Moody's Research Labs

MORTGAGE PORTFOLIO ANALYZER:

A QUASI-STRUCTURAL MODEL OF MORTGAGE PORTFOLIO LOSSES

TECHNICAL DOCUMENT[1]

Mar 4, 2011

Roger M. Stein
Ashish Das
Yufeng Ding
Shirish Chinchalkar

ABSTRACT

This document outlines the underlying research, model characteristics, data, and validation results for Mortgage Portfolio Analyzer, which is an analytic tool to assess credit risk measures, capital levels and stress scenarios for portfolios of residential mortgages. Mortgage Portfolio Analyzer comprises loan-level econometric models for default, prepayment, and severity. These models are integrated through common dependence on local macro-economic factors, which can be either simulated at national, state, and Metropolitan Statistical Area (MSA) levels or input in the form of stress scenarios. This integration produces correlation in behaviors of loans across the portfolio. The simulation incorporates a multi-step Monte Carlo approach and generates monthly P&I cash flows and losses which enables the model to be used for ALM applications or to be combined with an external cash flow waterfall tool and used for simulation of RMBS transactions. Scenario and stress testing is also done in a multi-period framework. Furthermore, the model accommodates both loan-level and portfolio-level mortgage insurance. The resulting tool can be used for analyzing the credit risk in both portfolios of whole loans and RMBS transactions.

---

[1] Many past and present members of the current research group made significant contributions to the building and implementation of the models underpinning the Mortgage Portfolio Analyzer. Xufeng (Norah) Qian and Weijian Liang helped build the prepayment models. Weijian Liang also helped estimating the severity model. Jordan Mann helped model MI. Pouyan Mashayekh helped model the macro economic factors. Samuel Ring, together with Grigoriy Enenberg and Tamara Lubomirsky helped in the validation exercises. Xufeng (Norah) Qian, Jipil Ha, and Aziz Lookman helped in drafting some sections of this paper. We are grateful for their contribution and for the comments of Navneet Agarwal, Jordan Mann, Albert Metz, Rajesh Shekhar, and numerous members of the Moody's Investors Service RMBS Surveillance team.

TABLE OF CONTENTS

1 INTRODUCTION

Mortgage Portfolio Analyzer (MPA) is an analytic tool to help measure and manage the credit risk of portfolios of whole mortgage loans as well as collateral pools of underlying RMBS transactions. This document describes the development of the models that form the analytic core of MPA, including the data used, the modeling techniques, the key challenges in building these models, and our efforts to address them. It also documents the model architecture, how the model may be used in practice and its capabilities.

MPA has been designed to provide insight and transparency into the measurement and management of credit risks of portfolios of U.S. residential mortgage loans. It has applications to risk-management, stress testing, portfolio-construction, and capital allocation. MPA provides a common and coherent framework for the analysis of whole-loan portfolios or, when integrated with a waterfall tool, of RMBS transactions. This single framework also accommodates both the analysis of newly originated portfolios and the monitoring of seasoned loan pools. MPA takes advantage of supplemental data that users may have including, loan- and portfolio-performance to date, loan status, mortgage insurance terms (at the loan- and portfolio-level) as well as other information.

Mortgage Portfolio Analyzer analyses mortgage portfolios in four steps. First, it generates trajectories of economic scenarios at a quarterly frequency over the specified horizon. Next, for each loan in the portfolio, the loan-level models calculate monthly default and prepayment probabilities over the target horizon as a function of loan-specific and economy-wide factors.[2] Given these probabilities, the software then simulates default events, prepayment events, and loss given default and aggregates the simulated losses across all loans in the portfolio for each trajectory. Finally, these simulated losses are themselves aggregated across all trajectories to produce an estimate of the distribution of portfolio-level losses. Historical economic data used for the simulations are updated quarterly. Users may also use pre-defined macro-economic scenarios or stress cases in place of simulation.[3]

[2] MPA uses a thirty-year horizon, by default.

[3] Users may construct their own macro-economic forecasts or stress scenarios or use forecasts produced by Moody's Analytics.

Mortgage Portfolio Analyzer's capabilities include:

1. **Estimating the impact of economic stresses on loan performance.** The loan-level models explicitly capture the relationships among loan characteristics and economic states of the world. These relationships can be reported in detail for any simulation. As a result, model output facilitates an understanding of the impact of key economic factors on a portfolio-specific basis. This is particularly useful since the same economic factor may have very different effects in different portfolios due to the interplay between key determinants of loan behavior.

2. **Modeling layered risks and heterogeneous portfolios.** In evaluating a mortgage portfolio, it is typically necessary to examine layers of risk (the combined effects of various factors, including the FICO scores, loan age, *CLTV*). The integrated structure of the models allows a user to examine the impact of simultaneously changing multiple characteristics, as well as the impact of factors that have competing effects on loan losses (e.g., on prepayment vs. default).

3. **Implementing multi-step analysis.** Because the simulation engine produces economic scenarios for each period between origination and the simulation horizon, a richer representation of mortgage behavior and risk is possible. For example, since the prepayment, default and recovery behavior of mortgages is time dependent, dips in home prices that occur in an early period, may have very different implications than those that occur in later periods. A multi-step approach can differentiate between such cases.

4. **Modeling mortgage insurance (MI).** MPA models the impact of loan-level (primary) and pool-level mortgage insurance on loss given default (LGD). Since LGD is stochastic and since mortgage policies may be written differently for different loans, the impact of MI is more richly captured than might be in the case of a model that assumed a constant haircut to LGD. Furthermore, the impact of MI on a particular loan may also vary significantly across economic states. By modeling MI as a feature of the individual loan, MPA more naturally reflects both the contingent (on the economic state) behavior of the insurance policy and the variability associated with each policy. This also captures the interaction between primary and pool MI for a given portfolio.

5. **Providing a framework that may be integrated with other tools for RMBS analysis, pricing or ALM.** The model simulates the losses and payments for each loan in each period of the simulation and produces cashflows that reflect these behaviors. Thus, the model can be integrated with cashflow waterfall tools to provide a detailed assessment of loss timing, interest cashflows and amortization payments. This is important for synchronizing loss behaviors across assets and liabilities.

6. **Providing increased transparency of mortgage portfolio risk**. The models offer risk managers, lenders, portfolio-managers, investors, securities issuers and intermediaries detailed insight into the specific economic factors and loan characteristics that affect a portfolio's credit risk profile. Users can calculate Value-at-Risk (VaR) and perform risk factor attribution including evaluating which loans are contributing most to the capital requirements for the portfolio. This can be used to determine hedging or diversification strategies. Users can also evaluate portfolio losses based on custom scenarios so that they may conduct stress testing under scenarios of particular interest for their applications.

Mortgage Portfolio Analyzer is a multi-step simulation tool that supports the credit risk measurement and management of whole loan mortgages portfolios and collateral pools underlying RMBS transactions.

1.1 Some key research findings

- Modeling default, prepayment, and severity processes at the loan-level (as opposed to the pool-level) significantly improves accuracy in estimating losses, particularly for portfolios with heterogeneous mortgage assets.[4]

- Modeling each loan behavior separately (i.e., default, severity and prepayment as well as mortgage insurance terms) provides far greater flexibility in calibration than using a single joint model. Although prepayment, default and severity are distinct processes, the modular approach uses common factor dependencies that permit it to capture the natural interrelationships among the processes.

- Prepayment can have a dominant effect in determining the distribution of losses during periods of home price appreciation and/or falling interest rates. While default and

[4] Interested readers can find a more detailed discussion of this point in Chinchalkar and Stein (2010).

severity are key processes in loan behavior, without adequately capturing prepayment, losses are difficult to understand over long cycles.

- In addition to loan-specific characteristics, the state of the local and national economy significantly impacts the performance of pools, with the key macro variables being levels of home prices, unemployment rates, and interest rates.
- Default, prepayment, and severity appear to be correlated through their joint dependence on common economic factors as well as through the timing of payment changes across similar loan types.
- The multi-step approach to simulation offers advantages when assets have time dependent behavior, as in the case with mortgages.

Mortgage Portfolio Analyzer is intended to help users better understand the risks of residential mortgage portfolios. It should be used as a supplement to rather than a substitute for a rigorous analytic process.

The remainder of this paper proceeds as follows: Section 2 provides an overview of the various econometric models that make up Mortgage Portfolio Analyzer. Section 3 describes the data and discusses the data collection and quality process. Section 4 discusses some of the validation methods used to test the models. Finally, Section 5 summarizes our results and provides some concluding thoughts on the models.

2 MODEL COMPONENTS

In this section, we describe the loan-level models for default, prepayment, and severity and the simulation framework used to integrate the output of these models in order to estimate the loss distribution for a mortgage portfolio or a pool of collateral underlying an RMBS transaction. Section 2.1 discusses the model framework in non-technical terms, and Sections 2.2 through 2.6 provide some general background on the modeling techniques. Sections 2.7 through 2.10 provide technical details on each of the component models. Sections 2.11 through 2.13 discuss additional modeling considerations.

2.1 A quasi-structural model of portfolio losses

To adequately estimate the loss behavior of mortgage pools, it is useful to model the processes that affect losses (default, prepayment, and severity) *at the loan level*. Many bankers and market participants intuitively contemplate default and severity when characterizing loan portfolio behavior. However, the prepayment behavior of borrowers also has a first-order impact on pool losses. This is because a loan with a lower prepayment probability is likely to remain outstanding for a longer period of time providing many more opportunities for it to default. That is, for a given level of default intensity, higher prepayment rates result in lower losses over the lifetime of the portfolio. Hence, MPA models prepayment as a risk that competes with default risk.

These processes are, often, naturally inter-related. For example, a drop in home prices can

- *increase* severity (lower recovery expected when the home value declines leading to lower liquidation prices in the foreclosure process);
- *decrease* prepayment likelihood (it is less likely that the borrower has enough remaining home equity to refinance and the borrower's incentive to do so is reduced); and
- *increase* default likelihood (home owners without equity in their homes are less likely to see the economic value in continuing to maintain their mortgages).

MPA incorporates these inter-relationships by using a common set of macro-economic factors (in addition to loan-specific characteristics) as inputs to the models of the default, prepayment, and severity processes. The framework makes use of macroeconomic factors at the national, state or Metropolitan Statistical Area (MSA) level.[5] This greater transparency of the drivers of portfolio losses and also results in straightforward mechanism for conducting macro-economic scenario analysis and provides more flexibility in modeling new types of collateral. It also offers an implicit way of modeling correlation in loan performance across the loans in a portfolio.

MPA combines information about the state of the economy with loan-specific information in order to assess the credit risk of an individual mortgage. It is the *interaction* of macro and loan-level factors that determines the performance of the loan. For example, lower interest rates (a macro factor) tend to lead to higher probabilities of prepayment. However, even when interest rates are low, loans for which the mortgage coupon (a loan-level factor) is high *relative to* the prevailing market rate are even more likely to prepay. Therefore, the *level* of interest rates alone is not a sufficient predictor of prepayment. Rather it is the level of interest rates relative to the borrower's current coupon rate that is required. Accordingly, MPA uses measures of *mortgage premium* that compare interest rates to coupon rates on a loan-by-loan basis rather than constructing a common measure based on interest rates alone.

---

[5] MPA's models currently cover 366 MSAs. For loans that do not fall into one of the 366 MSAs, MPA uses state-level information.

The default, prepayment, and severity models are combined using a simulation engine that generates scenarios of economic states of the world for the three broad sets of macro-economic factors. In addition, user specified macro-economic scenarios may be used as input to MPA.

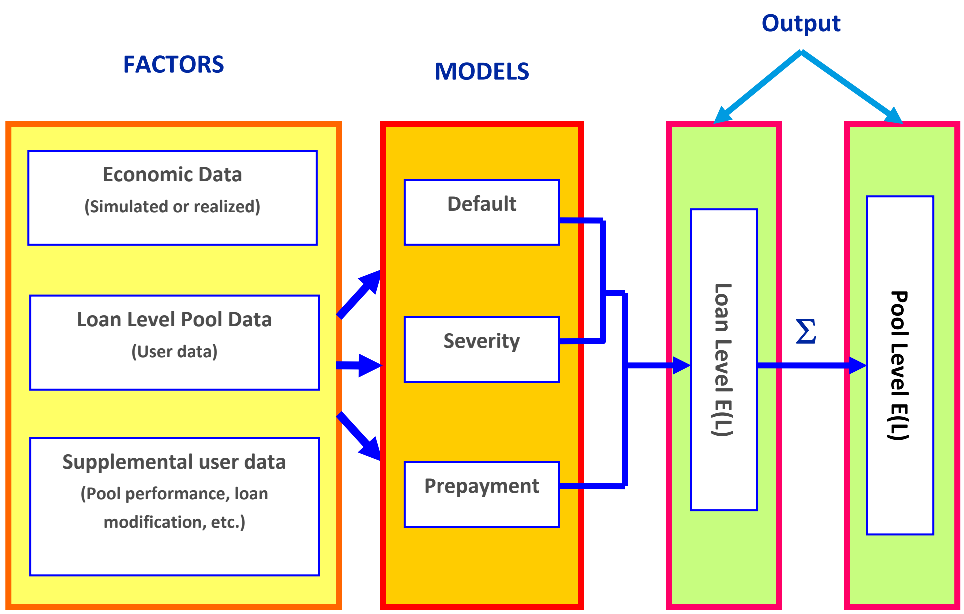
The three categories of macro-economic factors used by MPA are:

- National, state-, and MSA- level **home price changes;**
- National, state-, and MSA- level **unemployment rates**; and
- **Interest rates**, including the term structure of Treasury rates, the 6 month LIBOR rate, the Survey of Freddie Mac's Prime Mortgage 30-Year Fixed Rate and median subprime market rate.

2.2 Summary of the framework

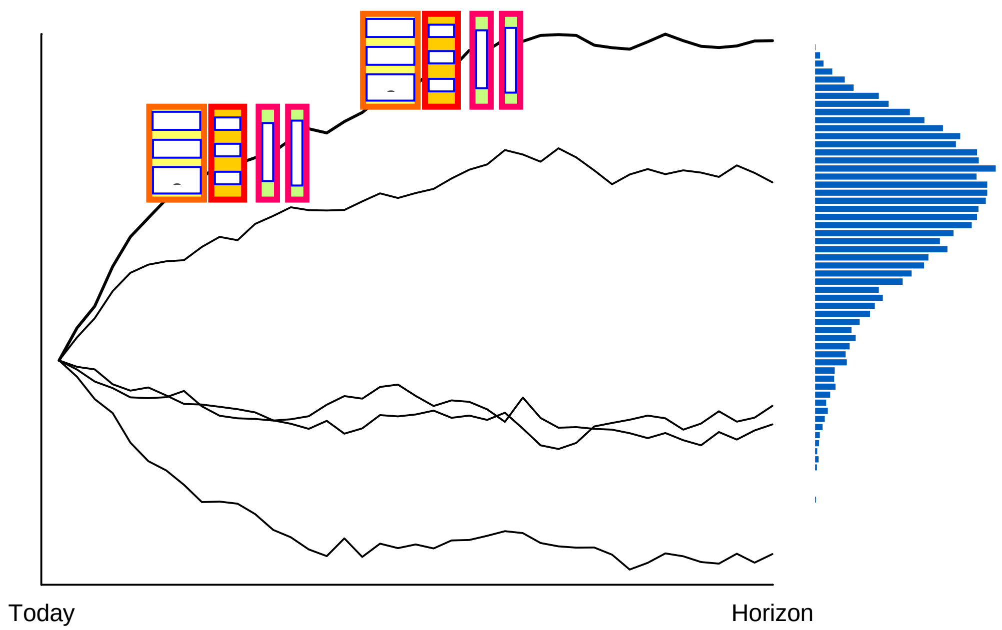
Mortgage Portfolio Analyzer is composed of a series of loan-level econometric models that are related through common dependence on macroeconomic as well as loan-specific factors. The macro factors used in the loan-level models are simulated at the national-level, state-level and MSA-level using econometric models developed for these factors. Figure 1, below, summarizes the process for a single period, schematically.

Figure 1: Interaction of econometric models



This computation is then repeated until the end of each path is reached. Using these loan-level models, MPA estimates the cash flows and losses for each loan given a realization of the economic factors in each period. Losses for this realization of economic factors are then aggregated across loans in the portfolio to produce an estimate of the portfolio credit loss and monthly cash flows under that specific economic trajectory. Portfolio losses are aggregated across the simulations to estimate the expected lifetime loss and the distribution of losses under all trajectories for the mortgage portfolio, as shown in Figure 2.

Figure 2: Calculation of losses based on simulation of different trajectories of economic factors



In the current implementation, by default, the simulation is conducted over a 30-year horizon at a monthly frequency using 10,000 equally-weighted simulated economic paths to produce default rates, prepayment rates and severity for each loan in the portfolio under each economic state of the world simulated.

In addition to computing losses across simulated economic paths, the tool can be used to estimate losses under one or more specific pre-defined forecast scenarios as well. Furthermore, user-supplied information (e.g., historical portfolio performance, loan modifications, etc.) may also be included as input.

The starting period for the economic simulations is period following that of the most recently available economic data, which is currently updated each quarter. The scenarios for economic factors are generated over the subsequent (future) 120 months. Briefly, the macro-economic factors are modeled as follows:

- **the base interest rate (U.S. Treasury)** is modeled using a **two factor Cox-Ingersoll-Ross (CIR)** term-structure model;
- **the home price and unemployment rate** projections are modeled using **auto-regressive models**, in which projections for each quarter are based on the previous two quarters of unemployment and home prices data, respectively, as well as the 10-year Treasury;

The algorithm (pseudo-code) of calculating the losses at the loan- and portfolio-level for each period using macro-economic factors is given in Box 1.

Box 1: Stylized outline of the multi-step Monte Carlo simulation

1. For each economic scenario
    a. Generate one realization of macro-economic variables over the next 10 years
        i. For each loan
            1. For each month
                I. If the loan has not defaulted, prepaid, or fully amortized
                    i. Calculate the loan level probability of default, *d*
                    ii. Calculate the loan level probability of prepayment, *p*
                    iii. Generate a uniform random number, *u*, in the range (0,1)
                    iv. Determine whether the loan defaults in this month ($u<=d$)
                        1. If default
                            a. mark loan as defaulted
                            b. determine loan-level severity
                            c. record a loss
                    v. Determine whether the loan prepays this month ($d<u<=d+p$)
                        1. If prepay, mark loan as prepaid
                II. End If
            2. End of month
        ii. End for each loan
    b. Sum losses for all loans to arrive at portfolio loss for this economic scenario
2. End for each economic scenario
3. Calculate loss distribution (every point in the distribution represents the portfolio loss under a single economic scenario)

2.3 Use of MPA for evaluating cashflow RMBS transactions and ALM applications

Since the simulator performs a multi-step Monte-Carlo (i.e., it computes losses at a monthly frequency for simulated paths), it can produce monthly cash flows for each loan under each simulated path over the specified horizon. These may be used for ALM analysis or in conjunction with an RMBS waterfall tool to produce detailed cash flow analyses for RMBS transactions.

Since the mortgage collateral pool losses are driven by economic factors, and the RMBS liabilities are also driven by these factors, the inter-relationships between the assets and the liabilities of an RMBS transaction are naturally modeled.

Consider two examples:

- High interest rates depress prepayments and lead to higher realized default rates in the portfolio. High interest rates also lead to higher interest payments on any floating-rate liabilities (or RMBS notes). Thus, at the same time the collateral cash flows are being reduced, the liability expense is increasing.
- If collateral losses occur early in an RMBS transaction, waterfall triggers will be tripped in the cash flow model (as they actually would, in such a setting) causing the payout to senior bond holders to be different than it would be were collateral losses to occur later in the transaction.

In contrast to the rather rich representation of the behavior of the assets and liabilities modeled in this multi-step simulation, simpler frameworks such as copula or single-step methods typically require the user to specify assumed default timing and severity curves in order to model the liability cashflows in structured transactions.

2.4 *The meaning of* "expected loss"

The expected loss for a mortgage portfolio is the *average loss* that the portfolio experiences over the horizon of analysis. The average is computed across the economic conditions simulated.

The expected loss is *not necessarily the most likely* loss outcome.

Confusion about what is meant by expected loss sometimes arises because the term *expected* is sometimes used differently in statistics than in when it is used in the vernacular. In statistics it means *on average* whereas in common conversation it is usually interpreted as meaning *typical* or *most likely*.

As with many other credit-risky assets, in the case of mortgage portfolios, the expected loss is often not the loss that will typically occur. In fact, MPA predicts a wide variety of potential losses, each of which will occur with some probability - the *loss distribution*. Averaging across these possible loss outcomes produces mean loss, which is the estimate of the expected loss. The *most likely outcome* is the *mode* of the distribution. For most credit-sensitive portfolios the modal loss will be smaller than the mean loss.[6]

Credit portfolios with correlated assets tend to have skewed distributions. This means that in most scenarios the observed loss will be less than the expected loss. However, there is a small probability of experiencing losses that are much higher than the expected loss. Generally, we might think of the lower losses happening in "normal" or "good" times and the larger losses happen during times of stress. We anticipate this when we see a mortgage loss distribution, as discussed in Section 2.14.

2.5 The general specification for default and prepayment models

MPA accommodates two types of loan exit: loans may exit a portfolio before maturity due to either default or prepayment. Both events are modeled using hazard rate models based on time-to-event data (the duration of time elapsed until the event, default or prepayment, happens for the first time).

Hazard rate models produce estimates of the duration of a loan's survival rather than the probability of prepayment or default over a specified horizon (although a transformation can convert the output from hazard rates to probabilities and *vice-versa*). In this setting, *all* loans are assumed to have the potential to prepay or default over a sufficiently long horizon. With time-to-event data, use of survival analysis can produce estimates of the *conditional* probability that a

[6] A rule of thumb for most credit portfolio distributions is: mode < median < mean. For a perfectly symmetrical unimodal distribution all three measures will be equal.

loan prepays or defaults over the next interval of time, where conditioning is done *subject to having survived up until that time*.

More specifically, MPA makes use of semi-parametric Cox hazard models (Cox, 1972); in some cases augmented with non-parametric techniques to better characterize the various behaviors of different types of loans. Box 2 below gives a brief overview of the basic model. The Cox hazard models used in this context make no assumptions about the *shape* of the baseline hazard rate function. This flexibility is important given the unique impact that loan characteristics (e.g., reset periods) can have on the underlying processes. This differs from simpler, but more constrained, parametric survival models, such as Weibull, loglogistic, and exponential. The use of a non-parametric method reduces the risk of model misspecification and provides much greater control over the baseline hazard shape, thus capturing the observed peaks and troughs in the empirical distributions more accurately. For example, it is known that different types of hybrid ARM loans (5/25, 7/1 etc.) have different prepayment peaks due to different reset timing, and thus warrant different baselines. It is straightforward to capture different baseline patterns by loan type using the semi-parametric Cox hazard model with stratification. This flexibility can be difficult to achieve using a simple parametric model.

In MPA's models, time-dependent variables are included (e.g., mortgage premium, home price changes, etc.). In this setting, the hazard ratios between loans are no longer constant over time due to the changing factors. However, the MPA models are constrained such that no interaction between time and other factors is used in the default/prepayment hazard models. This ensures that the effect of a unit change in the factors on the hazard ratio is still constant over time.

Some of the loans in the data set are not observed completely from origination because they are already seasoned when the data starts in 2001. Some others are not reported until a few months after their originations, and they are, thus, already seasoned when they appear in the data for the first time. This phenomenon, called left truncation, has been well studied in the analysis of survival data and a number of adjustments for it have been developed. In the case of MPA's models, the default/prepayment hazard rates are estimated in a time varying counting process manner; left truncated loans only enter into the risk sets when they are observed.

In addition, *right censoring* is also present in mortgage portfolio data. Two types of right censoring are addressed in our model:

- Type I: a loan stays in the data set during the entire observation period and, as of the last observation, has neither defaulted nor prepaid.
- Type II: a loan is terminated (due to default or prepayment) during the sample period but due to the “other” (competing) risk. That is, in the context of a default model, the loan leaves the data set because it prepaid. Since its exit is not due to default, this observation must be treated as a censored observation with respect to default. Similarly, a defaulted loan is treated as a censored observation in the estimation used in prepayment model.

An additional technical complication in the estimation involves the presence of possible "ties" in the data. Mortgage data are typically observed at a monthly frequency. When incidence rates are high and/or the number of observations is large, often multiple loans are reported as terminated at the same point in time (that is, in the same month). These are termed "tied events" since there are multiple events in a single “instant” of time. A key assumption of the standard Cox proportional hazard model is that no two events occur at the same time. This assumption would be violated by the ties. However, the standard Cox Proportional Hazard model can be extended to accommodate ties through a number of approaches. We use Efron’s method (Efron, 1977) to address the presence of ties.[7]

After the coefficients for the covariates have been estimated, the baseline hazard rates must also be estimated. This estimation involves a modified Bayesian framework that combines prior knowledge of prepayment and default behaviors for various loan types (with respect to loan age) with data on the empirical distributions of prepayment and default times. The baseline rates are estimated using spline functions and non-parametric density estimation.[8] The results suggest that for modeling prepayments, spline functions fit the data better and capture the known turning points more efficiently. Examples of such turning points include a peak around month 12, which reflects increased prepayments due to credit curing, and a peak around the reset date for ARMs (e.g., a peak is observed around month 60 for 5/25 ARM loans). (See Section 2.7.2 for examples of additional prepayment patterns due to different reset times and prepayment penalty terms.). Splines also represent the functional form more compactly making them attractive computationally.

[7] In the survival analysis literature a number of methods for addressing ties has been suggested. Three methods are commonly used by practitioners: Kalbfleisch and Prentice’s exact method (Kalbfleisch and Prentice, 1980), Breslow’s approximation (Breslow, 1974), and Efron’s approximation (Efron, 1977). When there are many such tied events, Kalbfleisch and Prentice’s (1980) exact method is computationally inefficient. On the other hand, Breslow’s (1974) approximation is most efficient computationally but its estimations are relatively more biased than the other two. We use Efron’s (1977) method since its results are closer to the exact results than Breslow’s but involve a trivial increase in the computation time.

[8] Interested readers can find references on these topics from many sources, e.g. Hastie and Tibshirani (2001) and Siminoff (1996).

Box 2 Cox Proportional Hazard Model

Survival specifications directly model the duration of survival (rather the presence or absence of a default or prepayment event at a given horizon). The fundamental quantity of a survival model is the hazard function, also known as conditional failure rate. $h_i(t)\Delta t$ is approximately the probability at time *t* that the event will occur for loan i in the next small time interval, Δt . For continuous survival data, the hazard rate is the instantaneous risk that an event will occur to subject i at time t, given that the subject survives to time t:

$$h_i(t) = \lim_{\Delta t \to 0} \frac{P(t \le T_i < t + \Delta t | T_i \ge t)}{\Delta t}$$

$h_i(t)$ is restricted to be nonnegative.

Survival models can be viewed as ordinary additive models in which the dependent variable is taken as the *time until the occurrence of an event*. However, the computation of the likelihood function is made more complicated by the censoring of the observations over time. Censoring is a common feature of survival data. Complete observation of the exit time T for each subject is not always possible. Unless a loan matures during the observation period, the loan's exit time is observed only if a loan is terminated: either prepaid or defaulted.

If a loan has not yet prepaid or defaulted by the end of the observation period, the event time is after the last period in which it could be observed as having prepaid or defaulted. For this loan, the time-to-event is treated as a censored observation. In addition to censoring at the end of the sampling period, observations can be censored at any time point during the sample period due to competing risks.

The Cox Proportional Hazards model is a popular survival model. Given the covariate vector x_i, the conditional hazard function for loan $i$is defined as

$$h_i(t) = h_0(t) e^{\beta' f(x_i)}$$

where $f(x)$ are the transformed factors, and $h_0(t)$ denotes the baseline hazard function. This is the hazard when all transformed covariates are zero. We use a single f to denote the transformations for notational simplicity. However they may be different for each factor.

Under this formulation, across all loans there is a baseline hazard rate that varies with the age of the loan (time since origination). A *particular* loan's hazard rate is based on the baseline but is increased or decreased due to the attributes of the loan, borrower and property and the state of the economy.

In the plain vanilla Cox model described above, the covariates are static over time, so the ratio of the estimated hazards over time will be constant; thus this model is referred as the "proportional" hazard model. It has been shown (Cox (1972,1975)) that, by eliminating the infinite dimensional nuisance parameter from the full likelihood, it is possible to estimate $\boldsymbol{\beta}$, the vector coefficients, based on a partial likelihood approach without explicit knowledge of the baseline function, $h_0(t)$.

2.6 The variable selection process

The research of academics, industry participants, RMBS analysts and modelers has generated long lists of factors that potentially explain different aspects of loan performance. MPA's factor selection draws on the insights from this rich body of research as well as, importantly, on the quantitative properties of the factors (e.g., robustness and predictive power).

In general, we have found that economic reasoning provides a more robust guide to factor selection than data-driven statistical correlations. For example, consider again the impact of the prevailing interest rate on prepayments, which we discussed in Section 2.1. As it turns out, the absolute level of interest rates is statistically related to prepayment through the time period of the mortgage data we examined. However, we find the simple "regression" relationship of prepayment on levels of interest rates to be less informative than the economic relationship that relates interest rates to the economic drivers of prepayment. In this case, interest rates affect prepayments through the spread between a borrower's current mortgage rate and the prevailing rate. When this spread is negative – that is the prevailing refinance rate is higher than or at least very close to the borrower's current mortgage rate – the borrower is relatively better off since the opportunity cost of not refinancing is small. Thus, in this case, there is lower incentive to prepay. On the other hand, when the spread between the current mortgage rate and the prevailing interest rate is negative (the borrower's rate is higher than the current refinance rate) there is a greater incentive to prepay. Therefore, the same *absolute* level of interest rates may affect different borrowers' propensity to prepay differently. This observation leads naturally to construction of loan-specific measures of the impact of interest rate on prepayments rather than measures based on the absolute *level* of market rates as the simple "regression" relationship would have suggested.

As another example, consider the macro-economic variable, home price change (*HPC*). In conjunction with information about the loan and property, *HPC* can be used to infer the borrower's *updated LTV*, which has a theoretically sound relationship to both default and severity. Rather than entering *LTV* and *HPC* independently into the model, it is more economically sensible to use the updated *LTV* factor to capture this effect. The use of transformed versions of the raw variables, when possible, results in models that are more stable and have greater predictive power.

It is not always practically possible to derive factors that combine loan-specific characteristics with a macro-economic factor. However, even when a macro-economic factor enters the model directly, rather than through other factors, economic reasoning provides guidance in deciding how this factor should be used as an explanatory variable (i.e., level, difference, lag, etc). For example, such reasoning motivates leads to the use of the *cumulative change* in home prices since the loan was originated rather than simply the *level* of home prices.

Of course, in addition to having attractive theoretical, statistical and predictive properties, factors must also be practically implementable from a business perspective.

In summary, the criteria for selection of explanatory factors are:

- **Common sense/Economic theory**. There should be economic rationale.
- **Statistical robustness.** A factor should:
  - have coefficient whose **direction (sign) and magnitude** are consistent with economic reasoning; and
  - have (relatively) **low correlation with other factors** in the model: a factor's inclusion should not introduce multicolinearity; For example, there are many factors that measure a borrower's stake in a property (borrower equity, updated *LTV*, and so forth). Since these are highly correlated, including several may introduce parameter instability.
- **Practical implementation**. A factor should be:
  - **available readily**: most institutions must be able to calculate the factor based on data they maintain and can easily access. As an example, neither updated FICO nor updated appraisal value is readily available to many market participants since many lenders do not update them regularly.
  - **defined relatively unambiguously**: most institutions must be able to calculate the factor in a similar manner or based on some simple rules; Debt-to-income ratio (DTI), for example, is not used in the models because it can be defined differently across different lenders.

Factor selection for the models in MPA is based first on univariate analysis that examines the ability of each factor to predict default, severity or prepayment on its own. It then proceeds to multivariate methods that assess which of these factors remain predictive in the presence of other factors in the model. The final step is a suite of out-of-sample tests for overfitting.

Univariate Analysis: Unlike multivariate regression analysis, univariate analysis does not control for the effects of other variables. This initial analysis can be used to examine the predictive power of individual variables by determining the univariate relationship between each variable and the frequency of the specified event (default or prepayment). The univariate factor screening makes use of techniques similar to those described in Falkenstein, Boral, and Carty (2000), Dwyer and Stein (2004) and Bohn and Stein (2009) for this analysis. These include use a number of non-parametric techniques (c.f., Hastie and Tibshirani (1990) and Siminoff (1996)) to control for both remaining outliers and non-linearity in the data.

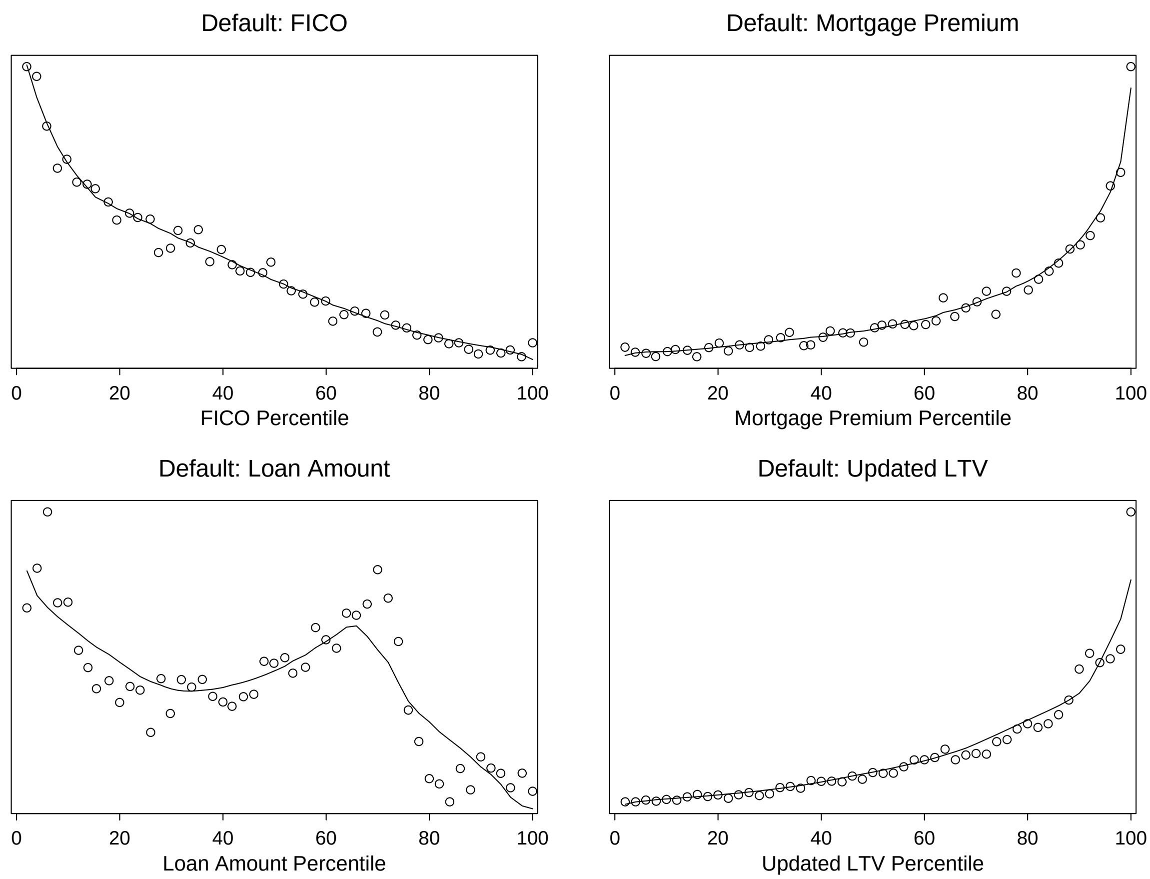
The general approach of the univariate analysis:

1. Rank all mortgage loans by a particular predictor variable. For time-varying variables, we rank the monthly observations.
2. Segment the loans into k bins, where k and the width (range of values for the predictor variable) of each bin are determined by the distribution of the variable.
3. Calculate the frequency of the event, default or prepayment, for each bin, which results in a granular mapping from the predictor variable to the frequency of the event. (This approach is commonly called bin smoothing.)
4. Generalize the estimates of relationship to by (non-parametrically) smoothing across the bins. This approach smoothes the default/prepayment rates across bins, while minimizing the prediction error for each bin.
5. Interpolate the values that fall between the endpoints of a bin using the density estimate.
6. Examine both the nature of the relationship (slope, monotonicity, etc.) and the univariate predictive power of this variable. (The latter can be achieved both by using the density estimates of the relationship and by calculating power curves.)

An analysis of the shape of the curve and the spacing of the bin boundaries provides insight into whether the relationship is in the expected direction or whether it is counterintuitive, suggesting data issues may be present or that (for the segment of the population) the theoretical relationship may not hold as strongly. It also gives a sense of how significantly a change in this variable affects the probability of an event. If the slope of the curve is steep, a small change in the levels of the factor will have a larger impact on the examined event than if the slope is flatter. Finally, the shape of the curves demonstrates the extent to which the relationship is non-linear, which would suggest using certain transformations.

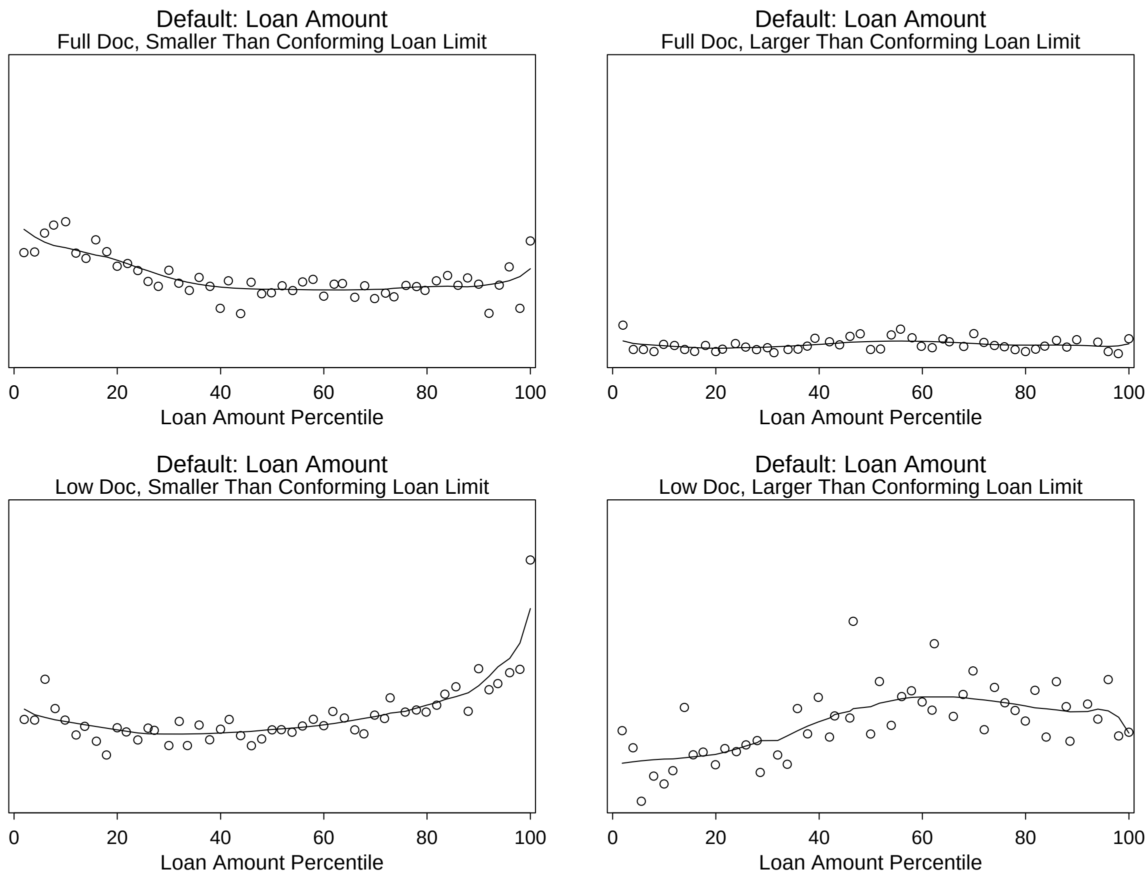
As an example of the usefulness of the univariate mapping, Figure 3 provides the mappings of some of the factors in the prime default model. Note how the transformation function is itself an estimate of the default rates at each quantile. The lack of perfect agreement between the actual points and the smoothed density estimate is not only due to sampling noise, but also due to the fact that the factor is not a perfect predictor.

Figure 3: Univariate relationship of selected variables in the prime default model



The relationship for FICO (upper left) is monotonic with a pronounced slope, suggesting high predictive power for FICO. The downward slope is economically reasonable: we expect borrowers with higher FICO score to have lower default rates. Similarly, the monotonic relationships between default and mortgage premium and updated *LTV* are also intuitive. In contrast, the relationship between default and loan amount suggests strong association, but also that loan amount by itself is probably not a good predictor. In such cases, it is often useful to examine interactions of two or more variables when they are economically sensible. For example, as shown in Figure 4, loan amount is found to have much more intuitive relationship with default rates for prime loans when it is considered jointly with documentation.[9]

Figure 4: The impact of prime loan amount on default hazard depends on documentation



Multivariate analysis: The next step in the factor selection process is multivariate analysis. Because many variables are correlated to some degree, it is important to evaluate whether the variables are *too* highly correlated to permit their joint inclusion in the model. This is done using both inspection (e.g., whether the added variable changes the sign of the coefficient of an existing variable in the model or whether the coefficient of the added variable has the wrong sign) and through the use of formal statistical methods such as calculating variance inflation factors (VIF).

[9] Together, loan amount and documentation indicate the degree to which a given loan agrees with prototypical notions of "Prime."

Variables that have a strong monotonic relationship in univariate analysis are often robust in multivariate analysis, provided they do not exhibit excessive colinearity. The variables which exhibit weak predictive power (an almost flat relationship) in the univariate analysis will typically not be statistically significant and may even have a coefficient with an unintuitive sign in the multivariate estimation.

Parameter stability: In the course of multivariate estimation, a useful check on model stability is to subset the data into smaller time periods and then re-estimates the coefficients for the multivariate model in each sub-period and to then examine the variation in coefficient value and significance across the different sub- periods. (In this context, the focus is on parameter stability, rather than on evaluating out-of-sample performance. Out-of-sample performance testing is discussed in Section 4).

2.7 The models for prime mortgages

In this section, we describe in detail the models (default, prepayment, and severity) for the prime residential mortgages. However, before going into the details, we summarize, in **Table 1**, the key factors and their effects in the default, prepayment, and severity models. The symbol "+" implies a positive relationship between the factor and the corresponding model whereas a "–" implies an inverse relationship.

2.7.1 Prime mortgage model variables: The default model factors

The default model is of the form of a semi-parametric Cox hazard model with time-dependent variables (see Box 2), which include mortgage premium and updated *LTV*. Ideally one might also wish to include time-dependent FICO in the default model, since the effect of original FICO dampens over time. However, updated FICO scores are often not available since many lenders do not refresh them regularly.

Table 1: The key factors and their effects on prime default, prepayment, and severity models

Factor		Default	Prepayment FRM	Prepayment ARM	Severity
LTV			–	–	
Updated *LTV*		+			
HPI Change			+	+	
Forward *LTV*[1]					+
Junior *LTV*		+	–	–	
Mortgage premium at origination		+	+	+	+
Mortgage premium change		+	+	+	
Unemployment rate change		+			
Unemployment rate		+			
Payment shock of initial reset		+			
FICO		–	+	+	
Judicial state		+			+
Loan amount		+/–	+/–	+/–	–
Documentation[1]		Bad Doc > Good Doc	+/–	Good Doc > Bad Doc	
Loan type		ARM > FRM			
Prepayment penalty term			–	–	
Burnout effect			–	–	
Occupancy type (relative to Owner occupied)	Investment	+	–	–	
	Second home	+	–	–	
Loan purpose (relative to purchase)	Rate-refi	+	–	+	
	Cashout	+	–	+	
	Debt consolidation		–	–	
	Home improvement		–	+	
Property Type (relative to single family)[1]	Condo	–	+	–	
	Co-op	–	–	–	
	PUD	–	+	+	
	Townhouse	–	+	–	

The definition of default

In MPA's default models, the definition of default is that

- a loan's reported status transitions to foreclosure;
- a loan's reported status transitions to real-estate-owned (REO);
- the borrower declares bankruptcy; or
- the borrower is more than 120 days past due on a mortgage payment and subsequently enters one of the three states listed above.

The timing of default is determined as the earlier of foreclosure, REO, borrower's bankruptcy or 120 days

The most important factors for predicting defaults for prime loans include FICO score, mortgage premium (both the mortgage premium at origination and the change over time), updated *LTV*, junior *LTV*, loan structure and loan size (**A complete list of the factors in the default model is given in Table 1**), and we discuss these briefly below.

FICO: FICO scores at loan origination (holding all other loan factors equal) are a strong predictor of default levels. While the impact of FICO at origination declines over time, possibly due to credit curing or FICO drift, it is a reasonable indicator of default risk even for seasoned loans.

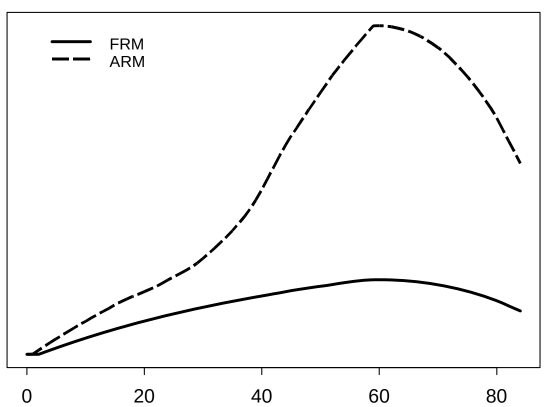
Mortgage Premium at origination: This variable is constructed as the spread between the coupon rate and the market rate at the time of loan origination. The market rate is measured by the survey of 30-year fixed rate mortgages given by Freddie Mac for prime borrowers. Loans with a higher mortgage premium are likely to default for two reasons. First, loans with a higher mortgage premium are associated with higher coupon rates and therefore have higher debt service burdens. Second, all else equal, higher mortgage premiums suggest a greater level of credit risk at the time of underwriting, perhaps due to borrower-specific factors that the lender may observe but that may not be formally reported.

Change in Mortgage Premium: As mentioned above, mortgage premium at origination carries additional risk information not captured by FICO, providing an additional measure of relative credit worthiness of the borrower. If nothing else changes over time, a borrower is expected to be able to avoid an increase in her mortgage premium (at least beyond transaction costs) through opportunistic refinancing. Observing an increased mortgage premium thus suggests a troubled borrower who is either unable to refinance or, for those same (unobserved) reasons, at greater risk of default. It also suggests that the borrower's financial burden may have increased, as monthly payments are higher than at the time of origination.

Updated Loan-to-value: *LTV* is a measure of leverage for the loan. For newly-originated purchase loans, *LTV* reflects the borrower's down payment, which in turn tends to be a function of the borrower's long-term financial capacity and discipline and the borrower's commitment to the property. Updated *LTV* is calculated by updating the house price using the state- or MSA- (when available) level housing price changes. This permits a more accurate time-varying measure of borrower equity. Higher values of the updated *LTV* imply lower equity in the home and, hence, reduced aversion to defaulting on the property.

The Loan Structure: Loan features can vary substantially across loans. These features, including coupon resets, coupon levels, the presence or absence of a teaser rate, and amortization terms, influence default behavior due to phenomena such as payment shock. (If, after a loan resets, the interest rate rises significantly due to changes in market rates or the presence of a teaser rate, this additional cost will impact default rates.)

Figure 5: Default baselines – prime loans



Baseline Hazard Rate: The baseline describes the general trend of the default rates for borrowers as a function of loan age. Default rates for prime mortgages tend to start low in the first year after loan origination and then rise, peaking around the fifth year after origination. This seasoning effect is reflected in the hump shape of the default baseline. All else equal, a loan seasoned past this peak will have a lower life-time default risk than that of a newer loan. We use different default baselines for ARM and fixed-rate mortgages (See Figure 5). The baselines of the default models are estimated using non-parametric density estimation techniques. This overall behavioral baseline is then modified in each period by macroeconomic and loan specific factors (see: Box 3) to project the default paths of each loan.

2.7.2 Prime mortgage model variables: The prepayment model factors

As in the case of the default model, the prepayment model takes the form of a semi-parametric Cox hazard model. The most important factors for predicting prepayments include home price change, mortgage premium (the premium at origination and the change over time), burnout effect, FICO, and *LTV* (**A complete list of the factors in the prepayment model is given in Table 1**.

Home Price Change is measured as the change of local home prices from the origination of the loan to the current date. When available, MSA-level home price is used to estimate the home price change for each loan. When MSA-level home price is not available, state-level is used. Home price change has a significant and positive effect on the rate of prepayment. When the home prices increase, the borrower's equity also increases. This leads to more favorable terms for rate-refinancing or cash-out refinancing with lower *LTV* (due to increased home price). Higher values of home equity also provide stronger economic incentives for the borrower to prepay in order to benefit from the accumulated equity. The opposite is true in a falling home price market as the borrower's equity will diminish or even become negative. In a falling home price market, therefore, both borrowers' ability and incentive to prepay can be significantly reduced.

Mortgage premium at origination is defined as the difference between the mortgage rate for a loan and the prevailing average mortgage rate for prime mortgages (e.g., FHLMC rate). Loans with a higher mortgage premium tend to have a higher prepayment rates. The reason for this is twofold. First, a borrower with a higher premium on his loan has a greater incentive to refinance, as his savings in financing costs (e.g., reduction in monthly payments) will be greater. Second, a higher premium is associated with higher credit risk at origination. Over time, the creditworthiness of borrowers that do not default typically improves because of credit curing, which means the borrower may be able to refinance at a lower rate that better reflects the improved creditworthiness.

Mortgage premium Change is defined as the change in the mortgage premium from the point of origination of the loan. A positive change could be due to a decrease in the market rate or an increase in the borrower's coupon rate or both. In either case, the borrowers have greater incentive to refinance.

Prepayment penalty clauses are provisions in the mortgage contract that require the borrower to pay a penalty if she pays off the loan before a certain date. The penalty is paid only if the loan is prepaid within a certain period of origination, referred to as the prepayment penalty period. Loans with prepayment penalty clauses are significantly less likely to be prepaid prior to the expiration of the prepayment penalty because the cost of this penalty may outweigh the long term benefit of reduced monthly payments. Note that prepayment penalty clauses are not as common for prime mortgages as for subprime or Alt-A mortgages.[10]

Burnout effect is meant to capture unobserved borrower-specific factors that may make it less likely for a mortgage-holder to refinance, even in environments that favor refinancing. We model this as a dummy variable which is set to value 1 when a borrower does not make use of, at least, two refinancing opportunities over a period of two years. In MPA's prepayment models, the definition of a refinancing opportunity is a month during which the loan is not in the prepayment penalty period and in which the prevailing market rate is lower than the current coupon rate by more than 200 bps. Loans that exhibit burnout are less likely to prepay in subsequent periods. If a borrower has decided not to refinance when doing so would reduce the borrower's financing costs, we assume that the borrower is either insensitive to changes in mortgage rates or has other reasons to avoid refinancing, and is therefore less likely to prepay the loan in the future.

FICO at origination is a significant predictor of prepayment, holding all other loan factors constant. FICO measures the credit quality of an individual borrower. The higher the FICO score, the higher the likelihood that the borrower could qualify for more attractive refinancing terms. Furthermore, in periods of tighter lending, lower credit quality borrowers may have more difficulty refinancing. The impact of FICO declines over time, possibly due to credit curing or FICO drift.

Loan-to-value (*LTV*) is a measure of leverage for the loan. *LTV* at origination is included in the model (together with Junior *LTV*). For newly originated purchase loans, *LTV* reflects the borrower's down payment, which in turn tends to be a function of the borrower's long-term financial capacity and discipline and the borrower's commitment to the property. The higher the *LTV*, the less equity a borrower has in his house and the less flexibility he has in refinancing or selling the property.

---

[10] In our data sample for prime loans less than 4% of the loans have prepayment penalty clauses.

Model Baseline Hazard Rates

Each model has an independent baseline. The baseline describes the general trend of the prepayment rates for various loan categories, such as 5/25, fixed, and so on, purely as a function of elapsed time. The loan specific factors and elements of the simulated economic paths are overlaid on the behavioral baseline pursuant to the Cox hazard framework to determine the modeled prepayment rate paths of each loan (see Box 2).

The observed prepayment behavior differs between fixed-rate and ARM loans. To accommodate this, MPA uses separate models for these two broad loan types. The ARM model is further refined by stratifying various baselines that differ by loan type within the ARM type depending on the structure (i.e. reset month). In general, there is a prepayment spike for both ARM and fixed loans after one year (resulting from the credit curing of some borrowers). For ARMS there is also generally second spike after a loan resets. Prepayment penalties can create additional spikes.

For example, in examining 5/25 loans with 24-month prepayment penalty terms, we observe a prepayment peak around month 24 (prepayment penalty expires) followed by a pronounced increase in prepayments in year 5, when the loan rate reset. If this heterogeneity is not addressed in the model--for example, if one common baseline were used for the different loans--the peak time would be biased and the bias would be proportional to the degree of heterogeneity in the population.

Figure 6 gives some examples of the uncalibrated prepayment baselines for FRM and ARM loans (the graphs have different scales). These can be interpreted as being the general shapes of the prepayment hazard rates ($h_i^p(t)$) before calibration. These baselines are scaled upward or downward depending on loan and borrower characteristics and the state of the modeled economy. Since some of the loan-level factors used in MPA's analysis depend on time-varying macro economic factors (for instance, home price appreciation or interest rate factors), the model scales the baseline differently in each month of an economic scenario. The effect of the scaling varying with time (due to changing macroeconomic conditions) is that the actual prepayment curve for a specific loan typically looks like a "warped" version of the baseline.

To give some sense of how different the actual prepayment hazard rate can be in different economic paths, Figure 7, below shows the hazard rate *h(t)* of a single 5/25 loan in different economic scenarios. Note that both the shape and the level of the prepayment curves vary substantially with the economy.

Figure 6: Baseline FRM and ARM prepayment rates for prime loans

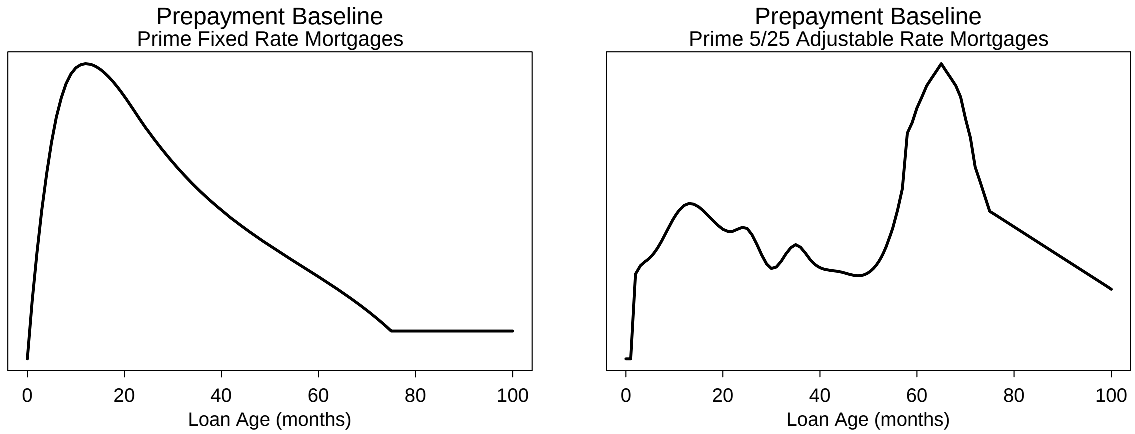
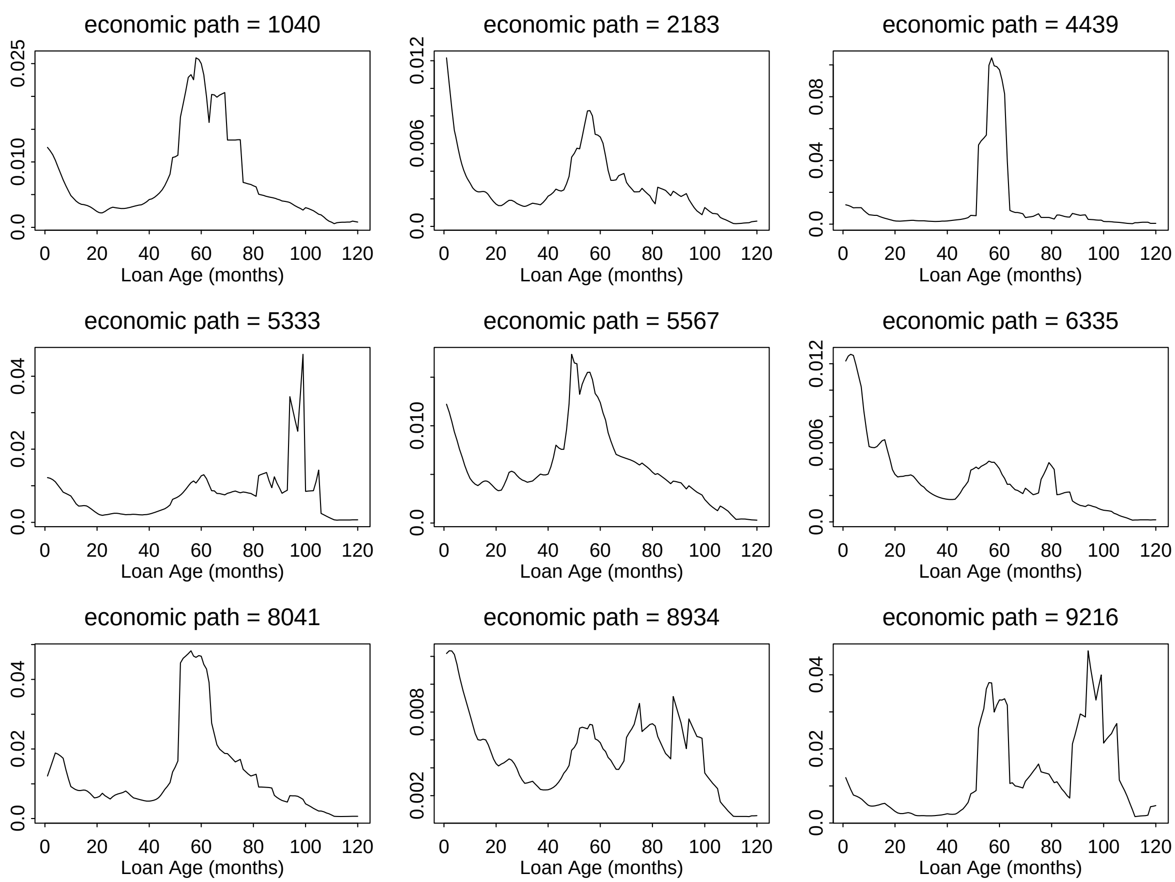


Figure 7 Prepayment hazard rate of a single prime loan in different economic paths

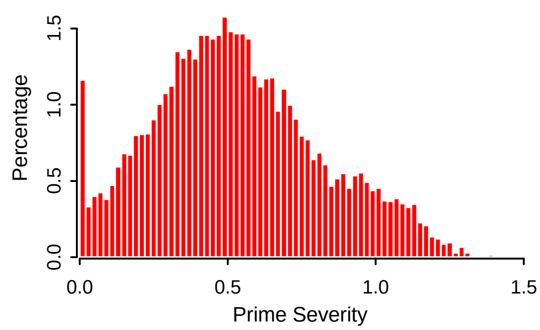


2.7.3 Prime mortgage severity model variables: LGD for prime loans

The severity or loss given default (LGD) is defined as the ratio of loss (as reported by originators and servicers) to original balance of the mortgage loan. MPA computes the dollar loss by multiplying severity by the original balance.

While a body of literature exists describing modeling of loss given default for corporate exposures (c.f., Gupton and Stein (2002, 2005)), Acharya, Sreedhar, and Srinivasan (2004)) less has been written on loss given default for retail exposures.[11] However, there are similarities. Clearly, both processes describe distributions of ratios. A common approach to modeling such a process is to assume that the data generating process follows a Beta distribution, which is chosen for several reasons. First, it can take a variety of shapes based on the choice of parameter values. Second, the parameter values can be readily estimated using only the sample mean and sample variance, which facilitates implementation of the model.

Figure 8: Distribution of loan severity for uninsured prime loans



The distribution of severity in our sample, as shown in Figure 8, is consistent with the skewed shape characteristic of the Beta distribution. A severity in excess of 100% seems improbably high, yet in our data, more than 10% of the defaulted loans have such high severities.

How could loans have greater than 100% severity? We find that these severities typically arise in cases where the loan amount is quite low and therefore the fixed costs associated with the recovery process, including servicing advances, are higher than the expected recovery amount, and thus severity exceeds 100%. This is consistent with the finding that smaller loans tend to have larger percentage losses as reported by Qi and Yang (2008) and others. We discuss this trend later in this article.

[11] This has recently begun to change. See for example Qi and Yang (2008) and references therein.

Given the observed beta-like distribution MPA's severity models use a methodology similar to that described in Gupton and Stein (2002).[12] First, the observed severities are transformed to follow an approximate Normal distribution by first applying Beta distribution and then applying inverse Normal transformation. Next, a linear regression is fit to the transformed values. Finally, an inverse beta transformation is performed on the predicted values from the liner model to compute the predicted severity.

The basic form of this model is:

$$s_i = Beta^{-1}(\Phi(z_i), \hat{\alpha}, \hat{\beta})$$

$$z_i = \hat{\delta}' q_i(X_i^s) + \varepsilon_i$$

where,

ε_i is distributed $N(0, \sigma_i^2)$, which makes the severity stochastic,

s_i is the severity of the i^{th} defaulted loan,

$\hat{\alpha}, \hat{\beta}$ are the estimated parameters of the Beta distribution,

$\Phi(.)$ and $Beta(.)^{-1}$ and are the standard cumulative Normal and inverse Beta distribution functions, respectively,

$\hat{\delta}$ is a vector of parameter estimates,

X_i^s is a set of loan specific and macro economic factors affecting severity for the i^{th}defaulted loan,

$q_i(\cdot)$ is a set of transformation functions of the individual factors (the form of the transformation may vary across factors.

The key factors for predicting severity are:

- **Judicial regulation:** This factor describes whether or not a loan is in a U.S. state that has judicial regulations relating to loan foreclosure and default (these make it harder to liquidate a property). This delay has a direct impact on the losses since the longer the process to liquidation, the more cost is likely to be incurred.

[12] Since the publication of Gupton and Stein (2002), more formal techniques have been introduced in the statistics literature (c.f., Ferrari and Cribari-Neto, 2004), though the stability of these techniques appears to be lower than that of Gupton and Stein (2002) in some empirical settings, such as this one.

- **Forward loan to value ratio:** Loan to value ratio is updated using the simulated value of the house as of the estimated liquidation date, which is assumed to be eighteen months from the time of default if the loan is in a state with judicial regulations on loan foreclosure, and twelve months after default otherwise.[13] Empirically, forward *LTV* appears to be more predictive than the updated *LTV* at the time of default. This is because liquidation takes time, and the house value can drop significantly between the time of default and the time when the house is actually sold as part of the liquidation process.
- **Loan amount:** Smaller loans tend to suffer a higher percentage loss, presumably due to the fixed cost associated with the recovery process. This is consistent with economic reasoning and some recent academic literature. Furthermore, our research and discussions with mortgage servicers suggest that they also recognize this effect and expend greater effort to recover losses on large loans than on smaller loans. In the limit, if the property value is less than the fixed cost of recovery, there is little incentive for the servicer to pursue the recovery.
- **Mortgage premium:** A higher mortgage premium is associated with lower principal payment (as a percentage of the fixed payments) and higher relative accruing loan coupon payments; it may also indicate lower borrower quality (see Section 2.7.1 for details on mortgage premium).

A complete list of the factors in the severity model is given in Table 1.

2.8 The models for subprime mortgages

In Table 2 below, we summarize the key factors and their effects in the default, prepayment, and severity models for subprime loans. As in Table 1, the symbol "+" implies a positive relationship between the factor and the corresponding model (default, prepayment, or severity) whereas a "–" implies an inverse relationship. Notice that there are differences in the factors included in the subprime models and the prime models. Furthermore, even in those cases in which the same factors are used, their effects may not be the same. Therefore, for completeness, and at the risk of some redundancy, we discuss in the following sections the subprime model factors at a full level of detail as we did for prime model factors. This also allows readers interested in only Subprime or Prime (but not both) to read the appropriate section without the need to cross-reference other models.

[13] The prediction for the value of the house in the future is done using our MSA-level model for house price appreciation. Section 2.10.2 provides more details on this.

Table 2: The key factors and their effects on subprime default, prepayment, and severity models

<table>
<tr><th colspan="2" rowspan="2">Factor</th><th rowspan="2">Default</th><th colspan="2">Prepayment</th><th rowspan="2">Severity</th></tr>
<tr><th>FRM</th><th>ARM</th></tr>
<tr><td colspan="2">LTV</td><td></td><td>–</td><td>–</td><td></td></tr>
<tr><td colspan="2">Updated LTV</td><td>+</td><td></td><td></td><td></td></tr>
<tr><td colspan="2">HPI Change</td><td>+</td><td>+</td><td>+</td><td></td></tr>
<tr><td colspan="2">Forward LTV</td><td></td><td></td><td></td><td>+</td></tr>
<tr><td colspan="2">Junior LTV</td><td>+</td><td>–</td><td>–</td><td></td></tr>
<tr><td colspan="2">Mortgage premium at origination</td><td>+</td><td>+</td><td>+</td><td>+</td></tr>
<tr><td colspan="2">Mortgage premium change</td><td>+</td><td>+</td><td>+</td><td></td></tr>
<tr><td colspan="2">Current loan rate</td><td></td><td></td><td></td><td>+</td></tr>
<tr><td colspan="2">Unemployment rate change</td><td>+</td><td>–</td><td>–</td><td></td></tr>
<tr><td colspan="2">Unemployment rate</td><td>+</td><td></td><td></td><td></td></tr>
<tr><td colspan="2">Lending standard at origination</td><td>–</td><td></td><td></td><td>–</td></tr>
<tr><td colspan="2">FICO</td><td>–</td><td>–</td><td>+</td><td></td></tr>
<tr><td colspan="2">Judicial state</td><td>+</td><td></td><td></td><td>+</td></tr>
<tr><td colspan="2">Beyond reset period?</td><td>+</td><td></td><td></td><td></td></tr>
<tr><td colspan="2">Payment shock</td><td>+</td><td></td><td></td><td></td></tr>
<tr><td colspan="2">Loan amount</td><td>+/–</td><td>+/–</td><td>+/–</td><td>–</td></tr>
<tr><td colspan="2">Documentation</td><td>Bad Doc > Good Doc</td><td>Good Doc > Bad Doc</td><td>Good Doc > Bad Doc</td><td></td></tr>
<tr><td colspan="2">Burnout effect</td><td></td><td>–</td><td>–</td><td></td></tr>
<tr><td rowspan="2">Occupancy type (relative to Owner occupied)</td><td>Investment</td><td>+</td><td>–</td><td>–</td><td rowspan="2"></td></tr>
<tr><td>Second home</td><td>+</td><td></td><td></td></tr>
<tr><td rowspan="4">Loan purpose (relative to purchase)</td><td>Rate-refi</td><td>–</td><td>–</td><td>–</td><td rowspan="4"></td></tr>
<tr><td>Cashout</td><td>–</td><td>–</td><td>–</td></tr>
<tr><td>Debt consolidation</td><td>–</td><td>–</td><td>–</td></tr>
<tr><td>Home improvement</td><td></td><td>–</td><td>–</td></tr>
<tr><td rowspan="4">Property Type (relative to single family)</td><td>Condo</td><td>–</td><td>+</td><td>+</td><td>–</td></tr>
<tr><td>Co-op</td><td>+</td><td>–</td><td>–</td><td></td></tr>
<tr><td>PUD</td><td></td><td>–</td><td></td><td>–</td></tr>
<tr><td>Townhouse</td><td>–</td><td>+</td><td>–</td><td>–</td></tr>
</table>

2.8.1 Subprime mortgage model variables: The default model factors

The definition of default and the form of default model are the same for subprime loans as for prime loans (See Section 2.7.1). But as mentioned above, the factors are sometimes different. In MPA's subprime default models, the most important factors include FICO score, mortgage premium (both the mortgage premium at origination and the change in mortgage premium over time), updated *LTV*, scaled junior *LTV*, unemployment rate, loan structure and loan size (**A complete list of the factors in the default model is given in Table 2**).

FICO: FICO scores at loan origination (holding all other loan factors equal) are a strong predictor of default levels. While the impact of FICO at origination declines over time, possibly due to credit curing or FICO drift, it is a reasonable indicator of default risk even for seasoned loans.

Mortgage Premium at origination: This variable is constructed as the spread between the coupon rate and the market rate at the time of loan origination. The market rate is calculated as the median origination rate of all the subprime FRM loans. Loans with a higher mortgage premium are likely to default for two reasons. First, loans with a higher mortgage premium are associated with higher coupon rates and therefore have higher debt service burdens. Second, higher mortgage premiums suggest a greater level of credit risk at the time of underwriting, perhaps due to borrower-specific factors that the lender may observe but that may not be formally reported.

Change in Mortgage Premium: As mentioned above, mortgage premium at origination carries additional risk information not captured by FICO, providing an additional measure of relative credit worthiness of the borrower. If nothing else changes over time, a borrower is expected to be able to avoid an increase in her mortgage premium (at least beyond transaction costs) through opportunistic refinancing. Observing an increased mortgage premium thus suggests a troubled borrower who is either unable to refinance or, for those same (unobserved) reasons, at greater risk of default. It also suggests that the borrower's financial burden may have increased, as monthly payments are higher than at the time of origination. The change in mortgage premium for FRM loans and ARM loans are calculated and modeled differently. The reason is that for FRM loans, change in mortgage premium is caused by the change in market rate. While for ARM loans, the different dynamics of the index rate and the median market rate (as defined by us) adds another source of the change.

Updated Loan-to-value: *LTV* is a measure of leverage for the loan. For newly-originated purchase loans, *LTV* reflects the borrower's down payment, which in turn tends to be a function of the borrower's long-term financial capacity and discipline and the borrower's commitment to the property. Updated *LTV* is calculated by updating the outstanding balance as well as the house price using the MSA-level housing price changes. This permits a more accurate time-varying measure of borrower equity. Higher values of the updated *LTV* imply lower equity in the home and, hence, reduced aversion to defaulting on the property.

Scaled Junior *LTV*: Similar to the updated *LTV*, junior *LTV* is updated using the MSA level (state level if MSA is not available) housing price. However, as the updated balance of the junior lien is often unavailable or not reported in some mortgage databases, unlike the updated *LTV*, we use the beginning junior lien balance throughout. To differentiate, this factor is named scaled (since the value of the house is adjusted for the home price changes), rather than updated, junior *LTV*. It has a similar affect on the default rate as the updated *LTV*, i.e. the higher the scaled junior *LTV*, the higher the risk of default.

Unemployment Rate: Both the level of the unemployment rate and change in the unemployment rate (from the loan origination) are used in the subprime default model. Both of these have a positive relationship with the default hazard. This is intuitive, as a higher unemployment rate indicates a worse state of the economy and thus higher default risks. Moreover, regardless of the absolute state of the economy, if the unemployment rate increases after a loan is originated, borrowers are likely worse off than they were when they took out the loan, and thus they have a higher risk of default.

Lending standard at origination: There is empirical evidence that all else being comparable, loans originated under stricter underwriting standard tend to have lower default hazard. MPA includes a custom series calculated from government statistics to proxy for the underwriting quality of typical loans during different origination periods. (See Das and Stein, 2009 for a fuller discussion).

Loan Structure: Loan features can vary substantially across loans. These features, including coupon resets, coupon levels, the presence or absence of a teaser rate, and amortization terms, can influence default behavior due to phenomena such as payment shock. (If, after a loan resets, the interest rate rises significantly due to changes in market rates or the presence of a teaser rate, this additional cost will impact default rates.)

Baseline Hazard Rate: The baseline describes the general trend of the default rates for borrowers as a function of loan age. MPA uses different default baselines for ARM and fixed-rate mortgages. This overall behavioral baseline is then modified in each period by macroeconomic and loan specific factors (see: Box 3) to project the default paths of each loan.

2.8.2 Subprime mortgage model variables: The prepayment model factors

As in the case of the default model, the subprime prepayment model takes the form of a semi-parametric Cox hazard model. The most important factors for predicting subprime prepayments include mortgage premium (the premium at origination and the change over time), burnout effect, FICO, and scaled *CLTV* (**A complete list of the factors in the prepayment model is given in Table 2**).

Mortgage premium at origination is defined as the difference between the mortgage rate for a loan and the market rate at the time of loan origination. The market rate is calculated as the median origination rate of all the subprime FRM loans. Loans with a higher mortgage premium tend to have a higher prepayment rates. The reason for this is twofold. First, a borrower with a higher premium on his loan has a greater incentive to refinance, as his savings in financing costs (e.g., reduction in monthly payments) will be greater. Second, all else equal, a higher premium is associated with higher credit risk at origination. Over time, the creditworthiness of borrowers that do not default typically improves because of credit curing, which means the borrower may be able to refinance at a lower rate that better reflects the improved creditworthiness.

Mortgage premium Change is defined as the change in the mortgage premium from the point of origination of the loan. A positive change could be due to a decrease in the market rate or an increase in the borrower's coupon rate or both. In either case, the borrowers have greater incentive to refinance.

Prepayment penalty clauses are provisions in the mortgage contract that require the borrower to pay a penalty if she pays off the loan before a certain date. The penalty is paid only if the loan is prepaid within a certain period of origination, referred to as the prepayment penalty period. Loans with prepayment penalty clauses are significantly less likely to be prepaid prior to the expiration of the prepayment penalty because the cost of this penalty may outweigh the long term benefit of reduced monthly payments.

Burnout effect is meant to capture unobserved borrower-specific factors that may make it less likely for a mortgage-holder to refinance, even in environments that favor refinancing. We model this as a dummy variable which is set to value 1 when a borrower does not make use of, at least, two refinancing opportunities over a period of eight quarters. MPA's prepayment models define a refinancing opportunity as a month during which the loan is not in the prepayment penalty period and the prevailing market rate is lower than the current coupon rate by more than 200 bps. Loans that exhibit burnout are less likely to prepay in subsequent periods. If a borrower has decided not to refinance when doing so would reduce the borrower's financing costs, we assume that the borrower is either insensitive to changes in mortgage rates or has other reasons to avoid refinancing, and is therefore less likely to prepay the loan in the future.

FICO at origination has different impact on FRM and ARM subprime loans in terms of prepayment hazard. For FRM loans, it has a negative relationship with prepayment hazard given all the other factors in the model. This indicates a credit curing effect. While a high FICO borrower is more likely to have already got the best rate, a low FICO borrower can improve his/her FICO score over the course of the loan and later find a better rate to refinance into and prepay the current loan. FICO effect in ARM loans is mixed and does not exhibit a clear trend in general. However, in the particular situation where borrowers may anticipate a spike in their loan rates (because the index rate for the loan increases), higher FICO borrowers tend to have higher prepayment hazard. This is likely because high FICO borrowers are more able to refinance at attractive rates.

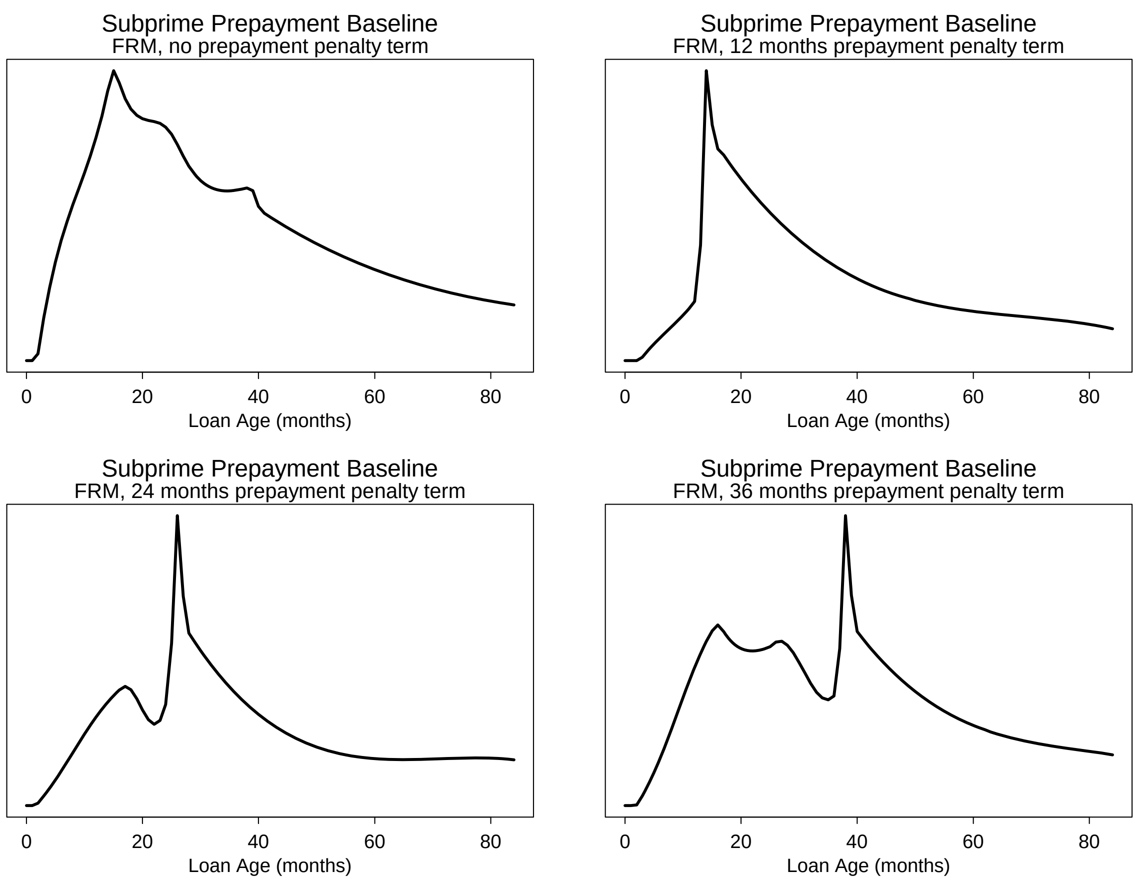
Scaled *CLTV* is a measure of the total liability of a borrower. While a substantial percentage of the prepayments result from refinancing and borrowers tend to consolidate their loans (at the least the loans on the same property), Combined *LTV* (*CLTV*) is a more intuitive measure than *LTV* and junior *LTV* separately. To capture the most updated picture, *CLTV* at origination is scaled using the MSA level (or state level if MSA level information is not available). MPA does not use outstanding balance because the junior lien balance is typically not available. The higher the scaled *CLTV*, the less equity a borrower has in his property and the less flexibility he has in refinancing or selling the property.

Model Baseline Hazard Rates

Each model has an independent baseline. The baseline describes the general trend of the prepayment rates for various loan categories, such as 2/28, fixed, and so on, purely as a function of elapsed time. The loan specific factors and elements of the simulated economic paths are overlaid on the behavioral baseline pursuant to the Cox hazard framework to determine the modeled prepayment rate paths of each loan (see Box 2).

The observed prepayment behavior differs between fixed-rate and ARM loans. To accommodate this, MPA makes use of separate models for these two broad loan types. Figure 9 and Figure 10 give some examples of the uncalibrated prepayment baselines for FRM and ARM loans. These can be interpreted as being the general shapes of the prepayment hazard rates ($h_i^p(t)$) before calibration. These baselines are scaled upward or downward depending on loan and borrower characteristics and the state of the modeled economy.

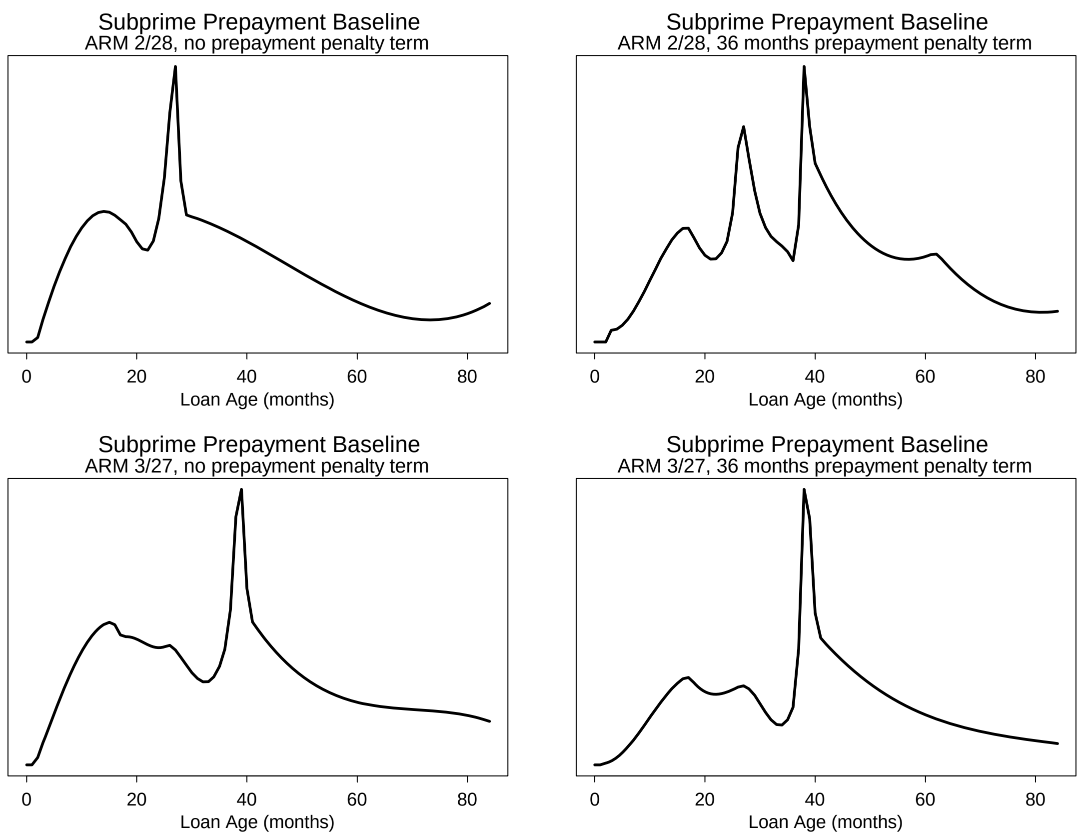
Figure 9: Subprime FRM baseline prepayment rates



As illustrated in Figure 9 and Figure 10, the ARM model was further refined by stratifying various baselines that differ by loan type within the ARM type depending on the structure (i.e. reset month). In general, there is a prepayment spike for both ARM and fixed loans after one year (resulting from the credit curing of some borrowers). For ARMs there is generally also a second spike after a loan resets. Prepayment penalties can create additional spikes.

For example, in examining 3/27 loans with 24-month prepayment penalty terms, we observe a prepayment peak around month 24 (prepayment penalty expires) followed by a pronounced increase in prepayments in year 5, when the loan rate reset. If this heterogeneity is not addressed in the model--for example, if one common baseline were used for the different loans--the peak time would have a bias that is proportional to the degree of heterogeneity in the population.

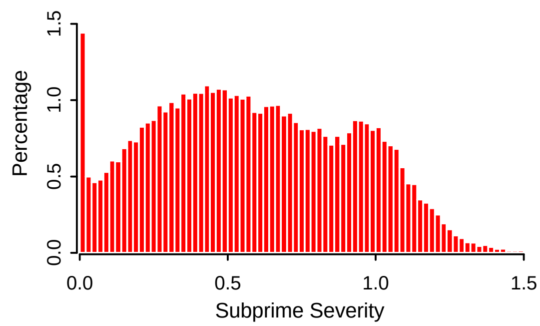
Figure 10: Subprime ARM baseline prepayment rates



2.8.3 Subprime mortgage severity model variables: LGD for subprime loans

The subprime severity model takes the same form as the prime severity model but uses different factors and coefficients. The distribution of subprime severity in our sample, as shown in Figure 11, is consistent with the skewed shape characteristic of the Beta distribution.

Figure 11: Distribution of loan severity for uninsured subprime loans



The key factors for predicting subprime severity are:

- **Judicial regulation:** This factor describes whether or not a loan is in a U.S. state that has judicial regulations relating to loan foreclosure and default (these make it harder to liquidate a property). This delay has a direct impact on the losses since the longer the process to liquidation, the more cost is likely to be incurred.
- **Forward loan to value ratio:** Loan to value ratio is updated using the simulated value of the house as of the estimated liquidation date, which is assumed to be eighteen months from the time of default if the loan is in a state with judicial regulations on loan foreclosure, and twelve months after default otherwise.[14] Empirically, forward *LTV* appears to be more predictive than the updated *LTV* at the time of default. This is because liquidation takes time, and the house value can drop significantly between the time of default and the time when the house is actually sold as part of the liquidation process.
- **Loan amount:** Smaller loans tend to suffer a higher percentage loss, presumably due to the fixed cost associated with the recovery process. This is consistent with economic reasoning and some recent academic literature. Furthermore, our research and discussions with mortgage servicers suggest that they also recognize this effect and expend greater effort to recover losses on large loans than on smaller loans. In the limit, if the property value is less than the fixed cost of recovery, there is little incentive for the servicer to pursue the recovery.

[14] The prediction for the value of the house in the future is done using our MSA-level model for house price changes. Section 2.10.2 provides more details on this.

- **Coupon rate at the time of default:** A higher coupon rate is associated with lower principal payment (as a percentage of the fixed payments) and higher relative accruing loan coupon payments; it may also indicate lower borrower quality (see Section 2.7.1 for a discussion on mortgage premium, but in the case of severity, coupon rate has a more direct relation with loss than mortgage premium).

- **Lending standard at origination:** Loans originated under stricter underwriting standard tend to have lower losses even if they default. MPA includes a custom series calculated from government statistics to proxy for the underwriting quality of typical loans during different origination periods. (See Das and Stein, 2009 for a fuller discussion).

A complete list of the factors in the severity model is given in Table 2.

2.9 Treatment of mortgage insurance (MI)

In addition to considering raw determinants of severity, MPA also accounts for the impact of mortgage insurance on loss given default. Mortgage insurance is a financial contract that pays beneficiary (a mortgage lender or an RMBS trust) a contracted amount when an insured borrower defaults on a mortgage loan. The purchaser of mortgage insurance typically makes periodic premium payments to the insurer in exchange for this protection. The presence of MI reduces the severity (and hence the losses) for the portfolio without affecting the incidence of default and prepayment.

We model two broad types of mortgage insurance:

1. *Primary* or *loan-level* mortgage insurance which covers a portion of the loss incurred on an *individual loan*; and
2. P*ool-level* mortgage insurance which covers a portion of the losses incurred by a *pool of mortgages* (protecting RMBS transactions is the most common use pool-level mortgage insurance).

In order to explain MPA's implementation of MI, it is useful to first illustrate how MI works by describing a simple mortgage insurance contract (the exact terms of specific contracts vary considerably).

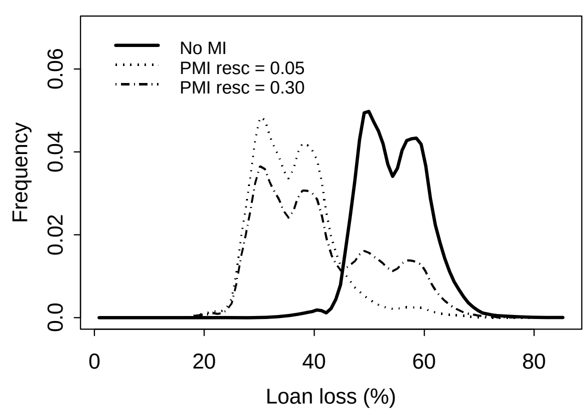
Consider a primary MI policy with a coverage level of 30%. When a loan defaults, the insurer pays out an amount equal to the *gross loss* times the *coverage*. The gross loss is calculated as the sum of the unpaid principal balance at the time of default, any unpaid interest and some additional costs. Even though the reimbursed amount is typically equal to the gross loss times the coverage level, if the property is sold before the claim is paid, the reimbursed amount is capped at the realized loss, which is the loss net of proceeds from the sale of the property.

For example, suppose a borrower with MI coverage of 30% defaults and the gross loss is $200,000. The insurer would pay $200,000 x 30% = $60,000. However, if the house were sold for $150,000 before the claim is paid, the net loss would be $50,000 (200,000 – 150,000 = 50,000) and the insurer would only pay this smaller amount ($50,000) rather than the full $60,000.

Mortgage insurance policies may also be *terminated* for a variety reasons, including the expiration of the term of the policy, the passage of half of the amortization period of the loan (e.g., for 30-year fixed loan, if 15 years have elapsed), and the reduction of the outstanding balance below a certain limit. Additionally, claims may be *rescinded* due to fraud or bad servicing. MPA permits user-defined non-payment probabilities (at the loan or portfolio-level) and randomly rescind claims during the simulation with that probability.[15]

It turns out that though the behavior of loan-level MI is fairly mechanical, its impact on the loss distribution for a loan can be complex and non-linear as it depends on not only the loan amount, but the timing of the default and the state of the economy. As an example, consider Figure 12, which shows the impact of primary MI (PMI) on a loan under different rescission rate assumptions. In this example, the loan is a $250,000, ARM loan with an *LTV*

Figure 12: An example of a loan with and without primary loan insurance (PMI) and with different rescission rate assumptions



[15] Note that non-payment may also occur due to the default of a mortgage insurer. The non-payment probability in MPA is designed to capture the probability of rescission and counterparty default.

of 90. It is to a borrower with a 605 FICO on a California property with partial documentation. (This example shows only the non-zero loss cases.)

Several things are notable in this example. First, as expected, one impact of the MI is to shift the mean loss to the left (lower) relative to the uninsured loan. However, note that the MI also changes the shape of the distribution. Thus, a simple shift in mean would not capture the richness of the dynamics of the MI. In this case the right tail of the distribution is thickened and lengthened as the distribution is shifted. This is due in part to the incidence of rescission and the effect is more pronounced for the case with higher rescission rate.

The dynamics of MI can become more involved when pool MI is also present. In contrast to primary MI which applies to a specific loan, as described above, pool MI applies to the entire pool balance and covers losses incurred after primary MI, if any, has been applied. There is usually a pool-level *deductible* which losses must exceed before the first dollar of insurance is paid. In addition, there are limits to the losses covered for any one loan and for the total pool. As with primary MI, claims may be rescinded due to fraud or bad servicing and they may be rescinded simply because the primary MI claim was rescinded.

Pool MI adds considerable complexity to the modeling of severity. When calculating the reimbursement to be paid by the pool MI policy for a given defaulted loan, it is necessary to evaluate whether the policy's pool-level deductible has been met yet, and if so, whether the pool-level loss limit has been met yet. Thus it is only possible to calculate the reimbursement (and hence the severity) for the given loan if with reference to the loss and reimbursement status of all the other loans in the pool up to the point in time that the claim for this loan is submitted. Few other aspects of loan behavior create this type of interdependence among the loans.

We note that at this time, we do not model default risk of the insurer, which would require consideration of the entire portfolio of the mortgage insurer, since the default risk of the insurer is correlated with the state of the economy. For example, an economy that results in a large drop in home prices will produce a large number of mortgage defaults in an insurer's portfolio and could increase the default risk of the insurer. However, users are able to generalize the rescission parameters (at a loan- or portfolio-level) to incorporate counterparty default risk.

2.10 Econometric models of the state of the economy

The key economic processes that are simulated in Mortgage Portfolio Analyzer are:

- **Interest rates** (10-year CMT & 6-month LIBOR)
- **Home Price Appreciation** (national, state, and MSA level)
- **Unemployment rates** (national, state, and MSA level)
- **Freddie Mac (FHLMC) mortgage rate (for the prime module)**
- **Subprime median market rate (for the subprime module)**

Briefly, Auto Regressive (AR(2)) processes are used to model changes in the unemployment rate and the log of the home price index at the national level. Subsequently the unemployment rate and home price index at the state and MSA level are modeled using the results at the national level, plus their own lags. These macro factors are correlated through common dependence on interest rates and, in the case of the local economic factors, on the national levels of unemployment and home prices, respectively. The risk-free rate is modeled through a two-factor CIR model.

The simulated interest rate, unemployment and home price movements serve as key inputs in determining the probabilities of a loan defaulting, prepaying or staying active in any month. Our simulation framework captures not only the evolution of interest rates, unemployment, and real estate market movements through time, but also the correlations of these movements across geographic regions to accurately reflect changes in mortgage portfolio credit quality.

We discuss the approach to modeling each macro-economic factor, below.

2.10.1 Interest rates (CMT 10 year & LIBOR Six-months)

The models produce estimates for the full term-structure of US Treasury rates as well as the six-month LIBOR rate. We take six-month LIBOR rate as a proxy for the reference rate for ARM coupon payments. Our approach to estimating LIBOR is to estimate the full term structure of U.S. Treasury rates and then extend this framework to model the 6-month LIBOR.

MPA uses a two-factor Cox-Ingersoll-Ross (1985) model (CIR) of the term structure of interest rates. This model has two desirable features. First the model incorporates mean-reversion, which reduces the likelihood of simulating unrealistically high levels of interest rate. Second, unlike some other interest rate models, this model ensures that the simulated rates are always non-negative.

While the CIR model has been well studied and written about extensively, for convenience we summarize the model here. The instantaneous interest rate, r is modeled as the sum of two vectors, namely

$$r(t) = x_1(t) + x_2(t)$$

The state vector follows CIR dynamics given by

$$dx_i = \kappa_i(\theta_i - x_i)dt + \sigma_i\sqrt{x_i}dW_i, \quad i = 1,2$$

where $x_1(t)$ and $x_2(t)$ are the (unobserved) state variables, θ_1 and θ_2 are the long term mean values of these two state variables, κ_1 and κ_2 are mean reversion rates, σ_1 and σ_2 are the volatilities, and dW_1 and dW_2 are Brownian motion increments. Box 3 describes how the estimated model is used in simulation.

Estimating the parameters of the CIR model is challenging because the state variables that govern the dynamics of the term structure are unobservable. We perform our estimation by representing the model in state-space. The parameters of the CIR model can be estimated using Maximum Likelihood Estimation (MLE) with UKF (Unscented Kalman filter) techniques to [Bolder (2001)]. MPA then uses the calibrated CIR model to simulate the term structure. Using the simulated term structure, MPA then calculates the LIBOR short rate by simulating the LIBOR-treasury spread.

Ongoing research continues to explore whether alternate term-structure models might further improve model performance.

Box 3 The Simulation of interest rates

Simulation of a standard CIR model involves the following steps:

1. generate two random shocks dW_1 and dW_2;
2. calculate the change in the two model factors based on the random shocks;
3. calculate the term structure of interest rates based on the changes in the factors;
4. calculate the LIBOR spread to Treasuries based on the changes in the factors.

2.10.2 House price change and Unemployment rate

House price change and unemployment rates vary with the local job and property markets. MPA makes use of national-, state-, and MSA-level house price change and unemployment rate models. We begin by discussing the form of the national level models and then go on to discuss the local models.

An important characteristic of U.S. national unemployment rate is that it moves asymmetrically over the business cycle. This means that it rises sharply in recession and it falls more slowly during subsequent expansion. As such, this movement is an example of a non-linear phenomenon, and this cyclical asymmetry cannot be well represented by standard linear time-series models such as those commonly used in the analysis of macroeconomic data (e.g. ARMA, ARIMA, vector auto-regressive (VAR), etc).[16] Accordingly, we model the unemployment rate through a threshold autoregressive (TAR) framework to capture the asymmetric dynamics of unemployment rate process. Home price changes, on the other hand, are modeled in a standard AR framework. Because we observe correlation between interest rates and unemployment we also include treasury rates in our specification at the national level. MPA's implementation also includes another term which tends to keep the unemployment levels within the range 4%–20%. The functional form of the model is shown below.

$$dU_{US,t} = \alpha_1 + \beta_1^{(1)} dU_{US,t-1} + \beta_2^{(1)} dU_{US,t-2} + \gamma \max\left(dU_{US,t-2} - 0.2, 0\right)$$
$$+ 1_{U_{US,t-1} \leq 4 \ OR \ U_{US,t-1} \geq 20}\left(a + bU_{US,t-1}\right) + \lambda_1 r_t + s_1 Z_t^{(1)}$$

where $dU_{US,t}$ is the first difference of U.S. level unemployment, r_t is the level of short interest rate, and $Z_t^{(1)}$ is an independent standard normal random variable.

---

[16] We note that alternative non-time-series approaches that represent the broad economy through a series of structural models may capture these effects through the interactions of many drivers of the individual macroeconomic factors. While such models can provide a much richer description of the economy as a whole, they can be complex to optimize and difficult to use in simulation due to the large number of factors and subsequent calculations they require. As our purpose is not to manipulate individual "levers" in the economy, but rather to simulate the distribution of correlated behaviors of the economy, a time series representation works well in our context. For a detailed example of a structural model of the U.S. economy, see Zandi and Pozsar (2006).

Home prices are modeled in a standard AR framework. Because empirically there is correlation between interest rates and home prices, treasury rates also enter the house price change model specification at the national level. The functional form of the model is shown below:

$$d\log H_{US,t} = \alpha_2 + \beta_1^{(2)} d\log H_{US,t-1} + \beta_2^{(2)} d\log H_{US,t-2} + \lambda_2 r_t + s_2 Z_t^{(2)}$$

where $d\log H_{US,t}$ is the first difference of the log of the US Home Price Index and $Z_t^{(2)}$ is an independent standard normal random variable.

The state and MSA-level models draw on the national-level effects. The simulator generates 51 state-level and 366 MSA level unemployment paths and 51 state-level and 366 MSA level home price index paths. Our research suggests that although the state unemployment rate and home price growth are local, they are influenced by national trends. The model projection of state unemployment rate is AR(2) and also is based on the national unemployment rate.

$$dU_{Region,t} = \hat{\alpha}_1 + \hat{\beta}_1^{(1)} dU_{US,t-1} + \hat{\beta}_2^{(1)} dU_{Region,t-2} + \hat{\gamma}_U \max\left(dU_{Region,t-2} - 0.2, 0\right) + 1_{U_{Region,t-1} \leq 4 \ OR \ U_{Region,t-1} \geq 20}\left(a + bU_{Region,t-1}\right) + \hat{s}_1 \hat{Z}_t^{(1)}$$

where $U_{Region,t}$ is the unemployment rate in the region (state/MSA) at time t. Similarly, the log growth rate of the state or MSA house price is an AR(2) process based on the national home price change rate.

$$d\log H_{Region,t} = \hat{\alpha}_2 + \hat{\beta}_1^{(2)} d\log H_{Region,t-1} + \hat{\beta}_2^{(2)} d\log H_{Region,t-2} + \hat{s}_2 \hat{Z}_t^{(2)}$$

Where, $H_{Region,t}$ is the HPI in the region at time t.

Box 4 describes how these models are used in simulation. Note that MPA does not model correlation between home price changes in different states or MSAs explicitly through, say, a correlation matrix. Correlation enters into the models due to their common dependence on the national level.

Box 4: Simulating HPA and Unemployment

Simulating an economic scenario for home price changes involves the following steps:

1. Simulate national level HPA using:
   a. the simulated value of interest rates (from Box 3)
   b. lagged values of national HPA (simulated or observed, see section 2.14.2)
   c. a random shock;

2. For each state or MSA (local region)

   a. Simulating local region home price changes using:
      i. the previous values of the local region changes (simulated or observed, see section 2.14.2))
      ii. the national level of HPA (from step 1, above)
      iii. the interest rate
      iv. a random shock.

2.10.3 Freddie Mac (FHLMC) mortgage rate and subprime median market rate

Calculation of the spread-related factors in MPA's models requires knowledge of the prime mortgage market rate, $F(t)$. MPA uses the Freddie Mac 30 year commitment rate as a proxy for the prime market rate. This rate is modeled as:

$$dF_t = b_1 dCMT10_t + b_2 ds_t + b_3 dY_t + b_4 dF_{t-1} + b_5 dF_{t-2} + \sigma_f Z_t^F$$

where CMT10 is the 10 year treasury rate, s_t is the spread between the 6 month LIBOR rate and the 6 month treasury rate, Y_t is the difference between the 10 year treasury rate and the 6 month treasury rate, which is a measure of the slope of the yield curve, and dF_{t-1} and dF_{t-2} are the lagged changes in the mortgage rate, F_t.

In contrast to the Freddie Mac rate, there is no published source for a subprime market rate. Therefore, MPA's model is based on the median origination rate of the fixed loans in the subprime data set as a measure of the subprime market rate.

2.11 *Incorporating delinquencies and realized pool performance*

Mortgage Portfolio Analyzer can be used to analyze the loss distribution of both new portfolios and seasoned pools for which historical performance information is available. For a seasoned pool, the losses consist of two components – those that have already been realized since the pool's inception or closing date (historical), and those that are projected from today (simulation). While there is only one historical realization up through the current period, by default MPA generates 10,000 economic simulations for periods after the current period to arrive at a distribution of future or projected losses.

Seasoned pools offer additional information for predicting lifetime losses. As a pool seasons, additional information becomes available:

- Updated value of the loan balance, including any curtailment;
- Up-to-date status (Current, 30DPD, 60DPD, 90DPD, Prepaid, Defaulted) for each loan in the pool; and
- For defaulted loans, the actual loss in cases where the losses have been realized.

2.11.1 Differentiating the performance of delinquent and current loans

The hazard rates for prepayment and default models described in Section 2.5 are estimated using a sample consisting of current and delinquent loans. However, for seasoned loans, MPA takes advantage of information about the loans' payment statuses, and this information helps to predict future default or prepayment. For example, a loan that is current is less likely to default relative to the general population, which includes both, current and delinquent loans; whereas a loan that is 90 days past due is more likely to default relative to the general population.

Accordingly, for seasoned loans, instead of using the sample wide (unconditional) hazard rate baselines to predict default and prepayment MPA uses a different baseline hazard rate for each delinquency status. These delinquency-specific baseline hazard rates are scaled versions of the sample-wide hazard rates.

2.11.2 Mid-course adjustments for unusual portfolios

MPA can use performance information to adjust for differences across portfolios in order to make the portfolio behavior more realistic for unusual portfolios. It is useful to consider why the realized loan performance might be different than the model's prediction. The loan-level models are developed based on loan and borrower attributes observable in the data set. However, there are also unobservable attributes that may drive performance but that cannot be included in the model (e.g., a particular new loan origination program may have targeted certain borrower types that turned out to be more or less risky than historical borrowers.) Because MPA was developed on a cross section of mortgages and borrowers, in most cases the unobserved attributes of a particular portfolio are similar to those that were captured in the estimation, and the prediction of MPA's models will be an unbiased estimate of a portfolio's performance. However, if the loans in a given portfolio were selected according to some criteria that are unobservable in the data, but different from the general population, MPA's predictions could be systematically biased for that special portfolio. For example, a pool of loans made to high net worth individuals might have a lower default rate than MPA would predict because of an unobservable (in the data) attribute, namely, the total wealth of the borrowers.

To adjust for this, MPA uses the differences in predicted versus actual loan performance as of the current date to calibrate its models in real time to account for portfolio-specific default, prepayment and severity drivers that are not observable. Specifically, after selecting an appropriate baseline for a loan, MPA can scale this baseline to account for differences in predicted versus observed performance to date. The scaling factor is chosen so that the predicted 5 year default and prepayment rates for a group of loans of the same delinquency status match closely their realized counterparts.

MPA estimates the scaling parameters as follows: it first predicts loan performance as of the current date for the current loans using the realized economy from loan origination to the present date as the input for the macro factors for the model. It then calculates the difference between the predicted and actual realized values and scales the baseline hazard rate to reduce the difference. MPA performs this process iteratively, adjusting the scaling factor until it finds a value that sets the predicted rate equal to the realized rate. The scaling scheme is designed to naturally bound the default and prepayment rates between 0 and 1.

When analyzing a seasoned portfolio, MPA uses the realized defaults, prepayments and actual losses, if available, to calculate the performance of the portfolio to-date, and only uses simulations to predict the performance of the portfolio going forward for those loans that have not already prepaid or defaulted.

Severity is handled slightly differently, since actual losses (severities) are often realized 12 to 18 months after default. In some cases, trustees may never report the actual losses. Provided a sufficient number of realized losses are available, MPA can incorporate that information into its severity models as follow: For each defaulted loan for which a realized loss is available, it calculates the expected severity. It then calculates an average realized severity over all loans for which realized loss is available. MPA compares this number to the expected severity computed by the model and then scales the model severity appropriately so that the scaled average severity matches the realized average severity. MPA uses this adjustment in the severity model, going forward.

The midcourse adjustment can be disabled at the user's discretion.

2.12 Enhancements based on expert judgment

Ideally, historical data used for calibrating the econometric models would be both fully descriptive of the underlying process and complete. In reality, this is seldom the case. The pace of innovation in the mortgage market has been rapid. New mortgage products, new underwriting practices and new approval technology all suggest behaviors that may not be fully captured in the historical data. This is true despite the recent market activity which has generated a wealth of new observations not previously available in historical data sets.

Because of this limitation, we augmented our statistical modeling with detailed discussions with RMBS analysts and other market participants. The goal is to ensure that MPA captures newer developments in the market as they manifest themselves in mortgage behavior particularly if historical data are not sufficient to model these phenomena.

The objective in incorporating qualitative analysis is to augment the models rather than to introduce arbitrary behavior. When there are neither sufficient data nor standardized reporting to permit quantitative estimation, we incorporate appropriate expert judgment to improve the models. In general, these adjustments serve to "fill in the gaps" in the relationships in our statistical analysis. These gaps are typically due to either the newness of some aspect of a market or economic states of the world that have not yet manifested themselves in the historical record.

We try to make these adjustments to our models in ways that are consistent with the modeling framework. For example, when we introduce augmented data (certain property types, certain documentation types, etc.) for which historical data are generally unavailable, we do so by introducing adjustments to our default model directly, rather than adding "hits" to expected loss. Furthermore, upon augmenting the models to reflect a specific characteristic of the market, we perform extensive validation to ensure that the results generated by the model following the enhancements are intuitive.

We use such adjustments sparingly, but include them when they are needed. Some examples of these adjustments are:

- The introduction of additional documentation (doc) type codes to permit better granularity for doc type;
- Sensitivity of models to unemployment levels outside of the range of the historical data; and
- Frailty parameters.

In addition, due to the difficulty in calibrating a model to properly predict extreme events, many of which are not present in the historical data, we also use expert judgment from RMBS analysts in calibrating the extreme loss levels predicted by the model.

2.13 Frailty

While any good model will endeavor to capture known risk factors through either statistical estimation or expert judgment, it is inevitable that some unknown factors and some known but impractical to quantify factors can influence default behavior. A recent stream of research in the corporate default literature (cf, Duffie, et al., 2009) has begun to explore models that expressly accommodate such latent *frailty* phenomena.[17]

Frailty need not relate only to macro-economic shocks. As a hypothetical example, consider how the widespread adoption of a defective fraud-detection system for mortgage underwriting might impact default rates. After adoption, many of the underwriters using the system may experience higher default rates as a result: they will experience correlated defaults that are explained by the common use of this system. A more realistic example would be the impact of a common

[17] The term *frailty* has been adopted from the health sciences literature where similar approaches are used to examine patient mortality.

underwriting practice among lenders. While such factors clearly induce correlation among mortgage defaults, it is unlikely that any model would include these factors, *a priori*, and given a very large number of such factors, the impact of which would only be realized *ex post*.

The estimation techniques for frailty models are involved and typically require long time series. As a first step in including frailty, MPA makes use of a simpler framework, similar to many parametric frameworks for imposing correlated defaults. In order to parameterize the model, we used a combination of expert judgment and empirical analysis.

The frailty component of the model works as follows: For each loan *i* the default model predicts a default probability, d_i. A correlation between the default events of different loans is introduced by the generation of a series of correlated Uniform [0,1] variables, X_i, for each loan (see equation below). Default occurs for loan *i* if X_i is less than the loan's default probability, d_i, where
$X_i = N\left(\sqrt{\rho}s + \sqrt{1-\rho}\varepsilon_i\right)$, $N(.)$ is the cumulative normal distribution, ρ is the correlation, and s and ε_i are independent standard normal random variables.

Thus, in addition to the correlation in default rates induced by the dependence of default rates on common interest rates and local macro-economic factors (themselves correlated with the national level), the default events of all loans are also correlated through their dependence on the latent frailty factor.

As noted, the estimation techniques for formal frailty models are involved and typically require long time series. Our implementation uses a simpler framework, and is similar to many parametric frameworks used for modeling correlated defaults. We estimate our frailty model, using a combination of expert judgment and empirical analysis.

Also note that the frailty framework is not the primary means of modeling correlation in MPA. Rather, frailty is used to *augment* the main correlation framework (see Section 2.10) which induces correlated behavior from common dependence on macro-economic factors.

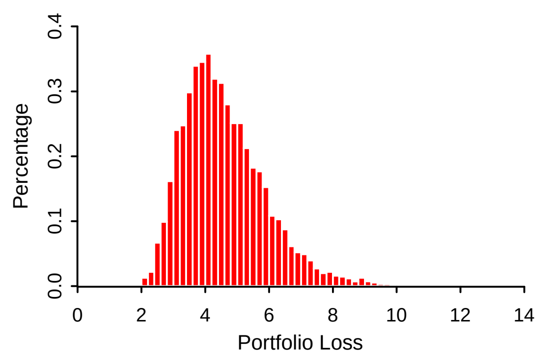
2.14 Estimating a Loss Distribution using Monte Carlo simulation

In this section, we describe the simulation module that Mortgage Portfolio Analyzer uses to generate the states of the economy used by the loan-level models in estimating default probably, prepayment probability, and severity. The engine uses a full multi-step Monte Carlo simulation to estimate the collateral loss distribution.

2.14.1 Estimation of the loss distribution

MPA estimates the collateral loss distribution in the following way: it simulates macro-economic variables for 10,000 different economic scenarios (by default), quarterly, over a ten year horizon. The quarterly macro-economic variables are interpolated to obtain monthly values. For each simulated economy, MPA determines if and when a loan defaults or prepays and the loss incurred on the loan if it defaults. For each economy, MPA adds the dollar amounts of the losses on all the loans in the portfolio to arrive at the dollar amount of loss for the portfolio. This quantity is divided by the notional balance of the pool to obtain the fractional loss (henceforth, referred to as the loss) for the portfolio. When this process is repeated for each of the 10,000 simulated economies, the result is a distribution of these 10,000 losses and, thus, the empirical loss distribution. This distribution for one portfolio is shown in Figure 13.

Figure 13: Portfolio loss distribution



2.14.2 The use of lagged values during simulation

Some of the macro-factor models use lagged values of the same or other variables as explanatory variables. For example, our current value of Home Price Change (*HPC*) is a function of the two previous lagged values of the variable. When lagged values are required to simulate the current value of a macro factor, MPA uses observed data (i.e., actual past *HPC*, interest rates and unemployment) when it is available. Since the models only use two lags, actual data can only be used, at most, for the first two quarters of the simulation. To simulate the macro variables further out into the future, MPA uses simulated values from previous time steps as the lagged values. We illustrate how we use actual versus simulated values in the simulation by considering how

HPC is calculated for the first four quarters of the simulation. Recall that the current value of *HPC* is modeled as a function of its two previous lagged values, in addition to other factors. Assume that the simulation is starting 2010Q3 and that all historical values of *HPC* up to and including 2010Q3 are available.

Table 3 shows how the simulation would proceed along a particular path. We denote the simulated values with a superscript "s" (^{s}). For the first quarter forward, MPA runs the simulation and computes *HPC* 2010Q4^s. Since *HPC* depends on its own two lags, MPA uses the actual *HPC* values for 2010Q3 and 2010Q2 to compute 2010Q4. Similarly to compute *HPC* 2011Q1^s, MPA uses the actual *HPC* value of 2010Q3 and simulated *HPC* value 2010Q4^s.

Table 3: Example of HPC simulation

Simulation period	Lag t-2	Lag t-1	Scenario value t
1 – 2010Q4	*HPC* 2010Q2	*HPC* 2010Q3	*HPC* 2010Q4^s
2 – 2011Q1	*HPC* 2010Q3	*HPC* 2010Q4^s	*HPC* 2011Q1^s
3 – 2011Q2	*HPC* 2010Q4^s	*HPC* 2011Q1^s	*HPC* 2011Q2^s
4 – 2011Q3	*HPC* 2011Q1^s	*HPC* 2011Q2^s	*HPC* 2011Q3^s

2.15 Analyzing the Loss Distribution: VaR, tranches and Expected Shortfall[18]

In addition to determining the expected loss for the entire mortgage portfolio, it is sometimes useful to consider the expected loss above or below a specific point on a distribution. In traditional risk management, these portions are commonly characterized as the *value at risk* (VaR) and *expected shortfall*. In a structured finance setting, variations on these quantities are typically thought of as *tranches*.

In a structured transaction, tranche defines an interval of the loss distribution bounded on the bottom by an *attachment point* and above by a *detachment point*. A specific tranche suffers no losses if the portfolio losses are less than the attachment point for the tranche. The tranche is, however, exposed to all portfolio losses beyond the level of the attachment point.

[18] This section is adapted from Das and Stein (2011) which also contains more detailed analysis of the properties of different tranching approaches as well as additional mathematical results.

Since the loss distribution describes the probability associated with each level of loss on the portfolio, the expected loss for a particular tranche is readily calculated as the probability weighted percentage losses to the tranche.

Agency ratings are also sometimes characterized using (idealized) expected losses or expected default rates, and in such settings it is convenient to compute the rating category associated with various points on the loss distribution.[19] These provide information about the tail risk in the mortgage portfolio or, in the case of an RMBS transaction, the tranching of the collateral pool.

When VaR is calculated, or in the case of calculating tranches based on a target default rate, the attachment point is simply the quantile of the distribution associated with the target default rate or the VaR. In the case of expected losses, the calculations are a bit more involved but the logic is similar. We describe these calculations in greater detail below.

Importantly, in the case of RMBS transactions, the expected loss on a tranche is typically a function not only of losses on the underlying portfolio of mortgages, but also of the manner in which the various mortgage cash flows (e.g., interest spread, recoveries, etc.) are apportioned in the transaction as prescribed by the cash flow waterfall. The analysis below assumes a simple cash flow waterfall where losses are apportioned strictly by tranche seniority and no other sources of credit support are available. While this analysis is suggestive of loss levels on tranches in general, it applies most closely to sequential pay structure or synthetic structure. (For cashflow transactions, various triggers, the use of excess spread and so forth can materially change the risk of specific tranches.)

In what follows, we discuss the calculation of VaR and Expected Shortfall useful for risk management of credit portfolios. For users interested in RMBS analysis, we also generalize these measures by recognizing that VaR is simply the definition of a single tranche using a target probability of default implying similar mathematics can be used for PD-based tranching. We similarly recognize that Expected Shortfall is a special case of EL-based tranching.
For readers only interested in the risk management applications of VaR and Expected Shortfall, Sections 2.15.1 through 2.15.3 describes the calculations. These sections, as well as 2.15.4 outline the calculations required to extend VaR and Expected Shortfall to tranching analysis of RMBS.

[19] Note that these are quantitative mappings of the agencies' published targets to portfolio losses and are *not* equivalent to agency ratings.

2.15.1 Notation

We use the following notation for our derivations:

L is a portfolio loss level
$F_L(L)$ is the cumulative distribution function (*CDF*) of L
$f_L(L)$ is the corresponding probability density function (*pdf*).

2.15.2 VaR and tranching based on Probability of Default

We begin by defining the probability of losses exceeding a given level of the portfolio loss distribution. If we choose a point, A, on the loss distribution (e.g., we might choose A to be 25% of the portfolio par), then under the conditions we describe at the beginning of this section the probability that losses are greater than losses at point A is simply the compliment of the cumulative probability distribution to A. We can call this probability *p(A)* which is defined as

$$p(A) = \int_A^1 f_L(L)dL = 1 - F_L(L).$$

Now we can ask the question: given a target probability, α and the loss distribution, what point A ensures that the probability of losses exceeding A (*pd(A)*) is less than or equal to α. This is the portfolio's 1-α VaR or $\mathrm{VaR}_{1\text{-}\alpha}$ level.

In the context of RMBS, it is also the tranche attachment point that results in the tranche having probability of default, *PD*, of α.

Given a target VaR probability of 1-*PD* or a target tranche default rate of *PD*, it is straightforward to determine $\mathrm{VaR}_{1\text{-}\alpha}$ level or the attachment point. From the formula for *p(A)*, we can see that the level can be obtained trivially as

$$VaR_{1-PD} = \text{Tranche attachment}(PD) = \min\{A|(1 - F_L(A)) \leq PD\}$$

In a ratings context, if the rating system were based on idealized PDs, the tranching would be done by setting various levels of PD to the idealized PDs.

2.15.3 Expected Shortfall

We now define the expected loss in the tail of the distribution beyond a point A. We note that whenever the pool experiences a loss (L) less than A, the tail experiences zero loss. Whenever the pool experiences a loss greater than A, the loss rate increases linearly from 0 at A to 100% (the portfolio is completely wiped out). For this region of portfolio loss, we get:

$$\text{Tranche loss rate} = \frac{L - A}{1 - A}$$

In practice, the exact distribution of the losses is not known. However, Mortgage Portfolio Analyzer provides a simulation-based estimate of the loss distribution which can be used to compute tail ELs.

The expected loss of the tail is often called the *Expected Shortfall* and is expressed with reference to a specific VaR level. Expected Shortfall can be expressed using the pdf of the loss distribution and the tail loss as:

$$\text{Expected Shortfall=Tail } EL = \int_A^1 \min\left(\frac{L - A}{1 - A}, 1\right) f_L(L) dL$$

$$= \int_A^1 \min\left(\frac{L - VaR_{1-PD}}{1 - VaR_{1-PD}}, 1\right) f_L(L) dL$$

The expected loss given default (LGD) in the tail is the ratio of the tail EL and the tail PD and can be written as

$$\text{Tail Expected } LGD = \frac{\text{Expected Shortfall}}{PD}$$

2.15.4 Expected Loss-based Tranching for RMBS

The expected loss of a *tranche* is defined by generalizing the definition of the expected loss in the tail in Section 2.15.2. Rather than assuming a single attachment point that defines the *tail*, we now consider an arbitrary *tranche*, anywhere in the capital structure, defined by an attachment point A and a detachment point D. Such would be the case for a very basic pass through RMBS transaction supported by a collateral pool of mortgages.

Whenever the pool experiences a loss (L) less than A, the tranche experiences zero loss. Whenever the pool experiences a loss greater than D, the tranche is completely wiped out, i.e., it experiences its maximum loss, $D - A$. When the collateral losses are greater than A but less than D, the tranche loss rate increases linearly from 0 at A to 100% at D. For this region of collateral loss, we get the tranche loss rate which is a generalization of the tail loss rate from Section 2.15.2:

$$\text{Tranche loss rate} = \frac{L - A}{D - A}$$

Similarly, by analogy to the expression in Section 2.15.2 for the expected loss in the tail, the expected loss of a tranche can be defined as:

$$Tranche\ EL = \int_A^1 \min\left(\frac{L - A}{D - A}, 1\right) f_L(L) dL$$

Now, given a target *EL* for a specific credit level, it is possible to infer the implied credit assessment for any tranche in a capital structure, based on the *EL* for the tranche by looking up the *EL* in a table that maps *EL* to credit levels. As an example, Table 4 below shows Moody's Investors Service's idealized *EL* values for different credit ratings, based on 10-year horizon, though any set of *EL* targets could be used.

Table 4: Moody's idealized (target) Expected Losses by rating

Rating	EL (%)
Aaa	0.0055
Aa1	0.055
Aa2	0.11
Aa3	0.22
A1	0.385
A2	0.66
A3	0.99
Baa1	1.43
Baa2	1.98
Baa3	3.355
Ba1	5.17
Ba2	7.425
Ba3	9.713

The minimum tranche credit enhancements needed to attain a specific rating grade can be calculated numerically.[20] The process starts from the highest rated tranche and proceeds to the lowest rated tranche as follows. The first tranche in our example is the **Aaa**. The detachment point for **Aaa** is 100%, and the idealized *EL* is 0.55 bps (Table 4). Therefore, the above formula for tranche *EL* becomes

$$Aaa_{CE} = \min\left\{A | \left[\int_A^1 \left(\frac{L-A}{1-A}\right) f_L(L)dL \leq 0.000055\right]\right\}$$

Although this equation has no closed-form solution, we can solve iteratively for the lowest value of A by reducing A from 1 until the above inequality is violated. The value of A which solves the above equation is the **Aaa** credit enhancement, Aaa_{CE}, consistent with the *EL* target.

The next tranche in this example is the Aa1. The detachment point for **Aa1** is equal to the **Aaa** attachment point, and its idealized *EL* is 5.5 bps. So, we can again obtain the **Aa1** attachment point by solving for value of B that solves the expression below for the **Aa1** tranche

$$Aa1_{CE} = \min\left\{B | \left[\int_B^1 \min\left(\frac{L-B}{Aaa_{CE}-B}, 1\right) f_L(L)dL \leq 0.00055\right]\right\}$$

This process is repeated until all tranche levels are determined.

Mortgage Portfolio Analyzer allows the user to determine the attachment and detachment points of tranches such that they can be assigned a given *PD* or *EL*. The analysis can be based on the tranche *EL* or they can be based on the tranche *PD*. MPA allows the user to enter values for the target tranche *EL* and the target tranche *PD*. These two methods of tranching are discussed in the sections below. Alternatively, for a fixed set of attachment and detachment points, MPA can calculate the *PD* and *EL* of each tranche.

Some properties of the losses experienced by tranches are worth noting. The same attachment point can result in quite different losses on a tranche depending on the size of the tranche (the location of the detachment point). For this reason, “thin tranches” tend to be much riskier than thicker tranches when both have the same level of subordination (the same attachment point). In general, *EL* can only be reduced by raising the attachment or detachment points (or both). Thickening a tranche by lowering the attachment point generally increases both the *EL* and *PD*.

[20] Note that these level approximations are not equivalent to ratings for a variety of reasons.

Also, it is not guaranteed that an attachment and detachment point can be found for target *EL*. In some cases, the shape of the distribution is such that some *EL*s cannot be achieved through subordination without reducing the size of (or eliminating) tranches above. (See Das and Stein, 2011)

2.16 Tail risk contribution (TRC)

A common task for credit portfolio managers is to assess the economic capital required to support a loan portfolio. Section 2.15 described measures of capital sufficiency, VaR and Expected Shortfall.

As economic capital is a constraint on lending (capital that is reserved to support a portfolio cannot be lent out or invested in other businesses), it is incumbent on the portfolio manager to find ways to minimize economic capital utilization. Hedging, loan sales or purchases and securitization are all mechanisms for reducing capital utilization. A key challenge in implementing such strategies is determining *which* loans to hedge, securitize or trade.

This is challenging since capital consumption occurs in the tail of the distribution and depends not only on the *EL* of a particular loan, but also on how highly correlated that loan is with the rest of the portfolio. It is not always the case that the riskiest few loans in the portfolio (high *EL*) contribute the most to the tail risk. Often, those scenarios that result in a high portfolio loss may result in very low individual losses on the riskiest loans. For example, if the riskiest loans were in Florida and the portfolio were concentrated in New York, then scenarios that result in the highest portfolio losses would be those that had high home price declines in New York. In many cases, these scenarios might not result in similarly large declines in Florida. In such a case, the risky Florida loans would contribute little to the tail risk of the portfolio since most of the action would occur when New York home prices declined rather than when Florida home prices did so.

To quantify each loan's contribution to the tail risk, Mortgage Portfolio Analyzer calculates the average contribution of each loan to tail losses for the portfolio.[21] These quantities are reported along with other loan-level results.

[21] This is actually done using a numerical approximation that enhances numerical stability by using information from regions of the loss distribution that neighbor the specific attachment point or VaR level.

As an example of the utility of this information, consider a whole-loan portfolio with a target *VaR* of 99.5%. The portfolio manager is given the opportunity to sell or hedge 100 mortgages in the portfolio with the objective of reducing the portfolio's *VaR*, and thus the economic capital needed to support the portfolio.

The manager evaluates two strategies. The first involves removing (either through sales or hedges) the 100 loans with the highest stand-alone risk (i.e., the highest *EL*s). The second strategy involves removing the 100 loans with the highest *TRC*. Table 5 shows a summary of each analysis with respect to both the *EL* and *TRC* of the resulting portfolio. The VaR and *TRC* of the original portfolio are provided for reference.

Table 5 *EL* and 99.5% *VaR* for different mortgage portfolio construction strategies

	EL	**99.5% *VaR* Level**
Original portfolio	4.0%	12.6%
With 100 highest *EL* loans removed	2.9%	10.2%
With 100 highest contributors to *VaR* removed	3.1%	9.7%

In examining Table 5, a number of things are notable. First, as expected, eliminating loans with high *EL*s from the portfolio produces the largest reduction in the portfolio *EL*. Secondly, eliminating the high *EL* loans also shifts the distribution of losses to the left, thus lowering the VaR number (by a bit more than the reduction in *EL*). However, note also that removing the highest *TRC* loans from the portfolio reduces the VaR further by another 50bps, thus saving a half percentage point of VaR over the *EL* reduced portfolio. Interestingly, though the portfolio *EL* is higher under the *TRC* reduction approach, it is only 30bps higher than in the *EL* reduced approach (versus the 50 bps saved in capital). Importantly, the hedges (sales) on the high *TRC* loans can likely be done at lower cost (higher prices) since the stand-alone quality of the high *TRC* loans is, on average, better than in the case of the high *EL* loans (the high *EL* loans, on average, have higher *EL* s than the high *TRC* loans.)

In addition to informing portfolio construction and asset allocation decisions, some banks use *TRC* measures as a form of incentive alignment. To see how, consider that *TRC* measures the cost, in terms of capital, to a bank of the addition of a credit risky asset to the institution's existing portfolio. Each new position changes the portfolio's risk and thus has an explicit "cost" in terms of additional capital that the institution must reserve to support the portfolio.

One mechanism that some institutions use for aligning the incentives of managers and lenders is to use the *TRC* of a loan as a form of transfer price. A lending officer can be made aware of the *TRC* of a prospective loan and this cost can be factored into the cost of lending. As concentrations in a particular asset class increase, the cost of making the next loan in that concentrated asset class increases thus providing economic incentives for diversification. A number of large institutions use this type of transfer pricing (particularly for corporate portfolios, see Bohn and Stein, 2009).

2.17 Calculations on custom economic scenarios

When performing loss simulations, Mortgage Portfolio Analyzer (by default) generates 10,000 economic paths and simulates loan defaults, *LGD* and prepayments one month at a time for each economic path. The modeling of loan behavior conditional on the economy is at the core of MPA's framework. This ability can also be extended to perform loss computations on a set of *user-specified* economic scenarios. This feature is useful for stress-testing a pool of loans under various adverse scenarios or for forecasting performance under a baseline scenario. For example, Mortgage Portfolio Analyzer could help the user evaluate the performance of a mortgage portfolio if US home prices were to decline at the rate of 2 percent per year for the next 3 years and then rise at the rate of 3 percent per year for the following 7 years.

An MPA economic scenario consists of a specification for one or more macro-economic factors that drive the default, prepayment, or severity behavior of individual loans. Mortgage Portfolio Analyzer permits the user to enter forecast values of the following six macro-economic series:

a. 6-month LIBOR rate
b. Freddie Mac 30-year commitment rate
c. US unemployment rate
d. Unemployment rate at the state and MSA level
e. US HPI
f. HPI at the state and MSA level

One feature of the custom scenario generation module is that it imputes missing economic data based on the data the user does provide. As a result a user may provide as much or as little detail as they wish when specifying a macro-economic path. For example, a user may choose to specify the US *HPC* for the next five years only. In such a case, MPA's custom scenario module will use the US *HPC* model described in Section 2.10.2 to simulate a *single path* of the national level *HPC* beyond five years (by assuming zero shocks, i.e., all random variables in the macro processes set to zero). Similarly, the user need not specify the values for all the macro inputs to MPA. In such cases, Mortgage Portfolio Analyzer uses its macro-economic model to compute a single path of values for those series that are not specified. For example, the user may choose to specify that the US unemployment rate stays at 10 percent for the first 3 years and that the US *HPC* drops by 2 percent per year for the first 3 years and increases by 3 percent for the next 7 years. MPA will calculate all other variables such as the state and MSA level *HPC* and unemployment rates using its macro-economic models, as in the previous example.

Once a single economic path is generated, MPA can determine the expected loss for the pool under that path as well as the loan-level results (for example, expected loss and default and prepayment probabilities for each loan) using the loan default, prepayment, and severity models.

2.18 Stress testing a portfolio of residential mortgages

Lately a substantial attention has been focused on the area of stress testing. A key benefit of stress testing is the ability of such analysis to provide transparency into both model behavior and portfolio risk.

MPA provides two mechanisms for stress-testing a portfolio of residential mortgages:

- The first method implements macro-economic stress testing in which a specific economic scenario (standard or user-defined) is used as input to MPA and losses and other risk measures under this scenario are calculated. For example, a user may wish to run MA's "Total Collapse Scenario" on their mortgage portfolio.

- The second method involves shocking default rates, prepayment rates and severity directly. For example, a user may wish to double default rates or reduce prepayments by 50%.

These two approaches may also be combined. For example, a user may wish to stress the default rates further under the MA's "Total Collapse Scenario." We discuss each of these approaches below.

2.18.1 Macro-economic stress testing

MPA allows a user to estimate losses under a specific macro-economic stress scenario. This is done in one of two ways. The user may either choose from a series of pre-defined stress scenarios provided by Moody's Analytics (MA), or the user may define a custom scenario either in its entirety or based on one of the pre-defined scenarios (see Section 2.17). In the first case, the loan portfolio loss estimate is run under any or all of the published MA macro economic scenarios. In the second case, the user may enter one or more customized scenarios under which the loss estimates are run. When MPA runs an economic scenario, it does not generate a full simulation. Rather it runs only the single path defined by the given scenario. As a result, some analytic outputs, such as economic sensitivities and loss distribution, are not available for that particular scenario analysis.

2.18.2 Stress testing defaults, prepayments, and severity directly

MPA provides the user with the ability to adjust default, prepayment, and severity levels across the entire portfolio. This can also be done at the individual loan level directly in the data input file. This is convenient in cases in which the user has additional information about certain loans or types of loans and believes this information may be relevant to the risk of the loans. For each loan in the portfolio, a user may input multipliers for the loan's prepayment rate, default rate and severity. The same multipliers can be used for entire segments of the portfolio that share common attributes.

Multipliers greater than 1.0 increase the quantities, while those less than 1.0 decrease them. Thus, if a user felt that all loans whose borrowers are employed in a certain industry (e.g., the auto industry) were more likely to default than otherwise identical loans whose borrowers are in a different industry, a multiplier greater than 1.0 could be used to represent this.

MPA produces a report that details the prevalence and magnitude of overrides for each run of the portfolio. This may be useful for documenting the analysis and to communicate the scope of overrides in a given analysis.

To use this functionality for stress testing a portfolio, the user can assign constant multipliers to all loans. For example, to stress test the portfolio by shocking default rates, a user might assign a multiplier of 1.5 to all loans in a portfolio. This type of stress testing may be useful when the user wishes to examine stress scenarios that might be driven by factors other than the macro-economic variables in the model or when the user does not have a strong view on the specifics of the macro-economy.

3 DATA

The development data sets used to estimate MPA's models drew on several different sources:

- A database constructed from on data files downloaded from RMBS trustees covering thousands of RMBS transactions;
- A database of individual mortgages dating back to 1998 containing loan-by-loan characteristics and performance provided by a leading RMBS originator;
- historical macroeconomic data and forecasts from Moody's Economy.com;
- lifetime loss estimates for RMBS asset pools from Moody's Investors Service's RMBS surveillance group;
- survey data published by the Federal Reserve; and
- summary statistics published by various academic and industry sources

Economic data, going back to 1980, was provided by Moody's Analytics (MA), which receives data from various sources. For example, the MA unemployment rate data are collected from the US Bureau of Labor Statistics. The home price index comes from the National Association of Realtors (NAR) and is the median home price of existing single family homes sold in each quarter. We use the unemployment rate and the home price index at the national, state, and Metropolitan Statistical Area (MSA) level.

We tested our dataset extensively for representativeness and completeness to ensure our model did not appear to be biased towards the idiosyncrasies of this particular sample. In general we found that from a statistical perspective the data are similar to data reported in other studies and to other data sets that we examined, both in terms of the frequency distributions of the independent variables (loan characteristics, etc.) and in terms of the timing, frequency and levels of the default and prepayment.

3.1 Data mapping and cleaning

In many cases, financial institutions maintain data sets to support business functions and not to develop quantitative models. As a result, data quality that is appropriate for their business applications may not be adequate for modeling. Accordingly, we perform a series of data standardization and mapping activities to address data issues.

The data-refinement process begins with a mapping of provider fields to a standard format and standard definitions. When we first receive data, either from trustees or from an originator, we work, where possible, with the providers to determine how best to map their data fields to the common definitions we use. As part of the mapping, we examine the granularity of the fields based on the providers' definitions. For example, some contributors maintain only three categories of documentation type while others have as many as twelve. In order to have a common data structure across all observations in our data set, we re-classify entries in such fields to a common number of categories that still provides sufficient discriminatory power.

Once the data have been mapped to a common format, we perform a formal data standardization and error trapping process in three steps that progressively deal with increasingly subtle data errors:

First, we eliminate obvious *typographical* errors by flagging records in which values of any field appear erroneous using a rule-base that we have created. The rule-base is tailored to each individual data contributor's database and reporting conventions. Where possible we also create rules to correct these errors. For example, a FICO score of 7200 may be converted to 720 where it is obvious that this is the intended value. However, a FICO of 23 or 2300 is erased and labeled "unknown" since it is not obvious what the corresponding correct value is. (23, 2300, and 230 are outside of the valid range of FICO.)

Next, we identify errors using *business rules* that flag values of certain fields that seem to be in conflict with basic business or accounting logic. For example, we verify that a loan's combined *LTV* (*CLTV*) is at least as great as its *LTV* and that the remaining term on a loan is no longer than its original term. Where possible, we correct errors by recalculating the incorrect fields using information from other fields in the record.

Finally, we use *statistical anomaly detection* techniques to identify potential errors in the data. We develop univariate and multivariate relationships between fields in the data set and then use these relationships to flag observations where predictions from the statistical relationships are outside of the expected range. We then investigate whether these outliers are legitimate observations or are bad data that should be removed from the sample.

In each level of refinement, we apply rules that we have developed in conjunction with data providers themselves or with our own analysts and researchers.

4 MODEL VALIDATION

Validation is an important step in the development process for a model used for commercial applications, and a significant amount of time was devoted to the MPA validation process.

The testing regimes can be separated into two broad categories:

- **Sensitivity test**s: These tests examine whether the sensitivity of losses to changes in individual factors (such as FICO or *LTV*) predicted by the model is in accord with economic intuition.

- **Prediction tests**: These tests examine whether the predicted loss performance of the model is consistent with realized values or other benchmarks. These include tests of the discriminatory power of models as well as tests of the predicted levels of default.

4.1 Sensitivity Tests

We examine the sensitivity of pool-level losses to key loan attributes and macro factors to ensure they are consistent with economic intuition.

Figure 10 shows an example of the sensitivity analysis. This figure illustrates how the pool-level expected loss varies with changes in national housing prices. As shown in the figure, losses increase as *HPC* decreases, which suggests that predictions made by MPA is consistent with economic reasoning.

The pool used for this example was a seasoned pool originated in 2006Q1. The average FICO score was 750 and the average *CLTV* was approximately 70%. The change in housing prices shown in this figure is the cumulative change over the first two years of the simulation. The sensitivity is calculated as follows:

- The simulation started at time *t* (2010Q1), with the macro factors and loan-specific attributes evolving according to the models described in earlier sections.

- At time *t*+2 years, we record the cumulative home price appreciation along each path from *t* to *t*+2, and group paths with similar *HPC* values into *n* bins. For this example, MPA segments home price movements into *n=12* distinct bins.

- At the end of the simulation period, which is *t*+10 years, MPA computes the pool-level losses for each path, average these losses across all paths belonging to a particular bin, and report these losses as the loss corresponding to a particular range of *HPC* values as shown in Figure 9. That is, the loss for each bin is *E[pool loss | HPChange* in *HPCrange(j)]*, where *HPCrange(j)* is the range of 2-year cumulative *HPC* values corresponding to bin *j*.

Since MPA uses the results of the simulation to compute sensitivity to *HPC*, all variables other than *HPC* vary according to the models specified earlier. MPA uses this approach to calculate sensitivity analysis since there are significant interaction effects between *HPC* and other macro factors as well as loan attributes on the pool loss, and the simulation approach provides a natural way of accounting for these interaction effects. Accordingly, MPA's measure of sensitivity of portfolio-level expected losses to change in *HPC* is

$$\text{sensitivity} = \frac{\Delta E[\text{pool losses}|HPC]}{\Delta HPC}$$

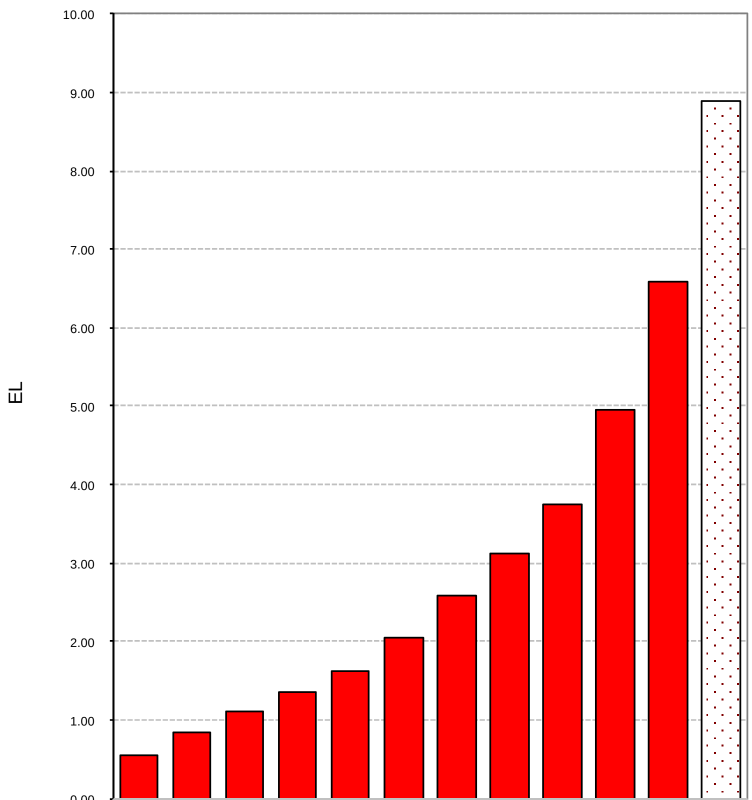
Sensitivity analysis for other variables of interest is performed in a similar manner. Another example of sensitivity analysis is provided in Figure 15. This figure summarizes the sensitivity of portfolio-level losses to a loan attribute, *CLTV*. The expected loss for each *CLTV* bin is calculated using an approach similar to that described earlier, except that the grouping is now based on loans that have similar *CLTV* at origination and averaging is done over all paths.

Recall that the model simulates losses on a loan-by-loan basis, which facilitates analyzing how different sub-groups of loans perform. As seen from the figure, the expected losses increase with *CLTV*, which again suggests that the predictions of the model are consistent with economic reasoning. Since the averages calculated for sparsely populated bins (less than 100 points) are likely to have large variance, we display conditional averages for such bins using a dotted instead of a solid background. This can be seen in Figure 15.

We described how we use sensitivity analysis for model validation. Because users of the model also find it useful to examine such sensitivities, either to better understand how the model behaves or to conduct scenario analysis, MPA also includes a number of analytic and graphical reports to characterize the behavior of the model for several key macro factors and loan attributes.

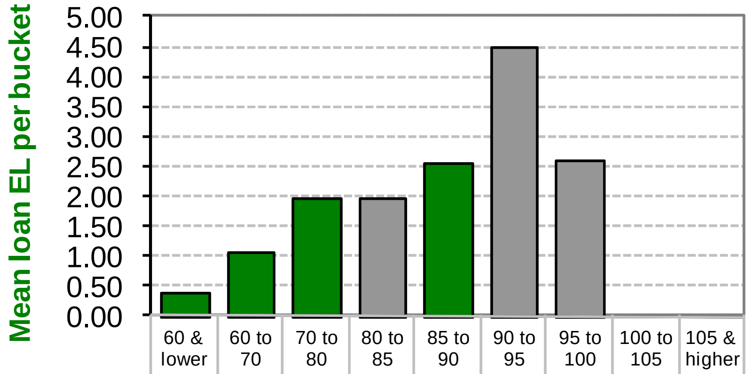
In fact, Figure 14 and Figure 15 are directly taken from the standard output the portfolio analysis software.

Figure 14: Model sensitivity to macroeconomic levels



	7.5% & higher	7.5% to 2.5%	2.5% to 0.0%	0.0% to -2.5%	-2.5% to -5.0%	-5.0% to -7.5%	-7.5% to -10.0%	-10.0% to -12.5%	-12.5% to -15.0%	-15.0% to -20.0%	-20.0% to -25.0%	-25.0% & lower
Sparse data	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	8.89
Fuller data	0.55	0.83	1.09	1.35	1.61	2.04	2.58	3.12	3.74	4.94	6.59	0.00

Figure 15: Model sensitivity to loan-specific factors fevels



	60 & lower	60 to 70	70 to 80	80 to 85	85 to 90	90 to 95	95 to 100	100 to 105	105 & higher
Sparse data	0.00	0.00	0.00	1.97	0.00	4.47	2.58	0.00	0.00
Fuller data	0.38	1.07	1.97	0.00	2.56	0.00	0.00	0.00	0.00

4.2 Predicting defaults for known macroeconomic paths

The sensitivity tests described in the previous section are conducted to examine whether the *variation* in model prediction because of a variation in macro factors or loan attributes is consistent with economic reasoning. We now examine whether the *level* of default rates predicted is consistent with realized values.

MPA was developed to estimate the *distribution* of pool-level losses. This distribution is generated by first simulating different states of the economy and then predicting the losses *conditional* on the states of the economy to arrive at a distribution of losses.

Since our effort focuses on predicting defaults conditional on the states of economy, we validate the predictive capability of the model by examining whether predicted default levels match realized levels, assuming we know the actual state of the economy, which we do for historical paths.

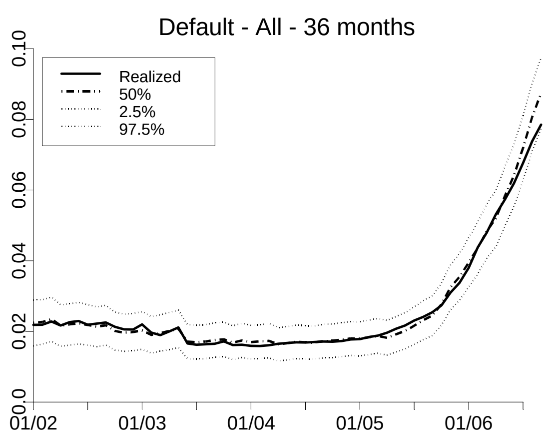
We conduct our tests using historical values of macroeconomic factors to predict defaults at different points in time and compare these predictions to the realized levels of defaults.

The procedure is detailed below:

1. Start in a month *t* and create a cohort of loans outstanding in month *t*
    a. For *T* months into the future, run the default model and the prepayment model to estimate the one-month default probability and prepayment probability for each loan for each month from time *t* to *t+T*. At the end of month *t+T*, calculate the *T*-month cumulative predicted default probability by compounding the one-month default and one-month prepayment probabilities.
    b. Calculate the pool-level cumulative default probability as the average default probability across all loans in the pool
2. Compare the predicted cumulative default probability to the realized default rate over the same period
3. Repeat Steps 1, 2 and 3 for each month in the data set (except for the last *T* months, as we will not have *T*-month-ahead realized default rates for these months)

Shown below in Figure 16 is the result of this validation exercise on the development sample (base sample) updated to include the most recent out-of-sample period. It shows the 36 month default rate (i.e., *T*= 36 months) for all loans in the base data set as a function of calendar time. The realized default rate is shown in solid line, the model forecast is shown in dashed line and the 95% confidence bounds are shown in dotted lines. The graph is representative of the general results we observed for various subsets of the data (with the exception of a few cases which we discuss in more detail below).

Figure 16: Realized and predicted default rates with confidence bounds

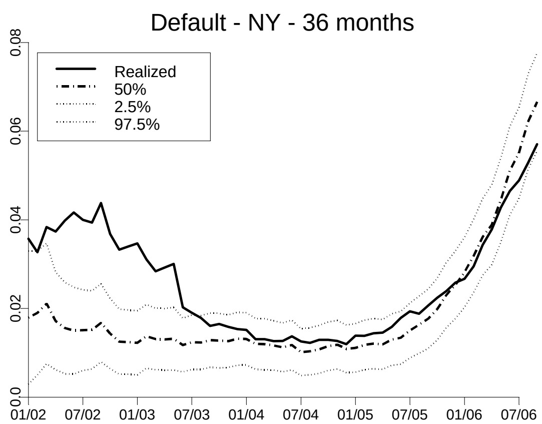


In considering these exhibits, a few features are worth noting. First, the observed data represents actual performance, net of the effects of Home Affordable Modification Program (HAMP), while the model output is meant to be a “pure-play” on loan quality and economic environment. Therefore, we should expect systematic deviations between the observed and predicted values during the HAMP period and find that the models’ estimates are higher than observed values for the last part of the sample period.

To gain more insight into the behavior of the models in different settings, we examine the performance of our model for different types of loans. We repeat our analysis on various sub-groups of loans, such as: ARM/Fixed, High/Medium/Low *LTV*, different states, different loan margins, large/small loan amount. In general, the results are consistent with those shown above, though, as expected, the confidence bounds widen as the number of loans used in the test pool shrinks because of subsetting.

Unsurprisingly, our model cannot match the realized performance of pools that are affected by a unique cohort-specific shock. For example, borrowers in the State of New York were particularly affected by the events of 9/11. Figure 17 plots default information for a pool of NY loans in our data set. As seen in Figure 17, realized default rates increased markedly for a period following 9/11. Since our model does not explicitly account for terrorist attacks, our prediction significantly under estimates defaults for about two years following the attack. However, as the economic impact of 9/11 abates, we again find good agreement between predicted and realized default rates.

Figure 17: Realized and predicted default rates for all NY state loans in our data set



4.3 Comparing model predictions to analyst estimates

Finally, we compare the model's loss performance estimates with those developed by analysts for RMBS collateral pools. The comparisons are made for seasoned pools. For each mortgage pool, MPA estimates the cumulative expected losses for a 10-year period, starting from the deal origination date. For seasoned pools, the model uses the actual historical values of economic factors to predict pool performance from origination to the current date, and then uses simulations to predict additional future losses from the current date to the end of the 10-year period used for the analysis. The model's estimates are compared with those developed by the RMBS surveillance group at Moody's Investors Service. This group's estimates are based on actual pool-level losses realized to-date plus an estimate of future losses. Their estimate of future losses is based on their expert judgment as well as on assumptions of appropriate roll-rates for different statuses of the loans (e.g., 60+ delinquent and 90+ delinquent). Note that since the model's estimates do not include adjustments for HAMP or other government programs, the analysts' estimates were adjusted to be net of HAMP as well.

We report, for example using the prime universe, results of the comparison in Table 6 for a set of mortgage pools originated at different points in time, starting from 2006Q1 to 2007Q4. For each quarter, we report the average pool-level losses estimated by MPA's default model versus those estimated by the analysts. The average is calculated across all the pools in the sample for the particular quarter of deal closing. These tests were performed based on analysis done in the beginning of the second quarter of 2010.

Table 6: A comparison of the average expected loss over prime pools: MPA's Prime vs. Surveillance

Comparison Date	Mortgage Portfolio Analyzer Prime estimates	MIS RMBS Surveillance Group estimates
2006Q1	8.4	8.3
2006Q2	9.6	9.3
2006Q3	9.1	8.9
2006Q4	8.7	8.5
2007Q1	13.0	12.4
2007Q2	14.1	13.7
2007Q3	13.7	12.6
2007Q4	15.7	15.3

We also note that the pool estimates demonstrated high correlations with analysts' estimates at the individual pool level as well as the vintage level. These results suggest that the model's predictions are consistent with those developed by analysts.

5 CONCLUSIONS

In this paper, we have described some of the details of the Mortgage Portfolio Analyzer models. This suite of quantitative models is integrated into an analytic portfolio tool for assessing the risk of residential mortgage portfolios. The tool calculates, in a correlated fashion, the default, prepayment and severity of (typically thousands of) individual mortgages in a pool and then aggregates these risks taking into account the loan-level characteristics as well as macroeconomic variables.

Mortgage Portfolio Analyzer is a quasi-structural model of mortgage portfolio loss: the individual drivers of losses are correlated but are treated as unique processes, and these processes are integrated in an economically coherent manner so that both the behavior and the results of the analytics are economically rational.

Our research has yielded a number of stylized facts. Firstly, our results suggest that there does not appear to be a single factor (or two) that can in general explain the losses on mortgages and mortgage pools; instead, we find that the recent industry trend towards "layered risk" analysis is justified and in fact required. As a result, we find that it is far more effective to model prepayment, default and severity at the loan level if the goal is to accurately capture the loss behavior of large pools of mortgages, particularly when the assets in the pools are distributed heterogeneously. Doing so also reveals that prepayment rates can at times be the dominant effect in determining the distribution of pool losses and that without adequately analyzing prepayment processes, losses are difficult to understand over the medium term.

We also find evidence that all three of the underlying processes that drive losses appear to be correlated through their joint dependence on economic factors such as the levels of interest rates and local home prices (there is a weaker dependence of some processes on local unemployment rates). This dependence in turn induces correlation among the loans in a portfolio which must be modeled.

Though the results of our analysis and validation are encouraging, this should not be taken to imply that the models cannot be further improved. Indeed, research on our next generation of models is already underway. While we feel that the Mortgage Portfolio Analyzer approach represents a useful tool for analyzing mortgage portfolio credit risk, we are also cognizant of the limitations of the model. We expect to introduce refinements to various aspects of the models as our data set continues to grow and our research reveals additional structure in mortgage behavior.

MPA has been designed to provide strong insight and transparency into the measurement and management of credit risks of portfolios of U.S. residential mortgage loans. It has applications to risk-management, portfolio-construction and capital allocation. MPA provides a common and coherent framework for the analysis whole-loan portfolios or, when integrated with a waterfall tool, for RMBS transactions. A single framework also accommodates both the analysis of newly originated portfolios and the monitoring of seasoned pools. MPA takes advantage of the data that users have available including, loan- and pool-performance to date, loan status, mortgage insurance terms (at the loan- and pool-level) as well as other information.

Finally, we would like to emphasize the spirit in which the model should be used: as an input to rather than a substitute for a rigorous analytic process. As an input to an analytic process, the approach can enhance users' understanding of the risks in mortgage portfolios they evaluate.

6 REFERENCES

Acharya, V., B. Sreedhar, and A. Srinivasan, 2004, *Understanding the Recovery Rates on Defaulted Securities*, and references therein.

Bohn, J. R. and R. M. Stein, 2009, *Active Credit Portfolio Management in Practice*, Wiley, Hoboken, NJ.

Bolder, D. J., 2001, Affine Term-Structure Models: Theory and Implementation, *Working paper*, Bank of Canada, 2001-15.

Breslow, N. E., 1974, Covariance Analysis of Censored Survival Data, *Biometrics*, 30, 89-99.

Chinchalkar, S. and R. M. Stein, 2010 Comparing loan-level and pool-level mortgage portfolio analysis, Working Paper 2010-11-1, Moody's Research Labs, New York.

Cox, C.J., J.E. Ingersoll, and Jr., S.A. Ross, 1985, A Theory of the Term Structure of Interest Rates, *Econometrica*, 53 (2), 385-408.

Cox, D.R., 1972, Regression Models and Life Tables (with discussion), *Journal of the Royal Statistical Society,* B34, 187-220.

Das, A. and R. Stein, 2009, Underwriting versus economy: a new approach to decomposing mortgage losses. *The Journal of Credit Risk*, Volume 5 issue 2, 19-41.

Das, A. and R. Stein, 2011, Differences in Tranching Methods: Some Results and Implications, *Working Paper*, Moody's Research Labs.

Dwyer, Douglas W., and Roger M. Stein, 2004, Moody's KMV RiskCalc v. 3.1 technical document, New York: Moody's KMV.

Duffie, D., A. Eckner, G. Hotel, and L. Saita, 2009, Frailty correlated default, *The Journal of Finance*, Volume 64 Issue 5, 2089-2123.

Efron, B., 1977, The Efficiency of Cox's Likelihood Function for Censored Data, *Journal of the American Statistical Association* 76, 312-319.

Falkenstein, Eric, A. Boral, and L.V. Carty, 2000, RiskCalc for private companies: Moody's default model. New York: Moody's Risk Management.

Ferrari, S., and F. Cribari-Neto, 2004, Beta regression for modeling rates and proportions, *Journal of Applied Statistics*, v31, 799-815.

Gupton, G. M., and R. M. Stein, 2002, LossCalc(TM): Moody's model for predicting Loss Given Default (LGD), *Moody's Rating Methodology*. Moody's Investors Service, New York.

Gupton, G. M., and R. M. Stein, 2005, LossCalc™V2: Dynamic prediction of LGD, *Moody's KMV Modeling Methodology*. Moody's KMV, New York. References therein.

Hastie, T. and R. Tibshirani, 1990, *Generalized additive models*, London; New York, Chapman and Hall.

Hastie, T., R. Tibshirani and J. Friedman, 2001, *The Elements of Statistical Learning*, New York, Springer-Verlag.

Kalbfleisch, J. D., and R. Prentice, 1980, *The Statistical Analysis of Failure Time Data*, John Wiley, New York.

Qi, M., and X. Yang, 2008, Loss Given Default of High Loan-to-Value residential mortgages, *Journal of Banking and Finance*.

Siminoff, J. S., 1996, *Smoothing Methods in Statistics*, New York, Springer-Verlag.

Zandi, M. M. and Z. Pozsar, 2006, U.S. Macro Model System, *Regional Financial Review*, Moody's Economy.com, West Chester, PA, December.