Assessment of influence on the score test statistic in non-linear regression models

Tatjana von Rosen[1] **and Karin Stål**[2]

[1]*Department of Statistics, Stockholm University, Sweden*
[2]*The National Agency for Education, Sweden*

Abstract

Regression analysis is a statistical technique for exploring the relationships between variables. Frequently, regression models are used to describe the dependence between a response variable and one or several explanatory variables. The parameters in the regression model are estimated based on observed data. However, some observations have a greater impact on the estimated model than others. The regression model considered in this work is the nonlinear model with an additive error term

$$\mathbf{y} = \mathbf{f}(\mathbf{X}, \boldsymbol{\theta}) + \boldsymbol{\epsilon},$$

where $\mathbf{f}(\mathbf{X}, \boldsymbol{\theta}) = (\mathbf{f}(\mathbf{X}_1, \boldsymbol{\theta}), \ldots, \mathbf{f}(\mathbf{X}_n, \boldsymbol{\theta}))^T = (\mathbf{f}_1(\boldsymbol{\theta}), \ldots, \mathbf{f}_n(\boldsymbol{\theta}))^T$, $\mathbf{X}$ is a $n \times p-$matrix of known explanatory variables, $\mathbf{y}$ is the n-vector of responses, $\boldsymbol{\theta}$ is a q-vector of unknown parameters, $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I}_n)$, and $\mathbf{I}_n$ denote the identity matrix of size n.

A well known example of a nonlinear model is the Michaelis-Menten model

$$y = \frac{\theta_1 x}{\theta_2 + x} + \varepsilon,$$

which is used in enzyme kinetics. It relates the initial velocity of an enzymatic reaction, y, to the substrate concentration, x. The parameter θ_1 is the maximum velocity of the enzymatic reaction, representing the asymptotic value of f as $x \to \infty$; θ_2 is the half-velocity parameter, representing the value of x when the velocity of the reaction reaches one-half of its ultimate value.

The existing influence measures in regression analysis are constructed to measure the impact of observations on the parameter estimates or the fitted values. However, it is of interest to assess the influence of observations on hypothesis testing. We will derive a diagnostic measure for assessing the influence of single and multiple observations on the score test statistic [**?**], both in linear and nonlinear regression. The proposed diagnostic measure is derived using the differentiation approach.

Keywords

non-linear regression, influential observations, Score test.

References

[1] Chen, C.F. (1983). Score tests for regression models. *JASA 78*, 158–161.

[2] Chen, C.F. (1985). Robustness aspects of score tests for generalized linear and partially linear regression models. *Technometrics 27*, 277–283.

[3] Li, B. (2001). Sensitivity of Rao's score test, the Wald test and the likelihood ratio test to nuisance parameters. *J Stat Plan Inference 97*, 57–66.

[4] Lustbader, E.D. and Moolgavkar, S.H. (1985). A diagnostic statistic for the score test. *JASA 80*, 375–379.

[5] Rao, C.R. (1948). Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation. *Proceedings of the Cambridge Philosophical Society 44*, 50–57.

[6] Vanegas, L.H., Rondón, L.M. and F.J.A. Cysneiros (2013). Assessing robustness of inference in symmetrical nonlinear regression models. *Commun. Stat. - Theory Methods 42*, 1692–1711.