Robust estimation for linear regression with asymmetric errors with applications to log-gamma regression

Ana Bianco[1], Marta García Ben[2] and Víctor J. Yohai[2]

[1] University of Buenos Aires, Instituto de Cálculo, Pabellón 2, Ciudad Universitaria 1428 Buenos Aires, Argentina
[2] University of Buenos Aires, Departamento de Matemáticas, Pabellón 1, Ciudad Universitaria 1428 Buenos Aires, Argentina

Keywords: log-gamma regression, robust estimates, MM-estimates, asymmetric errors

Abstract

We will focus our attention on the log-gamma regression model. However, our main results remain valid for a more general class of regression models with a continuous asymmetric errors. In our case it is convenient to parametrize the gamma family so that the family of densities is given by

$$f(z,\alpha,\mu)=\frac{\alpha^{\alpha}}{\mu^{\alpha}\,\Gamma(\alpha)}\,z^{\alpha-1}\,e^{-(\alpha/\mu)\,z}.$$

where μ is the expected value and α is the shape parameter. This distribution is denoted by $\Gamma(\alpha,\mu)$,

We observe response variables z_i and covariate vectors $\mathbf{x}_i$, $(1\le i\le n)$, such that $z_1,\ldots\, z_n$ are independent, $z_i\sim\Gamma(\alpha,\mu_i)$ and their conditional expected values and the covariates are related by the equation $\log(\mu_i)=\mathbf{x_i'}\beta,$. The parameter α does not depend on the covariates, hence it is the same for all the observations and β is the vector of regression coefficients to be estimated. If we consider the transformed variables $y_i=\log(z_i)$, we obtain the linear regression model

$$y_i=\mathbf{x}_i'\beta+u_i, \tag{1}$$

where the errors u_i are i.i.d. with distribution $\log\Gamma(\alpha,1)$.

The maximum likelihood estimator (MLE) of β can be obtained minimizing the deviance. Unfortunately, the MLE is not robust and can be upset by a few outliers.

Since the deviance components can be factorized as $d_i=2\alpha\ d^*(y_i,x_i,\beta)$,where $d^*(y_i,x_i,\beta)$ does not depend on α and its distribution does not depend on β, the MLE can be defined as the minimum of

$$\sum_{i=1}^{n}d^*(y_i,x_i,\beta).$$

We propose a family of robust estimates based on the same idea as the MM-estimators for the linear regression model introduced by Yohai (1987). Let us consider first M-estimators of β given by the minimization of

$$\sum_{i=1}^{n}\rho_1(\sqrt{d^*(y_i,x_i,\beta)}), \tag{2}$$

where ρ_1 is bounded, monotone increasing, differentiable and $\rho_1(0)=0$.

We prove that the estimates given by (2) are Fisher-consistent. Then, even if the distribution of d^* is asymmetric it is not necessary to introduce a correction term to get asymptotic unbiased estimates. This result is not only valid for a regression model with log-gamma residuals, but for any model of the form (1) where the u_i's have a density f_0 strictly unimodal and continuous.

The function ρ_1 can be chosen in a family defined by $\rho_1(t) = \rho_0(t/k)$ with ρ_0 satisfying the assumptions given above. One possible choice is the family of Tukey's bisquare functions.

The M-estimates are asymptotically normal under mild regularity conditions. Their asymptotic covariance matrix differs from that of the the MLE by a factor that only depends on k_1 and α. If $\rho_0'(0) = 0$ and $\rho_0''(0) > 0$, there exists a constant $k^*(\alpha)$ for which the M-estimate of β achieves a desired efficiency under the central model. So, in order to calibrate efficiency of the M- estimate it is necessary to have an estimate of α.

A way to compute simultaneously an estimate of α and an initial estimate of β is through an S-estimate. Consider the M-scale estimate $s(\beta)$ that solves

$$\frac{1}{n}\sum_{i=1}^{n}\rho_0\left(\frac{\sqrt{d^*(y_i, x_i, \beta)}}{s(\beta)}\right) = b \tag{3}$$

and define $s_o = \min_\beta\ s(\beta)$. Then the S-estimator of β is $\widehat{\beta} = \arg\min_\beta\ s(\beta)$. Since for any α, we have a value $s^*(\alpha)$ that solves the equation

$$E_\alpha\left(\rho_o\left(\frac{\sqrt{d^*}}{s^*(\alpha)}\right)\right) = b,$$

we define $\widehat{\alpha} = (s^*)^{-1}(s_o)$.

If we choose b in (3) so that the M-scale estimate has breakdown point 0.5 and then we compute the M-estimator in (2) taking $\rho_1(t) = \rho_0(t/k_1)$ with $k_1 = k_1(\widehat{\alpha})$, we will obtain a MM-estimator which has simultaneously high breakdown point and the desired efficiency under the central log-gamma regression model.

Please fill in this form and mail it together with your abstract.

My abstract fits best to topic number 20 (**“PLEASE INDICATE HERE THE NUMBER FROM THE LIST OF TOPICS BELOW WHICH BEST FITS TO YOUR ABSTRACT”**.).

List of Topics:

1. Algorithms
2. Applications
3. Biostatistics
4. Computing and graphics
5. Data analysis
6. Data mining
7. Economics, finance
8. Efficiency and robustness
9. Functionals and bias
10. Fuzzy statistics
11. Geostatistics
12. Inference for robust methods, model testing
13. Location depth and regression depth
14. Multivariate methods
15. Neural networks
16. Rank-based methods
17. Regression quantiles, trimming
18. Robust covariance
19. Robust designs
20. Robust regression
21. Time series analysis
22. Wavelets
23. Other (please specify)