The Performance of UK Securitized Subprime Mortgage Debt: 'Idiosyncratic' Behaviour or Mortgage Design?

Lanot, Gauthier and Leece, David

Keele Management School (Economics), Keele University, Manchester Metropolitan University Business School

15 November 2010

The Performance of UK Securitized Subprime Mortgage Debt: 'Idiosyncratic' Behaviour or Mortgage Design?

Professor David Leece (contact author), Professor of Financial Studies, Manchester Metropolitan University Business School, Aytoun Building, Aytoun Street, Manchester, M13GH, United Kingdom.

Contact: d.leece@mmu.ac.uk or tel. +44(0)161 247 3742

Professor Gauthier Lanot, Professor of Empirical Economics, Keele Management School, Keele University, Keele, Staffordshire, ST5 5BG, United Kingdom.

Contact: g.lanot@keele.econ.co.uk or tel. +44 (0)1782 733102

The Performance of UK Securitized Subprime Mortgage Debt: 'Idiosyncratic' Behaviour or Mortgage Design?

Professor David Leece, Manchester Metropolitan University Business School, UK

Professor Gauthier Lanot, Keele University Management School , Keele, UK

Abstract

The research estimates a competing risk model of mortgage terminations on a sample of UK securitized subprime mortgages. We consider whether the variety of mortgage contracts that were securitized explains the performance of subprime securities and their supposed 'idiosyncratic' behaviour. The methodological advance is the use of a general, flexible modelling of unobserved heterogeneity over several dimensions, controlling for both selection issues involving mortgage choice and dynamic selection over time. We conclude that securities consisting of subprime loans can be given meaningful valuations on bank balance sheets if the performance of the different types of loans can be better understood.

JEL Codes: G21 C13 C25 C51 D10 D14 E44

Key Words: Subprime mortgages, unobserved heterogeneity, loan performance, securitisation

The Performance of UK Securitized Subprime Mortgage Debt: 'Idiosyncratic' Behaviour or Mortgage Design?

1. INTRODUCTION

The paper reports the results of estimating a model of default and prepayment behaviour with competing risk and unobserved heterogeneity, using a sample of securitized United Kingdom (UK) subprime mortgages originated before the financial difficulties of 2007 and subsequently held by a global investment bank. We consider whether the variety of types of mortgage that were securitized including discounted debt, fixed mortgage interest rates (typically fixed for two or three years), and self certified (that is low or zero documentation) mortgages significantly influenced the default and prepayment behaviour of securitized subprime loans. Given the conventional wisdom that these securities are idiosyncratic and difficult to value this is an important question. Is the variation in loan performance between pools of subprime mortgage debt not only attributable to the contracts contained in the securities, but also the selectivity surrounding such choices? The broader puzzle is why such a variety of contracts were securitized and whether such contract variety is conducive to securitisation?

The motivation for the research is the recent credit crisis and its legacy. The 'aggressive' extension of mortgage lending to subprime borrowers leading to losses on pools of subprime 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 United States (US) prime borrowers, and to some extent subprime 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 report a variety of results concerning the effects of different contract design on loan performance, and the influence of variables representing the value of embedded options in a mortgage contract. Our work adds to this literature, covering mortgages originated and held in a more recent time period (2001-2008). This sample period is of interest in its own right, since a conventional explanation for the recent financial crisis orginates with the performance of subprime loans. An analysis of subprime loan performance is important for better understanding the nature and causes of the financial crisis.

The methodology of the research allows an extension of the modelling of behaviour beyond the data typically available to the securitisers. Unobserved individual influences, that is unobserved heterogeneity, would appear to be an obvious source of variation in loan/security performance and the basis of idiosyncratic behaviour. This paper adds to current research by introducing a general and flexible modelling of this heterogeneity over the relevant two dimensions of default and prepayment behaviour. This approach provides a framework capturing both the selection issues arising from initial mortgage choices, and the dynamic selection effects arising from changes in the population of individuals who remain in the pool of mortgages at each discrete point in time.

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

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

2. DEFAULT AND PREPAYMENT

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

The research into subprime loan performance emphasizes: the influence of contract features such as prepayment penalties and reduced documentation upon the probability of foreclosure (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 results in the research of the subprime market tend to reflect the research into the behaviour of prime mortgage loans, with the co-variates for subprime debt having larger marginal effects on the default and/or repayment probabilities

(Pennington Cross and Chomsisengphet, 2007). Thus the empirical modelling also draws upon the wider mortgage loan performance literature.

Option Theoretic and Empirical Prepayment Models

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

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

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

Mortgage Design

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

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

The mortgage contracts featuring in the current research involve self-certification, discounting, fixed rate contracts, and dates at which the interest rate on the contract reverts from a favourable rate of interest back to an higher index rate (typically London Interbank Offered Rate, LIBOR) plus margin (reset dates). The real estate economics and finance literature has examined the effects of these different aspects of contract design on the performance of mortgage loans (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. Ambrose and LaCour-Little (2001) apply this methodology and find a significant increase in prepayments at adjustment dates.

A further key feature of recent mortgage contracts has been low or zero documentation in the US and its equivalent the self certified mortgage in the UK[1]. This relaxed approach to mortgage underwriting attenuates or overrides prudential lending criteria and introduces information asymmetry, with the lender knowing less about the borrower's ability to pay and likelihood of default. There may be substantial adverse selection and borrowers may exhibit opportunistic behaviour (Brueckner, 2000; Leece, 2004). As a consequence self-certification will have a positive effect on the likelihood of defaulting. There is very little research in this area, an exception is Rose (2008) who estimates a multinomial logit model with unobserved heterogeneity using securitized subprime loans for the Chicago Metropolitan area from January 1999 and up to mid 2003[2]. Rose finds that the effects of variables on foreclosure depend upon loan features such as the level of documentation.

A key motivation for researching the impact of contract designs on loan performance is the question of which loans should be securitized and how such securities would be valued. Until recently, adjustable rate and other mortgage designs were less likely to have been securitized than long term fixed rate contracts (Ambrose and La-Cour Little, 2001). Furthermore, loan contracts which exhibit complex and less predictable default and repayment rates may be less suitable for securitisation. Research to date has typically controlled for different contract designs by analysing loans of a given type (typically long term fixed or adjustable rate). Even analysing one type of loan raises issues regarding

[1] Self-certified mortgages are designed for self-employed or employed individuals with uncertain incomes or incomes from multiple sources. These contracts may therefore also proxy this income uncertainty. There are also claims that self certification was used as an avoidance of due diligence and prudential lending; lending to households on social benefits, or with implausible income statements. In 2008 52% of all new mortgages were self certified (Financial Services Authority, 2010).

[2] Rose controls for unobserved heterogeneity with the use of robust standard errors which allows for clustering in loans. This differs from our generalised modelling of unobserved heterogeneity and analysis of correlations between unobserved heterogeneity for the two forms of exit, and survival.

selectivity (i.e. the borrowers associated with a particular loan specie may share observed or unobserved characteristics which may increase a specific risk). When several types of mortgage are pooled in the same security then the heterogeneity of these specific risk may multiply. The research reported in this paper addresses the selectivity that arises from the individual specific factors that generate the initial choice of contract type by modelling these factors as unobserved heterogeneity. We subsequently evaluate the effect of mortgage contract choice upon both default and prepayment behaviour in a competing risk framework.

3. ECONOMETRIC ESTIMATION

The use of econometric techniques in the mortgage loan literature has evolved from the use of qualitative dependent variable models to the application of Cox Proportional Hazard (CPH) and Multinomial Logit models (MNL) to incorporate competing risks. Further developments have recognised the potential importance of unobserved heterogeneity, particularly when modelling behaviour which from an option theoretic viewpoint appears sub optimal (Deng et al, 2000, 2002). The MNL has the advantage over CPH that the competing probabilities sum to unity modelling competing risk explicitly, because one risk is at the expense of the other, but it has the disadvantage of assuming the independence of irrelevant alternatives. There has been a tendency to favour the multinomial logit model with modifications to allow for the correlation between specific unobserved components, for example see Pennington-Cross and Chomsisengphet (2007).

The econometric methodology advances the literature in several ways. Firstly, we model unobserved heterogeneity as a continuous distribution, rather than estimating parameters for an arbitrary number of mass points that shift the base line hazard. Though there have been some plausible a-priori categorisations of groupings of unobserved heterogeneity for example: employment history; changes in marital status; household mobility (Clapp et al, 2006)- a specification that uses a more general form can cover a wide number of dimensions (unobserved attributes and selectivity) and is more flexible. There are also likely to be sources of selectivity bias in the choice of type of mortgage and other mortgage choices made by households (Nichols et al, 2005). The econometric methodology treats unobserved heterogeneity as a determinant for this selectivity.

The model presented here also allows for correlation between the sources of unobserved heterogeneity that effect 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). Though the empirical models could be viewed as parsimonious this research speeds up estimation and facilitates convergence of the likelihood function by using the methodology of Train (2003) and the adaptation by Lanot (2008). For an outline of the estimation procedure see Appendix A.

Overview

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

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

Finally we wish to account for unobserved differences between individuals which may affect both ”exit” decisions independently or jointly. Hence we denote $\varepsilon_i \equiv (\varepsilon_{i1}, \varepsilon_{i2})$ the vector of unobserved individual factors. We assume that the marginal joint distribution of 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

Firstly, we discuss the general assumptions we wish to maintain to estimate the parameters of interest. In general, given a set of initial exogenous variables, say w_{i0}, we can always express the probability density of a given history in the following general fashion; Equation (1):

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

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 practical and computational reasons the restriction on the conditional distribution of the time varying covariates represented by Equation (2) is maintained:

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

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

Contract Choices, Selectivity and Unobserved Heterogeneity

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

The density of the initial observed characteristics of the loan is assumed to be jointly normally distributed with a mean vector depending linearly on 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). This assumption is formally presented by Expression (3).

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

where K^0 is a parameter matrix and Δ_1^0 and Δ_2^0 are parameters vector 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); Equations (4) and (5). Firstly, in the case of the certification choice we assume that the probability is:

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

$$f_{c_1|x_0,w_0,\varepsilon}\left(1 \mid x_{i0}, w_{i0}, \varepsilon_i\right) = 1 - f_{c_1|x_0,w_0,\varepsilon}\left(0 \mid x_{i0}, w_{i0}, \varepsilon_i\right). \tag{5}$$

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 of Equation (6).

$$f_{c_2|c_1,x_0,\varepsilon}\left(c_2 \mid c_{1i}, x_{i0}, w_{i0}, \varepsilon_i\right) = \frac{e^{Z_i(c_2)}}{\sum_{k=1}^{3} e^{Z_i(k)}}, \text{ with } c_2 \text{ in } \{1,2,3\}, \tag{6}$$

Where:

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

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

Equation (7) represents the assumption that the conditional joint likelihood of the history of decisions can be decomposed into a product of conditional probabilities:

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

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; Equation (8):

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

(8)

where:

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

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 shown by Equation (9):

$$\delta_{31}^3 = \delta^3 \rho \text{ and } \delta_{32}^3 = \delta^3 \left(1-\rho^2\right)^{1/2}. \tag{9}$$

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:

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

The ratio of $\delta_{31}^3 / \delta_{32}^3$ is an increasing function of ρ only and therefore its inversion 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 is used in the case of the estimation of discrete mixtures to the case where the distribution of the unobserved component is known in order to obtain the maximum likelihood estimates. The advantage rests in the fact that both the E-step and the M-step are relatively straightforward in our case, and that in particular the M-step only requires the maximisation of the sum of standard (concave) logit likelihoods and of the likelihood for a multivariate normal variable, weighted by quantities which are directly calculated from the usual normal quadrature abscises and weights and the complete latent likelihood (see the appendix for some details). The precision of our estimates is calculated from the latent likelihood used in the EM algorithm using the results derived in Oakes (1982) and adapted to this case in Lanot (2003).

4. DATA AND EMPIRICAL MODEL

The Data

The data covers 100,000 mortgage contracts, and when constructed as a panel contains approximately two million observations (individual x time points). The mortgages were issued by a single US originator operating in the UK non-conforming residential mortgage market, but the pools also contain some prime and near prime debt. These issues remained the property of a major global investment bank with which one of the researchers undertook collaborative work. The research is subject to confidentiality agreements, and as such the identity of the data source cannot be disclosed[3].

[3] Confidentiality is maintained by not estimating models on particular tranches of securitized debt, but rather incorporating the whole issue for analysis. Though much of the data is now in the public domain the absence

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

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

Reference to the descriptive statistics by sample shown in Table 1 reveals that there is significant variation in the representation of different types of contract across and within issues. In particular the number of self-certified mortgages is significantly higher in the 04 sample. The proportion of fixed rate mortgages also increases in samples 05 and 06 compared to 03 and 04. Variations in contract terms increased significantly for later issues

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.

(samples); for example issue 203 has 49 different contracts and by issue 106 this has increased to 639. Insofar as different types of mortgage exhibit differences in loan performance then mortgage pools and their tranches of securities will also differ. The incidence of types of contract, across the issues/samples, led to low representation of some contracts, or co linearity problems such that not all contract choices could be modelled with each issue. This was not a problem as each contract choice could be modelled within at least one sample.

The investment bank selected the loans of this particular mortgage originator for analysis. Therefore there is concern that the pools may be subject to selectivity bias. Each of the issues used in the research originally had a triple A credit rating from Standard and Poor and Moody's, with an Aaa from Fitch. By March 2009 the ratings of some of the issues had fallen to BBB and BBB-. Though the data was from an originator in the top decile of average loan performance there was considerable variation by issue and between the yearly grouping of issues. This variation in performance can be seen in Figure 1 and Figure2 which show conditional prepayment and conditional default rates by issue. It is also the case that the legitimacy of credit ratings of AAA rated debt can be disputed (Lupica, 2008; Rosengren, 2010).

Insert Figure 1 and Figure 2

Conditional Prepayment Rates and Conditional Default Rates by Issue

We compared the default rates observed in our data with the mean and standard deviations of monthly repossession rates for 160 issues from 26 mortgage originators, evaluated by the same credit rating agency as the pools studied here. The comparison covered a period of twenty four months since the mortgages were pooled. The mean value for all issues was

0.81% with a standard deviation of 1.15%. The mean values for all the samples were within one standard deviation of the overall mean (0.60, 0.29, 0.75, and 0.70 for sample 03 to sample 06 respectively). Earlier dated pools performed better than those issued at a later date introducing some significant variation. Similar results were obtained for comparisons made at twelve and eighteen months. We conclude that the issues used in estimation can be taken as a "representative" sample, though not indicative of the worst performing subprime pools.

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 (*currentlv*) on a mortgage holder's property to represent the extent to which the put option is 'in the money'. The more likely that the put option to default is 'in the money' (negative loan to value ratios) the higher the probability of a household defaulting on the mortgage debt. The descriptive statistics for this and other variables are given in Table 1.

Insert Table 1

To indicate the extent to which the call option is 'in the money' we again follow Pennington-Cross and Chomsisengphet and use the change in interest rates since the date of origination (*libor change*). 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 (*stdlibor*-a moving standard deviation over 12 months) is included as an independent variable[4]. It is expected that the call option to prepay has a higher value when interest rate volatility is high and therefore there is an incentive to keep the option, the likelihood of prepayment then being less.

The empirical specification also includes variables that represent exogenous payment shocks that might influence default and prepayment. One measure of this is the interest rate shock, which is the change in the actual interest rates paid since the date of origination of the contract (*actual_shock*). There is no data on incomes which can be used with mortgage payments to represent the ability to pay. Given the absence of such data then the interest rate shock is used as a proxy for this, with larger shocks being more likely to lead to difficulty in paying. This may then lead to default, or prepayment to seek out discounts and cheaper rates. We also control for the general level of interest rates by including the current level of Libor (*libor*).

The data set does not contain indicators of personal characteristics, and to some extent the influence of these variables is attributed to unobserved heterogeneity in the samples. One variable that may proxy credit worthiness is the original loan to value ratio (Pennington-Cross and Chomsisengphet, 2007). This may also proxy the nature of the household balance sheet and its riskiness (Harrison et al, 2004). Though some studies have included both the current loan to value ratio and the original level of gearing in the same specification

[4] The period exhibits a continuous increase in nominal house prices and low levels of volatility and negative equity (stats). For this and the other reasons stated in the text exercising the option to default is considered less likely than exercising the option to prepay.

(Pennington-Cross and Chomsisengphet, 2007), a high degree of correlation between these two measures can be problematic. The research reported here addresses this problem by including the initial size of mortgage (*log initial loan balance*) and the value of the property at the date of origination of the mortgage (*log initial house value*) as independent variables. Purchasing a more expensive property and having a larger absolute size of loan is also a means of overcoming the prudential lending constraint of a maximum loan to value ratio (Brueckner, 2004) and may therefore increase the risk of facing payment difficulties at some point of time.

The key characteristics of mortgage contract design are indicated by dummy variables. Thus we have indicator variables to represent the choice of a fixed rate mortgage (*fixed*=1), and a discounted mortgage (*discounted*=1); mortgages with the standard variable rate or Libor are excluded for identification. Self-certification is also indicated by a categorical variable (*selfcert*=1). The review of previous research suggested that the sign on *discount* would be positive for both default and prepayment with the estimated parameter on *discount* being larger than that for *fixed*. Self- certification is taken as an indicator of information asymmetry and adverse selection and *selfcert*=1 is expected to lead to higher defaults and prepayments.

A further significant feature of mortgage design is the existence of the interest reset date on which discounts, or periods during which the rate of interest has been fixed, end. A dummy variable (*revert*) is used to represent the current mortgage month if it is within the period prior to the reset date (*revert=1*). Following (Ambrose et al, 2005) it is expected that both defaults and prepayments will be higher in the post reversion period. Defaults may increase because the increase in the interest rate leads to a payment shock that was not fully considered when the mortgage was taken out, that is consumers may have been myopic (Miles, 2004). 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.

5. EMPIRICAL RESULTS

The estimates for most parameters show consistent sign and magnitude across the four samples[5]. Irregular cases are discussed later on. We look first of all at the parameter estimates for those variables which identify features of mortgage contract design and discuss these separately for default and then refinancing behaviour. The results are then evaluated with respect to the time varying variables that largely represent the options embedded in the mortgage and/or reflect exogenous payment shocks and affordability. The parameter estimates for each of the four samples are reported in Table 2. The estimates represent shifts in the baseline hazard which is measured by time out from the date of origination of a mortgage (*log of months*).

Insert Table 2

Mortgage Contract Design

With respect to mortgage contract design the likelihood of default is increased by self-certification (*selfcert*=1). However, a discounted mortgage (*discount=1*) reduces the likelihood of default. There are no statistically significant effects on early repayment or default for holding a fixed rate mortgage (*fixed*=1). The sign and significance of *selfcert* is consistent with higher default rates. The information problems and adverse selection that

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

accompany the use of self-certified mortgages are a likely explanation for this observed pattern. The lower likelihood of default that accompanies holding a discounted mortgage reflects the favourable impact of teaser rates on affordability.

Default is found to be more likely the larger the original loan (*log initial loan balance*) and the lower the original house price (*log of initial house value*); these two variables have the largest parameters (5.1186 and 4.2209 respectively in Sample 03). While larger loans might reflect a good credit rating they also bear higher servicing costs and this may explain the higher likelihood of default indicated by our estimates. A low purchase price for a property (low value) may reflect other factors such as occupational status and wealth, but also indicates the possibility of less absolute value to use as collateral for further borrowing; liquidity constrained households with little or no equity in their property may have their borrowing constrained in the non housing loan market.

The likelihood of prepayment is increased when *selfcert=1* and with *discount*=1 . One possible explanation is that the higher rates of interest on self-certification may lead to risky borrowers, who do not disclose their incomes and who gradually repair their credit records, eventually seeking less expensive deals in the prime lending sector, or with a new subprime lender. There is also significant selectivity attached with those choosing discounted mortgages. Holders of discounted mortgages who may be cash constrained may have a greater incentive to shop around for other teaser rates.

The likelihood of refinancing is also increased when *log initial loan balance* is large, and when *log initial house value* is low; once again these exhibit the largest parameter values (0.4662 and 0.6191 respectively in Sample 03). Mortgage holders with larger loans can make higher absolute savings from searching for new mortgage deals. Households with lower valued houses may have less equity than those with higher valued properties, and therefore seek the release of cash through seeking more competitive mortgage deals, rather than through further borrowing. Again, the results are compatible with an emphasis by households upon cash flow and affordability.

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

Insert Figures 3.a to 3.d

Hazard and Sub Hazards by self certification and reversion=1

Figures 3.a and 3.b illustrate the effects of *self cert* upon the hazard rates and sub hazard rates for default (3.a) and prepayment (3.b). The estimates are based on the characteristics of the longest surviving observation that can be found in Sample 03. The simulations plot the reaction to the path of time varying variables such as *Libor* and are based upon the choice of a fixed rate mortgage (*Fixed*=1) prior to reversion. The hazard and sub hazard curves for default are shifted markedly upwards by the presence of self certification and significantly but less so for prepayment.

Figures 3.a and 3.b incorporate two alternative ways of measuring the risk in the population of exiting at a point of time t. The Hazard schedule depicts the ‘cause specific hazard’ where the population at risk (the risk set) contains survivors from all causes of exit up to a given time t. The sub-hazard differs from the hazard in so far as the relevant risk set includes those who have exited using any exit route other than the one being analysed, in addition to the survivors up to time t. For example, the sub hazard for default will have those individuals who prepaid their mortgage included in the risk set (see Fine and Grey, 1999; Lau, Cole and Gange, 2009). Hence the cause specific hazard measures the rate of exit at time t for a given cause given survival so far, while the sub-distribution hazard measures the exit rate at time t, but conditions both on survival up to time t and on the possibility that some individuals who have not survived up to t may have been at risk of the specific cause at time t. Qualitatively the impact of self certification is the same for prepayment and default though in the case of default the sub hazard schedules are much lower than the hazards and the rate of exit through default now fall away after twenty eight months.

Option Theoretic and Affordability Considerations

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

Change in Libor since origination is the proxy for the extent to which the call option to prepay is 'in the money'. The parameter estimates for this variable (*liborchange*) are not consistent across the four samples. The earlier samples Sample 03 and Sample 04 do have a statistically significant negative sign but has a positive effect in the two subsequent samples. The variable *stdlibor* provides a further test of the option theoretic explanation of refinancing behaviour. In this case the expected negative sign is found in Samples 05 and Sample 06 but the parameters are positive and statistically significant in Sample 03 and Sample 04. Thus the results for testing the option theoretic explanation of prepayment behaviour are ambiguous. The typically larger coefficient on *actual_shock* across samples, compared to that for *libor change* (0.7721 and -0.1858 respectively in Sample 03) and the lack of statistical significance of *libor change* again suggests the importance of affordability.

There are also possible exogenous shocks on prepayment behaviour. Large payment shocks may induce cash constrained borrowers to seek better deals. The change in interest payments on the mortgage since the origination of the debt (*actual_shock)* has a positive and statistically significant effect upon the likelihood of prepayment.

Those results where the estimates are inconsistent in sign across samples involve time dependent variables, in particular those variables reflecting upon the option theoretic interpretation of household mortgage choices. The inconsistency may be the result of complex interactions with mortgage contract features for which we do not control; alternatively households and/or credit market conditions may differ across the samples.

Credit market conditions deteriorated throughout 2007 and up to mid 2008. During this period Libor increased markedly, and with it mortgage rates, reflecting changed market perceptions of risk. Volatility in this case is likely to indicate the several rises in Libor which corresponded with a marked reduction in contracts available for refinancing. Though the four samples all have lives which extend into 2008, later issues have more mortgages that are likely to be affected by these changes. Samples 05 and 06 have a positive sign on *libor change* suggesting that higher interest rates induced a search for contracts with lower rates to minimise payments. There is a negative sign on *stdlibor* implying that tightening credit market conditions reduced prepayments, possibly through less competitive deals being available and/or access to mortgage finance being rationed. These results suggest the importance of affordability and ease of access to mortgage finance driving prepayment decisions[6].

Table 3 summarises the observed signs on the estimates of unobserved heterogeneity and the implied correlation between these estimates for each equation (*B1D, B1R, B2R*). The first point to note is the consistency of sign across the four samples. The sign on *B2R* implies a negative correlation between the unobserved components of default and prepayment so that a reduction in the likelihood of prepayment increases the likelihood of default. This is compatible with the perception that the latter part of 2008 stopped credit impaired households from improving their mortgage terms and thus increasing the risk of delinquency and mortgage default.

Insert Table 3

[6] Comparisons were made between simple logit estimates using the type of exit as the dependent variable with a logit with unobserved heterogeneity. There were no significant changes required in the interpretation of parameter estimates though controlling for unobserved heterogeneity did increase the size of parameter estimates.

6. CONCLUSION

This paper considered whether the variety of mortgage contract designs that were securitized explains the performance of subprime securities, and their supposed idiosyncratic behaviour. A model was estimated based upon the competing risk of mortgage defaults and prepayments controlling for individual unobserved heterogeneity. The unobserved individual specific factors were modelled in a flexible general form, allowing for their influence upon the initial choice of contract type (for example self certification; discounted; fixed). The mixing of a large number of different types of contract into pools of securitized subprime loans may be one reason for their supposed wide variability in risk and return, and presumed unpredictability. The pools of mortgages had different proportions of contract type and an increasing variety of contract terms. Without any assessment of the impact of these contract variations, particularly self-certified/low or zero documentation mortgages, it is not surprising that performance might be viewed as 'idiosyncratic'.

The significance and interpretation of the impact of mortgage contract design on loan performance depends upon whether mortgage choices can be seen as the exercise of embedded call and put options, or as empirically determined by exogenous shocks and factors influencing affordability. The estimation suggested that treating mortgages as embedded option contracts did not explain the default and prepayment behaviour of the sample. Given this then the impact of reversion periods, the information asymmetry and adverse selection associated with self-certification- and only fixing interest rates on mortgage contracts for short periods- largely operated though their impact upon affordability. These affordability issues resulted in adverse effects on default and generated highly active amounts of prepayment for periods where contract choices were plentiful.

There is little evidence, in the samples analysed here, of significant variations in unobserved heterogeneity between pools of mortgages. The main differences in mean loan performance between securities are most likely to arise from compositional effects resulting from having different proportions of contracts with different features. There was an observed change in the behaviour of later pools of debt, possibly arising from the change in credit market conditions which restricted refinancing by liquidity constrained households. To the extent that these considerations and unobserved heterogeneity were not taken into account when pricing securitized bonds then the behaviour of these pools would appear idiosyncratic. Thus securitized subprime loans may be given meaningful valuations on bank a balance sheet; that is if the behaviour resulting from the mix of mortgage designs used in the securities is better understood.

References

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

Alexander, W., S.D. Grimshaw, G.R., 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., M. LaCour-Little, 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., (1994), "The Demand for Mortgage Debt: Some Basic Results". *Journal of Housing Economics* 3, 251-262.

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.

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

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, (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.

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

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

Fine, J.P. and R. J. Gray (1999), "A Proportional Hazards Model for the Subdistribution of a Competing Risk*", Journal of the American Statistical Association*, Vol. 94, No. 446 (Jun., 1999), pp. 496-509.

Gouriéroux, C., A. Monfort, (1995), *Statistics and Econometric Models* , Cambridge : Cambridge University Press.

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.

Lau B., Cole, S.R., Gange, S.J., (2009), "Competing Risk Regression Models for Epidemiologic Data", *American Journal of Epidemiology*, 70, 2, 244-256

Lee, L.F., (2000), "A numerically Stable Quadrature Procedure", *Journal of Econometrics*, 95. 1, 117-129.

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

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.

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.

Press, W.H., (2007), *Numerical Recipes: The Art of Scientific Computing* (3rd edition), Cambridge University Press.

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.

Rosengren, E.S, (2010), "The Impact of Liquidity, Securitization and Banks on the Real Economy", Journal of Money Credit and Banking, 42, 6, pp. 221-228.

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

Train, K.E., (2003), " EM Algorithms for NonParametric 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.

Appendix

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

To simplify this presentation we express the complete latent likelihood for the i[th] observation as

$$\mathrm{L}_i(\theta) = L_i(\theta, \varepsilon_1, \varepsilon_2)\phi(\varepsilon_1)\phi(\varepsilon_2), \qquad \text{A.1}$$

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

$$L_i(\theta) = \iint L_i(\theta, \varepsilon_1, \varepsilon_2)\phi(\varepsilon_1)\phi(\varepsilon_2)d\varepsilon_1 d\varepsilon_2 \qquad \text{A.2}$$

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

$$\begin{aligned}\mathrm{L}_i(\theta) &= L_i(\theta, \varepsilon_1, \varepsilon_2)\phi(\varepsilon_1)\phi(\varepsilon_2) \\ &\approx L_i(\theta, \varepsilon_h, \varepsilon_{h'})\, p(h,h') \\ &= \prod_{\substack{h=1...H \\ h'=1...H}} \left(L_i(\theta, \varepsilon_h, \varepsilon_{h'})\, p(h,h')\right)^{\delta_i(h.h')}\end{aligned} \qquad \text{A.3}$$

for some *h* and *h'*, where we define $\delta_i(h,h')$ to be equal to 1 if observation *i* is of type *(h,h')* and 0 otherwise. In effect we augment the observed data with $\delta_i(h,h')$ the (latent) "type" of each individual observation. From A.3, the (approximate) latent log-likelihood can now be written as

$$\ln \mathrm{L}_i(\theta) \approx \sum_{\substack{h=1...H \\ h'=1...H}} \delta_i(h,h')\left\{\ln L_i(\theta, \varepsilon_h, \varepsilon_{h'}) + \ln p(h,h')\right\} \qquad \text{A.4}$$

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

$$\begin{aligned} & \mathrm{E}_\chi\left[\ln \mathrm{L}_i(\theta) \mid O_i\right](\equiv \ln L(\theta ; \chi)) \\ & \approx \sum_{\substack{h=1 \ldots H \\ h'=1 \ldots H}} \mathrm{E}_\chi\left[\delta_i\left(h, h'\right) \mid O_i\right]\left\{\ln L_i\left(\theta, \varepsilon_h, \varepsilon_{h'}\right)+\ln p\left(h, h'\right)\right\} \end{aligned} \quad \text{A.5}$$

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

$$\mathrm{E}_\chi\left[\delta_i\left(h, h'\right) \mid O_i\right]=\frac{L_i\left(\chi, \varepsilon_h, \varepsilon_{h'}\right) p\left(h, h'\right)}{\sum_{k, k'} L_i\left(\chi, \varepsilon_k, \varepsilon_{k'}\right) p\left(k, k'\right)} \equiv \pi_i\left(h, h' ; \chi\right) \quad \text{A.6}$$

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

$$\chi=\arg \max_\theta \sum_{i=1}^N \sum_{\substack{h=1 \ldots H \\ h'=1 \ldots H}} \pi_i\left(h, h' ; \chi\right)\left\{\ln L_i\left(\theta, \varepsilon_h, \varepsilon_{h'}\right)+\ln p\left(h, h'\right)\right\} . \quad \text{A.7}$$

The EM algorithm has good properties and if it converges it an be shown that it produces the maximum likelihood estimator (see Gouriéroux & Monfort,. 1995). The benefit of using the EM algorithm arises in practice since the objective in A.7 can be understood as the maximum likelihood based on the latent log-likelihood but weighted by the quantities $\pi_i\left(h, h' ; \chi\right)$ (which are treated as given within each M-step).

In our context the latent log-likelihood can be decomposed into the sum of several terms each involving a different set of parameters. Hence the M-step is obtained by the separate maximisation of each of the "independent" components of the properly weighted latent log-likelihood. To illustrate this property assume that we can write:

$$\ln L_i\left(\theta,\varepsilon_h,\varepsilon_{h'}\right)=\ln L_i^1\left(\theta^1,\varepsilon_h,\varepsilon_{h'}\right)+\ln L_i^2\left(\theta^2,\varepsilon_h,\varepsilon_{h'}\right)\ , \qquad \text{A.8}$$

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

$$\theta_M^1=\arg\ \max_{\theta^1}\sum_{i=1}^{N}\sum_{\substack{h=1\ldots H\\ h'=1\ldots H}}\pi_i\left(h,h';\chi\right)\left\{\ln L_i^1\left(\theta^1,\varepsilon_h,\varepsilon_{h'}\right)+\ln p\left(h,h'\right)\right\},$$

$$\theta_M^2=\arg\ \max_{\theta^2}\sum_{i=1}^{N}\sum_{\substack{h=1\ldots H\\ h'=1\ldots H}}\pi_i\left(h,h';\chi\right)\left\{\ln L_i^2\left(\theta^2,\varepsilon_h,\varepsilon_{h'}\right)+\ln p\left(h,h'\right)\right\}.$$

Table 1 **Descriptive Statistics Samples One to Four**				
	Sample 03 n=497570		**Sample 04** n=515654	
Variable	Mean	Std. Dev.	Mean	Std. Dev
Exit	2.971532	0.1693367	2.971211	.1728018
months	2 0.30177	14.45543	18.74752	12.66308
loantovalue	0.6955786	0.147224	0 .6858394	0.1445353
selfcert	0.5765159	0.494111	0 .7476021	0 .4343887
currentlv	0.5311837	0.153348	0 .5505707	0 .1462581
actual_shock	1.133196	1.085864	1.373973	0 .9628001
discount	0.8709388	0.335268	0.8820352	0.3225667
fixed	0.1234459	0.328947	0.1169214	0.3213269
currentlibor	4.313035	0.581544	4.817942	0 .4355623
liborchange	0.385928	0.621803	0.8091501	0.4684648
revert	0.3794903	0.485266	0 .4077017	0.4914077
stdlibor	0.2107883	0.1059696	0 .2403952	0 .1243599

Table 1 cont.

Descriptive Statistics Samples One to Four (cont.)

	Sample 05 n=455826		**Sample 06** n=465386	
Variable	Mean	Std. Dev	Mean	Std. Dev
Exit	2.974975	.1650379	2.981961	0.1412379
months	19.49204	15.99471	17.92071	15.84958
loantovalue	0.6991869	0.14986	0.7314664	0.164162
selfcert	0 .6167397	0.4861814	0.5950136	0.4908899
currentlv	0 .584231	0.1713933	0.5975107	0.1858507
actual_shock	0.5861474	1.060004	0.2322083	0.7776358
discount	0.4997236	0.5000005	0.3423244	0.474488
fixed	0.3742437	0.4839275	0.6202851	0.4853164
currentlibor	4.95956	0.5164024	5.168357	0.6255065
liborchange	0.0623435	0.5970122	0.4107663	0.8862877
revert	0.4647256	0.4987547	0.7453705	0.4356532
stdlibor	0.2083212	0.1276791	0.2629969	0.1319354

Table 2

Parameter Estimates: Competing Risk Model With Unobserved Heterogeneity

	Sample 03		Sample 04		Sample 05		Sample 06	
Default Equation								
	Parameter	Std Error	Parameter	St Error	Parameter	Std Error	Parameter	Std Error
constant	-9.9636	0.8048	-8.8620	0.2431	-10.7801	0.8021	-10.3914	0.8961
log months	2.1515	0.6026	2.2422	0.2299	3.9253	0.4897	2.9253	0.4661
actual_shock	0.7721	0.2914	0.5830	0.1610	0.3060	0.1241	0.5327	0.0708
currentlv	-0.2136	0.2179	0.1146	0.1366	-0.0566	0.1775	0.7016	0.2104
libor	-0.1257	0.3565	0.2205	0.1582	-1.4480	0.2885	-2.7527	0.4377
libor change	-0.1858	0.3153	-0.6101	0.1566	1.2295	0.3082	3.5136	0.5603
revert	0.0613	0.1085	-0.0178	0.0885	-0.1121	0.0619	0.0253	0.0355
stdlibor	-0.0172	0.0803	-0.1426	0.0965	-0.1865	0.1092	0.1116	0.1688
self cert	0.3800	0.0934	-	-	0.2898	0.0595	0.3323	0.0762
discount	-0.3218	0.1531	-0.1164	0.0800			-0.4626	0.0970
fixed					0.0725	0.0984		
libor mortgage					0.6211	0.1366	-0.2993	0.1147
Log initial loan balance	5.1186	0.9106	4.0890	0.5342	4.5993	0.6222	2.7639	0.6470
Log initial house value	-4.2209	0.7502	-3.4524	0.4365	-3.6623	0.4899	-2.4259	0.5387
Repayment Equation								
constant	-4.2544	0.0426	-4.6005	0.0790	-4.6542	0.0387	-6.1337	0.1616
log months	2.0366	0.0947	1.6595	0.1154	1.7099	0.0562	2.2951	0.1395
actual_shock	0.3351	0.0354	0.6032	0.0427	0.0611	0.0267	0.8050	0.0475
currentlv	-0.0364	0.0523	-0.2094	0.0549	-0.1502	0.0406	-0.2461	0.0659
libor	-0.4951	0.0664	0.5855	0.0448	-0.3255	0.0483	-2.0097	0.1414
libor change	-0.0318	0.0621	-0.5707	0.0535	0.4871	0.0487	2.4183	0.1572
revert	0.0866	0.0107	0.1376	0.0097	0.2044	0.0076	0.4096	0.0073
stdlibor	0.1010	0.0123	0.0450	0.0195	-0.1090	0.0261	-0.1587	0.0392
self cert	0.1073	0.0172	-	-	0.0307	0.0135	0.1952	0.0266
discount	-0.1804	0.0389	0.1104	0.0329			-0.6902	0.0474
fixed					0.0906	0.0238		
libor mortgage					-0.0291	0.0500	-0.3924	0.0447
log initial loan balance	0.4662	0.2595	0.6105	0.1396	0.3677	0.0860	0.0029	0.1527
log initial house value	-0.6191	0.1796	-0.6352	0.1089	-0.2445	0.0608	-0.2118	0.0966
Unobserved Heterogeneity								
B1D	-2.0668	0.5984	-1.0956	0.2823	-2.4980	0.4252	-2.1758	0.4766
B1R	-0.7609	0.1633	-1.1409	0.1081	-0.7120	0.0792	-1.1584	0.1519
B2R	0.8974	0.1590	1.4385	0.1106	0.3908	0.1007	1.3184	0.1809

Table 3 Parameter Estimates: Unobserved Heterogeneity

	Pool One	Pool Two	Pool Three	Pool Four
B1D	<0	<0	<0	<0
B1R	<0	<0	<0	<0
B2R	>0	>0	>0	>0
r	<0	<0	<0	<0

Figure 1

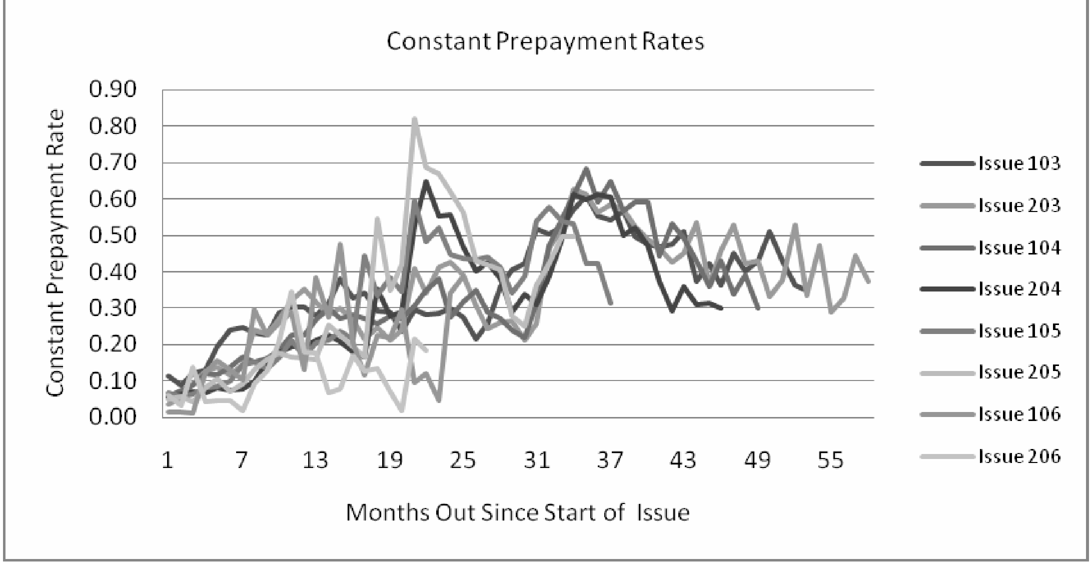
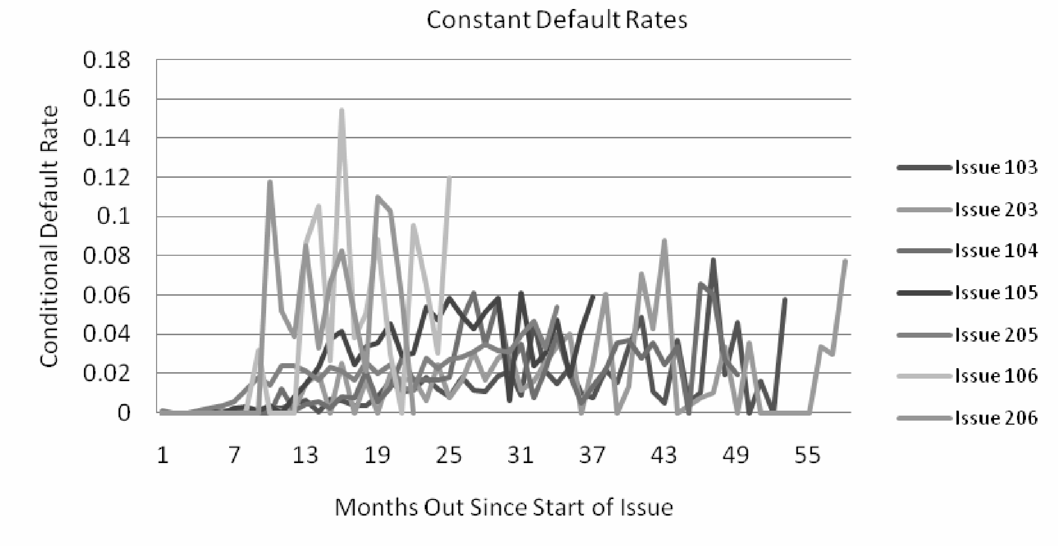


Figure 2



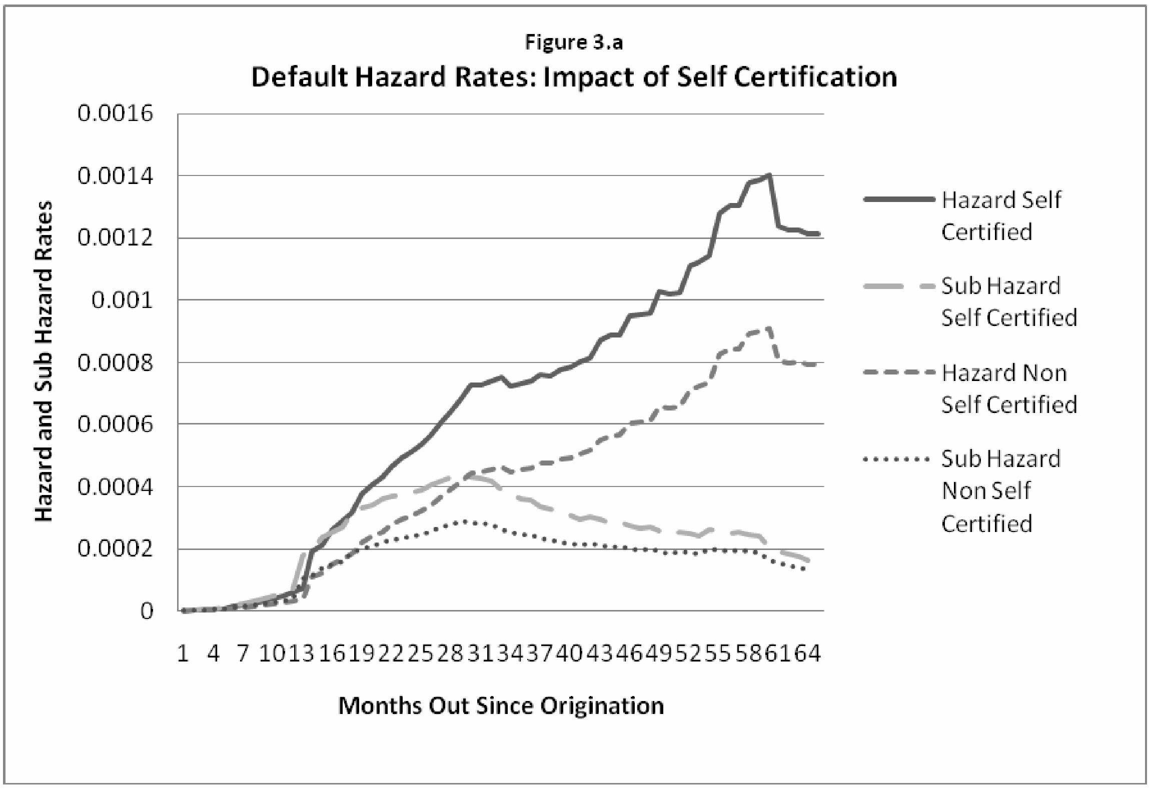


Figure 3.a
Default Hazard Rates: Impact of Self Certification
0.0016
0.0014
0.0012
0.001
0.0008
0.0006
0.0004
0.0002
0
Hazard and Sub Hazard Rates
1 4 7 10 13 16 19 22 25 28 31 34 37 40 43 46 49 52 55 58 61 64
Months Out Since Origination
Hazard Self Certified
Sub Hazard Self Certified
Hazard Non Self Certified
Sub Hazard Non Self Certified

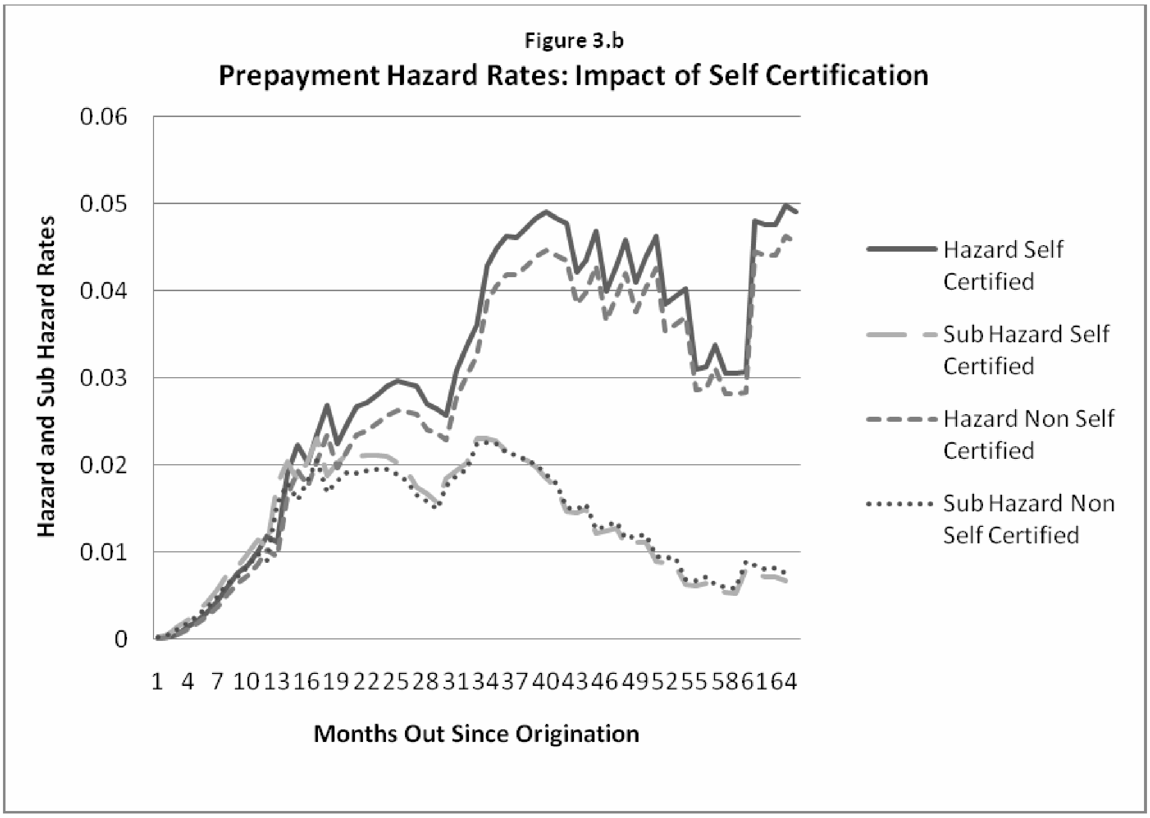


Figure 3.b
Prepayment Hazard Rates: Impact of Self Certification
0.06
0.05
0.04
0.03
0.02
0.01
0
Hazard and Sub Hazard Rates
1 4 7 10 13 16 19 22 25 28 31 34 37 40 43 46 49 52 55 58 61 64
Months Out Since Origination
Hazard Self Certified
Sub Hazard Self Certified
Hazard Non Self Certified
Sub Hazard Non Self Certified