THREE-WAY ROC ANALYSIS USING SAS SOFTWARE

Juraj Kapasný, Martin Řezáč

Received: August 28, 2013

Abstract


The most commonly used measure of model accuracy in medicine with three categories of target variable is the volume under ROC surface (VUS), which is the extension of the area under curve (AUC) for binary models (Le and Lili, 2013). This paper deals primarily with usage of the multinomial logistic regression and the three–way ROC analysis in the financial sector, especially in the credit risk management. Moreover, SAS system is very often used software in the financial sector. Therefore this paper is focused on ways of doing three–way ROC analysis in this statistical software, in particular on estimating the VUS.

We propose an estimate of the VUS based on the confusion matrix, which is compared to estimates based on Mann-Whitney statistic and on empirical distribution functions. We developed three SAS macros based on these approaches for computing the VUS. Furthermore, we developed some logistic models for three-value target variable based on the Loss Given Default (LGD). This was done on real financial data. Results obtained by the SAS macros on these models are presented and discussed in the paper.

multinomial logistic regression, LGD, Three-way ROC analysis, ROC surface, VUS, SAS Software

INTRODUCTION

In medicine, there are often used models with three categories target variable and three–way ROC analysis with VUS is proven to be one of the best measurement techniques for these models. It is quite different in financial mathematics or banking. Many banks use classic binary logistic regression for the prediction of the probability of default so they are able to use ROC curves and also AUC (see Řezáč and Řezáč, 2011). However, some banks have recently started using also multinomial logistic regression and therefore they also had to start using VUS and three-way ROC analysis as the measure of quality (predictive power) for theirs models. This paper reacts to these events and presents some ways of running three–way ROC analysis in SAS software, that is software very commonly used in financial sector, especially in banking. Scurfield (1996) and Mossman (1999) proposed and discussed the concept of the VUS, Ferri et al. (2003) discussed exact computation of the VUS compared with its approximations. Landgrebe and Duin (2006) did some experiments to find the lower bound of the VUS and to investigate integration approach to estimating the VUS. Waegeman at al. (2008) gave theoretical and experimental evidence of the advantages of the VUS compared to error rate, mean absolute error and other ranking-based performance measures. Li and Fine (2008) developed a method based on estimating the class probabilities. Wan (2012) proposed an empirical likelihood confidence interval for the VUS.

A typical area of usage of the multinomial logistic regression in banking is the credit risk management. Within this area, a very actual issue is the estimation of the LGD. When all defaulted clients are divided into three categories, first one around LGD = 0, second category with LGD ≤ 1 and third one with LGD higher than 1, the multinomial logistic regression can be used. The paper is devoted to this topic. We developed some logistic models for three-value target variable based on the LGD. Furthermore, we developed three SAS macros
Theory of multinomial logistic regression

The theory and all formulas in this section were retrieved from Hosmer and Lemeshow (2000). The main difference between multinomial logistic regression and classic binary logistic regression is that in multinomial regression with three categories there are two logit equations. Let \( x \) be the vector of values of independent variables, \( \beta \) the vector of parameters for every predictor plus intercept and \( Y \) the target variable with values 0, 1 or 2 then 2 logit equations can be written as

\[
\begin{align*}
g_1(x) &= \ln \frac{P(Y = 1 | x)}{P(Y = 0 | x)} = \beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p = x' \beta_1, \\
g_2(x) &= \ln \frac{P(Y = 2 | x)}{P(Y = 0 | x)} = \beta_{10} + \beta_{11} x_1 + \ldots + \beta_{1p} x_p = x' \beta_2,
\end{align*}
\]

(1)

and

\[
\begin{align*}
g_3(x) &= \ln \frac{P(Y = 1 | x)}{P(Y = 2 | x)} = \beta_{0} + \beta_{21} x_1 + \ldots + \beta_{2p} x_p = x' \beta_3, \\
g_4(x) &= \ln \frac{P(Y = 2 | x)}{P(Y = 1 | x)} = \beta_{10} + \beta_{11} x_1 + \ldots + \beta_{1p} x_p = x' \beta_4.
\end{align*}
\]

(2)

By solving equations (1) and (2) together with

\[
P(Y = 0 | x) + P(Y = 1 | x) + P(Y = 2 | x) = 1
\]

we get the solutions for probabilities for each category, which equal to

\[
P(Y = 0 | x) = \frac{1}{1 + e^{x' \beta_1} + e^{x' \beta_2}},
\]

(3)

\[
P(Y = 1 | x) = \frac{e^{x' \beta_1}}{1 + e^{x' \beta_1} + e^{x' \beta_2}},
\]

(4)

\[
P(Y = 2 | x) = \frac{e^{x' \beta_2}}{1 + e^{x' \beta_1} + e^{x' \beta_2}}.
\]

(5)

It can be observed that every probability is function of \( x \) and has \( 2(p+1) \) parameters \( \beta' = (\beta_1', \beta_2') \), where \( p \) is the number of independent predictors. Maximum likelihood estimation is then used to estimate values of parameters \( \beta \). For creating a likelihood function 3 dummy variables \( y \) are created according to following pattern: If \( Y = 0 \) then \( y_0 = 1, y_1 = 0 \) and \( y_2 = 0 \); if \( Y = 1 \) then \( y_0 = 0, y_1 = 1 \) and \( y_2 = 0 \); and if \( Y = 2 \) then \( y_0 = 0, y_1 = 0 \) and \( y_2 = 1 \). Then likelihood function can be created as

\[
\ell(\beta) = \prod_{i=1}^{n} \ln \left[ \pi_0(x_i)^{y_0} \pi_1(x_i)^{y_1} \pi_2(x_i)^{y_2} \right],
\]

where \( \pi_j(x_i) \) equals to \( P(Y = j | x_i) \) and \( y_j \) are dummy variables mentioned before for observation \( i \). Using the fact that \( \sum_{j=0}^{2} y_j = 1 \) for every observation and using logarithm on \( \ell(\beta) \), we get a log–likelihood function

\[
\ell(\beta) = \sum_{i=1}^{n} \left[ y_0 \ln \pi_0(x_i) + y_1 \ln \pi_1(x_i) + y_2 \ln \pi_2(x_i) \right] - \ln(1 + e^{x' \beta_1} + e^{x' \beta_2}).
\]

(6)

We get our likelihood equations as partial derivations of function (6) with respect to all \( 2(p+1) \) parameters. To simplify these equations we use denotation \( \pi_j = \pi_j(x_i) = P(Y = j | x_i) \). Final likelihood equations are in form

\[
\frac{\partial \ell(\beta)}{\partial \beta} = \sum_{i=1}^{n} x_i (y_j - \pi_j)
\]

(7)

for \( j = 1, 2 \) and \( k = 0, 1, 2, \ldots, \ p \) and \( x_0 = 1 \) for every observation \( i \). We get maximum likelihood estimation \( \hat{\beta} \) by setting equations (7) equal to 0 and solving them. To solve these equations the numerical method has to be employed. It can be found in much statistical software.

From classic ROC to Three–way ROC analysis

In classic binary case ROC analysis is based on confusion matrix of model, corresponding with that in the Tab. I.

<table>
<thead>
<tr>
<th>Predicted category</th>
<th>T</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>True Positive (TP)</td>
<td>False Positives (FP)</td>
</tr>
<tr>
<td>F</td>
<td>True Negatives (FN)</td>
<td>True Negatives (TN)</td>
</tr>
</tbody>
</table>

Things are bit more complicated in multinomial case. Similarly it is also based on confusion matrix, whose dimension, however, is now \( 3 \times 3 \), therefore there is no equivalent of False Positive Rate or False Negative Rate, but there are 6 different kinds of diagnostic errors in three–way logistic model.

In binary case true positive rate can be computed as

\[
TPR = \frac{TP}{TP + FN}
\]

and false positive rate as

\[
FPR = \frac{FP}{FP + TN}
\]

and when TPR vs. FPR is plotted, we get “reversed” ROC curve, which is starting at upper left corner and bending down to reach lower right corner. This approach can be extended to three dimensions and the three positives rates can be plotted on orthogonal axes. If observed points are then connected with line segments, the result is ROC surface. Fig. 1 provides an example of ROC surface for one of these multinomial logistic models.
Three-way ROC analysis using SAS Software

This picture was retrieved from work of Le and Lili (2013) and symbols $\delta_i$ on each axis are equivalent to true positive rate for category $i$.

One can see in the Fig. 1, that ROC surface has three peaks (1, 0, 0), (0, 1, 0) and (0, 0, 1). For the first one holds true, $P(Y = 0 \mid \text{target} = 0) = 1$, $P(Y = 1 \mid \text{target} = 1) = 0$ and $P(Y = 2 \mid \text{target} = 2) = 0$, where “target” is the target variable with three categories. Similar holds true for other two points. These points are part of every ROC surface for each model. In multinomial case volume under ROC surface (VUS), which is an extension of AUC for classic binary model, can be computed and it is equal to probability that three randomly selected observations, one from each class, are assigned correct classes. Consequently, VUS for useless test has to be smaller than AUC and it equals $\frac{1}{6} = 0.1667$. If one observation from first category is selected, there is $\frac{1}{3}$ chance of placing it in a right bin. If this is done correctly, there is $\frac{1}{2}$ chance of placing observation from second category to the right bin. These probabilities can be multiplied: $\frac{1}{3} \times \frac{1}{2} = \frac{1}{6}$ and that equals to volume under ROC surface for useless test, which is created by connecting points (1,0,0), (0,1,0) and (0,0,1) and can be seen in Fig. 2.

This picture was retrieved from work of Le and Lili (2013) and symbols $\delta_i$ on each axis are again equivalent to true positive rate for category $i$. 

1: ROC surface for model with three categories dependent variable

2: ROC surface for useless test
Application of multinomial logistic regression

In this section a logistic regression model with dependent variable with three categories is created using SAS software and real financial data.

Data

To create the model the data from one of the biggest banks in Czech Republic were used. These data consist of nine variables which are id variable, lgd and seven predictors (independent variables) and 8050 observations. The variable lgd contains a value of LGD for each customer and a new variable LGD_target is constructed where every observation is assigned a class of our target variable 0, 1, or 2, hence multinomial logistic regression can be used to predict the category of LGD for new customers if they default. After computing \( \chi^2 \)-test for all variables it was found out, that they all have statistically significant relationship with the target variable, so it has to be carefully decided which to use in our model. Before the data is used for the model, it has to be divided into training part and validation part. Once this is done, the similarity of these tables can be measured by PSI (see Řezáč, 2009) which can be found in Tab. II.

In general, PSI value under 0.1 is considered as very good for two similar tables. Identical table can be created for every variable in the dataset.

Transformation of Variables

There were many missing values in the dataset, so first thing to do was to replace them with value "NA" in case of categorical variables and with median in case of numerical interval variable. Linear relationship between predictors and dependent variable were looked for, so it can be said that if a client moves from one category to another, his LGD will raise. To find this relation equivalent of WOEs (weights of evidence) were used. A new variable bad was created following the pattern: bad = 0 if LGD is less than 0.8 and bad = 1 if LGD is more than 0.8. Then WOEs for independent predictors with respect to new variable bad were computed and all independent variables were ordered by their WOEs to create 7 new predictors.

Modelling

In this section the data from previous section were used and a multinomial logistic regression model with target variable with three categories was built. Three multinomial logistic regression models were created, first one (Model 1) with all independent variables as predictors, second without the predictors which p-value was bigger than 0.05 (this means that just statistically significant predictors were used) in at least one of partial logistic regression and third without the predictors which p-value was bigger than 0.05 in both partial logistic regressions. The results can be found in Tab. III. In the following section they are examined to determine, which one is the best using three-way ROC analysis.

\( \beta \) coefficients shown in Tab. III were used for scoring of validation data and models were compared by values of VUS for training and validation data in section denoted "results". An important feature of the multinomial logistic model is that it estimates models, where \( j \) is number of categories of target variable. In this case the category zero was treated as the referent one and models for category one relative to zero and category two relative to zero were estimated. According to Bruin (2011) the standard interpretation of multinomial logistic regression is that for a unit change in the predictor variable, the logit of outcome \( m \) relative to referent group, in this case logit of outcome 1 and 2 relative to 0, is expected to change by its respective parameter estimate.

II: PSI index for variable LGD_target

<table>
<thead>
<tr>
<th>LGD_target</th>
<th># of clients</th>
<th>ratio</th>
<th>psi</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>2000</td>
<td>0.355</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1668</td>
<td>0.296</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1967</td>
<td>0.349</td>
<td></td>
</tr>
<tr>
<td>total</td>
<td>5635</td>
<td></td>
<td>0.00056</td>
</tr>
</tbody>
</table>

III: \( \beta \) coefficients for computed models

<table>
<thead>
<tr>
<th>Variable</th>
<th>target</th>
<th>#1</th>
<th>#2</th>
<th>#3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>-1.1209</td>
<td>-1.0868</td>
<td>-0.8655</td>
</tr>
<tr>
<td>2</td>
<td>2.8186</td>
<td>1.5113</td>
<td>2.4885</td>
<td></td>
</tr>
<tr>
<td>Predictor4</td>
<td>1</td>
<td>-0.1259</td>
<td>-0.1325</td>
<td>-0.1260</td>
</tr>
<tr>
<td>WOE</td>
<td>2</td>
<td>-0.2804</td>
<td>-0.3008</td>
<td>-0.2810</td>
</tr>
<tr>
<td>Predictor1</td>
<td>1</td>
<td>-0.0700</td>
<td>-0.0741</td>
<td>-0.0715</td>
</tr>
<tr>
<td>WOE</td>
<td>2</td>
<td>-0.1363</td>
<td>-0.1422</td>
<td>-0.1324</td>
</tr>
<tr>
<td>Predictor5</td>
<td>1</td>
<td>1.3324</td>
<td>1.2189</td>
<td>1.3137</td>
</tr>
<tr>
<td>WOE</td>
<td>2</td>
<td>0.4089</td>
<td>0.1703</td>
<td>0.4325</td>
</tr>
<tr>
<td>Predictor2</td>
<td>1</td>
<td>0.0808</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WOE</td>
<td>2</td>
<td>-0.1022</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Predictor3</td>
<td>1</td>
<td>-0.1067</td>
<td>-0.1014</td>
<td>-0.1119</td>
</tr>
<tr>
<td>WOE</td>
<td>2</td>
<td>-0.2697</td>
<td>-0.2010</td>
<td>-0.2636</td>
</tr>
<tr>
<td>Predictor6</td>
<td>1</td>
<td>-0.0692</td>
<td></td>
<td>-0.0707</td>
</tr>
<tr>
<td>WOE</td>
<td>2</td>
<td>-0.1865</td>
<td></td>
<td>-0.1827</td>
</tr>
<tr>
<td>Predictor7</td>
<td>1</td>
<td>-0.0331</td>
<td></td>
<td>-0.0335</td>
</tr>
<tr>
<td>WOE</td>
<td>2</td>
<td>-0.2097</td>
<td></td>
<td>-0.2096</td>
</tr>
</tbody>
</table>
Computing VUS in SAS Software

In this section some ways for estimating VUS are presented. These methods were applied on models from Section 4 using SAS. Let $S_1$, $S_2$, and $S_3$ denote the scores resulting from logistic model for observations from categories one, two and three respectively. They indicate the probability that observation is assigned to appropriate category, therefore if there was an ideal model, $S_1$, $S_2$, and $S_3$ would always satisfy the condition $S_1 < S_2 < S_3$.

**Parametric approach**

According to Le and Lili (2013) the simple parametric approach is to assume that $S_i \sim N(\mu_i, \sigma_i), \ i = 1, 2, 3$. Then VUS under normality assumption can be expressed as

$$VUS = \int \phi(as - b)\phi(cs + d)\phi(s)ds,$$  
(8)

where

$$a = \frac{\sigma_3}{\sigma_1}, \ b = \frac{\mu_1 - \mu_2}{\sigma_1}, \ c = \frac{\sigma_1}{\sigma_3}, \ d = \frac{\mu_3 - \mu_2}{\sigma_3},$$

is standard normal distribution function and $\phi(.)$ is standard normal density function. The maximum likelihood estimate of VUS can be obtained by substituting sample $\hat{\mu}_i$ and sample standard deviations $\hat{\sigma}_i$ into (8). For some data this approach could achieve a very precise estimation of VUS. However, as can be seen in Fig. 3, in our data normal assumptions are not satisfied so this approach cannot be used.

Similar plots could be shown for $S_1$ and $S_3$. Nevertheless, due to the fact that VUS is invariant under monotonic transformation, Box–Cox type transformation can be applied to the data and then the normality-based method for estimation of the VUS can be used on the transformed data (see Le and Lili, 2013). Since we obtained very poor results by this approach, we focused on nonparametric approaches.

**Nonparametric approaches**

**Mann–Whitney U statistic**

In binary case AUC can be computed as Wilcoxon–Mann–Whitney statistic (see Cortes and Mohri, 2003) given by

$$\hat{AUC} = U = \frac{1}{n_+n_-} \sum_{i=1}^{n_+} \sum_{j=1}^{n_-} I(a_i < b_j),$$

where $a_i$ are results assigned by logistic regression to positive observations and $b_j$ are results assigned to negative ones, $n_+, n_-$ are respectively sample sizes for positive and negative observations and $I$ is indicator function. This is equal to probability that two randomly chosen observations, one from each class, are assigned with correct class.

The first approach of estimating VUS is based on the fact that, like in binary case, VUS can be estimated by Mann–Whitney U statistic (see Le and Lili, 2013). Let the sample sizes for categories one, two and three be $n_1$, $n_2$, and $n_3$, respectively. Then the estimation of the VUS is given by

$$\hat{VUS} = U = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} I(S_{ij} < S_{jk} < S_{il}).$$  
(9)
Approach based on the confusion matrix

The second approach is based on the confusion matrix, which is in case of three categories of the target variable in form presented in the Tab. IV.

<table>
<thead>
<tr>
<th>Predicted category</th>
<th>Current category</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>TP1</td>
</tr>
<tr>
<td>2</td>
<td>F12</td>
</tr>
<tr>
<td>3</td>
<td>F13</td>
</tr>
</tbody>
</table>

IV: Confusion matrix for model with target variable with three categories

Supposedly the client from category one is first to start with, then there is

\[
\frac{TP1}{TP1 + F12 + F13}
\]

chance that he will be assigned a correct class. That leaves the pair from categories two and three. Client from category two is next to be continued with, so there is

\[
\frac{TP2}{TP2 + F21}
\]

chance that he will be assigned a correct class and finally if this is done properly, the client from category three will be assigned to category three with probability equal to 1. This approach leads us to \(VUS\) for this method, which is marked as \(VUS^e\) and equals to

\[
VUS^e = \frac{TP1}{TP1 + F12 + F13} \times \frac{TP2}{TP2 + F21}
\]

(10)

There are six different ways of finding a correct class and each one of them is marked \(VUS^1\) – \(VUS^6\) respectively. Each of them can be chosen with probability \(1/6\). Therefore we propose the estimate \(\hat{VUS}_i\), which is given by formula (11)

\[
\hat{VUS}_i = \frac{1}{6} VUS^1 + \ldots + \frac{1}{6} VUS^6 = \frac{1}{6} (VUS^1 + \ldots + VUS^6).
\]

(11)

Approach based on empirical distribution functions

As mentioned several times before in previous sections, \(VUS\) equals to the probability that randomly selected three observations, each from one class, will be correctly sorted. Therefore \(VUS\) can be written as

\[
VUS = P(S_1 < S_2 < S_3).
\]

It is known that \(S_1, S_2\) and \(S_3\) are mutually independent therefore this approach can be based on the following calculation from Le and Lili (2013):

\[
VUS = P(S_1 < S_2 < S_3) = E_{S_1,S_2,S_3} [I(S_1 < S_2) | S_2 = s] =
\]

\[
E_{S_1,S_2,S_3} [I(S_1 < S_2) \cap I(S_1 > S_2) | S_2 = s] =
\]

\[
E_{S_1}[P(S_1 < s) | P(S_1 > s)] = \int_{-\infty}^{\infty} F_1(s)[1 - F_3(s)] f_3(s) ds.
\]

When estimates of these distribution functions \(\hat{F}_1, \hat{F}_3\) and \(\hat{f}_3\) are used, \(VUS\) can be estimated using formula (12)

\[
\hat{VUS}_2 = \int_{-\infty}^{\infty} \hat{F}_1(s)[1 - \hat{F}_3(s)] \hat{f}_3(s) ds.
\]

(12)

Instead of kernel estimates of distribution functions, which is recommended in Le and Lili (2013), empirical distribution functions were used.

\[
\hat{F}_1(s) = \frac{1}{n_1} \sum_{i=1}^{n_1} I(S_i \leq s \wedge \text{target} = 1),
\]

where \(n_1\) is sample size for category one, \(I\) is indicator function and \(\text{target}\) is target variable. This approach still leads to unbiased estimate of the \(VUS\), but computational demands are much lower. Functions \(\hat{F}_1\) and \(\hat{F}_3\) can be estimated similarly. \(\hat{f}_3\) was estimated using the estimate of the function \(\hat{F}_3\).

RESULTS

Values of the \(VUS\) estimated with techniques mentioned in previous sections can be found in Tab. V. \(\hat{VUS}_1\) is the value obtained by approach

V: Estimated \(VUS\) using SAS Software for created models

<table>
<thead>
<tr>
<th></th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimate</td>
<td>Train</td>
<td>Validation</td>
<td>Train</td>
</tr>
<tr>
<td>U</td>
<td>0.2170</td>
<td>0.1803</td>
<td>0.2618</td>
</tr>
<tr>
<td>(VUS^e)</td>
<td>0.2936</td>
<td>0.2678</td>
<td>0.2852</td>
</tr>
<tr>
<td>(\hat{VUS}_1)</td>
<td>0.2290</td>
<td>0.2240</td>
<td>0.2030</td>
</tr>
</tbody>
</table>
with formula (11), $VUS_2$ are obtained by approach with formulas (12) and $U$ is obtained by direct computation of Mann–Whitney U statistic given by (9). There can be seen that according to $VUS_1$ the model was slightly stronger. It was caused by the simplicity and roughness of the estimate based on confusion matrix. Better results were achieved by Mann–Whitney U statistic, but computation of this statistic, especially for big data, has very high computational demands. Therefore approach based on formula (12) looks like the best choice for estimation of the VUS. According to this method the best model is the model #1, because it has the biggest value of the VUS using both training and validation data.

**SUMMARY**

In this paper it has been showed one of many uses of multinomial logistic regression, estimation of LGD, which nowadays is very important task. Measure of quality of these models can be easily done with SAS software with created macros for estimation of the VUS. People and companies who are using SAS for prediction modelling should consider these methods in case of using multinomial logistic regression. Two approaches were used and both leads to similar results. Our estimates using first approach were around 0.28 and second approach around 0.22. These values have shown that these models are quite better than useless model, but in the future some improvements could be done to get even higher VUS. This paper was mainly concerned with which method use in SAS to compute the VUS, what is good complement to misclassification rate in terms of measuring quality. However, these methods can be used only with three categories of target variable therefore in future the research could be made to extent this article to even more than three dimensions because multinomial logistic regression with more than three categories is also becoming very popular.

**REFERENCES**


MOSSMAN, D., 1999: Three-way ROCs. *Medical Decision Making*, Volume 19, Issue 1, ISSN 0272-999X.


**Address**

Bc. Juraj Kapasný, Mgr. Martin Řezáč, Ph.D., Department of Mathematics and Statistics, Faculty of Science, Masaryk University, Kotlářská 2, 611 37 Brno, Czech Republic, e-mail: 356591@mail.muni.cz, mrezac@math.muni.cz