Estimating Conditional Mortgage Delinquency Transition Matrices

Qingqing Chen[1*]
Office of the Comptroller of the Currency
United States Department of the Treasury
400 7th Street SW, Mail Stop 6E-3
Washington, DC 20219
Phone: (202) 649-5532
E-mail: qingqing.chen@occ.treas.gov

Dennis Glennon
Office of the Comptroller of the Currency
United States Department of the Treasury
400 7th Street SW, Mail Stop 6E-3
Washington, DC 20219
Phone: (202) 205-6438
E-mail: dennis.c.glennon@gmail.com

Amos Golan
Director, Info-Metrics Institute and
Professor, American University
External Professor, Santa Fe Institute
Senior Associate, Pembroke College, Oxford
4000 Massachusetts Ave.
Washington, DC 20016
Phone: (202) 885-3783
E-mail: agolan@american.edu

First Version: October 21, 2020

JEL codes: C51 C53 G21

Keywords: Transition Matrices, Markov Chains, Mortgage Default and Prepayment, Credit Risk, Forecasting.

---

[*] Office of Comptroller of the Currency, 400 7th Street, SW, Washington, DC 20219. The American University, 4000 Massachusetts Ave., Washington, DC 20016. We thank Souphala Chomsisengphet and Irina Barakova for their helpful comments, and Regina Villasmil and Wan Peng for their excellent assistance. The views expressed in this paper do not necessarily reflect the views of the Office of the Comptroller of the Currency, the U.S. Department of the Treasury, or any federal agency and do not establish supervisory policy, requirements, or expectations. Corresponding Author: Qingqing.Chen@occ.treas.gov.

Estimating Conditional Mortgage Delinquency Transition Matrices

Abstract

In this paper, we outline alternative methods for constructing transition probabilities *conditional* on both borrower-/loan-specific and macroeconomic factors. We define the transition states in terms of monthly delinquency and prepayment behavior in which payment behavior is measured in terms of days past due over a reoccurring, fixed-length (monthly) billing cycle: a discrete-time model design that allows us to better capture consumers' decisions to continue servicing their debt, delay payment, prepay, or default. We find that for a large sample of active, first-lien, single-family, owner-occupied mortgages from 2004-2013 the assumptions of a Markov chain do not hold and by conditioning the estimates of the transition probabilities on loan-specific and macroeconomic factors, we generate more accurate out-of-time forecast over a time horizon typically used in practice (i.e., 24-months ahead) and are significantly more accurate during periods of changing economic conditions as was observed during the 2008-2010 financial crisis.

JEL code: C51 C53 G21

Keywords: Transition Matrices, Markov Chains, Mortgage Default and Prepayment, Credit Risk, Forecasting.

1. Introduction

Forecasted future expected loss is a key input into the calculation of reserves, the valuation of portfolios, and for pricing, capital, and stress-testing purposes. Expected loss in turn depends on the estimates of both the exposure to loss and the possibility that a loss will occur: the former conditioned on the probability of prepayment, the latter on the probability of default over the forecast horizon. Li (2014), An and Qi (2012) IFE (2007) and Campbell and Dietrich (1983) model default and prepayment as independent competing risks within an event-time analysis framework to estimate a lender's exposure to the likelihood a specific event (default or prepayment) will occur over time conditional on loan- and borrower-specific characteristics and changes in market and economic conditions. Alternatively, Smith et al. (1996), Molina Utrilla and Constantinou (2010), Grimshaw and Alexander (2011), and Leow, and Crook (2014) propose modeling payment behavior as a time-dependent Markov process in which individual borrowers migrate between various states of payment/nonpayment conditional on loan- and borrower-characteristic. Reasonable and supportable forecasts are conceptually possible using either approach.

Our objective in this paper, however, is not to assess all the various methods to determine which approach is best. We recognize data limitations and model purpose may weigh heavily in the decision to select one approach over the other. Instead, we focus on the question: is it possible to generate accurate medium to long-term (out-of-time) forecasts over a mix of economic conditions using these modeling methods? More specifically, in this paper, we investigate the ability of models that fall within the broader category of Markov chain models to accurately forecast default and prepayment rates over a mix of economic conditions.[2]

Using a large sample of single-family, owner-occupied mortgages from 2004 through 2013, we evaluate the predictive accuracy of various Markov-chain-based models to forecast cumulative default and prepayment rates over a 24-month out-of-time performance window. We develop several models following the methods outlined in the retail credit literature (i.e., Cyert, et al., 1962; Kallberg and Saunders, 1983; Smith and Lawrence, 1995; Smith et al., 1996; Molina Utrilla and Constantinou, 2010; Grimshaw and Alexander, 2011; and Leow and Crook, 2014) applied to data that includes the 2008-2010 mortgage crisis. We assess the predictive accuracy of each model utilizing a one-year-step-forward sample design in which we extend the sample period by an additional 12 months of data starting in 2006 and ending in 2011 generating a total of 6 timeframes that are used to develop models and test their 24-month, out-of-time forecast accuracy through 2013. We estimate a transition matrix for each of the 6 samples using the same source of data and set of covariates to isolate the impact of changes in the mix of economic conditions prior to, during, and after the 2008-2010 financial crisis. We find that the estimated transition probabilities vary greatly across individual loans and over time resulting in significant differences in the estimated transition matrices and the overall forecast accuracy of the models. Moreover, we find that (1) conditioning on both time invariant and time varying, loan- and borrower-specific, and macroeconomic factors significantly improves the predictive accuracy of the models especially during an economic downturn – consistent with the results reported in Li, 2014; Sarmiento, 2012; Crook and Banasik, 2012; and Demyanyk and Van Hemert, 2011, and (2) find weak support for the hypothesis that mortgage payment behavior is consistent with a 2nd order Markov process.

The remainder of the paper is outlined as follows. In the next section, we outline the data we use to assess the various models developed in Section III including the unconditional transition matrices based on the maximum likelihood (MLE) approach outlined in Anderson and Goodman (1957) and extended in Lee, et al. (1977). Insofar as that approach is still used in practice, we use the out-of-time forecast accuracy of the MLE models as benchmarks to assess the impact of conditioning on economic factors. In Section III,

[2] We address in a companion paper the same question using various modeling approaches that fall within the broader category of survival analysis models.

we also outline the multinomial logit (MNL) approach we adopt to estimate the conditional transition probabilities and demonstrate, in Section IV, that by relaxing the Markov property assumption and conditioning the transition probabilities on both time invariant and time-varying factors (e.g., economic and industry conditions) greatly increases out-of-time forecast accuracy over the economic cycle.

2. Data and Descriptive Analysis

We use the Black-Knight Securitized Mortgage (BKSM) database to construct various Markov/non-Markov state transition models. The BKSM database includes monthly loan-level data for non-agency mortgages compiled by MBSData, LLC. MBSData provides limited loan- and borrower-specific information at time of origination (e.g., LTV, FICO Score, loan terms, loan purpose, level of documentation product type, Fixed/ARM/Hybrid, IO, Balloon, etc. – static variables) and monthly updates on performance behavior (e.g., delinquency status, prepayment, scheduled payment/balance reflecting rate changes, current LTV, etc. – dynamic variables) for over 95% of public mortgage backed securities.[3] We augment the loan/borrower-specific data with macroeconomic and housing market-specific variables designed to capture the impact of systemic time-varying factors on borrowers' payment behavior.

We construct a monthly panel of first-lien, single-family, owner-occupied mortgages from 2004 through 2013 using a 15% stratified random sample, by year/month, for loans that existed as of January 2004 or enter the data set between January 2004 and December 2013.[4] The loans are tracked monthly from the time they are securitized until they either prepay, default, mature under the terms of the contract, or are censored. There are a total of 1,622,538 mortgage loans in our sample with an average of 34.1 months on book per loan resulting in a panel data set of over 55.3 million loan-month observations.

The descriptive statistics for a selected set of commonly used loan characteristics (Mayer et al., 2009; Sarmiento, 2012; Demyanyk and Van Hemert, 2012) over the full sample period (2004-2013) are reported in Table 1. In addition, to better understand the shift in the credit profile of the loans over the credit cycle, we also report the descriptive statistics for the loans existing prior to the crisis (2004-2007), during the crisis (2008-2010), and post-crisis (2011-2013). These statistics show that Alt-A and Subprime mortgages consistently default at higher rates over the sample period, approaching 50% during the crisis period (2008-

[3] We note that a large percentage of the securitized loans in our data set are identified as subprime or jumbo loans.

[4] We randomly selected a 15% sample of all the loans that existed as of January 2004 (i.e., 200401) and randomly sampled 15% of all new loans entering the data set each *month* through 201112. We then included all the monthly observations for each loan until it either defaults, prepays, matures, or is censored through 201312. We note that the public mortgage backed securities market grew dramatically from 2001 to 2007, slowed dramatically during the crisis (i.e., the number of *new* single-family loans entering the Black-Knight Securitized Mortgage (BKSM) database dataset in 2008 and 2009 were 3.2% and 3.3% of the level of *new* loans originated in 2006, respectively), and was effectively zero in 2010-2012 (i.e., the number of new loans entering the BKSM dataset was 0.1% or less than the number of loans in 2006). In 2013, the number of new loans entering the BKSM dataset increased significantly – over twice the number of new loans entering in 2008 – and by 2015 and 2016 the number of loans each year was just over 20% of the 2006 peak (see Goodman (2015) for a discussion of re-birth of mortgage securitization). The increase in the number of new single-family loans in the BKSM dataset beginning in 2013 was almost exclusively (99% of the loans) due to the credit risk transfer (CRT) program developed by Freddie Mac: a program designed to transfer credit risk to the private capital market (see https://crt.freddiemac.com/about-crt.aspx for a brief outline of the program). Under Freddie Mac's credit risk transfer (CRT) program, the securitized loans are subject to Freddie Mac's underwriting and quality control standards: conditions that fundamentally change the properties of the loans entering the data set into and after January 2013 relative to those entering prior to 2013. For that reason, we use only loans that entered the dataset prior to 2013 to estimate the transition probabilities in this study and, although we forecast performance through 2013, the forecasts use only loan cohorts prior to 2013. Mortgages underwritten through CRT program are excluded from our analysis.

10). The impact on the portfolio from those loans is significant insofar as roughly 20% of the loans in the sample are identified as subprime.[5] The decline in the percentage of subprime and alt-A loans post-2007 reflects the difficulty the market had underwriting for the risk of those loans. Interestingly, jumbo loans had the lowest default rates across the various loan types – a default rate that is notably lower than the overall default rate in each time period. The default rate of jumbo loans, however, also increased significantly during the crisis rising from 4.8% (pre-crisis) to 24.5% before falling back to 12.3% over the post-crisis period.

Fully documented loans consistently perform better than the average loan as reflected in the lower overall default rate (i.e., 26.3% vs 28.0%) and in each of the sub-periods. It is interesting to note that No Doc loans perform even better than the fully documented loans especially in the pre-crisis period; however, No Doc loans only represent 2% of the total number of loans in the sample. Stated Income loans, which represent nearly 13% of the total, however, consistently perform the worst relative to all the other documentation categories. The overall default rate for Stated Income loans was 40%, increasing to nearly 50% during the crisis period, then falling back to 23.6% in the post-crisis period. The percentage of Stated Income loans in the sample, however, declined sharply during the post-crisis period reflecting the tightening of underwriting standards and the restrictions imposed under the Dodd-Frank Act (2010) prohibiting banks from underwriting Stated Income loans for owner-occupied properties. Finally, Low, Limited, & Reduced (LLR) documented loans performed relatively well in the pre-crisis period (i.e., with a default rate of 6.6% relative to an overall average default rate of 9.9%), but less so during and after the crisis when their default rates aligned more closely to the overall average of those sub-periods.

The loans in our sample are overwhelmingly used for refinancing and new purchases (i.e., 61.1% and 37.4%, respectively). The loan-purpose distribution is relatively stable over the full sample period, although the percentage of loans used to refinance existing mortgages increased slightly post-crisis (2011-13). Purchase mortgages are slightly riskier than those in the other loan-purpose categories especially during the crisis period resulting in a large drop in the percentage of purchased loans in the post crisis sub-period.

There is a noticeable shift in the distribution of FICO scores-at-origination toward higher/better score bands over time. The percentage of loans with origination FICO scores below 620 (above 680) decreased (increased) from 2004 through 2013. Overall, the mean score increased from 663 in the pre-crisis period, 2004-07, to 695 in the post-crisis period, 2011-2013. The FICO score-at-origination performed reasonably well at ranking borrowers by their relative credit quality (i.e., loans with higher scores defaulted less often than those with lower scores) even though the definition of a 'bad' and the time horizon used to define a default in our data are not those used to construct the FICO score. The default rate within each score band, however, changed significantly over time reflecting the static design of a scoring model (Thomas, 2009) and the sensitivity of the default rate to changing economic conditions (Crook and Banasik, 2012). For example, the 16.3% default rate in the 680-720 score band during the post-crisis period is comparable to the 17.6% default rate in the 550-620 score range observed during the pre-crisis period; a result that suggests that even a shift up in the score distribution for a pool of securitized loans does not necessarily indicate the pool of loans are of higher/better credit quality.[6]

The distribution of the current LTV was increasing over time: a finding consistent with those observed by Mayer et al. (2012) and Demyanyk and Van Hement (2011). When combined with the observation that the LTV-at-origination was falling, suggests that declining house prices during the crisis was a key factor contributing to the dramatic decrease in the equity holdings of the borrowers in our sample and likely

---

[5] Default is defined as 120 days past due or worse.

[6] It was a common practice in the industry to use the FICO distribution as a measure of the credit quality of pool of securitized loans, a measure that influenced the pricing of the security backed by the loans (Koudinov, et al., 2019).

contributed significantly to the increase in default rates over the sample period (Andersson, et al., 2013). The percentage of loans with current LTVs below 80% was relatively stable over the 2004-2010 period (roughly 61% of the total); however, the percentage dropped significantly to 53% by 2011-13 as the percentage of loans with current LTV greater than 80% increased from 29.5% in 2004-07 to 37.6% during the crisis, and as high as 47.1% in 2011-13.

Mortgage loans with a balloon-payment feature defaulted at rates significantly higher, 60.4%, than average, 28.0%, over the full sample period (i.e., 2004-2013). Fortunately, the percentage of loans with a balloon-payment feature was relatively small, 4.7%; and, although the percentage increased during the crisis, the percentage of loans with a balloon-payment feature decreased significantly during the post-crisis period.

The percentage of mortgages in our sample with an interest-only (IO) feature was just over 20% and remained relatively stable over time. Prior to the crisis, IO mortgages performed better than average as reflected in a pre-crisis default rate of 7.7%, compared to the overall pre-crisis default rate of 9.9%. During the crisis (2008-10) and post-crisis (2011-13) periods, however, IO mortgages defaulted at higher than average rates – 38.0% and 18.6% relative to the overall default rates of 32.4% and 15.9%, respectively – suggesting that pre-crisis performance masked the inherent risk of those loans that was eventually revealed during the crisis.

The descriptive statistics for fixed rate and hybrid ARMs reveal that adjustable rate mortgages were, on average over the full sample, the preferred payment type. Moreover, a large percentage of the ARMs were hybrid ARMs.[7] This is especially true for the pre-crisis period in which 38.6% of the loans were hybrid ARMs and 38.7% were fixed rate mortgages. The distribution of payment type changed dramatically during and after the crisis. The percentage of fixed rate mortgages increased significantly during and after the crisis (i.e., 49.3% and 62.6%, respectively); the percentage of hybrid ARMs decreased to 24.2% and 13.4% during and after the crisis, respectively. Overall, fixed rate mortgages performed better than ARMs especially adjustable rate mortgages products that were designed to make mortgages more affordable yet exposing the borrower to repricing risk. The default rate on fixed rate mortgages was consistently lower than the overall averages and the default rate for hybrid ARMs was significantly higher than average increasing to 55.0% during the crisis period.

The descriptive statistics reported in Table 1 suggest that the informational content of the static variables (i.e., those known at time of origination and remain constant over the life of the loan) commonly used to evaluate the credit quality are useful for assessing relative performance across the portfolio. However, the variation in the default rate over a mix of economic conditions suggests static variables alone are inadequate for identifying and measuring risk especially during periods in which economic conditions are changing quickly. Models developed for valuation, pricing, stress testing, and loss forecasting will be especially sensitive to changes in systemic (i.e., time varying) factors associated with changing economic conditions. For example, the payment behavior of a loan originated during the pre-crisis period will perform much differently, on average, then a loan with the *same values for the static characteristics* originated during the crisis period. It is within that context that we construct default and prepayment models that are conditional on both time invariant (i.e., static) and time-varying (i.e., dynamic/systemic) variables within a Markov transition modeling framework in the next section.

[7] Hybrid ARMs – which are a subset of the adjustable rate mortgages – are defined as loans with fixed rates over the first few years then convert to adjustable rates over the remaining life of the loan (e.g., 2/28: a 30-year mortgage with a fixed rate for the first 2 years, then adjustable for the remaining 28 years; similarly, there are 3/27, 5/25, and other such products that are labelled as hybrid ARMs).

3. A Multistate Transition Model

A discrete-time Markov chain modeling framework (Anderson and Goodman, 1957) characterized by a finite number of states/categories and a finite number of equi-distant time points at which observations are made is a natural framework for modeling mortgage loan payment behavior in which the payment behavior is measured in terms of days past due over a reoccurring, fixed-length (i.e., monthly) billing cycle. A Markov chain consists of a set of transitions that satisfy the Markov property in which future delinquency/default states depend only on the current state, not on the events that occurred before it (i.e., memoryless). More formally, the Markov property is defined: for a stochastic process $\{Z_t\}$,

$$Pr(Z_{t+1} = j \mid Z_t = i, Z_{t-1} = i_{t-1}, \dots, Z_0 = i_0) \; = \; Pr(Z_{t+1} = j \mid Z_t = i) \; \equiv \; p_{ij} \qquad (1)$$

t≥0 and all states i_0, i_1,…, i_{t-1}, i, j, where the p_{ij} represents the probability of transitioning from state i in time t to state j in time $t+1$. The Markov property is convenient insofar as it adds structure and simplicity to a predictive model (or probability forecast) that otherwise might be intractable.

The simplest multistate transition model is a first-order, time-homogenous Markov model in which the maximum likelihood estimation (MLE) method can be used to generate the unconditional transition probabilities used to forecast (out-of-time) default and prepayment rates.[8] We use those forecasts as benchmarks to assess the impact of violating the assumptions of a Markov transition model. If the assumptions of this simple model do not hold or are inconsistent with process that generates the data, the forecasts are unlikely to reflect actual behavior and an alternative modeling approach should be considered. For that reason, we also outline below the estimation of conditional transition probabilities designed to capture the impact of trends in economic and market conditions on delinquency/default as observed using a multinomial logit (MNL) approach proposed in Smith et al. (1996), Molina Utrilla and Constantinou (2010), and Grimshaw and Alexander (2011).

3.1 1st-Order, Time-Homogenous Markov Model: Unconditional Transition Matrix – A Benchmark Model

We begin with the simplest and most straightforward approach to construct an unconditional, first-order Markov transition model: the maximum likelihood estimation (MLE) method (Anderson and Goodman, 1957). That approach was used in the early literature on modeling delinquency and default/loss behavior for retail loan products (Cyert, et al., 1962; and Kallberg and Saunders, 1983; Betancourt, 1999) and is still used in practice. The unconditional MLE transition probabilities, p_{ij}, are defined as:

$$\hat{p}_{ij} = \frac{n_{ij}}{n_i} \qquad (2)$$

where n_{ij} is the number of loans in state i that migrate to state j during the time period $(t, t+1)$ and n_i is the number of loans in a given state i at the beginning of period t for all i, j = 1, 2, …, I; such that, $P_{ij} = [p_{ij}]$ and $\sum_{j=1}^{J} p_{ij} = 1$, for all i=1,2,…, I. For our purpose, we define i over six financial states: four active delinquency states – current, 30-59 dpd, 60-89 dpd, and 90-120 dpd – and two terminal (or absorbing) states – prepay and default (120+ dpd).

[8] In this paper, we use the expression "unconditional" transition probabilities to mean the probabilities are not conditioned on time-varying factors (i.e., the Markov property holds).

Following Kallberg and Saunders (1983), we use equation (2) to calculate the unconditional monthly transition probabilities over the full sample through 2013. The results are reported in Table 2.

One advantage of this approach, i.e., if the assumption of a time-homogenous Markov model holds, is that, for a given distribution of loans (Z_0) across the six states as of a cohort date t_c, the probabilistic evolution of the Markov process across the states can be described by:

$$Z_{t,t+s} = Z_0 P_{ij}^{s} \tag{3}$$

where P_{ij} is the unconditional MLE transition matrix; $Z_{t,t+s}$ is the distribution of loans in time $t_c + s$ for $s = 1, 2, \dots, S$ - the forecast period. Even when the assumption of a time-homogenous process does not hold, equation (3) may still produce relatively accurate forecast over short forecast horizons. More specifically, if the assumption of a time-homogenous process does not hold, the condition that the two-periods ahead probability distribution for $Z_{t,t+2}$ is the one-period ahead distribution $Z_{t,t+1}$ times the transition matrix $P_{ij,(t+1,t+2)}$ still does, which we can write as:

$$Z_{t,t+2} = Z_{t+1} P_{ij,(t+1,t+2)} \tag{4}$$

where

$$Z_{t+1} = Z_0 P_{ij,(t,t+1)}. \tag{5}$$

Substituting eq (5) into eq (4), we get

$$Z_{t,t+2} = Z_0 P_{ij,(t,t+1)} P_{ij,(t+1,t+2)}, \tag{6}$$

which can be generalized for an a s-period ahead forecast horizon (Grimshaw and Alexander, 2011):

$$Z_{t,t+s} = Z_0 P_{ij,(t,t+1)} P_{ij,(t+1,t+2)} \cdots P_{ij,(t+s-1,t+s)}. \tag{7}$$

Unfortunately, to forecast beyond the most recent time period ($t_c = t_0$), the information required to estimate the MLE transition matrices, $P_{ij,(t+\tau-1,t+\tau)}$, for $t_\tau > t_c (= t_0)$ is not known. As a result, the unconditional MLE transition matrix approach is of limited use for forecasting out-of-time unless it is reasonable to assume the transition probabilities change slowly over time and can be approximated by a time-homogenous transition matrix based on historical data known at time of the forecast, t_c.

It is not always reasonable to assume that the transition probabilities change slowly over time. The change in the monthly, unconditional transition probabilities during the pre-crisis (2004-2007), crisis (2008-2010), and post-crisis (2011-2013) periods is a case in point. The MLE unconditional transition matrices for the three sub-periods are reported in Table 3. The transition probabilities from any one of the non-terminal states to the prepayment state (i.e., $Z_{i,t} \rightarrow Z_{pp,t+1}$) are much higher in the pre-crisis period (i.e., in the range of 2.1 to 2.8 percent per month) than observed during the crisis and post-crisis periods (i.e., 1 percent or lower per month). Moreover, the transition probabilities above (below) the main diagonal – which reflects the migration to higher (lower) delinquency states – are considerably higher (lower) during the crisis period relative to both the pre- and post-crisis periods. The increased rate of migration to higher delinquency states combined with a higher probability a loan that is 90dpd transitioning to default is consistent with the significantly higher overall default rate observed during the crisis period reported in Table 1.

The hypothesis that the transition probabilities are constant (i.e. time homogenous) is not typically formally tested in the literature or in practice. Either it is not addressed even when recognized as a potential problem (e.g., Cyert and Thompson, 1962; Campbell and Dietrich, 1983), evaluated visually by plotting the $p_{i,j}$s over time (e.g., Kallberg and Saunders, 1983; Grimshaw and Alexander, 2011), or assumed to be nonstationary by design (Smith, et al., 1996; Molina Utrilla and Constantinou, 2011; Malik and Thomas, 2012; and Leow and Crook, 2013). We, however, formally test the hypothesis that the transition probabilities are constant (i.e., that the data reflect a stationary Markov process) using the chi-square test of homogeneity suggested by Anderson and Goodman (1957). More specifically, we test the null hypothesis:

$$H_0: p_{ij}(t) = p_{ij}$$

for all j = 1,2,…,J and t = 1,2,…T. The results reported in Table 4 show, for each state i respectively, that the hypothesis of stationarity is rejected over all three subperiods 2004-2007, 2008-2010, and 2011-2013; results that are consistent with those reported for a much earlier timeframe - i.e., 198912 through 199312 - by Betancourt (1999).

To illustrate the potential impact these relatively small difference in the transition estimated transition probabilities can have on the forecast over a 24-month time horizon, we graph the distribution of loans in time t =0 (the initial distribution using the 2004Q1 cohort of active loans – 72,193 loans) across the six delinquency states and t =24 (i.e., the 24-month forecast period) using the four MLE transition matrices for the full sample 2004-2013, the pre-crisis period 2004-07, the crisis period 2008-10, and the post-crisis 2011-13.

The results in Figure 1 show that the 24-months ahead forecasts of the prepayment and default rates vary significantly depending on which MLE transition matrix is used. If the pre-crisis MLE transition matrix forecasts are used, the predicted number of prepaid (27,516 obs.; 38.1%) and defaulted (8,119 obs.; 11.2%) loans over a 24-month period is roughly twice the number of prepaid (13,271 obs.; 18.4%) and half the number of defaulted (16,963 obs.; 23.5%) loans forecasted using the crisis MLE transition matrix – a non-trivial difference. Interestingly, the forecasted prepayment (11,299 obs.; 15.7%) and defaulted (8,134 obs.; 11.3%) loans generated from the post-crisis MLE transition matrix are the lowest (prepaid) and near-lowest (defaults) respectively: a result consistent with the relatively high forecasted number of loans (47,477 obs.; 65.8%) that remain current (i.e., the highest across the set of MLE transition matrices): a result that is consistent with the relative benign economic environment that existed during the 2011-13 subperiod.

We conclude from the results in Table 3 and 4 that the unconditional MLE transition probabilities reported in Table 2 are not stationary (i.e., time invariant) and that changes in loan terms, borrower characteristics and economic/market conditions over time have more of an impact on the probability of migrating to an alternative state than the loan's current delinquency/payment state by itself. These results raise the concern that the assumptions required to model monthly mortgage payment behavior over a mix of economic conditions as a finite Markov chain do not hold, in general, nor, in particular, during the time periods prior to, during, and post the 2008-10 mortgage crisis.[9]

Our results support the hypothesis that the data generating process is neither stationary nor time homogenous and that the probability of transitioning from state i in time $t$to state j in time t+1 depends on more than just its current state.[10] The probability a mortgage migrates to any particular state in time

[9] Betancourt (1999) states: "If differences in loan characteristics are highly correlated with differences in payment behavior, then forecasts from a model that fails to incorporate these characteristics may be biased." (p. 310).

[10] Crook and Banasik (2012) and Demyanyk and Van Hemert (2011) find that trends in economic and market conditions impact delinquency/default transitions for retail portfolios, which suggests forecasting models developed using a transition matrix-based approach in which the Markov property holds will generate unconditional transition

t+1 is conditional on its previous history. To capture the impact of previous history, we estimate the transition probabilities conditional on borrower-/loan-specific characteristic (both time invariant and time varying) and time-varying macroeconomic factors that are selected to capture trends in market and industry conditions that impact payment behavior. We outline the estimation of the conditional transition probabilities in the next section.

3.2 Conditional Transition Matrix

We combine the modeling framework outlined in Smith and Lawrence (1995), Smith et al. (1996), Grimshaw and Alexander (2011), Molina Utrilla and Constantinou (2011), and Leow and Crook (2014) with the statistical methods outlined in Allison (1982), Begg and Gray (1984), IFE (2007), Bellotti and Crook (2013b), and Wang et al. (2017) to estimate monthly transition probabilities conditional on loan-/borrower-specific characteristics and macroeconomic/market factors. We utilize the structural property of a transition matrix in which, in any time t, there is a positive probability a loan in a non-terminal state i (i.e., current, 30dpd, 60dpd, and 90dpd) migrates from state i to any one of several feasible states j (j= 1,2,…,i,… J), subject to $\sum_{j=1}^{J} p_{ij} = 1$ for all initial states i=1,2,…,I` (I`=I-2) and to model the transition probabilities within a discrete-time multi-state/competing-risk modeling framework. More specifically, following the structural design of Grimshaw and Alexander (2011), Smith and Lawrence (1995), and Calhoun and Deng (2002), the monthly transition matrix has the following structure:

$$
P_{ij}(t) = \begin{bmatrix}
p_{c,c}(t) & p_{c,30}(t) & p_{c,60}(t) & p_{c,90}(t) & p_{c,p}(t) & p_{c,d}(t) \\
p_{30,c}(t) & p_{30,30}(t) & p_{30,60}(t) & p_{30,90}(t) & p_{30,p}(t) & p_{30,d}(t) \\
p_{60,c}(t) & p_{60,30}(t) & p_{60,60}(t) & p_{60,90}(t) & p_{60,p}(t) & p_{60,d}(t) \\
p_{90,c}(t) & p_{90,30}(t) & p_{90,60}(t) & p_{90,90}(t) & p_{90,p}(t) & p_{90,d}(t) \\
0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 1
\end{bmatrix}
$$

where the $p_{ij}(t)$ are the monthly conditional transition probabilities from state i in time t to state j in time t+1; and the transition probabilities in time t, $p_{ij}(t) = f(X_{ij,k};\theta)$ are conditional on $X_{ij,k} = \{w_k, x_{k,t}, v_t\}_{ij}$, for the k[th] borrower, where w_k, $x_{k,t}$, and v_t are the time-invariant loan/borrower attributes, time-varying loan/borrower attributes, and economic/market factors, respectively. As a result, there are a total of 24 transition probabilities that we estimate empirically below.

Following Grimshaw and Alexander (2011), Molina Utrilla and Constantinou (2011), and Smith, et al. (1996), we estimate the conditional transition probabilities $p_{ij}(t) = f(X_{ij,k};\theta)$ for each non-terminal state row, i = 1, 2, …, I-2, using a multinomial/competing-risk modeling framework. Smith, et al. (1996) estimate the unconditional transition probabilities of each of the three non-terminal rows (i.e., current, 30-89 dpd, and 90+dpd) using a multinomial $logit$ (MNL) modeling approach in which, for each initial state i, the transition probabilities a loan will migrate to each of the alternative states j = 1, 2, …, J at annual intervals is computed from:

$$p_{ij} = \frac{\exp(\alpha_{ij} + \beta_{ij}^{`} X_{i.})}{1 + \sum_{j \neq i} \exp(\alpha_{ij} + \beta_{ij}^{`} X_{i.})} \qquad \text{where } i \neq j \qquad (8)$$

---

matrices that are likely to perform poorly on out-of-time samples especially during periods of rapidly changing economic conditions.

$$p_{ii} = \frac{1}{1+\sum_{j \neq i} \exp\left(\alpha_{ij}+\beta_{ij}^{`} X_{i.}\right)} \tag{9}$$

The MNL approach, however, introduces two potential problems as outlined in Grimshaw and Alexander, (2011): (1) for each row j, the transitional probabilities p_{ij} are estimated conditional on the same set of covariates, $X_{i..}$;[11] and, (2) not all p_{ij} can or should be modeled (e.g., the probability of transitioning from current to 90 dpd).[12] To address those concerns, we estimate the multinomial transition probabilities of each row i= 1,2,…,4 from the set of binomial logit models (BNL) for each p_{ij} j = 1, 2, …, 6 using the approach discussed in Allison (1982) and Begg and Gray (1984). More specifically, for i = 1 (i.e., loans that are current in time t), the BNL models for each conditional transition probability, p_{1j}, for j = 2,..,6, are:

$$p_{1j} = \frac{\exp\left(\alpha_{1j}+\beta_{1j}^{`} X_{1j}\right)}{1+\exp\left(\alpha_{1j}+\beta_{1j}^{`} X_{1j}\right)} \qquad \text{for each } j = 2,3,..,6 \tag{10}$$

(Glennon and Nigro, 2005). Rearranging terms, equation (10) can be rewritten in terms of its corresponding odds ratio:

$$\frac{p_{1j}}{p_{11}} = \exp\left(\alpha_{1j} + \beta_{1j}^{`} X_{1j}\right) \qquad \text{for each } j = 2,..,6 \tag{11}$$

where $p_{11} = 1 - p_{1j}$. In other words, we restate the relationship between the covariates and the transition probabilities in terms of the probability a loan in state 1 in time t will migrate to state j in time t+1 relative to remaining in state 1 in time t+1.

For those cases in which it is not practical to model the p_{1j} conditional on X_{1j} (denoted $j = j_0$), we use an intercept only specification in equation (11).

Combine the binary logistic models into a multinomial logistic model for the ith row by first computing the sum (Grimshaw and Alexander, 2011):

$$d_{i=1} = 1 + \sum_{\substack{i \neq j, \\ j \neq j_0}} \exp\left(\alpha_{1j} + \beta_{1j}^{`} X_{1j,k}\right) + \sum_{\substack{i \neq j, \\ j = j_0}} \exp\left(\alpha_{1j}\right) \tag{12}$$

and then calculating each of the conditional transition probabilities for row i=1 as:

$$\begin{aligned} p_{1j} &= \frac{\exp\left(\alpha_{1j}+\beta_{1j}^{`} X_{1j,k}\right)}{d_{i=1}} && \text{for all } i \neq j \text{ and } j \neq j_0 \\ p_{1j} &= \frac{\exp\left(\alpha_{1j}\right)}{d_{i=1}} && \text{for all } j = j_0 \\ p_{11} &= \frac{1}{d_{i=1}} && \text{for } i = j \end{aligned} \tag{13}$$

[11] For example, it is unlikely the factors influencing the probability of transitioning from current to 30dpd are the same as those influencing the probability of transitioning from current to prepay; or, that the same factors influencing the transition from 30 dpd to current are the same as those from 30 dpd to 60dpd.

[12] There are thousands of loans in our data that transition from current in time i to 90dpd in time i+1: too many to ignore and delete (especially because many of those loans eventually default), but very difficult to explain much less model the conditional transition probability.

The method developed to calculate the transition probabilities in equation (13) is then applied to calculate the transition probabilities for rows 2, 3, and 4 of the transition matrix conditioned on the i, j-specific set of factors (X_{ij}).[13]

3.3 Model Construction and Estimation of the Transition Probabilities

The payment history of an individual borrower can best be portrayed as a sequence of events leading up to either full payment or default. The monthly delinquency status is a record of when the events of interest – especially for the interim past due events – take place over the life of the loan. As such, time is surely a key factor in identifying the cause of the events either directly through the process in which borrowers reveal their ability to manage their debts, or indirectly through their exposure to unexpected shocks resulting from changes in economic conditions or personal circumstances. Because the events (i.e., payment status) in our data are observed monthly, the events occur at discrete points in time. For those reasons, we model the transition probabilities in equation (13) using a discrete-time event history approach (Allison, 1982).

Under a discrete-time event-history approach, the number of observations per loan is expanded to one observation for every month the loan is active. For each BNL model, the dependent variable is defined as $Y_{ij} = 0$, if $j = i$; and $Y_{ij} = 1$, if $j \neq i$. For example, the dependent variable for the BNL transition probability model $p_{c,p}$– i.e., from current to prepay – is $Y_{c,p} = 0$, if in month t, a loan that is current remains current in time t+1, and $Y_{c,p}= 1$, if in month t, a loan that is current is prepaid in time t+1. Under this design, a loan that is current each month over the first 23 months on book and prepays on the 24th month would contribute 24 observations to the sample of which $Y_{c,p}= 0$ and one, the last one, $Y_{c,p} = 1$.

For each event type in equation (13), we used the event-history sample design, to estimate the (i, j)-specific BNL model by eliminating from the sample the observations in which event j' $(\neq j)$ occurs. As a result, each binomial logit (BNL) model, p_{ij}, is estimated using only the observations in which the transition to event j or no transition event occur (Allison, 1995) as a function of time/months-on-book (age), time invariant loan-/borrower-specific factors ($w_{ij,k}$), and time-varying factors ($x_{ij,k,t}, v_{ij,t}$) (e.g., Smith, et al, 1996; IFE, 2007; and Molina Utrilla and Constantinou, 2010).[14] The advantage of this approach is it allows us to focus on the factors that have the greatest impact on the specific event being modeled.

---

[13] We identified two conditional transition probabilities – $j_0 = \{p_{c,90}, p_{c,d}\}$ – for which a conceptually sound rational for modeling the transition probabilities as a function of a loan-/borrower-specific and economic factors does not exist.

[14] This adjustment is necessary to align the sample to reflect the true exposure-at-risk to event j before estimating the BNL probability of migrating from state i in time t to state j in time t+1. If we assume that event j' takes place at the beginning of the month in which it occurs, then the borrower that experiences event $j'$is no longer at risk of experiencing event j in that month and, therefore, that observation should not be included in the sample used to estimate the p_{ij} model (Allison, 1982, 1992). More specifically, the discrete-time sample used to estimate the BNL model for $p_{c,30}$ would include all the monthly observations in which the borrowers are current in time t and either become 30dpd or remain current in time t+1. If in time t the borrower is current, but prepays in time t+1, the borrower is no longer at risk to migrate to 30dpd in month t+1. For that borrower, the observation in which borrower migrates to prepayment status is excluded from the sample used to estimate the BNL model for $p_{c,30}$.

The estimated BNL conditional transition probabilities are reported in Table 5.[15] Although the specification of the BNL models are, in general, consistent with those found in the literature (e.g., Molina Utrilla and Constantinou, 2010; Smith, et al, 1996; and IFE, 2007), we use only covariates that are either static and known at time of origination, or vary over time in known (i.e., loan age) or predictable ways (i.e., LTV, selected economic/market variables). That restriction will allow us to evaluate the out-of-time forecast accuracy of our approach without using information that, in practice, would not be known or could not be forecasted by internal or external sources.[16] This restriction may result in our intentionally excluding specific variables commonly used in practice to model payment behavior. For example, a borrower's current credit score is often used as the most up-to-date summary of the borrower's overall credit quality; however, because future/out-of-time values of the current credit score are unknowable at time of forecast, we do not include the borrowers' current credit score in our specification of the transition probability models.

The macroeconomic variables are measures of the overall financial health of the economy and the mortgage market and are included to capture the effects of changes in systematic economic conditions on borrower behavior. It is unlikely that those changes have instantaneous effects and that the impact from a change in the systematic factors is likely to affect behavior as the impact of a change accumulates and is transmitted/propagates through the market over time (Malik and Thomas, 2012). We do, however, constrain the lag structure to include only a three- and a six-month lag of the macroeconomic variables to capture the cumulative effects over a reasonable length of time without introducing a large number of additional factors to the model as the length of the lag structure grows.

The binomial logistic regression results for each loan-level transition probability model (p_{ij}) for the full sample (i.e., 2004 thru 2011) are reported in Tables 5a and 5b. In general, the individual BNL models perform reasonably well at predicting the conditional transition probabilities as reflected in the relatively low Brier Scores and moderate to high AUC values, especially for the prepayment and default models (i.e., $p_{i,p}$and $p_{i,d}$ models).

The specifications of the models are generally consistent with those of Smith, et al. (1996), Grimshaw and Alexander (2011), and Molina Utrilla and Constaninous (2011). Moreover, the results in Table 5 support the conclusions of Crook and Banasik (2012), Demyanyk and Van Hemert (2011), and Crouhy, et al. (2008) that macroeconomic and market conditions play an important role in explaining borrowers' repayment behavior as reflected in the large percentage of models in which the estimated parameters for the macroeconomic variables are statistically significant (in bold).[17]

A key feature of the conditional transition matrix approach is the estimated transition probabilities are not only loan specific, but also time specific. To illustrate, we randomly selected two loans (neither of which

---

[15] For each row i, there are J possible events for which we can generate J-1 independent BNL models in which each p_{ij} is the probability the loan migrates to the $j^{th}(\neq i)$ outcome (i.e., event) rather than the i^{th} (i.e., non-event). The i^{th} outcome is defined to be the 'non-event' for estimation of all the BNL models. Using the condition that $\sum_{j=1}^{J} p_{ij} = 1$, we estimate the transition probability of remaining in the current state (i.e., j = i) as: $\widehat{p_{ii}} = 1 - \sum_{j\neq i}^{J} \widehat{p_{ij}}$. As a result, there are only 20 BNL models estimated in Table 5.

[16] Although we do not have access to out-of-time forecasts of the macroeconomic variables – i.e., unemployment rate, industrial production index, housing affordability index – or market variables – i.e., HPI, we include them knowing that forecasts of these variables are readily available in the market through third-party data providers and are commonly used in practice in the development of loss forecasting and stress testing models.

[17] We tested the joint hypothesis that the coefficients on the macroeconomic and market variables are all zero – i.e., $H_0: \beta_z = 0$, for all v_t using a likelihood ratio test $-2[lnL_r - lnL] \sim \mathcal{X}_{df-dfr}^2$ where lnL is the ln likelihood of the full model and lnL_r is the ln likelihood of the restricted (i.e., no macro variables) model – the restricted model is nested within the full model. The joint hypothesis is rejected for all the BNL models.

defaults or prepays over the full sample period) from the subset of current loans as of December 2004. We track the estimated, loan-specific monthly conditional transition probabilities from December 2004 through December 2011 that were generated from the regression models in Table 5a and 5b.

The initial values for the covariates of both borrowers are summarized in Table 6. Both loans were originated in 2003; neither was identified as a jumbo, alt-A, nor subprime loan; both were fully documented, and were used to refinance an existing mortgage. Although borrower B has a lower FICO score at origination (i.e., 628 vs 665), borrower B has a much lower LTV at origination (i.e., 53%) compared to borrower A's hybrid ARM loan with an LTV at origination of 80%. Borrower B borrows an amount (at origination) that is roughly half the amount of borrower A. Moreover, borrower B has a lower interest rate (7.4%) that is fixed over the life of the loan versus borrower A, who has a hybrid ARM loan with an original higher interest rate (7.9% at origination). The spread over the 30-yr Fixed Mortgage Rate on Loan B is lower, which suggest the market likely views borrower B as relatively less risky than borrower A.

The delinquency histories plotted in Figure 2a are consistent with that conclusion. Borrower B remains current in all 96 months between January 2005 and December 2011. Borrower A, however, is at least 30 dpd in 20 of the 96 months, at least 60 dpd in eight and as high as 90 dpd in two of the 96 months. Those results suggest that, even though both borrowers are current at the beginning of the observation period and have similar risk profiles at origination, the actual payment behavior varies significantly over time in ways that is consistent with borrower A being the more risky borrower: a result that should be reflected in the time path of the estimated conditional, loan-level, transition probabilities over a mix of economic conditions.

In Figures 2b – 2k, we plot a selected subset of the estimated monthly transition probabilities for both borrowers over the full sample period. We include the unconditional maximum likelihood estimator (MLE) in each of the graphs as a baseline for comparison purposes. We focus primarily on the relative relationship between the time paths of the p_{ij}s of (1) the off-diagonal transition probabilities (i.e., the $p_{i,j-1}$ and $p_{i,j+1}$), which reflect the transitions to a better or worse state, and (2) the transitions to default p_{id}. In general the relative location of the time paths of the p_{ij}s are consistent with the observation that borrower B is less risky as reflected in the significantly higher probability of remaining current over the full sample (i.e., Figure 2b) and the $p_{i,j+1}$ ($p_{i,j-1}$) curves that are systematically above – i.e., Figures 2d, 2g, and 2j (below – Figures 2c, 2e, and 2h) – those of borrower A. More specifically, although the paths of the transition probabilities are consistent with the changing economic conditions and time-varying loan and borrower-specific conditions, borrower B is relatively more (less) likely to recover (further decline) from one or more missed payments over all time periods. Moreover, borrower B is less likely to transition to default over the sample period as reflected in the relatively lower time path of the p_{id}s in Figures 2f, 2i, and 2k.

We also plotted the time paths from each state i to prepayment for both borrowers in Figure 3. Borrower B is relatively more likely to transition from states 2, 3, and 4 (i.e., 30dpd, 60dpd, and 90dpd) in time t to prepay in time t+1 over all time periods than borrower A; the exception is the transition from state 1 (i.e., current) to prepayment, in which the results are mixed.

The time path plots in Figures 2 and 3 are consistent with the expectation that the estimated conditional transition probabilities in Table 5 vary significantly over time due to changes in loan-/borrower-specific and economic conditions. Moreover, the estimated conditional transition probabilities capture the relative riskiness of the borrowers as reflected in the relative location of the time path of the p_{ij}s in Figure 2.

In Figures 4 and 5, we generalize the results of the analyses of the time paths of the transition probabilities of two specific individual borrowers to that of the monthly average p_{ij}s for the full sample relative to the unconditional MLE (time invariant) transition probabilities. Similar to the results for the time path of the loan-specific transition probabilities, the monthly averages for the full sample vary significantly over time even after accounting for a considerable amount of seasonal (month-specific) variation. It is, however, the trend in the time paths of the off-diagonal transition probabilities over a mix of economic conditions that is most important for our purpose.

During a decline in economic conditions, we would expect to observe (1) a decrease in the likelihood a delinquent account would transition to a lower delinquent state, and (2) an increase in the likelihood a delinquent account would transition to a higher delinquent state. That is, we should observe, during a downturn, a downward (upward) time trend in the monthly time path of the p_{ij}(t) for all $i > j$ ($i < j$).[18]

In Figure 4b - g, we plot the trends in the transition probabilities from higher to lower delinquency states (i.e., $i > j$; Figures 4.c, 4.e, and 4.g) and from lower to higher delinquency states (i.e., $i < j$; Figures 4.b, 4.d, and 4.f) during the 2008-2010 crisis and post-2010 recovery periods.[19] For the most part, the results are consistent with expectations (see E(Δp_{ij}) in Table 7). Only the trend in the time path for p_{34} during the 2008-10 downturn is clearly inconsistent with expectations. Moreover, the trends in the time paths for p_{43} during the downturn (2008-10) and p_{21} during the post-2010 recovery display no apparent trend (i.e., the graphs reveal no definitive upward or downward trend), and there is too much variation in the time path of p_{12} during the post-2010 recovery period to identify a trend. For the remaining eight possibilities, the trend in the time paths are consistent with our expectations.

The plots in Figures 5a-c, however, clearly show a downward trend in the time paths of the transition probabilities from each of the delinquency states (i.e., i = 2, 3, and 4) to default (j=6) over both the 2008-2010 crisis and post-2010 recovery periods.[20] Although the downward trend in the time paths to default during the post-crisis period is consistent with our expectations that the likelihood of transitioning to default during an recovery should decline as the economy improves, the downward trend in the time path of the transition probabilities p_{i6} for i=2,..,4, during the crisis is counterintuitive. The observed trend is most likely reflecting the dramatic increase in the average age of the loans in the portfolio in our data following the huge decline in new loans added to the MBS Data pool of securitized mortgages beginning around early-2008 (see the trend in the age distribution in Table 8), which in its self would lower the likelihood of transitioning to default, but would also likely increase the percentage of loans with larger equity positions to avoid default even from higher initial delinquency states.

In this section, we outline the design of a 1st-order Markov transition model in which the transition probabilities are conditional on loan/borrower-specific characteristics and economic/market conditions. We estimated the transition model using a large pool of securitized single-family, owner occupied, non-agency mortgages and find that, overall, the conditional Markov model holds up well both statistically and conceptually over a mix of economic conditions including the 2008-10 crisis period when evaluated on the (in-)sample data.

---

[18] Conversely, we would expect during an upturn that the trend in the monthly time path of the p_{ij}(t) for all $i > j$ ($i < j$) would decline (increase).

[19] The crisis and post-crisis recovery periods are identified by the dates between the two, and above the rightmost, vertical reference lines in each graph, respectively.

[20] As noted above in Section III.b, it is not practical to model all the transition probabilities (p_{1j}) conditional on the X_{1j}. In some cases, conceptually sound rational for modeling the transition probabilities does not exist. The transition probability from state i =1 (i.e., current) in time t to state j=6 (i.e., default) is one such case. For that reason, we do not estimate the conditional transition probability from state i =1 to j=6 nor report a time path for p_{16} in Figure 5.

3.4. Assessing Forecast Accuracy

In this section, we evaluate the out-of-time forecast accuracy of the conditional transition matrix using the conditional MLE-based transition matrix as a benchmark. We assess the accuracy of the models based on their ability to predict the cumulative, 24-months ahead default and prepayment rates for the December 2011 cohort: a forecast design similar to that used in practice for loss forecasting, stress testing, and portfolio valuation purposes. We forecast the cumulative default and prepayment rates from January 2012 through December 2013 for all active loan as of December 2011 (Z_0) using (1) the simple Markov process summarized in eq (3) in which the maximum likelihood estimators of the unconditional transition probabilities are used and (2) the process summarized in eq (7) in which the multinomial models are used to generate the conditional transition matrices $P_{ij,(\tau,\tau+1)}$ for each month τ, τ =0, 1, …, 23 over the forecast period.[21]

We compare the forecasted cumulative default and prepayment rates derived from both the unconditional and conditional transition matrices relative to the actual cumulative default and prepayment rates over the 24-month forecast period. The out-of-time forecast of the default and prepayment rates are summarized in Figure 6. The forecasts derived from the unconditional MLE transition matrices are relatively poor especially compared to those using the conditional transition matrices. Both the cumulative default and prepayment rates using eq (3) over predict actual rates over the 24-month forecast horizon; and in both cases the forecasts appear to diverge from the actual values as the forecast period increases. In contrast, the forecasted default rates derived from the conditional transition matrices (eq (7)) closely follow the actual default rates only slightly over predicting the default rate towards the end of the forecast period; and although the conditional transition matrices under predict the prepayment rates, they are significantly more accurate than the forecast from the unconditional transition matrix.

In addition to the visual comparison of the forecasted default and prepayment rates, we also compute the $Theil\text{-}U$ statistics for each event by model type.[22] The $Theil\text{-}U$ statistics derived from the forecasts generated using the unconditional transition matrix in eq (3) – 0.669 for defaults and 0.891 for prepayments – are much higher than those generated from the forecasts using the conditional transition matrices in eq (7) – 0.123 for defaults and 0.271 for prepayments – supporting the conclusion that the forecasts based on the conditional transition matrix approach are much more accurate.

We recognize that our use of the historical values of the macroeconomic variables – information not known at time of forecast – over the out-of-time forecasts horizon surely improves the accuracy of the conditional forecasts that would only be realized in practice if the forecasts of the macroeconomic variables are themselves accurately predicted over the forecast period. Unfortunately, we do not have access to forecasted values of the macroeconomic and market variables. For that reason, we follow the literature (Malik and Thomas, 2012; Bellotti and Crook, 2013a) and use the actual values of the macroeconomic variables over the forecast period to illustrate that potential benefits of using conditional transition matrices under the best possible conditions.

Our results show that indeed the transition matrix constructed using the conditional transition probabilities outperforms the unconditional MLE transition matrix approach: not an unexpected result given that the

---

[21] The conditional transition matrices used were derived using the actual values for the macroeconomic and market variables used to forecast the conditional transition probabilities and overstate the accuracy of the conditional model that, in practice, would have to use forecast of the macroeconomic and market variables, which would introduce error that are likely to reduce the overall accuracy of the models. Unfortunately, we do not have access to forecasted values of the macroeconomic and market variables.

[22] A Theil-U statistic closer to 0 indicates greater forecasting accuracy

MLE approach constructed over the full sample effectively averages out the variation in the p_{ij}'s over a mix of economic conditions to produce a more through-the-cycle (TTC) type estimate of the transition probabilities. If, however, we estimate the unconditional MLE transition matrix using more recent monthly data, say, the payment behavior from the previous three years, the unconditional transition probabilities may better reflect the economic conditions driving the migration behavior over the forecast period.

In Figure 7, we compare the forecasts using (1) the unconditional MLE transition matrix based on the full sample (2004-2011), (2) the conditional transition matrices (2004-2011), and (3) the MLE unconditional transition matrix base on the most recent 3-year period (2009-2011). The accuracy of the out-of-time forecasts of the prepayment rate improves significantly as reflected in the much lower $Theil\text{-}U$ statistic (0.064) when using only the most recent three years of performance history to construct the unconditional transition matrix. On the other hand, the accuracy of the forecasted out-of-time default rate decreases significantly as reflected in the much higher $Theil\text{-}U$ statistic (0.845): over predicting the default rate (i.e., forecast: 21.7%) by nearly twice the actual default rate of 11.3% after 24 months.

The results in Figure 7 indicate that by reducing the observation period used to construct the MLE transition probabilities to a relatively short timeframe that better reflect the current and future economic condition may improve the forecast accuracy of the unconditional Markov model. This is likely to be true during periods of economic and market stability. We will test this hypothesis below by comparing the out-of-time forecast accuracy of the conditional against the unconditional (based on the most recent 3-years of data) transition matrices over several selected sub-periods reflecting economic conditions leading up to, during, and after the 2008-10 financial crisis.

3.5. *Path-Depend Process: Second-order Markov Model*

The conditional transition matrix model developed in the previous section maintains the assumption that the migration process is not "path dependent". That is, conditional on the realized values for the time-dependent factors in time t, information about the previous delinquency states in, say, t-1, t-2…t-k, does not affect the estimates of the current transition probabilities. It is possible, however, that the likelihood of migrating to a higher delinquency state in t+1 depends not only on the loan/borrower, market conditions, and delinquency status in time t, but also the delinquency states in previous time periods, e.g., t-1. In that case, knowledge of the history of the borrower's delinquency status may improve model performance if the process is better represented by a higher-order Markov chain.

More specifically, we introduced time-varying factors and macroeconomic variables (i.e., systematic factors) into the model to indirectly capture the impact of broadly observed trends in economic conditions. They may not, however, capture all the borrower-specific information necessary to estimate the probability of the transition in time t+1 (Malik and Thomas, 2012). For that reason, we investigate the possibility that the migration process is path dependent by augmenting the specification of the BNL models with dummy variables that indicate the lagged delinquency status of each borrower as of time t-1 as time dependent covariates. In that way, we proxy a second-order Markov process using the lagged transition states as conditioning variables. More specifically, we introduce four additional indicator variables S_τ (τ= 1, 2, 3, 4) in which S_τ = 1 if the borrower is in state τ in the previous time period, t-1, and 0 otherwise to the set of covariates in Table 5,[23] and test the assumption that borrower's payment behavior over a mix of economic conditions follows a 1st-order of the Markov process.

---

[23] For example, in the BNL model for the transition probability of migrating from 30 dpd state in time t-1 to state 60 dpd in time t, when the mortgage was current in time t-2 is: $p_{30,60} = f(w_k, x_{k,t}, v_t, \Sigma)$ where $\Sigma = [1\ 0\ 0\ 0]$ and $w_k, x_{k,t}, v_t$ are defined above.

Because our primary purpose is to evaluate the application of the Markov framework to *forecast* losses over a mix of economic conditions, the in-sample analysis is only a first step in assessing the feasibility of using a Markov state transition approach for stress-testing or reserving purposes especially as market conditions are expected to change over the forecast period. In the next section, we evaluate the out-of-time forecast accuracy of the Markov state transition model outlined above to determine if allowing the transition probabilities to adjust to changing conditions over time improves forecast accuracy or decreases accuracy through the compounding of errors. Moreover, we evaluate the potential impact of modeling the migration process as a 2^{nd}-order Markov process on the out-of-time forecast accuracy.

4. Forecasting Performance over the Economic Cycle

There is always a concern that our results are sample-specific. To address this concern, we estimate the conditional transition matrices over five sub-periods of our sample and assess the out-of-time accuracy of the forecasts for various end-of-period cohorts from 2006 thru 2011. We isolate the impact of incorporating economic conditions on borrower payment behavior by maintaining a common model specification, $X_{ij,k}(t|T) = \{w_k, x_k(t), v(t)\}_{ij}$, for each transition probability p_{ij} conditioned on data from January 2004 until the end of each year from 2006 through 20011:

$$p_{ij}(t|T) = f\big(X_{ij}(t|T)\big)$$

for all T = 200612, 200712, …, 201112. For example, the first sub-sample begins in January 2004 and ends in December 2006 with an out-of-time forecast period from January 2007 through December 2008 using the delinquency distribution ($Z_{t=0}$) for the cohort of active loans as of December 2006; the second sub-sample begins in January 2004 and ends in December 2007 with an out-of-time forecast period from January 2008 through December 2009 using the delinquency distribution of active loans as of December 2007; and so on through 2011.

Each of the sub-sample models are developed using the same specification outlined in Table 5 to isolate the impact of increasing the sample with an additional 12 months of data. The objective is to capture the changes in market conditions on payment behavior over the financial cycle. We report the $Theil\text{-}U$ statistic for the 24-months out-of-time forecasts of the performance distribution in Table 9 for each of the sub-samples. The results for the unconditional model (i.e., eq (3)) are based on the MLE of the transition probabilities using the most recent three-year performance period. We also report the statistics for the forecasts of the performance distribution (Z_τ) using the BNL regression results to calculate the conditional transition probabilities used in eq (7) for models that both include and exclude macroeconomic variables. Similarly, we report the $Theil\text{-}U$ statistics for a 2^{nd}-order Markov model outlined above using the forecasted delinquency distribution S_τ for each loan in time t-1 to forecast the delinquency distribution in time t +1.

For loss forecasting purposes, the accuracy of the default and prepayment predictions are of most interest. For that reason, we focus primarily on the relative accuracy of the out-of-time forecasts of the default and prepayment rates by model type for each of the sub-samples. We first compare the forecast accuracy of the 1^{st}-order Markov model to the unconditional model. For five of the six sub-samples the modeling approach in which the transition probabilities are conditional on loan/borrower-specific characteristics *and* macroeconomic variable, the $Theil\text{-}U$ statistics are significantly lower than those derived from (1) conditioning on loan/borrower-specific characteristics only and (2) the unconditional models. The 24-month forecasts of the default rate for the 200912 cohort (forecast horizon: 2010-11) is only slightly less accurate ($Theil\text{-}U$: 0.2244) than the model in which the transition probabilities are conditional on loan/borrower-specific characteristics only ($Theil\text{-}U$: 0.1882); and, the prepayment rate for the 201112 cohort (forecast horizon: 2012-13) is the least accurate among the three methods ($Theil\text{-}U$: 0.2706).

Similar to the results for the 1st-order Markov model, including the macroeconomic variables in the 2nd-order Markov model improves forecast accuracy. More interestingly, the forecast accuracy of the 1st-order order (including macro variables) and the 2nd-order (including macro variables) conditional prepayment models are nearly the same across the six sub-samples. The $Theil\text{-}U$ statistics are very similar with the 1st order model is slightly more accurate at forecasting the *prepayment* rate for the 200812, 201012, and 201112 cohorts (i.e.,$Theil\text{-}U$ for 1st-order/2nd-order: **0.0488**/0.0505, **0.1605**/0.1685, and **0.2706**/0.2714) and the 2nd-order model slightly more accurate at forecasting the 200612, 200712, and 200912 cohorts (i.e.,$Theil\text{-}U$ for 1st-order/2nd-order: 0.2353/**0.2313**, 0.1061/**0.0946**, and 0.1829/**0.1739**).

The relative accuracy of the 1st-order Markov and 2nd-order conditional *default* models, however, tell a different story. A comparison of the $Theil\text{-}U$ statistics shows that the 1st-order Markov model (including macro variables) is more accurate at forecasting the default rate for the 200612, 200712, 200812, and 201112 cohorts (i.e., $Theil\text{-}U$ for 1st-order/2nd-order: **0.3495**/0.5040, **0.0208**/0.1002, **0.0500**/0.1468, and 0.1229/0.1528) and the 2nd-order model more accurate at forecasting the 200912 and 201012 cohorts (i.e., $Theil\text{-}U$for 1st-order/2nd-order: 0.2244/**0.1307** and 0.2544/**0.0794**).

To better understand the magnitude of the differences in the forecast accuracy of the models, we plot the forecasts of the default and prepayment rates against the actual rates for each sub-sample in Figure 8. The graphs suggest that through a period of rapid decline in the housing market (i.e., cohort graphs: 200712 and 200812 in which the forecast windows span the crisis period: 2008-2009), knowledge of the borrower's previous delinquency states did not improve the forecast accuracy of either the default or prepayment models. However, forecasts through a period in which the market was recovering from the housing crisis (i.e., cohort graphs: 200912 and 201012 in which the forecast windows span the post crisis period: 2011-2012), knowledge of a borrower's previous delinquency states generates more accurate out-of-time forecast of the default rate.

These results provide some support for the hypothesis that, at least in our data, knowledge of previous delinquency states may improve the out-of-sample forecasts of the default rate. These results are not conclusive. They suggest additional research on the impact of "momentum" in the transition from earlier time periods on the migration probabilities (i.e.., that the order of the migration process is greater than 1) is warranted.

VI. Conclusion

The MLE approach is still used to construct conventional *unconditional* transition matrices in practice, although more so for corporate than retail lending purposes. However, an unconditional transition matrix is useful only if borrower payment behavior satisfies the assumptions of a Markov chain: stationary and time homogenous. Those are very strong assumptions that are not always consistent with the process we wish to model. If the assumptions do not hold or are inconsistent with the process generating the data, predictions and/or out-of-time forecasts generated from the model are unlikely to represent actual behavior. We find that for our sample of first-lien, single-family, owner-occupied mortgages from 2004-2013 that the assumptions of a Markov chain do not hold and that the 24-month out-of-sample forecasts of default and prepayment rates are by-in-large nonstationary. That is especially true during the periods of changing economic conditions (i.e., 2008-2011) and would likely increase significantly if the forecast period is extended beyond 24-months (e.g., life-of-loan forecasts).

One important benefit of the *conditional* transition probability approach outlined in this paper is it would allow us to analyze the impact of changing macroeconomic and market conditions on the underlying credit quality of the loan portfolio for loss forecasting/ALLL, stress-testing, and portfolio valuation purposes. Because the transition probabilities adjust to reflect changes in economic and market conditions, the out-of-time forecast are more likely to closely track actual behavior over the financial cycle.

Our results are consistent with those of Smith, et al., (1996), Betancourt (1999), Moline Utrilla and Constantinou (2010), Grinshaw and Alexander (2011) and Leow and Crook (2013) in which they find borrowers' repayment behavior violate the assumptions of a Markov process and who also find that the fundamental requirement of time-homogenous condition fundamental to a Markov process is too strong an assumption. Moreover, and more importantly, we find the conditional transition approach generates more accurate out-of-time forecasts over a time horizon typically used in practice and performs well during periods of changing economic conditions.

Table 1: Loan Distribution and Default Rate by Attribute

Timeframe[1]	2004-2013		2004-2007		2008-2010		2011-2013	
	Distribution of Loans	Default Rate	Distribution of Loans	Default Rate	Distribution of Loans	Default Rate	Distribution of Loans	Default Rate
Overall Default Rate		28.0%		9.9%		32.4%		15.9%
Number of observations	1,622,334		1,570,156		752,234		346,917	
Loan Type[2]								
Jumbo	0.2321	21.6%	0.2336	4.8%	0.2824	24.5%	0.2729	12.3%
Alt-A	0.0119	38.5%	0.0121	14.3%	0.0108	49.8%	0.0078	23.4%
Subprime	0.2154	40.9%	0.2199	18.6%	0.1827	48.8%	0.1306	26.2%
Other	0.5547	25.8%	0.5488	8.7%	0.5348	31.3%	0.5938	15.2%
Documentation								
Full	0.4574	26.3%	0.4489	10.4%	0.4410	29.5%	0.4835	14.0%
No Doc	0.0210	21.6%	0.0212	4.8%	0.0263	22.7%	0.0275	13.5%
Stated Income	0.1289	40.0%	0.1293	13.1%	0.1332	48.4%	0.1028	23.6%
Low, Limited, or Reduced	0.1859	28.0%	0.1909	6.6%	0.2162	32.6%	0.2060	16.3%
Other	0.0668	26.2%	0.0682	6.0%	0.0847	27.7%	0.0846	14.5%
Missing	0.1399	24.2%	0.1415	12.1%	0.0987	30.1%	0.0956	17.6%
Loan Purpose								
Purchase	0.3736	30.9%	0.3759	12.1%	0.3705	34.9%	0.3424	15.8%
Refinance	0.6114	26.1%	0.6093	8.4%	0.6121	31.2%	0.6428	15.9%
Construction	0.0026	31.3%	0.0026	14.9%	0.0026	29.9%	0.0026	14.5%
Debt Consolidation	0.0027	31.1%	0.0025	11.7%	0.0030	30.8%	0.0031	19.3%
Other	0.0096	27.2%	0.0097	13.2%	0.0117	19.7%	0.0091	15.6%
Credit Score								
FICO (origination)	665	676/641	663	669/614	679	694/651	695	701/665
Missing	0.2372	17.8%	0.2422	9.2%	0.1421	24.3%	0.1559	14.1%
000 - 550	0.0586	40.8%	0.0598	21.0%	0.0440	48.8%	0.0319	26.3%
550 - 620	0.1607	41.0%	0.1641	17.6%	0.1427	48.8%	0.1017	26.0%
620 - 680	0.2092	37.3%	0.2120	11.1%	0.2290	43.5%	0.1917	22.0%
680 - 720	0.1308	29.0%	0.1304	5.5%	0.1662	32.4%	0.1727	16.3%
720 - 850	0.2035	15.4%	0.1915	2.1%	0.2760	16.3%	0.3461	9.0%
850 +	0.0000	22.2%	0.0000	4.5%	0.0000	13.6%	0.0000	14.3%
Current Loan-to-Value[3]								
LTV (origination)	76.9	75.4/80.7	77.0	76.4/82.6	75.4	73.1/80.0	74.2	73.5/78.1
00 - 60	0.1524	10.4%	0.1526	3.2%	0.2516	9.0%	0.2044	5.9%
60 - 80	0.4541	25.5%	0.4500	8.3%	0.3699	27.7%	0.3222	10.3%
80 - 80	0.0957	30.9%	0.0988	11.9%	0.0005	28.6%	0.0002	10.9%
80 - 95	0.2470	39.6%	0.2508	14.6%	0.2703	48.3%	0.2022	18.7%
95 - 105	0.0437	40.6%	0.0434	17.0%	0.0750	62.1%	0.0890	23.1%
105 +	0.0039	59.2%	0.0011	14.9%	0.0304	69.1%	0.1794	30.3%
Loan Characteristics								
Balloon Payment	0.0466	60.4%	0.0467	17.8%	0.0631	60.6%	0.0333	33.0%
Interest Only	0.2024	32.7%	0.2056	7.7%	0.2409	38.0%	0.2129	18.6%
Fixed Rate	0.3984	20.8%	0.3871	5.7%	0.4925	19.9%	0.6256	12.0%
Hybrid ARM	0.3770	35.2%	0.3864	16.8%	0.2420	55.0%	0.1335	29.4%
Prepayment Penalty	0.4095	40.4%	0.4154	13.9%	0.4292	46.2%	0.3490	23.3%
Teaser Rate Period	0.4490	33.5%	0.4594	12.4%	0.2885	41.8%	0.0890	19.6%
Loan Amount (origination)	$250,567		$249,988		$274,510		$256,842	
Interest Rate (origination)	7.054		7.109		7.330		6.033	
Spread over 30-yr Fix Mgt Rate	1.040		1.060		1.577		1.298	

[1] The number of loans observed in each timeframe represents all loans that existed during that period including loans that were originated prior to the start of the time frame that are still active and new loans originated durint that period.

[2] Because aproximately 1% of the jumbo loans are also altA or Subprime loans, the Loan Type percentages do not sum to 1.0 .

[3] The LTV percentages do not sum to 1.0 due to missing values for approximately 0.25% of the loans.

Table 2: MLE Transition Matrix (2004-2013) - $P_{f,ij}$

The unconditional monthly transition probabilities ($p_{ij} = n_{ij}/n_i$) are derived using all active loans from 2004 thru 2013

States	Current	30dpd	60dpd	90dpd	Prepay	Default
Current	0.9500	0.0337	0.0006	0.0001	0.0156	0.0001
30dpd	0.2762	0.4795	0.2246	0.0035	0.0156	0.0006
60dpd	0.0945	0.1537	0.3559	0.3791	0.0112	0.0056
90dpd	0.0551	0.0290	0.0737	0.2031	0.0110	0.6281
Prepay	0	0	0	0	1	0
Default	0	0	0	0	0	1

Table 3: Subperiods Unconditional MLE Transition Matrices

The unconditional monthly transition probabilities ($p_{ij}= n_{ij}/n_i$) are derived using all active loans during the subperiods: 2004 -2007 (pre-crisis), 2008-2010 (crisis), and 2011-2013 (post crisis).

	MLE Transition Matrix (2004-2007)				$P_{1,ij}$	
States	**Current**	**30dpd**	**60dpd**	**90dpd**	**Prepay**	**Default**
Current	0.9450	0.0330	0.0008	0.0001	0.0210	0.0001
30dpd	0.3126	0.4482	0.2054	0.0049	0.0280	0.0009
60dpd	0.1186	0.1851	0.3195	0.3461	0.0230	0.0078
90dpd	0.0646	0.0411	0.0902	0.1639	0.0221	0.6181
Prepay	0	0	0	0	1	0
Default	0	0	0	0	0	1

	MLE Transition Matrix (2008-2010)				$P_{2,ij}$	
States	**Current**	**30dpd**	**60dpd**	**90dpd**	**Prepay**	**Default**
Current	0.9509	0.0380	0.0005	0.0001	0.0105	0.0001
30dpd	0.2342	0.4899	0.2677	0.0028	0.0048	0.0005
60dpd	0.0744	0.1249	0.3623	0.4301	0.0033	0.0050
90dpd	0.0469	0.0214	0.0609	0.2188	0.0045	0.6475
Prepay	0	0	0	0	1	0
Default	0	0	0	0	0	1

	MLE Transition Matrix (2011-2013)				$P_{3,ij}$	
States	**Current**	**30dpd**	**60dpd**	**90dpd**	**Prepay**	**Default**
Current	0.9643	0.0275	0.0001	0.0000	0.0080	0.0001
30dpd	0.2657	0.5479	0.1820	0.0009	0.0034	0.0002
60dpd	0.0907	0.1561	0.4339	0.3139	0.0036	0.0017
90dpd	0.0636	0.0281	0.0823	0.2407	0.0083	0.5770
Prepay	0	0	0	0	1	0
Default	0	0	0	0	0	1

Table 4: Chi-Square Test of Stationarity[1]

The results of χ^2 tests show that, for each state i, the hypothesis of stationarity is rejected over all three subperiods 2004-2007, 2008-2010, and 2011-2013

	200401 to 200712		200801 to 201012		201101 to 201312	
State i	**Chi-Sqr**	**p-value**	**Chi-Sqr**	**p-value**	**Chi-Sqr**	**p-value**
Current	72588.8	0.0001	238387.4	0.0001	5627364.8	0.0001
30 dpd	38847.0	0.0001	216555.4	0.0001	814166.9	0.0001
60 dpd	16139.2	0.0001	96047.7	0.0001	46448.7	0.0001
90 dpd	6122.0	0.0001	27985.4	0.0001	22971842.0	0.0001

[1] H_0: test of $p_{ij}(t) = p_{ij}$ for all $j=1\ldots6$; $t = 1\ldots36$

Table 5a: Binomial Logit Estimates of the Conditional Transition Probabilities (P_{ij})

	Variable Name	$p_{c,30}$	$p_{c,60}$	$p_{c,90}$	$p_{c,p}$	$p_{c,d}$	$p_{30,c}$	$p_{30,60}$	$p_{30,90}$	$p_{30,p}$	$p_{30,d}$
	Intercept	-3.1138	-4.2213	-9.0128	-0.5399	-9.4074	0.0632	-0.3616	-2.8502	5.4327	-2.4012
Months-on-Book	[7 - 12] Months on Book	0.1562	0.0983		0.5381		-0.3718	-0.1817	-0.2625	-0.0233	
	[13 - 18] Months on Book	0.3310	0.1297		0.7749		-0.5860	-0.2860	-0.4848	-0.0443	
	[19 - 24] Months on Book	0.4472	0.1746		0.7635		-0.6986	-0.3152	-0.5505	-0.1959	-0.1857
	[25 - 36] Months on Book	0.5550	0.1168		0.7958		-0.8570	-0.3856	-0.8806	-0.1353	
	[27 - 48] Months on Book	0.5518	-0.0077		0.4657		-0.9552	-0.5346	-1.0123	-0.5374	
	[49 - 60] Months on Book	0.4987	-0.0607		0.3275		-1.0282	-0.6666	-1.0569	-0.8215	-0.3694
	[61 +] Months on Book	0.6073	-0.0812		0.4952		-1.1245	-0.8286	-1.3036	-0.9200	-0.4738
Loan/Borrower Characteristics											
Loan Type	Jumbo - Loan Type	-0.3293	0.1886		0.1305		0.0532	0.1143	0.4136	0.1719	0.4981
Loan Attribute	Prepayment Penalty	0.1494	0.1123		-0.4879		-0.0395	-0.0045	0.0185	-0.5583	-0.0401
Documentation	Full Documentation	-0.0128	-0.0796				-0.0885	-0.2457	-0.0947	-0.1626	-0.3742
	No Documentation	0.0447	0.0266		0.0859			0.0211		0.1444	-0.2226
	Stated Income	0.1621	0.1631		0.0213		-0.0538	-0.0100		-0.0262	-0.0240
	Low, Limited, or Reduced Doc	0.0142	-0.1998		-0.1141		0.0106	0.0145		-0.2291	-0.3899
	Missing Documentation	0.2171	0.5058		-0.1201		0.0213	-0.1167		-0.2546	0.0596
Loan Purpose	Loan Purpose - Purchase	-0.0321	0.1258		0.0326		0.0379	0.1787	0.1707	-0.1208	0.2633
	Loan Purpose - Construction	0.0754	0.2340		-0.1545		-0.1846	-0.2054			
	Loan Purpose - Debt Consolidation	0.0175						-0.0125			
Credit Score[a]	FICO Origination 720 - 800	0.9417	-1.9932		-0.5595		0.4808	0.1146	-0.6669	0.2597	-0.8606
	FICO Origination 680 - 720	0.5852	-0.8807		-0.4213		0.2810	0.1863	-0.3834	0.0697	-0.5330
	FICO Origination 550 - 620	-0.5655	0.7356		-0.1051		-0.2964	-0.1893	0.1673	-0.3551	0.1571
	FICO Origination 000 - 550	-1.5362	1.1936		-0.0231		-0.4557	-0.1433	0.3240	-0.3407	0.3874
Current Loan-to-Value[3]	Current LTV 60 - 80	0.3595	0.4974		-0.0836		-0.1438	0.1112	0.2237	-0.1628	0.2570
	Current LTV 80 - 95	0.5069	0.6657		-0.3791		-0.2511	0.1803	0.2615	-0.4844	0.3217
	Current LTV 95 - 105	0.6306	0.6745		-0.8367		-0.3179	0.2766	0.3032	-0.8717	0.2086
	Current LTV 105 +	0.8747	0.9077		-1.6058		-0.3865	0.4879	0.3878	-1.1911	0.2267
Loan Characteristics	Balloon Payment	0.2128	0.3707		-0.2029			0.0050	0.2472	-0.2278	0.2361
	Interest Only	0.0685					0.0528	0.2089	0.1459	0.0855	0.1295
	Fixed Rate	-0.2111	-0.2260		-0.6363		-0.0748	-0.3121	-0.3382	-0.6130	-0.1785
	Hybrid ARM	0.2065	0.4874		0.3258			-0.0163	0.3139	0.1115	0.2594
	Teaser Rate Period	-0.0705	-0.0514		-0.3006		-0.0941	-0.1453	-0.5216	-0.3390	0.1660
	Spread over 30-yr Fix Mgt Rate	0.2216	0.3186		0.1395		-0.0620	0.0237	0.0657	0.0293	0.0826
Macroeconomic Variables											
% Chg Since Origination	%ΔUS HPI(t_0 - t)	-0.5399	-1.1631				0.2860	-0.0772	-0.8084	0.0545	-1.9056
	%ΔUS Unemployment Rate(t_0 - t)	0.1472			-0.0264		-0.0735	0.0892		-0.1630	-0.0818
	%ΔHome Affordability Index(t_0 - t)	-0.2808	-0.4347		-0.4284		0.2217	0.1439		-1.2692	-1.4105
3-MonthLlag	US Unemployment Rate(t-3)		0.0692		-0.0048		-0.0173	0.0566			
	US HPI(t-3)		-0.0257		0.0294		0.0060	-0.0192	-0.0123	0.0543	-0.0105
	Home Affordability Index(t-3)		-0.0192		0.0051			-0.0012	-0.0222		
	Industrial Production Index(t-3)				-0.0375		0.0110	-0.0031		0.0068	0.0346
6-Month Lag	US Unemployment Rate(t-6)	-0.0091	-0.0786		0.0170		0.0167	-0.0491			
	US HPI(t-6)	0.0008	0.0286		-0.0242		-0.0058	0.0200	0.0146	-0.0450	0.0131
	Home Affordability Index(t-6)	-0.0080			-0.0133		0.0005	-0.0017	0.0069	-0.0181	-0.0084
	Industrial Production Index(t-6)	-0.0031	-0.0223		0.0046		-0.0056	0.0058		-0.0721	-0.0709
Seasonal Factor		Yes	Yes		Yes		Yes	Yes	Yes	Yes	Yes
	Brier Score	0.0330	0.0007	-	0.0170	-	0.2240	0.2100	0.0078	0.0312	0.0015
	AUC	0.7370	0.8230	-	0.6920	-	0.6020	0.6250	0.7030	0.7950	0.6940
	-2[lnL_r - *lnL*]; *H0:* $\beta_z = 0$, *for all* z_t	< 0.001	< 0.001		< 0.001		< 0.001	< 0.001	< 0.001	< 0.001	< 0.001

a Although the FICO score distribution, on average, may be relatively stable, an individual borrower's score may change quickly and significantly over time especially during periods of economic stress. The score as of the origination data may perform well at indicating a borrowe's relative credit quality, however, over time, the score at time of origination is less likely to represent the borrower's current expected payment performance. For that reason, we impose a constant decay function that reduces contribution of the origiantion FICO score each month until it's impact is eliminated 60 months after origination.

Table 5b: Binomial Logit Estimates of the Conditional Transition Probabilities (P_{ij})

	Variable Name	$p_{60,c}$	$p_{60,30}$	$p_{60,90}$	$p_{60,p}$	$p_{60,d}$	$p_{90,c}$	$p_{90,30}$	$p_{90,60}$	$p_{90,p}$	$p_{90,d}$
	Intercept	**-0.4354**	**0.3567**	**1.2036**	**7.2657**	**-0.0646**	**-1.2254**	0.3021	-0.2977	**2.4650**	**0.9976**
Months-on-Book	**[7 - 12] Months on Book**	**-0.5164**	**-0.1854**	**-0.5220**	**-0.8830**	**-0.2097**	**-0.6141**		-0.0995	**-1.4062**	**-0.6817**
	[13 - 18] Months on Book	**-0.7521**	**-0.2471**	**-0.7856**	**-1.0901**	**-0.5651**	**-0.9532**		**-0.2191**	**-1.9377**	**-1.1352**
	[19 - 24] Months on Book	**-0.8887**	**-0.2888**	**-0.9091**	**-1.2840**	**-0.6224**	**-1.0175**		**-0.2340**	**-2.1325**	**-1.2934**
	[25 - 36] Months on Book	**-1.0128**	**-0.3505**	**-0.9515**	**-1.2418**	**-0.8695**	**-1.0314**	**-0.1107**	**-0.3553**	**-2.1360**	**-1.4193**
	[27 - 48] Months on Book	**-1.0985**	**-0.3634**	**-1.1029**	**-1.4868**	**-1.0691**	**-1.1699**	**-0.1780**	**-0.4204**	**-2.4725**	**-1.6919**
	[49 - 60] Months on Book	**-1.1724**	**-0.3165**	**-1.2228**	**-1.6100**	**-1.2272**	**-1.1674**	-0.0399	**-0.3751**	**-2.4758**	**-1.7829**
	[61 +] Months on Book	**-1.2163**	**-0.2742**	**-1.3386**	**-1.7363**	**-1.3917**	**-1.2176**	**0.0648**	**-0.2847**	**-2.5984**	**-1.9218**
Loan/Borrower Characteristics											
Loan Type	**Jumbo - Loan Type**	**-0.0885**	**-0.1470**	**0.1724**	**0.1842**	**0.4057**	**-0.0720**	**-0.1860**		**0.2129**	**0.1113**
Loan Attribute	**Prepayment Penalty**	**-0.0600**	**-0.0182**	**-0.0514**	**-0.5054**	-0.0117	-0.0196	**-0.0948**		**-0.3940**	**-0.0420**
Documentation	**Full Documentation**	**-0.0757**	0.0232	**-0.1008**	**-0.2192**	**-0.1358**				**-0.2977**	**-0.1993**
	No Documentation		0.0024	**0.0924**	0.1077		**-0.1849**	-0.0814	0.0031		**-0.0736**
	Stated Income		-0.0189	**0.0858**	0.0522		**-0.0796**	**-0.0805**	**-0.0693**	**-0.1353**	**-0.0640**
	Low, Limited, or Reduced Doc	**-0.1305**	**-0.0457**	**0.0663**	**-0.2117**	**-0.3158**	**-0.1208**	**-0.1250**	**-0.1010**	**-0.1860**	-0.0308
	Missing Documentation		0.0290		**-0.2721**		-0.0187			**-0.2636**	**-0.1130**
Loan Purpose	**Loan Purpose - Purchase**	**-0.0613**	-0.0060	**0.1271**	**-0.2098**	**0.1587**	**-0.1009**		**0.0544**	-0.0363	**0.1297**
	Loan Purpose - Construction	**-0.1678**		**-0.1328**			**0.2355**	-0.1037		0.1960	
	Loan Purpose - Debt Consolidation	**-0.3177**	**-0.2858**	**-0.2385**			**0.1158**		**-0.1854**	-0.4039	
Credit Score	**FICO Origination 720 - 800**	**0.4227**		**0.9784**	**0.7342**		**0.4677**		**-0.2113**	**1.0452**	**1.0885**
	FICO Origination 680 - 720	**0.1641**		**0.5608**	**0.3648**		0.1080		**-0.2316**	**0.5065**	**0.5906**
	FICO Origination 550 - 620	**-0.4376**	**-0.0783**	**-0.4365**	**-0.3722**		**-0.3837**		**-0.0799**	**-0.5452**	**-0.5300**
	FICO Origination 000 - 550	**-0.5255**	-0.0308	**-0.5579**	**-0.2450**		**-0.4142**		-0.0250	**-0.3656**	**-0.7564**
Current Loan-to-Value[3]	**Current LTV 60 - 80**	**-0.2244**	**-0.0990**	**0.0661**	**-0.4312**		**-0.2094**	**-0.1484**	**-0.1006**	**-0.5592**	**0.0921**
	Current LTV 80 - 95	**-0.3644**	**-0.2100**	**0.1641**	**-0.8162**		**-0.3052**	**-0.3017**	**-0.2060**	**-0.8590**	**0.1967**
	Current LTV 95 - 105	**-0.4395**	**-0.3004**	**0.2361**	**-0.9866**		**-0.3549**	**-0.4014**	**-0.3056**	**-0.8235**	**0.2280**
	Current LTV 105 +	**-0.4831**	**-0.4320**	**0.3871**	**-0.9290**		**-0.2555**	**-0.5303**	**-0.3837**	**-0.5644**	**0.3791**
Loan Characteristics	**Balloon Payment**	**-0.1379**	**-0.0998**	**-0.0275**	**-0.2416**	**0.2013**		**-0.1050**	**-0.0816**	**-0.3298**	
	Interest Only		**-0.0920**	**0.2162**			0.0322		**-0.1125**	**0.1085**	**0.1855**
	Fixed Rate		**0.1112**	**-0.1436**	**-0.2860**	**-0.1081**	**0.1059**				
	Hybrid ARM	**0.0477**	**0.0449**		**0.0877**	**0.2786**	**0.2590**			**0.2294**	**0.1162**
	Teaser Rate Period		**0.0414**		**-0.1425**	**-0.2821**	**-0.1217**			**0.1280**	**0.0404**
	Spread over 30-yr Fix Mgt Rate	**-0.0472**	**-0.0126**	**-0.0132**		**0.0501**	**-0.0465**	**-0.0326**	-0.0065		**-0.0178**
Macroeconomic Variables											
% Change Since Origination	**%ΔUS HPI(t_0 - t)**	**0.3929**	**0.2667**	**-0.2482**	**0.2573**	**-1.3531**	**0.4862**	**0.5404**	**0.4171**	**1.5687**	**1.0152**
	%ΔUS Unemployment Rate(t_0 - t)	**-0.0815**	**-0.0858**	**0.1025**	-0.0561		**-0.0457**	**-0.1438**	**-0.0678**	**-0.1269**	**0.0913**
	%ΔHome Affordability Index(t_0 - t)				**-1.3776**		**-0.1673**		**0.1216**		**0.3453**
3-MonthLlag	**US Unemployment Rate(t-3)**	**-0.0408**	**-0.0367**	**0.0344**			**-0.0475**	**-0.0789**	**-0.0455**	**-0.0370**	**-0.0228**
	US HPI(t-3)	**0.0157**	**0.0039**	**-0.0277**	**0.0588**	**-0.0203**			**0.0055**	**0.0684**	**-0.0371**
	Home Affordability Index(t-3)		**-0.0041**	**-0.0016**		**-0.0179**		**-0.0084**			**-0.0016**
	Industrial Production Index(t-3)	**0.0046**	**0.0098**		0.0024		**-0.0318**	**-0.0033**	**0.0252**	**-0.0155**	
6-Month Lag	**US Unemployment Rate(t-6)**	**0.0257**	**0.0299**	**-0.0355**			**0.0401**	0.0510	**0.0389**	**0.0581**	**0.0138**
	US HPI(t-6)	**-0.0163**	**-0.0048**	**0.0273**	**-0.0506**	**0.0223**	**-0.0182**		**-0.0062**	**-0.0612**	**0.0355**
	Home Affordability Index(t-6)		0.0002	**-0.0028**	**-0.0216**		-0.0008		**-0.0049**	**-0.0139**	
	Industrial Production Index(t-6)	0.0041	**-0.0095**	**0.0041**	**-0.0685**	**-0.0133**	**0.0483**		**-0.0178**		**0.0197**
Seasonal Factor		Yes	Yes	Yes	Yes		Yes	Yes	Yes	Yes	Yes
	Brier Score	0.162	0.207	0.234	0.030	0.016	0.162	0.106	0.190	0.044	0.174
	AUC	0.623	0.596	0.642	0.816	0.691	0.614	0.660	0.610	0.787	0.652
	-2[lnL_r - *lnL*]; *H0: β_z = 0, for all z_t*	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001

Significant at the 5% level in bold.

Table 6: Borrower-Specific Profile at Origination

A comparison of a selected set of borrower and loan attributes of two randomly selected loans from the set of loans that neither defaulted nor prepaid over the full sample period.

Loan	A	B
Loan origination date	Apr-03	Aug-03
Characteristics		
Loan type	Other	Other
Documentation	Full	Full
Loan purpose	Refinance	Refinance
FICO at origination	665	628
LTV at origination	80	53
Mortgage type	Hybrid ARM	Fixed Rate
Prepayment Penalty	Yes	No
Teaser Rate	Yes	No
Loan amount	over $120K	Under $75K
Interest Rate	7.88	7.39
Spread over 30-yr Fixed Mtg Rate	2.07	1.09

Table 7: Trend in the Off-Diagonal P_{ijs}

A comparison of the expected impact of changes in economic conditions on the off-diagonal transition probabilities, $E(\Delta p_{ij})$, to the actual changes in the p_{ij}, $i \neq j$, observed in Figure 4.

Figure	Trans Prob	Trend	
		2008-10	post 2010
i > J	$E(\Delta p_{ij})$	(-)	(+)
4.c	p_{21}	(-)	(F)
4.e	p_{32}	(-)	(+)
4.g	p_{43}	(F)	(+)
		2008-10	post 2010
i < J	$E(\Delta p_{ij})$	(+)	(-)
4.b	p_{12}	(+)	(A)
4.d	p_{23}	(+)	(-)
4.f	p_{34}	(-)	(-)

$E(\Delta p_{ij})$ – Expected trend in p_{ij}. (-) decreasing, (+) increasing
(A) – Ambiguous: too much variation to determine trend.
(F) – Flat: no trend.

Table 8: Age Distribution by Year

The average age of the loans in our sample increased significantly following the large decline in new loans added to the MBS Data pool of securitized mortgages beginning 2008

Year	Mean	Median	Mode
2004	20.7	13	6
2005	19.9	14	7
2006	21.8	17	9
2007	26.4	22	10
2008	37	33	22
2009	50	46	34
2010	62.9	59	58
2011	75.3	72	70

Table 9. Theil-U Statistics: 24-month out-of-time forecast

The 24-month ahead out-of-time forecast of the default and prepayment rates for both the unconditional and conditional transition matrices using both a 1st-order and 2nd-order Markov approach with and without including systamtic (i.e., macroeconomic) factors are compared to the actual default and prepayment rates for six subperiods using a one-year-step-forward sample design (i.e., dev: 2004m01-2006m12, fcst: 2007m01-2008m12; dev: 2004m01-2007m12, fcst: 2008m01-2009m12; …; dev: 2004m01-2011m12, fcst: 20012m01-2013m12).

Cohort/ Forecast Horizon	Development Sample	Transition Matrix		Deliqnency Status: current	30 dpd	60 dpd	90 dpd	prepay	default
200612									
2007-08	**2004-06**	**unconditional**		0.0152	0.1582	0.3329	0.4798	0.3277	0.4641
	2004-06	**conditional**	*1st -order Markov*						
			include macro var	0.0134	0.0998	0.1553	0.2654	0.2353	0.3495
			exclude macro var	0.0180	0.1042	0.2898	0.4598	0.3772	0.4981
			2nd -order Markov						
			include macro var	0.0234	0.1101	0.2038	0.3490	0.2313	0.5040
			exclude macro var	0.0182	0.1610	0.3879	0.5655	0.3706	0.6365
200712									
2008-09	**2005-07**	**unconditional**		0.0215	0.1753	0.3929	0.5232	1.0516	0.5343
	2004-07	**conditional**	*1st -order Markov*						
			include macro var	0.0370	0.3032	0.1034	0.1408	0.1061	0.0208
			exclude macro var	0.0144	0.0966	0.2390	0.3739	0.6852	0.4179
			2nd -order Markov						
			include macro var	0.0130	0.2596	0.1015	0.1519	0.0946	0.1002
			exclude macro var	0.0256	0.1001	0.2887	0.4255	0.6841	0.5146
200812									
2009-10	**2006-08**	**unconditional**		0.0436	0.1261	0.2230	0.2814	0.7183	0.2508
	2004-08	**conditional**	*1st -order Markov*						
			include macro var	0.0311	0.3749	0.1676	0.1609	0.0488	0.0500
			exclude macro var	0.0384	0.1778	0.1088	0.1591	0.2744	0.1056
			2nd -order Markov						
			include macro var	0.0196	0.3680	0.1786	0.1524	0.0505	0.1468
			exclude macro var	0.0170	0.1384	0.0879	0.1605	0.2811	0.1908
200912									
2010-11	**2007-09**	**unconditional**		0.1253	0.0812	0.0514	0.1518	0.6701	0.2655
	2004-09	**conditional**	*1st -order Markov*						
			include macro var	0.0802	0.1412	0.2518	0.3776	0.1829	0.2244
			exclude macro var	0.1160	0.1957	0.1943	0.2457	0.4758	0.1882
			2nd -order Markov						
			include macro var	0.0570	0.1184	0.2164	0.3590	0.1739	0.1307
			exclude macro var	0.0875	0.1577	0.1678	0.2307	0.4833	0.0517
201012									
2011-12	**2008-10**	**unconditional**		0.1285	0.0704	0.2990	0.6435	0.2721	0.6591
	2004-10	**conditional**	*1st -order Markov*						
			include macro var	0.0454	0.1919	0.3346	0.5083	0.1605	0.2544
			exclude macro var	0.1297	0.3033	0.4115	0.5820	0.4003	0.3472
			2nd -order Markov						
			include macro var	0.0157	0.1272	0.2271	0.3856	0.1685	0.0794
			exclude macro var	0.1340	0.1751	0.3089	0.5237	0.5079	0.3964
201112									
2012-13	**2009-11**	**unconditional**		0.0987	0.1280	0.4517	0.9169	0.0638	0.8450
	2004-11	**conditional**	*1st -order Markov*						
			include macro var	0.0168	0.1321	0.2173	0.2573	0.2706	0.1229
			exclude macro var	0.1174	0.3418	0.5184	0.7404	0.2307	0.5584
			2nd -order Markov						
			include macro var	0.0515	0.0546	0.0813	0.0791	0.2714	0.1528
			exclude macro var	0.0796	0.2384	0.3962	0.5583	0.2413	0.2444
201112									
2012-13	**2004-11**	**unconditional**		0.2346	0.1876	0.2292	0.4749	0.8914	0.6688

Figure 1: Distribution of the 2004Q1 Cohort 24-month forecasts using the MLE Transition Matrices [$P_{ij|q1:q2}$] based on the Full, Pre-crisis, Crisis, and Post-crisis Subperiods

The dark blue bars represent the number of active loans (n=72,193) in each state at the start of 2004Q1. The yellow bars represent the forecasted distribution of the 2004Q1 cohort 24-months ahead using eq [3], $Z_{t+24} = Z_t \mathrm{P}^{24}$ where $\mathrm{P} = P_{ij|2004:2011}$. Similarly, the red, orange, and light-blue bars represent the forecasted distribution of the 2004Q1 cohort 24-months ahead using eq 3 with $\mathrm{P} = P_{ij|2004:2007}$, $\mathrm{P} = P_{ij|2008:2010}$, and $\mathrm{P} = P_{ij|2011:2013}$ respectively.

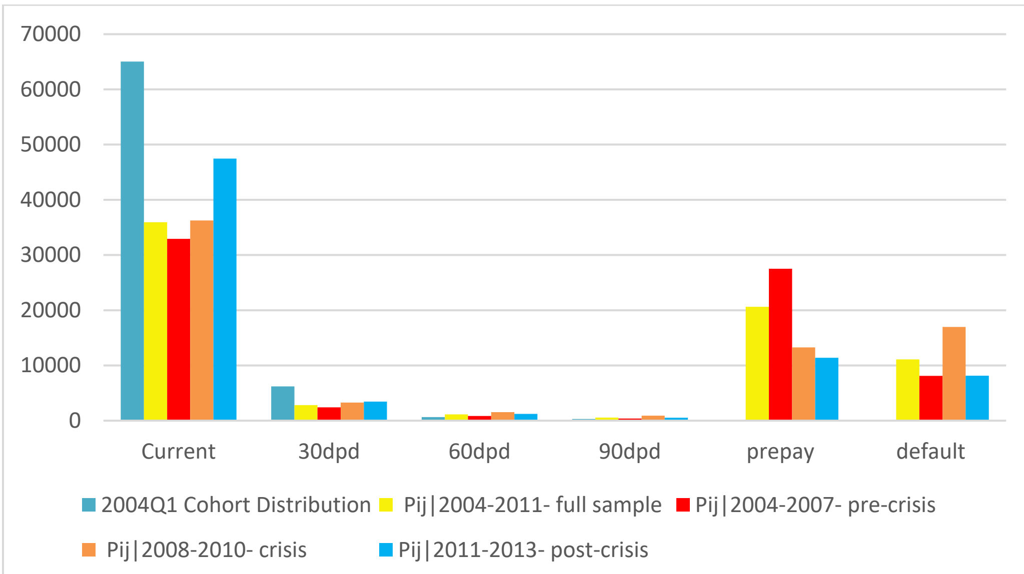
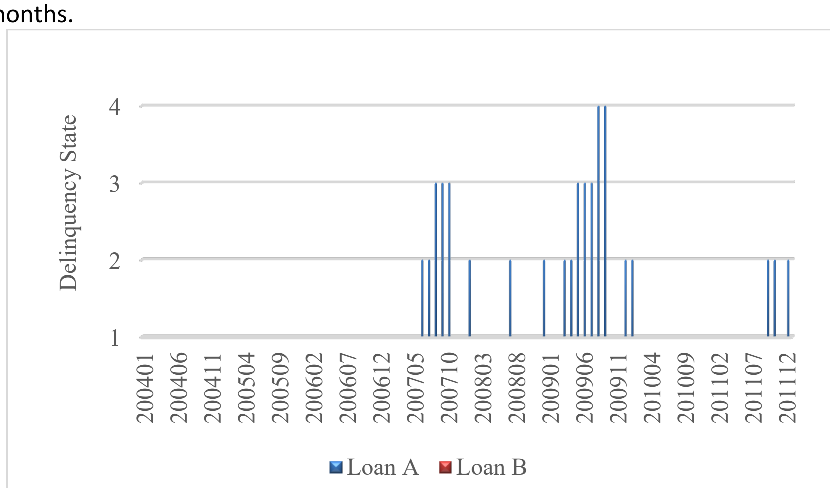


Figure 2a: Delinquency Status in Time t

A plot of the delinquency histories of Borrowers A and B. Borrower B remained current over the full sample period (January 2005 thru December 2011); Borrower A, however, was at least 30 dpd (delq status ≥ 2) in 20 of the 96 months and severely delinquent (delq status > 2) in 8 of the 96 months.



Figures 2b-2k: Plots of the Time Paths of Borrower-Specific Monthly Conditional Transition Probabilities (p_{ij}s)

The plots show the relative relationship between the time paths of the p_{ij}s of (1) the off-diagonal transition probabilities (i.e., the p_(i,j-1) and p_(i,j+1)), which reflect the transitions to a better or worse state, and (2) the transitions to default p_{id} for Borrowers A and B, where the loan to Borrower A is considered to be the more risky loan based on their risk profile outlined in Table 6). The MLE unconditional transition probabilities are included as baseline values for the p_{ij}s.

Figure 2b: P11(t) over time

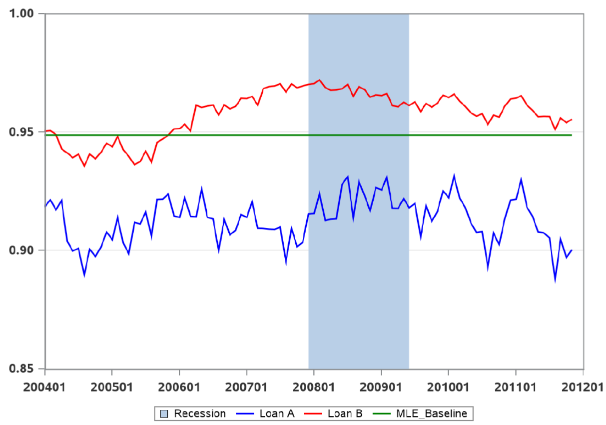


Figure 2c: P12(t) over time

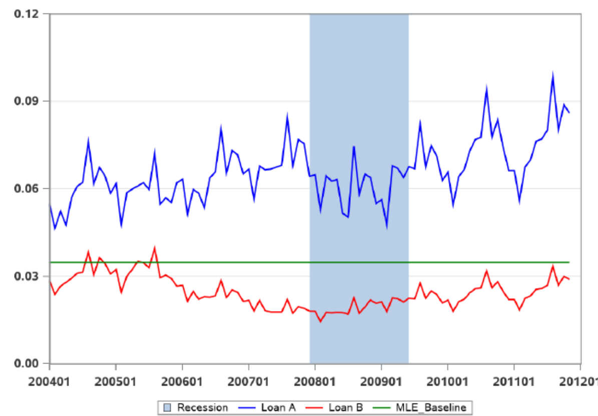


Figure 2d: P21(t) over time

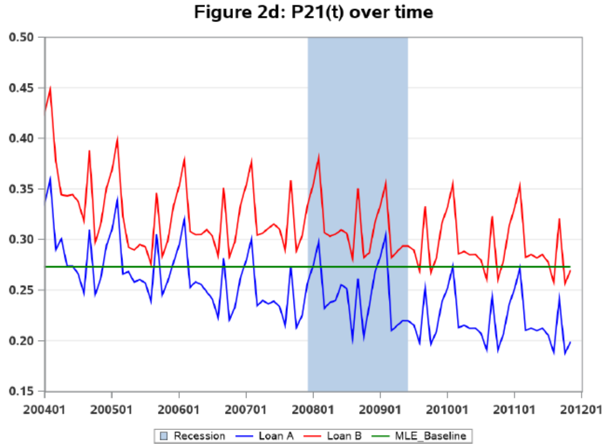


Figure 2e: P23(t) over time

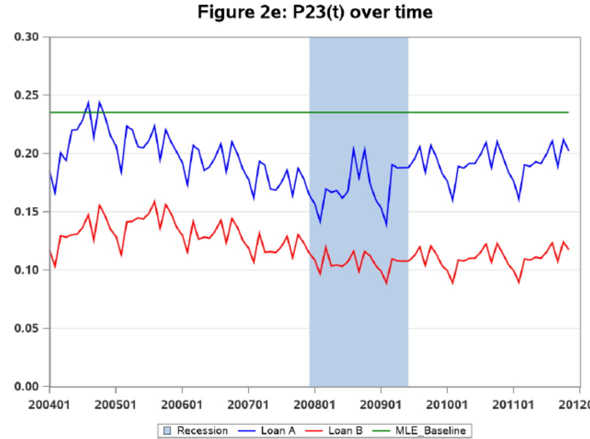


Figure 2f: P26(t) over time

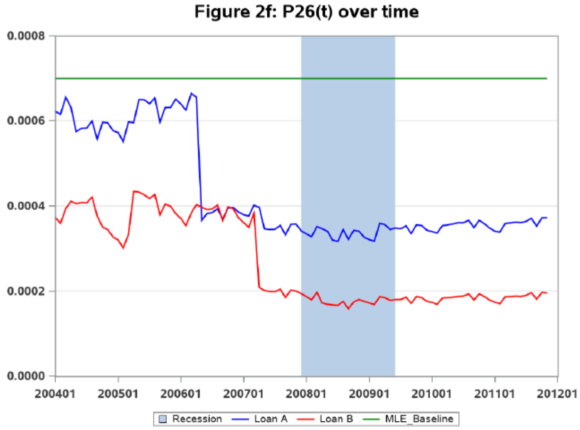


Figure 2g: P32(t) over time

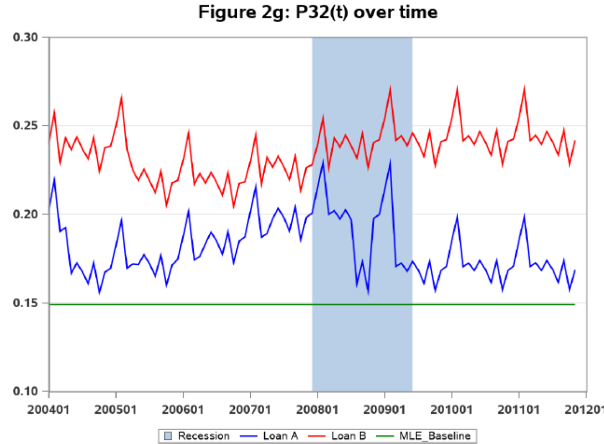


Figure 2h: P34(t) over time

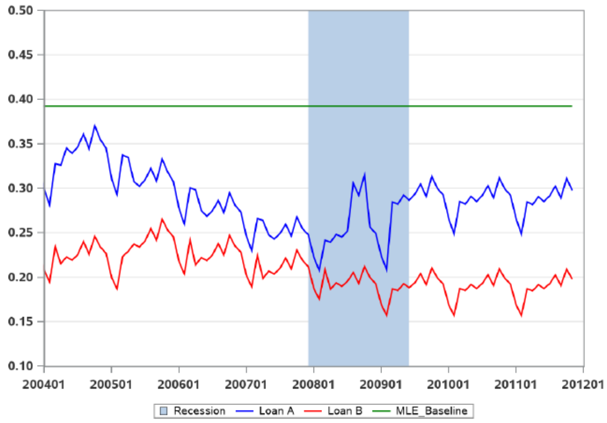


Figure 2i: P36(t) over time

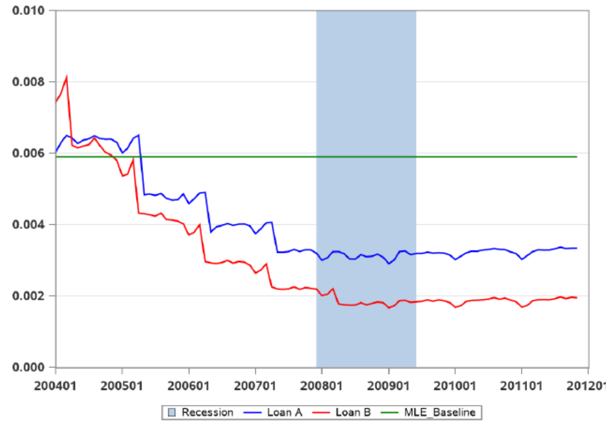


Figure 2j: P43(t) over time

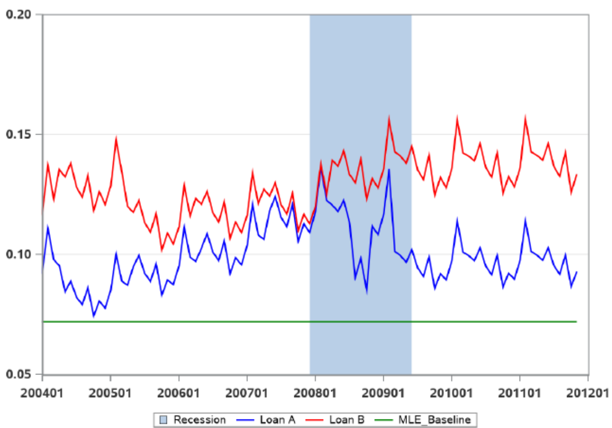
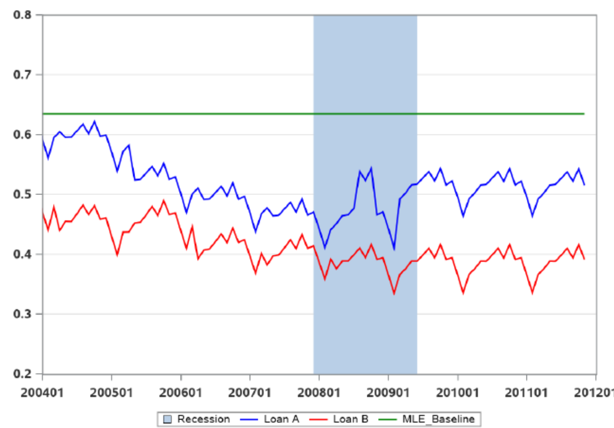


Figure 2k: P46(t) over time



Figures 3a-3d: Plots of Borrower-Specific Monthly Time Paths from each State i to Prepayment (p_{i5})

The plots show the relationship between the time paths of the probabilities of transitioning from each of the initial states to prepayment (p_{i5}; i=1,.2, …,4) for both Borrowers A and B. Borrower B is relatively more likely to transition from states 2, 3, and 4 (i.e., 30dpd, 60dpd, and 90dpd) in time t to prepay in time t+1 over all time periods than borrower A; the exception is the transition from state 1 (i.e., current) to prepayment, in which the results are mixed. The MLE unconditional transition probabilities are included as baseline values for the p_{i5}s.

Figure 3a: P15(t) over time

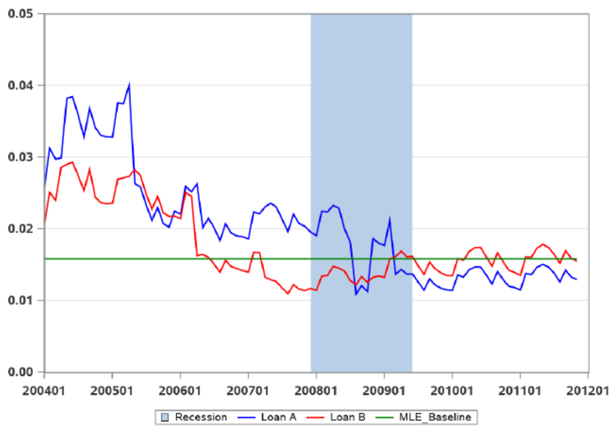


Figure 3b: P25(t) over time

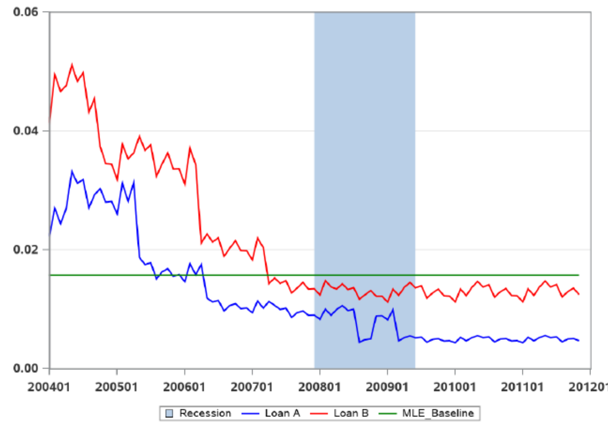


Figure 3c: P35(t) over time

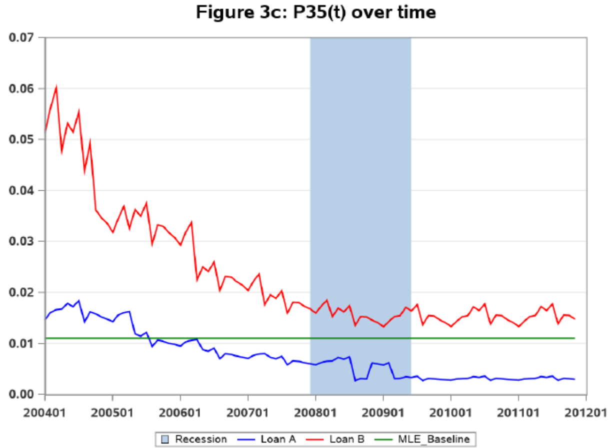
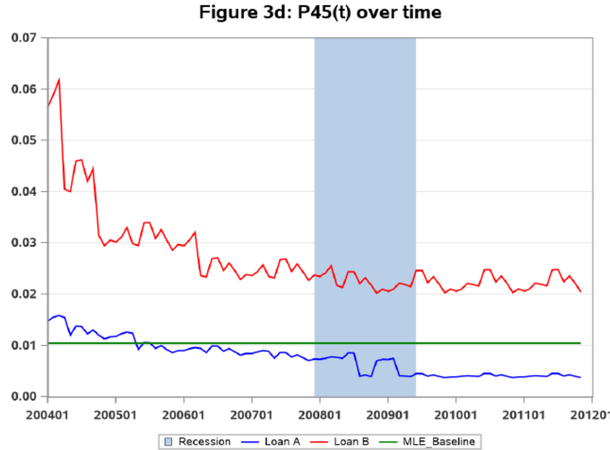


Figure 3d: P45(t) over time



Figures 4a-4g: Plots of the Time Paths in the Transition Probabilities from Higher to Lower Delinquency States (p_{ij}, i>j) and from Lower to Higher Delinquency States (p_{ij}, i<j)

The time paths of the p_{ij}s, averaged over the size of the sample in each time period, are reported in Figures 4a-4g. The p_{ij}s vary significantly over the cycle: during a decline in economic conditions, we expect (1) a decrease in the likelihood a delinquent account would transition to a lower delinquent state, and (2) an increase in the likelihood a delinquent account would transition to a higher delinquent state (i.e., we expect a downward (upward) time trend in the monthly time path of the p_{ij}(t) for all $i > j$ ($i < j$). The MLE unconditional transition probabilities are included as baseline values for the p_{ij}s.

Figure 4a: Monthly Transition Probabilities - P11

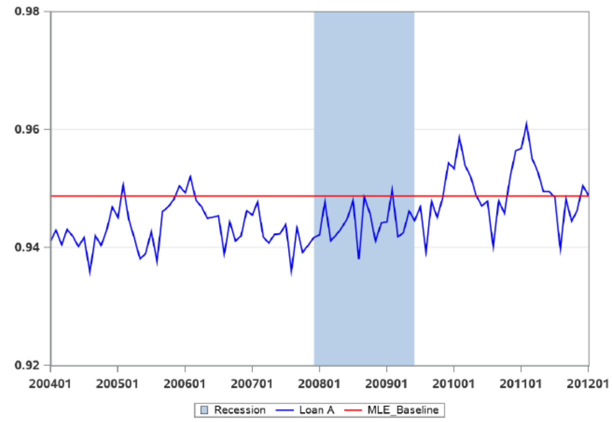


Figure 4b: Monthly Transition Probabilities - P12

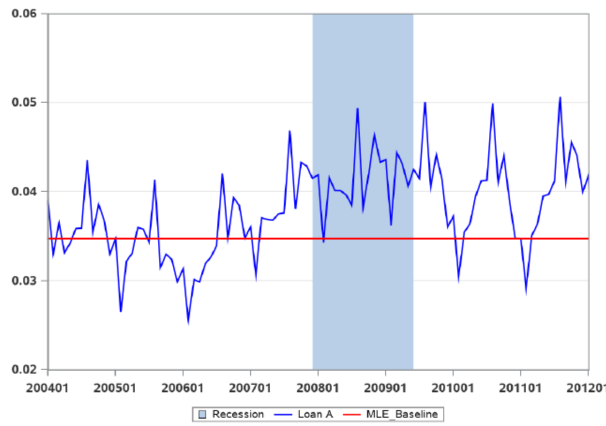


Figure 4c: Monthly Transition Probabilities - P21

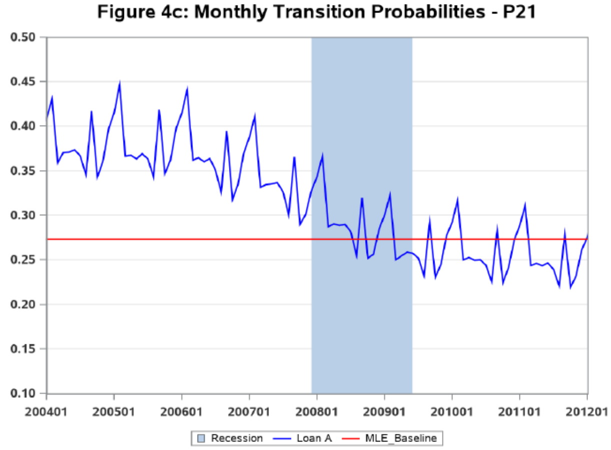


Figure 4d: Monthly Transition Probabilities - P23

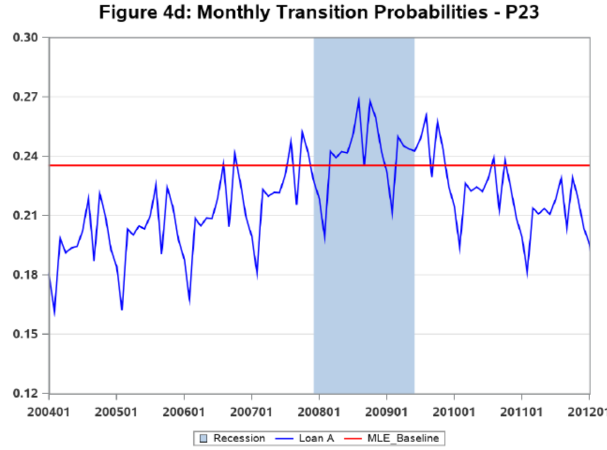


Figure 4e: Monthly Transition Probabilities - P32

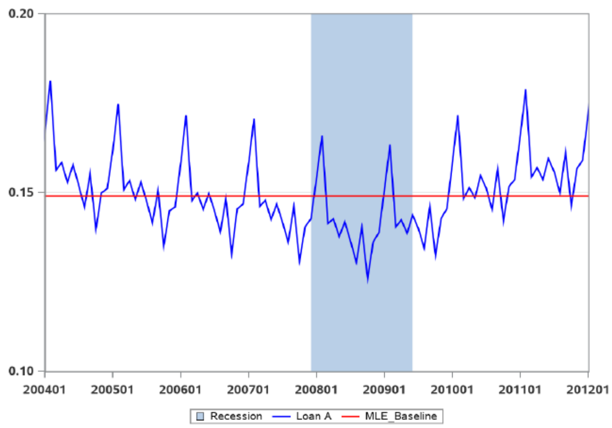


Figure 4f: Monthly Transition Probabilities - P34

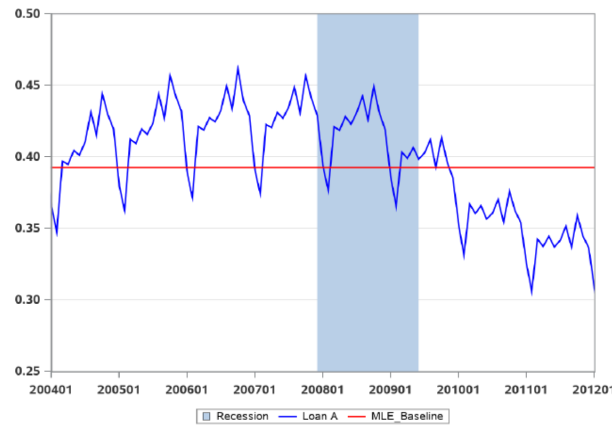
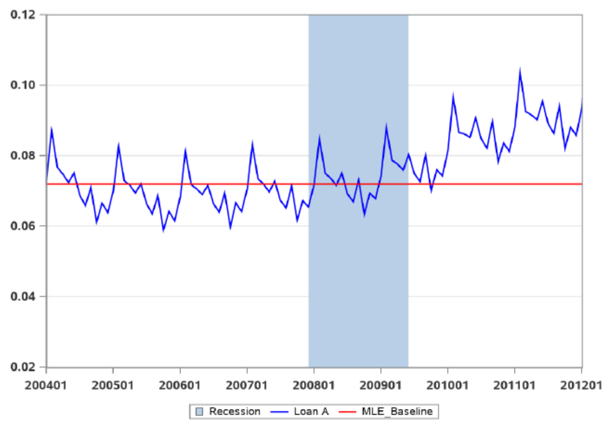


Figure 4g: Monthly Transition Probabilities - P43



Figures 5a-5c: Plots of the Time Paths in the Transition Probabilities from each Delinquency State i= 2, 3, and 4 to Default (p_{i6})

The time paths of the transition probabilities from each of the delinquency states to default, averaged over the size of the sample in each time period, are reported in Figure 5a-5c. There is a clear downward trend in the p_{i6}s beginning in 2007. Although the downward trend in the time paths to default during the post-crisis period is consistent with expectations that the likelihood of transitioning to default during a recovery should decline as the economy improves, the downward trend in the time path of the transition probabilities p_{i6} for i=2,..,4, during the crisis is counterintuitive. The MLE unconditional transition probabilities are included as baseline values for the p_{i6}s.

Figure 5a: Monthly Transition Probabilities - P26

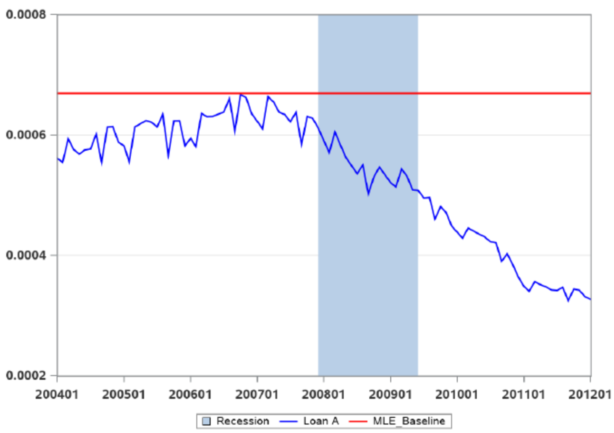


Figure 5b: Monthly Transition Probabilities - P36

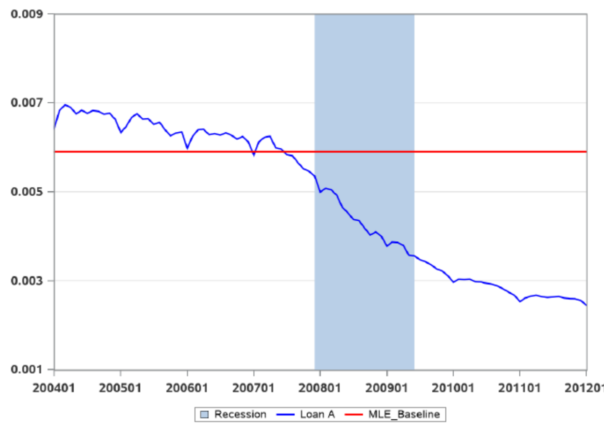


Figure 5c: Monthly Transition Probabilities - P46

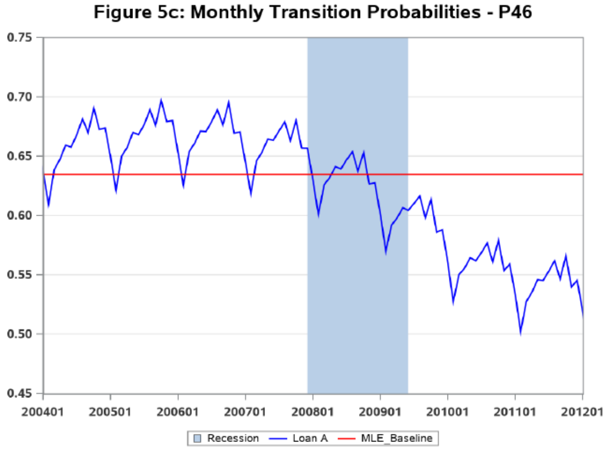


Figure 6: Out-of-Time Accuracy Plots – Conditional vs Unconditional

Out-of-time (forecasted) cumulative default and prepayment rates derived from both the conditional (eq 7) and unconditional (eq 3) transition matrices compared to the actual cumulative default and prepayment rates for loans active as of 201101. The p_{ij}s are estimated using the full sample (200401-201112); and the 24-month forecast period is 201201-201312.

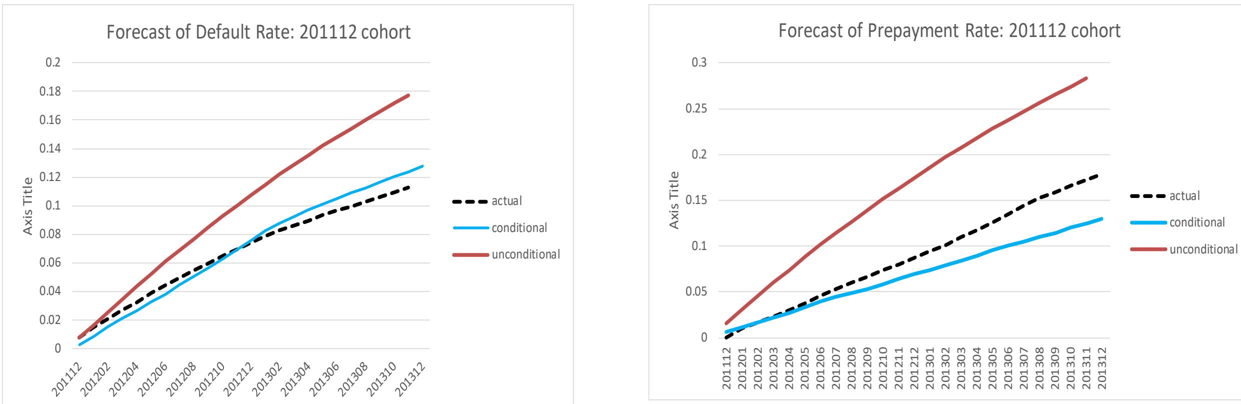


Figure 7: Alternative Out-of-Time Accuracy Plots – Conditional vs Unconditional

Out-of-time (forecasted) cumulative default and prepayment rates derived using the conditional (eq 7) and unconditional (eq 3) transition matrices compared to the actual cumulative default and prepayment rates for loans active as of 201101. The pijs for the conditional and first unconditional forecast are estimated using the full sample (200401-201112) as in Figure 6. The alternative unconditional p_{ij}s are derived using only data from 200901-201112. The 24-month forecast period is 201201-201312.

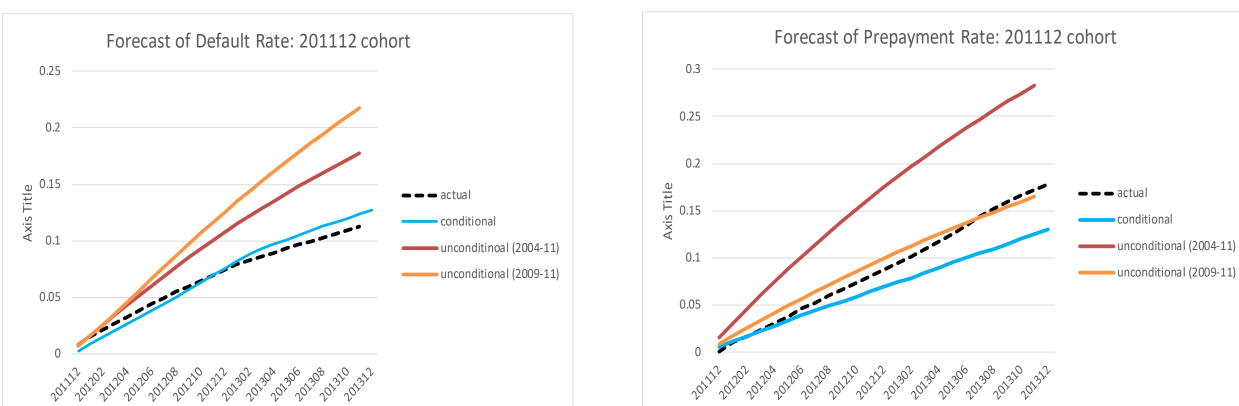
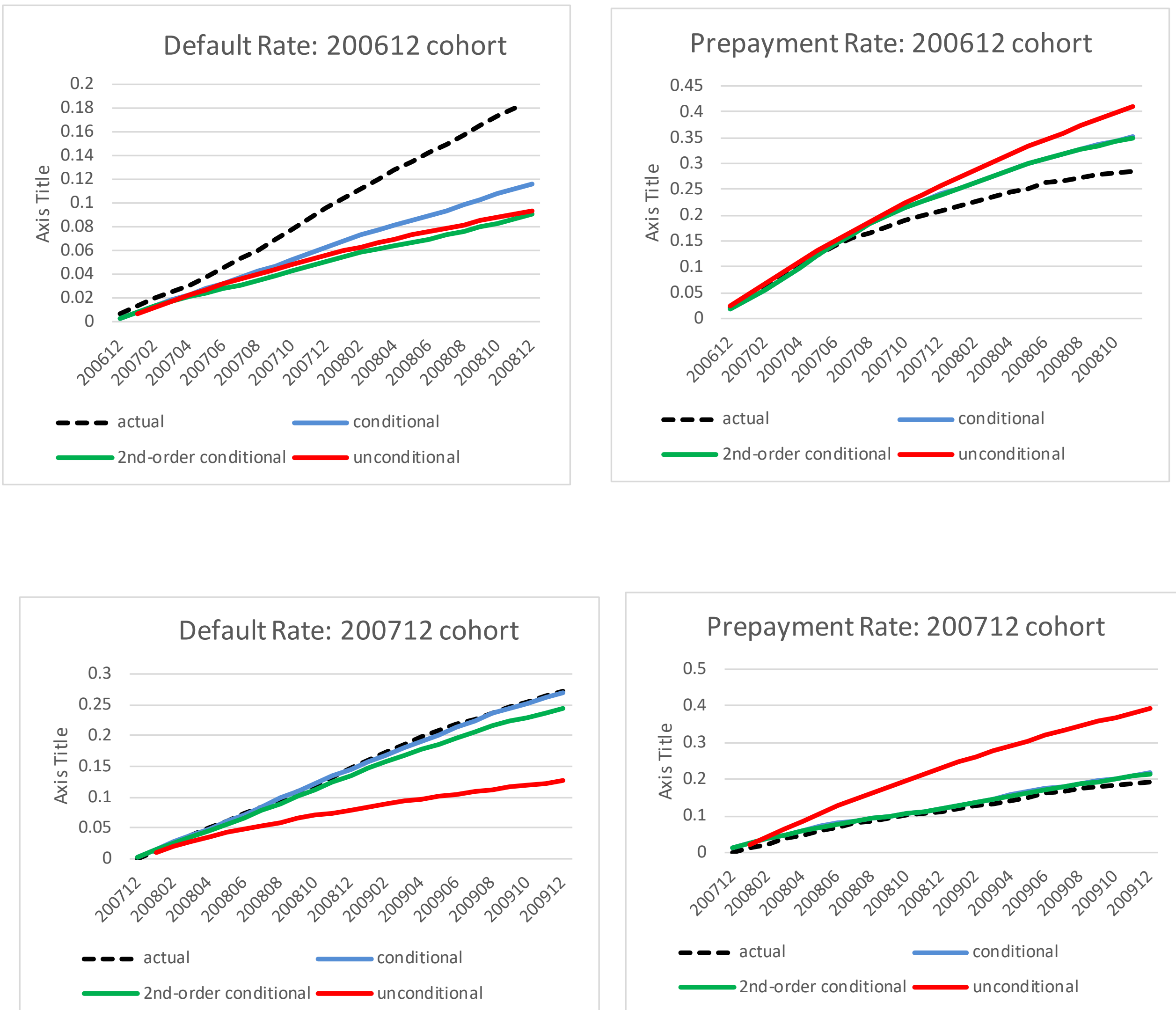
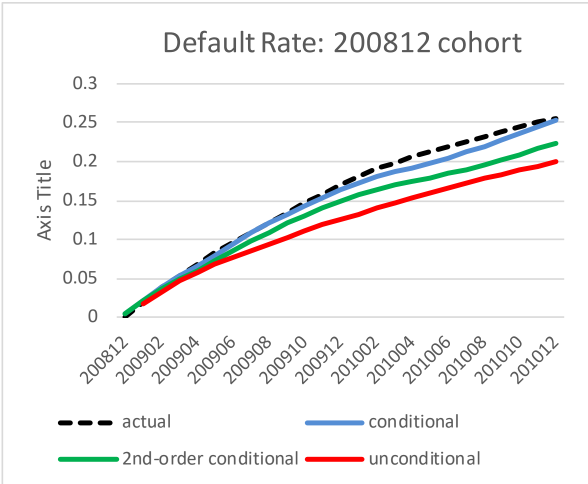
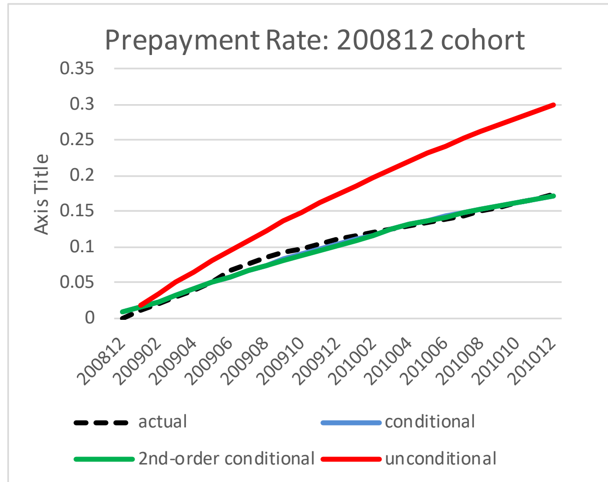


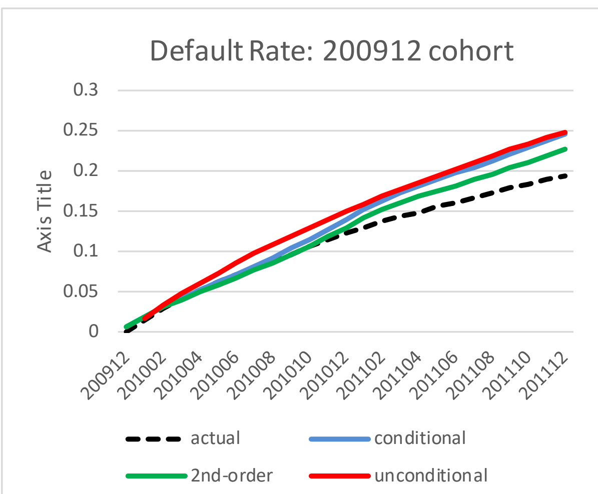
Figure 8: Out-of-Time Cohort-Based Accuracy Plots for Each of the Six Subsamples

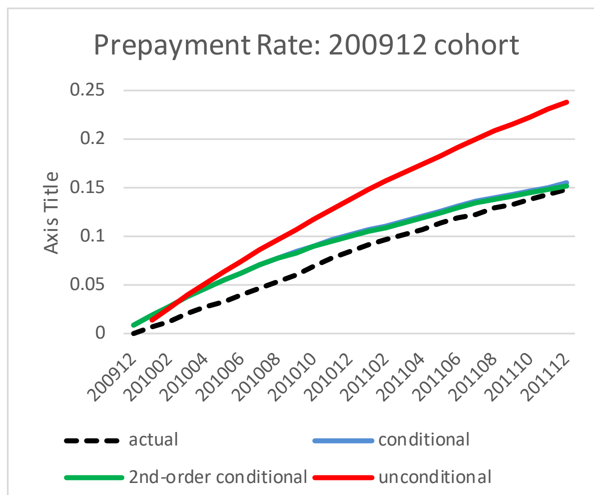
Out-of-time (forecasted) cumulative default and prepayment rates derived using the conditional (eq 7) and unconditional (eq 3) transition matrices compared to the actual cumulative default and prepayment rates for loans active as of 201101. We plot the 24-month ahead forecasts based on the 1st-order Markov conditional, 2nd-order Markov conditional, and the MLE unconditional transition matrices for each of the subperiods outlined in Table 9.

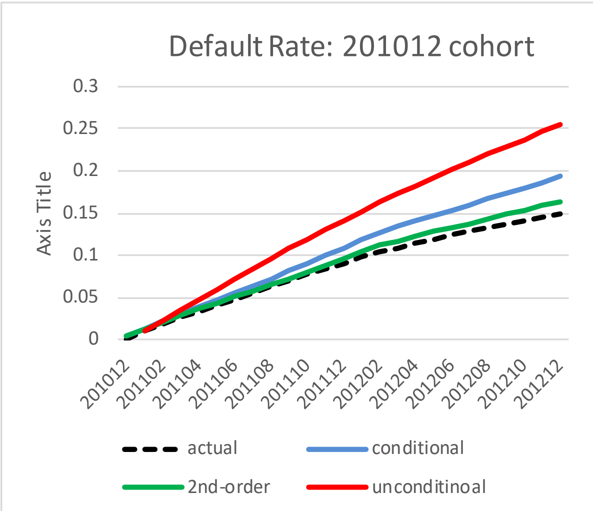


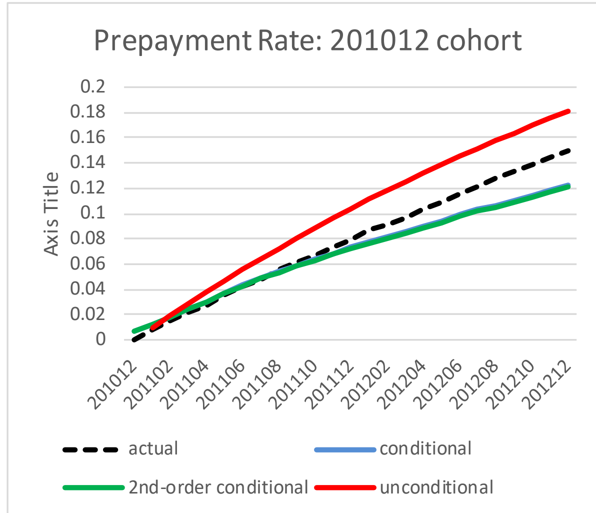

Default Rate: 200812 cohort
0.3
0.25
0.2
0.15
0.1
0.05
0
Axis Title
200812
200902
200904
200906
200908
200910
200912
201002
201004
201006
201008
201010
201012
actual
conditional
2nd-order conditional
unconditional

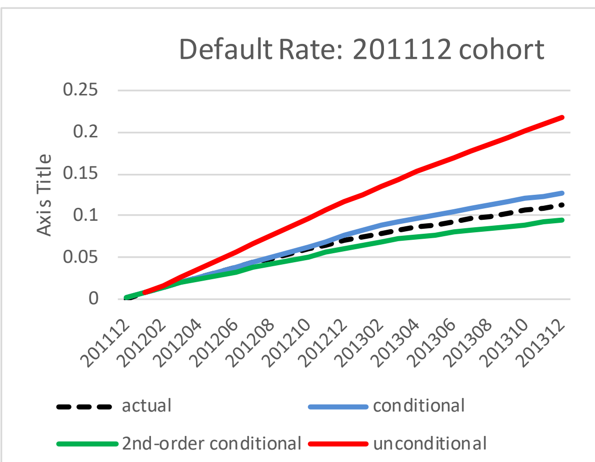

Prepayment Rate: 200812 cohort
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
Axis Title
200812
200902
200904
200906
200908
200910
200912
201002
201004
201006
201008
201010
201012
actual
conditional
2nd-order conditional
unconditional

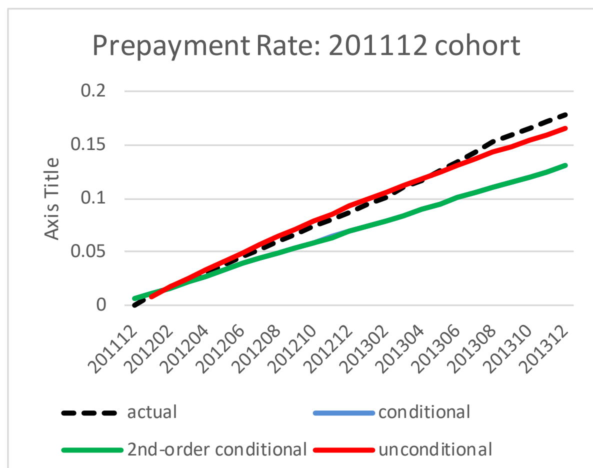

Default Rate: 200912 cohort
0.3
0.25
0.2
0.15
0.1
0.05
0
Axis Title
200912
201002
201004
201006
201008
201010
201012
201102
201104
201106
201108
201110
201112
actual
conditional
2nd-order
unconditional


Prepayment Rate: 200912 cohort
0.25
0.2
0.15
0.1
0.05
0
Axis Title
200912
201002
201004
201006
201008
201010
201012
201102
201104
201106
201108
201110
201112
actual
conditional
2nd-order conditional
unconditional


Default Rate: 201012 cohort
Axis Title
0.3
0.25
0.2
0.15
0.1
0.05
0
201012
201102
201104
201106
201108
201110
201112
201202
201204
201206
201208
201210
201212
actual
conditional
2nd-order
unconditinoal


Prepayment Rate: 201012 cohort
Axis Title
0.2
0.18
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
201012
201102
201104
201106
201108
201110
201112
201202
201204
201206
201208
201210
201212
actual
conditional
2nd-order conditional
unconditional


Default Rate: 201112 cohort
Axis Title
0.25
0.2
0.15
0.1
0.05
0
201112
201202
201204
201206
201208
201210
201212
201302
201304
201306
201308
201310
201312
actual
conditional
2nd-order conditional
unconditional


Prepayment Rate: 201112 cohort
Axis Title
0.2
0.15
0.1
0.05
0
201112
201202
201204
201206
201208
201210
201212
201302
201304
201306
201308
201310
201312
actual
conditional
2nd-order conditional
unconditional

VII. Reference

Allison, Paul D. (1982). "Discrete-time methods for the analysis of event histories" in *Sociological Methodology,* Samuel Leinhardt (ed.). San Francisco: Jossey-Bass, 61-98.

Allison, Paul D. (1995). Survival Analysis Using the SAS System: A Practical Guide. Cary, NC: SAS Institute Inc.

An, M.Y. and Z. Qi (2012). "Competing Risks Models using Mortgage Duration Data under the Proportional Hazards Assumption," *Journal of Real Estate Research*, 34, 1-26.

Anderson, T.W. and L.A. Goodman (1957). "Statistical Inference About Markov Chains," *The Annals of Mathematical Statistics*, 28(1), pp. 89-110

Andersson, F., Chomsisengphet, S., Glennon, D., and Li, F. (2013). "The Changing Pecking Order of Consumer Defaults," *Journal of Money, Credit, and Banking*, 45, 251-275.

Bellotti,Tony and Jonathan Crook (2013a). "Forecasting and Stress Testing Credit Card Default Using Dynamic Models," *International Journal of Forecasting*, 29, 563-574.

Bellotti,Tony and Jonathan Crook (2013b). "Retail Credit Stress Testing Using a Discrete Hazard Model with Macroeconomic Factors," *Journal of Operational Research Society*, 65, 340-350.

Begg, Colin and Robert Gray (1984). "Calculation of Polychotomous Logistic Regression Parameters Using Individualized Regressions," *Biometrika*, 71, 11-18.

Betancourt, Luis (1999). "Using Markov Chains to Estimate Losses from a Portfolio of Mortgages," *Review of Quantitative Finance and Accounting,* 12(3), 303-317.

Calhoun, Charles A. and Yongheng Deng. (2002) "A Dynamic Analysis of Fixed- and Adjustable-Rate Mortgage Terminations," *The Journal of Real Estate Finance and Economics*, 9-33.

Campbell, T.X. and K. Dietrich (1983). "The Determinant of Default on Insured Conventional Residential Mortgage Loans," *Journal of Finance*, 38, 1569-81.

Crook, J. and Banasik, J. (2012) 'Forecasting and explaining aggregate consumer credit delinquency behaviour', *International Journal of Forecasting*, 28, 145-160.

Crouhy, Michel, Robert Jarrow, and Stuart Turnbull (2008). "The Subprime Credit Crisis of 07," *The Journal of Derivatives,*16, 81-110.

Cyert, R.M., H.J. Davidson, and G.L. Thompson (1962) "Estimation of the Allowance for Doubtful Accounts by Markov Chains," *Management Science,* 8(3), 3-19.

Demyanyk, Y. and Van Hemert, O. (2011) 'Understanding the Subprime Mortgage Crisis', *The Review of Financial Studies*, 24 (6), 1847-1880.

Glennon, D. and P. Nigro (2005). "Measuring the Default Risk of Small Business Loans: A Survival Analysis Approach," *Journal of Money, Credit, and Banking*, 37(5), 923-947.

Goodman, L.S. (2016). "The Rebirth of Securitization: Where is the Private-Label Mortgage Market," *Journal of Structure Finance,* 22(1), pp. 8-19.

Grimshaw, Scott D. and William P. Alexander (2011). "Markov Chain Models for Delinquency: Transition Matrix Estimation and Forecasting," *Applied Stochastic Models in Business and Industry*, 27, 267-279.

IFE Group (2007). "An Actuarial Review of the Federal Housing Administration Mutual Mortgage Insurance Fund for Fiscal Year 2007." Report to the U.S. Department of Housing and Urban Development.

Kallberg, Jarl G., and Anthony Saunders, 1983, Markov Chain Approaches to the Analysis of Payment Behavior of Retail Credit Customers, *Financial Management*, Summer, 5-14

Koudinov, Youriy, Rohit Jadhav, and Olgay Cangur (2019) "First-Quarter 2019 RMBS Quarterly Update – High Prepayments Haven't Tarnished the Credit Quality of Non-QM RMBS." www.morningstarcreditratings.com.

Lee, T.C., G.G. Judge, and A. Zellner (1977) *Estimating the Parameters of the Markov Probability Model from Aggregate Time Series Data,* 2nd edition, North Holland Publication, Amsterdam.

Leow, Mindy and Jonathan Crook (2014). "Intensity Models and Transition Probabilities for Credit Card Loan Delinquencies," *European Journal of Operational Research*, 236, 685-694.

Li, Mingxin (2014). "Residential Mortgage Probability of Default Models and Methods," Research Paper, Financial Institutions Commission, British Columbia.

Malik Madhur, Lyn C. Thomas (2012). "Transition Matrix Models of Consumer Credit Ratings," *International Journal of Forecasting,* 28, 261-272.

Mayer, Christopher, Karen Pence, and Shane M. Sherlund (2009). "The Rise in Mortgage Defaults." *Journal of Economic Perspectives*, 23, 27-50.

Molina Utrilla, Jose and Nick Constantinou (2010). "Could the Trigger to the Subprime Crisis Have Been Predicted? A Mortgage Risk Modeling Approach." Available at SSRN: https://ssrn.com/abstract=1616697 or http://dx.doi.org/10.2139/ssrn.1616697

Sarmiento, C. (2012) 'The role of the economic environment on mortgage defaults during the Great Recession," *Applied Financial Economics*, 22, 243-250.

Smith, L Douglas and Edward Lawrence (1995). "Forecasting Losses on a Liquidating Long-Term Loan Portfolio," *Journal of Banking and Finance*, 19, 959-985.

Smith, L. Douglas, Susan Sanchez, and Edward Lawrence (1996). "A Comprehensive Model for Managing Credit Risk on Home Mortgage Portfolios," *Decision Sciences,* 27, 291-317.

Thomas, Lyn (2009) *Consumer Credit Models: Pricing, Profit, and Portfolio*, Oxford University Press, New York.

Wang, Yikai, Min Ding, Jiren Pan, Samuel Malone (2017). "Credit Transition Model 2017 Update: Methodology and Performance Review," *Moody's Analytics*.