Approaches in Loan and Vintage Level ROI Forecasting – A Consumer Lending Industry Perspective

Sridhar Sankranti, Axcess Financial, Cincinnati, OH

ABSTRACT

This paper discusses various commonly used approaches to loan level and vintage level ROI forecasting (cumulative cash-on-cash return) in consumer lending industry. Typical consumer lending products considered in this paper include personal loans (fixed term installment loans) and equipment leasing with product term ranging from 6- to 18-month duration. The expected returns can be forecasted at loan (customer) level or at origination vintage level (aggregation of loans within a common origination period). Furthermore, the forecasts can be developed at the time of origination itself or after the first payment performance becomes available (static forecasts) or on a continuing basis (adaptive forecasts) as the loans age to maturity. While this paper presents SAS code and discussion on different approaches to forecasting, it pays special attention to tracing the path of a vintage lifecycle by week-on-book and presents a SAS-based algorithm to construct dynamic cumulative ROI curves that are updated on a real-time basis and deployed in forecasting of unaged months on book.

INTRODUCTION

Not all loans in a lending product portfolio are expected to perform the same way even though all the loan approvals are subject to the same stringent underwriting policies and criteria. Customers vary quite widely within the approval range - as graded by a risk score - and this mix can gravitate to either side of risk spectrum. Added to this are external factors (unemployment rates, tax season effect, regulation etc.) and internal factors (IT, operational, and servicing issues affecting collections performance) that impact the loan repayment performance beyond the first payment.

Although terminal ROI (ROI realized by the end of the product term) is the eventual metric of interest for risk managers, this metric can be arrived by tracing the performance at every month on book up to the term end or by directly predicting the terminal performance by mapping the initial attributes such as risk score and first payment performance. For example, vintages with customers scoring above the historical average risk score or vintages with lower than historical first pay default (FPD) are expected to yield higher than average ROI.

Given the nature of inputs that go into the predictions, the approaches can be classified into two broad categories – Static vs. Adaptive forecasts. As the names imply, Static forecasts are fixed expectations on terminal ROI based on the loan application information while Adaptive forecasts 'adapt' by adjusting the forecast considering periodical repayment performance during the length of the product term.

Again, the Static and Adaptive forecasts can both be developed on loan level data or the aggregated loan pools depending on computational heavy lifting that the resources and data would permit.

SINGLE VARIABLE STATIC FORECASTS

Static forecasts most commonly use performance on first scheduled payment as an indicator of expected return on a given origination vintage. This involves developing a single variable regression equation on historical weekly or monthly vintages. The functional form of this equation depending on the line of best fit could take the of form of a straight line, logarithmic, polynomial etc.

Methodology:

1. Collect data on loan level first payment delinquency indicator (0/1), loan amount financed, total amount collected during the loan term, and the vintage month identifier
2. Calculated the loan level ROI by dividing total payment amount collected by the loan amount financed
3. Calculate the weighted mean of first payment delinquency (FPD) and ROI using the loan amount as the weight. The means are calculated by vintage month.
4. Develop appropriate regression equation:

$ROI_m = a + b*FPDRate_m$

5. Forecasting new vintages: Use the above developed equation to predict new vintage using the FPD mean as the input

SAS Code:

```
* Get Data;
data LoanData;
	set LoanData (keep=LoanDate LoanAmount SumOfPmts FPD);
	ROI = SumOfPmts/LoanAmount;
	VintageMonth=year(LoanDate)*100+month(LoanDate);
run;
* Vintage Aggregation - Mean ROI and FPD;
proc summary data=LoanData nway;
	class VintageMonth;
	var ROI FPD/weight=LoanAmount;
	output out=VintageSummary (drop=_type_ _freq_)
			mean(ROI)=VintageROI
			mean(FPD)=VintageFPD;
run;
* Regression;
ods graphics on;
proc reg data=VintageSummary outest=parms (rename=(intercept=parm_a
VintageFPD=parm_b1));
	where VintageROI ne .;
	model VintageROI=VintageFPD;
	output out=b p=yhat;
run;
* Scoring new vintages;
data scoring;
	if _n_=1 then set parms (keep=parm_a parm_b1);
	set VintageSummary (where=(VintageROI=.));
	P_VintageROI = parm_a + parm_b1*VintageFPD;
run;
```

MULTIPLE VARIABLE STATIC FORECASTS

In addition to the first payment performance, several other variables obtained from loan underwriting, seasonality, payment instrument type etc. can be used in the regression model in predicting the terminal ROI of the given loan. This differs from the single variable, FPD based static model by the fact that the equation would be developed on loan level dataset using a number of predictive variables.

$ROI = a + \Sigma(b1_i * FPD_i) + \Sigma(b2_i * RiskScoreBand_i) + \Sigma(b3_i * AgeGroup_i) + \Sigma(b4_i * IncomeGroup_i) + \Sigma(b5_i * PaymentType_i) + \Sigma(b6_i * NewRetCustType_i) + \Sigma(b7_i * TaxSeasonIndicator_i) + \Sigma(b8_i * HolidaySeason_i)$

The model can be estimated using Proc Reg or Proc GLM. However, while binned or categorical variables can be specified in class statement in Proc GLM, the former would require creation of dummy variables for all bin types.

While the predicted ROI should be bounded by 0 in the lower end and a meaningful higher bound such as 200% (depending on the interest rate), it is possible to see values out of this range that needs to be taken care of.

ADAPTIVE FORECASTS – STATE TRANSITION PROBABILITIES MODEL

Adaptive forecasts model cashflow by predicting cashflows in every month-on-book (MOB) for a given loan or vintage. Every loan's MOB_n can be classified into a finite number of possible states, such as – Current, DQ30, DQ60, DQ90, Default, and Paid-Off. By examining the historical roll-rates or transitions from one state to the other, it is possible to calculate probabilities of each state given the previous month's state. Multiplying these probabilities with scheduled payment amount in each month, we can arrive at expected total cashflow amount in each month of the loan term.

This methodology makes possible drill down approach to forecasting by modeling loan level cashflows over time. The loan level forecasts can in turn be neatly aggregated into store segments, regions, campaigns etc.

VINTAGE CURVES AND FORECASTING

<<WORK IN PROGRESS>>

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Sridhar Sankranti
Axcess Financial
513-229-6628
ssankranti@tempoe.com