THE JOURNAL OF FINANCE · VOL. XXXII, NO. 3 · JUNE 1977

PITFALLS IN THE APPLICATION OF DISCRIMINANT ANALYSIS IN BUSINESS, FINANCE, AND ECONOMICS

ROBERT A. EISENBEIS*

I. INTRODUCTION

OF THE APPLIED DISCRIMINANT analysis papers that have appeared in the business, finance, and economics literature to date, most have suffered from methodological or statistical problems that have limited the practical usefulness of their results. While it is not true that the statistical problems are unique to economics or finance, it does seem that the nature of the subject matter and data are such that one can expect to encounter statistical difficulties more frequently than in many other application areas. The problems are of several different types, among which are difficulties with (1) the distributions of the variables, (2) the group dispersions, (3) the interpretation of the significance of individual variables, (4) the reduction of dimensionality, (5) the definitions of the groups, (6) the choice of the appropriate *a priori* probabilities and/or costs of misclassification, and (7) the estimation of classification error rates. The purpose of this paper is to discuss these problems of application of discriminant analysis techniques. Ample references are made to the literature for examples to illustrate the pitfalls. Finally, a brief discussion of future problems and prospects for statistical research on the application of the techniques is provided.

II. THE DISTRIBUTION OF THE VARIABLES

The standard discriminant analysis procedures assume that the variables used to describe or characterize the members of the groups being investigated are multivariate normally distributed. In practice, deviations from the normality assumption, at least in economics and finance, appear more likely to be the rule rather than the exception. Violations of the normality assumptions may bias the tests of significance and estimated error rates. Hence, it is of interest to determine whether the assumption holds and what effects its relaxation may have on the tests and on the classification. In the applied literature, the problem of testing for the appropriateness of the distributional assumption has been largely ignored. This is due in part, one would presume, to the fact that most available normality tests are for univariate and not multivariate normality.[1] Malkovich and Afifi (1973) discuss several tests for multivariate normality, but these have not yet been used in discriminant analysis problems. Equally important, however, is that if the normal-

* Associate Research Division Officer, Division of Research and Statistics, Board of Governors of the Federal Reserve System. The author wishes to thank Drs. Joseph F. Sinkey, Jr. and Edward I. Altman for their helpful comments on this paper. However, all errors and omissions are the responsibility of the author. This work was completed while the author was Assistant Director of Research, FDIC.

1. For discussions and review of this literature see Shapiro, Wilk and Chen (1968), or Kowalski (1970), and more recent papers by D'Agostino (1973), and Shapiro and Francia (1972).

ity hypothesis is rejected, one is then faced with the practically impossible task of deriving the appropriate alternative joint probability density functions. The tactic which most researchers have adopted is simply to be satisfied that the more standard discriminant procedures yield reasonable approximations and proceed as if the normality assumption held.

The theoretical and statistical work dealing with the normality problems have been of two major types. Some researchers have investigated alternative schemes where specified types of nonnormality hold, while others have evaluated the robustness and bias introduced in the standard procedures when the normality assumptions are violated in known ways. In the case of the former, Hills (1967), Linhart (1959), and Chang and Afifi (1974) derived and examined classification methods where some or all the variables were discrete. In finance, and in particular in the development of credit scoring and bond rating schemes, many of the factors reflecting important attributes of the data tend to be categorical in nature. When continuous and discrete variables are mixed, procedures are proposed to split the samples based on the values of the discrete variables. Then standard discriminant analyses were employed on the subdivided samples. This procedure was used recently in an industrial and bond rating study by Pinches and Mingo (1975). Chang and Afifi (1974) have investigated Bayes' rules for the simple two-group case with one dichotomous variable and with one and with several continuous variables. Similar to Linhart (1959), their results indicate that it is appropriate to use the dichotomous variable to split the samples and then construct for the pairs of samples separate discriminant functions and rules.

In examining the robustness of the standard techniques when nonnormality holds Gilbert (1968) compared the performance of the linear discriminant function when applied to data where all the variables were discrete with the performance of two logit models and with a model which assumed mutual independence of the variables. She concluded that there was only a small loss in predictive accuracy using the linear function and that as the number of variables increased, the results should be quite stable.[2] More recently, a different aspect of the problem was studied by Lachenbruch, Sneeringer, and Revo (1973); they investigated the robustness of both linear and quadratic procedures for three specific nonmultivariate normal distributions. These distributions were transformations of normally distributed variables so that the true classification errors were known. The three distributions were the log normal, the logit normal, and the inverse hyperbolic sine normal. The authors concluded that the standard linear procedures may be quite sensitive to nonmultivariate normality. However, they also suggest that the problems may not be as great when the distributions were bounded, which was the case with the logit normal distribution, as compared with unbound distributions. They found that the estimated overall classification error rates were not affected as much as the individual group error rates. They also observed that even attempts to adjust for inequalities of the group dispersions by using quadratic classification

2. If one simply applies the standard techniques with the dichotomous variables as Gilbert (1968) did, it is quite easy to show that the assumption of equal dispersions is violated. Hence, linear techniques are always inappropriate when dichotomous variables are included in the analysis. This problem is described in detail in the next section.

rules did not significantly improve the results, and in many cases they were worse. The authors suggest that data should be first transformed, if possible, to approximate normality, and then standard tests for the equality of the group dispersion should be employed to determine whether linear or quadratic techniques should be used.

The use of certain transformations to transform data prior to estimating discriminant function has been a common procedure. The natural log and standard log are the most frequently used transformations. Pinches and Mingo (1973), Bates (1973), Carleton and Lerner (1969), Horton (1969), all used such transformations in their studies. This is because many of the variables employed, such as firm size, loan size, population, etc., tend to be highly skewed with a few large values and a number of small values. Hence, a much greater proportion of the values lie to the left of the variable mean than to the right. If a distribution is skewed to the right, the effect of the natural log transformation is to make the marginal distribution of the variable more symmetric and, of course, it is bounded from below by zero.[3,4] This procedure has intuitive appeal because it does make the distribution more symmetric, and probably more normal. But the cautions noted previously from the work of Lachenbruch, Sneeringer, and Revo (1973) still apply.

As a final point, it should also be recognized that the application of a tranformation may change the interrelationships among the variables and may also affect the relative positions of the observations in the group.[5] In the case of the log transformation, there is also an implicit assumption being accepted where such a transformation is employed. That is, the transformed variables give less weight to equal percentage changes in a variable when the values are larger than when they are smaller. If, for example, the variable being transformed was firm size, the implication would be that one does not believe that there is as much difference between a $1 billion and a $2 billion size firm as there is between a $1 million and a $2 million size firm. The percentage difference in the log will be greater in the latter than in the former case.

III. Equal Versus Unequal Dispersions

A second critical assumption of classical linear discriminant analysis is that the group dispersion (variance-covariance) matrices are equal across all groups. Relaxation of this assumption affects not only the significance test for the differences in group means but also the usefulness of the so-called "reduced-space transformations" and the appropriate form of the classification rules.

Little attention has been given to the effects of unequal dispersions on the hypothesis test of the equality of group means and related significance tests. Anderson (1958) presents an exact test for the equality of the means for the limited case when the sample sizes are equal. He provides an approximate test for the case

3. It should be noted that while the marginal distributions of a multivariate normal distribution are normal, simply making a variable's marginal distribution normal may not necessarily make the joint distribution more normal.

4. The disadvantage is that negative values cannot be transformed and are precluded.

5. Considerable attention is usually given in regression models about the implications that transformation of the variables have for the model being estimated. No attention has been given in discriminant analysis to this problem, but the same types of consideration apply.

when there are unequal sample sizes [see also Eisenbeis and Avery (1972)]. To date, however, such tests do not appear to have been programmed, or at least they have not been made readily available to researchers. More recently, Holloway and Dunn (1967) investigated the robustness of Hotelling's T^2 for the two group case with a particular type of inequality.[6,7] Drawing upon the work of Scheffe [1959], Hopkins and Clay [1963], and Ito and Schull [1964], their conclusion is that the robustness of the test depends upon both the number of variables and relative sample sizes in the groups. For the univariate case, the test seems to be affected very little as long as the sample sizes are equal or nearly equal, especially for larger samples. With widely disparate sample sizes, the actual significance level is greater than the hypothesized level, and therefore, the null hypothesis would be rejected more frequently when the means were in fact equal. When the number of variables increases, the significance level also increases and the sensitivity to unequal sample sizes increases. According to Holloway and Dunn (1967), "Equal sample sizes help in keeping the level of significance close to the supposed level, but do not help in maintaining the power of the test."

Considerable attention has been given by Cooley and Lohnes (1962), (1971), Rulon, Tiedeman, Tatsuoka and Langmuir (1967), and Tatsuoka (1971), among others, to the advantages of reduced-space discriminant analysis which can be used to reduce the original m dimensional variable test space to an r dimensional problem.[8,9] The transformation is selected as the matrix of eigenvectors associated with the roots of the determinantal equation $|T-\gamma W|=0$, which has r nonzero roots.[10] The reduction in dimensionality is possible because the linear transformation from test to reduced space preserves relative linear Eucledian distances among observations and leaves the significance tests and classification results unaffected. But this property holds if and only if the group dispersion matrices are equal. If the dispersions are not equal, then the transformation to reduced space is no longer distance perserving. The result is a warping of the relative positions of the observations in reduced space which affects both the significance tests and changes the resulting classifications. Lohnes (1961) has examined a portion of this problem by comparing linear test and linear reduced-space classifications when the group dispersions are unequal. However, his results are difficult to assess and are not particularly useful since he does not compare the linear results with the appropriate quadratic test space results. Since the assumption of equal dispersions is used in transforming the data from test to reduced space, the differences he observes are due solely to the warping of the reduced-space transformations in response to the unequal dispersions in his data.

6. With little loss in generality, they considered population of the form $N_1(0,I)$ and $N_2(M,\delta\cdot I)$. Because most of their examples had equal eigenvalues this tends to limit their conclusions somewhat.

7. See Porebski (1966), for a discussion of the relationships among the various statistics used in multivariate analysis.

8. r equals the minimum of m and one minus the number of groups (i.e., $k-1$).

9. Moreover, as m increases, the reduced-space variables are more likely to be multivariate normal because of the Central Limit Theorem [Lohnes (1961)].

10. T is the total deviation sums of squares matrix and W is the pooled within groups deviation of squares matrix. See Eisenbeis and Avery (1972).

Considerably more attention has been given to the effects of unequal dispersions on the classification procedures and results. It can be shown that the equality of the dispersions yields the standard linear classification rules. Unequal dispersions imply that a quadratic rule should be used. Gilbert (1969) has investigated and compared the effects on classification error rates and conditional probabilities if a linear rule is used (assuming equal dispersions) when, in fact, the dispersions are unequal. Only the two group case with known parameters was examined. The results indicate that significant differences can occur which are directly related to the differences in the dispersions, the number of variables and the separation among the groups. Agreement between the two procedures declines as the differences between the disperions and the number of variables increase. The further apart the groups are for given dispersions, the less important are the differences between the linear and quadratic results. Gilbert (1969) also provides tables which can serve as a useful guide in judging how the linear and quadratic rules may deviate for problems involving groups with particular parameter characteristics.

In a more recent study, Marks and Dunn (1974), investigated the performance in the two-group case of Fisher's linear discriminant function, the best linear function of Clunies-Ross and Riffenburgh (1960), (1960), and Anderson and Bahadar (1962), and the quadratic function. Sample estimates of population parameters were used and the overall probability of misclassification was used as the performance criterion. The authors conclude that for large samples the quadratic procedures performed better, the closer the groups were to each other, and as the number of variables increased. For small samples, they indicated that the quadratic performed worse for small numbers of variables and reasonably similar dispersion matrices, and this performance deteriorated further as the number of variables increased. However, as the dispersions became more dissimilar, the quadratic rules dominated. These results seem quite reasonable when one considers what is involved in constructing the sample functions and classification rules. For the two-group, m variable linear case $2m$ variable means and m^2 elements of the pooled within groups dispersion matrix must be estimated. This represents a total of $m(2+m)$ parameters. In the quadratic case, $2m(1+m)$ parameters are involved which is nearly twice the number as in the linear case. For fixed but small samples, the resulting degrees of freedom of the pooled within groups dispersion matrix used in the linear rules are substantially larger than for the individual group dispersions used in the quadratic rules. When the dispersions are similar, this leads to relatively more efficient linear classification results which are not swamped by the effects of averaging the group dispersions. However, as the differences in the dispersions increase, the efficiency of the linear parameter estimates begins to become dominated by the effects of the unequal dispersions. Marks and Dunn (1974) provide some tables indicating at what points the trade-offs between sample size and inequality of the dispersions occur. However, their tables are of limited practical usefulness because they investigated only a few specific types of dispersion inequality. It is not clear how general their conclusions are.

A number of the applied studies using quadratic classification techniques which have recently appeared in the literature tend to support the tentative observations

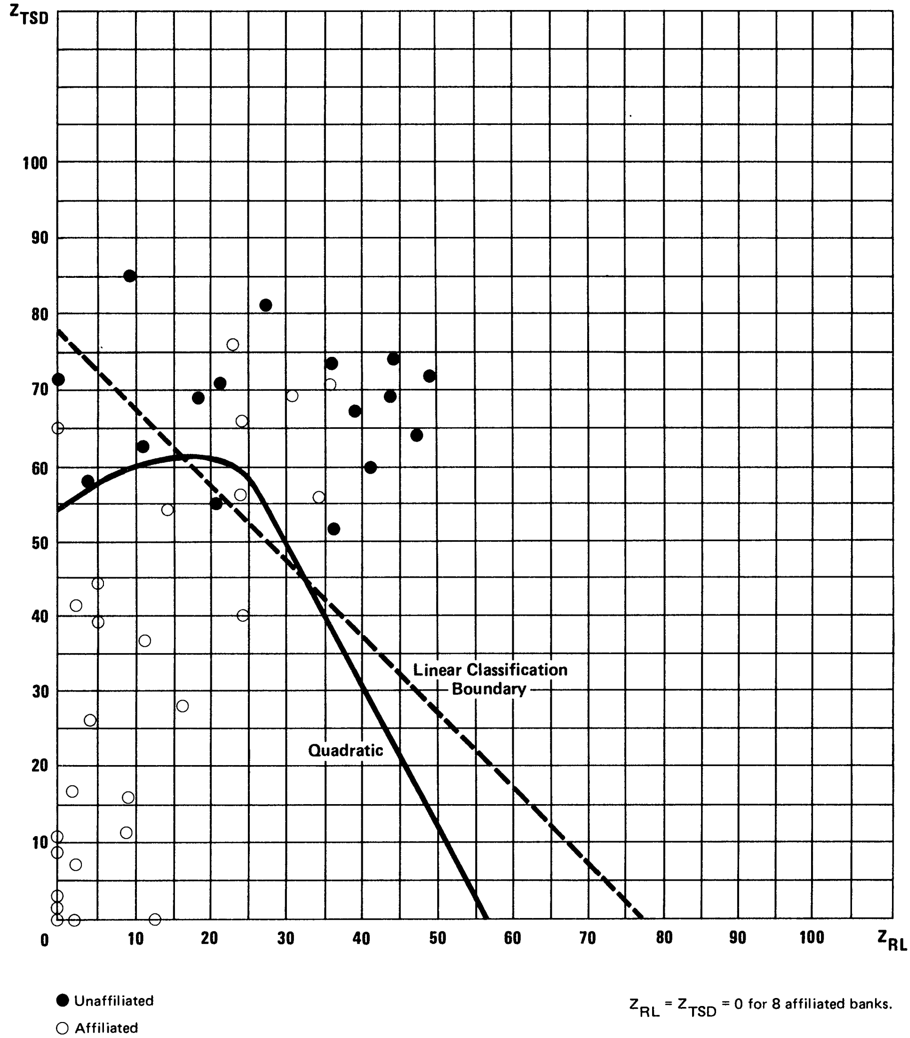

FIGURE 1. JOINT DISTRIBUTION OF Z_{TSD} AND Z_{RL}.

of Gilbert (1969).[11] For examples, in his classic three-group Iris problem, Fisher (1936) used linear procedures. Recently, Eisenbeis and Avery (1972) have shown that although the hypothesis of the equality of the dispersion matrices was rejected beyond any measurable level of significance, the use of quadratic classification rules, as shown in Table 1, yielded identical results to those using the linear rules. The reason for this becomes quite clear when the estimates of the group overlaps are examined in Table 2.[12] The group means are so far apart that there is almost no

11. See Gilbert (1975), Gilbert (1974), Eisenbeis and Avery (1972), Eisenbeis and McCall (1971), Bates (1973), Lane (1972), Eisenbeis and Murphy (1974), Sinkey (1975), for recent examples of the use of quadratic procedures.

12. See Cooley and Lohnes (1962) or Eisenbeis and Avery (1972) for explanation of the derivation of these tables.

significant overlap among the groups.[13] In a study of interlocking ownership and directorates among mutual savings banks and commercial banks in New Hampshire by Eisenbeis and McCall (1971), there was considerable overlap between the groups. In this two-group, two-variable problem, quadratic procedures significantly improved the classification results over the linear results as is shown in Table 3. Figure 1 shows the plots of the observations in (the two-variable) in test space with the commercial bank real-estate-loan ratio being plotted on the x axis and the time-and-savings-deposit ratio on the y axis. The linear and quadratic classification boundaries are shown as well. The variance of the real-estate-loan-ratio was greater in the unaffiliated group than in the affiliated group while the opposite held for the time-and-savings-deposit ratio. Equally important was the covariance which was negative in the unaffiliated group and positive in the affiliated group. The linear procedure, of course, averaged out these fairly diverse dispersions, whereas the quadratic classification rules used this additional information resulting in the improvement in the classification of the unaffiliated group.

TABLE 1

CLASSIFICATION RESULTS—IRIS DATA

		Linear Classification Results Predicated Groups[1]			Quadratic Classification Results Predicted Groups[1]		
Actual Groups	Total	Iris Setosa	Iris Versicolor	Iris Virginica	Iris Setosa	Iris Versicolor	Iris Virginica
Iris Setosa	50 (100.00)	50 (100.00)	0 (0.00)	0 (0.00)	50 (100.00)	0 (0.00)	0 (0.00)
Iris Versicolor	50 (100.00)	0 (0.00)	48 (96.00)	2 (4.00)	0 (0.00)	48 (96.00)	2 (4.00)
Iris Virginica	50 (100.00)	0 (0.00)	1 (2.00)	49 (98.00)	0 (0.00)	1 (2.00)	49 (98.00)

TABLE 2

EXPECTED GROUP OVER-LAP IRIS DATA*

	Percent Overlap Assuming Unequal Dispersions			Percent Overlap Assuming Equal Dispersions		
Equal Dispersions Groups	Iris Setosa	Iris Versicolor	Iris Virginica	Iris Setosa	Iris Versicolor	Iris Virginica
Iris Setosa	100.0000	$.2058\times 10^{-18}$	$.1927\times 10^{-32}$	100.0000	$.1401\times 10^{-15}$	$.1013\times 10^{-34}$
Iris Versicolor	$.1150\times 10^{-65}$	100.0000	.7830	$.1401\times 10^{-15}$	100.0000	1767
Iris Virginica	0.0**	.1311	100.0000	$.1013\times 10^{-34}$	.1767	100.0000

** less than 10^{-76}

* The table is read as follows: the means of the row group lie closer to the means of the column group than x% of the column group's observations.

13. A similar result was found by Altman who reestimated his original bankruptcy model Altman (1968) using quadratic procedures and the results were virtually identical.

TABLE 3

CLASSIFICATION RESULTS AND RULES—NEW HAMPSHIRE BANK DATA

Actual Groups	Total	Linear Classification Results Predicted Groups[1] Unaffiliated	Affiliated	Quadratic Classification Results Predicted Groups[1] Unaffiliated	Affiliated
Unaffiliated	16 (100.00)	12 (75.00)	4 (25.00)	15 (93.75)	1 (6.25)
Affiliated	33 (100.00)	6 (18.18)	27 (81.82)	6 (18.18)	27 (81.82)

Related to the work of Gilbert (1968) who investigated the performance of the linear function when all the variables were discrete, it can be shown that if a dummy or dichotomous variable is included in the variable set, the hypothesis of equal dispersions will most likely always be rejected. This again implies that quadratic and not linear classification procedures should be used.[14] For example, consider a Bernoulli variable X that takes on only two values $\{0,1\}$ with the probability $P(X=1)=p$ and $P(X=0)=1-p=q$. Then the mean of X in group 1 is $\mu_{x,1}=p_1$ and the variance is $\sigma^2_{x,1}=p_1(1-p_1)=p_1\cdot q_1$. For linear discriminant analysis to be appropriate in any problem including the variable X, the variance elements in the group dispersion matrices for X must be equal. That is, $\sigma^2_{x,1}=\sigma^2_{x,2}$. But this will hold only if (1) the mean in group one equals the mean in group two (i.e. $p_1=p_2$) in which case X is not an important discriminator in itself, or (2) if $q_2=p_1$ which represents a very limiting case.[15] Otherwise, if $p_1\neq p_2$ or $p_1\neq q_2$, then $\sigma^2_{x,1}\neq\sigma^2_{x,2}$. Therefore, if these two variance elements of the dispersions are unequal, then $\sum_1\neq\sum_2$, and quadratic rather than linear classification should be employed.

A priori there is little reason to believe that any one application area is likely to be more susceptible to this problem than any other, except to the extent that categorical variables may arise more frequently in certain types of problems than in others. The available evidence does indicate, however, that rejection of the hypothesis of equal group dispersions may have a significant and undesirable impact on the test for the equality of group means. More importantly, depending upon the sample sizes, number of variables, and differences in the dispersions, use of linear classification rules when quadratic rules are indicated may have drastic effects on the classification results. Logically then, the test for the equality of the dispersion matrices should precede both the test for the equality of group means and the estimation of classification errors.[16]

IV. INTERPRETATION OF THE SIGNIFICANCE OF INDIVIDUAL VARIABLES

One of the most widely misunderstood aspects of discriminant analysis relates to the problem of determining the relative importance of individual variables. Unlike

14. Note that Gilbert (1968) used only linear instead of quadratic procedures. Hence, her results reflect a mixture of problems of violation of not only the normality assumption but also the equality of dispersion assumption as well.

15. See Kshirsagar (1972) for a discussion of discrimination when there are zero mean differences of which this is a special case.

16. See Cooley and Lohnes (1962), (1971), Eisenbeis and Avery (1972), Box (1949) for the tests.

the coefficients in the classical linear regression model, the discriminant function coefficients are not unique; only their ratios are.[17] Therefore, it is not possible, nor does it make any sense to test, as is the case with regression analysis, whether a particular discriminant function coefficient is equal to zero or any other value. That is, there is no test for the absolute value of a particular variable. It is this aspect of discriminant analysis that may be more upsetting to economists than to others. It seems to be the nature of the behavioral hypotheses generated in economics and finance that they require that the influence of specific variables be isolated and quantified in a cardinal sense. Regression analysis seems particularly well suited for such problems, since it does allow one to test, *ceteris paribus*, whether specific coefficients are significantly different from a particular value.

A number of methods have been proposed in discriminant analysis which attempt to determine the relative importance of individual variables.[18] Five such methods were considered by Eisenbeis, Gilbert, and Avery (1973). These were to rank variables on the basis of (1) their univariate F-statistics, (2) their scaled discriminant function coefficients which were weighted by the appropriate diagonal elements of the pooled within groups deviation sums of squares matix, (3) stepwise forward methods based on the contribution to the multivarite F-statistic, (4) stepwise backward methods as in (3), and (5) a conditional deletion method which removed each variable in turn from the m-variable set, with replacement, and ordered variables according to the resulting reduction in overall discriminatory power as measured by the $(m-1)$ variable F-test.[19] Finally, a sixth method has been suggested in Mosteller and Wallace (1963) and Joy and Tollefson (1975) which weights each pairwise test space coefficient by the difference in the group means divided by the differences in the mean discriminant scores.[20] It represents the contribution of the ith variable to the Mahalanobis distance between the group means.

The problem with the first two methods rests with the fact that the variables are treated independently. Cochran (1964) has shown, however that seemingly insignificant or unimportant variables on a univariate basis may be very important when combined with other variables. In fact, he concluded that any negative correlation and extremely high positive correlations increase the discriminatory owner of a variable set while moderate or low positive correlations may not help much if at all. Some authors in applying discriminant analysis, such as Edmister (1972), Pinches and Mingo (1973), and Zumwalt (1975), excluded highly correlated variables because of their belief that "multicollinearity" was harmful. In fact, multicollinearity is a sample property that is largely an irrelevant concern in discriminant analysis except where the correlations are such that it is no longer possible to invert the dispersions matrices.[21]

17. See Kshirsagar (1972), Eisenbeis and Avery (1972), Ladd (1966), Anderson (1958) or Lachenbruch (1975) for discussions of this point.

18. See Cooley and Lohnes (1962), (1971), Eisenbeis and Avery (1972), Joy and Tollefson (1975), and Mosteller and Wallace (1963).

19. Weiner and Dunn (1966) examine four methods including (1), (2), and (3) above.

20. In the two-group case, where b_i is the coefficient of the ith variable, the measure of relative percentage contribution is $[b_i(\bar{X}_{i1}-\bar{X}_{i2})/(\bar{X}_1-\bar{X}_2)'B]$.

21. Eisenbeis, Gilbert, and Avery (1973), Altman (1968), and Altman and Katz (1975) also provide

The three other methods (3), (4), and (5) listed above are all conditional methods which do take into account correlations among the variables. In the case of the stepwise forward and backward methods, the relative contribution of a given variable is measured against an increasing (decreasing) number of variables. For example, in the stepwise forward method, the second variable to enter is the second most important variable, given that the first is already included. The conditional deletion method would seem to have the greatest appeal since the relative importance of each variable is conditional based on the inclusion of all other variables. Interestingly, Kshirsagar (1972) shows that this conditional deletion method has more than intuitive appeal. It was stated previously that while discriminant function coefficients are not unique, their ratios are; hence, their coefficients are unique up to a factor of proportionality.[22] Therefore, while it is not appropriate to test whether a given coefficient is significantly different from either zero or some other constant α, it does make sense to test whether the ratio of two coefficients is equal to α, i.e.,

$$H_0: \frac{b_i}{b_j} = \alpha.$$

When $\alpha = 0$, Kshirsager (1972) demonstrates that this is equivalent to the testing whether the addition of a given variable to an m-set significantly increases the overall discriminatory power of the set.[23] This is the essence of the conditional deletion method which ranks the m variables according to the discriminatory power each variable adds to the overall set given that the other $m-1$ variables have already been included. Kshirsagar (1972) also shows in the two-group case when a dichotomous dependent variable is being regressed on the independent variables to take into account the computational equivalence between regression and discriminant analysis [see Eisenbeis and Avery (1972), Anderson (1958), Ladd (1966)], that regression coefficient t-tests can also be used to test this hypothesis. Hence, they may be interpreted identically to the conditional deletion tests for the significance of the contribution of the variable given that the others have been included. The t-tests do not indicate whether a particular "coefficient" itself is zero because the coefficients are not unique.[24] Meyer and Pifer (1970) and others fall into the trap of assuming that the regression t-statistics are valid for determining the significance of individual coefficients.

The sixth ranking method proposed by Mosteller and Wallace (1963), which measures the gross contribution of a variable to the overall Mahalanobis distance between the group means, offers little to recommend it. First, the weights are difficult if not impossible to interpret because they (1) are signed (e.g. plus or minus), (2) can be greater than one, and (3) do not sum to one. Second, the method is not readily generalizable to more than two groups.

A final comment is that all the methods for investigating the relative importance

further empirical evidence that seemingly unimportant variables on a univariate basis may be important discriminators in a multivariate context.

22. See Eisenbeis and Avery (1972) for Kshirsagar (1972) for proofs.

23. If $\hat{H}_0: \alpha = 0$ is true, then it is true for all b_j, $j = 1_i, \ldots, m$ and $j \neq i$.

24. The test does indicate whether the coefficient is zero "relative to all the rest."

of variables that have currently been examined have assumed equal dispersions.[25] Rejection of that hypothesis implies that these methods are subject to the same limitations as the tests for the significance of the difference in group means.

V. Dimension Reduction

The two principal ways for reducing dimensionality in discriminant analysis are to eliminate (1) those variables or (2) those discriminant functions (in the reduced space K-group case) that do not contribute significantly to the overall ability to discriminate among groups. Objective assessment of the usefulness of particular dimension reducing procedures rests upon a knowledge of the trade-offs involved between the alternative reducing criteria and an understanding of how the reducing criteria relate to the research goals in performing a discriminant analysis. This can be particularly important for problems in business, economics, and finance where it is often possible to generate a large number of variables which need to be pared down to some manageable size.

To date, the dimension reducing methods used have focused solely on determining whether a variable or function contributed significantly to the Wilk's lambda or related statistics used in testing hypotheses about the equality of group means. (See Eisenbeis and Avery (1972), Eisenbeis, Gilbert, and Avery (1973) or Weiner and Dunn (1966).) Furthermore, they have assumed equal group dispersions.[26] The methods used have been based upon the univariate significance tests and the various stepwise procedures mentioned in Section IV. These are appropriate if the research goal is to maximize the separation among groups while minimizing the number of variables or functions used. If, however, the goal is to construct a classification scheme, then use of the aforementioned methods may not leave the classification results unaffected, even if seemingly insignificant variables or dimensions are eliminated. Eisenbeis and Avery (1973) have examined in an heuristic manner the relationship between the significance tests for the equality of group means and the problem of investigating group overlap through classification methods.[27] They argue that the existence of statistically significant differences among group means, especially when the sample sizes are large, does not convey much if any useful information about the ability to construct a successful classification scheme. They suggest, following Cooley and Lohnes (1962), that chi-square methods of describing group overlap be used.[28]

25. Except for the scaled-discriminant function method, the procedures mentioned in Section II for adjusting the test statistics could be directly applied to modify the statistics used to investigate the relative importance of variables.

26. For this reason, use of dimension reducing procedures when the group dispersions are unequal is subject to the same limitations noted in Section II pertaining to the significance tests of the equality of group means.

27. Joy and Tollefson (1975), Frank, Massy and Morrison (1965), and Morrison (1967) also touch upon this point.

28. Using the following chi-square measure $\chi^2_{m,gh} = (\bar{X}_g - \bar{X}_h)' S_w^{-1} (\bar{X}_g - \bar{X}_h)$ which is distributed with m degrees of freedom, the resultant significance level α indicates that the means of group g lie closer to the means of group h than $\alpha \cdot 100\%$ of the members of group h. If the group dispersions are equal, then the pooled within group dispersion matrix, S_w, is used in the χ^2, otherwise the individual dispersion matrix, S_h, is used.

Eisenbeis and Avery (1972) compare the classification results using one of two discriminant functions based upon a three group, twelve variable, security analysis program originally studies by Smith (1965). The analysis of the significance of the roots is shown in Table 4 and the one and two linear discriminant function classification results are shown in Table 5. The data clearly indicate that although the second root was statistically insignificant according to the usual standards, the omission of the second dimension significantly affected the classification results in two ways.

TABLE 4

SIGNIFICANCE OF THE ROOTS IN THE SECURITY ANALYSIS PROBLEM

Root Number	Value of Root	Percent of Total Variance Accounted for by Root	Value of Chi Square Statistic	Degrees of Freedom	Significance Level (Percent)
1	3.945974	83.462	40.76363	13	0.010407
2	0.7818336	16.537	14.72988	11	19.520000

TABLE 5

COMPARISON OF ONE AND TWO DIMENSIONAL LINEAR CLASSIFICATION RESULTS—SECURITY ANALYSIS PROBLEM

		Predicted Groups[1,2]					
Actual Groups		One Eigenvector Used			Two Eigenvectors Used		
Group	Total	Investment Group	Trading Group	Speculative Group	Investment Group	Trading Group	Speculative Group
Investment	11 (100.000)	7 (63.636)	4 (36.364)	0 (0.000)	11 (100.000)	0 (0.000)	0 (0.000)
Trading	11 (100.000)	1 (9.091)	10 (90.909)	0 (0.000)	0 (0.000)	9 (81.818)	2 (18.182)
Speculative	11 (100.000)	0 (0.000)	0 (0.000)	11 (100.000)	0 (0.000)	2 (18.182)	9 (81.818)

1. Percentages are shown in Parentheses.
2. Equal *A Priori* probabilities were assumed.

First, the overall accuracy fell off from 87.9% to 84.8%. More importantly, there were radical changes in the individual group error rates. Thus, if classification was a primary goal, dropping of even an insignificant dimension or variable may be undesirable.

Further work is needed in exploring the links between the significant tests and classification results. The available evidence suggests, however, that it may be unwise to drop dimensions or variables without first exploring in more detail what the possible effects may be. If classification accuracy is a primary goal, then the criterion for keeping or deleting variables and dimensions should be related to the overall efficiency of the classfication results. Therefore, the results using all variables should be compared with those based upon various subsets of variables.

It would seem inappropriate in such instances to discard variables such as Pinches and Mingo (1973), Edmister (1972), Meyer and Pifer (1970), Orgler (1970), Zumwalt (1975), and Bates (1973) did without first examining the overall classification results to determine what the effects or "costs" of dimension reduction really are. The implication is that concern for dimension reduction should "follow" and not precede the development and validation of alternative classification schemes as has been the case in most of the applied literature.

VI. The Definitions of Groups

Discriminant analysis procedures assume that the groups being investigated are discrete and identifiable. However, numerous examples occur in the literature which either violate this assumption or deal with classification schemes which limit the practical usefulness of the empirical results.

Perhaps the most extreme case occurs when an inherently continuous variable is segmented and used as a basis to form groups. Walter (1959) arbitrarily divided firms into quartiles based upon the distribution of their earnings price ratios and then used discriminant analysis to distinguish between firms in the first and fourth quartiles. Haslem and Longbrake (1971) constructed similar groups for banks based upon the distribution of their profitability.[29] There are four basic problems with this grouping criterion. First, the groups are arbitrary and not truly distinct. There is certainly nothing magical about quartiles as compared, for example, with deciles as the grouping criterion. In fact, an infinite number of different sets of groups could be formed and a given observation could be a member of each of several groups. Hence, neither the groups nor group membership were really distinct. Second, for classification purposes, two other groups, the second and third quartiles, were omitted. The discriminant analysis classification rules could only compare whether a given observation appeared relatively more like a first or fourth quartile firm and make assignments accordingly. The possibility of belonging to the second or third quartile would be excluded. Firms from that portion of the population would be forced into the first or second quartiles. Third, attempts to assess error rates for the two group model using samples drawn only from the first and fourth quartiles of the population are not particularly meaningful or realistic since in order to select the sample, one must already know which firms are from the first and fourth quartiles, but this is precisely what one is trying to predict. At the very least, a four group model should be estimated with each quartile representing a separate group. Finally, such problems do not really lend themselves to predictive discriminant analysis because they involve forming groups on the basis of a variable or factor that is, in fact, observable at the same time that the "independent" or "explanatory" variables are. For example, if one observes a firm's level of profits, it can be assigned to the proper quartile without reference to the other variables. Hence, there is no need to use other variables to predict in which quartile the criterion variable should be expected to lie.

As a practical matter, the only time it really makes sense to form groups based

29. Other studies with similar problems are Klemkosky and Petty (1973), Norgaard and Norgaard (1974), and Shick and Verbrugge (1975).

upon the distribution of a particular variable is if natural breaks or discontinuities appear. Otherwise, segmenting an inherently continuous variable effectively discards information about the relationships between the independent or explanatory variables and the grouping criterion variable that might more appropriately be captured in a regression or other causal model. In most instances, regression and not discriminant analysis is the more appropriate technique for such problems.

The problem of omitting groups or portions of populations has occurred to a greater or lesser degree in other studies such as Altman (1968), and similarly in Edmister (1972) where only relatively small firms were included. Joy and Tollefson (1975) have given particular attention to this problem. They emphasize that in order to obtain useful and interpretable results, particularly with respect to classification, it is important that the populations sampled to estimate discriminant functions and classification error rates correspond to the populations generating the new observation to which the model is to be applied. Divergence between the two population sets has been a common failing with certain types of application. For example, in the development of credit scoring systems, the objective is usually to develop models to discriminate between those *new* loan applicants who are and are not likely to default on their loans. Characteristically, however, loan performance data with which to estimate the discriminant function are not available on the population of *all* applicants, but rather on only the subset of applicants that were originally granted loans. Omitted are data on that portion of the population who applied for loans but did not receive them, some of which may or may not have defaulted. Technically, a model constructed in such a fashion should not be applied to new loans which are drawn from a broader population. Rather, a more appropriate use is as an internal loan review device since it was developed from, and the estimates of error rates are applicable to, the population of loans that were granted. The practitioner should keep in mind that the estimates of the error rates do not apply to new loan applicants. As a practical matter, however, if one is willing to assume (1) that the initial loan review process screened out the "bad" loan tail of the distribution of potential default loans and (2) relatively few of the good loans have been eliminated, it is likely that the resultant or screened populations of good and bad loans would have (1) closer centroids and (2) probably greater overlap than the true population. Hence, the estimated error rates from the "screened" populations are likely to represent conservative estimates of the true error rates.

A more extreme case of group identification occurred in the paper by Adelman and Morris (1968). They arbitrarily grouped countries into three groups according to their prospects for economic development. Eight criteria were used to construct the groups, and then discriminant functions were estimated using some of these same grouping variables as independent variables. The original samples were reclassified and group assignments were made based upon that classification. Finally, a second discriminant analysis was performed. Not surprisingly, the final results were quite acceptable and impressive. Of course, the authors were not really performing a discriminant analysis, but rather a crude cluster analysis was being employed to group the countries based upon their similarities.

Another common grouping problem in business and finance is to take arbitrarily defined groups such as bond classes as in Pogue and Soldofsky (1969), Pinches and Mingo (1973), Carleton and Lerner (1970), and Altman and Katz (1975), bank examiner capital adequacy ratings as in Dince and Fortson (1972), or problem bank groups as in Stuhr and Van Wicklen (1974), and Sinkey (1975). Here, the group definitions appear discrete as far as the researcher is concerned, in the sense that someone else has already made the grouping decision. However, the possibility still exists that the original assignments may have been in error. In such cases, the best that the researcher can hope to do is to replicate or simulate the original grouping decision. The effect of such grouping schemes is to introduce an additional source of error into the assessment of any classification results. That is, one can never be sure whether an observation has been misclassified (or correctly classified) because of overlaps among the groups or errors in the original assignment by an examiner, bond rater, etc. Lachenbruch (1966) has shown in the equal dispersion, two-group case that when each of the group assignment errors are random and equal, there is no effect on the classification errors. Otherwise, as long as the assignment errors are (1) small or (2) not too different from each other, there will not be a significant effect on the estimated error rates for both large and small samples. The assumption of random assignment errors is not particularly realistic, so Lachenbruch (1974) has also investigated two models with nonrandom assignment errors. As might be expected, the estimated error rates using the more common methods are biased and unreliable.

In summary, the best problems for discriminant analysis appear to be those in which the group definitions are distinct and non-overlapping as in the Fisher Iris problem. Every effort should be made to avoid arbitrary groupings such as those employed by Walter (1959) or Adelman and Morris (1968).

VII. THE SELECTION OF THE APPROPRIATE *A Priori* PROBABILITIES AND SAMPLES

The standard discriminant analysis classification rules incorporate *a priori* probabilities to account for the relative occurrence of observations in different populations in the universe and costs to adjust for the fact that some classification errors may be more serious (e.g. costly) than others. The importance of the *a priori* probabilities and/or costs of misclassification have been grossly overlooked, partly due to the lack of computer programs that easily incorporate them into the rules. As a result, most authors, at least up until 1973–74, ignored error costs and simply assumed that group membership was equally likely.

It can be easily shown by example that unless the groups are equally likely, the estimated error rates assuming equal *priors* might bear little relationship to what one might expect in the population. Consider the example in Table 6 taken from a recent business loan study which shows six group quadratic classification results assuming equal *a priori* probabilities and estimated population *a priori* probabilities. The overall expected probability of misclassification was 51.1 percent for the equal *a priori* probability case as compared with 45.5 percent for the unequal *a priori* probability case. More important, however, than the overall improvement between the unequal and equal *a priori* cases was the fact that some of the individual group error rates shifted radically. For example, 88% of group 2 was correctly assigned assuming equal *priors* whereas 100% were misclassified in the unequal *priors* case. This illustrates that one can be drastically misled about the effectiveness of his classification results if the correct *priors* are not employed.

In the absence of knowledge of the population *priors*, it has become common practice to use sample proportions as estimates. This is appropriate provided the pooled data represent a random sample from the population. Otherwise, of course, the resulting classifications would only minimize the classification errors in the sample rather than providing evidence on the population error rates. It is for this reason that Meyer and Pifer (1970) in their trial and error method found that a 50-50 rule minimize their classification errors. They simply discovered that they had equal members of observations in each sample.

Using independent estimates of the population *priors* relieves the researcher of the burden of worrying that the pooled sample represents a random sample from the population. It does not, however, relieve him of the problem of assuming that his samples are representative of the groups being investigated. Alternative sampling methods such as the paired sample methods of Meyer and Pifer (1970) or Sinkey (1975) are representative of non-random schemes that have been used. Nonrandom methods where certain factors are controlled (such as bank size, regulatory statutes, number of branches, and location in the study by Sinkey (1975) are appropriate for investigating the importance of certain variables but not for estimating classification error rates. To the extent that the control variables are not independent of the other included variables one would expect the dispersions and group means in the control sample to be different from a random sample drawn from the publication. If this is true, then the resulting classification rules would be different, and even use of appropriate *priors* would not yield valid estimates of the population error rates. An alternative method for such a study would have been to draw random samples from both groups, include the control variables as the α variables in the conditional Wilk's lambda statistics shown in Eisenbeis and Avery (1972) or Eisenbeis, Gilbert, and Avery (1973), and then examine the other variables.

Attempts to employ discriminant analysis in a time series context raise additional questions about the appropriate selection of *a priori* probabilities which have not yet been addressed in either the applied or theoretical literature. The first arises in studies where observations from a single period in time are used to form classification rules and make predictions about group membership in a future time period or periods. To the extent that the relative expected occurrences of the groups in the population may vary from period to period, it is not at all obvious what the population priors should be or how they should be estimated. For example, in the problem bank study of Sinkey (1975), the number of problem banks has varied quite significantly from year to year from a low of around 150 in some years to more than 349 in early 1976. Under such circumstances, it is not clear whether it is more appropriate to use the relative group frequencies from a given year as estimates of the *a priori* probabilities, or to attempt to use some average of past frequencies. In the absence of theoretical guidance, it would seem that the research design and goals should determine the method to be used. Clearly, in the problem bank study, the expected frequencies of problem and nonproblem banks are not

TABLE 6

EQUAL AND UNEQUAL *A Priori* QUADRATIC CLASSIFICATION RESULTS—BUSINESS LOAN STUDY

Actual Group Number	Total	Equal *A Priori* Probability Results: Predicted Groups 1	2	3	4	5	6	Unequal *A Priori* Probability Results: Predicted Groups 1	2	3	4	5	6	Estimated *A Priori* Probabilities
1	73 (100.00)	72 (98.63)					1 (1.37)	70 (95.89)					3 (4.11)	.04
2	125 (100.00)		110 (88.00)		5 (4.00)		10 (8.00)				125 (100.00)			.07
3	81 (100.00)	17 (20.99)		46 (56.79)			18 (22.22)	14 (17.28)		46 (56.79)			21 (25.93)	.05
4	676 (100.00)		213 (31.51)		73 (10.00)	390 (57.69)					263 (38.91)	413 (61.09)		.38
5	699 (100.00)		111 (15.88)		59 (8.44)	529 (75.68)			2 (.29)		138 (19.74)	559 (79.97)		.39
6	115 (100.00)	33 (28.70)	12 (10.43)	25 (21.74)	9 (7.83)		36 (31.30)	32 (27.83)		21 (18.26)	29 (25.21)		33 (28.70)	.07

independent of the state of the economy. During unstable times, then, a simple average of past years' frequencies might tend to understate the expected frequencies of the problem group.

A second aspect of the time series problem can be found in studies such as Pinches and Mingo (1973), Altman (1968), or Gilbert (1974) where the data on the groups are obtained by pooling observations from different time periods. This is usually done when one of the groups occurs relatively infrequently and pooling is necessary to get a large enough sample with which to work. Again, it is not clear what the appropriate *a priori* probabilities are or how they should be estimated. In such instances, gross sample proportions are probably inappropriate. Selection and estimation of the priors should be geared to the type of classification statements that the researcher wishes to make. For example, if a one period classification is to be made, then it would seem reasonable to use an average of the relative frequencies over several time periods to estimate the *priors*. On the other hand, if predictions are to be made over several time periods such as Gilbert (1974) did, then it may be more appropriate to pool data over the same length of time to estimate frequencies.

The standard discriminant analysis classification rules have been derived from minimizing loss functions of the form (for the two group case)

$$M = P(1|2)\pi_2 + P(2|1)\pi_1$$

$$L = C(1|2)\cdot P(1|2)\cdot\pi_2 + C(2|1)\cdot P(2|1)\cdot\pi_1$$

which take into account *a priori* probabilities (e.g. π_i's) and costs of misclassification (e.g. $C(g|h)$).[30]

For example, the linear form of the two group rules is[31]

Assign to group 1 if

$$X'B - \tfrac{1}{2}(\bar{X}_1 + \bar{X}_2)'B \leqslant \ln\frac{C(1|2)\pi_2}{C(2|1)\pi_1}$$

and the quadratic form is

Assign to group 1 if

$$X'(\Sigma_1^{-1} - \Sigma_2^{-1})X - 2(\bar{X}_1'\Sigma_1^{-1} - \bar{X}_2'\Sigma_2^{-1})X + \bar{X}_1'\Sigma_1^{-1}\bar{X}_1 - \bar{X}_2'\Sigma_2^{-1}\bar{X}_2$$
$$\leqslant \ln|\Sigma_2\cdot\Sigma_1^{-1}| - 2\ln\frac{C(1|2)\pi_2}{C(2|1)\pi_1}$$

However, little or no attempt has been made to explicitly incorporate costs into the models that have been developed to date. Dince and Fortson (1972) rationalized *ex post* the use of their cut off point based upon a poll of bank supervisors from whom they obtained subjective cost estimates. Finally, both Sinkey (1975), in

30. $C(g|h)$ is the cost of misclassifying an observation as a member of group g given that it came from group h. $P(g|h)$ is the conditional probability of misclassification.

31. See Eisenbeis and Avery (1972), Cooley and Lohnes (1962), (1971), or Anderson (1958) for derivations and formulations of these rules.

his study of problem banks, and Meyer and Pifer (1970) in developing their model to identify potential failing banks did discuss the development of rules that would maximize the returns from applying their screening model. However, these discussions were only general in nature. In neither study did they attempt to explicitly specify or estimate the model they proposed nor did they apply it in the empirical portion of their paper. As such, we must conclude that the general procedures for incorporating costs of misclassification into the classification procedures have been developed, but they have not yet been explored in the applied literature.

VIII. Assessment of Classification Error Rates

If one of the main purposes in conducting a discriminant analysis is to construct a classfication scheme, then a central problem involves assessing the performance of the estimated rules. It has been well publicized in the economics and finance literature by now (due in part to a paper by Frank, Massy, and Morrison (1965)) that reclassification of the original sample used in constructing the classification rules as a means to estimate expected error rates leads to a biased and overly optimistic prediction of how well rules would perform in the population. A number of alternative methods have been suggested, and evaluated, to estimate classification errors.[32] The alternatives are basically of three types: those using samples to estimate error rates, those using the assumption of normality, and those using jackknife procedures. Each of several such methods is briefly described in Table 7 and the limitations and strengths of each are noted.

Both Lachenbruch and Mickey (1968) and Cochran (1968) have evaluated some or all of these methods. As might be expected, the original sample method (1) and D method (4) performed poorly. The U or $\bar{U}$ methods appear to be the best for small samples. Overall, the OS method seemed the best. However, it should be noted that the U (9), original sample (1), and holdout (2) methods are most easily generalized to more than two group problems as well as the unequal dispersion cases.[33] In this respect then, the U (9) method would appear to be superior based upon current evidence, especially when coupled with its applicability to small samples and large dimension problems.[34] In fact, it is interesting to note that Lackenbruch and Mickey (1968) conclude that the holdout method has no clear superiority over the U method.

IX. Summary and Conclusions

This paper has discussed several of the more common problem areas appearing in the applied discriminant analysis literature. If one had to rank the problems

32. See Lachenbruch (1967), (1968), (1975), Lachenbruch and Mickey (1968), Frank, Massy, and Morrison (1965), Dunn (1971), Dunn and Vardy (1966), Sorum (1973), Hills (1966), Smith (1947), and Okamoto (1963).

33. Eisenbeis and Avery (1972) have programmed the U method for all cases including quadratic classification.

34. For an example of how sample sensitive and biased the original sample method is when compared with the U method even relatively large samples, see Eisenbeis and Murphy (1974). Their estimates of error rates between the original sample method and the U method in a two group, 12 variable problem increased more than 16 and 13 percentage points, respectively.

TABLE 7

ERROR ESTIMATION METHODS: TWO GROUP—EQUAL DISPERSION CASE

Method	Estimator	Properties and Other Comments
1. Original Sample Method	P_1 and P_2 are sample proportions of misclassified observations	Estimates are consistent but may be badly biased and overly optimistic for small samples in particular.
2. Holdout Method	Samples are split with one used to estimate functions which are then employed to classify the other (or holdout) sample and estimate P_1 and P_2	Estimates are consistent and unbiased but require large samples. Intuitively, the estimates may be less efficient than other methods below.
3. Population Method	When population parameters are known, i.e., $N_1(\mu_1, \Sigma)$ and $N_2(\mu_2, \Sigma)$, Δ in $\Phi(-1/2\Delta)$ is Mahalonobis' distance $\Delta^2 = (\mu_1 - \mu_2)'\Sigma^{-1}(\mu_1 - \mu_2)$	Errors are true errors in populations.
4. *D* Method	Estimator of Δ is D using sample estimates to form $D^2 = (\bar{X}_1 - \bar{X}_2)' S_W^{-1}(\bar{X}_1 - \bar{X}_2)$	Estimates of D are consistent but upward biased—Hence error estimates P_1 and P_2 are consistent but may be badly downward biased, particularly for small samples.
5. D^* Method	The D^* method simply substitutes an unbiased estimate of Δ for the biased estimate in 4 above. $D_2^* = (N_1 + N_2 - 3 - m/(N_1 + N_2 - 2))D^2 - (m(N_1 + N_2)/N_1 N_2)$	Estimates of P_1 and P_2 are unbiased and consistent, but if N_1 and N_2 are small relative to D^2 and m, then DS^2 may be negative which precludes its use.
6. The *DS* method	The *DS* method simply drops the last term in D^* such that $DS^2 = (N_1 + N_2 - 3 - m/(N_1 + N_2 - 2))D^2$	Estimates of P_1 and P_2 are consistent, slightly biased, and underestimates of the true rates.

7. The *O* Method	The estimate of $P_1=\Phi(-1/2D^*)$ + an expansion of Okamoto (1963) in terms of the parameters of group 1. $P_2=\Phi(-1/2D^*)$ + an expension in terms of the parameters of group 2	Allows separate estimates of P_1 and P_2. Where groups are close together and sample sizes are small, the method performs poorly. When D^2 is large, then *O* performs similar to *D* method because factors are proportional to $1/D^2$.
8. The *OS* Method	Substitute DS^2 for D^2 in 7 above.	Allows separate estimates of P_1 and P_2. Performs better than *O* method.
9. The "Lachenbruch" or *U* Method	The method holds out one observation at a time, estimates the discriminant functions based upon N_1+N_2-1 observations and classifies the held out observations. This is repeated until all observations are classified. Then the proposition of observations misclassified P_1 and P_2. Confidence intervals around P_1 and P_2 are given by $$\frac{2N_iP_i+Z^2\pm\left((2N_iP_i+Z^2)^2-4N_iP_i^2(N_i+Z^2)\right)^{1/2}}{2(N_i+Z^2)}$$ Where $Z=100\alpha/2$ percentile of the normal distributions	Method gives almost unbiased estimates of the confidence intervals. Moreover, it may be used to get around the sample size limitation associated with the Holdout method. Intuitively, one would expect the method to be more efficient than the hold out method. Kshirsagar feels the method may not be sensitive to the normality assumption. Eisenbeis and Avery (1972) have applied it to problems with unequal dispersion and more than two groups.
10. The $\overline{U}$ Method	Estimation of $P_j=\Phi(-D_1/S_1)$ and $P_1=\Phi(-D_2/S_2)$ where $\overline{D}_1=1/N_1$ times sum of the N_1 scores from group 1 in 9. $\overline{D}_2=1/N_2$ times sum of the N_2 scores from group 2 in 9.	Lachenbruch and Mickey (1968) conclude method performs quite well when compared to others using normal distribution. Method allows separate estimate of P_1 and P_2.

a. $P_1(1\,|\,\text{group } 2)=P_1$, $P_2(2\,|\,\text{group } 1)=P_2$ are the group error rates.

b. N_1 and N_2 are group sample sizes, and m is the number of variables.

c. In methods using normality assumption $P(1\,|\,\text{group } 2)=P_2(2\,|\,\text{group } 1)=\Phi(-1/2\,\Delta^2)$ where *a priori* probabilities are equal and Φ is the cumulative normal density functions.

according to severity of their affects on the usefulness of the analysis, it would seem that the problems related to classification are the most severe, with the issues surrounding the selection of the appropriate *a priori* probabilities being the most important followed in turn by the selection of the appropriate classification rules (linear vs. quadratic) and assessment of classification accuracy. In particular, the failure to relate the estimates of the *a priori* probabilities to the population *priors* by, for example, assuming equal *priors*, in fact limits the ability to make any meaningful inferences about the overall performance or accuracy of the classification scheme. Similarly, use of linear classification rules when the group dispersion matrices are equal nearly always results in an underassessment of the overall classification accuracy of the rules, moreover, there may be significant distortions in the individual group error rates.

Other problems such as nonnormality, the selection of subset variables and reducing dimensions, interpreting the significance of individual variables, are not as easy to remedy. Until further research is done, one must simply temper the conclusions reached by recognizing that the empirical results represent approximations that may be significantly biased in many cases.

There are several remaining problems that have either not been mentioned or only briefly touched upon. This would particularly include the important class of time series problems that frequently occur in business, economics, and finance. These relate to the application of discriminant analysis (1) to prediction one, two, or more time periods into the future as opposed to simply predicting the likelihood of an event occurring and (2) to samples of data that have been pooled across time periods.[35] Neither of these have yet been dealt with in the applied or theoretical literature.

In the first type of time series problem, the process by which an observation moves from one group to another might reasonably be expected to be a continuously evolving one. This suggests that the relationships among the relevant variables (e.g. means, variances and covariances) change over time. Hence, the appropriate model might be different, both in terms of parameters and variable composition, for each separate time period into the future for which prediction is to be made.

In the second type of time series problem, data from several time periods are often pooled to make predictions over the same span of time periods (e.g. Altman (1968)) or over future time periods (e.g. Pinches and Mingo (1973) or Gilbert (1974)). Joy and Tollefson (1975) have dealt with this problem heuristically in their discussion of Altman's (1968) work and they suggest that the predictive accuracy of his model could not be meaningfully verified by *ex post* classification of a sample selected from the same time period used to develop his model.[36] They incorrectly view the process of prediction only as making inferences into the future. Hence, they state that one should verify the predictive accuracy of a model *ex ante* using data drawn from outside the original sample period. However, if one is interested in only predicting whether an event will occur without reference to a particular

35. These two types of problems have often occurred together.

36. They state that successful *ex post* classification merely permits making inferences about the role of individual variables.

time period (which is usually the case with pooled samples such as Altman's (1973)) and one is willing to assume stationarity of the relationship among the variables over time, then intertemporal verification is clearly appropriate and meaningful. In this context, although Joy and Tollefson (1975) don't explicitly recognize it, divergence in *ex ante* and *ex post* classification results really constitute a crude test of the stationarity hypothesis.[37]

In addition to the time series problem, nothing has been said about linking discriminant analysis with other procedures such as factor analysis, Cooley and Lohnes (1971); principal components, Pinches and Mingo (1973); or regression analysis, Tatsuoka (1971). The latter may be particularly important since it relates to the investigation of the joint probability of group membership and relative performance as a member of a given group. Finally, it has not been uncommon to see some researchers make use of a so-called "doubtful" region in classification in order to avoid assigning an observation to any group, Rao (1952). Such a procedure essentially begs the assignment question given that the rules are already designed to minimize classification errors.

REFERENCES

Irma Adelman and Cynthia Taft Morris. "Performance Criteria for Evaluating Economic Development Potential: An Operational Approach," *Quarterly Journal of Economics*, 82, no. 2 (May 1968), 260–280.

Edward I. Altman. "Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy," *Journal of Finance*, 23, No. 4 (September 1968), 589–609.

———. "Predicting Railroad Bankruptcies in America," *The Bell Journal of Economics and Management Science*, Vol. 4, No. 1, Spring 1973.

——— and Steven Katz. "An Analysis of Bond Ratings in the Electric Public Utility Industry," following in *Relevant Research in Accounting*, Ed. by Michael Shiff and George Sirter, (New York, 1975).

T. W. Anderson. *An Introduction to Multivariate Statistical Analysis* (John Wiley and Sons: New York, 1958).

——— and R. R. Bahadur. "Classification into Two Multivariate Normal Distributions with Unequal Covariance Matrices," *The Annals of Mathematical Statistics*, 33 (June 1962).

Timothy Bates. "An Econometric Analysis of Lending to Black Businessmen," *Review of Economics and Statistics*, (August 1973).

G. E. P. Box. "A General Distribution Theory for a Class of Likelihood Criteria," *Biometrika*, 36 (1949), 317–346.

W. T. Carleton and E. M. Lerner. "Statistical Credit Scoring of Municipal Bonds," *Journal of Money, Credit, and Banking* (November 1969).

P. C. Chang and A. A. Afifi. "Classification Based on Dichotomous and Continuous Variables," *Journal of the American Statistical Association*, Vol. 69, No. 346, June, 1974.

C. W. Clunies-Ross and R. H. Riffenburgh. "Geometry and Linear Discrimination," *Biometrika*, 47 (June 1960).

———. "Linear Discriminant Analysis," *Pacific Science*, 14 (July 1960) 251–6.

W. G. Cochran. "Commentary on 'Estimation of Error Rates in Discriminant Analysis,'" *Technometrics*, Vol. 10, No. 1, February, 1968.

———. "On the Performance of the Linear Discriminant Function," *Technometrics*, 6 (May 1964), 179–190.

William W. Cooley and Paul R. Lohnes. *Multivariate Procedures for the Behavioral Sciences* (John Wiley and Sons: New York, 1962).

37. In the applied literature, Pinches and Mingo (1973) performed both cross and intertemporal classifications.

———. *Multivariate Data Analysis*, (John Wiley and Sons: New York, 1971).

R. B. D'Agostino. "Monte Carlo Power Comparison of the W' and D Tests of Normality for $n = 100$," *Communications in Statistics*, 1 (1973).

Robert R. Dince and James C. Fortson. "The Use of Discriminant Analysis to Predict the Capital Adequacy of Commercial Banks," *Journal of Bank Research*, Vol. 3, No. 1, Spring 1972.

Olive Jean Dunn. "Some Expected Values for Probabilities of Correct Classification in Discriminant Analysis," *Technometrics*, 13 (May 1971).

——— and Paul D. Vardy. "Probabilities of Correct Classification in Discriminant Analysis," *Biometrics*, 22, 1966.

R. O. Edmister. "An Empirical Test of Financial Ratio Analysis for Small Business Prediction," Paper presented at the Western Finance Association meetings, Vancouver, B. C., August 30, 1971, *Journal of Financial and Quantitative Analysis*, March 1972.

Robert A. Eisenbeis and Robert B. Avery. *Discriminant Analysis and Classification Procedures: Theory and Applications*, (Lexington, Mass.: D. C. Heath and Co., 1972).

———. "Two Aspects of Investigating Group Differences in Linear Discriminant Analysis," *Decision Sciences*, Vol. 4, No. 4, October, 1973.

———, Gary G. Gilbert and Robert B. Avery. "Investigating the Relative Importance of Individual Variables and Variable Subsets in Discriminant Analysis," *Communications in Statistics*, September, 1973.

——— and Alan S. McCall. "Some Effects of Affiliations Among Mutual Savings and Commercial Banks," *Journal of Finance*, 27 (September 1972), 865–877.

——— and Neil B. Murphy. "Interest Rate Ceilings and Consumer Credit Rationing: A Multivariate Analysis of a Survey of Borrowers," FDIC Working Paper No. 73-4, *Southern Economic Journal*, July, 1974.

R. A. Fisher. "The Use of Multiple Measures in Taxonomic Problems," *Annals of Eugenics*, 7 (September 1936).

R. E. Frank, W. F. Massy and G. D. Morrison. "Bias in Multiple Discriminant Analysis," *Journal of Marketing Research*, 2 (August 1965), 250–258.

Ethel S. Gilbert. "On Discrimination Using Qualitative Variables," *Journal of the American Statistical Association*, December, 1968.

———. "The Effect of Unequal Variance-Covariance Matrices on Fisher's Linear Discriminant Function," *Biometrics*, 25 (September 1969).

Gary G. Gilbert. "Predicting *de novo* Expansion in Bank Merger Cases," FDIC Working Paper No. 72-2, *Journal of Finance*, Vol. 29, No. 1, March, 1974.

———. "An Analysis of Federal Regulatory Decisions on Market Extension Bank Merger," *Journal of Money, Credit, and Banking*, Vol. 7, No. 1, February, 1975.

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

John A. Haslem and William A. Longbrake. "A Discriminant Analysis of Commercial Bank Profitability," *Quarterly Review of Economics and Business*, Vol. II, No. 3, 1971.

M. Hills. "Allocation Rules and Their Error Rates," *Journal of the Royal Statistical Society*, series B. vol. 28, no. 1 (1966): 1–20 (with discussion pp. 21–31).

———. "Discrimination and Allocation with Discrete Data," *Applied Statistics*, 16, No. 3 (1967).

L. N. Holloway and O. J. Dunn. "The Robustness of Hotelling's T^2" *Journal of the American Statistical Association*, 62 (March 1967).

J. W. Hopkins and P. P. F. Clay. "Some Empirical Distributions of Bivariate T^2 and Homoskedasticity Criterion M under Unequal Variance and Leptokurtosis," *Journal of the American Statistical Association*, 58 (1963).

Joseph J. Horton Jr. "Statistical Classification of Municipal Bonds," *Journal of Bank Research*, 1 (Autumn, 1970).

Koichi Ito and William J. Schull. "On the Robustness of the T^2 Test in Multivariate Analysis of Variance When Variance-Covariance Matrices Are Not Equal," *Biometrika*, 51 (June 1964).

O. Maurice Joy and John O. Tollefson, "On the Financial Applications of Discriminant Analysis," *Journal of Financial and Quantitative Analysis*, December, 1975.

Robert C. Klemkosky and J. William Petty. "A Multivariate Analysis of Stock Price Variability," *Journal of Business Research*, 1 (Summer 1973).

C. J. Kowalski. "The Performance of Some Rough Tests for Bivariate Normality Before and After Coordinate Transformations to Normality," *Technometrics*, 12 (1970), 517–44.

Anant M. Kshirsagar. *Multivariate Analysis*, (Marcel Dekker, Inc.: New York, 1972).

Peter A. Lachenbruch. "Discriminant Analysis When the Initial Samples Are Misclassified," *Technometrics*, Vol. 8, No. 4 (November, 1966).

———. "An Almost Unbiased Method of Obtaining Confidence Intervals for the Probability of Misclassification in Discriminant Analysis," *Biometrics*, 23 (December 1967), 639–645.

———. "On Expected Probabilities of Misclassification in Discriminant Analysis, Necessary Sample Size, and a Relation with the Multiple Correlation Coefficient," *Biometrics*, 24 (December 1968).

———. "Discriminant Analysis When the Initial Samples are Misclassified II: Non-Random Misclassification Models," *Technometrics*, Vol. 16, No. 3 (August, 1974).

———. *Discriminant Analysis*, (Hafner Press: New York, 1975).

——— and M. R. Mickey. "Estimation of Error Rates in Discriminant Analysis," *Technometrics*, Vol. 10, No. 1 (February, 1968).

———, Cheryl Sneeringer and Lawrence T. Revo. "Robustness of the Linear and Quadratic Discriminant Function to Certain Types of Non-Normality," *Communications in Statistics*, 1 (1), 39–56 (1973).

George W. Ladd. "Linear Probability Functions and Discriminant Functions," *Econometrica* 34, no. 4 (October 1966).

Sylvia Lane. "Submarginal Credit Risk Classification," *Journal of Financial and Quantitative Analysis*, 7 (January 1972), 1379–1386.

H. Linhart. "Techniques for Discriminant Analysis with Discrete Variables," *Metrica*, 2 (1959).

Paul R. Lohnes. "Test Space and Discriminant Space Classification Models and Related Significance Tests," *Educational and Psychological Measurement*, 21, no. 3 (1961).

J. F. Malkovich and A. A. Afifi. "On Tests for Multivariate Normality," *Journal of the American Statistical Association*, Vol. 68, No. 341, March, 1973.

Sidney Marks and Olive Jean Dunn. "Discriminant Functions when Covariance Matrices are Unequal," *Journal of the American Statistical Association*, Vol. 69, No. 346, June, 1974.

Paul A. Meyer and Howard W. Pifer. "Prediction of Bank Failures," *Journal of Finance*, 25 (September 1970), 853–868.

Donald F. Morrison. *Multivariate Statistical Methods* (McGraw-Hill Inc.: New York, 1967).

F. Mosteller and D. F. Wallace. "Inference in the Authorship Problem," *Journal of the American Statistical Association*, June, 1963.

R. Norgaard and C. Norgaard. "A Critical Examination of Share Repurchase," *Financial Management*, Spring, 1974.

M. Okamoto. "An Asymptotic Expansion for the Distribution of the Linear Discriminant Function," *Annals of Mathematical Statistics*, Vol. 34, (1963) p. 1286.

Y. E. Orgler. "A Credit Scoring Model for Consumer Loans," *Journal of Money, Credit, and Banking*, November 1970, 435–445.

George Pinches and Kent Mingo. "A Multivariate Analysis of Industrial Bond Ratings," *Journal of Finance*, March, 1973.

———. "The Role of Subordination and Industrial Bond Ratings," *Journal of Finance*, March, 1975.

Thomas F. Pogue and Robert M. Soldofsky. "What's in a Bond Rating," *Journal of Financial and Quantitative Analysis*, June, 1969.

O. R. Porebski. "On the Interrelated Nature of the Multivariate Statistics Used in Discriminatory Analysis," *British Journal of Mathematical and Statistical Psychology*, Vol. 19, pt. 2, November, 1966.

C. R. Rao. *Advanced Statistical Methods in Biometric Research* (John Wiley and Sons, New York: 1952).

Phillip J. Rulon; David V. Tiedeman; Maurice M. Tatsuoka; Charles R. Langmuir. *Multivariate Statistics for Personnel Classification*, (John Wiley and Sons, Inc., New York, 1967).

H. Sheffé. *The Analysis of Variance* (New York: John Wiley and Sons, 1959).

Richard A. Schick and James A. Verbrugge. "An Analysis of Bank Price-Earnings Ratios," *Journal of Bank Research*, Vol. 6, no. 2, Summer, 1975.

S. S. Shapiro and R. S. Francia. "Approximate Analysis of Variance Test for Normality," *Journal of the American Statistical Association*, 67 (1972).

———, M. B. Wilk and H. J. Chen. "A Comparative Study of Tests of Normality," *Journal of the American Statistical Association*, 63 (December 1968), 1343–72.

Joseph F. Sinkey, Jr. "A Multivariate Statistical Analysis of the Characteristics of Problem Banks," *Journal of Finance*, March, 1975.

C. A. B. Smith. "Some Examples of Discrimination," *Annals of Eugenics*, 18 (1947).

Keith V. Smith. *Classification of Investment Securities Using Multiple Discriminant Analysis*. Institute Paper No. 101 (Purdue University Institute for Research in the Behavioral, Economic and Management Sciences, 1965).

M. Sorum. "Estimating the Expected Probability of Misclassification for a Rule Based on the Linear Discriminant Function: Univariate Normal Case," *Technometrics*, Vol. 15, No. 2, May 1973.

David Stuhr and Robert Van Wicklin. "Rating the Financial Condition of Banks; A Statistical Approach to Aid Bank Supervision," *Monthly Review*, Federal Reserve Bank of New York, September, 1974.

Maurice M. Tatsuoka. *Multivariate Analysis: Techniques for Educational and Psychological Research* (New York: John Wiley and Sons, Inc., 1971).

J. E. Walter. "A Discriminant Function for Earnings Price Ratios of Industrial Corporations," *Review of Economics and Statistics*, 41 (February 1959), 44–52.

John M. Weiner and Olive Jean Dunn. "Elimination of Variables in Linear Discrimination Problems," *Biometrics* (June, 1966).

J. Kenton Zumwalt. "A Multivariate Analysis of the Preferred Stock Ratings of Electric Utilities," presented at the 1975 Midwest Finance Association meetings, April, 1975.