Ranking scientific journals via latent class models for polytomous item response data

Francesco Bartolucci, University of Perugia
Valentino Dardanoni, University of Palermo
Franco Peracchi, Tor Vergata University and EIEF

EIEF, Rome
November 29, 2012

Outline

Motivation

- Growing interest in issues related to the classification of scientific journals for evaluating research institutions or individual researchers
- Evaluation systems partially based on journal rankings have been recently introduced in various countries, such as Australia and France, and recently in Italy in connection with the Evaluation of the Quality of Research (VQR) 2004-2010
- There are by now many indicators which allow for a ranking of scientific journals, such as Impact Factor (IF), 5-year Impact Factor (IF5), Article Influence Score (AIS), and the h-index, which may be derived using commonly available databases (e.g., ISI-Thomson-Reuters, Scopus, Google Scholar)
- There is little agreement on whether there is a single best general indicator, and what this indicator is

What we do

- We propose a strategy for obtaining a complete ordering of scientific journals using a set of quantitative/qualitative indicators of the value of the journals in a chosen list
- Main features of the proposed strategy:
  - (*i*) may be simply implemented on the basis of a meaningful statistical model
  - (*ii*) is able to produce a complete ordering of the journals
  - (*iii*) provides a measure of the reliability of each indicator for classifying the journals in the chosen list
  - (*iv*) may be also applied with partially missing information
- Current approaches for reducing a set of journal value indicators into a single ranking are based on Principal Component Analysis or some type of average of the rankings induced by the different indicators (e.g., RePEc), but they do not assume a statistical model
- We apply our approach to data from the VQR 2004-2010

Our strategy

- Let n be the number of journals to rank, r the number of indicators on which the ranking is to be based, and x_{ij} the value of indicator j for journal i (some x_{ij} may be missing)
- We first discretize the x_{ij} and then apply a statistical model for polytomous Item Response Data (Hambleton & Swaminathan, 1985)
- We let $q_{j1}, \ldots, q_{j,s-1}$ be a set of cutoffs for the jth indicator x_{ij} (e.g., its quartiles or deciles) and define
$$y_{ij} = \sum_{m=0}^{s-1} m \cdot 1\{q_{jm} < x_{ij} \leq q_{j,m+1}\}, \quad i = 1, \ldots, n, j = 1, \ldots, r,$$
with $q_{j0} = -\infty$, $q_{js} = \infty$
- If the value of x_{ij} is missing for some i and j, then the value of y_{ij} is also missing
- Since discretization is essentially arbitrary, it is important to assess the sensitivity of the results

Model assumptions

- Discretizing allows us to use existing and easily interpretable models (in which the "scientific value" of a journal is considered as a latent trait) and it also offers some robustness to measurement errors
- Our model is based on the three assumptions:

A.1 For every journal i, $y_{i1}, \dots, y_{ir}$ are conditionally independent given a latent variable u_i ("scientific value" of the journal)

A.2 The conditional distribution of every y_{ij} given u_i satisfies

$$\log \frac{p(y_{ij} \geq m | u_i)}{p(y_{ij} < m | u_i)} = \alpha_j (u_i - \beta_{jm}), \quad m = 1, \dots, s-1,$$

as in the Graded Response Model (Samejima, 1969)

A.3 The latent variables $u_1, \dots, u_n$ are independent and have the same discrete distribution with k support points $\xi_1, \dots, \xi_k$ and corresponding probabilities $\pi_1, \dots, \pi_k$, with $\pi_h = p(u_i = \xi_h)$

Discussion of model assumptions I

- A.1 (*local independence*) allows us to interpret the latent variable u_i as the "intrinsic" scientific value of a journal: if we knew the value of u_i for the ith sample unit, then knowing the value of one indicator would not be useful to predict the value of any other indicator

- A.2 formalizes our interpretation of the latent variable u_i: if $\alpha_j > 0$ (discriminant index), then the distribution of y_{ij} stochastically increases with u_i, meaning that the probability distribution of y_{ij} moves its mass towards higher categories

- The discretization rule relating x_{ij} with y_{ij}, combined with the linear model

$$x_{ij} = \gamma_j + \delta_j u_i + \varepsilon_{ij},$$

  implies A.2 with $\alpha_j = \delta_j$ and $\beta_{jm} = (q_{jm} - \gamma_j)/\delta_j$, provided that ε_{ij} is a zero-mean random variable with a logistic distribution

Discussion of model assumptions II

- A.3 avoids the formulation of a parametric distribution for the latent variable and then our model is semiparametric in nature (Lindsay *et al.*, 1991)
- A.2 & A.3 imply that a journal's latent value is unidimensional, which is useful to obtain a unique ranking of journals; unidimensionality of the latent value may however be tested against multidimensionality
- A.3 further implies that the latent variables $u_1, \ldots, u_n$ are independent, so the response vectors $\mathbf{y}_i = (y_{i1}, \ldots, y_{ir})$, $i = 1, \ldots, n$, are also independent; this assumption may be restrictive in some cases, for example when the discretized outcomes y_{ij} are constructed using as cutoffs the sample quantiles; we expect that minor failures of this assumption should not significantly affect the results, especially when the sample size n is large

The statistical model

- The model parameters are:
  - support points ξ_h, $h = 1, \ldots, k$
  - corresponding probabilities π_h, $h = 1, \ldots, k$
  - discriminant indices α_j, $j = 1, \ldots, r$
  - cutoffs β_{jm}, $j = 1, \ldots, r$, $m = 1, \ldots, s-1$
- Since $\sum_{h=1}^{k} \pi_h = 1$ and due to the identifiability constraints $\alpha_1 = 0$ and $\beta_{11} = 1$, the number of free parameters is:
  $\#\mathrm{par}_k = k + (k-1) + (r-1) + [r(s-1) - 1] = 2k + rs - 3$
- An alternative identifiability constraint is that the latent distribution is standardized to have zero mean and unit variance ($\sum_{h=1}^{k} \xi_h \pi_h = 0$, $\sum_{h=1}^{k} \xi_h^2 \pi_h = 0$)
- ML estimation is based on the manifest distribution of $\mathbf{y}_i$:

$$p(\mathbf{y}_i) = \sum_{h=1}^{k} \pi_h \prod_{j=1}^{r} p(y_{ij} | u_i = \xi_h)$$

Prediction

- The posterior distribution of u_i has probability mass function

$$p(u_i = \xi_h|\mathbf{y}_i) = \frac{\pi_h \prod_{j=1}^{r} p(y_{ij}|u_i = \xi_h)}{p(\mathbf{y}_i)}$$

- The posterior distribution is used to assign every sample unit to a given group (latent class); once the model has been estimated, unit i is assigned to group h if

$$h = \underset{g=1,\dots,k}{\operatorname{argmax}}\, p(u_i = \xi_g|\mathbf{y}_i)$$

- The value of u_i may be predicted using the posterior mean

$$\hat{u}_i = \sum_{h=1}^{k} \xi_h p(u_i = \xi_h|\mathbf{y}_i)$$

Missing data

- With missing data, we compute the manifest distribution of the vector of observed outcome as

$$p(\mathbf{y}_i) = \sum_{h=1}^{k} \pi_h \prod_{j=1\ (m_{ij}=0)}^{r} p(y_{ij}|u_i = \xi_h),$$

where m_{ij} is equal to 1 if y_{ij} is missing and to 0 if it is observed

- This amounts to assuming that the data are Missing-at-Random (MAR); in our context, MAR implies that the event that the value of an indicator is missing may be predicted by the observable indicators (i.e., the h-index which we observe for all journals)

- MAR is rather realistic since missing values of certain indicators tend to exist for journals with a lower reputation and a lower level of the h-index

ML estimation

- Given observations on a set of n journals, consisting of the discrete outcomes y_{ij}, $i = 1, \ldots, n$, $j = 1, \ldots, r$, the log-likelihood is

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n} \log p(\mathbf{y}_i),$$

  where $\boldsymbol{\theta}$ is the vector of all model parameters and $p(\mathbf{y}_i)$ is the manifest probability of the observed response vector $\mathbf{y}_i$

- To maximize $\ell(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$, we use a version of the Expectation-Maximization (EM) algorithm (Dempster *et al.*, 1977) implemented in the `R` package `MultiLCIRT` (Bartolucci *et al.*, 2012)

- The EM algorithm is based on the *complete data log-likelihood*, that is, the log-likelihood that we could compute if we knew the value of u_i for every sample unit i

EM algorithm

- The complete data log-likelihood has expression

$$\ell^*(\boldsymbol{\theta}) = \sum_{h=1}^{k}\sum_{i=1}^{n} z_{hi} \log\left[\pi_h \prod_{j=1}^{r} p(y_{ij}|u_i = \xi_h)\right],$$

where z_{hi} is an (unobserved) indicator equal to 1 if $u_i = \xi_h$ and to 0 otherwise

- The algorithm alternates two steps until convergence in $\ell(\boldsymbol{\theta})$:

  E-step: compute the conditional expected value of $\ell^*(\boldsymbol{\theta})$ given the observed data and the current value of the parameters (this amount to computing the posterior probabilities $p(u_i = \xi_h|\mathbf{y}_i)$)

  M-step: maximize the above expected value with respect to $\boldsymbol{\theta}$ to get an updated estimate of the parameter vector

Choice of the number of latent classes (k)

- In applying the model we need to select the number of support points (or latent classes) of the distribution of u_i, denoted by k, when this number is not *a priori* fixed

- We use the Bayesian Information Criterion (BIC, Schwarz, 1978), which is based on the minimization of the index

$$BIC_k = -2\ell(\hat{\boldsymbol{\theta}}_k) + \log(n)\,\#\mathrm{par}_k,$$

  where $\hat{\boldsymbol{\theta}}_k$ is the ML estimate of $\boldsymbol{\theta}$ under the model with k latent classes

- The quality of the clustering may be measured by the entropy index

$$E_k = -\sum_{i=1}^{n}\sum_{h=1}^{k} \hat{p}(u_i = \xi_h|\mathbf{y}_i)\log[\hat{p}(u_i = \xi_h|\mathbf{y}_i)]$$

  which is alway between $n\log(k)$ and 0 (perfect clustering)

Data

- We consider the list of journals for the VQR sub-area Statistics and Applied Mathematics of GEV13
- The list was created starting from all journals in the Thomson Reuters Web of Science (WoS) that belong to the core subject categories for GEV13, plus a number of journals that belong to subject categories which are considered relevant
- It was expanded using the U-Gov list of journals in which at least one Italian researcher belonging to the area published in 2004–2010
- To avoid different rankings across sub-areas, each journal in the list has been assigned to one and only one of the 4 sub-areas covered by GEV13 (Business; Economics; Economic History; Statistics and Applied Mathematics)
- The list for the sub-area Statistics and Applied Mathematics contains $n = 445$ journals, excluding a very small set of journals (e.g., *Econometrica*) attributed to other sub-areas

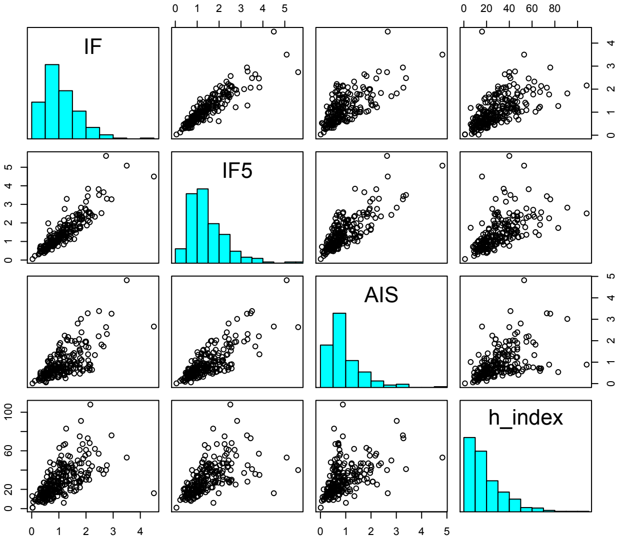
Definition of the indices

- For each journal, the list includes $r = 4$ indicators:
  - IF (2010): citations in 2010 to articles published in 2008-2009 divided by the number of articles published in 2008-2009
  - IF5 (2010): citations in 2010 to articles published in 2005-2009 divided by the number of articles published in 2005-2009
  - AIS (2010): it uses the same information as IF5, but it also considers which journals have contributed to these citations (highly cited journals influence the indicator more than lesser cited journals) and remove journal self-citations
  - h-index (collected in April 2012 for the period 2004-2010): a journal has h-index h if h published articles in that journal have at least h citations and the other articles have no more than h citations each

Descriptive statistics

		IF	IF5	AIS	h-index
Missing values	(%)	43.8	52.6	52.6	0.0
Mean		1.056	1.472	0.946	19.766
Variance		0.418	0.751	0.480	267.586
Skewness index		1.325	1.526	1.938	1.575
Quartile	1st	.586	.840	.506	7.0
	2nd	.954	1.284	.721	14.0
	3rd	1.381	1.867	1.203	28.0
Decile	1st	.370	.590	.313	4.0
	2nd	.521	.766	.454	6.0
	3rd	.643	.967	.553	9.0
	4th	.754	1.108	.660	12.0
	5th	.954	1.284	.721	14.0
	6th	1.088	1.467	.871	19.0
	7th	1.257	1.741	1.026	24.0
	8th	1.561	2.132	1.362	32.0
	9th	1.906	2.513	1.892	42.6

Scatterplot



Correlation matrix

	IF	IF5	AIS	*h*-index
IF	1.0000	0.8990	0.6925	0.5561
IF5	0.8990	1.0000	0.7950	0.5792
AIS	0.6925	0.7950	1.0000	0.4850
h-index	0.5561	0.5792	0.4850	1.0000

h-index distribution for ISI and non-ISI journals

Journals	#	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
non-ISI	195	3	4.000	7.000	8.446	12.000	26
new-ISI[1]	39	1	9.00	14.00	14.82	17.50	46
old-ISI[2]	211	1	19.00	28.00	31.14	40.00	108

- The great difference between the distribution of IF, IF5, and AIS for ISI and non-ISI journals offers the basis for the MAR assumption

[1] only IF is available
[2] also IF5 & AIS are available

Model fitting

- The first step of our journal ranking strategy consists of discretizing the observed values of the indicators
- We present the results based on two alternative discretizations: one uses as cutoffs the sample quartiles ($s = 4$), the other uses the sample deciles ($s = 10$)
- Given the discretized outcomes y_{ij}, we fit our model for increasing values of k (we increase the value of k until the BIC does not become smaller than that computed for the previous value of k)
- To prevent local maxima of the sample log-likelihood, following the current literature on latent class and finite mixture models we use two types of initialization (deterministic and random) of the EM algorithm

Preliminary model fit (with increasing values of k)

	$s = 4$ (quartiles)			$s = 10$ (deciles)		
k	$\ell(\hat{\boldsymbol{\theta}}_k)$	$\#\mathrm{par}_k$	BIC_k	$\ell(\hat{\boldsymbol{\theta}}_k)$	$\#\mathrm{par}_k$	BIC_k
1	-1544.9	12	3163.0	-2564.3	36	5348.0
2	-1343.8	17	2791.2	-2347.7	41	4945.4
3	-1293.0	19	2702.0	-2271.2	43	4804.6
4	-1273.6	21	**2675.2**	-2233.2	45	4740.7
5	-1271.0	23	2682.3	-2216.8	47	4720.2
6				-2206.5	49	4711.9
7				-2197.2	51	**4705.5**
8				-2194.7	53	4712.7

- with $s = 4$ categories BIC, leads to selecting $k = 4$ latent classes ($E_k = 135.714$, $\max E_k = 616.901$, 22.0%)
- with $10 = 4$ categories BIC, leads to selecting $k = 7$ latent classes ($E_k = 213.481$, $\max E_k = 1024.65$, 20.8%)

Estimated distribution of the latent variable

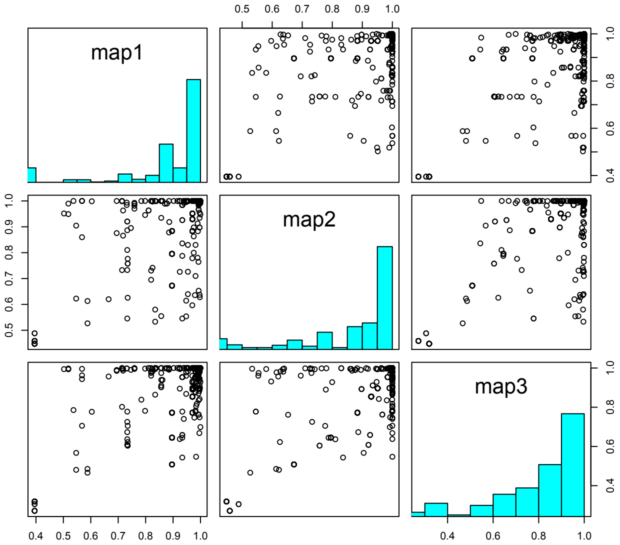
	$s = 4,\ k = 4$ (1)		$s = 10,\ k = 4$ (2)		$s = 10,\ k = 7$ (3)	
h	$\hat{\xi}_h$	$\hat{\pi}_h$	$\hat{\xi}_h$	$\hat{\pi}_h$	$\hat{\xi}_h$	$\hat{\pi}_h$
1	-0.840	.537	-0.924	.478	-0.997	.401
2	0.288	.178	.169	.216	-0.274	.156
3	1.092	.182	1.015	.194	.327	.123
4	1.941	.104	1.851	.113	.785	.123
5					1.229	.098
6					1.621	.058
7					2.208	.041

- The identifiability constraint is that the distribution of the latent trait is standardized

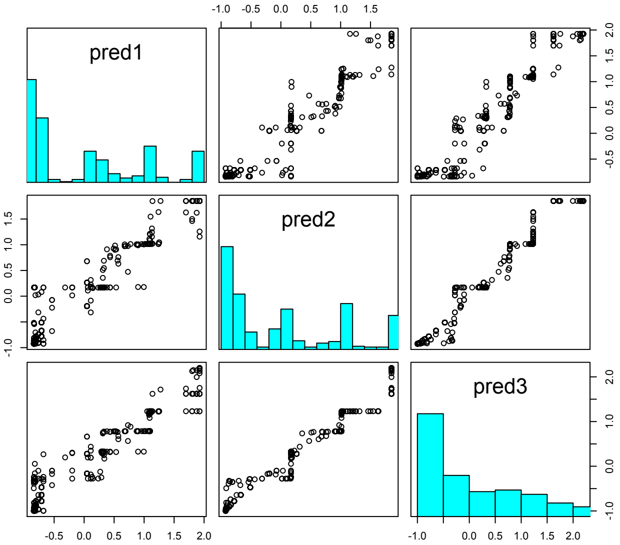
Comments

- When $k = 4$, the estimated distributions of the latent variable are rather similar using quartiles or deciles
- Four ordered groups of journals are found and their size is about equal to 50% for the first group, 20% for the second and third groups, and 10% for the last group (the best journals)
- These percentages are close to the ones suggested by the Italian VQR except that, in the VQR, the class of best journals is expected to have size 20% and the class of medium level journals (the third class) is expected to have size 10%
- The distribution when $s = 10$ and $k = 7$ is not directly comparable with the other ones and thus we compare the different discretizations by contrasting the corresponding predicted values of u_i based on the posterior probabilities $\hat{p}(u_i = \xi_h|\mathbf{y}_i)$

Scatterplot of maximum posterior probabilities (map)



Scatterplot of predicted values (pred)



Correlation of the predicted values of the latent variable

	Pearson			Spearman		
	pred1	pred2	pred3	pred1	pred2	pred3
pred1	1.000	.974	.968	1.000	.944	.928
pred2	.974	1.000	.985	.944	1.000	.985
pred3	.968	.985	1.000	.928	.985	1.000

- The results suggest that, apart from rescaling, the predicted values of the latent variables are very similar and provide a very similar ranking of the journals

Cross-classification (with Cohen's κ index)

- Comparison of the classification of the journals under the different models and with that provided by GEV13 (GEV13 adopts a classification based on four groups of journals of size 216, 36, 81, and 112)

		$s = 10,\ k = 4$				$s = 10,\ k = 7$				GEV13			
		c1	c2	c3	c4	c1	c2	c3	c4	c1	c2	c3	c4
$s = 4$	c1	208	1	7	0	203	10	3	0	189	17	10	0
	c2	8	22	6	0	13	16	7	0	27	5	4	0
	c3	0	13	65	3	0	10	67	4	0	14	54	13
	c4	0	0	3	109	0	0	4	108	0	0	13	99
$s = 10$	c1					211	5	0	0	197	13	6	0
$k = 4$	c2					5	21	10	0	19	12	5	0
	c3					0	10	70	1	0	11	56	14
	c4					0	0	1	111	0	0	14	98
$s = 10$	c1									200	12	4	0
$k = 7$	c2									14	11	10	1
	c3									2	13	52	14
	c4									0	0	15	97

		$s=10,\ k=4$	$s=10,\ k=7$	GEV13
$s=4$	unweighted κ	0.8607	0.8267	0.6670
		(0.0207)	(0.0228)	(0.0297)
	linear weights κ	0.9211	0.9113	0.8225
		(0.0126)	(0.0123)	(0.0173)
	quadratic weights κ	0.9574	0.9587	0.9120
		(0.0609)	(0.0614)	(0.0619)
$s=10$	unweighted κ		0.8913	0.7214
$k=4$			(0.0185)	(0.0278)
	linear weights κ		0.9474	0.8554
			(0.0091)	(0.0153)
	quadratic weights κ		0.9780	0.9312
			(0.0609)	(0.0618)
$s=10$	unweighted κ			0.7112
$k=7$				(0.0282)
	linear weights κ			0.8488
				(0.0157)
	quadratic weights κ			0.9271
				(0.0615)

Comments

- Strong agreement between the classifications of the journals obtained under different values of s and k
- In particular, the percentage of journals that change classification ranges from 7.9% (comparison between $k = 4$ and $k = 7$, with $s = 10$ in both cases) to 11.5% (comparison between $s = 4$ with $k = 4$ and $s = 10$ with $k = 7$); this is confirmed by the values of the Cohen's κ indices
- As for the comparison between these classifications and that set up by GEV13, the percentage of disagreement is somewhat higher and ranges from 18.4% (comparison with $s = 10$ and $k = 4$) to 22.0% (comparison with $s = 4$ and $k = 4$)
- The classification produced by GEV13 does not use IF among the indicators, and uses the h-index alone as a predictor of IF5 and AIS when these indicators are missing; the size of each class is chosen to be in agreement with the VQR rules

Discriminant indices

- What is the quality of an indicator as a measure of the latent scientific value of a journal? We can answer this question by comparing the estimated discriminant indices ($\hat{\alpha}_j$)

j	$s = 4$	$s = 10$ $(k = 4)$	$s = 10$ $(k = 7)$
1 IF	3.772	4.194	5.150
2 IF5	6.740	12.645	39.424
3 AIS	2.103	2.199	2.438
4 h-index	2.626	2.251	2.264

- The identifiability constraint is that the distribution of the latent trait is standardized

- As frequently happens with these models, the estimate of some discriminant index assumes extreme values; in fact, the distribution of the estimator is known to be highly skewed

- Confidence intervals (obtained with a parametric bootstrap) for the discriminant indices (case $s = 4,\ k = 4$)

j	est.	95%-interval	
1 IF	3.772	3.140	5.921
2 IF5	6.740	5.156	88.164
3 AIS	2.103	1.449	2.489
4 h-index	2.626	2.090	3.256

- It is also possible to obtain confidence intervals for the ratios $\alpha_{j_2}/\alpha_{j_1}$ which may be used to compare different bibliometric indicators (case $s = 4,\ k = 4$)

j_1	j_2	comparison	est.	95%-interval	
1	2	IF5 vs. IF	1.787*	1.089	23.885
1	3	AIS vs. IF	0.558*	0.340	0.660
1	4	h-index vs. IF	0.696*	0.414	0.938
2	3	AIS vs. IF5	0.312*	0.018	0.394
2	4	h-index vs. IF5	0.390*	0.031	0.498
3	4	h-index vs. AIS	1.248	0.928	1.803

- The null hypothesis $H_0 : \alpha_{j_2}/\alpha_{j_1} = 1$ is that indicators j_1 and j_2 have the same discriminant power and may be tested by checking if 1 is contained in the confidence interval
- IF5 has a significantly higher discriminant power than other bibliometric indicators; AIS and h-index have not a significantly different discriminant power

Comparison with mixture model for non-discretized data

- The proposed approach is based on discretizing the observations x_{ij} which are transformed in the categorical responses y_{ij}
- It is possible to explain the assumption about the conditional distribution of every y_{ij} given the latent variable u_i, which is formulated through cumulative logits, on the basis of a linear model for x_{ij}
- It is natural to compare the proposed approach with an approach in which a finite mixture model is assumed directly for the distribution of the outcomes x_{ij}

Comparison with mixture model for continuous data

- We consider a finite mixture model which is the counterpart for continuous data of the assumed model; it assumes that:

  A.1 For every journal i, the latent variables $x_{i1}, \dots, x_{ir}$ are conditionally independent given u_i

  A.2 Given u_i, each variable x_{ij} satisfies the linear model

  $$x_{ij} = \gamma_j + \delta_j u_i + \sigma_j \varepsilon_{ij}, \quad i = 1, \dots, n,\ j = 1, \dots, r,$$

  where ε_{ij} are independent error terms with distribution $N(0, 1)$

  A.3 The latent variables $u_1, \dots, u_n$ are independent and have the same discrete distribution with k support points $\xi_1, \dots, \xi_k$ and corresponding probabilities $\pi_1, \dots, \pi_k$

- This is a particular finite mixture model of normal distributions which may be fitted by an EM algorithm similar to that used for the initial model for the categorical outcomes

Preliminary model fit (with increasing values of k)

- The model is applied to the original data (without any transformation of the data) and to the transformed data (log-transformation for the skewness)

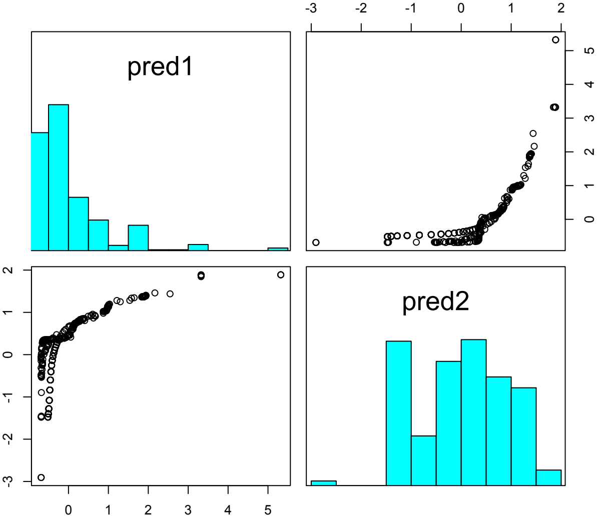
	original scale			log-transformed data		
k	$\ell(\hat{\boldsymbol{\theta}}_k)$	$\#\text{par}_k$	BIC_k	$\ell(\hat{\boldsymbol{\theta}}_k)$	$\#\text{par}_k$	BIC_k
1	-2612.2	8	5273.1	-1394.6	8	2838.0
2	-2396.9	13	4873.2	-1120.6	13	2320.4
3	-2309.5	15	4710.5	-1019.3	15	2130.1
4	-2268.0	17	4639.6	-942.7	17	1989.1
5	-2237.5	19	4590.8	-909.5	19	1934.9
6	-2217.1	21	**4562.3**	-880.8	21	1889.7
7	-2211.5	23	4563.3	-857.9	23	1856.1
8				-842.0	25	1836.4
9				-835.0	27	**1834.7**
10				-830.9	29	1838.7

Estimated distribution of the latent variable

- BIC leads to selecting $k = 6$ components with the original data and $k = 9$ components with the log-transformed data

	original (1)		log-transformed (2)	
h	$\hat{\xi}_h$	$\hat{\pi}_h$	$\hat{\xi}_h$	$\hat{\pi}_h$
1	-0.687	0.539	-2.906	0.008
2	0.149	0.259	-1.471	0.217
3	0.961	0.109	-0.504	0.113
4	1.939	0.069	-0.063	0.170
5	3.323	0.017	0.358	0.156
6	5.321	0.007	0.730	0.141
7			1.049	0.095
8			1.396	0.078
9			1.889	0.023

Scatterplot of predicted values



Agreement of the predicted values of the latent variable

	Pearson		Spearman	
	pred1	pred2	pred1	pred2
pred1	1.000	0.723	1.000	0.810
pred2	0.723	1.000	0.810	1.000

		log-transformed				GEV13			
		c1	c2	c3	c4	c1	c2	c3	c4
original	c1	187	14	15	0	173	22	20	1
	c2	28	0	8	0	28	1	6	1
	c3	1	22	57	1	15	13	40	13
	c4	0	0	1	111	0	0	15	97
log-transformed	c1					196	14	6	0
	c2					20	11	5	0
	c3					0	11	55	15
	c4					0	0	15	97

		log-transformed	GEV13
original	unweighted κ	0.6942	0.5447
		(0.0288)	(0.0329)
	linear weights κ	0.8258	0.7173
		(0.0684)	(0.0671)
	quadratic weights κ	0.9051	0.8281
		(0.0613)	(0.0614)
log-transformed	unweighted κ		0.7078
			(0.0283)
	linear weights κ		0.8488
			(0.0690)
	quadratic weights κ		0.9285
			(0.0619)

- The results are not particularly robust with respect to transformations of the response outcomes

Conclusions

- The main advantage of our approach is that it relies on a statistical model that has some nonparametric features
- The mean of the posterior distribution of the latent variable provides a prediction on a continuous scale of the latent value of each journal in the given list, so journals can be univocally ordered and the distance between any pair of journals can be evaluated
- It is possible to classify journals in any arbitrary number of classes of a given size
- We can assess the discriminant power of each indicator; in the data we analyze, IF5 appears to be the most reliable indicator of the value of a journal among the available indicators
- Because our approach is based on a discretization of quantitative indicators, results may be sensitive to the choice of cutoffs, so sensitivity analysis is important; our results appear to be relatively robust

Further developments

- Use of a different number of categories for different indicators (so as to also include, for instance, binary indicators) and use of discriminant indices which are category-dependent
- Investigation of the reliability of the estimates of the discriminant indices, given the high skewness in the bootstrap samples
- Application to other lists of journals in different fields
- Deeper comparison with the approach based on the normal mixture model for the original outcomes x_{ij} which uses less parameters
- Our preliminary results suggest that the normal mixture approach is rather sensitive to the parametric assumptions and this is in agreement with studies (Shentu & Xie, 2010) according to which discretizing continuous observations may increase the robustness of the model with respect to model misspecifications and contaminations

References

- Bartolucci, F., Bacci, S., and Gnaldi, M. (2012), MultiLCIRT: Multidimensional latent class Item Response Theory models. R package, URL http://CRAN.R-project.org/package=MultiLCIRT.
- Dempster, A.P., Laird, N.M., and Rubin, D. B. (1977), Maximum likelihood from incomplete data via the EM algorithm (with discussion), *J. of the Royal Statistical Society, Series B*, **39**, 138.
- Hambleton, R.K. and Swaminathan, H. (1985), *Item Response Theory: Principles and Applications*, Kluwer Nijhoff, Boston.
- Lindsay, B., Clogg, C., and Greco, J. (1991), Semiparametric estimation in the Rasch model and related exponential response models, including a simple latent class model for item analysis, *J. of the American Statistical Association*, **86**, 96-107.
- Samejima, F. (1969), Estimation of ability using a response pattern of graded scores, *Psychometrika Monograph*, **17**.
- Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, **6**, 461-464.
- Shentu, Y. and Xiem M. (2010), A note on dichotomization of continuous response variable in the presence of contamination and model misspecification, *Statistics in Medicine*, **29**, 2200-2214.