preamble

Ordinal rating models for financial evaluation

Paolo Giudici and Paola Cerchiello
University of Pavia,Italy
giudici@unipv.it

18 December

Outline

- Motivation;
- Integration of qualitative and quantitative variables;
- Non parametric indexes: MGI, SDI, QBI;
- Non parametric bayesian model;
- Semi-parametric bayesian model;
- Empirical Evidence;
- Conclusions and further research.

The Problem

Motivation

We want to classify companies in groups (i.e. rating classes) in a supervised way.

Such groups to comply with BASEL requirements, have to be:

- homogeneous with regard to target variable (i.e. default- not default);
- order preserving (i.e. ordering ability);
- stable with regard to horizon time.

The issue of different data sources

In this context we are typically provided with databases of various origin, often not transparent and made of qualitative and quantitative variables.

Our proposal is to build integrated, effective (easy to explain) ordinal rating models integrated by means of Bayes theorem.

Applications

- Institutional rankings (Universities)
- Company reputation
- Quality of services
- Financial evaluations

The Models

Model preview: A proposal

$$E(\theta_i \mid Data) = \sum_{k=1}^{K} E(\theta_j | g^k) \cdot p(g^k | Data)$$

where g^k is a partition induce by each variable to be combined that classify each unit i into one and only one level j

If a variable is quantitative (as it occurs when scorings are available) it must be discretized in level classes (ratings) using for example quantiles.

Measurement of variables

Variables must shown the same granularity for the final rating.

For continuous variables $\Rightarrow$ Easy just cut in quantiles

For ordinal qualitative variables $\Rightarrow$ Needs a method to quantify items maintaining their order. Method should be distribution free.

Non parametric indexes

- MGI
- SDI
- QBI

MGI and SDI

MGI: We recall that the nature of available data is ordinal thus we suggest to employ a simple scorecard model based on the median and the Gini Index (see Cerchiello et al., 2010)
SDI: on the basis of the cumulative distribution function, a summary index, that we name SDI (Stochastic dominance index) can be calculated as follows:

$$SDI = \sum_{i=1}^{K} F_i \tag{1}$$

Where F is the cumulative distribution function, and K the number of classes. A normalized version is obtained dividing it by its maximum value, J, which is attained when all data points are concentrated in the lowest class.

QBI

QBI: the drawback of MGI index is that it is based only on the median (besides the Gini index) and not on the other location measures.

$$QBI = \sum_{k=1}^{K} q_k + \left[1 - \frac{\sum_{k=1}^{K-1} (F_{(q_k)} - \frac{k}{K})}{0.5(K-1)} \right] \qquad (2)$$

or

$$QBI = SC + \left[1 - \frac{TEQ}{\max(TEQ)} \right] \qquad (3)$$

where K is the number of points of the measurement scale, $F_{(.)}$ is the cumulative distribution function, $\sum_{k=1}^{K} q_k$ (named SC) is the sum of the K points scale that contain the predefined quantiles, TEQ is the normalized sum of the total frequency excesses at each predefined quantile.

An example:MGI vs SDI vs QBI

	Q1	Q2	Q3	Q4	Q5
A	0	100	25	75	18.8
B	0	0	25	25	37.5
C	0	0	25	0	41.7
D	100	0	25	0	2.1
Total	100	100	100	100	100
MAX	100	100	25	75	41.7
MEDIAN	D	A	C	A	B
GINI INDEX G^*	0.0	0.0	1.0	0.5	0.867
GINI RATING	DDD	AAA	C	AA	B
QBI	17	4	10	5.16	7.8
SDI	4	1	2.5	3.75	2.92

Parametric Model

Marginal Likelihood distribution:

$$P(Y|g) = \left[\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\right]^J \prod_{j=1}^J \frac{\Gamma(\alpha+d_j)\Gamma(\beta+nd_j)}{\Gamma(\alpha+\beta+n_j)}$$

where J is the number of level of the covariate, d_j in the number of default in the j-th level, nd_j is the number of not default in the j-th level and n_j is the total number of observation in the j-th level.

$$E(\theta|Y) = \frac{\alpha}{\alpha+\beta}\frac{\alpha+\beta}{\alpha+\beta+n_j} + \frac{d_j}{n_j}\frac{n_j}{\alpha+\beta+n_j}$$

Non Parametric Model

$$P(Y|g) = \prod_{j=1}^{J} \frac{M^{d_j}}{M^{[d]}} \prod_{i=1}^{r} (n_{j(i)} - 1)! \times \left[\frac{\beta_j}{\alpha_j + \beta_j} I_{[0,1)}(x_j) + 1 I_{[1]}(x_j) \right]$$

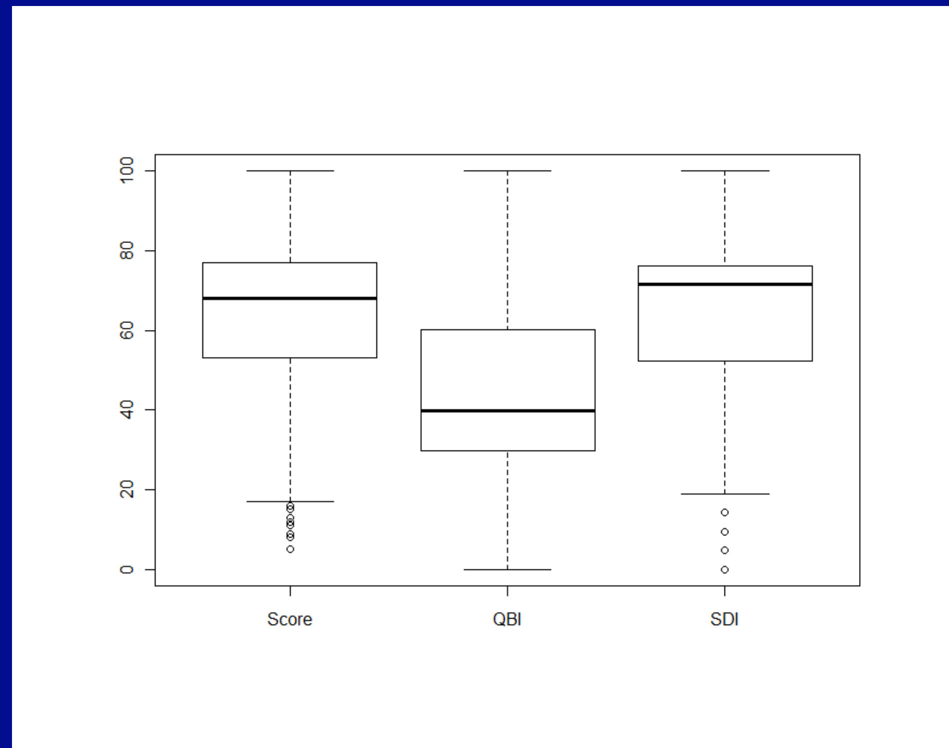
$$E(\theta|Y) = \frac{M}{M+n_j} \left[\theta I_{[0,1)}(x_j) + 1 I_{[1]}(x_j) \right] + \frac{n_j}{M+n_j} \hat{F}$$

The Empirical Evidence

Results:Parametric vs Non parametric model

Models	Parametric	Non parametric
AI	3.64e-21	1.10e+15
Cr	1.41e-23	3.83e+12
Dir	3.28e-15	2.07e+14
Cebi	1.02e-14	5.50e+22
Score	2.78e-15	1.91e+14
SDI	1.19e-26	1.73e+14
QBI	2.02e-28	3.19e+13

Comparison of quantifications



Reference

- Outline

The Problem

The Models

The Empirical Evidence

- Results:Parametric vs Non parametric model
- Comparison of quantifications
- **Reference**