Credit Scoring Models and their Quality

Jan Koláček, Martin Řezáč

Department of Mathematics and Statistics
Faculty of Science
Masaryk University
Brno, Czech Republic
www.muni.cz

Contents

- Introduction
- Quality indexes
- Proposed index
- Simulation study
- Example
- References

Credit Scoring Model

Construction of the test

- $\mathcal{G}_0$ group of n_0 bad clients
- $\mathcal{G}_1$ group of n_1 good clients
- S – the score for each client (one-dimensional absolutely continuous random variable)
- $D = 0,\ 1$ random variable denotes bad or good client
- c – given cutoff point, $c \in \mathbb{R}$
- The client is classified as $\mathcal{G}_1$ if $S \geq c$ and $\mathcal{G}_0$ otherwise for given cutoff point c

Measure of Diagnostic Accuracy

- $T = 1$ positive test result
- $T = 0$ negative test result

Test results: Confusion matrix

	Positive test, $T = 1$	Negative test, $T = 0$	Total
$\mathcal{G}_1$ $(D = 1)$	True positive (TP)	False negative (FN)	$TP + FN$
$\mathcal{G}_0$ $(D = 0)$	False positive (FP)	True negative (TN)	$FP + TN$
Total	$TP + FP$	$FN + TN$	$n = n_0 + n_1$

The *sensitivity* (Se) of the test is its ability to detect good client when he is good. $Se = P(T = 1|D = 1)$ is a probability P that the test result is positive $(T = 1)$, given that the client is good $(D = 1)$.

The *specificity* (Sp) of the test is its ability to exclude the solidity of client when it is absent. $Sp = P(T = 0|D = 0)$ is a probability P that the test result is negative $(T = 0)$, given that the client is bad $(D = 0)$.

Extreme models

Ideal model: $Se = Sp = 1$

Random model: $Se = Sp = 1/2$.

Notation

Assume the realization $s \in \mathbb{R}$ of random value S (score) is available for each client.
Let F_0, F_1 denote cumulative distribution functions of score of bad and good clients, i.e.

$$F_0(a) = P(S \leq a \,|\, D = 0),$$

$$F_1(a) = P(S \leq a \,|\, D = 1),\ a \in \mathbb{R}.$$

Assumption: F_0, F_1 and their corresponding densities f_0, f_1 are continuous on $\mathbb{R}$.

Practice

Empirical estimators of distribution functions

$$\widehat{F}_0(a) = \frac{1}{n_0} \sum_{i=1}^{n} I(s_i \leq a \wedge D = 0)$$

$$\widehat{F}_1(a) = \frac{1}{n_1} \sum_{i=1}^{n} I(s_i \leq a \wedge D = 1), \ a \in [L, H],$$

where

$I(A)$ ... the indicator of event A

s_i ... the score of i-th client

$n_0, \ n_1$ ... number of bad and good clients, $n = n_0 + n_1$

L ... the minimum value of given score

H ... the maximum value of given score

Lorenz curve

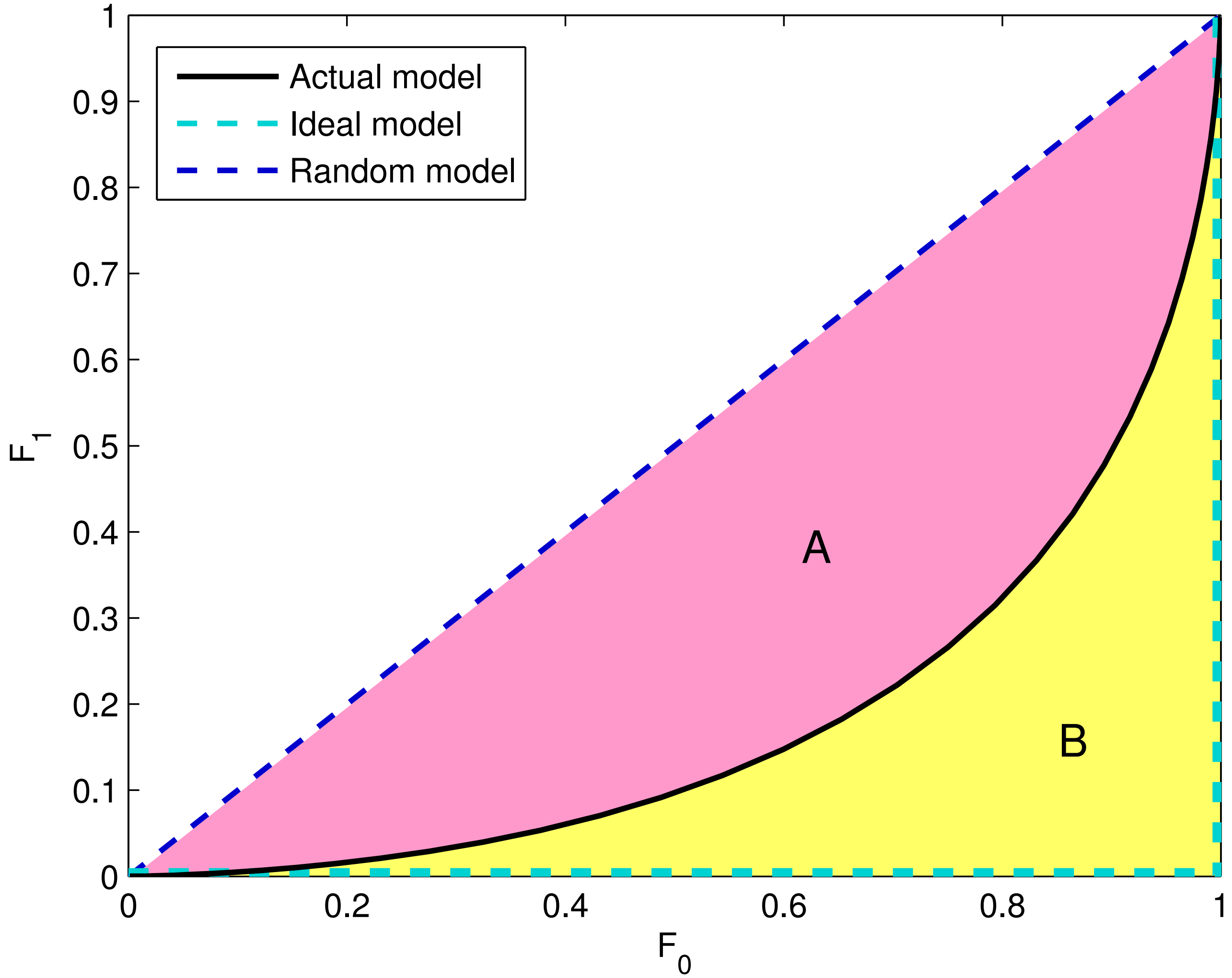
The curve is given parametrically by

$$x = F_0(a)$$

$$y = F_1(a),\ \ a \in \mathbb{R}.$$

Notation: $x = F_0(a),\ \ R(x) = F_1(F_0^{-1}(x))$

we can write the Lorenz curve as $R(x),\ \ x \in [0,1]$.



Lorenz curve, Gini index

Gini index

Definition

$$Gini = \frac{A}{A+B} = 2A,$$

where

A . . . area between the diagonal and Lorenz curve for actual model

$A+B$ . . . area between the diagonal and Lorenz curve for ideal model

Properties

$Gini \in [0,1]$

random model $\Rightarrow Gini = 0$

ideal model $\Rightarrow Gini = 1$

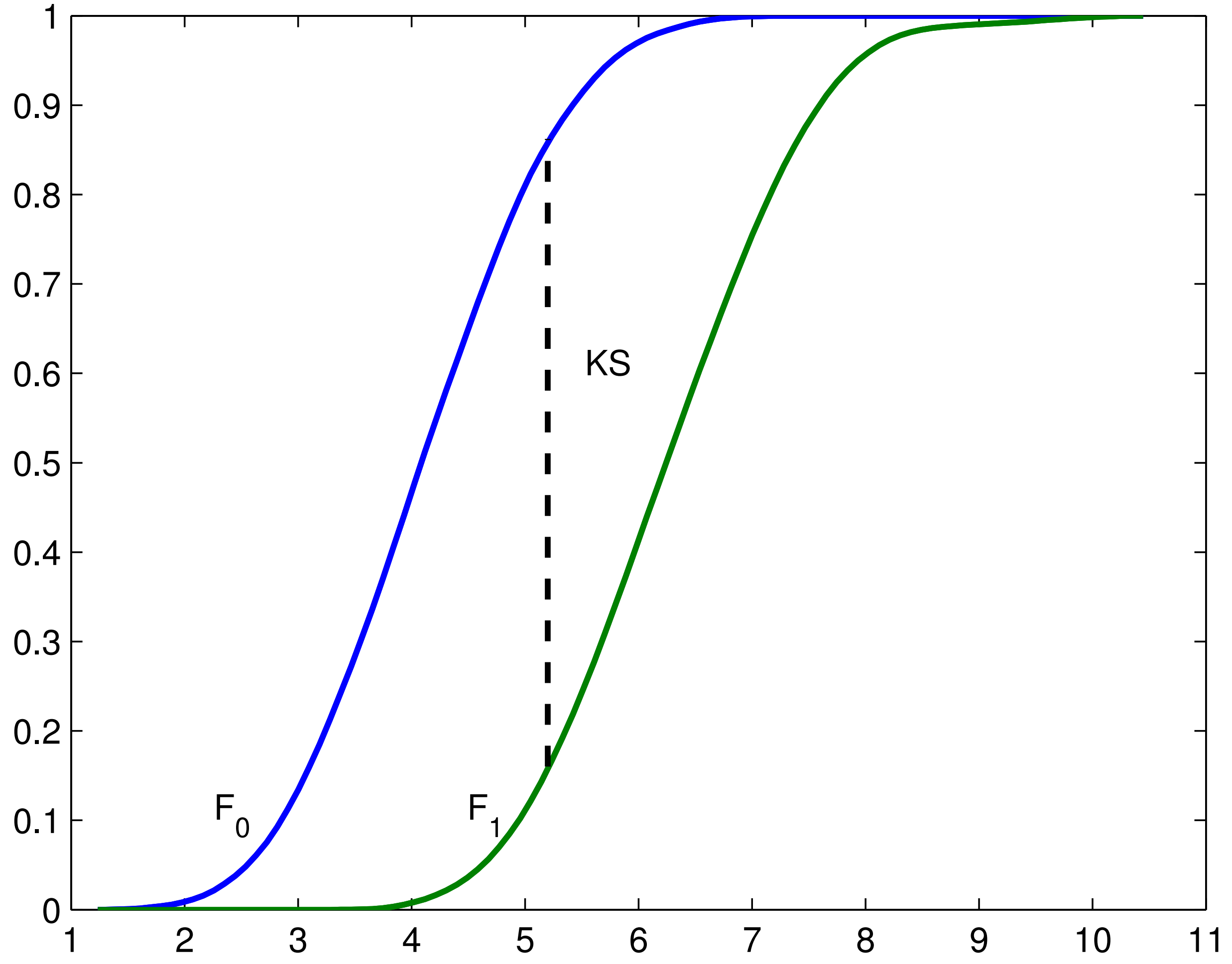
Kolmogorov-Smirnov statistics

Definition

$$KS = \max_{a \in \mathbb{R}} \left| F_0(a) - F_1(a) \right| .$$

Remark In context with notation $R(x)$ for the Lorenz curve we can express K-S statistics as

$$KS = \max_{x \in [0,1]} \left| x - R(x) \right| .$$



K-S statistics

The Lift, QLift

Definition

$$Lift(a) = \frac{P(D=0 \,|\, S \le a)}{P(D=0)} = \frac{P(S \le a \,|\, D=0)}{P(S \le a)} = \frac{F_0(a)}{F_{ALL}(a)},$$

where

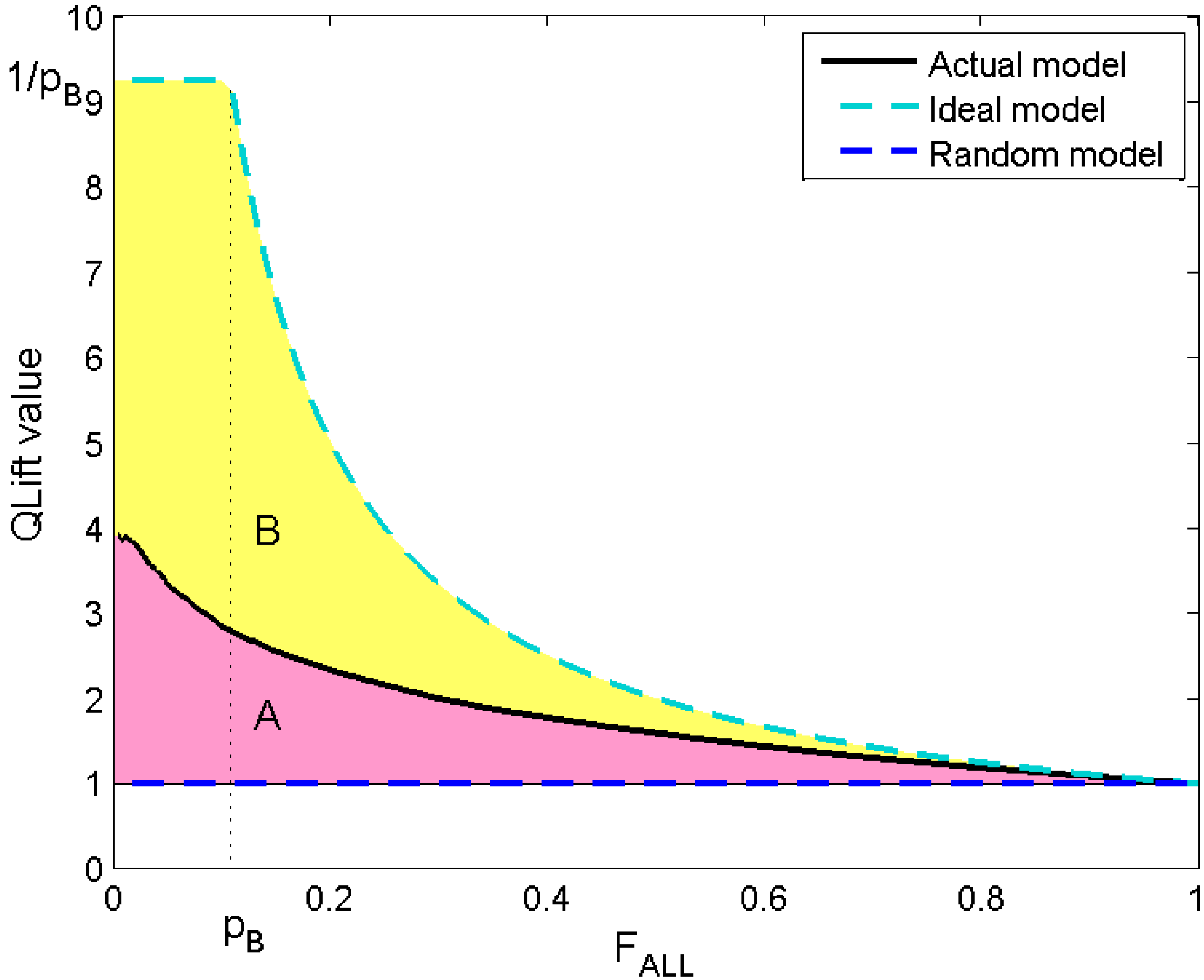
$$F_{ALL}(a) = P(S \le a) = P(S \le a \wedge D = 0) + P(S \le a \wedge D = 1).$$

If we denote $p_B = P(D=0)$, we can write

$$Lift(a) = \frac{F_0(a)}{p_B F_0(a) + (1-p_B)F_1(a)}, \quad a \in \mathbb{R}.$$

Remark The transformation $q = F_{ALL}(a)$ leads to $QLift$

$$QLift(q) = \frac{1}{q} F_0(F_{ALL}^{-1}(q)),\ q \in (0,1],$$


10
1/p_B 9
8
7
6
5
4
3
2
1
0
QLift value
Actual model
Ideal model
Random model
B
A
0
0.2
0.4
0.6
0.8
1
p_B
F_ALL

The QLift

Lift Ratio

As analogy to Gini index, we can choose a similar approach to derive the Lift Ratio (LR) index for Lift

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

where $QLift_{ideal}(q)$ represents the value of $QLift(q)$ for the case of ideal model.

For more detailed description of LR index, see Řezáč and Koláček [3].

Proposed index

Let $a \in \mathbb{R}$ be a cut-off point. Let us consider the classical contingency table of given discrimination problem

			Σ
	$P(S>a \mid D=1)P(D=1)$	$P(S\leq a \mid D=1)P(D=1)$	$n_{1\cdot}$
	$P(S>a \mid D=0)P(D=0)$	$P(S\leq a \mid D=0)P(D=0)$	$n_{2\cdot}$
Σ	$n_{\cdot 1}$	$n_{\cdot 2}$	1

We can express the probabilities in the table by cumulative distribution functions F_0, F_1. The table takes the form

			Σ
	$(1-F_1(a))(1-p_B)$	$F_1(a)(1-p_B)$	$n_{1\cdot}$
	$(1-F_0(a))p_B$	$F_0(a)p_B$	$n_{2\cdot}$
Σ	$n_{\cdot 1}$	$n_{\cdot 2}$	1

Pearson's Chi-square test of independence for contingency table:

$$\chi^2(a) = \frac{(n_{11}n_{22}-n_{12}n_{21})^2}{n_{\cdot 1}n_{\cdot 2}n_{1\cdot}n_{2\cdot}}$$

$$= \frac{(F_0(a)-F_1(a))^2}{(F_0(a)-F_1(a))^2+\frac{1}{p_B}F_1(a)(1-F_1(a))+\frac{1}{1-p_B}F_0(a)(1-F_0(a))}$$

The value $\chi^2(a)$ describes the power of dependence of both groups (good and bad clients) for given score value a.

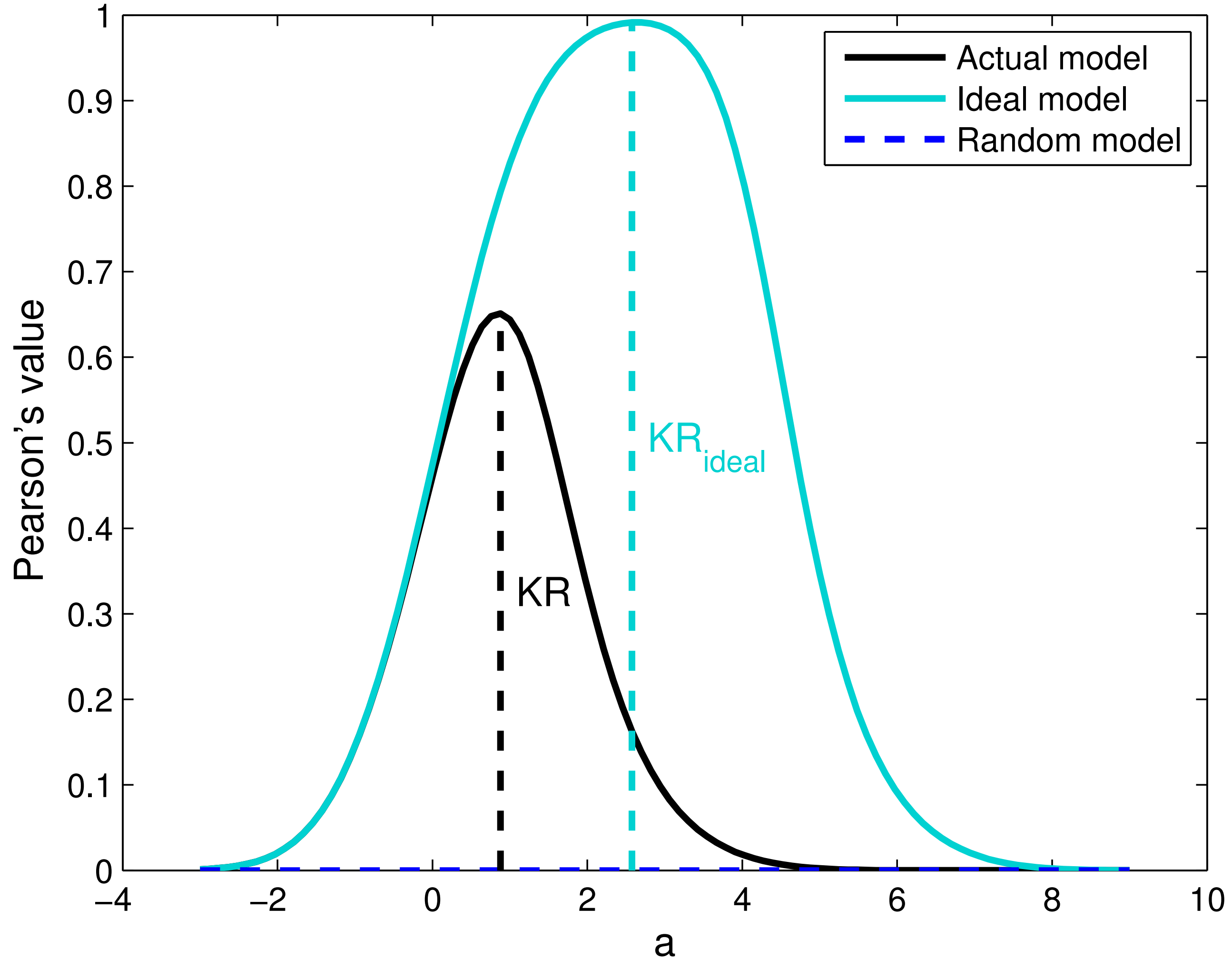
Definition The proposed index *KR*

$$KR = \max_{a \in \mathbb{R}} \chi^2(a).$$

Properties of $\chi^2(a)$:

- $\chi^2(a) \in [0, 1], \quad \forall a \in \mathbb{R}$
- $\chi^2(a) \to 0$ for $a \to \pm\infty$
- For ideal model $\Rightarrow \exists a \in \mathbb{R}$ such that $\chi^2(a) = 1$
- For random model $\chi^2(a) = 0, \ \forall a \in \mathbb{R}$

The KR index is a type of “generalization” of KS index. However, it takes some advantages. Moreover, it reflects the proportion of bad clients, so it gives more information about actual model then KS index.



Proposed Index

Simulation Study

Parameters of simulation

- distribution of bad clients $N(\mu_0, \sigma^2)$
- distribution of good clients $N(\mu_1, \sigma^2)$
- $\mu_0 < \mu_1$

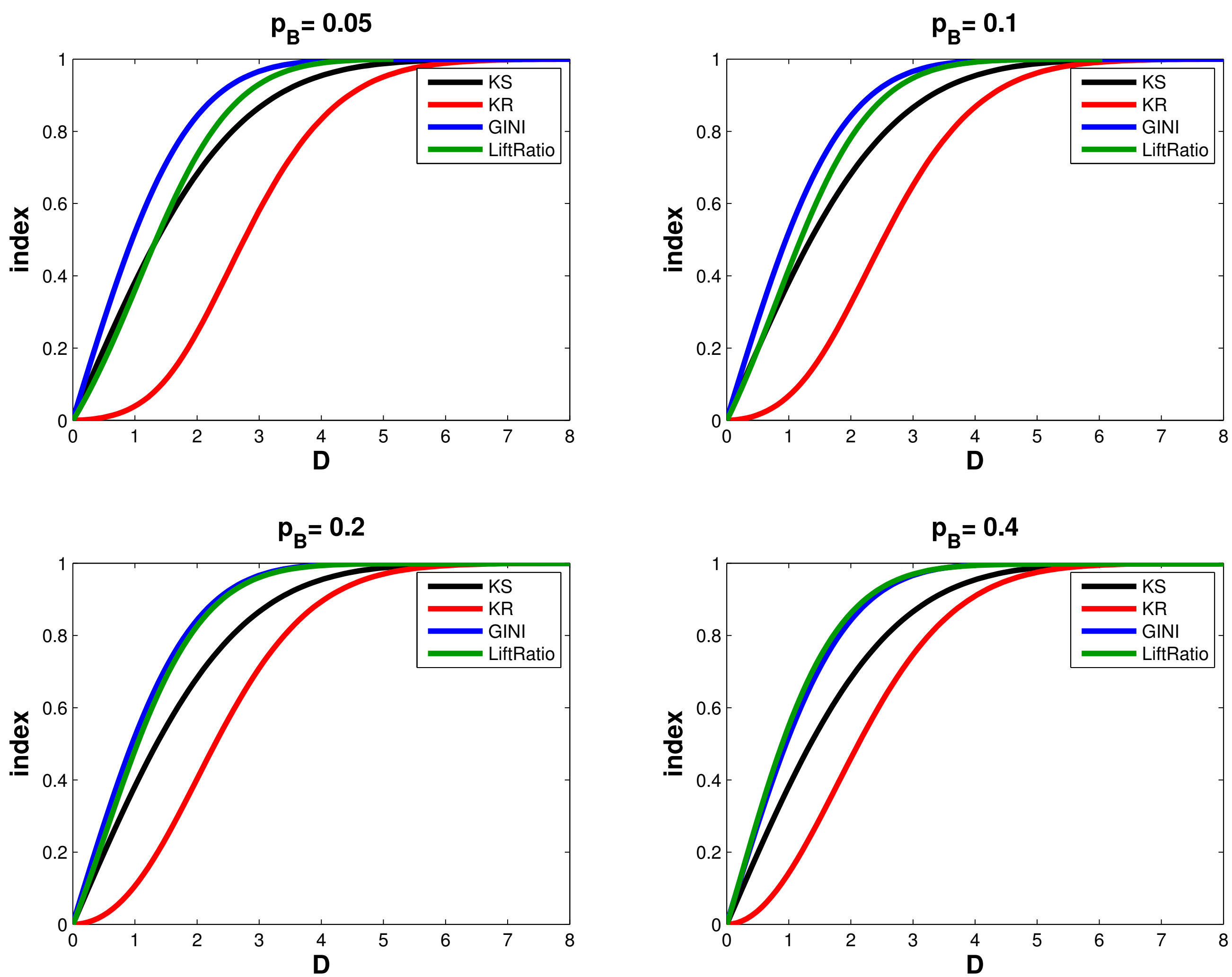
Let us define Mean Difference D (*Mahalanobis distance*)

$$D = \frac{\mu_1 - \mu_0}{\sigma}.$$

It describes the difference between score of groups of bad and good clients. It takes values from 0 to ∞. In our simulation study, we have calculated all quality indexes for each value of D.

Four cases of models:

$$p_B = 0.05,\ 0.1,\ 0.2,\ 0.4.$$



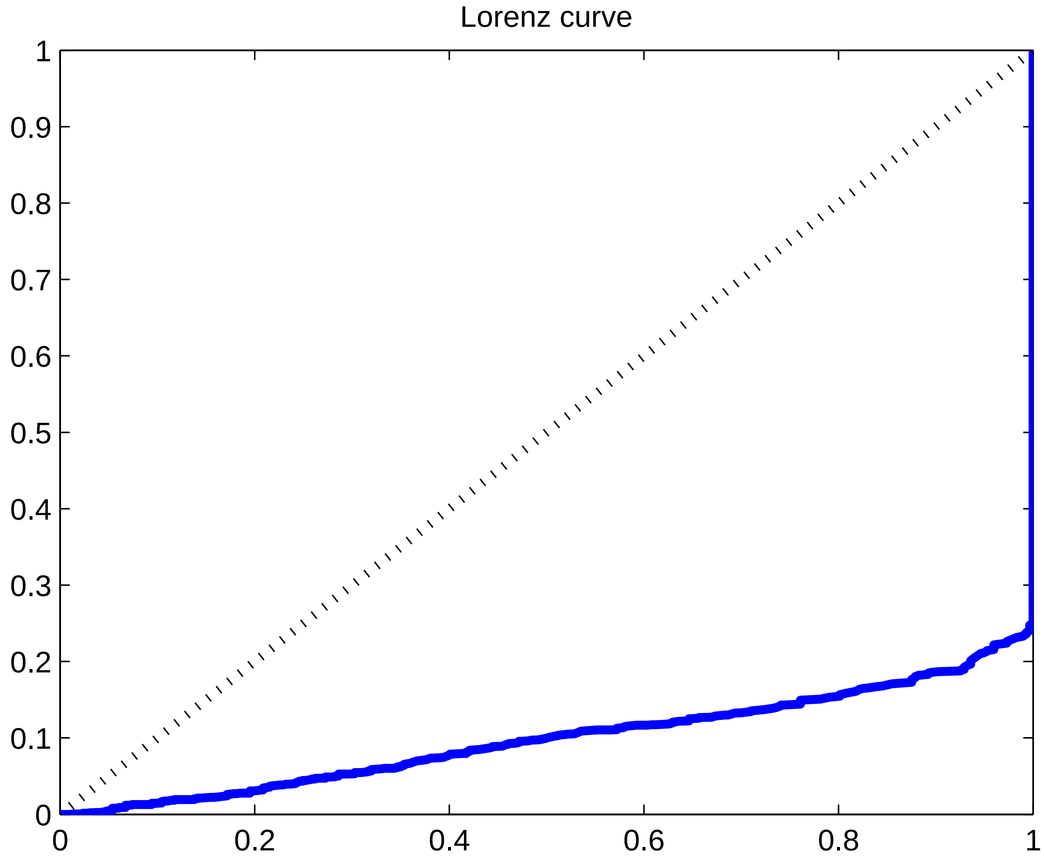
Dependence on D for all indexes.

Real data

Consumer loans data

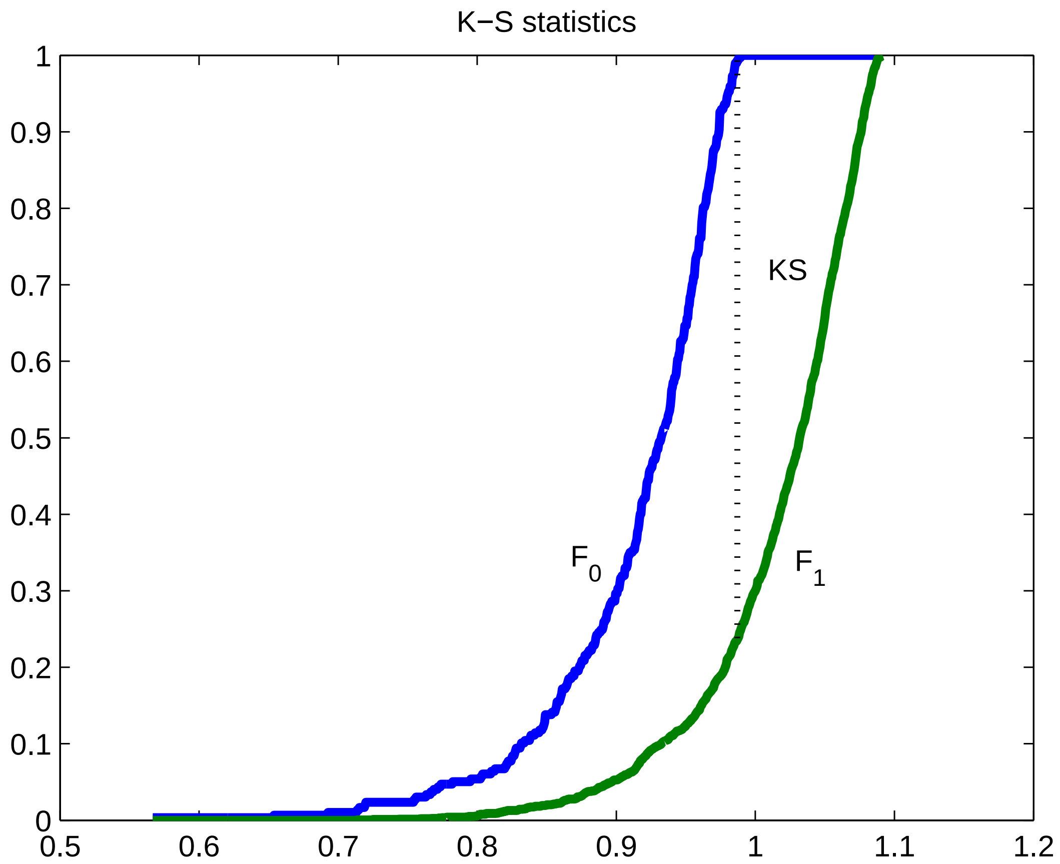
- The use of some (not specified) scoring function for predicting the likelihood of repayment of a client.
- We are interested in determining which clients are able to repay their loans.
- A test set: 2327 clients – 2030 have repaid their loans (group $\mathcal{G}_1$) and 297 had problems with payments or did not pay (group $\mathcal{G}_0$). Thus $p_B \doteq 0.13$.
- We use mentioned indexes to assess the discrimination power of given scoring function.

The empirical estimate of Lorenz curve, $Gini = 0.803$



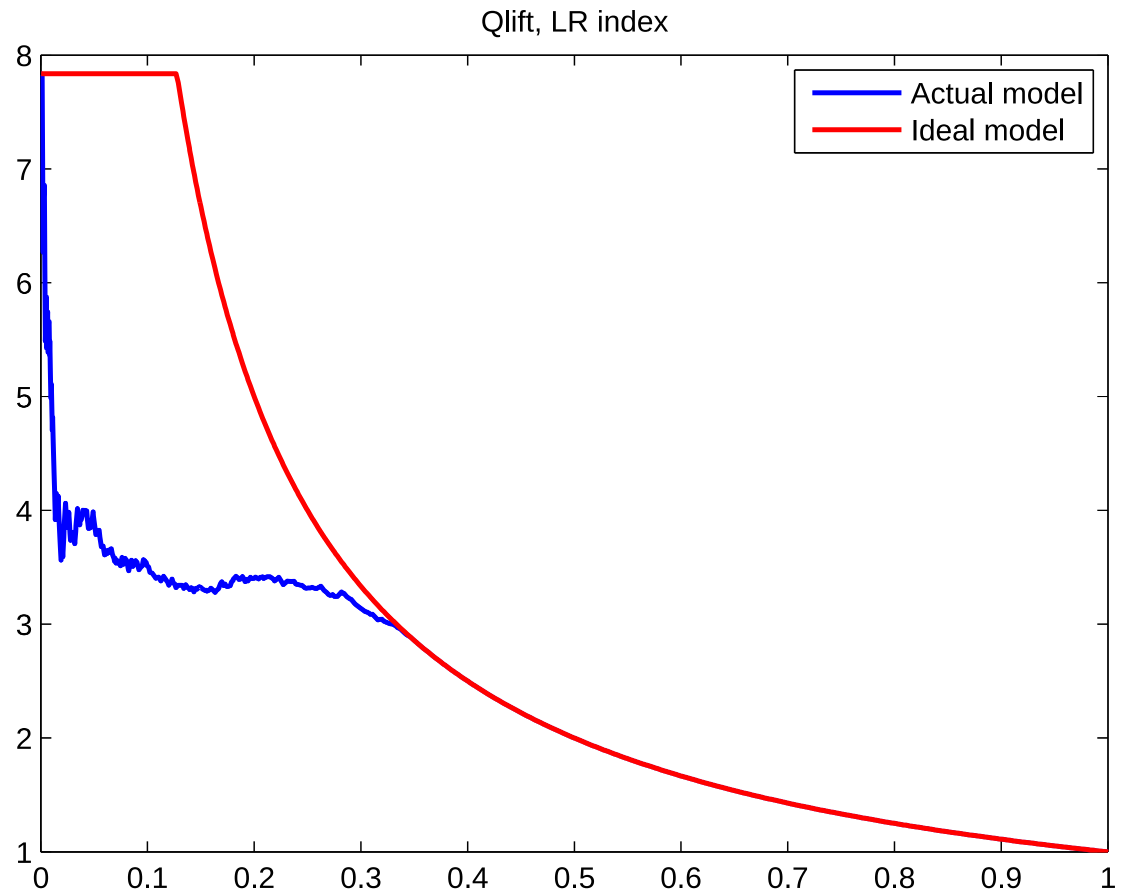
Lorenz curve

The empirical estimates of $F_0,\ F_1,\ KS = 0.757$



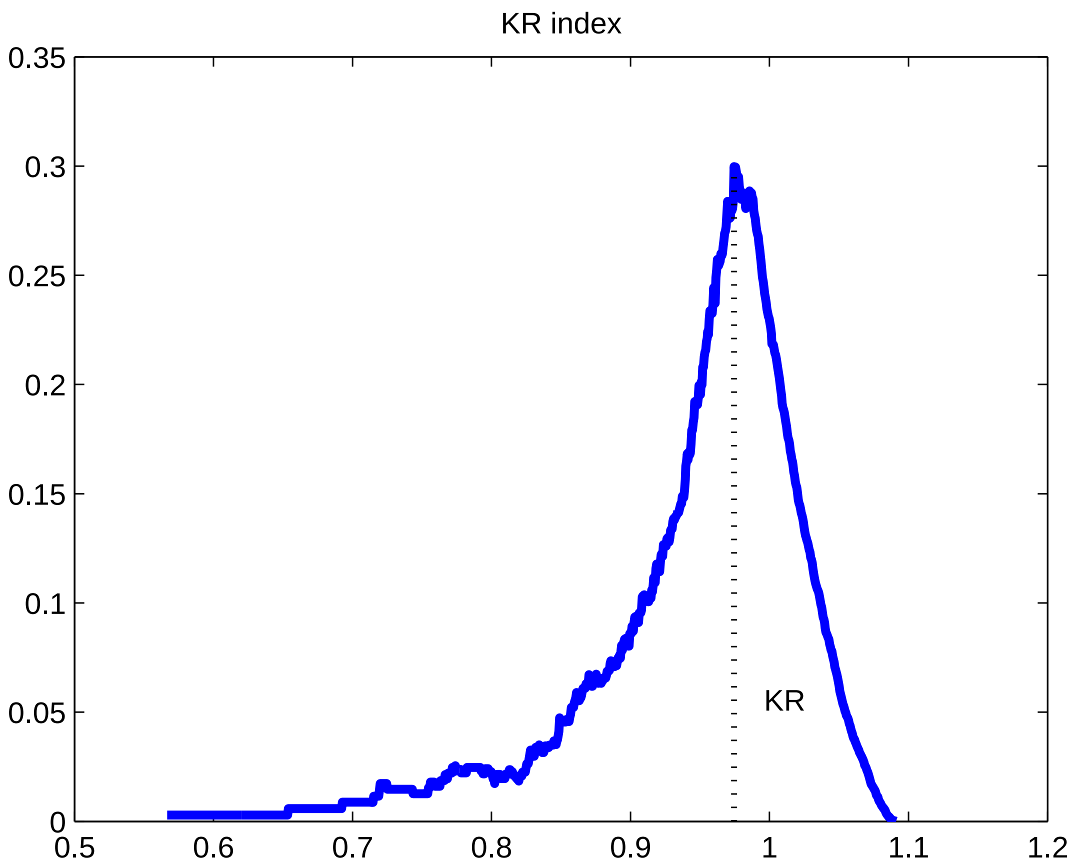
K-S statistics

The empirical estimate of QLift, $LR = 0.615$



QLift and Lift Ratio

The empirical estimate of χ^2, $KR = 0.300$



KR index

Summary of measures

	Gini	K–S	LR	KR
Index for the data	0.803	0.757	0.615	0.300

Conclusions

- all described indexes are widely used in practice
- we developed a new approach to measure power of scoring models
- the proposed index is more conservative

References

[1] Anderson, R. *The Credit Scoring Toolkit: Theory and Practice for Retail Credit Risk Management and Decision Automation*. Oxford University Press, Oxford, 2007.

[2] Hand, D.J., Henley, W.E. Statistical Classification Methods in Consumer Credit Scoring: a review. *Journal of the Royal Statistical Society*, Series A. 160 (3), 523-541, 1997.

[3] Řezáč, M., Koláček, J. On Aspects of Quality Indexes for Scoring Models. *19th International Conference on Computational Statistics, Paris France, August 22-27, 2010 Keynote, Invited and Contributed Papers* 1, 1517-1524, 2010.

[4] Siddiqi, N. *Credit Risk Scorecards: developing and implementing intelligent credit scoring*. Wiley, New Jersey, 2006.

[5] Thomas, L.C. A survey of credit and behavioural scoring: forecasting financial risk of lending to consumers. *International Journal of Forecasting* 16 (2), 149-172, 2000.

[6] Thomas, L.C. *Consumer Credit Models: Pricing, Profit, and Portfolio*. Oxford University Press, Oxford, 2009.

[7] Thomas, L.C., Edelman, D.B., Crook, J.N. *Credit Scoring and Its Applications*. SIAM Monographs on Mathematical Modeling and Computation, Philadelphia, 2002.

[8] Xu, K. How has the literature on Gini's index evolved in past 80 years?. *economics.dal.ca/RePEc/dal/wparch/howgini.pdf*, 2003. Accessed on 1 December 2009.