

# Robust Variable Selection Based on Least Angle Regression

S. Van Aelst<sup>1</sup>, J. Khan<sup>2</sup> and R.H. Zamar<sup>2</sup>

<sup>1</sup> Ghent University, Dept. of Applied mathematics and Computer Science, Krijgslaan 281 S9, B-9000 Gent, Belgium

<sup>2</sup> University of British Columbia, Dept. of Statistics, 6356 Agricultural, Vancouver, BC, Canada V6T 1Z2

**Keywords:** Robustness, Variable selection, Regression.

## Abstract

Robust model selection is a crucial but difficult part of a successful robust data analysis. Recently, selection techniques have been proposed by Ronchetti *et al.* (1997) and Morgenthaler *et al.* (2004). Most robust model selection methods are computationally intensive and time consuming.

Efron *et al.* (2004) proposed Least Angle Regression (*LARS*), a computationally fast variable selection procedure based on pairwise correlations which is closely related to (stagewise) forward selection and LASSO (see Hastie *et al.*, 2001). Last year we presented some techniques to construct a robust least angle regression method. A first class of methods was based on robust correlation estimates which could be obtained by extending one-dimensional Huberization (Huber, 1981) to the bivariate setting or by using bivariate M-estimators (Maronna, 1976). A second class of methods was based on data cleaning using multivariate Huberization or multivariate M-estimators.

The main goal of our algorithms is an initial reduction of a large set of candidate predictors to a smaller set of 'good' predictors from which an optimal model can be selected using more refined methods such as robust Mallow's  $C_p$  (Ronchetti and Staudte, 1994). In this talk we further explore the robust proposals by measuring their performance and comparing them in simulation studies. We also address the problem of selecting the size of the smaller set from which the refined model can be selected. This is a trade-off between speed and effectiveness to capture the important predictors.

B.E. Efron, T. Hastie, I. Johnstone and R. Tibshirani (2004). Least Angle Regression. *Ann. Statist.*, 32, 407-451.

S. Morgenthaler, R.E. Welsch and A. Zenide (2004). Algorithms for Robust Model Selection in Linear Regression. In: M. Hubert *et al.*, editors, *Theory and Applications of Recent Robust Methods*, pp. 195-206, Birkhäuser Verlag Basel, Switzerland.

E. Ronchetti, C. Field and W. Blanchard (1997). Robust Linear Model Selection by Cross-Validation, *J. Amer. Statist. Assoc.*, 92, 1017-1023.

E. Ronchetti and R.G. Staudte (1994). A Robust Version of Mallow's  $C_p$ , *J. Amer. Statist. Assoc.*, 89, 550-559.