

# A Study of Error Variance Estimation in Lasso Regression

Stephen Reid<sup>1</sup>, Robert Tibshirani<sup>2</sup> and Jerome Friedman<sup>1</sup>

<sup>1</sup> Department of Statistics, Stanford University, <sup>2</sup> Departments of Health, Research & Policy, and Statistics, Stanford University

## Abstract

Variance estimation in the linear model when  $p > n$  is a difficult problem. Standard least squares estimation techniques do not apply. Several variance estimators have been proposed in the literature, all with accompanying asymptotic results proving consistency and asymptotic normality under a variety of assumptions.

It is found, however, that most of these estimators suffer large biases in finite samples when true underlying signals become less sparse with larger per element signal strength. One estimator seems to be largely neglected in the literature: a residual sum of squares based estimator using Lasso coefficients with regularisation parameter selected adaptively (via cross-validation).

In this paper, we review several variance estimators and perform a reasonably extensive simulation study in an attempt to compare their finite sample performance. It would seem from the results that variance estimators with adaptively chosen regularisation parameters perform admirably over a broad range of sparsity and signal strength settings. Finally, some initial theoretical analyses pertaining to these types of estimators are proposed and developed.

**Keywords:** *cross-validation, error variance estimation, lasso*