CREDIT LIFE CYCLE FORECASTING MODEL

Final Presentation

Research practice 3

Carolina González-Restrepo & Milton Alfonso Martinez-Negrete

June 07, 2016

Bancolombia - EAFIT University

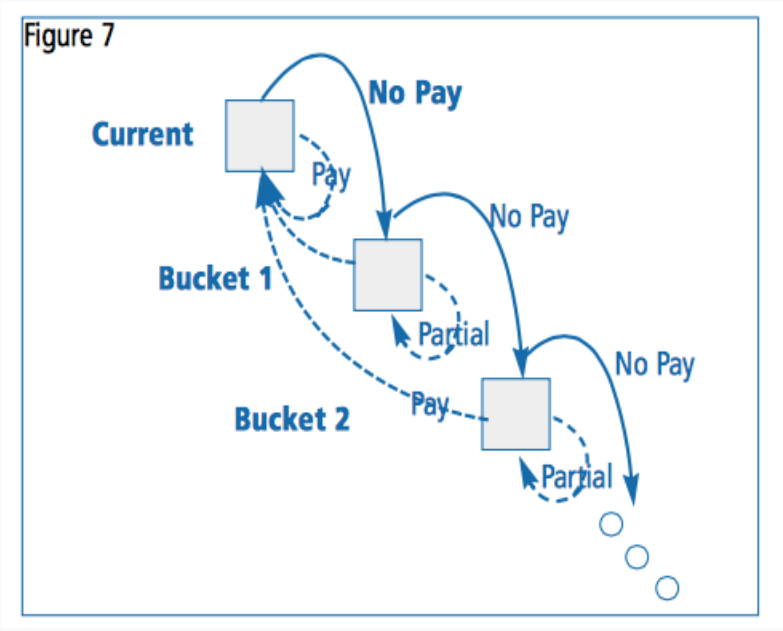
Introduction

Bancolombia counts with a strategy for an integral administration of risks, that points towards the identification, measurement, monitoring and mitigation of the inherited risks of the organization.

- **Credit risk**: it is the possibility of incurring in losses when a third party fails to fulfill its obligations partial or totally. This translates in deterioration of the credit quality.

Forecasting Model

The model considers tendency and seasonality of the events occurred in the past to forecast the future using **Markov Chains**, its important to precise that the model has quarterly forecasts



Discrepancies were noticed and external parameter was included, this parameter is moved in a very empirical way.

figure:www.strategicanalytics.com/pdf/RMAJ200310ForecastTools.pdf

Objectives

General: Develop an intervened a forecasting model for the credit life cycle concept(s).

Specific:

- Measure the quality of the actual
- Study different models capable of forecasting.
- Study external variables
- Extract all the information needed
- Implement an improved model
- Measure the accuracy of the model

Model Diagnosis

MAD

Type of error which measures the mean of the absolute deviations of the forecast errors

$$MAD = \frac{\sum_{i=1}^{n} |x_i - \hat{x}|}{n}$$

where:
x_i: real value
$\hat{x}$: forecasted value

RMSE

Quantitatively evaluates the accuracy of forecasts. This calculation, compared with MAD, amplifies and strongly penalizes those errors of greater magnitude.

$$RMSE = \sqrt[2]{\frac{\sum (x_i - \hat{x})^2}{n}}$$

Model Diagnosis

MAPE

As an measure of error independent of any scale is commonly used to evaluate and compare accuracy.

$$MAPE = \frac{\sum_{i=1}^{n} |\frac{e_i}{x_i}|}{n} * 100$$

where

$$e_i = y_{real_i} - y_{forecast_i}$$

Theil's U

U_1 which evaluates the forecast accuracy. U_1 is bound between 0 and 1, with values closer to 0 indicating greater forecasting accuracy.

$$U_1 = \frac{\sqrt[2]{\sum (x_i - \hat{x})^2}}{\sqrt{\sum x_i^2} + \sqrt{\sum \hat{x}_i^2}}$$

Diagnosis Results

Errors	REAL/FORECAST 2014-2015			
	RMSE	MAD	MAPE (%)	U1
Balance	128.269	85.407	2.8	0.084
Performing Loans	127.693	80.705	3.0	0.094
Disbursement	55.490	37.775	15.5	0.541
$>$ 30 past due	9.907	7.583	5.4	0.156
31-60 past due	5.037	4.582	11.1	0.249
$>$ 60 past due	6.816	5.486	5.8	0.160
A. Expense	3.509	2.692	27.0	0.686
Written offs	3.030	2.584	37.9	0.696

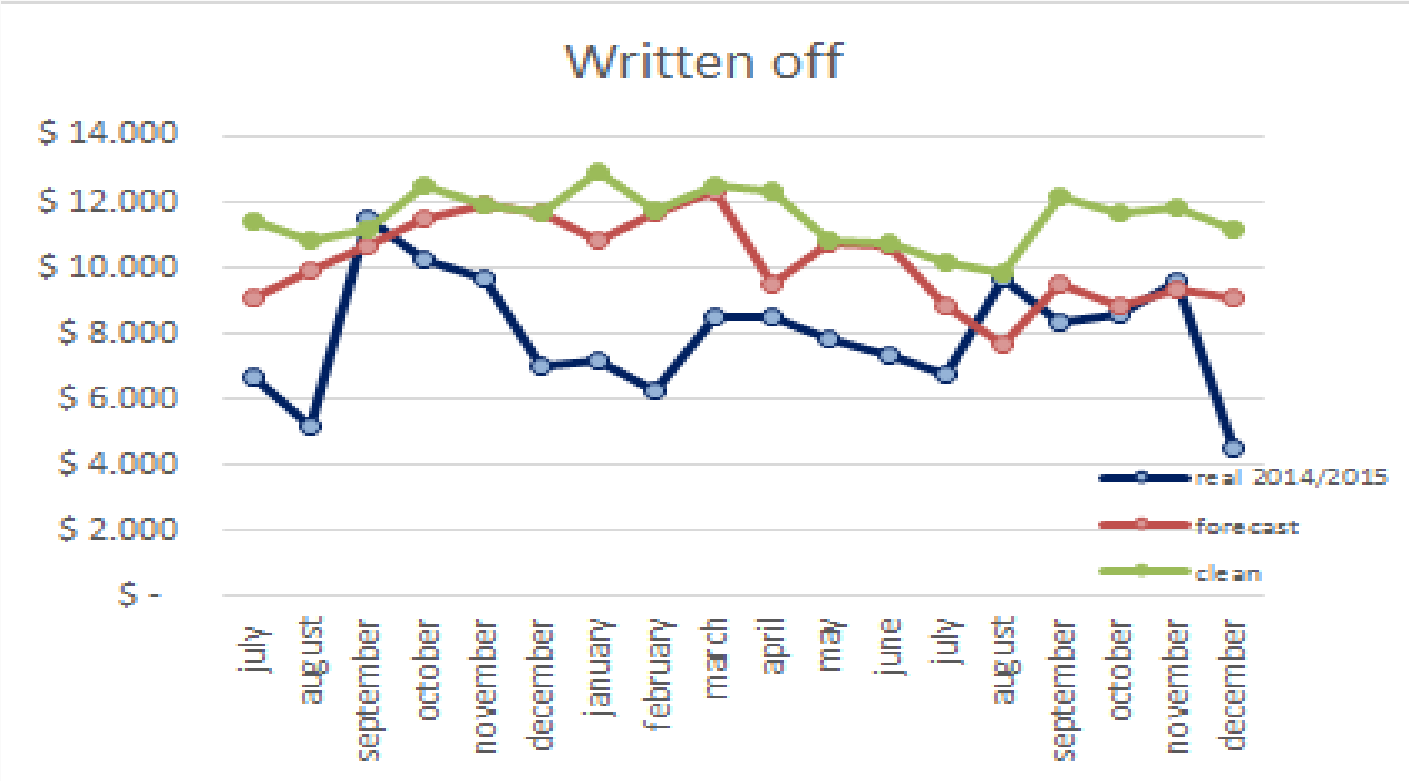
Table: Errors of the forecasting model

Diagnosis Results

Errors	REAL/CLEAN 2014-2015			
	RMSE	MAD	MAPE (%)	U1
Balance	128.269	85.407	2.8	0.084
Performing Loans	127.693	80.705	3.0	0.094
Disbursement	55.490	37.775	15.5	0.541
$>$ 30 past due	8.319	5.707	4.1	0.132
31-60 past due	5.085	4.500	10.7	0.250
$>$ 60 past due	5.556	4.829	5.2	0.129
A. Expense	3.902	3.294	35.9	0.719
Written offs	3.979	3.594	52.2	0.856

Table: Errors of the forecasting clean model

Diagnosis Graphic Results



Diagnosis Results

To evaluate the impact, a weighted error was calculated based on the allowance expense it generates. As the only precise information obtained is the total amount of allowance expense a historical distribution of the same across the concepts had to be calculated:

%	Disburs.	P. loans	31-60	>60	Cancel	W. offs
2014	65.6	-62.5	24.7	122.6	-50.3	-104.3
2015	70.4	-48.4	20.8	104.5	-47.3	-81.8
total	67.8	-56.2	22.9	114.4	-49	-94.1
Avg.	67.1	-52.4	22.3	109.4	-49.8	-88.9

Table: Historical distribution of the allowance expense

Diagnosis Results

There is a spending released each time a credit is canceled for this reason should take into account cancellations C_t:

$$C_t = B_{t-1} + D_t - B_t$$

where:
B_{t-1}: real balance $t-1$
B_t : forecasted balance t
D_t : forecasted disbursements in t

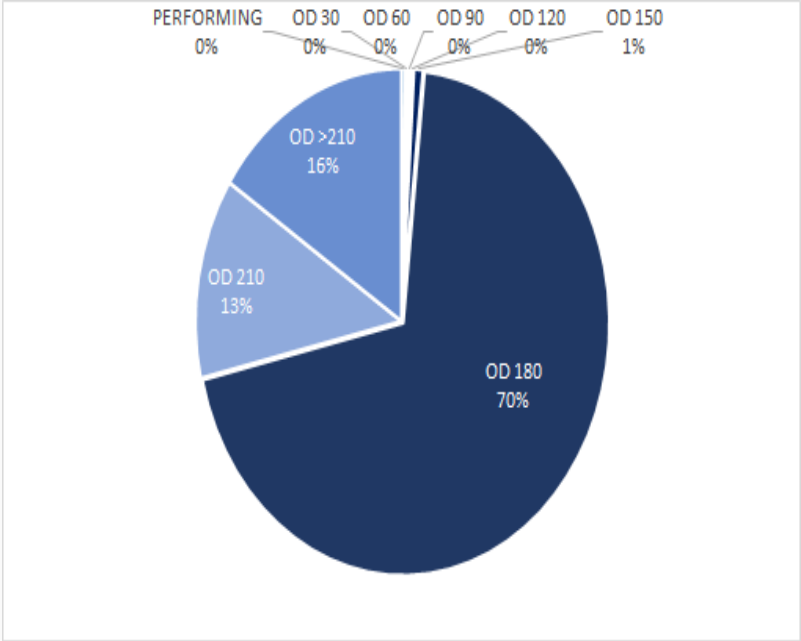
Diagnosis Results

CONCEPTS	MAPE	Expense (Millions)	Participation	W. Error
P. Loans	3.0%	-16.505	54%	1.6%
Cancel	18.5%	-15.717	51%	9.4%
Disburs.	15.5%	17.698	58%	9.0%
31-60	11.1%	7.571	25%	2.8%
>60	5.8%	37.506	123 %	7.1%
Written off	37.9%	32.641	107%	40.6%
Total		30.553		

Table: Weighted Error

Diagnosis Results

After analyzing the calculated errors and the weighted error, it can be clearly seen that the concept that has a greater error while being forecasted are the written offs. For these reason this will be the first concept to be intervened in the model.



where OD is the overdue loan in the different default buckets

Are normally used to make forecasts, due to the strong assumption that the values variables takes are the result of a tendential, seasonal and random component present in past observations.

Moving averages method

Used as a forecast average of n latest observations. where n is the number of periods backwards to be considered in the average.

$$M_t^n = \frac{y_t + y_{t-1} + y_{t-2} + \ldots + y_{t-(n-1)}}{n}$$

For forecasting through moving averages simply follow:

$$\hat{y}_t = M_{t-1}^n$$

Exponential smoothing method

The greatest weight is called α and is assigned to the immediately preceding observation from this assignment so on weights of $(1-\alpha)$, $(1-\alpha)^2$, $(1-\alpha)^3$ and so on until the last observation that will be consider [1].

$$P_{t+1} = \alpha Y_t + (1-\alpha)P_t$$

where

Y_t : value of the series in t

P_{t+1}: forecast in $t+1$

P_t : forecast in t

Time series

Before continuing implementing other time series forecasting techniques a test is performed in *R* using the function *aunto.arima*

$$result = ARIMA(0, 0, 0)$$

Linear Regression

$$y = b_0 + b_1 * x_1 + b_2 * x_2 + b_3 * x_3 + ...b_n * x_n + u$$

A collection of internal bank and macroeconomic variables such as DTF, CPI, GDP, among other was performed. In total there were 50 base variables that generated 500 variables which includes their respective annual, previous year and absolute variations and the remnants of themselves.

The biggest problem that may be found when performing linear regression is the collinearity; that can be expressed in terms of the correlation coefficients[6].

Correlation

Once raised the problem, a correlation test is considered necessary to determine which variables should be eliminated because they are already being represented by others within the model.

	TRM	TRM(t-1)	TRM(t-2)	TRM(t-3)
TRM	1	*0,9832*	*0,95621*	0,93439
TRM(t-1)	*0,9832*	1	0,9832	0,95621
TRM(t-2)	*0,95621*	0,9832	1	0,9832
TRM(t-3)	0,93439	0,95621	0,9832	1

Figure: Correlation matrix

The above process is performed for the 500 variables and a total of 196 variables were removed. The remaining variables will be considered in the linear regression.

Analysis of variance	
Source	p-value
model	< 0.0001

As observed in the last column, the p-value is less than 0.05 indicating a rejection in the null hypothesis, and therefore indicating the linear model is significant.

Forward selection method		
Step	Variable	P-value
1	V. abs OD>180	< 0.0001
2	V. py TRM_2	< 0.0001
3	V. abs DTF_3	< 0.0001
4	$CV60_2$	0.0342
5	V. abs OD> 180_2	0.0500
6	V. annual expense	0.0106
7	V. abs TRM	0.0112
8	V. abs $REPO_1$	0.0019
9	V. abs TES_3	0.0094
10	V. abs employment	0.0252
11	V. py petroleum	0.0016
12	V. annual employment	0.0043
13	V. annual GDP	0.0091

Table: Results of SAS model

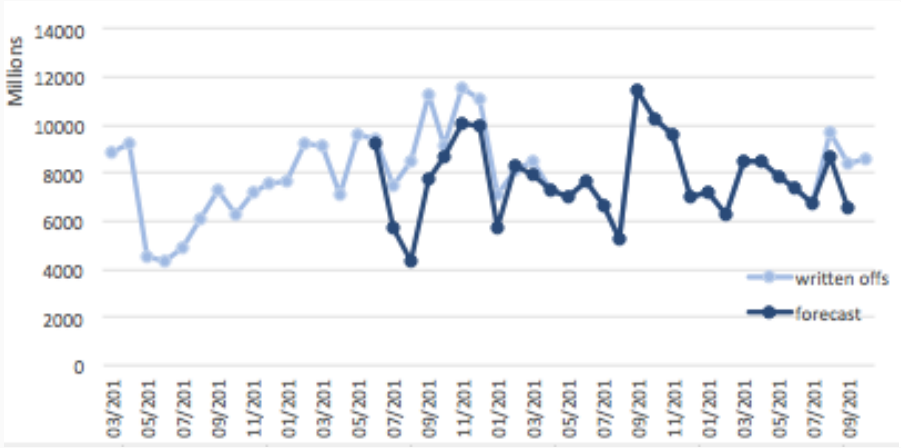
Analysis of variance	
Source	p-value
model	< 0.0001

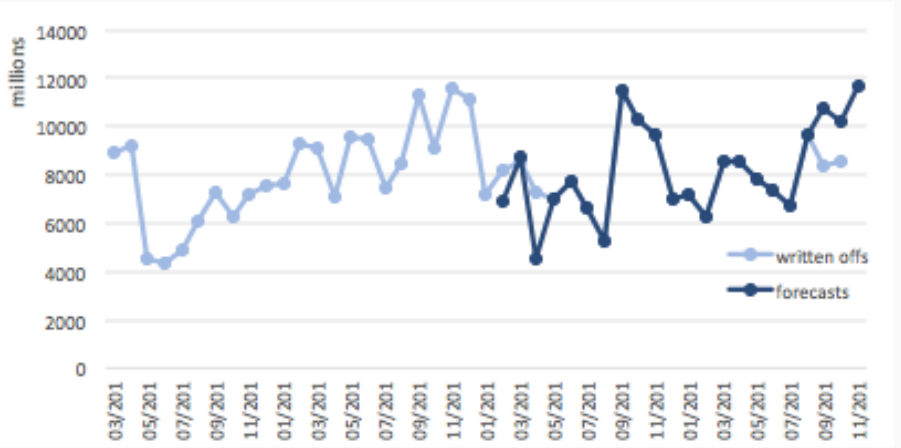
As observed in the last column, the p-value is less than 0.05 indicating a rejection in the null hypothesis, and therefore indicating the linear model is significant.

Forward selection method		
Step	Variable	P-value
1	V. py $colcap_2$	0.0015
2	V. py $colcap_3$	0.0085
3	V. abs TES_2	< 0.0001
4	V. annual TES_1	< 0.0001
5	v. annual $employ_2$	< 0.0001
6	V. abs $unemploy_3$	< 0.0001
7	V. abs $LICC_2$	< 0.0001
8	V. abs $LICC_3$	< 0.0001
9	V. py GDP_2	< 0.0001
10	V. abs OD_1	< 0.0001
11	OD 30_3	< 0.0001
12	V. abs OD 30_1	< 0.0001
13	V. abs OD 120_2	< 0.0001
14	$writtenoffs_8$	< 0.0001

Table: Results of SAS model

Linear Regression Results





Linear Regression Results

The following table contains the error measurements for the concept being estimated for both proporsed linear models.

Errors	Real/forecast intervened model			
	RMSE	MAD	MAPE (%)	U1
Written offs (contemponary)	786	634	5.46	0.234
Written offs (lagged model)	904	389	4.87	0.207

Table: Errors of the forecasting models proposed

Artificial Neural Networks Results

An artificial neural network was trained with the resultant significant variables obtained from the past linear model. The neural network was trained using 37 patterns and 5000 epochs but it did not ended in the expected time nor learned as expected. This might be explained because of the volatility of the series or output and the little amount of patterns or input information had.

Conclusions

The term with greater systematic error are the written offs, which generate a high cost to the bank as seen in the weighted error.

The written offs series of can not be modeled by a time series model since it lacks auto regressive integrated moving average. This makes written offs a random walk.

The models obtained with SAS are statistically significant as the p-values reject the null hypothesis in the ANOVA test, which support the fact of modeling and forecasting written offs with a linear model.

The linear model proposed is a feasible alternative due to its simple replicability for other bank products.

The results obtained by the unrestricted model in terms of lag are accurate but require forecasting other concepts as input to forecasts written offs. The results obtained by the lagged variables model are an alternative because they retain the accuracy and do not require contemporary variables.

The errors decreased significantly with the intervened forecasting model, evidenciating the fact written offs could be better estimated using methods different from Markov chains.

For the neural network it is recommended to generate more patterns, the generation could be by using a bootstrapping method or by running another linear regression model with a database containing variables with more observations.

REFERENCES I

L Allen.

Métodos de pronósticos.

Técnicas de Suavización, 1998.

J.L. Breeden.

Portfolio forecasting tools: What you need to know.

The RMA journal, pages 6–10, 2003.

J.C. ChambersSatinder, K. MullickDonald, and D. Smith.

How to choose the right forecasting technique.

Harvard Bussiness Review, pages 1–4, 1971.

Primitivo Reyes Aguilar.

Métodos de pronósticos.

Administración de operaciones, Agosto 2009.

Jose Manuel Rojo.

Resgresión lineal múltiple.

Laboratorio de Estadística, 2007.

Mario Orlando Suárez Ibujes.

Coeficiente de correlación de karl pearson.

Questions?