Affirm Loan Level Transition Model

David Tan

A technical paper for Affirm's Head of Risk

Overview

One of the biggest problems faced by Affirm's risk management team is to measure the performance of the loans including how much to lose to charge-offs, how quickly a loan is repaid, and how the future cash flows look like. In this report, we present a new loan level transition model to project paid-off /charged-off rates and analyze the resulting distribution of cash flows over the loan lifetime. This helps us to manage risk and estimate how much money we expect to receive month by month.

Affirm's loan level transition model takes each delinquency state (condition of the loan) into account as well as borrower's payment history, loan level attributes, and macroeconomic conditions to determine the probability of transitioning from one state to another (Figure 1). Traditional econometric models were solely relying on loan origination attributes and separated loans into different groups based on origination terms which treated all loans within one group homogenous. By design, our loan level transition model can produce more accurate predictions by scoring individual loans along with a Cash Flow Engine to generate expected principal and interest payment according to loan amortization logic.

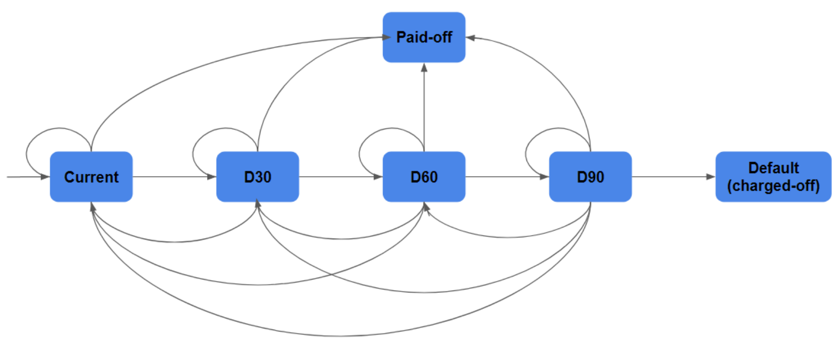

Figure 1: Loan transition changing state

Model Methodology

Affirm's new loan level transition model consists of four parts:

- Delinquency: The borrower has missed payments and are 30, 60, or 90 days beyond payment schedule. We model loan state from current to D30, D30 to D60, D60 to D90, and D90 to D120+ (charged-off). The term "current" refers to the borrower has made all the payments as scheduled thus far.
- Charged-off (default): The borrower is more than 120 days late on the payment schedule, at which point we write the loan amount off as a loss and the account is closed to future charges.
- Curing: We model D30 to current, D60 to current, D60 to D30, D90 to current, D90 to D60, and D90 to D30. While other events do happen, the majority curing events are D30 to current.
- Paid-off (including prepayment): The loan is fully paid off by the borrower. Note that for each month we expect there is a finite probability that a borrower will pay off all the loan early.

We first construct the loan delinquency pipeline to cover the entire life cycle from origination to termination (either paid-off or charged-off). We model the probability of making or missing a payment, along with fully paid-off and catching up on late payment (curing), conditional on the latest state that a loan is on. For each state, our model estimates the conditional probability of transitioning from a given row state to a column state (Figure 2). We develop an explicit model for each transition and each model can be developed by various machine learning methods. The probability of remaining in the current state is the residual probability after considering all other transitions. For example, if the borrower is current, he/she can make a scheduled payment (stays at current), miss a payment (becomes D30), or prepay (paid-off early). The likelihood of remaining current is the residual probability.

		State at Month t+1					
		Current	D30	D60	D90	Charged-off	Paid-off
State at Month t	Current	R	f(x)	0%	0%	0%	f(x)
	D30	f(x)	R	f(x)	0%	0%	f(x)
	D60	f(x)	f(x)	R	f(x)	0%	f(x)
	D90	f(x)	f(x)	f(x)	R	f(x)	f(x)
	D120+ (charged-off)	0%	0%	0%	0%	100%	0%

Figure 2: Transition matrix. f(x) = explicitly modeled. R = residual probability.

Model Variables

We include three suites of variables: origination (static), dynamic loan variables (time varying), and macroeconomic variables. Each suite of variables plays a particular role depending on the transition. For example, we have observed that the origination variables are strong indicator in early delinquencies but less so in later transitions like D60 or D90. The following figure (Figure 3) summaries the primary variables that are being used in our models. We also include interaction terms to capture different effects. Please note that the variable list is subject to change according to the actual model performance.

Model Variables	Cure	Delinquency	Default	Prepayment
Static Variables				
Original FICO score			x	x
Documentation			x	x
Original loan size	x	x	x	x
Seasonality index	x	x	x	x
Mortgage holder indicator				x
Geographic state		x	x	x
Dynamic variables				
Loan age (in months)	x	x	x	x
Remaining loan balance	x	x	x	x
Months in state		x	x	x
Total balance at the end of the month	x	x	x	x
Number of missed payments		x	x	x
Recent Payment behavior	x	x	x	x
Payment to due ratio	x	x	x	x
Payment to ending period balance ratio	x	x	x	x
Credit limit increase indicator	x	x	x	x
Out of the top 2 loan balances for the borrower			x	x
Average daily balance on accts for D30		x		x
Average daily balance on accts for D60		x		x
Average daily balance on accts for D90		x		x
Macroeconomic variables				
Interest rates month-over-month change				x
House price index month-over-month change				x
Unemployment rate	x	x	x	x
Wages inflation	x	x	x	x
Existing home sales volumn	x	x	x	x

Figure 3: A list of variables affecting different components of the model

Model Implementation

The model is implemented through Monte Carlo simulation at the loan level, which generates forward delinquency paths. This is necessary because our model is based on a borrower's entire payment history and months in state (number of missed payments). Below is an example to illustrate how the simulation works. According to the transition matrix (Figure 2), a borrower who has entered D60 is facing 5 options: make a payment (stays at D60), miss a payment (moves to D90), cure - make back one additional payment (moves to D30), cure - make back two additional payments (moves to dirty current), and paid-off. If we assign probabilities to these cases, it might look like:

P(make a payment) = 0.05, P(miss a payment) = 0.9, P(cure 1 month) = 0.03, P(cure 2 months) = 0.01, P(paid-off) = 0.01.

The next step is to randomly choose a transition by drawing a random number between 0 and 1 and compare it to the above probabilities. If the number is less than or equal to 0.9, the borrower misses a payment and enters D90. If the number is greater than 0.9 but less than or equal to 0.95, the borrower makes a payment and stays in D60. If the number is greater than 0.95 but less than or equal to 0.98, the borrower makes additional payment and is cured to D30. If the number is greater than 0.98 but less than or equal to 0.99, the borrower makes two additional payment and is cured to current. If the number is greater than 0.99, the borrower pays off the loan entirely. Once the borrower's new delinquency state is determined and payment history is updated, the simulation process is repeated until the loan reaches termination. One complete sequence of draws leading to termination represents a single path. We choose to simulate 150 paths per loan and then average across paths and loans to calculate the total expected loss and total cash flows. Please note that for interest-bearing loans, we apply loan amortization logic to estimate the expected interest payment month-by-month.

Model limitation

One advantage of simulation is the ability to use a borrower's full payment history and track months in state without increasing the dimensionality of the transition matrix. However, given 150 paths per loan and multiple delinquency buckets for millions of loans over long terms, we easily run into the computation scaling challenges. Each transition matrix represents a random number draw in one path and requires us make model predictions four times, once for each previous state. In addition, when running Cash Flow Engine, suppose we have 2 million of loans and average maturity term is 12 months, the size of our training dataset would be 2 million * 12. Hence, we need to run a billion of model predictions for the simulation. Some possible solutions to this computation challenge include open a large data container and leverage the latest Spark techniques.

Summary

Affirm's loan level transition model tracks the progression of an individual loan through the entire chain of borrower actions – from origination to termination. The model structure is based on a transition matrix and each transition is modelled explicitly. The model features include static and dynamic loan attributes, as well as macroeconomic variables. We also do backtesting on historical snapshots to check how well the model predicts. It is implemented through Monte Carlo simulation at the loan level, which allows us to project future payment behavior and cash flows.