Choosing the Number of Factors in Independent Factor Analysis Model

Cinzia Viroli[1]

Abstract

Independent Factor Analysis (IFA) has recently been proposed in the signal processing literature as a way to model a set of observed variables through linear combinations of hidden independent ones plus a noise term. Despite the peculiarity of its origin the method can be framed within the latent variable model domain and some parallels with the ordinary Factor Analysis can be drawn. If no prior information on the latent structure is available a relevant issue concerns the correct specification of the model. In this work some methods to detect the number of significant latent variables are investigated. Moreover, since the method defines a probability density function for the latent variables by mixtures of gaussians, the correct number of mixture components must also be determined. This issue will be treated according to two main approaches. The first one amounts to carry out a likelihood ratio test. The other one is based on a penalized form of the likelihood, that leads to the so called information criteria. Some simulations and empirical results on real data sets are finally presented.

1   Introduction

Independent Factor Analysis (IFA) has recently been introduced by Attias (1999) in the context of signal processing as a way to model a set of observed variables through linear combinations of hidden independent ones plus a noise term. Despite the peculiarity of its origin Independent Factor Analysis is indeed a generative latent variable model (Montanari and Viroli, 2005) whose structure closely resembles the one of ordinary factor model but it assumes that the latent variables are mutually independent and not necessarily Gaussians. The assumption of *non-gaussianity* of the factors represents the most appealing idea of this approach with respect to the most common latent variable models that are instead based on normally distributed latent variables. In the IFA model the probability density function of the factors is assumed to be a mixture of Gaussians. This choice allows to model arbitrary probability density functions.

Implicit in the IFA estimation problem are the assumptions regarding the number of significant common factors and the number of mixture components for modeling each factor. The aim of this paper is to propose some methods to detect the correct specification of the IFA model. In doing this, we have moved from the more traditional approaches: the likelihood ratio statistics and some information criteria. The methods discussed are illustrated using simulated and empirical data sets.

---

[1] Statistics Department, University of Bologna; cinzia.viroli@unibo.it

2   Independent Factor Analysis

The aim in Independent Factor Analysis is to describe p observed variables x_j, which are generally correlated, in terms of a smaller set of k unobserved independent latent variables y_i and an additive specific term u_j:

$$x_j = \sum_{i=1}^{k} \lambda_{ji} y_i + u_j,$$

where $j = 1, ..., p$, $i = 1, ..., k$. In a more compact form the model is

$$\mathbf{x} = \Lambda \mathbf{y} + \mathbf{u} \tag{2.1}$$

where the factor loading matrix $\Lambda = \{\lambda_{ji}\}$ is also termed as *mixing matrix*. Its structure closely resembles the classical factor model but it differs from it as far as the properties of the latent variables it involves is concerned. The random vector $\mathbf{u}$ representing the noise is assumed to be normally distributed, $\mathbf{u} \sim \mathcal{N}(\mathbf{0}, \Psi)$ with Ψ allowing for correlations between the error terms. The latent variables $\mathbf{y}$ are assumed to be mutually independent and not necessarily normally distributed; their probability density functions are indeed modeled as mixtures of Gaussians. The independence assumption allows to model the density of each y_i in the latent space separately. In more formal terms each factor is thus described as a mixture of m_i Gaussians with mean μ_{i,q_i}, variance ν_{i,q_i} and mixing proportions w_{i,q_i} ($q_i = 1, ..., m_i$):

$$f(y_i) = \sum_{q_i=1}^{m_i} w_{i,q_i} \mathcal{N}\left(\mu_{i,q_i}, \nu_{i,q_i}\right) \tag{2.2}$$

The mixing proportions w_{i,q_i} are constrained to be non-negative and sum to unity.

A particular characterization of the IFA model is that it involves two layers of latent variables: besides the factors, $\mathbf{y}$, an *allocation variable*, $\mathbf{z}$, must be introduced, as always when dealing with mixture models. With reference to a particular factor i, the mixture can be thought of as the density of an heterogeneous population consisting of m_i subgroups. For each observation the allocation variable denotes the identity of the subgroup from which it is drawn. In the k-dimensional space, the multivariate allocation variable, $\mathbf{z}$, follows a multivariate multinomial distribution. The density of the observed data can be constructed by conditioning to these two latent layers:

$$
\begin{aligned}
f(\mathbf{x}|\Theta) &= \sum_{\mathbf{z}} \int f(\mathbf{x},\mathbf{y},\mathbf{z}|\Theta) d\mathbf{y} \\
&= \sum_{\mathbf{z}} \int f(\mathbf{z}|\Theta) f(\mathbf{y}|\mathbf{z},\Theta) f(\mathbf{x}|\mathbf{y},\mathbf{z},\Theta) d\mathbf{y} \\
&= \sum_{\mathbf{z}} f(\mathbf{z}|\Theta) f(\mathbf{x}|\mathbf{z},\Theta)
\end{aligned}
\tag{2.3}
$$

where Θ denotes the whole set of the IFA model parameters.

It is not difficult to derive that the conditional density $f(\mathbf{x}|\mathbf{z}, \Theta)$ follows a Gaussian distribution since it is the convolution of two Gaussians:

$$f(\mathbf{x}|\mathbf{y}, \mathbf{z}, \Theta) = \mathcal{N}(\Lambda \mathbf{y}, \Psi) \tag{2.4}$$

and

$$f(\mathbf{y}|\mathbf{z}, \Theta) = \mathcal{N}(\boldsymbol{\mu}_{\mathbf{z}}, \mathbf{V}_{\mathbf{z}}) \tag{2.5}$$

where $\boldsymbol{\mu}_{\mathbf{z}}$ and $\mathbf{V}_{\mathbf{z}}$ are respectively defined as:

$$\boldsymbol{\mu}_{\mathbf{z}} = \left[\prod_{q_1=1}^{m_1} \mu_{1,q_1}^{z_{1,q_1}}, ..., \prod_{q_k=1}^{m_k} \mu_{k,q_k}^{z_{k,q_k}} \right] \qquad \mathbf{V}_{\mathbf{z}} = \mathrm{diag} \left[\prod_{q_1=1}^{m_1} \nu_{1,q_1}^{z_{1,q_1}}, ..., \prod_{q_k=1}^{m_k} \nu_{k,q_k}^{z_{k,q_k}} \right].$$

Therefore the expression (2.3) indicates that the density of the observed data given the IFA model, *i.e.* the likelihood function $f(\mathbf{x}|\Theta)$, is a finite mixture of p-variate normals. Its generic component is given by

$$f(\mathbf{x}|\mathbf{z}, \Theta) = \mathcal{N}\left(\Lambda \boldsymbol{\mu}_{\mathbf{z}}, \Lambda \mathbf{V}_{\mathbf{z}} \Lambda^T + \Psi \right) \tag{2.6}$$

3 The IFA model estimation

The parameter estimates in the IFA model can not be obtained by maximizing directly the likelihood function, $f(\mathbf{x}|\Theta)$, obtained in the (2.3) since this quantity is analytically hard to deal with.

In such situations, it is generally convenient resorting to some additional unobserved variables, such that if they *were known* the optimal value Θ could be computed easily. In the IFA model the introduction of the two hidden layers (the factors $\mathbf{y}$ and the multivariate allocation variable $\mathbf{z}$) allows to rephrase the likelihood function of the observed data (the "incomplete" data) in terms of the "pseudo-complete" data density:

$$f(\mathbf{x}|\Theta) = \sum_{\mathbf{z}} \int f(\mathbf{z}, \mathbf{y}, \mathbf{x}|\Theta)\, d\mathbf{y}, \tag{3.1}$$

with

$$f(\mathbf{x}, \mathbf{y}, \mathbf{z}|\Theta) = f(\mathbf{x}|\mathbf{y}, \Theta) f(\mathbf{y}|\mathbf{z}, \Theta) f(\mathbf{z}|\Theta), \tag{3.2}$$

The objective of maximizing the incomplete log-likelihood $\ln f(\mathbf{x}|\Theta)$ can be now achieved by the EM algorithm: the incomplete-data likelihood problem is solved indirectly by proceeding iteratively in terms of the complete-data log-likelihood function $\ln f(\mathbf{z}, \mathbf{y}, \mathbf{x}|\Theta)$. As it is unobservable it is replaced by its conditional expectation given the observable data, using the current fit for the parameters, Θ':

$$\arg\max_{\Theta} E_{\mathbf{z}, \mathbf{y}|\mathbf{x}, \Theta'} \left[\ln f(\mathbf{z}, \mathbf{y}, \mathbf{x}|\Theta) \right].$$

By making use of the previous decomposition, the two steps of the EM algorithm can be analytically derived. After laborious but straightforward calculations (for more details

see Montanari and Viroli, 2005), the following estimates for the new parameters Θ in terms of the old ones Θ' are obtained:

$$\hat{\Lambda} = \mathbf{x}E[\mathbf{y}^T|\mathbf{x}]E\left[\mathbf{y}\mathbf{y}^T|\mathbf{x}\right]^{-1} \qquad \hat{\Psi} = \mathbf{x}\mathbf{x}^T - \mathbf{x}E[\mathbf{y}^T|\mathbf{x}]\Lambda^T \qquad (3.3)$$

$$\hat{\mu}_{i,q_i} = \frac{f(z_i|\mathbf{x})E[y_i|z_i,\mathbf{x}]}{f(z_i|\mathbf{x})} \quad \hat{\nu}_{i,q_i} = \frac{f(z_i|\mathbf{x})E[y_i^2|z_i,\mathbf{x}]}{f(z_i|\mathbf{x})} - \mu_{i,q_i}^2 \quad \hat{w}_{i,q_i} = f(z_i|\mathbf{x}).$$

The EM algorithm has the appealing property to converge monotonically to a local maximum of the log-likelihood.

The number of parameters to be estimated is a consequence of an implicit assumption regarding the number of significant common factors and the number of mixture components for each. Assessing the correct specification of the model is thus an important phase of the exploratory analysis. As far as the number of mixture components is concerned, its increase improves the factor density approximation, but also involves a remarkable increase in the total parameters of the model, in contrast with the empirical principle of parsimony. This issue is consequent and strictly connected with another relevant aspect: the detection of the correct number of factors.

The identification of the latent structure dimension has been traditionally achieved according to two main approaches. The first one is to carry out a hypothesis test, using the likelihood ratio as the test statistic. The other one is based on a penalized form of the likelihood. As the likelihood increases with the addition of a factor, the likelihood is "penalized" by the subtraction of a term that considers the number of parameters in the model. This leads to the so called information criteria. These two approaches will be presented in the next sections.

4   Likelihood ratio test statistic (LRTS)

The conventional application of the LRTS to test the null hypothesis $H_0 : k = k_0$ versus $H_1 : k = k_1$ for some $k_1 > k_0$ is not possible in the IFA model, since the classic regularity conditions do not hold for the transformed likelihood ratio $-2\log\lambda$ to have its usual asymptotic null distribution of chi-squared. In order to demonstrate it, one should observe that the likelihood function $f(\mathbf{x}|\Theta)$ in the IFA model can be rephrased as a finite mixture of p-variate normals, as shown in the expression (2.3). The classic regularity conditions are not fulfilled for finite mixture models, because under the null hypothesis the last $k_1 - k_0$ mixing proportions lie on the boundary of the parameter space (under H_0 they are all zero).

Several simulation results have been reported in literature on the null distribution of the LRTS when standard regularity conditions do not hold. In particular, Wolfe (1971), on the basis of a MonteCarlo study, suggested a modified test for a finite mixture of p-variate normals in which

$$-\frac{2}{n}\left(n - 1 - p - \frac{1}{2}k_1\right)\log\lambda \sim \chi_g^2 \qquad (4.1)$$

is distributed as a chi-square with degrees of freedom $g = 2p(k_1 - k_0)$.

In the IFA model, the number of mixture components, m, is identified by the domain dimension of the multivariate allocation variable **z**, that is

$$m = \prod_{i=1}^{k} m_i \tag{4.2}$$

where k is the number of factors and m_i is the number of mixture components for each factor. As clear from the expression (4.2), testing an hypothesis concerning the number of factors is equivalent to assessing an hypothesis concerning the number of mixture components which depends on k_0

$$H_0 : k = k_0 \quad \Rightarrow \quad H_0 : m = m_0 = \prod_{i=1}^{k_0} m_i. \tag{4.3}$$

Moreover using the (4.2) it is also possible to perform a test in order to assess the number of mixture components for a specific factor in the same spirit of the (4.3). Adopting a more complex strategy with a series of tests the correct specification of the IFA model can thus be completely investigated.

From an extensive simulation study we derived that Wolfe's approximation for the IFA model gives satisfactory results only if the sample size is adequately large, at least about $n > 10h$ where h is the number of free parameters (Viroli, 2003). Unfortunately the IFA model is characterized by the use of a large number of unknown parameters: with p observed variables and k factors the free parameters are $h = pk + p + (3 \sum_{i=1}^{k} m_i - k)$. As a consequence the applicability of Wolfe's solution is limited, since it requires more observations than those generally available in the empirical context.

An alternative way to carry out a hypothesis test on the number of factors could be to approximate the null distribution of the LRTS by boostrapping the data. However, it must be pointed out that the implementation of the boostrapped LRTS on the IFA model is actually computationally prohibitive.

5   Information criteria

Two commonly-used information criteria for the comparison and selection between different models are the Akaike's Information Criterion (AIC; Akaike, 1973) and the Bayesian Information Criterion (BIC; Schwarz, 1978). Akaike's information criterion is constructed on the log-likelihood

$$AIC = -2 \log \max L + 2h$$

where h is the number of the model free parameters. The second term of the criterion gives a growing penalty in correspondence of an increasing number of factors, according with the parsimony principle. With respect to the AIC, Schwarz's information criterion is characterized by a higher penalty term since it involves also the sample dimension n

$$BIC = -2 \log \max L + h \log n.$$

Table 1: Values of the root mean squared error and the bias for BIC and AIC.

$p=7$		BIC			AIC		
		$k=1$	$k=2$	$k=3$	$k=1$	$k=2$	$k=3$
		$(h=22)$	$(h=37)$	$(h=52)$	$(h=22)$	$(h=37)$	$(h=52)$
$n=100$	Bias	0.2	-0.9	-1.8	0.6	-0.9	-1.1
	MSE	0.6	1.0	1.9	1.0	1.0	1.3
$n=200$	Bias	0.5	-0.8	-1.2	0.7	-0.5	-0.6
	MSE	0.8	0.9	1.4	1.0	0.8	0.8
$n=500$	Bias	0.6	-0.5	-0.7	0.6	-0.2	-0.5
	MSE	0.9	0.9	0.9	0.9	0.9	0.7

Table 2: Values of the root mean squared error and the bias for BIC when $p=10$.

$p=10$		BIC					
		$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$
		$(h=28)$	$(h=46)$	$(h=64)$	$(h=82)$	$(h=100)$	$(h=118)$
$n=100$	Bias	0.2	-0.8	-1.6	-2.4	***	***
	MSE	0.5	0.9	1.8	2.6	***	***
$n=200$	Bias	0.7	-0.5	-0.7	-1.8	-1.9	-3.9
	MSE	1.1	0.7	1.2	2.2	2.3	4.2
$n=500$	Bias	1.0	0.2	1.3	1.0	0.2	-1.3
	MSE	1.4	1.0	1.7	1.3	0.7	1.7

The performances of the two criteria in the IFA model have been analyzed through a wide simulation study performed by using several initialization of the parameters in the EM-algorithm.

The situations considered differ according to the sample sizes n=100, 200, 500, the values $p = 7$ and $p = 10$ of the number of observed variables, the number of latent factors $k = 1, \ldots, K$ factors, where K is the maximum number of estimable latent variables identified by the Lederman's condition

$$k \leq \frac{1}{2}\{2p + 1 - \sqrt{8p + 1}\}.$$

Each factor is here modelled by a constant number of mixture components, with $m_i = 3 \ \forall i$. For each of the 50 simulations, the root mean squared error (MSE) and the bias ($Bias$) were calculated:

$$MSE = \left(\frac{1}{50}\sum_{i=1}^{50}(k_i^* - k)^2\right)^{\frac{1}{2}} \qquad Bias = \frac{\sum_{i=1}^{50}k_i^*}{50} - k.$$

The results are summarized in the Tables 1, 2 and 3. The BIC criterion generally shows a negative bias for both $p = 7$ and $p = 10$ because of its high penalty term but this

Table 3: Values of the root mean squared error and the bias for AIC when p=10.

p=10		AIC					
		$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$
		($h=28$)	($h=46$)	($h=64$)	($h=82$)	($h=100$)	($h=118$)
n=100	Bias	0.8	-0.2	0.3	-0.3	***	***
	MSE	1.4	0.8	1.2	1.3	***	***
n=200	Bias	1.4	0.7	2.3	1.4	0.4	-1.0
	MSE	1.9	1.2	2.5	1.5	0.7	1.3
n=500	Bias	1.7	1.3	2.6	1.7	0.6	-0.6
	MSE	2.2	1.7	2.7	1.8	0.8	0.8

tendency to underestimate the number of factors decreases as the sample size n increases. When compared to BIC the penalty term of AIC penalizes complex models less heavily, since its penalty term does not depend on the sample size. In fact, the Akaike's criterion is more favourable to factor inclusion but it tends to fit too many factors for the case $p = 10$, and this tendency gets worse as n increases. Although the joint results of the two information criteria seem to offer appreciable indications on the correct number of factors, the appearance of improper solutions is not in principle excluded, because as explained by Titterington et al. (1985) they rely essentially on the same regularity conditions needed for the $-2 \log \lambda$ to have its usual asymptotic distribution under the null hypothesis.

6 Empirical results

6.1 Thyroid data

The data are taken from the study by Coomans *et al.* (1983) on the tyroid disease. The example consists of 5 measurements (T3-resin uptake test, Total Serum thyroxin, Total serum triiodothyronine, Basal thyroid-stimulating hormone and maximal absolute difference of TSH value after injection of 200 micro grams of thyrotropin-releasing hormone as compared to the basal value) on 215 patients, that are distinguished in three groups on the basis of their thyroid status (normal, hyper and hypo).

Before performing the IFA estimation, the previous criteria has been applied in order to detect the correct specification of the model. The AIC and BIC criteria indicate that two factors are enough to describe the latent space. The application of the likelihood ratio to detect the number of components for each factor finally lead to two factors with two components each.

The research of independent factors with distribution non necessarily gaussian offers some advantages with respect to the ordinary Factor Analysis solution, since the description of the latent space could be sometimes warped by the assumption of gaussianity. It the case of this example. The first graph of Figure 1, shows the scatter plot of the factor scores in the ordinary Factor Analysis obtained by the well known iterated principal factor method. The second graph of the Figure 1 shows instead the IFA solution. It seems more interesting. In fact, the first factor clearly captures the variability of the normal and hyper
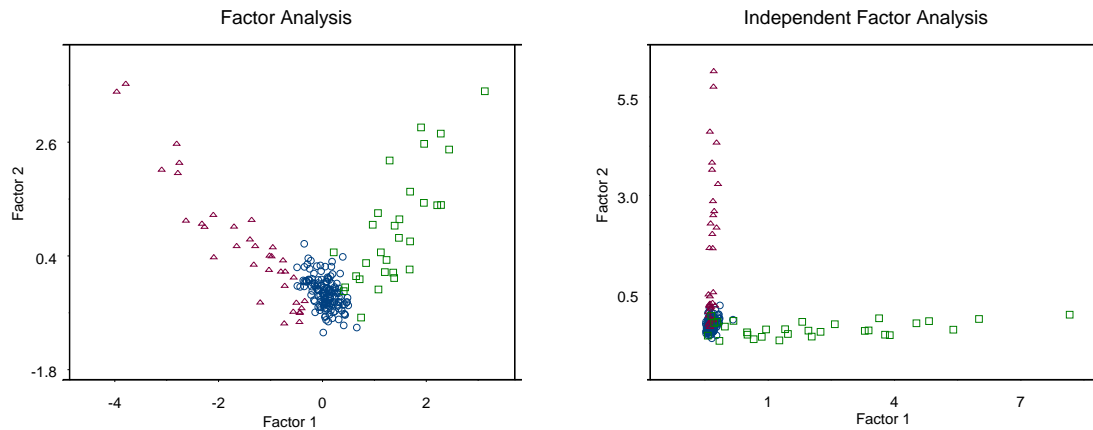
Figure 1: Scatter plot of the factor scores in ordinary Factor Analysis and Independent Factor Analysis. In the graphs the circles denote the normal group of patients, the triangles indicates the hypo group of patients and the boxes the hyper group.

groups of patients while the second one describes those of the patients with normal and hypo thyroid.
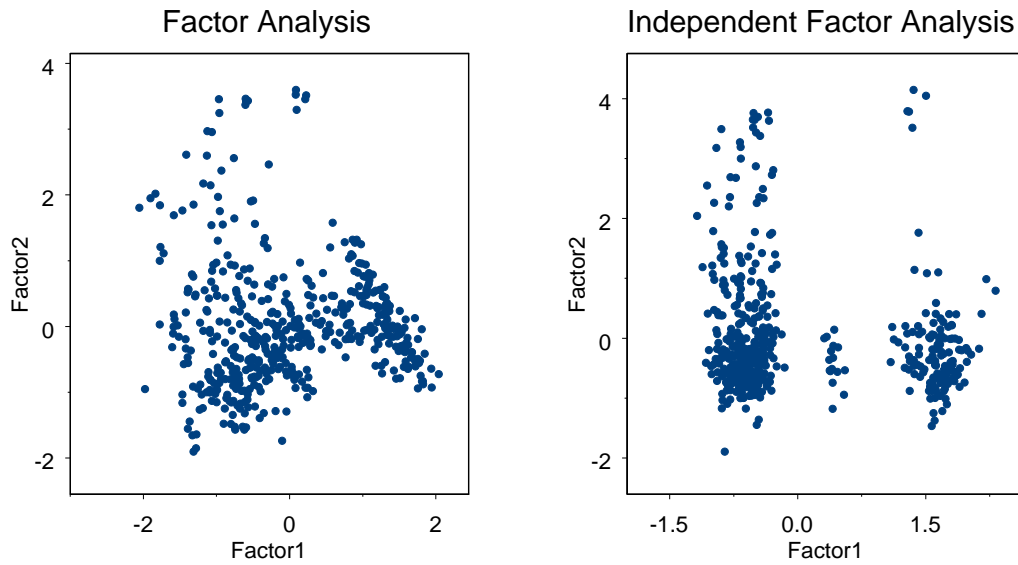
6.2   Boston neighborhood data

The data set has been entirely published in Belsley *et al.* (1980). Each observation is a standard metropolitan tract in the Boston area. For each zone 13 summary variables have been measured: median value of owner-occupied homes; per capita crime rate by town; proportion of residential land zoned for big lots; proportion of non-retail business acres per town; nitric oxides concentration; average number of rooms per dwelling; proportion of owner-occupied units built prior to 1940; weighted distances to five Boston employment centres; index of accessibility to radial highways; full-value property-tax rate; pupil-teacher ratio by town; proportion of blacks by town; fraction of lower status of the population.

Exploratory Projection Pursuit (Friedman, 1987) has showed striking structure of the data and non gaussian projected directions.

In Figure 2 the ordinary Factor Analysis solution and the Independent Factor Analysis one are represented. The specification of the IFA model has been accomplished by a two step forward procedure. In the first step the likelihood ratio test statistics has been applied to determine the number of factors, with a fixed number of mixture components. In the second step, with fixed $k = 2$ factors the number of mixture components has been chosen applying the LRTS test in a forward strategy. This procedure has indicated a model with two factors both of them modeled by 3-dimensional gaussian mixtures. The IFA solution exhibits a clear clustering structure of the data. These groups are not captured by the ordinary factor analysis solution, although the differences in the factor loading estimates are not so relevant (Table 4).

Figure 2: Boston Neighborhood Data: Factor Analysis and Independent Factor Analysis solutions.

Table 4: Factor loading estimates of IFA, FA unrotated solution and FA with varimax rotation.

	IFA		FA		Varimax	
MEDV	-0.36	0.82	-0.69	0.65	-0.21	0.93
CRIM	0.67	0.03	0.57	-0.01	0.47	-0.33
ZN	-0.29	0.18	-0.58	-0.11	-0.54	0.24
INDUS	0.64	-0.16	0.84	0.13	0.77	-0.36
NOX	0.73	-0.04	0.83	0.29	0.85	-0.22
RM	-0.20	0.61	-0.50	0.51	-0.12	0.70
AGE	0.50	0.00	0.74	0.25	0.75	-0.20
DIS	-0.55	-0.09	-0.76	-0.43	-0.87	0.07
RAD	0.97	0.03	0.74	0.10	0.67	-0.33
TAX	0.90	-0.05	0.81	0.07	0.71	-0.39
PTRATIO	0.37	-0.33	0.49	-0.25	0.26	-0.48
B	-0.50	0.04	-0.45	-0.01	-0.38	0.25
LSTAT	0.52	-0.34	0.78	-0.28	0.49	-0.67

Conclusions

In this paper the Independent Factor Analysis has been introduced and some criteria have been employed in order to detect the correct specification of the model. The proposed methods seem to perform quite well both on the simulated and real data. However some issues are still open and the problem deserves further attention.

In fact, the approximation for the likelihood ratio test statistics gives satisfactory re-

sults only if the sample size is adequately large, at least about $n > 10h$ where h is the number of free parameters. The boostrap methods could offer a valid strategy to better approximate the null distribution of the LRTS, although it is computationally intensive.

Moreover the information criteria do not offer a strategy for the choice of the number of mixture components. Also they rely essentially on the same regularity conditions needed for the $-2 \log \lambda$ to have its usual asymptotic distribution under the null hypothesis and therefore despite their easy and immediate applicability they are not robust enough to violations of standard assumptions.

Some future research could be developed in the direction of the bayesian approach which provides a natural framework for considering the case where the number of components k is unknown. By allowing k to vary with the other parameters and specifying their joint prior distribution $f(k, \boldsymbol{\theta})$, inference may be based on the posterior distribution $f(k, \boldsymbol{\theta}|\mathbf{x})$. In this spirit, an interesting approach is based on the reversible jump Markov chain Monte Carlo (Green, 1995), that is capable of jumping between the parameter subspaces of differing dimensionality.

References

[1] Attias, H. (1999): Independent factor analysis. *Neural Computation*, **11**, 803-851.

[2] Bartholomew, D.J. and Knott, M. (1998): *Latent Variable Models and Factor Analysis*. New York: Oxford University Press.

[3] Belsley, D.A., Kuh, E., and Welsh, R.E. (1980): *Regression Diagnostics*, New York: John Wiley.

[4] Bentler, P.M. (1980): Multivariate analysis with latent variables. *Annual Review of Psychology*, **31**, 419-416.

[5] Coomans, D., Broeckaert, M. and Broeckaert, D.L. (1983): Comparison of multivariate discriminant techniques for clinical data – application to the tyroid functional state. *Meth. Inform. Med.*, **22**, 93-101.

[6] Dempster, N.M., Laird, A.P., and Rubin, D.B. (1977): Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society B*, **39**, 1-38.

[7] Friedman, J. (1987): Exploratory projection pursuit, *Journal of the American Statistical Association*, **82**, 249-266.

[8] Green, P.J. (1995): Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, **82**, 711-732.

[9] Harman, H. (1960): *Modern Factor Analysis*, Chicago: The University of Chicago Press.

[10] Hyvärinen, J., Karhunen, A., and Oja, E. (2001): *Independent Component Analysis*, John Wiley & Sons.

[11] Jöreskog, K.G. (1973): A general method for estimating a Linear Structural Equation System. In Goldberger, A.S. and Duncan O.D. (Eds.): *Structural Models in the Social Sciences*. New York: Academic Press.

[12] Montanari, A. and Viroli, C. (2005): The Independent Factor Analysis approach to latent variable modeling. *Submitted*.

[13] Redner, R.A. and Walker, H.F. (1984): Mixture densities, maximum likelihood and the EM algorithm. *SIAM Review*, **26**, 195-239.

[14] Sundberg, R. (1976): An iterative method for solution of the likelihood equations for incomplete data from exponential families. *Comm. Statist.-Simula. Computa.*, **B5**, 55-64.

[15] Titterington, D., Smith, A.F., and Makov, U.E. (1985): *Statistical Analysis of finite mixture distributions*. Great Britain: John Wiley & Sons.

[16] Viroli, C. (2003): *Latent Structure Exploratory Analysis based on Independent Component and Independent Factor approaches*, Ph.D. Dissertation, University of Bologna, Italy.

[17] Wolfe, J. (1971): A montecarlo study of the sampling distribution of the likelihood ratio for mixtures of multinormal distributions. *Technical Bullettin, STB*. **72** San Diego: Naval Personnel and Training Research Laboratory.