2001-4

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DRAFT COPY – NOT FOR CITATION OR QUOTATION WITHOUT THE AUTHORS' PERMISSION

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).

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

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, Mallow'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 Mallow's C_p are clearly inferior and should be avoided.

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

I. Introduction

Model selection strategies – selecting the set of variables to be included in a statistical model, transformations of variables, and functions linking independent to dependent variables – have a somewhat shadowy existence in the sociological research literature. Quantitative research articles conventionally adhere to the following sequence: An introductory statement of the problem is followed by discussion of theory and prior research; then, a discussion of the authors' contributions to this literature is followed by hypotheses to be tested, data sources, and analytic methods; finally, results are presented and interpreted, and conclusions discussed. If not explicitly asserted, it is implicit in this sequence that models to be estimated are derived from theory and prior research. Hypothesis tests are presented in the context of this sequence: following classical statistical theory, significance levels are interpreted as Type I error rates. Although it is rare for a published research article to describe in detail the procedures actually followed in selecting models, there is compelling anecdotal evidence that the implicit classic hypothesis-testing sequence is equally rare[1].

We know that if many models are estimated, and/or many tests conducted, nominal significance levels drastically understate Type I error rates (e.g. Toothaker, 1991). Without explicit information on actual model selection procedures, it is impossible to evaluate Type I error rates in published research. Still, it may be feasible to evaluate the statistical properties of various model selection procedures. This is possible, however, only if there is an explicit discussion of the model selection strategies utilized in practice. This paper is a step in that direction. In the following section, we review a variety of formal model selection procedures discussed in the statistical literature and/or offered as options in widely utilized statistical software packages. We then evaluate several of these procedures using monte-carlo simulation.

II. Model Selection Procedures

The following model selection options are available in the SAS regression procedure (SAS Institute 1990, pp. 1367, 1397-99):

Forward: starting with no variables, sequentially adds variables to the model if their F-statistics exceed a pre-selected value. The variable with the largest F-statistic is added at each step.

Backward: Starting with all variables, sequentially removes variables from the model if their F-statistics are below a pre-selected value. The variable with the smallest F-statistic is removed at each step.

Stepwise: Combination of Forward and Backward. Variables added to the model will be removed if their F-statistics later fall below the pre-selected value. Variables removed from the model will be added if their F-statistics later exceed the pre-selected value.

---

[1] An admission that "through trial and error, we stumbled on the following results" would surely doom any submission to a refereed journal. Yet, it is not uncommon for authors to note that various transformations or functional forms were estimated , that only pared-down models are presented, or that insignificant variables are excluded from tables.

Maxr: Selects the "best" 1-variable, 2-variable, etc. model by adding variables that produce the largest increment to R^2. Similar to Stepwise.

Rsquare: Calculates all possible subset regressions (for each subset of k variables). For each subset size, selects the model with the highest R^2.

Adjrsq: The same as Rsquare, but uses adjusted R^2 as the criterion.

C_p: The same as Rsquare, but uses Mallow's C_p as the criterion.

In addition, the following options are available in one or more of the above procedures: Aikaike's information criterion (AIC), Sawa's Bayesian information criterion (BIC), Mallow's C_p (C_p), Amemiya's prediction criterion (PC), and Schwarz's Bayesian criterion (SBC). These calculate the following statistics (p. 1369):

$R^2 = 1 - [SSE/SST]$
Adjusted $R^2 = 1 - [(n-1)/(n-p)][1-R^2]$
$AIC = n[\log(SSE/n)] + 2p$
$BIC = n[\log(SSE/n)] + 2(p+2)q - 2q^2$
$C_p = (SSE/\hat{\sigma}^2) + 2p - n$
$PC = (1-R^2) [(n+p)/(n-p)]$
$SBC = n[\log(SSE/n)] + p[\log(n)]$

where n = sample size
p = number of variables included in the model
$q = \hat{\sigma}^2/(SSE/n)$
SSE = sum of squared residuals
SST = total sum of squares
$\hat{\sigma}^2$ = estimated error variance from fitting the full model
log = natural logarithm

These and other criteria are discussed by Judge, et al. (1985, Chapter 21). All balance model fit (relative to either an intercept-only model or a saturated model) with model parsimony. Differences are with regard to the measure of fit and the penalties for additional variables and observations. All are related to the F-statistic and/or the likelihood ratio. Judge, et al.(1985, pp. 888-889) notes that these criteria are ad-hoc, justified by practical utility, and that their sampling properties are unknown.

Recently, Raftery (1995) introduced a model selection criterion derived from Bayesian statistical theory. For linear regression models, this Bayesian Information Criterion can be expressed as $BIC = n[\log(1-R^2)]+ p[\log(n)]$.[2] According to Raftery (1995) BIC approximates the ratio of posterior probabilities for two competing models. In contrast to classical hypothesis testing, which only offers evidence against a null model, BIC weighs the evidence in favor of competing models. In addition, the

[2] We note that when expressed as a function of the log-likelihood, BIC = SBC, and Judge et al (1985, p. 874) show that AIC, C_p, and PC are all closely linked to the posterior odds ratio.

classical hypothesis testing approach to model selection can evaluate only nested models.[3] BIC is equally applicable to a large class of non-nested models because it is a ratio of posterior odds relative to a common null model (Raftery 1995).

Weakliem (1999) points out that BIC depends not only on posterior probabilities of models, but also on the prior beliefs of researchers and properties of the data, and that "the prior distributions that yield [BIC] change depending on the distribution of the variables and the presence of irrelevant observations" (p. 388). He argues that the simplifying assumptions Raftery (1995) used to derive BIC are neither reasonable nor inconsequential, and shows that alternative but reasonable prior beliefs can result in dramatically different conclusions.

III. Empirical Evaluation

Evaluation of statistical procedures typically follow one of two strategies: 1) Use real data that have already been extensively analyzed, or 2) use monte-carlo simulation to construct artificial datasets that meet specified criteria. The advantage of the first strategy is that it allows comparisons to previous findings based on alternative procedures. The disadvantage is that the true data generating process is unknown. The advantage of the second strategy is that the true data generating process is known. The disadvantage is that the structure of the data may be far removed from data structures encountered in practice.

Lovell (1983) combined these two strategies. In his monte-carlo study of model selection techniques, he selected 20 variables from a large publicly available dataset, and estimated nine different models for personal consumption expenditure. He then used the estimated parameter values and the same independent variables to generate nine dependent variables (i.e., the predicted values of each of the nine dependent variables). Finally, a random draw from a standard normal distribution was used to calculate the nine "pseudo-consumption" variables used in the analyses. Lovell (1983) used 50 replications of this procedure to illustrate the pitfalls of three model selection strategies: stepwise regression, maximum adjusted R-square, and max-min |t|.

We use this method, but improve on Lovell's (1983) study in a number of respects. First, we include additional selection techniques, most importantly Raftery's BIC, which has attained some prominence in sociological research, including a special issue of Sociological Methods and Research (Vol. 27, No. 3). Second, whereas Lovell (1983) focused his attention on the ability of selection techniques to correctly include or exclude one or two variables, we focus on the ability of these techniques to correctly include and exclude entire sets of independent variables. This more closely approximates actual research applications, where collinearity among independent variables is an important source of ambiguity. Finally, we use a dataset that has already been analyzed for both substantive and methodological research purposes, and use specific models that social scientists might reasonably be expected to consider. Of these, two were formulated by the original researcher (Erlich 1973), and three were selected from the original data by the different model selection techniques (Raftery 1995).

---

[3] Judge, et al. (1985) discuss all the above criteria only in the context of nested models.

We use Lovell's (1983) method to evaluate the empirical performance of model selection based on maximum adjusted R-square, Mallow's C_p, stepwise regression[4], and Raftery's BIC. All were used by Raftery (1995) with Erlich's (1973) crime dataset[5] to illustrate the superior performance of BIC. We use the same crime data analyzed by Raftery (1995)[6], but we follow Lovell (1983) in using specific models to generate the dependent variables. The specific models we use are those considered by Raftery (1995). This procedure allows us to know the true data generating process while also remaining firmly grounded in data structures actually encountered in empirical research.

The Erlich (1973) dataset contains 47 observations on 16 variables (observations are U.S. states). The following variable descriptions are from Raftery (1995, Table 4):

Y: state crime rate in 1960
X_1: percent of males, 14-24
X_2: southern state
X_3: mean years of schooling
X_4: police expenditure in 1960
X_5: police expenditure in 1959
X_6: labor force participation rate
X_7: males per 100 females
X_8: state population
X_9: nonwhites per 1000 population
X_{10}: unemployment rate of urban males, 14-24
X_{11}: unemployment rate of urban males, 35-39
X_{12}: GDP
X_{13}: income inequality
X_{14}: probability of imprisonment
X_{15}: average time served in state prison

We used these data to specify the six linear models presented and discussed by Raftery (1995, pp. 120-124):

Model 1: $Y_1 = f(X_1,X_2,X_3, X_4,X_5,X_6,X_7,X_8,X_9,X_{10},X_{11},X_{12},X_{13},X_{14},X_{15}, U_1)$
Model 2: $Y_2 = f(X_1,X_3, X_4,X_9,X_{11},X_{13},X_{14},U_2)$
Model 3: $Y_3 = f(X_1,X_3, X_4,X_9,X_{11},X_{12},X_{13},X_{14},X_{15},U_3)$
Model 4: $Y_4 = f(X_1,X_3, X_4,X_7,X_8,X_9,X_{11},X_{12},X_{13},X_{14},X_{15},U_4)$
Model 5: $Y_5 = f(X_9,X_{12},X_{13},X_{14},X_{15},U_5)$

4 For the stepwise regression, we used p = .05 for entry and for removal.

5 Raftery (1995, p. 121) notes that he used "corrected data".

6 Adrian Raftery kindly provided these data to us.

Model 6: $Y_6 = f(X_1,X_6,X_9,X_{10},X_{12},X_{13},X_{14},X15,U_6)$

Coefficients values for each model were obtained by OLS regression on the original data.[7] The stochastic components (U_1 through U_6) were randomly drawn from zero-mean normal distributions (standard deviations are discussed below). We then used four common model selection techniques (adjusted R^2, Mallow's C_p, stepwise regression, and BIC) to select the best model from the entire pool of fifteen independent variables. We replicated this procedure 1,000 times, and tabulated the percent of selected models that included all of the correct variables (i.e., all of the variables in the true DGP), and the percent of selected models that included only the correct variables (i.e., all of the correct variables but no incorrect variables).

We repeated this procedure for three different stochastic distributions and two different sample sizes. For the latter replications, OLS regressions were run from correlation matrices (regression coefficients are therefore standardized, but none of the statistics of interest are affected). To simulate the effect of different amounts of noise in the data, we used error standard deviations of 0.74, 0.41, and 0.18 (corresponding to population R^2 of 0.45, 0.83, and 0.97, respectively). To simulate the effect of different sample sizes, we used both the original N=47, and N=1000.

Results for F = .74 (corresponding to a population $R^2 = 0.45$) are presented in Table 1. The top panel uses the observed sample of N=47. The bottom panel uses N=1000. For each true model, we present the percent of 1000 replications that included all of the variables in the true model, and the percent that included only the variables in the true model (all of the correct variables and no others). Thus, the "All Correct Variables" column provides an indication of the Type-II error rate (excluding a correct variable is failing to reject the false hypothesis that its parameter is zero), and the "Only Correct Variables" column provides an indication of the Type-I error rate (including an irrelevant variable is rejecting the true hypothesis that the parameter is zero) . For example, with N=47, BIC includes all of the correct Model 2 variables in only 0.3% of the replications, but also included irrelevant variables in a third of those. Adjusted R^2 includes all of the correct Model 2 variables in only 9.5% of replications, but also included irrelevant variables in almost all of those (91 of the 95, or 9.1% of the 1000). Results for Models 1, 3, 4, and 6 are equally grim. It is exceedingly rare for any of the selection procedures to correctly identify all of the variables in the true model, and even then most include irrelevant variables also.

All of these model-selection procedures are more successful with Model 5. All but Stepwise include all of the correct variables in more that a quarter of the 1000 simulations, and adjusted-R^2 includes all of the correct variables in over half. However, these procedures also include irrelevant variables at a very high rate. Adjusted-R^2, for example, includes only the correct variables in 2.4% of the 1000 replications -- i.e. (59.2 - 2.4)/ 59.2 = 96% of the models with all of the correct variables also include irrelevant variables. Here, BIC does somewhat better than the other procedures, correctly identifying the true model in 13.4% of replications.

[7] Our regression estimates reproduced the quantities in Raftery's Table 5 (1995, p.123).

Table 1 about here

With sample size increased to N=1000, success rates increase dramatically. With the exception of Model 1 (which includes all fifteen variables) and Model 6, all procedures include all correct variables in most simulations (the lowest is 65.1% for BIC with Model 4). Adjusted R^2, although it always includes all of the correct variables in Models 2 through 4, also includes irrelevant variables at a very high rate. Indeed, Adjusted R^2 performs worst by this criterion, correctly identifying the correct model in less than 25% of replications. BIC excludes irrelevant variables at a comparatively high rate, surpassed only by Stepwise in Model 4 and Model 6.

Results for F = .41 (corresponding to a population R^2 = 0.83) are presented in Table 2. Success rates are considerably higher than in Table 1. With N=47, all of the correct model 2 variables are included by BIC in more than a fourth (27.3%) of the replications, by Adjusted R^2 in almost half (47.2%), and by Mallow's C_p in over a third (35.2%). For Model 5, all of the selection procedures include all of the correct variables in over 80% of the simulations (97.8% for Adjusted R^2). As in Table 1, however, these procedures also include irrelevant variables at a very high rate. The selection procedures were most successful with Model 5, but all included irrelevant variables in over 40% of Model 5 simulations. For Model 2, irrelevant variables were included by all procedures in over 80% of the simulations.

Table 2 about here

With sample size increased to N=1000, success rates are much higher. With the exception of Model 1 and Model 6, all procedures include all correct variables in virtually all simulations (the lowest is 99.3% for BIC with Model 3). Compared to Table 1, there is also increased success in eliminating irrelevant variables. With the exception of Model 1 and Model 6, BIC identifies the correct model in over 90% of simulations, and Stepwise succeeds in more than two thirds. Adjusted R^2 again performs worst by this criterion, identifying the correct model in less than 25% of replications.

Table 3 presents results for F = .18 (corresponding to a population R^2 = 0.97). The reduction in noise substantially improves success rates. With the exception of Model 1 (which includes all fifteen variables) and Model 6, all selection procedures include all the correct variables in well over half of the simulations with the original sample size (the minimum is 56.6% for Stepwise). However, the rates at which irrelevant variables are included are rarely under 50% (BIC for Model 4 and Model 5), lower than in Table 2 but still high in absolute terms.

Table 3 about here

With the sample size increased to 1000, all procedures include all the correct variables in all models except Model 1 and Model 6. With the exception of these two models, the rates at which irrelevant variables are included are now below 10% for BIC and below 33% for Stepwise, but still high for Adjusted R^2 (above 75%) and Mallow's C_p (above 50%).

Clearly the reduction in noise and the increase in sample size make a difference. With a very

small stochastic component and a large sample, all four model selection procedures select from the pool all of the correct independent variables. There is substantial variability, however, in the rates at which these procedures select the correct model. Even under the most favorable of our simulated conditions, adjusted R^2 includes irrelevant variables in more than 75%
of the replications. BIC, on the other hand almost always selects the correct model under these favorable conditions, and Stepwise does so more than two-thirds of the time.[8]

Regardless of the stochastic variance and sample size, all selection procedures fared worst with Model 1 and tended to do best with Model 4 or Model 5. Model 1 includes all of the eligible independent variables, so inter-correlations can be expected to severely muddy the waters, especially in small samples. Model 5 is the simplest in that it has the fewest independent variables, but Model 4 has more variables than all except Model 1. Evidently, patterns of inter-correlation among independent variables substantially affects the performance of these model selection procedures.

The relative performance of the different model selection procedures is summarized in Table 4. With a large noise component and a small sample, none of the procedures performs well. Adjusted R^2 does locate all of the correct Model 5 variables in just over 25% of simulations, but also includes irrelevant variables in most of those (see Table 1). With a large sample, Adjusted R^2 selects the correct variables in Models 2, 3, 4, and 5 in over 90% of simulations (and in just under half of Model 6 simulations), but also includes irrelevant variables at a very high rate. BIC is much more successful in selecting the correct Model 2, Model 3, and Model 5, and is a close second for Model 4 (63.4% vs. 69.8% for Stepwise).

Table 4 about here

Reducing the noise component substantially improves success rates. In small samples, Adjusted R^2 is more successful than other procedures at selecting all of the correct variables, but (except for Model 5), success rates are below 50%. None of the procedures is notably better than others at excluding irrelevant variables, and only Stepwise succeeds in over half of the simulations. In large samples, Adjusted R^2 is again best at selecting all of the correct variables, although (except for Model 1 and Model 6) all procedures have success rates over 90%. BIC is clearly superior at eliminating irrelevant variables. With the exception of Model 1 (with no irrelevant variables to remove) and Model 6, BIC's success rates of over 90% greatly surpass those of the alternative procedures, and BIC is a close second to Stepwise in excluding irrelevant variables from Model 6 (44.8% vs. 47.6%).

With the noise component only 3% of total variation, Adjusted R^2 is again most successful at locating all of the correct variables in small samples (although, with a few exceptions, the alternatives do almost as well). Stepwise is most successful at selecting the correct Model 2, Model 3, and Model 5 (in two-thirds of simulations), BIC is slightly more successful for Model 4 (57.3%). Except for Model 1

[8] The obvious exception in these analyses is Model 1, which contains all 15 of the variables in the pool. It is therefore not possible to include an irrelevant variable in Model 1. Because the "Only Correct Variables" column cannot exceed the "All correct Variables" column, the rate at which a procedure selects the correct model is limited by the rate at which it includes all correct variables.

and Model 6, all procedures are equally successful (100%) selecting all of the correct variables in large samples, and all have success rates over 98% for Model 6. Except for Model 1 (with no irrelevant variables to exclude) BIC is far superior to the alternative procedures in excluding irrelevant variables (over 90% of simulations).

IV. Summary and Conclusions

Model selection strategies play an important, but unfortunately not an 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. This paper contributes to the empirical evaluation of various model selection procedures. We evaluate several commonly used procedures using monte-carlo simulation of model selection when the true data-generating processes are known.

Following the example of Lovell (1983), we first compile a dataset that is similar to those typically used in social science research. Using these data, we specify six models that a social science researcher might reasonably be expected to consider. Next, we create 1,000 datasets generated by each of these models by adding independent draws from a normal distribution. Finally, we select the 'best' regression model from each of these datasets using four standard model selection procedures: BIC, Adjusted R^2, Mallow's C_p, and Stepwise regression. We repeated this procedure for three different stochastic distributions (corresponding to population R^2 of 0.45, 0.83, and 0.97), and two different sample sizes (the original N=47 and N=1000). Comparing the results of these model selection procedures to the true data-generating processes allows us to evaluate the extent to which each strategy includes the correct variables and excludes those that are irrelevant under conditions that approximate those actually encountered in social research.

Results show that both the amount of noise in the model and the size of the sample are important. For all four selection procedures, success rates for including important variables increases with sample size and with a reduction in the stochastic variance, although Adjusted R^2 is the most powerful in all the circumstances we simulated. Success rates for selecting the correct model (including relevant variables <u>and</u> excluding irrelevant variables) also varies with sample size and stochastic variance, but there is no best procedure. In small samples, Stepwise has an advantage when the noise component is small, but even with stochastic variance only 3% of the total, the best success rate for selecting the correct model was only 68%. Success rates for selecting the correct model are higher in large samples, and here BIC has a clear advantage. Even with a large stochastic component, BIC selected the correct model in most simulations, and with a small stochastic component BIC success rates exceed 90%. Surprisingly, Stepwise did nearly as well as BIC in large samples. The other two procedures are clearly inferior in this regard.

Of the conditions that we simulated, only sample size is known in practice. 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. Adjusted R^2 and Mallow's C_p are clearly inferior and should be avoided.

BIC's advantage depends on the size of the stochastic component and particulars of the true model, which are not known in practice. Like the other selection procedures, BIC does worse with Model 6 than with the other data generating processes. Model 6 is not noteworthy in any particular respect – it is neither the largest nor the smallest model in terms of the number of independent variables – so we speculate that the pattern of correlations among independent variables is an important factor in model selection success rates. This speculation is reinforced by the relatively high success rates with Model 4 (the largest) and Model 5 (the smallest), and needs to be investigated in a systematic manner. In addition, our criteria for entry and removal of variables in stepwise regression are likely idiosyncratic (SAS defaults for both are p=.15), and the performance of Stepwise with alternative entry and removal criteria needs systematic study.

References

Ehrlich, Isaac. 1973. “Participation in Illegitimate Activities: A Theoretical and Empirical Investigation”. Journal of Political Economy 81:521-565.

Judge, George G., W.E. Griffiths, R. Carter Hill, Helmut Lutkepohl, and Tsoung-Chao Lee. 1985. The Theory and Practice of Econometrics. New York: Wiley.

Lovell, Michael. 1983. “Data Mining”. Review of Economics and Statistics 65:1-12.

Raftery, Adrian E. 1995. “Bayesian Model Selection in Social Research”. Sociological Methodology 25:111-163.

SAS Institute. 1990. SAS/STAT User’s Guide, Volume 2. Version 6, 4th Edition. Cary, NC: SAS Institute.

Toothaker, Larry E. 1991. Multiple Comparisons for Researchers. Newbury Park, CA: Sage.

Weakliem, David L. 1999. “A Critique of the Bayesian Information Criterion for Model Selection”. Sociological Methods and Research 27:359-397.

Table 1. Success Rates (%) of Four Model Selection Procedures: 1000 Replications, s =.74

	Selection Procedure							
	BIC		Adjusted R^2		Mallow's C_p		Stepwise	
	Only Correct Variables	All Correct Variables	Only Correct Variables	All Correct Variables	Only Correct Variables	All Correct Variables	Only Correct Variables	All Correct Variables
N=47								
Model 1	0.0	0.0	0.1	0.1	0.0	0.0	0.0	0.0
Model 2	0.2	0.3	0.4	9.5	0.8	2.1	0.0	0.0
Model 3	0.0	0.0	0.4	4.2	0.1	0.4	0.0	0.0
Model 4	0.0	0.0	0.1	1.4	0.1	0.1	0.0	0.0
Model 5	13.4	26.5	2.4	59.2	11.5	41.7	3.6	5.4
Model 6	0.1	0.7	0.2	6.6	0.1	2.0	0.0	0.0
N=1000								
Model 1	0.0	0.0	8.3	8.3	0.7	0.7	0.0	0.0
Model 2	87.9	93.4	6.1	98.5	27.6	96.5	67.5	95.7
Model 3	76.3	78.7	10.9	93.1	36.2	87.6	75.3	93.1
Model 4	63.4	65.1	20.4	94.6	45.8	88.7	73.3	84.3
Model 5	92.7	100.0	3.0	100.0	21.5	100.0	68.3	100.0
Model 6	11.2	11.8	5.3	49.1	13.8	36.3	13.9	17.5

<u>Notes:</u>

Model 1: includes all 15 variables
Model 2: includes X1, X3, X4, X9, X11, X13, X14
Model 3: includes X1, X3, X4, X9, X11, X12, X13, X14, X15
Model 4: includes X1, X3, X4, X7, X8, X9, X11, X12, X13, X14, X15
Model 5: includes X9, X12, X13, X14, X15
Model 6: includes X1, X6, X9, X10, X12, X13, X14, X15

Table 2. Success Rates (%) of Four Model Selection Procedures: 1000 Replications, s =.41

	Selection Procedure							
	BIC		Adjusted R^2		Mallow's C_p		Stepwise	
	Only Correct Variables	All Correct Variables	Only Correct Variables	All Correct Variables	Only Correct Variables	All Correct Variables	Only Correct Variables	All Correct Variables
N=47								
Model 1	0.0	0.0	0.5	0.5	0.0	0.0	0.0	0.0
Model 2	15.3	27.3	3.2	47.2	11.1	35.2	19.3	25.5
Model 3	4.9	8.2	3.1	26.8	5.4	12.4	2.6	3.0
Model 4	1.8	3.4	4.3	18.1	3.4	6.9	0.1	0.1
Model 5	49.7	93.8	2.9	97.8	24.6	96.0	55.8	80.0
Model 6	2.0	4.2	1.8	23.5	3.0	9.5	0.5	0.8
N=1000								
Model 1	0.1	0.1	26.0	26.0	10.4	10.4	1.5	1.5
Model 2	93.2	100.0	6.1	100.0	27.6	100.0	68.0	100.0
Model 3	95.2	99.3	10.9	100.0	37.5	100.0	77.5	99.7
Model 4	95.9	99.5	20.5	100.0	47.1	100.0	82.1	99.8
Model 5	92.7	100.0	3.0	100.0	21.5	100.0	68.2	100.0
Model 6	44.8	47.1	8.5	76.0	26.4	70.0	46.3	58.4

Notes:

Model 1: includes all 15 variables
Model 2: includes X1, X3, X4, X9, X11, X13, X14
Model 3: includes X1, X3, X4, X9, X11, X12, X13, X14, X15
Model 4: includes X1, X3, X4, X7, X8, X9, X11, X12, X13, X14, X15
Model 5: includes X9, X12, X13, X14, X15
Model 6: includes X1, X6, X9, X10, X12, X13, X14, X15

Table 3. Success Rates (%) of Four Model Selection Procedures: 1000 Replications, s =.18

	Selection Procedure							
	BIC		Adjusted R^2		Mallow's C_p		Stepwise	
	Only Correct Variables	All Correct Variables	Only Correct Variables	All Correct Variables	Only Correct Variables	All Correct Variables	Only Correct Variables	All Correct Variables
N=47								
Model 1	0.6	0.6	6.1	6.1	1.5	1.5	0.0	0.0
Model 2	48.2	91.7	4.8	96.7	27.7	93.8	67.1	94.8
Model 3	49.8	79.3	10.7	89.0	35.8	82.2	69.0	86.1
Model 4	57.3	78.0	20.2	89.7	47.3	82.8	50.1	58.2
Model 5	52.2	100.0	2.9	100.0	25.6	100.0	68.0	100.0
Model 6	14.9	26.6	3.7	50.6	12.6	35.5	11.2	13.4
N=1000								
Model 1	1.8	1.8	38.4	38.4	22.0	22.0	8.7	8.7
Model 2	93.2	100.0	6.1	100.0	27.6	100.0	68.0	100.0
Model 3	95.4	100.0	10.9	100.0	37.5	100.0	77.5	100.0
Model 4	96.0	100.0	20.5	100.0	47.1	100.0	82.1	100.0
Model 5	92.7	100.0	3.0	100.0	21.5	100.0	68.3	100.0
Model 6	93.8	98.3	9.5	99.5	33.2	99.1	74.0	99.0

Notes:

Model 1: includes all 15 variables
Model 2: includes X1, X3, X4, X9, X11, X13, X14
Model 3: includes X1, X3, X4, X9, X11, X12, X13, X14, X15
Model 4: includes X1, X3, X4, X7, X8, X9, X11, X12, X13, X14, X15
Model 5: includes X9, X12, X13, X14, X15
Model 6: includes X1, X6, X9, X10, X12, X13, X14, X15

Table 4. Summary of Most Successful Selection Procedures

	F = .74		F = .41		F = .18	
	All Correct Variables	Only Correct Variables	All Correct Variables	Only Correct Variables	All Correct Variables	Only Correct Variables
N=47						
Model 1	Adj. R^2 *	Adj. R^2 *	Adj. R^2 *	Adj. R^2 *	Adj. R^2 *	Adj. R^2 *
Model 2	Adj. R^2 *	C_p *	Adj. R^2 **	Stepwise *	Adj. R^2	Stepwise
Model 3	Adj. R^2 *	Adj. R^2 *	Adj. R^2 **	C_p *	Adj. R^2	Stepwise
Model 4	Adj. R^2 *	Adj. R^2, C_p *	Adj. R^2 *	Adj. R^2 *	Adj. R^2	BIC
Model 5	Adj. R^2	BIC *	Adj. R^2	Stepwise	All	Stepwise
Model 6	Adj. R^2 *	Adj. R^2 *	Adj. R^2 *	C_p *	Adj. R^2	BIC *
N=1000						
Model 1	Adj. R^2 *	Adj. R^2 *	Adj. R^2 **	Adj. R^2 **	Adj. R^2 **	Adj. R^2 **
Model 2	Adj. R^2	BIC	All	BIC	All	BIC
Model 3	Adj. R^2, Stepwise	BIC	Adj. R^2, C_p	BIC	All	BIC
Model 4	Adj. R^2	Stepwise	Adj. R^2, C_p	BIC	All	BIC
Model 5	All	BIC	All	BIC	All	BIC
Model 6	Adj. R^2 **	Stepwise *	Adj. R^2	Stepwise **	Adj. R^2	BIC

Note: * Success rate < 25%, ** Success rate < 50%