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

Gauthier Lanot
Economics Department, Umea University, Sweden
Contact: e-mail gauthier.lanot@econ.umu.se

David Leece (contact author)
Keele Management School, Keele University, United Kingdom.
Contact: e-mail d.leece@keele.ac.uk

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 and empirical 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. The paper demonstrates the need for researchers and practitioners to jointly estimate household choices whiles controlling for selectivity through unobserved heterogeneity.

JEL Codes: G21 C13 C25 C51 D10 D14 E44

Key Words: Subprime mortgages, unobserved heterogeneity, household behaviour, loan performance

Introduction

The (arguably) 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 UK mortgages originated and held over a more recent time period (2001-2008). Specifically the research tests the proposition that the duration of mortgage contracts, through a process of household self selection, varies according to the characteristics of the contract chosen. Moreover we establish the importance of unobserved heterogeneity in this process of self selection.

To our knowledge, there is little econometric work on subprime loan performance for the United Kingdom which would be 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 significant institutional differences between United States and United Kingdom mortgage markets lead to variations in the contract choices available to households and possibly different default and prepayment behaviour. Non-recourse mortgages in the United Kingdom mean that rational/selfish default is less likely as the defaulting borrower is still liable to any outstanding debt. This requires an empirical specification which accounts for affordability concerns while allowing for option theoretic influences. Secondly, the funding of UK mortgages is largely through retail deposits, and or short term interest rate swaps, leading to a prevalence of variable rate mortgages and mortgages where interest rates are fixed for a short period of time only (see Miles, 2004). Together with self-certified mortgages which minimise documentation and proof of ability to pay, the institutional framework constrains the range of mortgage choices available in the United Kingdom to largely short run interest rate fixes; typically spanning 1 to 5 years (see Miles, 2004, 2005; Leece 2000, 2004). It is of interest therefore to analyse the selectivity issues surrounding default and prepayment behaviour where households are likely to be focussed upon affordability and where the mortgage contract choices are characterised by heavy discounting (teaser rates) and interest rates fixed for a short duration..

The securitization of mortgage debt involved the pooling of many contract types; this is true for example in the UK data we are studying. The absence of precise knowledge of the self selection of individuals among these contracts may lead to unexpected and unexplained variations in loan performance. Finkelstein and Poterba (2004) in their seminal study on annuity contracts note how the choice of characteristics reflects private information and attitudes to risk. In the annuity case the ability for an individual to alter her life time (here the risk for the issuer derives from longer lives) is presumably small and therefore observed contract characteristics arguably captures all the information asymmetry. In the context of mortgage contracts, individuals have a greater ability to control the date and manner of their exit (through default or prepayment) and hence will determine the initial contract characteristics in order to reduce the costs of these later choices. In an earlier paper Dunn and Spatt (1988) discuss the links between mortgage contract design and expected prepayment, and observes the presence of private information, self selection and clientele effects in mortgage markets; using UK data, Leece (2004) also highlights the interdependence of choices in mortgage market. Elliehausan and Hwang (2010) note that the sub prime mortgage market provides a unique opportunity to study the choices of mortgage rate type over a range of risk characteristics and note the higher rates of default for adjustable rate mortgages. In further theoretical work Campbell and Cocco (2012) find that the choice of type of mortgage influences default behaviour.

In a departure from the existing literature, the econometric model presented in this paper accounts for the effect of unobserved heterogeneity arising out of contract choices and other sources of private information on the duration to/probability of, default or prepayment (see Nichols et al, 2005). The implications of these findings are firstly that the characteristics of mortgage contracts were neglected in the pooling and securitisation of UK subprime mortgage debt and secondly that there is evidence of adverse selection and selectivity bias arising from significant unobserved heterogeneity.

The paper begins with a review of the academic literature on the default and prepayment behaviour of mortgage loans and subprime loans where applicable. This is followed by an outline of the econometric methodology and the modelling approach to unobserved heterogeneity and estimation. The sample and the empirical specification of the model are discussed in the section which follows. Parameter estimates are then reported and discussed. The paper concludes with a summary and the implications of the findings.

Household Mortgage Choices

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 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 in the literature is largely United States specific, since 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 where the interest rate is discounted, but eventually reverts to a 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). This work has focused on studying loan level data where borrower characteristics can be analysed. Furthermore, the apparent persistence of sub optimal prepayment and default decisions has encouraged the modelling 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. Furthermore, 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 exogenous shocks.

Mortgage Design

There are several prevalent types of United Kingdom mortgage contract that feature in the data used for this research. The fixed rate contract typically has the interest rate fixed for one to three years, with some longer term contracts. UK variable rate contracts have interest rates that change at irregular periods depending upon the prevailing rate of interest (LIBOR or Base Rate). US fixed rate contracts have the interest rate fixed for longer periods (15 or 30 years) and the adjustable rate contract is the equivalent of the UK one year fixed rate. Those

UK contracts which have the rate fixed for 2 to 3 years are equivalent to the so called 'hybrid mortgages' in the US (see Ambrose et al, 2005).

The US literature argues that alternative mortgage contracts may elicit different patterns of behaviour. For example, there is evidence 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, though this has yet to be established.

Empirically the real estate economics and finance literature has examined the effects of different aspects of contract design on the performance of mortgage loans (Pennington-Cross et al, 2010; Phillips et al, 1996; VanderHoff, 1996; Green and Shilling, 1997; Ambrose and LaCour-Little, 2001). 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 statistically significant increase in prepayments at interest rate adjustment dates.

A further key feature of mortgage contracts, within the sample period, has been low or zero documentation mortgages in the US, and its equivalent, the self certified mortgage in the UK. In 2008, 52% of all new mortgages were self certified (Financial Services authority, 2010). 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 is expected to have a positive effect on the predictability of defaulting.

Rose (2008) 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 and finds that the effects of variables on foreclosure indeed depend upon loan features such as the level of documentation.

Research to date has typically controlled for different contract designs by analysing loans of a given type (for example long term fixed or adjustable rate mortgages). Even analysing one type of loan raises issues regarding selectivity (i.e. the borrowers associated with a particular type of loan may share observed or unobserved characteristics which may increase a specific risk). When several types of mortgage are pooled in the same security it is likely that the heterogeneity of these specific risks will increase. Here we address 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.

Modeling Framework

The use of econometric techniques in the mortgage loan literature has evolved from the use of limited dependent variable models (Probit) 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 control of unobserved heterogeneity when using CPH has been based upon the estimation of discrete mass points (see Pennington-Cross et al, 2010).

The econometric methodology reported here differs in several ways. 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 arguably more flexible. The model 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/expensive and is not available in commercial software (Clapp et al, 2006). The approach reported in this paper speeds up estimation and facilitates convergence of the likelihood function by using the methodology developed by Train (2008) (see the online technical appendix for a description).

Overview of the Econometric Model

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}, here c_i is 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 rate mortgages.

Finally, we account for the unobserved differences between individuals which may affect both "exit" decisions independently or jointly. Hence we denote $\varepsilon_i \equiv (\varepsilon_{1i}, \varepsilon_{2i})$ the vector of unobserved individual factors. We assume that the marginal joint distribution of ε_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 of the Econometric Modelling

We describe the general assumptions we maintain to estimate the parameters of interest. In general, given a list of initial exogenous variables, say w_{i0} , we can always express the probability density of a given history as a product of conditional densities:

$$
\begin{aligned}
& f\left(\varepsilon_i, c_i, x_{i0}, \{x_{it}\}_{t=1}^{T_i}, \{d_{it}\}_{t=1}^{T_i} \mid w_{i0}\right) = \\
& f_{\varepsilon|w_0}\left(\varepsilon_i \mid w_{i0}\right) f_{c,x_0|w_0,\varepsilon}\left(c_i, x_{i0} \mid w_{i0}, \varepsilon_i\right) f_{x|c,x_0,w_0,\varepsilon}\left(\{x_{it}\}_{t=1}^{T_i} \mid c_i, x_{i0}, w_{i0}, \varepsilon_i\right) \qquad (1) \\
& f_{d|x,c,x_0,w_0,\varepsilon}\left(\{d_{it}\}_{t=1}^{T_i} \mid \{x_{it}\}_{t=1}^{T_i}, c_i, x_{i0}, w_{i0}, \varepsilon_i\right),
\end{aligned}
$$

This decomposes the joint density of all the quantities of interest into a product of marginal and conditional densities. We denote by f_s each density, and the subscript indicates the identity (its argument and the set of conditioning variables) of each density. For example, $f_{d|x,c,x_0,w_0,\varepsilon}$ is the density for the type of exit conditional on other choices, a vector of independent variables, and the unobserved differences.

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

$$
f_{x|c,x_0,w_0,\varepsilon}\left(\{x_{it}\}_{t=1}^{T_i} \mid c_i, x_{i0}, w_{i0}, \varepsilon_i\right) = f_{x|c,x_0,w_0}\left(\{x_{it}\}_{t=1}^{T_i} \mid c_i, x_{i0}, w_{i0}\right), \qquad (2)
$$

hence we assume the information contained in the individual specific effect ε_i is captured by the initial value of the characteristics of the loan, x_{i0} , and the type of loan c_i. This assumption is plausible and 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 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 loan contracts (see equation 1) 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 the model determines 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 allows for a more flexible correlation structure between the diverse elements of the model. In particular it is more flexible than a specification (with the factor loadings for the individual specific effect set to 0) which would rely only on conditional independence.

The density of the observed characteristics of the loan, that is the amount borrowed and the value of the asset, is assumed to be jointly normally distributed with a mean vector depending linearly on ε_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 ε_i). Formally we assume:

$$x_{i0} \mid w_{i0}, \varepsilon_i \sim N\left(w_{i0}\mathrm{K}^0 + \varepsilon_{1i}\Delta_1^0 + \varepsilon_{2i}\Delta_2^0, \Sigma\right) , \qquad (3)$$

where K^0 is a parameter matrix and Δ_1^0 and Δ_2^0 are parameter vectors conformable to the dimensions of x_{i0} . Σ is the variance-covariance matrix for x_{i0} given w_{i0} and ε_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:

$$f_{c_1|x_0,w_0,\varepsilon}\left(0\,|\,x_{i0},w_{i0},\varepsilon_i\right)=\frac{1}{1+\exp\left(-\left(x_{i0}\alpha_0^1+w_{i0}\kappa_0^1+\delta_1^1\varepsilon_{1i}+\delta_2^1\varepsilon_{2i}\right)\right)}\ , \qquad (4)$$

$$f_{c_1|x_0,w_0,\varepsilon}\left(1\,|\,x_{i0},w_{i0},\varepsilon_i\right)=1-f_{c_1|x_0,w_0,\varepsilon}\left(0\,|\,x_{i0},w_{i0},\varepsilon_i\right)\ .$$

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

$$f_{c_2|c_1,x_0,\varepsilon}\left(c_2\,|\,c_{1i},x_{i0},w_{i0},\varepsilon_i\right)=\frac{e^{Z_i(c_2)}}{\sum_{k=1}^{3}e^{Z_i(k)}}\ \text{, with } c_2 \text{ in } \{1,2,3\}, \qquad (5)$$

Where:

$$\begin{aligned}
&Z_i(1)=0,\\
&Z_i(2)=\lambda_2^2c_{1i}+x_{i0}\alpha_2^2+w_{i0}\kappa_2^2+\delta_{21}^2\varepsilon_{1i}+\delta_{22}^2\varepsilon_{2i}, \qquad (6)\\
&Z_i(3)=\lambda_3^2c_{1i}+x_{i0}\alpha_3^2+w_{i0}\kappa_3^2+\delta_{31}^2\varepsilon_{1i}+\delta_{32}^2\varepsilon_{2i}.
\end{aligned}$$

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:

$$f_{d|x,c,x_0,w_0,\varepsilon}\left(\{d_{it}\}_{t=1}^{T_i} \mid \{x_{it}\}_{t=1}^{T_i}, c_i, x_{i0}, w_{i0}, \varepsilon_i\right) = \prod_{t=1}^{T_i} f_{d|x,c,x_0,w_0,\varepsilon}\left(d_{it} \mid \{x_{it}\}_{t=1}^{T_i}, c_i, x_{i0}, w_{i0}, \varepsilon_i, t\right)$$
$$= \prod_{t=1}^{T_i} f_{d|x,c,x_0,w_0,\varepsilon}\left(d_{it} \mid x_{it}, c_i, x_{i0}, w_{i0}, \varepsilon_i, t\right). \quad (7)$$

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

$$f_{d|x,c,x_0,w_0,\varepsilon}\left(d \mid x_{it}, c_i, x_{i0}, w_{i0}, \varepsilon_i, t\right) = \frac{e^{V_{it}(d)}}{\sum_{k=1}^{3} e^{V_{it}(k)}}, \text{ with } d \text{ in } \{1,2,3\}, \quad (8)$$

where:

$$\begin{aligned} V_{it}(1) &= 0, \\ V_{it}(2) &= x_{it}\beta_2^3 + c_{2i}\gamma_2^3 + \lambda_2^3 c_{1i} + x_{i0}\alpha_2^3 + w_{i0}\kappa_2^3 + \delta_{21}^3\varepsilon_{1i}, \\ V_{it}(3) &= x_{it}\beta_3^3 + c_{2i}\gamma_3^3 + \lambda_2^3 c_{1i} + x_{i0}\alpha_3^3 + w_{i0}\kappa_3^3 + \delta_{31}^3\varepsilon_{1i} + \delta_{32}^3\varepsilon_{2i}. \end{aligned} \quad (9)$$

The parameters of interest which appear in the conditional probabilities are therefore $\left(\mathrm{K}^0, \Sigma, \alpha_0^1, \alpha_k^j, \lambda_k^j, \kappa_0^1, \kappa_k^j, \beta_k^3, \gamma_k^3, \Delta_1^0, \Delta_2^0, \delta_{k1}^j, \delta_{k2}^j\right)$ with *j=2,3* and *k=2,3*. The parameters $(\Delta_1^0, \Delta_2^0, \delta_{k1}^j, \delta_{k2}^j)$ 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 ε_i are uncorrelated; this is however only a matter of presentation. Indeed, we are always able to define the parameters δ_{31}^3 and δ_{32}^3 as functions of an unrestricted parameter δ^3 and a correlation coefficient ρ as follows:

$$\delta_{31}^3 = \delta^3 \rho \text{ and } \delta_{32}^3 = \delta^3 \left(1-\rho^2\right)^{1/2} . \tag{10}$$

The term $\delta_{31}^3 \varepsilon_{1i} + \delta_{32}^3 \varepsilon_{2i}$ 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 ε_{1i} . We therefore have the correspondence:

$$\begin{aligned}
&\delta_{31}^3 > 0, \delta_{32}^3 > 0 \Leftrightarrow \delta^3 > 0, \rho > 0; \\
&\delta_{31}^3 > 0, \delta_{32}^3 < 0 \Leftrightarrow \delta^3 < 0, \rho < 0; \\
&\delta_{31}^3 < 0, \delta_{32}^3 > 0 \Leftrightarrow \delta^3 > 0, \rho < 0; \\
&\delta_{31}^3 < 0, \delta_{32}^3 < 0 \Leftrightarrow \delta^3 < 0, \rho > 0.
\end{aligned} \tag{11}$$

The ratio of $\delta_{31}^3 / \delta_{32}^3$ is an increasing function of ρ only and therefore inverting it gives an estimate of ρ , and given this estimate it is then straightforward to obtain an estimate for the parameter δ^3 (applying the delta method to the transformation given the precision for δ_{31}^3 and δ_{32}^3 will provide an easy way to obtain the precision for ρ and δ^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 deals with 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. For our model the advantage rests in the fact that both the E-step (the Expectation step, E-step, entails the prediction of the probability to observe a given value of the unobserved heterogeneity components) and the M-step (the maximisation step, M-step, follows on from the E-step and maximises the latent log-likelihood) are relatively straightforward. Furthermore, 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. 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 (2008).

To summarise the modelling approach, we estimate the parameters of the model jointly. We estimate the parameters of the initial characteristics of the loan and of the value of the house, the parameters of the logit model describing the self certification status as well as the parameters of the logit model describing the kind of interest rate deal the mortgage holder choses, and finally we estimate the parameters of the logit model which describes the default and repayment behaviour.

Data and Empirical Model

The Data

The data covers 60,000 mortgage contracts, and contains approximately 1.75 million observations (individual × 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[1].

[1] Confidentiality is further 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

Subprime mortgage holders in the United Kingdom are generally defined as those borrowers who are unable to obtain a loan from high street banks or Building Societies, who have had poor credit histories or received County Court judgements (bankruptcy). In this study the classification of which loans were non-conforming/subprime was undertaken by the securitising bank following recommendations from the lender.

The mortgages are classified by date of 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 calendar period over which the performance of the loans is analysed runs from January 2003 for the earlier issue to the end of May 2008 for all issues. 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.

We remove from the analysis the comparatively small number of prime and near prime loans (7.89% and 4.32% of total observations respectively) concentrated in particular issues. Buy to let mortgages - typically second loans for the purchase of property to rent out - (8.42% of observations but with 5.08% 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 as well as the distinct economic conditions which affect each issue, 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

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.

assisted estimation and allowed a comparison between the parameter estimates for mortgages issued across different periods of time.

Insert Table 1

Table 1 presents the basic descriptive statistics of the sample and it 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. Overall we are able to model all contract choices though in some samples the incidence of particular types of contract was low or zero. Hence the contract choices we model depend on the particular sub sample.

The investment bank selected the loans of this particular mortgage originator for analysis. Therefore we are concerned that the pools may be subject to selectivity/originator bias (see Ciochetti et al, 2003). Unfortunately data is unavailable to directly correct for this potential bias. 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 in loan performance by issue and between the yearly grouping of issues. This variation in performance can be observed in Figure 1 and Figure 2 which show conditional prepayment and conditional default rates by issue over time (monthly). We observe that both observed prepayment and default rate vary between issues over time: Figure 1 suggests that default rate among surviving mortgages have increased in aggregate with later issues, while Figure 2 suggests that prepayment takes place around some specific dates (i.e. the annual anniversary at 13, 25 and 37 months of the date of the initial contract completion) and later issues have higher prepayment rates among surviving mortgages than earlier issues. Figure 3 shows that these jumps in the repayment rates correspond to the specific dates when the individual mortgage contract reverts to the default variable rate. In a further regression analysis of the default and repayment aggregate rates, we show that the proportion of mortgages reaching a reversion date does not explain default but that the specific issues do (as expected later issues have on average higher default rates). The same analysis shows that prepayment rates are not explained by the year of issue but are significantly explained by the proportion of mortgages reaching a reversion date (see Table A1 in the appendix for full details).

Insert Figure 1, Figure 2 and Figure 3

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 default rate 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 this 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 (*current loan/value ratio*) on a mortgage holder’s property to represent the extent to which the put option is ‘in the money’[2]. The more likely the put option to default is ‘in the money’ (loan to value ratios greater than one), the higher the probability of a household defaulting on the mortgage debt.

To indicate the extent to which the call option is ‘in the money’ we again follow Pennington-Cross and Chomsisengphet (2007) and use the change in interest rates since the date of origination (*libor change*). 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. Given that the typical index rate for subprime 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 (*std dev libor*-a moving standard deviation over 12 months) is included as an independent variable[3].

It is important to note the difference here between United States and United Kingdom contract designs as this determines the empirical specification of the model and the expected results. The United Kingdom mortgage market is dominated by variable rate debt or mortgages with interest rates fixed for short periods of time; compared to a United States market were longer term fixed rates are more prevalent. This might be thought to limit the effect of interest rate changes in the UK on the likelihood of prepayment. However, even

---

[2] The Department of the Environment (2008 and later) weighted house price index was used to update house prices to compute the current loan to value ratio. See table 507, https://www.gov.uk/government/statistical-data-sets/live-tables-on-housing-market-and-house-prices

short term fixes (typically 3 years) can lead to prepayment behaviour. In addition the practice of heavy discounting (teaser rates) may lead to refinancing just prior to interest reset dates.

The empirical specification also includes variables that represent payment shocks that might influence default and prepayment. For example, we measure the interest rate shock, that is the change in the *actual* interest rates paid since the date of origination of the contract (*actual shock*). There is no data on household incomes to directly measure the ability to pay. The interest rate shock is likely to have a negative impact upon the ability to pay. It differs from changes in the general level of interest rates as it is a function of the initial interest rate, the length of the interest rate agreement, whether fixed or variable and the post reversion interest rate premium which varied by risk assessment. Unfavourable changes via the actual shock may then lead to default, or 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*).

In the data for the first two issues, the relationship between *libor* and *libor change* given the completion month dummies is nearly perfect, whereas for the last two issues while we can explain a large proportion of the libor rate given the completion month dummies the explanation is not longer nearly exact. Furthermore for the first two issues the actual interest shock is also almost perfectly explained in a regression on changes to the libor rate (*libor change*) and the completion dummies (R2 greater than 0.9), while for the last two issues the same regression has less explanatory power (but still large with R2 greater than 0.7). To resolve the co-linearity issue we estimate the competing risk models conditioning on *libor* only for the first two issues. For the last two issues, we estimate the competing risk model conditioning on *actual shock*, the *libor* and *libor change*.

We do not observe any personal characteristics, and to some extent the influence of these variables can be attributed to unobserved heterogeneity. Neither is there any data available to assess the credit rating of individual borrowers, such as the FICO score used in US studies.

The initial loan to value ratio is often included in studies to reflect the willingness to pay. Though some researchers have included both the current loan to value ratio and the original level of gearing in the same specification (Pennington-Cross and Chomsisengphet, 2007), for our study the high degree of correlation between these two measures is problematic. This problem was overcome by replacing the initial loan to value ratio with both the initial size of mortgage (*log initial loan value*) and the value of the property at the date of origination of the mortgage (*log initial house value*) as independent variables. Importantly these variables correspond to the initial characteristics of the loan signified by x_{i0} in the econometric modelling. This specification using absolute values is also more consistent with the perspective on affordability required of United Kingdom studies.

The key characteristics of mortgage contract design signified by c_i in the econometric modelling 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 (*discount*=1); mortgages with the standard variable rate or Libor are the excluded category. 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 we expect that self-certification leads to higher defaults and prepayments.

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) we expect that both defaults and prepayments are 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 option is more likely to be 'in the money', or households may simply be augmenting their cash flow by seeking cheaper alternative mortgage deals. Thus we cannot control for the ability to pay as used in most United States studies of default but rather we proxy these effects through payment shocks. It is also the case that the "securitiser" and holders of this mortgage debt could not directly observe these variables when attempting to forecast mortgage termination and price the securitised debt.

Regarding the estimation of models for the initial contract choices we do not observe many independent covariates, and so we control only for dummy variables which indicate the month of completion of the mortgage contract (i.e. in that particular month the lender made the money available to the borrower). The coverage in time of each sample varies, hence the number of months dummies varies between samples. In sample 06 the issue extends into the next calendar year, so that we include a 13th month dummy. The estimations control for other contract choices and for unobserved heterogeneity (see Tables A2 and A3 in the appendix).

The model specification assumes that the joint distribution of the unobserved heterogeneity components is normal. Hence we need to establish the identification of the parameters of the model which explain the effects of the components of unobserved heterogeneity on the initial contract choices and on the durations until exit. To a first approximation, these parameters are identified by the covariance between the (endogenous) characteristics of the initial conditions given the exogenous variables (in our case, the months of contract completion) and the covariance between the durations until exit given the exogenous variables and the (predetermined) initial conditions (i.e. the size of the loan, the value of the property, etc). We can then show that all the parameters of the model are then identified if we observe exits to each destination (prepayment or default) in at least two periods after the initial contract choice[4].

Empirical Results

To be consistent with the focus of the paper we look first at the parameter estimates for those variables which identify features of mortgage contract design and discuss these separately for their impact upon default and then refinancing behaviour. The results are then discussed with respect to the time varying variables that largely represent the options embedded in the mortgage and/or reflect exogenous payment shocks and affordability. We then consider the impact of unobserved heterogeneity on the above exit probabilities and on the jointly estimated initial mortgage contract choices.

After experimentation with various flexible forms of the time trend, shifts in the baseline hazard were best captured by log linear time from the date of origination of a mortgage (*log of months*). The parameter estimates for default and prepayment equations covering each of the four samples are reported in Table 2 [5]. We present the estimation results for the default and prepayment equation with and without accounting for unobserved heterogeneity. However, the evidence summarised in Table 3 clearly suggests that the restriction to the model without unobserved heterogeneity is rejected by the data since the likelihood ratio statistics are clearly larger than the relevant χ^2 critical values (in this case it is the 95% percentile of the χ^2 distribution with 11 [samples 03 and 04] or 13 [samples 05 and 06], i.e.

[4] A formal description of this intuition is available with the paper online appendix

[5] For estimation purposes the data has been standardized with mean zero and a standard deviation of one.

19.67 and 22.36). The detailed estimates of the initial choice equations relating to interest rate type and loan characteristics are presented in Tables A.2 to A.4 in Appendix A.

Insert Table 2

Insert Table 3

The Effect on Loan Performance of Initial Choices (Default)

We focus our discussion on the parameter estimates accounting for unobserved heterogeneity. Across all samples, the likelihood of default is increased by self-certification (*selfcert*=1) The information problems facing lenders and the adverse selection of borrowers that accompany the use of self-certified mortgages are likely explanations for this observed pattern.

Discounted mortgages (*discount*=1) exhibit a lower likelihood of default in sample 03 and 05 and 06. The parameter estimate in sample 04 is positively signed and statistically significant at the 1% level. The availability of data on the alternative mortgage design *fixed* for the later samples may explain this inconsistency. Tentatively, we conclude that discounted rates reduce the likelihood of default.

Fixed rate contracts (*fixed*=1) were only available in sample 05 and 06. In these cases, a fixed rate (*fixed*) contract reduces the likelihood of default although significantly so only for sample 05; allowing for unobserved heterogeneity changes the parameter sign in sample 06, but does not make the parameter significant at 5% level. Thus again there is tentative evidence of a negative effect on default from choosing a fixed rate mortgage.

Default is more likely the larger the original loan (*log initial loan balance*) and the lower the original house price (*log of initial house value*). Larger loans bear higher servicing costs and this may explain the higher likelihood of default indicated by our estimates. A low purchase price for a property (low value) may reflect other factors such as occupational status and wealth, but may also indicate less collateral for additional borrowing to overcome liquidity problems.

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*) is not statistically significant in the default equations though it is statistically significant and negatively signed for sample 05. Thus the increase in mortgage payment at the end of a teaser rate period does not appear to induce additional default. This may result in part because of the increased likelihood of the competing risk, i.e. prepayment and obtaining cheaper mortgage deals.

The Effect on Loan Performance of Initial Choices (Prepayment)

The likelihood of prepayment increases generally when the borrower self-certifies his/her situation (*selfcert=1),* for samples 03, 05 and 06. In sample 04 the effect of self-certification is insignificant. We suggest a possible explanation: higher rates of interest on self-certification may attract risky borrowers, (who do not disclose their incomes) who gradually repair their credit records, and who will eventually seek less expensive deals in the prime lending sector, or with a new subprime lender.

Holders of discounted mortgages (*discount=1*) exhibit a significant reduction of their likelihood of prepayment in sample 03 and 06 relative to the baseline. In sample 04 discounted mortgages prepay earlier relative to the baseline, while in sample 05 the effect of a discounted mortgage is insignificant. Tentatively, we conclude that discounted rates reduce the likelihood of prepayment. However, the positive and significant coefficient on *revert* in the prepayment equations indicates that an increase in interest payments after a deal has finished induced re-contracting and reduced default.

The likelihood of refinancing is also increased when the *log initial loan balance* is large (the negative parameter estimate for 04 is not significant), and when the *log initial house value* is low. Mortgage holders with larger loans can make higher absolute savings from searching for new mortgage deals. As with default low house value may be indicative of lower occupational status and income and affordability driven movements from deal to deal.

Option Theoretic and Affordability Considerations

The current (i.e. measured at time t) loan to value ratio (*current loan/value ratio*) is used to proxy the extent to which the put option to default is 'in the money'. We expect that its associated parameter will have a positive sign. For default this variable is not statistically significant at the 5% level in three of the samples and it is large, positive and significant for sample 04. Given, that UK mortgages are debt with recourse then there is a lower likelihood of observing ruthless default in the United Kingdom compared to the United States.

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 at the 1% level, on the likelihood of default in the two samples (05 and 06) in which it is included.

Change in the Libor rate since origination (*libor change)* proxies the extent to which the call option to prepay is 'in the money' with an expected negative sign. The parameter estimates for this variable are positive and statistically significant in both samples, possibly reflecting affordability driven renegotiation of existing contracts or the search for new discounted deals. This interpretation is reinforced by (*actual shock)* having a positive and statistically significant effect upon the likelihood of prepayment in both sample 05 and sample 06. The ambiguous results for *std dev libor* (having two positively signed parameters and two negative) are further evidence of the weakness of the option theoretic explanation in the case of United Kingdom originated subprime mortgages.

The Effect of Unobserved Heterogeneity on Estimates and Initial Contract Choices

There are some significant differences in parameter estimates in Table 2 for equations estimated without unobserved heterogeneity and those estimated with 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) relative to the parameter estimates of the "homogenous" model. Controlling for unobserved heterogeneity affects the parameter estimates of mortgage contract choice, so for example the impact of choosing a discounted mortgage on default for samples 03, 04 and 05 increases relative to the parameter estimates of the "homogenous" model and changes its sign for sample 06. Furthermore controlling for unobserved heterogeneity markedly increases, in most cases, the magnitude of the parameter estimate for self certification. Similar effects are observed for parameter estimates of the prepayment equation.

Insert Table 4

Table 4 summarizes the observed signs on the estimates of the parameters measuring the effect of unobserved heterogeneity. The first point to note is the consistency of sign across the four samples. The sign on δ_{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 change in credit market conditions during 2008 which stopped credit impaired households from improving their mortgage terms and thus increased the risk of delinquency and mortgage default.

To further illustrate the effect of heterogeneity we added to the conventional homogenous multinomial logit model of competing exits, we plot in Figure 3 the probabilities of default and prepayment, the corresponding hazards and the survivor probability (the probability of no exit up to time *t*), for five alternative values of the unobserved heterogeneity terms. In each case we calculate the hazard of the probabilities assuming that the value of the unobserved heterogeneity terms are given and observed. The estimates are based on the characteristics and the history of the interest rate and other economic variables that are observed for the longest surviving observation in Sample 03 (64 months). In the case of the probability of prepayment we furthermore indicate (the dashed lines) the effect of removing the reversion option (in this particular case it takes place after 12 months). We chose the particular values of the unobserved components so that the survivor probabilities are distinct.

The plot of the survivor probabilities (the probability that the loans is not pre-paid or defaulted on at the given date) shows that alternative values of the heterogeneity component make a observable difference. For example, for an individual observed with $\varepsilon_i \equiv (\varepsilon_{1i}, \varepsilon_{2i}) = (-1,-1)$ the expected survival time is estimated to be about 10 months, while for $\varepsilon_i = (0,0)$ the expected survival time is about 26 months. For the values of the unobserved component $\varepsilon_i = (0.5, 0.5)$ and $\varepsilon_i = (1,1)$ we do not observe the history of the time dependent variables long enough to give a reliable estimate of the average duration in the sample (although we can give the lower bounds 34 and 20 months respectively).

Insert Figure 4

The heterogeneity terms affects the default and prepayment probabilities of exit and hazards as well. The difference between the probabilities of exit because of default and because of prepayment are clear, in particular note the difference in scale: the probabilities of exit because of prepayment are about 100 times larger than the probabilities of default. Their dependence on the unobserved heterogeneity components are different: the probability of default is increasing with time when $\varepsilon_i = (1,1)$ while the probability of repayment peaks between 10 and 20 months when $\varepsilon_i = (-1, -1)$ or $\varepsilon_i = (-0.5. -0.5)$. The discrete hazard rates are increasing over the observed period and again suggest that unobserved heterogeneity determines the exit rates.

Insert Table 5

In Table 5, we illustrate how the values of the unobserved components affect the (expected) initial characteristics of the loans, i.e. the value of the initial loan, the value of the house purchased, the probability that the loan is self-certified and the probability that the interest rate for the loan is fixed. In all cases we observe that the differences in parameter estimates are substantial. This is obvious in particular when we consider the effect of the unobserved component on the initial balance of the loan and value of the house purchased.

The results of the regression for the initial balance and initial value equation with and without heterogeneity that are estimated jointly are to be found in table A2 in the Appendix. The parameter estimates for the logit models for the choice of initial contract characteristics (fixed/variable, Libor, Self Certification) with and without heterogeneity that are jointly estimated are found in Tables A3 and A4 in the Appendix.

Conclusion

This paper reports the estimates from an econometric model of loan performance using a sample of United Kingdom subprime mortgages that were securitised between 2003 and 2006 with performance statistics available up to and including May 2008. The focus of the empirical analysis was the impact of initial choices of mortgage contracts of different design and with different features, such as the initial loan balance and property value, on the duration to default or prepayment exits from mortgage contracts. Thus the joint estimation recognises the possibility of selectivity surrounding initial choices and its conditional effects on jointly estimated exits from mortgage contracts.

The United Kingdom mortgage market provides an interesting contrast to the United States in that mortgages are largely financed through short term funding, such as retail deposits, leading to a menu of contracts with short rather than longer term fixed interest rates, or variable rates. This together with the continuing liability of mortgage holders for debt on default creates a focus on affordability that is largely borne out by the empirical results. Given that the mortgages securitised were based on contracts with short term fixed interest rates or teaser rates this raises questions regarding the advisability of having securitised such contracts, or at least doing so without a fuller understanding of the selectivity issues and the effect of unobserved heterogeneity on subsequent default and prepayment behaviour.

The estimates show that unobserved heterogeneity has a significant impact upon initial contract choices such as mortgage balance and house value and also upon the characteristics of the loan, that i.e. whether it is a fixed rate, discounted or a self-certified mortgage. Unobserved heterogeneity also explains the default and prepayment probabilities. The significant effect of self-certification on both default and prepayment suggests that there was significant information asymmetry and adverse selection in the United Kingdom subprime mortgage market over the sample period.

The research has several implications. The results suggest the importance of unobserved heterogeneity for initial contract choices and the impact that this selectivity has upon subsequent loan performance. It is advisable to take such factors into account when modelling, valuing and pricing mortgage debt; particularly for the purposes of securitisation where the behaviour of subprime debt has been described as 'idiosyncratic'. UK mortgage securitisers may have packaged the mortgages based on a limited information and understanding of the basis of the mortgage choices made and the selectivity issues involved. Thus the application of our analysis to cover a period which includes the onset of the credit crisis in 2007 and 2008 further adds to our knowledge of the drivers of that financial crisis.

References

Alexander, W., Grimshaw, S. D., McQuen. G. R. and B.A. Slade, (2002). "Some Loans are More Equal than Others: Third-party Originations and Defaults in the Subprime Mortgage Industry". *Real Estate Economics, 30,* 2, 667–697.

Ambrose, B.W. and M. LaCour-Little, (2001), "Prepayment Risk in Adjustable Rate Mortgages Subject to Initial Year Discounts: Some New Evidence", *Real Estate Economics*, 29,2, 305-327.

Ambrose, B.W., LaCour-Little, M., and Z.R. Husza, (2005), "A note on hybrid mortgages", *Real Estate Economics,* 33,4, 765-782.

Ambrose, B.W., A. Sanders, (2003), "Commercial Mortgage Backed Securities: Prepayment and Default". *The Journal of Real Estate Finance and Economics* 26(2/3), 175-192.

Archer, W. and D. Ling, (1993), "Pricing Mortgage-backed Securities: Integrating Optimal Call and Empirical Models of Prepayment", *AREUEA Journal* 21(4), 373-404.

Bank of England, (2008), *Financial Stability Report*, Bank of England, London.

Brueckner, J.K., (1995), "Borrower Mobility, Adverse Selection and Mortgage Points". *Journal of Financial Intermediation*, 3, 416-441.

Brueckner, J.K., (2000), "Mortgage Default and Asymmetric Information". *Journal of Real Estate Finance and Economics*, 20(3), 251-274.

Campbell, J.Y. and J.F. Cocco, (2012). "A Model of Mortgage Default", American Economics Association, 2012 Conference.

Chomsisengphet, S., A. Pennington-Cross, (2004), "Borrower Cost and Credit Rationing in the Subprime Mortgage Market", Office of Federal Housing Enterprise Oversight Working Paper.

Ciochetti, B. , Y, Deng, G. Lee, J. Shilling, and R. Yao, (2003), "A Proportional Hazards Model of Commercial Mortgage Default with Originator Bias", *Journal of Real Estate Finance*, 27. 1, 5-23.

Clapp, J., Y. Deng, A. Xudong, (2006), "Unobserved Heterogeneity in Models of Competing Mortgage Termination Risks", *Real Estate Economics,*34 , 2, 243-273.

Courchane, M.,B. Surette, P. Zorn, (2004), "Subprime Borrowers: Mortgage Transitions and Outcomes", *Journal of Real Estate Finance and Economics, 29.* 4, 365–392.

Cowan, A., C. Cowan, (2004), "Default Correlation: An Empirical Investigation of a Subprime Lender". *Journal of Banking and Finance.28.* 4, 753–771.

Danis, A., A. Pennington-Cross, (2008), "The Delinquency of Subprime Mortgages", *Journal of Economics and Business*. 60, 67-90.

Deng, Y., J.M. Quigley, Van Order, R., (2000), "Mortgage Terminations, Heterogeneity, and the Exercise of Mortgage Options", *Econometrica*. 68 . 2, 275-307.

Deng, Y., J.M. Quigley, (2002), "Woodhead Behaviour and the Pricing of Residential Mortgages", USC Lusk Center for Real Estate, Working Paper No. 2003-1005, University of Southern California.

Deng, Y., S.A. Gabriel, (2006), "Risk-based Pricing and the Enhancement of Mortgage Credit Availability Among Underserved and Higher Credit-Risk Populations", *Journal of Money, Credit and Banking*, 38. 6, 1431-1460.

Dunn, K.B. and C.S. Spatt, (1988). "Private Information and Incentives: Implications for Mortgage Contract Terms and Pricing", *Journal of Real Estate Finance and Economics*, 1:47-60.

Elliehausen. G and M. Hwang, (2010). "Mortgage Contract Choice in Subprime Mortgage Markets", Finance and Economics Discussion Series, Division of Research, Statistics and Monetary Affairs, Federal Reserve Board, Washington D.C. 2010-53.

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

Finkelstein, A., J. Poterba, (2004), "Adverse Selection in Insurance Markets: Policy Holder Evidence from the UK Annuity market", *Journal of Political Economy*, 112, 1, 183-208.

Green, R.K ., J.D. Shilling, (1997), "The Impact of Initial Year Discounts on ARM Prepayments", *Real Estate Economics*, 25.3, 373-386.

Harrison, D.M., T.G. Noordewier, A. Yavas, (2004), "Do Riskier Borrowers Borrow More?", *Real Estate Economics*,32.3, 385-411.

Kau, J.B. , D. Keenen, W. Mueller, J. Epperson, (1993), "Option Theory and Floating Rate Securities with a Comparison of Adjustable- and Fixed-rate Mortgages", *Journal of Business. 66(4)*,595-618.

Lambrecht, B. M., W.R. Perraudin, S.E., Satchell, (2003), "Mortgage Default and Possession: A Competing Hazards Approach", *Journal of Money, Credit and Banking*, 35.2, 425-442.

Lanot, G., (2008), "On the Variance Covariance Matrix of the Maximum Likelihood Estimator of a Discrete Mixture", Working Paper, University of Keele. UK.

Leece, D., (2000), "Household Choice of Fixed versus Floating Rate Debt: A Binomial Probit Model with Correction for Classification Error", *Oxford Bulletin of Economics and Statistics*, 62:1, 61-82.

Leece, D., (2004), *Economics of the Mortgage Market: Perspectives on Household Decision Making*, Blackwell, Oxford.

Lupica, L. R., (2008), "Credit Rating Agencies, Structured Securities and the Way out of the Abyss", *Review of Banking and Financial Law*, 28: 639.

Miles, D. (2004), **"The UK Mortgage Market: Taking a Longer Term View",** HM Treasury, March, United Kingdom.

Miles, D., (2005), "Incentives Information and Efficiency in the UK Mortgage Market", *Economic Journal*, 115,502, C82-C98.

Nichols, J., A. Pennington-Cross, A. Yezer, (2005), "Borrower Self-selection, Underwriting Costs, and Subprime Mortgage Credit Supply". *Journal of Real Estate Finance and Economics*, *30*.2, 119–219.

Oakes, D., (1982), " Direct Calculation of the Information Matrix via the EM Algorithm", *Journal of the Royal Statistical Society*.61.2, 479:482.

Pennington-Cross, A., (2003), "Credit History and the Performance of Prime and Nonprime Mortgages", *Journal of Real Estate Finance and Economics, 27,* 3, 279–301.

Pennington-Cross, A., S. Chomsisengphet, (2007), "Subprime Refinancing: Equity Extraction and Mortgage Termination, *Real Estate Economics*, 35, 2, 233-263.

Pennington-Cross, A and A.G. Ho (2010), “The Termination of Subprime Hybrid and Fixed Rate Mortgages”, *Real Estate Economics*, 38, 3, 399-426.

Pereira, J.A.A., D.P. Newton, D.A. Paxson, (2002), “UK Fixed Rate Repayment Mortgage and Mortgage Indemnity Valuation”, *Real Estate Economics.* 30(2).185-211.

Phillips, R.A., E. Rosenblatt, J.H. VanderHoff, (1996), “The Probability of Fixed and Adjustable Rate Mortgage Termination”, *Journal of Real Estate Finance and Economics*, 13(2), 95-104.

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

Quigley, J.M., R. Van Order, (1990), “Efficiency in the Mortgage Market the Borrowers Perspective”, *AREUEA Journal*, 18(3), 237:252.

Quigley, J.M., R. Van Order, (1995), “Explicit Tests of Contingent Claims Models of Mortgage Default”, *Journal of Real Estate Finance and Economics*, *11, 99-117.*

Rose, M.J., (2008), “Predatory Lending Practices and Subprime Foreclosures: Distinguishing Impacts by Loan Category”, *Journal of Economics and Business*, 13-32.

Sherlund, Shane M. (2008), “The Past, Present, and Future of Subprime Mortgages.” Finance and Economics Discussion Series 2008-63, Federal Reserve Board.

Stephens, M., and D. Quigars, (2008), “Subprime Mortgage Lending in the UK”, *European Journal of Housing Policy*, 8(2), 197-215.

Train, K.E., (2008), “EM Algorithms for Non-Parametric Estimation of Mixing Distributions”, *Journal of Choice Modelling*, 1.1, 40:69.

VanderHoff, J., (1996), “Adjustable and Fixed Rate Mortgage Termination, Option Values and Local Market Condition: An Empirical Analysis”. *Real Estate Economics.* 24(3),379-406.

Figure 1■Default Rates

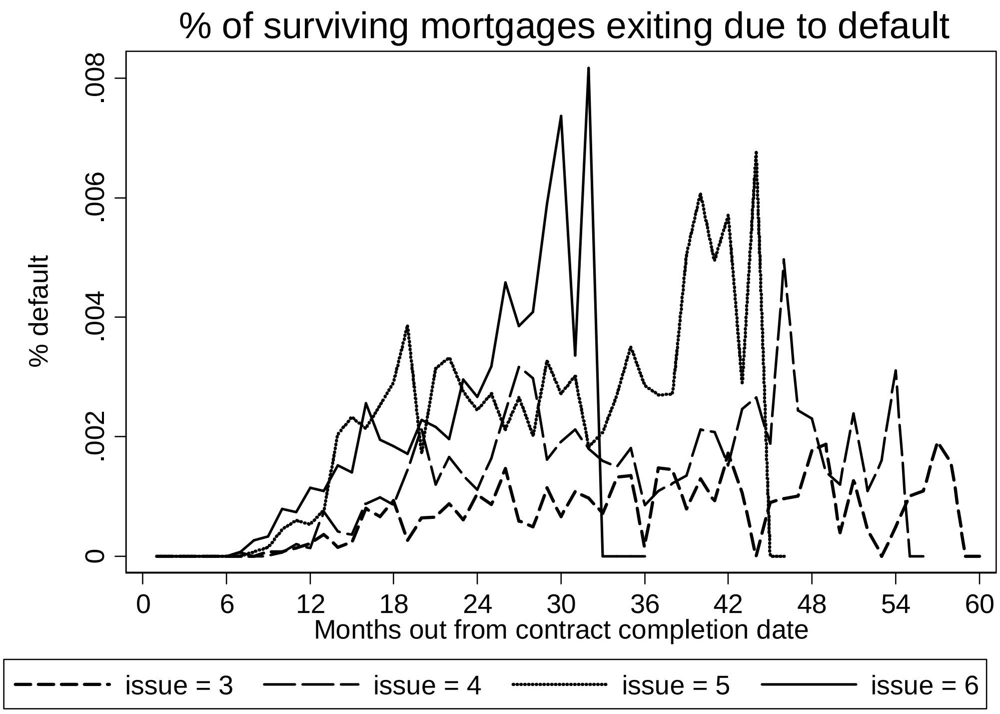

Figure 2 ■ Prepayment Rates

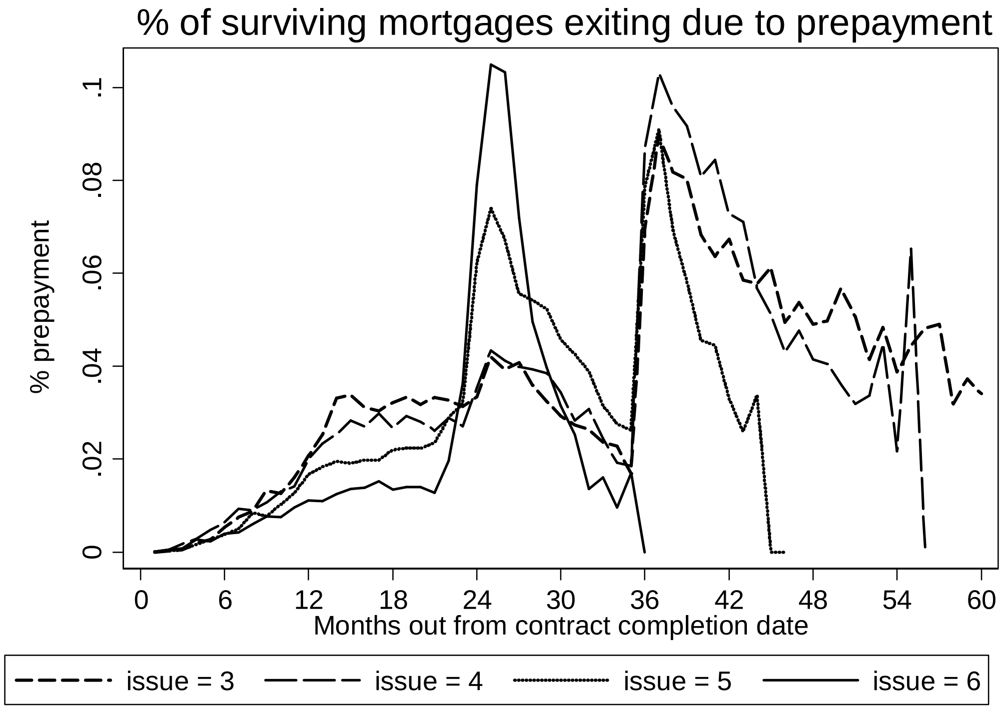

Figure 3 ■ Reversion Rates

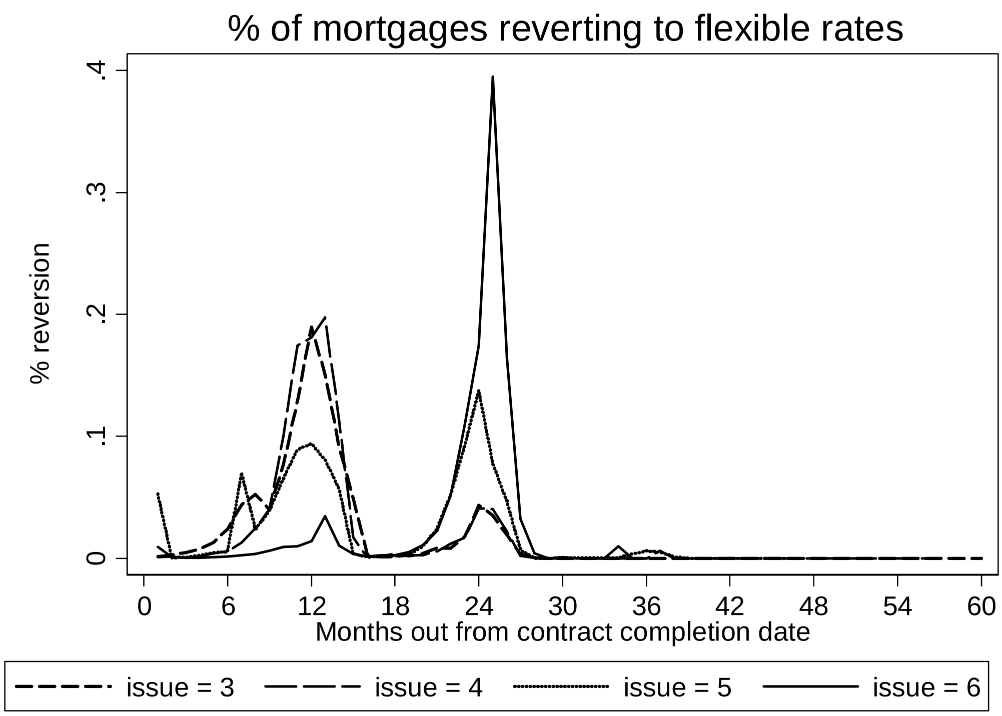

Figure 4 ■ The Effect of Unobserved Heterogeneity on Exit Probabilies

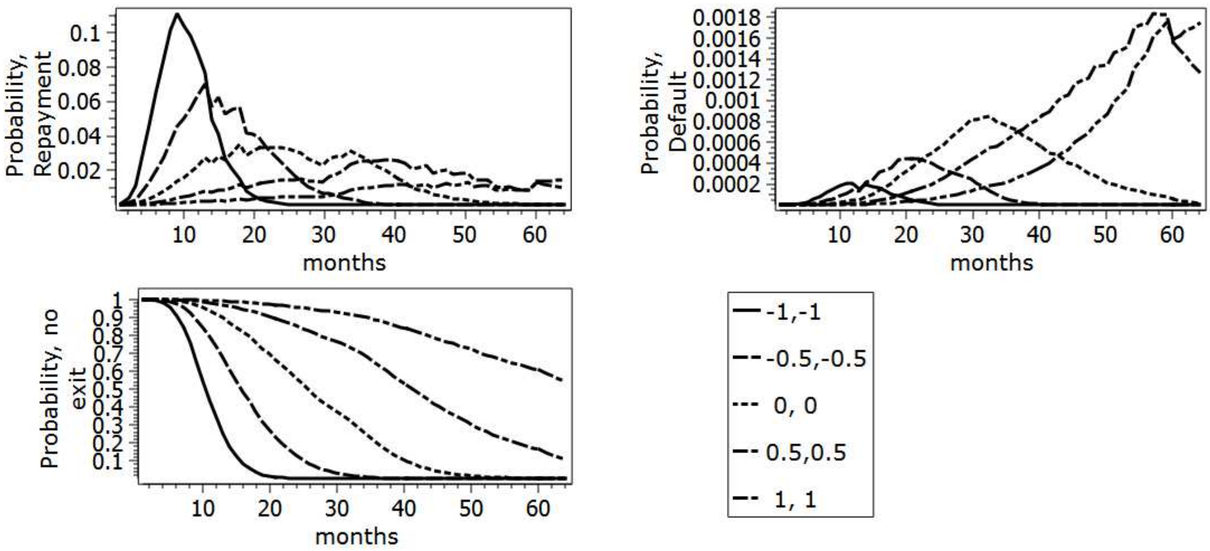

Table 2 ■ Descriptive Statistics

Variable	Sample 03		Sample 04		Sample 05		Sample 06	
	Mean	Std. Dev.	Mean	Std. Dev	Mean	Std. Dev	Mean	Std. Dev
exit: default, % (a)	0.0164	0.127	0.0295	0.1692	0.0454	0.2083	0.0322	0.1766
exit: prepayment, % (a)	0.8792	0.3259	0.8257	0.3794	0.6833	0.4652	0.429	0.495
self-certification, % (a)	0.4194	0.4935	0.2477	0.4317	0.3503	0.4711	0.3799	0.4853
discount, % (a)	0.8738	0.3321	0.8826	0.3219	0.5432	0.4981	0.2695	0.4437
fixed, % (a)	0.1262	0.3321	0.1174	0.3219	0.4045	0.4908	0.7213	0.4484
Initial loan value, £1000 (a)	84.5453	53.6982	84.0951	52.5259	95.1115	54.8472	112.1901	60.508
Initial house value, £1000 (a)	123.6372	82.562	123.6917	75.5366	137.1794	75.5614	152.2807	80.7525
initial loan/value ratio, (a)	0.7061	0.1427	0.6926	0.1411	0.6991	0.1467	0.7449	0.161
current loan/value ratio	0.5313	0.1533	0.5503	0.1462	0.6067	0.1556	0.6414	0.1566
months in sample	20.2446	14.2886	18.8391	12.6766	16.5001	10.4745	13.2205	7.7354
actual shock	1.1397	1.0814	1.3886	0.9612	0.6817	0.9785	0.3072	0.6973
libor, % *100	4.3062	0.577	4.8125	0.43	4.9802	0.473	5.2881	0.5466
libor Change	0.3857	0.609	0.813	0.467	0.1306	0.4726	0.7173	0.5522
std dev libor	0.2099	0.1049	0.2401	0.1243	0.1988	0.1181	0.2635	0.1271
reversion, % all obs.	0.0286	0.1667	0.0295	0.1693	0.0311	0.1737	0.0306	0.1721
Number of (Obs.,Time)		482808		500122		399008		377475
Number of Obs		15118		16232		13955		15570

Note: (a) as a mean or proportion of individual mortgages.

Table 2 ■ Competing Risk Model With and Without Unobserved Heterogeneity

	Sample 03		Sample 04		Sample 05		Sample 06	
	no Unobs. Het.	Unobs. Het.	no Unobs. Het.	Unobs. Het.	no Unobs. Het.	Unobs. Het.	no Unobs. Het.	Unobs. Het.
DEFAULT EQUATION								
Constant	-8.615	-10.607	-8.580	-10.262	-8.046	-10.241	-7.981	-8.753
	(0.135)	(1.362)	(0.133)	(0.656)	(0.119)	(0.938)	(0.114)	(0.605)
log months	1.465	3.458	2.126	3.400	2.373	3.559	1.788	2.238
	(0.250)	(1.198)	(0.168)	(0.447)	(0.179)	(0.599)	(0.174)	(0.359)
actual shock					0.307	0.327	0.217	0.386
					(0.100)	(0.120)	(0.043)	(0.072)
current loan/value ratio	0.188	-0.304	0.367	0.888	0.278	-0.176	0.398	0.044
	(0.174)	(0.237)	(0.124)	(0.198)	(0.120)	(0.171)	(0.159)	(0.204)
libor	0.005	0.134	-0.087	0.033	-1.888	-2.676	-0.794	-1.119
	(0.165)	(0.206)	(0.103)	(0.105)	(1.602)	(2.228)	(1.850)	(2.013)
libor change					1.473	2.248	0.514	0.749
					(1.601)	(2.233)	(1.864)	(2.020)
revert	-0.037	-0.055	-0.081	-0.093	-0.112	-0.135	-0.012	0.005
	(0.098)	(0.104)	(0.086)	(0.089)	(0.067)	(0.068)	(0.037)	(0.038)
std dev libor	0.032	-0.010	-0.273	-0.203	-0.245	-0.123	0.250	0.176
	(0.078)	(0.092)	(0.093)	(0.105)	(0.100)	(0.107)	(0.187)	(0.198)
self cert	0.243	0.462	0.199	0.729	0.294	0.269	0.100	0.219
	(0.073)	(0.115)	(0.052)	(0.121)	(0.045)	(0.058)	(0.047)	(0.064)
discount	0.274	-0.353	0.177	0.954	-0.197	-1.089	0.159	-0.560
	(0.071)	(0.205)	(0.048)	(0.156)	(0.126)	(0.243)	(0.208)	(0.335)
fixed					-0.189	-0.563	0.141	-0.299
					(0.097)	(0.149)	(0.208)	(0.277)
log initial loan value	2.669	6.141	2.805	0.619	2.558	4.651	1.910	3.266
	(0.462)	(1.207)	(0.377)	(0.544)	(0.336)	(0.639)	(0.417)	(0.648)
log initial house value	-2.383	-5.015	-2.649	-1.872	-2.450	-3.469	-1.874	-2.419
	(0.437)	(0.958)	(0.345)	(0.358)	(0.291)	(0.434)	(0.355)	(0.426)
REPAYMENT EQUATION								
Constant	-4.042	-4.179	-3.985	-4.590	-4.321	-4.510	-4.905	-5.330
	(0.015)	(0.031)	(0.013)	(0.057)	(0.020)	(0.049)	(0.025)	(0.085)
log months	1.662	2.121	0.885	1.856	1.339	1.581	1.478	1.801
	(0.036)	(0.076)	(0.025)	(0.090)	(0.037)	(0.069)	(0.048)	(0.081)
actual shock					0.095	0.080	0.368	0.504
					(0.024)	(0.026)	(0.014)	(0.021)
current loan/value ratio	0.093	-0.136	0.057	-0.035	-0.058	-0.185	-0.317	-0.553
	(0.026)	(0.049)	(0.024)	(0.057)	(0.029)	(0.044)	(0.038)	(0.069)
libor	-0.569	-0.489	0.227	0.330	-0.848	-1.186	-1.071	-1.258
	(0.025)	(0.028)	(0.018)	(0.020)	(0.461)	(0.624)	(0.475)	(0.577)
libor change					0.746	1.106	0.662	0.767
					(0.460)	(0.623)	(0.478)	(0.581)
revert	0.055	0.046	0.104	0.072	0.226	0.211	0.367	0.381
	(0.009)	(0.010)	(0.008)	(0.009)	(0.008)	(0.008)	(0.006)	(0.007)
std dev libor	0.184	0.141	-0.151	-0.038	0.017	0.088	-0.036	-0.086
	(0.012)	(0.013)	(0.015)	(0.017)	(0.027)	(0.030)	(0.050)	(0.054)
self cert	0.050	0.135	0.044	0.275	0.060	0.050	0.007	0.091
	(0.010)	(0.017)	(0.009)	(0.038)	(0.011)	(0.016)	(0.013)	(0.027)
discount	-0.002	-0.260	-0.012	0.730	0.017	-0.063	-0.071	-0.624
	(0.009)	(0.023)	(0.009)	(0.045)	(0.034)	(0.110)	(0.065)	(0.252)
fixed					0.147	0.082	0.181	-0.161
					(0.028)	(0.043)	(0.065)	(0.133)
log initial loan value	0.252	1.566	0.130	-0.230	0.142	0.485	0.376	1.227
	(0.049)	(0.126)	(0.048)	(0.171)	(0.055)	(0.113)	(0.066)	(0.280)
log initial house value	-0.331	-1.348	-0.225	-0.116	-0.179	-0.296	-0.290	-0.547
	(0.044)	(0.098)	(0.042)	(0.082)	(0.045)	(0.057)	(0.054)	(0.115)
δ_{21}^{3}		-2.793		-2.428		-2.552		-1.539
		(0.869)		(0.478)		(0.547)		(0.539)
δ_{31}^{3}		-1.094		-1.150		-0.739		-1.175
		(0.107)		(0.149)		(0.145)		(0.131)
δ_{32}^{3}		0.167		1.312		0.479		-0.029
		(0.469)		(0.133)		(0.193)		(0.649)

Note: The table shows for each sample the parameters estimates for the competing risk model without unobserved heterogeneity and then accounting for unobserved heterogeneity. The top part of the table shows the parameter estimates for the default equation the bottom part shows the parameters estimates for the repayment equation. The standard errors are obtained from the estimated information matrix. For sample 03 and 04 the "fixed interest rate" option was not available, hence we do not control for this initial characteristic. The full set of month of completion dummies is estimated but not shown.

Table 3 ■ Likelihoods and Likelihood Ratio Estimates for the Competing Risk Model

	Sample 03	Sample 04	Sample 05	Sample 06
Homogenous Likelihood	-114141.4	-118501.4	-94706.9	-85873.7
Heterogenous Likelihood	-113949.0	-118158.4	-94315.4	-85647.8
Likelihood Ratio	384.8	685.8	783.0	451.8
Number of Observations (individual)	15118.0	16232.0	13955.0	15570.0

Note: this table shows the log likelihood values for the homogenous and heterogenous model for each sample. The likelihood ratio entry show the LR statistic of the null hypothesis that the two models are identical.

Table 4 ■ Parameter Signs: Unobserved Heterogeneity

	Sample 03	Sample 04	Sample 05	Sample 06
δ_{21}^{3}	<0	<0	<0	<0
δ_{31}^{3}	<0	<0	<0	<0
δ_{32}^{3}	>0	>0	>0	<0
ρ	<0	<0	<0	0>

Table 5 ■ The Effect of the Unobserved Heterogeneity on the Initial Characteristics of the Loan

$$\varepsilon_i \equiv (\varepsilon_{1i}, \varepsilon_{2i})$$

	(-1,-1)	(-0.5,-0.5)	(0,0)	(0.5,0.5)	(1,1)
%Deviation from average initial loan	-0.5038	-0.277	-0.0503	0.1765	0.4032
%Deviation from average initial value	-0.4064	-0.2247	-0.043	0.1387	0.3204
Probability of Self Certification	0.5163	0.5506	0.5845	0.6176	0.6497
Probability, Interest Rate Fixed	0.9658	0.9449	0.9125	0.8637	0.7939
Probability, Interest Rate Discount	0.0342	0.0551	0.0875	0.1363	0.2061

Table A1 ■ Monthly default and repayment rates regressions

Variable		default	prepayment
reversion, %		0.0002	0.0996
		(0.0013)	(0.0432)
log months		0.0008	0.018
		(0.0001)	(0.0019)
issue			
	4	0.0008	-0.001
		(0.0001)	(0.0046)
	5	0.0018	-0.0037
		(0.0002)	(0.0044)
	6	0.0017	-0.0081
		(0.0004)	(0.0045)
constant		-0.002	-0.0208
		(0.0003)	(0.0051)
N		203	203
df_m		5	5
F		19.4255	21.7184
F-issue		24.9028	1.4348
p-value F-issue		0.0000	0.2338
R2		0.382	0.3362

Note: The table shows the results of a regression of the relevant rate on the proportion of mortgage reaching the reversion date, the logarithm of months elapsed since the initial completion date and dummies for each issues. The robust standard errors are shown between brackets below the parameter estimates.

Table A2 ■ Initial Mortgage Characteristics

	Sample 03		Sample 04		Sample 05		Sample 06	
	no Unobs.	Unobs. Het.	no Unobs.	Unobs. Het.	no Unobs.	Unobs. Het.	no Unobs.	Unobs. Het.
INITIAL BALANCE EQUATION								
constant	0.000	-0.018	-0.000	-0.024	0.000	-0.000	0.000	-0.000
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.009)	(0.008)	(0.008)
month 3	0.017	-0.012	-0.034	-0.016				
	(0.013)	(0.013)	(0.017)	(0.016)				
month 4	-0.011	0.018	-0.025	-0.031			0.021	-0.000
	(0.013)	(0.013)	(0.018)	(0.017)			(0.011)	(0.010)
month 5	-0.009	0.019	-0.029	-0.029	-0.081	-0.000	0.014	0.021
	(0.013)	(0.013)	(0.018)	(0.017)	(0.012)	(0.012)	(0.013)	(0.013)
month 6	-0.014	-0.010	-0.016	-0.018	-0.065	-0.081	-0.001	0.022
	(0.013)	(0.013)	(0.018)	(0.017)	(0.012)	(0.012)	(0.014)	(0.013)
month 7	-0.010	0.001	-0.036	-0.022	-0.078	-0.068	0.015	0.014
	(0.013)	(0.013)	(0.019)	(0.018)	(0.011)	(0.011)	(0.023)	(0.022)
month 8	-0.009	-0.009	-0.010	-0.020	-0.070	-0.065	0.046	0.000
	(0.012)	(0.012)	(0.019)	(0.018)	(0.011)	(0.011)	(0.023)	(0.022)
month 9	0.011	-0.002	0.005	-0.017	-0.067	-0.057	0.033	-0.001
	(0.011)	(0.011)	(0.019)	(0.018)	(0.011)	(0.011)	(0.023)	(0.022)
month 10	0.013	-0.014	-0.026	-0.006	-0.050	-0.078	0.035	-0.007
	(0.012)	(0.012)	(0.013)	(0.013)	(0.012)	(0.012)	(0.023)	(0.023)
month 11	0.019	0.001			-0.055	-0.061	0.052	0.015
	(0.012)	(0.012)			(0.010)	(0.010)	(0.025)	(0.024)
month 12	0.000	-0.010					0.062	-0.006
	(0.010)	(0.009)					(0.023)	(0.022)
month 13							0.045	0.046
							(0.022)	(0.022)
Δ_{11}^{0}		0.001		-0.005		-0.070		0.015
		(0.019)		(0.015)		(0.014)		(0.014)
Δ_{21}^{0}		-0.010		-0.027		-0.054		0.033
		(0.123)		(0.016)		(0.020)		(0.033)
INITIAL VALUE EQUATION								
constant	0.000	0.012	-0.000	-0.024	0.000	-0.066	0.000	0.008
	(0.008)	(0.009)	(0.008)	(0.008)	(0.008)	(0.009)	(0.008)	(0.008)
month 3	0.018	0.010	-0.031	-0.006				
	(0.013)	(0.013)	(0.017)	(0.017)				
month 4	0.000	0.025	-0.026	-0.002			0.022	0.035
	(0.013)	(0.013)	(0.018)	(0.018)			(0.011)	(0.010)
month 5	-0.003	0.011	-0.024	0.008	-0.068	-0.055	0.000	0.009
	(0.013)	(0.013)	(0.018)	(0.017)	(0.012)	(0.012)	(0.013)	(0.012)
month 6	0.001	0.019	-0.012	0.012	-0.057	-0.050	-0.008	0.052
	(0.013)	(0.013)	(0.018)	(0.018)	(0.012)	(0.011)	(0.014)	(0.013)
month 7	0.002	0.016	-0.030	-0.022	-0.061	-0.036	-0.006	0.013
	(0.013)	(0.013)	(0.019)	(0.019)	(0.011)	(0.011)	(0.023)	(0.021)
month 8	0.012	0.031	-0.004	-0.011	-0.054	-0.055	0.015	0.062
	(0.012)	(0.013)	(0.019)	(0.018)	(0.011)	(0.011)	(0.023)	(0.021)
month 9	0.025	-0.001	0.011	-0.518	-0.056	-0.038	0.008	0.020
	(0.011)	(0.011)	(0.019)	(0.018)	(0.011)	(0.011)	(0.023)	(0.021)
month 10	0.020	0.007	-0.014	-0.421	-0.036	0.395	0.009	0.045
	(0.012)	(0.012)	(0.013)	(0.013)	(0.012)	(0.012)	(0.023)	(0.021)
month 11	0.033	0.365			-0.038	0.380	0.013	0.021
	(0.012)	(0.013)			(0.010)	(0.010)	(0.025)	(0.023)
month 12	0.008	0.245					0.020	0.486
	(0.010)	(0.010)					(0.023)	(0.021)
month 13							0.021	0.416
							(0.022)	(0.021)
Δ_{12}^{0}		0.089		-0.259		0.163		0.205
		(0.022)		(0.016)		(0.014)		(0.014)
Δ_{22}^{0}		0.118		-0.295		0.138		0.145
		(0.138)		(0.016)		(0.020)		(0.033)
σ_{11}^{2}	0.998	0.856	0.998	0.660	0.995	0.815	0.998	0.725
	(0.000)	(0.035)	(0.000)	(0.022)	(0.000)	(0.015)	(0.000)	(0.016)
σ_{12}	0.893	0.792	0.893	0.597	0.864	0.694	0.832	0.604
	(0.000)	(0.074)	(0.000)	(0.041)	(0.000)	(0.027)	(0.000)	(0.028)
σ_{22}^{2}	0.998	0.924	0.998	0.734	0.997	0.836	0.999	0.808
	(0.000)	(0.042)	(0.000)	(0.021)	(0.000)	(0.014)	(0.000)	(0.014)

Note Table A2: The measurements of the initial balance and value have been standardised so that they have mean 0 and variance 1. For each model, the exogenous characteristics are dummy variables for the month of completion for the particular mortgage contract within each sample. The parameters Δ_{jk}^{0} measure the effect of the unobserved components ε_j , $j=1,2$ on the initial balance, $k=1$ and on the initial value, $k=2$. The parameters $\sigma_{11,}^{2}\sigma_{22,}^{2}$ and σ_{12} are estimates of the variances and covariance component of Σ. The standard errors are obtained from the estimated information matrix.

Table A3 ■ Logit Model of Choice of Self Certification Mortgage

	Sample 03		Sample 04		Sample 05		Sample 06	
	no Unobs.	Unobs. Het.	no Unobs.	Unobs. Het.	no Unobs.	Unobs. Het.	no Unobs.	Unobs. Het.
SELF-CERTIFICATION EQUATION								
constant	0.375	0.369	1.192	1.360	0.647	0.653	0.517	0.520
	(0.018)	(0.018)	(0.019)	(0.036)	(0.018)	(0.019)	(0.017)	(0.017)
month 3	-0.020	-0.018	0.003	0.012				
	(0.027)	(0.027)	(0.042)	(0.045)				
month 4	-0.032	-0.033	0.027	0.026			0.027	0.029
	(0.027)	(0.027)	(0.044)	(0.048)			(0.023)	(0.023)
month 5	-0.032	-0.032	0.040	0.037	-0.038	-0.031	0.010	0.011
	(0.027)	(0.027)	(0.043)	(0.047)	(0.026)	(0.026)	(0.027)	(0.028)
month 6	-0.016	-0.017	0.045	0.034	-0.045	-0.039	0.058	0.058
	(0.027)	(0.027)	(0.044)	(0.048)	(0.025)	(0.025)	(0.029)	(0.029)
month 7	0.004	0.003	0.004	0.002	0.002	0.010	0.015	0.015
	(0.027)	(0.027)	(0.046)	(0.050)	(0.025)	(0.025)	(0.047)	(0.048)
month 8	0.020	0.019	-0.029	-0.034	-0.015	-0.009	0.015	0.017
	(0.027)	(0.027)	(0.045)	(0.049)	(0.024)	(0.025)	(0.048)	(0.049)
month 9	0.039	0.039	-0.078	-0.092	-0.059	-0.053	0.026	0.028
	(0.024)	(0.024)	(0.044)	(0.048)	(0.024)	(0.025)	(0.047)	(0.048)
month 10	0.028	0.028	-0.115	-0.120	-0.062	-0.058	0.026	0.028
	(0.026)	(0.026)	(0.031)	(0.034)	(0.025)	(0.026)	(0.049)	(0.050)
month 11	0.051	0.050			-0.046	-0.042	0.037	0.040
	(0.027)	(0.027)			(0.022)	(0.022)	(0.052)	(0.053)
month 12	0.019	0.019					-0.014	-0.011
	(0.021)	(0.021)					(0.048)	(0.049)
month 13							0.101	0.104
							(0.046)	(0.048)
log initial loa	-0.377	-0.477	-0.191	0.417	0.084	0.154	0.253	0.211
	(0.043)	(0.045)	(0.047)	(0.077)	(0.037)	(0.038)	(0.030)	(0.041)
log initial hou	1.064	1.129	0.725	0.630	0.356	0.391	0.224	0.213
	(0.045)	(0.066)	(0.048)	(0.047)	(0.037)	(0.036)	(0.031)	(0.031)
δ_1^1		0.131		0.818		-0.210		0.151
		(0.072)		(0.094)		(0.038)		(0.036)
δ_2^1		-0.049		0.148		-0.079		-0.087
		(0.375)		(0.069)		(0.047)		(0.083)

Note Table A3: The measurements of the initial balance and value have been standardised so that they have mean 0 and variance 1. For each model, the exogenous characteristics are dummy variables for the month of completion for a particular mortgage contract within each sample. The parameters δ_j^1 measure the effect of the unobserved components ε_j, $j=1,2$ on the probability to self-certify the mortgage contract. The standard errors are obtained from the estimated information matrix.

Table A.4 ■ Logit Model of Choice of Interest Rate

	Sample 03		Sample 04		Sample 05		Sample 06	
	no Unobs.	Unobs. Het.	no Unobs.	Unobs. Het.	no Unobs.	Unobs. Het.	no Unobs.	Unobs. Het.
FIXED INTEREST RATE								
constant	-2.064	-2.424	-2.510	-3.676	-0.444	-0.434	0.901	0.983
	(0.036)	(0.055)	(0.037)	(0.095)	(0.024)	(0.029)	(0.023)	(0.501)
month 3	-0.108	-0.105	-0.160	-0.221				
	(0.044)	(0.047)	(0.043)	(0.061)				
month 4	-0.074	-0.091	-0.341	-0.436			0.065	0.079
	(0.043)	(0.047)	(0.047)	(0.066)			(0.023)	(0.050)
month 5	-0.016	-0.024	-0.565	-0.712	-0.661	-0.787	0.141	0.152
	(0.042)	(0.046)	(0.051)	(0.070)	(0.027)	(0.044)	(0.028)	(0.073)
month 6	0.025	0.019	-0.722	-0.885	-0.600	-0.710	0.127	0.130
	(0.042)	(0.045)	(0.056)	(0.073)	(0.026)	(0.040)	(0.029)	(0.056)
month 7	0.083	0.083	-0.471	-0.605	-0.575	-0.687	0.276	0.286
	(0.041)	(0.044)	(0.051)	(0.071)	(0.026)	(0.039)	(0.048)	(0.126)
month 8	0.090	0.086	-0.166	-0.215	-0.571	-0.678	0.353	0.379
	(0.040)	(0.043)	(0.046)	(0.065)	(0.025)	(0.039)	(0.049)	(0.176)
month 9	0.108	0.112	-0.127	-0.148	-0.467	-0.540	0.380	0.403
	(0.035)	(0.038)	(0.046)	(0.065)	(0.026)	(0.035)	(0.049)	(0.178)
month 10	0.165	0.177	-0.198	-0.249	-0.395	-0.438	0.427	0.453
	(0.037)	(0.041)	(0.036)	(0.051)	(0.027)	(0.035)	(0.050)	(0.198)
month 11	0.240	0.257			-0.467	-0.522	0.480	0.515
	(0.037)	(0.041)			(0.025)	(0.033)	(0.053)	(0.231)
month 12	0.093	0.098					0.499	0.541
	(0.029)	(0.032)					(0.050)	(0.246)
month 13							0.561	0.600
							(0.049)	(0.250)
log initial loan value	1.910	1.306	1.283	0.396	0.513	0.292	0.202	-0.141
	(0.093)	(0.080)	(0.077)	(0.127)	(0.039)	(0.046)	(0.032)	(0.703)
log initial house valu	-1.666	-1.327	-0.757	-0.292	-0.099	-0.197	-0.154	-0.140
	(0.095)	(0.107)	(0.077)	(0.073)	(0.039)	(0.044)	(0.032)	(0.099)
self cert	-0.286	-0.188	0.692	0.045	0.416	0.340	0.280	0.305
	(0.056)	(0.062)	(0.057)	(0.117)	(0.040)	(0.046)	(0.039)	(0.067)
δ_{21}^2		0.952		-1.718		0.640		0.476
		(0.135)		(0.198)		(0.086)		(0.396)
δ_{22}^2		-0.091		1.087		0.692		0.550
		(0.626)		(0.110)		(0.120)		(2.241)
LIBOR INTEREST RATE								
constant					-3.965	-4.142	-4.627	-5.338
					(0.130)	(0.202)	(0.245)	(2.110)
month 4							0.065	0.104
							(0.035)	(0.070)
month 5					0.075	-0.102	-0.075	-0.063
					(0.288)	(0.292)	(0.054)	(0.104)
month 6					0.162	0.011	-0.130	-0.137
					(0.270)	(0.275)	(0.058)	(0.113)
month 7					0.087	-0.068	-0.580	-0.600
					(0.263)	(0.268)	(0.099)	(0.259)
month 8					0.429	0.281	-0.737	-0.718
					(0.240)	(0.243)	(0.115)	(0.305)
month 9					1.191	1.093	-0.848	-0.836
					(0.233)	(0.235)	(0.133)	(0.304)
month 10					1.547	1.493	-1.246	-1.234
					(0.252)	(0.254)	(0.211)	(0.367)
month 11					1.098	1.029	-1.774	-1.747
					(0.182)	(0.184)	(0.371)	(0.527)
month 12							-1.301	-1.248
							(0.249)	(0.406)
month 13							-1.376	-1.329
							(0.330)	(0.486)
log initial loan value					0.337	0.016	0.257	-0.753
					(0.081)	(0.102)	(0.167)	(1.044)
log initial house value					0.279	0.120	-0.065	-0.025
					(0.081)	(0.087)	(0.162)	(0.248)
self cert					0.724	0.588	-0.018	0.141
					(0.086)	(0.097)	(0.190)	(0.304)
δ_{31}^2						1.077		1.707
						(0.173)		(0.690)
δ_{32}^2						0.822		1.069
						(0.247)		(2.990)

Note Table A4: The measurements of the initial balance and value have been standardised so that they have mean 0 and variance 1. For each model, the exogenous characteristics are dummies for the completion of the particular mortgage contract within each sample. The parameters δ_{kj}^2 measure the effect of the unobserved components ε_j , $j=1,2$ on the probability to demand a fixed interest rate contract, $k=1$, an interest rate linked to libor, $k=2$. The omitted category is a discounted mortgage. The standard errors are obtained from the estimated information matrix.

Identification (general discussion):

Our model has two parts: a set of equations for the initial conditions of the loan and the competing risk model which describes the duration to exit as default or pre-payment. The characteristics of the initial conditions of the loan (loan size, asset value, certification, interest rate choice) act as predetermined variables for the description of the duration to exit. Because of the nature of the data, we have little specific information about the borrower herself. We control in all cases for the month of contract completion (i.e. the month in a particular issue when the mortgage contract was agreed between the lender and the borrower) and for the individual specific (time invariant) unobserved heterogeneity terms. The unobserved heterogeneity allows for added correlation between the initial conditions of the loan and the duration to exit. Since we limit ourselves here to normally distributed unobserved components, the question of the identification of the joint distribution of the unobserved component is not relevant here. Hence it is the identification of the parameters which capture the effects of the unobserved heterogeneity components on the characteristics of the initial conditions of the loan and on the duration to exit that is of concern. Intuitively, these parameters are identified by the covariance between the (endogenous) characteristics of the initial conditions given the exogenous variables (the months of contract completion) and the correlation between the duration to exit given the exogenous variables and the (predetermined) initial conditions. To form this intuition, consider the (much) simplified model where the initial characteristics of the loan is a single continuous variable y_{0i} and the duration part of the model is reduced to whether or not the individual exits either after the first period, measured by y_{1i}, or a second period, measured by y_{2i}. The simplified model can then be described as:

$$\mathrm{E}_i\left[y_{0i} \mid \varepsilon_i\right] = m_{0i} + \gamma_0 \varepsilon_i,\ \mathrm{V}_i\left[y_{0i} \mid \varepsilon_i\right] = \sigma_0^2,\ \mathrm{V}_i\left[\varepsilon\right] = 1 \text{ and } \mathrm{V}_i\left[y_{0i}\right] = \sigma_0^2 + \gamma_0^2;$$

$$\mathrm{E}_i\left[y_{1i} \mid \varepsilon_i, y_{0i}\right] = F\left(m_{1i} + \gamma\varepsilon_i + \beta y_{0i}\right) \equiv F_{1i},\ \mathrm{V}_i\left[y_{1i} \mid \varepsilon_i, y_{0i}\right] = F_{1i}\left(1 - F_{1i}\right)$$

$$\mathrm{E}_i\left[y_{2i} \mid \varepsilon_i, y_{0i}\right] = \left(1 - F_{1i}\right)F_{2i},\ \mathrm{V}_i\left[y_{2i} \mid \varepsilon_i, y_{0i}\right] = \left(1 - F_{1i}\right)F_{2i}\left(1 - \left(1 - F_{1i}\right)F_{2i}\right)$$

with $F_{2i} \equiv F\left(m_{2i} + \gamma\varepsilon_i + \beta y_{0i}\right)$. The parameter β allows for the duration to exit to depend on the initial characteristic y_{0i}. The expression for the expectation of y_{2i} captures the idea that only observations which do not exit in the first period are potentially exiting in the second period. In keeping with our more elaborate model we assume that the form for the distribution function $F(\ )$ is known. This implies for example that identifying m_{1i} identifies $F\left(m_{1i}\right)$ as well and reciprocally identifying $F\left(m_{1i}\right)$ identifies m_{1i}. The overall question is whether the parameters $m_{0i}, m_{2i}, m_{2i}, \gamma_0, \gamma, \beta$ as well as σ_0^2 are identified from the observation of a sample drawn from the joint distribution of $\left(y_{0i}, y_{1i}, y_{2i}\right)$. We show here that all the parameters of the model are identified when γ and β take small values.

The formal difficulty concerns the description of the reduced form model for y_{1i} and y_{2i}. We approximate F_{1i} and F_{2i} around $\gamma = \beta = 0$ (using first order Taylor expansions) we find:

$F_{1i} \approx F_{1i}^{0} + F_{1i}^{0\prime}\gamma\varepsilon_i + F_{1i}^{0\prime}\beta y_{1i}$, with $F_{1i}^{0} \equiv F\left(m_{1i}\right)$ and $F_{1i}^{0\prime} \equiv F'\left(m_{1i}\right)$,

$$\begin{aligned}\left(1-F_{1i}\right)F_{2i} &= F_{2i} - F_{1i}F_{2i} \\ &\simeq \left(1-F_{1i}^{0}\right)F_{2i}^{0} + \left[\left(1-F_{1i}^{0}\right)F_{2i}^{0\prime} - F_{2i}^{0}F_{1i}^{0\prime}\right]\left(\gamma\varepsilon_i + \beta y_{0i}\right).\end{aligned}$$

Hence we can rewrite the system of equations around $\gamma = \beta = 0$ as a set of simultaneous equations:

$$\begin{aligned}y_{0i} &= m_{0i} + \gamma_0\varepsilon_i + u_i, \\ y_{1i} &= F_{1i}^{0} + F_{1i}^{0\prime}\gamma\varepsilon_i + F_{1i}^{0\prime}\beta y_{0i} + v_{1i}, \\ y_{2i} &= \left(1-F_{1i}^{0}\right)F_{2i}^{0} + \left[\left(1-F_{1i}^{0}\right)F_{2i}^{0\prime} - F_{2i}^{0}F_{1i}^{0\prime}\right]\left(\gamma\varepsilon_i + \beta y_{0i}\right) + v_{2i}\end{aligned}$$

with $\mathrm{E}_i\left[u_i\right] = 0$,

$\mathrm{E}_i\left[{u_i}^2\right] = \sigma_0^2$,

$\mathrm{E}_i\left[v_{ki}\right] = 0, \;\; k = 1,2$,

$\mathrm{E}_i\left[{v_{1i}}^2\right] = F_{1i}^{0}\left(1-F_{1i}^{0}\right)$,

$\mathrm{E}_i\left[{v_{2i}}^2\right] = \left(1-F_{1i}^{0}\right)F_{2i}^{0}\left(1-\left(1-F_{1i}^{0}\right)F_{2i}^{0}\right)$

$\mathrm{E}_i\left[u_i v_{ki}\right] = 0$[1], $\;k = 1,2$,

$\mathrm{E}_i\left[u_i \mid \varepsilon_i\right] = \mathrm{E}_i\left[v_{ki} \mid \varepsilon_i\right] = 0 \;,\;\; k = 1,2$,

and $\mathrm{E}_i\left[v_{1i}v_{2i} \mid \varepsilon_i\right] = -F_{1i}^{0}\left(1-F_{1i}^{0}\right)F_{2i}^{0}$.

Clearly, the reduced form for this approximate model becomes:

[1] This property is present in our more elaborate model since we assume that the unobserved component which determine the duration to exit to be conditionally independent of the unobserved components which determine the initial characteristics of the loan.

$$y_{0i} = m_{0i} + \gamma_0 \varepsilon_i + u_i,$$
$$y_{1i} = F_{1i}^0 + F_{1i}^{0\prime} \beta m_{1i} + F_{1i}^{0\prime}(\gamma + \beta\gamma_0)\varepsilon_i + v_i + F_{1i}^{0\prime}\beta u_i,$$
$$y_{2i} = (1 - F_{1i}^0) F_{2i}^0 + \left[(1 - F_{1i}^0) F_{2i}^{0\prime} - F_{2i}^0 F_{1i}^{0\prime}\right]\beta m_{0i}$$
$$+\left[(1 - F_{1i}^0) F_{2i}^{0\prime} - F_{2i}^0 F_{1i}^{0\prime}\right](\gamma + \beta\gamma_0)\varepsilon_i$$
$$+v_{2i} + \left[(1 - F_{1i}^0) F_{2i}^{0\prime} - F_{2i}^0 F_{1i}^{0\prime}\right]\beta u_i.$$

This allows us to characterise the expectations, variances and covariances of the observed variables given individual characteristics (which we assume are captured by m_{0i}, m_{1i} and m_{2i}). We find:

$$\mathrm{E}_i[y_{0i}] = m_{0i}, \tag{1}$$

$$\mathrm{E}_i[y_{1i}] = F_{1i}^0 + F_{1i}^{0\prime}\beta m_{0i}, \tag{2}$$

$$\mathrm{E}_i[y_{2i}] = H^0 + H^{0'}\beta m_{0i}, \tag{3}$$

$$\mathrm{V}_i[y_{0i}] = \gamma_0^2 + \sigma_0^2, \tag{4}$$

$$\mathrm{V}_i[y_{1i}] = F_{1i}^0(1 - F_{1i}^0) + \left(F_{1i}^{0\prime}\right)^2\left[(\gamma + \beta\gamma_0)^2 + \beta^2\sigma_0^2\right], \tag{5}$$

$$\mathrm{V}_i[y_{2i}] = H^0(1 - H^0) + H^{0\prime 2}\left[(\gamma + \beta\gamma_0)^2 + \beta^2\sigma_0^2\right], \tag{6}$$

$$\mathrm{Cov}_i[y_{1i}, y_{2i}] = -F_{1i}^0 H_i^0 + F_{1i}^{0\prime} H_i^{0\prime}\left[(\gamma + \beta\gamma_0)^2 + \beta^2\sigma_0^2\right], \tag{7}$$

$$\mathrm{Cov}_i[y_{1i}, y_{0i}] = F_{1i}^{0\prime}\left[\gamma_0(\gamma + \beta\gamma_0) + \beta\sigma_0^2\right], \tag{8}$$

$$\mathrm{Cov}_i[y_{2i}, y_{0i}] = H_i^{0\prime}\left[(\gamma + \beta\gamma_0)\gamma_0 + \beta\sigma_0^2\right]. \tag{9}$$

where

$$H^0 \equiv (1 - F_{1i}^0) F_{2i}^0,$$
$$H^{0'} \equiv (1 - F_{1i}^0) F_{2i}^{0\prime} - F_{2i}^0 F_{1i}^{0\prime}.$$

This appears to be a system of nine equations with seven unknowns ($m_{0i}, m_{2i}, m_{2i}, \gamma_0, \gamma, \beta$ and σ_0^2) given our assumptions. However, the variance equations for the exit dates convey the same information as the expectation equations for the same exit dates, hence we are left with a system of seven independent equations for the same number of unknowns. Assume that for each of the 9 moments we observe an empirical equivalent θ_k with $k = 1,...,9$ corresponding to the quantity of the left hand side of each moment equation above. Clearly m_{i0} is identified by θ_1 directly. Assuming that $F_1^{0\prime} > 0$ and $H_i^{0\prime} \neq 0$, Equations (2) and (3) imply:

$$H_i^{0\prime}\left(\theta_2 - F_{1i}^0\right) = F_{1i}^{0\prime}\left(\theta_3 - H_i^0\right),$$

And similarly equations (8) and (9) imply:

$$H_i^{0\prime}\theta_8 = F_i^{0\prime}\theta_9,$$

and therefore $F_{1i}^0 = \theta_2 - \theta_8$ and $H_i^0 = \theta_3 - \theta_9$, and we determine m_{1i} and m_{2i} as

$$m_{1i} = F_1^{-1}\left(\theta_2 - \theta_8\right),$$

$$m_{2i} = F_2^{-1}\left(\frac{H_i^0}{1 - F_{1i}^0}\right)$$

Provided $\theta_1 \neq 0$ and $F_i^{0\prime} + H_i^{0\prime} \neq 0$, we can then determine an estimate for β :

$$\beta = \frac{1}{\theta_1}\frac{\theta_8 + \theta_9}{F_i^{0\prime} + H_i^{0\prime}},$$

where we use

$$F_i^{0\prime} + H_i^{0\prime} = F_1'\left(F_1^{-1}\left(\theta_2 - \theta_8\right)\right) + \left(1 - \theta_8 + \theta_2\right)F_2'\left(F_2^{-1}\left(\theta_3 - \theta_9\right)\right) - \left(\theta_3 - \theta_9\right)F_1'\left(F_1^{-1}\left(\theta_2 - \theta_8\right)\right).$$

Having determined β, m_{0i}, m_{1i} and m_{2i} we observe that equations (7) and (9) imply :

$$\frac{\theta_7 + F_{1i}^0 H_i^0}{F_{1i}^{0\prime} H_i^{0\prime}} = \gamma^2 + 2\gamma\gamma_0\beta + \beta^2\theta_4,$$

$$\frac{\theta_8 + \theta_9}{F_i^{0\prime} + H_i^{0\prime}} = \gamma\gamma_0 + \beta\theta_4,$$

These two equations can be solved for γ and γ_0. Finally equation (4), given the known parameter values and in particular γ_0, determines σ_0^2. Hence for values of γ and β close to 0 all the parameters of the model are exactly identified in this simplified model. As the duration to exit increases the number of covariances which contribute to the identification of the parameters γ γ_0 and β increases and therefore the parameters of the model become over-identified. The complete model in the paper has a structure similar to this simplified model (with a small set of initial characteristics and long durations) and the identification of the parameters (in particular the parameters which capture the effect of the unobserved components on the initial characteristics of the loan and on the duration to exit as well as the parameters capturing the dependence on the initial characteristics) follows from a similar (but lengthier argument).