Could the subprime crisis have been predicted?
A mortgage risk modeling approach

Nick Constantinou
University of Essex

José Molina Utrilla
CCFEA, University of Essex

Address for correspondence

Nick Constantinou
Essex Finance Centre
University of Essex
Colchester
Essex CO4 3SQ, UK

E-mail: nconst@essex.ac.uk

Updated November, 2010

Could the subprime crisis have been predicted? A mortgage risk modeling approach

José Molina Utrilla*

Centre for Computational Finance and Economic Agents, University of Essex,
Wivenhoe Park, Colchester CO4 3SQ, UK.

Nick Constantinou†

Essex Business School, University of Essex,
Wivenhoe Park, Colchester CO4 3SQ, UK.

November 22, 2010

Abstract

The abnormally high mortgage default rates that became apparent in early 2007 were not foreseen in June 2005, when mortgage production reached its peak. Could the significant increase in mortgage defaults and the resultant subprime crisis have been predicted? This paper develops a mortgage-level predictive model for mortgage default and delinquency rates, based on a logistic regression and Markov chain framework. The results are compared against actual mortgage level default data and provide strong evidence that the high US nonprime mortgage default rates which triggered the crisis were already predictable in mid-2005 using historical data only available at the time.

JEL: G01, G17, G32

Keywords: Default Probabilities, Logistic Regression, Markov Chain, Credit Risk, Mortgage Defaults

*e-mail: jmolin@essex.ac.uk.

†Corresponding author. e-mail: nconst@essex.ac.uk, Tel.: +44 1206 87 3919, Fax: +44 1206 87 3429.

1. Introduction

The subprime crisis which emerged in 2007 in the nonprime ('Alt-A' and 'Subprime') US housing market, was not predicted and led to the near collapse of the global financial system, unprecedented government (tax-payer) intervention and a global recession. It triggered the worst financial crisis since at least the 1930s whose consequences will likely remain for a substantial number of years. The catalyst for the crisis was the significant increase in default rates (and delinquencies) in US nonprime mortgages (see e.g. Gramlich 2007, Goodman et al. 2008, Mayer et al. 2009, and Crouhy et al. 2008 for a review).

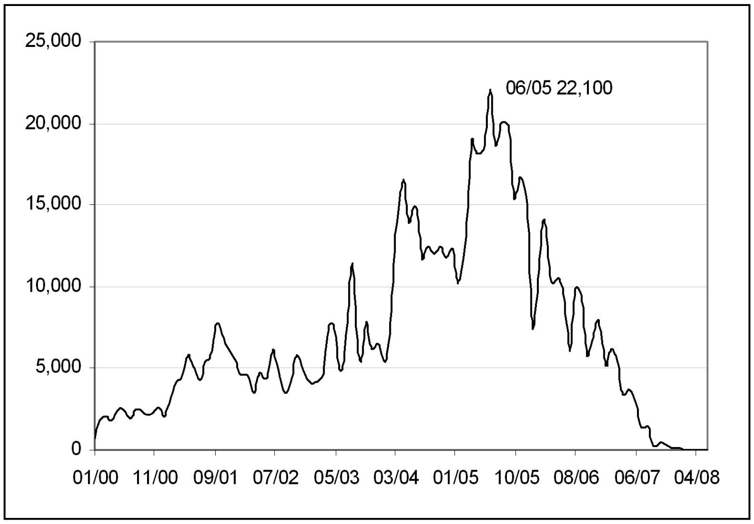
From 2004[1], the mortgage production for sale in secondary markets was the main driver of lender income. The peak in mortgage production, reached in June 2005 (see Figure 1), was driven by a greater liberalization of mortgage underwriting as well as the proliferation of non-traditional mortgages with high-risk features (i.e. 'Alt-A' and 'Subprime'[2]). The liquidity arising from packaging mortgages into Mortgage-Backed Securities (MBS) encouraged lenders to continue loosening credit practices so mortgage production kept up with rising interest rates and further inflated home prices. The flood of MBS into the investment market was a significant driver of the subprime crisis. These MBS experienced much higher than predicted default rates for their underlying mortgages. By the end of 2006, the total amount of MBS outstanding was $6.4 trillion, 49% larger than the market for Treasury debt (Fabozzi et al. 2007). This dynamic stopped abruptly when the housing bubble burst in late 2006 and early 2007.

Mortgage risk modeling failed to predict the high delinquency and default rates that occurred from late 2006 in the USA. Rating agencies and mortgage lenders had very little historical experience in evaluating credit risk on a significant number of mortgages with high-risk features: high combined loan-

[1] In 2004, approximately 64% of new mortgage product was securitised in the secondary market, a rate of securitization that was slightly above the long-term average. See Kramer and Sinha 2006, for a description of the US Mortgage Market.

[2] 'Subprime' mortgages are loans with lower initial credit quality that are more likely to experience significantly higher levels of default. 'Alt-A' loans have some attributes (e.g. unverifiable income) that either increase their perceived credit riskiness or cause them to be difficult to categorize and evaluate (Fabozzi et al. 2007).

Figure 1: Daily Mortgage Production from January 2000 to June 2008 (Closing balance, Million of US Dollars)



Source: First American CoreLogic LoanPerformance (henceforth: LoanPerformance) database.

to-value (CLTV) mortgages, teaser-note mortgages[3], or stated-income mortgages[4]. CLTV[5] is a key ratio used by lenders to determine the risk of default by prospective homebuyers when more than one loan is used (Bhattacharya and Berliner 2005). In general, lenders are willing to lend at CLTV ratios of 80% and above to borrowers with a high credit rating. The lower the CLTV, the lower the default rate. More details on the relationship between CLTV and defaults are found in Zimmerman and Neelakantan (2001).

The main contribution of this paper is the development of a predictive model of mortgage delinquency and default rates that clearly demonstrates that projections for the default rates underlying MBS, calcu-

[3]A teaser rate is an initial rate on an adjustable-rate mortgage (ARM). This rate will typically be below the going market rate, and is used by lenders to persuade borrowers to choose ARMs over traditional mortgages. The teaser rate will be in effect for only a few months, at which point the rate will gradually climb until it reaches the full indexed rate, which will be a static margin rate plus the floating rate index to which the mortgage is tied, usually the LIBOR index (Hayre 2001).

[4]A type of reduced documentation mortgage program which allows the borrower to state on the loan application what their income and assets are without verification by the lender ('Alt-A' loans); however, the source of the income is still verified. These loans might carry a higher interest rate than a 'Prime' (high quality) mortgage.

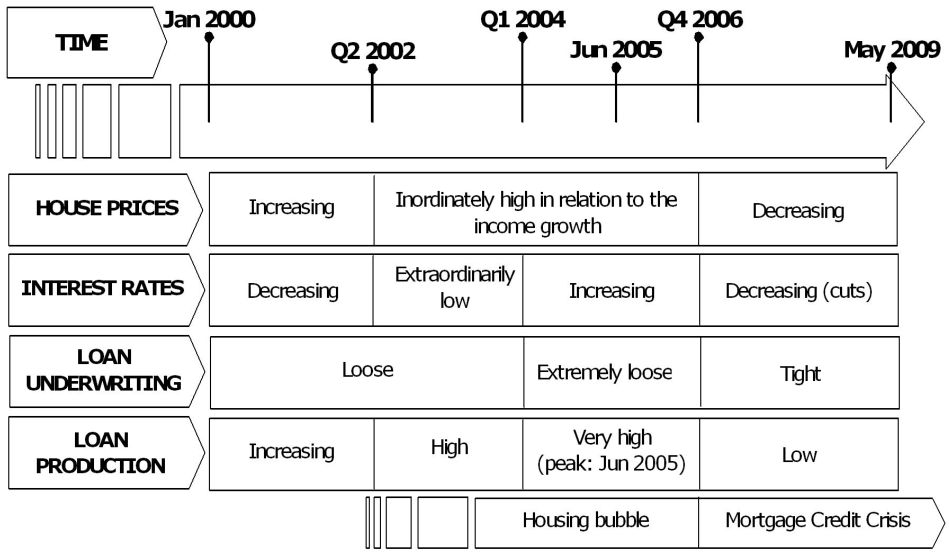
[5]$CLTV = \frac{1}{V}\sum_{l=1}^{n} L_l$, where L_l denotes the closing balance for loan l, V is the property value, and n is the number of loans

lated on mortgages issued on 1 June 2005 (prior to the subprime crisis) would have been closer to actual rates if mortgage level data *only* available at the time were employed for calibration. These include underlying borrower and loan characteristics such as age, balance, loan-to-value, house price, interest rate, employment, occupancy, documentation, loan purpose, and credit score. Models based only on the seasonal nature of default (e.g. the PSA[6] benchmark) have shown a weak predictive power under 'extreme economic' conditions. For the purpose of this study 'extreme economic' conditions arise as a result of a specific economic environment where sharply decreasing house prices coexists with a significant amount of loans with high-risk features. These conditions emerged in early 2007 after a 3-year period characterised by (1) inflated home prices, (2) increasing interest rates, (3) extremely loose loan underwriting, and (4) high production of loans with high-risk features. Figure 2 depicts a chronology of the mortgage phenomenon in the USA from January 2000.

The predictive framework described here combines a logistic regression model based on loan-level data with transition probability matrices (or Markov chain analysis) building on previous studies by Smith and Lawrence (1995), Smith et al. (1996), Calhoun and Deng (2002), Kaskowitz et al. (2002), and Demyanyk and Van Hemert (2010). Smith et al. (1996) present a forecasting model with a Markovian structure and nonstationary transition probabilities to model the life of a mortgage. They use logistic regression models to estimate the severity of losses for a major financial institution and the models calculate annual transition probabilities in contrast to this paper which computes monthly rates in line with amortization schedules. Calhoun and Deng (2002) provide a broad spectrum of approaches to estimate the loss probability distribution for mortgages. Although the focus of their paper is the concept of economic capital, it introduces the state-transition model among the best-practices in mortgage default risk measurement. Kaskowitz et al. (2002) estimate logit models to compare mortgage terminations for fixed- and adjustable-rate loans. The empirical analysis is based on a discrete-time framework that utilizes data on the event histories of individual mortgage loans. Demyanyk and Van Hemert (2010)

[6]The PSA (Public Securities Association), a US trade organization of securities firms and banks created in 1976, underwrites, distributes and trades debt securities both domestically and internationally. It was renamed the Bond Market Association in 1997 and merged with the Securities Industry Association to form the Securities Industry and Financial Markets Association later that year (BMA 1999).

Figure 2: Chronology leading to the Mortgage Crisis in the USA: How the key factors behaved before and after the credit meltdown arose in late 2006



Source: SMR (2007), Gramlich (2007), Goodman et al. (2008), Crouhy et al. (2008) and Mayer et al. (2009).

use LoanPerformance historical data to analyse the quality of subprime mortgages by adjusting their performance for differences in borrower characteristics, loan characteristics, and some macroeconomic variables. These differences were estimated using odds models to show that loan quality for subprime mortgages deteriorated monotonically between 2001 and 2007 prior to the subprime crisis.

In contrast, this paper develops a *predictive* framework to forecast default rates for *all* types of loans 'Prime', 'Alt-A', and 'Subprime'. In particular, previous approaches have tended to focus on subprime mortgages; however, it is important to note that the total number of Alt-A mortgages originated in June 2005 is similar to those of subprime mortgages, and the value of Alt-A loans is 1.5 times that of subprime loans. Both experienced high default rates. The parameter estimation employs mortgage level data obtained from LoanPerformance from January 2000 until May 2005 to forecast defaults and delinquencies for mortgages issued at the hight of US mortgage production (1 June 2005) using both logistic regression and state-transition models. This approach provides strong evidence that the high default rates for both Alt-A and Subprime, starting in late 2006 in the USA could have been detected a year and a half prior to the start of the mortgage crisis.

The remainder of the paper is structured as follows. Section 2 provides a brief description of the PSA benchmark, a simple but realistic model for the typical default rates experienced on thirty-year conventional pools under normal circumstances. This is then followed by a detailed description of the logistic transition method approach. Section 3 applies the theoretical framework to loan level data to forecast default rates based on the simulation of all possible paths a loan can follow from its current status until defaulting or prepaying. Section 4 concludes the paper.

2. Mortgage default risk modeling

The conditional default rate (CDR) is the key measure for quantifying default risk on pools of mortgages underlying MBS. It is defined as the annualized value of the unpaid principal balance of newly defaulted loans over the course of a month as a percentage of the outstanding balance of the pool at the beginning of the month (Fabozzi et al. 2007, Fabozzi 2010). The CDR is computed by first calculating the default rate for the month t, as shown below:

$$DR_t = \frac{DLB_t}{BB_t}, \tag{1}$$

where DLB_t is the default loan balance and BB_t the beginning balance for month t on the underlying pool of mortgages. The annualized CDR is obtained as follows:

$$CDR_t = 1 - (1 - DR_t)^{12}. \tag{2}$$

It should be stressed that the CDR measures only the rate of defaults and not the amount of losses, since the actual losses depend upon the amount that can be recovered on loans in default (adjusted for the costs of collection, and servicer advances, if applicable).

Accurate CDR forecasting methods are crucial in order to provide a better understanding of borrower behavior. In the following section the standard PSA method is described and demonstrated to be a poor predictor of the CDR during the recent credit crisis. An alternative methodology, a combination of logistic regression and Markov chain theory is then proposed.

2.1. The PSA Standard Default Assumption (PSA SDA)

With the increase in the issuance of MBS, a standard benchmark for forecasting CDR was introduced by the Public Securities Association (which was renamed as the Bond Market Association in 2006) in May 1993 (Goodman and Lowell 2000, Fabozzi 2010). The PSA SDA benchmark, or 100 SDA, specifies the following (BMA 1999, Fabozzi 2010):

(a) The CDR in month 1 is 0.02% and increases by 0.02% up to month 30, so that the CDR is 0.6%,

(b) From month 30 to month 60, the CDR remains at 0.6%,

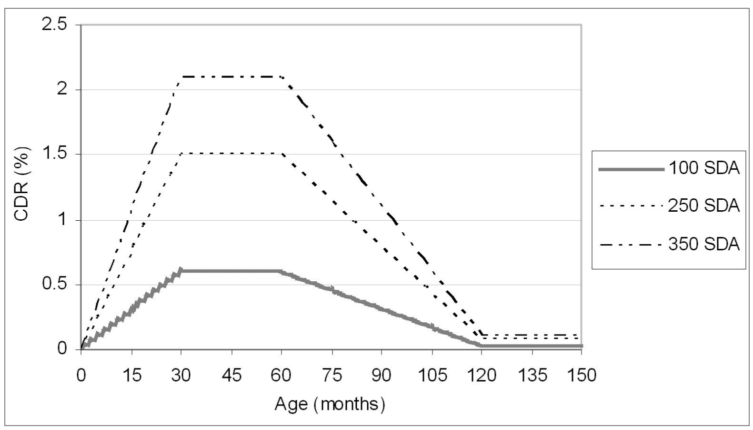
(c) From month 61 to month 120, the CDR declines from 0.6% to 0.03%,

(d) From month 120 onwards, the CDR remains constant at 0.03%,

(e) Monthly Default Rates (see DR_t in equation 1) can be derived from equation 2.

Although the 100 SDA curve is a reasonable model for the typical default rates experienced on thirty-year conventional pools under normal circumstances, it is not an appropriate benchmark for either default rates experienced on conventional mortgage in regions undergoing prolonged recessions or default rates seen on non-conventional loans (i.e. subprime mortgages). Nevertheless, the 100 SDA has been widely used as a benchmark for mortgage default (see e.g. Deng and Gabriel 2006, and An and Deng 2007), and simple linear multiples of 100 SDA, such as 250 SDA and 350 SDA (with maximum CDR of 1.5% and 2.1% respectively), are employed for stress testing the benchmark (Goodman and Lowell 2000). The appropriate SDA curve for a pool of mortgage loans depends strongly on the loan type. For example, default rates on conventional prime loans will typically be much lower than 100 SDA, while the default rates on subprime loans will be much higher than 100 SDA (Hayre et al. 2000). As illustrated in Figure 3, the 100 SDA benchmark constitutes simply a reference to quantify the risk of individual borrowers (the stressed 250 SDA and 350 SDA are shown for comparison). However, the prediction of default risk under extreme economic conditions requires the use of relevant covariables within a framework built on empirical mortgage payment history at the loan level. This framework is developed in the following section.

Figure 3: PSA Standard Default Assumption Benchmark (100 SDA together with the 250 SDA and 350 SDA)



2.2. Logistic Transition Matrix Approach (LTMA)

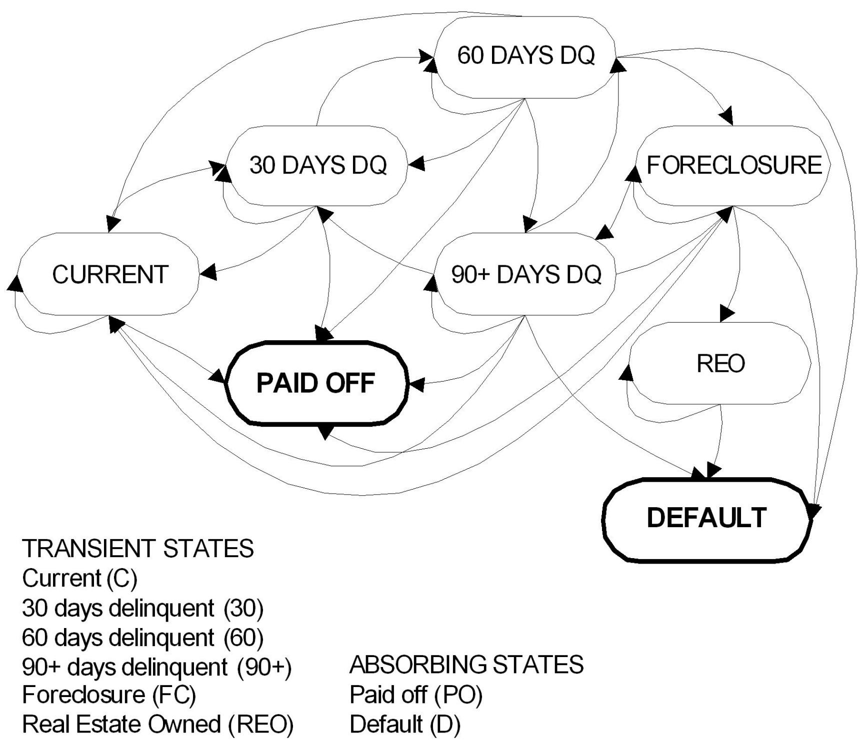
This approach is developed from Markov chain analysis and determines the probability of a borrower making a transition from one credit status to another (Brown and Wadden 2001). This methodology relies on historical *individual* mortgage-level data. The data are categorized according to their characteristics at origination (see section 3) into three asset classes; ‘Prime’, ‘Alt-A’ and ‘Subprime’. Once a mortgage is designated ‘Prime’, ‘Alt-A’ or ‘Subprime’ it remains within that asset class -there are no further changes unless the debt is refinanced, in which case the original loan terminates and another loan is created with the new characteristics at origination. ‘Prime’ (high-credit quality) mortgages have high

FICO scores (credit scores based on the Fair Isaac and Company model), generally above 680, and loan-to-value ratios less than 95%. ‘Alt-A’ mortgages are offered to borrowers with ‘Prime’ characteristics who are unable or unwilling to provide full documentation of assets or income; some of these borrowers are investing in real estate rather than occupying the properties they purchase. ‘Subprime’ mortgages typically have low FICO scores, usually in the low 600s and below, and borrowers have little savings available for down payments (see Kramer and Sinha 2006, for more details on how loans are designated).

Transitions are modeled for individual loans within each asset class. To illustrate the approach, consider a borrower in a particular asset class whose status is current (i.e. no payments overdue). A transition matrix contains the complete set of mathematical probabilities of this borrower rolling from his or her current status to any of the other possible states: (1) remaining current, (2) 30 day delinquent (30 days in arrears), or (2) paid off (PO). A borrower who is 30, 60 or 90+ days delinquent will have a more extensive set of possible transitions. Foreclosure (FC) is the legal proceedings that a lender can initiate after 90 days of delinquency which ends with an auction to sell the property. If the property is not sold (unsuccessful foreclosure auction), the lender regains possession of the property which is then listed on their books as real estate owned (REO), prior to the final default status (D). Before or during the foreclosure process, the servicer may encourage the borrower to sell the property, even if the selling price is less than the mortgage amount; this is called a short sale and might take place after reaching 60 days delinquent. Short sales are default states (D) and mainly occur under adverse economic conditions. Figure 4 displays the possible transitions for loans serviced in the USA (see Fabozzi et al. 2007, Smith et al. 1996, and Fabozzi and Dunlevy 2001).

In order to describe the transition probabilities mathematically, a discrete time Markov chain analysis is employed. Markov chains have been extensively employed in credit modeling (Lando 2004). The mathematical framework is now briefly described. Let $\eta = (\eta_0, \eta_1, \ldots)$ be a stochastic process defined on some probability space $(\Omega, \mathcal{F}, P)$. Assume that the state space is finite and equal to $(1, \ldots, k)$. The process η is said to be a Markov chain if, for every time t and every combination of states $\{i_0, i_1, \ldots, i_t\}$, then

Figure 4: Transition paths for mortgage loans in the USA



$$Pr(\eta_t = i_t \mid \eta_0 = i_0, \eta_1 = i_1, \ldots, \eta_{t-1} = i_{t-1}) = Pr(\eta_t = i_t \mid \eta_{t-1} = i_{t-1}). \tag{3}$$

The Markov chain is said to be time-inhomogeneous if $Pr(\eta_{t+1} = j \mid \eta_t = i)$ depends on t, hence the following one-period transition probability from status i to status j is defined as

$$p_{ij} = Pr(\eta_{t+1} = j \mid \eta_t = i), \tag{4}$$

for a specific t. The transition probabilities at time t are conveniently expressed in a transition matrix

$$
\mathbf{P}_t = \begin{bmatrix} p_{11} & \cdots & p_{1k} \\ \vdots & \ddots & \vdots \\ p_{k1} & \cdots & p_{kk} \end{bmatrix}, \tag{5}
$$

where $\sum_{j=1}^{k} p_{ij} = 1$ for all i (i.e. rows sum to 1). The connection between multi-period and one-period transition probabilities results from $\mathbf{P}_T = \mathbf{P}_0 \cdot \mathbf{P}_1 \ldots \mathbf{P}_{T-1}$. The transition matrix corresponding to figure 4 is described in detail in Section 3.

The probability of a transition from one status to another is estimated using binary logistic regression. Logit or binary logistic regression analysis models the relationship between a binary response variable y, with the possible dichotomous values of 1 or 0, and one or more explanatory variables. For $i = 1, \cdots, k$ states at month t moving to $j = 1, \cdots, k$ states at month $t + 1$, the linear logistic regression model has the form

$$
\log \left[\frac{p_{ij}}{1 - p_{ij}} \right] = \alpha_{ij} + \boldsymbol{\beta}'_{ij} \cdot \mathbf{x}_{ij}, \tag{6}
$$

where $p_{ij} = \text{Prob}(y_{ij} = 1 \mid \mathbf{x}_{ij})$, is the response probability to be modeled, α_{ij} is the intercept parameter, $\boldsymbol{\beta}_{ij}$ is the vector of slope parameters, and $\mathbf{x}_{ij}$ is the vector of explanatory variables related to the transition from status i to status j. In contrast to other applications based on annual transition estimates (Smith et al. 1996), t refers to months in the parameter estimation period.

The expression on the left-hand side of equation 6 is usually referred to as the *logit* or *log-odds*. The logit equation can be solved for p_{ij} to obtain

$$
p_{ij} = \frac{\exp(\alpha_{ij} + \boldsymbol{\beta}'_{ij} \cdot \mathbf{x}_{ij})}{1 + \exp(\alpha_{ij} + \boldsymbol{\beta}'_{ij} \cdot \mathbf{x}_{ij})} \tag{7}
$$

A detailed description of the theoretical framework underpinning logistic regression can be found in standard textbooks (Hosmer and Lemeshow 2000, Allison 2006). The studies by Westgaard and Van der Wijst (2001), and Campbell and Dietrich (1983) are applications of the logistic approach to credit risk. The former presents a method of estimating default probabilities of a retail bank portfolio. The latter provides a multinomial logit model for privately insured conventional mortgage loans using data from 1960 to 1980.

3. Forecasting CDRs under extreme economic conditions

This is the key section of the paper where the theoretical framework described above is applied to actual loan level data. Using the PSA Standard Default Assumption for comparison purposes, it is demonstrated that the LTMA could have predicted the high mortgage default rates that occurred from late 2006 in the USA. The investigation takes the following steps:

1. Calculate the actual CDRs for loans originated on 1 June 2005 forward for a 5-year time window,

2. Plot CDRs for the same forward time window using the PSA SDA benchmark,

3. Forecast CDRs for the same population and forward time window using LTMA calibrated with mortgage level data from 1 January 2000 to 31 May 2005, which was the parameter estimation period, and

4. Integrate both the parameter estimation and the forecasting routines into a computational framework capable of running simulations of all possible transitions that a particular loan can follow. Delinquency and default predictions can be produced at any level of aggregation (i.e. loan, pool, asset class, etc.).

This is achieved by using the SAS (Release 8.2) programming language and statistical package along with the LoanPerformance source of data, a database containing loan-level information on approximately

85 percent of all non-agency mortgage securities[7] in the USA.

3.1. Projected CDRs using the PSA SDA benchmark

The standard PSA SDA benchmark is now compared with the actual CDR for those loans originating on the 1st June 2005. The loans originated on this date are chosen as the benchmark set of mortgages as they occur at the peak of loan production. Loans with missing or insufficient data were excluded (approximately 8.6%). Table 1 displays the loan population employed. The *actual* CDRs for these benchmark loans are obtained from the database from 1st June 2005 to 31st June 2010. The loan-level delinquency status provided by LoanPerformance is used to identify defaults at different points within the time window. By definition, a default occurs when (1) the borrower loses title to the property and (2) the loan is no longer on the books, that is to say, when the previous state was REO, or the previous state was delinquent or foreclosure and there was a loss (see Fabozzi et al. 2007, and Smith et al. 1996). In other words (Hayre et al. 2008), a default only occurs when a loan is liquidated. If a loan becomes delinquent, or goes into foreclosure, but is cured (that is, becomes current again) or is prepaid without a loss to the servicer, there is no default. Figure 5 shows the actual CDRs for the 1 June 2005 benchmark mortgage population compared against the PSA SDA benchmark.

Table 1: Non-agency mortgages originated on 1 June 2005 in the USA (loan production peak)

Asset Class	# Loans	%	Original Balance (Million of USD)	%
Prime	7,483	9.5	3,811.42	18.8
Alt-A	35,100	44.3	10,090.03	49.7
Subprime	36,596	46.2	6,408.72	31.6
TOTAL	79,179	100.0	20,310.17	100.0

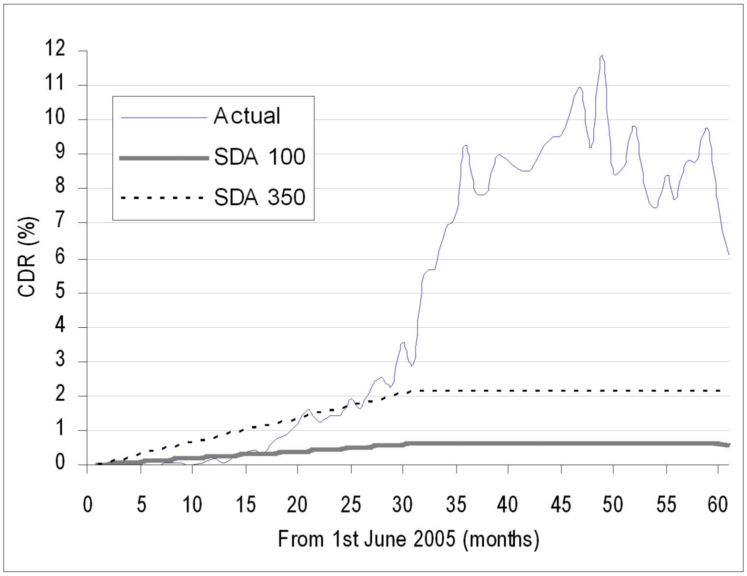
Source: LoanPerformance.

The 100 SDA projection fits the actual CDRs well until the ‘bursting’ of the housing bubble in late

[7] Non-agency mortgage securities are those instruments issued by entities without the full faith and credit of either agencies with close ties to the US government (government-sponsored enterprises) or the US government itself. In 2007, non-agency mortgages represented approximately 34% of the securitized market (Goodman et al. 2008).

2006 (month 17). From this point forward, actual CDRs rise sharply reaching 2% in October 2007 (month 27) and going beyond 7% in June 2008 (month 35). The 350 SDA never exceeds 2.1% during the whole time window. Clearly, the PSA SDA did not work as an acceptable benchmark under the new marktet conditions.

Figure 5: Actual vs Projected CDRs using the PSA SDA Benchmark (100 SDA) and 350 SDA



Source: LoanPerformance (primary data for actual CDRs).

3.2. Projected CDRs using the LTMA

A set of loan-level transition models is developed to predict the probability that a loan makes a transition from one status to another based on historical data of the borrower and loan characteristics. Logistic regression models linked together through their relationship in a transition matrix are the foundation for loan-level predictions. The monthly transition matrix has the following structure (for each asset class):

$$\mathbf{P}_t = \begin{array}{rc} \textit{From} \backslash \textit{To} & \begin{array}{ccccccccc} \textit{C} & \textit{30} & \textit{60} & \textit{90+} & \textit{FC} & \textit{REO} & & \textit{PO} & \textit{D} \end{array} \\ \begin{array}{r} \textit{C} \\ \textit{30} \\ \textit{60} \\ \textit{90+} \\ \textit{FC} \\ \textit{REO} \\ \textit{PO} \\ \textit{D} \end{array} & \left[\begin{array}{ccccccc|cc} p_{11} & p_{12} & 0 & 0 & 0 & 0 & & p_{17} & 0 \\ p_{21} & p_{22} & p_{23} & 0 & 0 & 0 & & p_{27} & 0 \\ p_{31} & p_{32} & p_{33} & p_{34} & p_{35} & 0 & & p_{37} & p_{38} \\ p_{41} & p_{42} & p_{43} & p_{44} & p_{45} & 0 & & p_{47} & p_{48} \\ p_{51} & 0 & 0 & p_{54} & p_{55} & p_{56} & & p_{57} & p_{58} \\ 0 & 0 & 0 & 0 & 0 & p_{66} & & 0 & p_{68} \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right] \end{array}, \tag{8}$$

where p_{ij} are transition probabilities across states from one month to the other. For instance, the first row (C) implies that any current account this month may next month either become thirty days delinquent (30), pay the outstanding balance in full (PO), or stay current; zeros imply that it is impossible for a current account to move to one of these states within 30 days, and ones denote liquidated accounts. Since the outcome of an initial status generally have more than two results, all p_{ij} in equation 8 are calculated extending the binary logistic model (equation 7) to polychotomous logistic regression[8] as follows:

$$p_{ij} = \frac{\exp(\alpha_{ij} + \boldsymbol{\beta}'_{ij} \cdot \mathbf{x}_{ij})}{1 + \sum_{j \neq i} \exp(\alpha_{ij} + \boldsymbol{\beta}'_{ij} \cdot \mathbf{x}_{ij})} \text{ where } i \neq j, \tag{9}$$

$$p_{ii} = \frac{1}{1 + \sum_{j \neq i} \exp(\alpha_{ij} + \boldsymbol{\beta}'_{ij} \cdot \mathbf{x}_{ij})}. \tag{10}$$

Parameters α_{ij} and β_{ij} are estimated using binary logits (equation 7). A key advantage of the binary

[8]See Smith et al. (1996) for an application of the polychotomous (or multinomial logistic) regression to credit risk on home mortgage portfolios.

logits (individualized regressions) is that they allow different variables to influence different transitions. This leads to more accurate predictions as it is unlikely that the same set of variables provide the best specification for every transition within a row of the matrix. Begg and Gray (1984) provide a justification for this approach of calculating polychotomous logistic regression parameters using individualized regressions.

For the estimation of the logistic regression parameters, data from LoanPerformance of mortgages originated between 1 January 2000 to 31 May 2005 are employed (1.9% of the loans were excluded since they were originated but did not have servicing records over the observation period). Table 2 summarizes this historical loan population by asset class.

Table 2: Non-agency mortgages originated between 1 January 2000 to 31 May 2005 in the USA (prior to the loan production peak that occurred on 1 June 2005)

Asset class	# Loans	%	Original Balance (Million of USD)	%
Prime	1,646,197	16.0	732,071.43	33.7
Alt-A	2,129,616	20.7	529,625.33	24.4
Subprime	6,528,429	63.3	909,851.49	41.9
TOTAL	10,304,242	100.0	2,171,548.25	100.0

Source: LoanPerformance.

The explanatory variables defined for the parameter estimation are mostly borrower and loan characteristics, although unemployment was also included along with house price indices to calculate current loan to value ratios (a detailed definition of all the explanatory variables is provided in Appendix A). All these variables are as follows:

AGE - Number of months between the first payment date and the current date (as of date). It captures the seasonality of loans, and equals 0 at origination (note according to the PSA SDA benchmark, risk is higher for loans with ages ranging from 30 to 60 months).

BALANCE - Original principal balance (closing balance).

MTMCLTV - Mark to Market CLTV. It is the current CLTV based upon the house price appreciation. The Case-Shiller House Price Index published by Moody's Economy.com on a quarterly basis is employed. The industry formula is expressed as follows (Zimmerman and Neelakantan 2001):

$$MTMCLTV_t = CLTV_0 \left(\frac{b_t}{b_0}\right)\left(\frac{HPI_0}{HPI_t}\right), \tag{11}$$

where $CLTV_0$ is the Combined Loan-to-Value at origination, b_t is the current balance at the beginning of month t, b_0 is the original balance, HPI_t is the current house price index, and HPI_0 is the original house price index. Qi and Yang (2009) suggest that the current loan-to-value ratio is the most important determinant of default.

OWNER_OCCUPY - Categorical variable where 1 indicates that the owner occupies the property, and 0 indicates non-owner occupied homes (i.e. purchases often in the anticipation of making a profit by reselling the properties after prices rise). Speculation is deemed riskier than owner occupancy.

FULLDOC - Categorical variable where 1 indicates that the borrower provided all the documentation required by the underwriting process, especially proof of monthly income, source of the funds for the down payment and closing cost, and other reports to determine his or her creditworthiness. A value of 0 means insufficient documentation and, in turn, higher risk (e.g. 'Alt-A' loans).

PURPOSE - Categorical variable where 1 indicates purchase and 0 indicates refinance. Purchase involves less risk than refinance. Daglish (2009) analyzes the relationship between refinance and default through a real options approach including periods when interest rates rises and the housing market declines in value.

FICO - Credit score based on the Fair Isaac and Company model which spans from 300 to 850 (a numerical grade of the credit history of the borrower). The lower the grade the greater the risk. The credit scores resulting from this model are broadly used in the USA. FICO scores above 680 correspond to ‘Prime’ borrowers. Borrowers with FICOs from 680 to 620 are considered A- borrowers (‘Alt-A’ for loans with unverifiable income). FICO scores below 620 place borrowers squarely in the ‘Subprime’ category (McElravey 2005).

UNRATE - Unemployment rate by state in the USA (from the US Bureau of Labor Statistics). The higher unemployment rate the higher risk.

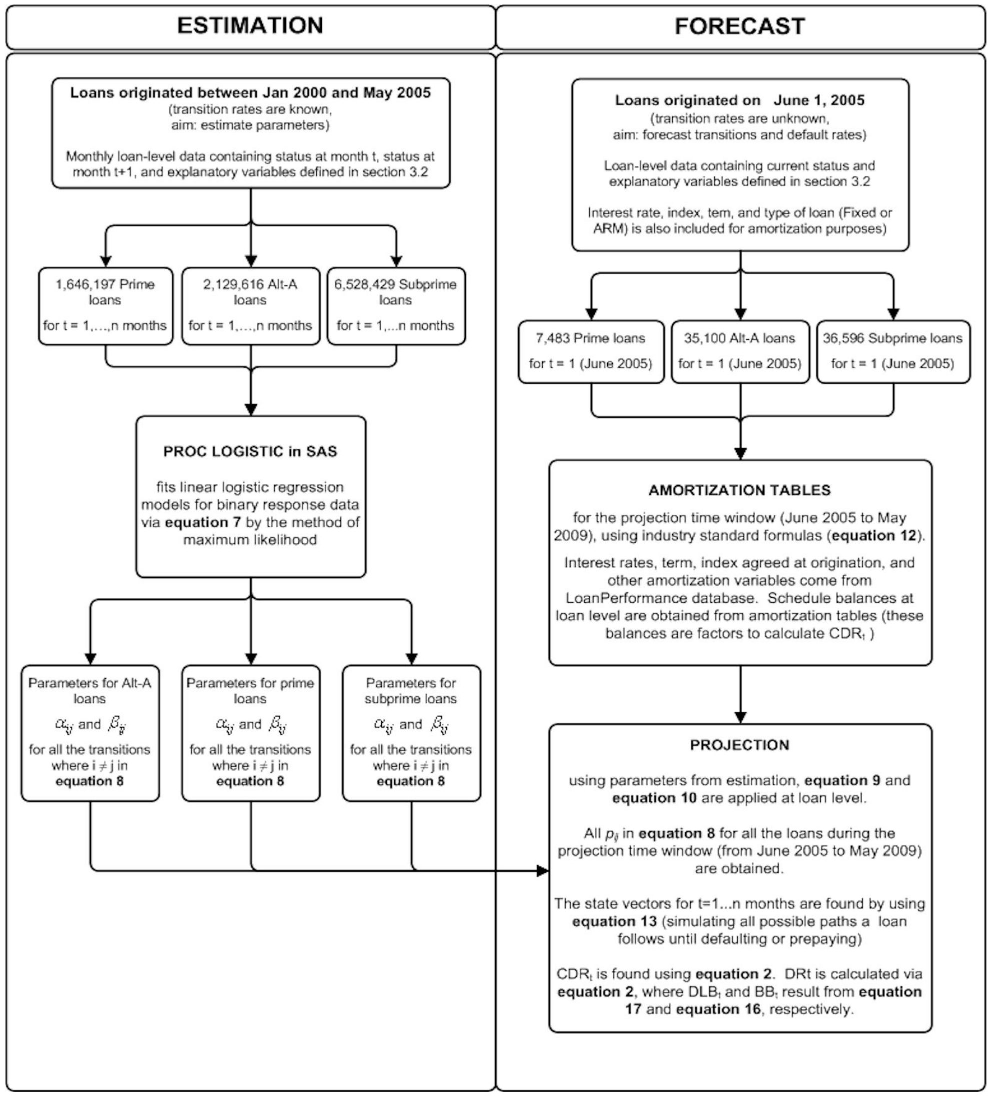
AGE, BALANCE, MTMCLTV, and FICO were grouped into data classes (binning process) to capture the outcome of specific ranges and reduce the effect of minor observation errors (e.g. outliers).

A total of 66 PROC LOGISTIC procedures[9] (off-diagonal p_{ij} elements in equation 8 for every asset class) were run iteratively until attaining the maximum likelihood estimates with p-values less than 0.05 (the probabilities p_{38}, p_{48}, and p_{58} were ignored for ‘Prime’ loans since for the historical period considered here, the number of transition to default from states prior to REO is insignificant). Appendix B and C tabulate the final output for all the transitions for ‘Alt-A’ and ‘Subprime’ assets that are statistically significant at the 95% level (the output for ‘Prime’ assets are omitted for brevity as the resultant CRD’s are not too different from the benchmark 100 SDA). A positive coefficient means that, *ceteris paribus*, the average value of the ratio of the transition probabilities increases as the value of the corresponding explanatory variable increases. It is clear that, consistent with Qi and Yang (2009), delinquency status and loan size relative to market value of the collateral (MTMCLTV) are the prime driver for default. Other important determinants are FICO score, the amount of documentation (FULLDOC, particularly for ‘Alt-A’ loans), whether the property is occupied by the owner (OWNER_OCCUPY), and whether the loan is for purchase or refinance (PURCHASE).

The Kolmogorov-Smirnov (K-S) statistic is employed to quantify the error (or ‘goodness of fit’) and for measuring predictive power (Anderson 2007, Thomas et al. 2002, D’Agostino and Massaro 1992). The K-S statistic calculates how far apart the distribution functions of the scores (transition rates) of

[9]PROC LOGISTIC is a SAS procedure that estimates binary logit models. See SAS (1999), SAS (1995) and Stokes et al. (2000) for explanations in detail. A recent application of PROC LOGISTIC for historical delinquency trends can be found in Demyanyk and Van Hemert (2010).

Figure 6: Estimation and Implementation the LTMA



the ’goods’ (desired outcome) and ’bads’ (undesired outcome) are. Most of the key transition models in Appendix B and C show K-S statistics ranging from 20% to 40%. A general rule for predictive models is that a K-S in the 20%s is good, in the 30%s is very good, and 40% or above is excellent. As a reference, using a transition mean rate from the past as a predictor would have a K-S of approximately zero per cent. Thus, a model with K-S of 5% is still superior to a simple mean (Mays 2004).

Once the regression parameters for all the transitions and asset classes are determined, the intercepts and coefficients in the 66 forecasting models were employed to predict CDRs simulating all the possible paths every loan from the population in Table 1 takes forward (monthly) for the projected 5-year time window. The process is undertaken as if *today* was 1 June 2005, without making *any* assumptions on either actual data for later periods or projected economic variables. This has two implication:

1. The beginning balance for month t (see BB_t in equation 1) was obtained from amortizing the loans over the projected time window. Using industry standard formulas, the following equation is applied to determine the beginning balance for any month t for an individual loan l_a, of asset type $a \in \{Prime, Alt-A, Subprime\}$, (Fabozzi 1997, Fabozzi 2010):

$$b_t^{(l_a)} = b_0^{(l_a)} \left[\frac{(1+r_{l_a}^{(t)})^{n_{l_a}} - (1+r_{l_a}^{(t)})^t}{(1+r_{l_a}^{(t)})^{n_{l_a}} - 1} \right], \tag{12}$$

   where $b_0^{(l_a)}$ is the original balance, $r_{l_a}^{(t)}$ the monthly interest rate (LoanPerformance provide the rate agreed at origination for each loan) and n_{l_a} is the original term of the loan in months. The beginning monthly balance for every loan is provided by LoanPerformance, and this data is employed in the analysis.

2. The $MTMCLTV$ variable was not affected by the changes in house prices for projections from 1 June 2005 forward since no projected HPI data are assumed. The same is assumed for unemployment rates. By doing so, the effect of the underwriting practices (borrower and loan characteristics) is isolated.

The computational framework produces forecast probabilities for *every* single mortgage (within an asset class) issued on the 1 June 2005 at any month t based on the following finite Markov process for loan l_a (Kemeny and Snell 1983):

$$\boldsymbol{\pi}_t^{(l_a)} = \boldsymbol{\pi}_{t-1}^{(l_a)} \cdot \mathbf{P}_{t-1}^{(a)}(l_a), \tag{13}$$

where $\boldsymbol{\pi}_t^{(l_a)}$ is the (1×8) state vector at month *t* for transient (first six elements) and absorbing (last two elements) states. It is noted that the initial state vector is $\boldsymbol{\pi}_0^{(l_a)}$=[1,0,0,0,0,0,0,0] for all loans since on the 1 June 2005 all loans have 'current' status. $\mathbf{P}_{t-1}^{(a)}(l_a)$ is the transition matrix with the form shown in equation 8 for loan l_a, that is the transition probabilities of moving from state i to state j between months $t-1$ and t. The time dependence of the transition matrix arises purely from the change in the age of the mortgage, which is of course, deterministic. The framework now allows for the determination of the default rates for the various asset classes of loans that originated on the 1 June 2005 (as well as all the loans together). The calculations and the connection to equation 9 and equation 10 is as follows. The proportion of the balance at the beginning of month t for loan l_a which is not in default or paid off is given by

$$bb_t^{(l_a)} = b_t^{(l_a)} \cdot \sum_{i=1}^{6} \pi_t^{(l_a)}(i), \tag{14}$$

whilst the proportion in default

$$dlb_t^{(l_a)} = b_t^{(l_a)} \cdot \pi_t^{(l_a)}(8) \tag{15}$$

(it is noted that $\pi_t^{(l_a)}(7)$, proportion of the loan paid off, provides information on prepayment rates which is not considered in this paper). The total balance not in default is obtained by summing over the individual loans

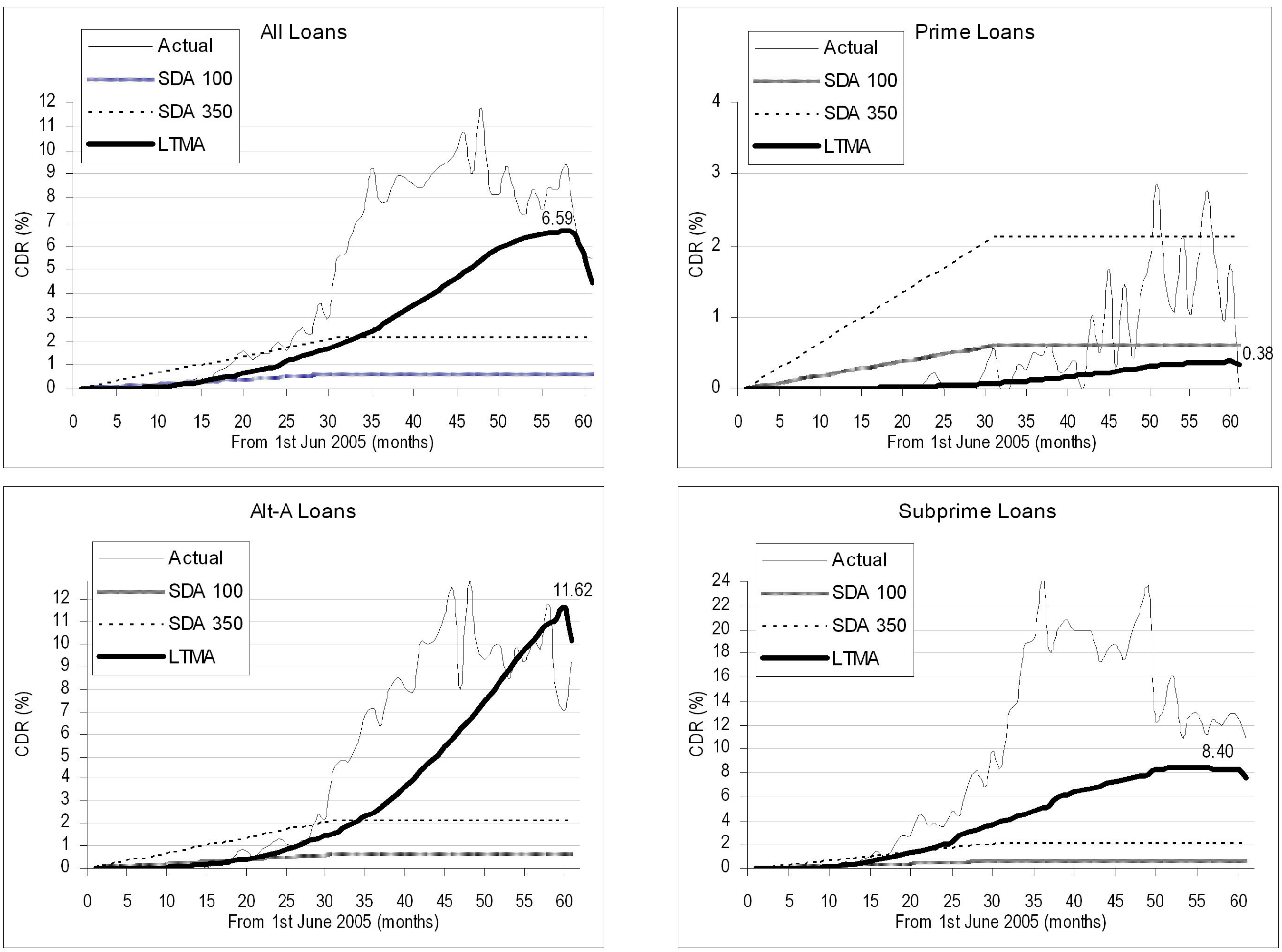
$$BB_t^{(a)} = \sum_{l_a} bb_t^{(l_a)}, \tag{16}$$

whilst the total balance in default is similarly determined

$$DLB_t^{(a)} = \sum_{l_a} dlb_t^{(l_a)}. \tag{17}$$

The conditional default rate ($CDR_t^{(a)}$) for the asset class is then obtained via equation 1. The CDR_t for the set containing all the loans originating on the 1 June 2005 is obtained in an analogous manner. For convenience, Figure 6 presents an illustrative flow diagram of both the estimation and forecast process.

Figure 7: Projected CDRs using the LTMA for loans originated on 1 June 2005. Also depicted are the actual CDRs and 100 and 350 SDA. The LTMA maximum prediction is also indicated.



Source: LoanPerformance (primary data for actual CDRs).

The numerical results are now presented. Figure 7 shows the CDR projection for the benchmark pool of mortgages from June 2005 to June 2010 obtained using the LTMA. The interpretation of these projections is as follows. The LTMA model has been calibrated using mortgage level data from January 2000 to May 2005. The mortgages issued on 1 June 2005, at the peak in US mortgage production and *before* the subprime crisis are the benchmark pool of mortgages to project forward CDRs, using the LTMA model, over the next five years assuming *only* data available on the 1 June 2005. Actual rates as well as projected CDRs using 100 SDA and 350 SDA are also plotted for comparison purposes. The LTMA *clearly* predicts the sharp increase in defaults from January 2007 (month 18) forward, particularly for 'Alt-A' and 'Subprime' assets. The maximum forecast CDRs for 'Alt-A' and Subprime' loans reached 1900 SDA and 1400 SDA respectively, and for all loans 1100 SDA which is greater than three times the stressed 350 SDA. A forecast CDR of 1100 SDA would have been a *significant* prediction viewed from the standpoint of June 2005. This is the key result of the paper and is the basis for the conclusion that the subprime crisis was predictable.

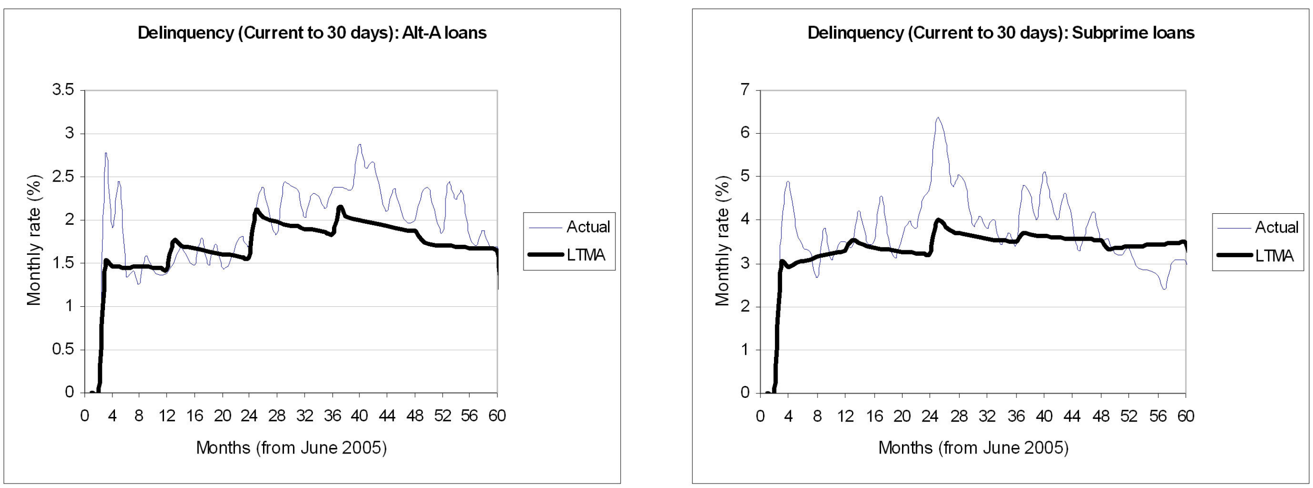
In figure 7, the 'Alt-A' projection is closer to the actual rates than the prediction for 'Subprime' loans. This is due to the strong effect of the borrower and loan characteristics (mainly lack of documentation and non-occupancy) on the former, relative to a greater weight of exogenous variables (e.g. interest rate resets, house depreciation, and decreasing income) affecting the lower-income borrowers of the latter[10] during the economic recession of 2007. Some authors (Hayre et al. 2008) suggest adding a weighting factor to subprime models to capture the abnormally high default rates of subprime loans from early 2007 to Summer 2009. They argue that (1) the weakening economy deteriorated the situation for struggling borrowers, (2) servicers decided to cut their losses when home prices fell and provided incentives to delinquent borrower undertake short sales (that is, sell the home for less than the mortgage balance), and (3) declining home prices had a psychological impact on struggling homeowners who decided that it was not worthwhile to try to hold on to the property. It is not, however, the aim in the LTMA model to adjust its predictions (i.e. by adding external weights); the aim is to predict with available information as of 1 June 2005. Although the LTMA understimates the actuals CDRs, particularly for 'Subprime' loans,

[10]See Mayer et al. (2009) for an analysis of the macroeconomic factors of the mortgage credit crisis in the USA.

both the pattern and the upward trend of actual CDRs is clearly captured. As for 'Prime' loans, the 100 SDA and LTMA give similar projections since these experienced low incidences of default during the 5-year time window (less than 2.9%).

The LTMA also predicts delinquency rates for all the transitions in the matrix of equation 8. It is particularly important for researchers and practitioners to identify potential defaults from loans transitioning to delinquency states, especially borrowers missing one payment after making all the previous payments on time. Figure 8 shows the projected and actual monthly transition rates from current to 30 days delinquent for 'Alt-A' and 'Subprime' loans in table 1 from June 2005 to June 2010.

Figure 8: Monthly transition rates from current to 30 days delinquent for loans originated on June 1, 2005 (predicted vs actual rates)



Source: LoanPerformance (primary data for actual rates).

As shown, the LTMA model predicts the current to 30 day delinquent transition rates well across time for both 'Alt-A' and 'Subprime' loans. The spikes from month 3 to month 5 in the actual rates indicate the presence of very poor or even fraudulent underwriting practices[11]. Current to 30 delinquency rates for traditional loans are generally lower than 1.7% (Brown and Wadden 2001). The dynamic between seriously delinquent states and default is left for future research.

[11]Known as EPDs (Early Payment Delinquency); a significant problem of the non-prime market has been loans that go to delinquency after making just one or two payments (Hayre and Saraf 2008).

4. Conclusions

The subprime crisis highlighted significant failures in financial modeling. Standard mortgage risk modeling failed to predict the high default rates that occurred from late 2006 in the USA and triggered the subprime crisis. These models, mainly based on the seasonal nature of default, overlooked the empirical evidence for underlying borrower and loan characteristics, especially for non-traditional loans with high-risk features (i.e. 'Alt-A' and 'Subprime'). Default projections would have been closer to actual rates if underlying borrower and loan characteristics such as age, balance, loan-to-value, occupancy, documentation, loan purpose, and credit score had been taken into consideration. The LTMA model developed in this paper and applied to a benchmark set of non-agency mortgages that originated at the height of the underwriting excess on 1 June 2005 (and prior to the subprime crisis) is shown to provide a much better predictor of default rates using data *only* available on 1 June 2005. In fact, the LTMA prediction for CDRs attained maximums of 1900 SDA and 1400 SDA for 'Alt-A' and 'Subprime' loans respectively. This is a significant prediction from the standpoint of June 2005.

It is demonstrated, therefore, that modeling mortgage defaults under extreme economic conditions requires sophisticated techniques based on loan level data, with the LTMA model providing a promising approach. This is the key conclusion of this paper; irrespective of the effect of other economic variables (e.g. interest rate resets, house depreciation, decreasing income), both the pattern and upward trend of actual default rates from late 2006 to May 2010 could have been detected on 1 June 2005. The answer, therefore, to the question posed in the title of this paper is yes, the subprime crisis could have been predicted.

Acknowledgements

The authors thank First American CoreLogic for authorizing the use of the First American CoreLogic LoanPerformance database for this research. They also thank Jerry Coakley, Wing Lon Ng, and John Nankervis for useful comments that made the paper clearer. Any opinions expressed or implied are

solely those of the authors.

References

Allison, P. D., 2006. Logistic Regression Using SAS: Theory and Application. Cary, NC: SAS Institute Inc.

An, X., and Deng, Y., 2007. Subordination Levels in Commercial Mortgage-backed Securities (CMBS). http://www.reri.org/research/article_pdf/wp136.pdf.

Anderson, R., 2007. The Credit Scoring Toolkit, Theory and Practice for Retail Credit Risk Management and Decision Automation. Oxford University Press:UK.

D'Agostino, R. B., and Massaro, J. M., 1992. Goodness-of-Fit Tests. In Balakrishnan, N. editor, Handbook of The Logistic Distribution, Statistics: Textbooks and Monographs, Marcel Dekker, Inc. NY.

Begg, C. B. and Gray, R., 1984. Calculation of Polychotomous Logistic Regression Parameters Using Individualized Regressions. Biometrika, 71, 11-18.

Bhattacharya, A. K., and Berliner, W. S., 2005. An Overview of Mortgages and the Mortgage Market. In Fabozzi, F. J. and Mann, S. V., The Handbook of Fixed Income Securities, McGraw-Hill,487-512.

The Bond Market Association, 1999. Standard Formulas for the Analysis of Mortgage-Backed Securities and Other Related Securities, 2nd edition, The Bond Market Association.

Brown, J. W., and Wadden, W. M., 2001. Mortgage Credit Analysis. In Fabozzi, F. J. (Ed.), The Handbook of Mortgage-Backed Securities, McGraw-Hill,315-328.

Calhoun, C. A. and Deng, Y., 2002. A Dynamic Analysis of Fixed- and Adjustable-Rate Mortgage Terminations. Journal of Real Estate Finance and Economics, 24, 9-33.

Campbel, T. S., and Dietrich, J. K., 1983. The Determinants of Default on Insured Conventional Residential Mortgage Loans. Journal of Finance, 38, 1569-1581.

Crouhy, M. G., Jarrow, R. A., and Turnbull, S. M., 2008. The Subprime Credit Crisis of 2007. Journal of Derivatives, 16, 81-110.

Daglish, T., 2009. What motivates a subprime borrower to default. Journal of Banking and Finance, 33, 681-693.

Deng, Y., and Gabriel, S. A., 2006, Risk-Based Pricing and the Enhancement of Mortgage Credit Availability among Undeserved and Higher Credit-Risk Populations. Journal of Money, Credit, and Banking, 38, 1431-1460.

Demyanyk, Y., and Van Hemert, H. O., 2010. Understanding the Subprime Mortgage Crisis. Review of Financial Studies, doi:10.1093/rfs/hhp033.

Fabozzi, F. J., 2010. Bond Markets, Analysis and Strategies, 7th edition, Pearson Education Inc., Upper Saddle River, New Jersey.

Fabozzi, F. J., 1997. Fixed Income Mathematics, Analytical and Statistical Techniques, 3rd edition, McGraw-Hill.

Fabozzi, F. J., Bhattacharya, A. K., and Berliner, W. S., 2007. Mortgage-Backed Securities: Products, Structuring and Analytical Techniques, The Frank J. Fabozzi Series, John Wiley & Sons, Inc., Hoboken, New Jersey.

Fabozzi, F. J., and Dunlevy, J. N., 2001. Real Estate-Backed Securities, Frank J. Fabozzi Associates, New Hope, Pennsylvania.

Goodman, L. and Lowell L., 2000. A risk-Return Framework for Evaluating Non-Investment-Grade Subordinated MBS. in Fabozzi, F. J., Ramsey, C., and Marz, M., The Handbook of Nonagency MortgageBacked Securities, 2nd edition, Frank J. Fabozzi Associates, New Hope, Pennsylvania.

Goodman, L. S., Li, S., Lucas, D. J., Zimmerman, T. A., and Fabozzi, F. J., 2008. Subprime Mortgage Credit Derivatives, The Frank J. Fabozzi Series, John Wiley & Sons, Inc., Hoboken, New Jersey.

Gramlich, E. M., 2007. Subprime Mortgages: America's Latest Boom and Bust, The Urban Institute Press, Washington, D.C.

Hayre, L., 2001. A Concise Guide to Mortgage-Backed Securities (MBSs). In Hayre, L. (Ed.), Salomon Smith Barney Guide to Mortgage-Backed and Asset-Backed Securities, John Wiley & Sons, Inc., New York, NY, 9-68.

Hayre, L., Chaudhary, S., and Young, R., 2000. Anatomy of Prepayments. The Journal of Fixed Income, 10 (1), 19-31.

Hayre, L., and Saraf, M., 2008. A Loss Severity Model for Residential Mortgages. The Journal of Fixed Income, 18(2), 5-31.

Hayre, L., Saraf, M., Young, R, and Chen, J., 2008. Modeling of Mortgage Defaults. The Journal of Fixed Income, 17(4), 6-30.

Hosmer, D. W., and Lemeshow, S., 2000. Applied Logistic Regression, 2nd edition, Wiley Inter-Science, John Wiley & Sons, Inc., Hoboken, New Jersey.

Kaskowitz, D., Kipkalov, A., Lundstedt, K, and Mingo, J., 2002. Best-Practices in Mortgage Default Risk Measurement and Economic Capital. The RMA Journal, 84, 80-85.

Kemeny, J. G., and Snell J. L., 1983. Finite Markov Chains. Springer-Verlag, Princeton, NJ.

Kramer, B., and Sinha, G., 2006. Bear Stearns Quick Guide to NonAgency MortgageBacked Securities. Bear Stearns & FAST Fixed Income Research.

Lando, D., 2004. Credit Risk Modeling. Princeton Series in Finance, Princeton University Press, Princeton, NJ.

Mayer, C., Pence, K., and Sherlund, S. M., 2009. The Rise in Mortgage Defaults. Journal of Economic Perspectives, 23, 27-50.

Mays, E., 2004. Credit Scoring for Risk Managers: The Handbook for Lenders. South-Western Publishing: Mason, OH.

McElravey, J., 2005. Residential Asset-Backed Securities. In Fabozzi, F. J. editor with Mann, S. V., The Handbook of Fixed Income Securities, McGraw-Hill.

Qi, M and Yang, X., 2009. Loss given default of high loan-to-value residential mortgages. Journal of Banking and Finance, 33,788-799.

SAS Institute Inc., 1995. Logistic Regression Examples Using the SAS System, Version 6, First Edition. Cary, NC: SAS Institute Inc.

SAS Institute Inc., 1999. SAS/STAT User's Guide, Version 8. Cary, NC: SAS Institute Inc.

Smith, L. D., and Lawrence, E. C., 1995. Forecasting losses on liquidating long-term loan portfolio. Journal of Banking and Finance, 19, 959-985

Smith, L. D., Sanchez, S. M., and Lawrence, E. C., 1996. A Comprehensive Model for Managing Credit Risk on Home Mortgage Portfolios. Decision Sciences, 27, 291-317.

SMR Research Corporation 2007. 2007 - The Mortgage Credit Crisis. SMR Research Corporation, Hackettstown, NJ.

Stokes, M. E., Davis, C. S., and Koch, G. G., 2000. Categorical Data Analysis Using The SAS System, 2nd edition. Cary, NC: SAS Institute Inc., 183-322.

Thomas, L. C., Edelman, D. B., and Crook, J. N., 2002. Credit Scoring and Its Applications. Monographs on Mathematical Modeling and Computation, Society for Industrial and Applied Mathematics (SIAM), PA.

Westgaard, S., and Van der Wijst, N., 2001. Default probabilities in a corporate bank portfolio: A logistic model approach. European Journal of Operational Research, 135, 338-349.

Zimmerman, T., and Neelakantan, K., 2001. Credit Performance of High LTV Loans. In Fabozzi, F. J. editor, The Handbook of Mortgage-Backed Securities, McGraw-Hill, 329-362.

Appendix A. Definition of Explanatory Variables

	Definition		
Variable	Prime	Alt-A	Subprime
AGE	Current Date - First Paymt Date (mths)	Current Date - First Paymt Date (mths)	Current Date - First Paymt Date (mths)
AGE1	0 < AGE <= 12	0 < AGE <= 12	0 < AGE <= 12
AGE2	12 < AGE <= 24	12 < AGE <= 24	12 < AGE <= 24
AGE3	24 < AGE <= 36	24 < AGE <= 36	24 < AGE <= 36
AGE4	36 < AGE <= 48	36 < AGE <= 48	36 < AGE <= 48
AGE5	48 < AGE <= 60	48 < AGE <= 60	48 < AGE <= 60
AGE6	60 < AGE	60 < AGE	60 < AGE
BAL1	CLOSE_BAL <= 400000	CLOSE_BAL <= 50000	CLOSE_BAL <= 50000
BAL2	400000 < CLOSE_BAL <= 500000	50000 < CLOSE_BAL <= 100000	50000 < CLOSE_BAL <= 100000
BAL3	500000 < CLOSE_BAL <= 600000	100000 < CLOSE_BAL <= 150000	100000 < CLOSE_BAL <= 150000
BAL4	600000 < CLOSE_BAL <= 700000	150000 < CLOSE_BAL <= 200000	150000 < CLOSE_BAL <= 200000
BAL5	700000 < CLOSE_BAL <= 800000	200000 < CLOSE_BAL <= 250000	200000 < CLOSE_BAL <= 250000
BAL6	800000 < CLOSE_BAL <= 900000	250000 < CLOSE_BAL <= 300000	250000 < CLOSE_BAL <= 300000
BAL7	900000 < CLOSE_BAL <= 1000000	300000 < CLOSE_BAL <= 350000	300000 < CLOSE_BAL <= 350000
BAL8	1000000 < CLOSE_BAL	350000 < CLOSE_BAL <= 500000	350000 < CLOSE_BAL
BAL9	Not Defined	500000 < CLOSE_BAL	Not Defined
CLOSE_BAL	Principal Balance at Origination	Principal Balance at Origination	Principal Balance at Origination
FICO1	FICO <= 580	FICO <= 540	FICO <= 480
FICO2	580 < FICO <= 620	540 < FICO <= 580	480 < FICO <= 540
FICO3	620 < FICO <= 660	580 < FICO <= 620	540 < FICO <= 600
FICO4	660 < FICO <= 700	620 < FICO <= 660	600 < FICO <= 640
FICO5	700 < FICO <= 740	660 < FICO <= 700	640 < FICO <= 680
FICO6	740 < FICO	700 < FICO <= 720	680 < FICO
FICO7	Not Defined	720 < FICO	Not Defined
FULLDOC	Full Borrower Documentation	Full Borrower Documentation	Full Borrower Documentation
MTMCLTV	Mark to Market CLTV	Mark to Market CLTV	Mark to Market CLTV
MTMCLTV_1	MTMCLTV <= .65	MTMCLTV <= .65	MTMCLTV <= .65
MTMCLTV_2	.65 < MTMCLTV <= .95	.65 < MTMCLTV <= .95	.65 < MTMCLTV <= .95
MTMCLTV_3	.95 < MTMCLTV <= 1.05	.95 < MTMCLTV <= 1.05	.95 < MTMCLTV <= 1.15
MTMCLTV_4	1.05 < MTMCLTV <= 1.25	1.05 < MTMCLTV <= 1.25	1.15 < MTMCLTV <= 1.45
MTMCLTV_5	1.25 < MTMCLTV <= 1.45	1.25 < MTMCLTV <= 1.45	1.45 < MTMCLTV
MTMCLTV_6	1.45 < MTMCLTV	1.45 < MTMCLTV	Not Defined
OWNER_OCCUPY	Owner Occupancy	Owner Occupancy	Owner Occupancy
PURCHASE	Loan Purpose (Not Refinance)	Loan Purpose (Not Refinance)	Loan Purpose (Not Refinance)
UNRATE	Not included	Unemployment Rate by State	Unemployment Rate by State

Appendix B. Parameter estimates for Alt-A loans using logistic regression models (transitions p_{12} to p_{38}). The iterative process was implemented in SAS using PROC LOGISTIC. All p-values are significant (less than 0.05).

PARAMETER	p_{12}	p_{17}	p_{21}	p_{23}	p_{27}	p_{31}	p_{32}	p_{34}	p_{35}	p_{37}	p_{38}
Intercept	-5.1166	-4.6378	2.5644	0.1745	-1.3283	2.7382	0.6606	2.6998	1.1443	-0.1988	-7.3102
	(0.0185)	(0.0141)	(0.0365)	(0.0454)	(0.0680)	(0.1038)	(0.0756)	(0.0946)	(0.1542)	(0.1264)	(0.4030)
AGE											0.0316
											(0.0139)
AGE1	0.0144	0.1416	-0.0990	-0.0122	0.0515	-0.1023		-0.0621	-0.0396		
	(0.000928)	(0.000807)	(0.00215)	(0.00291)	(0.00487)	(0.00761)		(0.00690)	(0.0121)		
AGE2	0.4102	1.8738	-1.2315	-0.0881	0.6249	-1.4481	-0.1645	-0.8697	-0.7815	-0.2362	
	(0.00798)	(0.00730)	(0.0188)	(0.0255)	(0.0437)	(0.0684)	(0.0261)	(0.0623)	(0.1087)	(0.0415)	
AGE3	0.7337	2.0735	-1.4863	-0.1298	0.4898	-1.7428	-0.3286	-1.1160	-1.1687	-0.3666	
	(0.00976)	(0.00845)	(0.0207)	(0.0274)	(0.0470)	(0.0712)	(0.0307)	(0.0640)	(0.1141)	(0.0508)	
AGE4	0.9348	2.5482	-1.5870	-0.1757	0.8644	-1.8138	-0.3139	-1.1577	-1.3931		
	(0.0144)	(0.0110)	(0.0264)	(0.0333)	(0.0536)	(0.0813)	(0.0462)	(0.0715)	(0.1362)		
AGE5	0.9186	2.8525	-1.5095	-0.3163	1.4345	-1.7985	-0.4148	-1.2579	-1.9992		
	(0.0333)	(0.0217)	(0.0537)	(0.0665)	(0.0837)	(0.1404)	(0.1151)	(0.1208)	(0.3068)		
AGE6		2.3564	-1.6975								
		(0.1587)	(0.2920)								
BAL1	-0.0632	-1.0378	-0.1527	-0.2202	-0.8428			-0.2160	-0.4251	-0.5576	
	(0.0137)	(0.0117)	(0.0272)	(0.0336)	(0.0489)			(0.0642)	(0.1061)	(0.1079)	
BAL2	-0.0345	-0.8221	-0.2894	-0.3947	-0.8478	-0.1953		-0.4104	-0.4018	-0.6682	
	(0.00999)	(0.00735)	(0.0203)	(0.0254)	(0.0249)	(0.0357)		(0.0475)	(0.0570)	(0.0542)	
BAL3	-0.0561	-0.5565	-0.3052	-0.4423	-0.5610	-0.2470		-0.4597	-0.3657	-0.3623	
	(0.00978)	(0.00690)	(0.0200)	(0.0252)	(0.0222)	(0.0352)		(0.0468)	(0.0543)	(0.0478)	
BAL4	-0.0674	-0.3283	-0.2655	-0.4155	-0.2538	-0.2712		-0.4849	-0.4421		
	(0.0102)	(0.00708)	(0.0209)	(0.0264)	(0.0239)	(0.0397)		(0.0493)	(0.0637)		
BAL5	-0.0592	-0.1743	-0.2275	-0.3548		-0.3313		-0.4304	-0.4130		
	(0.0110)	(0.00757)	(0.0224)	(0.0283)		(0.0461)		(0.0527)	(0.0737)		
BAL6	-0.0271	-0.0611	-0.2137	-0.3027		-0.2810		-0.3952	-0.1752		
	(0.0114)	(0.00774)	(0.0234)	(0.0297)		(0.0493)		(0.0551)	(0.0743)		
BAL7	-0.0571	0.0344	-0.1547	-0.2158		-0.2511		-0.3611			
	(0.0118)	(0.00764)	(0.0241)	(0.0303)		(0.0514)		(0.0565)			
BAL8	-0.0957	0.0381	-0.1010	-0.1855				-0.2039			
	(0.0105)	(0.00673)	(0.0218)	(0.0276)				(0.0522)			
CLOSE_BAL											
FICO1	2.5070	0.3357	-1.0323	0.2216	-0.8632	-0.8063		-0.8386			
	(0.0342)	(0.0546)	(0.0567)	(0.0591)	(0.1315)	(0.1288)		(0.1095)			
FICO2	2.2041	0.3341	-0.9570	0.1398	-0.6219	-0.6427		-0.6774	-0.3873		
	(0.0213)	(0.0292)	(0.0357)	(0.0378)	(0.0741)	(0.0850)		(0.0708)	(0.1354)		
FICO3	1.9445	0.2146	-0.7562		-0.6963	-0.5943		-0.4968	-0.7295	-0.4903	
	(0.00906)	(0.0105)	(0.0164)		(0.0342)	(0.0452)		(0.0342)	(0.0718)	(0.0698)	
FICO4	1.4578	0.1825	-0.5892		-0.5472	-0.4919	-0.0545	-0.4094	-0.4077	-0.3258	
	(0.00623)	(0.00503)	(0.0123)		(0.0241)	(0.0368)	(0.0222)	(0.0264)	(0.0490)	(0.0487)	
FICO5	0.8489	0.1134	-0.3422		-0.3293	-0.2262		-0.2183	-0.1271	-0.1016	
	(0.00600)	(0.00401)	(0.0122)		(0.0235)	(0.0374)		(0.0276)	(0.0498)	(0.0495)	
FICO6	0.4282	0.0537	-0.1591		-0.1412	-0.1206					
	(0.00786)	(0.00500)	(0.0161)		(0.0307)	(0.0504)					
FULLDOC	-0.3709	-0.1059	0.1310	-0.1491	-0.0875			-0.0934		-0.2446	0.7701
	(0.00464)	(0.00349)	(0.00878)	(0.0113)	(0.0184)			(0.0212)		(0.0440)	(0.3193)
MTMCLTV	0.7683	0.0989	-0.6749	0.1216	-0.7439	-0.8404	-0.3407			-0.9402	
	(0.0119)	(0.00821)	(0.0198)	(0.0249)	(0.0364)	(0.0535)	(0.0512)			(0.0811)	
MTMCLTV_1											
MTMCLTV_2											
MTMCLTV_3											
MTMCLTV_4											
OWNER_OCCUPY	0.0913	0.0110	-0.4018	-0.3105	-0.4635	-0.4781	-0.1616	-0.6883	-0.6128	-0.6925	
	(0.00587)	(0.00438)	(0.0120)	(0.0153)	(0.0238)	(0.0354)	(0.0366)	(0.0314)	(0.0556)	(0.0573)	
PURCHASE	0.2699	0.3285	-0.0990	0.1261	0.1028	-0.0693	0.0484			0.1045	
	(0.00452)	(0.00346)	(0.00834)	(0.0105)	(0.0169)	(0.0235)	(0.0224)			(0.0384)	
UNRATE	-0.1039	-0.0656	0.0393	-0.0475	0.0959	-0.0310	-0.0796	-0.0538	-0.1902	0.0588	
	(0.00210)	(0.00161)	(0.00396)	(0.00498)	(0.00803)	(0.0111)	(0.0107)	(0.00931)	(0.0117)	(0.0184)	
K-S Statistic (%)	28.91	26.32	21.83	7.63	21.00	21.04	7.52	18.42	23.13	21.30	33.77

For each parameter estimated, one degree of freedom (DF) is required, and the DF defines the Chi-Square distribution to test whether the individual regression coefficient is zero given the other variables are in the models.

Standard errors of the individual regression coefficients are given in brackets.

Appendix B. Parameter estimates for Alt-A loans using logistic regression models (transitions p_{41} to p_{68}). The iterative process was implemented in SAS using PROC LOGISTIC. All p-values are significant (less than 0.05).

PARAMETER	p_{41}	p_{42}	p_{43}	p_{45}	p_{47}	p_{48}	p_{51}	p_{54}	p_{56}	p_{57}	p_{58}	p_{68}
Intercept	-0.5541 (0.1414)	-2.1030 (0.1981)	-3.0154 (0.0957)	-0.4846 (0.0893)	-1.1297 (0.1294)	-8.3055 (1.0967)	-2.1356 (0.2403)	0.3520 (1.2347)	-4.3430 (0.2568)	-3.5750 (0.1115)	-5.8654 (0.1518)	-3.7674 (0.1458)
AGE								-0.00486 (0.00160)		0.0137 (0.00165)		0.0289 (0.00236)
AGE1	-0.1274 (0.0121)	-0.0994 (0.0196)		0.0181 (0.00841)	-0.0878 (0.00943)	0.2238 (0.1076)	-0.1027 (0.0216)		0.1223 (0.0240)			
AGE2	-1.8747 (0.1111)	-1.3122 (0.1795)	-0.2376 (0.0499)	-0.2136 (0.0805)	-1.1794 (0.0842)	2.8735 (1.0987)	-1.5114 (0.2138)		1.4936 (0.2499)		0.4559 (0.1540)	
AGE3	-2.2543 (0.1142)	-1.5773 (0.1825)	-0.3700 (0.0555)	-0.4767 (0.0817)	-1.2652 (0.0861)	3.1181 (1.0994)	-1.7134 (0.2154)		1.5963 (0.2504)		0.6321 (0.1564)	
AGE4	-2.1900 (0.1236)	-1.6720 (0.1956)	-0.5835 (0.0813)	-0.5560 (0.0860)	-0.6931 (0.0912)	3.1857 (1.1069)	-1.5934 (0.2218)		1.7047 (0.2527)		0.6679 (0.1830)	
AGE5	-2.4124 (0.1995)	-1.2774 (0.2532)	-0.6680 (0.1920)	-0.7389 (0.1198)		2.6261 (1.2383)	-1.7675 (0.2851)		1.4607 (0.2805)		0.9693 (0.3235)	
AGE6				-1.4783 (0.6110)								
BAL1					-0.8408 (0.0976)				0.2522 (0.0587)			
BAL2					-0.6220 (0.0490)				0.3211 (0.0344)			
BAL3					-0.4250 (0.0456)		0.1928 (0.0531)		0.3124 (0.0356)		0.3671 (0.0996)	0.1132 (0.0538)
BAL4	0.2584 (0.0456)						0.1940 (0.0629)				0.4087 (0.1217)	0.1570 (0.0718)
BAL5	0.1445 (0.0579)			0.1174 (0.0329)			0.1800 (0.0737)				0.5692 (0.1416)	
BAL6	0.2674 (0.0587)			0.1156 (0.0357)			0.2036 (0.0755)					
BAL7				0.1686 (0.0372)			0.2607 (0.0754)					
BAL8				0.0983 (0.0299)								
CLOSE_BAL			-2.62E-7 (1.025E-7)			5.549E-7 (1.774E-7)				2.867E-7 (6.717E-8)		
FICO1	-0.5840 (0.1930)			-0.6666 (0.1124)				1.3694 (0.1588)				
FICO2	-0.3487 (0.1263)		0.3127 (0.1223)	-0.4925 (0.0756)			0.5381 (0.1443)	0.8861 (0.1238)				
FICO3	-0.5381 (0.0585)			-0.4546 (0.0348)	-0.4688 (0.0658)	-0.6796 (0.2099)	-0.1654 (0.0681)	0.4435 (0.0620)		-0.3278 (0.0602)		-0.2077 (0.0732)
FICO4	-0.4324 (0.0420)			-0.4054 (0.0276)	-0.2779 (0.0444)	-0.4306 (0.1330)	-0.0961 (0.0430)	0.3397 (0.0502)	-0.1420 (0.0287)	-0.1343 (0.0357)		-0.1156 (0.0490)
FICO5	-0.2151 (0.0426)			-0.2099 (0.0281)	-0.1321 (0.0454)			0.2629 (0.0515)				
FICO6				-0.1185 (0.0373)				0.1557 (0.0695)				
FULLDOC	0.1786 (0.0363)			0.1801 (0.0200)	-0.1584 (0.0422)	0.4183 (0.1258)		-0.4795 (0.0380)	-0.1364 (0.0298)		0.8885 (0.0825)	0.6472 (0.0457)
MTMCLTV	-0.9295 (0.0624)	-0.4349 (0.0899)	-0.3052 (0.0781)	-0.2547 (0.0368)	-0.6816 (0.0677)		-1.1360 (0.0725)		0.5553 (0.0670)	-0.7648 (0.0663)		-0.4255 (0.0792)
MTMCLTV_1								-3.3382 (1.2336)				
MTMCLTV_2								-3.5399 (1.2335)				
MTMCLTV_3								-3.6718 (1.2366)				
MTMCLTV_4								-4.8887 (1.5926)				
OWNER_OCCUPY		0.3205 (0.0728)	0.5498 (0.0666)	-0.2079 (0.0239)	-0.3070 (0.0505)		0.3799 (0.0584)	0.4363 (0.0460)	-0.1687 (0.0347)			
PURCHASE	-0.1312 (0.0327)				0.1016 (0.0360)					0.1540 (0.0349)		
UNRATE	0.1177 (0.0155)				0.0950 (0.0168)		0.1358 (0.0191)			0.1449 (0.0159)		0.0749 (0.0207)
K-S Statistic (%)	21.89	13.40	8.95	14.27	21.03	23.51	18.95	15.13	11.99	13.26	23.85	19.19

For each parameter, one degree of freedom (DF) is required, and the DF defines the Chi-Square distribution to test whether the individual regression coefficient is zero given the other variables are in the models.

Standard errors of the individual regression coefficients are given in brackets.

Appendix C. Parameter estimates for subprime loans using logistic regression models (transitions p_{12} to p_{38}). The iterative process was implemented in SAS using PROC LOGISTIC. All p-values are significant (less than 0.05).

PARAMETER	p_{12}	p_{17}	p_{21}	p_{23}	p_{27}	p_{31}	p_{32}	p_{34}	p_{35}	p_{37}	p_{38}
Intercept	-5.1689	-5.2876	1.7230	0.7692	-1.5153	2.0292	0.6434	3.0217	2.5248	-0.1583	-5.5969
	(0.0560)	(0.00835)	(0.0109)	(0.1151)	(0.0292)	(0.0306)	(0.0260)	(0.0262)	(0.0409)	(0.0471)	(0.2142)
AGE											
AGE1	0.0381	0.1743	-0.0724	-0.0304	0.0521	-0.0705	-0.0222	-0.0844	-0.0723	-0.0377	
	(0.000327)	(0.000474)	(0.000651)	(0.000810)	(0.00169)	(0.00206)	(0.00206)	(0.00182)	(0.00250)	(0.00381)	
AGE2	0.5574	2.1145	-0.9193	-0.3638	0.5358	-1.0038	-0.3241	-1.0667	-1.1410	-0.4492	
	(0.00283)	(0.00441)	(0.00574)	(0.00713)	(0.0155)	(0.0187)	(0.0188)	(0.0164)	(0.0226)	(0.0344)	
AGE3	0.8009	2.3038	-1.1589	-0.4134	0.3979	-1.3556	-0.4747	-1.2625	-1.6217	-0.5676	
	(0.00325)	(0.00479)	(0.00618)	(0.00754)	(0.0163)	(0.0193)	(0.0192)	(0.0168)	(0.0237)	(0.0355)	
AGE4	0.8652	2.2162	-1.2586	-0.4637	0.2283	-1.5671	-0.6009	-1.4166	-1.8962	-0.7075	
	(0.00411)	(0.00586)	(0.00715)	(0.00849)	(0.0186)	(0.0209)	(0.0202)	(0.0179)	(0.0259)	(0.0384)	
AGE5	0.7894	2.0308	-1.2739	-0.5054	0.1244	-1.6299	-0.6880	-1.5415	-2.1389	-0.6660	
	(0.00613)	(0.00872)	(0.00962)	(0.0111)	(0.0251)	(0.0252)	(0.0234)	(0.0211)	(0.0329)	(0.0460)	
AGE6	0.6785	1.8800	-1.2726	-0.5835		-1.6952	-0.7280	-1.7129	-2.4481	-0.7571	
	(0.0232)	(0.0328)	(0.0338)	(0.0386)		(0.0808)	(0.0676)	(0.0660)	(0.1205)	(0.1498)	
BAL1	0.0338	-0.3749		0.0676	-0.8248				-0.1999	-0.8521	
	(0.00259)	(0.00436)		(0.00742)	(0.0156)				(0.0192)	(0.0192)	
BAL2	0.1292	-0.4922		-0.0733	-0.7612			-0.2673	-0.0449	-0.8715	
	(0.00225)	(0.00425)		(0.00702)	(0.0145)			(0.00669)	(0.0176)	(0.0154)	
BAL3	0.0866	-0.2799		-0.1076	-0.4060			-0.3653	-0.1633	-0.5639	
	(0.00241)	(0.00432)		(0.00723)	(0.0146)			(0.00771)	(0.0182)	(0.0158)	
BAL4	0.0274	-0.0610		-0.1058	-0.0994			-0.3564	-0.1225	-0.2207	
	(0.00279)	(0.00447)		(0.00781)	(0.0151)			(0.0101)	(0.0197)	(0.0177)	
BAL5		0.0689		-0.0821	0.0718			-0.2882	-0.1258		
		(0.00479)		(0.00883)	(0.0161)			(0.0134)	(0.0224)		
BAL6		0.0993		-0.0485	0.1277			-0.2499	-0.0771		
		(0.00525)		(0.0101)	(0.0176)			(0.0171)	(0.0255)		
BAL7		0.0619			0.1307			-0.1983			
		(0.00599)			(0.0203)			(0.0218)			
CLOSE_BAL			-3.26E-7			-8.49E-7	-6.55E-7	-9.55E-7			-2.31E-6
			(1.394E-8)			(3.604E-8)	(3.315E-8)	(3.872E-8)			(7.41E-7)
FICO1	2.7331	-0.2840	-0.9786	-0.4126	-1.2215	-0.9223	-0.0434	-1.0045	-0.9933	-0.9091	-1.5849
	(0.00787)	(0.0175)	(0.0119)	(0.0130)	(0.0304)	(0.0294)	(0.0194)	(0.0226)	(0.0379)	(0.0526)	(0.5123)
FICO2	2.3725	-0.00896	-0.7343	-0.3383	-0.8768	-0.6081		-0.9626	-0.5387	-0.5916	-1.6747
	(0.00342)	(0.00333)	(0.00630)	(0.00744)	(0.0131)	(0.0169)		(0.0137)	(0.0215)	(0.0275)	(0.1815)
FICO3	1.9192	-0.0592	-0.5489	-0.3029	-0.7583	-0.4806		-0.8263	-0.4535	-0.6587	-1.0584
	(0.00322)	(0.00250)	(0.00604)	(0.00714)	(0.0124)	(0.0163)		(0.0132)	(0.0207)	(0.0264)	(0.1363)
FICO4	1.4222	-0.0816	-0.3566	-0.2692	-0.5691	-0.3421		-0.6616	-0.2696	-0.5299	-0.5693
	(0.00332)	(0.00254)	(0.00625)	(0.00739)	(0.0128)	(0.0170)		(0.0138)	(0.0214)	(0.0276)	(0.1485)
FICO5	0.8320	-0.0546	-0.2022	-0.1656	-0.3275	-0.2347		-0.3865	-0.1777	-0.2866	
	(0.00357)	(0.00257)	(0.00676)	(0.00799)	(0.0138)	(0.0184)		(0.0149)	(0.0232)	(0.0296)	
FULLDOC	-0.2354	-0.0741	-0.0261	-0.1517	-0.0793	-0.1237	-0.0382	-0.0898	-0.0669	-0.2228	0.5739
	(0.00161)	(0.00189)	(0.00275)	(0.00318)	(0.00609)	(0.00680)	(0.00618)	(0.00575)	(0.00843)	(0.0116)	(0.1356)
MTMCLTV											
MTMCLTV_1	0.3317			-0.3526							
	(0.0557)			(0.1143)							
MTMCLTV_2	0.5295	-0.1831	-0.1924	-0.4499	-0.5655				-0.4663		
	(0.0557)	(0.00193)	(0.00277)	(0.1143)	(0.00603)				(0.00852)		
MTMCLTV_3	0.6547	-0.7440	-0.2325	-0.2460	-1.2802				-1.1774		
	(0.0558)	(0.00445)	(0.00579)	(0.1144)	(0.0192)				(0.0212)		
MTMCLTV_4	-0.2348	-1.2847	-0.0904	0.7038	-0.9704				-3.5315		
	(0.0581)	(0.0160)	(0.0405)	(0.1202)	(0.1310)				(0.5057)		
OWNER_OCCUPY	-0.1446	0.2098	-0.1568	-0.2346	0.0341	-0.1883	-0.0722	-0.3027	-0.3976	-0.0905	-0.5395
	(0.00283)	(0.00355)	(0.00495)	(0.00564)	(0.0121)	(0.0124)	(0.0117)	(0.0102)	(0.0145)	(0.0232)	(0.1779)
PURCHASE	0.1884	0.1568		0.1556	-0.0177	-0.1886	-0.0567	-0.0271	-0.0406	-0.2891	
	(0.00154)	(0.00181)		(0.00301)	(0.00621)	(0.00637)	(0.00555)	(0.00524)	(0.00811)	(0.0117)	
UNRATE	-0.0309	0.0580	-0.0367	-0.0213	0.0512	-0.1354	-0.0743	-0.0911	-0.2313		
	(0.000666)	(0.000805)	(0.00111)	(0.00130)	(0.00260)	(0.00285)	(0.00258)	(0.00241)	(0.00353)		
K-S Statistic (%)	29.02	24.73	13.00	8.66	22.86	15.93	6.31	16.15	21.81	20.21	29.81

For each parameter, one degree of freedom (DF) is required, and the DF defines the Chi-Square distribution to test whether the individual regression coefficient is zero given the other variables are in the models.

Standard errors of the individual regression coefficients are given in brackets.

Appendix C. Parameter estimates for subprime loans using logistic regression models (transitions p_{41} to p_{68}). The iterative process was implemented in SAS using PROC LOGISTIC. All p-values are significant (less than 0.05).

PARAMETER	p_{41}	p_{42}	p_{43}	p_{45}	p_{47}	p_{48}	p_{51}	p_{54}	p_{56}	p_{57}	p_{58}	p_{68}
Intercept	-1.2384 (0.0501)	-2.6242 (0.0623)	-1.6443 (0.3015)	0.3455 (0.0231)	-1.0339 (0.0592)	-6.1216 (0.1735)	-2.5972 (0.0569)	-5.1665 (0.4137)	-4.4916 (0.0657)	-2.8891 (0.0349)	-6.7295 (0.1859)	-4.3346 (0.3065)
AGE												
AGE1	-0.1172 (0.00400)	-0.0903 (0.00536)	-0.0544 (0.00440)	-0.1027 (0.00224)	-0.0607 (0.00528)	0.1418 (0.0159)	-0.0381 (0.00466)	0.0232 (0.00371)	0.1073 (0.00597)	-0.0422 (0.00253)	0.1779 (0.0169)	0.1232 (0.0278)
AGE2	-1.7306 (0.0371)	-1.4254 (0.0501)	-0.9415 (0.0418)	-1.5661 (0.0211)	-0.7926 (0.0498)	1.6286 (0.1612)	-0.6305 (0.0455)	0.1846 (0.0368)	1.5044 (0.0607)	-0.4727 (0.0231)	2.2879 (0.1758)	1.7691 (0.3038)
AGE3	-2.1272 (0.0377)	-1.7588 (0.0508)	-1.1772 (0.0421)	-1.9757 (0.0214)	-0.8910 (0.0501)	1.5704 (0.1615)	-0.8305 (0.0460)	0.1606 (0.0370)	1.6760 (0.0608)	-0.4642 (0.0237)	2.5207 (0.1759)	1.9602 (0.3038)
AGE4	-2.2342 (0.0390)	-1.8428 (0.0523)	-1.1882 (0.0430)	-2.2124 (0.0220)	-0.9105 (0.0509)	1.4848 (0.1627)	-0.8907 (0.0472)	0.1883 (0.0377)	1.7117 (0.0613)	-0.3321 (0.0259)	2.6599 (0.1766)	2.0416 (0.3038)
AGE5	-2.2198 (0.0429)	-1.7841 (0.0569)	-1.1214 (0.0457)	-2.2852 (0.0241)	-0.8433 (0.0534)	1.5341 (0.1662)	-0.9118 (0.0515)	0.3296 (0.0399)	1.6829 (0.0628)		2.7470 (0.1793)	2.0815 (0.3042)
AGE6	-2.3674 (0.1072)	-2.0712 (0.1445)	-1.1272 (0.0913)	-2.4898 (0.0589)	-1.2073 (0.1114)	1.5277 (0.2418)	-0.7761 (0.1097)		1.6407 (0.0961)		2.8249 (0.2357)	2.2903 (0.3132)
BAL1			-0.2021 (0.0121)				-0.2830 (0.0144)	-0.2597 (0.0131)	0.4544 (0.0152)	-1.0761 (0.0196)		0.2847 (0.0440)
BAL2			-0.1473 (0.00974)		-0.8048 (0.0119)		-0.4210 (0.0115)	-0.2057 (0.0110)	0.3275 (0.0138)	-0.9518 (0.0149)		0.3105 (0.0433)
BAL3					-0.4984 (0.0139)		-0.1799 (0.0128)	-0.0668 (0.0119)	0.2284 (0.0149)	-0.4031 (0.0153)	0.2417 (0.0219)	0.3697 (0.0441)
BAL4					-0.1829 (0.0173)			-0.0351 (0.0141)	0.1155 (0.0180)	-0.1478 (0.0174)	0.3424 (0.0294)	0.3068 (0.0466)
BAL5											0.2998 (0.0409)	0.3022 (0.0513)
BAL6											0.2509 (0.0520)	0.2818 (0.0567)
BAL7												0.2735 (0.0639)
CLOSE_BAL	3.272E-7 (5.269E-8)	5.623E-7 (6.675E-8)		1.485E-6 (2.601E-8)		-9.88E-6 (2.314E-7)						
FICO1	-0.5307 (0.0414)		0.3582 (0.0393)	-0.2850 (0.0168)	-2.2124 (0.0457)	-1.2926 (0.1020)		0.5344 (0.0298)			-1.0149 (0.0968)	-0.5814 (0.0387)
FICO2	-0.3253 (0.0232)		0.4125 (0.0282)	-0.1869 (0.00637)	-1.9643 (0.0174)	-1.0372 (0.0321)	0.3126 (0.0222)	0.4736 (0.0202)	-0.0830 (0.0117)		-0.6123 (0.0420)	-0.3732 (0.0137)
FICO3	-0.1183 (0.0220)		0.4048 (0.0275)	-0.1684 (0.00551)	-1.8061 (0.0154)	-0.5947 (0.0226)	0.2917 (0.0214)	0.3306 (0.0196)	-0.1263 (0.0105)		-0.4202 (0.0392)	-0.1348 (0.0113)
FICO4	-0.1417 (0.0233)		0.2939 (0.0285)		-1.4917 (0.0170)		0.1521 (0.0226)	0.1500 (0.0203)	-0.0970 (0.0115)	-0.1446 (0.0131)	-0.2338 (0.0405)	
FICO5	-0.0893 (0.0253)		0.1736 (0.0309)		-0.5937 (0.0164)		0.0581 (0.0255)	0.1153 (0.0220)		-0.0496 (0.0170)	-0.1430 (0.0443)	
FULLDOC	-0.1756 (0.0102)	-0.1461 (0.0131)	-0.0520 (0.0101)			-0.0852 (0.0236)	-0.1542 (0.00985)	-0.0708 (0.00788)	-0.0868 (0.00856)	-0.3532 (0.0109)		0.0595 (0.0120)
MTMCLTV												
MTMCLTV_1			-1.2731 (0.2966)					1.5771 (0.4111)				
MTMCLTV_2			-1.2275 (0.2966)	-0.0416 (0.00546)				1.5379 (0.4111)	0.4547 (0.00870)			
MTMCLTV_3			-1.2184 (0.2970)	-0.8447 (0.0138)				1.8535 (0.4117)	0.8079 (0.0258)			
MTMCLTV_4			-2.1311 (0.3349)	-2.6952 (0.1453)				2.5521 (0.4662)				
OWNER_OCCUPY	-0.0995 (0.0175)	-0.0779 (0.0233)		-0.3406 (0.00852)	0.3046 (0.0206)	-0.1456 (0.0357)	0.1167 (0.0170)	0.1371 (0.0131)	-0.1258 (0.0118)		-0.1472 (0.0291)	
PURCHASE	-0.1597 (0.00972)	-0.0906 (0.0123)	-0.0447 (0.00939)	-0.0561 (0.00502)	-0.3489 (0.0105)	0.4016 (0.0209)	-0.2392 (0.00982)	0.0429 (0.00728)	0.1430 (0.00742)	0.0605 (0.0107)		-0.0436 (0.0103)
UNRATE	0.0506 (0.00434)	0.1001 (0.00563)	0.1074 (0.00423)		0.0395 (0.00455)	0.1906 (0.00996)	0.0454 (0.00414)	0.1191 (0.00321)	-0.0222 (0.00339)	0.1392 (0.00489)	0.0413 (0.00839)	
K-S Statistic (%)	18.73	16.71	13.31	20.11	29.93	33.67	13.42	9.22	12.41	22.32	14.08	8.55

For each parameter, one degree of freedom (DF) is required, and the DF defines the Chi-Square distribution to test whether the individual regression coefficient is zero given the other variables are in the models.

Standard errors of the individual regression coefficients are given in brackets.