MULTIPLE LINEAR REGRESSION VIEWPOINTS
VOLUME 19, NUMBER 1, SUMMER 1992

A Graphical Method for Selecting the Best Sub-set Regression Model

Mark Alexander Constas
University of Northern Colorado

Joe D. Francis
Cornell University

Abstract

The purpose of the present paper is to provide an empirical example of a graphical procedure that may be used for parameter selection in multiple linear regression..An all possible sub-sets approach is utilized in order to accomplish this objective. Elements related to the model construction process, such as parsimony, and explanation value are explored in the context of this approach..A graphical presentation of the findings is discussed as a way of facilitating the understanding of how R^2, adjusted R^2 and the number of parameters in a model are related to one another.

A Graphical Method for Selecting the Best Sub-set Regression Model

Stated in general terms, the purpose of many studies using regression analysis is to determine which variables or combination of variables offers the best prediction for a given dependent variable. Most will agree that there does not exist a single indicator or method of determining the absolute goodness of a given model. It is often suggested that researchers employ the tripartite criteria of theory, economy, and explanation, where theory is concerned with the set of theoretical rationalizations for proceeding with a selected group of variables, economy is centered around issues such as simplicity or efficiency of explanation (i.e. how many variables does it take to achieve some reasonable level of prediction) and explanation is concerned with the amount of variance explained by a given model.

The primary purpose of the present paper is to describe a graphical method for applying the criteria of economy and explanation in the process of constructing a prediction model using regression analysis. Of central importance here is illustrating the relationship between R^2, adjusted R^2 and the number of parameters in a model. After providing a brief description of the empirical context of the analysis, the paper will proceed to illustrate the way in which a graphical comparsion of R^2 and adjusted R^2 makes clear an important concept in the parameterization of multiple linear regression problems. It is argued that the graphical clarity of the method helps explain the utility of using adjusted R^2 in $k + 1$ regression problems.

Empirical Context

The data used to demonstrate the method being discussed here were derived from a study that sought to determine the relative importance of two types of cognitive variables in predicting the clinical skills of medical students. While one domain of cognitive functioning (cognitive preference) was composed of four variables, the second domain (knowledge competencies) was composed of two. Stated more specifically, the cognitive preferences variable was composed of four independent scores that represented an individual's preference for four different kinds of cognitive functioning (Recall, Principles, Application, and Questioning). The knowledge competency domain was composed of two grade point averages that reflected a given student's level of academic achievement for two distinct periods of his/her medical education. In all cases a total of 14 terms were included in the model construction process. The total of 14 was accumulated by having four terms from the cognitive preferences domain, two terms from the knowledge competencies domain and eight interaction terms that were products of the simple terms.

Analytical Framework

While there are a number of different methods for generating prediction models in the context of multiple linear regression, the most comprehensive and obviously the most exhaustive method involves running regressions between the dependent variable and all possible subsets of the independent variables. With k regressors one may generate 2^{k-1} models. As one can see, the number of models to consider will grow to a large number when trying to construct a model with only a small number of variables. With $k=14$, as in the case for the present empirical example, the number of models generated exceeds 8,000. While the development of high speed computers has almost trivialized calculation procedures, subsequent decisions about which variables to include in the initial runs and which models to select for further analysis are not simplified.

Within the framework of multiple regression, R^2 is often used as a general indicator of the power of a given model. Although R^2 exists as a convention for model selection and evaluation there are some rather fundamental limitations of relying on that statistic. For example, it is important to note that R^2 will continue to increase as a direct function of the number of parameters (k) in the model. It could be argued that a strong reliance on R^2 is inappropriate, given the illusory effects of increasing k. One can see the way in which R^2, being a partial artifact of k, may be misleading.

As an antidote to the problems associated with R^2, the adjusted R^2 has a built in discounting factor that counters this rather serious flaw in R^2 by attaching a penalty clause for increasing the value of k (see Darlington, 1968; Kerlinger and Pedhazur, 1982). The equation takes the following form:

$$R^2\,(\text{adj}) = 1 - (1 - R^2)\ \frac{(N-1)}{(N-k-1)}$$

where
N = sample size
k = number of parameters

The presence of the "N-k-1" component in the equation has an attenuating effect that provides a correction for increments in R^2 that are associated with simply increasing the number of parameters in a given model.

A Graphic Demonstration

A basic consideration in model construction often concerns the number of parameters to include in the model. When an a priori decision has not been made regarding parameterization one must proceed in an inductive fashion where empirical outcomes more actively determine the number of variables to include in a given model. In the present example $2^{14}-1$ equations of varying combinations and lengths were generated in order to find the maximal prediction equation.

The graphic approach for selecting the maximum value for k involves plotting the R^2 <u>and</u> adjusted R^2 values against k. This procedure gives one a visual display to help determine the point at which the incremental value of R^2 is insufficiently large to counter the unwanted effects of increasing k. In a typical plot of R^2 against k, the curve rises more steeply or less steeply, depending on the nature of model specification. After the addition of a certain number of parameters the curve will usually begin to flatten. The notion of using a flattening area as a termination point for adding additional predictors is often employed as a decision rule in model construction. To many, this decision rule may appear questionable, since the perception of flatten may seem subjective.

A plot of R^2 against k fails to reveal a definite turning point. By comparison the adjusted R^2 aginst k plot demonstrates a distinct point of descent. More than a mere perturbation, there is a very real turning point to be observed. This point may serve as a ceiling for the number of parameters to be used in model construction.

Insert figure 1 about here

Figure 1 provides an illustration of the relationship between R^2, adjusted R^2 and the number of parameters in the model. The point of descent mentioned above appears in a rather clear way. One should also notice the way in which R^2 continues to increase in relation to R^2 (adj.) This information suggest that one should not proceed beyond a certain level of k. Although the area between k=3 to k=7 should be considered more closely, the level of k selected should certainly not exceed seven.

Subsequent Procedures

Having decided on the number of parameters to include in the model, issues such as simplicity, theoretical relevance, and ease of explanation may be considered more closely. The next step is may be to obtain the combinatoric options for k=1 to k=7. For purposes of illustration, permutations of variables, for only the top two candidates at each level of k are presented in Table 1.

Insert Table 1 about here

The primary criterion that one may apply at this point is often invoked under the term "parsimony." A "parsimonious" model is one that contains the parent terms of any interaction terms that may appear in the model while simultaneously using the fewest number of parameters to achieve the greatest amount of explanation. Application of this criterion led to the selection of the model marked as "tested" in Table 1. Although there is a gain of approximately .4 when moving from three to seven parameters, it was decided that the value of this increment is dubious, given the cost. The necessity of using four more parameters does not support the notion of parsimony. The model selected could then be subject to more detailed statistical scrutiny such as tests of significance.

Summary Statement

There are a wide variety of methods for constructing models in multiple regression. In the case where one has chosen to use the all possible regressions approach some defensible procedure is needed to help make decisions about the size and contents of a final model. Admittedly, The model construction procedure followed here was not informed by an incredibly strong theoretical base, hence the decidsion to proceed with the all possible sub-sets approach. Such a situation is not uncommon in social and educational research. Results of the kind obtained here may provide one with enough empirical evidence to perform a replication or to forge an inductively derived theoretical base. Some progress may be realized.

The relation between R^2, adjusted R^2 and the number of parameters in the model is an important one to understand. Although a tabular display of these date will reveal the relationship, a graphical expression may make the association more explicit. In summary, one may argue that the present approach to generating a regression model is useful in at least two areas. Firstly, it provides one with a reasonably objective method for defining the upper limits for model construction. Secondly, the graphical method has proven to be quite useful in instructional settings for demonstrating the weaknesses associated with the R^2 selection method. The procedure is also useful in that it provides a rather telling illustration of the relationship that exists between R^2 adjusted R^2, and the number of parameters in a model.

References

Darlington, R. B. (1968). Multiple regression in psychological research and practice. Psychological Bulletin, 69, 161-182.

Kerlinger, F., and Pedhazur, E. (1973). Multiple Regression in Behavioral Research. New York: Holt, Rinehart and Winston.

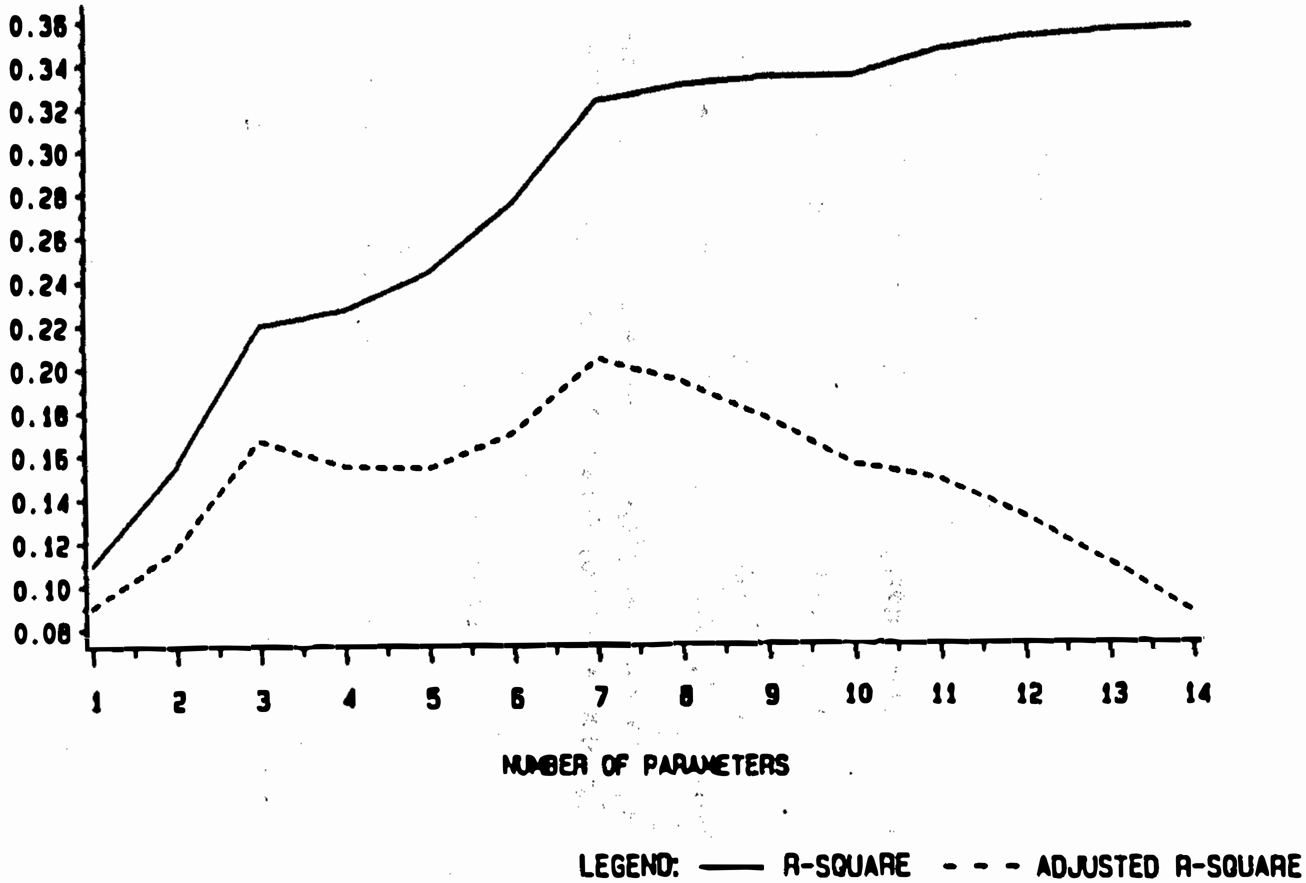

Figure 1. Graphical relationship between R^2, adjusted R^2, and the number of parameters in the model

Table 1

Model Specifications for Prediction of Clinical Skills

k	R^2 (adj.)	Terms in the Equation
1	.090	(principles x GPA^a)
	.089	(Application x GPA^b)
2	.117	(Questioning), (Questioning X GPA^b)
	.107	(Application), (GPA^b)
3	.167*	(Application), (GPA^a), (Application X GPA^a)
	.129	(Application), (GPA^b), (Application X GPA^b)
4	.155	(Application), (GPA^a), (Principles X GPAb), (Application X GPA^a)
	.152	(Application), (GPA^a), (GPA^b), (Application X GPA^a)
5	.154	(Principles), (Questioning), (GPA^b), (Principles X GPA^b), (Questioning X GPA^b)
	.153	(Principles), (Application), (GPA^a), (Principles X GPA^a), (Application X GPA^b)
6	.170	(Principles, (Application), (Principles X GPA^a), (Application X GPA^a), (Questioning X GPA^a), (Questioning X GPA^b)
	.163	(Principles, (Application), (GPA^a), (GPA^b), (Principles X GPA^b), (Application X GPA^a), (Questioning X GPA^a)
7	.204	(Principles), (Questioning), (GPA^a),(Recall X GPA^a), (Recall X GPA^b). (Principles X GPA^b), (Questioning X GPA^a)
	.187	(Principles), (Questioning), (GPA^b), (Principles X GPA^b), (Application X GPA^a), (Application X GPA^b), (Questioning X GPA^a)

Key:

k	Number of parameters in the model
R^2 (adj.)	Adjusted R^2
*	Model tested in subsequent analysis