Moody’s Analytics DPLC: Methodology for Forecasting and Stress-Testing U.S. Vehicles ABS Deals

PREPARED BY
Moody’s Analytics

Contact
Juan Carlos Calcagno
Director Credit Analytics
Email: juancarlos.calcagno@moodys.com

Jian Hu
Senior Economist
Email: jian.hu2@moodys.com

Benjamin Kanigel
Economist
Email: Benjamin.kanigel@moodys.com
Abstract

The Moody's Analytics Default, Prepayment and Loss Curves product is an econometric model to forecast and stress-test the collateral backing global asset-back securities/residential mortgage-backed securities deals. In this paper, we outline the modeling approach we use to forecast and stress-test the cash flow backing U.S. vehicles ABS deals. Our econometric approach considers loan characteristics, economic conditions at loan origination, past pool performance, and dynamics in the macroeconomic environment over time to explain changes in pool-level performance. The main outputs of these models are off-the-shelf, scenario-based default, prepayment and severity vectors, all consistent with macroeconomic assumptions generated by Moody's Analytics Chief Economist Mark Zandi. These vectors provide all necessary data to run waterfall valuation engines and thus compute fair value and expected loss under baseline as well as stressed economic conditions.
1. Overview

Since the inception of the asset-backed securities market, vehicles securitization has accounted for a substantial portion of total ABS volume. Vehicles securitizations were the largest component of the ABS market issuance volume from inception until 2001. They were relegated to the second spot during the home equity products boom and until the market collapse in 2007-2008. Since 2008, vehicles issuance as a percentage of total ABS issuance has been increasing and is expected to remain a key component of the market in the years to come (Figure 1). Bond issuance volume for these assets has been particularly robust in recent years because of several appealing features including strong vehicles sales combined with short- and medium-term loans, high asset quality and ease in liquidation of delinquent receivables, and steady and predictable collateral performance.

Figure 1: Vehicle ABS Market Volume (June 2012)

Sources: Sifma, Moody's Analytics
Notes: Vehicles data include auto loans and leases, truck, RV, motorcycle, and auto dealer floorplans.

As the U.S. ABS vehicles market regains its status within the ABS sector and the residential mortgage-backed securities market remains dormant, it becomes crucial to understand and project cash flows and assess bond value of vehicles deals. The accuracy of forecasts of these cash streams should be the objective of any model used to assess bond value. This technical document explains Moody's Analytics econometric model to solve the first part of the puzzle—the collateral projections, vectors or assumptions—that can later be used to run waterfall valuation engines and thus compute fair value and expected loss. To build our econometric model, we use standardized pool-level data available for all outstanding and paid-off U.S. vehicles deals Moody's Investors Service has rated since its inception. The models consider average loan characteristics, economic conditions at loan origination, past pool performance, and dynamics in the macroeconomic environment over time to forecast prepayment, default and recovery vectors. The main purpose of the econometric models is forecasting and stress testing; hence our overarching goal is to maximize precision and accuracy while taking into account causal relationships.
This technical report about the Moody’s Analytics Default Prepayment and Loss Curves product applied to U.S. vehicles ABS deals should be important for market participants interested in monitoring and valuating bonds for regulatory compliance and/or buy-hold/sell strategies. The direct application of DPLC is valuation of outstanding deals, though it also can be used indirectly to generate industry-level numbers and/or for benchmarking. The off-the-shelf monthly projections are available for deals outstanding and are based on an array of deal-specific factors; thus it cannot be used for pricing to-be-structured bonds unless additional work is conducted. Finally, DPLC provides off-the-shelf scenario-based collateral projections, all consistent with macroeconomic assumptions generated by a team of economists led by Moody’s Analytics Chief Economist Mark Zandi.

2. Data

2.1. The case for data aggregation

In this paper, we consider the vehicles ABS market to be composed of auto loans (prime and subprime), auto leases, motorcycles, boats, and RVs. Broadly speaking, underwriting standards in the vehicles sector have been consistent over the last 15 years. Although loan-to-value ratios and length of loan terms have being trending up, other standards such as percentage of subprime borrowing, distribution channels, income and employment remain roughly constant. Final bookings, however, vary over time because they are a function of supply and demand for credit. As a result of this economic process, the final makeup of average loans typically shifts over time as lenders and borrowers adjust through economic cycles. Once vintage origination composition is observed, the future performance of vehicles loans and leases, either the whole loan or as collateral for an ABS, is highly predictable and stable.

In the retail market, participants interested in projecting portfolio losses commonly use their own historical monthly data after aggregating by cohorts or segments and apply statistical tools to generate projections. Cohorts are defined based on month or quarter of origination (depending on volume). The cohorts are composed of a fixed number of loans that tend to be homogenous because underwriting standards and policies are clearly set by a central unit and because they are all affected by the same economic conditions. Good amounts of data on sources of heterogeneity are observable and known (i.e., credit scores, terms of loans, geographical and collateral information, etc.). The homogeneity within cohorts and certainty about credit quality make future performance highly predictable within reasonable bounds. Moreover, data requirements and statistical techniques are relatively simple.

Conditional on resources, technology, and size of lender, some lenders have developed statistical models for probability of default and loss-given-default using account/loan level data. The objective of the statistical model is very different here: Models are needed for scoring/loan origination and/or to supplement credit bureau scores using factors important for the vehicles industry. When a lender is pricing a loan, it becomes extremely important to rank individuals from highest to lowest

1 Dealer floorplan deals are not included here because those share more structural similarities with credit card ABS than with other vehicle deals.
2 See Davis and Frank (2011); Date and Reed (2009), Moody’s Investors Service (2008)
credit risk, thus effectively deciding on a case-by-case basis whether a loan should be originated given underwriting standards. Loan-level models also provide valuable information for collections policies. For example, individuals with a particular mix of characteristics may respond to collections activity in one way, while others with a different mix of characteristics may respond entirely differently. If one's aim is to manage collections at such a granular level, loan-level modeling approaches are necessary to achieve one's objectives, and model discriminatory power becomes the overarching goal.

Translating retail lending practices to the securitization market is straightforward. At deal closing or during monitoring, the most important objective of a statistical model aimed to value the bond must be forecast accuracy at deal/pool level (default, recovery and prepayment). There is no need to price or score individual loans anymore because loans have already been originated and now belong to a single structure. As long as (i) the condition of homogeneity within the pool is met, (ii) source of variations can be measured or explained using observable factors, and (iii) aggregated data are high quality (created using loan level files), future performance of vehicles ABS deals is highly predictable using aggregated data. Furthermore, during monitoring, actual aggregated performance is observed and can be helpful to project future performance.

There is, however, one fundamental difference between collateral backing a vehicles deal and cohort of loans kept on lenders' books: The loans backing securitizations are not necessarily originated in the same month or quarter. At deal closing, the loans cut to serve as collateral for the bond can be seasoned, unseasoned, or a combination of both. This could lead to slightly more heterogeneity than would be expected given uniform origination times. However, it is standard practice that once the pool of loans is “cut” the number of loans remains fixed or decreases as loans are paid off. This feature is the same as for a vintage of loans kept on lenders' books. Although less uniform than whole loans kept on lenders' books, vehicles ABS deal composition tends to be generally homogenous and simple, so aggregate-level statistical models provide extremely useful information to project pool-level losses after seasoning at origination is accounted for.

2.2. U.S. vehicles ABS performance data PDS

The performance data used to build econometric models backing DPLC is drawn from the Moody's Analytics Performance Data Services product. The PDS data collection process starts with raw data collected when Moody's Investors Service rates an ABS/RMBS deal and the bond is originated. During the deal-monitoring process, data files are drawn from major servicer/trustee reports and are scrubbed and standardized by a team of specialists. Fields collected include both origination conditions and continuously updated performance information from servicer and trustee monitoring reports. PDS coverage matches the market share of MIS in rating the underlying deals. The data are consistent across regions, allowing for international analysis if so desired.

The data are of excellent quality and of a perfect multidimensional structure for the forecasting models considered in this article. Included in the files are more than 100 performance statistics at the deal, pool, and tranche level. Our analysis typically concentrates on key pool performance metrics such as delinquency rates as a proportion of original balance, longer-term indicators of default, prepayment rates, and severity of losses or loss given default. For the special case of the asset groups discussed in this paper, we obtained information on 244 outstanding pools and 1,278 other pools originated since 1989 that have already paid off or terminated (total 1,522 pools, Table 1).
Table 1: PDS Data Summary Statistics (June 2012)

<table>
<thead>
<tr>
<th>Type of Vehicle Deal</th>
<th>Total Pools Available</th>
<th>Pools Outstanding (as June 2012)</th>
<th>Closing Date of Oldest Deal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Auto loan prime</td>
<td>690</td>
<td>148</td>
<td>Mar-1990</td>
</tr>
<tr>
<td>Auto loan marginal</td>
<td>211</td>
<td>8</td>
<td>Mar-1993</td>
</tr>
<tr>
<td>Auto loan subprime</td>
<td>401</td>
<td>42</td>
<td>Mar-1993</td>
</tr>
<tr>
<td>Auto leases</td>
<td>110</td>
<td>24</td>
<td>Dec-1992</td>
</tr>
<tr>
<td>Motorcycle</td>
<td>51</td>
<td>10</td>
<td>Mar-1994</td>
</tr>
<tr>
<td>Boats and RVs</td>
<td>59</td>
<td>12</td>
<td>Sep-1989</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td><strong>1,522</strong></td>
<td><strong>244</strong></td>
<td></td>
</tr>
</tbody>
</table>

An important feature of the dataset is its length: MIS began collecting performance data for most of these groups in early 1990. For a substantial number of deals, we observed origination and complete performance until maturity. This feature of the data is fundamental for forecasting, because it allows the model to capture the entire life cycle of the deals. More important, the data cover three full business cycles, allowing us to correctly measure the impact of cyclical factors and to weigh these factors against internal pool-specific information.

If we consider both dimensions in the data, pools and time periods, a total of 71,057 potential observations are available for modeling purposes. Although coverage and quality of data are high, missing data vary by vector of interest. For example, coverage for delinquencies, default and severity of losses is more than 70% of total potential observations (Table 2). Auto leases has the most data missing among the groups but is still well populated. The average coverage for the prepayment vector is close to 65%, lower than for the other vectors. This could be due to extremely low or no prepayments that are coded as missing and not as zero. The systematic missing data pattern in auto leases and prepayment could cause misleading inferences for deals not included in the estimation sample (i.e. new deals, not MIS-rated, etc.) but we do not expect biases for deals that participate in the estimation process.

Table 2: Data Summary for Key Vectors (June 2012)

<table>
<thead>
<tr>
<th>Type of Vehicle Deal</th>
<th>Total Potential Observations (Pools x Time)</th>
<th>30-Day Delinquency</th>
<th>Total <strong>Non-Missing</strong> Data by Dependent Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Auto loan prime</td>
<td>29,323</td>
<td>78.47%</td>
<td>77.67% 77.36% 90.78% 65.35% 91.06%</td>
</tr>
<tr>
<td>Auto loan marginal</td>
<td>11,011</td>
<td>77.85%</td>
<td>76.04% 75.82% 91.07% 72.15% 92.90%</td>
</tr>
<tr>
<td>Auto loan subprime</td>
<td>18,503</td>
<td>86.95%</td>
<td>77.66% 76.94% 91.01% 66.79% 93.83%</td>
</tr>
<tr>
<td>Auto leases</td>
<td>3,643</td>
<td>70.88%</td>
<td>67.50% 72.60% 72.06% 64.97% 56.90%</td>
</tr>
<tr>
<td>Motorcycle</td>
<td>2,504</td>
<td>95.93%</td>
<td>95.93% 95.93% 95.41% 53.99% 94.57%</td>
</tr>
<tr>
<td>Boats and RVs</td>
<td>6,073</td>
<td>79.29%</td>
<td>76.77% 76.77% 83.58% 45.69% 80.19%</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td><strong>71,057</strong></td>
<td><strong>80.88%</strong></td>
<td><strong>77.46% 77.37% 89.47% 64.68% 89.51%</strong></td>
</tr>
</tbody>
</table>
In terms of outliers, two procedures were put in place to avoid extreme values. First, we restrict the values of key rates to their natural boundary between zero and 100. Second, we compute the first and 99th percentiles of the vector-specific distribution and remove extreme observations if the value of the dependent variable is strictly less than the first percentile or strictly greater than the 99th percentile.

2.3. Macroeconomic series

Moody's Analytics maintains one of the largest repositories of macroeconomic, demographic and financial data from a multitude of government and private sources. The data set covers the national accounts, banking and finance, demographics, personal income, prices, retail sales, labor markets, energy, financial markets, and many other indicators. For this study, we concentrated on a group of about 25 variables measuring income, economic activity, labor market and price indexes. Anecdotal evidence suggests dealers blamed “larger economic factors” outside their control for the peak in charge-offs and repossessions that occurred during 2008 (see Davis and Frank, 2011). Table 3 shows a selection of variables chosen for consideration in this study based on forecast availability, economic intuition, and prior literature on the topic (i.e., Agarwal, Ambrose, and Chomsisengphet, 2008; Heitfield, and Sabarwal, 2004).

Table 3: List of Potential Economic Variables Impacting Vehicles Performance

<table>
<thead>
<tr>
<th>Income</th>
<th>Activity</th>
<th>Labor Market</th>
<th>Interest Rates</th>
<th>Price Indexes</th>
<th>Others</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total personal income</td>
<td>Gross domestic product</td>
<td>Unemployment rate</td>
<td>Bank prime rate</td>
<td>CPI new vehicles</td>
<td>Personal bankruptcies</td>
</tr>
<tr>
<td>Disposable income</td>
<td>Retail sales</td>
<td>Total employed</td>
<td>Discount window borrowing</td>
<td>CPI used vehicles</td>
<td>Underwriting standards</td>
</tr>
<tr>
<td>Wages and salaries</td>
<td>Vehicle sales</td>
<td>Total unemployed</td>
<td>Federal funds rate</td>
<td>Manheim used vehicle</td>
<td></td>
</tr>
<tr>
<td>Debt service burden</td>
<td>New vehicle registrations</td>
<td>UI initial claims</td>
<td>New car loans</td>
<td>Petroleum crude oil</td>
<td></td>
</tr>
<tr>
<td>Net worth</td>
<td></td>
<td></td>
<td></td>
<td>Retail gasoline</td>
<td></td>
</tr>
</tbody>
</table>

Moody's Analytics produces monthly forecasts for each variable, both baseline and stressed, for many countries across the globe using large-scale structural macroeconometric models. The scenarios are internally consistent and take into account cross-correlations between macroeconomic variables. The model and subsequent forecast are generated by Moody's Analytics Chief Economist Mark Zandi and overseen by a team of about 60 economists who cover the performance of the economy in real time. The forecasts undergo continual revision and adjustment to reflect new trends and changing data.
Moody’s Analytics produces one upside and several downside alternative scenarios each month including the Federal Reserve’s supervisory stress scenario for its Comprehensive Capital Analysis and Review. Figure 2 shows historical trends and alternative scenarios for four key variables used for vehicles econometric models: unemployment rate, GDP (year-over-year changes), vehicles sales, and the Manheim Used Vehicle Index.
Figure 2: Scenario Description, Graphical Analysis

- Unemployment Rate (%)
- GDP (SAAR, y-o-y)
- Vehicles Sales (SAAR, y-o-y)
- Manheim Used Vehicle Value Index (SA)
3. Econometric Issues

3.1. Intuition behind econometric model

Factors influencing collateral pool performance can be conceptually divided into three classifications: (i) life cycle trends depending on a pool’s age-on-books (seasoning); (ii) factors indicating the quality of a pool; and (iii) characteristics of the current economic environment that depend only on calendar time. Other effects also operate across more than one of these main categories and can be modeled as interactions between them. An examination of each of these types of effects in isolation is key to understanding the multidimensional nature of the data and the models used to forecast it.

3.1.1. Life cycle

Performance metrics, whether they measure delinquency, default, prepayment, or another attribute of a collateral pool, often tend to follow a predictable baseline life cycle common to all pools of loans. Delinquency on pools of auto loans, for example, tends to rise from near-zero levels in early months after origination, then reach a peak and gradually decline as the pool matures. Borrowers who enter delinquency tend to do so at fairly similar times relative to loan origination, with the timing and magnitude depending on the product and the particular measure of delinquency being considered. After the peak, most borrowers who would become delinquent have done so already, and with a significant portion of balances paid off, the pool risk declines over time. Figure 3 charts an example data set of delinquency rates on auto loans, with each line representing one pool of loans and aligned by age-on-books to show the similar shape of the life cycle across pools. One fundamental reason this behavior is predictable as time passes is that the number of loans backing a pool is fixed at origination and can go down only as loans paid off (voluntary or involuntary).
This nonlinear maturation cycle is not a function of any observable factors; rather, it is purely a function of age-on-books. Even if no characteristics of a pool or its environment change, it will still perform differently and in a predictable fashion as it matures. Estimating this baseline maturation cycle is the first step in predicting pool performance. The life cycle for a given variable can be loosely understood as its average level across all pools at each value of age. Although more sophisticated econometric techniques are actually used to estimate the life cycle, this approximation is useful to demonstrate the basic shape. Aligning pools by age and averaging removes environmental and quality effects, leaving only the life cycle. Figure 4 shows an example data set with the average life cycle graphed as a single line.

3.1.2. Credit quality

Individual pool performance does not always follow its baseline-predicted life-cycle contour closely. One reason that individual performance strays from the average is heterogeneity in credit/pool quality. Consider two pools of loans, one with an average borrower FICO score of 750 (red line in charts below) and one with a score of 580 (blue line). Delinquencies in both pools might peak and begin to decline at about the same age relative to origination, but we would expect the magnitude of the peaks to differ substantially. This is borne out in the data, and it can be incorporated into models most simply by including time-invariant effects measuring the origination quality. This is demonstrated in the left side of Figure 5, which charts two life cycles that are simple level-shifts away from the average life cycle in Figure 4.

Alternatively, interactions between the life cycle function and certain quality variables can be included in order to model disparate groups of loans that have fundamentally different life cycles. For example, the peak delinquency rates for subprime and prime auto loans might not only have different magnitudes, but different timing as well. This is shown in Figure 6.
There is no single reliable metric that completely accounts for credit quality at origination, so an array of factors is used to capture it. Some of these factors are intended to directly measure borrower risk at origination, such as FICO score or loan-to-value ratio. Others are only tangentially related to intrinsic quality, such as the average coupon rate of the loans in a pool or the issuer of the loans. This set of observable quality variables is often still inadequate to explain a significant amount of variance in quality, and another approach is needed to augment them. One method is to use economic conditions at the time of origination as a proxy for quality. It is well established that loan managers at lending institutions adjust underwriting standards based on current macroeconomic conditions. Amid an expanding economy, lenders loosen standards in competition for borrowers, and loans originated during booms are therefore likely to be of lower intrinsic quality, on average. During recessions, lenders tighten standards in an attempt to stem future losses, so new loans are extended only to exceptionally safe borrowers. Somewhat counterintuitively, it follows that loan quality can be thought of as being inversely correlated to origination economic conditions.

3.1.3. The economy

The inverse relationship between credit quality and economic conditions at origination should not be confused with current economic impacts on performance of existing loan pools. Just as a rising tide lifts all boats, an expanding economy boosts the performance of loan pools of all vintages and qualities. Low unemployment and strong income growth ease the burden of debt payments on the average borrower, lessening the risk tied to their loans. Of course, the converse is true also: A recession leads to higher joblessness and tepid wage growth, sending more borrowers into delinquency and default. Macroeconomic variables, often transformed or lagged, are included in models to account for these time-dependent effects. Unlike economic variables at origination of loans, these time-varying effects influence all loan pools contemporaneously—indeed independently of life-cycle stage—and have a positive correlation with pool performance. This is illustrated in Figure 7: Three pools of different origination dates are charted across calendar time, and all three are impacted in mid-2013 by a forecast economic shock.

Figure 7: Shifts Due to Economic (Time) Shock
Finally, certain factors explain performance variation in more than one of the dimensions discussed above. More typically, the statistical model includes interactions between any of the effects in the three primary categories of age (life cycle), quality (pool specific), and time (macroeconomic). All these types of connections between groups and dimension reduction can be handled using multivariate analysis explained in the next section.

### 3.2. Mathematical specification

The intuitive approach described above can be formalized using a mathematical specification and estimated using statistical techniques to construct high-quality forecasts. As our general approach we consider a model of the form:

$$ y_{ivt} = \mu_i + X_{ivt}' \beta + \omega_i' u_{ivt} $$

where $i = 1, \cdots, N$ indicates a series of pools contained in the data set of interest, $v = 1, \cdots, V$ is a time series defining the cohort, or vintage, in which the pool was originated, $t = 1, \cdots, T$ is a standard time series indicator of the period in which the pool’s performance is observed and $y_{ivt}$ is the dependent variable of interest, be it a delinquency rate, a default rate, a prepayment rate, or loss severity. The vector $u_{ivt}$ is an unknown random error term and $\mu_i$ defines the unobserved, time-invariant, pool-specific heterogeneity in the data. Decomposing $X_{ivt}$ into the components explained in the previous section:

$$ y_{ivt} = \mu_i + a_i' \theta + b_i' \gamma + c_i' \tau + d_i' \delta + w_{ivt}' \varphi + \omega_i' u_{ivt} $$

The vectors $a_{ivt}, b_i, c_V, d_t$ and $w_{ivt}$ contain independent variables thought to explain the behavior of $y_{ivt}$.

The **life cycle** is represented by $a_{ivt}$, since age of pools is a simple function of time and vintage origination date. We model this inherently nonlinear behavior using standard cubic spline functions with a small number of knots to best approximate the shape changes across the pool’s life cycle. We use Stata ado commands `rc_spline` or `mkspline` that create $a_{ivt}$ variables based on seasoning. A number of $k$ knots are defined by the econometrician after visual inspection of the data or using optimization routines such as Stata `uvrs` (Royston and Sauerbrei, 2007). The $a_{ivt}$ set of variables is defined to be a continuous smooth function that is linear before $k_1$; is a piecewise cubic polynomial between adjacent knots; and is linear again after the last knot $k_T$. There is always one fewer variable than there are knots.

The vectors $b_i$ and $c_V$ contain factors that pertain to credit **quality** and do not change over time. The $b_i$ component specifically relates to pool-specific factors such as average characteristics at deal origination or the issuer of the loans. Pool quality is a direct function of underwriting systems and strict norms common within the same originator. The $c_V$ vector, meanwhile, contains factors specific to a time period and that do not change over time, the main factors that would describe economic conditions that existed when the pools in question were being originated. The $d_t$ vector contains factors that change over time but which are common to all pools. This is therefore the component that explains the way the economy affects the performance of all legacy pools.
The vector of observables factors also include $$w_{tvt}$$ as a measure for any factor that varies across in more than one of the dimensions of age, quality and time. The vector may include interaction terms between any or all the individual components described above. For example, pools with different origination characteristics might respond to macroeconomic shocks in different ways, or pools backed by assets in different product lines might have differently shaped average life cycles. It can also include roll rates—forecasts of one performance metric used to explain and forecast another. They tend to have a high degree of explanatory power, and additional effects to be included in a model together with roll rates are carefully vetted to avoid redundancy.

The last component of the forecasting model is the error term. We assume $${\omega'}$$ is the estimated vector for the unknown random error term where $$\omega = E[\mu_{t+n}, \mu]$$ is the covariance between future disturbance $$\mu_{t+n}$$ and $$\mu$$. For the estimation of the error term and especially the covariance between disturbance $$\mu_t$$ we use an autoregressive model of order 1 (AR1). For each vector, we first estimate the residual and then obtain $$\omega_t$$ assuming an AR1 process by groups in the dataset. Values of the parameters are bounded in order that the model remains stationary ($$|\omega| \leq 1$$).3

3.3. Estimation and practical problems

In this section, we turn to estimation issues, followed by some practical matters of model specification for the problem at hand. Note that all outcomes discussed in this paper are considered as proportion data with values that fall between zero and one. It is important to have the predicted values also fall between zero and one. For this purpose, we use a generalized linear model rather than the naïve linear regression model to perform the estimation. The GLM generalizes ordinary linear regression by relating the model to the dependent variable via a link function and allowing for dependent variables that have arbitrary distributions rather than simply normal distributions. The GLM is more flexible than ordinary linear regression. In this study, we use a GLM with a logistic link function to estimate model parameters as follows:

$$y_{tvt}^* = \ln \frac{y_{tvt}}{1 - y_{tvt}}$$

As the equation above shows, $$y_{tvt}^*$$ is a logistic transformation of $$y_{tvt}$$, while $$y_{tvt}$$ is a linear combination of independent variables on the right-hand side. The dependent variable is assumed to follow binomial distribution, which ensures that the probability of events (prepayment and default) ranges from zero to one. Maximized likelihood optimization is carried out to estimate the model.

3.3.1. Variable selection

A key aspect of model development is variable selection—identifying which credit and economic variables best explain the dynamic behavior of the dependent variable in question. Aligned with principles

---

3 The combined estimator is the best linear unbiased predictor; see Goldberger (1962) or Baltagi (2008).
of modern econometrics, we prefer to choose the variables based on a combination of economic theory or intuition, together with a consideration of the statistical properties of the estimated model. We believe models built using pure data-mining techniques or principles such as machine learning, though they may fit the existing data well, are more likely to fail in a changing external environment because they lack theoretical underpinnings. The best prediction models employ a combination of statistical rigor with a healthy dose of economic principle. Models built this way enjoy the additional benefit of ease of interpretation.

For explanatory purposes, we focus on economic variables here, but the same principles apply for all other model inputs. Among the economic variables listed in Table 3, we select the most suitable set of economic variables for each vector according to the following principles:

1. The economic variables can explain the variation of dependent variable according to economic theory, and the correlations are statistically and economically meaningful:
   a. The economic variables have the expected signs implied by economic theory and intuition. For example, higher unemployment should lead to a higher default rate. Therefore, the sign of unemployment rate should be positive in the default model.
   b. Adding each economic variable helps the model improve predictive power. Generally speaking, the economic variables should be useful in both producing accurate out-of-sample forecasts and providing good in-sample fit. However, we sometimes have to make tradeoff decisions to balance out between these two goals when they are conflicting. If the discrepancy is unavoidable and very significant, we prioritize forecast accuracy rather than in-sample fit, as forecasts are end results of our models.

2. The impacts of economic variables should be robust over time. More specifically, the correlations between vectors and explanatory variables do not break from month to month. We monitor correlations on an ongoing basis and make corrections using updated information.

3. The economic variable should be stationary. For instance, GDP is a nonstationary series; hence we use GDP growth rate as an alternative because the growth rate is stationary.

In the variable selection process, we allow the sensitivity between economic variables and products to vary by allowing for interactions between product line dummy and a few economic variables. For example, the expected loss of subprime auto ABS deals could be more sensitive to a weak labor market than that of prime auto ABS deals because subprime borrowers are more likely to lose jobs and have limited financing opportunities during an economic downturn.

Variable selection is more art than science. The criteria mentioned above are not black or white. The bottom line is to build a theoretically sound and empirically workable model and get reasonable and consistent forecasts that are supported by both economic intuition and statistical significance.
3.3.2. Using origination or updated characteristics

One topic under consideration in the literature is whether origination conditions provide enough information for forecasting or whether refreshed characteristics should be used (i.e. refreshed FICO, LTV, etc.). Recent developments on the mortgage data market have made updated data available from the modeling process and it is being used for improving forecasting modeling. There are, however, problems and limitations with using refreshed characteristics that might outweigh the benefits, particularly when used for nonmortgage assets.

In this paper and for all assets covered by DPLC models, we do not use updated characteristics for the following reasons:

1. Limited amount of data available to build robust econometric models
2. Cost of data systems to generate this information
3. Availability of alternative ways to measure time-varying credit quality
4. Homogeneity and predictability of vehicles performance

From a theoretical standpoint, instead of relying on refreshed characteristics, time-varying quality can be proxy based on observed historical performance. As discussed above, credit quality cannot be measured using a single metric, but it fundamentally impacts performance. Thus, observed performance indirectly measures credit quality and is the best proxy for future performance after accounting for other factors such as observable metrics, life cycle and economic conditions. Because the main goal of the model is forecasting for valuation and monitoring and not scoring at origination, previous performance for the same loans can be used and has an important role in the models.

Yet, let’s assume updated data such as a refreshed FICO score or LTV as of the current time period are available regardless of cost (i.e. mortgages). In order to introduce the data in the models, it would be required first to forecast the characteristic in order to be used to forecast performance. More generally, if a model input is time-varying, a forecast for the input is required to generate a forecast of performance. Our selection of the explanatory variables, or inputs, is based on our ability to forecast them easily and sufficiently accurately. In econometric forecasting, the future realizations of the dependent variable rely on accurate forecasts for the independent variables. Forecast errors can easily be multiplied if input variable forecasts are not closely monitored or if the data backing the forecasts are noisy and thus inherently difficult to project.
4. Model Results

4.1. Estimation

In this section, we discuss important modeling issues. Our focus is the right-hand-side variables. The explanatory variables fall into three categories: Life cycle (age-related), pool quality (cohort-related), and economic variables (period-related).

1. Life cycle variables

The life cycle variables are represented by age splines in the econometric model. We purposely create different age splines for different product lines to reflect the unique features of each line. Not only the life span but also the shape and pattern of life cycle vary from line to line. For example, the average age of auto prime ABS pools is about 60 months in contrast to 84 months for pools backed by recreational vehicles. Another example is that the life cycle pattern of auto prime deals deviates from that of auto subprime deals because the borrower behaviors differ from one another.

In the model development process, we customize the equations to accommodate different consumer behaviors by creating line-specific age splines. Technically speaking, we first graph the vector against age using a local polynomial smoothing function. Visual inspection of the resulting fitted curve guides the process of determining optimal knot location. In accordance with the mathematical properties of the restricted cubic spline function, the first knot is the point in terms of pool age where the trend in the vector initially becomes nonlinear (this can be very close to age zero if there does not appear to be any initial linear portion of the curve). Likewise, the last knot is the point at which the trend appears to become linear in perpetuity. For the knots in between, the optimal locations are the points where the peak or trough occurs or the curvature changes (i.e. inflection points). Mathematically, the knots should be at the points where the third derivative of the vector with respect to age changes or, equivalently, where the second derivative changes direction (from increasing to decreasing or vice versa). In theory, more knots always allow for a better model fit. In practice, however, we often find three to five knots to be optimal in order to avoid over-parameterization. If the visual inspection cannot give us clear indication of optimal knot locations, we generally choose evenly distributed knots such as 12, 24, 36, 48 for the product lines with average maturity of 60 months (auto prime, auto near-prime, auto subprime, motorcycle, auto leases) or more knots for the product lines with longer maturity (boats and RVs) as a starting point.

Once an initial set of knots is chosen, we then look at the in-sample fit by checking the statistical significance of each age spline variable and graphically examine the fitted curve against the actual data, adjusting the number and locations of the knots for each product line accordingly. Finally, we look into out-of-sample performance and, if necessary, make further minor adjustments. This process allows us to strike a balance between a good model fit and accurate prediction.

\[ \text{Roughly speaking, the smoothing technique computes the local averages of the series. We use Stata command “lpoly” to do the job.} \]

\[ \text{In differential calculus, an inflection point is a point on a curve at which the curvature or concavity changes sign from plus to minus or from minus to plus.} \]
2. Credit quality variables

Quality variables can be pool-specific or vintage origination-specific, both of them time-invariant. The first set of variables are those defining the different groups or product lines in the data set: auto loans prime, subprime, marginal, leases, boats, RVs, and motorcycles. We also use the weighted average coupon as the key pool-specific quality variable to represent loan credit quality. Empirical research shows that dealers tend to mark up interest rates more for borrowers with weaker credit (see Davis and Frank, 2011). Plus, the average interest rate of auto loans backing ABS deals has a direct impact on the average coupon rate of the ABS securities. Hence, WAC is the final outcome of risk-based pricing on observable factors because interest rates are more favorable for individuals who are observably lower risks. Finally, different originators price risk at origination and collect debt after default in different ways. To account for heterogeneity among originators, we create dummies for the top 15 originators (in terms of the number of deals originated in the past five years) and incorporate them into the model in order to account for the heterogeneous behaviors. We pick only the top 15 originators rather than all originators (188) because we want to model heterogeneity parsimoniously and avoid multi-collinearity issues as much as we can.

Deals that are originated in the same month are considered in the same vintage. The macroeconomic quality indicators \( (c_j) \) should affect all deals originated at that time. For example, the unemployment rate in the month when the deal is originated is indicative of the creditworthiness of borrowers, since a high unemployment rate is associated with high underwriting standards and leads to high credit quality of borrowers at origination. These macroeconomic variables reflect the credit quality of the loans backing the ABS deals. They take on the same value for the same vintage across all deals. For almost all vectors, we include unemployment rate at origination as quality variable.

3. Economic variables

Economic variables are time-varying and are matched with all deals by calendar time. The performance of auto ABS deals is very sensitive to many of these economic indicators representing important aspects of the economy. As an example, the delinquency rate is positively correlated with the unemployment rate. These indicators could be lagged because of the delay between when the economic condition changes and the impact on performance metrics takes effect. The scenario-based forecasts enable us to stress-test the performance of active ABS deals.

We use two types of economic variables in models: 1) Absolute variables measuring the current economic performance in absolute terms. For example, the number of unemployment claims in the current period, the price index of used cars; 2) Relative variables measuring the current economic performance relative to the origination period. They are usually ratios of the difference between origination period and current period to the level in origination period.

Table 4 presents a summary of model inputs per vector. As mentioned before, we use WAC, originator and product line dummies to account for deal quality and heterogeneity at origination. The unemployment rate is also used in all models (except severity) to represent the economic condition at origination and how that affects underwriting standards and loan quality. New-car prices are useful predictors of severity of losses, because the higher the asset value at origination the more difficult it is to
recover later on. Other control variables are seasonality dummies (month) and roll rate (90-day-plus delinquency rate). Finally, the life cycle component is represented by product line-specific age splines, and the number and location of knots vary across vector and product line.

Variables used to represent current economic conditions vary from vector to vector. In order to explain the variation of delinquencies, we use activity (GDP growth rate), income (personal disposable income), labor market (unemployment rate), price index (used cars), and personal bankruptcies. They are correlated with delinquency one way or another. For instance, if personal disposable income increases, borrowers are more likely to make timely payments. For the CDR (default) model, activity (GDP growth rate) and labor market (unemployment rate) factors are major contributors after controlling for roll rates coming from the lagged 90-plus delinquency rate. One can argue that the default rate should rise if the economy slows down as indicated by a decreasing GDP growth rate. As for the CPR (prepayment) model, interest rate relative to WAC and activity (GDP growth rate) provide the largest in-out-of-sample boost. The intuition is that when the economy starts to accelerate, borrowers will expect a more positive economic outlook and are more inclined to prepay. As the market lending rate continues to rise relative to the interest rate on the loans, people will be less likely to prepay, as the outside financing opportunities are more costly. These economic variables are picked based on principles and procedures described in section 3.2.

<table>
<thead>
<tr>
<th>Group</th>
<th>Variable</th>
<th>30-day Delinq</th>
<th>60-day Delinq</th>
<th>90-day+ Delinq</th>
<th>CDR</th>
<th>CPR</th>
<th>Loss Severity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Current Origin Conditions</td>
<td>WAC</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Originator dummies</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Product line</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Economics at Origination</td>
<td>Unemployment rate</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>New-car prices</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Current Economic Conditions</td>
<td>GDP (y/y)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Unemployment rate</td>
<td>t</td>
<td>t</td>
<td>t</td>
<td>t</td>
<td></td>
<td>t-6</td>
</tr>
<tr>
<td></td>
<td>Used-car prices (y/y)</td>
<td>t</td>
<td>t</td>
<td>t</td>
<td>t</td>
<td></td>
<td>t-9</td>
</tr>
<tr>
<td></td>
<td>Car sales (y/y)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Personal bankruptcies (y/y)</td>
<td>t</td>
<td>t</td>
<td>t</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Disposable income (y/y)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>t</td>
</tr>
<tr>
<td></td>
<td>Total income (y/y)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>t</td>
</tr>
<tr>
<td></td>
<td>Bank prime rate</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>t</td>
</tr>
<tr>
<td>Other Variables</td>
<td>Month seasonality</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>60-day delinquency</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>t-1</td>
</tr>
<tr>
<td></td>
<td>90-day-plus delinquency</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>t-1</td>
</tr>
<tr>
<td>Life cycle</td>
<td>Number of knots</td>
<td>3 or 4*</td>
<td>4 or 5*</td>
<td>3 or 4*</td>
<td>N/A</td>
<td>3 or 4*</td>
<td>5</td>
</tr>
</tbody>
</table>

Notes: All models include seasonality factors (month) and dummy for zeros. *Knot locations may vary across product lines. N/A means age splines are not included.
For expository purposes, Table 5 presents partial estimation results for CDR. WAC is positively correlated with the default rate because a higher coupon rate at origination indicates lower credit quality, which leads to a higher default rate. The sign of unemployment rate at origination is negative because in an environment of high unemployment rates lenders usually tighten the underwriting standards, leading to higher credit quality and hence a lower default rate later on. The contemporaneous unemployment rate is associated with borrowers’ job prospects and income expectations. Therefore, the higher the current unemployment rate, the lower the income expectation and the higher the default rate. For brevity, we report only partial regression results for one vector. Other results are available upon request.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>t-statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>WAC</td>
<td>0.087</td>
<td>0.002</td>
<td>39.490</td>
</tr>
<tr>
<td>Unemployment rate at origination</td>
<td>-0.076</td>
<td>0.006</td>
<td>-11.800</td>
</tr>
<tr>
<td>Unemployment rate in current period</td>
<td>0.059</td>
<td>0.003</td>
<td>20.300</td>
</tr>
<tr>
<td>GDP, % change year ago (t-6)</td>
<td>-0.006</td>
<td>0.002</td>
<td>-2.780</td>
</tr>
<tr>
<td>Used-car price, % change year ago</td>
<td>-0.003</td>
<td>0.001</td>
<td>-3.930</td>
</tr>
<tr>
<td>Observations</td>
<td></td>
<td>39,812</td>
<td></td>
</tr>
</tbody>
</table>

Note: Model includes 90-day-plus delinquency, seasonality, dummies for originators, and a constant term.

4.2. Model validation

We follow the Supervisory Guidance on Model Risk Management issued by the Federal Reserve Board and the Office of the Comptroller of the Currency to conduct our model validation. Specifically, our validation framework includes three core components:

1. *Evaluation of conceptual soundness:* We assess the quality of the model design and the robustness of the model specification on regular basis. This involves the examination of overall theoretical construction, key assumptions, input data, and mathematical calculations. Specifically, a) we first check completeness and integrity of performance data and economic data by looking at summary statistics and making sure they are within a reasonable interval. For example, the probability of event (delinquency rate, default rate, etc.) should be within a 0 and 1 interval and the maximum of the delinquency rate in the current month should be in line with the maximum of historical delinquency rate. b) We then compare the signs of the parameter estimates in the model with the signs of correlations between independent variables and dependent variable and make sure they are consistent with economic and financial theory. Although the signs of correlations are not necessarily the same as the signs of estimated parameters, a discrepancy at least triggers a re-examination of the validity of the explanatory variable. c) We check robustness of the model by comparing the parameter estimates between full sample results and subsample results. A robust model should give us fairly close coefficients regardless the sample size.

2. *Ongoing monitoring:* The purpose of monitoring is to confirm that the model is appropriately implemented and is performing as intended. In this step, we evaluate whether changes in performance data and market conditions necessitate adjustment, redevelopment or replacement of the original model. This involves process verification and a consistency check. Process verification checks that
all model components are functioning as expected. Input data should be accurate, complete and consistent over time. Computer code should be correct and changes are logged. Forecast statistics reports should be reviewed carefully and confirm that the model is performing well and warrants no changes. The consistency check is the comparison of model performance across different production runs. We keep track of the model performance by comparing the forecast statistics over time. The results of the analysis may suggest revisions to the model. However, differences do not necessarily indicate that the model is in error. We should look into what causes the discrepancy and how this affects the end results. If the statistics get really worse and fall into an unacceptable range, we should modify the original model to accommodate revised performance data and changing economic conditions and make sure that the model reflects the most recent development in the auto ABS market.

3. **Outcome analysis:** One of the most important findings in the forecasting literature is that the model that best fits the data is not necessarily the one that will provide the most accurate out-of-sample forecast (Fildes and Makridakis, 1995). Typically, to assess accuracy, the data will be split into two data sets: The first set, a development sample, is used to specify the model and estimate its coefficients; the second set is used to evaluate forecast accuracy and is known as the hold-out sample. To validate our preferred models, we hold out the last six available data points at the end of the time series for each pool. The models are fitted to the remaining historical data. The forecasts generated using the models are then compared with actual values observed in the hold-out sample, and aggregate statistics are computed. In addition, we break down the statistics by product lines so that we can clearly assess how our models perform for the different groups. Some thresholds and ranges are established. If forecast statistics fall outside those confidence intervals or tolerance ranges, we will analyze the discrepancies and investigate the causes that are significant in terms of magnitude or frequency.

For the forecast statistics, we compute root mean squared errors, mean absolute error, and median proportional prediction error. Mathematically, we define these stats as follows:

The RMSE is defined as:

$$RMSE = \sqrt{\frac{\sum_{t=1}^{T}(\hat{y}_t - y_t)^2}{T}}$$

where $T$ is sample size, $\hat{y}_t$ is the forecast for period $t$, and $y_t$ is the actual value observed at time $t$.

The MAE is defined as:

$$MAE = \frac{\sum_{t=1}^{T}|\hat{y}_t - y_t|}{T}$$

where $|x|$ denotes the absolute value of $x$. 
The MPPE is defined as:

$$MPPE = \text{median}_t \left( \frac{\hat{y}_t - y_t}{y_t} \right)$$

These accuracy measures evaluate forecast performance from different perspectives and have their own positives and negatives. For example, RMSE and MAE serve to aggregate absolute individual differences (residuals) into one single measure of predictive power. Although these two statistics of forecast accuracy are extensively reported in the literature, they do not carry any information on the direction of bias (always positive), and they are not comparable across vectors/groups as the average of the actuals varies by vector and group. MPPE is a popular relative measure of forecast accuracy based on percentage errors. A positive sign indicates an upward bias while a negative sign shows a downward bias. For example, a MPPE of 3.89% for auto prime prepayment rate means that the forecast is, on average, 3.89% higher than the actual. Hence the lower these three statistics, the better the forecast performance. The advantage of MPPEs is that they are intuitive and easy to explain and can be compared across all vectors/models/samples.

Each month, we compute and report all these statistics for a full picture of the out-of-sample performance. Table 6 presents a summary of select forecast statistics for the key vectors, CPR, CDR and loss severity, or LGD. Recall RMSE and MAE are absolute measures that can be compared only over time within each vector and product line. MPPE is a relative measure and can be compared across vectors and lines. The last row in each panel reports overall statistics across all product lines.

Table 6. Forecast Statistics by Product Line for All Vectors

<table>
<thead>
<tr>
<th>Product line</th>
<th>CPR</th>
<th>CDR</th>
<th>LGD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>RMSE</td>
<td>MAE</td>
</tr>
<tr>
<td>Prime</td>
<td>712</td>
<td>0.20</td>
<td>0.09</td>
</tr>
<tr>
<td>Marginal</td>
<td>38</td>
<td>0.13</td>
<td>0.05</td>
</tr>
<tr>
<td>Subprime</td>
<td>180</td>
<td>0.27</td>
<td>0.19</td>
</tr>
<tr>
<td>Motorcycles</td>
<td>53</td>
<td>0.28</td>
<td>0.21</td>
</tr>
<tr>
<td>RVs</td>
<td>46</td>
<td>0.67</td>
<td>0.40</td>
</tr>
<tr>
<td>Leases</td>
<td>91</td>
<td>0.81</td>
<td>0.47</td>
</tr>
<tr>
<td>All</td>
<td>1130</td>
<td>0.36</td>
<td>0.16</td>
</tr>
</tbody>
</table>

Note: Due to small sample bias, "boats" is dropped from the table.

In addition to the monthly cross section out-of-sample validation, we keep track of the time series evaluation of these statistics. If model performance, as measured by these statistics, deteriorates over time, we reexamine the model and make adjustments, recalibrations or redevelopment as needed. MPPE figures above 10% (absolute terms) are closely monitored for explanations, while MPPEs above 20% guarantee a model respecification. Table 7 presents the forecast statistics of CDR over time. The model performs well over the last six months: All statistics are fairly stable and consistent over time. For example, RMSEs and MAEs are very similar (even decreasing) in the last six executions. Although MPPEs are a little bit volatile, the numbers are still reasonable. No correction or reexamination of the model is needed at this point. Monthly or quarterly reviews are conducted, but major redevelopments are conducted only after models demonstrate a long-term deviation from trend.
In certain situations, it is also important to compare the forecast performance over time for single deals. These types of graphs allow us to identify directional breaks, keep track of long-run averages, and note deviations from trends. Taking an individual pool of auto subprime deals as an example, Figure 8 presents the forecast series for CDR starting from different time periods in the past six months. We can see that the out-of-sample forecasts are fairly close to one another for this pool, confirming our forecast statistic results.

**Table 7: Time Series of Forecast Statistics for CDR**

<table>
<thead>
<tr>
<th>Sample End Date</th>
<th>Product Line</th>
<th>N</th>
<th>RMSE</th>
<th>MAE</th>
<th>MPPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jul-2011</td>
<td>Auto Subprime</td>
<td>214</td>
<td>0.34</td>
<td>0.22</td>
<td>3.29%</td>
</tr>
<tr>
<td>Aug-2011</td>
<td>Auto Subprime</td>
<td>212</td>
<td>0.28</td>
<td>0.19</td>
<td>-2.84%</td>
</tr>
<tr>
<td>Sep-2011</td>
<td>Auto Subprime</td>
<td>206</td>
<td>0.25</td>
<td>0.17</td>
<td>1.61%</td>
</tr>
<tr>
<td>Oct-2011</td>
<td>Auto Subprime</td>
<td>195</td>
<td>0.24</td>
<td>0.17</td>
<td>-3.04%</td>
</tr>
<tr>
<td>Nov-2011</td>
<td>Auto Subprime</td>
<td>190</td>
<td>0.24</td>
<td>0.16</td>
<td>-8.88%</td>
</tr>
<tr>
<td>Dec-2011</td>
<td>Auto Subprime</td>
<td>174</td>
<td>0.27</td>
<td>0.17</td>
<td>4.04%</td>
</tr>
</tbody>
</table>

4.3. Macroeconomic stress testing

The first part of the modeling process is the estimation, testing and validation. A sound econometric model is the foundation for an accurate forecasting and stress-testing model. Once explicit dependence between inputs and outputs has been estimated, four types of stresses can be considered: i) structural stress or shock on parameters; ii) macroeconomic or outside environment; iii) policy changes or internal events; and iv) business strategy decisions. In this paper, we concentrate on stress testing based on macroeconomic shocks, also called economic scenario-based forecasting, similar to the U.S. Federal Reserve Comprehensive Capital Analysis and Review.
Once the link between performance and economic variables is estimated, macroeconomic forecasts are applied to the models in order to generate pool-level forecasts that are based on forecasts of economic drivers. Extending this principle to stress tests—which really constitute little more than pessimistic economic forecasts—is equally simple. This process delivers a set of vectors under each alternative macro scenario, including a baseline projection. The models can be applied to custom scenarios provided all necessary macroeconomic variables are covered by the scenario. The particular shape of these forecasts and scenarios depends on the estimated elasticity of the risk vector to the included macroeconomic series. As mentioned earlier, the elasticities have been found to be, in many cases, heterogeneous across different collateral groups.

To illustrate these features, we present forecasts and scenarios for a small number of pools drawn from auto prime and subprime collateral types under two alternative macroeconomic scenarios—the current baseline forecast (a recovery) and an alternative, depression-like event. These projections are depicted in Figures 9 and 10 for two deals from auto prime and subprime lines.

Under the baseline scenario, prepayment rates are projected to increase in 2013 while they will initially decrease under depression scenario, S4, and then come back to the normal level as the life cycle suggests. Under baseline assumptions, we find that the default rate will stay relatively flat or slightly increase in 2013. A depression scenario, however, generates a significant increase in the next three or four years. Note that starting levels and stress vary by product and are a function of deal-specific characteristics for the deals chosen (i.e. age, WAC, etc.).

**Figure 9. Auto prime**

- **60-day Delinquency Rate**
- **CPR**
- **CDR**
- **Loss Severity**
Figure 10. Auto subprime

60-day Delinquency Rate

CPR

CDR

Loss Severity

[Graphs showing trends over time for 60-day Delinquency Rate, CPR, CDR, and Loss Severity for Baseline and S4]
5. Final Comments

In this paper, we outline the modeling approach that the Moody's Analytics DPLC product uses to forecast and stress-test the cash flow backing U.S. vehicles ABS deals including auto loans (prime and subprime), auto leases, motorcycles, boats and RVs. To build our econometric model, we use standardized pool-level data available for all outstanding and paid-off U.S. vehicles deals that Moody’s Investors Service has rated since its inception. The data files includes more than 1,500 pools and more than 70,000 time series/cross section observations that allow us to build a robust process.

The econometric approach separates the inputs into three factors: life cycle, credit quality and the economic environment. We consider average loan characteristics, economic conditions at loan origination, past pool performance, and dynamics in the macroeconomic environment over time to forecast prepayment, default and recovery vectors. The main purpose of the econometric models is forecasting and stress testing; hence our overarching goal is to maximize precision and accuracy while taking into account causal relationships. We highlight our variable selection process and explain the details surrounding the selection of economic factors. Standard out-of-sample validation results are presented as well as the process in place to re-examine the model and make adjustments, recalibrations or redevelopment as needed.

This technical report should be important for market participants interested in monitoring and valuating bonds either for regulatory compliance and/or buy-hold/sell strategies. The direct application of DPLC is valuation of outstanding deals: These vectors provide all necessary data to run waterfall valuation engines and thus compute fair value and expected loss under baseline as well as stressed economic conditions. The vectors can also be used indirectly to generate industry-level numbers and/or for benchmarking. The off-the-shelf monthly projections are available for deals outstanding and are based on an array of deal-specific factors; thus they cannot be used for the purpose of pricing bonds to-be-structured unless additional work is conducted. Finally, DPLC provides off-the-shelf, scenario-based collateral projections, all consistent with macroeconomic assumptions generated by a team of economists led by Moody’s Analytics Chief Economist Mark Zandi.
References


Moody’s Analytics
Economic & Consumer Credit Analytics

Moody’s Analytics helps capital markets and credit risk management professionals worldwide respond to an evolving marketplace with confidence. Through its team of economists, Moody’s Analytics is a leading independent provider of data, analysis, modeling and forecasts on national and regional economies, financial markets, and credit risk.

Moody’s Analytics tracks and analyzes trends in consumer credit and spending, output and income, mortgage activity, population, central bank behavior, and prices. Our customized models, concise and timely reports, and one of the largest assembled financial, economic and demographic databases support firms and policymakers in strategic planning, product and sales forecasting, credit risk and sensitivity management, and investment research. Our customers include multinational corporations, governments at all levels, central banks and financial regulators, retailers, mutual funds, financial institutions, utilities, residential and commercial real estate firms, insurance companies, and professional investors.

Our web and print periodicals and special publications cover every U.S. state and metropolitan area; countries throughout Europe, Asia and the Americas; and the world’s major cities, plus the U.S. housing market and other industries. From our offices in the U.S., the United Kingdom, and Australia, we provide up-to-the-minute reporting and analysis on the world’s major economies.

Moody’s Analytics added Economy.com to its portfolio in 2005. Its economics and consumer credit analytics arm is based in West Chester PA, a suburb of Philadelphia, with offices in London and Sydney. More information is available at www.economy.com.