

# Selection of variables in multivariate regression models for large dimensions

Muni S. Srivastava<sup>1</sup> and Tatsuya Kubokawa<sup>2</sup>

<sup>1</sup>*University of Toronto, Canada*

<sup>2</sup>*University of Tokyo, Japan*

## Abstract

The Akaike information criterion, AIC, and Mallows'  $C_p$  statistic have been proposed for selecting a smaller number of regressor variables in the multivariate regression models with fully unknown covariance matrix. All these criteria are, however, based on the implicit assumption that the sample size is substantially larger than the dimension of the covariance matrix. To obtain a stable estimator of the covariance matrix, it is required that the dimension of the covariance matrix be much smaller than the sample size. When the dimension is close to the sample size, it is necessary to use ridge type of estimators for the covariance matrix. In this paper, we use a ridge type of estimators for the covariance matrix and obtain the modified AIC and modified  $C_p$  statistic under the asymptotic theory that both the sample size and the dimension go to infinity. It is numerically shown that these modified procedures perform very well in the sense of selecting the true model in large dimensional cases.

## Keywords

Akaike information criterion, Mallows'  $C_p$ , Large dimension, Multivariate linear regression model, Selection of variables.

## References

- Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In Petrov, B.N., Csaki, F. (Eds.) *2nd International Symposium on Information Theory*, (pp. 267–281). Akademia Kiado, Budapest.
- Akaike, H. (1974). A new look at the statistical model identification. System identification and time-series analysis. *IEEE Trans. Autom. Contr.* 19, 716–723.
- Mallows, C.L. (1973). Some comments on  $C_p$ . *Technometrics* 15, 661–676.