Nonlinear elliptical models for correlated data

Cibele M. Russo[1], Gilberto A. Paula[1] and Reiko Aoki[2]

[1] Instituto de Matemática e Estatística, Universidade de São Paulo, Brazil
[2] Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, Brazil

The variety of distributions encompassed by the elliptical class represents an alternative for robust modeling, since it includes light and heavy tailed distributions such as Student-t, power exponential, logistic, among others. Particularly for nonlinear models, few works can be found in the literature considering elliptical errors. For linear models recent references are the works by Savalli et al. (2006) and Osorio et al. (2007). In this work we discuss the development and analysis of nonlinear elliptical models for correlated data, which represents a new approach for obtaining robust estimates against outlying observations, in order to avoid sensitivity to extreme points. Let $\mathbf{y}_i$ be an m_i-dimensional vector such that E$(\mathbf{y}_{ij}) = \boldsymbol{\mu}_{ij} = \mathbf{f}(\mathbf{x}_{ij}, \boldsymbol{\beta})$, for $i = 1, \ldots, n$. A possible mixed-effects model for $\mathbf{y}_i$ is given by

$$\mathbf{y}_i = \mathbf{f}(\mathbf{x}_i, \boldsymbol{\beta}) + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i, \quad i = 1, \ldots, n,$$

with $\mathbf{f}(\mathbf{x}_i, \boldsymbol{\beta}) = (f(\mathbf{x}_{i1}, \boldsymbol{\beta}), \ldots, f(\mathbf{x}_{im_i}, \boldsymbol{\beta}))^T$ being an m_i-dimensional nonlinear function of $\boldsymbol{\beta}$, $\mathbf{x}_{ij}$ is a vector of explanatory variable values, $\mathbf{x}_i = (\mathbf{x}_{i1}^T, \ldots, \mathbf{x}_{im_i}^T)^T$, $\mathbf{Z}_i$ is a matrix of known constants, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^T$ a vector of unknown parameters, $\mathbf{b}_i = (b_{i1}, \ldots, b_{ir})^T$ a vector of unobserved random regression coefficients such that

$$\begin{bmatrix} \mathbf{y}_i \\ \mathbf{b}_i \end{bmatrix} \sim \mathrm{El}_{m_i+r}\left\{ \begin{pmatrix} \mathbf{f}(\mathbf{x}_i, \boldsymbol{\beta}) \\ \mathbf{0} \end{pmatrix}; \begin{bmatrix} \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i^T + \sigma^2\mathbf{I}_{m_i} & \mathbf{Z}_i\mathbf{D} \\ \mathbf{D}\mathbf{Z}_i^T & \mathbf{D} \end{bmatrix} \right\},$$

where the matrices $\boldsymbol{\Sigma}_i = \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i^T + \sigma^2\mathbf{I}_{m_i}$, $\mathbf{D}$, and $\mathbf{Z}_i\mathbf{D}$ are proportional to the variance-covariance matrices $\mathrm{Var}(\mathbf{y}_i)$, $\mathrm{Var}(\mathbf{b}_i)$, and $\mathrm{Cov}(\mathbf{y}_i, \mathbf{b}_i)$, respectively, by a quantity $\alpha_i > 0$ which depends on the assumed elliptical distribution (for the normal case one has $\alpha_i = 1$). We will work with the marginal model, namely $\mathbf{y}_i \sim \mathrm{El}_{m_i}(\mathbf{f}(\mathbf{x}_i, \boldsymbol{\beta}); \boldsymbol{\Sigma}_i)$. Because estimation is not straightforward in this case, we propose an estimating iterative algorithm, implemented in R language. Some inferential results are derived and robust aspects of the maximum likelihood estimates are discussed. For sensitivity investigation, we present diagnostic tools including residual analysis and local influence graphics, which can help on validation and model choice. As numerical illustration, we apply the obtained results to a kinetics longitudinal data set (Vonesh and Carter, 1992), which was previously analysed under normality but presented outlying observations and can be better accommodated using elliptical non-Gaussian distributions. We conclude that the Student-t distribution with few degrees of freedom presents a better fit in this case, providing robustness of the maximum likelihood estimates against outlying observations.

References

Osorio F.; Paula, G. A. and Galea, M. (2007). Assessment of local influence in elliptical linear models with longitudinal structure. *Computat. Statist. Data Anal.* **51**, 4354-4368.

Savalli, C.; Paula, G. A. and Cysneiros, F. J. A. (2006). Assessment of variance components in elliptical linear mixed models. *Statist. Modelling* **6**, 59-76.

Vonesh, E. F. and Carter, R. L. (1992). Mixed-effests nonlinear regression for unbalanced repeated measures. *Biometrics* **48**, 1-17.