

# INDETERMINATE VALUES OF TARGET VARIABLE IN DEVELOPMENT OF CREDIT SCORING MODELS

Martin Řezáč, Lukáš Toma

**Received: June 19, 2013**

## Abstract

ŘEZÁČ MARTIN, TOMA LUKÁŠ: *Indeterminate values of target variable in development of credit scoring models.* Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis, 2013, LXI, No. 7, pp. 2709–2716

In the beginning of every modelling procedure, the first question to ask is what we are trying to predict by the model. In credit scoring the most frequent case is modelling of probability of default; however other situations, such as fraud, revolving of the credit or success of collections could be predicted as well. Nevertheless, the first step is always to define the target variable.

The target variable is generally an 'output' of the model. It contains the information on the available data that we want to predict in future data. In credit scoring it is commonly called good/bad definition. In this paper we study the effect of use of indeterminate value of target variable in development of credit scoring models. We explain the basic principles of logistic regression modelling and selection of target variable. Next, the focus is given to introduction of some of the widely used statistics for model assessment. The main part of the paper is devoted to development and assessment of 27 credit scoring models on real credit data, which are built up and assessed according various definitions of target variable. We show that there is a valid reason for some target definitions to include the indeterminate value into the modelling process, as it provided us with convincing results.

credit scoring, indeterminate value, target variable, Gini, KS, lift

In these times of uncertainty, one branch of business application of mathematics has grown on importance – credit scoring. Nowadays, a widely believed starter of the crises was expansive credit politics and lack of risk management in the US and by consequence growing number of bad clients, which were not able to repay their liabilities. It is for these events, that new regulations regarding risk management were developed. Furthermore, the inner processes were being re-evaluated in financial companies, seeking improvement and trying to survive the hard times. This has opened the space for new techniques in credit scoring and innovation of the old ones.

In this paper, we deal with the modelling of the scoring function and estimation of client's creditworthiness using logistic regression model. It is focused on the definition of the target variable, especially on indeterminate values. In the second half of this paper, the most often used measuring statistics for the assessment of credit scoring model

are presented and their values, obtained on real credit data, are compared for different developed models.

The literature devoted to the topic of indeterminate values of target variable in credit scoring is not very extensive. Beardsell (2004) dealt with the phenomenon of indeterminates. According to his results, the use of indeterminate value of target variable for credit scoring development provides no extra value. However, this was proved only by one test. Siddiqi (2006) discussed the definition of the indeterminates and stated that indeterminates are only used where the "bad" definition can be established several ways, and are usually not used where the definition is clear-cut, e.g. bankrupt. Furthermore he suggested that indeterminates should not exceed more than 10% to 15% of the portfolio.

Sarlija *et al.* (2006) tried to find an efficient model for consumer credit scoring using neural networks in comparison with logistic regression. A specific

characteristic of the examined data set was that the credit repayment period was not completed. Applicants were divided into three categories: “good”, “bad”, and “indeterminate” applicants who influenced the model accuracy. Five different modelling strategies were tested: (1) multinomial model with three categories of applicants, (2) binomial model using only good and bad applicants, (3) binomial model including indeterminate applicants as good, (4) binomial model including indeterminate applicants as bad, and (5) binomial model in which indeterminate credit applicants were estimated by model 2 and then included in the dataset. The results suggested that the best strategy to deal with indeterminate applicants is to estimate them as good and bad, and then include into the model or to exclude them from the data set.

Anderson (2007) stated that another potential indeterminate group for credit scoring models is early settlements. But at the same time, he added that this may be contentious, as they are good accounts, but may be justified, because their inclusion can bias credit scoring model in favour of applicants that are often unprofitable. Bolton (2009) presented in his work that some developers of credit scoring models deleted indeterminates from a sample with the hope that eliminating gray credits will produce a model that can better distinguish between good and bad credits. But simultaneously he added that other developers found this practice quite useful.

The re-evaluation of the statement of Beardsell, Sarlija et al. and Bolton will be the primary goal of this paper as it could be possible improvement in management of credit risk. Secondly, the strength of multiple models will be compared and therefore the strongest possible definition of good and bad client could be suggested.

### Modelling

In the beginning of every modelling procedure, the first question to ask is what we are trying to predict by the model. In credit scoring the most frequent case is modelling of probability of default; however other situations, such as fraud, revolving of the credit or success of collections could be predicted as well. Nevertheless, the first step is always to define the target variable.

Target variable is generally an ‘output’ of our model. It contains the information on the available data that we want to predict in future data. In credit scoring it is commonly called good/bad definition. There are four outcome-performance statuses used in credit scoring associated with good/bad definition (see Anderson,2007).

Good	Desired state, something to be welcomed in future.
Bad	Unwanted state, something to be avoided in future.
Indeterminate	In between, greeted with mild reluctance and not revulsion.
Excludes	Any outcome outside of the intended purpose of the scorecard.

Traditionally, the target variables with binary response are used (e.g. default/non-default or fraud/no fraud) and therefore only two categories – good and bad – are used for the model. The other two outcomes are omitted, as it gives better prediction results. Or does it? In this paper we beg to differ and try to prove that indeterminate value has its justification as a part of modelling process.

As our goal was to describe the influence of the indeterminate value of response variable, we were using multiple target definitions so to have more models to compare with and support our observations. We used nine different definitions explained in the following Tab. I, where DPD stands for Days past default and “on first” means on the first payment.

The values of bad rates (on real credit data, described in chapter 4) were between 0.32% and 3.28% for different definitions, which is quite small number; however, as there was sufficient number of observations, it was not an issue. The shares of indeterminate clients were between 0.5% and 5%. This did not exceed 15%, which is the recommended maximum share in the portfolio (Siddiqi, 2006). On the other hand, this was enough, so that operations with them will influence the discriminating power of the model.

We have used logistic regression model as it is the most widely accepted technique for credit scoring. The model is defined (see Hosmer and Lemeshow, 2000) by

I: Good/bad definitions

model	Good	Indeterminate	Bad
3090	never 30 DPD	ever 30-89 DPD	ever 90 DPD
3060	never 30 DPD	ever 30-59 DPD	ever 60 DPD
6090	never 60 DPD	ever 60-89 DPD	ever 90 DPD
3090onfirst	never 30 DPD	30 DPD not on first	90 DPD on first
3060onfirst	never 30 DPD	30 DPD not on first	60 DPD on first
6090onfirst	never 60 DPD	60 DPD not on first	90 DPD on first
30onfirst90onfirst	not 30 DPD on first	30-89 DPD on first	90 DPD on first
30onfirst60onfirst	not 30 DPD on first	30-59 DPD on first	60 DPD on first
60onfirst90onfirst	not 60 DPD on first	60-89 DPD on first	90 DPD on first

$$\text{logit}(p) = \log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k, \quad (1)$$

where  $p$  is the modelled probability of default,  $\beta_0, \dots, \beta_k$  are coefficients of the model and  $x_1, \dots, x_k$  are input variables of the model. More precisely, we use SAS procedure Logistic with stepwise selection method with the Wald Chi-square test with 0.95 level of significance (we tried out also backward and forward methods, but the best results, considering Gini coefficient and Lift, were reached with the stepwise method).

We made three models for each definition on our sample. First model was a model, in which the indeterminate clients were erased from the sample and the modelling was done on the remaining clients. Such a model uses information about the indeterminates, but the power of the model is assessed on data sample, which is not representative – does not correspond with the overall population (indeterminate clients were erased). In the second model, the indeterminate clients were marked as good and again model was developed on the binary target variable. This is a classical modelling approach. Potential drawback of this approach lies in losing information about the indeterminates – they are put together with good in the same bag. In the third model, regression coefficients from the first model were used and data with the Indeterminates marked as good were scored. This model uses information about the indeterminates (as well as the first model), but the power of the model is assessed (unlike the first model) correctly on the whole population – indeterminate clients were put back to the data sample before assessing the power of the model.

Naming convention model $x$ (onfirst) $y$ (onfirst) was used, where  $x$  are number of days past due reached never or not on the first payment respectively for client to be good,  $y$  are number of DPD reached either ever or on the first payment for client to be bad.

For the three models, following suffixes were used:

_indet_erased	model developed and assessed without indeterminates,
_indet_dev	model developed and assessed with indeterminates (classical modelling approach),
_indet_assessed	model developed without indeterminates, but assessed with indeterminates.

### Assessment of quality

After the development of multiple scoring functions, the next important step is to assess them and determine the strongest one in some sense. In credit scoring, many statistics of quality are used for this assessment, working each on different principle. When put all together they give us a complex picture of the power of the model and

unable us to make informed decision, regarding what scoring function to use or whether there are things that should be improved. The most widely used statistics are Kolmogorov-Smirnov statistic (KS), Gini coefficient and Lift. For further available statistics and appropriate remarks see Wilkie (2004), Giudici (2003), Siddiqi (2006) or Řezáč and Kolářček (2012).

Assume that score  $s$  is available for each client and put the following markings:

$$D_K = \begin{cases} 1, & \text{client is good} \\ 0, & \text{otherwise.} \end{cases} \quad (2)$$

The empirical cumulative distribution functions (CDFs) of the scores of good (bad) clients are given by the relationships

$$F_{n,GOOD}(a) = \frac{1}{n} \sum_{i=1}^n I(s_i \leq a \wedge D_K = 1) \quad (3)$$

$$F_{m,BAD}(a) = \frac{1}{m} \sum_{i=1}^m I(s_i \leq a \wedge D_K = 0) \quad a \in [L, H],$$

where  $s_i$  is the score of the  $i^{\text{th}}$  client,  $n$  is the number of good clients,  $m$  is the number of bad clients, and  $I$  is the indicator function, where  $I(\text{true}) = 1$  and  $I(\text{false}) = 0$ .  $L$  is the minimum value of a given score,  $H$  is the maximum value. We denote the proportion of bad clients by

$$p_B = \frac{m}{n+m}$$

and the proportion of good clients by

$$p_G = \frac{n}{n+m}.$$

The empirical distribution function of the scores of all clients is given by

$$F_{N,ALL}(a) = \frac{1}{N} \sum_{i=1}^N I(s_i \leq a) \quad a \in [L, H], \quad (4)$$

where  $N = n + m$  is the number of all clients.

An often-used characteristic in describing the quality of the model (scoring function) is the Kolmogorov-Smirnov statistic (KS). It is defined as

$$KS = \max_{a \in [L, H]} |F_{m,BAD}(a) - F_{n,GOOD}(a)|. \quad (5)$$

However, the main disadvantage of this statistic is, that it only measures the biggest vertical gap, meaning that its realisation is in very specific value. Therefore, it is not a reliable reference tool for the entire population, as all we can say according to this statistic about the entire distribution functions

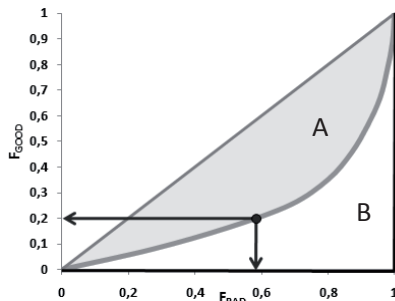
is their maximal distance and not the overall behaviour.

The Lorenz curve (LC), sometimes confused with the ROC curve (Receiver Operating Characteristic curve), can also be successfully used to show the discriminatory power of the scoring function, i.e., the ability to identify good and bad clients. The curve is given parametrically by

$$\begin{aligned} x &= F_{m,BAD}(a) \\ y &= F_{n,GOOD}(a), a \in [L, H]. \end{aligned} \quad (6)$$

The definition and name (LC) is consistent with Müller and Rönz (2000). One can find the same definition of the curve, but called the ROC, in Thomas *et al.* (2002). Siddiqi (2006) used the name ROC for a curve with reversed axes and LC for a curve with the CDF of bad clients on the vertical axis and the CDF of all clients on the horizontal axis. For a short summary of currently used credit scoring methods and the quality testing thereof by using the ROC on real data with interpretations, see Kočenda and Vojtek (2009).

Each point on the curve represents some value of a given score. If we assume this value to be the cut-off value, we can read the proportion of rejected bad and good clients. An example of a Lorenz curve is given in Fig. 1. We can see that by rejecting 20% of good clients, we reject almost 60% of bad clients at the same time.



1: Lorenz Curve, Gini coefficient

In connection with the LC, we will now consider the next quality measure, the Gini coefficient. This index describes the global quality of a scoring function. It takes values between -1 and 1. The ideal model, i.e. a scoring function that perfectly separates good and bad clients, has a Gini coefficient equal to 1. On the other hand, a model that assigns a random score to the client has a Gini coefficient equal to 0. Negative values correspond to a model with reversed meanings of scores. Using Fig. 1 the Gini coefficient can be defined as

$$Gini = \frac{A}{A+B} = 2A. \quad (7)$$

The actual calculation of the Gini coefficient can be made using

$$Gini = 1 - \sum_{k=2}^{n+m} [(F_{m,BAD_k} - F_{m,BAD_{k-1}}) \times (F_{n,GOOD_k} + F_{n,GOOD_{k-1}})], \quad (8)$$

where  $F_{m,BAD_k}$  ( $F_{n,GOOD_k}$ ) is the  $k^{\text{th}}$  vector value of the empirical distribution function of bad (good) clients. For further details see Thomas *et al.* (2002), Siddiqi (2006) or Xu (2003).

The third considered indicator of the quality of a scoring model was (*cumulative*) *Lift*, which states how many times, at a given level of rejection, the scoring model is better than random selection (the random model). More precisely, it indicates the ratio of the proportion of bad clients with a score of less than  $a$ ,  $a \in [L, H]$ , to the proportion of bad clients in the overall population. In practice, the calculation is done for Lift corresponding to 10%, 20%, ..., 100% of clients with the worst score (see Coppock, 2002). It was shown in Řezáč, M. and Řezáč, F. (2011) that the Lift can be expressed by CDFs of scores bad and all clients as

$$Lift(a) = \frac{F_{n,BAD}(a)}{F_{N,ALL}(a)}, a \in [L, H]. \quad (9)$$

In connection with Coppock's approach, we define

$$Lift_q = \frac{F_{n,BAD}(F_{N,ALL}^{-1}(q))}{F_{N,ALL}(F_{N,ALL}^{-1}(q))} = \frac{1}{q} F_{n,BAD}(F_{N,ALL}^{-1}(q)), \quad (10)$$

where  $q$  represents the score level of 100% of the worst scores (i.e. 100% cutoff level) and  $F_{N,ALL}^{-1}(q)$  can be computed as  $F_{N,ALL}^{-1}(q) = \min\{a \in [L, H], F_{N,ALL}(a) \geq q\}$ .

## RESULTS

We obtained a data from financial company<sup>1</sup> operating on the Czech market. Data originally contained 47 socio-demographic characteristics for each observation. First group of variables described client's personal characteristics, such as age, ZIP code, education, employment, number of children. Second group were the behavioural characteristics, describing client's relationship with the company, such as credit amount, time of ratification, defaults in the past, annuity date, annuity amount and others. The sample window consisted of clients who were registered between 1<sup>st</sup> January 2009 and 1<sup>st</sup> April 2011 as clients, which were included in

<sup>1</sup> The company wishes to be undisclosed according to the Privacy contract.

II: KS statistics for scoring models (bold is the stronger one from 2<sup>nd</sup> and 3<sup>rd</sup> column)

Target definition	Indet_erased	Indet_dev	Indet_assessed
model3090	0.5868	<b>0.5771</b>	0.5628
model3060	0.5515	0.5432	<b>0.5438</b>
model6090	0.5805	0.5771	<b>0.5772</b>
model3090onfirst	0.6011	0.5832	<b>0.5848</b>
model3060onfirst	0.6045	0.5868	<b>0.5872</b>
model6090onfirst	0.5941	<b>0.5832</b>	0.5830
model30onfirst90onfirst	0.6246	0.6234	<b>0.6237</b>
model30onfirst60onfirst	0.5884	0.5868	<b>0.5871</b>
model60onfirst90onfirst	0.6084	<b>0.6085</b>	0.6082

III: Gini coefficients (bold is the strongest among all columns)

Target definition	Indet_erased	Indet_dev	Indet_assessed
model3090	<b>0.746</b>	0.734	0.720
model3060	<b>0.714</b>	0.704	0.704
model6090	<b>0.738</b>	0.734	0.734
model3090onfirst	0.734	0.716	<b>0.738</b>
model3060onfirst	<b>0.760</b>	0.744	0.742
model6090onfirst	0.728	0.716	<b>0.738</b>
model30onfirst90onfirst	0.756	0.756	<b>0.774</b>
model30onfirst60onfirst	0.744	0.744	<b>0.774</b>
model60onfirst90onfirst	0.716	0.716	<b>0.732</b>

the database later have not even had a possibility to reach the 90 days default in the set performance window and therefore would bias the sample.

As we were working with real raw data, we had to deal with considerable amount of missing and extreme values. In some instances we dealt with data entries, which were logically inconsistent with the variable's requirements, such as negative ZIP code, age value under 18 years or text entry in numerical variables. All such cases had to be erased or solved by some imputation method. Overall, we have selected 346,322 observations over the mentioned sample window.

In the practical applications in this chapter, obtained statistics are in the tables sorted in rows by different definitions of target variable and in columns there are always the three models for each target variable as explained in the previous chapter.

Kolmogorov-Smirnov statistic was calculated for all 27 models and their values compared among three models for each good/bad definition as well as among different target definitions.

Data in the Tab. II show the KS statistic for all three models made for each good/bad definition (written in the left column). The most obvious trend is that the first model, where indeterminate clients are taken out off the sample, is almost always the strongest, except of the model60onfirst90onfirst. This can be easily explained, as with indeterminate clients, we eliminate data, which are most difficult to evaluate. However, more interesting is the fact that for six out of nine target definitions was the

third model, where indeterminate clients were used (in the sense described above) for scoring function development, stronger than model developed without these clients. This trend is shown in bold in the Tab. II.

The quality of the models was also assessed among different target definitions. From this point of view, the most favourable seems to be model30onfirst90onfirst, which has by far the highest values of KS statistic. On the other hand, the weakest target definition according to KS statistic is model3060. As concerned with in previous part, the KS statistic is measured only as a maximum distance, therefore provided results are not values of overall quality of the model.

Gini coefficient was also used as a primary evaluator of the model during development and deciding, which type of variable selection to use or how to transform data for better scoring performance.

The results of this statistic can be split into two groups by the target definitions (in the Tab. III the split is illustrated by dashed horizontal line). First group contains the target variables, which are defined using only 'never DPD' and 'ever DPD' limitations. For these, the use of indeterminate clients for model development proved clearly unhelpful and it is ought to be used only with respect to the other possibilities.

For the second group, however, containing the models, which target variables were defined using also/only the default on the first payment, the

IV: Lift for 20% cutoff level (bold is the stronger one from 2<sup>nd</sup> and 3<sup>rd</sup> column)

Target definition	Indet_erased	Indet_dev	Indet_assessed
model3090	3.8307	<b>3.7770</b>	3.6797
model3060	3.6261	3.5838	<b>3.5892</b>
model6090	3.7921	3.7770	<b>3.7781</b>
model3090onfirst	3.9599	3.8647	<b>3.8682</b>
model3060onfirst	3.9828	<b>3.9036</b>	3.8984
model6090onfirst	3.9208	3.8647	<b>3.8664</b>
model30onfirst90onfirst	4.0759	<b>4.0760</b>	4.0758
model30onfirst60onfirst	3.9101	3.9036	<b>3.9053</b>
model60onfirst90onfirst	3.9973	3.9972	<b>3.9973</b>

V: Lift for 50% cutoff level (bold is the stronger one from 2<sup>nd</sup> and 3<sup>rd</sup> column)

Target definition	Indet_erased	Indet_dev	Indet_assessed
model3090	1.8895	1.8821	<b>1.8830</b>
model3060	1.8572	1.8519	<b>1.8528</b>
model6090	1.8841	<b>1.8821</b>	1.8819
model3090onfirst	1.9028	<b>1.8939</b>	1.8925
model3060onfirst	1.9139	<b>1.9071</b>	1.9070
model6090onfirst	1.9001	1.8939	<b>1.8953</b>
model30onfirst90onfirst	1.9309	<b>1.9301</b>	1.9292
model30onfirst60onfirst	1.9071	1.9071	1.9071
model60onfirst90onfirst	1.9311	<b>1.9312</b>	1.9310

situation is dramatically different. Not only was the third model, using the indeterminate clients (in the sense described above), stronger than the model without them, it was also in almost all cases stronger than the first model, developed only on good and bad clients with indeterminates taken out of the sample (the strongest model is shown in bold in the table). This can, of course, be valid only on this particular data. However, trend has been confirmed also by testing it on validation sample.

For the determination of discriminating power of the model by Lift statistic, the values of Cumulative Lift statistic were compared for 20% and 50% cutoff levels. The values were once again compared for three models among one target definition as well as for different definitions.

According to the Tab. IV, which shows the Cumulative Lift statistic values for the 20% cutoff level, we can again see the first obvious trend that the first model (without the indeterminate clients) is the strongest. This can be, however, easily explained as it is only logical, that when we take out the clients, which practically fall to the 'not-sure' category, the discriminative power of the model will become considerably higher. Nevertheless, for the purpose of this paper more important fact is, that again, for the 6 out of 9 models (again shown in bold in the table), the third model presents higher values than second model. This means the use (in the sense described above) of indeterminate clients for

development of the model has a positive influence on its power as measured by Cumulative Lift statistic.

Regarding the discriminating power of different target definitions, the Cumulative Lift statistic states, that the strongest is again model30onfirst90onfirst. The most widely used target definition in model3090 presents, however, one of the weakest results according to this statistic.

For the Cumulative Lift statistic at 50% cutoff level, the results presented in Tab. V do not follow the same trends as in the previous cutoff level. The first model still stays as the strongest one among the three models for each target definition. However, at this cutoff level, the third model, using the indeterminate clients for modelling of the scoring function, does not appear to follow as persuasive trend of positive influence on the power of the model (stronger of the two models again in bold). Nevertheless, there are still three target definitions for which their use is justified. Among different target definitions is on this occasion strongest model60onfirst90onfirst.

Overall, the most interesting fact is, that with the rising cutoff level, the influence of the indeterminate clients on the model development is almost perfectly reversed (except for model model3060). This fact and its causes are however left by the author for further exploration.

## SUMMARY

Credit scoring industry, in the light of events from last few years, seeks innovation. New procedures are tried out and old ones need improvement. As the improvement of procedures, which are already set up, is usually less expensive, their improvement can be implemented more easily and requires less additional resources. Therefore, this paper was mainly concerned with how the indeterminate value of target variable in credit scorecard development influences the quality of the model predictions, as there is a possibility to enhance the predictive power of logistic regression models, which are most widely used in credit scoring. This was researched on twenty seven logistic regression models for nine good/bad definitions for sufficient backup for the claims presented. There is a brief introduction to the foundations of logistic regression modelling and its mathematical principles are explained and illustrated. After the model development, quality of the models were assessed by standard quality measuring techniques used in credit scoring, such as Gini coefficient, Kolmogorov-Smirnov statistic or Lift. From the comparison of these data interesting facts appeared. Firstly the results show, that there is a valid reason for some target definitions to include the indeterminate value into the modelling process, as it provided us with convincing results. On the other hand, this was mainly for the 'less usual' definitions of good and bad client, which are not that often (if at all) used in real practice. This can be the issue, as in most real-world firms, target definitions are either set or at least bounded by the mother company or some kind of regulator. Nevertheless, the statistics have also shown, that these definitions give usually models with similar or even slightly higher predictive power, therefore they should be also considered into practice. These are, however, more strict definitions, so there can be a problem with rising level of rejection and also lack of observations for defaulted customers for development process. This was left by authors for further exploration as it was not a part of the hypothesis. In conclusion, the indeterminate clients in the data sample for model development can be a source of valuable information and they should not be omitted entirely in the modelling process. Its usage at least as a support tool is recommended.

## REFERENCES

- ANDERSON, R., 2007: *The credit scoring toolkit: theory and practice for retail credit risk management and decision automation*. Oxford: Oxford University Press. 731, ISBN 978-0-19-922640-5.
- BEARDSSELL, M., 2004: Impact of Indeterminate Performance Exclusions on Credit Score Model Development. In: Mays, E. (ed.), *Credit Scoring for risk managers: The handbook for lenders*. Mason: South-Western Publishing, 147–156. ISBN 978-1450578967.
- BOLTON, CH., 2009: *Logistic regression and its application in credit scoring*, <http://upetd.up.ac.za/thesis/available/etd-08172010-202405/unrestricted/dissertation.pdf>. Accessed on January 12, 2013.
- COPPOCK, D. S., 2002: Why Lift? DM Review Online. <http://www.dmreview.com/news/5329-1.html>. Accessed on December 1, 2009.
- GIUDICI, P., 2003: *Applied Data Mining: statistical methods for business and industry*. Chichester: Wiley. 364, ISBN 978-0-470-84679-7.
- HOSMER, D. W. and Lemeshow, S., 2000: *Applied Logistic Regression*. 2nd ed., New York: John Wiley & Sons Inc. 375, ISBN 978-0-471-35632-5.
- KOČENDA, E. and VOJTEK, M., 2009: *Default Predictors and Credit Scoring Models for Retail Banking*. CESifo Working Paper, No. 2862.
- MÜLLER, M. and RÖNZ, B., 2000: Credit Scoring using Semiparametric Methods. In: FRANKE, J., HÄRDLE, W., STAHL, G. (Eds.): *Measuring Risk in Complex Stochastic Systems*. New York: Springer-Verlag, 85–102. ISBN 978-0-387-98996-9.
- SARLIJA, N., BENSIC, M., ZEKIC-SUSAC, M., 2006: A neural network classification of credit applicants in consumer credit scoring. In: *AIA'06 Proceedings of the 24th IASTED international conference on Artificial intelligence and applications*. Innsbruck: ACTA Press, Anaheim, 205–210. ISBN 0-88986-556-6.
- ŘEZÁČ, M. and KOLÁČEK, J., 2012: Lift-Based Quality Indexes for Credit Scoring Models as an Alternative to Gini and KS. *Journal of Statistics: Advances in Theory and Applications* 7, 1: 1–23. ISSN 0975-1262.
- ŘEZÁČ, M. and ŘEZÁČ, F., 2011: How to Measure the Quality of Credit Scoring Models. *Finance a úvěr – Czech Journal of Economics and Finance*, 61, 5: 486–507. ISSN 0015-1920.
- SIDDIQI, N., 2006: *Credit Risk Scorecards: developing and implementing intelligent credit scoring*, New Jersey: Wiley, 196. ISBN 978-0-471-75451-0.
- THOMAS, L. C., EDELMAN, D. B., CROOK, J. N., 2002: *Credit Scoring and Its Applications*. Philadelphia: SIAM Monographs on Mathematical Modeling and Computation, 248. ISBN 0-89871-483-4.
- WILKIE, A. D., 2004: Measures for comparing scoring systems. In: THOMAS, L. C., EDELMAN, D. B., CROOK, J. N. (Eds.): *Readings in Credit Scoring*. Oxford: Oxford University Press, pp. 51–62. ISBN 978-0-19-852797-8.
- XU, K., 2003: *How has the literature on Gini's index evolved in past 80 years?* <http://economics.dal.ca/RePEc/dal/wparch/howgini.pdf>. Accessed on December 1, 2009.

Address

Mgr. Martin Řezáč, Ph.D., Department of Mathematics and Statistics, Faculty of Science, Masaryk University, Kotlářská 2, 602 00 Brno, Czech Republic, Bc. Lukáš Toma, Polytech Nantes, Université de Nantes 1, quai de Tourville BP 13522, 44035 Nantes Cedex 1, France, e-mail: [mrezac@math.muni.cz](mailto:mrezac@math.muni.cz), [toma.lukas@gmail.com](mailto:toma.lukas@gmail.com)