FACULTE DES SCIENCES ECONOMIQUES ET SOCIALES

HAUTES ETUDES COMMERCIALES

Generalized monotone additive latent variable models

Sylvain Sardy
Maria-Pia Victoria-Feser

HEC GENÈVE

UNIVERSITÉ DE GENÈVE

Generalized monotone additive latent variable models

Sylvain Sardy
and
Maria-Pia Victoria-Feser
University of Geneva

May 4, 2010

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 (Jöreskog 1969) (FA) which can be used when the data are multivariate normal. 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 given the latter, there is no more correlation between the manifest variables. 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 $q < p$ (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 the $\alpha_{jl}, l = 1, \dots, q$'s in $\boldsymbol{\alpha}_j, j = 1, \dots, p$ (together with the intercepts α_{j0}) must be estimated. If the model holds, i.e., the correlations between the manifest variables are explained by a smaller number of latent ones, then one assumes that the manifest variables are conditionally independent given the latent ones. This means that $\boldsymbol{\epsilon} = (\epsilon^{(1)}, \dots, \epsilon^{(p)})^{\mathrm{T}} \sim (\mathbf{0}, \boldsymbol{\Psi})$ 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*. One further assumes that $\boldsymbol{z} \sim \mathrm{N}(\mathbf{0}, \boldsymbol{R})$ where the independence assumption is often stated (i.e., $\boldsymbol{R} = \boldsymbol{I}$), and that $\mathrm{cov}(z^{(l)}, \epsilon^{(j)}) = 0$ for all j and l. Hence the FA model satisfies that $E[\boldsymbol{x} \mid \boldsymbol{z}] = \boldsymbol{\alpha}_0 + \boldsymbol{\alpha}\boldsymbol{z}$ and $\mathrm{Var}(\boldsymbol{x}) = \boldsymbol{\alpha}\boldsymbol{\alpha}^{\mathrm{T}} + \boldsymbol{\Psi}$, with $\boldsymbol{\alpha}_0 = [\alpha_{j0}]^{\mathrm{T}}_{j=1,\ldots,p}$ and $\boldsymbol{\alpha} = [\boldsymbol{\alpha}_j]^{\mathrm{T}}_{j=1,\ldots,p}$. The FA model's 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 $\widehat{\boldsymbol{z}}_i$ can also be estimated using for example Bartlett's factor scores given by

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

with $\boldsymbol{\alpha}$ and $\boldsymbol{\Psi}$ replaced by their estimates (for example, see Mardia, Kent, and Bibby 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. The number of latent variables q is 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, Ronchetti, and Victoria-Feser (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, Bai, Brignell, and Shen 2009), chemometrics, and biomedical studies.

In practice, such an FA model does not always fit in a "reasonable way" the data, as may reveal 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 e.g. 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 general model is given by one that allows the relationship between the manifest and the latent variables to be also nonlinear.

In this paper we propose an additive nonparametric estimation of the associations η_j, $j = 1, \ldots, p$, in (1) between latent and manifest variables. As opposed to the parametric linear case (2), nonparametric techniques allow to fit data better without imposing too strong of a structure. It covers both cases when either normality of the latent variables or linearity of the predictor are not appropriate assumptions. The nonparametric estimation can be used for exploratory data analysis to assess the linearity of η_j, the Gaussianity of the data and if necessary propose nonlinear parametric associations and then estimate the model in a parametric fashion. The nonparametric estimation is challenging in latent variable modeling since the latent variables are not observed.

Additive nonparametric regression is a well established research field (see Hastie, Tibshirani, and Friedman 2001 for an overview). Nonparametric extensions to settings using latent variables have been proposed such as in longitudinal data (Proust, Jacqmin-Gadda, Taylor, Ganiayre, and Commenges 2006) and in mixed linear models (Ghidey, Lesaffre, and Eilers 2004) or generalized mixed linear models (Hall, Mller, and Yao 2008). The aim of our paper is to develop a nonparametric extension to factorial analysis as an exploratory data analysis tool to help practitioners choose a correct latent variable model. The methodology described in Section 2 is simple and fast to compute. We evaluate its advantage in terms of mean squared errors compared to the standard parametric FA on a Monte Carlo simulation study in Section 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 Section 4 .

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 Section 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 Figure 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.

Figure 1 here

In such a situation, a parametric 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 Figure 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 can not 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.

Figure 2 here

We now describe our methodology that is based on the additive model framework (Hastie and Tibshirani 1986; Hastie and Tibshirani 1990). The idea consists in 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

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

where η_{jl} are univariate functions, pq in total, that will be estimated nonparametrically. They can then be plotted, as in Figure 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

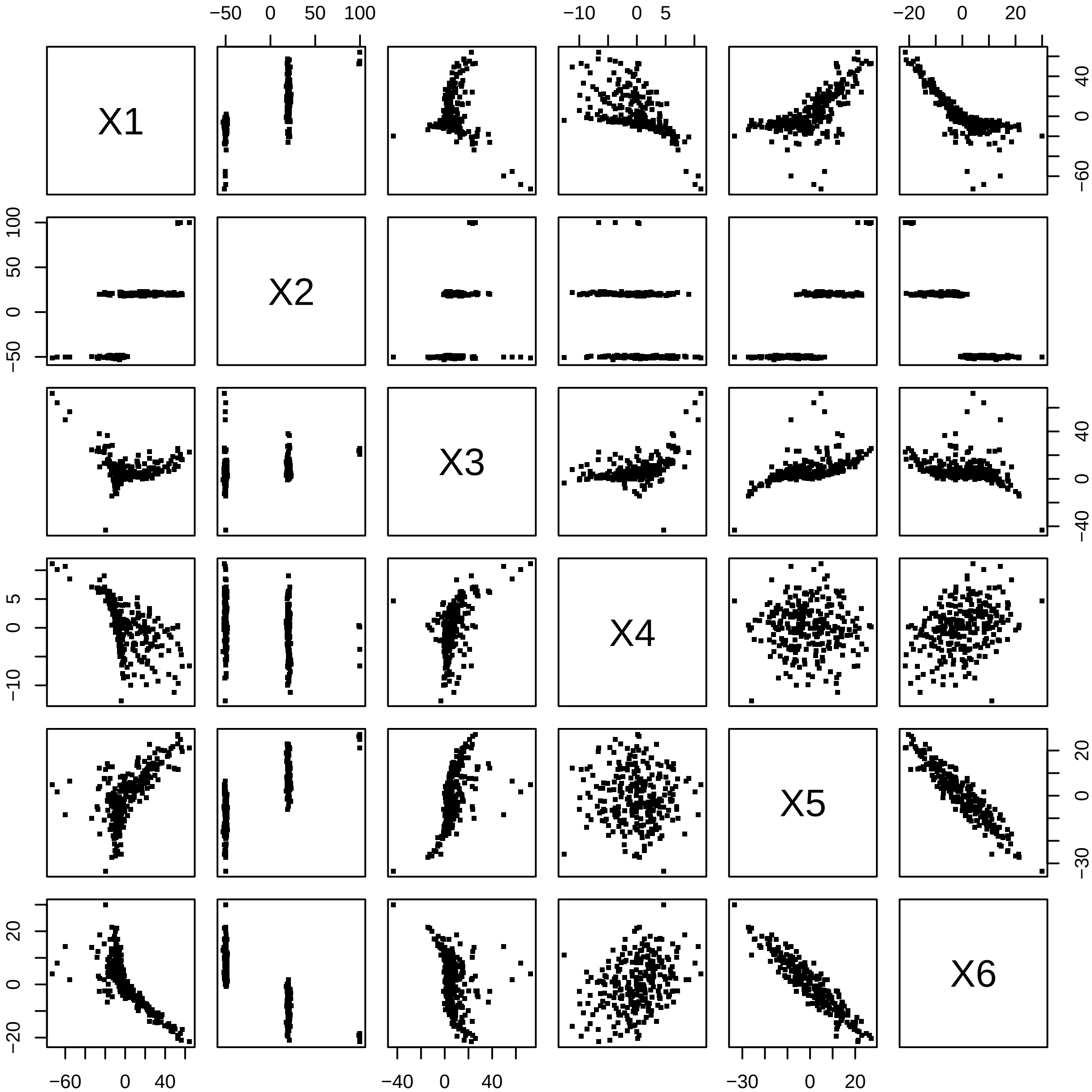

Figure 1: Scatter diagram of simulated manifest variables from two Gaussian latent variables and three nonlinear associations (first three) and 3 linear associations (last three). Data simulated using $n = 250, p = 6, q = 2$.

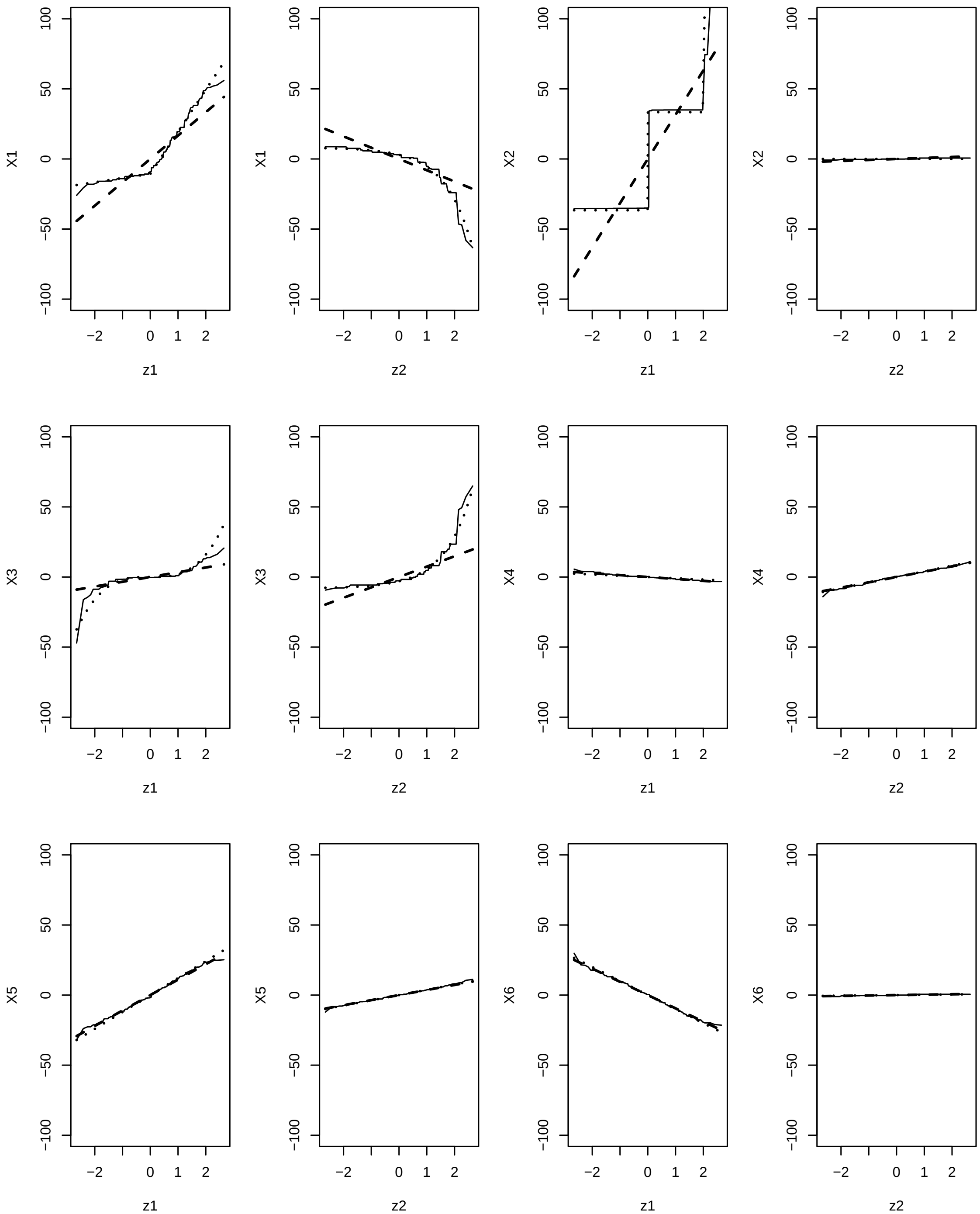

Figure 2: Parametric linear (dashed lines), nonparametric (solid lines) and true (dotted lines) association between latent and manifest variables. Data simulated using $n = 250, p = 6, q = 2$.

additive model regression, not only the η_{jl} but also the latent variables $\boldsymbol{z}$ 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 an ordering of the observations on the latent space.

Hence, the nonparametric estimation of the relationship between the manifest and latent variables we propose here 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 steps 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, the standard normal Φ:

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

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, Hastie, and Tibshirani 1989; Friedman and Silverman 1989; Wood 2000) or wavelets (see Sardy and Tseng 2004 and references therein). To constrain mono-

tonicity here, we use the `isotone` function of the `EBayesThresh` library in `R` 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.

The pq estimated functions corresponding to the estimated latent variables scores, possibly with confidence intervals, can then be visualized. 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 $\check{z}_i^{(l)} = \eta_l^{-1}(\tilde{z}_i^{(l)})$.

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 according to

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

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

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

In order to study the performance of the nonparametric estimator and compare it to the linear parametric one, we simulate 200 samples of size $n = 250$ from the FA model with $p = 6$ manifest variables and $q = 2$ latent 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 MSE for increasing sample sizes $n = 250$, $n = 500$ and $n = 1000$ for the nonparametric and the parametric linear estimators. For the nonlinear associations (first three rows) the MSE of the nonparametric estimator decreases with the sample size, while it remains constant with the parametric linear estimator, due to bias. For the linear associations (last three rows), both estimators have decreasing MSE as the sample size increases, with smaller MSE for the parametric model as expected. We conclude that the gain of using the isotone additive nonparametric estimator is important when some associations are potentially nonlinear.

Figure 3 here

4 Application

The data come from a large study in psychology (de Ribaupierre, Borella, and Delaloye 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

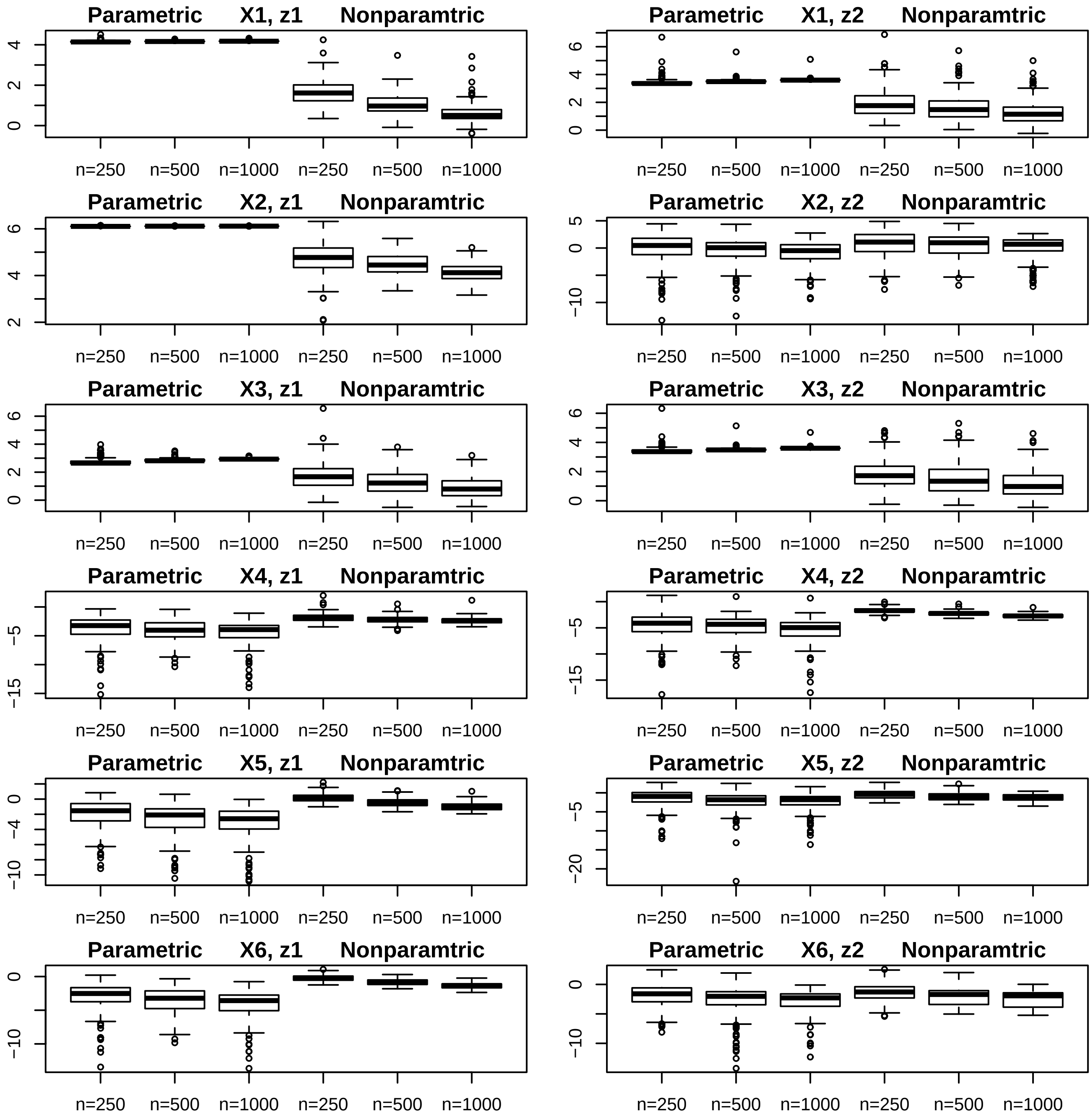

Figure 3: MSE of prediction (in logarithm) for the nonparametric and the parametric linear estimated association between latent and manifest variables. Data simulated using respectively $n = 250, 500, 1000$ and $p = 6, q = 2$, 200 samples.

median response times in milliseconds for two conditions of a negative priming task (Tipper and Cranston 1985), the third and forth 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 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 Figure 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 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 Figure 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.

Figures 4 and 5 here

So we estimate a parametric and nonparametric FA on both the original and the log-transformed data. In Figure 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 (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

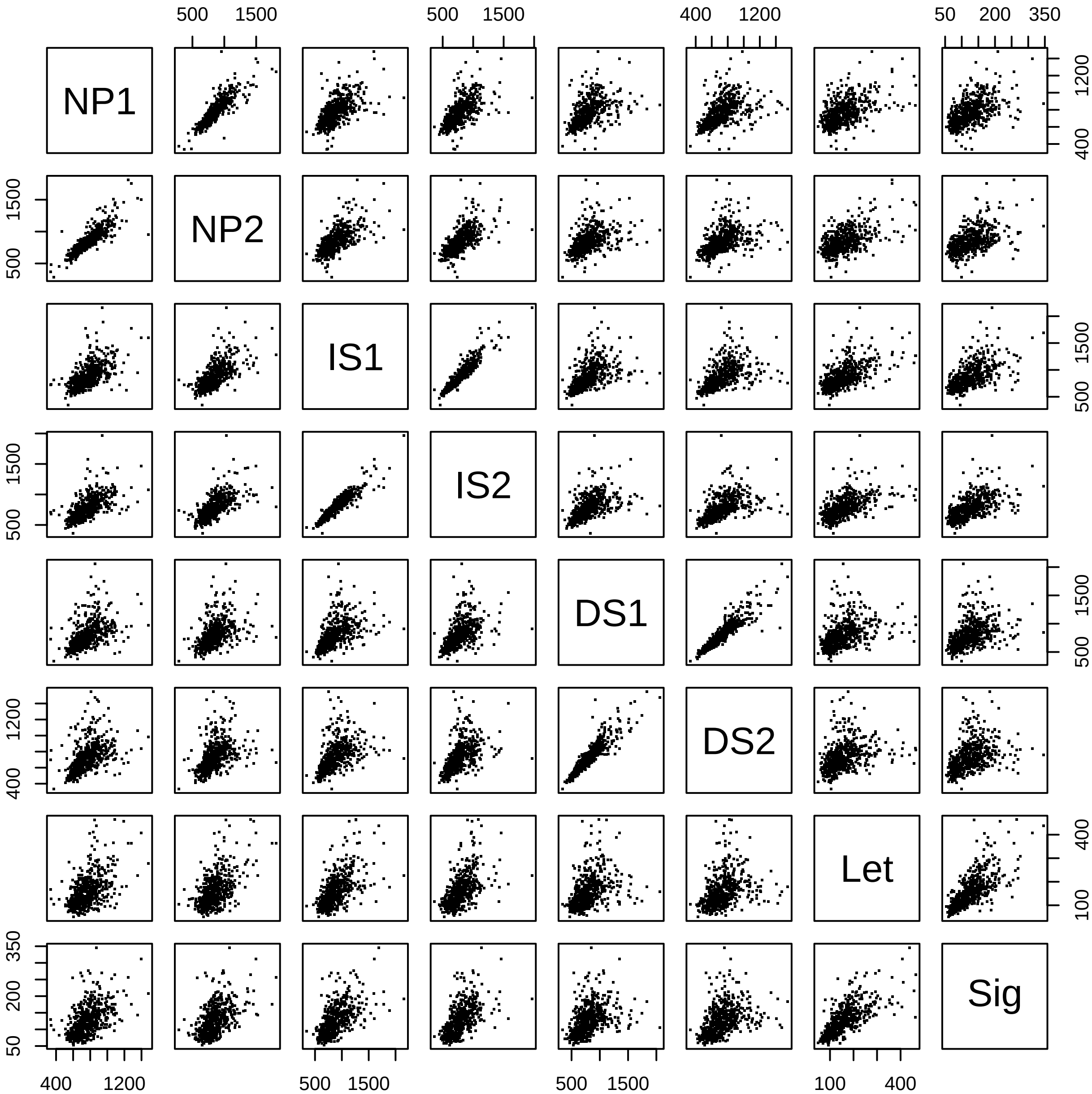

Figure 4: Scatter diagram of the selective attention data

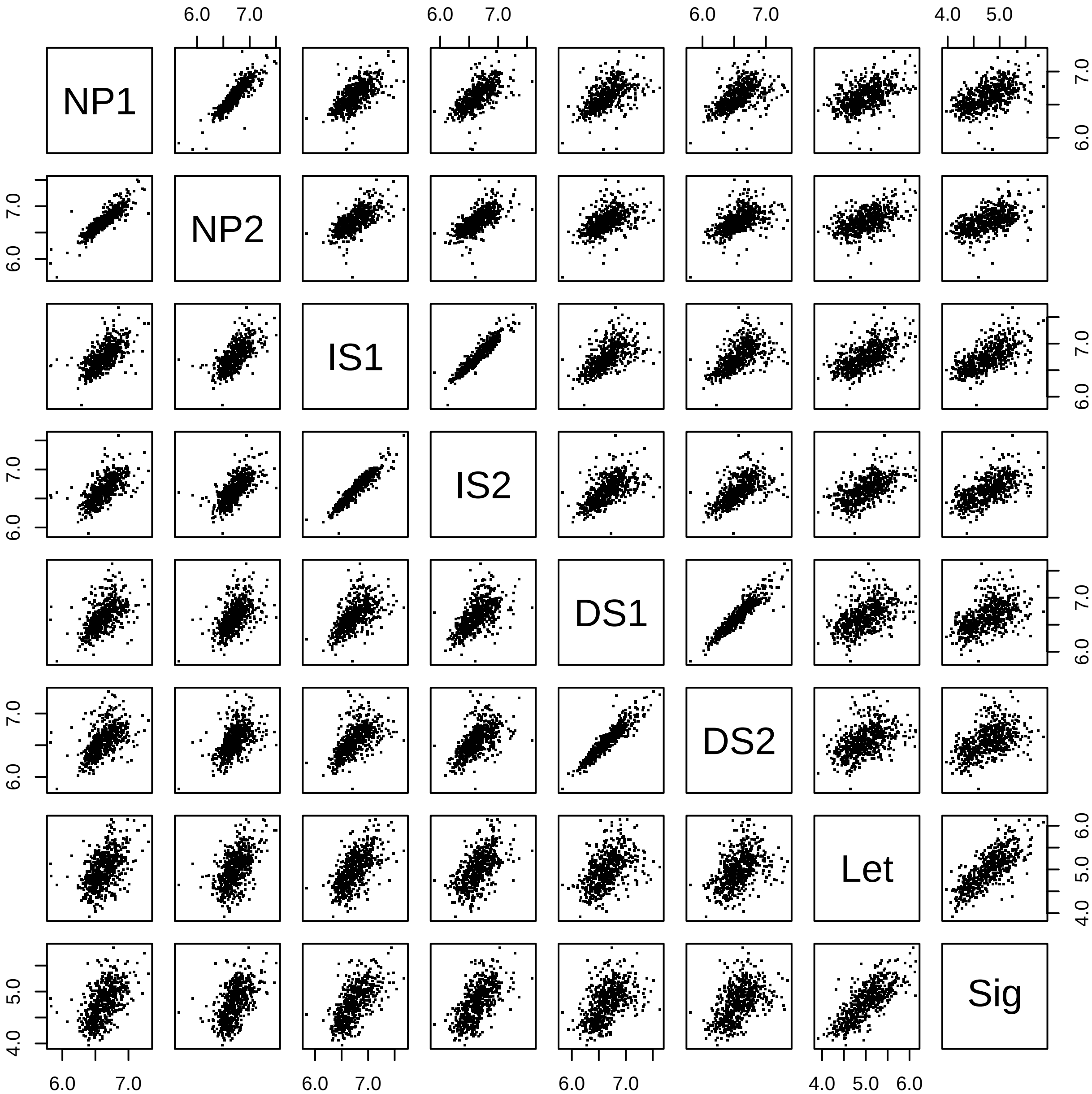

Figure 5: Scatter diagram of the log-transformed selective attention data

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.

Figures 6 to 8 here

5 Conclusion

FA models are very widely used in many disciplines. They suppose linear associations between manifest and latent variables, an 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 model.

The model can also be extended to the generalized case when the conditional distribution of the manifest variables belongs to the exponential family. For non-Gaussian distributions, Moustaki and Knott (2000) and Moustaki and Knott (2000) develop Generalized Linear Latent Variable Models (GLLVM) by extending the work of Moustaki (1996) and Sammel, Ryan, and Legler (1997) for mixed binary and metric variables (the latter with covariate effects as well) and Bartholomew and Knott (1999) for categorical variables. In that case, the framework of generalized linear models (McCullagh and Nelder 1989) is borrowed and the conditional distribution of the manifest variables given the latent ones is taken from the exponential family. The distributional assumptions on the latent variables remains the same. Our methodology could also be implemented in this more general setting.

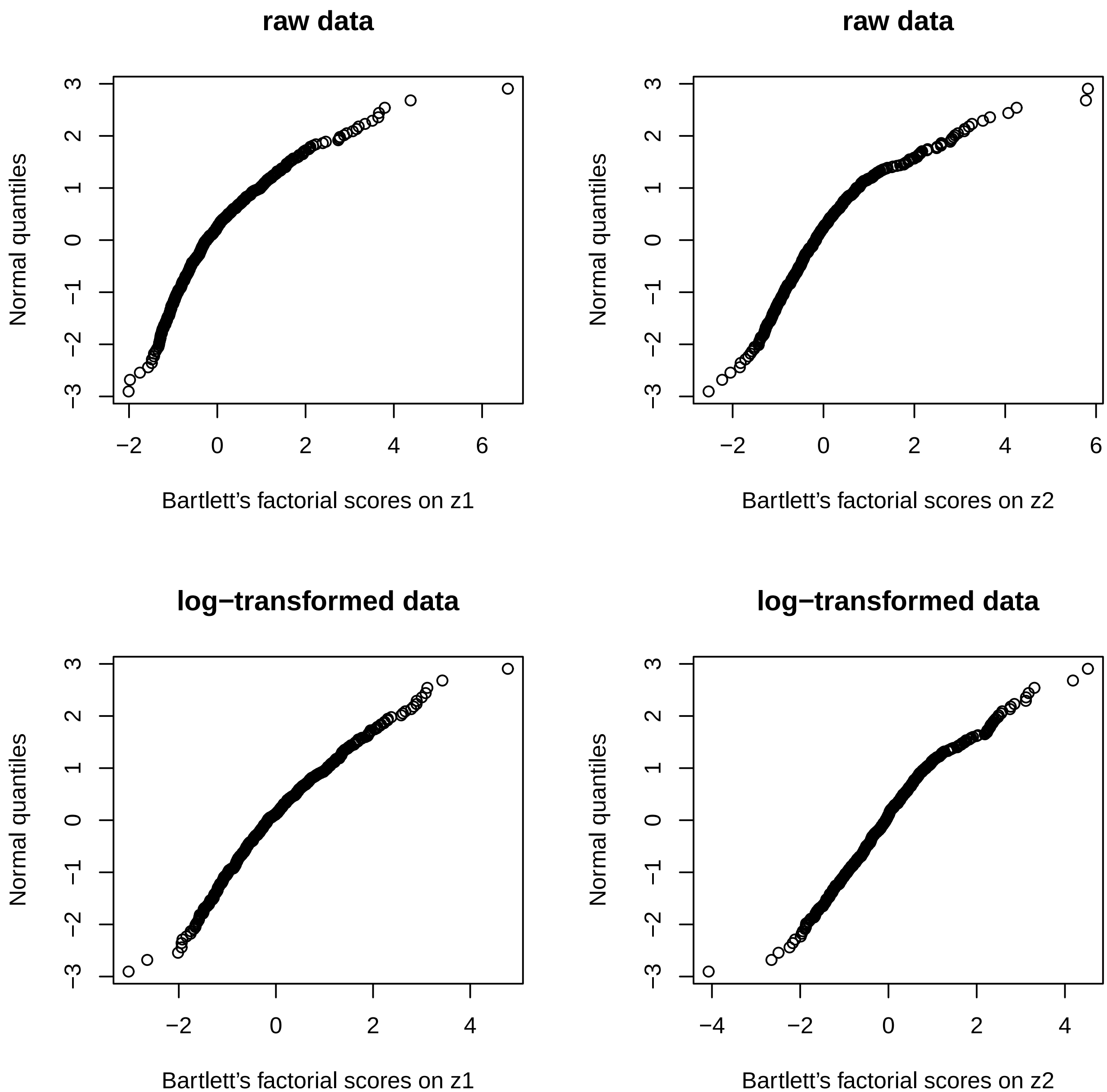

Figure 6: Scatter diagram of normalized versus raw Bartlett's factorial scores for the psychology example.

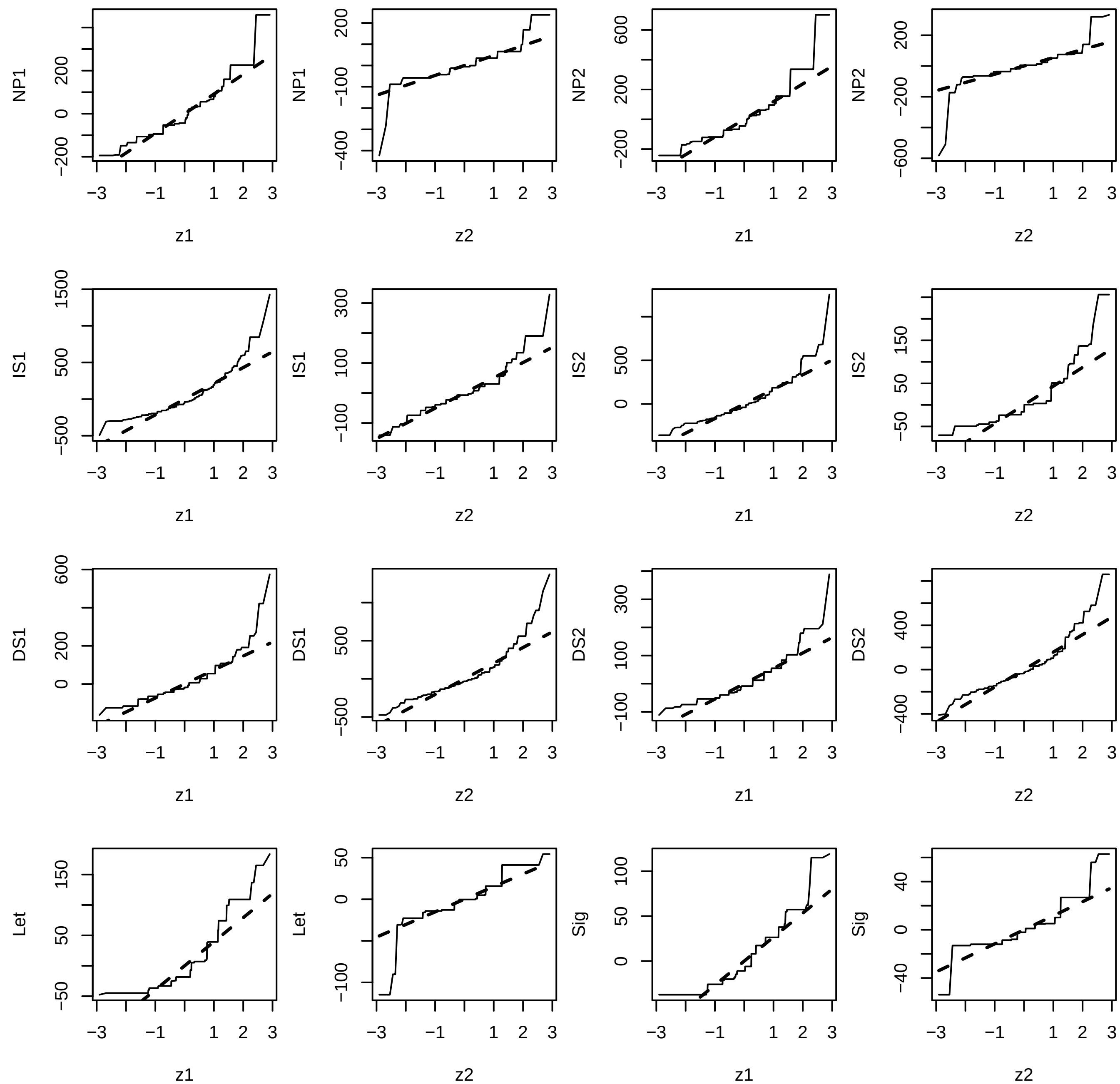

Figure 7: Nonparametrically (solid line) and parametrically (dashed line) estimated relationships between the latent and the manifest variables of the psychology example.

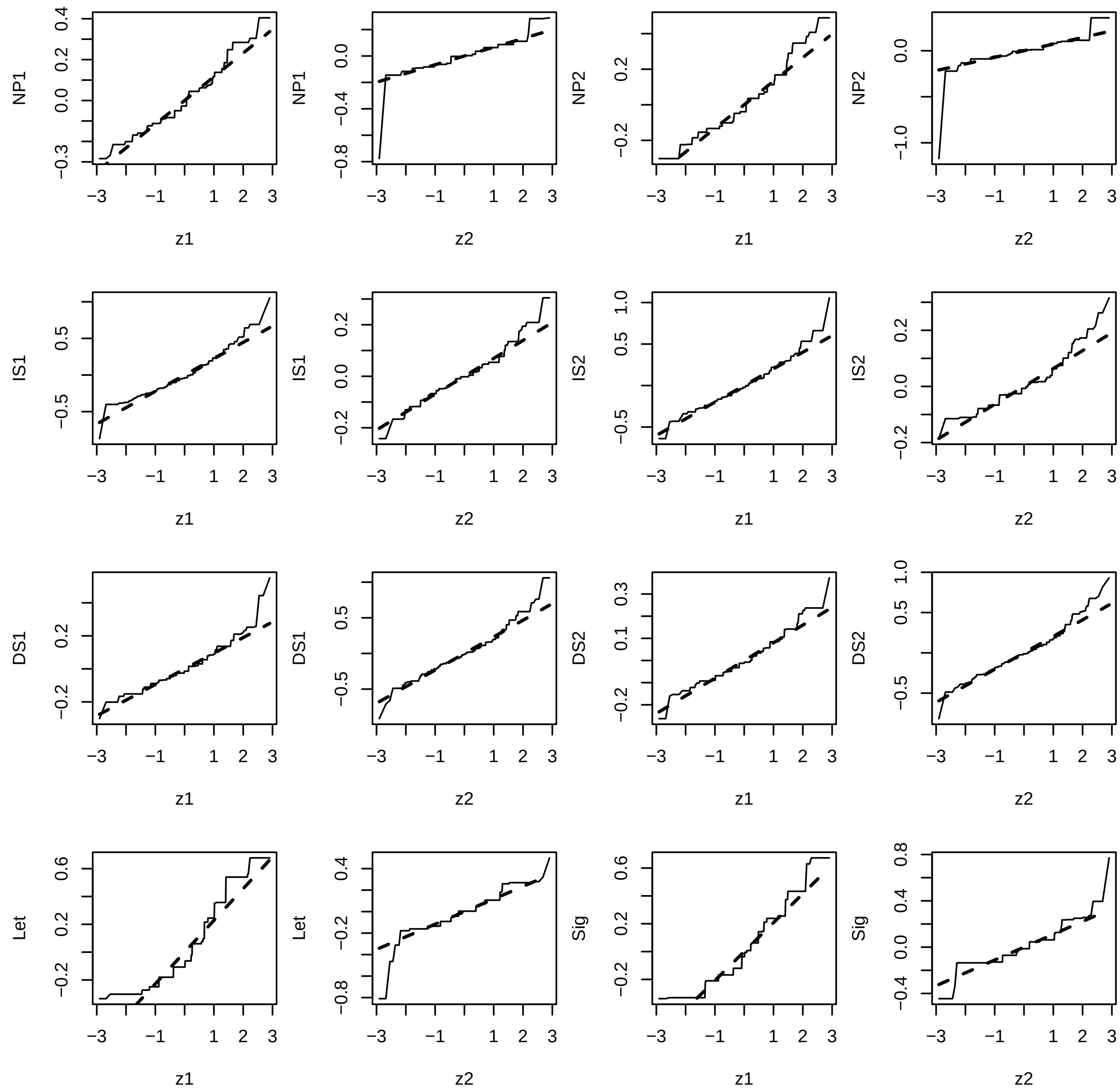

Figure 8: Nonparametrically (solid line) and parametrically (dashed line) estimated relationships between the latent and the manifest variables of the psychology example (log-transformed).

References

Bartholomew, D. J. (1988). The sensitivity of latent trait analysis to choice of prior distribution. *British Journal of Mathematical and Statistical Psychology 41*, 101–107.

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

Browne, M. W. (1984). Asymptotically distribution-free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology 37*, 62–83.

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

Conne, D., E. Ronchetti, and M.-P. Victoria-Feser (2010). Goodness-of-fit for generalized linear latent variables models. *Journal of the American Statistical Association*. To appear.

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

Dempster, F. N. and C. J. Brainerd (1995). *Interference and inhibition in cognition*. San Diego: Academic Press.

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

Dupuis Lozeron, E. and M.-P. Victoria-Feser (2010). Robust estimation of constrained covariance matrices for confirmatory factor analysis. *Computational Statistics and Data Analysis*. Forthcoming.

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

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

Hall, P., H.-G. Mller, and F. Yao (2008). Modelling sparse generalized longitudinal observations with latent gaussian processes. *Journal of the Royal Statistical Society, Series B 70*, 703–723.

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

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

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

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

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

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

Mammen, E. and K. Yu (2007). Additive isotone regression. In J. A. Wellner, E. A. Cator, G. Jongbloed, C. Kraaikamp, and H. P. Lopuhaä (Eds.), *Asymptotics: Particles, processes and inverse problems – Institute of mathematical statistics lecture notes – Monograph Series 55*, pp. 179–195. Institute of Mathematical Statistics.

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

Mavridis, D. and I. Moustaki (2009). The forward search algorithm for detecting aberrant response patterns in factor analysis for binary data. *Journal of Computational and Graphical Statistics 18*, 1016–1034.

McCullagh, P. and J. A. Nelder (1989). *Generalized Linear Models*. London: Chapman and Hall. Second edition.

Moustaki, I. (1996). A latent trait and a latent class model for mixed observed variables. *British Journal of Mathematical and Statistical Psychology 49*, 313–334.

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

Moustaki, I. and M.-P. Victoria-Feser (2006). Bounded-bias robust inference for generalized linear latent variable models. *Journal of the American Statistical Association 101*, 644–653.

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

Salthouse, T. A. (1991). *Theoretical perspectives on cognitive aging*. Hillsdale: Lawrence Erlbaum.

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

Sammel, M. D., L. M. Ryan, and J. M. Legler (1997). Latent variable models for mixed discrete and continuous outcomes. *Journal of the Royal Statistical Society, Series B 59*, 667–678.

Sardy, S. and P. Tseng (2004). AMlet, RAMlet and GAMlet: Automatic nonlinear fitting of additivemodel, robust and generalized with wavelets. *Journal of Computational and Graphical Statistics 13*, 283–309.

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

Stone, C. J. (1985). Additive regression and other nonparametric models. *The Annals of Statistics 13*, 689–705.

Tipper, S. P. and M. Cranston (1985). Selective attention and priming:inhibitory and facilitatory effects of ignored primes. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology 37A(4)*, 591–611.

Tipping, M. E. and C. M. Bishop (1999). Probabilistic principal component analysis. *Journal of the Royal Statistical Society, Series B 61*, 611–622.

Wood, S. N. (2000). Modelling and smoothing parameter estimation with multiple quadratic penalties. *Journal of the Royal Statistical Society, Series B 62*, 413–428.

Yuan, K.-H. and P. M. Bentler (1998). Robust mean and covariance structure analysis. *British Journal of Mathematical and Statistical Psychology 51*, 63–88.