INTERNATIONAL CENTRE
OF EXCELLENCE FOR
EDUCATION IN
MATHEMATICS

A Sensitivity Analysis of the Ordinary Least Squares Slope Vector

Nathan Bock, Department of Mathematics and Statistics, La Trobe University

The ordinary least squares (OLS) method is commonly used to estimate predictor coefficients under an assumed multiple linear regression model. Whilst the OLS method can be useful, it can be highly influenced by certain observational types. Some of these observations can be identified visually, but what if there are others that seemingly conform to the rest of the data? How can they be identified?

I sought answers to these questions as part of my AMSI summer vacation project. In particular, influence diagnostics were studied for the estimate of the ordinary least squares slope vector. To achieve this outcome the bulk of the study was spent looking at the influence function (IF) and two of its derivations, the Sample and the Empirical influence functions. The purpose of the SIF and the EIF is to identify which observations in a data set may be influential.

Consider a data set of n observations denoted $(y_1,x_1),\ldots,(y_n,x_n)$. We begin by calculating the slope estimate with and without the i^{th} observation and denote these estimates as β and β_i. The SIF for the slope estimate is then defined to be,

$$SIF(\beta, F_n; x_i) = (n-1)[\beta - \beta_i]$$

If the removal of the i^{th} observation results in a large SIF value, then this observation must be influential. On the other hand, if the SIF value is small then the i^{th} observation has little influence on the data set.

A large part of the AMSI scholarship was spent finding the IF for the OLS slope vector estimator. The result of this study enabled what is known as the EIF to be obtained. The EIF for the i^{th} observation is,

$$EIF(\beta, F_n; y_i, x_i) = r_i S_{xx}^{-1}(x_i - \bar{x})$$

Where r_i is the i^{th} OLS residual, S is the inverse covariance matrix between the x's and $\bar{x}$ is the average of the observations.

Through a long mathematical process it can be shown that for a sufficiently large n,

$$EIF(\beta, F_n; y_i, x_i) \approx SIF(\beta, F_n; y_i, x_i)$$

So the final part of the project was spent looking at simulated and real-life examples that compared the SIF and EIF.

There are two key advantages when using the EIF as an influence diagnostic in that it is computationally efficient and can be used to understand why an observation may be influential. A disadvantage of the EIF method is that it is only an approximation to the SIF and the closeness of this approximation will depend on the sample size.

Stephen received an ICE-EM Vacation Scholarship in December 2006.
See http://www.ice-em.org.au/students.html#scholarships2007

Through the use of the simulations it became apparent that the EIF could successfully detect influential observations even when the sample size is small. This is important in practice because the requirement of large data sets can restrict the usefulness of an influence diagnostic. In conclusion the EIF is a useful method to use because of the fact that it is both computationally efficient and may be used to understand why observations can be influential.

Overall this scholarship has greatly enhanced my research and presentation skills and will form the basis of what I will be doing for my honors thesis. Here I will consider influence diagnostics for OLS under the more general single-index model (see, for e.g., Brillinger, 1983 and Li & Duan, 1989). I would like to thank my supervisor, Dr Luke Prendergast, AMSI and CSIRO for giving me this great experience.

References:

Brillinger, D.R. 1983. A Generalized Linear Model with “Gaussian” Regression Variables. Pages 97-114 of: Bickel, P.J., Doksum, K.A., & Hodges Jr. J.L. (eds). *A Festschrift for Erik L Lehmann*. Wadsworth International Group Belmont. California.

Li, K,-C., Duan N. 1989. Regression Analysis Under Link Violation. *The Annuls of Statistics* 17(3), 1009-1052.

Stephen received an ICE-EM Vacation Scholarship in December 2006.
See http://www.ice-em.org.au/students.html#scholarships2007