

LIFT-BASED QUALITY INDEXES FOR CREDIT SCORING MODELS AS AN ALTERNATIVE TO GINI AND KS

MARTIN ŘEZÁČ and JAN KOLÁČEK

Department of Mathematics and Statistics
Masaryk University
Kotlářská 2, 61137 Brno
Czech Republic
e-mail: mrezac@math.muni.cz

Abstract

Assessment of risk associated with the granting of credits is very successfully supported by techniques of credit scoring. To measure the quality, in the sense of the predictive power, of the scoring models, it is possible to use quantitative indexes such as the Gini index (Gini), the K-S statistic (KS), the c -statistic, and lift. They are used for comparing several developed models at the moment of development as well as for monitoring the quality of the model after deployment into real business. The paper deals with the aforementioned quality indexes, their properties and relationships. The main contribution of the paper is the proposal and discussion of indexes and curves based on lift. The curve of ideal lift is defined; lift ratio (LR) is defined as analogous to Gini index. Integrated relative lift (IRL) is defined and discussed. Finally, the presented case study shows a case when LR and IRL are much more appropriate to use than Gini and KS.

2010 Mathematics Subject Classification: 62P05, 90B50.

Keywords and phrases: credit scoring, quality indexes, Gini index, lift, lift ratio, integrated relative lift.

Received February 14, 2012

1. Introduction

Banks and other financial institutions receive thousands of credit applications every day (in the case of consumer credits, it can be tens or hundreds of thousands every day). Since it is impossible to process them manually, automatic systems are widely used by these institutions for evaluating the credit reliability of individuals, who ask for credit. The assessment of the risk associated with the granting of credits has been underpinned by one of the most successful applications of statistics and operations research: credit scoring.

Credit scoring is the set of predictive models and their underlying techniques that aid financial institutions in the granting of credits. These techniques decide who will get credit, how much credit they should get, and what further strategies will enhance the profitability of the borrowers to the lenders. Credit scoring techniques assess the risk in lending to a particular client. They do not identify “good” or “bad” (negative behaviour is expected, e.g., default) applications on an individual basis, but forecast the probability that an applicant with any given score will be “good” or “bad”. These probabilities or scores, along with other business considerations such as expected approval rates, profit, churn, and losses, are then used as a basis for decision making.

Several methods connected to credit scoring have been introduced during last six decades. The most well-known and widely used are logistic regression, classification trees, the linear programming approach, and neural networks.

The methodology of credit scoring models and some measures of their quality have been discussed in surveys including Hand and Henley [7], Thomas [14] or Crook et al. [4]. Even if ten years ago the list of books devoted to the issue of credit scoring was not extensive, the situation has improved in the last decade. In particular, this list now includes Anderson [1], Crook et al. [4], Siddiqi [11], Thomas et al. [15], and Thomas [16].

The aim of this paper is to give an overview of widely used techniques used to assess the quality of credit scoring models, to discuss the properties of these techniques, and to extend some known results. We review widely used quality indexes, their properties and relationships. The main part of the paper is devoted to lift. The curve of ideal lift is defined; lift ratio is defined as analogous to Gini index. Integrated relative lift is defined and discussed.

2. Measuring the Quality

We can consider two basic types of quality indexes: first, indexes based on a cumulative distribution function like the Kolmogorov-Smirnov statistic, Gini index or lift; second, indexes based on a likelihood density function like the mean difference (Mahalanobis distance) or informational statistic. For further available measures and appropriate remarks, see Wilkie [17], Giudici [6] or Siddiqi [11].

Assume that the realization $s \in R$ of a random variable S (score) is available for each client and put the following markings:

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

Distribution functions, respectively, their empirical forms, of the scores of good (bad) clients are given by

$$F_{n.GOOD}(a) = \frac{1}{n} \sum_{i=1}^N I(s_i \leq a \wedge D = 1),$$

$$F_{m.BAD}(a) = \frac{1}{m} \sum_{i=1}^N I(s_i \leq a \wedge D = 0), \quad a \in [L, H], \quad (2)$$

where s_i is the score of i -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. The empirical distribution function of the scores of all clients is given by

$$F_{N.ALL}(a) = \frac{1}{N} \sum_{i=1}^N I(s_i \leq a), \quad a \in [L, H], \quad (3)$$

where $N = n + m$ is the number of all clients. We denote the proportion of bad (good) clients by

$$p_B = \frac{m}{n + m}, \quad p_G = \frac{n}{n + m}. \quad (4)$$

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

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

It takes values from 0 to 1. Value 0 corresponds to a random model, value 1 corresponds to the ideal model. The higher the KS, the better the scoring model.

The Lorenz curve (LC), sometimes called the ROC curve (receiver operating characteristic curve), can also be successfully used to show the discriminatory power of a scoring function, i.e., the ability to identify good and bad clients. The curve is given parametrically by

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

Each point of the curve represents some value of a given score. If we consider this value as a cut-off value, we can read the proportion of rejected bad and good clients. An example of a Lorenz curve is given in Figure 1. We can see that by rejecting 20% of good clients, we also reject 50% of bad clients at the same time.

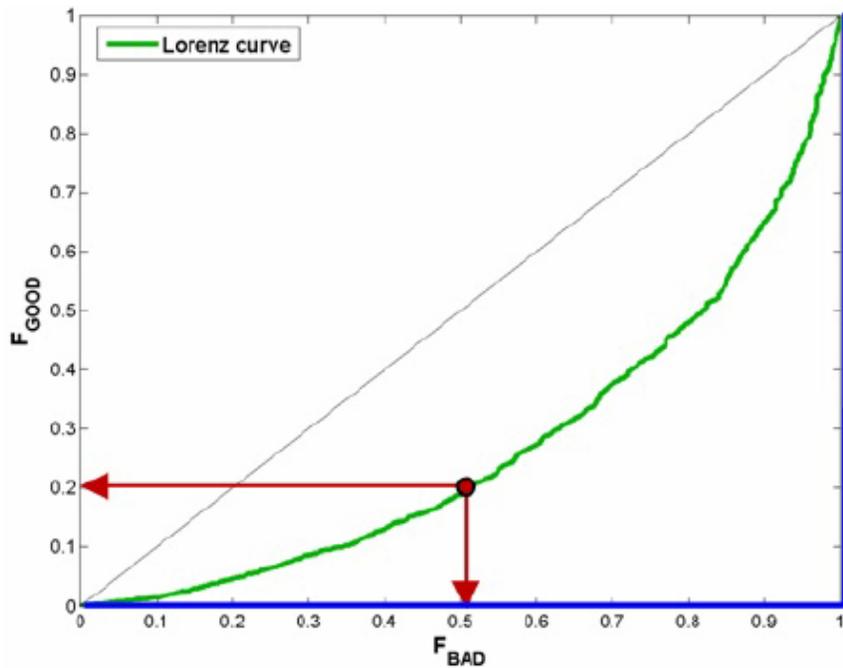


Figure 1. Lorenz curve (ROC).

The LC for a random scoring model is represented by the diagonal line from $[0, 0]$ to $[1, 1]$. It is the polyline from $[0, 0]$ through $[1, 0]$ to $[1, 1]$ in the case of an ideal model. It is obvious that the closer the curve is to the bottom right corner, the better is the model.

The definition and name (LC) is consistent with Müller and Rönz [8]. One can find the same definition of the curve, but called ROC, in Thomas et al. [15]. Siddiqi [11] 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. This curve is also called the CAP (*cumulative accuracy profile*) or lift curve, see Sobehart et al. [12] or Thomas [16]. Furthermore, it is called a *gains chart* in the field of marketing; see Berry and Linoff [2]. An example of CAP is displayed in Figure 2. The ideal model is now represented by a polyline from $[0, 0]$

through $[p_B, 1]$ to $[1, 1]$. The advantage of this figure is that, one can easily read the proportion of rejected bads against the proportion of all rejected. For example, in the case of Figure 2, we can see that if we want to reject 70% of bads, we have to reject about 40% of all applicants.

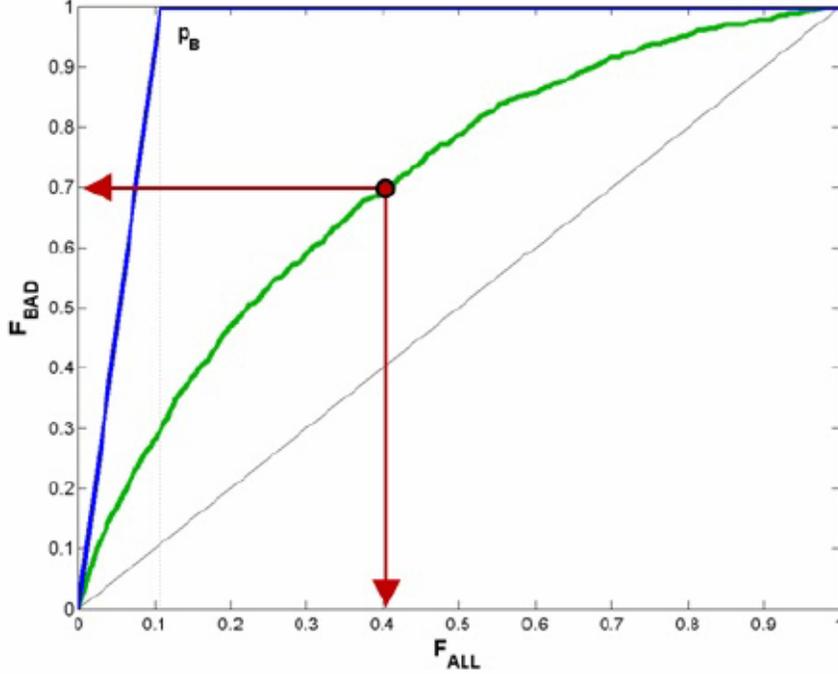


Figure 2. CAP.

In connection to LC, we consider the next quality measure, the Gini index. This index describes a global quality of the scoring model. It takes values from 0 to 1 (it can take negative values for contrariwise models). The ideal model, i.e., the scoring function that perfectly separates good and bad clients, has a Gini index equal to 1. On the other hand, a model that assigns a random score to the client, has a Gini index equal to 0. It can be shown that the Gini index is greater than or equal to KS for any scoring model. Using Figure 3, it can be defined as follows:

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

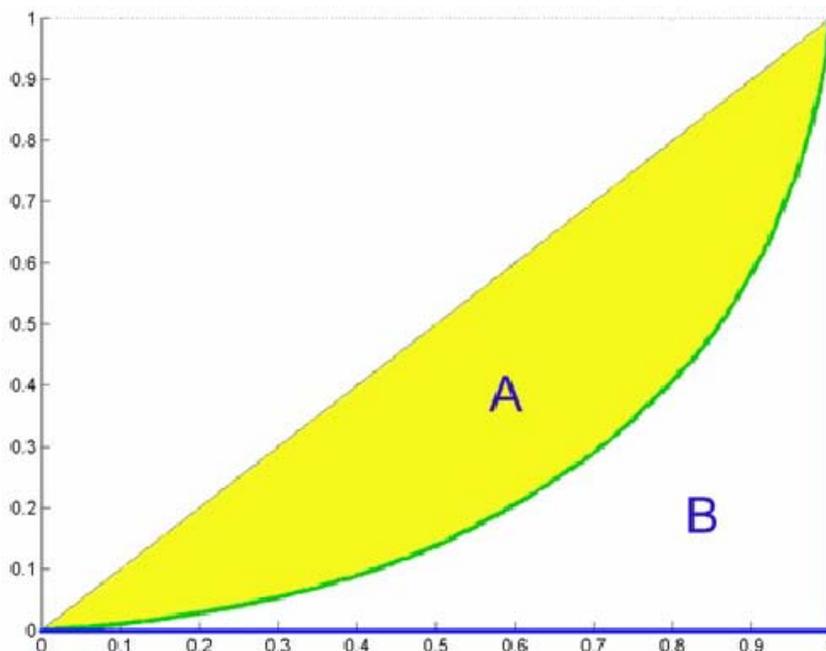


Figure 3. Lorenz curve, Gini index.

This means that, we compute the ratio of the area between the curve and the diagonal (which represents a random model) to the area between the ideal model's curve and the diagonal. Since the axes describe a unit square, the area $A + B$ is always equal to 0.5. Therefore, we can compute the Gini as two times the area A . Using previous markings, the computational formula of the Gini index is given by

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

where $F_{m.BAD_k}(F_{n.GOOD_k})$ is the k -th vector value of the empirical distribution function of bad (good) clients. For further details, see Anderson [1] or Xu [18]. The Gini index is a special case of Somers' D (Somers [13]), which is an ordinal association measure. According to Thomas [16], one can calculate the Somers' D as

$$D_S = \frac{\sum_i g_i \sum_{j < i} b_j - \sum_i g_i \sum_{j > i} b_j}{n \cdot m}, \quad (9)$$

where $g_i(b_j)$ is the number of goods (bads) in the i -th interval of scores. Furthermore, it holds that D_S can be expressed by the Mann-Whitney U -statistic; see Nelsen [9] for further details.

When we use CAP instead of LC, we can define the accuracy rate (AR); see Thomas [16] or Sobehart et al. [12], where it is called the accuracy ratio. Again, it is defined by the ratio of some areas. We have

$$\begin{aligned} AR &= \frac{\text{Area between CAP curve and diagonal}}{\text{Area between ideal model's CAP and diagonal}} \\ &= \frac{\text{Area between CAP curve and diagonal}}{0.5(1 - p_B)}. \end{aligned} \quad (10)$$

Although the ROC and CAP are not equivalent, it is true that Gini and AR are equal for any scoring model. Proof for discrete scores is given in Engelmann et al. [5]; for continuous scores, one can find it in Thomas [16].

In connection to the Gini index, the c -statistic (Siddiqi [11]) is defined as

$$c_stat = \frac{1 + Gini}{2}. \quad (11)$$

It represents the likelihood that a randomly selected good client has a higher score than a randomly selected bad client, i.e.,

$$c_stat = P(s_1 \geq s_2 | D_1 = 1 \wedge D_2 = 0). \quad (12)$$

It takes values from 0.5, for the random model, to 1, for the ideal model. An alternative name for the c -statistic can be found in the literature. It is known also as Harrell's c , which is a reparameterization of Somers' D (Newson [10]). Furthermore, it is called AUROC, e.g., in Thomas [16] or AUC, e.g., in Engelmann et al. [5].

3. Lift

Another possible indicator of the quality of scoring model is lift, which determines the number of times that, at a given level of rejection, the scoring model is better than random selection (the random model). More precisely, the ratio is the proportion of bad clients with a score less than a (where $a \in [L, H]$) to the proportion of bad clients in the general population. Formally, it can be expressed by

$$\begin{aligned}
 Lift(a) &= \frac{CumBadRate(a)}{BadRate} = \frac{\frac{\sum_{i=1}^N I(s_i \leq a \wedge D = 0)}{\sum_{i=1}^N I(s_i \leq a)}}{\frac{\sum_{i=1}^N I(D = 0)}{\sum_{i=1}^N I(D = 0 \vee D = 1)}} \\
 &= \frac{\frac{\sum_{i=1}^N I(s_i \leq a \wedge D = 0)}{\sum_{i=1}^N I(s_i \leq a)}}{\frac{m}{N}}. \tag{13}
 \end{aligned}$$

It can be easily verified that the lift can be equivalently expressed as

$$Lift(a) = \frac{F_{n.BAD}(a)}{F_{N.ALL}(a)}, \quad a \in [L, H]. \tag{14}$$

Now, we would like to discuss the form of the lift function for the case of the ideal model. This is the model for which sets of output scores of bad and good clients are disjoint. So there exists a cut-off point, for which

$$P(S \leq a) = \begin{cases} P(S \leq a \wedge D = 0), & a \leq c, \\ P(D = 0) + P(S \leq a \wedge D = 1), & a > c. \end{cases} \quad (15)$$

Thus, we can derive the form of the lift function

$$Lift_{ideal}(a) = \begin{cases} \frac{1}{p_B}, & a \leq c, \\ \frac{1}{F_{N.ALL}(a)}, & a > c. \end{cases} \quad (16)$$

In practice, lift is computed corresponding to 10%, 20%, ..., 100% of clients with the worst score (see Coppock [3]). Usually, it is computed by using a table with the numbers of both all and bad clients in given score bands (deciles). An example of such a table is given by Table 1.

Table 1. Lift (absolute and cumulative form) computational scheme

Decile	Absolutely			Cumulatively			
	#Clients	# Bad clients	Bad rate	Abs. Lift	#Bad clients	Bad rate	Cum. Lift
1	100	35	35.0%	3.50	35	35.0%	3.50
2	100	16	16.0%	1.60	51	25.5%	2.55
3	100	8	8.0%	0.80	59	19.7%	1.97
4	100	8	8.0%	0.80	67	16.8%	1.68
5	100	7	7.0%	0.70	74	14.8%	1.48
6	100	6	6.0%	0.60	80	13.3%	1.33
7	100	6	6.0%	0.60	86	12.3%	1.23
8	100	5	5.0%	0.50	91	11.4%	1.14
9	100	5	5.0%	0.50	96	10.7%	1.07
10	100	4	4.0%	0.40	100	10.0%	1.00
All	1000	100	10.0%				

It is possible to compute the lift value in each decile (absolute lift in the fifth column in Table 1), but usually, and in accordance with the definition of $Lift(a)$, the cumulative form is used. It holds that the value of lift has an upper limit of $1/p_B$ and tends to a value of 1 when the score tends to infinity (or to its upper limit). In our case, we can see that the

best possible value of lift is equal to 10. We obtained the value 3.5 in the first decile, which is nothing excellent, but high enough for the model to be considered applicable in practice. Results are further illustrated in Figure 4.

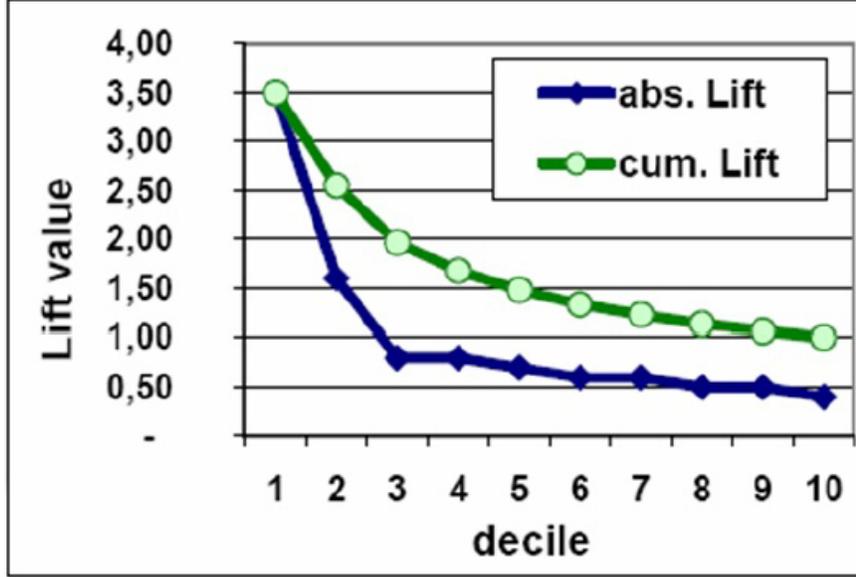


Figure 4. Lift value (absolute and cumulative).

In the context of this approach, we define

$$\begin{aligned} \mathcal{Q}Lift(q) &= \frac{F_{m.BAD}(F_{N.ALL}^{-1}(q))}{F_{N.ALL}(F_{N.ALL}^{-1}(q))} \\ &= \frac{1}{q} F_{m.BAD}(F_{N.ALL}^{-1}(q)), \quad q \in (0, 1], \end{aligned} \quad (17)$$

where q represents the score level of $100q\%$ of the worst scores and $F_{N.ALL}^{-1}(q)$ can be computed as

$$F_{N.ALL}^{-1}(q) = \min\{a \in [L, H], F_{N.ALL}(a) \geq q\}. \quad (18)$$

It can be easily shown that the lift function for the ideal model is now

$$QLift_{ideal}(q) = \begin{cases} \frac{1}{p_B}, & q \in (0, p_B], \\ \frac{1}{q}, & q \in (p_B, 1]. \end{cases} \quad (19)$$

Figure 5, below, gives an example of the lift function for ideal, random, and actual models.

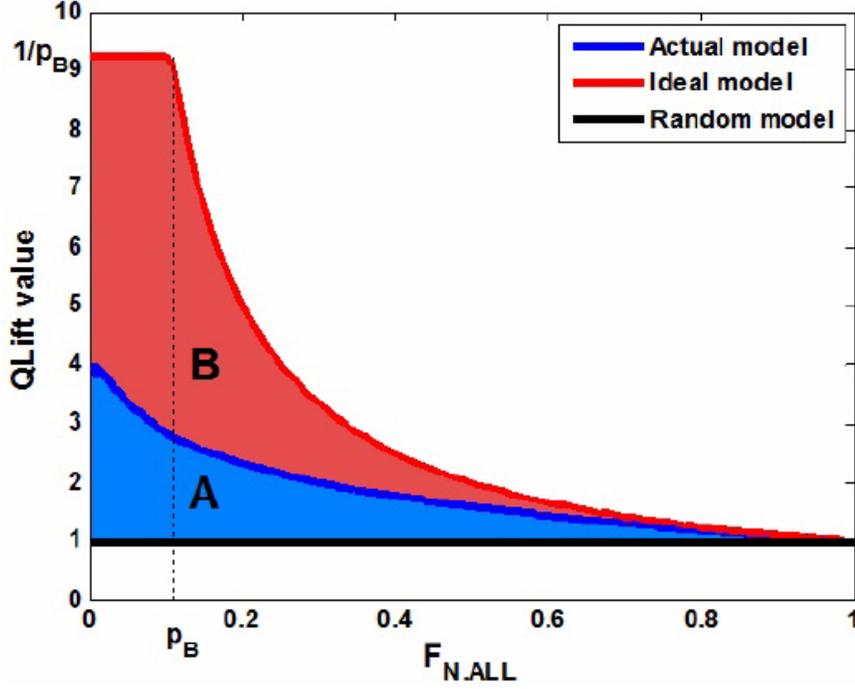


Figure 5. QLift function, lift ratio.

Using the previous Figure 5, we define lift ratio as analogous to Gini index

$$LR = \frac{A}{A+B} = \frac{\int_0^1 QLift(q) dq - 1}{\int_0^1 QLift_{ideal}(q) dq - 1}. \quad (20)$$

It is obvious that, it is a global measure of a model's quality and that it takes values from 0 to 1. Value 0 corresponds to the random model, value 1 matches the ideal model. The meaning of this index is quite simple: the higher, the better. An important feature is that lift ratio allows us to fairly compare two models developed on different data samples, which is not possible with lift.

Since lift ratio compares areas under the lift function corresponding to actual and ideal models, the next concept is focused on the comparison of lift functions themselves. We define the relative lift function by

$$RLift(q) = \frac{QLift(q)}{QLift_{ideal}(q)}, \quad q \in (0, 1]. \quad (21)$$

An example of this function is presented in Figure 6. The definition domain of the function is $[0, 1]$; the range is a subinterval of $[0, 1]$. The graph starts at point $[q_{\min}, p_B \cdot QLift(q_{\min})]$, where q_{\min} is a positive number near to zero. Then, it falls to a local minimum in point $[p_B, p_B \cdot QLift(p_B)]$ and then rises up to point $[1, 1]$. It is obvious that the graph of relative lift function for a better model is closer to the top line, which represents the function for the ideal model.

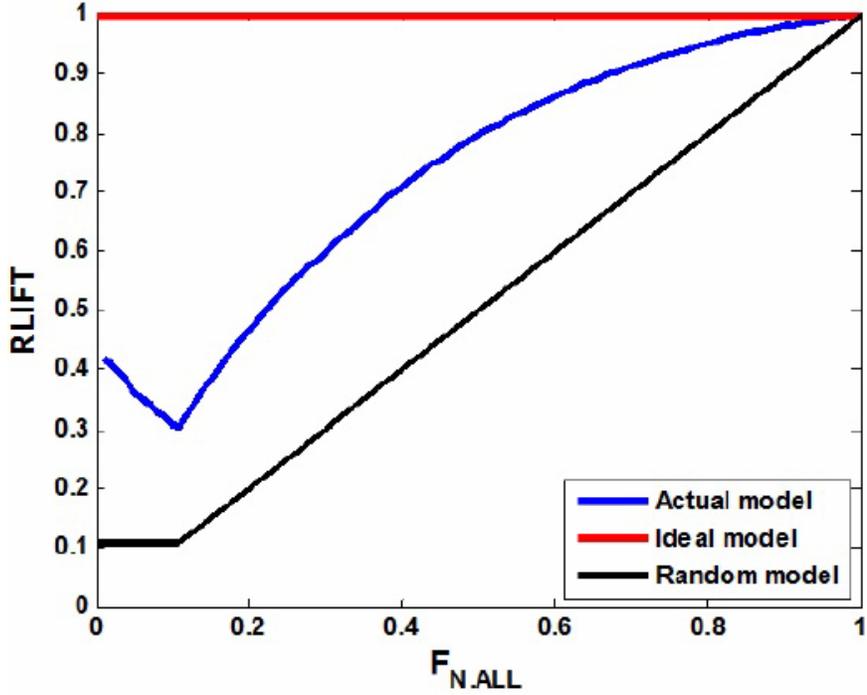


Figure 6. Relative lift function.

Now, it is natural to ask what we obtain when we integrate the relative lift function. We define the integrated relative lift (IRL) by

$$IRL = \int_0^1 RLift(q) dq. \quad (22)$$

It takes values from $0.5 + \frac{p_B^2}{2}$, for the random model, to 1, for the ideal model. Again the following holds: the higher, the better. This global measure of scoring a model's quality has an interesting connection to the c -statistic.

We made a simulation with scores generated from a normal distribution. The scores of bad clients had a mean equal to 0 and a variance equal to 1. The scores of good clients had a mean and variance

from 0.1 to 10 with a step equal 0.1. The number of samples and sample size were 1000, p_B was equal to 0.1. IRL and the c -statistic were computed for each sample and each value of the mean and variance of a good clients' scores. Finally, means of IRL and the c -statistic were computed. The results are presented in Figure 7. Part (b) represents the contour plot of the figure in part (a).

The simulation shows that IRL and the c -statistic are approximately equal when the variances of good and bad clients are equal. Furthermore, it shows that they significantly differ when the variances are different and the ratio of the mean and variance of good clients is near to 1.

4. Case Study

To illustrate the advantage of the proposed indexes, we introduce a simple case study. We consider two scoring models with a score distribution given in Table 2.

Furthermore, we consider the standard meaning of scores, i.e., a higher score band means better clients (clients with the lowest scores, i.e., clients in score band 1, have the highest probability of default).

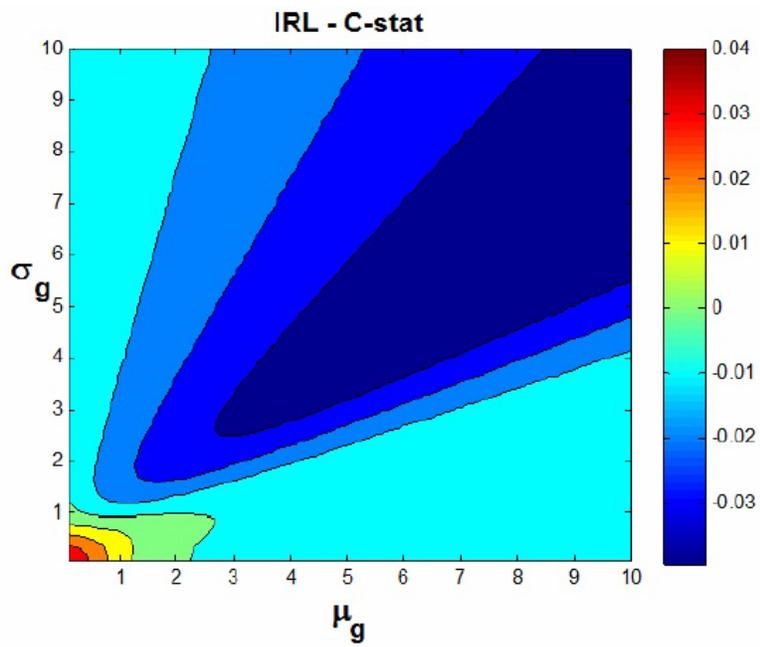
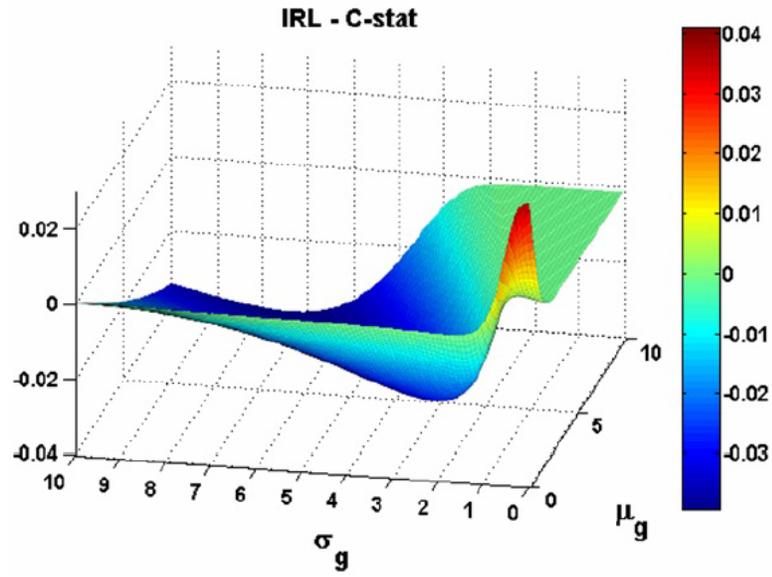


Figure 7. Difference of IRL and c -stat (a) and its contour plot (b).

Table 2. Score distribution and QLift of given scoring models

Score band	#Clients	q	Scoring model 1			Scoring model 2		
			# Bad clients	Cumul. bad rate	QLift	#Bad clients	Cumul. bad rate	QLift
1	100	0.1	20	20.0%	2.00	35	35.0%	3.50
2	100	0.2	18	19.0%	1.90	16	25.5%	2.55
3	100	0.3	17	18.3%	1.83	8	19.7%	1.97
4	100	0.4	15	17.5%	1.75	8	16.8%	1.68
5	100	0.5	12	16.4%	1.64	7	14.8%	1.48
6	100	0.6	6	14.7%	1.47	6	13.3%	1.33
7	100	0.7	4	13.1%	1.31	6	12.3%	1.23
8	100	0.8	3	11.9%	1.19	5	11.4%	1.14
9	100	0.9	3	10.9%	1.09	5	10.7%	1.07
10	100	1.0	2	10.0%	1.00	4	10.0%	1.00
All	1000		100			100		

The Gini index for each model is equal to 0.420. KS is equal to 0.356 for model 1 and to 0.344 for model 2. According to these numbers, one can say that both models are almost the same, maybe the first one is slightly better. However, if we look at the models in more detail, we find that they differ significantly. We get the first insight from their Lorenz curves in Figure 8.

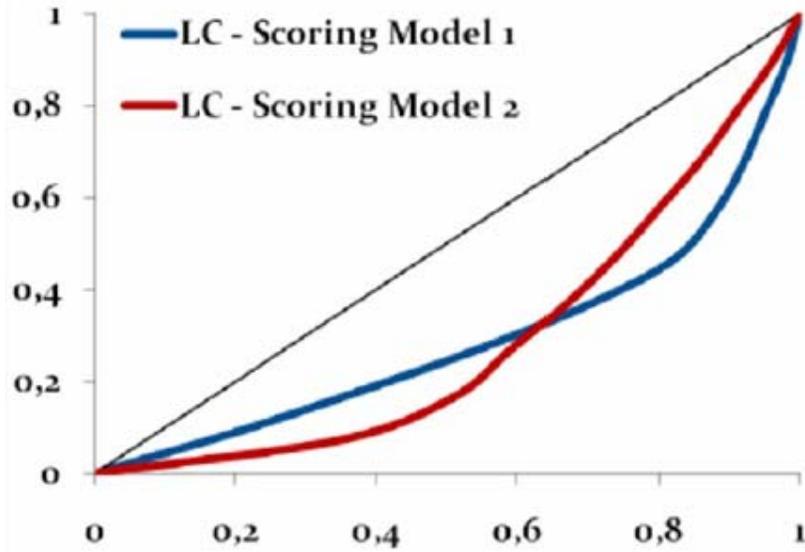


Figure 8. Lorenz curves for model 1 and model 2.

We can see that model 1 is stronger for higher score bands. This means that this model better separates the good from the best clients. On the other hand, model 2 is stronger for lower score bands, which means that it better separates the bad from the worst clients. We can read the same result from the figures of QLift and RLift in Figure 9.

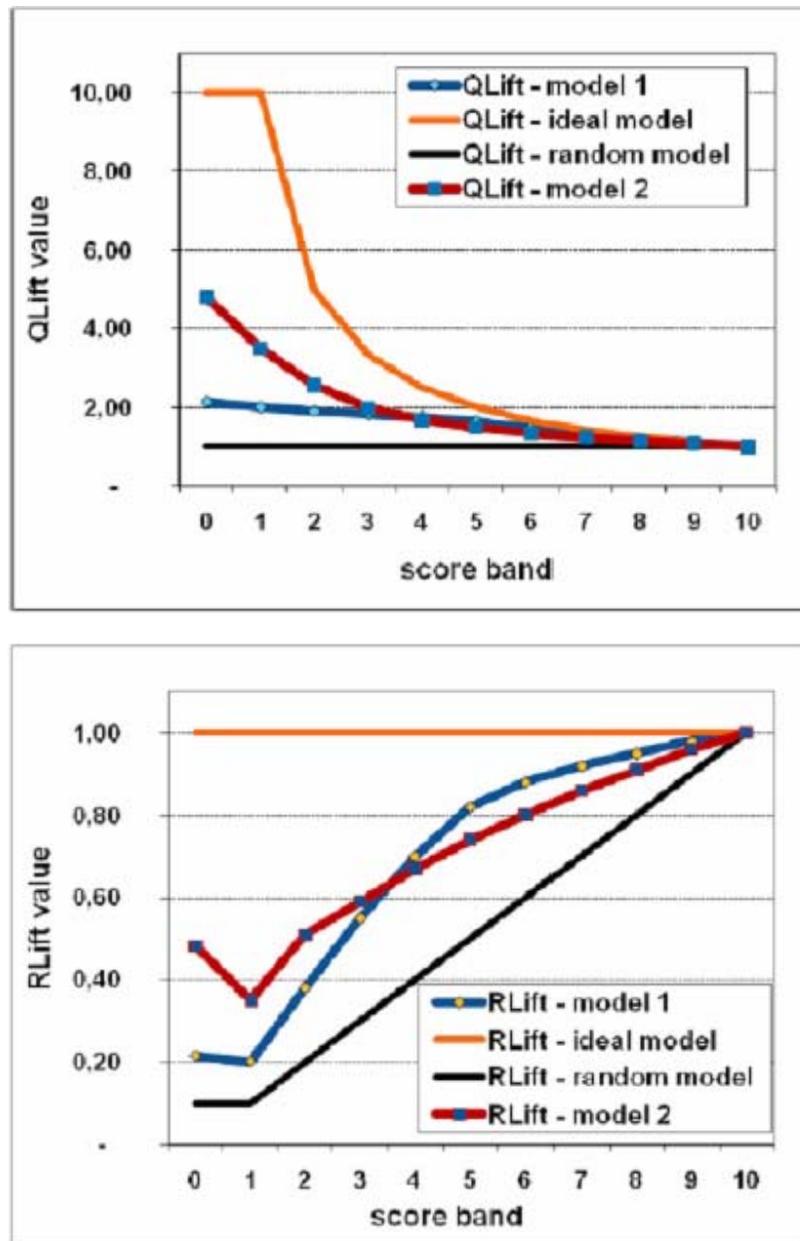


Figure 9. QLift and RLift for model 1 and model 2.

It is necessary to mention one computational problem at this point. In the discrete case, as in the case of Table 2, we do not know the value of QLift for q less than 0.1. Since QLift is not defined for $q = 0$, we need to extrapolate it somehow. According to the shape of the QLift curve, we propose using quadratic extrapolation, which yields

$$QLift(0) = 3 \cdot QLift(0.1) - 3 \cdot QLift(0.2) + QLift(0.3). \quad (23)$$

When we have a full data set, we can use formula (17). In this case, the extrapolation is not needed. Of course, we still do not have the value QLift (0). However, if we start the computation of QLift in some positive value of q , which is sufficiently near to zero, the final result is precise enough.

Overall, we can compare our two scoring models. Table 3, below, contains values of Gini indexes, K-S statistics, values of QLift(0.1), LR indexes, and IRL indexes. QLift(0.1) is a local measure of a model's quality; model 2 was designed to be better in the first score bands, hence it is natural that the value of QLift(0.1) is significantly higher for model 2, concretely 3.5 versus 2.0. On the other hand, all remaining indexes are global measures of a model's quality. Models were designed to have the same Gini index and similar KS. However, we can see that LR and IRL significantly differ for our models, 0.242 versus 0.372 and 0.699 versus 0.713, respectively.

Table 3. Quality indexes of two assessed scoring models

	Scoring model 1	Scoring model 2
Gini	0.420	0.420
KS	0.356	0.344
QLift(0.1)	2.000	3.500
LR	0.242	0.372
IRL	0.699	0.713

Finally, if the expected reject rate is up to 40%, which is a very natural assumption, using LR and IRL, we can state that model 2 is better than model 1 although their Gini indexes are equal and even their KS are in reverse order.

5. Conclusion

In Section 2, we presented widely used indexes for the assessment of credit scoring models. We focused mainly on the definitions of Lorenz curve, CAP, Gini index, AR, and lift. The Lorenz curve is sometimes confused with ROC. The discussion of their definitions is given within the paper. We suggest using the definition of the Lorenz curve given in Müller and Rönz [8], the definition of ROC given in Siddiqi [11], and the definition of CAP given in Sobehart et al. [12].

The main part of the paper, Section 3, was devoted to lift. Formulas for lift in basic and quantile form were presented as well as their forms for ideal models. These formulas allow the calculation of the value of lift for any given score and any given quantile level and comparison with the best obtainable results.

Lift ratio was presented as analogous to Gini index. An important feature is that LR allows the fair comparison of two models developed on different data samples, which is not possible with lift or QLift. Furthermore, a relative lift function was proposed, which shows the ratio of the QLifts of the actual and ideal models. Finally, integrated relative lift was defined. The connection to the c -statistic was presented by means of a simulation by using normally distributed scores. This simulation showed that IRL and the c -statistic are approximately equal in the case when the variances of good and bad clients are equal.

Despite the high popularity of the Gini index and KS, we conclude that the proposed lift based indexes are more appropriate for assessing the quality of credit scoring models. In particular, it is better to use them in the case of an asymmetric Lorenz curve. In such cases, using the Gini index or KS during the development process could lead to the selection of a weaker model.

Acknowledgement

This research was supported by our department and by The Jaroslav Hájek Center for Theoretical and Applied Statistics (grant No. LC 06024).

References

- [1] R. Anderson, *The Credit Scoring Toolkit: Theory and Practice for Retail Credit Risk Management and Decision Automation*, Oxford University Press, Oxford, 2007.
- [2] M. J. A. Berry and G. S. Linoff, *Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management*, 2nd Edition, Wiley, Indianapolis, 2004.
- [3] D. S. Coppock, Why Lift? *DM Review Online*, (2002). [Accessed on 1 December 2009].
www.dmreview.com/news/5329-1.html
- [4] J. N. Crook, D. B. Edelman and L. C. Thomas, Recent developments in consumer credit risk assessment, *European Journal of Operational Research* 183(3) (2007), 1447-1465.
- [5] B. Engelmann, E. Hayden and D. Tasche, Measuring the Discriminatory Power of Rating System, (2003). [Accessed on 4 October 2010].
http://www.bundesbank.de/download/bankenaufsicht/dkp/200301dkp_b.pdf
- [6] P. Giudici, *Applied Data Mining: Statistical Methods for Business and Industry*, Wiley, Chichester, 2003.
- [7] D. J. Hand and W. E. Henley, Statistical classification methods in consumer credit scoring: A review, *Journal of the Royal Statistical Society, Series A* 160(3) (1997), 523-541.
- [8] M. Müller and B. Rönz, Credit Scoring using Semiparametric Methods, In: J. Franke, W. Härdle and G. Stahl (Eds.), *Measuring Risk in Complex Stochastic Systems*, Springer-Verlag, New York, 2000.
- [9] R. B. Nelsen, Concordance and Gini's measure of association, *Journal of Nonparametric Statistics* 9(3) (1998), 227-238.
- [10] R. Newson, Confidence intervals for rank statistics: Somers' D and extensions, *The Stata Journal* 6(3) (2006), 309-334.
- [11] N. Siddiqi, *Credit Risk Scorecards: Developing and Implementing Intelligent Credit Scoring*, Wiley, New Jersey, 2006.
- [12] J. Sobehart, S. Keenan and R. Stein, *Benchmarking Quantitative Default Risk Models: A Validation Methodology*, Moody's Investors Service, (2000). [Accessed on 4 October 2010].
<http://www.algorithmics.com/EN/media/pdfs/Algo-RA0301-ARQ-DefaultRiskModels.pdf>

- [13] R. H. Somers, A new asymmetric measure of association for ordinal variables, *American Sociological Review* 27 (1962), 799-811.
- [14] L. C. Thomas, A survey of credit and behavioural scoring: Forecasting financial risk of lending to consumers, *International Journal of Forecasting* 16(2) (2000), 149-172.
- [15] L. C. Thomas, D. B. Edelman and J. N. Crook, *Credit Scoring and its Applications*, SIAM Monographs on Mathematical Modelling and Computation, Philadelphia, 2002.
- [16] L. C. Thomas, *Consumer Credit Models: Pricing, Profit, and Portfolio*, Oxford University Press, Oxford, 2009.
- [17] A. D. Wilkie, Measures for Comparing Scoring Systems, In: L. C. Thomas, D. B. Edelman and J. N. Crook (Eds.): *Readings in Credit Scoring*, Oxford University Press, Oxford, (2004), 51-62.
- [18] K. Xu, How has the literature on Gini's index evolved in past 80 years? (2003). [Accessed on 1 December 2009].

economics.dal.ca/RePEc/dal/wparch/howgini.pdf

