

# Variable Selection With Robust Least Angle Regression

J. Khan, S. Van Aelst, and R.H. Zamar

February 13, 2004

Robust model selection has not received much attention in the robustness literature. A few papers that address this issue include Ronchetti (1984) and Ronchetti & Staudte (1994), where the authors robustify the normal-theory selection criteria  $AIC$  and  $C_p$ , respectively. Morgenthaler *et al.* (2003) propose a selection technique to identify the correct model structure as well as unusual observations. Ronchetti *et al.* (1997) propose robust model selection by cross-validation.

One major drawback of robust model selection tools is that they are in general computationally intensive and time consuming, as they require the fitting of all submodels. One exception is the model selection based on the Wald test (Sommer & Huggins, 1996) which requires the computation of estimates only from the full model. However, fitting the ‘full’ model may not be reasonable or computationally feasible.

In this study, we focus our attention on the robustification of *Stepwise Regression*. This will provide us with a robust ordering of the covariates so that we can choose a number of predictors from the top of the list. Efron *et al.* (2003) proposes Least Angle Regression (*LARS*), a promising normal-theory algorithm that has clear advantages over the *Forward Selection* and the *Forward Stagewise* procedures. We illustrate the sensitivity of *LARS* to outliers and present two different approaches to its robustification. The *robust LARS* is computationally suitable because we can avoid the fitting of all the submodels and the full model.

## References

- Efron, B.E., Hastie, T., Johnstone, I. and Tibshirani, R.** (2004). Least Angle regression. Unpublished manuscript.
- Morgenthaler, S., Welsch, R.E. and Zenide, A.** (2003) Algorithms for robust model selection in linear regression. *ICORS 2003* proceedings.
- Ronchetti, E.** (1985). Robust model selection in regression. *Statistics and Probability Letters*, 3, 21-23.
- Ronchetti, E. and Staudte, R.G.** (1994). A robust version of Mallor's  $C_p$ . *JASA*, 89, 550-559.
- Ronchetti, E., Field, C. and Blanchard, W.** (1997) Robust linear model selection by cross-validation. *JASA*, 92, 1017-1023.
- Sommer, S. and Huggins, R.M.** (1996). Variable selection using the Wald Test and Robust  $C_p$ . *J.R. Statist. Soc.*, B, 45, 15-29.