

Model Selection Procedures in Social Research: Monte-Carlo Simulation Results

Lawrence E. Raffalovich
Glenn D. Deane
David Armstrong
Hui-shien Tsao

Department of Sociology
University at Albany
SUNY

Draft
October 2001

**DRAFT COPY – NOT FOR CITATION OR QUOTATION WITHOUT THE AUTHORS’
PERMISSION**

Abstract

Model selection strategies play an important, if not explicit, role in quantitative research. The inferential properties of these strategies are largely unknown, so there is little basis for recommending (or avoiding) any particular set of strategies. In this paper we evaluate several commonly used model selection procedures (BIC, adjusted R^2 , Mallows’s C_p , and stepwise regression) using monte-carlo simulation of model selection when the true data-generating processes are known.

We find that the ability of these selection procedures to include important variables and exclude irrelevant variables increases with the size of the sample and decreases with the amount of noise in the model. None of the model selection procedures do well in small samples unless the true DGP is almost entirely deterministic; so data mining in small samples should be avoided entirely. In large samples, BIC is better than the other procedures at correctly identifying most of the generating processes we simulated, and Stepwise does even better with some models and almost as well with others. In the absence of strong theory, both BIC and Stepwise appear to be reasonable model selection strategies in large samples. Adjusted R^2 and Mallows’s C_p are clearly inferior and should be avoided.

A previous version of this paper was presented at the 96th annual meetings of the American Sociological Association, Anaheim, CA, August 2001. Direct all correspondence to Lawrence E Raffalovich, Department of Sociology, University at Albany, Albany, NY 12222 (L.raffalovich@albany.edu). We thank Richard Alba and Dana Haynie for comments on a previous draft. This research was partially supported by a grant to the first author from the National Science Foundation (SBR-9730029). The Center for Social and Demographic Analysis, University at Albany, provided technical and administrative support through grants from NICHD (P30 HD32041) and NSF (SBR-9512290).