

On High-Dimensional Cross-Validation

C.K. Ing¹, W.C. Hsiao¹ and W.Y. Wu^{2*}

¹ *Institute of Statistical Science, Academia Sinica, Taiwan.*

² *Department of Applied Mathematics, National Dong Hwa University, Taiwan.*

**Presenting author*

Keywords. *High-dimensional Linear Model; Model Selection, Cross-validation, Consistency*

Cross-validation (CV) is one of the most popular methods for model selection. By splitting n data points with $n_v/n \rightarrow 1$ and $n_c \rightarrow \infty$ into a training sample of size n_c and a validation sample of size n_v , Shao (1993) showed that subset selection based on CV is consistent in a regression model of p candidate variables with $p \ll n$. However, in the case of $p \gg n$, not only does CV's consistency remain undeveloped, but subset selection is also practically infeasible. In this paper, we fill this gap by using CV as a backward elimination tool for eliminating variables that are included by high-dimensional variable screening methods possessing sure screening property. By choosing an n_v such that n_v/n converges to 1 at a rate faster than the one in Shao's (1993) paper, we establish the consistency of our selection procedure. We also illustrate the finite-sample performance of the proposed procedure using Monte Carlo simulation.