Journal of Multivariate Analysis 134 (2015) 82–98

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Robust linear functional mixed models

Marco Riquelme [a], Heleno Bolfarine [b], Manuel Galea [c,*]

[a] Facultad de Ciencias Básicas, Universidad Católica del Maule, Talca, Chile

[b] Departamento de Estatística, Universidade de São Paulo, São Paulo, Brazil

[c] Departamento de Estadística, Pontificia Universidad Católica de Chile, Chile

ARTICLE INFO

Article history:
Received 5 December 2012
Available online 8 November 2014

AMS subject classifications:
62H12
62F12
62J20

Keywords:
Corrected score estimators
Local influence
Elliptical distributions
Linear functional mixed models

ABSTRACT

In this paper we propose a linear functional model with normal random effects and elliptical errors, thus extending the standard normal models considered previously. The corrected score approach (Nakamura, 1990) is used for parameter estimation and the resulting estimators are shown to be consistent and asymptotically normal. The local influence approach (Cook, 1986) is used for assessing influence of small perturbations on the parameter estimates. A simulation study is presented illustrating the good performance of the proposed approach, including the robustness property for the heavier tail models.

1. Introduction

The linear mixed normal model (LMNM) has been frequently used in repeated measures, grouped and longitudinal data. Applications have been reported in different areas such as agriculture, biology, economy, geophysics and social sciences [13]. The relevance of this model seems in part explained by its flexibility in explaining intraunits and interunits correlations, typically encountered in longitudinal [25] and grouped data [23]. It is also suitable for fitting balanced as well as unbalanced data using software specifically designed for its implementation [42]. Moreover, as special cases of the mixed models we have the classical linear model, the variance components model and the hierarchical (multilevel) model [33].

On the other hand, the measurement error problem appears frequently in several areas of knowledge. Fields, such as agriculture, medicine, engineering, psychology, education, and finance are some disciplines presenting situations where covariates are contaminated by measurement errors. For this reason there has been extensive research in the measurement error problem, as seen, for example, in the books by Fuller [16], Carroll et al. [9], Buonaccorsi [8], and Wu [44]. If the covariates are measured with errors and these are not properly accounted for, it can lead us to incorrect statistical inferences. For example, a significant covariate can be considered not significant, as seen in [44]. If the measurement errors are not accounted for (naive approach), then the parameter estimates are biased and not consistent, as seen, for example, in [5,37,16,8]. Therefore, it is important to investigate the combination of random effects and measurement errors in mixed effects models. As pointed out in [12], open problems (Chapter 12, p. 328) section, "it is often the case in practical problems that covariate values are typically measured with nonnegligible measurement error and that inference for such models are not well developed". Wang et al. [41] and Lin and Carroll [29] studied the bias for estimators of the variance components for the

* Correspondence to: Departamento de Estadística, Pontificia Universidad Católica de Chile, Avenida Vicuña Mackenna 4860, Macul, Santiago, Chile.
E-mail address: mgalea@mat.puc.cl (M. Galea).

http://dx.doi.org/10.1016/j.jmva.2014.10.008
0047-259X/

mixed models with measurement errors using the simulation–extrapolation approach (SIMEX). Zhong et al. [45] proposed a mixed-effects model with covariates measured with errors, under the normality assumption using the corrected score approach [32]. More recently Wu [44] proposed a normal mixed effects model with measurement errors using the likelihood approach. See also [8].

In this paper we extend the model proposed by Zhong et al. [45], relaxing the normality assumption for the measurement errors. It is well known that the normal distribution is not always suitable for modeling multivariate continuous data, as stated in [26,27]. In this paper we consider the family of elliptical distributions, which provides greater flexibility in modeling heavier tails or extreme observations frequently observed in multivariate symmetric data sets. Apart from this flexibility, it preserves several of the well-known properties of the normal distribution, allowing one to derive attractive explicit solution forms. This class of distributions, which contains the normal, Student-t, contaminated normal, power exponential and slash distributions, among others, has received greater attention in the statistical literature, as can be seen in [24,14,15]. More specifically, in this paper, we study and develop fixed (functional) mixed linear measurement error models where measurement errors follow an elliptical distribution, generalizing then previous normal linear mixed models considered in the literature. We call it the "*elliptical functional linear mixed models*" family. The corrected score approach [32] is used for parameter estimation which guarantees consistent and asymptotic normal distributed estimators.

A semiparametric Bayesian approach to elliptical measurement error models is developed in [10]. Ultrastructural, structural and functional elliptical measurement error models are investigated in [7,3,2,39]. It is also the case that recent statistical literature has experienced growing interest in elliptical generalizations of normal models [34,4], perhaps influenced by the development of more efficient computational procedures.

The paper is organized as follows. In Section 2, we present basic results on elliptical models. In Section 3, we define the linear functional mixed model and implement the correct score approach [32,18]. In Section 4 we study large sample properties for the corrected score estimators. In Section 5 we develop a study on local influence [46], using the corrected likelihood function under different perturbation schemes. Section 6 presents results of a simulation study where the main object is to illustrate the performance of the estimation approach and also study robustness of the model proposed using the local influence approach. Finally, Section 7 contains some concluding remarks.

2. Elliptical distributions

The class of elliptical distributions has received increasing attention in recent statistical literature (see [15,14,1,34,4]). We present in the following some properties that are crucial in developing the main results of this paper.

Definition 2.1. A random variable $X \in \mathbb{R}^p$ $(p \geq 2)$ is said to follow a distribution with elliptical contours if its characteristic function can be written as

$$\psi_X(t) = \exp\{it^T\mu\}\phi(t^T\Omega t), \tag{1}$$

in which $\mu \in \mathbb{R}^p$ denotes the location parameter, $\Omega \in \mathbb{R}^{p\times p}$ denotes the scale parameter (symmetric and positive definite matrix), $\phi : \mathbb{R}^p \to \mathbb{R}$ is a generator of characteristic functions, $i = \sqrt{-1}$ and $t \in \mathbb{R}^p$.

If X has an elliptical distribution with characteristic function given in (1), we write $X \sim El_p(\mu, \Omega, \phi)$ or, simplifying, $X \sim El_p(\mu, \Omega)$. In general terms, there are situations in which the random vector X does not possess a density function, as is the case when $\operatorname{rank}(\Omega) = r(<p)$ (singular case), the density function is not defined in $\mathbb{R}^p$. However, it is always possible to define a density function in a lower dimensional space (smaller than the rank of the scale matrix). On the other hand, if $\operatorname{rank}(\Omega) = p$ (nonsingular), the density function is well defined in relation to the Lebesgue measure on the space $\mathbb{R}^p$.

Definition 2.2. Assuming that $\operatorname{rank}(\Omega) = p$, we have that the random vector X has density function given by

$$f_X(x) = |\Omega|^{-1/2} g((x-\mu)^T\Omega^{-1}(x-\mu)), \quad x \in \mathbb{R}^p, \tag{2}$$

for some function g, $(g(u) \geq 0, u \geq 0)$. The function g is typically known as the density generator function satisfying

$$\int_0^\infty u^{\frac{p}{2}-1} g(u)du < \infty. \tag{3}$$

If X has elliptical distribution with density function given by (2), we write $X \sim El_p(\mu, \Omega, g)$. Some examples of elliptical distributions, densities and function g are given in the following assuming that $\operatorname{rank}(\Omega) = p$ (see also [31,17]).

Example 2.1. The p-variate normal distribution, denoted by $N_p(\mu, \Omega)$:

$$f_X(x) = (2\pi)^{-p/2}|\Omega|^{-1/2} \exp\left\{-\frac{1}{2}(x-\mu)^T\Omega^{-1}(x-\mu)\right\}, \tag{4}$$

and

$$g(u) = (2\pi)^{-p/2}e^{-u/2}, \quad u \geq 0.$$