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

Francesco Bartolucci,

*University of Perugia, Italy*

Valentino Dardanoni

*University of Palermo, Italy*

and Franco Peracchi

*University of Rome Tor Vergata and Einaudi Institute for Economics and Finance, Rome, Italy*

[Received November 2013. Final revision November 2014]

**Summary.** We propose a model-based strategy for ranking scientific journals starting from a set of observed bibliometric indicators that represent imperfect measures of the unobserved ‘value’ of a journal. After discretizing the available indicators, we estimate an extended latent class model for polytomous item response data and use the estimated model to cluster journals. We illustrate our approach by using the data from the Italian research evaluation exercise that was carried out for the period 2004–2010, focusing on the set of journals that are considered relevant for the subarea statistics and financial mathematics. Using four bibliometric indicators (IF, IF5, AIS and the $h$-index), some of which are not available for all journals, and the information contained in a set of covariates, we derive a complete ordering of these journals. We show that the methodology proposed is relatively simple to implement, even when the aim is to cluster journals into a small number of ordered groups of a fixed size. We also analyse the robustness of the obtained ranking with respect to different discretization rules.

**Keywords**: Clustering; Finite mixture models; Graded response model; Item response models; Research evaluation; Valutazione della Qualità delle Ricerca

1. **Introduction**

There is growing interest in issues surrounding the classification of scientific journals as an intermediate step towards evaluating research institutions or individual researchers. In fact, evaluation systems partially based on journal rankings have recently been introduced in various countries, e.g. in Australia by the Australian Research Council, in France by the Agence d’Évaluation de la Recherche et de l’Enseignement Supérieur and in Italy by the Agenzia di Valutazione del Sistema Universitario e della Ricerca. Since many of these evaluation systems aim at clustering journals into merit classes, the problem arises of how many classes should be considered and how they should be constructed given the available information.

A large number of bibliometric indicators is now available, typically derived from either citation or usage log data, and any of them can in principle be employed to rank journals. Recently
Zimmermann (2012) considered seven indicators, five of which—four different versions of the impact factor (see, for example, Garfield (2006)) plus the h-index (Hirsch, 2005)—are based on alternative transformations of the number of citations that a journal receives, and the other two are based on the number of abstract views and downloads. Chang et al. (2012) considered a larger set of 15 indicators for 299 leading international journals in economics. Bollen et al. (2009) considered an even larger set of 39 indicators that include traditional measures based on citation counts and other measures based on social network analysis and usage log data. Many of these bibliometric indicators are now routinely computed in commonly available citation databases such as the ISI–Thomson Reuters Web of Science, SciVerse Scopus, Google Scholar or Microsoft Academic Search, which helps to explain their widespread use.

Although rankings of journals generally differ depending on what indicator or database is used, there is disagreement on whether there is a single best general indicator or, more generally, on how to combine the information that is contained in the available indicators to rank journals. One approach is principal component analysis, which aims at extracting from the data the orthogonal components that account for most of the observed correlation between the indicators considered (see, for example, Bollen et al. (2009)). The problem with principal component analysis is that interpretation is not simple, especially when more than one component is needed to account for a significant fraction of total variance, so the association of the components with elusive concepts such as ‘value’, ‘popularity’ or ‘prestige’ of a journal remains rather speculative. An alternative approach is to take some kind of average of the rankings induced by the different indicators. For example, the ‘Research papers in economics’ ranking of economic journals (which is available from https://ideas.repec.org/top/top.journals.all.html) employs the harmonic mean of ranks after dropping the best and worst values. An important drawback of both approaches is that they are based neither on an explicit model for measuring intellectual influence (Palacios-Huerta and Volij, 2004) nor on a well-defined statistical model, which makes it difficult to interpret the results that are obtained and to carry out inference.

In this paper we propose a model-based strategy for ranking scientific journals starting from a set of observed bibliometric indicators that represent imperfect measures of the unobserved scholarly influence of a journal, treated initially as a unidimensional latent trait. After discretizing these indicators to avoid strong parametric assumptions, we estimate several latent class models for polytomous item response data and use them to cluster journals. Our models are latent class versions of the graded response model (Samejima, 1969, 1996), which is commonly used in educational research and is one of the most popular item response theory (IRT) models (Hambleton and Swaminathan, 1985). Relative to other approaches, our strategy offers several advantages. First, it is based on a well-defined statistical model and is easy to implement. Second, it produces a complete ordering of the journals. Third, it is semiparametric in nature, since it requires no parametric assumption on the distribution of the latent traits. Fourth, it provides a measure of the reliability of each indicator for classifying or clustering journals. Finally, it can be applied even when the bibliometric indicators contain missing values.

We illustrate our approach by using data from the recent Italian research evaluation exercise (‘Valutazione della Qualità della Ricerca’ (VQR)). The VQR involved all state universities, all private universities granting publicly recognized academic degrees and several public research centres. Researchers affiliated to these institutions were asked to submit for evaluation a number of research products published during the period 2004–2010 and participating institutions were then ranked on the basis of the average score received by the products submitted by their researchers.

The VQR was co-ordinated by a public agency (the Agenzia di Valutazione del Sistema Universitario e della Ricerca) through groups of experts (‘Gruppi di Esperti della Valutazione’
(GEV), one for each of 14 broadly defined scientific areas. In most areas, journal papers were the main type of research products submitted for evaluation. For these areas, an important input to the evaluation process was the ranking of the journals where these papers were published. In our application, we focus on the set of journals that were considered relevant for the subarea statistics and financial mathematics, as defined in the VQR. We discuss the robustness of the estimated journal ranking to different rules for discretizing the available indicators, and how to handle the VQR requirement that journals must be classified in ordered groups of \textit{a priori} fixed size.

To be as close as possible to the actual problems that are faced in the VQR, we base our ranking of journals on exactly the same bibliometric indicators as used by GEV 13, namely the 2-year and 5-year impact factors IF and IF5, the article influence score AIS and the $h$-index. IF and IF5 measure the average number of citations to papers published in a given journal during the 2 and 5 previous years respectively. AIS 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 removes journal self-citations. A journal has an $h$-index of $h$ if $h$ of the $N$ papers that it published during a given period receive at least $h$ citations each, and the other $N-h$ receive no more than $h$ citations each.

The remainder of the paper is organized as follows. Section 2 describes the ranking strategy proposed, focusing in particular on the statistical models on which the strategy is based. Section 3 deals with estimation of these models and selection of the proper model specification. Section 4 illustrates our methodology by presenting the results that were obtained by using the data from the Italian VQR. Finally, Section 5 provides some conclusions.

2. Ranking strategy proposed

Let $n$ denote the number of journals to be ranked and let $r$ denote the number of available indicators on which the ranking is to be based. In our application (Section 4), the available indicators are IF, IF5 and AIS obtained from the \textit{Web of Science} plus the $h$-index obtained from Google Scholar, so $r=4$. Also let $x_{ij}$ be the value of indicator $j$ for journal $i$, with $i=1,\ldots,n$ and $j=1,\ldots,r$. Note that the value of an indicator may be missing for some journals; in our application this occurs for IF, IF5 and AIS, but never for the $h$-index. Our strategy for ranking scientific journals is based on a preliminary discretization of the available indicators in $s$ categories. A statistical model for polytomous item response data is then applied to the discretized indicators.

Denoting by $q_{j1},\ldots,q_{js-1}$ the cut-offs or threshold values for the $j$th indicator (e.g. its quartiles), the discretized value of indicator $j$ for journal $i$ is defined as

$$y_{ij} = \sum_{v=0}^{s-1} v \{q_{jv} < x_{ij} \leq q_{j,v+1}\}, \quad i=1,\ldots,n,$$

where $q_{j0} = -\infty$, $q_{js} = \infty$ and $\{A\}$ is the indicator function of the event $A$. Thus, $y_{ij}$ is equal to 0 if $x_{ij} \leq q_{j1}$, is equal to 1 if $q_{j1} < x_{ij} \leq q_{j2}$, and so on until $y_{ij} = s-1$ if $x_{ij} > q_{js-1}$. Clearly, if the value of $x_{ij}$ is missing for some $i$ and $j$, then the value of $y_{ij}$ is also missing. We collect the $r$ indicators corresponding to the $i$th journal in the $r$-dimensional vector $y_i = (y_{i1},\ldots,y_{ir})$.

Discretizing the available indicators, rather than working directly with their original values, implies some loss of information but allows us to avoid strong parametric assumptions. It also offers some robustness to measurement errors. However, since the way in which the available indicators are discretized is essentially arbitrary, it is important to assess the sensitivity of the results to the discretization assumed.

In Section 2.1 we first describe a baseline model, relying on typical IRT assumptions, to
analyse the data that were described above. These assumptions may be relaxed, giving rise to the extended models that are described in Section 2.2. Then, in Section 2.3, we show how to use these models to cluster journals into ordered groups and to predict their latent scientific impact.

2.1. Baseline item response theory model

Our baseline IRT model assumes the existence of \( n \) latent variables \( u_1, \ldots, u_n \), which we interpret as the scholarly impact of each journal. In interpreting these latent variables we may take into account that IF, IF5 and AIS are measures of the immediate impact of a journal (respectively 2 and 5 years after publication for IF and IF5, and 5 years after for AIS), whereas the \( h \)-index may be seen as a measure of lifetime impact. Consequently, in this baseline IRT model, a journal has a high impact (i.e. a high value of \( u_1 \)) if it has both a high immediate impact and a high lifetime impact. Distinguishing between these two dimensions may be important but would require two latent variables and, therefore, a bidimensional model of the type that is described in detail in Section 2.2. Note, however, that, whereas the unobserved impact of a journal may be multi-dimensional, as argued for example by Bollen et al. (2009), unidimensionality is essential if we want to obtain a unique ranking of journals.

Our baseline IRT model relies on the following three assumptions.

Assumption 1. For every sample unit (i.e. journal) \( i = 1, \ldots, n \), the discretized values \( y_{i1}, \ldots, y_{ir} \) are conditionally independent given the latent variable \( u_i \).

Assumption 2. The conditional distribution of every \( y_{ij} \) given \( u_i \) satisfies

\[
\log \left\{ \frac{p(y_{ij} \geq v | u_i)}{p(y_{ij} < v | u_i)} \right\} = \alpha_j (u_i - \beta_j v), \quad v = 1, \ldots, s - 1,
\]

as in the graded response model (Samejima, 1969).

Assumption 3. The latent variables \( u_1, \ldots, u_n \) are independent and have the same discrete distribution with \( k \) support points \( \xi_1, \ldots, \xi_k \) and corresponding probabilities \( \pi_1, \ldots, \pi_k \), with \( \pi_h = p(u_i = \xi_h) \). We then denote by \( \lambda_{jh}(v) = p(y_{ij} = v | u_i = \xi_h) \) the probability of the \( v \)-th category for the \( j \)-th discretized indicator \( y_{ij} \) conditionally on the latent variable \( u_i \) taking value \( \xi_h \). Obviously, these conditional probabilities satisfy condition (2).

Assumption 1, which is known as local independence, is typical of IRT models (Hambleton and Swaminathan, 1985). It also characterizes latent class models (Lazarsfeld and Henry, 1968; Goodman, 1974) in which the conditional response probabilities \( \lambda_{jh}(v) \) are not constrained to take a specific parametric form but are free parameters to be estimated. In practice, this assumption means that, if we knew the value of \( u_i \), then knowing the value of one indicator would not be useful to predict the value of any other indicator, as all the relevant information to capture the latent impact of a journal is already contained in \( u_i \).

Assumption 2 formalizes our interpretation of the latent variable \( u_i \). In particular, if the parameter \( \alpha_j \), which is known in the IRT literature as the discriminating index, is positive, then the distribution of \( y_{ij} \) stochastically increases with \( u_i \). In fact, parameterization (2) is based on a version of cumulative (or global) logits (see Agresti (2002) among others), which generalize the standard logits for binary outcomes to the case of ordinal outcomes. Therefore, assumption 2 means that the probability distribution of \( y_{ij} \) moves its mass towards higher categories as \( u_i \) increases. It is worth noting that, in terms of the original indicators \( x_{ij} \), it may equivalently be expressed as
This shows that the parameter $\alpha_j$ measures the sensitivity of the distribution of $x_{ij}$ to changes in $u_i$; that, in our context, represents the latent impact of a journal. The interpretation of the parameters $\beta_{jv}$, which are known in the IRT literature as the difficulty parameters, is context specific. In fact, in the educational context they are interpreted as the levels of difficulty of the various test item categories. In the present context, $\beta_{jv}$ is interpreted as a measure of severity in reaching category $v$ for indicator $j$, in the sense that, with $\alpha_j > 0$, the probability that journal $i$ falls in category $v$ or in a higher category is larger than the probability that it falls in one of the previous categories if and only if $u_i > \beta_{jv}$. Consequently, these difficulty parameters are in increasing order, i.e. $\beta_{j1} < \ldots < \beta_{j,s-1}$, for all $j$. For another interpretation of the model parameters, suppose that the original indicators satisfy the linear model $x_{ij} = \gamma_j + \delta_j u_i + \varepsilon_{ij}$, where $\delta_j \neq 0$ and $\varepsilon_{ij}$ is a zero-mean random variable distributed independently of $u_i$ with a logistic distribution. Combining this linear model with the discretization rule (1) gives model (2) with $\alpha_j = \delta_j$ and $\beta_{jv} = (q_{jv} - \gamma_j)/\delta_j$.

According to assumption 3, the latent variables $u_1, \ldots, u_n$ follow the same discrete distribution. Since both the support points $\xi_1, \ldots, \xi_k$ and the corresponding probabilities $\pi_1, \ldots, \pi_k$ are parameters to be estimated, this assumption avoids the need to specify a parametric distribution for the latent variables. Thus, our model is semiparametric in nature; see Lindsay et al. (1991) for a simpler semiparametric model for binary outcomes formulated along the same lines.

Assuming a discrete distribution for the latent variables is quite natural if the aim of the model is to cluster journals, giving rise to a model-based clustering (Fraley and Raftery, 2002). In this case, $k$ may be fixed a priori, provided that the size of each cluster is not constrained in advance. When not fixed a priori, $k$ may be chosen through some statistical criterion, such as the Bayesian information criterion BIC that was introduced by Schwarz (1978).

Assuming a discrete distribution for the latent variables makes sense even if we believe that their distribution is continuous. In fact, the discrete distribution assumed is better seen as a convenient approximation to the unknown continuous distribution, the quality of the approximation increasing with $k$. In this regard, it has been established that even moderate values of $k$ may provide an adequate approximation and even a better fit with respect to alternative models based on a parametric continuous distribution (Heckman and Singer, 1985; Lindsay et al., 1991).

By assumption 3 the latent variables $u_1, \ldots, u_n$ are mutually independent, so the response vectors $y_1, \ldots, y_n$ are also independent across sample units. This independence assumption may be restrictive in some cases, e.g. when the discretized indicators $y_{ij}$ are constructed by using as cut-offs a set of sample quantiles, which necessarily depend on the joint distribution of the data. However, such minor failures are unlikely to matter much, especially when the sample size $n$ is large. In contrast, relaxing the independence assumption would lead to a much more complex model.

Given the above three assumptions, the model parameters are the support points $\xi_h$, the corresponding probabilities $\pi_h$, $h = 1, \ldots, k$, the discriminant indices $\alpha_j$, $j = 1, \ldots, r$, and the difficulty parameters $\beta_{jv}$, $j = 1, \ldots, r$, $v = 1, \ldots, s - 1$. However, because of the identifiability constraints $\alpha_1 = 1$ and $\beta_{11} = 0$ and the fact that $\sum_{h=1}^k \pi_h = 1$, the number of free parameters is

$$\# \text{par} = k + k - 1 + r - 1 + r(s - 1) - 1 = 2k + rs - 3. \quad (3)$$

To make the model identifiable we can alternatively impose, as we do in our application, that the latent distribution has zero mean and unit variance. This identifiability constraint is equivalent to that directly expressed on the parameters ($\alpha_1 = 1, \beta_{11} = 0$), in the sense that the same maximum
of the likelihood is reached under both constraints. However, fixing the mean and the variance of the latent distribution makes the interpretation of the results easier and helps in comparing results that are obtained under different model specifications, as will be clear in our application (see in particular Section 4.4).

2.2. Extended item response theory models

In our application to the Italian VQR, the assumptions of the baseline IRT model may be too restrictive. In this section we show how to relax these assumptions, while retaining the main features of the proposed IRT model in terms of interpretability.

Local independence (assumption 1) is certainly restrictive, as some of the observed indicators (in particular IF and IF5) are constructed starting from the same information and using a similar method. Relaxing this assumption requires allowing for dependence between the indicators even after conditioning on the latent journal impact. This in turn raises the issue of how to model the conditional association between the indicators. For this aim, we adopt a Plackett formulation (Plackett, 1965). Its main advantage is that the association between each pair of indicators depends on only one parameter that has a straightforward interpretation and is the analogue of the correlation coefficient for the bivariate normal distribution. Further, in the resulting extended IRT model, the interpretation of the latent variables that was suggested at the beginning of Section 2.1 does not change.

For each pair \((j_1, j_2)\) of discretized indicators, the Plackett association parameter corresponds to the following global log-odds ratio (Douglas et al., 1991):

\[
\log \left\{ \frac{p(y_{ij_1} < v_1, y_{ij_2} < v_2 | u_i) p(y_{ij_1} > v_1, y_{ij_2} > v_2 | u_i)}{p(y_{ij_1} < v_1, y_{ij_2} > v_2 | u_i) p(y_{ij_1} > v_1, y_{ij_2} < v_2 | u_i)} \right\} = \tau_{j_1 j_2}, \quad v_1, v_2 = 1, \ldots, s - 1. \tag{4}
\]

Note that expression (4) does not depend on the specific categories \(v_1\) and \(v_2\) but only on the pair of indicators that are involved. When \(\tau_{j_1 j_2}\) is equal to 0, we have conditional independence between the pair of indicators given the latent variable \(u_i\), so local independence holds when \(\tau_{j_1 j_2} = 0\) for all possible pairs \((j_1, j_2)\). Moreover, \(\tau_{j_1 j_2} > 0\) corresponds to a positive association (as the first indicator increases, the second also tends to increase), whereas \(\tau_{j_1 j_2} < 0\) corresponds to a negative association. Testing the null hypothesis that \(\tau_{j_1 j_2} = 0\) against the alternative that \(\tau_{j_1 j_2} \neq 0\) may be performed by using a standard likelihood ratio test, whose test statistic has an asymptotic \(\chi^2(1)\) distribution under the null hypothesis. It is clear that the result of this test, as well as the results of testing all the other assumptions that were discussed in Section 2.1, may depend on the chosen number of classes \(k\). Moreover, relative to expression (3), the number of additional parameters that are required by this extension is equal to the number of pairs of responses for which the association parameter defined in expression (4) is not constrained to be 0.

Another assumption that may be problematic is unidimensionality, which has already been discussed at the beginning of Section 2.1. This assumption may be tested against multi-dimensionality (see Bartolucci (2007), among others). In particular, given the number \(k\) of support points, we can use a likelihood ratio test to compare the unidimensional model against a multi-dimensional model. The multi-dimensional model may be formulated by assuming that for each journal \(i\) there is a latent vector \(u_i = (u_{i1}, \ldots, u_{it})\) of \(t > 1\) (instead of a single latent variable \(u_i\)) such that

\[
\log \left\{ \frac{p(y_{ij} \geq v | u_i)}{p(y_{ij} < v | u_i)} \right\} = \alpha_j (u_{id} - \beta_{dv}), \quad v = 1, \ldots, s - 1,
\]
where the index $d_j$ denotes the latent variable measured by indicator $j$ and ranges from 1 to $t$. The latent vector $\mathbf{u}_i$ is still assumed to have a discrete distribution with support points $\xi_1, \ldots, \xi_k$ and mass probabilities $\pi_1, \ldots, \pi_k$. Considering the identifiability constraints, the number of additional free parameters that are required by this extension is equal to $(k - 2)(t - 1)$, which therefore makes sense only when $k > 2$. Further details on this extension may be found in Bacci et al. (2014).

Finally, the assumption that all latent variables $u_i$ (or vectors of latent variables $\mathbf{u}_i$ in the multidimensional case) have the same distribution may be relaxed by introducing a vector of covariates $\mathbf{w}_i$, for $i = 1, \ldots, n$. In the unidimensional case, there is no loss of generality in reordering the support points of the latent distribution so that $\xi_1 < \ldots < \xi_k$. We then assume a parameterization of the conditional distribution of each latent variable $u_i$ given $\mathbf{w}_i$ based on cumulative logits:

$$
\log \left\{ \frac{p(u_i > \xi_h | \mathbf{w}_i)}{p(u_i \leq \xi_h | \mathbf{w}_i)} \right\} = \log \left( \frac{\pi_{h+1,i} + \cdots + \pi_{ki}}{\pi_{1i} + \cdots + \pi_{hi}} \right) = \phi_h + \mathbf{w}_i' \psi, \quad h = 1, \ldots, k - 1, \tag{5}
$$

where $\pi_{hi} = p(u_i = \xi_h | \mathbf{w}_i)$, $\phi_1, \ldots, \phi_{k-1}$ are ordered intercepts and $\psi$ is a vector of regression coefficients. In practice, this is a proportional odds model (McCullagh, 1980) for the probability of belonging to the different latent classes. Unlike the multinomial logit parameterization, the vector $\psi$ is the same for each latent class $h$, making the interpretation easier. Moreover, the additional number of free parameters due to the inclusion of covariates in expression (5) is simply equal to the dimension of the vector $\mathbf{w}_i$.

### 2.3. Manifest and posterior distributions

As already mentioned, our model is of the IRT type. In fact, it may be seen as a finite mixture version of the graded response model. Its finite mixture nature derives from considering the distribution of the latent variables as discrete.

Under the basic assumptions 1–3 in Section 2.1, the manifest distribution of $y_i$ may be expressed as

$$
p(y_i) = \sum_{h=1}^k \pi_h \prod_{j=1}^r \lambda_{j|h}(y_{ij}). \tag{6}
$$

This manifest distribution is key for maximum likelihood estimation of the model parameters. The posterior distribution of the latent variable $u_i$, namely its conditional distribution given the vector $y_i$ of observed indicators, has probability mass function

$$
p(u_i = \xi_h | y_i) = \frac{\pi_h \prod_{j=1}^r \lambda_{j|h}(y_{ij})}{p(y_i)}. \tag{7}
$$

This is used to assign every sample unit (i.e. journal) to a given latent class or cluster. Specifically, once the model has been estimated, unit $i$ is assigned to latent class $h$ if

$$
h = \arg \max_{g=1,\ldots,k} p(u_i = \xi_g | y_i). \tag{8}
$$

Moreover, we can predict the value of $u_i$ by using the mean of the posterior distribution of $u_i$, or posterior mean, defined as

$$
\hat{u}_i = \sum_{h=1}^k \xi_h p(u_i = \xi_h | y_i). \tag{9}
$$
It is worth noting that posterior probabilities also provide a measure of the classification error when a journal is assigned to a particular latent class by using rule (8). In particular, the amount of classification error for journal $i$ may be directly measured by the distance of $\max_{g=1,\ldots,k} p(u_i = \xi_g|y_i)$ from 1. Moreover, even if the latent journal impact is modelled as discrete, its predicted value computed by equation (9) ranges on a continuous scale. To clarify this point, let $i$ be a journal for which we do not have a ‘neat’ assignment, i.e. for this journal there are two latent classes, say the first two, with posterior probabilities that are significantly greater than 0. Then, the predicted value $\hat{u}_i$ will be intermediate between the support points $\xi_1$ and $\xi_2$, closer to $\xi_1$ when $p(u_i = \xi_1|y_i) > p(u_i = \xi_2|y_i)$ and closer to $\xi_2$ when the opposite is true. This is of particular interest when we regard the discrete distribution assumed for the latent variables as an approximation to a continuous distribution.

When there are missing data, we compute the manifest distribution of the discretized indicators as

$$p(y_i) = \sum_{h=1}^{k} \pi_h \prod_{j=1}^{r} \lambda_{jh}(y_{ij}),$$

where $m_{ij}$ is a binary indicator 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 in the sense of Little and Rubin (2002). In our context, missingness at random implies that the event that the value of an indicator—say IF5—is missing may be predicted by the observable indicators, in our case the $h$-index, and the available covariates. We consider this assumption to be sufficiently realistic because, as discussed in Section 4, missing values of certain indicators tend to be observed for journals with a lower level of the $h$-index or for specific covariate values.

The manifest distribution (10) may be simply modified to take into account failure of some of the model assumptions, as discussed in Section 2.2. For example, when we allow for conditional dependence between a pair $(j_1, j_2)$ of indicators, the manifest distribution becomes

$$p(y_i) = \sum_{h=1}^{k} \pi_h \lambda_{j_1j_2|h}(y_{i_{j_1}}, y_{i_{j_2}}) \prod_{j=1}^{r} \lambda_{j|h}(y_{ij}),$$

where $\lambda_{j_1j_2|h}(y_{i_{j_1}}, y_{i_{j_2}})$ refers to the conditional (bivariate) distribution of $(y_{i_{j_1}}, y_{i_{j_2}})$ given $u_i = \xi_h$. This distribution depends on the conditional (univariate) probabilities $\lambda_{j_1|h}(y_{i_{j_1}})$ and $\lambda_{j_2|h}(y_{i_{j_2}})$ and the association parameter $\tau_{j_1j_2}$ that is defined in expression (4). A similar extension is available for the posterior distribution of $u_i$. More complex expressions arise when more than one pair of responses are allowed to be conditionally associated, in which case we may use the rules in Colombi and Forcina (2001), or when some indicators have missing values.

Finally, in the multi-dimensional case the above expressions for the manifest and posterior distributions remain unchanged, because we essentially have the same finite mixture model, except that a vector of supports points $\xi_h$ is now associated with each latent class $h = 1, \ldots, k$. Moreover, when the distribution of the latent variables depends on covariates, we only need to replace the probabilities $\pi_h$ with unit-specific probabilities $\pi_{hi}$. Given the parameters $\phi_h$ and $\psi_h$, these probabilities are obtained by simply inverting expression (5), and the manifest and posterior distributions are denoted by $p(y_i|w_i)$ and $p(u_i|y_i, w_i)$ respectively.

### 3. Likelihood inference

Given $n$ vectors of discrete indicators $y_1, \ldots, y_n$, one for each journal, inference is based on the sample log-likelihood
where $\theta$ is the vector containing all the model parameters and $p(y_i)$ is the manifest probability of $y_i$, which depends on $\theta$ and is computed according to expression (6), or its extended version which depends on the specific model formulation. When unit-specific covariates collected in the vector $w_i$ are available, we substitute $p(y_i)$ with $p(y_i|w_i)$ in the above expression for $l(\theta)$.

To maximize $l(\theta)$ with respect to $\theta$, we use the version of the EM algorithm (Dempster et al., 1977) that was described in Bacci et al. (2014), to which we refer for details. This implementation is available in the R package MultiLCIRT (available from http://CRAN.R-project.org/package=MultiLCIRT) that, in its current version, also addresses multi-dimensionality and allows for covariates. We also implemented an R extension of some of the functions that are included in this package to deal with the model that allows for local dependence; the extended functions are available from http://wileyonlinelibrary.com/journal/rss-datasets

In the remainder of this section we first provide a brief description of the EM algorithm. Then, we address the problem of model selection.

3.1. Estimation algorithm

The EM algorithm is based on the complete-data log-likelihood. For the baseline IRT model that was described in Section 2.1, this log-likelihood is equal to

$$l^*(\theta) = \sum_{h=1}^{k} \sum_{i=1}^{n} z_{hi} \log \left\{ \pi_h \prod_{j=1}^{r} \lambda_{j|h} (y_{ij}) \right\},$$

(12)

where $z_{hi}$ is an (unobserved) binary indicator equal to 1 if $u_i = \xi_h$ and to 0 otherwise. Slightly more complex versions of this expression are used for the complete-data log-likelihood of the extended IRT models that were described in Section 2.2, taking into account the results that were derived in Section 2.3.

The EM algorithm alternates between the following two steps until convergence.

(a) **E-step**: compute the conditional expected value of $l^*(\theta)$ given the observed data and the current value of the parameters.

(b) **M-step**: maximize the above expected value with respect to $\theta$ to obtain an updated estimate of the parameter vector.

The E-step computes the expected value of $z_{hi}$ given $y_i$ (and possibly $w_i$) for every $h$ and $i$ through the posterior probabilities, defined in expression (7) for the initial model, and then substitutes these expected values in equation (12). At the M-step, the resulting function is maximized with respect to $\theta$. When the distribution of the latent variable is the same for $i = 1, \ldots, n$, the existence of a closed form solution for the probabilities $\pi_h$ makes the maximization problem easier, whereas updating the other parameters requires only simple iterative algorithms. In contrast, with unit-specific covariates we need an iterative algorithm also to update the parameters in expression (5).

A crucial point is the initialization of the EM algorithm, as the likelihood function may present several local maxima. For the IRT model that was defined in Section 2.1, we rely on two different types of initialization: deterministic and random, both based on guessing initial values of the class weights $\pi_h$ and the conditional response probabilities $\lambda_{j|h}(v)$. In the deterministic initialization, we set $\pi_h = 1/k$, $h = 1, \ldots, k$, and choose $\lambda_{j|h}(v)$ so that, for $j = 1, \ldots, r$, these probabilities are increasingly ordered in $h$ for $v > (s - 1)/2$ and are decreasingly ordered in $h$.
for $v < (s - 1)/2$, with $h = 1, \ldots, k$ and $v = 0, \ldots, s - 1$. This rule guarantees that the adding-up constraint $\Sigma_{v=0}^{s-1} \lambda_{jvh}(v) = 1$ is satisfied for all $h$ and $j$. In the random initialization, first we draw every $\pi_h$ and $\lambda_{jvh}(v)$ from a uniform distribution between 0 and 1; then we normalize them in a suitable way. Similar rules are implemented when an extended model, formulated as described in Section 2.2, is adopted.

### 3.2. Model selection

To apply the ranking strategy proposed, we must select the number $k$ of support points (or latent classes) of the distribution of $u_i$. Of course, this is not needed when this number is fixed a priori. Among the model selection criteria that are available in the literature on finite mixture and latent class models (McLachlan and Peel (2000), chapter 6), we suggest the Bayesian information criterion (BIC) of Schwarz (1978). This criterion is based on minimization of the index

$$
BIC = -2 \hat{l}(\hat{\theta}) + \log(n) \#par,
$$

where $\hat{\theta}$ is the maximum likelihood estimate of $\theta$ under the model of interest and $\#par$ is the associated number of parameters, which is defined in equation (3) for the baseline IRT model. The BIC aims at selecting the model that is the best compromise between goodness of fit (measured by the log-likelihood) and complexity (measured by the number of parameters).

In practice, the BIC is similar to the Akaike information criterion (AIC) of Akaike (1973), which is based on the index

$$
AIC = -2 \hat{l}(\hat{\theta}) + 2 \#par.
$$

The main difference of the AIC with respect to the BIC is the smaller penalty for lack of parsimony when $\log(n) > 2$. Consequently, the AIC tends to select larger models with respect to the BIC, and for this reason the latter is usually preferred. Moreover, many simulation studies (e.g. Dias (2006)) show that the BIC tends to outperform the AIC in selecting the correct number of classes of a latent class model.

A different principle is behind another selection criterion, the normalized entropy criterion (NEC) that was proposed by Celeux and Soromenho (1996), which is specifically applied to the choice of the number of latent classes. This criterion takes into account the separation between the latent classes or, equivalently, the quality of the classification. For $k \geq 2$, the NEC is based on minimization of the index

$$
NEC = -\frac{\sum_{i=1}^{n} \sum_{h=1}^{k} \hat{p}(u_i = \xi_h|y_i) \log\{\hat{p}(u_i = \xi_h|y_i)\}}{\hat{l}(\hat{\theta}) - \hat{l}_1},
$$

where the numerator is an entropy measure based on the posterior probabilities of the latent classes and $\hat{l}_1$ is the log-likelihood of the model with only one latent class, i.e. the independence model. For $k = 1$, we conventionally set $NEC = 1$; see also Biernacki et al. (1999). The use of the NEC typically leads to selecting a very small number of latent classes that are strongly separated.

Finally, in selecting a suitable model for the analysis of the available data, we suggest beginning from the choice of the number of latent classes, $k$, under the initial assumptions that were described in Section 2.1. Then, using the same model selection criterion, say BIC, we suggest selecting some of the possible extensions that were described in Section 2.2.

### 4. Empirical application

We illustrate our approach by considering the problem that is faced in the Italian VQR of ranking
the journals that are considered relevant for a given scientific area. In most areas covered by the VQR, journal papers were the main type of research products submitted for evaluation. For these areas, an important input to the evaluation process was the ranking of the journals where these papers were published.

4.1. The Italian research assessment exercise

The VQR involved all state universities, all private universities granting publicly recognized academic degrees and several public research centres. Researchers who were affiliated to these institutions were asked to submit for evaluation various research products (i.e. journal papers, books, book chapters and patents) produced during the period 2004–2010. The typical number of products submitted by each researcher was six for public research centres and three for universities (under the convention that half of the time of the academic staff is devoted to teaching). The evaluation process assigned each product to one of six merit classes, namely ‘excellent’ (corresponding to ‘the highest 20% of the quality ranking shared by the international scientific community’), ‘good’ (corresponding to the top 20–40% segment), ‘acceptable’ (corresponding to the top 40–50% segment), ‘limited’ (corresponding to the lowest 50%), ‘not assessable’ and ‘fraud or plagiarism’. On the basis of its classification, each product received a numerical score equal to 1 for excellent, 0.80 for good, 0.50 for acceptable, 0 for limited, $-1$ for not assessable and $-2$ for fraud or plagiarism. Finally, participating institutions (and their departments) were ranked on the basis of average score received by the products that were submitted by their researchers. Individual researchers were not ranked.

In our application, we focus on a subset of the journals that were considered relevant for the area economics and statistics, whose products were evaluated by one of the groups of experts formed by Agenzia di Valutazione del Sistema Universitario e della Ricerca: GEV 13 as the area economics and statistics is 13 in the Italian classification of scientific areas. Journal papers in this area were evaluated via a bibliometric analysis based on a preliminary classification of the journals where they had been published and the number of citations that they had received. For a random sample of journal papers, the bibliometric analysis was complemented by informed peer review. (We refer to Bertocchi et al. (2014) for more detail on the procedures that were adopted by GEV 13.) To avoid different rankings across subareas, GEV 13 associated each journal with one and only one of its four subareas: business, management and finance; economics; economic history and history of economic thought; statistics and financial mathematics. For simplicity we focus on the set of $n = 445$ journals that were associated with the subarea statistics and financial mathematics.

Some of the journals are listed in Tables 1 and 2, along with the four bibliometric indicators that we used. The full data set is available from http://www.anvur.org/index.php?option=com_content&view=article&id=92&Itemid=390&lang=it.

4.2. Data

To be as close as possible to the actual problems that are faced in the VQR, we base our ranking on exactly the same bibliometric indicators as used by GEV 13, namely IF, IF5, AIS and the $h$-index.

GEV 13 obtained IF, IF5 and AIS from ISI–Thomson Reuters with reference to the year 2010 and collected the $h$-index from Google Scholar in April 2012 with reference to the period 2004–2010. Whereas IF, IF5 and AIS have missing values for some journals, the $h$-index is available for all journals considered by GEV 13. Note that IF, IF5 and AIS are based on the particular selection of journals by one specific commercial supplier and measure visibility in recent years (2 years for IF and 5 years for IF5 and AIS), not lifetime impact. In addition,
Table 1. List of journals assigned to the third class under model 1†

<table>
<thead>
<tr>
<th>Title</th>
<th>IF</th>
<th>IF5</th>
<th>AIS</th>
<th>h-index</th>
<th>post 3</th>
<th>post 4</th>
<th>Predicted value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Advances in Applied Probability</td>
<td>0.720</td>
<td>0.967</td>
<td>0.980</td>
<td>29</td>
<td>0.9591</td>
<td>0.0362</td>
<td>0.9036</td>
</tr>
<tr>
<td>Algorithmica</td>
<td>1.239</td>
<td>1.227</td>
<td>0.721</td>
<td>34</td>
<td>0.9157</td>
<td>0.0828</td>
<td>0.9398</td>
</tr>
<tr>
<td>American Mathematical Monthly</td>
<td>—‡</td>
<td>—‡</td>
<td>—‡</td>
<td>—‡</td>
<td>19</td>
<td>0.4871</td>
<td>0.1026</td>
</tr>
<tr>
<td>Annales de l’Institut Henri Poincaré—Probabilités et Statistiques</td>
<td>0.759</td>
<td>0.900</td>
<td>1.083</td>
<td>18</td>
<td>0.5712</td>
<td>0.0007</td>
<td>0.4642</td>
</tr>
<tr>
<td>Annals of Operations Research</td>
<td>0.675</td>
<td>1.223</td>
<td>0.741</td>
<td>48</td>
<td>0.9641</td>
<td>0.0301</td>
<td>0.8981</td>
</tr>
<tr>
<td>Applied Mathematics &amp; Optimization</td>
<td>0.881</td>
<td>1.061</td>
<td>0.822</td>
<td>20</td>
<td>0.9188</td>
<td>0.0034</td>
<td>0.8090</td>
</tr>
<tr>
<td>Applied Mathematics Letters</td>
<td>1.155</td>
<td>1.127</td>
<td>0.481</td>
<td>41</td>
<td>0.8649</td>
<td>0.0118</td>
<td>0.7704</td>
</tr>
<tr>
<td>Applied Psychological Measurement</td>
<td>1.137</td>
<td>1.497</td>
<td>0.854</td>
<td>24</td>
<td>0.9360</td>
<td>0.0630</td>
<td>0.9263</td>
</tr>
<tr>
<td>Biometrical Journal</td>
<td>1.438</td>
<td>1.273</td>
<td>0.822</td>
<td>24</td>
<td>0.9392</td>
<td>0.0284</td>
<td>0.8711</td>
</tr>
<tr>
<td>British Journal of Mathematical and Statistical Psychology</td>
<td>1.419</td>
<td>1.413</td>
<td>0.920</td>
<td>21</td>
<td>0.9304</td>
<td>0.0691</td>
<td>0.9309</td>
</tr>
<tr>
<td>Canadian Journal of Statistics—Revue Canadienne de Statistique</td>
<td>0.689</td>
<td>1.175</td>
<td>1.163</td>
<td>20</td>
<td>0.5395</td>
<td>0.0007</td>
<td>0.4331</td>
</tr>
<tr>
<td>Chaos, Solitons &amp; Fractals</td>
<td>1.268</td>
<td>1.729</td>
<td>0.538</td>
<td>83</td>
<td>0.8984</td>
<td>0.1009</td>
<td>0.9533</td>
</tr>
<tr>
<td>Combinatorics Probability &amp; Computing</td>
<td>0.990</td>
<td>1.008</td>
<td>1.465</td>
<td>28</td>
<td>0.7059</td>
<td>0.2935</td>
<td>1.0894</td>
</tr>
<tr>
<td>Complexity</td>
<td>1.367</td>
<td>1.190</td>
<td>0.528</td>
<td>43</td>
<td>0.9668</td>
<td>0.0165</td>
<td>0.8780</td>
</tr>
<tr>
<td>Computational Optimization and Applications</td>
<td>1.274</td>
<td>1.470</td>
<td>0.916</td>
<td>34</td>
<td>0.6216</td>
<td>0.3784</td>
<td>1.1501</td>
</tr>
<tr>
<td>Computational Statistics &amp; Data Analysis</td>
<td>1.089</td>
<td>1.363</td>
<td>0.754</td>
<td>55</td>
<td>0.5562</td>
<td>0.4437</td>
<td>1.1963</td>
</tr>
<tr>
<td>Demographic Research</td>
<td>1.531</td>
<td>1.582</td>
<td>0.708</td>
<td>38</td>
<td>0.8730</td>
<td>0.1267</td>
<td>0.9718</td>
</tr>
<tr>
<td>Differential and Integral Equations</td>
<td>—‡</td>
<td>—‡</td>
<td>—‡</td>
<td>—‡</td>
<td>18</td>
<td>0.4595</td>
<td>0.0770</td>
</tr>
<tr>
<td>Discrete Event Dynamic Systems</td>
<td>0.872</td>
<td>1.011</td>
<td>0.719</td>
<td>24</td>
<td>0.5196</td>
<td>0.0004</td>
<td>0.4132</td>
</tr>
<tr>
<td>Econometric Theory</td>
<td>1.015</td>
<td>1.264</td>
<td>1.541</td>
<td>40</td>
<td>0.7217</td>
<td>0.2776</td>
<td>1.0781</td>
</tr>
<tr>
<td>Econometrics Journal</td>
<td>0.691</td>
<td>1.166</td>
<td>1.253</td>
<td>30</td>
<td>0.8656</td>
<td>0.1312</td>
<td>0.9722</td>
</tr>
<tr>
<td>Educational and Psychological Measurement</td>
<td>0.831</td>
<td>1.434</td>
<td>0.658</td>
<td>34</td>
<td>0.8402</td>
<td>0.1573</td>
<td>0.9914</td>
</tr>
<tr>
<td>Electronic Journal of Probability</td>
<td>0.946</td>
<td>1.044</td>
<td>1.343</td>
<td>13</td>
<td>0.5582</td>
<td>0.0052</td>
<td>0.4592</td>
</tr>
<tr>
<td>Electronic Journal of Statistics</td>
<td>1.025</td>
<td>1.208</td>
<td>1.411</td>
<td>21</td>
<td>0.9074</td>
<td>0.0150</td>
<td>0.8173</td>
</tr>
<tr>
<td>Environmental and Ecological Statistics</td>
<td>1.645</td>
<td>1.641</td>
<td>0.872</td>
<td>22</td>
<td>0.9372</td>
<td>0.0622</td>
<td>0.9260</td>
</tr>
<tr>
<td>European Journal of Ageing: Social, Behavioural and Health Perspectives</td>
<td>1.119</td>
<td>—‡</td>
<td>—‡</td>
<td>24</td>
<td>0.7299</td>
<td>0.0380</td>
<td>0.6822</td>
</tr>
<tr>
<td>European Journal of Population</td>
<td>1.049</td>
<td>1.966</td>
<td>1.088</td>
<td>27</td>
<td>0.6473</td>
<td>0.3518</td>
<td>1.1303</td>
</tr>
<tr>
<td>Extremes</td>
<td>1.053</td>
<td>—‡</td>
<td>—‡</td>
<td>—‡</td>
<td>17</td>
<td>0.7830</td>
<td>0.0562</td>
</tr>
<tr>
<td>IEEE Transactions on Reliability</td>
<td>1.288</td>
<td>1.698</td>
<td>0.701</td>
<td>38</td>
<td>0.8012</td>
<td>0.1985</td>
<td>1.0227</td>
</tr>
<tr>
<td>IEE Transactions</td>
<td>1.186</td>
<td>1.535</td>
<td>0.777</td>
<td>42</td>
<td>0.6875</td>
<td>0.3125</td>
<td>1.1034</td>
</tr>
<tr>
<td>IMA Journal of Management Mathematics</td>
<td>0.608</td>
<td>1.980</td>
<td>1.006</td>
<td>16</td>
<td>0.7544</td>
<td>0.2377</td>
<td>1.0430</td>
</tr>
<tr>
<td>Informs Journal on Computing</td>
<td>1.172</td>
<td>1.450</td>
<td>0.903</td>
<td>34</td>
<td>0.5562</td>
<td>0.4437</td>
<td>1.1963</td>
</tr>
<tr>
<td>Insurance: Mathematics and Economics</td>
<td>1.178</td>
<td>1.451</td>
<td>0.780</td>
<td>41</td>
<td>0.6397</td>
<td>0.3603</td>
<td>1.1373</td>
</tr>
<tr>
<td>International Migration</td>
<td>0.556</td>
<td>1.047</td>
<td>0.520</td>
<td>33</td>
<td>0.6853</td>
<td>0.0024</td>
<td>0.5786</td>
</tr>
<tr>
<td>International Statistical Review</td>
<td>0.860</td>
<td>0.852</td>
<td>0.625</td>
<td>20</td>
<td>0.5492</td>
<td>0.0004</td>
<td>0.4422</td>
</tr>
<tr>
<td>Journal of Agricultural, Biological and Environmental Statistics</td>
<td>0.722</td>
<td>1.220</td>
<td>0.744</td>
<td>20</td>
<td>0.9299</td>
<td>0.0044</td>
<td>0.8214</td>
</tr>
<tr>
<td>Journal of Applied Probability</td>
<td>0.768</td>
<td>0.866</td>
<td>0.767</td>
<td>29</td>
<td>0.9591</td>
<td>0.0362</td>
<td>0.9036</td>
</tr>
<tr>
<td>Journal of Biopharmaceutical Statistics</td>
<td>1.073</td>
<td>1.285</td>
<td>0.602</td>
<td>26</td>
<td>0.9769</td>
<td>0.0124</td>
<td>0.8809</td>
</tr>
<tr>
<td>Journal of Biosocial Science</td>
<td>1.217</td>
<td>1.330</td>
<td>0.454</td>
<td>25</td>
<td>0.8981</td>
<td>0.0078</td>
<td>0.7961</td>
</tr>
</tbody>
</table>

(continued)
<table>
<thead>
<tr>
<th>Title</th>
<th>IF</th>
<th>IF5</th>
<th>AIS</th>
<th>h-index</th>
<th>post3</th>
<th>post4</th>
<th>Predicted value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Journal of Chemometrics</td>
<td>1.377</td>
<td>1.858</td>
<td>0.539</td>
<td>29</td>
<td>0.8773</td>
<td>0.1221</td>
<td>0.9684</td>
</tr>
<tr>
<td>Journal of Classification</td>
<td>1.100</td>
<td>1.225</td>
<td>0.704</td>
<td>18</td>
<td>0.7098</td>
<td>0.0015</td>
<td>0.6804</td>
</tr>
<tr>
<td>Journal of Computational and Applied Mathematics</td>
<td>1.030</td>
<td>1.299</td>
<td>0.571</td>
<td>51</td>
<td>0.8556</td>
<td>0.1440</td>
<td>0.9840</td>
</tr>
<tr>
<td>Journal of Empirical Finance</td>
<td>0.807</td>
<td>—‡</td>
<td>—‡</td>
<td>46</td>
<td>0.7506</td>
<td>0.1375</td>
<td>0.8702</td>
</tr>
<tr>
<td>Journal of Ethnic and Migration Studies</td>
<td>1.041</td>
<td>1.424</td>
<td>0.646</td>
<td>45</td>
<td>0.8773</td>
<td>0.1221</td>
<td>0.9684</td>
</tr>
<tr>
<td>Journal of Financial Econometrics</td>
<td>0.846</td>
<td>—‡</td>
<td>—‡</td>
<td>36</td>
<td>0.7148</td>
<td>0.2298</td>
<td>0.9909</td>
</tr>
<tr>
<td>Journal of Forecasting</td>
<td>0.655</td>
<td>0.866</td>
<td>0.525</td>
<td>28</td>
<td>0.9446</td>
<td>0.0073</td>
<td>0.8407</td>
</tr>
<tr>
<td>Journal of Global Optimization</td>
<td>1.160</td>
<td>1.433</td>
<td>0.698</td>
<td>38</td>
<td>0.8688</td>
<td>0.1307</td>
<td>0.9746</td>
</tr>
<tr>
<td>Journal of Mathematical Analysis and Applications</td>
<td>1.174</td>
<td>1.345</td>
<td>0.718</td>
<td>66</td>
<td>0.8556</td>
<td>0.1440</td>
<td>0.9840</td>
</tr>
<tr>
<td>Journal of Mathematical Psychology</td>
<td>1.582</td>
<td>1.833</td>
<td>1.017</td>
<td>27</td>
<td>0.9304</td>
<td>0.0691</td>
<td>0.9309</td>
</tr>
<tr>
<td>Journal of Mathematical Sciences</td>
<td>—‡</td>
<td>1.737</td>
<td>—‡</td>
<td>23</td>
<td>0.4595</td>
<td>0.0770</td>
<td>0.4792</td>
</tr>
<tr>
<td>Journal of Multivariate Analysis</td>
<td>1.010</td>
<td>1.180</td>
<td>0.917</td>
<td>37</td>
<td>0.9380</td>
<td>0.0599</td>
<td>0.9229</td>
</tr>
<tr>
<td>Journal of the Operational Research Society</td>
<td>1.102</td>
<td>1.481</td>
<td>0.603</td>
<td>46</td>
<td>0.8556</td>
<td>0.1440</td>
<td>0.9840</td>
</tr>
<tr>
<td>Journal of Optimization Theory and Applications</td>
<td>1.011</td>
<td>1.209</td>
<td>0.706</td>
<td>36</td>
<td>0.9623</td>
<td>0.0270</td>
<td>0.8913</td>
</tr>
<tr>
<td>Journal of Quality Technology</td>
<td>1.377</td>
<td>2.132</td>
<td>1.026</td>
<td>26</td>
<td>0.5081</td>
<td>0.4914</td>
<td>1.2295</td>
</tr>
<tr>
<td>Journal of the Royal Statistical Society Series C—Applied Statistics</td>
<td>0.645</td>
<td>1.284</td>
<td>0.961</td>
<td>29</td>
<td>0.5689</td>
<td>0.4309</td>
<td>1.1870</td>
</tr>
<tr>
<td>Journal of Time Series Analysis</td>
<td>0.678</td>
<td>0.888</td>
<td>0.871</td>
<td>25</td>
<td>0.9188</td>
<td>0.0034</td>
<td>0.8090</td>
</tr>
<tr>
<td>Lifetime Data Analysis</td>
<td>0.873</td>
<td>1.014</td>
<td>0.857</td>
<td>18</td>
<td>0.8982</td>
<td>0.0026</td>
<td>0.7873</td>
</tr>
<tr>
<td>Networks</td>
<td>0.991</td>
<td>1.167</td>
<td>0.728</td>
<td>32</td>
<td>0.9157</td>
<td>0.0828</td>
<td>0.9398</td>
</tr>
<tr>
<td>Networks and Heterogeneous Media</td>
<td>0.909</td>
<td>1.381</td>
<td>0.836</td>
<td>17</td>
<td>0.9406</td>
<td>0.0253</td>
<td>0.8672</td>
</tr>
<tr>
<td>Nonlinear Analysis, Theory, Methods &amp; Applications</td>
<td>1.279</td>
<td>1.409</td>
<td>0.565</td>
<td>44</td>
<td>0.8773</td>
<td>0.1221</td>
<td>0.9684</td>
</tr>
<tr>
<td>Operations Research Letters</td>
<td>0.743</td>
<td>1.007</td>
<td>0.661</td>
<td>34</td>
<td>0.9446</td>
<td>0.0073</td>
<td>0.8407</td>
</tr>
<tr>
<td>Operations Research—Spektrum</td>
<td>2.030</td>
<td>1.864</td>
<td>0.714</td>
<td>33</td>
<td>0.8474</td>
<td>0.1523</td>
<td>0.9901</td>
</tr>
<tr>
<td>Pharmaceutical Statistics</td>
<td>1.630</td>
<td>1.467</td>
<td>0.729</td>
<td>19</td>
<td>0.9394</td>
<td>0.0599</td>
<td>0.9243</td>
</tr>
<tr>
<td>Population Bulletin</td>
<td>1.182</td>
<td>1.741</td>
<td>1.007</td>
<td>6</td>
<td>0.9381</td>
<td>0.0426</td>
<td>0.8938</td>
</tr>
<tr>
<td>Population Studies</td>
<td>0.974</td>
<td>1.723</td>
<td>0.902</td>
<td>28</td>
<td>0.5949</td>
<td>0.4051</td>
<td>1.1689</td>
</tr>
<tr>
<td>Population, Space and Place</td>
<td>1.429</td>
<td>1.500</td>
<td>0.581</td>
<td>30</td>
<td>0.8730</td>
<td>0.1267</td>
<td>0.9718</td>
</tr>
<tr>
<td>Probabilistic Engineering</td>
<td>1.252</td>
<td>1.306</td>
<td>0.721</td>
<td>28</td>
<td>0.6397</td>
<td>0.5603</td>
<td>1.1373</td>
</tr>
<tr>
<td>Mechanics</td>
<td>0.971</td>
<td>0.966</td>
<td>0.754</td>
<td>20</td>
<td>0.9521</td>
<td>0.0051</td>
<td>0.8444</td>
</tr>
<tr>
<td>Probability in the Engineering and Informational Sciences</td>
<td>0.590</td>
<td>0.968</td>
<td>0.687</td>
<td>38</td>
<td>0.9276</td>
<td>0.0056</td>
<td>0.8212</td>
</tr>
<tr>
<td>Quantitative Finance</td>
<td>0.802</td>
<td>1.207</td>
<td>0.855</td>
<td>28</td>
<td>0.9538</td>
<td>0.0421</td>
<td>0.9084</td>
</tr>
<tr>
<td>Queueing Systems</td>
<td>1.418</td>
<td>1.120</td>
<td>0.860</td>
<td>19</td>
<td>0.9147</td>
<td>0.0126</td>
<td>0.8204</td>
</tr>
<tr>
<td>Set-valued Analysis</td>
<td>1.000</td>
<td>1.239</td>
<td>0.409</td>
<td>48</td>
<td>0.8240</td>
<td>0.0080</td>
<td>0.7238</td>
</tr>
<tr>
<td>Social Indicators Research</td>
<td>0.714</td>
<td>1.021</td>
<td>0.756</td>
<td>20</td>
<td>0.8941</td>
<td>0.0024</td>
<td>0.7832</td>
</tr>
<tr>
<td>Statistical Modelling</td>
<td>0.956</td>
<td>1.020</td>
<td>0.969</td>
<td>32</td>
<td>0.9549</td>
<td>0.0276</td>
<td>0.8851</td>
</tr>
<tr>
<td>Statistica Sinica</td>
<td>1.777</td>
<td>1.700</td>
<td>0.460</td>
<td>22</td>
<td>0.9279</td>
<td>0.0086</td>
<td>0.8266</td>
</tr>
<tr>
<td>Stochastic Environmental Research and Risk Assessment</td>
<td>1.778</td>
<td>1.818</td>
<td>0.698</td>
<td>27</td>
<td>0.9702</td>
<td>0.0262</td>
<td>0.8976</td>
</tr>
<tr>
<td>Studies in Family Planning</td>
<td>0.667</td>
<td>1.586</td>
<td>0.506</td>
<td>26</td>
<td>0.9211</td>
<td>0.0120</td>
<td>0.8258</td>
</tr>
<tr>
<td>System Dynamics Review</td>
<td>0.412</td>
<td>1.768</td>
<td>1.022</td>
<td>54</td>
<td>0.5789</td>
<td>0.4211</td>
<td>1.1803</td>
</tr>
<tr>
<td>Systems &amp; Control Letters</td>
<td>1.036</td>
<td>1.108</td>
<td>1.176</td>
<td>18</td>
<td>0.7323</td>
<td>0.0018</td>
<td>0.6237</td>
</tr>
</tbody>
</table>

†post3 and post4 stand for the posterior probability that a certain journal is in the third and in the fourth class respectively.
‡Not applicable.
### Table 2. List of journals assigned to the fourth class under model 1†

<table>
<thead>
<tr>
<th>Title</th>
<th>IF</th>
<th>IF5</th>
<th>AIS</th>
<th>h-index</th>
<th>post3</th>
<th>post4</th>
<th>Predicted value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ageing and Society</td>
<td>1.309</td>
<td>1.900</td>
<td>0.559</td>
<td>39</td>
<td>0.3858</td>
<td>0.6140</td>
<td>1.3164</td>
</tr>
<tr>
<td>American Statistican</td>
<td>0.981</td>
<td>1.322</td>
<td>0.924</td>
<td>29</td>
<td>0.4994</td>
<td>0.5006</td>
<td>1.2364</td>
</tr>
<tr>
<td>Annals of Applied Probability</td>
<td>1.120</td>
<td>1.447</td>
<td>1.595</td>
<td>44</td>
<td>0.2143</td>
<td>0.7857</td>
<td>1.4380</td>
</tr>
<tr>
<td>Annals of Applied Statistics</td>
<td>1.746</td>
<td>2.443</td>
<td>2.072</td>
<td>28</td>
<td>0.0211</td>
<td>0.9789</td>
<td>1.5746</td>
</tr>
<tr>
<td>Annals of Probability</td>
<td>1.470</td>
<td>1.665</td>
<td>1.996</td>
<td>45</td>
<td>0.1388</td>
<td>0.8612</td>
<td>1.4914</td>
</tr>
<tr>
<td>Annals of Statistics</td>
<td>2.940</td>
<td>3.274</td>
<td>3.260</td>
<td>76</td>
<td>0.0061</td>
<td>0.9939</td>
<td>1.5851</td>
</tr>
<tr>
<td>Bayesian Analysis</td>
<td>1.213</td>
<td>2.756</td>
<td>2.237</td>
<td>25</td>
<td>0.3744</td>
<td>0.6245</td>
<td>1.3230</td>
</tr>
<tr>
<td>Bernoulli</td>
<td>1.000</td>
<td>1.284</td>
<td>1.577</td>
<td>33</td>
<td>0.2675</td>
<td>0.7325</td>
<td>1.4003</td>
</tr>
<tr>
<td>Biometrics</td>
<td>1.764</td>
<td>2.204</td>
<td>1.594</td>
<td>44</td>
<td>0.0620</td>
<td>0.9380</td>
<td>1.5457</td>
</tr>
<tr>
<td>Biostatistics</td>
<td>2.769</td>
<td>3.303</td>
<td>2.312</td>
<td>45</td>
<td>0.0134</td>
<td>0.9866</td>
<td>1.5800</td>
</tr>
<tr>
<td>Chaos, an Interdisciplinary Journal of Nonlinear Science</td>
<td>2.081</td>
<td>2.134</td>
<td>1.082</td>
<td>37</td>
<td>0.0576</td>
<td>0.9424</td>
<td>1.5487</td>
</tr>
<tr>
<td>Chemometrics and Intelligent Laboratory Systems,</td>
<td>2.222</td>
<td>2.415</td>
<td>0.645</td>
<td>44</td>
<td>0.2109</td>
<td>0.7891</td>
<td>1.4404</td>
</tr>
<tr>
<td>Computational Biology and Bioinformatics, IEEE, ACM Transactions</td>
<td>1.664</td>
<td>2.171</td>
<td>0.950</td>
<td>33</td>
<td>0.0598</td>
<td>0.9402</td>
<td>1.5472</td>
</tr>
<tr>
<td>Computers &amp; Operations Research</td>
<td>1.769</td>
<td>2.250</td>
<td>0.886</td>
<td>75</td>
<td>0.0620</td>
<td>0.9380</td>
<td>1.5457</td>
</tr>
<tr>
<td>Decision Support Systems</td>
<td>2.135</td>
<td>2.573</td>
<td>0.708</td>
<td>68</td>
<td>0.2109</td>
<td>0.7891</td>
<td>1.4404</td>
</tr>
<tr>
<td>Demography</td>
<td>2.465</td>
<td>3.817</td>
<td>2.207</td>
<td>57</td>
<td>0.0061</td>
<td>0.9939</td>
<td>1.5851</td>
</tr>
<tr>
<td>Ecological Modelling</td>
<td>1.769</td>
<td>2.439</td>
<td>0.753</td>
<td>61</td>
<td>0.0620</td>
<td>0.9380</td>
<td>1.5457</td>
</tr>
<tr>
<td>Econometric Reviews</td>
<td>1.088</td>
<td>1.400</td>
<td>1.346</td>
<td>31</td>
<td>0.1722</td>
<td>0.8278</td>
<td>1.4677</td>
</tr>
<tr>
<td>European Journal of Operational Research</td>
<td>2.159</td>
<td>2.513</td>
<td>0.886</td>
<td>108</td>
<td>0.0620</td>
<td>0.9380</td>
<td>1.5457</td>
</tr>
<tr>
<td>Finance and Stochastics</td>
<td>1.326</td>
<td>1.870</td>
<td>0.216</td>
<td>43</td>
<td>0.0310</td>
<td>0.9690</td>
<td>1.5778</td>
</tr>
<tr>
<td>Fuzzy Sets and Systems</td>
<td>1.875</td>
<td>2.250</td>
<td>0.591</td>
<td>67</td>
<td>0.2109</td>
<td>0.7891</td>
<td>1.4404</td>
</tr>
<tr>
<td>Games and Economic Behavior</td>
<td>1.017</td>
<td>1.503</td>
<td>1.817</td>
<td>63</td>
<td>0.1722</td>
<td>0.8278</td>
<td>1.4677</td>
</tr>
<tr>
<td>International Family Planning Perspectives</td>
<td>2.118</td>
<td>2.575</td>
<td>0.902</td>
<td>30</td>
<td>0.0446</td>
<td>0.9554</td>
<td>1.5580</td>
</tr>
<tr>
<td>International Migration Review</td>
<td>1.188</td>
<td>2.145</td>
<td>1.236</td>
<td>47</td>
<td>0.0519</td>
<td>0.9481</td>
<td>1.5528</td>
</tr>
<tr>
<td>Journal of the American Statistical Association</td>
<td>2.063</td>
<td>3.439</td>
<td>3.280</td>
<td>73</td>
<td>0.0061</td>
<td>0.9939</td>
<td>1.5851</td>
</tr>
<tr>
<td>Journal of Applied Econometrics</td>
<td>1.341</td>
<td>2.268</td>
<td>2.172</td>
<td>55</td>
<td>0.0179</td>
<td>0.9821</td>
<td>1.5768</td>
</tr>
<tr>
<td>Journal of Business &amp; Economic Statistics</td>
<td>1.693</td>
<td>2.433</td>
<td>2.804</td>
<td>50</td>
<td>0.0077</td>
<td>0.9923</td>
<td>1.5840</td>
</tr>
<tr>
<td>Journal of Computational Biology</td>
<td>1.600</td>
<td>2.033</td>
<td>0.907</td>
<td>45</td>
<td>0.0576</td>
<td>0.9424</td>
<td>1.5487</td>
</tr>
<tr>
<td>Journal of Computational and Graphical Statistics</td>
<td>1.206</td>
<td>1.848</td>
<td>1.576</td>
<td>39</td>
<td>0.2143</td>
<td>0.7857</td>
<td>1.4380</td>
</tr>
<tr>
<td>Journal of Econometrics</td>
<td>1.815</td>
<td>2.823</td>
<td>3.016</td>
<td>91</td>
<td>0.0108</td>
<td>0.9892</td>
<td>1.5818</td>
</tr>
<tr>
<td>Journal of Educational and Behavioral Statistics</td>
<td>1.644</td>
<td>2.474</td>
<td>1.862</td>
<td>27</td>
<td>0.0840</td>
<td>0.9160</td>
<td>1.5301</td>
</tr>
<tr>
<td>Journal of Risk and Uncertainty</td>
<td>1.558</td>
<td>1.953</td>
<td>1.317</td>
<td>36</td>
<td>0.0134</td>
<td>0.9866</td>
<td>1.5800</td>
</tr>
<tr>
<td>Journal of the Royal Statistical Society, Series A—Statistics in Society</td>
<td>2.570</td>
<td>2.527</td>
<td>1.822</td>
<td>40</td>
<td>0.0090</td>
<td>0.9910</td>
<td>1.5831</td>
</tr>
<tr>
<td>Journal of the Royal Statistical Society, Series B—Statistical Methodology</td>
<td>3.500</td>
<td>5.086</td>
<td>4.822</td>
<td>53</td>
<td>0.0090</td>
<td>0.9910</td>
<td>1.5831</td>
</tr>
<tr>
<td>Journal of Statistical Mechanics: Theory and Experiment</td>
<td>1.822</td>
<td>2.169</td>
<td>1.088</td>
<td>39</td>
<td>0.0778</td>
<td>0.9222</td>
<td>1.5345</td>
</tr>
<tr>
<td>Journal of Statistical Physics</td>
<td>1.447</td>
<td>1.534</td>
<td>0.950</td>
<td>47</td>
<td>0.4358</td>
<td>0.5642</td>
<td>1.2814</td>
</tr>
<tr>
<td>Journal of Statistical Software</td>
<td>2.647</td>
<td>3.654</td>
<td>1.735</td>
<td>36</td>
<td>0.0100</td>
<td>0.9900</td>
<td>1.5824</td>
</tr>
<tr>
<td>Mathematical Finance</td>
<td>1.052</td>
<td>1.801</td>
<td>1.892</td>
<td>47</td>
<td>0.2143</td>
<td>0.7857</td>
<td>1.4380</td>
</tr>
<tr>
<td>Mathematical Programming</td>
<td>1.970</td>
<td>2.781</td>
<td>1.951</td>
<td>55</td>
<td>0.0108</td>
<td>0.9892</td>
<td>1.5818</td>
</tr>
<tr>
<td>Mathematics of Computation</td>
<td>1.382</td>
<td>1.565</td>
<td>1.276</td>
<td>42</td>
<td>0.1137</td>
<td>0.8863</td>
<td>1.5091</td>
</tr>
<tr>
<td>Mathematics of Operations Research</td>
<td>1.145</td>
<td>1.478</td>
<td>1.423</td>
<td>34</td>
<td>0.1722</td>
<td>0.8278</td>
<td>1.4677</td>
</tr>
<tr>
<td>Multivariate Behavioral Research</td>
<td>1.290</td>
<td>3.295</td>
<td>2.062</td>
<td>28</td>
<td>0.0143</td>
<td>0.9857</td>
<td>1.5794</td>
</tr>
</tbody>
</table>

(continued)
Table 2 (continued)

<table>
<thead>
<tr>
<th>Title</th>
<th>IF</th>
<th>IF5</th>
<th>AIS</th>
<th>h-index</th>
<th>post3</th>
<th>post4</th>
<th>Predicted value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nonlinear Analysis Real World Applications</td>
<td>2.138</td>
<td>2.039</td>
<td>0.625</td>
<td>40</td>
<td>0.2488</td>
<td>0.7512</td>
<td>1.4136</td>
</tr>
<tr>
<td>Operations Research</td>
<td>2.000</td>
<td>2.708</td>
<td>1.928</td>
<td>61</td>
<td>0.0061</td>
<td>0.9939</td>
<td>1.5851</td>
</tr>
<tr>
<td>Oxford Bulletin of Economics and Statistics</td>
<td>1.182</td>
<td>1.622</td>
<td>1.225</td>
<td>46</td>
<td>0.1960</td>
<td>0.8040</td>
<td>1.4509</td>
</tr>
<tr>
<td>Pattern Recognition Letters</td>
<td>1.235</td>
<td>1.897</td>
<td>0.696</td>
<td>67</td>
<td>0.3073</td>
<td>0.6926</td>
<td>1.3721</td>
</tr>
<tr>
<td>Perspectives on Sexual and Reproductive Health</td>
<td>2.075</td>
<td>3.842</td>
<td>1.367</td>
<td>35</td>
<td>0.0104</td>
<td>0.9896</td>
<td>1.5821</td>
</tr>
<tr>
<td>Population and Development Review</td>
<td>1.507</td>
<td>2.381</td>
<td>1.362</td>
<td>43</td>
<td>0.0077</td>
<td>0.9923</td>
<td>1.5840</td>
</tr>
<tr>
<td>Probability Theory and Related Fields</td>
<td>1.590</td>
<td>1.625</td>
<td>1.985</td>
<td>40</td>
<td>0.1588</td>
<td>0.8412</td>
<td>1.4772</td>
</tr>
<tr>
<td>Psychometrika</td>
<td>1.778</td>
<td>1.804</td>
<td>1.181</td>
<td>28</td>
<td>0.4358</td>
<td>0.5642</td>
<td>1.2814</td>
</tr>
<tr>
<td>Reliability Engineering &amp; System Safety</td>
<td>1.899</td>
<td>2.023</td>
<td>0.660</td>
<td>52</td>
<td>0.2109</td>
<td>0.7891</td>
<td>1.4404</td>
</tr>
<tr>
<td>Risk Analysis</td>
<td>2.096</td>
<td>2.344</td>
<td>0.788</td>
<td>57</td>
<td>0.0446</td>
<td>0.9554</td>
<td>1.5580</td>
</tr>
<tr>
<td>Scandinavian Journal of Statistics</td>
<td>0.835</td>
<td>1.326</td>
<td>1.354</td>
<td>28</td>
<td>0.2094</td>
<td>0.7905</td>
<td>1.4414</td>
</tr>
<tr>
<td>SIAM Journal on Applied Mathematics</td>
<td>1.529</td>
<td>1.824</td>
<td>1.602</td>
<td>37</td>
<td>0.4358</td>
<td>0.5642</td>
<td>1.2814</td>
</tr>
<tr>
<td>SIAM Journal on Control and Optimization</td>
<td>1.297</td>
<td>1.666</td>
<td>1.270</td>
<td>39</td>
<td>0.1421</td>
<td>0.8579</td>
<td>1.4890</td>
</tr>
<tr>
<td>SIAM Journal on Imaging Sciences</td>
<td>4.500</td>
<td>4.500</td>
<td>2.661</td>
<td>16</td>
<td>0.2197</td>
<td>0.7801</td>
<td>1.4339</td>
</tr>
<tr>
<td>SIAM Journal on Mathematical Analysis</td>
<td>1.797</td>
<td>1.744</td>
<td>1.454</td>
<td>36</td>
<td>0.1137</td>
<td>0.8863</td>
<td>1.5091</td>
</tr>
<tr>
<td>SIAM Journal on Optimization</td>
<td>2.091</td>
<td>2.566</td>
<td>1.686</td>
<td>47</td>
<td>0.0100</td>
<td>0.9900</td>
<td>1.5824</td>
</tr>
<tr>
<td>Sociological Methods &amp; Research</td>
<td>2.000</td>
<td>2.448</td>
<td>1.722</td>
<td>30</td>
<td>0.0077</td>
<td>0.9923</td>
<td>1.5840</td>
</tr>
<tr>
<td>Stata Journal</td>
<td>2.000</td>
<td>3.142</td>
<td>1.964</td>
<td>37</td>
<td>0.0104</td>
<td>0.9896</td>
<td>1.5821</td>
</tr>
<tr>
<td>Statistical Applications in Genetics and Molecular Biology</td>
<td>1.842</td>
<td>2.182</td>
<td>1.100</td>
<td>30</td>
<td>0.0778</td>
<td>0.9222</td>
<td>1.5345</td>
</tr>
<tr>
<td>Statistical Methods in Medical Research</td>
<td>1.768</td>
<td>2.541</td>
<td>1.535</td>
<td>29</td>
<td>0.0100</td>
<td>0.9900</td>
<td>1.5824</td>
</tr>
<tr>
<td>Statistical Science</td>
<td>2.480</td>
<td>3.504</td>
<td>3.383</td>
<td>41</td>
<td>0.0077</td>
<td>0.9923</td>
<td>1.5840</td>
</tr>
<tr>
<td>Statistics and Computing</td>
<td>1.851</td>
<td>2.339</td>
<td>1.838</td>
<td>31</td>
<td>0.0100</td>
<td>0.9900</td>
<td>1.5824</td>
</tr>
<tr>
<td>Statistics in Medicine</td>
<td>2.328</td>
<td>2.334</td>
<td>1.330</td>
<td>62</td>
<td>0.0077</td>
<td>0.9923</td>
<td>1.5840</td>
</tr>
<tr>
<td>Stochastic Processes and Their Applications</td>
<td>0.951</td>
<td>1.381</td>
<td>1.368</td>
<td>38</td>
<td>0.2094</td>
<td>0.7905</td>
<td>1.4414</td>
</tr>
<tr>
<td>Structural Equation Modeling</td>
<td>2.738</td>
<td>5.611</td>
<td>2.633</td>
<td>40</td>
<td>0.0134</td>
<td>0.9866</td>
<td>1.5800</td>
</tr>
<tr>
<td>Technometrics</td>
<td>1.560</td>
<td>1.985</td>
<td>1.424</td>
<td>34</td>
<td>0.0061</td>
<td>0.9939</td>
<td>1.5851</td>
</tr>
</tbody>
</table>

†post3 and post4 stand for the posterior probability that a certain journal is in the third and in the fourth class respectively.

the selection of journals by ISI–Thomson Reuters is not completely transparent and does not follow quality criteria alone. In contrast, the h-index is based on a much larger but also much more heterogeneous set of indexed sources, also including books, book chapters and publicly available conference proceedings. Although the h-index may better measure the impact of a journal, it does favour older journals.

Table 3 shows descriptive statistics for our $r = 4$ indicators. Note that IF is available for only 56.2% of the journals and IF5 and AIS for only 47.4%, whereas the h-index is always available. Also note the lower mean h-index for journals with IF missing compared with those with IF5 and AIS missing. All four indicators are available for a subset of 211 journals. For this subset of journals Table 3 shows also the mean, the variance, the index of skewness, the quartiles and the deciles of the available indicators. Fig. 1 presents their distribution and their scatter plot matrix. In general, we note large differences between the distribution of the h-index for ISI–Thomson Reuters and non-ISI–Thomson Reuters journals, suggesting that the h-index, together with the covariates that are included in the data set, may be a good predictor of the probability that another indicator is missing for a certain journal.
Table 3. Descriptive statistics for the indicators observed

<table>
<thead>
<tr>
<th>Results for the following indicators:</th>
<th>IF</th>
<th>IF5</th>
<th>AIS</th>
<th>h-index</th>
</tr>
</thead>
<tbody>
<tr>
<td>Missing values (%)</td>
<td>43.8</td>
<td>52.6</td>
<td>52.6</td>
<td>0</td>
</tr>
<tr>
<td>Mean</td>
<td>1.056</td>
<td>1.472</td>
<td>0.946</td>
<td>19.766</td>
</tr>
<tr>
<td>Mean (given IF missing)</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>8.446</td>
</tr>
<tr>
<td>Mean (given IF5 or AIS missing)</td>
<td>0.492</td>
<td>—</td>
<td>—</td>
<td>9.509</td>
</tr>
<tr>
<td>Variance</td>
<td>0.418</td>
<td>0.751</td>
<td>0.480</td>
<td>267.585</td>
</tr>
<tr>
<td>Skewness index</td>
<td>1.325</td>
<td>1.526</td>
<td>1.938</td>
<td>1.575</td>
</tr>
</tbody>
</table>

Quartiles

| 1st        | 0.586| 0.840| 0.506| 7.0    |
| 2nd        | 0.954| 1.284| 0.721| 14.0   |
| 3rd        | 1.381| 1.867| 1.203| 28.0   |

Deciles

| 1st        | 0.370| 0.590| 0.313| 4.0    |
| 2nd        | 0.521| 0.766| 0.454| 6.0    |
| 3rd        | 0.643| 0.967| 0.553| 9.0    |
| 4th        | 0.754| 1.108| 0.660| 12.0   |
| 5th        | 0.954| 1.284| 0.721| 14.0   |
| 6th        | 1.088| 1.467| 0.871| 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   |

As for the covariates, we include the age of each journal (i.e. the number of years since the journal was first published), its language and the country of the publisher. For age, we consider six classes (1–3, 4–6, 7–10, 11–20, 21–40 and 41 or more years), for language we distinguish only between English and other languages, and for the country of the publisher we consider three categories: one for the USA (USA), one for the UK, Germany or the Netherlands (UK, D, NL), and one for all other countries (other). Table 4 shows how the distribution of the four indicators depends on these covariates. As expected, the percentage of missing values of IF, IF5 and AIS is much higher for more recent than for older journals. The mean value of the available indicators does not show a clear pattern but tends to be lower for journals in the age class 4–6 years. In this regard note that IF, IF5 and AIS are available for only two of the 17 journals in the first age category (1–3 years), so, the exceptionally high value of the mean of these indicators is better regarded as an anomaly. The pattern for the $h$-index shows instead a clear tendency for this indicator to increase with the age of a journal. Another clear difference also emerges in connection with language, as English journals have on average much higher values of all indicators compared with journals in other languages. The percentage of missing values is also much higher for this second category. Similarly, for journals that are published in the USA, the UK, Germany or the Netherlands we observe a smaller percentage of missing values and a higher mean for all indicators compared with journals that are published in other countries. However the comparison between journals that are published in the USA and those published in the UK, Germany or the Netherlands is not so clear, as the first perform better in terms of IF, IF5 and AIS, but perform a little worse in terms of $h$-index.
4.3. Model fitting

As discussed at the beginning of Section 2, the first step of our strategy consists of discretizing the indicators observed. We present two alternative types of discretization: the first uses as cut-offs the sample quartiles \( s = 4 \); the second is based on fixed cut-offs giving \( s = 5 \) categories. In the latter case, the cut-offs are \((0.50, 0.75, 1.00, 1.50)\) for IF and AIS, \((0.75, 1.00, 1.50, 2.00)\) for IF5 and \((5, 10, 20, 30)\) for the \( h \)-index. To avoid categories containing too few journals, these sets of cut-offs are not equally spaced, i.e. the relationship between the value of the discretized indicators \( y_{ij} \) and the cut-offs \( q_{jk} \) is not restricted to be linear.

Given the discretized indicators \( y_{ij} \), to select an appropriate model for our data we proceed as discussed in Section 3.2. First we choose the number \( k \) of latent classes for the baseline IRT model which assumes local independence, unidimensionality and no covariates. To avoid
Table 4. Descriptive statistics for the indicators observed given the available covariates

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Category</th>
<th>Number</th>
<th>Missing values (%)</th>
<th>Mean</th>
<th>IF</th>
<th>IF5</th>
<th>AIS</th>
<th>h-index</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>1–3</td>
<td>17</td>
<td>88.2</td>
<td>2.487</td>
<td>88.2</td>
<td>2.487</td>
<td>88.2</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>4–6</td>
<td>43</td>
<td>74.4</td>
<td>0.791</td>
<td>79.1</td>
<td>1.324</td>
<td>79.1</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>7–10</td>
<td>50</td>
<td>66.0</td>
<td>1.112</td>
<td>78.0</td>
<td>1.731</td>
<td>78.0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>11–20</td>
<td>112</td>
<td>48.2</td>
<td>1.082</td>
<td>59.8</td>
<td>1.516</td>
<td>59.8</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>21–40</td>
<td>138</td>
<td>26.8</td>
<td>0.982</td>
<td>36.2</td>
<td>1.362</td>
<td>36.2</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>⩾41</td>
<td>85</td>
<td>28.2</td>
<td>1.137</td>
<td>34.1</td>
<td>1.549</td>
<td>34.1</td>
<td>0</td>
</tr>
<tr>
<td>Language</td>
<td>English</td>
<td>430</td>
<td>42.8</td>
<td>1.065</td>
<td>51.2</td>
<td>1.475</td>
<td>51.2</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Other</td>
<td>15</td>
<td>73.3</td>
<td>0.490</td>
<td>93.3</td>
<td>0.992</td>
<td>93.3</td>
<td>0</td>
</tr>
<tr>
<td>Country</td>
<td>USA</td>
<td>130</td>
<td>27.7</td>
<td>1.286</td>
<td>31.5</td>
<td>1.694</td>
<td>31.5</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>UK, D, NL</td>
<td>143</td>
<td>27.3</td>
<td>1.091</td>
<td>36.4</td>
<td>1.483</td>
<td>36.4</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Other</td>
<td>172</td>
<td>69.8</td>
<td>0.568</td>
<td>82.0</td>
<td>0.804</td>
<td>82.0</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 5. Fit of the baseline IRT models with discretization based on quartiles ($s = 4$) and fixed cut-offs ($s = 5$) as a function of the number $k$ of latent classes†

<table>
<thead>
<tr>
<th>$k$</th>
<th>Results for quartiles</th>
<th>Results for fixed cut-offs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$l(\hat{\theta})$</td>
<td>$#\text{par}$</td>
</tr>
<tr>
<td>1</td>
<td>$-1544.9$</td>
<td>12</td>
</tr>
<tr>
<td>2</td>
<td>$-1343.8$</td>
<td>17</td>
</tr>
<tr>
<td>3</td>
<td>$-1293.0$</td>
<td>19</td>
</tr>
<tr>
<td>4</td>
<td>$-1273.6$</td>
<td>21</td>
</tr>
<tr>
<td>5</td>
<td>$-1271.0$</td>
<td>23</td>
</tr>
<tr>
<td>6</td>
<td>$-1271.0$</td>
<td>25</td>
</tr>
</tbody>
</table>

†The smallest values of AIC, BIC and NEC are in italics.

spurious estimates due to multimodality of the sample log-likelihood, we use two types of initialization (deterministic and random) of the EM algorithm. The results, in terms of the AIC, BIC and NEC, are reported in Table 5 for the cases of $s = 4$ (quartiles) and $s = 5$ (fixed cut-offs). As shown in Table 5, the BIC suggests two models for the data: one with $k = 4$ latent classes in the case of quartiles, and one with $k = 5$ latent classes in the case of fixed cut-offs. The AIC gives a similar result for the first type of discretization and exactly the same result for the second. The NEC suggests instead only two latent classes for both types of discretization.
Next we select the covariates on the basis of the parameterization (5) for the latent class probabilities and a stepwise forward covariates selection scheme based on the BIC; the results are displayed in Table 6. For both discretizations that were considered (quartiles and fixed cut-offs), the specification preferred includes a journal’s age and country, but not its language.

To check the unidimensionality assumption, we look for evidence of bidimensionality. Table 7 shows the values of the BIC for various specifications of bidimensionality. For instance, the first model considered, which is denoted by \{1\},\{2, 3, 4\}, assumes that the first indicator measures a latent trait that is different from that for the other three, whereas the fifth model considered, which is denoted by \{1, 2\} \{3, 4\}, assumes that the first two indicators measure a latent trait that is different from that for the other two. Under both discretizations, the smallest value of the BIC in Table 7 is greater than the smallest value of the BIC in Table 6. We conclude that the assumption of unidimensionality is consistent with the data, so our four indicators represent different measurements of the same latent trait.

Finally, we consider the assumption of local independence by fitting some extended models that include a set of association parameters for specific pairs of responses, as defined in expression (4), under Plackett’s formulation of local dependence. The results are presented in Table 8. For the first type of discretization (quartiles), there is evidence of conditional association (given the latent trait) between IF and IF5. In fact, the extended model that includes

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Results for quartiles</th>
<th></th>
<th>Results for fixed cut-offs</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( l(\hat{\theta}))</td>
<td>#par</td>
<td>BIC</td>
<td>( l(\hat{\theta}))</td>
</tr>
<tr>
<td>Age</td>
<td>–1259.0</td>
<td>26</td>
<td>2676.6</td>
<td>–1426.4</td>
</tr>
<tr>
<td>Language</td>
<td>–1269.7</td>
<td>22</td>
<td>2673.5</td>
<td>–1441.9</td>
</tr>
<tr>
<td>Country</td>
<td>–1203.4</td>
<td>23</td>
<td>2547.0</td>
<td>–1375.1</td>
</tr>
<tr>
<td>Age + country</td>
<td>–1180.0</td>
<td>28</td>
<td>2530.7</td>
<td>–1345.9</td>
</tr>
<tr>
<td>Language + country</td>
<td>–1203.3</td>
<td>24</td>
<td>2552.9</td>
<td>–1374.8</td>
</tr>
<tr>
<td>Age + language + country</td>
<td>–1179.1</td>
<td>29</td>
<td>2535.1</td>
<td>–1344.0</td>
</tr>
</tbody>
</table>

†The smallest values of the BIC in each column are in italics.

<table>
<thead>
<tr>
<th>Groups of items</th>
<th>Results for quartiles</th>
<th></th>
<th>Results for fixed cut-offs</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( l(\hat{\theta}))</td>
<td>#par</td>
<td>BIC</td>
<td>( l(\hat{\theta}))</td>
</tr>
<tr>
<td>{1}, {2, 3, 4}</td>
<td>–1178.7</td>
<td>30</td>
<td>2540.4</td>
<td>–1344.9</td>
</tr>
<tr>
<td>{2}, {1, 3, 4}</td>
<td>–1179.2</td>
<td>30</td>
<td>2541.4</td>
<td>–1345.9</td>
</tr>
<tr>
<td>{3}, {1, 2, 4}</td>
<td>–1179.9</td>
<td>30</td>
<td>2542.7</td>
<td>–1345.9</td>
</tr>
<tr>
<td>{4}, {1, 2, 3}</td>
<td>–1179.9</td>
<td>30</td>
<td>2542.7</td>
<td>–1345.0</td>
</tr>
<tr>
<td>{1, 2}, {3, 4}</td>
<td>–1179.7</td>
<td>30</td>
<td>2542.4</td>
<td>–1344.9</td>
</tr>
<tr>
<td>{1, 3}, {2, 4}</td>
<td>–1179.2</td>
<td>30</td>
<td>2541.4</td>
<td>–1345.0</td>
</tr>
</tbody>
</table>

†The smallest values of the BIC in each column are in italics.
this association has a lower BIC (namely 2500.1) compared with the model with local independence, and no association between other pairs of indicators needs to be considered. In contrast, for the second type of discretization (fixed cut-offs), there is evidence of conditional association both between IF and IF5 and between IF5 and AIS. However, although the larger model including these two associations and the model including only the association between IF and IF5 have a very similar BIC, the former seems to have some numerical instability in the estimation.

In conclusion, although the two types of discretization imply some differences in terms of local independence and some changes in the ranking of journals in terms of predicted impact, the IRT models that were selected by the procedure that was suggested in Section 3.2 are based on exactly the same features and differ only in the number of latent classes, namely $k = 4$ for the quartiles case and $k = 5$ for the fixed cut-offs case. The features that are common to both models are unidimensionality, local dependence only between IF and IF5, and latent distribution affected by two covariates: age of the journal and country of the publisher.

4.4. Latent distributions, clustering and prediction of journal impact

The maximum likelihood estimates that were obtained for the parameters of the selected models under the two different discretizations are presented in Tables 9–11. In particular, Table 9 shows the estimated distribution of the latent variable (support points and probabilities). Note that the support points correspond to classes of journals characterized by increasing impact, whereas the mass probabilities are averaged over all the sample units. Also note that we use the standardization of the latent variables as an identifiability constraint. The first discretization (quartiles) selects four classes of journals, the class with highest impact being the smallest (it includes 15.8% of the journals) and the class of journals with lowest impact being the largest (it includes 33.6% of the journals). The second discretization (fixed cut-offs) selects five classes, with the added class consisting of a relatively small number of journals with very low impact (9.8% of the total) and the top two classes having size and support that are comparable with the top two classes estimated under the first model.

Table 10 reports the estimated intercepts and regression coefficients for the covariates included, namely the age of the journal (the reference category is aged 41 or more years) and the
Table 9. Estimated distribution of the latent variable under the models selected

<table>
<thead>
<tr>
<th>h</th>
<th>Results for quartiles</th>
<th></th>
<th>Results for fixed cut-offs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \hat{\xi}_h )</td>
<td>( \hat{\pi}_h )</td>
<td>( \xi_h )</td>
</tr>
<tr>
<td>1</td>
<td>-1.158</td>
<td>0.336</td>
<td>-2.763</td>
</tr>
<tr>
<td>2</td>
<td>-0.096</td>
<td>0.317</td>
<td>-0.172</td>
</tr>
<tr>
<td>3</td>
<td>0.883</td>
<td>0.190</td>
<td>0.279</td>
</tr>
<tr>
<td>4</td>
<td>1.589</td>
<td>0.158</td>
<td>0.632</td>
</tr>
<tr>
<td>5</td>
<td>1.002</td>
<td>0.141</td>
<td></td>
</tr>
</tbody>
</table>

Table 10. Estimated regression coefficients in the proportional odds model for the latent class probabilities

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Results for quartiles</th>
<th>Results for fixed cut-offs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>Standard error</td>
</tr>
<tr>
<td>Intercept 1</td>
<td>3.506</td>
<td>0.444</td>
</tr>
<tr>
<td>Intercept 2</td>
<td>1.125</td>
<td>0.306</td>
</tr>
<tr>
<td>Intercept 3</td>
<td>-0.260</td>
<td>0.375</td>
</tr>
<tr>
<td>Intercept 4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age 1–3 years</td>
<td>-3.974</td>
<td>1.030</td>
</tr>
<tr>
<td>Age 4–6 years</td>
<td>-2.905</td>
<td>0.553</td>
</tr>
<tr>
<td>Age 7–10 years</td>
<td>-0.801</td>
<td>0.424</td>
</tr>
<tr>
<td>Age 11–20 years</td>
<td>-0.734</td>
<td>0.318</td>
</tr>
<tr>
<td>Age 21–40 years</td>
<td>-0.315</td>
<td>0.290</td>
</tr>
<tr>
<td>UK, D, NL</td>
<td>-0.553</td>
<td>0.256</td>
</tr>
<tr>
<td>Other countries</td>
<td>-3.650</td>
<td>0.395</td>
</tr>
</tbody>
</table>

country of the publisher (the reference category is published in the USA), in the proportional odds model (5) for the latent class probabilities. The estimated coefficients suggest that the probability of being in a higher class (i.e. having a higher impact) increases with a journal’s age and, compared with journals published in the USA, it is slightly lower for those published in the UK, Germany and the Netherlands, and is much lower for those published in all other countries.

Table 11 presents estimates of the discriminant indices \( \alpha_j \) in equation (2), which allow us to assess the quality of each indicator as a measure of the impact of a journal. Table 11 also presents the estimates of the difficulty parameters \( \beta_{jv} \). It is worth noting that we reach the same conclusion under both types of discretization: the indicator that best measures the impact of a journal is IF5 because it corresponds to the highest estimate of the discrimination parameter. Moreover, the estimates of the discrimination parameter for the other indicators are quite similar under both types of discretization. This agrees with the result in Chang et al. (2010), based on the ISI–Thomson Reuters database of citations from all fields in the sciences and social sciences, that AIS does not add much compared with more traditional indicators such as IF5.

We then assign each journal to a latent class by using the rules that were discussed in Section 2.3. Under the first type of discretization (quartiles), we assign 159 journals to the first class (lowest impact), 138 to the second, 79 to the third and 69 to the fourth (highest impact). The
Table 11. Estimated discriminant indices and difficulty parameters

<table>
<thead>
<tr>
<th>j</th>
<th>Results for quartiles</th>
<th>Results for fixed cut-offs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\hat{\alpha}_j$</td>
<td>$\hat{\beta}_{j1}$</td>
</tr>
<tr>
<td>1</td>
<td>3.124</td>
<td>0.000</td>
</tr>
<tr>
<td>2</td>
<td>5.132</td>
<td>0.709</td>
</tr>
<tr>
<td>3</td>
<td>3.234</td>
<td>0.614</td>
</tr>
<tr>
<td>4</td>
<td>3.523</td>
<td>−2.818</td>
</tr>
</tbody>
</table>

Fig. 2. Scatter plot of the predicted latent variables under the discretization based on (a) quartiles and (b) fixed cut-offs

journals that were assigned to the top two classes are listed in Tables 1 and 2, together with information of interest, such as the estimated posterior probabilities and the predicted impact under both models. Fig. 2 presents the scatter plot of the predicted impact of all journals. These results show that, although we are assuming a discrete distribution for the latent variable with a relatively small number of support points, there is a high variety of different predicted values, confirming our previous remarks in Section 2.3. This is because for some journals the estimated posterior probabilities are not strongly unbalanced towards a certain class, as confirmed by the plots in Fig. 3.
On the basis of the above journal clustering, Table 12 reports the average conditional correlation between the indicators given the latent class under the two types of discretization, together with the observed (marginal) correlation between the indicators. Note that the conditional correlation between the pairs of indicators for which we assume local independence is very low compared with both the observed correlation and the conditional correlation between the pair of indicators (IF and IF5) for which we admit local dependence. This provides indirect support for the assumptions underlying the models adopted.

5. Conclusions

We propose an approach to ranking scientific journals in a given list based on a latent variable model for polytomous item response data. The latent variables, which are assumed to be discrete and are interpreted as the unobserved scholarly impact of the journals in the list, are predicted on the basis of a set of bibliometric indicators that are discretized to avoid strong parametric assumptions.

Our approach has the advantage of relying on a well-defined statistical model, which simplifies the interpretation of the results and inference. The model is semiparametric in nature as it does not need parametric assumptions on the distribution of the latent variables. This is instead approximated by a discrete distribution with a finite number of support points corresponding to a set of latent classes, i.e. groups of journals with similar unobservable characteristics.
Table 12. Correlation matrix of the observed indicators and average of the conditional correlation given the assigned cluster

<table>
<thead>
<tr>
<th>Indicator</th>
<th>Results for the following indicators:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>IF</td>
</tr>
<tr>
<td><strong>Observed</strong></td>
<td></td>
</tr>
<tr>
<td>IF</td>
<td>1.000</td>
</tr>
<tr>
<td>IF5</td>
<td>0.899</td>
</tr>
<tr>
<td>AIS</td>
<td>0.692</td>
</tr>
<tr>
<td>h-index</td>
<td>0.556</td>
</tr>
<tr>
<td><strong>Conditional (quartiles)</strong></td>
<td></td>
</tr>
<tr>
<td>IF</td>
<td>1.000</td>
</tr>
<tr>
<td>IF5</td>
<td>0.724</td>
</tr>
<tr>
<td>AIS</td>
<td>0.147</td>
</tr>
<tr>
<td>h-index</td>
<td>0.108</td>
</tr>
<tr>
<td><strong>Conditional (fixed cut-offs)</strong></td>
<td></td>
</tr>
<tr>
<td>IF</td>
<td>1.000</td>
</tr>
<tr>
<td>IF5</td>
<td>0.671</td>
</tr>
<tr>
<td>AIS</td>
<td>0.120</td>
</tr>
<tr>
<td>h-index</td>
<td>0.010</td>
</tr>
</tbody>
</table>

Journals are assigned to a latent class on the basis of a maximum a posteriori probability rule. The posterior mean of the latent variable provides a prediction on a single continuous scale of the impact of each journal, so journals can be ranked, the distance between any pair of them can be compared and journals can be clustered into any arbitrary number of classes of a given size. One can also assess the discriminant power of each indicator, i.e. the sensitivity and reliability of each indicator in its relationship to the latent variable. For example, in our data we find that IF5 is more reliable than IF, AIC and the h-index as an indicator of the impact of a journal. Finally, our approach has the advantage of easily handling missing values in the available indicators.

A key aspect of our approach is that it requires a preliminary discretization of the available bibliometric indicators. It is therefore important to assess the sensitivity of the results of an analysis to the discretization adopted. As shown in our application, this can be assessed by replicating the analysis with different discretizations and then comparing the results that are obtained. In our empirical application, the results appear to be fairly robust to the choice of discretization.

Our analysis can be easily extended to handle the case where different indicators have different numbers of categories (to include, for instance, binary indicators), and to employ discriminant indices which are category dependent. Finally, our method could be applied to other lists of journals in different fields.

Acknowledgements

We thank Sergio Benedetto, Tullio Jappelli, Daniele Terlizzese, the Associate Editor, two referees and seminar participants at the Einaudi Institute for Economics and Finance 2013 and the 5th Italian Congress of Econometrics and Empirical Economics for helpful comments.
References


