

# Multiple Regression with a Qualitative Dependent Variable

Daniel L. Rubinfeld

**This paper describes several methods by which a single set of scores for a qualitative (usually ordinal) dependent variable can be estimated simultaneously with the coefficients of the explanatory variables of a model. The canonical correlation and multiple discriminant analysis approaches are well known in the statistics literature.**

The problem of scale construction has received considerable attention outside the discipline of economics.<sup>1</sup> At the same time the use of maximum-likelihood estimation techniques such as probit and logit analysis to study qualitative choice problems has become quite popular. Within the purview of economics, relatively little attention has been paid to the relationship between scaling techniques and the analysis of models with qualitative dependent variables. This paper attempts to fill in some missing ground by stressing the link between multiple regression and models in which one seeks to determine a scale to represent a qualitative dependent variable. We describe methods by which a single set of scores for the dependent variable can be estimated simultaneously with the coefficients of the "independent" variables of a model.<sup>2</sup> Although the results may be interpreted in a multiple regression framework (e.g., as an extension of the linear probability model), we stress that the estimation technique need not involve multiple re-

gression calculations. However, the paper goes on to show an iterative least-squares multiple regression technique can provide a useful approximation to these more general procedures. The techniques are illustrated with labor force participation and voter turnout examples.

gression calculations. Because most of these techniques involve nonlinear estimation techniques that can be time-consuming and expensive for large data sets or large models, we propose an ad hoc multiple regression scaling technique that is relatively inexpensive to use. The multiple regression technique provides a useful approximation of some of the more general multivariate statistical techniques.

Assume that we know that a given individual unit of study, family, firm, city, and so on is characterized by a vector of attributes. Each individual population member is assumed to belong to one of several mutually exclusive groups, and the attribute vectors associated with each group are known to be normally distributed with different means but identical variance-covariance matrices. The problem is to find a *single* linear decision rule that predicts the group "score" of an individual after the vector of attributes describing that individual is observed. The predicted group scores can be interpreted as providing for a qualitative "dependent" variable in a multiple regression procedure. The predicted scores also provide for a method that allows for the classification of individuals into groups. The procedure is related in terms of distributional assumptions to multiple discriminant analysis and is identical when the sample means of the group attribute vectors lie on a straight line.

The techniques described here should be of use in economics as well as in related social science disciplines. Examples of some relevant applications are as follows:

1. In survey analysis, respondents to a questionnaire might be classified into one of several

---

<sup>1</sup> The problem of scale construction has received substantial treatment in the statistics literature. One procedure similar to the scaling technique described in this paper was devised by Guttman (1941, 1950). Guttman deals with the case in which the independent variables take the form of responses to a series of survey questions. Other discussions of scale construction appear in Shepard, Romney, and Nerlove (1972) and Green and Rao (1973).

<sup>2</sup> The search for a single set of scores is a restrictive one and thus substantially limits the scope of our analysis. We shall return to this issue later in the discussion.

From the University of Michigan, Ann Arbor, Michigan.

Address reprint requests to Dr. Daniel L. Rubinfeld, Associate Professor of Economics and Law, University of Michigan, Ann Arbor, Michigan 48109.

groups (e.g., Agree, Disagree, No response). Attribute data are available for each of the respondents and all respondents in each group are assumed to be drawn from a population with the same mean attribute vector. One can attach a score to each of the groups in an attempt to determine whether those who do not answer a given question are more like those who agree or more like those who disagree.

2. In a study of the work status of a certain segment of the labor force, one may wish to emphasize the distinction between those who work part-time and those who work full-time. If the individuals sampled can be properly classified into distinct groups such as unemployed, part-time employed, and full-time employed, and if attribute and labor market data are available, the scaling technique can provide a useful mode of analysis. The technique will be particularly valuable if one wishes to determine those individual attributes that best distinguish between the labor force categories.
3. Assume that corporate or municipal bond ratings have been attached to a list of bonds to be studied.<sup>4</sup> The scaling procedure can be used in an attempt to replicate the behavior of the rating agencies. One can determine weights for each of the attributes that determine ratings as well as a quantitative score for each of the rating categories. Classification rules can be obtained and used in the sample to test the validity of using a single linear decisions rule to describe the rating process.

The remainder of the paper is divided into five sections. The first introduces the formulation of the multiple regression model, with the unknown group scores interpreted artificially as the dependent variable in the regression model. Least-squares minimization subject to a constraint on the estimated parameters leads to the simultaneous determination (using eigenvectors and eigenvalues) of a set of group scores and the weights attached to the vector of individual attributes. The second discusses the interpretation of the regression model with particular emphasis on its use for classification purposes. The third describes an alternative view of the identical scoring problem through a generalized analysis of variance approach. The fourth describes a method by which ordinary least squares can be used to estimate the dependent variable scale and the attribute

weights. The fifth contains two examples of the application of the ordinary least-squares regression technique.

## GROUP SCORING IN A MULTIPLE REGRESSION SETTING

We assume that each individual unit under study is represented by an attribute vector, and that the population of attribute vectors may be partitioned into  $G$  groups. We also assume a quantitative score  $\alpha_g$  can be attached to each group, with no presumption that the  $\alpha_g$ s will necessarily be distinct. We proceed in this section to determine a set of scores and attribute weights that minimize the sum of squared residuals of an artificially defined regression problem. We stress the artificiality of the procedure because the "dependent" variable is not known and is clearly not normally distributed. We shall see that the multiple regression approach yields a technique that is identical to other ad hoc scaling procedures computationally, and has the advantage of providing a useful means of interpreting the estimated scale and attribute weights.

We define the unknown dependent variable  $y$  as follows:<sup>4</sup>

$$y_t = \begin{cases} \alpha_1 & \text{if the observation is in group 1} \\ \alpha_2 & \text{if the observation is in group 2} \\ \vdots & \\ \alpha_G & \text{if the observation is in group } G. \end{cases} \quad (1)$$

$$t = 1, 2, \dots, N$$

There are  $N_g$  observations in each of the  $G$  groups, and a total of  $N$  observations in the sample.

It will be helpful to represent the vector  $y$  as the product of a grouping matrix and a vector of unknown scores, i.e.,

$$y = D\alpha$$

where

$$\alpha = (\alpha_1, \dots, \alpha_G)' \text{ a } G \times 1 \text{ vector}$$

<sup>4</sup> If one were able to specify a priori the probabilities with which  $y_t$  takes each of the  $G$  possible values, then further pursuit of a linear probability model approach might be justified. The linear probability model is described in Ladd [8]. An attempted extension of the linear probability appears in Warner [12]. Since our primary objective is to estimate a set of scores or scale that can be interpreted in a regression setting, we shall proceed in a different direction.

<sup>5</sup> See Rubinfeld (1973) for details concerning this application.

$D = [D^{(1)}, \dots, D^{(G)}]$  an  $N \times G$  matrix

$D^{(g)}$   $g = 1, 2, \dots, G = 1$  if the observation is in group  $g$  and 0 otherwise.

The artificial linear regression model is  $y = D\alpha = X\beta$ . Our objective is to find estimates of the parameter vectors  $\alpha$  and  $\beta$  that minimize the sum of squares  $(D\alpha - X\beta)'(D\alpha - X\beta)$ , where  $X$  is an  $N \times k$  matrix of attribute observations (measured as deviations from the mean), and  $\beta$  is a  $k \times 1$  vector of attribute weights. Throughout the paper, superscripts within parentheses will be used to represent group labels ( $g$ ), while subscripts will refer to observations ( $t$ ) and variables ( $i, j, k, h, r$ ).

The parameters of the scaled dependent variable will not be uniquely determined, but such a procedure does yield a unique solution when a normalization is made on the vector of group attributes. We choose the normalization that the variance of the predicted values  $\hat{y}'\hat{y}$  be constant.<sup>5</sup> The Lagrangian expression to be minimized is the following:

$$L = (D\alpha - X\beta)'(D\alpha - X\beta) - (1 - \lambda)(\beta'X'X\beta - k). \quad (1)$$

Differentiating equation (1) with respect to  $\alpha$ ,  $\beta$ , and  $(1 - \lambda)$  yields the following first-order conditions ( $\mathbf{b}$  and  $\mathbf{a}$  refer to the vectors of parameter estimates of  $\beta$  and  $\alpha$ ):

$$D'D\alpha - D'X\beta = 0$$

$$X'X\beta - X'D\alpha - (1 - \lambda)(X'X\beta) = 0$$

$$\beta'X'X\beta - k = 0$$

which yields

$$\mathbf{b} = (1/\lambda)(X'X)^{-1}D'\mathbf{a} \quad (2)$$

$$\mathbf{a} = (D'D)^{-1}D'X\mathbf{b} = \bar{X}'\mathbf{b} \quad (3)$$

where  $\bar{X}$  is a  $k \times G$  matrix of the means of independent variables in each group. Notice that the vector of scores  $\mathbf{a}$  is a simple weighted average of the within group means of  $X\mathbf{b}$ . To solve explicitly for  $\mathbf{b}$ , substitute equation (3) into equation (2) to get

$$[A - \lambda I]\mathbf{b} = \mathbf{0} \quad (4)$$

<sup>5</sup> This normalization has been chosen primarily because it yields an estimation procedure that is tractable and relates closely to the scaling work of Guttman and others.

where  $I$  is a  $k \times k$  identity matrix and  $A = (X'X)^{-1}X'D(D'D)^{-1}D'X$ .

In general a solution for  $\mathbf{b}$  must involve the search for the largest eigenvalue of the matrix  $A$  and the corresponding eigenvector  $\mathbf{b}$ .<sup>6</sup> We shall return to this point later, but for the moment we shall concentrate on the use of the estimated vector of group scores. Once the group scores are known they may be interpreted in light of the regression approach used. In particular, if we utilize the scores determined from equation (4) as the dependent variable in an ordinary least-squares regression, then we obtain a set of estimated coefficients that are identical up to a scalar multiple to the vector  $\mathbf{b}$ .<sup>7</sup> This can be most easily seen by recalling the form of a least-squares estimator and comparing this to equation (2). Equation (2) also makes it clear that the equality of the estimated vector of attribute weights can be guaranteed by the appropriate renormalization of the group scores (each group score must be divided by  $\lambda$ ). To simplify matters we often utilize the vector  $(1/\lambda)\mathbf{a}$  as the vector of group scores.

## INTERPRETATION OF THE REGRESSION MODEL

The  $\mathbf{b}$ s measure the effect of a change in one or more of the independent variables on the normalized numerical index of the dependent variable, but the normalization renders any cardinal interpretation of the coefficients meaningless. The most appropriate use for the regression results is to classify observations into groups in the spirit of multiple discriminant analysis by calculating the residuals between the measured index  $a$  and the fitted values of the original equation  $X\mathbf{b}$ . The residual vector for each of the groups can be used to calculate an estimate of the variance associated with each category of the dependent variable.

The  $R^2$  statistic calculated from the multiple regression procedure provides one measure of good-

<sup>6</sup> There will, in general, be more than one eigenvalue associated with the matrix  $A$ . The choice of the largest eigenvalue is a result of our desire to emphasize the "one-dimensional" aspect of the scaling problem. If the largest eigenvalue does not give sufficient explanatory power, then it may be advisable to utilize more than one. See Guttman (1941, 1950) for details.

<sup>7</sup> Some normalization must be made when the scoring problem is set up. Without a normalization, it is easy to see that if  $\mathbf{a}$  and  $\mathbf{b}$  satisfy the first-order conditions that  $k\mathbf{a}$  and  $k\mathbf{b}$  will also. Thus our interest in the estimated vector of weights  $\mathbf{b}$  must be relative, not absolute.

ness of fit of the scoring procedure.  $R^2$  measures the proportion of the variation in  $\mathbf{y} = D\mathbf{a}$  explained by the variation in the weighted average of group attributes  $X\mathbf{b}$ . The calculated  $R^2$  will be identically equal to  $\lambda$ , the value of the largest eigenvalue of the matrix  $A$  in equation (4). This follows directly from the fact that

$$R^2 = \frac{\mathbf{b}'X'X\mathbf{b}}{\mathbf{y}'\mathbf{y}} = \frac{\mathbf{b}'X'X\mathbf{b}}{\mathbf{a}'D'D\mathbf{a}}$$

where  $\mathbf{a}$  has been renormalized so that:

$$\mathbf{a} = \begin{pmatrix} 1 \\ \vdots \end{pmatrix} \bar{X}'\mathbf{b}.$$

Then

$$R^2 = \lambda \left[ \frac{\mathbf{b}'X'X\mathbf{b}}{\mathbf{b}'X'D(D'D)^{-1}D'X\mathbf{b}} \right]$$

$= \lambda$  from equation (4)

Perhaps a more proper measure of success of the procedure would involve a comparison with alternative techniques of estimation and classification. One reasonable approach would be to compare the multiple regression classification errors with the number of classification errors associated with a multiple discriminant analysis procedure. This measure of success is deceptive, because multiple discriminant analysis involves the estimation of  $G - 1$  independent equations, and thus uses more degrees of freedom than the regression technique.

### GROUP SCORING IN AN ANALYSIS OF VARIANCE SETTING

The group scaling or scoring problem has frequently been described in the literature in terms of generalized analysis of variance and canonical correlation. We shall describe the former approach here and leave the latter derivation to the reader.<sup>8</sup> Assume that we wish to find a vector  $\mathbf{b}$  which maximizes the variance among group means relative to the total variance within groups. To accomplish this we define

1.  $S = (S_{ij}) =$  matrix of pooled sum of cross-products using deviations about overall means

$$S_{ij} = \sum_{g=1}^G \sum_{t=1}^{N_g} X_{ij}^{(g)} X_{jt}^{(g)} - N \bar{X}_i \bar{X}_j.$$

where  $X^{(g)}$  refers to the  $N_g \times k$  matrix of observations associated with the  $g$ th group.

Then with variables measured as deviations about means  $S = X'X$ ,

2.  $C = (C_{ij}) =$  matrix of pooled sum of cross-products using deviations about within group means

$$C_{ij} = \sum_{g=1}^G \sum_{t=1}^{N_g} X_{ij}^{(g)} X_{jt}^{(g)} - \sum_{g=1}^G N_g \bar{X}_i^{(g)} \bar{X}_j^{(g)}. \quad (5)$$

Then

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

The variance among group means can be represented by  $\mathbf{b}'V\mathbf{b}$  where

$$V = S - C = X'D(D'D)^{-1}D'X,$$

while the total variance is given by  $\mathbf{b}'S\mathbf{b}$ .

We can maximize  $\mathbf{b}'V\mathbf{b}$  subject to  $\mathbf{b}'S\mathbf{b}$  being constant by writing the Lagrangian

$$L = \mathbf{b}'X'D(D'D)^{-1}D'X\mathbf{b} - \lambda(\mathbf{b}'X'X\mathbf{b} - k).$$

Differentiating with respect to  $\mathbf{b}$  and solving we obtain

$$X'D(D'D)^{-1}D'X\mathbf{b} = \lambda X'X\mathbf{b}$$

or

$$(A - \lambda I)\mathbf{b} = 0 \quad (6)$$

where  $A = (X'X)^{-1}X'D(D'D)^{-1}D'X$  and  $I$  is a  $k \times k$  identity matrix. This can be seen to be identical in form to equation (4). The solution to equation (4) is once again obtained by choosing the largest eigenvalue  $\lambda$  (which satisfies the constraints).<sup>9</sup>

<sup>9</sup> Not infrequently the derivation just described is given as one that maximizes  $\mathbf{b}'V\mathbf{b}$  subject to  $\mathbf{b}'C\mathbf{b}$  being constant. The attribute weights and group scores obtained will be equivalent up to a scalar multiple, but the new eigenvalue obtained will not be equal to  $\lambda$ . In particular, it is possible to show that  $\theta = \lambda/(1 + \lambda)$  where  $\theta$  is the eigenvalue of the matrix  $C^{-1}V$ . See Cooley and Lohnes (1971) for some details. This has relevance here because the canonical correlation package used in the application provides an estimate of  $\theta$ , not  $\lambda$ .

<sup>8</sup> For further details concerning both approaches, see Guttman (1950), Anderson (1958), Cooley and Lohnes (1962, 1971), and Rao (1952).

### ORDINARY LEAST SQUARES WITH A QUALITATIVE DEPENDENT VARIABLE (OQDV)

The scoring procedure just described suffers from the disadvantage that a solution cannot be obtained using a standard multiple regression package. In this section we briefly outline an estimation process that will enable us to obtain an estimate of the vector of weights  $\beta$  (and the scores  $\alpha$ ) using ordinary least squares.

Recall that the original formulation of the model was  $D\alpha = X\beta$ . We can utilize ordinary least squares by normalizing the vector  $\mathbf{a}$ , arbitrarily choosing  $a_1 = 0$  and  $a_G = 1$ . Ordinary least-squares estimation can be used if we rewrite the model in the following form:

$$D^{(G)} = [X \quad -D^{(2)} \quad -D^{(3)} \quad \dots \quad -D^{(G-1)}] \begin{pmatrix} \beta \\ \alpha_2 \\ \alpha_3 \\ \vdots \\ \alpha_{G-1} \end{pmatrix} \quad (7)$$

The least-squares technique applied to equation (7) provides for an estimate  $\mathbf{b}$  of the parameter  $\beta$ . The estimated scale is simply the vector  $\mathbf{a} = \bar{X}'\mathbf{b}$ . There is no guarantee that the estimation procedure will yield a scale consistent with one's prior notions about the ordering of the groups. More importantly, there is no guarantee that OQDV will yield scores identical (up to a suitable transformation) to those derived earlier. This can be realized intuitively by noting that the calculated scores for all groups other than the 1st and  $G$ th will be identical to those obtained by the ordinary least-squares procedure. But the scores for the 1st and the  $G$ th groups will not necessarily equal 0 and 1, respectively.<sup>10</sup>

The scale estimate obtained for the set of group scores can be used to obtain improved parameter estimates through the use of an iterated least-squares procedure. The second iteration is accomplished by regressing the vector  $D\mathbf{a}$  on the vector of attributes  $X$ . This will yield a new set of estimated attribute weights  $\mathbf{b}'$  and a new set of group scores  $\mathbf{a}'$ . Then the new group scores yield a new set of attribute weights, and so on. The iterated set of group scores and attribute weights will remain unchanged from the previous set only when the estimated scale and weights corre-

spond to the generalized analysis of variance solution (and when the  $R^2$  is identically equal to the eigenvalue of the matrix  $A$ ). We have been unable to prove convergence under a general set of conditions, but the lack of a general set of convergence conditions is not likely to be of practical consequence. The reason is that when convergence does occur, the estimated scale and weights will correspond to the generalized analysis of variance solution, with  $R^2$  being equal to the largest eigenvalue of  $A$ . And conversely when the scale is equal to the analysis of variance scale, the iterative process will end—additional iterations will yield the same scale.<sup>11</sup> Thus convergence guarantees that the “optimal” solution has been reached, whereas nonconvergence will be rapidly apparent if it occurs. In practice, in a number of experiments conducted with different data sets, the estimated scale converged rapidly with the generalized analysis of variance scale.<sup>12</sup>

We have chosen to focus on a single linear function of the group attributes, but the techniques discussed should be viewed as a special case of the more general decision rule used for multiple classification. In fact, the scoring procedures (both OQDV and the more general iterative procedure) can be shown to be identical to multiple discriminant analysis when the sample group means of  $X$  lie on a straight line.<sup>13</sup> This suggests a set of conditions under which the use of a *single* set of group scores involves no loss of explanatory power as well as the conditions under which OQDV is likely to approximate closely the more general scoring procedure.

### Statistical Testing

As a final item, it is reasonable to ask whether the statistical tests associated with the ordinary least-squares regression are valid in the scoring procedure described in this paper. Assume that we are working

<sup>11</sup> Recall that the scale is  $\mathbf{a} = \bar{X}'\mathbf{b}$ . The dependent variable is  $D\mathbf{a}$  and the regressor  $\mathbf{b}$ . The newly calculated coefficient vector is then

$$\begin{aligned} \mathbf{b}^* &= (X'X)^{-1}X'D\mathbf{a} = (X'X)^{-1}X'D\bar{X}\mathbf{b} \\ &= (X'X)^{-1}X'D(D'D)^{-1}D'X\mathbf{b} = \mathbf{A}\mathbf{b} \end{aligned}$$

From equation (4), if  $\mathbf{b}$  is the generalized analysis of variance solution  $(A - \lambda I)\mathbf{b} = 0$ . Therefore  $\mathbf{b}^* = \lambda I\mathbf{b} = \lambda\mathbf{b}$ , so that  $\mathbf{b}^*$  is a scalar multiple of  $\mathbf{b}$ .

<sup>12</sup> The initial set of group scores obtained will depend on the particular 0-1 normalization chosen. This can be seen most clearly when the labor force example is discussed below.

<sup>13</sup> A detailed proof of this result is given in the Appendix.

<sup>10</sup> The estimated scores will be equivalent to those derived earlier when the sample means of the attribute vectors lie on a straight line. See Appendix for proof.

under the set of conditions described by multiple discriminant analysis (the  $X$ s are assumed to be joint-normally distributed). Then it is clear that under the condition that the true means of the group attribute vectors lie on a straight line, the distributional results of multiple discriminant analysis hold.<sup>14</sup> In this case, small sample tests are appropriate, but since the conditions of the previous theorem are likely to hold approximately at best, any tests based on the methodology of discriminant analysis are likely to be inexact. If we view the scoring model as an approximation to the logit or probit model, large sample tests can be appropriate.

## TWO EXAMPLES

### Labor Force Participation

The following application is based on a study of the labor force participation of married female teachers.<sup>15</sup> The focal point of the study was the breakdown of labor supply status into three groups—those working full-time, those working part-time, and those not working at all. One of the objectives of the study was to determine those variables that best help to distinguish between the three labor force groups. Discriminant analysis (two types) was the sole technique used.

Using the same data set, we have attempted to estimate a scale of set of scores for the three labor force groups. A brief description of the data set and a list of the relevant variables appear in Table 1. The reader is referred to the original source article for more complete details about the data set.

Tables 2, 3, and 4 contain the results of several of the estimation experiments. Each of the scales listed in Table 2 was renormalized (through an appropriate affine transformation) to make the results of the alternative estimation procedures comparable. Such a renormalization has no effect on the relative magnitude of attribute coefficients and no effect on the statistical tests used. The complete multiple regression results for three of the experiments are listed for illustrative purposes.

Scale 1 was determined from a canonical correlation package as a solution to the original scaling problem. The regression results using Scale 1 to form the ordered dependent variable are given in Table 3.

**TABLE 1.** Definition of Variables

*Sample:* 414 married women consisting of 254 full-time teachers, 118 substitute teachers and 42 nonworking teachers

$W(S)$	=	part-time wage of wife in \$1000
$W(F)$	=	full-time wage of wife in \$1000
$W(H)$	=	wage of husband in \$1000
$T$	=	household age in years
$ASSETS$	=	household assets in \$1000
$\#CH$	=	number of children under six years of age
$D1$	=	dummy variable equal to 1 if not working and zero otherwise
$D2$	=	dummy variable equal to 1 if part-time and zero otherwise
$D3$	=	dummy variable equal to 1 if full-time and zero otherwise

*Source:* Gramm (1973).

What value does such a scaling procedure have in the context of this labor force problem? First, recall that the implicit assumption here is that one set of regression parameters can be used to explain the full-time, part-time, no work decision. Given this assumption, the estimated scale of 1, 0.77, 0 suggests that the characteristics or attributes of part-time workers is such that they are more similar or like those not working than those working part-time. This may have implications for policy not apparent in the Gramm study because it suggests that labor force participation responses to policy changes may be substantial as workers move into the labor force to part-time jobs and out of the labor force with relative ease. The regression results themselves make it much easier to evaluate the importance of individual characteristics in explaining the ordered labor force choice. The use of asymptotic tests in Table 3 allows us to say that the full-time wage of the wife has a significant positive effect on the probability of the woman's entrance into the labor force (recall that 1 represents the nonworking choice). Significant but in the opposite direction is the wage of the husband, as one would expect. All of the coefficients have the expected sign, and only one is insignificant at the 10 percent level. The overall results suggest that taking into account the part-time status of women teachers does not alter the behavior that would be predicted by labor economics theory. These results are also implicit in the more general discriminant analysis approach, but the multiple regression does simplify the analysis. In our case, the lost precision or

<sup>14</sup> These tests are described in some detail by Anderson (1958) and Rao (1952).

<sup>15</sup> See Gramm (1973). The author wishes to thank Ms. Gramm for her assistance in making the data available.

TABLE 2. Scale Variables Defined<sup>a</sup>

Labor Force Status	Scale 1 <sup>b</sup>	Scale 2 <sup>c</sup>	Scale 3 <sup>d</sup>	Scale 4 <sup>e</sup>	Scale 5 <sup>f</sup>	Scale 6 <sup>g</sup>
Not working	1.000	1.000	1.000	1.000	1.000	1.000
Part-time	0.770	0.858	0.662	0.399	0.699	0.757
Full-time	0.000	0.000	0.000	0.000	0.000	0.000

<sup>a</sup> All scale variables have been renormalized for purposes of comparison by means of suitably chosen affine transformations.

<sup>b</sup> Obtained through the solution of equation (6) using canonical correlation computer program.

<sup>c</sup> Obtained by regression of D2 on D1.

<sup>d</sup> Obtained by regression of D3 on D2.

<sup>e</sup> Obtained by regression of D1 on D3.

<sup>f</sup> Obtained through second iteration using least-squares regression of D1 on D3.

<sup>g</sup> Obtained through third iteration using regression of D1 on D3.

explanatory power of the model is small, since our results predict almost as well as do those obtained from the multiple discriminant analysis approach.

How does our iterative least-squares approach to estimation work in this case? To evaluate this question, the remaining scale variables were obtained using the OQDV approximation and a standard multiple regression package. What is most striking about the OQDX approximation is the degree of accuracy associated with the first iteration of the procedure. The scale obtained through the regression D1 on D3 and the vector of attributes (Scale 4) yields the poorest results, but even in this case  $R^2$  is only 0.04

below the maximal  $R^2$ , all signs are identical, and only one additional coefficient (the  $T$  variable) is insignificant. When the Scale 4 variable was used iteratively to obtain new scale variables (Scale 5 and Scale 6) the results were again very promising. By the third iteration the difference in scales was small and no substantive difference in regression results could be seen.

### Voter Turnout

In the 1978 congressional election, Michigan voters passed a tax limitation amendment that limited the

TABLE 3. Regression Results

	Constant	$W(S)$	$W(F)$	$W(H)$	$T$	Assets	#CH
Scale 1 (St. Err.)	0.422 <sup>a</sup> (0.116)	-0.0063 (0.022)	-0.056 <sup>a</sup> (0.008)	0.018 <sup>a</sup> (0.002)	0.0045 <sup>a</sup> (0.0020)	0.0013 <sup>b</sup> (0.0007)	0.356 <sup>a</sup> (0.035)
	$R^2 = 0.443$		S.E. = 0.308		Percent variance explained by first eigenvalue = 91.1		
Scale 2 (St. Err.)	0.455 <sup>a</sup> (0.124)	-0.0061 (0.024)	-0.061 <sup>a</sup> (0.009)	0.019 <sup>a</sup> (0.002)	0.0054 <sup>a</sup> (0.0022)	0.0014 <sup>b</sup> (0.0007)	0.370 <sup>a</sup> (0.037)
	$R^2 = 0.442$		S.E. = 0.330				
Scale 4 (St. Err.)	0.279 <sup>a</sup> (0.094)	-0.0072 (0.018)	-0.035 <sup>a</sup> (0.007)	0.013 <sup>a</sup> (0.002)	0.0005 (0.0016)	0.0010 <sup>b</sup> (0.0005)	0.300 <sup>a</sup> (0.028)
	$R^2 = 0.399$		S.E. = 0.250				

Notes: 1. The standard error is an estimate.

2. Regression coefficients and standard errors of the regression are not directly comparable, since the variance of the dependent variable has not been fixed.

3. The results of Scale 1 are not identical (up to a scalar multiple) to those given in Gramm (1973), owing most likely to an error in the transmission of the data.

<sup>a</sup> Significant at the 5 percent level (using standard  $t$  test).

<sup>b</sup> Significant at the 10 percent level.

TABLE 4. Classification Matrices

	Actual Labor Force Status	Predicted Status		
		No Work	Part-time	Full-time
Scale 1	No Work	6	28	8
	Part-time	10	76	32
	Full-time	0	27	227
Percent observations correctly classified = 73				
Scale 2	No Work	5	29	8
	Part-time	10	72	36
	Full-time	0	27	227
Percent of observations correctly classified = 73				
Scale 4	No Work	6	33	3
	Part-time	7	88	23
	Full-time	0	41	213
Percent of observations correctly classified = 74				

Note: Classifications were made using cutoff points midway between the group scores.

growth of state and local government to a fixed fraction of state personal income. The vote on the "Headlee" amendment is analyzed in Courant, Gramlich, and Rubinfeld (1980) and the behavior of nonvoters is described in Gramlich, Rubinfeld, and Swift (1980). The example that follows uses a substantially simplified version of the model presented in the latter paper. The simplification is for pedagogic reasons and does not alter the qualitative nature of the results.

Assume that voters favor tax limitation because they want less public spending, but oppose it when they would like less. Of special importance for us are the views of those who do not vote. Are nonvoters typically those who are indifferent in the sense of being relatively happy about the current level of service provision, or do they have strong views for or against spending—sufficiently different from most of the rest of the population to alienate them and encourage them not to vote at all?

To pursue this issue we estimated a model using responses from an interview survey of Michigan voters taken immediately after the November election. The canonical correlation regression results are presented in Table 5.

Overall, the simple preference model of voting is not a particularly good one (note the low  $R^2$ ). The scaled regression result does illustrate a point that remains true when more elaborate voting models are estimated. The estimated scale variable "Vote" suggests that nonvoters are similar to no voters (in terms

of voting preferences). Whereas their individual characteristics may be (and are) different, their predicted voting behavior suggests that they would oppose the tax limitation amendment. Since turnout was in the neighborhood of 50 percent and the tax limitation amendment passed with only 54 percent of the vote, our policy conclusions are strong. Nonvoting on the amendment was not random—yes-leaning voters turned out at a higher rate than those leaning toward no. Had turnout in the election been substantially higher, our results suggest that the tax limitation amendment would have failed.

We might note also that the iterative least-squares approximation to the canonical correlation scale

TABLE 5. Voter Turnout Results

Vote	=	1 if vote yes, 0.096 if not voted, 0 if vote no
More	=	1 if desired more public spending, 0 otherwise
Less	=	1 if desired less public spending, 0 otherwise
DK	=	1 if don't know; 0 otherwise
NA	=	1 if response was not available; 0 otherwise
Vote	=	.304 - .094 More + .124 Less - .039 DK - 100 NA
		(.013) <sup>a</sup> (.032) (0.21) (.064) (.097)
$R^2$	=	0.03

<sup>a</sup> Standard errors are in parentheses.



converged rapidly to the solution. To obtain the scale, we regressed a yes-no dummy on the right-hand variables including a nonvoting dummy and obtained an estimated scale of 1, 0.155, 0. Using this scale to form a new dependent variable, we estimated a new regression and formed a new scale, 1, 0.102, 0. The third iteration yielded a scale of 1, 0.096, 0, essentially the same as the canonical correlation scale. For large models, this iterative technique can save substantially on computer cost.

## CONCLUSIONS

We have described a procedure whereby a set of scores or scale for a "dependent" variable can be estimated simultaneously with the coefficients of the independent variables of an equation. Although the estimation procedure involves the search for the eigenvalues of a matrix, the resulting scores and attribute weights may be interpreted in a regression setting. The scale for the dependent variable provides the basis for a decision rule that allows for the prediction or classification of individual observations into categories and also provides for a measure of success of the regression procedure. We have seen that the scaling procedure can be viewed alternatively as a process that involves the maximization of the among group variance relative to the within group variance. The resulting scale is equivalent to one that would be obtained by the suggested scaling techniques of Guttman and others.

The ordinary least-squares regression package can be used to approximate the scale obtained through the previously discussed technique. The OQDV regression technique and the more general scaling procedure are seen to be equivalent to multiple discriminant analysis when the means of the attribute vectors for each category lie on a straight line in attribute space.

There are several limitations to the kind of scaling procedure described in this paper. A decision must be made as to whether an inherently multidimensional problem should be reduced to a unidimensional one. The benefits of determining a single linear set of weights for the group attributes and interpreting the calculated scores in a regression framework may be outweighed by the costs associated with the loss of predictive power in the model. If the decision to use a single linear function is made, a choice of computational techniques is available. The OQVD technique has the advantage of simplicity of computation, but it may cause misleading results if the assumption that the sample group means lie on a straight line is not

approximated. Further research is needed into the convergence properties of the iterative least-squares procedure as well as the tradeoffs involved in the choice of a one-dimensional scaling objective.

---

The author wishes to thank Franklin Fisher for stimulating his interest in this problem. Franklin Fisher, W. Locke Anderson, Saul Hymans, and an anonymous referee made helpful comments at several points in the development of the paper.

---

## APPENDIX

*The relationship between the iterative least-squares scale estimation procedure (OQVD) and multiple discriminant analysis.*

In the discriminant analysis approach for the multiple group case (and equal costs of misclassification) the attribute vectors associated with each group are assumed to be normally distributed with different means but identical variance-covariance matrices. Let  $\mu^{(g)}$  be the mean of  $X$  in group  $g$  and  $\Sigma$  the variance-covariance matrix of  $X$  and let  $X = (X^{(1)}, X^{(2)}, \dots, X^{(g)})'$  where  $X^{(g)}$  is an  $N_g \times k$  matrix for  $g = 1, 2, \dots, G$ . Taking the logarithm of the ratio of the probability density functions for two arbitrary groups, one can obtain regions for classification. The decision rule that minimizes the costs of misclassification involves the evaluation of  $G(G-1)/2$  functions in which all pairwise comparisons of groups are made. These functions are as follows:

$$D_{gh}(X) = [X - (\mu^{(g)} + \mu^{(h)})/2] \\ \times \Sigma^{-1}(\mu^{(g)} - \mu^{(h)}) \\ g, h = 1, 2, \dots, G \quad g \neq h. \quad (A1)$$

From these discriminant functions hyperplanes are calculated that span a  $(G-1)$  dimensional space. If the *a priori* probabilities of an observation falling into any category are equal, then the classification rule is to place an observation in category  $g$  if  $D_{gh} \geq 0$  for all  $h$ .<sup>1</sup> In actual practice sample estimates of the true  $\mu^{(g)}$  and  $\Sigma$  must be utilized.<sup>2</sup> With this background, it is now

---

<sup>1</sup> For a complete derivation utilizing this approach, see Anderson (1958).

<sup>2</sup> The matrix of pooled sum of cross-products using deviations about within group means  $C$  can be used to obtain a consistent estimator of the unknown variable covariance-matrix  $\Sigma$ .

possible to compare multiple discriminant analysis with our scoring processes.

**Lemma 1.** *If  $\bar{X}^{(g)}$ , the mean vectors of each of the categories, fall on a straight line (in the parameter space), then the discriminant analysis hyperplanes are parallel and only one discriminant function is needed to represent the process.*

**PROOF.** The fact that the means lie on a straight line may be written algebraically as  $\bar{X}^{(g)} = \gamma + k_g \delta$  where the  $k_g$  are scalars and  $\delta, \gamma$  are vectors. The discriminant functions to be calculated may be written

$$D_{gh}(X) = X' C^{-1} (\bar{X}^{(g)} - \bar{X}^{(h)}) \\ - [(\bar{X}^{(g)} + \bar{X}^{(h)})/2] C^{-1} (\bar{X}^{(g)} - \bar{X}^{(h)}) \quad (\text{A2})$$

where the second term of the equation is a scalar representing the intercept of the hyperplane. Substituting for  $\bar{X}^{(g)}$  and  $\bar{X}^{(h)}$  in the equation above, we find that

$$D_{gh}(X) = (k_g - k_h) X' C^{-1} \delta \\ - (k_g - k_h) \left( \gamma + \left( \frac{k_g + k_h}{2} \right) \delta \right)' V^{-1} \delta. \quad (\text{A3})$$

Since the  $k_g$ s are scalars and the second term is a scalar, all  $D_{gh}$ s will represent parallel hyperplanes. Only one discriminant function and a set of cutoff points or rules is needed for the classification procedure. The coefficients in the discriminant function  $D_{12}(X)$  may be written as  $C^{-1}(\bar{X}^{(1)} - \bar{X}^{(2)})$ . Coefficients in  $D_{gh}(X)$  may be represented as  $Q_{gh} V^{-1}(\bar{X}^{(1)} - \bar{X}^{(2)})$  where  $Q_{gh}$  is a scalar whose value depends on  $g$  and  $h$ . Coefficients in each discriminant function are a scalar multiple of coefficients in every other discriminant function.

In discriminant analysis the pooled cross-products of deviations of  $X_i$  and  $X_j$  about the within class means are used to approximate  $\Sigma$  ( $i$  and  $j$  represent independent variables here). But ordinary least squares involves the use of pooled cross-products of deviations of  $X_i$  and  $X_j$  about the overall mean of the sample. To compare discriminant analysis with OQDV, a lemma relating these two matrices is necessary. Since the lemma is standard in the analysis of variance literature a proof is not presented here.<sup>3</sup>

### Lemma 2

$$C_{ij} = S_{ij} - \sum_{h=1}^G \sum_{r=1}^G N_h N_r N (\bar{X}_i^{(h)} - \bar{X}_i^{(r)}) \\ \times (\bar{X}_j^{(h)} - \bar{X}_j^{(r)}) \quad (\text{A4}) \\ h, r = 1, 2, \dots, G.$$

**Corollary.** *If the means of the groups lie on a straight line (i.e.,*

$$(\bar{X}^{(g)} = \gamma + k_g \delta) \text{ then } C_{ij} = S_{ij} - Q_2 (\bar{X}_i^{(1)} - \bar{X}_i^{(2)}) \\ \times (\bar{X}_j^{(1)} - \bar{X}_j^{(2)}) \quad (\text{A5})$$

where  $Q_2$  is a scalar and  $\bar{X}_i^{(1)}, \bar{X}_j^{(1)}$ , and  $\bar{X}_i^{(2)}, \bar{X}_j^{(2)}$  are the means of the  $i$ th and  $j$ th variables in the first two categories.

**PROOF.** If the means lie on a straight line then for any groups  $h$  and  $r$ ,

$$(\bar{X}_i^{(h)} - \bar{X}_i^{(r)}) = Q_{hr} (\bar{X}_i^{(1)} - \bar{X}_i^{(2)})$$

where

$$Q_{hr} = \frac{k_h - k_r}{k_1 - k_2} \quad (\text{A6})$$

Then

$$C_{ij} = S_{ij} - \sum_{h=1}^G \sum_{r=1}^G (Q_{hr} N_h N_r / N) (\bar{X}_i^{(1)} - \bar{X}_i^{(2)}) \\ \times (\bar{X}_j^{(1)} - \bar{X}_j^{(2)}) \\ = S_{ij} - Q_2 (\bar{X}_i^{(1)} - \bar{X}_i^{(2)}) (\bar{X}_j^{(1)} - \bar{X}_j^{(2)}).$$

One final lemma is necessary to make comparisons possible between the regression techniques and multiple discriminant analysis.

Let

$y_i^{(g)}$  = the  $i$ th observation of the  $g$ th group  
 $i = 1, 2, \dots, N_g$  and  $g = 1, 2, \dots, G$

$\bar{y}$  = mean of  $y = \sum_{g=1}^G a_g N_g / N$

$S_{iy} = \sum_{g=1}^G \sum_{i=1}^{N_g} X_{ii}^{(g)} y_i^{(g)} - N \bar{y} \bar{X}_i$

$S_{iy}$  = covariance of  $y$  with  $X_i$ .

### Lemma 3

$$S_{iy} = (1/N) \sum_{h=1}^G \sum_{r=1}^G N_h N_r (a_h - a_r) \\ (\bar{X}_i^{(h)} - \bar{X}_i^{(r)}) \quad (\text{A7}) \\ h > r$$

where  $a_h, a_r, h, r = 2, \dots, G$  represent the estimated numerical scores described earlier.<sup>4</sup>

<sup>3</sup> This is equivalent to the statement in equation (5).

<sup>4</sup> The covariances are calculated with the estimated scores in preparation for a description of the regression estimation procedure.

$$\begin{aligned}
S_{iy} &= \sum_{h=1}^G N_h a_h \bar{X}_i^{(h)} - \bar{y} \sum_{h=1}^G N_h \bar{X}_i^{(h)} \\
&= \sum_{h=1}^G N_h \bar{X}_i^{(h)} (a_h - \bar{y}) \\
&= (1/N) \left[ \sum_{h=1}^G \left( \sum_{r=1}^G N_r \right) N_h \bar{X}_i^{(h)} a_h \right. \\
&\quad \left. - N_h \bar{X}_i^{(h)} \left( \sum_{r=1}^G N_r a_r \right) \right] \\
&= (1/N) \left[ \sum_{h=1}^G \sum_{r=1}^G (N_h N_r a_h \bar{X}_i^{(h)} - N_h N_r a_r \bar{X}_i^{(h)}) \right] \\
&= (1/N) \sum_{h=1}^G \sum_{r=1}^G [N_h N_r \bar{X}_i^{(h)} (a_h - a_r)] \\
&= (1/N) \sum_{k=1}^G \sum_{r=1}^G N_h N_r (a_h - a_r) (\bar{X}_i^{(h)} - \bar{X}_i^{(r)}) \\
&\quad k > r
\end{aligned}$$

**Corollary.** *If the means of the groups lie on a straight line, then*

$$S_{iy} = Q_3 (\bar{X}_i^{(1)} - \bar{X}_i^{(2)})$$

where  $Q_3$  is a scalar and  $\bar{X}_i^{(1)}$  and  $\bar{X}_i^{(2)}$  are the means of the first two categories.

**PROOF.** If the means lie on a straight line, recall from equation (A6) that

$$(\bar{X}_i^{(h)} - \bar{X}_i^{(r)}) = Q_{hr} (\bar{X}_i^{(1)} - \bar{X}_i^{(2)})$$

$$\text{where } Q_{hr} = \frac{k_h - k_r}{k_1 - k_2}.$$

Then

$$\begin{aligned}
S_{iy} &= (1/N) \sum_{h=1}^G \sum_{r=1}^G N_h N_r (a_h - a_r) Q_{hr} (\bar{X}_i^{(1)} - \bar{X}_i^{(2)}) \\
&= Q_3 (\bar{X}_i^{(1)} - \bar{X}_i^{(2)})
\end{aligned}$$

**Theorem.** *The coefficients of discriminant functions in multiple discriminant analysis are identical (up to scalar multiples) to the regression coefficients of the multiple regression procedures of the sample means of the  $G$  categories lie on a straight line.<sup>5</sup>*

**PROOF.** Discriminant analysis coefficients are equal to

$$\mathbf{b}^* = Q_{gh} C^{-1} (\bar{X}^{(1)} - \bar{X}^{(2)})$$

where  $Q_{gh}$  is a scalar and

$$\bar{X}^{(g)} = (\bar{X}_1^{(g)}, \dots, \bar{X}_k^{(g)})$$

if the sample means lie on a straight line (by lemma 1). Regression estimates are best viewed if deviations around the mean are used. The regression coefficients are  $\mathbf{b}^{**} = A^{-1} S_y$  where  $S_y = (S_{1y}, \dots, S_{ky})$ . Proof of the theorem involves proof that  $\mathbf{b}^{**} = Q_1 \mathbf{b}^*$  where  $Q_1$  is a scalar, if the group means lie on a straight line. If the means do lie on a straight line, then writing the results of the corollaries to lemmas 2 and 3 in matrix notation, we obtain

$$C = S - Q_2 \mathbf{d} \mathbf{d}'$$

$$S_y = Q_3 \mathbf{d}$$

where

$$\mathbf{d}' = (d_1, \dots, d_k) \text{ and } d_i = (\bar{X}_i^{(1)} - \bar{X}_i^{(2)})$$

$Q_2$  and  $Q_3$  are scalars,  $C$  and  $S$  are  $k \times k$  matrices, and

$S_y = (S_{1y}, \dots, S_{ky})$ ,  $S_y$ ,  $\mathbf{d}$ ,  $\mathbf{b}^*$ , and  $\mathbf{b}^{**}$  are  $k \times 1$  vectors.

Since  $\mathbf{b}^{**} = (S^{-1} S_y)$ , it follows by substitution that

$$\begin{aligned}
C \mathbf{b}^{**} &= (S - Q_2 \mathbf{d} \mathbf{d}') S^{-1} S_y = (S - Q_2 \mathbf{d} \mathbf{d}') S^{-1} Q_3 \mathbf{d} \\
&= Q_3 \mathbf{d} - Q_3 Q_2 \mathbf{d} (\mathbf{d}' S^{-1} \mathbf{d}) = \mathbf{d} (Q_3 Q_2 Q_4)
\end{aligned}$$

where  $Q_4 = \mathbf{d}' S^{-1} \mathbf{d}$  is a scalar. Then  $\mathbf{b}^{**} = Q_5 C^{-1} \mathbf{d} = Q \mathbf{b}^*$  where  $Q_5$  and  $Q$  are scalars that are properly chosen. Q.E.D.

The theorem that has been proved is stated in terms of the sample means of the categories, but an equivalent theorem is true for the population parameters. This can be seen by retracing the steps of the lemmas and the theorem while replacing sample means and variances with their population counterparts.<sup>6</sup> If the sample means do lie on a straight line, then the normalization is unimportant in the sense

<sup>6</sup> It might be valuable to compare the scaling procedure described earlier to the multivariate regression technique described by Warner (1963). Warner's regression approach is distributionally quite similar to that of multiple discriminant analysis. It is not difficult to show that if prior probabilities associated with each category are equal and there is a common covariance matrix associated with each of the group attribute vectors, then Warner's results are comparable to that of multiple discriminant analysis. When the sample means of the attribute vectors associated with each group do lie on a straight line, a scale identical to the scale described previously will result. There does not appear to be an obvious method of collapsing Warner's multidimensional model to a unidimensional one (and determining an appropriate scale) when the means of the group attribute vectors do not lie on a straight line.

<sup>5</sup> The theorem applies to the coefficients of the independent variables in the regression procedure, not to the constant term or to the estimated numerical indices.

that the estimated coefficient vector will be uniquely determined (up to a scalar multiple) no matter which categories are arbitrarily assigned the values of 0 and 1. This can best be seen by noticing in the corollary to Lemma 3 that if the sample means lie on a straight line, then the estimated covariance vector  $S_y$  is uniquely determined (again, up to a scalar multiple) whatever the values of  $a_y$ . If the sample means do not lie on a straight line, then the estimated coefficient vector will not be uniquely determined.<sup>7</sup> In practice, the normalization is unlikely to matter very much if the terms of the theorem are closely approximated if the fit of the equation is very good, or if the iterative least-squares process is used.

---

Note that the choice of 0-1 values for the normalization rather than other constants is not what causes the difficulty; it is the choice of which numerical indices will be unknowns and which corresponding dummy variables will appear as "independent" variables in the regression equation

## REFERENCES

- Anderson, T. W. 1958. *An Introduction to Multivariate Statistical Analysis*. New York: John Wiley and Sons.
- Cooley, W. and Lohnes, P. 1962. *Multivariate Processes for the Behavioral Sciences*. New York: John Wiley, Inc.
- Cooley, W. and Lohnes, P. 1971. *Multivariate Data Analysis*. New York: John Wiley, Inc.
- Courant, P. N., Gramlich, E. M., and Rubinfeld, D. L. March 1980. Why voters support tax limitation amendments: the Michigan case. *National Tax Journal* 33.
- Gramlich, E. M., Rubinfeld, D. L., and Swift, D. 1980. Voter and nonvoter preferences for public spending and spending limits. Institute of Public Policy Studies Discussion Paper.
- Gramm, W. L. Aug. 1973. The labor force decision of married female teachers: a discriminant analysis approach. *Review of Economics and Statistics* IV(3).
- Green, P. and Rao, V. 1973. *Applied Multidimensional Scaling*. New York: Holt, Rinehart and Winston, Inc.
- Guttman, L. 1941. The quantification of a class of attributes: a theory and method of scale construction. In *The Prediction of Personal Adjustment*, P. Horst et al. (eds.) New York: Social Science Research Council.
- Guttman, L. 1950. Components of scale analysis. In *Measurement and Prediction*, Stouffer et al. (eds.) Chapt. 9 Princeton, New Jersey: Princeton University Press.
- Ladd, G. Oct. 1966. Linear probability functions and discriminant functions. *Econometrica*.
- Rao, C. R. 1952. *Advanced Statistical Methods in Biometric Research*. New York: John Wiley, Inc.
- Rubinfeld, D. March 1973. Credit ratings and the market for general obligation municipal bonds. *National Tax Journal* XXVI(1).
- Salkever, D. S. and Seidman, R. L. 1978. A comparison of OLS and two-limit probit estimates for models with ordered trichotomous dependent variables. Unpublished.
- Shepard, R. N., Romney, A. K., and Nerlove, S. B. 1972. *Multidimensional Scaling I and II*. New York: Seminary Press.
- Warner, S. 1963. Multivariate regression of dummy variates under normality assumptions. *Journal of the American Statistical Association* 58:1054-1063.

Accepted 5 December 1980