MASARYK UNIVERSITY
Czech Republic

ECDM' 12
Lisbon

INFORMATION VALUE ESTIMATOR FOR CREDIT SCORING MODELS

Martin Řezáč
Dept. of Mathematics and Statistics, Faculty of Science,
Masaryk University

Introduction

- Information value is widely used to assess discriminatory power of credit scoring models.
  - I.e. models that try to predict a probability of client's default.
- Moreover it is very often used to assess the discriminatory power of variables that enter into these models.
  - This means that the Information value is used as a filter for variable selection.
- However, empirical estimate using deciles of scores, which is the common way how to compute it, may lead to strongly biased results.
- The main aim of this paper is to give an alternative estimator of the Information value, which leads to lowered bias and MSE.

MASARYK UNIVERSITY
Czech Republic

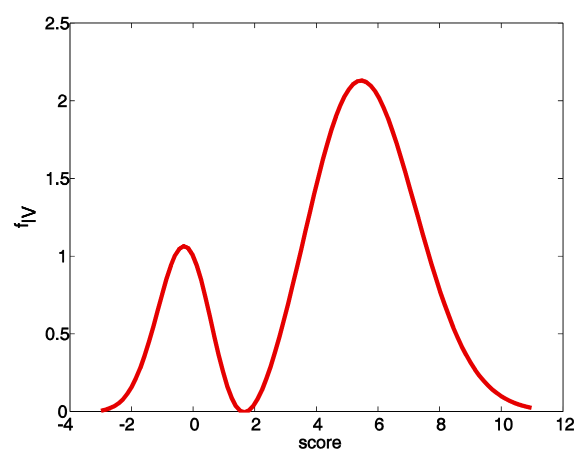
Information value

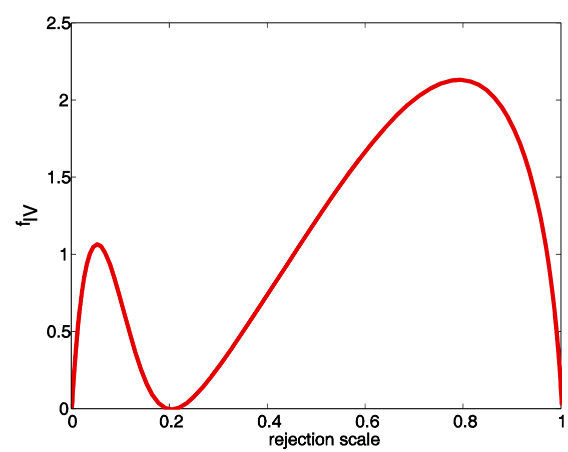
- The special case of Kullback-Leibler divergence given by:

$$I_{val} = \int_{-\infty}^{\infty} f_{IV}(x)dx \quad \text{where } f_{IV}(x) = (f_1(x) - f_0(x)) \ln\left(\frac{f_1(x)}{f_0(x)}\right)$$

  - f_0, f_1 are densities of scores of bad and good clients.

The example of $f_{IV}(x)$ for 10% of bad clients with $f_0 \sim N(0,1)$ and 90% of good clients with $f_1 \sim N(4,2)$





MASARYK UNIVERSITY
Czech Republic

Kernel estimator, empirical est.

- The kernel density estimators are defined by $\tilde{f}_0(x,h_0) = \frac{1}{n_0}\sum_{i=1}^{n_0} K_{h_0}(x - s_{0_i})$ $\tilde{f}_1(x,h_1) = \frac{1}{n_1}\sum_{i=1}^{n_1} K_{h_1}(x - s_{1_i})$

where $K_{h_i}(x) = \frac{1}{h_i} K\left(\frac{x}{h_i}\right), i = 0,1$ and K is the Epanechnikov kernel.

$$\tilde{f}_{IV}(x) = \left(\tilde{f}_1(x,h_1) - \tilde{f}_0(x,h_0)\right)\ln\left(\frac{\tilde{f}_1(x,h_1)}{\tilde{f}_0(x,h_0)}\right).$$

- for given $M+1$ equidistant points $L = x_0, x_1, \ldots, x_M = H$

$$\widehat{I}_{val,KERN} = \frac{H-L}{2M}\left(\tilde{f}_{IV}(L) + 2\sum_{i=1}^{M-1}\tilde{f}_{IV}(x_i) + \tilde{f}_{IV}(H)\right).$$

- The main idea of empirical estimators is to replace unknown densities by their empirical estimates. Consider counts of bad (good) clients in given intervals of score:

$$n_{0_j} = \sum_{i=1}^{n_0} I\left(s_{0_i} \in (q_{j-1}, q_j]\right)$$
$$n_{1_j} = \sum_{i=1}^{n_1} I\left(s_{1_i} \in (q_{j-1}, q_j]\right), \quad j = 1, \ldots, r$$

$$I(A) = \begin{cases} 1 & A \text{ is true} \\ 0 & A \text{ is false} \end{cases}$$

- For intervals given by deciles of score we have:

$$\widehat{I}_{val,DEC} = \sum_{j=1}^{r}\left(\frac{n_{1_j}}{n_1} - \frac{n_{0_j}}{n_0}\right)\ln\left(\frac{n_{1_j} n_0}{n_{0_j} n_1}\right)$$

Empirical estimate with supervised interval selection (ESIS)

- We want to avoid zero values of n_{0j} or n_{1j} .
- We propose to require to have at least *k*, where *k* is a positive integer, observations of scores of both good and bad clients in each interval.
- Intervals of score are given by

$$q_0 = L - 1$$
$$q_i = \widehat{F_0^{-1}}\left(\frac{k \cdot i}{n_0}\right), i = 1, \ldots, \lfloor \frac{n_0}{k} \rfloor$$
$$q_{\lfloor \frac{n_0}{k} \rfloor + 1} = H,$$

where $\widehat{F_0^{-1}}(\cdot)$ is the empirical quantile function appropriate to the empirical cumulative distribution function of scores of bad clients.

Empirical estimate with supervised interval selection (ESIS)

- Usage of quantile function of scores of bad clients is motivated by the assumption, that number of bad clients is less than number of good clients.

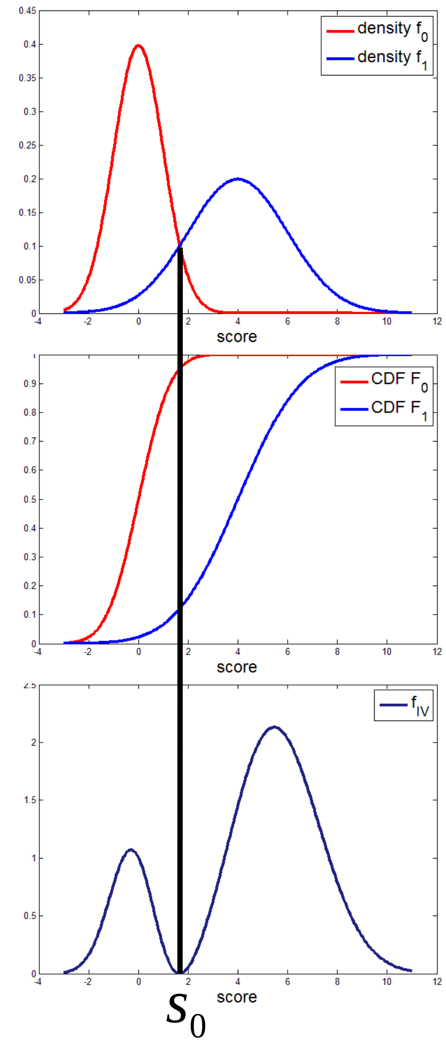
- If n_0 is not divisible by k, it is necessary to adjust our intervals, because we obtain number of scores of bad clients in the last interval, which is less than k. In this case, we have to merge the last two intervals.

- Furthermore we need to ensure, that the number of scores of good clients is as required in each interval. To do so, we compute n_{1j} for all actual intervals. If we obtain $n_{1j} < k$ for j^{th} interval, we merge this interval with its neighbor on the right side.

- This can be done for all intervals except the last one. If we have $n_{1j} < k$ for the last interval, than we have to merge it with its neighbor on the left side, i.e. we merge the last two intervals.

ESIS.2

- The original ESIS often merge calculated intervals in the second phase of the algorithm.
- Only $F_0^{-1}(.)$ is used for computation.
- But it is clear that in order to meet the condition $n_{11} > k$, the border of the first interval has to be greater or equal to $F_1^{-1}\left(\frac{k}{n_1}\right)$.
- This directly leads to idea to use F_1 firstly, and then, from some value of the score, to use F_0.
- A suitable value of the score for this purpose would be the value of s_0, in which intersect the density functions of the scores, difference of distribution functions of the scores takes its maximum value and also the function F_{IV} becomes zero.



Point of intersection of densities

=

Point of maximal difference of CDFs

=

Point of zero value of f_{IV}

ESIS.2

- Algorithm for the modified ESIS:

1) $s_0 = mean\left(\arg\max_s \left|\hat{F}_1(s) - \hat{F}_0(s)\right|\right)$

it is necessary to solve the case when $\left|\hat{F}_1(s) - \hat{F}_0(s)\right|$ takes its maximal value for more than one value *s*.

2) $q_{1_j} = F_1^{-1}\left(\frac{j \cdot k}{n_1}\right),\ j = 1, \ldots, \left\lfloor \frac{n_1}{k} \cdot F_1(s_0) \right\rfloor$

3) $q_{0_j} = F_0^{-1}\left(\frac{j \cdot k}{n_0}\right),\ j = \left\lceil \frac{n_0}{k} \cdot F_0(s_0) \right\rceil, \ldots, \left\lfloor \frac{n_0}{k} \right\rfloor - 1$

4) $\mathbf{q} = [\min(s) - 1, \mathbf{q}_1, \mathbf{q}_0, \max(s) + 1]$

5) Merge intervals given by $\mathbf{q}_1$ where number of bads is less than *k*.

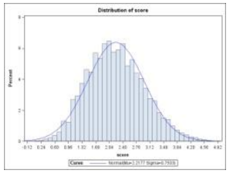
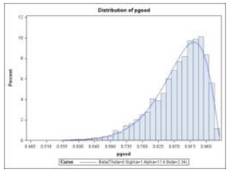
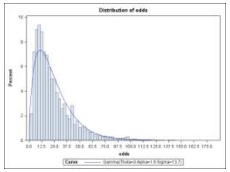
6) Merge intervals given by $\mathbf{q}_0$ where number of goods is less than *k*.

$\Rightarrow$ $\hat{I}_{val,ESIS2}$

where $k = \lceil \sqrt{n_0} \rceil$

Simulation results

- Consider n clients, $100p_B\%$ of bad and $100(1-p_B)\%$ of good clients with 1) normally, 2) beta and 3) gamma distributed scores:

1) $$f_0(x)=\frac{1}{\sqrt{2\pi}\sigma_0}\exp\left(\frac{-(x-\mu_0)^2}{2\sigma_0^2}\right) \qquad f_1(x)=\frac{1}{\sqrt{2\pi}\sigma_1}\exp\left(\frac{-(x-\mu_1)^2}{2\sigma_1^2}\right)$$

2) $$f_0(x)=\frac{1}{B(\alpha_0,\beta_0)}x^{\alpha_0-1}(1-x)^{\beta_0-1} \qquad f_1(x)=\frac{1}{B(\alpha_1,\beta_1)}x^{\alpha_1-1}(1-x)^{\beta_1-1}$$

3) $$f_0(x)=\frac{\lambda_0^{\alpha_0}}{\Gamma(\alpha_0)}x^{\alpha_0-1}\exp(-\lambda_0 x) \qquad f_1(x)=\frac{\lambda_1^{\alpha_1}}{\Gamma(\alpha_1)}x^{\alpha_1-1}\exp(-\lambda_1 x)$$

- The exact value is then

1) $$I_{val}=(A+1)D^2+A-1,\text{ where } A=\frac{1}{2}\left(\frac{\sigma_0^2}{\sigma_1^2}+\frac{\sigma_1^2}{\sigma_0^2}\right),\ D=\frac{\mu_1-\mu_0}{\sqrt{\sigma_0^2+\sigma_1^2}}$$

2) $$I_{val}=(\alpha_1-\alpha_0)(\psi(\alpha_1)-\psi(\alpha_0))+(\beta_1-\beta_0)(\psi(\beta_1)-\psi(\beta_0))+(\alpha_1-\alpha_0+\beta_1-\beta_0)(\psi(\alpha_0+\beta_0)-\psi(\alpha_1+\beta_1))$$

3) $$I_{val}=(\alpha_1-\alpha_0)\left(\psi(\alpha_1)-\psi(\alpha_0)+\ln\left(\frac{\lambda_0}{\lambda_1}\right)\right)+\alpha_0\left(\frac{\lambda_1}{\lambda_0}-1\right)+\alpha_1\left(\frac{\lambda_0}{\lambda_1}-1\right)$$

Gamma function $\Gamma(x)=\int_0^\infty t^{x-1}e^{-t}dt$ Beta function $B(x,y)=\frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$ Digamma function $\psi(x)=\frac{\Gamma'(x)}{\Gamma(x)}$

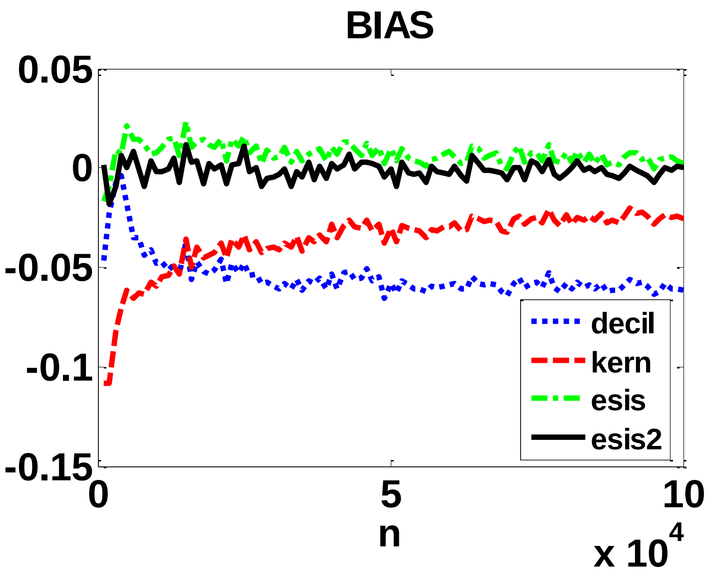
Simulation results

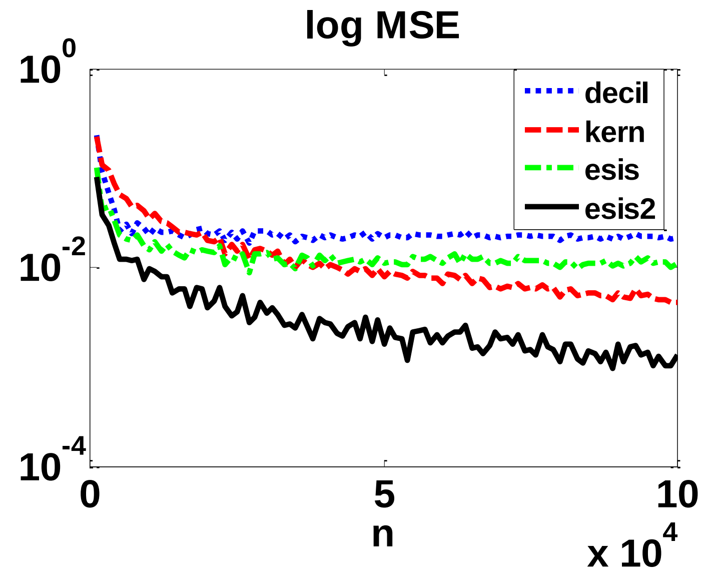
- Consider following values of parameters:
  - n = 1000 to 100 000
    - **Very small to large data set.**
  - $(\mu_0, \mu_1, \sigma_0, \sigma_1)$, $(\alpha_0, \alpha_1, \beta_0, \beta_1)$, $(\alpha_0, \alpha_1, \lambda_0, \lambda_1)$ that lead to I_{val} = 0.25, 1, 2.25
    - **Weak to very high performance of a model.**
  - p_B = 0.02, 0.05, 0.1, 0.2
    - **Portfolios with very low risk (mortgages) to very high risk (subprime cash loans).**

- Very common way how to assess quality of the Information value estimators is to compute bias and mean square error (MSE), or its logarithm.
  - $Bias = \hat{I}_{val} - I_{val}$
  - $\log MSE = \log\left(E(\hat{I}_{val} - I_{val})^2\right)$

Simulation results

- Bias and MSE for normally distributed scores:
  - p_B=0.1, I_{val}=1 for bias, p_B=0.2, I_{val}=2.25 for MSE

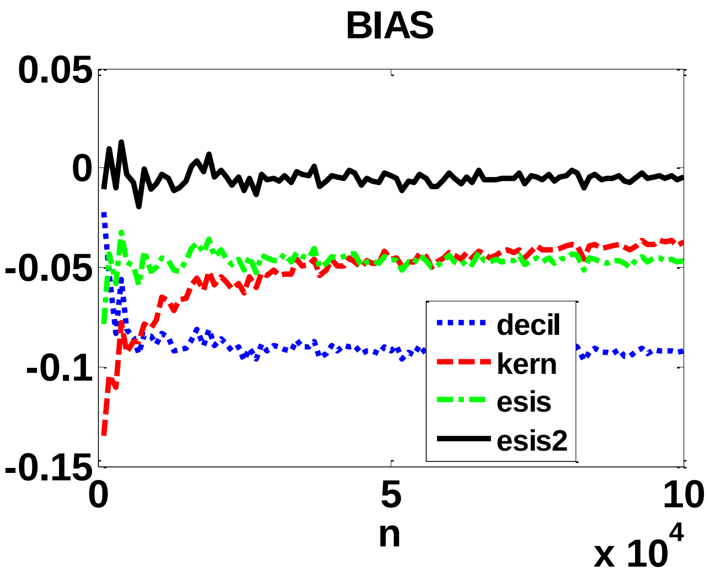


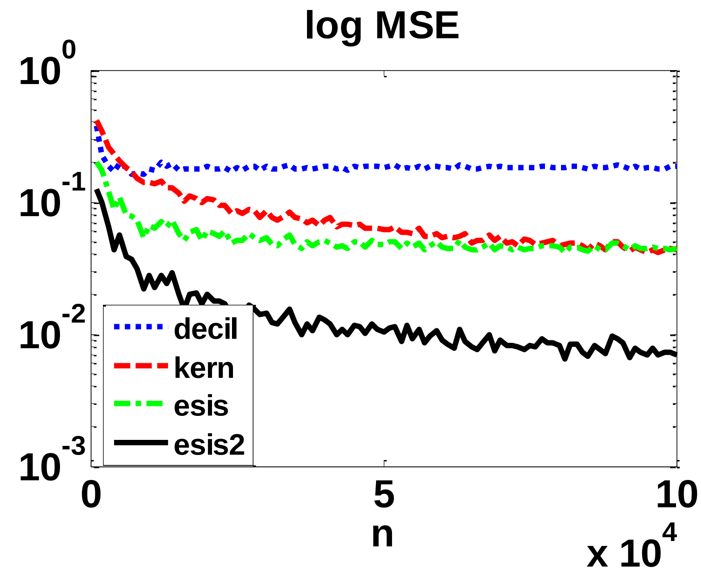


- Estimator using deciles is strongly biased – I_{val} was underestimated – for all considered values of parameters.
- Overall, the proposed algorithm ESIS2 had the best performance, followed by ESIS and kernel estimator.

Simulation results

- Bias and MSE for beta distributed scores:
  - p_B=0.2, I_{val}=1 for bias, p_B=0.1, I_{val}=2.25 for MSE

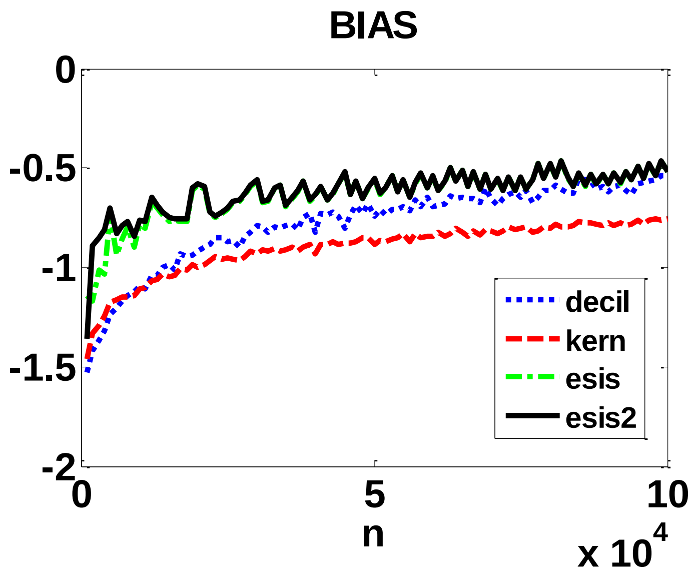


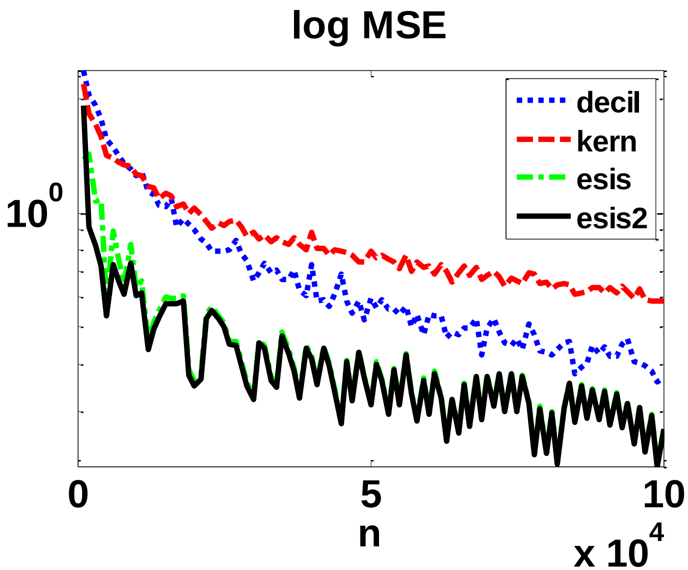


- Estimator using deciles is even worse than for normally distributed scores.
- The proposed algorithm ESIS2 had the best performance again.

Simulation results

- Bias and MSE for gamma distributed scores:
  - p_B=0.02, I_{val}=2.25 for bias, p_B=0.02, I_{val}=2.25 for MSE





- ESIS and ESIS2 had very similar performance considering both bias and MSE. However, all estimators were biased in case of gamma distributed scores – I_{val} was underestimated for all considered values of parameters.
- Overall, considering bias and MSE, ESIS and ESIS2 had the best performance.

Conclusions

- The classical way of computation of the Information value, i.e. empirical estimator using deciles of scores, is easy to implement, but may lead to strongly biased results. We conclude that kernel estimator and empirical estimators with supervised interval selection (ESIS) are much more appropriate to use. In total, the new algorithm ESIS2 outperformed all other considered estimators.

- The ESIS2 seems to be almost unbiased for normally and beta distributed scores. Its bias is negligible for all considered sizes of data samples, MSE tends to zero much more quickly compared to other estimators. However, it is biased in case of gamma distributed scores. Still, it is better than other considered estimators.

- Consequently, ESIS2 seems to be the right choice to estimate the Information value when assessing discriminatory power of credit scoring models. Moreover the Information value is very often used to assess the discriminatory power of variables that enter into these models. This means that ESIS2 may lead to more appropriate, compared to the empirical estimator using deciles, filter for variable selection.

Thank you for your attention