MULTICOLLINEARITY: EFFECTS, SYMPTOMS, AND REMEDIES

Cleve E. Willis and Robert D. Perlack, et. al.

ABSTRACT

Multicollinearity is one of several problems confronting researchers using regression analysis. This paper examines the regression model when the assumption of independence among the independent variables is violated. The basic properties of the least squares approach are examined, the concept of multicollinearity and its consequences on the least squares estimators are explained. The detection of multicollinearity and alternatives for handling the problem are then discussed. The alternative approaches evaluated are variable deletion, restrictions on the parameters, ridge regression and Bayesian estimation.

It is probably safe to conclude that while the proportion of the readership of this Journal which would claim mastery of econometric theory is relatively low, most of us in the Northeast make at least occasional use of least squares procedures for estimating relations of interest. It is a healthy sign that, led by the increasing accessibility of easily understood software packages for regression analysis, applied researchers have adopted least squares and other regression techniques *en masse* as an important research tool. This trend is accompanied, however, by the growing danger that users of the technique may unwittingly introduce bad information in the form of misinterpretation of results arising from undetected violations of the (at least implicit) assumptions behind the analysis. Although said in reference to another topic, Bellman's [p. 15] words are cogent here as well, "The fault of so many mathematical studies of this type is not so much in sinning as in the lack of realization that one is sinning, or even a lack of acknowledgment of any conceivable type of sin."

In what follows, we have attempted to synthesize some of the literature on multicollinearity and to reorganize and present this material in a less technical form and, hopefully, in a way more understandable to many applied researchers. Our approach will emphasize readability for the occasional user of econometric methods, and where this is at odds with technical completeness, the latter will be relegated to "footnote status." This paper should be regarded less as a dispositive treatment of the subject than as a "guided tour" through some of the important concepts relevant to the applied economist.

As the title suggests, we shall focus on describing the consequences of multicollinearity, on methods for detecting its presence, and on methods for reducing its level and its associated adverse consequences. The section immediately below provides the background for this undertaking. It sets out the linear model, the usual assumptions made, and the properties which result. The subsequent section defines the condition of multicollinearity, summarizes consequences, suggests ways of detecting it, and lists methods which have been advanced for overcoming multicollinearity difficulties. Since the methods generally introduce bias while reducing variance, a criterion is needed for comparing these alternatives. The mean square error criterion is described in the subsequent section. The alternative methods are then described in some detail in the following section, and finally comparisons are made and conclusions are drawn.

Cleve E. Willis is Associate Professor and Robert D. Perlack is Visiting Assistant Professor, Food and Resource Economics, University of Massachusetts.

This paper emanated in part from a course in Econometrics at the University of Massachusetts. The following students are co-authors of this article: David Himelfarb, Thomas McBride, Philip Sczerzenie, Vivien Singer, Franklin Tirsch, Stuart Westin, and Yisehac Yohannes. This paper was made possible by support from the University of Massachusetts Experiment Station.

LEAST SQUARES APPROACH

In this section, we set out the general linear mode 1 and the assumptions usually made in using this model—this combined specification is termed the "maintained hypothesis". The least squares technique is then described and, following this, the properties of least square estimators, given the usual assumptions, are reviewed.

Maintained Hypothesis

For present purposes, assume we are interested in learning about a single relation given by:

$$(1)\ Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \ldots + \beta_k X_{ki} + \mu_i,$$

where:

Y_i is the value of the dependent variable for the i^{th} observation, $i = 1, \ldots, n$,

X_{ji} is the value of the j^{th} independent variable for the i^{th} observation,

β's are the unknown parameters which give the expected impact of a small unit change in that particular X on Y, and

μ_i is the value of the i^{th} error or disturbance.

In words, the value of some dependent variable (Y) is presumed to be influenced by a set of independent variables (X's). But this relation is not exact or "deterministic". Even if you knew all the β's and the i^{th} values of each of the X's, you would not necessarily know the exact value of Y_i. For a variety of reasons (errors of measurement, neglected data, etc.) Y may differ from what you might *expect* it to be given particular values of the X's and of the β parameters. This difference for the i^{th} observation is the random variable μ_i.

Further assumptions about (1) are that the disturbances have an expected value of zero (any observation has a positive, zero, or negative μ_i, this value is as likely to be positive as negative; and if we were to *average* a sufficiently large number of these random variables, this mean would be zero), and a variance which is some constant and hence which is independent of the values of the X's. In the occasionally mystical parlance of the econometricians, this assumption of constant variance is called *homoscedasticity*. The disturbance terms are also assumed to be independent of, or uncorrelated[1] with, one another—violation of this assumption is called *autocorrelation*. Finally, the X's and μ_i are presumed independent (uncorrelated) and the X's are independent of one another in the sense that there is no way to express the data (n observations) on X_2, for example, as a linear combination of one or more of the other X's. Violation of this assumption is termed *multicollinearity*, and results in a breakdown of the least squares procedures. We return to this topic in the next section.

Least Squares Approach

Referring back to relation (1), the β's and the μ_i (and its constant variance) are never known. Ultimately, we wish to use what is known (the Y's and the X's) to estimate these β's, to make inferences based on probabilities about these β's and perhaps to predict values of Y given values of the X's.

Since relation (1) is unknown, let us replace it by:

$$(2)\ Y_i = b_1 + b_2 X_{2i} + b_3 X_{3i} + \ldots + b_k X_{ki} + e_i$$

The b's are just numbers we choose to replace the β's and e_i is the difference between the i^{th} value of Y and the sum of the b's times the i^{th} values of the X's. This e_i is often called the *residual* for fairly obvious reasons. We are free, of course, to select the b's (our estimates of β) any way we like. It might make sense, however, to select the b's, given the X and Y data, in such a way as to make the values of those e's in our sample small. The *least squares* criterion says to select the set of b's so as to make the squared values of e_i summed over all sample observations as small as possible.

Mechanically, the least squares method for estimating b's is quite simple. If we move all terms other than e_i in (2) to the left side of the equation, square both sides and sum over the sample, we have the sum of squares we seek to minimize expressed in terms of X's, Y's, and b's. The X's and Y's are given data and the b's are variables. It is a straight-forward matter to solve for a set of equations for the b's in terms of the X's and Y's which minimize the sum of squares of the residuals. These equations, called *normal* equations, are obtained by calculus procedures. It is then simply a matter of substituting the X and Y data into these equations and solving for the least squares estimators of $\beta_1, \ldots, \beta_k$ —namely, $b_1, \ldots, b_k$. Other information, liked standard deviations (synonymously, standard errors) of the b's, is routinely produced by computer programs in the process.

Properties of Least Squares Estimators

As a final part of this section, the properties of least squares estimators, given the assumptions made earlier, are briefly described. These become relevant in comparing alternative procedures for reducing the effects of multicollinearity below.

Least squares estimators are often cryptically described as BLUE. These letters stand for *best linear unbiased estimators*. Several definitions are in order:

i) Unbiased—If we regard the X's as a set of fixed numbers[2] and we could repeatedly draw samples of sets of X's and Y's, where the X's remain the same and the Y's vary due to the different disturbances (μ_i), then we could find, repeatedly, different estimators (b's) of the β's. Some of these would be too high, some too low, and, as in Goldilocks and the Three Bears, some would be just right. The problem is that we don't repeatedly sample—that would be prohibitively expensive. It can be shown, however, that if we were to do so, given the assumptions made earlier, the *average* of all the b_2's, for example, would be the unknown β_2 This is the essence of the unbiasedness property. What consolation does this property provide us? Merely that while the b_2 which is estimated from our sample of Y and X's may be greater or less than the real but unknown β_2, it had an equal chance of being above or below β_2. The same cannot be said for estimators which are biased.

ii) Linear—The least squares estimators of the β's are linear estimators in the sense that the normal equations express the b's as linear combinations of the Y's. The constants or weights in expressing these linear combinations are formed from the fixed X values. The import of this is that the b's are easy to calculate and to interpret relative to nonlinear estimators.

iii) Best—This term refers to the minimum variance of the b's property. If one considers again the paradigm of repeatedly drawing samples of Y and X's and re-estimating the b's each time, then the b's are random variables with expected values of the β's (given the unbiased property) and with some variance for each of the b's. While we don't repeatedly draw samples in practice, it is possible to derive unbiased estimates of these variances from a single sample. The square roots of these variances, generally termed standard errors, are often reported in parentheses below the estimated values of the b's.

In brief the BLUE properties of least squares estimators, given the usual assumptions, state that of the set of estimators which are both unbiased and simple linear functions of the Y data, are least squares estimators having variances which are minimum.

MULTICOLLINEARITY

One of the requirements made in the previous section was that the explanatory variables, the X's, be independent of (not perfectly correlated with) one another. The opposite situation, where the X's are linearly dependent and hence can be expressed as linear combinations of one another, is called *perfect multicollinearity*. In this sense, the condition of perfect multicollinearity is either *existent* or *non-existent*. If it exists, the least squares procedure simply breaks down and is incapable of estimating the β's. If multicollinearity does not exist in this sense, there may still be problems stemming from high, but not perfect, correlation among the X's.

It seems more useful, then, to speak in terms of the multicollinearity problem's *severity* rather than its existence. A case of perfect multicollinearity is rare, as is a zero correlation among explanatory variables (X's). Accordingly, multicollinearity will be defined here in terms of departures from independence, or from non-correlation, of the X's with one another.

[1]Technically, statistical independence is a stronger assumption than uncorrelated. For present, we use the terms interchangeably to facilitate exposition.

[2]We need not make this strong as assumption. However, for present purposes the gains from making a weaker assumption are more than offset by the costs of increasing the complexity of this discussion. See Johnston.

Having defined multicollinearity in a relative rather than an absolute sense, we must face the increased problem of detection. There are no longer standard test statistics which permit statements of differences from hypothesized values of parameters; as, for example, the Durbin-Watson statistic for rejecting hypothesized zero autocorrelation. Nevertheless, rules of thumb for detection of "serious" levels of multicollinearity are important, and some of these are suggested in the second part of this section below. Once severe multicollinearity is diagnosed, its resolution requires the generation of additional information. The form of this additional information should depend upon the specifics of the particular problem. The final part of this section briefly describes the more common forms of additional information, and a later section shows how the problems *specifics* should guide choice of information form. But first, we summarize the consequences of high degrees of multicollinearity.

Consequences

The limiting case of perfect linear dependence within the set of explanatory variables leads to a completely indeterminate set of parameter estimates (b's). In its "brute" and "tactless" way, as Farrar and Glauber [1967, p. 93] put it, the mathematics ". . .tells us that explained variance can be allocated completely arbitrarily between linearly dependent members of a *perfectly correlated* set of variables, and almost arbitrarily between members of an almost *perfectly correlated*[3] set."

In a sense, the investigator whose set of explanatory variables is perfectly correlated may be more fortunate than one whose variables are nearly so. The former's problems are soon discovered by the mechanical inability to derive b's, while the latter's problems may never be fully understood. For the latter, multicollinear explanatory variables will result in large variances (and standard errors) on regression coefficients; indicating the low information content of data observed and hence the low quality of the resulting parameter estimates.

The presence of multicollinearity alone does not lead to bias in estimating the parameters, and this is perhaps a major factor leading to complacency on the parts of investigators. However, investigators are sometimes led to incorrectly delete explanatory variables because their coefficients are not statistically different from zero. The true situation may be not that the variable has no influence on Y, but that the particular set of sample data does not enable us to detect it. Thus, to the extent that multicollinearity leads to a misspecification of the model, least squares estimators will be biased.

A final consequence of multicollinearity is that estimates of coefficients become particularly sensitive to given sets of sample data. Often the addition of several observations on these explanatory variables can yield dramatic changes in estimated coefficients.

In sum, the primary consequence of multicollinearity is an inability to distinguish the separate contributions of explanatory variables which exhibit little truly independent variation in explaining variance of the dependent variable. If our purpose is less in explaining separate influences as in predicting values of the dependent variable, however, these effects are not so severe, provided the values of the explanatory variables for which we wish to predict dependent variables are correlated in the same way as the sample of X's.

[3]The expressions in italic have been substituted for Farrar and Glauber's more technical expression, singular.

Detection

Since we have chosen to deal with *degrees* of multicollinearity rather than *existence,* tests are made difficult. The simplest, and admittedly arbitrary, rule of thumb would involve defining "harmful" multicollinearity to exist when simple pairwise correlations between the exogenous variables exceed, say, r = .8. This would help avoid the most obvious form of pairwise interdependence, although correlations of a more complicated form may be severe, even while pairwise correlations are slight.

A more elaborate definition is suggested by Klein [1962, p. 101], who defines harmful multicollinearity to exist when the magnitude of pairwise correlation between two explanatory variables is as large as, or larger than, the coefficient of multiple correlation between the dependent variable and the set of explanatory variables. Again, a set of X data passing inspection by this criterion may still be highly correlated in a more complex way than pairwise. A classic and an aggravatingly common example has been termed by Johnston [1972, p. 178], the "dummy variable trap".

Farrar and Glauber extend the concept of simple correlation between explanatory variables to include multiple correlation within a set of explanatory variables. The test of multicollinearity they provide is based on the F distribution and is summarized in Johnston [1972, pp. 163-4]. Essentially, it involves regressing, in turn, each of the explanatory variables against the remaining set of explanatory variables and observing the coefficient of multiple determination (R^2) for each regression. If it is high, then the explanatory variables are closely related in some linear way and vice versa. While the formal test will nearly always reject the null hypothesis of independence of the X's, at least the values of the test statistics may provide an indication of which explantory variables are most affected by multicollinearity, and where to begin the search for better data.

A final note on diagnosis of "harmful" multicollinearity derives from the first sympton mentioned above—viz., the resulting inordinately large standard errors of the b's. It would be a good bet that multicollinearity is present if, for example, all standard errors were larger than their corresponding estimated b's, while the coefficient of multiple determination was high.

Alternatives

As indicated earlier, when severe conditions of multicollinearity within the sample data on the explanatory variables are diagnosed, the usual econometric prescription is acquisition of additional information, although this information can take many forms. The most common procedure adopted, however, is deletion of variables. Some of these approaches are briefly sketched below, and expanded upon in a subsequent section.

1. Variable Deletion

The simplest approach to resolving the problem of severe multicollinearity is to simply drop from the analysis explanatory variables which appear to be highly correlated. While this may resolve the multicollinearity problem, the general result is misspecification of the model and biased estimation. Stepwise regression procedures are formal techniques which achieve roughly the same thing.

2. Restrictions

An alternative which is sometimes used is to impose additional information in the form of restrictions on a value of b or some combination of regression coefficients. The restrictions may be exact or of the inequality form. In either form, as will be seen below, these restrictions can be used to reduce the effects of multicollinearity.

3. Principal Components

Another appoach involves the generation of "artificial" new variables (principal components) which are truly independent of one another, but which explain as much of the variation of the dependent variable as possible. These new variables are formed as weighted combinations of the explanatory variables.

4. Ridge Regression

Ridge regression involves a mechanical augmenting of, essentially, the matrix of correlations of the explanatory variables used in calculating the b's. In the process, precision is improved (standard errors are reduced).

5. Bayesian Estimation

Bayesian estimation involves combining (prior) information about the b's, in the form of probability distributions, with the sample data to obtain *posterior* estimates of the b's. This information is combined in a well-defined mathematical way which weights each source of information inversely with its variance.

Each of the approaches indicated above may be useful in reducing the standard errors of the estimators and hence in estimating the parameters more precisely. Unfortunately, the unbiasedness property of least squares is generally sacrified in the process. If we are to compare these procedures to decide which is better, we need a criterion of "better" which considers both bias and precision. The discussion below offers one such criterion to be used in the comparisons made in the final two sections.

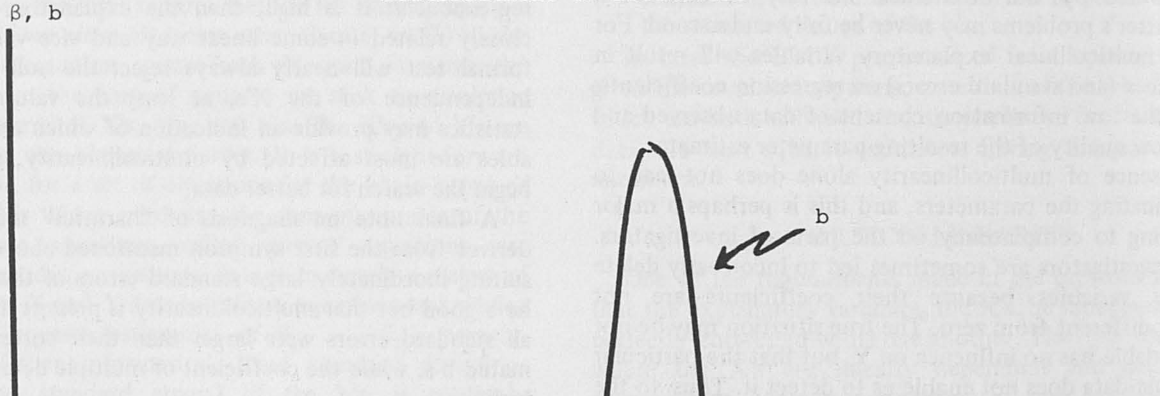
Mean Square Error

As indicated, the alternatives above can be regarded as adding additional information by imposing restrictions in the estimation process. For example, in choosing to delete a variable, one is in effect restricting the β parameter corresponding to the deleted variable to be zero. If these restrictions are true, precision will be improved (standard errors lowered) and the estimators remain unbiased. In the more usual situation in which these restrictions are not exactly true, imposing them still reduces standard errors, but at the expense of introducing some bias. Thus, the premise behind the alternative biased estimators discussed here is that the introduction of a small amount of bias may be a reasonable price to pay in order to achieve a major reduction in variance. Visually, biased estimator b in Figure 1 may be preferred to unbiased estimator $\hat{\beta}$, since we are more likely to be "closer" to β with estimator b than with $\hat{\beta}$.

Figure 1

Distributions of Estimators

Probability of $\hat{\beta}$, b



This notion of a reasonable trade-off between variance and bias, of course, remains nebulous until we define "reasonable". Mean square error (MSE) criteria have been most commonly applied for this purpose. For the case of a single estimator, MSE is variance plus the bias squared.[4] Thus, according to this criterion, switching from an unbiased estimator (like β) to a biased estimator (like b) is "reasonable" if the square of the bias introduced is less than the reduction in variance (square of the standard error) achieved.

A problem with this criterion is that it is dependent upon the value of the unknown parameter β, such that for some values of β a particular estimation procedure may be better than another in the MSE sense, while for other values of β, the reverse may be true. This does not mean that the MSE criteria are useless. Rather, it suggests that the alternative estimators discussed here will out-perform each other and ordinary least squares in some circumstances and not in others. The section to follow describes these alternative estimators in somewhat more detail and outlines the existing evidence regarding the *circumstances* favoring each method in the presence of multicollinearity.

ALTERNATIVES

The alternative procedures which might be employed in the presence of multicollinearity, which were listed in the second section above, are: deleting variables; principal components; restrictions on linear combinations of parameters; ridge regression; and Bayesian estimation.

Variable Deletion

Easily the most common approach to "resolving" situations in which multicollinearity is present is to delete variables which were included in the original model specification. This deletion is often at the arbitrary discretion of the investigator, where high correlations, low significance, or both, are used as the basis for such a decision. Other times a more formal procedure, the leading example of which is the use of stepwise regression programs to add or drop regressors, is followed. Whether formal or informal procedures are followed, whenever the same set of data is used to specify the final model as well as to estimate the parameters of this specification, the process is termed "pretest" or "sequential estimation". In an excellent survey article on the subject, Wallace [1977] describes the consequences of pretesting and compares these with the consequences of using OLS on the original specification and the alternative of applying restrictions on the original specification (restricted least squares RLS) before estimation is undertaken. Since pretest and RLS estimators are biased in general, the MSE criterion was used for comparison. Wallace's results show that for very accurate prior information (restrictions), RLS is best, OLS is worst, and pretest is in between. For bad priors, the ordering is reversed, and in all cases pretest (deletion) estimators are inferior to the best of OLS or RLS by the MSE criterion. While Wallace's survey was not written expressly for the situation of multicollinearity, the results are instructive.

Freund and Debertin [1975] provide additional insight into the consequences of pretesting. They use Monte Carlo experiments to make the consequences of variable deletion more graphic. In brief, a host of irrelevant variables were subjected to stepwise regression and as predicted an overly large proportion of these appeared to be statistically significant by the usual procedure (the t-test). The demonstration verifies the analytical result that in the presence of pretest estimation, the usual tests do not provide valid indicators of significance. Their conclusion is not that sequential estimation should never be used, but rather that when *exploratory data analysis* is used, the results should be clearly labeled as such so as not to mislead the user of this information.

Principal Components

Principal Components (PC) techniques are sometimes used when the number of regressors is so large as to severely limit degrees of freedom and/or when these regressors are highly intercorrelated. Essentially, PC transforms a set of highly correlated regressors into a set of uncorrelated variables,[5] where each new PC is an artificial variable expressed as some linear combination of the original regressors. These new variables (components) can now be tested with greater accuracy since there is no distortion due to multicollinearity among these PC's.

Since the PC's are artificial constructs, it is usually not possible to attach an economic meaning to the effect of any particular component. To surmount this difficulty, principal components regression can be used. This involves deleting one or more of the PC's[6] and then solving back for the implied effects of the separate original regressors.

The PC estimator has a smaller variance than the OLS estimator and for some values of the true parameters will have smaller MSE. In particular, McCallum's work concludes that the PC estimator out-performs the OLS estimator the higher is the regressor correlation, the smaller is the number of observations, and the more alike are the parameters.

Restrictions

Economic theory or previous empirical studies may provide an investigator with some information concerning:

i) a specific value[7] of some of the regression parameters,
ii) proportional relationships between a set of the parameters, including a specific value of some linear combination of parameters,[8] and
iii) relations in which individual coefficients or linear combinations lie within some bounds such that restrictions of an inequality type can be expressed.

In this case, exact restrictions (i and ii) and inequality restrictions (iii) can be used to incorporate this information.

In the former case, the RLS procedure discussed earlier can be used. The variance of this estimator will always be smaller than for the OLS estimator, although the estimates are biased except when the restrictions are precisely correct. Whether the RLS estimator is better in MSE than OLS depends upon how inaccurate these restrictions are.

[4]See Wallace for a fuller discussion of the MSE criteria.

[5]See Johnston or Kendall for a more detailed description of this technique.

[6]See Mittlehammer and Baritelle.

[7]In this Journal, an example of this procedure is provided by Lindsay and Willis.

[8]For example, in the Cobb-Douglas production function, linear homogeneity requires that parameters sum to one.

If, on the other hand, the prior information about the parameters is less complete and exists in linear inequality form as in case (iii) above, then we may combine the information with the sample data to obtain better estimates of the parameters. However, this estimation procedure involves the use of quadratic programming (Judge and Takayama [1966]). This technique can handle both the equality and inequality restrictions together in the estimation of the parameters.

Ridge Regression

Ridge regression attacks the principal multicollinearity symptom, high estimator variance, by systematic data manipulation. Specifically, this procedure involves the addition of successively larger constant terms to the matrix of correlations among the explanatory variables used in calculating the parameter coefficient estimates. Each addition is followed by the derivation of a set of estimates. Brown and Beattie [1975] experimentally suggest that the variance reduction potential of this process is an increasing function of the magnitude of the constant term. Their experiment both provides the general range of ridge regression usefulness and, given usefulness, offers an intuitive technique for optimum constant term selection.

Ridge regression can confidently be used when the true values of the estimators are known to be of the same sign and roughly the same magnitude. Additionally, this procedure is preferable to others in this situation if the number of regressors is small and all should be retained. The Brown and Beattie experiment demonstrates large MSE reductions for all ridge regression applications, regardless of the true nature of the estimates. However, artificial augmentation of the explanatory variable correlation matrix introduces bias into the estimation process. Bias is extreme for situations which do not fulfill the true estimator sign and magnitude criteria.

Bayesian Estimation

Sometimes the prior information available can more nearly be regarded as stochastic, or subject to a probability distribution, than as exact or inequality restrictions. These cases lend themselves well to Bayesian estimation. In brief, Bayesian estimation involves combining sample data with a prior probability distribution in a well-defined mathematical way. It often involves numerical integration and can become increasingly difficult as the number of parameters increases. The offsetting advantages include the explicit formulation of stochastic side conditions on parameters and the incorporation of these priors with the sample data to produce "posterior" distributions on these estimates,[9] as well as the fact that Bayesian estimates are expressed in terms of a specific probability distribution, including means and variances.

Clearly, Bayesian estimation is most appropriate when the number of parameters is small and side conditions on some or all of these parameters are readily expressible as probability densities. Zellner, and Tiao and Zellner can be used for general background, while Rausser, *et al.* and Chowdhury, *et al.* serve as illustrative applications.

For present purposes, the Chowdhury article is most interesting. Preceding their work, the various production function estimations of a Cobb-Douglas form had been fairly unsuccessful. Some of the investigators were baffled; others recognized that the intuitively unsatisfactory results were the result of multicollinearity. Finally, Doll discovered that if the conditions generally assumed by economists regarding the production function were true, then multicollinearity would be the logical consequence. Indeed, exact multicollinearity would result from these economic conditions being met. Rather than being discouraged by the presence of conditions of multicollinearity, Doll (p. 558) comments, the investigators ". . .should be pleased because the presence of multicollinearity serves as a verification of their economic model."

Chowdhury's work involves using the data from one of these previous studies which produced counter-intuitive results and estimating the production function using Bayesian procedures. As expected, the Bayesian estimates were consistent with prior knowledge about the parameters.

SUMMARY

The sections above described the condition of multicollinearity, its detection, and its effects. Some alternative procedures were described which lead to a reduction in variance of the estimators. Since these alternatives are not unbiased, a trade-off between bias and variance reduction is involved. MSE was advanced as one means of evaluating this trade-off.

The alternatives considered here include: variable deletion (and stepwise regression); principal components regression; RLS and quadratic programming procedures for inequality restrictions on parameters; ridge regression; and Bayesian analysis. None of these alternatives is best in all situations–a summary of the situations which favor each is provided.

Variable deletion procedures should generally be reserved for situations characterized as exploratory data analysis. The results should be clearly labeled as such and statistical tests should not be made with the results, since the tests will be misleading.

Principal components regression and ridge regression procedures are suitable candidates as alternatives to OLS in cases in which the parameters are expected to be of similar sign and roughly similar magnitude. Principal components is best in cases of high correlations, large number of regressors, and few observations. To the contrary, ridge regression is at its best when the number of regressors is small. For cases of a large number of regressors, parameters of opposite sign and dissimilar size, ridge regression estimators are subject to extreme bias. Software packages for both techniques are easy to obtain and use.

If information about specific values of parameters or linear combinations is known with fair precision, RLS is a useful approach, being clearly superior to OLS in an MSE context. Likewise, when inequality information is present (e.g., a parameter lies between zero and one), quadratic programming can be used. RLS software programs are ubiquitous; adaptations of quadratic programming for the second case are not.

If the prior information can be expressed as a probability density, Bayesian estimation is feasible. It is capable of incorporating more information than the quadratic programming procedure above and exhibits an intuitively pleasing method of combining information. It is at its best when side conditions can be expressed by probabilities and the number of parameters is small. Unfortunately, there are not general software packages and application will involve some sophisitication with statistical methods.

[9]Essentially, the Bayesian method for combining priors and sample information weights each inversely proportional to its variance.

We conclude with two comments. First, there are glaring gaps in this presentation. This is not necessarily bad–if this article proves useful for some of the readers, follow-ups on topics not covered may be as well. Second, the discussion of alternatives may be academic if software packages for their use are not available. At least one new program, called SHAZAM, is available which provides, among many others, OLS, RLS, Principal Components regression, and Ridge regression options. The user's manual is available upon request from Kenneth White at Rice University.[10]

REFERENCES

Bellman, R., *Adaptive Control Processes: A Guided Tour,* Santa Monica: RAND Corporation, 1961.

Brown, W. and B. Beattie, "Improving Estimates of Economic Parameters by Use of Ridge Regression with Production Function Applications", *AJAE,* 22:21-32, February 1975.

Chowdhury, S. R., *et al.*, "A Bayesian Application on Cobb-Douglas Production Function," *AJAE,* May 1975, pp. 361-363.

Doll, J., "On Exact Multicollinearity and the Estimation of the Cobb-Douglas Production Function", *AJAE,* 56:3:556-563, August 1974.

Farrar, D. E. and R. R. Glauber, "Multicollinearity in Regression Analysis: The Problem Re-visited", *Review of Economics and Statistics,* 49:92-107, 1967.

Freud, R. D. Oebertin, "Variable Selection and Statistical Significance: A Sampling Experiment", *AJAE,* 57:4:721-722, November 1975.

Johnston, Jr., *Econometric Methods,* New York: McGraw-Hill, 1972.

Judge, G. and T. Takayama, "Inequality Restrictions in Regression Analysis", *JASA,* 61:166-181, March 1966.

Kendall, M., *A Course in Multivariate Analysis,* London: Charles Griffin, 1957.

Klein, L. R., *An Introduction to Econometrics,* Prentice-Hall, 1962.

Lindsay, Bruce E. and Cleve E. Willis, "Factors Influencing Land Values in the Presence of Suburban Sprawl", *Journal of the Northeastern Agricultural Economics Council,* 3:1:112-124, May 1974.

McCallum, B., "Artificial Orthogonalization in Regression Analysis", *Review of Economics and Statistics,* 52:110-113, 1970.

Mittlehammer, R. and J. Baritelle, "On Two Strategies for Choosing Principal Components in Regression Analysis," *AJAE,* 59:2:336-343, May 1977.

Rausser, G; C. Willis; and P. Frick, "Learning, External Benefits, and Subsidies in Water Desalination," *Water Resources Research,* 8:6:1385-1400, December 1972.

Tiao, G. and A. Zellner, "Bayes' Theorem and the Use of Prior Knowledge in Regression Analysis", *Biometrica,* 51:219-230, 1964.

Wallace, T., "Pretest Estimation in Regression: A Survey", *AJAE,* August 1977, pp. 431-443.

White, K., "A General Computer Program for Econometric Methods–SHAZAM", *Econometrica,* 46:1:239-240, January 1978.

Zellner, A., *An Introduction to Bayesian Inference in Econometrics,* New York: John Wiley and Sons, 1971.

[10]See White.