Best Regression Model Using Information Criteria

Phill Gagné
University of Maryland – College Park

C. Mitchell Dayton
University of Maryland, cdayton@umd.edu

Early Scholars
Best Regression Model Using Information Criteria

Phill Gagné
Department of Measurement
Statistics and Evaluation
University of Maryland – College Park

C. Mitchell Dayton
Department of Measurement
Statistics and Evaluation
University of Maryland – College Park

The accuracy of AIC and BIC is evaluated under simulated multiple regression conditions, varying number of total and valid predictors, R², and n. AIC and BIC were increasingly accurate as n increased and as total predictors decreased. Interactions of the ratio of valid/total predictors affected accuracy.

Key words: AIC, BIC, simulated regression, information criteria

Introduction

Exploratory model building is often used within the context of multiple regression (MR) analysis. As noted by Draper and Smith (1998), these undertakings are usually motivated by the contradictory goals of maximizing predictive efficiency and minimizing data collection/monitoring costs. A popular compromise has been to adopt some strategy for selecting a "best" subset of predictors.

Many different definitions of best can be found in the literature, including incremental procedures such as forward selection MR, backward elimination MR, stepwise MR, all-possible subsets MR with criteria related to residual variance, multiple correlation, Mallows C_p, etc. Incremental procedures are efficient, computationally, but do not necessarily result in

Phill Gagné is a doctoral student in the Department of Measurement, Statistics and Evaluation at the University of Maryland – College Park. E-mail: blueplanet88@hotmail.com. His research interests include factor analysis, SEM, and personality assessment. C. Mitchell Dayton is Professor of Measurement, Statistics and Evaluation at the University of Maryland, College Park, Maryland. E-Mail: cd4@umail.umd.edu. His current research interests are focused on the areas of latent variable analysis and model comparison procedures.

the selection of an unconditionally best model. For example, as usually implemented, forward selection MR includes additional variables in the regression model based on maximizing the increment to R-squared from step to step. At the third step, for example, the model contains the best three predictors only in a conditional sense. Also, the modifications to forward selection incorporated into stepwise MR do not guarantee finding the best three predictors.

In contrast to incremental procedures, all-possible subsets does choose a best model for a fixed number of predictors but not necessarily an overall best model. For the mth model based on p_m out of a total of p independent parameters, Mallows C_p, for example, utilizes a criterion of the form $SS_m / \hat{\sigma}_e^2 - [n - 2(p_m + 1)]$ where $\hat{\sigma}_e^2$ is the residual variance estimate based on the full model (i. e., the model with all p predictors). Models with values close to $p_m + 1$ are best in a final prediction error (FPE) sense. Thus, a best model can be identified for fixed values of p_m, but there is no general method for selecting an overall best model.

Akaike (1973) adopted the Kullback-Leibler definition of information, $I(f;g)$, as a natural measure of discrepancy, or asymmetrical distance, between a true model, $f(y)$, and a proposed model, $g(y|\beta)$, where β is a vector of parameters. Based on large-sample theory, Akaike derived an estimator for $I(f;g)$ of the form:

$$AIC_m = -2Ln(L_m) + 2 \cdot k_m \qquad ,$$

where L_m is the sample log-likelihood for the mth of M alternative models and k_m is the number of independent parameters estimated for the mth model. The term, $2 \cdot k_m$, may be viewed as a penalty for over-parameterization. The derivation of AIC involves the notion of loss of information that results from replacing the true parametric values for a model by their maximum likelihood estimates (MLE's) from a sample. In addition, Akaike (1978b) has provided a Bayesian interpretation of AIC.

   A min(AIC) strategy is used for selecting among two or more competing models. In a general sense, the model for which AIC_m is smallest represents the "best" approximation to the true model. That is, it is the model with the smallest expected loss of information when MLE's replace true parametric values in the model. In practice, the model satisfying the min(AIC) criterion may or may not be (and probably is not) the "true" model since there is no way of knowing whether the "true" model is included among those being compared. Unlike traditional hypothesis testing procedures, the min(AIC) model selection approach is holistic rather than piecemeal. Thus, for example, in comparing four hierarchic linear regression models, AIC is computed for each model and the min(AIC) criterion is applied to select the single "best" model. This contrasts with the typical procedure of testing the significance between models at consecutive levels of complexity. An excellent and more complete introduction to model selection procedures based on information criteria is presented by Burnham and Anderson (1998).

   Typically, for regression models, the number of independent parameters, k_m, is equal to the number of predictor variables in the equation plus two since, in addition to partial slope coefficients, an intercept and residual variance term are estimated. It should be noted that the maximum likelihood estimator for the residual variance is biased (i. e., the denominator is the sample size, n, rather than $n - p_m - 1$ for a p_m-predictor model). In particular, for p predictors based on a normal regression model (i. e., residuals assumed to be normally distributed with homogeneous variance), the log(likelihood) for the

model is: $-.5n \cdot (\ln(2\pi) + \ln(SS_e / n) + 1)$ where SS_e is the sum of squared residuals. Then, the Akaike information measure is:

$$AIC = n(\ln(2\pi) + \ln(SS_e / n) + 1) + 2(p_m + 2).$$

The Akaike model selection procedure entails calculating AIC for each model under consideration and selecting the model with the minimum value of AIC as the preferred, or "best," model. In the context of selecting among regression models, a "best" model can be selected for each different size subset of predictors as well as overall.

   AIC, which does not directly involve the sample size, n, has been criticized as lacking properties of consistency (e.g., Bozdogan, 1987; but see Akaike, 1978a for counter arguments). A popular alternative to AIC presented by Schwarz (1978) and Akaike (1978b) that does incorporate sample size is BIC where:

$$BIC_m = -2Ln(L_m) + \ln(n) \cdot k_m.$$

   BIC has a Bayesian interpretation since it may be viewed as an approximation to the posterior odds ratio. Note that BIC entails heavier penalties per parameter than does AIC when the sample size is eight or larger. When the order of the model is known and for reasonable sample sizes, there is a tendency for AIC to select models that are too complex and for BIC to select models that are too simple. In fact, the relative tendencies for the occurrence of each type of misspecification can be derived mathematically as shown by McQuarrie and Tsai (1998). The tendency for AIC to select overly complex models in cases where complexity is known has been interpreted as a shortcoming of this measure. Hurvich and Tsai (1991), for example, argue for a modified version of AIC that incorporates sample size. In practical applications, however, the performance of criteria such as AIC and BIC can be quite complex.

   AIC was originally developed by Akaike within the context of relatively complex autoregressive time series models for which he presented some simulation results (Akaike, 1974). Bozdogan (1987) compared rates of successful model identifications for AIC and CAIC (a close kin of BIC) for a single cubic model with various

error structures. Hurvich and Tsai (1991) compared AIC and their own consistent estimator, AICC, for a normal regression case and for a complex time series. Bai et al. (1992) compared AIC and several modifications of AIC within the context of multinomial logistic regression models. Although each of these previous studies has investigated the use of AIC and related criteria in exploratory frameworks, the present study expands the focus to applications of multiple regression analysis that are more typical of a behavioral science setting. More specifically, AIC and BIC were investigated under a variety of realistic scenarios.

Methodology

AIC and BIC were evaluated under several simulated multiple regression conditions. Data were collected regarding the accuracy of both information criteria for each condition and the nature of the incorrect choices. The accuracy of an information criterion was defined as the percentage of iterations in which it selected the correct model. Incorrect model selections fell into one of three categories: 1) Low: The chosen model had too few predictors in it; 2) High: The chosen model had too many predictors in it; 3) Off: The chosen model had the correct number of predictors but included one or more that had a correlation of 0 with the criterion without including one or more that had a nonzero correlation with the criterion.

The number of total predictors, the number of valid predictors, R-squared, and sample size were manipulated. For total number of predictors, p, the values of 4, 7, and 10 were chosen. These values are a reasonable representation of the number of predictors found in applied research settings and they are sufficiently different to illustrate potential relationships between p and accuracy of the information criteria. With 4 total predictors, conditions with 2, 3, and 4 valid predictors (v) were simulated; with 7 total predictors, conditions with 2 through 7 valid predictors were simulated; and with 10 total predictors, conditions with 2 through 8 valid predictors were simulated. For $p = 10$, 9 and 10 valid predictors were not included because predictor-criterion correlations for a ninth and tenth valid predictor at $R^2 = .1$, after controlling for the first eight predictors would have been trivially

small. Furthermore, research contexts rarely incorporate 9 or 10 valid predictors for a single criterion.

Three values of R-squared, .1, .4, and .7, were evaluated. These values were chosen to represent small, moderate, and large multiple correlations, respectively. They were also chosen to allow for consideration of accuracy trends that were a linear function of R-squared.

Each combination of the above factors was tested with sample sizes that were 5, 10, 20, 30, 40, 60 and 100 times the number of total predictors. Relative sample sizes were used rather than absolute sample sizes, because sample size recommendations in multiple regression are typically a function of the number of predictors in the model. These values for relative sample size were chosen to simulate conditions that were below generally accepted levels, at or somewhat above generally accepted levels, and clearly above generally accepted sample sizes.

All simulations were carried out by programs written and executed using SAS 8.0, and 1000 iterations were conducted for each condition. The simulated data were generated for each condition based on a correlation matrix with the designated number of nonzero correlations between predictors and the criterion. The correlations in each combination increased from zero in a linear fashion based on their squared values, such that the r^2-values summed to the designated R^2-value. All correlations among predictors were set at 0. Although, in applied work, predictors are not independent of each other, this design does not lose generalizability since this is equivalent to residualizing the predictor-criterion correlations for all but the strongest predictor to compute R-squared, which results in all these intercorrelations becoming 0, regardless of their original values.

Results

Best Overall Models

The valid predictor ratio, VPR = v/p, is defined as the ratio of valid predictors (v) to total predictors (p). For purposes of interpreting accuracy in selecting true models, values of at least 70% were considered satisfactory. The percentage of correct selection is presented for AIC and BIC in Tables 1 and 2 (see Appendix A). Results based on sample size sorted by total numbers of variables

equal to 4, 7 and 10 are summarized as graphs in Figures 1, 2, and 3, respectively (shown following tables in Appendix A).

BIC

The accuracy of BIC for selecting the best overall model consistently improved as sample size increased and as R-squared increased. In general, accuracy declined with increases in the total number of predictors, p, with an exception being the behavior for two valid predictors, where accuracy steadily improved as p increased. The relationship of accuracy to VPR was not as straightforward, being complicated by interactions with sample size, R-squared, and *p*. For all combinations of R-squared and total number of predictors, there was an inverse relationship between accuracy and VPR for values of *p* at n = 5p. For R^2 = .1, this relationship held across all sample sizes, with the differences between VPR's generally increasing with sample size. For R^2 = .4, the differences in accuracy between the VPR's within *p* slowly decreased, with the mid-range VPR's consistently being superior to the others at the two largest relative sample sizes. For R^2 = .7, there was an inverse relationship between VPR and accuracy at the lowest sample sizes; the relationship became direct, however, by n = 30p with p = 7, and n = 20p at 4 and 10 total predictors.

For R^2 = .1, the accuracy of BIC was generally low. In only 10 of the 112 combinations in the simulation design did BIC achieve acceptable accuracy, doing so when n ≥ 400 with two valid predictors, n ≥ 600 with three valid predictors, and at n = 1000 with a VPR of 4/10. For R^2 = .4, the accuracy of BIC improved. For v = 2, sample sizes of 10p were adequate to achieve acceptable accuracy. As VPR increased within *p*, and as *p* increased, the sample size necessary for acceptable accuracy also increased. At VPR's of 7/7 and 8/10, for example, acceptable accuracy was not achieved until n = 60p, while at VPR = 4/4, BIC was 69.2% accurate at n = 30p and 80.5% accurate at 40p.

For R^2 = .7, BIC was quite accurate at all but the smallest relative sample size. At n = 5p, BIC's accuracy was only acceptable with VPR = 2/4. At n = 10p, only VPR's of 7/7, 7/10, and 8/10 failed to achieve acceptable accuracy. For the remaining relative sample sizes with R^2 = .7, BIC was at least 80% accurate.

AIC

Like BIC, the accuracy of AIC at selecting the best overall model consistently declined as the total number of predictors was increased. This was the only similarity in the pattern of results for AIC and BIC. The change in accuracy of AIC was not stable across any other single variable.

AIC was consistently at its worst at the smallest sample sizes, with improved accuracy attained with medium sample sizes. For larger sample sizes, AIC behaved nearly at its asymptote, although rarely at or near 100% accuracy. Only VPR's of 4/4 and 7/7 approached 100% accuracy, doing so at the higher relative sample sizes with R^2 = .4, and doing so for n ≥ 30p with R^2 = .7. As R-squared increased, each VPR behaved asymptotically at gradually smaller relative sample sizes. Lower VPR's stabilized around their asymptotes sooner, in terms of sample size, than higher VPR's due to a general tendency for the higher VPR's to be less accurate at the smaller sample sizes and due to the fact that higher VPR's consistently had higher asymptotes.

For the combinations with R^2 = .1, AIC achieved acceptable levels of accuracy even less frequently than did BIC, breaking the 70% barrier in only two cases: n = 400 at VPR's of 2/4 and 3/4. With R^2 = .4, AIC did poorly for p = 10 with only the v = 8, n = 1000 case reaching satisfactory accuracy. At VPR = 7/7, AIC performed well for sample sizes of at least 30p.

AIC achieved acceptable accuracy at VPR's of 2/4, 3/4, and 4/4 by n = 20p (albeit asymptotically for 2/4). For R^2 = .7, all VPR's with p = 4, reached acceptable accuracy by 10p (again asymptotically for 2/4). With VPR = 5/7, the accuracy of AIC again appeared asymptotic at 70%, but the VPR's 6/7 and 7/7 demonstrated acceptable accuracy for all but the smallest sample size With eight valid predictors out of 10 total, AIC's accuracy seemed to be asymptotic for a value just above 70% at n ≥ 30p.

Comparison of BIC and AIC

At VPR's of 4/4 and 7/7, AIC was consistently as good as or better than BIC at selecting the correct overall model regardless of sample size and R-squared. With R^2 = .1, AIC

outperformed BIC at all sample sizes when the VPR > .5. For R^2 = .4, AIC consistently outperformed BIC only at n = 5p and n = 10p in conjunction with VPR's above .5. For R^2 = .7 and VPR > .5, AIC outperformed BIC only at n = 5p and for all other cases BIC outperformed AIC.

Patterns of Misselection

Unlike the accuracy patterns of BIC and AIC, patterns of incorrect choices are nearly identical and relatively straightforward. The incorrect decisions made by both AIC and BIC tended to be in the direction of more complex models when sample size was large and valid predictor ratio was low. At lower sample sizes and higher valid predictor ratios, both criteria tended to select models that were too simple.

The rates of change from errors of complexity to errors of simplicity, however, were appreciably different for AIC and BIC. As sample size increased with decreasing VPR, incorrect decisions by BIC tended toward simpler models until reaching the higher relative sample sizes with the lower VPR's. AIC, by contrast, made more errors of simplicity than of complexity only at the combination of the lower sample sizes and higher VPR's.

Results were also obtained for incorrectly selecting models with the correct number of predictors but not the actual best predictors. This type of error occurred more often with AIC than with BIC and in general, it happened more often at smaller sample sizes, smaller R^2-values, and for more total predictors. The relationship between VPR's and the frequency of this type of incorrect selection interacted with R-squared and sample size. For R^2 = .1, these errors occurred predominantly at lower relative sample sizes with lower VPR's. As VPR increased, the distribution became slightly quadratic, with the error occurring most at the moderate sample sizes and tapering to either side of the middle. At the higher values of VPR, the larger relative sample sizes contained the highest frequencies of this type of error.

For R^2 = .4, incorrectly selecting the right number but wrong set of predictors was generally limited to the lower sample sizes with the overall frequency dropping off rapidly after VPR = .5. For R^2 = .7, this type of error was rare; at no sample size above 5p was the frequency greater than 4.3% of the iterations, the frequency never exceeded

10% for BIC and only at VPR's of 7/10 (.136) and 8/10 (.139) did it exceed 10% for AIC.

Conclusion

The results of the present study suggest that different multiple regression scenarios in applied research call for different information criteria for selecting the best set of predictors. As is so often the recommendation in research, the larger the sample sizes the better; both BIC and AIC were increasingly more accurate as sample size increased. The information criteria were also generally more accurate as the number of total predictors decreased, although the reverse was true of BIC with two valid predictors. The results also provide some unfortunately complex recommendations for accuracy based on interactions of VPR with other facets of model conditions.

When all, or nearly all, predictors in a set are valid predictors, AIC is as good as or better than BIC at selecting the best overall model at every sample size and R^2-value tested. When R-squared is low, the advantage of AIC at higher valid predictor ratios is essentially moot, because at higher VPR's neither information criterion reached satisfactory accuracy (except AIC at VPR = 3/4 and n = 100p). With higher multiple correlations, however, AIC was at least 70% accurate at high VPR's and sample sizes of 20 to 30 times the number of predictors (with a negative relationship between sample size and R-squared required for good accuracy). For VPR's above .5 but below .8, sample size affects the relative performance of BIC and AIC. AIC is the better choice for relative sample sizes below 30p when R^2 < .7. BIC is generally the better choice for relative sample sizes of at least 30p or when $R^2 \geq$ .7, with one exception in the current study at VPR = 3/4 and R^2 = .1 in which AIC is better across sample size. It should be noted, however, that with VPR's in the .5 to .8 range and relative sample sizes below 30p, neither AIC nor BIC reached satisfactory accuracy with R^2 < .7, so AIC's advantage in such situations may not have practical importance.

For VPR's $\leq$ .5, BIC performed uniformly better than AIC. The importance of this advantage was related to R-squared. With small multiple correlations, BIC only achieved satisfactory

accuracy at low VPR's for relatively large sample sizes ($n \geq 400$). At moderate levels of R-squared, BIC begins to perform well at lower relative sample sizes (20p with 3 valid predictors and 10p at $v = 2$, with $R^2 = .4$) when the VPR is low. At extremely high values of R-squared, BIC is at least 70% accurate with sample sizes that are 10 times the number of predictors when VPR is low.

The sample sizes chosen for the present study seemed to provide a reasonable illustration of the patterns of accuracy at fixed relative sample sizes. There were, however, very few conclusions that could be made based on absolute sample size. Restructuring the tables and charts to line up sample sizes would line up only similar sample sizes, the conclusions of which would be confounded by having only similar valid predictor ratios. It might therefore be fruitful to investigate patterns of the accuracy of information criteria as a function of absolute sample size.

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov & F. Csake (eds.), *Second International Symposium on Information Theory*. Budapest: Akademiai Kiado, 267-281.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control, AC-19*, 716-723.

Akaike, H. (1978a). A new look at the Bayes procedure. *Biometrika, 65*, 53-59.

Akaike, H. (1978b). A Bayesian analysis of the minimum AIC procedure. *Ann. Inst. Statist. Math.*, *30*, 9-14.

Bai, Z. D., Krishnaiah, P. R., Sambamoorthi, N., & Zhao, L. C. (1992) Model selection for log-linear model. *Sankhya B*, *54*, 200-219.

Bozdogan, H. (1987). Model-selection and Akaike's information criterion (AIC): The general theory and its analytical extensions. *Psychometrika, 52*, 345-370.

Burnham, K. P. & Anderson, D. R. (1998*). Model Selection and Inference: A Practical Information-Theoretic Approach*. New York: Springer-Verlag.

Draper, N. R. & Smith, H. (1998) *Applied Regression Analysis (*3rd ed*.)*. New York: Wiley.

Hurvich, C. M. & Tsai, C-L. (1991). Bias of the corrected AIC criterion for underfitted regression and time series models. *Biometrika, 78*, 499-509.

McQuarrie, A. D. R. & Tsai, C.-L. (1998). *Regression and Time Series Model Selection*. Singapore: World Scientific.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics, 6*, 461-464.

Appendix A: Tables & Figures

Table 1. Percentage of correct model selection for AIC.

p				4					7							10
R²	v	2	3	4	2	3	4	5	6	7	2	3	4	5	6	7	8
	n=5p	9.7	3.9	1.8	8.8	2.2	0.7	0.4	0.1	0	8.7	3.3	0.4	0.2	0.1	0.1	0
	10p	17.2	4.8	2.1	19	6	1.9	0.5	0.1	0.1	14.5	8.6	2.8	1.6	0.5	0	0
	20p	35.5	18.2	5.7	29.3	20.2	10.5	5.6	2	1.3	20.9	18.1	10.6	6.5	3.4	1.7	0.2
	30p	48	30.5	13.3	38	33	20.5	11	4.6	2.7	23	21.6	19.2	13.5	7.8	4.6	2.3
	40p	57.8	42.1	23	37.6	39.8	30.3	20.3	12	6.1	23.6	26.5	23.6	20.3	13.2	8.3	7
	60p	66.6	59.5	39.6	43.5	43.3	42.4	37.7	25.9	19.7	27.1	28.6	28.8	30.6	25.7	20.1	16.7
.1	100p	72.7	73.6	69.3	40.1	47.7	53.1	53.7	51.1	40.3	25.1	29.1	34.3	37	40.1	39.4	35.8
	n=5p	44.4	25.3	13.3	28.5	21.9	13.4	9.7	6	2.5	16.6	19.6	13.8	8.1	6.6	3.9	2.2
	10p	61.1	54	39.4	35.9	41.4	40.2	32.6	24.2	16.9	22.9	25.4	29.1	25.9	21.9	19.5	15.6
	20p	68.7	77.2	74.7	43	46.5	55.3	59.2	56.6	49.2	24.3	31.8	33.1	39.1	39.4	41.6	40.5
	30p	69.3	81.5	91.8	39.9	49.1	59.7	66.2	70.4	72.7	23.5	29.7	37.7	38.9	48.3	49.9	56.1
	40p	70.1	82.6	95.9	43.9	48.4	57.1	67	79.9	88.2	25	28.7	33.7	39.6	47	54.6	64.8
	60p	71.4	84.2	99.2	39.7	49.7	61.1	71.7	83.4	95.8	22	28.4	33.9	39.3	48.1	60.1	67.3
.4	100p	69.6	82.2	100	43.4	48	59.7	68.3	85.5	99.5	24.5	32.2	35	44.5	48.5	61.7	70.6
	n=5p	60	62.4	61.8	35	41.6	41.5	45	45.4	38.7	19.1	26	29.5	27.5	33.1	31.3	28.2
	10p	69.3	77.7	92.5	38.2	48.1	56.8	66.2	72.9	76.8	22.8	26.2	31.3	39.1	42.9	50.1	54.3
	20p	67	81.9	100	41.6	52	58.6	66.8	84.4	95.4	23.5	29.7	33.2	40.7	47.6	56.9	68.3
	30p	69.2	84.5	100	40.3	49.3	59.5	68.8	84.8	99.7	22.1	30.8	35.3	40.2	49.2	61.2	70.3
	40p	71	83.6	100	39.9	47.7	61.3	68.6	83.1	100	27.4	28.8	33.7	40.6	49.6	60.6	72.6
	60p	70.9	83.2	100	39.2	48.6	59.3	71.3	84.5	100	26.7	32.2	35.1	42.2	50	59.9	70.5
.7	100p	71.8	83.8	100	44.2	48	61.3	70.7	82.2	100	25.7	30.9	35.6	43.2	49.2	58.3	73.4

Table 2. Percentage of correct model selection for BIC.

p				4					7							10
R²	v	2	3	4	2	3	4	5	6	7	2	3	4	5	6	7	8
	n=5p	6	2	0.3	7.7	0.9	0.1	0.1	0	0	8.2	1.4	0.1	0	0	0	0
	10p	8.8	1.5	0	16.3	1.5	0.1	0	0	0	21.5	3.9	0.1	0.2	0	0	0
	20p	22.3	4.1	0.3	34.6	8.5	1.4	0.1	0	0.1	48.5	17.6	3.8	0.1	0.1	0	0
	30p	34.3	7.5	0.9	57.9	21.3	4.2	0.5	0	0	69	32.4	11.6	2.2	0.4	0.3	0
	40p	45.8	12.4	1.7	69	29.2	13.3	1.9	0.4	0.1	82.9	46.8	19.5	5.6	1.3	0.2	0
	60p	68.3	26.8	7.2	86.8	53.6	25.1	6.8	1.2	0.3	89.5	71.2	40.5	18	4.9	1.4	0.7
0.1	100p	88.2	53.2	21.1	94.2	79.8	49.9	24.3	8.9	2.9	91.8	91	70.4	46.9	23.5	10.4	3.4
	n=5p	44.5	17.9	6.5	48.4	23.5	8.3	4.1	1.4	0.2	50.4	35.5	15.2	5.7	2.1	0.2	0.2
	10p	72.2	40.1	17.7	73.1	58.8	33.8	17.5	5.7	2.7	74.8	63.4	48.3	27.3	11.7	5.3	2.2
	20p	88.8	74.1	46.4	86.9	85.9	69.8	46.2	29.1	13.6	83.1	86	74.9	63.5	44.7	28.2	14.9
	30p	93.1	87.9	69.2	89.8	90.4	84.3	71.9	51.2	33.1	85.9	87.4	88.8	81.4	66.7	52.4	36
	40p	94.7	94.1	80.5	91.7	92.9	90.5	82.2	67.3	50.7	88.5	90	90.5	89.3	79.1	65.9	52.5
	60p	95.5	97.2	93.3	91.8	94.6	96.7	92.3	87.3	70.6	90.6	91.1	94.3	94.4	92.1	86.2	75.6
0.4	100p	97	98.6	99.3	94.6	94.1	97.1	97.4	96.7	91.9	92.2	94.6	94.3	95.8	96.8	97.3	93.4
	n=5p	72.7	62.8	45.7	65.4	67.2	54.1	43.7	33.1	17.3	58	61.6	61.3	50.6	38.5	26.8	18.2
	10p	88.2	87.4	80.6	79.2	83.1	80.8	77.3	67.9	51.6	74.8	77.2	77.4	79	72.1	63.3	51
	20p	91.7	95.2	97.9	86.8	88	90.9	91.5	91.1	85	83.2	84.7	87.1	89.9	89.9	89.4	84.9
	30p	92.6	96.8	99.9	89.7	90.7	92.5	96.1	97.1	96.5	85.5	87.2	90.2	91.5	93.6	93.8	93.2
	40p	94.6	96.6	100	89.9	92.2	94.3	95.6	98.4	99.2	88.2	90.6	91.8	92.7	95.1	95.2	96.3
	60p	95.5	97.9	100	92.5	94.5	94.4	97.7	98.3	100	90.3	91.7	92.8	94.4	95	95.8	97.2
0.7	100p	97.2	98.6	100	94.2	96.9	96.6	98.1	99	100	93.4	94.5	95.7	95.2	96.4	97.7	98.4

Figure 1. Percentage of correct model selection for BIC and AIC; four total predictors
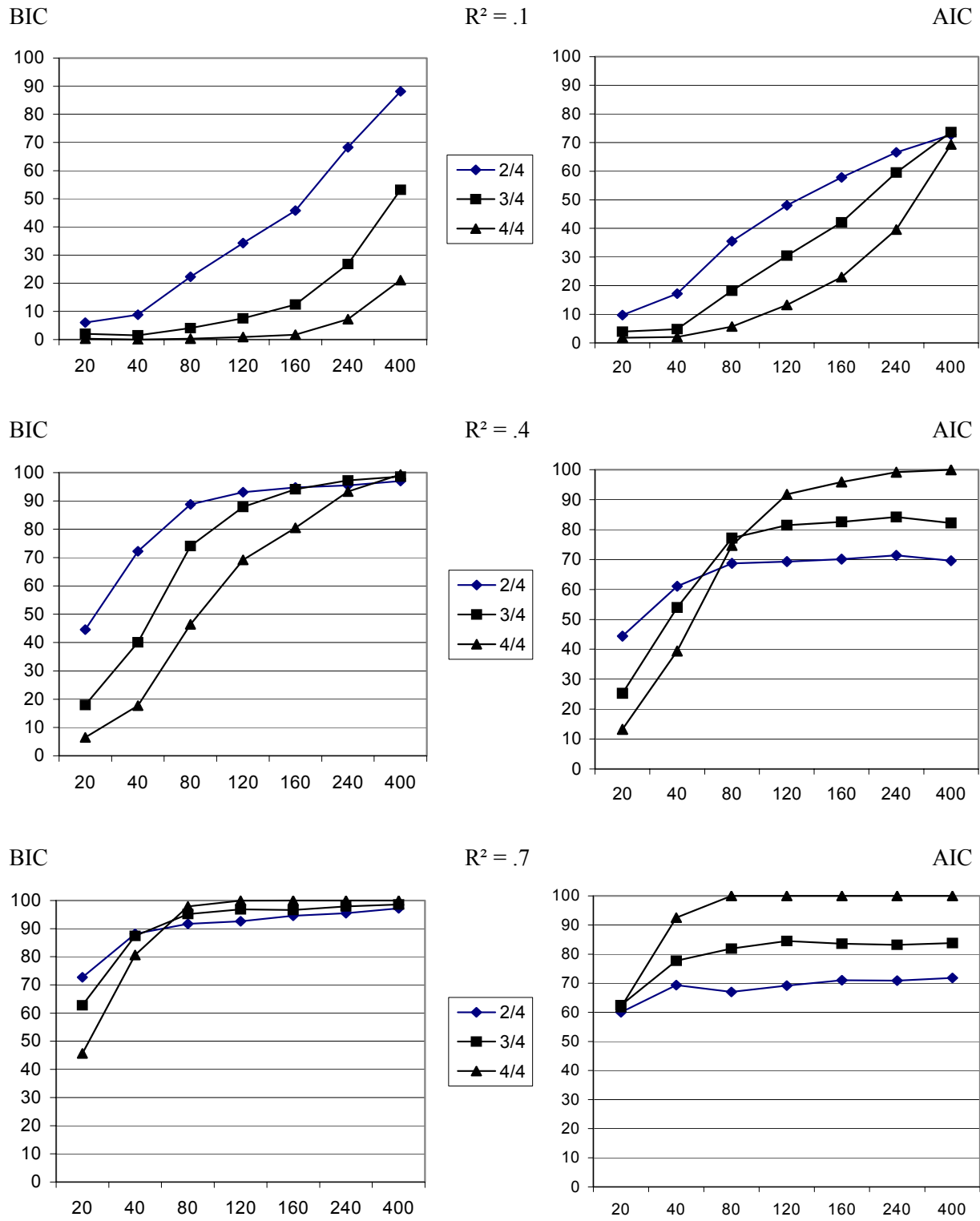
Figure 2. Percentage of correct model selection for BIC and AIC; seven total predictors
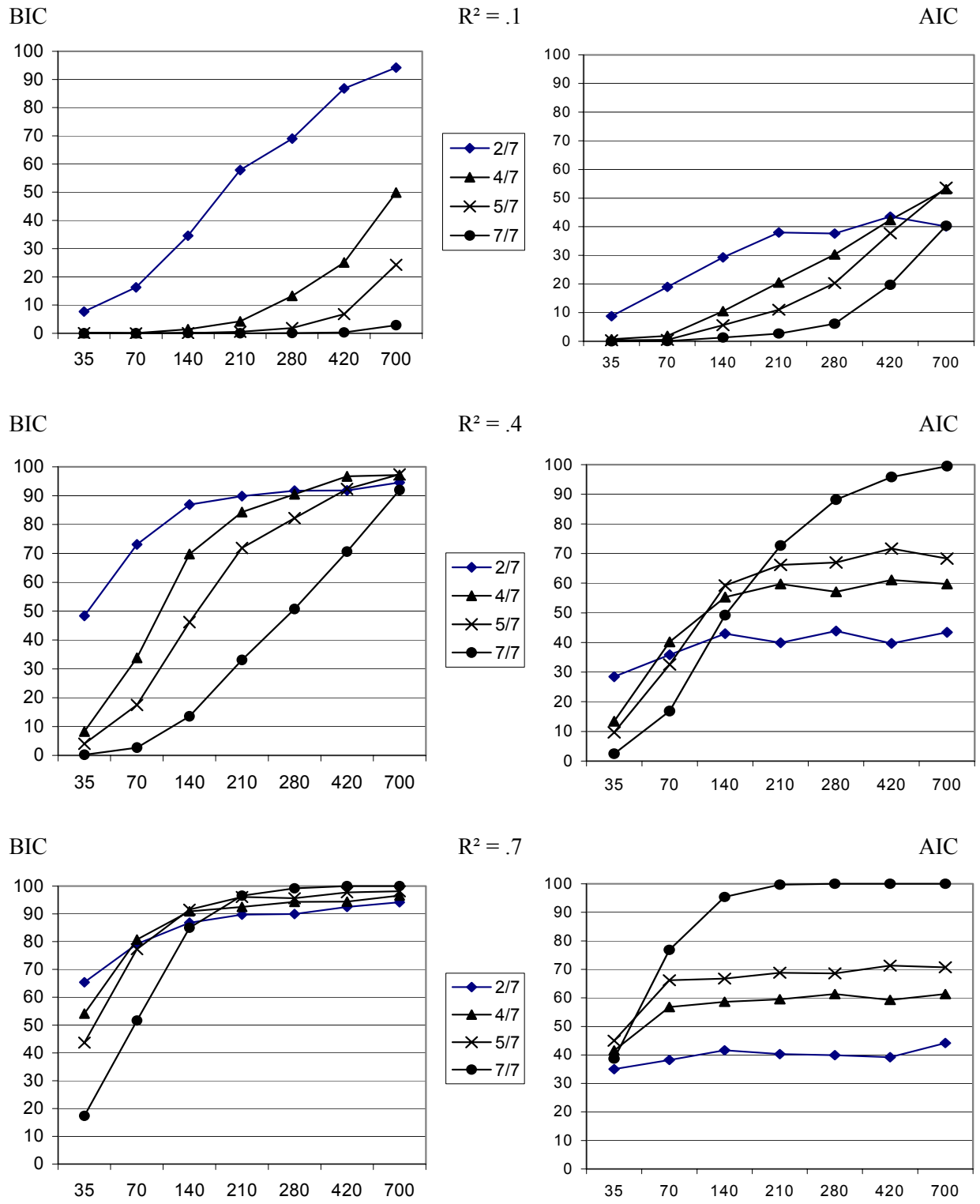
Figure 3. Percentage of correct model selection for BIC and AIC; ten total predictors