Loan Characteristics and Selectivity, Unobserved Heterogeneity and the Performance of United Kingdom Securitised Sub-Prime Loans

Gauthier Lanot
Professor of Empirical Economics, Keele Management School, Keele University
Contact: e-mail g.lanot@econ.keele.ac.uk, tel. +44(0) 1782 733102

David Leece (contact author)
Professor of Financial Studies, Manchester Metropolitan University Business School.
Contact: e-mail d.leece@mmu.ac.uk, tel. +44 (0)161 247 3742, Fax +44 (0)0161 247 6303
A Competing Risk Model of United Kingdom Sub Prime Mortgage Loan Performance with Controls for Unobserved Heterogeneity

Abstract

The research estimates a competing risk model of mortgage terminations on samples of UK securitised subprime mortgages. Given the argued role of these types of loan in the recent financial crisis then it is important to better understand their performance and supposed ‘idiosyncratic’ behaviour. The methodological advance is the use of a general, flexible modelling of unobserved heterogeneity over several dimensions, controlling for both selection issues involving initial mortgage choices and dynamic selection over time. Moreover, we estimate specific coefficients for this unobserved heterogeneity and determine the correlation between the unobserved components of default and prepayment. We conclude that securities consisting of subprime loans can be given meaningful valuations on bank balance sheets if the performance of loans with different chosen features and the associated unobserved heterogeneity can be better understood.

JEL Codes: G21 C13 C25 C51 D10 D14 E44

Key Words: Sub prime mortgages, unobserved heterogeneity, loan performance, securitisation

I. INTRODUCTION

The paper reports the results of estimating a model of default and prepayment behaviour with competing risk and unobserved heterogeneity, using a sample of UK (United Kingdom) subprime mortgages, originated and subsequently securitised before the financial difficulties of 2007 and held by a global investment bank. The research extends the econometric literature in this area by introducing an estimation that provides a more flexible approach to modelling unobserved heterogeneity which incorporates selectivity surrounding initial contract choices, in addition to presenting individual coefficients for these effects.

The motivation for the research is the recent credit crisis and its legacy. The ‘aggressive’ extension of mortgage lending to sub prime borrowers leading to losses on pools of sub prime debt are seen as the proximate causes of the global credit crunch that began in 2007. The analysis of the loan performance of quits from mortgage pools of US (United States) prime borrowers, and to some extent sub prime mortgage holders, are reported elsewhere (Alexander et al, 2002; Chomsisengphet and Pennington Cross, 2004; Courchane et al, 2004; Cowan and Cowan, 2004; Danis and Pennington Cross, 2008; Stephens and Quigars, 2008) and offer a variety of results concerning the effects of different contract design on loan performance, and the influence of variables representing the value
of embedded options in a mortgage contract. Our work adds to this literature, covering mortgages originated and held in a more recent time period (2001-2008).

Unobserved individual influences, that is unobserved heterogeneity, would appear to be a probable source of variation in loan/security performance (Deng et al, 2000). This paper introduces a general and flexible modelling of this heterogeneity over the relevant two dimensions of default and prepayment behaviour. This approach provides a framework capturing both the selection issues arising from initial mortgage choices, and the dynamic selection effects arising from changes in the population of individuals who remain in the pool of mortgages at each discrete point in time. The econometric estimation is novel in estimating coefficients for unobserved heterogeneity for a multinomial logit model, and reporting the correlation between the unobserved influences upon both mortgage refinancing and default. In addition the modelling and the data base which involves a variety of mortgage contracts facilitating a focus upon the selectivity issues relating to initial mortgage choices.

A further distinctive feature of the research is the use of data relating to sub prime loans originated in the United Kingdom. To our knowledge, there is no econometric work on sub prime loan performance for the United Kingdom which is comparable with that undertaken for the United States mortgage market. This is a major omission given the extent to which UK mortgages were securitised, together with the variety of and distinctive contractual features of UK housing debt (see Leece, 2004; Miles, 2005). The Bank of England estimated that the stock of outstanding non conforming securitised debt in the UK to be £39 billion (Bank of England, 2008). An understanding of the factors underpinning UK sub prime loan performance, and the problematic nature of pricing risk on these securities, contributes to the debates on valuation; on future requirements of any reinvigorated securitised mortgage market; on mortgage contract design; on risk management; and on financial regulation.

The paper begins with a review of the academic literature on the default and prepayment behaviour of mortgage loans and sub prime loans where applicable. This is followed by an outline of the econometric methodology and the modelling approach to unobserved heterogeneity. The sample and the empirical specification of the model are discussed in the section which follows. Parameter estimates are then reported and analysed.
II. DEFAULT AND PREPAYMENT

This section of the paper presents the key theoretical and empirical approaches adopted in the mortgage loan performance literature and positions the research in relation to that work. The discussion facilitates the identification of the key influences upon default and prepayment behaviour which informs the empirical model that follows.

The majority of empirical studies focus on the prime lending market in the United States rather than subprime or non-conforming loans. However, there is a growing body of literature focussing upon sub prime mortgage loan performance, again in the US (Alexander et al, 2002; Pennington Cross, 2003; Cowan and Cowan, 2004; Deng and Gabriel, 2006; Pennington Cross and Chomsisengphet, 2007; Danis and Pennington Cross, 2008). The literature emphasizes: the influence of contract features such as prepayment penalties and reduced documentation upon the probability of foreclosure (Pennington-Cross et al, 2010; Quercia, Stegman and Davis, 2005; Rose, 2008); the effect on default rates of originating loans from third parties (Alexander et al, 2002; Pennington Cross, 2003); and how default rates vary by loan classification (Cowan and Cowan, 2004). Econometric specifications and findings in the research of the subprime market tend to reflect the research into the behaviour of prime mortgage loans and finds that relative to the finding on prime loans, co-variates have larger marginal effects on the default and/or repayment probabilities (Pennington Cross and Chomsisengphet, 2007).

Option Theoretic and Empirical Prepayment Models

The literature on the options embedded in mortgage contracts and their impact upon loan performance is well established, both theoretically and empirically (Kau et al, 1993; Deng et al, 2000; Ambrose and LaCour-Little, 2001; Ambrose and Sanders, 2003). The possibility of defaulting on a loan is treated as a put option (selling the house back) while prepayment is considered as a call option (buying back the mortgage). The analysis is to some extent United States specific, for example in the UK the borrower retains liability for the outstanding mortgage debt on default (Leece, 2004), i.e. the put option may not be so valuable. The prepayment option also applies more readily to long term fixed interest rate mortgages more typical of the US (Leece, 2004). However, short term fixed rates and periods were the interest rate is discounted, but eventually reverts to an higher rate, provide boundary conditions for valuing the call option (Kau, 1993).

Previous work suggests that borrowers do not always take systematic advantage of the embedded options they hold, such as not prepaying when favourable alternative contracts are available (the call option is in the money) or not defaulting when the put option is well into the money. This has led to several developments. One is to estimate empirical prepayment models that recognise the
importance of exogenous effects (surprises) on default and prepayment behaviour, for example the
effect of payment shocks (Quigley and Van Order, 1990, 1995; Archer and Ling, 1993). In particular
some work has focused on studying loan level data where borrower characteristics can be analysed.
Furthermore, the literature has further emphasized the role of unobserved heterogeneity among
borrowers (Deng and Quigley, 2002; Alexander et al, 2002).

The majority of recent empirical studies of mortgage loan performance now incorporate modelling
of both the embedded options in mortgage contracts and variables typical of empirical prepayment
models. There is also a recognition that the embedded options represent a competing risk in that
the exercise of one option precludes the exercise of the other (Deng et al, 2000; Lambrecht et al,
2006). The research reported in this paper uses loan level data and estimates an empirical model of
mortgage default and prepayment which incorporates both an option theoretic specification and
includes variables that impact upon affordability, or reflect exogeneous shocks. Our study is unusual
in incorporating a wide variety of mortgage contracts in the sample which makes explicit issues
regarding selection bias which have not been treated in previous work looking at single types of
contract.

**Mortgage Design**

The mortgage contracts studied in the US are typically fixed rate mortgages with the interest rate
fixed for 15 or 30 years, and adjustable rate mortgages (ARM) where the rate of interest changes
annually. This compares to UK mortgage contracts where the majority have interest rates fixed for
one to three years or the interest rate changes at irregular periods (a variable rate mortgage). The
UK one year fixed rate contracts are similar to the US adjustable rate mortgage and the two to three
year interest rate fixes are equivalent to the so called 'hybrid mortgage' in the US (see Ambrose et
al, 2005). The details of these contracts can have an impact upon prepayment and/or default
behaviour, for example prepayment penalties (Pereira et al, 2002). For the purposes of discussion in
this paper we retain the label fixed rate for UK mortgages that have the rate of interest fixed for any
period and discounted mortgages for those contracts with lower initial rates (teaser rates in the US
nomenclature).

The literature finds that alternative contracts may elicit different patterns of behaviour. For
example, there is evidence for the United States that adjustable rate mortgages prepay at a faster
rate than fixed rate mortgages (Ambrose and LaCourLittle, 2001). Households holding adjustable
rate mortgages may be more mobile than those with fixed rate contracts and will start refinancing
after a short period of time (Brueckner, 1995). Borrowers choosing a discounted mortgage may
chase new (better) deals. Thirdly ARM borrowers may refinance into fixed rate mortgages on the
reset date, depending upon interest rate expectations. Hence discounted mortgage holders will tend
to have a higher likelihood of prepayment than fixed rate mortgage holders. The payment shock
which arises around the date of adjustment of ARM interest rates might induce a higher level of
mortgage defaults (Ambrose et al, 2005), an effect that might also be apparent with UK short term
fixed rate debt.

The mortgage contracts featuring in the current research involve self-certification, discounting,
fixed rate contracts, and dates at which the interest rate on the contract reverts from a favourable
rate of interest back to an higher index rate (typically London Interbank Offered Rate, LIBOR) plus
margin. The real estate economics and finance literature has examined the effects of these different
aspects of contract design on the performance of mortgage loans (Pennington-Cross et al, 2010;
The majority of papers have considered the effect of discounting the initial interest rate (teaser
rates). The empirical results have been mixed and contradictory. Later work should be credited with
the use of a competing risk framework and controlling for unobserved heterogeneity. For example,
Ambrose and LaCour-Little (2001) apply this methodology and find a significant increase in
prepayments at adjustment dates.

A further key feature of recent mortgage contracts has been low or zero documentation in the US
and its equivalent the self certified mortgage in the UK\(^1\). This relaxed approach to mortgage
underwriting attenuates or overrides prudential lending criteria and introduces information
asymmetry, with the lender knowing less about the borrower’s ability to pay and likelihood of
default. There may be substantial adverse selection and borrowers may exhibit opportunistic
behaviour (Brueckner, 2000; Leece, 2004). As a consequence self-certification will have a positive
effect on the likelihood of defaulting. For an exception see Rose (2008) who estimates a multinomial
logit model with unobserved heterogeneity using securitised sub prime loans for the Chicago
Metropolitan area from January 1999 and up to mid 2003. Rose finds that the effects of variables on
foreclosure depend upon loan features such as the level of documentation.

A key motivation for researching the impact of contract designs on loan performance is the question
of which loans should be securitised and how such securities would be valued. Until recently,
adjustable rate and other mortgage designs were less likely to have been securitised than long term

\(^1\) Self-certified mortgages are designed for self-employed or employed individuals with uncertain incomes or
incomes from multiple sources. These contracts may therefore also proxy this income uncertainty. There are
also claims that self certification was used as an avoidance of due diligence and prudential lending; lending to
households on social benefits, or with implausible income statements. In 2008 52% of all new mortgages were
self certified (Financial Services Authority, 2010).
fixed rate contracts (Ambrose and La-Cour Little, 2001). Furthermore, loan contracts which exhibit complex and less predictable default and repayment rates may be less suitable for securitisation. Research to date has typically controlled for different contract designs by analysing loans of a given type (typically long term fixed or adjustable rate). Even analysing one type of loan raises issues regarding selectivity (i.e. the borrowers associated with a particular loan specie may share observed or unobserved characteristics which may increase a specific risk). When several types of mortgage are pooled in the same security then the heterogeneity of these specific risk may multiply. The research reported in this paper addresses the selectivity that arises from the individual specific factors that generate the initial choice of contract type by modelling these factors as unobserved heterogeneity. We subsequently evaluate the effect of mortgage contract choice upon both default and prepayment behaviour in a competing risk framework.

III. ECONOMETRIC ESTIMATION

The use of econometric techniques in the mortgage loan literature has evolved from the use of qualitative dependent variable models to the application of Cox Proportional Hazard (CPH) and Multinomial Logit models (MNL) to incorporate competing risks. Further developments have recognised the potential importance of unobserved heterogeneity, particularly when modelling behaviour which from an option theoretic viewpoint appears sub optimal (Deng et al, 2000, 2002). The MNL has the advantage over CPH that the competing probabilities sum to unity modelling competing risk explicitly, because one risk is at the expense of the other, but it has the disadvantage of assuming the independence of irrelevant alternatives. There has been a tendency to favour the multinomial logit model with modifications to allow for the correlation between alternative specific unobserved components. The control of unobserved heterogeneity when using CPH has been based upon the estimation of mass points (see Pennington-Cross et al, 2010).

The econometric methodology reported here advances the literature in several ways. Firstly, we model unobserved heterogeneity as a continuous distribution, rather than estimating parameters for an arbitrary number of mass points that shift the base line hazard. Though there have been some plausible a-priori categorisations of groupings of unobserved heterogeneity for example: employment history; changes in marital status; household mobility (Clapp et al, 2006)- a specification that uses a more general form can cover a wide number of dimensions (unobserved attributes and selectivity) and is more flexible. There are also likely to be sources of selectivity bias in the choice of type of mortgage and other mortgage choices made by households (Nichols et al, 2005). The econometric methodology treats unobserved heterogeneity as a determinant for this selectivity while retaining the advantages of modelling competing risk with a MNL analysis.
The model presented here also allows for correlation between the sources of unobserved heterogeneity that affect the various decisions (to remain current, to default or to prepay). Though this has been used in a limited number of studies it is generally used in the context of the Cox Proportional Hazard model (see Alexander et al, 2002). It has been suggested that modelling unobserved heterogeneity with a more general functional form is too time consuming/expense and is not available in commercial software (Clapp et al, 2006). The technique reported in this paper speeds up estimation and facilitates convergence of the likelihood function by using the methodology of Train (2003) and the adaptation by Lanot (2008). For an outline of the estimation procedure see Appendix A.

Overview

We wish to model the individual (discrete time) history of the decisions of default or early repayment, \( \{d_{it}\}_{t=1}^{T_i} \), jointly or conditionally on the history of a set of time dependent regressors, say \( \{x_{it}\}_{t=1}^{T_i} \). Here we think of \( t \) as the history time and not the calendar time. \( T_i \) is the first of three possible times, either it is the date when the individual decides to repay or decides to default, or it is the end of the observation period. \( d_{it} \) can take three possible values: 1 if the individual decides to keep paying the mortgage in period \( t \), 2 if the individual “decides” to default and 3 if the individual decides to repay the mortgage early.

Furthermore, we wish to account for the difference at the time of contracting between mortgage/contract type and initial characteristics of the loan, say \( c_i, x_{i0} \) with \( c_i \) is potentially a vector of qualitative variables indicating the type of contract chosen, while \( x_{i0} \) measures the more quantitative aspects of the loan (the amount borrowed, the value of the property on which the loan is based, etc). For example, it seems natural to distinguish between certificated mortgage and self certificated mortgages and/or between fixed and variable mortgages.

Finally we wish to account for unobserved differences between individuals which may affect both “exit” decisions independently or jointly. Hence we denote \( \varepsilon_i \equiv (\varepsilon_{i1}, \varepsilon_{i2}) \) the vector of unobserved individual factors. We assume that the marginal joint distribution of \( \varepsilon_i \) is normal with means 0, unit variances and zero correlation. We discuss later on how this specification captures the likely dependence between the two individual factors.

Assumptions
Firstly, there is a discussion of the general assumptions we wish to maintain to estimate the parameters of interest. In general, given a set of initial exogenous variables, say $w_{i0}$, we can always express the probability density of a given history in the following general fashion:

$$f(e_i, c_i, x_{i0}, \{x_{it}\}_{t=1}^T, \{d_{it}\}_{t=1}^T | w_{i0}) =$$

$$f_{\theta_{00}}(e_i | w_{i0}) f_{c, x, w, \theta_{01}}(c_i, x_{i0} | w_{i0}, e_i) f_{x_{it}, x_{i0}, w_{i0}, \theta_{02}}(\{x_{it}\}_{t=1}^T | c_i, x_{i0}, w_{i0}, e_i)$$

This decomposes the joint density of all the quantities of interest into a product of marginal and conditional densities. We denote by $f_\sigma$ each density, and the subscript indicates the identity (its argument and the set of conditioning variables) of each density.

For practical and computational reasons the following restriction on the conditional distribution of the time varying covariates is maintained:

$$f_{x_{it}, x_{i0}, w_{i0}, \theta_{03}}(\{x_{it}\}_{t=1}^T | c_i, x_{i0}, w_{i0}, e_i) = f_{x_{it}, x_{i0}, w_{i0}}(\{x_{it}\}_{t=1}^T | c_i, x_{i0}, w_{i0})$$

All the information contained in the individual specific effect $\varepsilon_i$ is captured by the the initial value of the characteristics of the loan, $x_{i0}$, and the type of loan $c_i$. This assumption is plausible in fact since it suggests that the regressor’s history from time 1 onward is independent from the individual specific effect conditional on the initial characteristics of the loan $c_i, x_{i0}$ and the predetermined variables $w_{i0}$. This implies that we believe the evolution of the time varying covariate is mostly determined outside of the individual time invariant circumstances and/or decisions (this means that conditional on the choice of interest rate the joint distribution of future interest rates is independent of individual specific unobserved components).

**Contract Choices, Selectivity and Unobserved Heterogeneity**

Conditioning on the individual specific factors in the description of the distributions of the initial characteristics of the individual loan contracts allows us to specify the dependence between the initial characteristics and the outcome of interest, i.e. the timing and the nature of the exit decisions. This provides a “natural” parametrisation for the (dynamic) selection effects we would expect to observe. In principle we are able to evaluate quantities such as the distribution of the duration until defaults/repayment given the observed initial characteristics or the distribution of the expected time to default/repayment given the observed initial characteristics. The introduction of the unobserved
component allow for a more “flexible” correlation structure between the diverse elements of the models. In particular it is more flexible than a comparable specification (with the factor loadings for the individual specific effect set to 0) which would rely only on conditional independence.

The density of the initial observed characteristics of the loan is assumed to be jointly normally distributed with a mean vector depending linearly on \( \varepsilon_i \) and a constant variance covariance (although it would be feasible to make the variance covariance matrix dependent on some exogenous observed characteristics as well as dependent on the individual specific effects in \( \varepsilon_i \)). Formally we assume:

\[
x_{i0} \mid w_{i0}, \varepsilon_i \sim N \left( x_{i0}^0 + \varepsilon_i \Delta_1^0 + \varepsilon_i \Delta_2^0, \Sigma \right),
\]

where \( x_{i0}^0 \) is a parameter matrix and \( \Delta_1^0 \) and \( \Delta_2^0 \) are parameters vector conformable to the dimensions of \( x_{i0} \). \( \Sigma \) is the variance-covariance matrix for \( x_{i0} \) given \( w_{i0} \) and \( \varepsilon_i \).

We assume that the probability density of the mortgage type is of the logit or multinomial logit form in the various dimensions of choice (i.e. certification on the one hand and interest rate choice).

Firstly, in the case of the certification choice we assume that the probability is:

\[
f_{c_1 \mid x_{i0}, w_{i0}, \varepsilon_i} (0 \mid x_{i0}, w_{i0}, \varepsilon_i) = \frac{1}{1 + \exp \left( -\left( x_{i0} \alpha_0^1 + w_{i0} \kappa_0^1 + \delta_1^1 \varepsilon_i + \delta_2^1 \varepsilon_i \right) \right)},
\]

\[
f_{c_1 \mid x_{i0}, w_{i0}, \varepsilon_i} (1 \mid x_{i0}, w_{i0}, \varepsilon_i) = 1 - f_{c_1 \mid x_{i0}, w_{i0}, \varepsilon_i} (0 \mid x_{i0}, w_{i0}, \varepsilon_i).
\]

Since there are three potential choices for the interest rate choice (fixed, discount or Libor), given the certification choice we assume that probabilities take the form

\[
f_{c_2 \mid c_1, x_{i0}, w_{i0}, \varepsilon_i} (c_2 \mid c_1, x_{i0}, w_{i0}, \varepsilon_i) = \frac{e^{Z\left(c_2\right)}}{\sum_{k=1}^{3} e^{Z\left(k\right)}}, \text{ with } c_2 \in \{1,2,3\},
\]

Where:
In some cases one of the three options (option 3) is not available at a particular time, and the model above collapses to a simple binary logit model.

**Estimation**

We assume that the conditional joint likelihood of the history of decisions can be decomposed into a product of conditional probabilities:

\[
    f_{d|x,c,x_0,w_0,e} \left( \{d_a \}^T_{a=1}, \{x_a \}^T_{a=1}, c_i, x_{i0}, w_{i0}, \varepsilon_i, t \right) = \prod_{a=1}^T f_{d|x,c,x_0,w_0,e} \left( d_a \mid \{x_a \}^T_{a=1}, c_i, x_{i0}, w_{i0}, \varepsilon_i, t \right)
\]

Practically we assume that each probability \( f_{d|x,c,x_0,w_0,e} \left( d \mid x_a, c_i, x_{i0}, w_{i0}, \varepsilon_i, t \right) \) is of the multinomial logit form:

\[
    f_{d|x,c,x_0,w_0,e} \left( d \mid x_a, c_i, x_{i0}, w_{i0}, \varepsilon_i, t \right) = \frac{e^{V_i(d)}}{\sum_{k=1}^3 e^{V_i(k)}}, \text{ with } d \in \{1,2,3\},
\]

where:

\[
    V_i(1) = 0,
\]

\[
    V_i(2) = x_{i0}\beta_2 + c_{2i}\gamma_2 + \lambda_2^3c_{i1} + x_{i0}\alpha_2 + w_{i0}\kappa_2^3 + \delta_{21}\varepsilon_{i1},
\]

\[
    V_i(3) = x_{i0}\beta_3 + c_{3i}\gamma_3 + \lambda_3^3c_{i1} + x_{i0}\alpha_3 + w_{i0}\kappa_3^3 + \delta_{31}\varepsilon_{i1} + \delta_{32}\varepsilon_{2i}.
\]

The parameters of interest which appear in the conditional probabilities are therefore \( \left( K^0, \Sigma, \alpha_0, \alpha_j, \lambda_j, \kappa_j, \beta_j, \gamma_j, \Delta_{k1}^i, \Delta_{k2}^i, \xi_{k1}, \xi_{k2} \right) \) with \( j=2,3 \) and \( k=2,3 \). The parameters \( \left( \Delta_{k1}^i, \Delta_{k2}^i, \xi_{k1}, \xi_{k2} \right) \) are the loadings of the individual specific component in the conditional densities/probabilities and capture the dependence between the dependent variables and the unobserved individual specific unobserved components.
Of particular interest is the interpretation of the sign and magnitude of the parameters $\delta_{21}^3, \delta_{31}^3, \delta_{32}^3$ in the conditional density of the repayment/default decisions. Recall that we assumed that the components of $\varepsilon_i$ are uncorrelated; this is however only a matter of presentation. Indeed, we are always able to define the parameters $\delta_{31}^3$ and $\delta_{32}^3$ as functions of an unrestricted parameter $\delta^3$ and a correlation coefficient $\rho$ as follows:

$$\delta_{31}^3 = \delta^3 \rho \quad \text{and} \quad \delta_{32}^3 = \delta^3 \left(1 - \rho^2\right)^{1/2}.$$  

The term $\delta_{31}^3 \varepsilon_{i1} + \delta_{32}^3 \varepsilon_{i2}$ can then be written as $\delta^3 \left\{ \varepsilon_{2i} \left(1 - \rho^2\right)^{1/2} + \rho \varepsilon_{1i} \right\}$ and the term in braces is then a normal variate correlated, by construction, with $\varepsilon_{1i}$. We therefore have the correspondence:

$$\delta_{31}^3 > 0, \delta_{32}^3 > 0 \iff \delta^3 > 0, \rho > 0;$$
$$\delta_{31}^3 > 0, \delta_{32}^3 < 0 \iff \delta^3 < 0, \rho < 0;$$
$$\delta_{31}^3 < 0, \delta_{32}^3 > 0 \iff \delta^3 > 0, \rho < 0;$$
$$\delta_{31}^3 < 0, \delta_{32}^3 < 0 \iff \delta^3 < 0, \rho > 0.$$

The ratio of $\delta_{31}^3 / \delta_{32}^3$ is an increasing function of $\rho$ only and therefore its inversion gives an estimate of $\rho$, and given this estimate it is then straightforward to obtain an estimate for the parameter $\delta^3$ (applying the delta method to the transformation given the precision for $\delta_{31}^3$ and $\delta_{32}^3$ will provide an easy way to obtain the precision for $\rho$ and $\delta^3$).

The difficulty with the estimation of these kind of models resides in the fact that the individual specific effect are not observed, and therefore the observed likelihood is derived from the latent likelihood described above by integrating out the individual specific effects. This observed likelihood is potentially difficult to evaluate (and therefore optimise), since it requires to integrate a product of terms over the multidimensional density of the unobserved component. Instead of taking this direct route, we adapt the EM algorithm described in Train (2008), which is used in the case of the estimation of discrete mixtures to the case where the distribution of the unobserved component is known in order to obtain the maximum likelihood estimates. The advantage rests in the fact that both the E-step and the M-step are relatively straightforward in our case, and that in particular the M-step only requires the maximisation of the sum of standard (concave) logit likelihoods and of the likelihood for a multivariate normal variable, weighted by quantities which are directly calculated.
from the usual normal quadrature abscises and weights and the complete latent likelihood (see the appendix for some details). The precision of our estimates is calculated from the latent likelihood used in the EM algorithm using the results derived in Oakes (1982) and adapted to this case in Lanot (2003).

IV. DATA AND EMPIRICAL MODEL

The Data

The data covers 100,000 mortgage contracts, and when constructed as a panel contains approximately two million observations (individual x time points). The mortgages were issued by a single US originator operating in the UK non-conforming residential mortgage market, but the pools also contain some prime and near prime debt. These issues remained the property of a major global investment bank with which one of the researchers undertook collaborative work. The research is subject to confidentiality agreements, and as such the identity of the data source cannot be disclosed.2

The mortgages are classified by issue and there were two issues per year from January 2003 to May 2006, offering eight issues over four years (labelled 103, 203, 104, 204, 105, 205, 106, and 206). The data was collected by issue rather than by the tranches of loans packaged into different securities. For example, the January 2003 issue (103) is a mortgage portfolio available in eight tranches but we were not given the means to identify these. The data covers the full range of types of loan that were securitised, including fixed rate, discounted interest rates, buy to let and self certified mortgages.

The data was reduced by removing the comparatively small number of prime and near prime loans (0.0789 and 0.0432 of total observations respectively) concentrated in particular issues. Buy to let mortgages (0.0842 of observations but with 0.508 of buy to let also prime loans) were also excluded so that the sample included only first lien loans on owner occupied property. Given the magnitude of the data set, and the difficulties of achieving convergence with the econometric models used, the data was broken down into four sub samples. Sample 03 contains issues 103 and 203, those for 2004 are in sample 04, 2005 sample 05 and 2006 sample 06. Using these year by year samples assisted estimation and allowed a comparison of changes in parameter estimates and unobserved heterogeneity for mortgages issued and pooled over different periods of time.

2 Confidentiality is maintained by not estimating models on particular tranches of securitised debt, but rather incorporating the whole issue for analysis. Though much of the data is now in the public domain the absence of dates of redemption and repossession on investor reports means that the timing of exits from the pool cannot be reconstructed. Therefore public domain data is not fully useable beyond May 2008.
Reference to the descriptive statistics by sample shown in Table 1 reveals that there is significant variation in the representation of different types of contract across and within issues. In particular the number of self-certified mortgages is significantly higher in the 04 sample. The proportion of fixed rate mortgages also increases in samples 05 and 06 compared to 03 and 04. Variations in contract terms increased significantly for later issues (samples); for example issue 203 has 49 different contracts and by issue 106 this has increased to 639. Insofar as different types of mortgage exhibit differences in loan performance then mortgage pools and their tranches of securities will also differ. The incidence of types of contract, across the issues/samples, led to low representation of some contracts, or colinearity problems such that not all contract choices could be modelled in each estimation. This was not a problem as each contract choice could be modelled within at least one sample.

The investment bank selected the loans of this particular mortgage originator for analysis. Therefore there is concern that the pools may be subject to selectivity/originator bias (see Ciochetti et al, 2003). Data is unavailable to directly correct for this potential bias and therefore there needs to be caution in generalising any empirical results. However, there is some significant within sample variability in performance across the tranches of debt. Each of the issues used in the research originally had a triple A credit rating from Standard and Poor and Moody’s, with an Aaa from Fitch. By March 2009 the ratings of some of the issues had fallen to BBB and BBB-. Though the data was from an originator in the top decile of average loan performance there was considerable variation by issue and between the yearly grouping of issues. This variation in performance can be seen in Figure 1 and Figure2 which show conditional prepayment and conditional default rates by issue. It is also the case that the legitimacy of credit ratings of AAA rated debt can be disputed (Lupica, 2008).

**Insert Figure 1 and Figure 2**

**Conditional Prepayment Rates and Conditional Default Rates by Issue**

We compared the default rates observed in our data with the mean and standard deviations of monthly repossession rates for 160 issues from 26 mortgage originators, evaluated by the same credit rating agency as the pools studied here. The comparison covered a period of twenty four months since the mortgages were pooled. The mean value for all issues was 0.81% with a standard deviation of 1.15%. The mean values for all the samples were within one standard deviation of the overall mean (0.60, 0.29, 0.75, and 0.70 for sample 03 to sample 06 respectively). Earlier dated pools performed better than those issued at a later date introducing some significant variation. Similar results were obtained for comparisons made at twelve and eighteen months. We tentatively
conclude that the issues used in estimation may be a “representative” sample, though not indicative of the worst performing sub prime pools on a national basis.

Variable Definitions and Measures

Empirically it is important to assess how far the call option to prepay and the put option to default are ‘in the money’. Given that the value of embedded options is a complex function of stochastic variables then it is difficult to measure precisely the intrinsic value of an option; and so ‘indirect’ measures are used to evaluate the likelihood of the call or put option being ‘in the money’. We follow Pennington-Cross and Chomsisengphet (2007), and use an estimate of the current loan to value ratio \((\text{currentlv})\) on a mortgage holder’s property to represent the extent to which the put option is ‘in the money’\(^3\). The more likely that the put option to default is ‘in the money’ (negative loan to value ratios) the higher the probability of a household defaulting on the mortgage debt. The descriptive statistics for this and other variables are given in Table 1.

Insert Table 1

To indicate the extent to which the call option is ‘in the money’ we again follow Pennington-Cross and Chomsisengphet and use the change in interest rates since the date of origination \((\text{libor change})\). Given that the typical index rate for sub prime mortgages is 3 month Libor we use the Libor index as the representative rate. For the UK it is expected that endogenously determined financial behaviour is more likely in the case of prepayment than default, so as a further measure of the value of a call option to prepay the standard deviation of Libor \((\text{stdlibor})\) - a moving standard deviation over 12 months) is included as an independent variable\(^4\). It is expected that the call option to prepay has a higher value when interest rate volatility is high and therefore there is an incentive to keep the option, the likelihood of prepayment then being less.

The empirical specification also includes variables that represent payment shocks that might influence default and prepayment. One measure of this is the interest rate shock, which is the change in the actual interest rates paid since the date of origination of the contract \((\text{actual_shock})\). There is no data on incomes which can be used with mortgage payments to represent the ability to pay. Given the absence of such data then the interest rate shock is used as a proxy for this, with larger shocks being more likely to lead to difficulty in paying. This may then lead to default, or

\(^3\) The Department of the Environment weighted house price index was used to update house prices to compute the current loan to value ratio.

\(^4\) The period exhibits a continuous increase in nominal house prices and low levels of volatility and negative equity (stats). For this and the other reasons stated in the text exercising the option to default is considered less likely than exercising the option to prepay.
prepayment to seek out discounts and cheaper rates. We also control for the general level of interest rates by including the current level of Libor (libor).

The data set does not contain indicators of personal characteristics, and to some extent the influence of these variables is attributed to unobserved heterogeneity in the samples. Neither was there any data available to assess the credit rating of individual borrowers, such as the FICO score used in US studies. One variable that may proxy credit worthiness is the original loan to value ratio (Pennington-Cross and Chomsisengphet, 2007). This may also proxy the nature of the household balance sheet and its riskiness (Harrison et al, 2004). Though some studies have included both the current loan to value ratio and the original level of gearing in the same specification (Pennington-Cross and Chomsisengphet, 2007), a high degree of correlation between these two measures can be problematic. The research reported here addresses this problem by including the initial size of mortgage (log initial loan balance) and the value of the property at the date of origination of the mortgage (log initial house value) as independent variables.

The key characteristics of mortgage contract design are indicated by dummy variables. Thus we have indicator variables to represent the choice of a fixed rate mortgage (fixed=1), and a discounted mortgage (discounted=1); mortgages with the standard variable rate or Libor are excluded for identification. Self-certification is also indicated by a categorical variable (selfcert=1). The review of previous research suggested that the sign on discount would be positive for both default and prepayment with the estimated parameter on discount being larger than that for fixed. Self-certification is taken as an indicator of information asymmetry and adverse selection and selfcert=1 is expected to lead to higher defaults and prepayments. Though there is an increase in the heterogeneity of contract terms over our time period, for practical reasons and compatibility with existing studies the econometric modelling uses the binary variables noted here to represent the main classifications of debt.

A further significant feature of mortgage design is the existence of the interest reset date on which discounts, or periods during which the rate of interest has been fixed, end. A dummy variable (revert) is used to represent the current mortgage month if it is within the period prior to the reset date (revert=1). Following (Ambrose et al, 2005) it is expected that both defaults and prepayments will be higher in the post reversion period. Defaults may increase because the household is more vulnerable at a higher interest rate. Prepayments may increase because post reversion the call

---

5 To this extent there is an empirical equivalence between rational prepayment and affordability shocks though the expectation might be that beyond the threshold of transactions costs the latter would be more likely linked to the extent of interest rate changes.
option is more likely to be ‘in the money’, or households may simply be augmenting their cash flow by seeking cheaper alternative mortgage deals.

V. EMPIRICAL RESULTS

The estimates for most parameters show consistent sign and magnitude across the four samples. Irregular cases are discussed where relevant, in passing. We look first of all at the parameter estimates for those variables which identify features of mortgage contract design and discuss these separately for default and then refinancing behaviour. The results are then evaluated with respect to the time varying variables that largely represent the options embedded in the mortgage and/or reflect exogenous payment shocks and affordability. The parameter estimates for each of the four samples are reported in Table 2 (a,b,c); the default equations are in Table 2 (a), Prepayment in Table 2(b) and the parameter estimates for unobserved heterogeneity in Table 2(c). The estimates represent shifts in the baseline hazard which is measured by time out from the date of origination of a mortgage (log of months).

Insert Table 2

Mortgage Contract Design and Initial Choices

With respect to mortgage contract design the likelihood of default is increased by self-certification (selfcert=1). The information problems and adverse selection that accompany the use of self-certified mortgages are a likely explanation for this observed pattern. A lower likelihood of default is expected to accompany holding a discounted mortgage, reflecting the favourable impact of teaser rates on affordability. The results for holding a discounted mortgage (discount=1) or a fixed rate contract (fixed=1) are ambiguous. So for discount the sign is negative in two cases and positive and insignificant in the other two samples. In the case of a fixed rate (fixed) contract both samples containing such mortgages (05 and 06) have parameters with negative signs though this is not statistically significant at the 5% level in sample 06. Given the short term nature of fixed rate contracts and the duration of discounts then these forms of contract may be considered close if not perfect substitutes, so that from the perspective of default there is little benefit from choosing different contract designs, unless there is a very large increase in the index rate.

Default is found to be more likely the larger the original loan (log initial loan balance) and the lower the original house price (log of initial house value); these two variables have the largest parameters (5.4218 and 4.4275 respectively in Sample 03). While larger loans might reflect a good credit rating they also bear higher servicing costs and this may explain the higher likelihood of default indicated

---

6 For estimation purposes the data has been standardised with mean zero and a standard deviation of one.
by our estimates. A low purchase price for a property (low value) may reflect other factors such as occupational status and wealth, but also indicates the possibility of less absolute value to use as collateral for further borrowing; liquidity constrained households with little or no equity in their property may have their borrowing constrained in the non housing loan market.

The likelihood of prepayment is generally increased when selfcert=1 (for 03, 05 and 06) and decreases with discount=1 (for 03, 05, 06). One possible explanation is that the higher rates of interest on self-certification may lead to risky borrowers, who do not disclose their incomes and who gradually repair their credit records, eventually seeking less expensive deals in the prime lending sector, or with a new sub prime lender. Holders of discounted mortgages who may be cash constrained may have a restricted ability to shop around for other teaser rates in the prime lending market.

The likelihood of refinancing is also increased when log initial loan balance is large, and when log initial house value is low; once again these exhibit the largest parameter values (0.5918 and 0.6846 respectively in Sample 03). Mortgage holders with larger loans can make higher absolute savings from searching for new mortgage deals. Households with lower valued houses may have less equity than those with higher valued properties, and therefore seek the release of cash through seeking more competitive mortgage deals, rather than through further borrowing. Again, the results are compatible with an emphasis by households upon cash flow and affordability.

The existence of prepayment penalties and a date at which the favourable contract rates revert back to the higher index interest rate are other important features of mortgage contract design. The dummy for post reversion decisions (revert=1) was not statistically significant in the default equations. Thus the increase in mortgage payment did not induce default. This results in part because of the increase likelihood in the competing risk, i.e. prepayment. The positive and significant coefficient on revert in the prepayment equations signals the possibility that an increase in interest payments induced re-contracting.

A comparison of parameter estimates for discount with fixed across equations (05, 06), for both default and prepayment, shows the estimated impact of having a discounted mortgage significantly greater than that for having a fixed rate contract, a result consistent with theoretical expectations.

Insert Figures 3. a to 3. d

**Hazard and Sub Hazards by self certification and reversion=1**

Figures 3.a to 3.d illustrate the effects of self certification and reversion upon the hazard rates and sub hazard rates for default (3.a and 3.b) and prepayment (3.c and 3.d). The estimates are based on
the characteristics of the longest surviving observation that can be found in Sample 03. The simulations in all cases are based upon the household choosing a fixed rate mortgage and the impact of revert is assessed for the case where selfselect=1. The hazard and sub hazard curves for default and prepayment are shifted significantly upwards by the presence of self certification. There is an increase in both defaults and refinancing post reversion although for defaults the effect is not statistically significant.

*Option Theoretic and Affordability Considerations*

The current (i.e. measured at time t) loan to value ratio (currentlv) is used to proxy the extent to which the put option to default is ‘in the money’ with the expectation is that its associated parameter will have a negative sign. For default this variable is not statistically significant at the 5% level in three of the samples. Given, that UK mortgages are debt with recourse then there is less likelihood of observing ruthless default in the United Kingdom than in the United States mortgage market. Other parameter estimates suggest that affordability may be a more critical issue for default than endogenous financial calculation. For example, the extent to which interest payments changed since origination of the mortgage contract (actual_shock) has a positive and statistically significant effect on the likelihood of default.

Change in Libor since origination is the proxy for the extent to which the call option to prepay is ‘in the money’. The parameter estimates for this variable (liborchange) are not consistent across the four samples. Only sample 04 has the expected and statistically significant negative sign. The variable stdlibor provides a further test of the option theoretic explanation of refinancing behaviour. In this case the expected negative sign is found in Sample 06 but the parameters are positive and statistically significant in Sample 03 and Sample 04. Thus the results for testing the option theoretic explanation of prepayment behaviour are ambiguous. There are also possible affordability shocks on prepayment behaviour. Large payment shocks may induce cash constrained borrowers to seek better deals. The change in interest payments actually made on the specific mortgage contract since the origination of the debt (actual_shock) has a positive and statistically significant effect upon the likelihood of prepayment across all four mortgage pools.

Credit market conditions deteriated throughout 2007 and up to mid 2008. During this period Libor increased markedly, and with it mortgage rates, reflecting changed market perceptions of risk. Volatility in this case is likely to indicate the several rises in Libor which corresponded with a marked reduction in contracts available for refinancing. Though the four samples all have lives which extend into 2008, later issues have more mortgages that are likely to be affected by these changes. Samples 05 and 06 have a positive sign on libor change suggesting that higher interest rates induced a search
for contracts with lower rates to minimise payments. There is a negative sign on \( stdlibor \) implying that tightening credit market conditions reduced prepayments, possibly through less competitive deals being available and/or access to mortgage finance being rationed. These results suggest the importance of affordability and ease of access to mortgage finance driving prepayment decisions\(^7\).

Table 2 (a,b,c) shows the corresponding estimates for the equations without control for unobserved heterogeneity. For default behaviour, the pattern of signs generally remains the same without requiring any major re-interpretation of the results. However, there are some significant differences in parameter estimates that suggest the need to control for unobserved heterogeneity. For example, the impact of initial loan balance and log of initial house value are increased in three case (03, 05 and 06) while significantly reduced in one other case (04). There was also some impact upon the parameter estimates of mortgage contract choice, so for example the control increased the impact of choosing a discounted mortgage on default for 03, 04 and 05 while changing the sign for 06. The impact upon the size of parameter for self certification is particularly marked.

The effects were similar for the prepayment equations. Table 3 summarises the observed signs on the estimates of unobserved heterogeneity and the correlation between these estimates for each equation \( d_{31}^3, d_{31}^3, d_{32}^3 \). The first point to note is the consistency of sign across the four samples. The sign on \( d_{32}^3 \) implies a negative correlation between the unobserved components of default and prepayment so that a reduction in the likelihood of prepayment increases the likelihood of default. This is compatible with the perception that the latter part of 2008 stopped credit impaired households from improving their mortgage terms and thus increasing the risk of delinquency and mortgage default.

The main contrast in results compared to previous predominantly United States research relates to the findings on defaults. The two economies have different housing finance systems, and prevalent contract designs. Consequently, while there is evidence of ‘ruthless default’ in many US studies this is not the case for our research. The current loan to value ratio is an important explanatory variable in US default research (see Pennington_Cross and Ho, 2010) but is not found to be statistically significant for this paper, possibly reflecting the continuing liability of UK defaulters for outstanding mortgage debt. A further contrast is the lack of statistical significance of the reversion of interest rates for explaining default. US research finds this to be significant at times of falling house prices (Sherlund, 2008; Pennington-Cross and Ho, 2010). The lack of statistical significance here may reflect the fact that nominal house prices did not fall over his period.

Insert Table 3
VI. CONCLUSION

This paper estimated an econometric model of loan performance on a sample of sub prime mortgages that were securitised between 2003 and 2006. The focus of the empirical analysis was the impact of variations in contract design, including self certified mortgages, and the selectivity issues surrounding initial choices. A theoretical consideration was the applicability of an option theoretic approach to the key mortgage choices of default and prepayment on United Kingdom originated sub prime debt, that is compared to the importance of affordability and exogenous influences.

A model was estimated based upon the competing risk of mortgage defaults and prepayments controlling for individual unobserved heterogeneity. The unobserved individual specific factors were modelled in a flexible general form, allowing for their influence upon the initial choice of contract type (for example self certification; discounted; fixed) and dynamic selection effects over time. The research also differed from other studies using the multinomial logit competing risk model by estimating parameters for unobserved heterogeneity and the correlation between them.

The mixing of a large number of different types of contract into pools of securitised subprime loans may be one reason for their supposed wide variability in risk and return, and presumed unpredictability. The pools of mortgages had different proportions of contract type. Without any assessment of the impact of these contract variations on household behaviour, particularly self-certified/low or zero documentation mortgages, it is not surprising that performance might be viewed as ‘idiosyncratic’.

The estimation suggested that treating mortgages as embedded option contracts did not explain the default and prepayment behaviour of the sample. Given this then the impact of reversion periods, the information asymmetry and adverse selection associated with self-certification- and only fixing interest rates on mortgage contracts for short periods- largely operated though their impact upon affordability. These affordability issues resulted in adverse effects on default and generated highly active amounts of prepayment for periods where contract choices were plentiful.

There is some evidence of variation in unobserved heterogeneity between pools of mortgages, and also significant differences in parameter estimates when unobserved heterogeneity was controlled for. Though we do not observe the risk/return behaviour of individual securities the variations in loan performance suggest that differences are most likely to arise from compositional effects resulting from having different proportions of contracts with different features and different initial choices (self certification, original loan to value ratios, reversion periods). To the extent that these considerations and unobserved heterogeneity were not taken into account when pricing
securitised bonds then the behaviour of these pools would appear idiosyncratic. There was an observed change in the behaviour of later pools of debt, possibly arising from the change in credit market conditions which restricted refinancing by liquidity constrained households. Thus securitised subprime loans may be given meaningful valuations on bank a balance sheet; that is if the effects of the composition of the securities, selectivity and the role of unobserved heterogeneity are better understood.

References

Abramowitz, M., I.A. Stegun. eds. (1972), Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables, New York: Dover Publications.


Financial Services Authority, (2010), Mortgage Market Review-Responsible Lending, CP/10/16, July.


Quercia, R., M. Stegman, W. Davis, (2005), *The Impact of Predatory Loan Terms on Subprime Foreclosures: The Special Case of Prepayment Penalties and Balloon Payments*. University of North Carolina–Chapel Hill, Center for Community Capitalism


Appendix

In this appendix we describe the E-step and the M-step of the optimisation algorithm we use to estimate the parameters of the model.

To simplify this presentation we express the complete latent likelihood for the $i^{th}$ observation as

$$L_i(\theta) = L_i(\theta, \varepsilon_1, \varepsilon_2) \phi(\varepsilon_1) \phi(\varepsilon_2),$$

where $L_i(\theta, \varepsilon_1, \varepsilon_2)$ is the likelihood of observation $i$ given $(\varepsilon_1, \varepsilon_2)$ and where $\phi(\varepsilon)$ is the normal density function. In principle the observed likelihood is deduced from A.1 by integrating the complete latent likelihood over the range of $(\varepsilon_1, \varepsilon_2)$:

$$L_i(\theta) = \int \int L_i(\theta, \varepsilon_1, \varepsilon_2) \phi(\varepsilon_1) \phi(\varepsilon_2) d\varepsilon_1 d\varepsilon_2$$

Instead of considering the continuum of possible values for $(\varepsilon_1, \varepsilon_2)$ we limit ourselves to $H$ values for $\varepsilon_i$, say $\{e_1, e_2, ..., e_H\}$ and we associate with each couple $(e_h, e_{h'})$ a positive weight $p(h, h')$ such that $\sum_{h=1}^{H} \sum_{h'=1}^{H} p(h, h') = 1$, and such that $\sum_{h=1}^{H} \sum_{h'=1}^{H} e_h p(h, h') = 0$, $\sum_{h=1}^{H} \sum_{h'=1}^{H} e_h^2 p(h, h') = 1$ and $\sum_{h=1}^{H} \sum_{h'=1}^{H} e_h e_{h'} p(h, h') = 0$. We deduce such the values $\{e_1, e_2, ..., e_H\}$ and the weights $p(h, h')$ from the Gauss-Hermite quadrature abscissas and weights for a given $H$ (see Press et al., 1986, for an introduction and Abramowitz and Stegun, 1964). We therefore approximate A.1 with

$$L_i(\theta) \approx \prod_{h=1}^{H} \left( L_i(\theta, e_h, e_{h'}) p(h, h') \right)^{\delta_i(h, h')},$$

for some $h$ and $h'$, where we define $\delta_i(h, h')$ to be equal to 1 if observation $i$ is of type $(h, h')$ and 0 otherwise. In effect we augment the observed data with $\delta_i(h, h')$ the (latent) “type” of each individual observation. From A.3, the (approximate) latent log-likelihood can now be written as
\[ \ln L_i(\theta) \approx \sum_{h=1, h' = 1}^{H} \delta_i(h, h') \{ \ln L_i(\theta, \varepsilon_h, \varepsilon_{h'}) + \ln p(h, h') \} \]  

(A.4)

For some values for the parameters, say \( \chi \), the EM algorithm process first (Expectation step) by calculating the Expected latent log-likelihood given what is observed (which we represent by \( O_i \))

\[ E_{\chi}[\ln L_i(\theta) | O_i] \equiv \ln L(\theta; \chi) \]

\[ \approx \sum_{h=1, h' = 1}^{H} E_{\chi}[\delta_i(h, h') | O_i] \{ \ln L_i(\theta, \varepsilon_h, \varepsilon_{h'}) + \ln p(h, h') \} \]  

(A.5)

where \( E_{\chi}[Z | O_i] \) evaluates the conditional expectation of a random variable \( Z \) given \( O_i \) and evaluated with the distribution parametrised by the vector \( \chi \). The key to (A.5) is the fact that \( E_{\chi}[\delta_i(h, h') \ln L_i(\theta, \varepsilon_h, \varepsilon_{h'}) | O_i] = E_{\chi}[\delta_i(h, h') | O_i] \ln L_i(\theta, \varepsilon_h, \varepsilon_{h'}) \) (given the type \((h, h')\) the value of the log-likelihood \( L_i(\theta, \varepsilon_1, \varepsilon_2) \) is constant given \( O_i \). In the present case since we specify that \( \phi(\varepsilon) \) is the standard normal distribution, the weights \( p(h, h') \) and abscissas \( \{e_1, e_2, \ldots, e_H\} \) are known (for any given \( H \)) and don’t need to be estimated (this would be the case if instead we assumed that the joint distribution \((\varepsilon_1, \varepsilon_2)\) was unknown. Then the present EM algorithm could be amended to estimate this unknown joint distribution). Hence we can evaluate \( E_{\chi}[\delta_i(h, h') | O_i] \)

from the value of \( \ln L_i(\theta, \varepsilon_h, \varepsilon_{h'}) \) for all types

\[ E_{\chi}[\delta_i(h, h') | O_i] = \frac{L_i(\chi, \varepsilon_h, \varepsilon_{h'}) p(h, h')}{\sum_{k, k'} L_i(\chi, \varepsilon_k, \varepsilon_{k'}) p(k, k')} \equiv \pi_i(h, h'; \chi) \]  

(A.6)

The second stage (Maximisation Step) we maximise with respect to \( \theta \) the Expected latent log-likelihood given what is observed and given some initial value \( \chi \) for the parameters. This procedure is repeated until convergence, where \( \theta \) becomes the next value for \( \chi \), that is until

\[ \chi = \arg \max_{\theta} \sum_{i=1}^{N} \sum_{h=1, h' = 1}^{H} \pi_i(h, h'; \chi) \{ \ln L_i(\theta, \varepsilon_h, \varepsilon_{h'}) + \ln p(h, h') \} \]  

(A.7)

The EM algorithm has good properties and if it converges it an be shown that it produces the maximum likelihood estimator (see Gouriéroux & Monfort,. 1995). The benefit of using the EM algorithm arises in practice since the objective in (A.7) can be understood as the maximum likelihood based on the latent log-likelihood but weighted by the quantities \( \pi_i(h, h'; \chi) \) (which are treated as given within each M-step).
In our context the latent log-likelihood can be decomposed into the sum of several terms each involving a different set of parameters. Hence the M-step is obtained by the separate maximisation of each of the “independent” components of the properly weighted latent log-likelihood. To illustrate this property assume that we can write:

\[ \ln L_i(\theta, \varepsilon_h, \varepsilon_{h'}) = \ln L_{i1}(\theta^1, \varepsilon_h, \varepsilon_{h'}) + \ln L_{i2}(\theta^2, \varepsilon_h, \varepsilon_{h'}) , \]

with \( \theta = (\theta^1, \theta^2) \) with \( \theta^1 \) and \( \theta^2 \) distinct. Then the M-step is amounts to two separate maximisations

\[
\theta^1_M = \arg \max_{\theta^1} \sum_{i=1}^{N} \sum_{\substack{h=1 \ldots H \atop h'=1 \ldots H}} \pi_i(h, h'; \chi) \left\{ \ln L_{i1}(\theta^1, \varepsilon_h, \varepsilon_{h'}) + \ln p(h, h') \right\},
\]

\[
\theta^2_M = \arg \max_{\theta^2} \sum_{i=1}^{N} \sum_{\substack{h=1 \ldots H \atop h'=1 \ldots H}} \pi_i(h, h'; \chi) \left\{ \ln L_{i2}(\theta^2, \varepsilon_h, \varepsilon_{h'}) + \ln p(h, h') \right\}.
\]
<table>
<thead>
<tr>
<th>Variable</th>
<th>Sample 03 n=497570</th>
<th></th>
<th>Sample 04 n=515654</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Std. Dev</td>
<td>Mean</td>
<td>Std. Dev</td>
</tr>
<tr>
<td>Exit</td>
<td>2.971532</td>
<td>0.1693367</td>
<td>2.971211</td>
<td>0.1728018</td>
</tr>
<tr>
<td>months</td>
<td>2 0.30177</td>
<td>14.45543</td>
<td>18.74752</td>
<td>12.66308</td>
</tr>
<tr>
<td>loantovalue</td>
<td>0.6955786</td>
<td>0.147224</td>
<td>0.6858394</td>
<td>0.1445353</td>
</tr>
<tr>
<td>selfcert</td>
<td>0.5765159</td>
<td>0.494111</td>
<td>0.7476021</td>
<td>0.4343887</td>
</tr>
<tr>
<td>currentlv</td>
<td>0.5311837</td>
<td>0.153348</td>
<td>0.5505707</td>
<td>0.1462581</td>
</tr>
<tr>
<td>actual_shock</td>
<td>1.133196</td>
<td>1.085864</td>
<td>1.373973</td>
<td>0.9628001</td>
</tr>
<tr>
<td>discount</td>
<td>0.8709388</td>
<td>0.335268</td>
<td>0.8820352</td>
<td>0.3225667</td>
</tr>
<tr>
<td>fixed</td>
<td>0.1234459</td>
<td>0.328947</td>
<td>0.1169214</td>
<td>0.3213269</td>
</tr>
<tr>
<td>currentlibor</td>
<td>4.313035</td>
<td>0.581544</td>
<td>4.817942</td>
<td>0.4355623</td>
</tr>
<tr>
<td>liborchange</td>
<td>0.385928</td>
<td>0.621803</td>
<td>0.8091501</td>
<td>0.4684648</td>
</tr>
<tr>
<td>revert</td>
<td>0.3794903</td>
<td>0.485266</td>
<td>0.4077017</td>
<td>0.4914077</td>
</tr>
<tr>
<td>stdlibor</td>
<td>0.2107883</td>
<td>0.1059696</td>
<td>0.2403952</td>
<td>0.1243599</td>
</tr>
<tr>
<td>Variable</td>
<td>Sample 05 n=455826</td>
<td>Sample 06 n=465386</td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------------</td>
<td>--------------------</td>
<td>--------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>Std. Dev</td>
<td>Mean</td>
<td>Std. Dev</td>
</tr>
<tr>
<td>Exit</td>
<td>2.974975</td>
<td>.1650379</td>
<td>2.981961</td>
<td>0.1412379</td>
</tr>
<tr>
<td>months</td>
<td>19.49204</td>
<td>15.99471</td>
<td>17.92071</td>
<td>15.84958</td>
</tr>
<tr>
<td>loantovalue</td>
<td>0.6991869</td>
<td>0.14986</td>
<td>0.7314664</td>
<td>0.164162</td>
</tr>
<tr>
<td>selfcert</td>
<td>0.6167397</td>
<td>0.4861814</td>
<td>0.5950136</td>
<td>0.4908899</td>
</tr>
<tr>
<td>currentlv</td>
<td>0.584231</td>
<td>0.1713933</td>
<td>0.5975107</td>
<td>0.1858507</td>
</tr>
<tr>
<td>actual_shock</td>
<td>0.5861474</td>
<td>1.06004</td>
<td>0.2322083</td>
<td>0.7776358</td>
</tr>
<tr>
<td>discount</td>
<td>0.4997236</td>
<td>0.5000005</td>
<td>0.3423244</td>
<td>0.474488</td>
</tr>
<tr>
<td>fixed</td>
<td>0.3742437</td>
<td>0.4839275</td>
<td>0.6202851</td>
<td>0.4853164</td>
</tr>
<tr>
<td>currentlibor</td>
<td>4.95956</td>
<td>0.5164024</td>
<td>5.168357</td>
<td>0.6255065</td>
</tr>
<tr>
<td>liborchange</td>
<td>0.0623435</td>
<td>0.5970122</td>
<td>0.4107663</td>
<td>0.8862877</td>
</tr>
<tr>
<td>revert</td>
<td>0.4647256</td>
<td>0.4987547</td>
<td>0.7453705</td>
<td>0.4356532</td>
</tr>
<tr>
<td>stdlibor</td>
<td>0.2083212</td>
<td>0.1276791</td>
<td>0.2629969</td>
<td>0.1319354</td>
</tr>
</tbody>
</table>
Figure 1
Constant Prepayment Rates

Figure 2
Constant Default Rates
Figure 3.a  
Default Hazard Rates: Impact of Self Certification

Figure 3.b  
Prepayment Hazard Rates: Impact of Self Certification
### Table 2 (a)

#### Parameter Estimates: Competing Risk Model With and Without Unobserved Heterogeneity

<table>
<thead>
<tr>
<th></th>
<th>Sample 03</th>
<th>Sample 04</th>
<th>Sample 05</th>
<th>Sample 06</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Simple Logit</td>
<td>Simple Logit</td>
<td>Simple Logit</td>
<td>Simple Logit</td>
</tr>
<tr>
<td></td>
<td>Logit Model with Heterogeneity</td>
<td>Logit Model with Heterogeneity</td>
<td>Logit Model with Heterogeneity</td>
<td>Logit Model with Heterogeneity</td>
</tr>
<tr>
<td><strong>Default Equation</strong></td>
<td>-----------</td>
<td>-----------</td>
<td>-----------</td>
<td>-----------</td>
</tr>
<tr>
<td><strong>Sample 03</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Sample 04</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Sample 05</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Sample 06</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Constant</strong></td>
<td>-8.6623</td>
<td>-10.714</td>
<td>-8.0498</td>
<td>-9.6469</td>
</tr>
<tr>
<td></td>
<td>0.1439</td>
<td>1.0437</td>
<td>0.119</td>
<td>0.113</td>
</tr>
<tr>
<td><strong>Log months</strong></td>
<td>1.1542</td>
<td>2.1882</td>
<td>1.8909</td>
<td>3.256</td>
</tr>
<tr>
<td></td>
<td>0.2698</td>
<td>0.7924</td>
<td>0.2555</td>
<td>0.7023</td>
</tr>
<tr>
<td><strong>Year dummy</strong></td>
<td>-0.1067</td>
<td>-1.1286</td>
<td>0.0621</td>
<td>-0.1196</td>
</tr>
<tr>
<td></td>
<td>0.0769</td>
<td>0.0863</td>
<td>0.096</td>
<td>0.0551</td>
</tr>
<tr>
<td><strong>Actual shock</strong></td>
<td>0.8649</td>
<td>0.7577</td>
<td>0.3518</td>
<td>0.3176</td>
</tr>
<tr>
<td></td>
<td>0.2842</td>
<td>0.2718</td>
<td>0.1703</td>
<td>0.134</td>
</tr>
<tr>
<td><strong>Current tv</strong></td>
<td>0.1453</td>
<td>-0.2476</td>
<td>0.1882</td>
<td>-0.104</td>
</tr>
<tr>
<td></td>
<td>0.1728</td>
<td>0.2154</td>
<td>0.1203</td>
<td>0.1624</td>
</tr>
<tr>
<td><strong>Libor</strong></td>
<td>-0.3091</td>
<td>-0.2177</td>
<td>0.1947</td>
<td>-1.1309</td>
</tr>
<tr>
<td></td>
<td>0.3421</td>
<td>0.4214</td>
<td>0.1588</td>
<td>0.581</td>
</tr>
<tr>
<td><strong>Libor change</strong></td>
<td>-0.1176</td>
<td>-0.0817</td>
<td>-0.6489</td>
<td>1.5229</td>
</tr>
<tr>
<td></td>
<td>0.2963</td>
<td>0.3408</td>
<td>0.1521</td>
<td>0.5742</td>
</tr>
<tr>
<td><strong>Revert</strong></td>
<td>0.0852</td>
<td>0.059</td>
<td>-0.0229</td>
<td>-0.1102</td>
</tr>
<tr>
<td></td>
<td>0.1072</td>
<td>0.1105</td>
<td>0.0909</td>
<td>0.0671</td>
</tr>
<tr>
<td><strong>Std libor</strong></td>
<td>0.0069</td>
<td>-0.0188</td>
<td>-0.2761</td>
<td>-0.2462</td>
</tr>
<tr>
<td></td>
<td>0.078</td>
<td>0.0906</td>
<td>0.1054</td>
<td>0.1001</td>
</tr>
<tr>
<td><strong>Self cert</strong></td>
<td>0.239</td>
<td>0.0732</td>
<td>0.0516</td>
<td>0.2951</td>
</tr>
<tr>
<td></td>
<td>0.0485</td>
<td>0.1015</td>
<td>0.1001</td>
<td>0.0449</td>
</tr>
<tr>
<td><strong>Discount</strong></td>
<td>-0.0339</td>
<td>-0.3408</td>
<td>0.0752</td>
<td>-0.1882</td>
</tr>
<tr>
<td></td>
<td>0.1207</td>
<td>0.1661</td>
<td>0.0928</td>
<td>0.1224</td>
</tr>
<tr>
<td><strong>Fixed</strong></td>
<td>_</td>
<td>_</td>
<td>_</td>
<td>-0.1664</td>
</tr>
<tr>
<td></td>
<td>_</td>
<td>_</td>
<td>_</td>
<td>0.0949</td>
</tr>
<tr>
<td><strong>Log initial loan</strong></td>
<td>2.7643</td>
<td>0.4609</td>
<td>2.8048</td>
<td>3.354</td>
</tr>
<tr>
<td></td>
<td>5.4218</td>
<td>0.9801</td>
<td>0.3773</td>
<td>0.43</td>
</tr>
<tr>
<td><strong>Log initial house value</strong></td>
<td>-2.4653</td>
<td>0.4366</td>
<td>-4.4275</td>
<td>-1.9789</td>
</tr>
</tbody>
</table>
Table 2 (b)
Prepayment Equation

<table>
<thead>
<tr>
<th></th>
<th>Sample 03</th>
<th>Sample 04</th>
<th>Sample 05</th>
<th>Sample 06</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Simple Logit</td>
<td>Logit Model with Heterogeneity</td>
<td>Simple Logit</td>
<td>Logit Model with Heterogeneity</td>
</tr>
<tr>
<td></td>
<td><strong>β</strong> s.e.</td>
<td><strong>β</strong> s.e.</td>
<td><strong>β</strong> s.e.</td>
<td><strong>β</strong> s.e.</td>
</tr>
<tr>
<td>Constant</td>
<td>-4.0488 0.0152</td>
<td>-4.2091 0.0315</td>
<td>-3.9845 0.0131</td>
<td>-5.2621 0.1032</td>
</tr>
<tr>
<td>Log months</td>
<td>1.4932 0.0384</td>
<td>1.9814 0.0824</td>
<td>0.7093 0.0283</td>
<td>-2.416 0.1345</td>
</tr>
<tr>
<td>Year dummy</td>
<td>-0.0439 0.0108</td>
<td>-0.0435 0.0152</td>
<td>-0.0203 0.0151</td>
<td>-0.0133 0.0117</td>
</tr>
<tr>
<td>Actual shock</td>
<td>0.3763 0.032</td>
<td>0.3514 0.0356</td>
<td>0.3359 0.0268</td>
<td>0.0947 0.024</td>
</tr>
<tr>
<td>Currentiv</td>
<td>0.0924 0.0254</td>
<td>-0.0218 0.0463</td>
<td>0.0583 0.0237</td>
<td>-0.4639 0.0779</td>
</tr>
<tr>
<td>Libor</td>
<td>-0.811 0.0503</td>
<td>-0.7227 0.068</td>
<td>0.1449 0.0296</td>
<td>0.2586 0.0807</td>
</tr>
<tr>
<td>Libor change</td>
<td>0.1148 0.043</td>
<td>0.1081 0.0616</td>
<td>-0.0954 0.0298</td>
<td>-0.3016 0.085</td>
</tr>
<tr>
<td>Revert</td>
<td>0.1048 0.0104</td>
<td>0.092 0.0109</td>
<td>0.1473 0.009</td>
<td>0.1517 0.0105</td>
</tr>
<tr>
<td>Std Libor</td>
<td>0.1646 0.0116</td>
<td>0.1299 0.0128</td>
<td>-0.1027 0.0159</td>
<td>0.0928 0.0194</td>
</tr>
<tr>
<td>Self cert</td>
<td>0.0493 0.0095</td>
<td>0.1134 0.0162</td>
<td>0.0418 0.0092</td>
<td>-0.0019 0.0674</td>
</tr>
<tr>
<td>Discount</td>
<td>-0.1313 0.0141</td>
<td>-0.2453 0.0271</td>
<td>-0.1373 0.0135</td>
<td>-0.0289 0.0587</td>
</tr>
<tr>
<td>Log initial loan</td>
<td>0.2493 0.0482</td>
<td>0.5918 0.2355</td>
<td>0.1281 0.0479</td>
<td>0.1456 0.2665</td>
</tr>
<tr>
<td>Log initial house value</td>
<td>-0.3272 0.0442</td>
<td>-0.6846 0.158</td>
<td>-0.2176 0.0419</td>
<td>-0.8393 0.1549</td>
</tr>
<tr>
<td>Parameter</td>
<td>Estimate 1</td>
<td>Estimate 2</td>
<td>Estimate 3</td>
<td>Estimate 4</td>
</tr>
<tr>
<td>-----------</td>
<td>------------</td>
<td>------------</td>
<td>------------</td>
<td>------------</td>
</tr>
<tr>
<td>$d_{21}$</td>
<td>-0.711</td>
<td>0.1373</td>
<td>-0.711</td>
<td>0.1373</td>
</tr>
<tr>
<td>$d_{31}$</td>
<td>-0.6929</td>
<td>0.1347</td>
<td>-0.6929</td>
<td>0.1347</td>
</tr>
<tr>
<td>$d_{3}$</td>
<td>-0.7062</td>
<td>0.4245</td>
<td>-0.7062</td>
<td>0.4245</td>
</tr>
</tbody>
</table>

Notes:
The data has been standardised (with mean 0 and variance 1), the parameters correspond to the standardised data. The parameters presented here are Maximum Likelihood parameters. The parameter for the year dummy correspond to a dummy variable for the "early" year in each issue (for example in issue 3, the year dummy corresponds to the year 2002, 2003 for issue 4 etc...). In the case of the Logit Model with Heterogeneity, the EM algorithm is used to obtained the parameter estimates. When it is possible we improve the EM algorithm with Newton-Raphson updates using the expected Hessian of the observed likelihood. The Standard Error presented here are calculated using the expected Hessian of the observed likelihood.
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Pool One</th>
<th>Pool Two</th>
<th>Pool Three</th>
<th>Pool Four</th>
</tr>
</thead>
<tbody>
<tr>
<td>$d_{31}$</td>
<td>&lt;0</td>
<td>&lt;0</td>
<td>&lt;0</td>
<td>&lt;0</td>
</tr>
<tr>
<td>$d_{32}$</td>
<td>&lt;0</td>
<td>&lt;0</td>
<td>&lt;0</td>
<td>&lt;0</td>
</tr>
<tr>
<td>$d_3$</td>
<td>&gt;0</td>
<td>&gt;0</td>
<td>&gt;0</td>
<td>&gt;0</td>
</tr>
<tr>
<td>$r$</td>
<td>&lt;0</td>
<td>&lt;0</td>
<td>&lt;0</td>
<td>&lt;0</td>
</tr>
</tbody>
</table>