The choice of variables in multivariate regression: a non-conjugate Bayesian decision theory approach

P. J. Brown[*], T. Fearn[†] and M. Vannucci[‡]

SUMMARY

We consider the choice of explanatory variables in multivariate linear regression. Our approach balances prediction accuracy against costs attached to variables in a multivariate version of a decision theory approach pioneered by Lindley (?). We also employ a non-conjugate proper prior distribution for the parameters of the regression model, extending the standard normal-inverse Wishart by adding a component of error which is unexplainable by any number of predictor variables, thus avoiding the determinism identified by Dawid (?). Simulated annealing and fast updating algorithms are used to search for good subsets when there are very many regressors. The technique is illustrated on a near infrared spectroscopy example involving 39 observations and 300 explanatory variables. This demonstrates the effectiveness of multivariate regression as opposed to separate univariate regressions. It also emphasises that within a Bayesian framework more variables than observations can be utilised.

Some key words: Bayesian decision theory; Determinism; Multivariate regression; Near infrared spectroscopy; Non-conjugate distribution; Prediction; Quadratic loss; Simulated annealing; Utility.

Acknowledgement

This work is supported by the UK Engineering and Physical Sciences Research Council with a grant under the Stochastic Modelling in Science and Technology Initiative. We are grateful to the Flour Milling and Baking Research Association for providing the data.

[*]University of Kent at Canterbury, Institute of Mathematics and Statistics, Cornwallis Building, Canterbury, CT2 7NF, UK. FAX 01227-827932, email Philip.J.Brown@ukc.ac.uk

[†]University College London, UK

[‡]Texas A & M University, USA