MULTIPLE LINEAR REGRESSION VIEWPOINTS
VOLUME 15, NUMBER 1, SUMMER 1986

Microcomputer Selection of a Predictor Weighting Algorithm

John D. Morris
Florida Atlantic University

An empirical method (PRESS) for examining and contrasting the cross-validated prediction accuracies of some popular algorithms for weighting predictor variables was advanced and examined. The weighting methods that were considered were ordinary least squares, ridge regression, regression on principal components, and regression on an equally weighted composite. PRESS was executed on several data sets having varied characteristics, with each of the weighting techniques obtaining the greatest accuracy under some conditions. The degree of advantage or disadvantage offered by these alternate weighting algorithms relative to ordinary least squares was considered. As it was not possible to determine _a priori_ which weighting technique would be most accurate for a particular data set from theoretical knowledge or from simple sample data characteristics, the sample specific PRESS method was proffered as possibly most appropriate when the researcher wishes to select from among the several alternate predictor weighting algorithms in order to achieve maximum cross-validated prediction accuracy. The feasibility of the use of a microcomputer for the computation intensive PRESS algorithm was also considered.

Many empirical and theoretical studies (Darlington, 1978; Dempster, Schatzoff, and Wermuth, 1977; Gibbons, 1981; Morris, 1979; Pruzek and Frederick, 1978; Wainer, 1976) have suggested that there are more accurate (in the sense of cross-validatior predictor weighting strategies than the traditionally used ordinary least squares (OLS).

Much of the effort has concentrated on ridge regression, with Darlington's (1978) recommendations being by far the strongest in the behavioral sciences. However, some more recent results (Morris, 1982, 1983) suggest a less enthusiastic outlook toward ridge regression in the specific situations considered by Darlington (1978), but a possibly more promising outlook under other data conditions (Morris, 1981). Additional evidence and reservations of others about ridge regression may b found in Egerton and Laycock (1981), Pagel and Lunneborg (1985), Rozeboom (1979), a Smith and Campbell (1980).

Similar controversy spanning at least a quarter of a century (Claudy, 1972; Dawes and Corrigan, 1974; Dorans and Drasgow, 1978; Einhorn and Hogarth, 1975; Gabriel, 1980; Laughlin, 1978; Lawshe and Schucker, 1959; Pruzek and Frederick, 197 Schmidt, 1971; Trattner, 1963; Wainer, 1976, 1978; Wesman and Bennett, 1959) has surrounded the use of equally weighted predictors as a substitute for OLS weights. In addition, several investigators have proposed the use of reduced-rank prediction methods to enhance cross-validation prediction accuracy, possibly beginning with Burkett (1964), to more recently (Morris and Guertin, 1977; Pruzek and Frederick, 1978).

It seems clear that claims for a "panacea" weighting technique to fit all data configurations, such as ridge coefficients "will undoubtedly be closer to (the true parameters) and are more stable for prediction than the least squares coefficients" (Hoerl and Kennard, 1970, p. 72), or "Ridge regression is the best technique for a broad range of intermediate values of validity concentration and is little worse th alternative techniques at the extremes" (Darlington, 1978, p. 1250) are unrealistic Equally clear is that many simulation results strongly suggest that non-OLS weighti strategies offer the researcher enhanced cross-validation prediction accuracy in ma data configurations. The most important next step seems to be to determine the frequency with which such data configurations that are conducive to non-OLS methods occur in the behavioral sciences and to examine the importance of the gain or loss resultant from using these strategies. Given encouraging gains in a reasonable proportion of available data sets, another step would be to generate mechanisms for helping the researcher decide which of the alternate weighting techniques are best for which data situations, and for estimating how much improvement or degradation might be realized by using an alternate technique instead of OLS in a specific data set.

Some simulation results (Morris, 1981, 1982; Pagel and Lunneborg, 1985;) have yielded some <u>general</u> suggestions for when to use which technique. One major factor suggested by Pruzek and Frederick (1978) and explicated more explicitly by Darlingt (1978), is validity concentration, the degree to which predictive validity is concentrated in the first few principal components of the predictors. From simulation results and theory (Darlington, 1978; Morris, 1982; Pagel and Lunneborg, 1985), it is know that as predictor variable collinearity and validity concentratio increase, non-OLS methods usually become more accurate than OLS at some point. In addition, Cattin (1981) has argued that in typical behavioral science data small eigenvalues from the predictor variable intercorrelation matrix tend to explain mor noise than signal. Thus as the validity concentration is high, non-OLS methods are usually most accurate. However, this tendency is diminished by an opposite trend in favor of OLS regression as sample size and population multiple correlation increase. How these trends balance out with real data is not immediately apparent.

These effects also depend on the type of prediction accuracy of concern. Many simulation studies have concentrated on the error in estimating population regression weights. Instead, the interest in this paper is on the accuracy of criterion score prediction. This accuracy criterion seems more reasonable than that of the accuracy of estimating population regression weights because such techniques as ridge regression may be inappropriate when the sizes of regression weights are of primary concern (Darlington, 1978; Pagel and Lunneborg, 1985). Moreover, the same analytic strategy illustrated in this paper is generalizable to the task of examining errors in estimating regression weights.

However, even when limiting consideration to prediction, one must consider both "relative" and "absolute" types of prediction accuracy. Is the researcher interested in generating a prediction equation that yields predicted scores that are maximally correlated with the actual criterion score (relative), or is the goal to minimize the differences between the actual and predicted criterion scores (absolute)? These are not the same goals, and the comparative accuracies of the methods are partially a function of which one is considered.

Some theoretical (Thisted and Morris, 1980) as well as empirical (Musgrage, Marquette, and Newman, 1982) rules have been offered for determining when various types of ridge regression may be helpful in enhancing prediction accuracy. These rules do not specifically consider the effects either of validity concentration or of sample size, both of which have been shown in simulation studies to affect the relative performance of OLS and non-OLS methods. Also, as operating characteristics for these theoretical rules have not been examined through simulation, it is difficult to know how they would perform with real data. As well, the rules due to Thisted and Morris consider only ridge regression as an alternative to OLS regression.

Although some general trends and suggestions may be gleaned from these studies, it is at best difficult to suggest to an applied researcher what method to select given the specific data characteristics of a sample. The results are useful theoretically, but they are just not sufficiently simple to allow easily applicable rules to be generated to use for specific data sets. Also, such rules would require unknown population information for which one has no sample estimate, as in the case of validity concentration.

More important, very little, if any, information is available about how much gain or loss in prediction accuracy one might expect by using non-OLS weighting with <u>real data</u>. What is the potential payoff or loss for the researcher in trying these non-traditional methods?

Purpose

The purpose of this paper was to advance and examine an empirical sample-based method (PRESS) to be used for exploring the comparative performance of several predictor weighting methods on a specific data set to aid in selection, and most important, to assist in judging the probable resulting gain or loss in prediction accuracy in selecting a weighting algorithm. Although the specific technique is different, the use of an empirical sample-based method to aid in selecting a predictor weighting method is parallel with the suggestion of Dempster, Schatzoff, and Wermuth (1977, p. 106) that "it would seem that comparison of the predictive capabilities of various methods from one subset to another would provide a reasonable empirical basis for selecting a particular method in a given situation." To demonstrate the technique, the PRESS algorithm was executed on several typical, although not necessarily completely representative, sets of data. The feasibility of the use of a microcomputer for the computation intensive PRESS algorithm was also considered.

The PRESS Algorithm

Allen (1971) introduced a technique that he labeled PRESS (PRedicted Error Sum of Squares) to be used to select a multiple regression variable subset that would yield a minimum sum of squared errors in prediction on cross-validation. This algorithm is executed by alternately predicting each subject's criterion score from the regression equation generated from the predictor and criterion scores of all other subjects. The resulting squared errors of prediction over all subjects are accumulated and the sum obtained serves as a criterion for cross-validation accuracy.

Although most of the multiple regression literature dealing with this "round-robin" subject deletion strategy references Allen and terms the technique PRESS, it is not original with Allen. Perhaps the earliest explicit description of the technique was in a paper by Gollob (1967). Many researchers, however, have recommended the procedure for both multiple regression and discriminant analysis-type classification cross-validation (Allen, 1971; Allen and Cady, 1982; Lachenbruch and Mickey, 1968; Mosteller and Tukey, 1968; Stone, 1974). Additionally, the technique has also been descriptively termed "leave-one-out" (Huberty, 1984; Huberty and Mourad, 1980).

Allen (1971) also provided a derivation for a computational simplification used in calculating PRESS that requires only one matrix inversion, rather than the implied n inversions, where n is the total number of subjects. This derivation was based on a matrix identity often attributed to Bartlett (1951), although no mention was made of Bartlett's work. However, one also can find the same identity in Horst (1963, p. 428) with no mention of Bartlett. Whether all three authors independently derived the same matrix identity is unknown.

Although this algorithm was introduced to help select a subset of predictors that would yield the smallest sum of squared errors upon OLS cross-validation and to give an estimate of the resulting cross-validated prediction accuracy, the same logic and algorithm can be used to judge the cross-validated prediction accuracies (relative or absolute) of alternate predictor weighting methods; the idea is completely general across any weighting strategy. PRESS can be performed for each competing predictor weighting method, and the most accurate method can be chosen as the one most probable to be most accurate on use in replicate samples, or the researcher may decide that the gain, if offered by a non-OLS strategy, is not important enough to warrant selection of a method that may not be well known.

The computational simplification offered by Allen (1971) is rather straightforward for OLS. If one considers the usual model for multiple linear regression,

$$Y = BX + e,$$

where X is an nxp matrix of p - 1 predictor variable values and the usual unit vector, Y is the vector of criterion scores, and e is the vector of error terms, the al solution for B, the vector of regression weights, is

$= (X'X)^{-1} X'Y.$

deleting a subject would change both X'Y, and X'X, it would seem that both X'Y, the matrix inverse $(X'X)^{-1}$ would need to be recalculated as each subject is leted.

However, if $\hat{Y}_{(i)}$ is a subject i's predicted criterion score when that subject's ctor of predictor scores, X_i, and criterion score, Y_i, are excluded from X and Y, len (1971, p. 11) showed that

$) = (1 - Q_i)^{-1}\hat{Y}_i - Q_i(1 - Q_i)^{-1}Y_i,$

ere $Q_i = X'_i(X'X)^{-1}X_i$, $\hat{Y}_i$ is the subject's criterion score predicted from the gression weights based on all the sample, and Y_i is the subject's actual criterion ore. Although this formulation avoids the numerous matrix inversions, it still quires the calculation of the predicted criterion score and the Q_is for every bject. This calculation route, which was found to be as much as an order of gnitude faster than actually calculating the inverses in a recent comparison orris, 1984), requires very little extra computation if one ordinarily calculates siduals.

The most obvious step would then seem to be to try to adapt this computational ortcut for use with the non-OLS methods of interest. In fact, by recognizing the lationship between OLS, principal component, and ridge regression, one not only can opt the algorithm, but also can do the calculations for the methods essentially multaneously. As well, the Allen formulation obviously fits the case of regression an equally weighted composite, as regression on such a composite just turns out to a case of simple regression.

In fact, in a later publication, Allen (1972) provided a version of the shortcut rmula for ridge regression. Given the usual simple ridge regression model of

$= (X'X + kI)^{-1} X'Y,$

len showed that it followed that PRESS can be calculated from the same formulation with OLS except that the kI would be added to the X'X matrix before inversion in e calculation of $\hat{Y}_i$ and Q_i.

However, there is a problem with this formulation. When the researcher decides on a biasing "k" in ridge regression, it is added to the <u>correlation matrix</u> rather an to X'X. Although one can center and scale the score vectors such that X'X = R, e formulation is still incorrect since kI is being added not to the correlation trix, but to the correlation matrix decreased by the contribution of one subject. an illustrative problem with five subjects and one predictor variable (X = 2, 0, 3, 9; Y = 3, 4, 4, 7, 6; and a Dempster, Schatzoff, and Wermuth [1977] RIDGEM k 2.73), the PRESS cross-validated correlation calculated by the shortcut formula as -.92, but the true PRESS cross-validated correlation calculated by actually nverting n correlation "matrices" augmented by kI was -.07. This example is ertainly not purported to be representative. Moreover, the difference would clearly e less for samples of even moderate size and with smaller ks. However, it does llustrate that the Allen shortcut formulation for ridge regression gives incorrect esults.

Another difficulty, however, stems from the fact that for ridge regression, the used is often derived from characteristics of the sample. Thus it is also a random ariable. As the accuracy of the choice of k affects the accuracy of the resulting rediction equation, the algorithm for that choice must also be cross-validated. his task is clearly not accomplished in the Allen shortcut formulation. The same argument can be advanced for any choice made using information from the data of the sample that affects the prediction equation. Thus one also must cross-validate the algorithm for selecting the number of components in regressing the criterion on principal components, and for choosing the algorithm for deciding which variables are "salient" enough to be included in an equally weighted composite, if such judgments are to be made from sample information.

If one adopts this philosophy of cross-validating the <u>total</u> choice process involved in constructing a prediction model from sample data, then the only computational route possible is to calculate *n* versions of each equation by actually leaving a subject out each time.

A Pascal computer program was written that cross-validates OLS, ridge regression, regression on principal components, and regression on an equally weighted composite via PRESS for any input data set. One of the difficulties with such techniques as PRESS, bootstrapping (see Efron, 1979; 1983), and other resampling plans is the extreme amount of computation required. When using a mainframe or minicomputer, this translates into costly run times. As microcomputers are a "one-time" expense, such computation costs essentially nothing given the availability of the machine and software. A disadvantage of the microcomputer is that it is slower than mainframes and minicomputers. However, the degree of difference in speed is rapidly decreasing with the continuing introduction of faster and more powerful microprocessors. With this in mind, this program was used with an MS DOS microcomputer to illustrate and to examine the method on several sets of data, and to assess the performance of the microcomputer in accomplishing these relatively demanding computational tasks.

Method

Weighting Techniques

There are many possible choices for a k for ridge regression. Because of its excellent performance and its ease of calculation, the Lawless and Wang (1976) k, which is the inverse of the F ratio resulting from a test of the OLS R^2, was used for ridge regression.

Because of its ubiquity, the Kaiser (1960) rule of selecting components with roots larger than one was used to select the number of components in regressing a criterion variable on principal components. One might also consider using a significance test (e.g., Bartlett, 1950) to determine the number of predictor components to use. One should note, however, that a subjective decision would be necessary even though a significance test is used as the researcher must select a significance level.

As is often the practice, equal weighting was accomplished by specifying a threshold predictor-criterion correlation for inclusion of a predictor. The predictor then received either a +1 or -1 weight depending on the sign of the predictor-criterion correlation. The resultant composite was then used to predict the criterion. For the example data sets presented in this paper, predictor variables with a correlation significant at the .05 level were included.

Obviously, if other non-OLS strategies were used, different results might have been obtained. Likewise, with other data sets, results might have been different. The purpose, however, was a demonstration of a method for examining and comparing the accuracies of the weighting methods for specific data sets rather than a general comparison of the weighting methods.

The Demonstration Data Sets

~~Twenty-one~~ data sets of widely varying characteristics from the behavioral and natural sciences were used in this demonstration. An attempt at sampling a variety of types of data was made; however, the data sets are not advanced as representative. The results were not intended and should not be interpreted as generalizable to all behavioral science data sets. The intent was to explore and to demonstrate a strategy for estimating what one might expect for a _specific_ data set.

It also is important to note that the actual "real" data sets were used rather than Monte Carlo simulations from covariance structures as has been done in some studies mentioned previously. This procedure not only allows the characteristics of the data structures to vary as they do in nature, but also affords the unique distributional characteristics of a sample to affect the results, contrary to the situation in simulation studies in which multivariate normality is usually assured.

These data sets actually have been used in regression analyses. They are from journal articles, paper presentations, or text books. Therefore any aberrant score vectors are assumed to have been deleted. Before applying the PRESS strategy (or any other analytic method), the researcher probably would wish to consider the removal of "outliers" that manifest appreciable leverage. One may find it helpful to consider the excellent review by Hocking (1983), as well as associated comments for information on methods for detecting such score vectors.

Results

Tables 1 and 2 show the performance ~~of the~~ four weighting techniques for each of the 21 data sets. In concentrating on relative prediction accuracy Table 1 furnishes cross-validated correlations; Table 2 provides absolute accuracy as the mean squared error in predicting the criterion score. In both tables there appears (a) a short description of the origin of each data set (exact citations being available on request), (b) the OLS squared multiple correlation calculated in the total sample (RSQ), (c) the multicollinearity index due to Thisted and Morris (1980) (MI), (d) the ratio of the number of subjects to predictor variables (_n/p_), and (e) the performance of the methods, with the performance of the non-OLS methods shown as a percent of the OLS performance. It should be noted that the MI criterion proposed by Thisted and Morris is different when one considers relative and absolute accuracy.

The number of subjects ranged from 16 to 293, and the number of predictor variables varied from 3 to 17. The largest raw score matrix analyzed had 271 subjects with 12 predictors.

An interesting characteristic exhibited in the results is the amount of variety obtained. The comparative performance of the methods is clearly dependent on which data set is being considered and on whether the criterion of accuracy of concern is relative or absolute. In addition, the fact that the different methods performed better with differing real data sets may lend some credibility to such differences found in simulated data sets.

Relative Accuracy (Table 1)

Relative accuracy is discussed first. In 16 of the data sets of Table 1 (1, 2, 3, 4, 6, 7, 9, 10, 11, 12, 13, 14, 15, 18, 20, and 21) ridge performance was about the same as that of OLS (within 2%). However, within these same data conditions, the accuracies of regressing the criterion variable on principal components, and of regressing the criterion variable on an equally weighted composite were much less consistent. Sometimes these procedures were also very close to OLS performance. In one data set (10) they were about 10% better than OLS. Moreover, they ranged down to being appreciably inferior to OLS regression for equal weighting (as evidenced in 11, 13, 14, 18, 20, and 21) to drastically inferior for regression on principal components (6, 13, 14).

Ridge regression was appreciably superior to OLS regression in relative accuracy on four data sets (8, 16, 17 and 19) ranging from 11% up to 44% better than OLS regression. However, for all these four data sets, at least one (in two cases both) of the other non-OLS methods were considerably superior to ridge - - an outcome much like that provided by the results reported in a previous simulation study of relative accuracy (Morris, 1982).

In one data set (5), ridge did very poorly on relative prediction accuracy, as evidenced by yielding a negative cross-validated correlation (as principal components did in data set 16). Yet regressing the criterion variable on principal components or on an equally weighted composite performed much better than OLS. However, the importance of this particular result must be viewed in context; even though the squared multiple correlation was an appreciable .817, the cross validated OLS correlation was only .028 so that no meaningful prediction could take place on replicate samples in any case.

<u>Absolute Accuracy</u> (Table 2)

As for absolute accuracy (Table 2), the results were different. In 12 of the data sets, (1, 2, 3, 4, 6, 10, 12, 13, 14, 15, 20, and 21) ridge was within about two percent of the mean squared error produced by OLS regression. (It should be noted that <u>smaller</u> is superior for this measure of accuracy.) These data sets constituted a subset of the 16 meeting this same criterion for relative accuracy. On these same 12 data sets regressing the criterion variable on an equally weighted composite followed the results of ridge fairly closely; although superior (ranging from very slightly to appreciably) to ridge regression on three data sets (3, 10 and 12) regressing on an equally weighted composite was inferior on the rest. Regressing the criterion variable on principal components displayed much more variety within these 12 data sets. Performance was about the same as ridge on five of the data sets (2, 3, 15, 20, and 21), superior on three data sets (1, 10 and 12), and ranged to drastically inferior (4, 6, 13, and 14).

On eight of the data sets (5, 7, 8, 9, 16, 17, 18, and 19) ridge was appreciably better than OLS regression in absolute accuracy, with the decrease in mean squared error of prediction ranging from about 4% (data set 18) up to nearly 70% (data set 5). In four (5, 8, 16, and 19) of these eight data sets both regressing the criterion variable on principal components and on an equally weighted composite were in turn considerably better than ridge.

In only one data set (11) did ridge not perform at least about as well as OLS on absolute accuracy, with a mean squared error of about 21% more than that for OLS regression. Both principal components and equal weighting also performed very poorly on this data set. It is quite interesting and possibly important to note that this is not the same data set as the one on which ridge was so poor in relative accuracy; on that data set (5), ridge exhibited its <u>best</u> absolute accuracy performance (only 31% of the mean squared error of OLS regression).

Although the results from this data set may need to be considered especially cautiously because of the very small cross-validated correlation, the results also did not agree between relative and absolute accuracy in other instances. The decision of whether one is primarily interested in relative or absolute accuracy is an important one.

For these data sets, the number of subjects per variable, multicollinearity, and sample OLS multiple correlation all appeared to be of no use in helping the researcher decide whether one of the non-OLS methods would be worth pursuing. The question of identifying the most accurate prediction method is really one of classification. Can one "classify" a data set to the method yielding the greatest accuracy from sample characteristics? Using the "leave-one-out" strategy of Lachenbruch and Mickey (1968), these three sample characteristics were unable to classify the data sets into the most accurate strategy (OLS or non-OLS) any better than chance assignment would have for both relative and absolute accuracy. In fact, when combining the results for both relative and absolute accuracy, the number of correct classifications was exactly the same as one would expect by chance. For this reason, it would not seem possible to construct rules for deciding _a priori_ from these statistics arising from a specific sample which method would be likely to be most accurate on application to a replicate sample.

Discussion

Any summative comments that ~~could be made~~ related to the relative performance of the methods are necessarily only relevant to these data sets. Moreover, the purpose of this study was not to declare a best method, or even to derive rules based on sample characteristics for deciding which strategy to use. Indeed, the inability to explain easily the behavior of the weighting techniques from the sample characteristics presented argues for just such a sample specific approach as has been used and is being proffered.

One generalization that probably can be made from the results, however, is that none of the non-OLS methods offers a panacea for achieving maximum accuracy across _all_ data sets as some reports in the literature might suggest. The researcher stands to lose a lot of prediction accuracy by choosing _any_ of the non-OLS strategies under some data conditions. Likewise, the researcher stands to gain a great deal in some data conditions if a superior algorithm can be selected. The problem is that it is not easy to specify under what circumstances the realization of a superior algorithm will occur from simple sample data characteristics; thus, the more complicated PRESS procedure may be called for.

Although the data sets utilized in this paper may not be representative, it may still be reasonable to suggest that the performance of none of the non-OLS methods was good enough often enough to recommend routine application of them in the same way that OLS regression is used. At the same time, moreover, there are appreciable accuracy gains possible in _some_ cases. If prediction accuracy is sufficiently important for the data set and situation at hand, the researcher may wish to take the trouble to ferret out those occasions for which a more accurate non-OLS procedure can deliver greater accuracy; the PRESS algorithm is suggested as a viable strategy for that task.

The computation times for all the runs are included in Table 3. Most of the runs only took a few seconds, with several taking a few minutes. The two largest jobs in which the Project Talent data was analyzed separately by sex each took more than an hour to run. Whether the times are reasonable or not is clearly a subjective decision. However, even times of more than an hour don't compare unfavorably with the batch job turn-around time that can be expected when using many large computers.

The microcomputer used was a Sanyo MBC 550. This is an MS DOS machine with an 8088 microprocessor. It is similar in many ways to an IBM PC, but the 8088 clock rate is slower (3.6) than that of the IBM PC (4.77). An 8087 arithmetic coprocessor was also installed to aid in speed and accuracy. Because of the slower clock rate, almost all IBM PC "clones" would run these jobs faster than the times represented.

The computer language used was Turbo Pascal. While a good performer in general, it is certainly not the fastest "number crunching" language available. For example, a recent article in BYTE found the Microsoft Pascal compiler to run a computation intensive program utilizing the 8087 nearly twice as fast as Turbo Pascal. Microsoft Pascal, however, was unavailable to test. The Pascal program should run with no modification.

It should also be noted that newer, faster, and more powerful microprocessors are now commonplace. The 8086, 80186, and the 80286 of the IBM AT should all perform better than the times represented here. Therefore, for all these reasons, the times presented should be considered as quite conservative. Moreover, a 32 bit 80386 has recently been released and will be much faster (probably by a factor of more than four) than the fastest of these (the 80286). Super microcomputers with the power of a VAX mini should be on our desks very soon.

While microcomputer time is essentially free, a deficit in a long running job is that the machine is generally lost for other uses. However, there are now some good multitasking systems available that will allow the use of the computer for other purposes, i.e. word processing, while such a computation laden job is number-crunching in the "background." Such multitasking systems will almost certainly be a standard part of the operating system of the more powerful microcomputers that will be common in the very near future.

Although several strategies can be employed to make the computing algorithm as efficient as possible, a large amount of computation may result in any case. In general, in judging whether the PRESS technique is worth pursuing a researcher would need to consider the size of the prediction problem and resulting costs of PRESS in relation to the relative importance of the goal of maximizing prediction accuracy. It is important to note, however, that most prediction problems seen in the behavioral science literature are not excessively large and that in any case the non-OLS methods are really only contenders with relatively small samples. Further, the trend of the decreasing cost of computational power is accelerating; researchers need to plan their methods such that they can capitalize on this resource. Tukey's (1985) comments relating to our need to make sure that the statistical techniques we invent anticipate the incredible resources of computational power that we will have in the near future seem especially relevant.

A copy of the Pascal computer program is available for those wishing it. It is a COM file and should work on any MS DOS microcomputer with a microprocessor in the Intel 8088, 86, 286, etc. line. In requesting the program, please specify whether the program can expect to find an 8087 arithmetic processing unit available. If the program is of interest, send a blank DSDD diskette to:

John D. Morris
College of Education - IRDTE
Florida Atlantic University
Boca Raton, Florida 33431

References

Allen, D. M. (1971). *The prediction sum of squares as a criterion for selecting predictor variables* (Tech. Rep. No. 23). University of Kentucky, Department of Statistics.

Allen, D. M. (1972). *Biased prediction using multiple linear regression* (Tech. Rep. No. 36). University of Kentucky, Department of Statistics.

Allen, D. M., and Cady, F. B. (1982). *Analyzing experimental data by regression.* Belmont, CA: Wadsworth.

Bartlett, M. S. (1950). Tests of significance in factor analysis. *British Journal of Psychology, 3,* 77-85.

Bartlett, M. S. (1951). An inverse matrix adjustment arising in discriminant analysis. *Annals of Mathematical Statistics, 22,* 107-111.

Burkett, G. R. (1964). A study of reduced rank models for multiple prediction. *Psychometric Monographs,* No 12.

Cattin, P. (1981). The predictive power of ridge regression: Some quasi-simulation results. *Journal of Applied Psychology, 66,* 282-290.

Claudy, J. G. (1972). A comparison of five variable weighting procedures. *Educational and Psychological Measurement, 32,* 311-322.

Darlington, R. B. (1978). Reduced-variance regression. *Psychological Bulletin, 85,* 1238-1255.

Dawes, R. M., and Corrigan, B. (1974). Linear models in decision making. *Psychological Bulletin, 81,* 95-106.

Dempster, A. P., Schatzoff, M., and Wermuth, N. (1977). A simulation study of alternatives to ordinary least squares. *Journal of the American Statistical Association, 72,* 77-91.

Dorans, N., and Drasgow, F. (1978). Alternate weighting schemes for linear prediction. *Organizational Behavior and Human Performance, 21,* 316-345.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *Annals of Statistics, 7,* 1-26.

Efron, B. (1983). Estimating the error of a prediction rule: Improvement on cross-validation. *Journal of the American Statistical Association, 78,* 316-331.

Egerton, M. F., and Laycock, P. J. (1981). Some criticisms of stochastic shrinkage and ridge regression, with counterexamples. *Technometrics, 23,* 155-159.

Einhorn, H. J., and Hogarth, R. M. (1975). Unit weighting schemes for decision making. *Organizational Behavior and Human Performance, 13,* 171-192.

Gabriel, R. M. (1980, September). *Using composite variables in multivariate analysis: Why weight?* Paper presented at the meeting of the American Psychological Association, Montreal, Quebec, Canada.

Gibbons, D. G. (1981). A simulation study of some ridge estimators. *Journal of the American Statistical Association, 76,* 131-139.

Gollob, H. F. (1967). *Cross-validation using samples of size one.* Paper presented at the meeting of the American Psychological Association, Washington, D.C.

Hocking, R. R. (1983). Developments in linear regression methodology: 1959-1982. *Technometrics, 25,* 219-230.

Hoerl, A. E., and Kennard, R. W. (1970). Ridge regression: Applications to nonorthogonal problems. *Technometrics, 12,* 69-82.

Horst, P. (1963). *Matrix algebra for social scientists.* New York: Holt, Rinehart, and Winston.

Huberty, C. J. (1984). Issues in the use and interpretation of discriminant analys: Psychological Bulletin, 95, 156-171.

Huberty, C. J., and Mourad, S. A. (1980). Estimation in multiple correlation/prediction. Educational and Psychological Measurement, 40, 101-11:

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. Educational and Psychological Measurement, 20, 141-151.

Lachenbruch, P. A., and Mickey, M. R. (1968). Estimation of error rates in discriminant analysis. Technometrics, 10, 1-11.

Laughlin. J. E. (1978). Comment on "Estimating coefficients in linear models: It don't make no nevermind." Psychological Bulletin, 85, 247-253.

Lawless, J. F., and Wang, P. (1976). A simulation study of ridge and other regressi estimators. Communications in Statistics - Theory and Methods, A5(4), 307-323.

Lawshe, C. H., and Schucker, R. E. (1959). The relative efficiency of four test weighting methods in multiple prediction. Educational and Psychological Measurement, 19, 103-114.

Morris, J. D. (1979). A comparison of regression prediction accuracy on several typ of factor scores. American Educational Research Journal, 16, 17-24.

Morris, J. D. (1981, April). An extension to "A comparison of regression accuracy o several types of factor scores." Paper presented at the meeting of the America Educational Research Association, Los Angeles, CA.

Morris, J. D. (1982). Ridge regression and some alternate weighting techniques: A comment on Darlington. Psychological Bulletin, 91, 203-210.

Morris, J. D. (1983). Stepwise ridge regression: A computational clarification. Psychological Bulletin, 94, 363-366.

Morris, J. D. (1984). Cross-validation with Gollob's estimator: A computational simplification. Educational and Psychological Measurement, 44, 151-154.

Morris, J. D., and Guertin, W. H. (1977). The superiority of factor scores as predictors. Journal of Experimental Education, 45, 41-44.

Mosteller, F., and Tukey, J. W. (1968). Data analysis, including statistics. In G. Lindzey & E. Aronson (Eds.), Handbook of Social Psychology: Vol. 2. (pp. 80-203). Reading Mass.: Addison-Wesley.

Musgrage, B., Marquette, J., and Newman, I. (1982). On using the average intercorrelation among predictor variables and eigenvalue orientation to choos a regression solution. Multiple Linear Regression Viewpoints, 11, 1-21.

Pagel, M. D., and Lunneborg, C. E. (1985). Empirical evaluation of ridge regressior Psychological Bulletin, 97, 342-355.

Pruzek, R. M., and Frederick, B. C. (1978). Weighting predictors in linear models: Alternatives to least squares and limitations of equal weights. Psychological Bulletin, 85, 254-266.

Rozeboom, W. W. (1979). Ridge regression: Bonanza or beguilement? Psychological Bulletin, 86, 242-249.

Schmidt, F. L. (1971). The relative efficiency of regression and simple unit predictor weights in applied differential psychology. Educational and Psychological Measurement, 31, 699-714.

Smith, G., and Campbell, F. (1980). A critique of some ridge regression methods. Journal of the American Statistical Association, 75, 74-81.

Stone, M. (1974). Cross-validatory choice and assessment of statistical prediction: Journal of the Royal Statistical Society, Series B, 36, 111-133.

.isted, R. A., and Morris, C. N. (1980). Theoretical results for adaptive ordinary ridge regression estimators (Tech. Rep. No. 94). Chicago: University of Chicago.

attner, M. R. (1963). Comparison of three methods for assembling aptitude test batteries. Personnel Psychology, 16, 221-232.

.key, J. W. (1985). Comment. The American Statistician, 39, 12-14.

.iner, H. (1976). Estimating coefficients in linear models: It don't make no nevermind. Psychological Bulletin, 83, 213-217.

.iner, H. (1978). On the sensitivity of regression and regressors. Psychological Bulletin, 1978, 85, 267-273.

:sman, A. G., and Bennett, G. K. (1959). Multiple regression vs. simple addition of scores in prediction of college grades. Educational and Psychological Measurement, 19, 243-246.

Table 1

Weighting Methods' Relative Performance (Cross-Validated Correlation) for Several Data Sets

Numerical Designator and Data Set Description	RSQ	MI	OLS	Method: As a % of OLS: Ridge	PC	Equal
1 Marquardt's Acetylene Data	.920	1.0	.920	100.01	102.05	99.71
2 Chew LP(5) Predicts MRT	.591	1.0	.750	100.64	100.60	100.30
3 Hoerl's Kansas Corn Yield	.800	1.4	.854	100.24	100.23	100.32
4 Draper and Smith (p. 204)	.914	1.1	.927	100.03	92.67	99.52
5 Drehmer Data (EPM)	.817	1.1	.028	-192.43	379.01	131.28
6 Golf score from Task Perf	.848	1.6	.912	99.99	47.40	99.98
7 Hald Data (D & S, p. 366)	.982	1.0	.980	100.32	99.11	100.27
8 Hocking & Dunn RR Symp. '82	.620	1.0	.318	132.67	230.95	230.59
9 Hoerl RR-1980 Paper	.986	1.1	.979	100.21	100.17	100.17
10 Kerlinger and Pedhazur, 292	.640	1.6	.690	100.46	109.69	109.80
11 Longley D&W p. 312	.996	1.0	.992	99.83	95.61	92.57
12 Journal of Exp. Education	.475	1.1	.635	101.27	104.27	104.42
13 Rulon: Pref & Success - Mech	.261	1.9	.441	98.59	26.49	92.62
14 Rulon: Pref & Success - Oca	.323	2.4	.494	98.68	28.51	74.81
15 Rulon: Pref & Success - Pas	.252	1.5	.432	99.01	94.61	97.25
16 Retention from Demos & WISC	.388	1.2	.058	144.11	-40.94	520.58
17 Piers-Harris from IQ & Ach	.185	2.2	.108	111.96	61.88	142.59
18 D & S Steam Data (p.352)	.949	1.1	.925	99.06	89.44	86.92
19 D & S Data (p. 233)	.816	1.2	.691	111.05	121.17	118.50
20 Female Talent Data C&L p. 345	.331	1.9	.520	100.58	97.78	88.15
21 Male Talent Data C&L p. 349	.411	1.7	.577	101.17	100.54	94.97

Note. Additional information about data sources is available from the author; the abbreviated headings at the top of each data column are described in the text at the beginning of the section concerned with results.

e 2

<u>hting Methods' Absolute Performance (Mean Squared</u>
<u>r) for Several Data Sets</u>

merical Designator and ta Set Description	RSQ	MI	OLS	Method: As a % of OLS: Ridge	PC	Equal
Marquardt's Acetylene Data	.920	1.1	21.0	99.02	75.60	101.32
Chew LP(5) Predicts MRT	.591	1.0	63.9	98.25	98.25	99.10
Hoerl's Kansas Corn Yield	.800	2.1	14.2	98.27	98.37	97.78
Draper and Smith (p. 204)	.914	1.4	13.9	99.16	186.42	106.79
Drehmer Data (EPM)	.817	1.5	1.12	31.20	21.20	26.51
Golf score from Task Perf	.848	2.3	1.98	100.06	485.98	100.21
Hald Data (D&S, p. 366)	.982	1.0	8.49	83.49	141.20	106.61
Hocking & Dunn RR Symp. '82	.620	1.1	53.7	72.57	31.25	31.38
Hoerl RR-1980 Paper	.986	1.4	2.98	90.31	92.58	92.62
1 Kerlinger and Pedhazur, p.292	.640	2.2	.19	97.63	79.39	79.29
Longley D&W p. 312	.996	1.2	.18E+6	121.39	641.15	1066.6
Journal of Exp. Education	.475	1.4	9.89	97.89	93.62	93.41
3 Rulon: Pref & Success - Mech	.261	2.4	2.42	100.09	123.62	103.65
4 Rulon: Pref & Success - Oca	.323	2.7	2.65	100.21	130.28	117.00
5 Rulon: Pref & Success - Pas	.252	2.0	309.7	99.89	102.17	100.86
6 Retention from Demos & WISC	.388	1.7	.18E+5	84.15	81.97	58.07
17 Piers-Harris from IQ & Ach	.185	3.4	209.9	92.12	93.89	93.33
18 D & S Steam Data (p.352)	.949	1.5	.43	94.83	190.84	208.86
19 D & S Data (p. 233)	.816	1.7	.007	68.58	48.35	53.23
20 Female Talent Data C&L p. 345	.331	3.7	2.10	99.16	101.20	107.77
21 Male Talent Data C&L p. 349	.411	3.2	1.63	98.31	98.79	104.17

<u>Note</u>. The information presented in the Note of Table 1 is appropriate for this table.

Table 3

Score Matrix Size and Computation Times for Several Data Sets

Numerical Designator and Data Set Description	n	p	Time(M:S)
1 Marquardt's Acetylene Data	16	3	:06
2 Chew LP(5) Predicts MRT	293	5	6:51
3 Hoerl's Kansas Corn Yield	51	6	2:02
4 Draper and Smith (p. 204)	21	3	:08
5 Drehmer Data (EPM)	14	9	2:03
6 Golf score from Task Perf	120	4	1:51
7 Hald Data (D&S, p. 366)	13	4	:11
8 Hocking & Dunn RR Symp. '82	20	3	:07
9 Hoerl RR-1980 Paper	15	5	:21
10 Kerlinger and Pedhazur, p.292	30	4	:23
11 Longley D&W p. 312	16	6	:36
12 Journal of Exp. Education	83	4	:58
13 Rulon: Pref & Success - Mech	93	3	:38
14 Rulon: Pref & Success - Oca	66	3	:24
15 Rulon: Pref & Success - Pas	86	3	:33
16 Retention from Demos & WISC	29	10	1:30
17 Piers-Harris from IQ & Ach	55	7	3:40
18 D & S Steam Data (p.352)	25	9	3:29
19 D & S Data (p. 233)	16	4	:13
20 Female Talent Data C&L p. 345	271	12	93:39
21 Male Talent Data C&L p. 349	234	12	80:09

Note. The information presented in the Note of Table 1 is appropriate for this table.