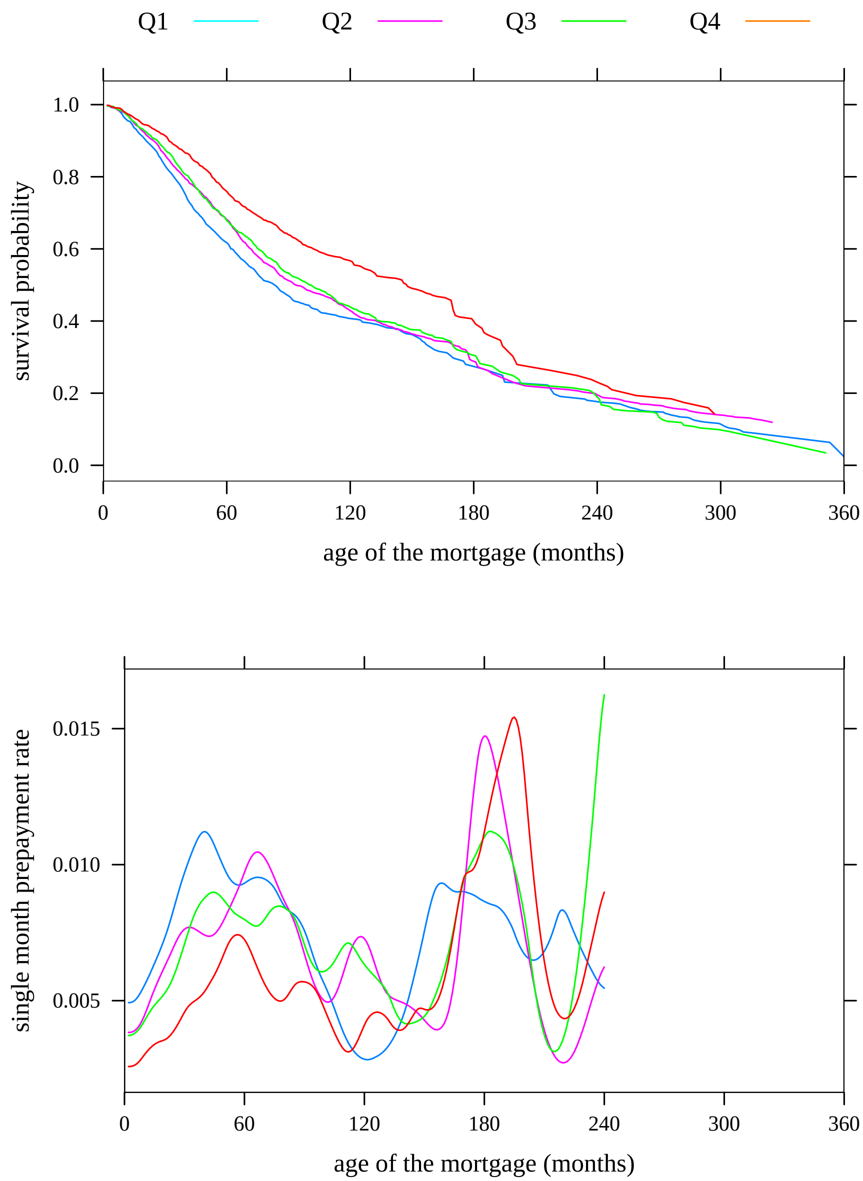


Figure 4: Estimated survival and hazard curve by age of the mortgagor at the initiation of the contract.

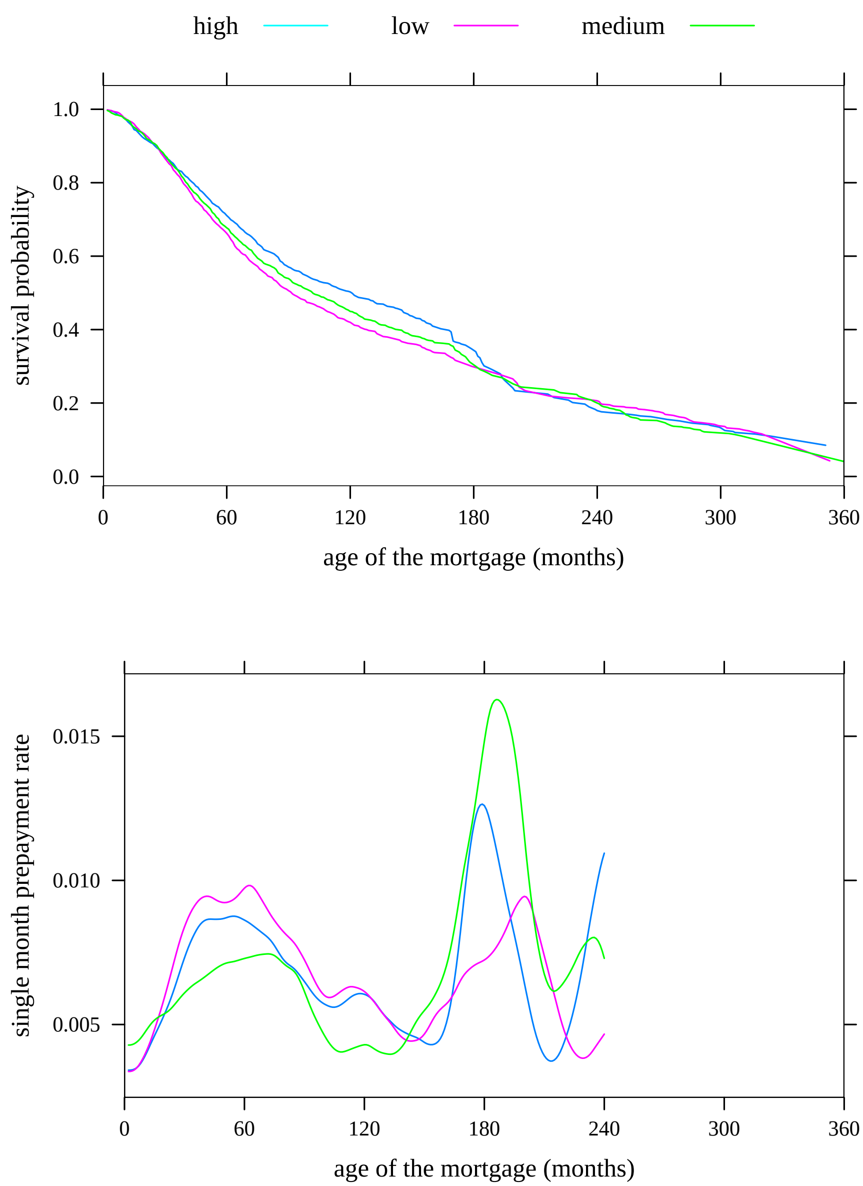


Figure 5: Estimated survival and hazard curve by socio-economic status of zip-code area.

Table 3: Estimation results hazard models by type of mortgage.

	coef.	exp(coef)	sd(coef)	z	p-value
saving					
loan.size.Q2	−0.009	0.991	0.094	−0.091	0.927
loan.size.Q3	−0.116	0.890	0.091	−1.277	0.202
loan.size.Q4	−0.419	0.658	0.094	−4.445	0.000
age.Q2	−0.281	0.755	0.064	−4.410	0.000
age.Q3	−0.372	0.690	0.067	−5.504	0.000
age.Q4	−0.312	0.732	0.080	−3.899	0.000
ses.medium	0.107	1.113	0.062	1.713	0.087
ses.low	0.133	1.143	0.062	2.136	0.033
rfi	0.001	1.001	0.003	0.404	0.686
red.free					
loan.size.Q2	−0.158	0.854	0.064	−2.451	0.014
loan.size.Q3	−0.083	0.920	0.091	−0.915	0.360
loan.size.Q4	−0.302	0.740	0.107	−2.808	0.005
age.Q2	−0.185	0.831	0.087	−2.128	0.033
age.Q3	−0.202	0.817	0.085	−2.387	0.017
age.Q4	−0.158	0.854	0.076	−2.063	0.039
ses.medium	0.055	1.057	0.068	0.812	0.417
ses.low	0.078	1.081	0.068	1.141	0.254
rfi	0.036	1.036	0.004	9.717	0.000

from this baseline hazard are captured through the proportional effect $\exp(\beta RFI_{it})$. The point estimate for β is 0.013, assumed to be the same for both product types. The exponentiated effect is 1.013, so a 1 percentage point increase of the refinance incentive raises the baseline hazard by 1.3%. This estimate differs significantly from zero. In a second attempt to estimate the effect of the refinance incentive, we use $\max(0, RFI_{it})$ as the explanatory variable, as this corresponds to the interest penalty to be paid if the mortgage is prepaid for other reasons than moving or interest reset. This truncated version of the refinance incentive does not have any significant effect on the survival of contracts.

In our analysis so far we have looked at one variable at a time only. It is more interesting to model prepayment with a multivariate model, as some of the variables are correlated. A multivariate model enables us to identify partial effects which are not measured through univariate modelling. Hence, we estimate hazard models of the form

$$\lambda_{it} = \lambda_{0t}^{PT} \exp(\beta' x_{it}), \tag{13}$$

where x_{it} is the vector of characteristics of contract i at time t (note, t is process time and not calendar time). Some experiments showed that the effects β differ by product type, so we present estimation results separately by product type. Considering the number of observations in our sample, this does not entail a significant loss of precision.

Estimation results are presented in table 3, the top half contains the results for savings mortgages, the bottom half the ones for redemption free mortgages. As in figures 3-5, loan size, age of the mortgagor at initiation, and socio-economic status score are discretized. Q1 refers to the first quartile of the applicable distribution, etc. The reference group is the first quartile of the loan size and age distributions, and a house in a high status zip code area. We also include refinance incentive as a covariate.

Qualitatively, the estimation results are similar for both types of mortgage. A higher loan size (compared to the first quartile of the loan size distribution) and an older mortgagor (again, compared to the first quartile of the age distribution) decrease the prepayment rate. Houses in medium or low status areas have a higher risk of prepayment. The refinance incentive is significant only for redemption free mortgages, which is perhaps not surprising considering the fact that these mortgages do not have capital accumulation. A 1 percentage point increase of the refinance incentive leads to a 3.6% higher hazard rate. Not all effects are significantly different from zero. Most notably, the socio-economic status indicator is insignificantly different from 0 in both equations. Probably, this effect has been captured by the loan size.

5 Summary and Conclusion

In this paper we have discussed a fruitful approach to model prepayment of mortgage contracts. Three variables known at this initiation of the contract (type of mortgage, size of the loan, and age of the mortgagor) provide useful information that can be used to estimate the median duration of the contract. The mortgagee can use this information to price prepayment risk, to refinance his portfolio, or to focus marketing efforts.

Even though the principal case of this paper is mortgage prepayment, we believe that the methodology discussed can be used in other cases as well. Considering the fact that insurance companies and financial institutions are required to value loans on an ongoing basis, prepayment information is crucial.

References

Alink, B.J. (2002). *Mortgage Prepayments in The Netherlands*. Ph. D. thesis, University of Twente.

Archer, W.R., D.C Ling, and G.A. McGill (1996). The effect of income and collateral constraints on residential mortgage termination. *Regional Science and Urban Economics 26*, 235–261.

Brennan, M. and E.S. Schwartz (1985). Determinants of GNMA mortgage prices. *AREUEA Journal 13*, 209.

Bussel, A.P.J.M. van (1998). *Valuation and interest rate risk of mortgages in The Netherlands*. Ph. D. thesis, University of Maastricht.

Charlier, E. (2001). Quantifying refinancing incentives and losses from prepayments. *The Journal of Fixed Income 10*, 55–64.

Charlier, E. and A. van Bussel (2003). Prepayment behavior of Dutch mortgages: an empirical analysis. *Real Estate Economics 31*, 165–204.

Dunn, K. and J. McConnell (1981). Valuation of GNMA mortgage-backed securities. *The Journal of Finance 36*, 599–616.

Golub, W.B. and L. Pohlman (1994). Mortgage prepayments and an analysis of the Wharton prepayment model. *Interfaces 24*, 282–296.

Green, J. and J.B. Shoven (1986). The effects of interest rates on mortgage prepayments. *Journal of Money, Credit, and Banking 18*, 41–59.

Harrell, Jr, F.E. (2001). *Regression Modelling Strategies*. New York: Springer.

Hayre, L. (2003). Prepayment modeling and valuation of Dutch mortgages. *The Journal of Fixed Income 12*, 25–47.

Kang, P. and A.S. Zenios (1992). Complete prepayment models for mortgage-backed securities. *Management Science 38*, 1665–1685.

Knol, F. (2003). Rangorde naar sociale status van postcodegebieden in Nederland. `www.scp.nl/bronmateriaal/statusscores/index.html`.

Koning, R.H. and G. Ridder (2000). Utility maximization and mode of payment. In R. Heijmans, D.S.G Pollock, and A. Satorra (Eds.), *Innovations in Multivariate Statistical Analysis*, Dordrecht, pp. 117–136. Kluwer Academic Publishers.

Koning, R.H. and G. Ridder (2003). Discrete choice and stochastic utility maximization. *The Econometrics Journal 6*(1), 1–27.

Maxam, C.L. and M. Lacour-Little (2001). Applied nonparametric regresson techniques: Estimating prepayments on fixed-rate mortgage-backed securities. *Journal of Real Estate Finance and Economics 23*(2), 139–160.

Rabobank (2003). Dutch residential mortgage market from hype to pain. Rabobank Securities, May 2003.

Schwartz, E.S. and W.N. Torous (1992). Prepayment, default, and the valuation of mortgage pass-through securities. *Journal of Business 65*, 221–239.

Smith, P.J. (2002). *Analysis of Failure and Survival Data*. London: Chapman & Hall/CRC.

Therneau, T.M. and P.M. Grambsch (2000). *Modeling Survival Data*. New York: Springer.

Venables, W.N. and B.D. Ripley (2002). *Modern Applied Statistics with S* (Fourth ed.). New York: Springer.

AUTHOR'S ADDRESSES

dr. J.P.A.M. Jacobs, Dept. of Economics, University of Groningen, PO Box 800, The Netherlands, email: j.p.a.m.jacobs@rug.nl.
prof. dr. R.H. Koning, Dept. of Economics, University of Groningen, PO Box 800, The Netherlands, email: r.h.koning@rug.nl.
prof. dr. E. Sterken, Dept. of Economics, University of Groningen, PO Box 800, The Netherlands, email: e.sterken@rug.nl.