An Example Using Excel Stepwise Regression to Forecast High Risk Automobile Loan Losses

Kris K. Moore and Jonathan K. Trower

Information Systems Department
Baylor University
One Bear Place #98005
Waco, TX 76798-8005

Abstract
A multiple regression model to predict losses on high risk automobile loans is found using stepwise regression and seven independent variables for 769 loans. The independent variables used were A.P.R. (annual percentage rate), amount financed, automobile mileage, down payment, credit rating (Auto Zoom), pass time device, and whether or not the customer had a previous repossession.

Key Words: multiple regression, stepwise regression, automobile loan losses

1. Introduction

An automobile dealership in central Texas sells used automobiles to high risk customers that are unable to buy automobiles by standard procedures because of their bad credit. The business makes a profit on 78.31% of the loans, but the owner is concerned about the 21.89% (n=769) that result in a loss. The owner would like to forecast the loss and also determine the factors that are related to the loss on the loans.

2. Data

The problem described above can be solved using Excel Stepwise Regression. The dependent variable for the regression is the Repossessions Losses (Repo Loss). The loss on the loan in dollars is given as a positive number.

The independent variables are:
- A.P.R. -- Annual percentage rate charged for the loan
- Amount financed—dollar amount.
- Mileage—Mileage on the used automobile at date of the sale
- Down Payment—Amount customer pays at time of sale in dollars
- Payment Amount — Monthly payment on car loan
- Auto Zoom — Credit rating by the automobile industry (Higher the number the better the credit rating)
- Total Paid—Total paid at the completion of the loan or the time of repossession of the car
- Past Due —The number of late payments made by the customer over the life of the loan.
- Payoff Amount --- The amount to pay off the remaining balance owed on the automobile loan
- Pass Time Device—1 if a pass time device is applied to the automobile, 0 otherwise. The pass time device allows the dealer to disable the automobile so it

will not start if a customer fails to make payments for an extended time. The cost to the dealer is about $300.

- Repo—1 if the customer has had a repossession in the past, 0 otherwise

Table 1: Descriptive Statistics

Variable	Mean	St. Dev.	Minimum	Maximum
Repo Loss	$10,120.80	3,058.89	549.55	22,008.72
A.P.R.	20.8182%	2.3706%	11.71%	25%
Payment Amount	$343.7596	103.4852	79.35	636.74
Amount Financed	$13,901.0937	2,678.4580	6,481.85	23,916.03
AutoZoom Score	631.94	72.14	431	899
Down Payment	$1,010.44	757.57	0	4,400
Mileage	72,764.79	36,838.40	15	89,544

For the 769 loans with a loss, 136 of the customers had a previous repossession and 152 received a pass time device on their purchased automobiles. No one had both a repossession and a pass time device.

3. Methodology

Stepwise regression was used to build the model.

Stepwise chooses the best one variable, then the best two, and continues till all the variables have been used on the last step. In step one the variable with the highest correlation is selected because the variable has the highest coefficient of determination. On step two, all combinations of two independent variables are selected and the pair with the highest coefficient of determination is selected. Step three all pairs of three variables are used to find the three that produce the highest value for the coefficient of determination. This continues until all the independent variables are used on the last step.

The variables total paid, pay off amount and past due were not used in the stepwise because their values cannot be determined until the loan is finalized. The other eight variables for which values can be determined at the time of the loan were used in stepwise regression and mileage and payment amount were eliminated. The other six variables had the smallest standard error of regression (s=2063.063) and these six produced a prediction interval that was narrower than any other combination of the possible variables that could be used in the model. In other words, mileage and payment amount or significantly correlated with the other six variables and they did not contribute significantly to the explained variation. If the two eliminated variables are included, multicollinerity affects the interpretation of the model coefficients in addition to the problem of increasing the value of the standard error of the regression.

4. The Model

The model is valid significantly and the six independent variables explain 55.7% of the variability of repossession losses. In addition, the independent variables amount financed, pass time device, A.P.R., repo in past and down payment are significant in the presence of the other predictor. This means the variables can be interpreted regarding the effect of these variables individually on the repo loss.

Table 2: Regression Model

Repo Loss = -4181.96 + .08294* Amount Financed
$\qquad$ + 987.22 * Pass Time Device + 140.49* A.P.R.
$\qquad$ + 663 *Repo in Past – .02455 *Down Payment
$\qquad$ - .3862 *Auto Zoom

Regression Analysis

R^2	0.557		
Adjusted R^2	0.553	n	769
R	0.746	k	6
Std. Error	2063.063	Dep. Var. **REPO LOSS**	

ANOVA table

Source	SS	Df	MS	F
Regression	4,077,206,034.4393	6	679,534,339.0732	159.66
Residual	3,243,245,738.2391	762	4,256,228.0029	
Total	7,320,451,772.6784	768		

Regression output

variables	coefficients	std. error	t (df=762)	p-value
Intercept	-4,181.9564	1,154.0851	-3.624	.0003
Amount Fin.	0.8294	0.0310	26.754	1.04E-111
Pass Time Device	987.2236	208.4021	4.737	2.59E-06
A.P.R.	140.4900	33.4715	4.197	3.02E-05
Repo?	663.2045	206.9431	3.205	.0014
Down Payment	-0.2455	0.1012	-2.425	.0155
AutoZoom	-0.3862	1.0668	-0.362	.7174

5. Conclusions

The larger the down payment the lower the repossession losses for this dealership. For every dollar of down payment, the repo loss is reduced 24.55 cents. As the amount financed increases by a dollar the potential repo loss increases by 82.94 cents. If a customer has had a repossession in the past, the repo loan increases by $663.20 on the average. Also the pass time device averages $987.22 of contribution to the overall loss. For every percentage point increase in A.P.R. the repo loss increases by $140.49.

The model indicates loans should not be given to customers with a repossession in the past and loans should be refused to those customers qualifying for a pass time device. Also high A.P.R. is working against the overall profit. The owner should consider reducing the A.P.R. below the current 11.71% minimum that he has charged in the past. This would give the high risk customers a better chance of paying off their loan and reduce the owners overall loss.

References

Anderson, David R., Sweeney, Dennis J., and Williams, Thomas A. (2010) Modern Business Statistics, Cengage Southwestern, 3rd edition.

Keller, Gerald. (2010) Statistics for Management and Economics, Cengage Southwestern.

Krehbiel, Timothy C., Ng, Pin T., Berenson, Mark L., Stephan, David, and Levine, David M. Statistics for Managers Using Microsoft Excel, (2008) 5th Edition, Prentice-Hall Inc.

Lind, Douglas A., Marchal, William G., and Wathen, Samuel A. (2011) Statistical Techniques in Business and Economics. McGraw-Hill, 14th edition.

Ragsdale, Cliff T. (2007) Spreadsheet Modeling and Decision Analysis, Southwestern, 5th Edition.