Multistate loan transition model

When constructing a “healthy” loan portfolio, probability of default (PD) is one of the primary factor to consider. PD describes the likelihood of default for a loan. In the long period, we want to estimate the probability of default by tracking the loan status’s movement at each time t>1 until the loan being paid off or terminated. This is determined by multi-state loan transition model. In this write up, it provides a high level of overview of probability of default transition framework, together with its use case.

**Methodology**

Before digging into multi-state loan transition framework, two concepts are described here which constitute the foundation of multi-state loan transition model: Markov model and multinomial logistic model.

Markov model

The assumption of Multi-State loan transition model is it follows Markov process. Markov model assumes future event depend only on the current state but not the past state. The model can be formalized as following:

Where represents the value current state, and represents value at any state prior to .

Multinomial logistic regression

Multinomial logistic regression generalize the logistic model with two outcomes to more than two outcomes. A general form of multinomial logistic model is as follows:

……

where K represent outcome level and X represents a vector of covariates. β is the logodds when covariate increase by 1 unit.

Multi-state loan transition model

Putting Markov model and multinomial logistic regression model together, the multi-state loan transition model can be constructed to model Affirm loan portfolio. To start with, we defined the following state of a loan:

1. Current: all of the principal and interest that due being paid on time; there is no miss-payment.
2. Delinquent: one or more payments are missed. Delinquency is measured in days, such as 30 days delinquent, 60 days delinquent, .etc.
3. Charge off: the loan is in outstanding status, this often happens when loan hits 120/180 day past due.
4. Prepaid: the loan is prepaid before its official due date.

We want to model the transition path of a loan along the timeline until it is paid off or terminated. Below is a full transition cycle for a loan starting from current. For simplicity, we consider a loan as charge-off once it passed 60 days delinquency.

Prepay

CurrentNNTNNT

Dq30

Dq60NNTNNT

ChgoffNNTNNT

In that, we want to track all the possible transition state for a loan. For example: a loan in current state can either stay in current, or transit into Dq30, or being prepaid. The table below provides the corresponding possible delinquency transitions from time t to t+1.

**Transition matrix**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Current | Dq30 | Dq60 | Dq90 | Dq120 | Charge-off | Prepaid |
| Current | current | C- Dq30 |  |  |  |  | C-Prepaid |
| Dq30 | Dq30-C | Dq30- Dq30 | Dq30- Dq60 |  |  |  | Dq30-Prepaid |
| Dq60 | Dq60-C | Dq60- Dq30 | Dq60- Dq60 | Dq60- Dq90 |  |  | Dq60- Prepaid |
| Dq90 | Dq90-C | Dq90- Dq30 | Dq90- Dq60 | Dq90- Dq90 |  |  | Dq90- Prepaid |
| Dq120 | Dq120-C | Dq120- Dq30 | Dq120- Dq60 | Dq120- Dq90 | Dq120- Dq120 | Dq120- Charge-off | Dq120- Prepaid |

For a current loan, it can either transition into Dq30, or pay the balance and remain in current, or prepays and close the account.

For the loan in Dq30 state, the loan can either cures by making two payment, or stays at Dq30 by making one payment, or transition to Dq60, or prepays.

The multi-state loan transition model is then model the transition probabilities from any time , to each of the possible state at t+1. And it assumes markov process hold, that the next state of loan only depends on current state regardless of previous state, which means

For each of current/delinquency state at time t, a multinomial logistic regression which described in previous section, or tree based models are used to predict transition probabilities. A separate multinomial logistic model is fit for each starting point. Using the above transition matrix as an example, at time t, there are 5 possible states, that state (current, Dq30, Dq60, Dq90, Dq120), for each state, a multinomial logistic model is built, thus there are 5 multinomial logistic regressions in total. When building classification models, loan attributes will be added into model as covariates. The loan attributes including lender’s credit score, loan term length, and loan interest rate, which can be dynamic or static.

**Implications**

Predicting transition probabilities can help company optimize the loan portfolio and reduce credit risk.

With loan transition probabilities, we are able to estimate future cash flows, further calculate the potential gain or loss with each loan application. For each loan, conditional on the probabilities that predicted in previous section, the cash flow at t = 2, 3, … can be iteratively calculated until all the loans being prepaid or terminated. Lastly, the charge off amount can be obtained for each loan and consider as the loss amount. By that, we sort loss amount from minimal to maximum, and select the loans with low loss.