A comparison of least squares model averaging estimators

Erkki P. Liski[1] **and Antti Liski**[2]

[1] *University of Tampere, Finland, email: erkki.liski@uta.fi*
[2] *Tampere University of Technology, Finland, email: antti.liski@tut.fi*

Keywords: mean square error, model selection, restricted least squares, risk profile, shrinkage estimator.

Our framework is the linear regression model

$$y = \mathbf{X}\beta + \mathbf{Z}\gamma + \varepsilon, \qquad \varepsilon \sim N(0, \sigma^2\mathbf{I}_n), \tag{1}$$

where $\mathbf{X}$ is an $n \times p$ matrix of explanatory variables that we want to keep in the model on theoretical or other grounds. An $n \times m$ matrix $\mathbf{Z}$ contains m additional explanatory variables which we add in the model only if they are supposed to improve estimation of β. To minimize the mean square error of estimation, a balance must be attained between the bias due to omitted variables and the variance due to parameter estimation. Magnus et al. (2010) presented a technique of averaging least squares estimators over models such that the resulting estimator can be presented as a shrinkage estimator. Among this type of estimators we wish to find those that have good risk profile, i.e. the risk is close to the efficiency bound. In general, shrinkage estimators have better risk profile over Post Model Selection (PMS) estimators, and they avoid an unbounded risk.

Shrinkage estimators are computationally superior over the PMS estimators and the model averaging estimators which require estimation of models weights. Computing time of shrinkage estimators increases only linearly with m, the number of auxiliary regressors, while computing time of the PMS estimators is of order 2^m. Thus the shrinkage technique can be easily applied to large data sets when the number of auxiliary regressors is large. We apply the technique on hip fracture patients data to compare treatment costs between hospital districts in Finland.

References

[1] Magnus, J. R., Powell, O., Prüfer, P. (2010). A comparison of two model averaging techniques with an application to growth empirics. *Journal of Econometrics* **154**, 139–153.