Bayesian analysis of flexible measurement error models

Luz Marina Rondon[1,a], *Heleno Bolfarine*[2]

[1]lumarp@gmail.com, Departamento de Estadística, Universidad Nacional de Colombia
[a] Instituto de Matemática e Estatística, Universidade de São Paulo
[2]hbolfar@ime.usp.br, Instituto de Matemática e Estatística, Universidade de São Paulo

Regression models under the assumption of independent and normally distributed errors are very useful statistical tools. However, in practice, there are data sets in which the presence of explanatory variables measured with error can substantially affect the good performance of the parameter estimators. Therefore, the regression models where the measurement errors are taken into account, termed measurement error models, have been studied by many authors. Most of the available proposals are based on the assumption of a random term that exhibits a normal distribution, which is known to be vulnerable in the presence of outlying observations. Nonetheless, most of these works do not allow the simultaneous presence of vectors of explanatory variables with and without measurement error as well as the presence of nonlinear effects whose functional form is unknown. Thus, in this paper, a measurement error model which admits vectors of explanatory variables with and without measurement error as well as the presence of nonlinear effects approximated by using B-splines, is introduced. The model investigated is the structural version, as the error-prone variables follow scale mixtures of normal distributions ($\mathcal{SMN}$) such as Student-t, slash, contaminated normal, Laplace and symmetric hyperbolic distributions. In addition, the model considered assumes an error term whose distribution also belongs to the $\mathcal{SMN}$ class. Some of these distributions present heavier tails than the normal ones. Accordingly, the regression models based on them seem to be reasonable choice for robust inference. However, given the complexity of the model investigated, the Bayesian approach is considered.

The rest of this paper is organized as follows: Section 2 describes the $\mathcal{SMN}$ class. Section 3 formulates the flexible measurement error models that allow explanatory variables with and without measurement error, as well as the presence of a nonlinear effects, which are approximated by using B-splines. Section 4 is dedicated to the specification of the prior distributions and the Gibbs sampler to draw samples from the posterior distribution of the interest parameters. Model selection is also discussed in that section. Section 5 presents a simulation study where the performance of the proposed MCMC algorithm is illustrated. That section also presents the function `fmem()` of the R package **BayesGESM**, which has been developed to provide a easy way to apply the statistical methodology presented in this paper. In Section 6, the proposed methodology is applied to a real data set. It reveals that the onset of measurement errors in one of the covariates is a better model than the model which does not take into account measurement errors.