A New Proposal to Improve Credit Scoring Model Predictive Accuracy

Arianna Agosto and Paolo Giudici

Department of Economics and Management
University of Pavia, Italy

Email: arianna.agosto@unipv.it
paolo.giudici@unipv.it

Emanuela Raffinetti

Department of Economics,
Management and Quantitative Methods
University of Milan, Italy

Email: emanuela.raffinetti@unimi.it

***Abstract*—Machine Learning models and Artificial Intelligence algorithms are required to provide powerful predictions to support the decision process of operators in the FinTech sector, characterised by an extensive use of credit scoring models and digitalised financial services. In such a context, the model predictive accuracy assessment represents a basic requirement. On the one hand, literature provides several predictive accuracy measures but, on the other hand, these measures are typically computationally intensive or are based on subjective criteria. In this paper a solution is provided through a novel predictive accuracy measure, we called Rank Graduation Accuracy (RGA), which is based on the distance between the predicted and observed ranks of the response variable. The RGA presents properties which allow to fulfill the need of ensuring reliable predictions improving the model predictive accuracy assessment in highly complex situations.**

Keywords–Machine Learning models; Artificial Intelligence-based systems; Predictive accuracy; Credit Scoring models.

I. INTRODUCTION

A very key point in the application of Machine Learning (ML) and Artificial Intelligence (AI) methods is the evaluation of their predictive accuracy. The predictive accuracy requirement is basic especially in banking and FinTech sectors where data have to be exploited in order to draw conclusions from them and predict future trends. To do this, more accurate results have to be performed to allow organizations to detect new risks (in terms of default predictive accuracy). This objective is more evident when dealing with AI systems, which have a black-box nature resulting in automated decision making which in turn can classify a user into a class associated with the prediction of the individual behavior, without explaining the underlying rationale. In order to avoid that wrong actions are taken as a consequence of "automatic" choice, predictive accuracy measures have to be as much as possible reliable.

Several researchers have proposed, along the years, statistical measures aimed at evaluating predictive accuracy [4] [7]. Likewise, the increasing availability of computational power has allowed to translate these measures in statistical softwares giving rise to direct comparisons between different types of predictive models on the same data. But model comparison methods are not universal, since depending on the nature of response variable to be predicted. Our proposal is motivated by the several applications of machine learnings models in credit rating, where the response variable usually takes a binary nature. In this case, predictive accuracy can be assessed in terms of false positive and false negative predictions providing, for a given set of cut-off points, the Receiver Operating Characteristic (ROC) curve, whose main summary measure corresponds to the area under it (Area Under the Receiver Operating Characteristic curve - AUROC). If on the one hand, the AUROC is widely employed, on the other hand it suffers from some drawbacks due to subjective choice of the cut-off points. With the aim of overcoming this restriction, [10] proposes to resort to the Somers' D measure [11] in the context of credit rating accuracy measurement. Even if the Somers' D is independent on the subjective choice of cut-off points, Somers' D is highly computational intensive.

Our purpose is to introduce a new predictive accuracy measure which, due to its construction, is based on objective criteria and less computational intensive than its main competitors. The novel predictive accuracy measure appears as a derivation of a recent research contribution in the field of dependence analysis illustrated by [3] and is based jointly on the comparison between the observed and the predicted response variable ranks and on the employment of the actual values of the response variable corresponding to both ranks.

The rest of the paper is organized as follows. In Section II, an overview of the mainly used predictive accuracy measures is introduced. In Section III, our new proposed predictive accuracy measure is presented and discussed. In Section IV, an application to credit scoring data is illustrated. In Section V, concluding remarks, together with details on future works, are provided.

II. BACKGROUND

Credit scoring models typically involve a binary response variable denoting the borrower's default. Given the binary nature of the response variable, the most commonly employed predictive accuracy measure is the AUROC [2] [7].

Let n denote the total number of borrowers, such that $n = n_D + n_{ND}$, where D and ND are the defaulting and non-defaulting borrower sets. Let S_D and S_{ND} be the credit score random variable, for the defaulting and non-defaulting borrowers, respectively.

For a specific cut-off value c, $F_D(c)$ and $F_{ND}(c)$ are the sensitivity (true positive rate) and 1-specificity (false positive rate) of a credit scoring model based on the cut-off value c. Let $F_D(c)$ and $F_{ND}(c)$ be the sensitivity (true positive rate) and 1-specificity (false positive rate) of a credit scoring model based on the cut-off value c, such that $F_D(c) = P(S_D \leq c)$ and $F_{ND}(c) = P(S_{ND} \leq c)$ [1].

For a given set of cut cut-off values $c = \{1, \ldots, C\}$, the ROC curve is characterised by the set of points with coordinates $(F_D(c), F_{ND}(c))$ or, equivalently, by (G_{ND_i}, G_{D_i}),

 ISBN: 978-1-61208-816-7

where $G_{ND_i} = \sum_{i=1}^{n} p_{ND_i}$, $G_{D_i} = \sum_{i=1}^{n} p_{D_i}$, $p_{ND_i} = P(S_{ND_i} = s_i)$, $p_{D_i} = P(S_{N_i} = s_i)$ and $i = 1, \ldots, n$. From this, it follows that the AUROC is computed as

$$AUROC = \frac{1}{2} \sum_{i=1}^{n} (G_{D_i} + G_{D_{i-1}})(G_{ND_i} - G_{ND_{i-1}}).$$

The AUROC maesure is equal to 0.5 for a random model without any predictive accuracy and is equal to 1 for a perfect model. In the intermediate situations, AUROC takes values in the range $(0.5, 1)$.

An alternative measure of predictive performance is the Somers' D measure [11]. Let Y be a response variable and X be a predictor variable, and let us denote with n the total number of borrowers. Let the variable Y values be arranged in a non-decreasing sense, i.e., $Y_i \leq Y_j$ for $i < j$. Thus, we can define the quantity c_{ij} as follows

$$c_{ij} = \begin{cases} +1, & \text{if } X_i < X_j, Y_i < Y_j \\ -1, & \text{if } X_i > X_j, Y_i < Y_j \\ 0, & \text{otherwise.} \end{cases} \quad (1)$$

The Somers' D measure, pointed out with D_{XY}, is formalized as follows:

$$D_{XY} = \frac{1}{n_u} \sum_{i=1}^{n} \sum_{j>i} c_{ij}, \text{ with } n_u = \sum_{i=1}^{n} \sum_{j>i} 1_{[Y_i \neq Y_j]}. \quad (2)$$

Some specifications are needed:

- D_{XY} takes values in the close range $[-1, +1]$ and does not depend on the chosen cut-off points;
- D_{XY} is computational intensive since, given N observations to be predicted, it computes $\binom{N}{2}$ combinations;
- in the case of a multivariate model, depending on more than one explanatory variable, the Somer's D can be extended by replacing the values of X with the predictions derived from the model, which can be function of all the explanatory variables. Let us denote this extension with $D_{\hat{Y},Y}$;
- since the focus is not on the direction of concordance, the absolute value of $D_{\hat{Y},Y}$ has to be considered.

III. METHODOLOGY

Let $\boldsymbol{y}$ be a vector of the observed values to be predicted and let $\hat{\boldsymbol{y}}$ be the vector of the corresponding predicted values, computed through a specific model $f(\boldsymbol{X})$, where $\boldsymbol{X}$ is the matrix containing the observations on the explanatory variables. Our goal is to compare different models: $\hat{y} = f^1(\boldsymbol{X})$, $\hat{y} = f^2(\boldsymbol{X})$, ..., using a general methodology based on the concordance curve.

A. The concordance curve

Let Y be a target variable and let $X_1, X_2, \ldots, X_p$ be a set of p explanatory variables. The Y values, re-ordered in non-decreasing sense, can be used to build the Y Lorenz curve, denoted with L_Y. More formally, the curve is characterised by the following pairs: $(i/n, \sum_{j=1}^{i} y_{r_j})$, for $i = 1, \ldots, n$, where r_i indicates the (non-decreasing) ranks of Y.

The same Y values can also be used to build the Y dual Lorenz curve, denoted with $L_Y^{'}$, obtained by re-ordering the Y variable values in a non-increasing sense. More formally, the curve is characterised by the following pairs: $(i/n, \sum_{j=1}^{i} y_{d_j})$, for $i = 1, \ldots, n$, where d_i indicates the (non-increasing) ranks of Y.

Likewise, the predicted $\hat{Y}$ values can also be re-ordered, in a non-decreasing sense. Let $\hat{r}_i$, for $i = 1, \ldots, n$, indicate the (non-decreasing) ranks of $\hat{Y}$. A third curve, named concordance curve and denoted with C_Y, can be provided by ordering the Y values with respect to the ranks of the predicted $\hat{Y}$ values, $\hat{r}_i$. Formally, the concordance curve is characterised by the pairs: $(i/n, \sum_{j=1}^{i} y_{\hat{r}_j})$, for $i = 1, \ldots, n$, where $\hat{r}_i$ indicates the (non-decreasing) ranks of $\hat{Y}$.

To illustrate the previous concept, Figure 1 reports, for a given set of test values Y, and the corresponding predictions $\hat{Y}$: the Lorenz curve, the dual Lorenze curve and the concordance curve, together with the bisector curve $(i/n, i/n)$, for $i = 1, \ldots, n$. To ease the illustration, all values have been normalised using the sum of all Y values: $(n\bar{y})$, where $\bar{y}$ indicates the mean of Y.

From Figure 1, we note that the Lorenz curve and its dual are symmetric around the bisector curve, and that the concordance curve lies between them. Note also that, when $\hat{r}_i = r_i$, for all $i = 1, \ldots, n$, the concordance curve is equal to the Lorenz curve, and a perfect concordance between the Y values and the corresponding predictions arises. On the other hand, when $\hat{r}_i = d_i$, the concordance curve is equal to the dual Lorenz curve and a perfect discordance between the Y values and the corresponding predictions emerges. In general, for any given point, a discrepancy between the Lorenz curve and the concordance curve arises only when the predicted rank is different from the observed one. We finally remark that, when the $\hat{Y}$ values are all equal each other, the concordance C_Y curve perfectly overlaps with the bisector curve. In this case, the model has no predictive capability, as it coincides with a random prediction of the Y values.

B. Our proposal: the RGA predictive accuracy measure

The concordance curve, and its relationship with the Lorenz and the dual Lorenz curve can be exploited to summarise the "distance" between the Y and the $\hat{Y}$ values, in terms of the "discrepancy" between their corresponding ranks. In this way, we fully address the ordinal requirement for credit scores. A novel predictive accuracy measure, we call Rank Graduation Accuracy (RGA), is introduced starting from a function C of C_Y and L_Y defined as:

$$C = \frac{\sum_{i=1}^{n} \left\{ i/n - (1/(n\bar{y})) \sum_{j=1}^{i} y_{\hat{r}_j} \right\}}{\sum_{i=1}^{n} \left\{ i/n - (1/(n\bar{y})) \sum_{j=1}^{i} y_{r_j} \right\}}, \quad (3)$$

where y_{r_j} are the Y variable values ordered according to the ranks r_j; $y_{\hat{r}_j}$ are the same values but ordered according to the ranks $\hat{r}_j$.

From (3), we note that the C index is a function of the y-axis values of the points lying on the concordance curve C_Y and of the y-axis values of the points lying on the Lorenz curve L_Y. Indeed the numerator of the index in (3) compares the distance between the set of points lying on the bisector curve and the set of points lying on the concordance curve

 ISBN: 978-1-61208-816-7

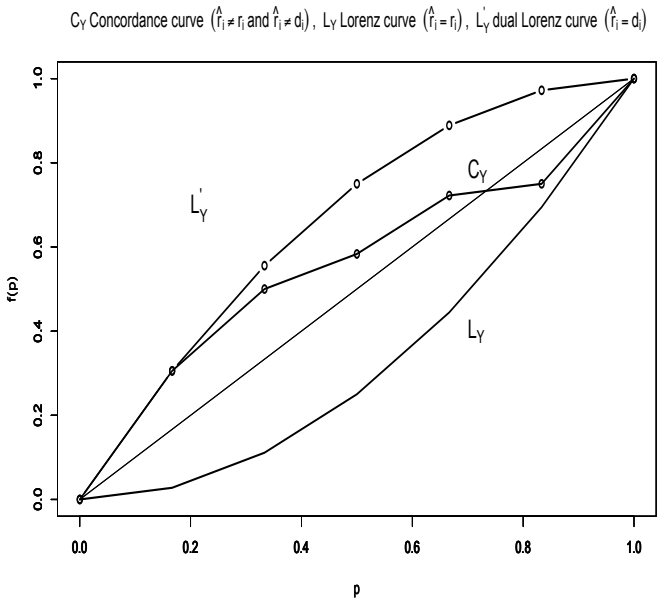

Figure 1. The L_Y and $L_Y^{'}$ Lorenz curves and the C_Y concordance curve.

C_Y, while the denominator compares the distance between the set of points lying on the bisector curve and the set of points lying on the Lorenz curve L_Y.

It can be shown that C fulfills the following properties, whose proofs can be found in [5]:

- $-1 \leq C \leq +1$: specifically, when $0 < C \leq +1$, Y and $\hat{Y}$ are concordant and when $-1 \leq C < 0$ they are discordant;
- $C = +1$ if and only if $C_Y = L_Y$ (full concordance): the concordance curve C_Y overlaps with the Lorenz curve L_Y;
- $C = -1$ if and only if $C_Y = L_Y^{'}$ (full discordance): the concordance curve C_Y overlaps with the dual Lorenz curve $L_Y^{'}$.

Remark Note that, when some of the $\hat{Y}$ values are equal to each other, the original Y values associated with the equal $\hat{Y}$ values can be substituted by their mean, as suggested by [3]. This adjustment is coherent with the definition of a model without predictive capability. To illustrate this point, suppose to consider a general model $f(X)$ with only one explanatory variable, such that $\hat{Y} = E(Y|X) = E(Y) = \bar{y}$ holds for any value of X. Since a re-ordering problem arises if the response variable values are associated with equal estimated values, the response variable values corresponding to the same estimated values are replaced by their mean. As a result, the resulting concordance curve C_Y overlaps with the bisector curve, whose co-ordinates are given by the set of pairs $(i/n, i/n)$. This can be easily shown considering the normalised set of pairs $(i/n, \sum_{j=1}^{i} y_{\hat{r}_j}/n\bar{y})$, characterising the concordance curve C_Y. In the case in which $\hat{y}_i = \bar{y}$, $\forall i = 1, \ldots, n$, we obtain $(i/n, \sum_{j=1}^{i} y_{\hat{r}_j}/n\bar{y}) = (i/n, \sum_{j=1}^{i} \bar{y}/n\bar{y}) = (i/n, i\bar{y}/n\bar{y}) = (i/n, i/n)$.

Looking more closely at (3) note that, when different models are compared, the denominator does not change, while the numerator does. It is therefore intuitive to compare models in terms of differences between the distances expressed by the numerator of formula (3), leading to the following:

$$C_{num} = \sum_{i=1}^{n} \left\{ i/n - (1/(n\bar{y})) \sum_{j=1}^{i} y_{\hat{r}_j} \right\}. \tag{4}$$

The above measure suffers from a drawback: positive values of the index may be compensated by negative values, leading C_{num} to take a value equal to zero. To overcome this problem, we resort to the squared distance between the set of points lying on the concordance curve C_Y and the set of points lying on the bisector curve. Indeed, as the bisector curve defines the situation of a random, non predictive model, for which the Y values are independent on the $\hat{Y}$, we can interpret the squared distance as the difference between the observed and the expected concordance values of Y, where by expected we mean the concordance values that we would have with a random model. If we divide the difference by the expected values themselves, we obtain the RGA (Rank Graduation Accuracy) measure as:

$$RGA = \sum_{i=1}^{n} \frac{\left\{ (1/(n\bar{y})) \sum_{j=1}^{i} y_{\hat{r}_j} - i/n \right\}^2}{i/n}. \tag{5}$$

Through some manipulations, an equivalent version of (5) can be further derived as

$$RGA = \sum_{i=1}^{n} \frac{\left\{ C(y_{\hat{r}_i}) - i/n \right\}^2}{i/n}, \tag{6}$$

which emphasises the role of the quantity $C(y_{\hat{r}_j}) = \frac{\sum_{j=1}^{i} y_{\hat{r}_j}}{\sum_{i=1}^{n} y_{r_i}}$, that represents the cumulative values of the (normalised) response variable.

Note that the RGA index takes values between 0 and its maximum value RG_{max}, which is obtained when the predicted ranks order the response variable values in full concordance (full discordance) with the observed ranks. It can be used to normalise the values of the RGA index, obtaining a measure that is bounded between 0 and 1. It is worth remarking that all models with the same predicted ranks provide the same value of the RGA index. This issue has not to be intended as a limitation of our proposal being the goal of the measure to assess the model attitude in providing a re-ordering of the observed values, which is as much as possible similar to the original ordering.

C. The RGA for scoring models

When assessing the predictive accuracy of credit scoring models in terms of our diagnostic measure, the response variable Y values can be re-ordered according to the predicted values $P(y_i = 1)$, which indeed take real values. Thus, the computation of the RGA index involves only the values 0 or 1, according to the absence or the presence of the attribute of interest, which in this case is the non-default or default occurance.

The possible behaviors of the concordance curve in the binary case is illustrated in Figure 2. Figure 2 illustrates the

three alternative scenarios that can arise, if Y and $\hat{Y}$ are: a) perfectly concordant, b) perfectly discordant and c) partially concordant (discordant). Looking more closely at Figure 2 note that the C_Y concordance curve has a behavior which is similar to the ROC curve. However, while the ROC curve is built ordering cut-off points in an arbitrary way, the C_Y concordance overcomes this subjectivity issue, as the ordering is based on the predicted values themselves. This is indeed a further advantage of our proposal presenting as an objective predictive accuracy diagnostics.

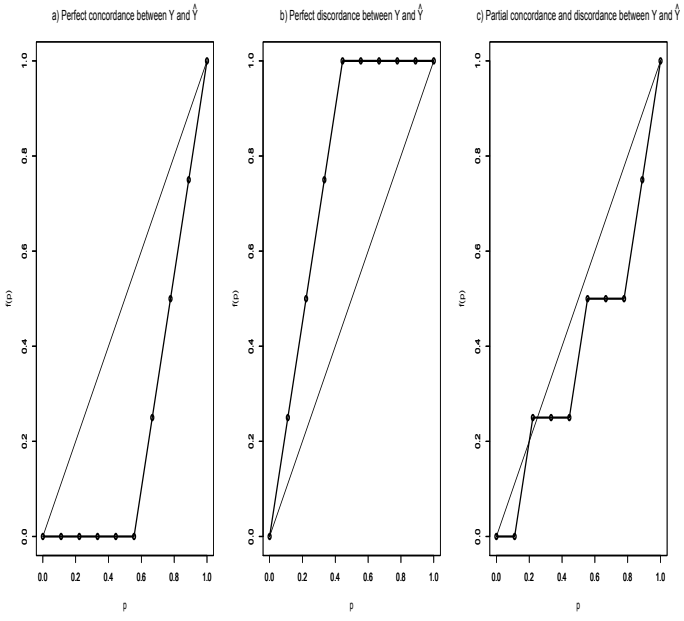

Figure 2. The L_Y and $L_Y^{'}$ Lorenz curves and the C_Y concordance curve.

We finally remark that, in the binary case, the number of points on which the concordance curve is constructed is equal to the number of observations. For each observation, the RGA index compares the values of the actual response, which in the binary case can be either 0 or 1, ordered in one case according to the ranks of the observed response, in the other according to the ranks of the predicted response. We have perfect concordance (Figure 2 a)) when the ranks coincide on all observations; perfect discordance (Figure 2 b)) when the ranks are in reverse correspondence.

IV. Application to credit scoring models

The aim of this section is to show the RGA measure behavior when assessing the predictive accuracy of alternative logistic regression models employed in credit scoring applications. The models are applied to data supplied by a European External Credit Assessment Institution (ECAI), specialized in credit scoring for P2P platforms and focused on SME commercial lending. The dataset includes a set of information on the end-of-year 2015 financial ratios (calculated from balance-sheet variables) related to 15,045 South-European SMEs, for which the specification about the status (0 = active, 1 = defaulted) one year later (2016) is provided. For more details about the data, see [6].

In Table I, the financial ratios employed to predict company's status are reported. Table I shows that, to predict the company's status in 2016, 23 financial ratios from 2015 are available.

TABLE I. LIST OF FINANCIAL RATIOS EMPLOYED AS EXPLANATORY VARIABLES.

ID	Formula or Description
1	Total Assets/Equity
2	(Long term debt + Loans)/Shareholders Funds
3	Total Assets/Total Liabilities
4	Current Assets/Current Liabilities
5	(Current assets - Current assets: stocks)/Current liabilities
6	Shareholders Funds + Non current liabilities)/Fixed assets
7	EBIT/interest paid
8	(Profit or Loss before tax + Interest paid)/Total assets
9	Return on Equity (ROE)
10	Operating revenues/Total assets
11	Sales/Total assets (Activity Ratio)
12	Interest paid/(Profit before taxes + Interest paid)
13	EBITDA/interest paid (Solvency ratio)
14	EBITDA/Operating revenues
15	EBITDA/Sales
16	EBIT Dummy (=1 if EBIT<0, 0 otherwise)
17	Profit before tax Dummy (=1 if Profit before tax<0, 0 otherwise)
18	Financial Profit Dummy (=1 if Financial Profit<0, 0 otherwise)
19	Net Profit Dummy (=1 if Net Profit<0, 0 otherwise)
20	Trade Payables/Operating Revenues
21	Trade Receivables/Operating Revenues
22	Inventories/Operating Revenues
23	Turnover

Following the standard cross-validation approach, the dataset is split into a training and a test subsample, corresponding to 70% and 30% of the sample. A stepwise logistic regression is performed on the training dataset. From Figure 3, which provides the `R` output of the stepwise procedure, it results that that 17 variables over the original 23 variables are selected with $\alpha = 5\%$. For each variable, the corresponding estimated coefficients are also reported. In order to fulfill the model parsimony requirement, variables which are not significant at a level of 1%, are removed leading to select only 9 variables.

By using the estimated coefficient values reported in Figure 3 and derived from the implemented stepwise procedure, the predicted response values $\hat{Y}$ are computed for a set of models that are obtained considering all the subsets of the 9 selected predictors, whose number of predictors is let vary from 1 to 8. For each model, the RGA, Somers' D and AUROC are determined.

A comparison of the measures in terms of model selectivity is also provided by assessing the capability to order models by performance and choose the best among them. We first assess selectivity, for a given model dimension. To this aim, the boxplots in Figure 4 represent the distribution of the three measures for different model dimensions: from 1 predictor to 8 predictors. From the boxplots in Figure 4, it arises that the variability of the RGA measure across the models of the same dimension is always larger than that associated with the other measures, except in the case of only one predictor. This result shows the attitude of the RGA measure to order models and discriminate between them, by resorting to their predictive accuracy. On the contrary, Somers' D works better in the one dimensional case and this is motivated by the usual use of Somers' D as an exploratory tool for variable

selection. From a methodological viewpoint, the better model selection power of the RGA measure, with respect to the AUROC, stems from the different construction. While the AUROC is calculated at a selected set of cut-off points, the RGA is calculated at all response values making it more sensible to model variations. Moreover, if on the one hand increasing the number of cut-off points would improve the AUROC performance making it similar to that associated with the RGA, this "modus operandi" may lead the AUROC-based approach more computationally intensive. Somers' D is also calculated at all response values but, differently from RGA, employs an additional data transformation, based on the binarisation of model errors, which makes it less sensible than the RGA.

As the last step, model selectivity is assessed by comparing different model dimensions. To do this, RGA, Somers' D and AUROC measures are computed on the best model - the one for which the analyzed measure is maximum - for model dimensions that go from 1 to 8. Figure 5 displays the relative change in the maximum value of the three measures, as the number of predictors increases. From Figure 5, it arises that the RGA measure dominates the others in terms of relative change, for all dimensions, allowing us to further show the RGA measure superiority in ordering models and discriminating between them.

```
Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.768e+00  1.715e-01 -10.309  < 2e-16 ***
id_1         3.212e-03  1.521e-03   2.112  0.03465 *
id_3        -5.538e-01  1.144e-01  -4.839 1.30e-06 ***
id_4        -3.351e-01  6.943e-02  -4.826 1.39e-06 ***
id_6         4.775e-03  1.584e-03   3.015  0.00257 **
id_7         3.418e-03  1.645e-03   2.078  0.03773 *
id_8        -2.829e+00  3.588e-01  -7.884 3.18e-15 ***
id_9        -6.001e-02  4.175e-02  -1.438  0.15058
id_10       -2.745e-01  1.615e-01  -1.699  0.08923 .
id_11        3.717e-01  1.602e-01   2.320  0.02034 *
id_13       -3.029e-03  1.347e-03  -2.249  0.02451 *
id_15       -4.749e-01  1.973e-01  -2.407  0.01608 *
id_20        1.849e-03  2.954e-04   6.260 3.85e-10 ***
id_21        7.038e-04  2.681e-04   2.625  0.00867 **
id_23       -2.245e-05  7.658e-06  -2.932  0.00337 **
id_17        4.770e-01  1.631e-01   2.924  0.00345 **
id_19        6.236e-01  1.538e-01   4.055 5.01e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 3. Logistic regression output for the model selected trough the R stepwise procedure.

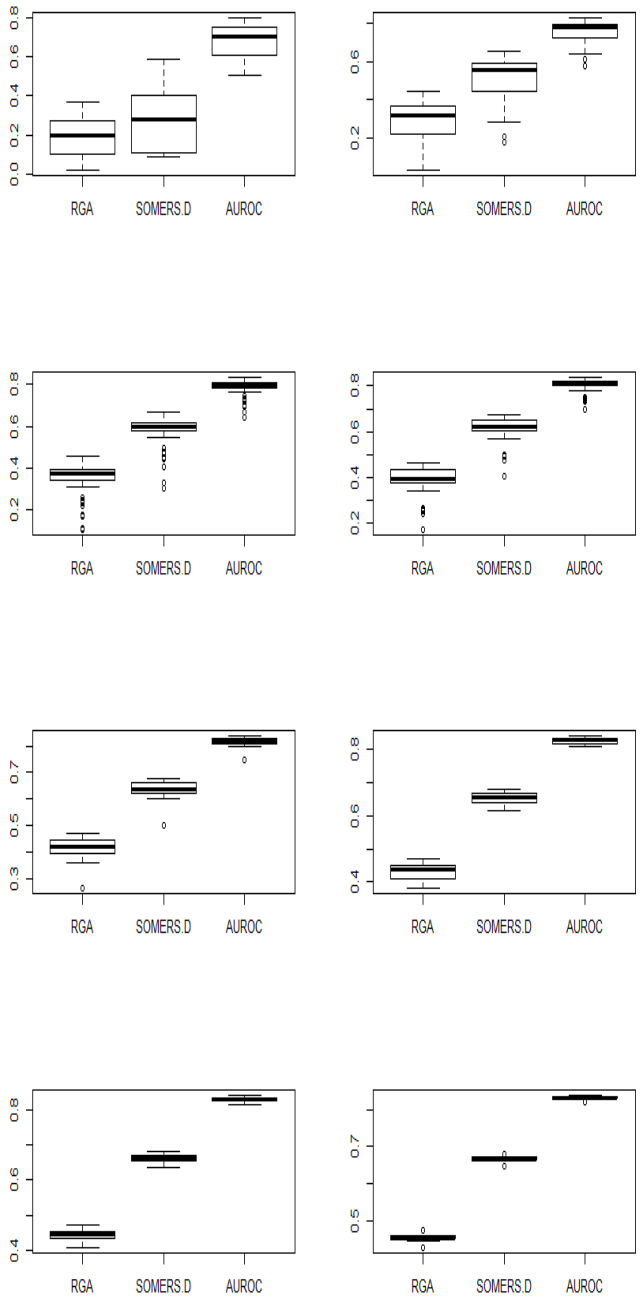

Figure 4. Distribution of RGA, Somers' D and AUROC over the models estimated on credit rating data. The eight plots correspond to different model dimensions; reading from left to right, and from top to bottom: models with 1, 2, 3, 4, 5, 6, 7 and 8 predictors.

V. Conclusions and Future Work

In this paper, a new measure to evaluate the predictive accuracy of a credit scoring model was presented.

The new measure, called RGA, is based on the computation of the cumulative values of the response variable, re-ordered according to the ranks of the values predicted by a given model.

Compared with the other most commonly used predictive accuracy measures, the RGA has the advantage of respecting the ordering requirement for borrowers, and of being independent on the choice of cut-off points, differently from the AUROC, and similarly to Somers' D. Nevertheless, on the contrary of the Somers' D, the RGA is less computational intensive.

 ISBN: 978-1-61208-816-7

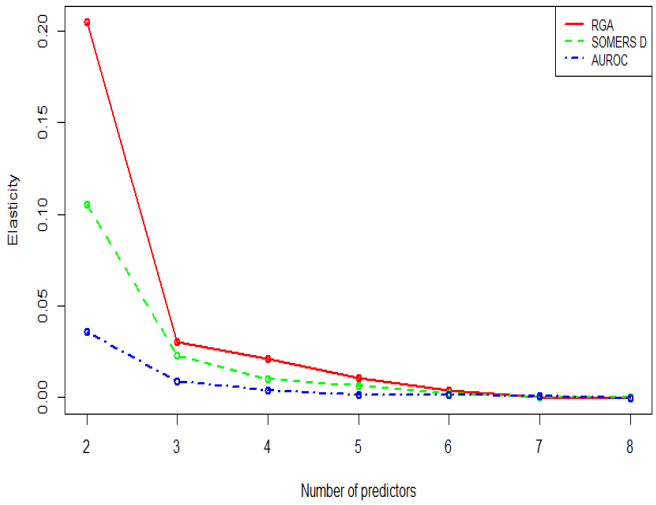

Figure 5. Relative change of RGA, Somers' D and AUROC, as a function of the number of predictors.

The proposed measure appears mathematically sound and easy to implement. Moreover, it has been found quite effective in both a real and a simulated credit scoring application. It overperforms Somers' D and AUROC in model ordering and in discriminating between "good" and "bad" models.

Due to its properties, we believe that the main beneficiaries of the proposed measure may be regulators and supervisors, interested in assessing and validating the credit risk models employed by banks and financial technology companies.

Future extensions of the research will be addressed both to the methodological and application contexts. In the former case, the development of a statistical testing procedure would provide to the predictive accuracy assessment a significance measure. In the latter case, the extensive application to several other application fields, involving the implementation of other machine learning models, would further shed light on the adequacy of our proposal as a suitable criterion for the evaluation of the predictive accuracy in multiple scenarios.

ACKNOWLEDGMENT

The work in the paper has received support from the European Union's Horizon 2020 training and innovation programme "FIN-TECH", under the grant agreement No. 825215 (Topic ICT-35-2018, Type of actions: CSA, https://www.fintech-ho2020.eu).

REFERENCES

[1] B. Engelmann, Measures of a Rating's Discriminative Power-Applications and Limitations. The Basel II Risk Parameters, Springer, 2006, ISBN: 978-3-540-33087-5.

[2] T. Fawcett, "An introduction to ROC analysis," Pattern Recognition Letters, vol. 27, 2006, pp. 861–874, ISSN: 0167-8655.

[3] P. A. Ferrari and E. Raffinetti, "A Different Approach to Dependence Analysis," Multivariate Behavioral Research, vol. 50, 2015, pp. 248–264, ISSN: 1532-7906.

[4] P. Giudici, Applied Data Mining: Statistical Methods for Business and Industry. Wiley, Hoboken, 2003, ISBN: 978-0470871409.

[5] P. Giudici and E. Raffinetti, *Multivariate Ranks-Based Concordance Indexes*, Advanced Statistical Methods for the Analysis 465 of Large Data-Sets, Studies in Theoretical and Applied Statistics, Springer-Verlag Berlin Heidelberg, 2012, ISBN: 978-3-642-21037-2.

[6] P. Giudici, B. Hadji-Misheva, A. Spelta, "Network based credit risk models," Quality Engineering, vol. 32, 2020, pp. 199–211, ISSN: 1532-4222.

[7] D. Hand, H. Mannila and P. Smyth, Principles of data mining. Adaptive Computation and Machine Learning Series. MIT Press, 2001, ISBN: 978-0262082907.

[8] M. O. Lorenz, "Methods of Measuring the Concentration of Wealth," Journal Publications of the American Statistical Association, vol. 9, 1905, pp. 209–219.

[9] D. McFadden, Conditional logit analysis of qualitative choice behavior. In Frontiers in Econometrics, ed. P. Zarembka, New York: Academic Press, 1974, ISBN: 0127761500.

[10] W. Orth, "The predictive accuracy of credit ratings: Measurement and statistical inference," International Journal of Forecasting, vol. 28, 2012, pp. 288–296, ISSN: 0169-2070.

[11] R. H. Somers, "A new asymmetric measure of association for ordinal variables," American Sociological Review, vol. 27, 1962, pp. 799–811, ISSN: 1939-8271.

 ISBN: 978-1-61208-816-7