The Rise and Fall and Rise of Multiple Regression

Tianqi Han
Northern Illinois University

Dennis Leitner
Southern Illinois University

The rise and fall and rise of multiple regression is chronicled in the literature by examining its initial impetus and popularity, followed by the acknowledgement of potential problematic issues such as violation of assumptions and overzealous usage, and the subsequent resurgence of the technique as the problems are addressed and procedures clarified. Jacob Cohen brought to the attention of many researchers that multiple regression can be used as a general data-analytic system. With the increasing availability of mainframe computers and programs to perform statistical analysis, journal editors were inundated with an avalanche of regression analyses. The assumptions underlying the analyses were emphasized, considered, and often found to be unmet. Two major problems of using stepwise regression were identified: (1) incorrect degrees of freedom were specified when evaluating changes in explained variance and, (2) incorrect interpretation of stepwise resultswhen a few variables are selected from many. Subsequently, many different regression models have been developed for different situations, especially when assumptions are violated. These models include ridge regression, robust regression, and nonlinear regression.

Like the length of skirts or the cuffs on pant legs, the popularity of statistical tests rises and falls and rises. In his Presidential Address to the Mid-Western Educational Research Association, Leitner (1990) traced this rise and fall and rise of three statistical tests in the literature. First, the initial presentation and use was followed by a second period of the acknowledgement of potential problematic issues such as violation of assumptions and overzealous usage, which resulted in a third period characterized by a resurgence of the technique as the problems are addressed and procedures are clarified. The three statistical techniques he examined were the t-test/analysis of variance, factor analysis and meta-analysis. This same approach is used in this paper to examine the rise and fall and rise of multiple regression.

While this review of literature is largely chronological, it is not strictly so. Some of the statistical aspects are reported in the mathematical and statistical literature long before they appear in the psychological and educational literature. It is the latter which forms the principal basis of the chronology.

The Initial Rise

In one of the first references to multiple regression in the social science literature, Goldberger (1964), having recognized the nature of multiple regression, pointed out that

> ...[T]he whole point of multiple regression as contrasted with simple regression is to try to isolate the effects of the individual regressors, by 'controlling' on the others. Still, when orthogonality is absent the concept of the contribution of an individual regressor remains inherently ambiguous. (p. 201)

A large impetus for the use of multiple regression came from the work in the late 1960's of the distinguished statistician, Jacob Cohen. Cohen (1968) pointed out that multiple regression and analysis of variance and covariance are special cases of the general linear model.

> If you should say to a mathematical statistician that you have discovered that linear multiple regression analysis and the analysis of variance (and covariance) are identical systems, he would mutter something like, 'Of course--general linear model,' and you might have trouble maintaining his attention. If you should say this to a typical psychologist, you would be met with incredulity, or worse. Yet it is true, and in its truth lie possibilities for more relevant and therefore more powerful exploitation of research data. (Cohen, 1968, p. 426)

He showed that through use of indicator variables (i.e., dummy variable coding), an equivalence between multiple regression and analysis of variance, in fact, exists. In addition, through use of contrast coding, powers and products of variables, and comparisons of appropriate regression equations, multiple regression can be used as a general data-analytic system.

At about the same time, Richard Darlington (1968) emphasized that, besides providing the partial correlation between the dependent variable and each of the independent variables, regression weights rather than correlation coefficients have the interpretative advantage in prediction allowing statements like "Increasing X_j by 1 unit increases the dependent variable by β_j units" (p. 167). He discussed the logical fallacies involved in using variance-apportionment techniques for any purpose when the independent variables in a set are intercorrelated. He pointed out that the notion of "independent contribution to variance" is meaningless especially when multicollinearity is a problem (p.169).

In perhaps the first text devoted exclusively to the use of multiple regression, Kelly, Beggs, McNeil, Eichelberger, and Lyon (1969) took advantage of the growing presence of high-speed digital computers by freeing the researcher from simplistic designs that can be handled computationally with ease on a desk calculator. By forcing researchers to use "...a series of factorial designs, Type I, Type II models, etc., derived to ease computation with a desk calculator,"...the researchers were either "confused" or had to "impose such constraints on his design that he is forced to ask a limited research question." (Kelly et al., 1969, p. vii).

In addition, Kelly et al. (1969) emphasized that the availability of multiple regression procedures and programs allowed the researcher to ask meaningful research questions.

> The multiple regression analysis presented in this book is designed to prepare the research investigator to construct statistical models which will reflect his original research question rather than limiting that question. Regression analysis will be shown to be the generalized case of analysis of variance. These discussions shall be intimately related to a computer program so that the simple elegance of the generalized analysis of variance is not obscured and so that the investigator can circumvent the anachronistic desk calculator. (p. vii)

Four years later, another popular text of multiple regression was written by Kerlinger and Pedhazur (1973). The book, which listed a different computer program in the appendix than did the Kelly et al. (1969) text, promoted the advantages of multiple regression analysis.

> Multiple regression analysis [is] a most important branch of multivariate analysis... It is a powerful analytic tool widely applicable to many different kinds of research problems. It can be used effectively in sociological, psychological, economic, political, and educational research. It can be used equally well in experimental or nonexperimental research. It can handle continuous and categorical variables. It can handle two, three, four, or more dependent variables. In principle, the analysis is the same. Finally, multiple regression analysis can do anything the analysis of variance does ... (Kerlinger & Pedhazur, 1973, p. 2-3). [In addition], multiple regression analysis not only gives more information about the data, it also applicable to more kinds of data. (p. 6)

Multiple regression not only provides a way to analyze the relations of one variable with a set of variables, but it, with the stepwise method, also can be used for purposes of parsimony. Efroymson (1960) first advanced stepwise regression in an article in which he presented an algorithm which performed a true stepwise (as distinguished from FORWARD or BACKWARD methods) regression.

> An important property of the stepwise procedure is based on the facts that (a) a variable may be indicated to be significant in any early stage and thus enter the equation, and (b) after several other variables are added to the regression equation, the initial variable may be indicated to be insignificant. The insignificant variable will be removed from the regression equation before adding an additional variable. Therefore, only significant variables are included in the final regression. (p. 192)

Efroymson's (1960) article presented computer output from an example, as well as estimates of how much space and time would be needed to run problems based on the number of variables and sample size. Stepwise regression has received considerable attention in reducing the number of independent variables in the prediction equation or selecting the best subset of the variables from a set of independent variables.

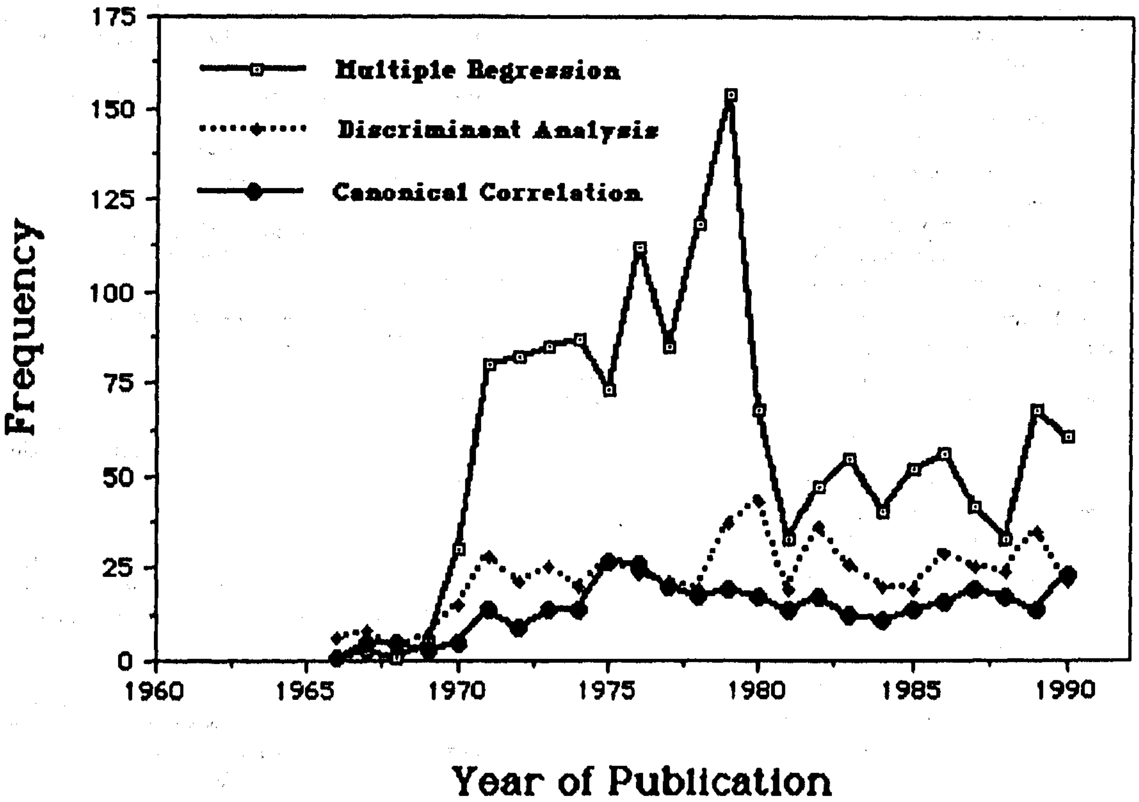
Following Cohen and Darlington's work, the 1970's saw a great increase in research on the theory as well as application of multiple regression. For example, see Heise (1969, 1970) who used multiple regression in causal relation research using social science and panel data.

As you will see in the next section, the middle of the 1970s saw the peak in the number of applications of multiple regression. Questions about assumptions being met and appropriate uses come to the forefront of researchers' use of the statistical methodology.

The Subsequent Fall

With the increasing availability of mainframe computers and programs to perform statistical analysis, journal editors were inundated with an avalanche of regression analyses. Figure 1 demonstrates the growth of multiple regression, discriminant analysis, and canonical correlation from article references by the Educational Resources Information Center (ERIC). The ERIC database consists of the Resources in Education (RIE) file of document citations and the Current Index to Journals in Education (CIJE) file of journal article citations from over 750 professional journals.
Questions were raised about whether assumptions were being met and the use of stepwise regression was strongly criticized. Attention was given to whether the regression models were correctly specified. Confusion between multiple correlation and prediction estimation began to be identified.

Figure 1 Number of Citations of Multiple Regression, Discriminant Analysis, and Canonical Correlation in ERIC Journals from 1965 to 1991



Violation of the Assumptions

During 1970s there were many criticisms related to the assumption of normal distribution of errors of measurement which is, in many cases, not likely to be true with variables in behavioral research.

The classical linear model $Y = \beta x + \varepsilon$ assumes that y, an N x 1 vector, is a random variable, X is an N x (k+1) matrix with fixed (not random) values', (i.e., X is matrix of known constants); β, a (k+1) x 1 vector, contains the k unknown parameters, or regression weights, plus an intercept parameter; and ε, an N x 1 vector, is a random variable. It further assumes that the errors have the properties of normality, linearity, independence, and homoscedasticity. This expression of the classical model is from Sockloff (1976), pp. 268-9.

It seemed that multiple regression does not have any requirement for the data except meeting those assumptions described above. Box (1966) alerted the mathematical community to a possible concern in treating data collected from "field research" (without controls on variables or manipulation of independent variables) in the same manner as data from "lab experiments" (with random assignment of subjects to groups).

The method of least squares is used in the analysis of data from planned experiments and also in the analysis of data from unplanned happenings. ... It is the tacit assumption that the requirements for the validity of least squares analysis are satisfied for unplanned data that produces a great deal of trouble. Whether the data are planned or unplanned the quantity ε, which is usually quickly dismissed as a random variable having the very specific properties mentioned above, really describes the effect of a large number of 'latent' variables $x_{k+1}, x_{k+2}, \ldots, x_m$, which we know nothing about. (Box, 1966, p. 625)

For the unplanned data, suppose k independent variables are input in the model, ε includes a combination of some latent variables, say, x_{k+1}, $x_{k+2}, \ldots x_m$. Therefore, the regression model contained two components:

$$Y = [\beta_0 + \beta_1 x_{1u} + \beta_2 x_{2u} + \ldots \beta_k x_{ku}] + [\beta_{k+1} x_{k+1} + \ldots \beta_m x_m]$$

$$Y = x_1 \beta_1 x_2 \beta_2$$

As an example of analysis of unplanned data, Box (1966) discussed a possible situation in industry.

> In the operation of an industrial process past experience often shows that certain variables are of major importance. In order to control fluctuations in the process, therefore, care is taken to hold precisely these variables very close to fixed values. As the "statistical significance" of any variable is greatly affected by the range it covers, there is a strong probability, therefore, that the most important variables will be dubbed "not significant" by a standard regression analysis. A further difficulty is that with unplanned data regression variables will frequently be highly correlated only because of operating policy. (p. 628)

Although presented here as a violation of the assumption of the errors being normally and identically distributed, the problem identified by Box (1966) may also be considered as misspecification of the regression model and multicollinearity resulting from unplanned data.

Some people questioned the robustness of least square estimation when the assumption(s) was(were) not met. Wainer and Thissen (1976) concluded:

> In this paper we have explored a variety of schemes for estimating coefficients of linear functions with respect to their ability to yield reasonable answers when the form of the data distribution ranges broadly. Our strongest finding is that the most commonly applied methodology, least squares estimators (LSE), are the worst performers in general. (pp. 30-31)

Earlier in this article, Wainer and Thissen discussed the assumptions in multiple regression and using equal weights (β s).

> The robustness of equal weights is beyond question, since their estimation does not involve the data at all; the shape of the sample distribution is irrelevant. Least squares estimates are another story. They are used without distributional assumptions and are identical to maximum likelihood estimates with Gaussian assumptions, provided that one assumes independence of error. If this assumption is violated the least squares estimates overestimate the betas. This is only one thing that can go wrong and is indicative of the "capitalization on chance" that has become the hallmark of least squares regression. (p. 12)

In another article advocating the use of robust regression methods, Wainer (1976) wrote:

> It is noted that the usual estimates that are optimal under a Gaussian assumption are very vulnerable to the effects of outliers. ... Normality assumptions are very useful theoretically, but have sometimes proved unrealistic in practice. (p. 285)

> In a 1976 article, Sockloff noted that the assumptions under which analyses are conducted are not always specified.

> Recent works by Cohen (1968), Kelly, Beggs, McNeil, Eichelberger, and Lyon (1969), Kerlinger and Pedhazur (1973), and Bottemberg and Ward ... have attested to the flexibility of the General Linear Model. These publications have shown the capabilities of a single approach to the solution of correlation, regression, and the Fisherian analysis of variance problems. It is noteworthy that all six of these publications claim, more or less, to be using the General Linear Model, but in no case has the particular linear model and its assumptions been clearly specified and consistently applied.
>
> The General Linear Model is a name given to the family of models possessing a common characteristic, namely, linearity in the parameters of the equation specifying the model. The members of this family are distinguishable in terms of their various assumptions, and it is the contention of this author that the distinctions among these different linear models are of more than just passing interest.
>
> The above publications, plus those of Digman (1966) and of McNeil and Spaner (1971), have shown the capabilities of the General Linear Model in handling the analysis of nonlinear data....[T]he interest of this paper is to show that the analysis of nonlinearity via polynomial and product variables in a linear model has limitations far more stringent than have been realized by educational and psychological researchers. (pp. 267-268)

Sockloff (1976) distinguished between three linear models (fixed, random, and provisional) and emphasized the differences between a fixed model and a random model and the limitation of general linear model in handling nonlinear data. In the "fixed" model, the matrix X consists of "regressors that are observable and are fixed (determined a priori) values of random variables" (p. 269). In the random model, X is a matrix of regressors that are observable and random variables.

> The Random Normal Model requires the additional assumptions: (a) in the population, X and y are distributed multivariate normal, and X and ε are uncorrelated; and (b) in the sample, each multivariate observation corresponding to a row of X and y is randomly drawn. If X and y are distributed multivariate normal, the $\varepsilon = y - X\beta$ is independently distributed multivariate normal with common variance σ^2 as in the Fixed Normal Model, and X and ε are not only uncorrelated but also independent. The population to which inferences are made under the Random Normal

Model covers the total multivariate population from which the validation sample is randomly drawn." (pp. 269-270).

Kelly et al. (1969), Kerlinger and Pedhazur (1973), and Bottenberg and Ward devote most of their respective texts to multiple regression and capitalize on the similarity of computational procedures required for the solution of analysis of variance, multiple correlation, and polynomial regression problems. Whereas Bottenberg and Ward fail to specify models or assumptions, Kelly et al. and Kerlinger and Pedhazur work under an apparent Fixed Normal Model insofar as distributional assumptions are not made about the regressors. Although Kerlinger and Pedhazur never distinguish the two classical models, Kelly et al. make a distinction, but this distinction is made late in the book at which point the reader cannot easily determine the appropriate model for each of the problems presented earlier. (p. 272)

He pointed out that the computational similarity between the fixed and random models was the initial source of the confusion of the two models. He argued that "regarding the analysis of nonlinearity in observational data under the Random Model, the Random Normal Model cannot be used, and contrary to the various publications extolling the generality of the General Linear Model, the appropriate counterpart inferential model does not currently exist." (p. 288)

Multicollinearity

Statistical analysts using multiple regression have known for some time about the problems caused by intercorrelations among the independent variables. High intercorrelations among the predictors, but not complete linear dependency, has been called "collinearity" or "ill conditioning" of the correlation matrix, or for the purposes of this paper, "multicollinearity". Gordon (1969) alerted us to the potential problems:

> Although the warnings concerning multicollinearity are to be found in statistics texts, they are insufficiently informative to prevent the mistakes described here. This is because the problem is essentially one of substantive interpretation rather than one of mathematical statistics per se. (p. 592)

The effects of multicollinearity on the least squares estimates of the regression coefficients were pointed by Johnstone in 1972 as follows:

> 1. The precision of estimation falls so that it becomes very difficult, if not impossible, to disentangle the relative influence of various x variables. This loss of precision has three aspects; Specific estimates may have very large errors; these error may be highly correlated, one with another; and the sampling variances of the coefficients will be very large.
>
> 2. Investigators are sometimes led to drop variables incorrectly from an analysis because their coefficients are not significantly different from zero, but the true situation may be not that a variable has no effect but simply that the set of sample data has not enabled us to pick it up.
>
> 3. Estimates of coefficients become very sensitive to particular sets of sample data, and the addition of a few more observations can sometimes produce dramatic shifts in some of the coefficients. (p. 160)

Gordon (1969) concluded:

> ...[W]e have not been condemning the method of multiple regression in general. There remain many situations in sociology for which regression is an excellent tool of analysis. We do condemn, however, those applications of regression coefficients that seek to determine the relative importance of variables in the manner of the examples we have cited. (pp. 615-6)

Abuse of Stepwise Regression

One of the most common uses of regression has been model-building automatically, that is, determining the relative importance of variables by the order in which they are entered (or deleted) to find the "best" regression model. Pope and Webster (1972) pointed out that:

> The methods generally known as stepwise procedures are, however, the most widely used data analysis methods; in particular by non-professional statisticians. This has come about through the availability of computer programs.
>
> This paper was stimulated by this widespread use of the stepwise procedures and the lack of understanding (by the non-statistician) of their weaknesses. (p. 328)

Huberty (1989) listed three intended uses of stepwise regression.

> Stepwise analyses have basically been used for three purposes: (1) selection or deletion of variables, (2) assessing relative variable importance; or (3) both variable selection and variable ordering. (p. 45)

Stepwise regression has been commonly used for selecting the best subset for any specified number of retained independent variables. Among a total of k (k+1) / 2 fits, "as observed by Gorman and Toman (1966), it is unlikely that there is a single best subset but rather several equally good ones" (Hocking, 1960, p. 9). Mantel (1970) criticized forward selection by illustrating a situation in which an excellent model would be overlooked because of the restriction of adding only one variable at a time and pointed out the disadvantage of forward selection needs k (k+1) / 2 fits k where backward elimination only needs k fits for testing among k variables. Hocking (1960) also expressed concern about the limited number of solutions for the "best" regression equation.

> Another criticism of FS [forward selection] and BE [backward elimination] often cited is that they imply an order of importance to the variables. This can be misleading since, for example, it is not uncommon to find that the first variable included in FS is quite unnecessary in the presence of other variables.... The lack of satisfaction of any reasonable optimality criterion by the subsets revealed by stepwise methods, although a valid criticism, may not be as serious a deficiency as the fact that typical computer routines usually reveal only one subset of a given size. (p. 9)

Pope and Webster (1972) pointed out the "pseudoness of the F-statistic" for testing the significance of independent variables in linear prediction equation" (p. 327). "Unfortunately, the most widely used computer programs print this statistic at each step without any warning that it does not have the F distribution under automated stepwise selection" (Wilkinson, 1979, p. 168). Using a Monte Carlo simulation, Wilkinson (1979) constructed the tables of the upper 95th and 99th percentage points of the sample R^2 distribution in forward selection. He examined 71 articles published in psychology from 1969 to 1977 which used stepwise regression.

> Out of these articles 66 forward selection analyses reported as significant by the usual F test were found. Of these 66 analyses, 19 were not significant [using Wilkinson table].. . (p. 172)

The severe consequences of abuse of stepwise regression were emphasized by Thompson in a 1989 editorial entitled, "Why Won't Stepwise Methods Die?"

> First, most researchers, thanks to "canned" computer programs, do not employ the correct degrees of freedom when evaluating changes in explained variance (i.e., usually changes in squared R or lambda). ... Second, some researchers incorrectly interpret stepwise results in which *q* predictor variables have been selected as indicating that the predictor variables are the best variables to use if the predictor variable set is limited to size *q*. ... Third, some researchers incorrectly consult order of entry information to evaluate the importance of various predictor variables." (pp. 146-147)

In one of the most serious and thorough critiques of stepwise regression, Huberty (1989) postulated that:

> (1) ..stepwise analysis should not generally be used for variable selection purposes. A basic detect of stepwise procedures is attributable to 'their consideration of variables one-at-time... direct tests for the additional information supplied jointly by several variables are not made' (McKay & Campbell, 1982, pp. 13, 45)
>
> (2)...order of variable entry in a stepwise analysis should not be used to assess relative variable contribution/importance." because "the inter-relationship of the response variables are completely ignored when the most 'important' [first variable entered] is determined... and the dependence [of following variable on preceding variable] or conditionality truly makes variable importance as determined by stepwise analysis very question". (pp. 46-47)

Kachigan (1986), warned researchers that sampling error can seriously distort stepwise results.

> There is a danger that we might selected variables for inclusion in the regression equation based on chance relationship. Therefore, as stressed in our discussion of multiple correlation, we should apply our chosen regression equation to a fresh sample of objects to see how well it does in fact predict values on the criterion variable. This validation procedure is absolutely essential if we are to have any faith at all in the future applications of the regression equation. (p. 265)

We will see in the Second Rise section that Huberty proposed alternative methods to address these problems.

Misspecification of Regression Models

Included in our definition of misspecification of regression models are specification errors by using the "wrong" independent variables as well as expressing the wrong relationship among the independent variables or the relationship between the independent variables with the dependent variable. This first type was identified in 1971 by Borhnstedt and Carter.

> When one has mistakenly either omitted or included variables in an equation assumed to capture the true causal structure to Y, or when the functional form chosen to represent the variables is incorrect, we say that one has made a specification error. (p. 128)

The second type would include following: (a) specifying a linear model though a nonlinear model is more appropriate, (b) postulating an additive model even though a nonadditive model is more appropriate, and (c) applying a linear additive model when a nonlinear or nonadditive one is called for (Pedhazur, 1982, pp. 225-229).

When any of the assumptions are violated or when the stepwise regression technique is not correctly used, misspecification of the regression model is an inevitable outcome. However, researchers often ignore such errors.

Gordon (1969) contended that the theoretical context of research should determine the nature of importance of the variables controlled. Since R^2 was the most often used criterion to judging prediction models and (partial) regression coefficients were often used as indicators of the relative importance of variables, Gordon (1969) showed the interrelationship of the multicollinearity and misspecification problems. He provided the examples showing that:

> [S]mall variation among the correlations of a highly related set can be create large variations among their regression coefficients" (p. 612). In addition "the values of regression coefficients are not immutable and that they can be greatly affected by changes in the selection of independent variables to be included in an analysis" (p. 613). He warned us that "multiple regression is not an all-purpose methods for data reduction" (p. 163) and emphasized going "beyond simple examination of the regression coefficients". (p. 615)

Bohrnstedt and Carter (1971) discussed the effect of specification errors:

> specification errors can seriously affect our estimates of the true structural parameters operating in the system. ... if we hypothesize the wrong model, then our estimation of that model will yield meaningless estimates. (p. 141)

They concluded that "we can only come to the sobering conclusion, then, that many of the published results based on regression analysis... are possible distortions of whatever reality may exist" (p. 143).

Confusion Between Multiple Correlation and Prediction Estimation

The prediction model and the correlation model were seldom to be distinguished. Huberty and Mourad (1980) emphasized the difference of the parameters estimated in the multiple correlation and prediction estimation.

> All of the statistical techniques associated with the prediction model are applicable with the correlation model. However, from a correlation estimation viewpoint, different parameters are associated with the two models. With the correlation model, the population multiple correlation coefficient of interest is ρ, which reflects the correlation between Y and the optimal linear composite of $X_1, X_2, ... X_p$ in the population as a whole. The optimal linear composite is that composite determined so as to maximize this correlation in the population. With the prediction model, the population multiple correlation coefficient of interest is ρ_v which reflects the correlation between Y and the linear composite of the X's which is optimal for the calibration sample. With each calibration sample is associated a ρ_v, which is a type of validity coefficient. Values of ρ_v are coefficients of correlation between a criterion Y and a linear composite of the predictors, the weights of which will vary across repeated sampling. (p. 102)

They also criticized the deficiencies in reporting estimates of correlation coefficient in the literature and the inflated predictive validity of the studies, overestimation of the parameter ρ for prediction using R_w and R_e. They discuss two estimation procedures for the parameters ρ and ρ_v: cross-validation and usage of a "shrinkage" formula.

The Second Rise

In this period which these authors call the 'second rise', comparatively new techniques are recognized for handling the problems identified during the period of "the fall." Some of those techniques are robust regression, ridge regression and nonlinear regression. These methods were introduced to behavioral scientists in the late 70's and early 80's. Also, new methods using multiple and/or categorical dependent variables, such as canonical correlation and discriminant analysis, have been popularized.

Nonlinear Regression

When the assumption of linearity is violated, an appropriate nonlinear regression model should be considered. Since regression weights in nonlinear regression equations can be changed by changing the means of the independent variables, and the means are often chosen arbitrarily, the coefficients of nonlinear regression models can not be interpreted causally. A general solution to the importance of each independent variable in the linear and nonlinear models was attempted by Darlington and Rom (1972). For the sake of the difficulty of the interpretation of the nonlinear regression model, the effects on the transformation of polynomial regression equations into a format that is readily interpretable were made.

Robust Regression

In 1976, Howard Wainer wrote an article published in *Psychological Bulletin* entitled "Estimating Coefficients in Linear Models: It Don't Make No Nevermind." In his article, he stated:

> It is proved that under very general circumstances coefficients in multiple regression models can be replaced with equal weights with almost no loss in accuracy on the original data sample. It is then shown that these equal weighs will have greater robustness than least squares regression coefficients. (p 213)

The general conditions given are "all predictor variables should be oriented properly" and "the predictor variables should be intercorrelated positively" (Wainer, 1976, p 213).

Wainer's approach essentially ignores the sample data. A less radical solution to the problems with ordinary least squares solutions (OLS) to the estimate of parameters in multiple regression in light of non-normality or outlier problems has been addressed by Huyhn (1982), who referenced the sources of the alternatives for handling outliers and explained the concept and functions of Least Absolute Residual (LAR), first introduced by Gentle (1977):

> LAR estimates are the maximum-likelihood

> estimates when the errors follow a double exponential structure. Because large residuals are given smaller weights in LAR estimation than in OLS [ordinary least squares] estimation, LAR estimates are less influenced than OLS estimates by those residuals. (Huyhn, 1982, p. 506)

Huyhn reviewed each of the four robust regression techniques provided by Huber (M-estimate), Hample (psi function), Andrew (sine estimate) and Tukey (biweight estimate), respectively, provided an example of using these four robustness regression methods, and compared them with the results from employing the ordinary least square method. The reader should refer to Hogg (1979) for a discussion of the last four estimators. Huyhn (1982) summarized the conclusions about robust regression against OLS.

> First, if the data do not contain any outlying observations, then OLS and robust regressions provide estimates that do not differ markedly from each other. Second, for data with suspected or abnormal observations, OLS estimates may differ substantially from the robust estimates; third, observations considered as outliers by OLS regression may not be outliers at all under robust regressions. Fourth, robust regression procedures, as proposed by Hampel, Andrews, or Tukey, may be able to detect outliers automatically by giving each one a weight that is zero or very small as compared with other weights. (p. 511)

He re-emphasized the recommendations provided by Hogg (1979).

> Perform the usual OLS analysis along with a robust procedure such as that used by Andrews. If the resulting estimates are in essential agreement, report the OLS estimates and relevant statistics. If substantial differences occur, however, take a careful look at the observations with large robust residuals and check to determine whether they contain errors of any or if they represent significant situations under which the postulated regression model is not appropriate. (pp. 511-512)

Ridge Regression

Knowledge of the potential problems caused by multicollinearity has alerted researchers to avoid misinterpretations. Many alternatives have been proposed. A researcher might first try to eliminate the variables that contribute to the high degree of multicollinearity. However, we should not have considered a logically redundant variable initially. Removal of any one variable may lead to misspecification of the model. Pedhazur (1982) noted other remedies:

> One of the proposed remedies is the collection of additional data in the hope that this may ameliorate the condition of high multicollinearity. Another set of remedies relates to the grouping of variables either in blocks on the basis of a priori judgements or by the use of such methods as principal components analysis and factor analysis.... Another set of proposals...is to abandon Ordinary Least-Squares analysis and use instead other methods of estimation. One such method that has been gaining in popularity is Ridge Regression.... [N]one of the proposed methods of dealing with high multicollinearity constitutes a cure. High multicollinearity is symptomatic of insufficient, or deficient, information, which no amount of data manipulation can rectify. (p. 247)

Reduced variance regression, as a compromise between ordinary regression and some other techniques such as weighted least squares, was advocated for its potential solution of dealing with problems of multicollinearity, ratio of number of predictors to sample size, as well as validity issues. Ridge regression, introduced by Hoerl and Kennard in 1970, is an application of reduced variance regression. "Ridge regression is a controversial procedure that attempts to stabilize estimates of regression coefficients by inflating the variance that is analyzed" (Tabachnick & Fidell, 1989, p. 130).

In late 70's and early 80's, ridge regression was reemphasized in the psychology and social sciences. For example, Price (1977) and Darlington and Boyce (1982) highlighted the function of ridge regression in exploring and extracting information from multifactor data. Price (1977) gave an example of how to use ridge regression, introduced the criterion of choosing a value of k (see below) from inspection of the ridge trace, and emphasized the nature of ridge regression in reducing total mean square error by introducing some degree of bias.

Darlington and Boyce (1982) also provided the behavioral scientist with a very comprehensible explanation about ridge regression using the concept of regression to the mean.

> It is well known that estimates for many independent parameter values can be improved by regressing the unbiased estimates of those values toward the grand mean of all the values. ... If the investigator assumes that on the average, each observed correlation exceeds the true value by a proportion k, then the ratio between average observed and true values is (1+k) / 1. ... Ridge regression essentially consists of adjusting all the correlations in the matrix (both the X - X and the X - Y correlations) by this factor 1/(1+k), and then deriving regression weights in the ordinary way. ... Thus adjustment of the X - X correlations produces the largest increases in apparent independence (and hence increases in beta weights) for those regressors which correlate most highly with the other regressors. This is how ridge regression takes advantage of validity concentration -- regressors correlating

highly with the total set of regressors are upgraded in importance relative to the others. (pp. 84 - 85)

They informed researchers that about a dozen formulae for estimating k have been proposed and the ridge trace was no longer recommended by the statisticians. The alternative for estimating k, an iteration procedure was introduced in this paper. They also provided recommendations about when ridge regression should be used.

Alternatives to Stepwise Regression

Concerning the possible distorted results from careless use of stepwise regression, many researchers tried to find better alternatives to stepwise regression. Huberty (1989) provided the alternative approaches and suggested that "a 'natural' criterion to use to determine the best subset size in the context of prediction and estimation is to minimize the residual sum-of-squares value" (p. 50). For selecting the variables from a set of initial variables, SAS PROC RSQUARE (SAS Institute, Inc., 1990) procedure was recommended to assess $2^{p}-1$ equations, where p is the number of predictors (Huberty, 1989, p. 50). For determining the final subset size of the independent variables, Huberty (1989) recommended adjusted R^2 or scree test --- "plot[ing] the adjusted R^2 values for the 'best' subset of each size (determined by the researcher using information from computer output plus sound judgment) against subset size" (p. 51).

Thompson (1989) proposed that a possible alternative to the misleading results of stepwise regression would be to "employ a cross-validation procedure such as one recommended by Huck, Cormier, and Bounds (1974, p. 159)". Huck, Cormier and Bounds (1974) proposed a four-step method.

> (1) The original group of people (for whom both predictor and criterion scores are available) is randomly divided into two subgroups. (2) Just one of the subgroups is used to develop the prediction equation.. (3) The equation is used to predict a criterion score for each person in the second subgroup, i.e., the subgroup that was not used to develop the prediction equation). (4) The predicted criterion scores for people in the second subgroup are correlated with their actual criterion scores. A high correlation (that is significantly different from zero) means that the prediction equation works for people other than those who were used to develop the equation. If the individuals in future studies are not too much different from those in the cross-validation procedure, the researcher is justified in using the prediction equation for groups other than the original. (pp. 159-160)

Henderson and Velleman (1981) illustrated the superiority of substantively guided data analysis over automatic model building. "Automated multiple regression model-building techniques often hide important aspects of data from the data analyst. Such feature as nonlinearity, collinearity, outliers, and points with high leverage can profoundly affect automated analyses, yet remain undetected." Henderson and Velleman (1981) proposed an alternate method integrating "interactive computing and exploratory methods to discover unexpected features of the data." (p 391). They illustrated their alternative method using two examples, one from Hocking (1973) involving variables on 32 automobiles and a second example on air pollution and mortality from McDonald and Schwing (1973).

Henderson and Velleman (1981) stated a fundamental axiom of their philosophy of data analysis "The data analyst knows more than the computer" (p. 391).

Checking for the Assumptions

Following the concern for possible violation of assumptions, methods to check for whether assumptions were tenable or not were developed using computer programs. Some of these methods were nicely summarized in a paper by Elmore, Woehlke, and Spearing (1990). They also compared the procedures available in SAS and SPSS^{X}. Leitner (1990) provided examples of how multicollinearity among independent variables can be detected using the SAS and SPSSX computer packages, and recommended procedures for reducing the extent of multicollinearity. In addition, Pohlmann (1990) presented some methods using SAS (version 6) check for outliers.

Multivariate Technique

Although it was originally developed in the 30's (Hotelling, 1935), canonical correlation was not realized as the most general case of the general linear model until the late 70's or early 80's.

> ...Baggaley (1981) has noted that canonical correlation analysis, and not regression analysis, is the most general case of the general linear model. Knapp (1978) demonstrated this in detail and concluded that "virtually all of the commonly encounter parametric tests of significance can be treated as special cases of canonical correlation analysis, which is the general procedure for investigating the relationships between two sets of variables." In a similar vein Fornell (1978) notes, "Multiple regression, MANOVA and ANOVA, and multiple discriminant analysis can all be shown to be special cases of canonical analysis...." (Thompson, 1984)

Extended from a single dependent variable in the model to multiple dependent variables, canonical correlation could be used at least to predict or explain a set of dependent variables by a set of independent variables. When the dependent variables are categorical, the procedure is called discriminant analysis. The roles of discriminant analysis include that separation, discrimination, and estimation of the populations of objects (Huberty, 1975). Since a great deal of research in the behavioral sciences involves these three aspects, discriminant analysis has been considered as, follow-up technique to MANOVA, one of the most significant development in multivariate analysis.

Conclusion

While this journey through the literature was not exhaustive (although it may have been tiring to many readers) and strictly chronological, the authors feel that a similar trend of introduction, questioning, and resolution of the problems for the statistical technique of multiple regression existed as with t-test, factor analysis and meta-analysis. Perhaps other statistical procedures could similarly be documented.

References

Baggaley, A. R. (1981). Multivariate analysis: An introduction for consumers of behavioral research. *Evaluation Review, 5*, 123-131.

Bohrnstadt, R.A., & Carter, T.M. (1971). Robustness in regression analysis. In H.L. Coster, (Ed.) *Sociological methodology* . San Francisco: Jossey-Bass.

Box, G. E. P. (1966). Use and abuse of regression. *Technometrics, 8*, 625-629.

Cohen, J. (1968). Multiple regression as a general data-analytic system. *Psychological Bulletin, 70*, 426-443.

Darlington, R. B., & Boyce, C. M. (1982). Ridge and other new varieties of regression. In G. Keren (Ed.), *Statistical And Methodological Issues In Psychology And Social Science Research*. Hillsdale, NJ: Erlbaum.

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

Darlington, R. B., & Rom, J. F. (1972). Assessing the importance of independent variables. *American Educational Research Journal, 9*, 449-462.

Elmore, P., Woehlke, P., & Spearing, D. (1990, April) *Testing assumptions in multiple regression: a comparison of procedures available in SAS and SPSSX*. Paper presented at the Annual meeting of the American Educational Research Association, Boston, MA.

Efroymson, M. A. (1960). Multiple regression anlysis. In A. Ralston & H. S. Wilf (Eds.), *Mathematical Methods for Digital Computers*, New York: Wiley.

Fornell, C. (1978). Three approaches to canonical analysis. *Journal of the Market Research Society, 20*, 166-181.

Gentle, J. E. (1977). Least absolute values estimation: An introduction. *Communications in Statistics, B6*, 313-328.

Goldberger, A. S. (1964). *Econometric theory*. New York: Wiley.

Gordon, R. A. (1969). Issues in multiple regression. *American Journal of Sociology, 73*, 592-616.

Heise, D. R. (1970). Causal inference from panel data. In E. F. Borgatta, & G. W. Bohrnstedt (Eds.), *Sociological methodology: 1970*, pp. 3-27. San Francisco: Jossey-Bass.

Heise, D. R. (1969). Problems in path analysis and causal inference. In E. F. Borgatta & G. W. Bohrnstedt (Eds.), *Sociological Methodology: 1969*, (pp. 38-73). San Francisco: Jossey-Bass.

Henderson, H., & Velleman, P. F. (1981). Building multiple regression models interactively. *Biometrics, 37*, 391-411.

Hocking, R. R. (1960). Selection of the best subset of regression variables. In K. Enslein, A. Ralston, & H. S. Wilf (Eds.), *Statistical Methods For Digital Computers*. Vol. 3, (pp. 39-57). New York: Wiley.

Hogg, R. V. (1979). Statistical robustness: One view of its use in application today. *The American Statistician, 33*, 108-115.

Hotelling, H. (1935). The most predictable criterion. *Journal of Experimental Psychology, 26*, 139-142.

Huber, P. J. (1973). Robust regression: Asymptotics, conjecture and Monte Carlo. *Annals of Statistics, 1*, 799-821.

Huberty, C. J. (1975). Discriminant analysis. *Review of Educational Research, 45*, 543-598.

Huberty, C. J. (1989). Problems with stepwise methods– better alternatives. In B. Thompson (Ed.), *Advances In Social Science Methodology*. Vol. 1, (pp. 43-70). Greenwick, CT: JAI Press.

Huberty, C. J., & Mourad, S. A. (1980). Estimation in multiple correlation/prediction. *Educational and Psychological Measurement, 40*, 101-112.

Huynh, H. (1982). A comparison of four approaches to robust regression. *Psychological Bulletin, 92*, 505-512.

Johnstone, J. (1972). *Econometric methods (2nd ed.)*. New York: McGraw-Hill.

Kachigan, S. K. (1986). Statistical analysis: An interdisciplinary introduction to univariate and multivariate methods (2nd ed.). New York: Holt, Rinehart and Winston.

Kelly, F. J., Beggs, D. L., McNeil, K. A., Eichelberger, T., & Lyon, J. (1969). *Multiple regression approach*. Carbondale, IL : Southern Illinois University Press.

Kerlinger, F., & Pedhazur, E. J. (1973). *Multiple regression in behavioral research*. New York: Holt, Rinehart and Winston.

Knapp, T. R. (1978). Canonical correlation anlysis: A general parametric significance-testing system. *Psychological Bulletin, 85*, 410-416.

Leitner, D. W. (1990, April). Detecting multicollinearity in multiple regression: A comparison of procedures in SPSSX and SAS. Paper presented at the Annual meeting of the American Educational Research Association. Boston, MA.

Leitner, D. W. (1990). The rise and fall and rise of selected statistical procedures. *MWERA Researcher, 3*(1), 8-17.

Mantel, N. (1970). Why stepdown procedures in variable selection. *Technometrics, 12*, 621-625.

McDonald, G. C., & Schjwing, R. C. (1973). Instabilities of regession estimates relating air pollution to mortality. *Technometrics*, 463-481.

McNeil, K. A., & Spaner, S. D. (1971). Highly correlated predictor variables in multiple regression models. *Multivariate Behavioral Research, 6*, 117-125.

Pedhazur, E. J. (1982). *Multiple regression in behavioral research*. New York: Holt, Rinehart and Winston.

Pohlmann, J., (1990, April). Case influence statistics available in SAS (Version 5). Paper presented at the Annual meeting of the American Educational Research Association, Bosten, MA.

Pope, P. T., & Webster, J. T. (1972). The use of an F-statistic in stepwise regression procedure. *Technometrics, 14*, 327-340.

Price, B. (1977). Ridge regression: application to nonexperimental data. *Psychological Bulletin, 84*, 759-766.

SAS Institute Inc. (1985). *SAS user's guide: Statistics, Version 5 edition*. Cary: NC: SAS Institute Inc.

Sockloff, A. L. (1976). The analysis of nonlinearity via linear regression with polyomial and product variables: An examination. *Review of Educational*

Research, 46, 267-291.

Tabachnick, B. G., & Fidell, L. S. (1989). *Using multivariate statistics (2nd ed.)*. New York: Harper & Row.

Thompson, B. (1984). Canconical correlation analysis: Use and interpretation. Beverly Hills, CA: Sage.

Thompson, B. (1989). Why won't stepwise methods die? *Measurement and Evaluation in Counseling and Development, 21*, 146-148.

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

Wainer, H., & Thissen, D. (1976). Three steps towards robust regression. *Psychometrika, 41*, 9-34.

Wilkinson, L. (1979). Tests of significance in stepwise regression. *Psychological Bulletin, 86*, 168-174.