Stat Comput (2012) 22:647–659
DOI 10.1007/s11222-011-9262-z

Isotone additive latent variable models

Sylvain Sardy · Maria-Pia Victoria-Feser

Received: 4 May 2010 / Accepted: 29 May 2011 / Published online: 29 June 2011

Abstract For manifest variables with additive noise and for a given number of latent variables with an assumed distribution, we propose to nonparametrically estimate the association between latent and manifest variables. Our estimation is a two step procedure: first it employs standard factor analysis to estimate the latent variables as theoretical quantiles of the assumed distribution; second, it employs the additive models' backfitting procedure to estimate the monotone nonlinear associations between latent and manifest variables. The estimated fit may suggest a different latent distribution or point to nonlinear associations. We show on simulated data how, based on mean squared errors, the nonparametric estimation improves on factor analysis. We then employ the new estimator on real data to illustrate its use for exploratory data analysis.

Keywords Factor analysis · Principal component analysis · Nonparametric regression · Bartlett's factor scores · Dimension reduction

1 Introduction

Latent variable models are widely used in the social sciences for studying the interrelationships among observed variables. More specifically, latent variable models are used for reducing the dimensionality of multivariate data, for assigning scores to sample members on the latent dimensions identified by the model as well as for the construction of measurement scales, for instance in educational testing and psychometrics. They are therefore very important in practical data analysis.

The simplest latent variable model is factor analysis (Spearman 1904; Jöreskog 1969; Bartholomew 1984) (FA) which assumes multivariate normal data. The FA model states that for a set of manifest variables $\boldsymbol{x}$ that are correlated, it is possible to construct a smaller space made of the latent variables $\boldsymbol{z}$, such that conditionally on $\boldsymbol{z}$, there is no more correlation between the manifest variables. More precisely, given a vector of observed (manifest) variables $\boldsymbol{x} = (x^{(1)}, \dots, x^{(p)})^{\mathrm{T}}$, it is supposed that there exists a vector of size $q < p$ of (unobserved) latent variables $\boldsymbol{z} = (z^{(1)}, \dots, z^{(q)})^{\mathrm{T}}$ such that the manifest variables $x^{(j)}$ are linked to the latent variables through the additive noise model

$$x^{(j)} \mid \boldsymbol{z} = \eta_j(\boldsymbol{z}) + \epsilon^{(j)}, \quad j = 1, \dots, p, \tag{1}$$

with parametric linear associations

$$\eta_j(\boldsymbol{z}) = \alpha_{j0} + \boldsymbol{\alpha}_j^{\mathrm{T}} \boldsymbol{z}, \tag{2}$$

where $\boldsymbol{\alpha}_j = (\alpha_{j1}, \dots, \alpha_{jq})$, $j = 1, \dots, p$. If the model holds, i.e., the correlations between the manifest variables are explained by a smaller number of latent variables, then the manifest variables are conditionally independent given the latent ones. This means that $\epsilon^{(j)}$, $j = 1, \dots, p$, in (1) are independent, and it is usually assumed that $\boldsymbol{\epsilon} = (\epsilon^{(1)}, \dots, \epsilon^{(p)})^{\mathrm{T}} \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Psi})$ with $\boldsymbol{\Psi} = \mathrm{diag}(\psi_1, \dots, \psi_p)$. Note that the elements ψ_j are called *uniquenesses*, and the slopes $\boldsymbol{\alpha}_j^{\mathrm{T}}$ are called *factor loadings*. Finally Gaussian independent distribution is assumed for $\boldsymbol{z} \sim \mathrm{N}(\mathbf{0}, \boldsymbol{I})$, and $\mathrm{cov}(z^{(l)}, \epsilon^{(j)}) = 0$ for all j and l. Hence the FA model satisfies $E[\boldsymbol{x} \mid \boldsymbol{z}] = \boldsymbol{\alpha}_0 + \boldsymbol{\alpha}\boldsymbol{z}$ and $\mathrm{Var}(\boldsymbol{x}) = \boldsymbol{\alpha}\boldsymbol{\alpha}^{\mathrm{T}} + \boldsymbol{\Psi}$, with $\boldsymbol{\alpha}_0 = [\alpha_{j0}]^{\mathrm{T}}_{j=1,\dots,p}$ and $\boldsymbol{\alpha} = [\boldsymbol{\alpha}_j]^{\mathrm{T}}_{j=1,\dots,p}$. The FA model's

S. Sardy (✉) · M.-P. Victoria-Feser
University of Geneva, Geneva, Switzerland
e-mail: Sylvain.Sardy@unige.ch

parameters $\boldsymbol{\Psi}$ and $\boldsymbol{\alpha}_j$ can be estimated using the maximum likelihood estimator (MLE) based on the sample covariance matrix $\boldsymbol{S}$ of the manifest variables which is supposed to have a Whishart distribution (Jöreskog 1967). Generalized least squares estimation is an alternative (Browne 1984). Latent scores $\widehat{z}_i$ can also be estimated using for example Bartlett's factor scores given by

$$z_i = \left(\boldsymbol{\alpha}'\boldsymbol{\Psi}^{-1}\boldsymbol{\alpha}\right)^{-1}\boldsymbol{\alpha}'\boldsymbol{\Psi}^{-1}\boldsymbol{x}_i$$

with $\boldsymbol{\alpha}$ and $\boldsymbol{\Psi}$ replaced by their estimates (for example, see Mardia et al. 1979, p. 274).

Applications of FA are numerous. In psychometrics for instance, FA is used to construct measurement scales: given the answers to questionnaire items (manifest variables), the scores of the participants are reduced to scores on the latent variables for a better understanding of the phenomenon under investigation. Recent publications include Donncha et al. (2008) who examine the structure of the statistics anxiety rating scale among UK psychology students and Boelen et al. (2008) who study the factor structure of posttraumatic stress disorder symptoms.

An important issue is the number of latent variables q which is, in general, unknown and needs to be estimated in practice; see for instance Bartholomew and Knott (1999) and Skrondal and Rabe-Hesketh (2004) as general references and Conne et al. (2010) for a recent treatment of model selection in generalized FA. Selection of q is a separate issue that we do not treat here.

It should also be noted that FA, as a dimension reduction method, has close links to another well-known dimension reduction method called principal component analysis. Tipping and Bishop (1999) propose probabilistic PCA which results as a special case of FA in which the uniquenesses are all equal, i.e., $\boldsymbol{\Psi} = \psi \boldsymbol{I}$. PCA and its extensions are used in many applications such as pattern recognition (see e.g. Dryden et al. 2009), chemometrics, and biomedical studies.

In practice, a (linear) FA model does not always fit in a "reasonable way" the data, as may be revealed by diagnostic plots. Departures from the assumption are of two types: functional or structural. The first is concerned with small model deviations, such as outliers in the data set, that can be dealt with using robust statistics; see for instance Yuan and Bentler (1998) and Dupuis Lozeron and Victoria-Feser (2010) for robust methods in confirmatory FA, Moustaki and Victoria-Feser (2006) for robust inference in generalized linear latent variable models, and Mavridis and Moustaki (2009) for diagnostic methods for FA with binary data.

Structural departure from the assumptions is the one we are interested in. It is concerned with wrong model specification, which can take the following forms:

- the latent variables are not normal,
- the relationship between the manifest and the latent variables is nonlinear,
- a combination of the two.

Since the latent variables are not observed, it is difficult to distinguish between the two types of violations. Indeed, a non-normal latent variable could be transformed into a normal one (by means of the normal quantiles on the corresponding order statistics) and the linear relationship between the manifest and latent variables changed into a nonlinear one. In the other case, however, a nonlinear relationship cannot always be made linear by changing the distribution of the normal latent variable, unless the transformation from the nonlinear to the linear relationship is the same across all manifest variables. Since the choice of the distribution for the latent variables is arbitrary (at least for the interpretation of the results, see for instance Bartholomew 1988), it can be fixed to be normal so that when the FA model does not fit the data at hand, a more flexible model can be chosen that allows for nonlinear relationships between the manifest and the latent variables.

A general framework for (parametric) nonlinear FA is given by generalized linear latent variable models (GLLVM) in the spirit of generalized linear models (McCullagh and Nelder 1989). Bartholomew and Knott (1999) and Moustaki and Knott (2000) have set the framework by extending the work of Moustaki (1996) and Sammel et al. (1997); see also Skrondal and Rabe-Hesketh (2004). Extensions of FA to nonlinear relationships have also been proposed outside this framework by allowing $\eta_j(z)$ in (1) to take a (parametric) nonlinear form. Gibson (1960) proposes to discretize the factors to allow for nonlinear relationships, while McDonald (1962) (and McDonald 1965; Etezadi-Amoli and McDonald 1983) propose for $\eta_j(z)$ a parametric expansion on orthonormal polynomials of low degree. Mooijaart and Bentler (1986) further assume that the factors are random normal variables. Kenny and Judd (1984) consider for $\eta_j(z)$ a linear combination of the latent variables, their pairwise interactions and possible quadratic effects. Several estimators have then been proposed for the (parametric) nonlinear FA model, among which we find product indicator methods (see e.g. Ping 1996; Wall and Amemiya 2001), two stage least squares (Bollen 1996), maximum likelihood (see e.g. Klein and Moosbrugger 2000; Lee and Zhu 2002), and the method of moments (Wall and Amemiya 2000). A general overview of other (parametric) nonlinear methods can be found in Yalcin and Amemiya (2001) who also discuss possible identification problems in the nonlinear case. Other nonlinear extensions are also possible, and for example Proust et al. (2006) have proposed a parametric transformation of the manifest variables together with a linear form for η_j.

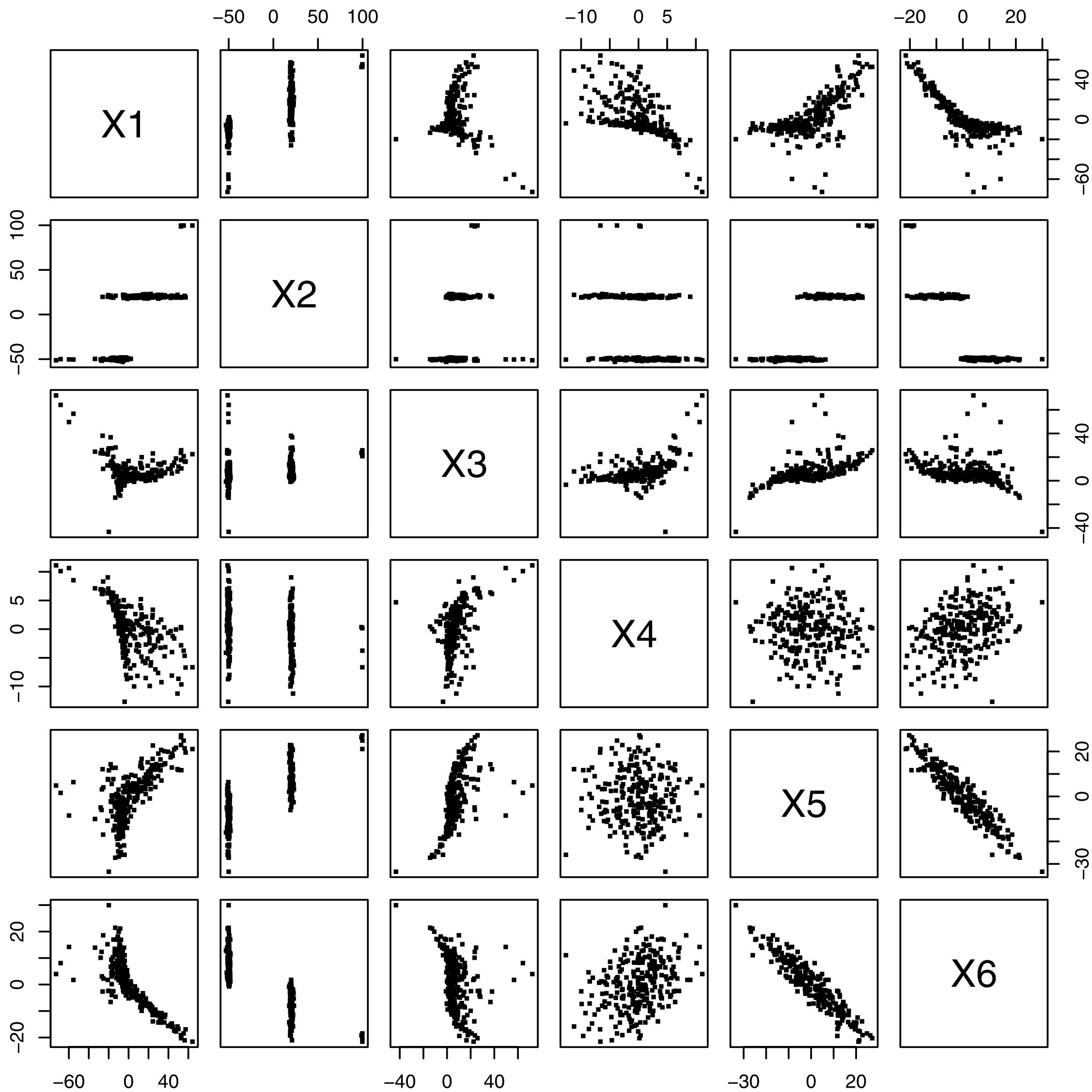

Fig. 1 Scatter diagram of simulated manifest variables from two Gaussian latent variables and three nonlinear associations (first three) and 3 linear associations (last three). Data simulated using $n = 250$, $p = 6$, $q = 2$

In this paper we propose an additive nonparametric estimation of the associations η_j, $j = 1, \dots, p$, in (1) between latent and manifest variables. In some ways, we extend the work of McDonald (1962) who considered polynomials of high degrees. A nonparametric approach provides a more flexible extension to not only the parametric linear case (2), but also to the parametric nonlinear extensions. Nonparametric techniques allow to fit data better without imposing too strong of a structure whether linear or nonlinear. The nonparametric estimation can be used for exploratory data analysis to assess the linearity of η_j, the Gaussianity of the data, hence to help practitioners choose a correct latent variable model that can then be estimated in a parametric fashion. The nonparametric estimation is challenging in latent variable modeling since the latent variables are not observed. Nonparametric extensions to settings using latent variables have been proposed such as in mixed linear models Ghidey et al. (2004) or generalized mixed linear models Hall et al. (2008); see also Ramsay and Silverman (2005). The methodology described in Sect. 2 is simple and fast to compute. It is based on additive nonparametric regression which is a well established research field (see Hastie et al. 2001 for an overview). We evaluate its advantage in terms of mean squared errors compared to the standard (parametric linear) FA on a Monte Carlo simulation study in Sect. 3. Finally we illustrate the method on a data set made of psychological measurements and show that a nonparametric approach can give another insight to the data analysis in Sect. 4.

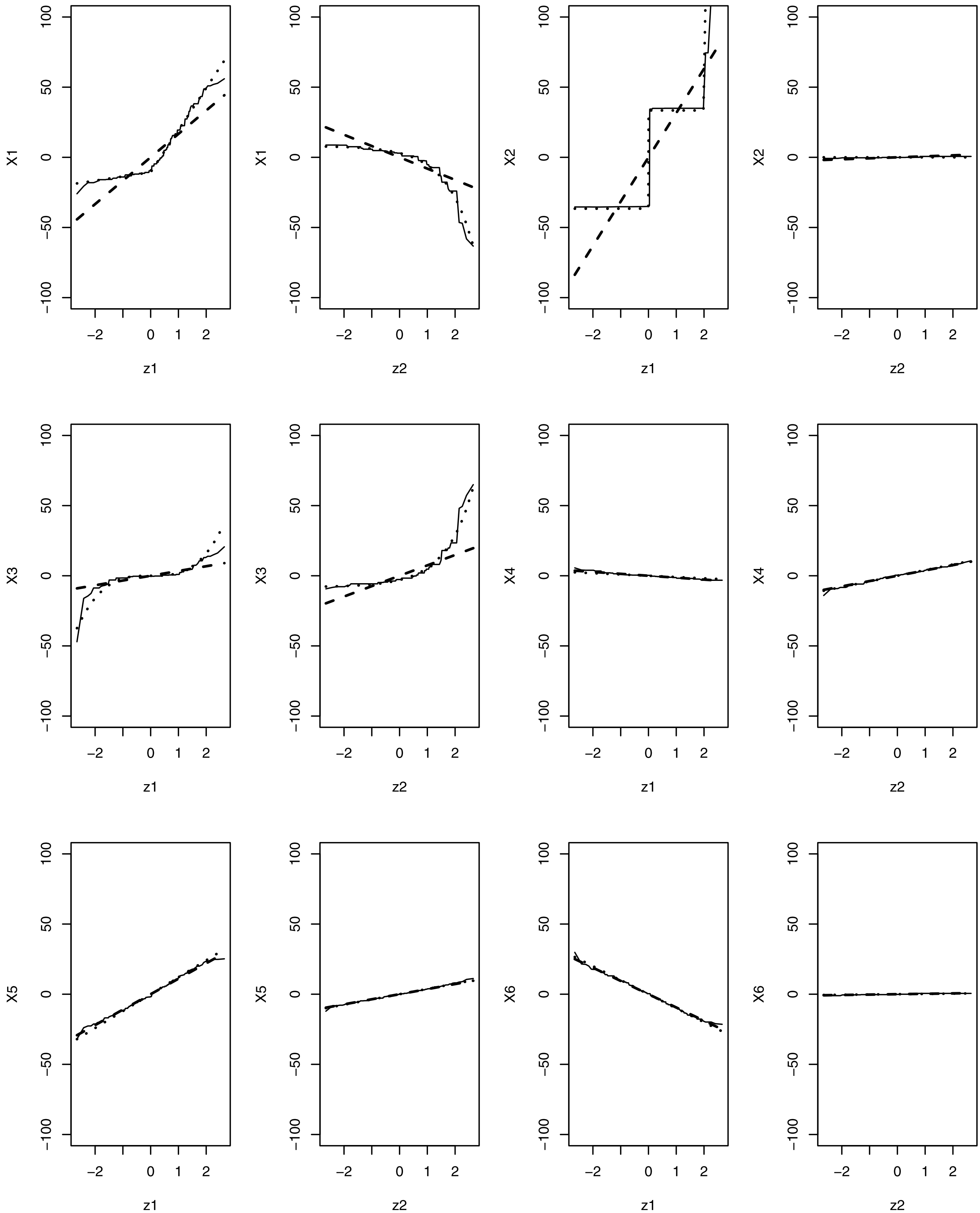

Fig. 2 Parametric linear (*dashed lines*), nonparametric (*solid lines*) and true (*dotted lines*) association between latent and manifest variables. Data simulated using $n = 250$, $p = 6$, $q = 2$

2 Nonparametric estimation

The classical FA model assumes that the associations between latent and manifest variables are monotone in a linear way. This might not be true and cause severe bias in the estimation of the latent variables and the associations between latent and manifest variables. For example, consider the setting in which we have $p = 6$ manifest and $q = 2$ latent variables; out of the six associations, the first three are nonlinear, but additive and monotone, as defined by (4) in Sect. 3, including an elbow, exponential, cubic and step functions. The last three are linear with loadings defined by (5). We generate $n = 250$ observations per manifest variables from this model. We then look at the scatter diagram of the manifest variables in Fig. 1: it is clear that the relationship between the first three manifest variables is not linear, while linear associations seem to be true for the last three. Without knowledge of the true nonlinear model, it is however unclear if an underlying latent structure can explain the observed relationships.

In such a situation, a FA would not be able to capture the underlying relationship between the manifest and the latent variables. To illustrate the danger of linear parametric estimation on our simulated data, the plots of Fig. 2 represent with dashed curves the estimated associations assuming linearity: while the fit is good for the six linear associations, the nonlinear ones are badly estimated. A clear improvement is observed with the solid curves that represent the nonparametric estimate we are proposing: we cannot only see the linear associations, but the nonlinear ones are clearly recognizable as elbow, step, cubic and exponential trends. Section 3 quantifies the advantage of using our nonparametric estimator on a Monte-Carlo simulation.

We now describe our methodology that is based on the additive model framework (Hastie and Tibshirani 1986, 1990). The idea consists of allowing the model to go beyond a rigid parametric linear model. To avoid the curse of dimensionality in high dimension, we assume an additive model, which offers a compromise between flexibility and estimation efficiency (Stone 1985). Additive models assume that the multivariate association has the special additive structure

$$\eta_j(z) = \alpha_{j0} + \sum_{l=1}^{q} \eta_{jl}(z^{(l)}),$$

where η_{jl} are univariate functions, pq in total, that will be estimated nonparametrically. They can then be plotted, as in Fig. 2, to visualize the univariate trends. Additive models have the advantage that they can be plotted on one-dimensional graphs and can be interpreted. As opposed to the standard additive model regression problem, not only the η_{jl} but also the latent variables z are unknown in FA. To develop a nonparametric estimator in this challenging setting, the additional assumption we make is that the associations η_{jl} between manifest and latent variables are monotone. This enables one to use any estimator of the latent scores that respects this monotonicity as a starting point for estimating the latent variables. More precisely, given estimated latent scores $\widehat{z}_i^{(l)}$, $l = 1, \ldots, q$, $i = 1, \ldots, n$ based on a standard FA model, their order can be used to estimate the quantile by taking Gaussian theoretical quantiles. They are then used as regressors in a monotone additive model framework to estimate η_{jl} in a nonparametric fashion. If the relationships between the manifest and latent variables are monotone, then the FA model provides a linear approximation of these relationships and the factor scores provide an ordering of the observations on the latent space.

Hence, the nonparametric estimation of the relationship between the manifest and latent variables we propose uses the ordering of the latent scores for constructing normalized scores with the normal quantiles transformation that preserves this ordering. We then use these normalized scores as covariates. In other words, to fit a monotone additive latent variable model nonparametrically, we propose the following two step estimator:

1. Compute the latent scores as if the associations were linear. Assuming the associations are monotone, this provides an estimate of the order $o_1, \ldots, o_n$ between (true) latent scores $z_1^{(q)}, \ldots, z_n^{(l)}$, such that the ith order statistics is $z_{[i]}^{(l)} = z_{o_i}^{(l)}$. Using these estimated orders $\hat{o}_i$, fix the estimated latent variables as the theoretical quantiles of the assumed latent variable distribution, e.g., the standard normal Φ:

$$\tilde{z}_i^{(l)} = \Phi^{-1}(\hat{o}_i/(n+1)); \quad (3)$$

2. Estimate the nonlinear associations using backfitting with a univariate isotone nonparametric estimator. For additive model regression, various estimators have been proposed based on splines (Buja et al. 1989; Friedman and Silverman 1989; Wood 2000) or wavelets (see Sardy and Tseng 2004 and references therein). To constrain monotonicity here, we use the `isotone` function of the `EBayesThresh` library in `R`. The complexity of this least squares fit with monotonicity constraint is $O(n)$. Note that, although `isotone` belongs to the `EBayesThresh` library, it has nothing to do with thresholding and empirical Bayes, but is rather based on the pool adjacent violator algorithm. Mammen and Yu (2007) studied some properties of isotone regression with backfitting, which is the procedure we are proposing here. In particular they show convergence of the algorithm for given predictors; our pre-

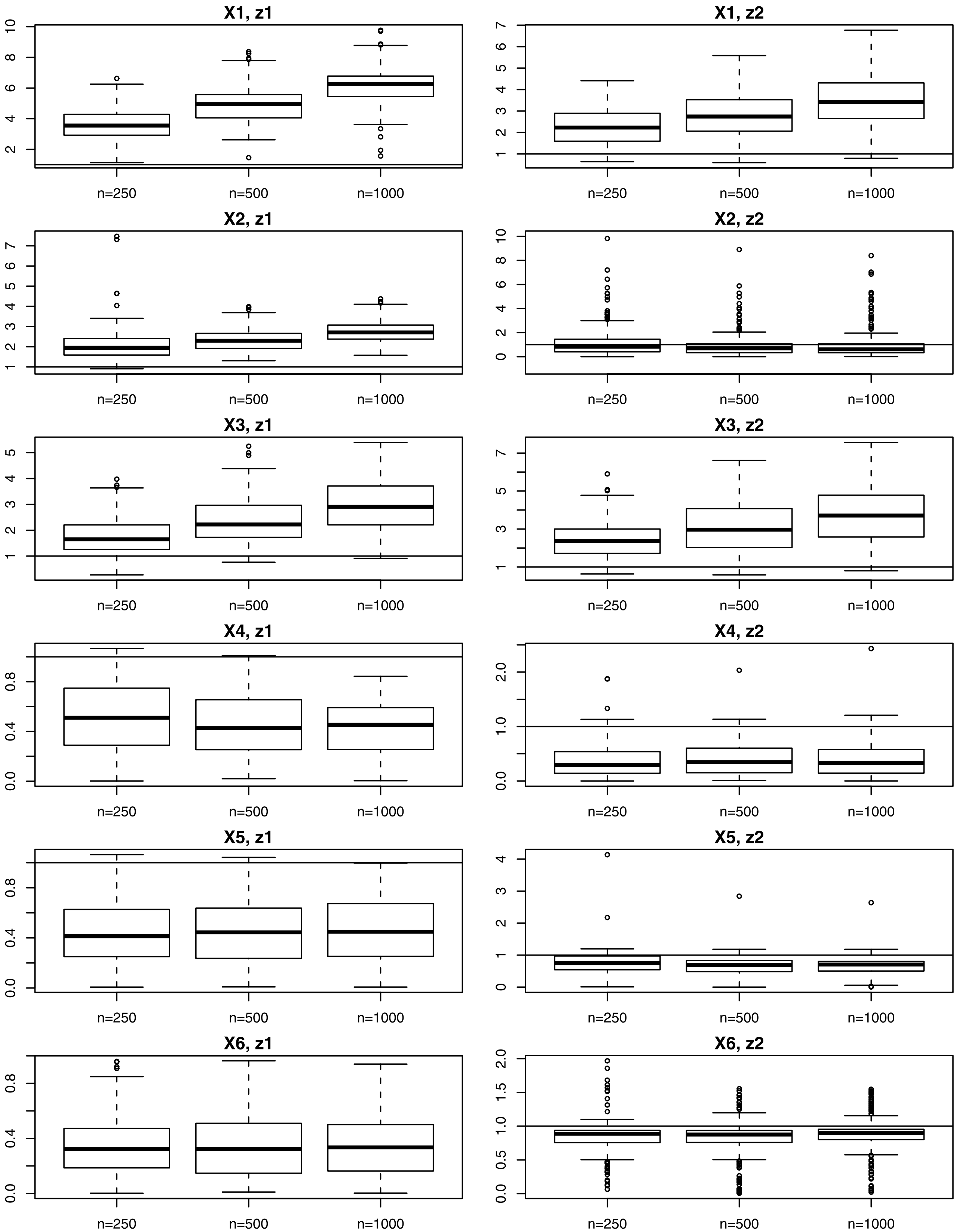

Fig. 3 Root MSE of prediction ratio of the parametric linear relative to the nonparametric estimator of the association between latent and manifest variables. Data simulated using respectively $n = 250, 500, 1000$ and $p = 6, q = 2$, 200 samples. The boxplots for $(x^{(2)}, z^{(2)})$ have been upper truncated at 10 living out respectively 8, 6 and 1 root MSE ratios

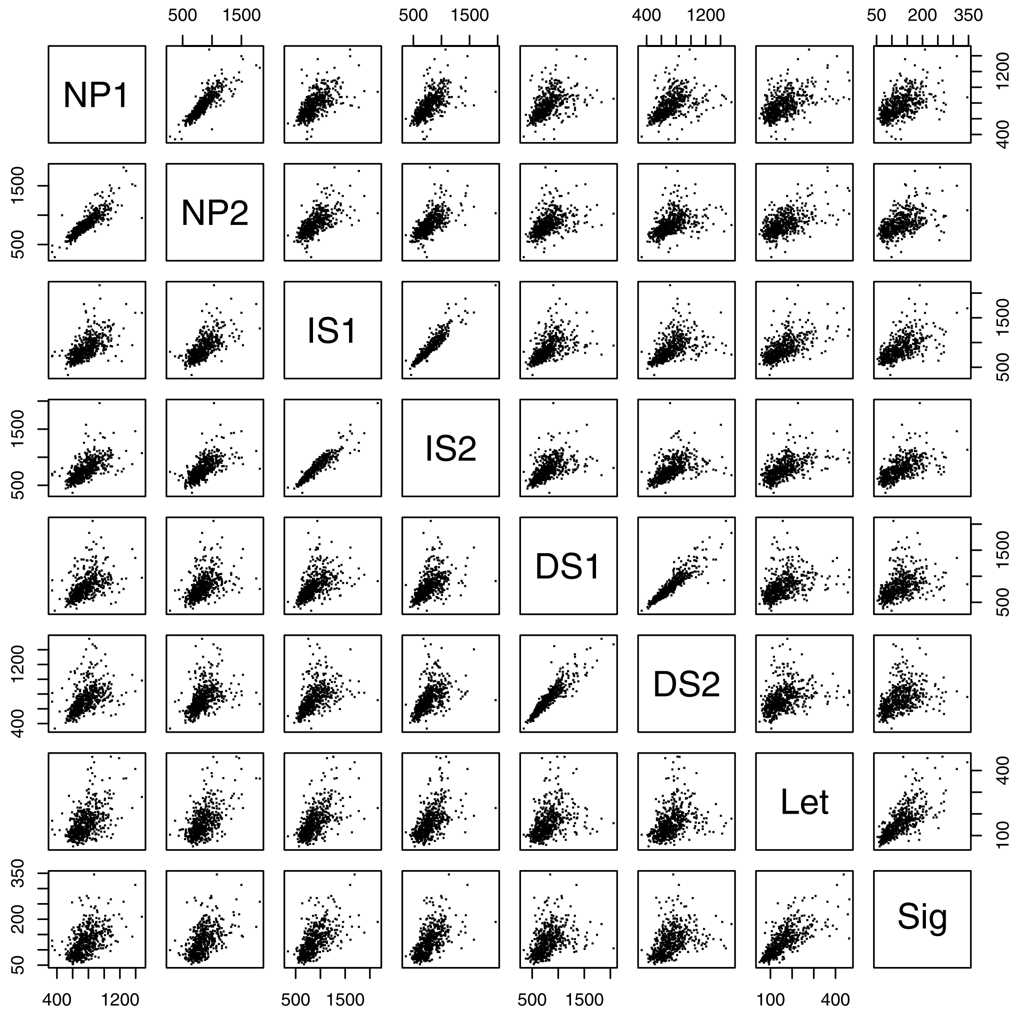

Fig. 4 Scatter diagram of the selective attention data

dictors here are the quantiles (3) obtained from estimating the latent variables with FA, and are considered fixed.

The pq estimated functions corresponding to the estimated latent variables scores, possibly with confidence intervals, can then be visualized. The confidence intervals can be computed using the bootstrap (Efron and Tibshirani 1993) as is done for the FA model for instance by Ichikawa and Konishi (1995). This enables the data analyst to check whether Gaussian and linear associations is an appropriate model, or if such assumptions of the standard FA model are strongly violated. If the relationship is not linear, then either the assumed latent variables distribution and/or the linearity assumption are wrong. Interestingly, if the associations $\eta_{jl}(\cdot)$ are identical and strictly increasing for all j (i.e., $\eta_l := \eta_{jl}$), a linear fit can be achieved by bending the corresponding estimated latent variable according to $\check{z}_i^{(l)} = \eta_l^{-1}(\tilde{z}_i^{(l)})$.

Finally, in latent variable models, the identification problem is an important issue (see for instance Yalcin and Amemiya 2001). The estimation method we propose inherits the identification problems of the parametric linear FA model since the latter provides the (transformed) latent variables scores that are then used as fixed covariates in the nonparametric regression. To avoid multiple solutions in FA models, it is well-known that a rotation, such as the varimax rotation, can be used for the factor loadings matrix.

3 Simulation

In this section, we illustrate the advantage of using a nonparametric estimator on simulated data. We consider the setting in which we have $p = 6$ manifest and $q = 2$ latent variables; out of the six associations, the first three are nonlinear

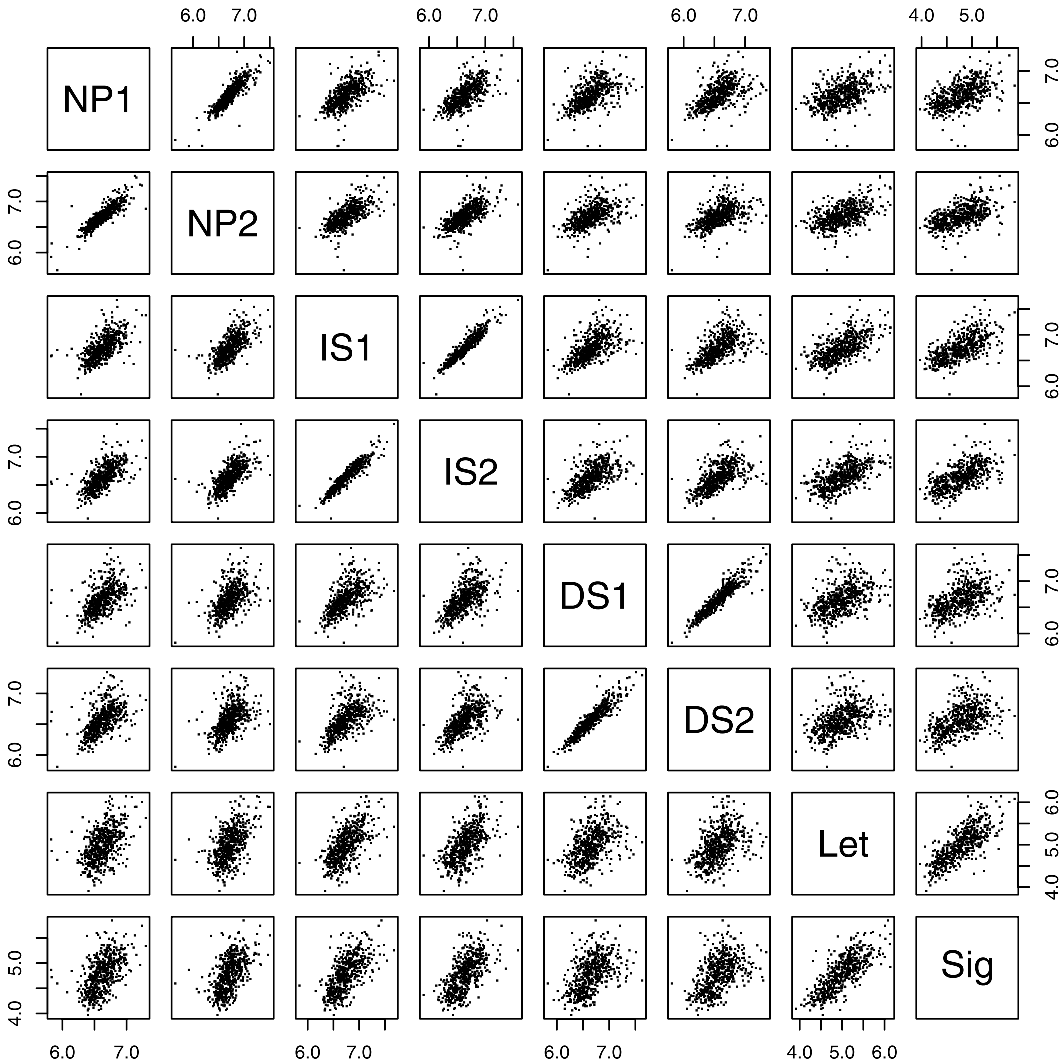

Fig. 5 Scatter diagram of the log-transformed selective attention data

according to

$$
\begin{aligned}
&\eta_1(z^{(1)}, z^{(2)}) \\
&\quad = \{3z^{(1)} \cdot \mathbf{1}(z^{(1)} < 0) + 30z^{(1)} \cdot \mathbf{1}(0 \le z^{(1)})\} - 5\exp(z^{(2)}) \\
&\quad = \eta_{11}(z^{(1)}) + \eta_{12}(z^{(2)}) \\
&\eta_2(z^{(1)}, z^{(2)}) \\
&\quad = \{-50 \cdot \mathbf{1}(z^{(1)} < 0) + 20\,\mathbf{1}(0 \le z^{(1)} < 2) \\
&\qquad + 100 \cdot \mathbf{1}(2 \le z^{(1)})\} + 0 \\
&\quad = \eta_{21}(z^{(1)}) + \eta_{22}(z^{(2)}) \\
&\eta_3(z^{(1)}, z^{(2)}) \\
&\quad = 2(z^{(1)})^3 + 5\exp(z^{(2)}) \\
&\quad = \eta_{31}(z^{(1)}) + \eta_{32}(z^{(2)}),
\end{aligned}
\tag{4}
$$

where $\mathbf{1}(\cdot)$ is the indicator function which takes the value of one if the argument is true and zero otherwise. Note that the associations are additive and monotone but nonlinear, with nonlinear terms like an elbow function η_{11}, two exponential functions η_{12} and η_{32}, a cubic term η_{31} and a step function η_{21}. They have been scaled to have comparable signal to noise ratio. The other three associations are linear with loadings values

$$
\boldsymbol{\alpha} = [\alpha_{jl}]_{j=4,\dots,6,l=1,2} = \begin{bmatrix} -0.91 & 4.02 \\ 12.08 & 3.74 \\ -10.00 & 0.21 \end{bmatrix}. \tag{5}
$$

In order to study the performance of the nonparametric estimator and compare it to the parametric linear one, we simulate 200 samples of size $n = 250$ from the nonlinear FA model with $p = 6$ manifest variables and $q = 2$ latent

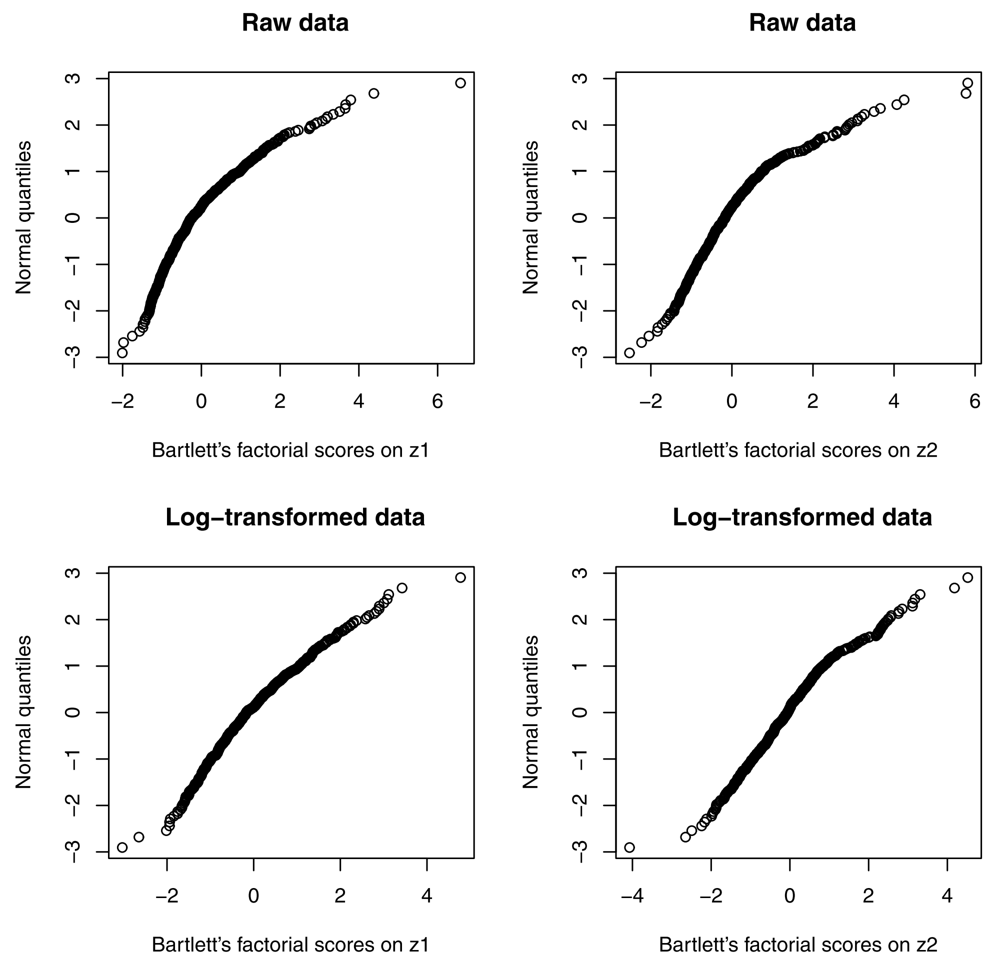

Fig. 6 Scatter diagram of normalized versus raw Bartlett's factorial scores for the psychology example

variables associated using the relationships given in (4) and with the loadings given in (5). We estimate the mean squared errors (MSE) taking the average over the samples of the squared differences between the estimated associations (either nonparametric or parametric linear) and the true associations. Figure 3 presents the boxplots of the ratio of the root MSE between the parametric linear and nonparametric estimates for increasing sample sizes $n = 250$, $n = 500$ and $n = 1000$. The horizontal line is set to the value of one, so that ratios above this line denote a better estimation with the proposed nonparametric method than with the parametric linear one. For the nonlinear associations (first three rows) the root MSE ratios show a better performance for the nonparametric estimator the larger the sample size n. For example, with $(x^{(1)}, z^{(1)})$, the median root MSE ratio goes from 3.55 with $n = 250$, to 4.95 with $n = 500$ and up to 6.27 with $n = 1000$. For the linear associations (last three rows), the performance of the parametric linear estimator is better, and the root MSE ratio remains generally stable across sample sizes. For example, with $(x^{(6)}, z^{(1)})$, the median root MSE ratio is 0.887 with $n = 250$, 0.876 with $n = 500$ and 0.898 with $n = 1000$. As far as running time is concerned, the average CPU time of the nonparametric estimator is 1.0, 1.4, 2.2 for $n = 250, 500, 1000$, respectively, which tends to show a linear association with the number n of observations. We conclude that the gain of using the isotone additive nonparametric estimator is important when some associations are potentially nonlinear.

4 Application

The data come from a large study in psychology (de Ribaupierre et al. 2003) aimed at the study of selective attention (Dempster and Brainerd 1995) and processing speed (Salthouse 1996). Within this study and for the purpose of illustrating our method, we have selected 8 variables measured on 544 participants. The first two variables (NP1 and NP2) are median response times in milliseconds for two conditions of a negative priming task (Tipper and Cranston 1985),

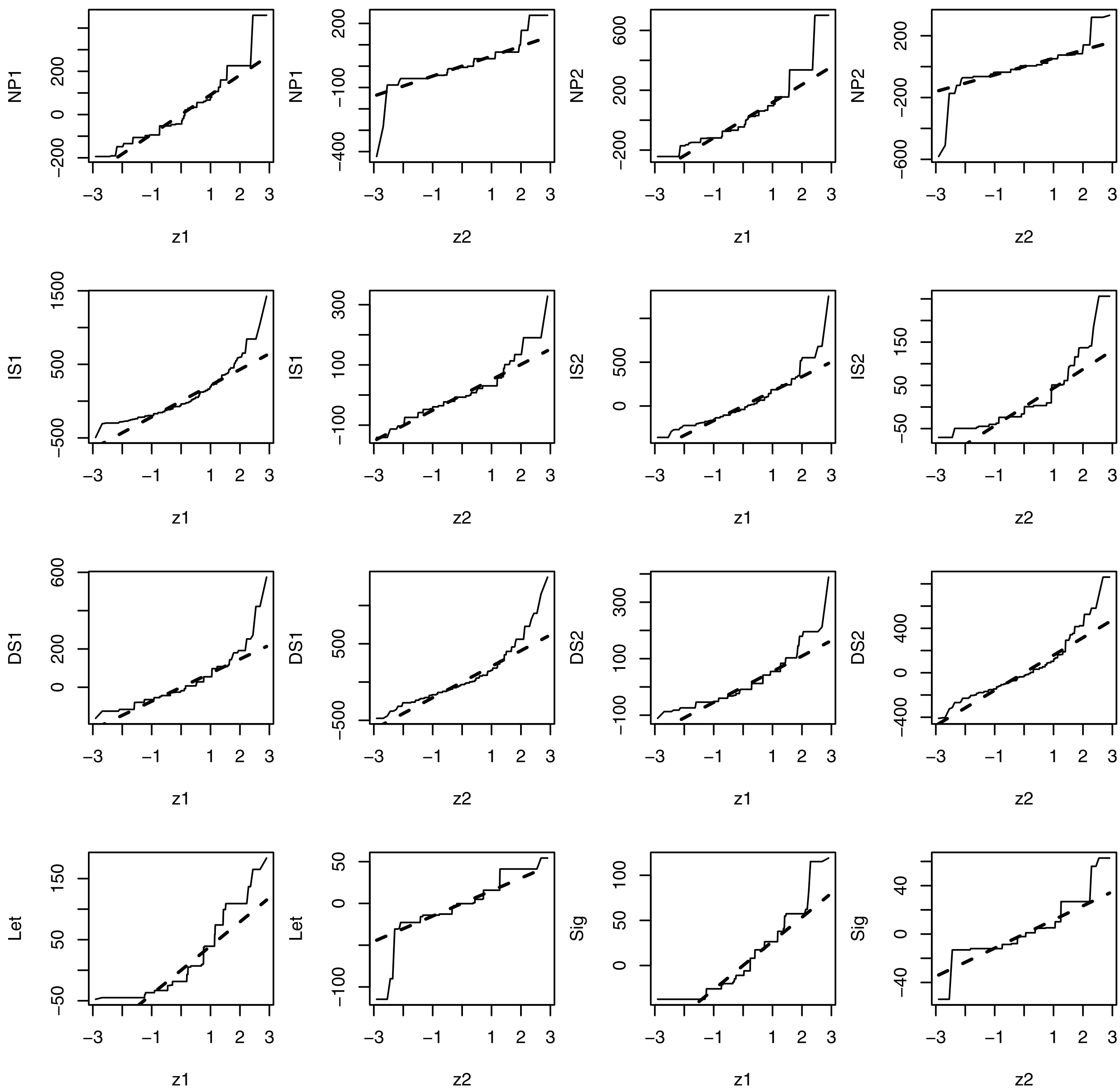

Fig. 7 Nonparametric (*solid line*) and parametric linear (*dashed line*) estimated relationships between the latent and the manifest variables of the psychology example

the third and fourth variables (IS1 and IS2) are median response times in milliseconds for two conditions of an integrated stroop task, and the fifth and sixth variables (DS1 and DS2) are median response times in milliseconds for two conditions of a dissociated stroop task (MacLeod 1991). These six variables are supposed to measure indirectly and at different levels the selective attention of the subjects. The last two variables (Let and Sig) are the total time in seconds for completing two tasks for respectively a comparison of letters and a comparison of signs (Salthouse 1991). They are supposed to measure indirectly and at different levels the processing speed of the subjects.

The scatter of the data is given in Fig. 4. The data clouds do not really fit into ellipsoids, hence the multivariate normality of the manifest variables, implied by the multivariate normality of the latent variables and the linearity of the conditional mean, does not look like a reasonable assumption, so that standard FA should not be employed here. In such

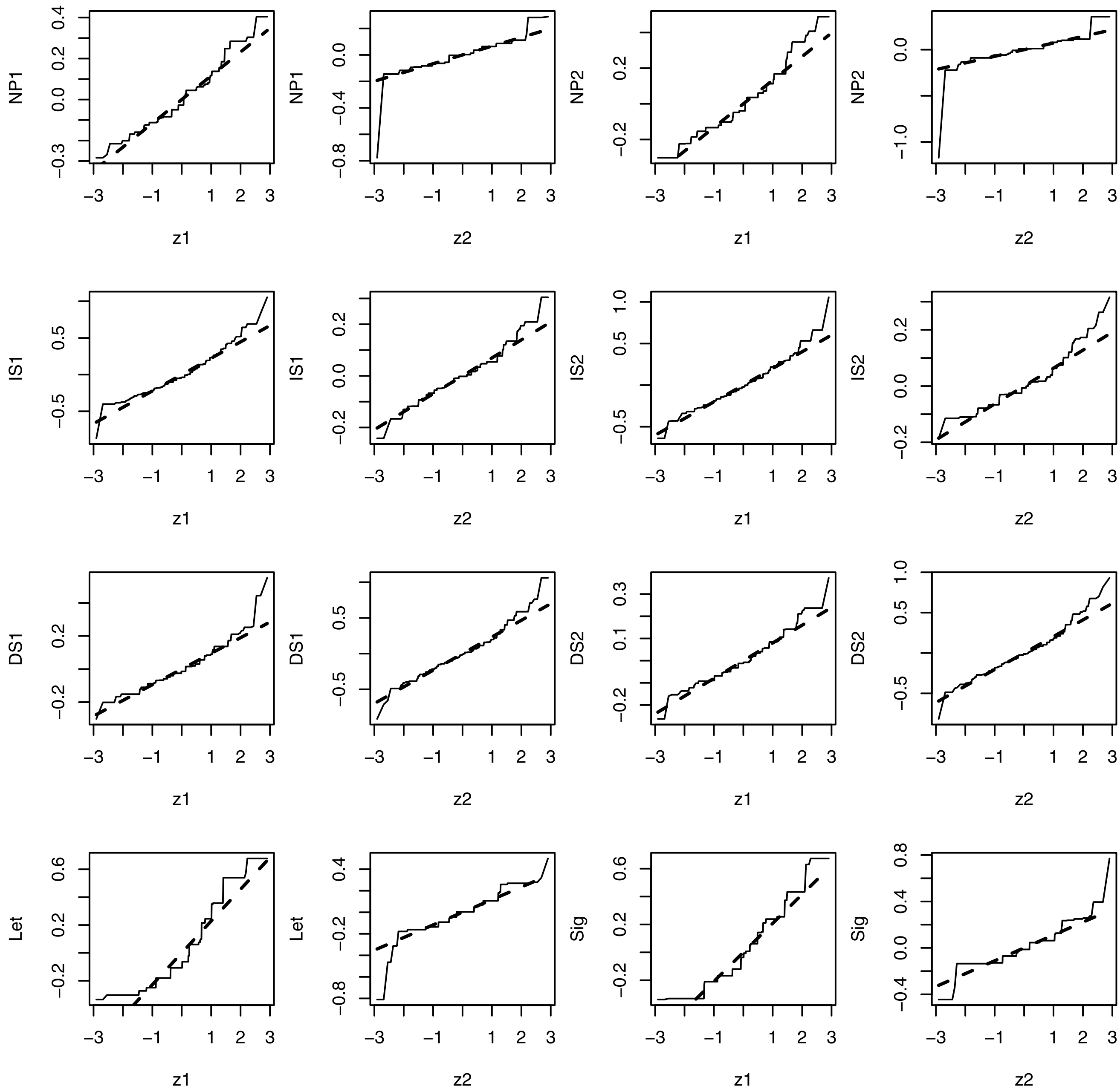

Fig. 8 Nonparametric (*solid line*) and parametric linear (*dashed line*) estimated relationships between the latent and the manifest variables of the psychology example (log-transformed)

situation, an alternative is to hunt for a good transformation of the variables. With measures taken on the time scale, one can resort to a log-transformation of the manifest variables. The scatter diagram of the log-transformed data is given in Fig. 5. The data clouds are now more in conformity with the normality assumption and a comparison between a classical FA and our nonparametric approach will confirm whether the former model is suitable. It should be noted that the approach we adopt here is an exploratory data analysis in that the fit of the different models are assessed graphically rather than quantified with a goodness-of-fit measure, which in the present case is an open research area. The aim is to investigate whether a given parametric model is reasonable for a given dataset.

So we estimate a parametric linear and nonparametric FA on both the original and the log-transformed data. In Fig. 6 are presented the scatter plots of the normalized versus raw Bartlett's factorial scores, i.e., $\tilde{z}_i^{(l)}$ versus $\hat{z}_i^{(l)}$. With this example, we see the normalization has an effect on the transformed latent scores, suggesting that the type of nonlinear relationship might be the same across manifest variables. Figures 7 and 8 show the plots of the parametric linear (dashed lines) and nonparametric (solid lines) estimated associations between each manifest and latent variable on the original and on the log-transformed data: when the data are not transformed, the associations are overall nonlinear and similar for all pairs of manifest/latent variables. This suggests that the same transformation can be used for the latent variables or for the manifest variables. When the data are log-transformed, the relationships become linear for some associations, but not quite linear for a few of them. Hence in this example, the isotone additive fit shows that a log-transformation may not be suitable yet.

5 Conclusion

FA models are very widely used in many disciplines. They suppose linear associations between manifest and latent variables, a hypothesis that can be violated in practice. In this paper we extend FA models to monotone additive latent variable models and propose a procedure for estimating nonparametrically the nonlinear relationships between manifest and latent variables. The resulting analysis assesses the FA model and can possibly propose transformations then appropriately fit a parametric nonlinear model. Moreover, it is possible to extend the method to the more general framework of GLLVM, which is the scope of future research.

Acknowledgements We thank the associate editor and two anonymous referees for their thorough revision of the paper leading to a significant improved version.

References

Bartholomew, D.J.: The foundations of factor analysis. Biometrika **71**, 221–232 (1984)

Bartholomew, D.J.: The sensitivity of latent trait analysis to choice of prior distribution. Br. J. Math. Stat. Psychol. **41**, 101–107 (1988)

Bartholomew, D.J., Knott, M.: Latent Variable Models and Factor Analysis. Kendall's Library of Statistics, vol. 7. Arnold, London (1999)

Boelen, P.A., van den Hout, M.A., van den Bout, J.: The factor structure of posttraumatic stress disorder symptoms among bereaved individuals: A confirmatory factor analysis study. J. Anxiety Disord. **22**, 1377–1383 (2008)

Bollen, K.: An alternative two stage least squares (2SLS) estimator for latent variable equation. Psychometrika **61**, 109–121 (1996)

Browne, M.W.: Asymptotically distribution-free methods for the analysis of covariance structures. Br. J. Math. Stat. Psychol. **37**, 62–83 (1984)

Buja, A., Hastie, T.J., Tibshirani, R.J.: Linear smoothers and additive models (with discussion). Ann. Stat. **17**, 453–555 (1989)

Conne, D., Ronchetti, E., Victoria-Feser, M.-P.: Goodness-of-fit for generalized linear latent variables models. J. Am. Stat. Assoc. **105**, 1126–1134 (2010)

de Ribaupierre, A., Borella, E., Delaloye, C.: Inhibition et variabilité individuelle: généralité ou spécificité des processus inhibiteurs [inhibition and individual variability: generality or specificity of inhibitory processes]. In: Moutier, S. (ed.) Inhibition Neurale et Cognitive. Hermès, Paris (2003)

Dempster, F.N., Brainerd, C.J.: Interference and Inhibition in Cognition. Academic Press, San Diego (1995)

Donncha, H., Shevlin, M., Dempster, M.: The structure of the statistics anxiety rating scale: A confirmatory factor analysis using UK psychology students. Pers. Individ. Differ. **45**, 68–74 (2008)

Dryden, I.L., Bai, L., Brignell, C.J., Shen, L.: Factored principal components analysis, with applications to face recognition. Stat. Comput. **19**, 229–238 (2009)

Dupuis Lozeron, E., Victoria-Feser, M.-P.: Robust estimation of constrained covariance matrices for confirmatory factor analysis. Comput. Stat. Data Anal. **54**, 3020–3032 (2010)

Efron, B., Tibshirani, R.: An Introduction to the Bootstrap. Chapman & Hall, London (1993)

Etezadi-Amoli, J., McDonald, R.P.: A second generation nonlinear factor analysis. Psychometrika **48**, 315–342 (1983)

Friedman, J.H., Silverman, B.W.: Flexible parsimonious smoothing and additive modeling (with discussion). Technometrics **31**, 3–39 (1989)

Ghidey, W., Lesaffre, E., Eilers, P.: Smooth random effects distribution in a linear mixed model. Biometrics **60**, 945–953 (2004)

Gibson, W.A.: Nonlinear factors in two dimensions. Psychometrika **25**, 381–392 (1960)

Hall, P., Müller, H.-G., Yao, F.: Modelling sparse generalized longitudinal observations with latent gaussian processes. J. R. Stat. Soc., Ser. B, Stat. Methodol. **70**, 703–723 (2008)

Hastie, T., Tibshirani, R.: Generalized additive models (C/R: P310-318). Stat. Sci. **1**, 297–310 (1986)

Hastie, T.J., Tibshirani, R.J.: Generalized Additive Models. Chapman & Hall, London (1990)

Hastie, T.J., Tibshirani, R.J., Friedman, J.: Elements of Statistical Learning: Data Mining, Inference and Prediction. Springer, New York (2001)

Ichikawa, M., Konishi, S.: Application of the bootstrap methods in factor analysis. Psychometrika **60**, 77–93 (1995)

Jöreskog, K.G.: Some contributions to maximum likelihood factor analysis. Psychometrika **32**, 443–482 (1967)

Jöreskog, K.G.: A general approach to confirmatory maximum likelihood factor analysis. Psychometrika **34**, 183–203 (1969)

Kenny, D.A., Judd, C.M.: Estimating the nonlinear and interactive effects of latent variables. Psychol. Bull. **96**, 201–210 (1984)

Klein, A.G., Moosbrugger, H.: Maximum likelihood estimation of latent interaction effects with the LMS method. Psychometrika **65**, 457–474 (2000)

Lee, S.Y., Zhu, H.T.: Maximum likelihood estimation of nonlinear structural equation models. Psychometrika **67**, 189–210 (2002)

MacLeod, C.M.: Half a century of research on the stroop effect: An integrative review. Psychol. Bull. **109**(2), 163–203 (1991)

Mammen, E., Yu, K.: Additive isotone regression. In: Wellner, J.A., Cator, E.A., Jongbloed, G., Kraaikamp, C., Lopuhaä, H.P. (eds.) Asymptotics: Particles, Processes and Inverse Problems. Institute of Mathematical Statistics Lecture Notes—Monograph Series, vol. 55, pp. 179–195. Institute of Mathematical Statistics, Beachwood (2007)

Mardia, K.V., Kent, J.T., Bibby, J.M.: Multivariate Analysis. Academic Press, London (1979)

Mavridis, D., Moustaki, I.: The forward search algorithm for detecting aberrant response patterns in factor analysis for binary data. J. Comput. Graph. Stat. **18**, 1016–1034 (2009)

McCullagh, P., Nelder, J.A.: Generalized Linear Models, 2nd edn. Chapman & Hall, London (1989)

McDonald, R.P.: A general approach to nonlinear factor analysis. Psychometrika **27**, 397–415 (1962)

McDonald, R.P.: Difficulty factors and nonlinear factor analysis. Br. J. Math. Stat. Psychol. **18**, 11–23 (1965)

Mooijaart, A., Bentler, P.: Random polynomial factor analysis. In: Diday, E. et al. (eds.) Data Analysis and Informatics. North-Holland, Amsterdam (1986)

Moustaki, I.: A latent trait and a latent class model for mixed observed variables. Br. J. Math. Stat. Psychol. **49**, 313–334 (1996)

Moustaki, I., Knott, M.: Generalized latent trait models. Psychometrika **65**, 391–411 (2000)

Moustaki, I., Victoria-Feser, M.-P.: Bounded-bias robust inference for generalized linear latent variable models. J. Am. Stat. Assoc. **101**, 644–653 (2006)

Ping, R.A.: Latent variable regression: A technique for estimating interaction and quadratic coefficients. Multivar. Behav. Res. **31**, 95–120 (1996)

Proust, C., Jacqmin-Gadda, H., Taylor, J.M.G., Ganiayre, J., Commenges, D.: A nonlinear model with latent process for cognitive evolution using multivariate longitudinal data. Biometrics **62**, 1014–1024 (2006)

Ramsay, J.O., Silverman, B.W.: Functional Data Analysis. Springer Series in Statistics. Springer, Berlin (2005)

Salthouse, T.A.: Theoretical Perspectives on Cognitive Aging. Erlbaum, Hillsdale (1991)

Salthouse, T.A.: The processing-speed theory of adult age differences in cognition. Psychol. Rev. **103**(3), 403–428 (1996)

Sammel, M.D., Ryan, L.M., Legler, J.M.: Latent variable models for mixed discrete and continuous outcomes. J. R. Stat. Soc., Ser. B, Stat. Methodol. **59**, 667–678 (1997)

Sardy, S., Tseng, P.: AMlet, RAMlet and GAMlet: Automatic nonlinear fitting of additive model, robust and generalized with wavelets. J. Comput. Graph. Stat. **13**, 283–309 (2004)

Skrondal, A., Rabe-Hesketh, S.: Generalized Latent Variable Modeling: Multilevel, Longitudinal, and Structural Equation Models. Chapman & Hall, London (2004)

Spearman, C.: General intelligence objectively determined and measured. Am. J. Psychol. **15**, 201–293 (1904)

Stone, C.J.: Additive regression and other nonparametric models. Ann. Stat. **13**, 689–705 (1985)

Tipper, S.P., Cranston, M.: Selective attention and priming: inhibitory and facilitatory effects of ignored primes. Q. J. Exp. Psychol., A Hum. Exp. Psychol. **37**(4), 591–611 (1985)

Tipping, M.E., Bishop, C.M.: Probabilistic principal component analysis. J. R. Stat. Soc., Ser. B, Stat. Methodol. **61**, 611–622 (1999)

Wall, M.M., Amemiya, Y.: Estimation for polynomial structural equation models. J. Am. Stat. Assoc. **95**, 929–940 (2000)

Wall, M.M., Amemiya, Y.: Generalized appended product indicator procedure for nonlinear structural equation analysis. J. Educ. Behav. Stat. **26**, 1–29 (2001)

Wood, S.N.: Modelling and smoothing parameter estimation with multiple quadratic penalties. J. R. Stat. Soc. B **62**, 413–428 (2000)

Yalcin, I., Amemiya, Y.: Nonlinear factor analysis as a statistical method. Stat. Sci. **16**, 275–294 (2001)

Yuan, K.-H., Bentler, P.M.: Robust mean and covariance structure analysis. Br. J. Math. Stat. Psychol. **51**, 63–88 (1998)