LOGIT VERSUS DISCRIMINANT ANALYSIS: A SPECIFICATION TEST

by

Andrew W. Lo

11-85

RODNEY L. WHITE CENTER FOR FINANCIAL RESEARCH
The Wharton School
University of Pennsylvania
Philadelphia, PA 19104

Logit Versus Discriminant Analysis:

A Specification Test

Andrew W. Lo

University of Pennsylvania

April 1984

Revised May 1985

Two of the most widely used statistical techniques for analyzing discrete economic phenomena are discriminant analysis (DA) and logit analysis. For purposes of parameter estimation, logit has been shown to be more robust than DA. However, under certain distributional assumptions both procedures yield consistent estimates and the DA estimator is asymptotically efficient. This suggests a natural Hausman specification test of these distributional assumptions by comparing the two estimators. In this paper, such a test is proposed and an empirical example is provided. The finite-sample properties of the test statistic are also explored through some sampling experiments.

This is a revised version of Chapter 4 of my dissertation. I am grateful to Craig MacKinlay, Whitney Newey, and two anonymous referees for many helpful comments. I would like to thank Chris Cavanagh, Jerry Hausman, Daniel McFadden, and Jim Powell for comments on an earlier version of this paper. I also thank Data Resources Inc. for use of their computing facilities and Stephanie Hogue, Gillian Speeth, and Madhavi Vinjamuri for preparing the manuscript. The National Science Foundation and the Alfred P. Sloan Foundation provided much appreciated financial support. Any errors are of course my own.

1. Introduction

Two of the most widely used statistical procedures in empirical studies of discrete economic phenomena are discriminant analysis (DA) and logit analysis. Although distinct, these two methods are closely related as McFadden (1976) has shown. In particular, if y is a discrete variable and X is a vector of "explanatory" continuous variables, logit and DA are alternate means of characterizing the joint distribution of (y, X). DA focuses on the distribution of the X variates conditional on y and, in practice, it is almost always assumed that the distribution of $X|y$ is normal with a common covariance matrix across the y's. In contrast to DA, logit analysis involves the distribution of y conditional on the X's which is assumed to be logistic.

In addition to the essential distinction between causal and conjoint models which McFadden (1976) pointed out, logit and DA are distinguished by another characteristic: logit is more robust than DA. More specifically, it is easily demonstrated that logit analysis is applicable for a wider class of distributions of (y, X) than is normal DA. However, Efron's (1975) study indicates that if the normality of $X|y$ does obtain, then DA is considerably more efficient than logit. Indeed, since the normal DA procedure is in fact maximum likelihood estimation when $X|y$ is normally distributed, it is asymptotically efficient in this case. Nevertheless if normality does not obtain, then the normal DA estimator is inconsistent in general whereas the logit estimator maintains its consistency under a wide class of alternative distributions of (y, X). In a related study, Amemiya and Powell (1983) show that, for purposes of classification, DA does quite well even if the X's are binary (in which case, $X|y$ is clearly not normally distributed). However, they conclude that this may be more likely to hold for discrete X's than for continuous X's which are conditionally nonnormal. Therefore, an important consideration in choosing between logit and DA estimation is whether or not the assumption of conditional normality obtains.

In this paper, a simple Hausman-type specification test for such departures from normality based on the logit and DA estimators is proposed. Under the null hypothesis that the explanatory variables are conditionally normal, the logit and DA estimators should be numerically close. Under the alternative joint hypothesis that $X|y$ is not normal and $y|X$ is logistic, the two estimators should differ since logit is consistent and DA is not. Due to the asymptotic efficiency of the DA estimator, the usual Hausman (1978) specification test is applicable.

Since it seems that the normality of $X|y$ is at issue, why not apply the more standard tests of normality such as the Shapiro-Wilks, skewness, or kurtosis tests? The primary reason is that a rejection of normality on the basis of such tests does not leave the econometrician with a clear alternate method of analysis. Although discriminant analysis may in principle be performed for distributions of $X|y$ other than the normal, this has little practical value due to the intractibility of alternative multivariate distributions. It will be shown in Section 2, however, that logit analysis is appropriate for <u>any</u> distribution of $X|y$ which is a member of the exponential family. Therefore if normality is rejected, logit analysis is the natural alternative. Since a rejection of normality will usually entail the estimation of the logit model anyway, it may be more convenient to estimate the logit model first and use those estimates to test for normality rather than perform an additional test. Furthermore, the proposed Hausman test does not involve additional manipulation of the data as do the normality tests mentioned above, but requires only the parameter estimates from logit and DA.

In Section 2 the relation between logit and DA is reviewed and the parameters of interest are defined. The test statistic is derived in Section 3. In Section 4 an empirical example involving corporate bankruptcies is presented for which the proposed test is performed. For illustrative purposes, two simple sets of sampling experiments were performed and the results are reported in Sections 5 and 6. The first set considers the finite-sample properties of the test under the null hypothesis that $X|y$ is normally distributed, and the second set studies the finite-sample power of the test under the alternative hypothesis that the distribution of $X|y$ is gamma. We conclude in Section 7.

2. The Relation between Logit and Discriminant Analysis

For simplicity, only the bivariate case is considered here although the results extend readily to the general multivariate case. Let y denote a discrete dichotomous random variable which takes the values 0 or 1 and let X be a $(k \times 1)$-vector of related continuous random variables. Denote by $F(y, X)$ the joint distribution function of (y, X).

Although the joint distribution F contains all available information concerning the relation between y and X, it is often more convenient to focus on the conditional distribution of $X|y$. This is the case, for example, when dealing with the standard statistical problem of classification: Given an observation x of attributes X which is generated by one of two probability models indexed by y, decide in an optimal fashion which population x belongs to. The standard DA procedure assumes that the conditional distribution of $X|y$ is multivariate normal with mean μ_y and common covariance Σ. More formally, let $F_D(X|y)$ denote the conditional distribution function of $X|y$ and let $f_D(X|y)$ be the corresponding density function. Then normal DA requires that:

$$f_D(X|y) = (2\pi)^{-k/2}|\Sigma|^{-1/2}\exp[-1/2\,(X - \mu_y)'\Sigma^{-1}(X - \mu_y)] \;. \tag{1}$$

Under these conditions, the solution of the general classification problem takes the particularly simple form based on the well-known linear discriminant function.

The DA procedure may be related to logit analysis through a simple application of Bayes' formula. Let $f_X(X)$ denote the marginal density function of X and let

$$\pi_y = \int f_X(X)P(y|X)dX \;. \tag{2}$$

π_y is simply the marginal distribution of y or, in DA terminology, π_y is the <u>a priori</u> probability of being a member of population y. Letting $F_L(y|X)$ denote the conditional distribution function of $y|X$ and applying Bayes' formula yields:

$$F_L(y|X) = \frac{f_D(X|y)\pi_y}{f_X(X)} \;. \tag{3}$$

Since $f_X(X) = \sum_y \pi_y f_D(X|y)$, equation (3) may be rewritten as:

$$F_L(y=1|X) = \frac{f_D(X|y=1)\pi_y}{\sum_y \pi_y f_D(X|y)} = \frac{1}{1 + \frac{\pi_0}{\pi_1}\frac{f_D(X|y=0)}{f_D(X|y=1)}} \;. \tag{4}$$

Substituting the conditional densities of (1) into equation (4) and simplifying then yields:

$$F_L(y=1|X) = \frac{1}{1 + \exp[-(\alpha + \beta'X)]} \tag{5a}$$

$$\alpha = 1/2\,(\mu_0 - \mu_1)'\Sigma^{-1}(\mu_0 + \mu_1) - \ln\frac{\pi_0}{\pi_1} \tag{5b}$$

$$\beta = \Sigma^{-1}(\mu_1 - \mu_0) . \quad (5c)$$

Equation (5) demonstrates that the required assumptions for normal DA insure that the conditional distribution of $y|X$ is logistic. Because the converse is not true, logit analysis is a more robust procedure. In fact, as Efron (1975) observes logit analysis is appropriate under general exponential family assumptions on $F_D(X|y)$. Specifically, let

$$f_D(X|y) = g(\theta_y, \eta)h(X, \eta)\exp[\theta_y'X] \quad (6)$$

where η is an arbitrary nuisance parameter. Note that (1) is a special case of (6). The conditional density of $y|X$ under (6) is then given by:

$$F_L(y|X) = \frac{1}{1 + \exp\left[-(\alpha + \beta'X)\right]} \quad (7a)$$

$$\alpha = \ln \frac{g(\theta_1, \eta)}{g(\theta_0, \eta)} - \ln \frac{\pi_0}{\pi_1} \quad (7b)$$

$$\beta = \theta_1 - \theta_0 . \quad (7c)$$

Since logit analysis is appropriate for a wider class of distributions than normal DA, a natural test of the normality assumption against other distributions of the exponential family is a comparison of the logit and DA estimators of (α, β) using (5b) and (5c). In particular, the null and alternative hypotheses may be stated explicitly as:

H_0: $f_D(X|y)$ is multivariate normal with parameters μ_y, Σ.

H_1: $f_D(X|y)$ is of the exponential family with parameters θ_y, η.

Such a specification test is developed explicitly in the next section.

3. A Specification Test

Suppose we have T i.i.d observations $[y_1, X(y_1)], \ldots, [y_T, X(y_T)]$ where $X(y_i)$ denotes the vector of attributes associated with the response variable y_i. Define $T_1 = \sum_{i=1}^{T} y_i$ and $T_0 = T - T_1$ and the index sets $I_0 = \{i | y_i = 0\}$ and $I_1 = \{i | y_i = 1\}$. Then under H_0 the joint log-likelihood function of y and X is given by the sum of the conditional and marginal log-likelihood functions:

$$L(y, X) = K_0 - \frac{T}{2} \ln|\Sigma| - \tfrac{1}{2} \sum_{i \varepsilon I_0} [X(y_i) - \mu_0]'\Sigma^{-1}[X(y_i) - \mu_0] - \tfrac{1}{2} \sum_{i \varepsilon I_1} [X(y_i) - \mu_1]'\Sigma^{-1}[X(y_i) - \mu_1] + T_0 \ln\pi_0 + T_1 \ln(1 - \pi_0) . \tag{8}$$

In this case, the DA estimators coincide with the full information maximum likelihood estimators and are given by:

$$\hat{\mu}_0 = \frac{1}{T_0} \sum_{i \varepsilon I_0} X(y_i) \qquad \hat{\mu}_1 = \frac{1}{T_1} \sum_{i \varepsilon I_1} X(y_i) \tag{9a}$$

$$\hat{\Sigma} = \frac{1}{T} [\sum_{i \varepsilon I_0} [X(y_i) - \hat{\mu}_0][X(y_i) - \hat{\mu}_0]' + \sum_{i \varepsilon I_1} [X(y_i) - \hat{\mu}_1][X(y_i) - \hat{\mu}_i]'] \tag{9b}$$

$$\hat{\pi}_0 = \frac{T_0}{T} \qquad \hat{\pi}_1 = \frac{T_1}{T} . \tag{9c}$$

By the principle of invariance, the maximum likelihood estimators of (α, β) under H_0 may be obtained from substituting the estimators (9) into equations (5b, c). Denoting such estimators by $\hat{\alpha}_{DA}$ and $\hat{\beta}_{DA}$, where the subscript indicates that they were obtained from DA estimators, we have:

$$\hat{\alpha}_{DA} = (\hat{\mu}_0 - \hat{\mu}_1)'\hat{\Sigma}^{-1}(\hat{\mu}_0 + \hat{\mu}_1) + \ln \frac{\hat{\pi}_1}{\hat{\pi}_0} \tag{10a}$$

$$\hat{\beta}_{DA} = \hat{\Sigma}^{-1}(\hat{\mu}_1 - \hat{\mu}_0) \ . \tag{10b}$$

Although $\hat{\alpha}_{DA}$ and $\hat{\beta}_{DA}$ are consistent and asymptotically efficient under H_0, they are generally inconsistent under the alternative hypothesis H_1. A Hausman-type specification test of H_0 may then be constructed by taking the difference of the DA estimator and an alternative estimator which is consistent under both hypotheses. One such estimator may be obtained by maximizing the conditional logistic log-likelihood function of y conditioned on X:

$$L(y|X) = \sum_{i=1}^{T} \left[(\alpha + X_i'\beta)(y_i - 1) - \ln\left[1 - e^{-(\alpha+X_i'\beta)}\right]\right] \tag{11}$$

Denote these estimators $\hat{\alpha}_L$ and $\hat{\beta}_L$. For analytic convenience, we choose to ignore the intercept term α in the remaining analysis and base our test only on β. Let $\hat{V}_{DA}$ and $\hat{V}_L$ be the estimated asymptotic covariance matrices of the β estimators from DA and logit respectively. Following Hausman (1978), we let $\hat{q} = \hat{\beta}_L - \hat{\beta}_{DA}$, and form the χ^2 statistic:

$$J = T\,\hat{q}'[\hat{V}_L - \hat{V}_{DA}]^{-1}\hat{q} \overset{A}{\sim} \chi_k^2 \ . \tag{12}$$

A consistent estimator of the asymptotic covariance matrix for $\hat{\beta}_L$ may be obtained from the estimated Hessian of the likelihood function in the usual way. A consistent estimator of V_{DA} may be constructed by first calculating the asymptotic covariance matrix of $\hat{\beta}_{DA}$. This is done in Appendix 1 and is given by the formula:

$$V_{DA} = \delta\Sigma^{-1} + 2(\mu' \otimes I)(\Sigma^{-1} \otimes \Sigma^{-1})R'Q(\Sigma \otimes \Sigma)Q'R(\Sigma^{-1} \otimes \Sigma^{-1})(\mu \otimes I) \tag{13}$$

where $\mu \equiv \mu_1 - \mu_0$, $\delta \equiv \frac{1}{\pi_0\pi_1}$, and R and Q are standard selection matrices as defined in Richard (1975) (see Appendix 1). A consistent estimator $\hat{V}_{DA}$ is then obtained by substituting consistent estimators $(\hat{\mu}, \hat{\Sigma}, \hat{\pi}_0, \hat{\pi}_1)$, of $(\mu, \Sigma, \pi_0, \pi_1)$ into equation (13). Note that once the DA estimators of μ and Σ are computed, the computation of $\hat{V}_{DA}$ using (13) involves only simple matrix multiplications requiring few additional steps in standard econometric software packages such as RATS or TSP. In addition, since the estimate $\hat{V}_L$ is standard output in most maximum likelihood and logit computer programs, the proposed specification test (12) is in practice quite easy to construct. Of course, the power of this test against H_1 will differ across different members of the exponential family. To illustrate the practical value of this test, the next section considers an empirical example involving an analysis of corporate bankruptcies.

4. Empirical Analysis of Corporate Bankruptcies.

In this section, we apply the proposed specification test for logit versus DA estimation of corporate bankruptcies. Previous empirical studies of business failures using DA are too numerous to cite and the reader is referred to Scott (1981) for an excellent review and critique of the empirical literature. Logit analysis however has only recently been applied to the study of default and, to this author's knowledge, Martin (1977), Ohlson (1980) and Zavgren (1980) have been the only studies.

Because there is no single canonical source of data for bankrupt firms, the procedure for compiling the sample of failed firms requires some explanation. An initial sample of 184 firms was extracted from Standard and Poor's COMPUSTAT Industrial Research File using the bankruptcy deletion code '02' as the extraction criterion.[6] From this sample, firms in the financial industries (SIC code 6000 to 6999) were excluded leaving 168 firms. This subset was then reduced to 77 firms by including only those firms for which some data was available between the years 1975 and 1983 inclusively. The firms in this subset were then checked for the type of bankruptcy proceedings filed using the Wall Street Journal Index (WSJI) and the Directory of Obsolete Securities (DOS) and only those firms which explicitly filed under Chapter X or XI were included, yielding the final sample of 38 bankrupt firms. The actual year of bankruptcy filing was tabulated for each firm at this point using the WSJI and DOS. The final set of financial rations for each firm was then constructed by using data at least one year and at most three years prior to the actual year of bankruptcy.[7]

Most DA studies of bankruptcy use samples composed of pairs of bankrupt and non-bankrupt firms matched by industry and year of failure. This procedure clearly introduces much sample-selection bias (for example, the maximum likelihood estimate of the unconditional probability of bankruptcy $\hat{\pi}_1$ will always be $\frac{1}{2}$ for a matched sample) and is discussed at greater length in Martin (1977) and Zavgren (1980). This method of constructing the sample is followed so as to render the results more readily comparable to the existing literature. A matching sample of 38 non-bankrupt firms was extracted from the COMPUSTAT Industrial Annual File where firms were matched by year of observation, industry and, when possible, total sales. The final sample thus consists of 38 bankrupt firms and 38 solvent firms yielding 76 data points in all. Tables 1 and 2 list these firms, the year in which the data was drawn, and the year of bankruptcy. Appendix 2 provides summary statistics.

TABLE 1. SAMPLE OF BANKRUPT FIRMS

Firm	Industry	Year of Data	Year of Bankruptcy
Frigitemp	1700	1976	1978
Tobin Packing	2010	1980	1981
CS Group	2300	1980	1982
Garland	2300	1979	1980
Lynnwear	2300	1979	1981
Nelly Don	2300	1977	1978
Poloron Products	2450	1979	1981
Brody (B.) Seating	2510	1979	1981
Saxon Industries	2600	1980	1982
Supronics	2844	1975	1977
Acme-Hamilton Mfg.	3069	1977	1978
Frier Industries	3140	1975	1978
Maule Industries	3270	1975	1976
Universal Containers	3410	1976	1978
Randal Data Systems	3573	1979	1980
Advent	3651	1979	1981
GRT	3652	1977	1979
Allied Technology	3662	1979	1980
Gladding	3662	1976	1977
Hy-Gain Electronics	3662	1977	1978
Multronics	3662	1979	1980
DAIG	3693	1980	1981
Medcor	3693	1980	1981
Allied Artists Industries	3716	1977	1979
Reinell Industries	3730	1976	1979
Gruen Industries	3870	1975	1977
Mego Industries	3940	1980	1982
Miner Industries	3940	1975	1977
Auto-Train	4013	1979	1980
Cooper-Jarrett	4210	1981	1982
Nelson Resource	4210	1978	1981
Pacific Far East Line	4400	1977	1978
Shulman Transport Enterprise	4700	1977	1978
Fireco Sales	5099	1980	1982
Gilman Services	5120	1980	1982
Ormont Drug & Chemical	5120	1975	1977
Filigree Foods	5140	1975	1976
Research Fuels	5199	1978	1979

TABLE 2. MATCHING SAMPLE OF SOLVENT FIRMS

Southwest Forest Industries	Industry
Amelco	1700
Sunstar Foods	2010
Wolf (Howard B)	2300
Movie Star	2300
Beeline	2300
Madison Industries	2300
De Rose Industries	2450
Jensen Industries	2510
Southwest Forest Industries	2600
Roffler Industries	2844
Mark IV Industries	3069
Lama (Tony)	3140
Florida Rock Industries	3270
Plan Industries	3410
Access	3570
Esquire Radio & Electron	3651
Electrosound Group	3652
Alarm Products Intl.	3662
Communications Industries	3662
AEL Industries	3662
Watkins-Johnson	3662
Staodynamics	3693
Healthdyne	3693
Executive Industries	3716
Uniflite	3730
Talley Industries	3870
Ohio Art	3940
Empire of Carolina	3940
Falls City Industries	4210
Eazor Express	4210
Arnold Industries	4210
Overseas Shipholding Group	4400
Dereco	4700
Ronco Teleproducts	5099
Napco Industries	5120
Krelitz Industries	5120
Distribuco	5140
Nolex	5199

In addition to a constant term, six explanatory variables were included in the logit estimation and are defined in Table 3.

Variable Names and Definitions

Variable	Definition
SIZE	Log(Total Assets/GNP price deflator)[8]
CLTA	Current Debt Liabilities divided by Total Assets
OLTA	Other Debt Liabilities divided by Total Assets
CATA	Current Assets divided by Total Assets
NITA	Net Income divided by Total Assets
BANK	Bankruptcy Index suggested by Lo (1984)

Table 3.

The first five variables were chosen because of the frequency with which they appear in other empirical studies of bankruptcy.[9] The sixth variable included was suggested by the theoretical model developed in Lo (1984) in which the key determinant of bankruptcy was whether or not the value of cash y plus the value of intangible assets or goodwill G exceeded the current debt obligations c. The BANK variable is essentially the ratio of cash plus intangibles to current debt liabilities $\frac{G + y}{c}$.[10]

Maximum likelihood logit estimation was performed on the 76 data points with the likelihood function given in equation (10) using the MLOGIT computer package developed by Bronwyn H. Hall.[11] The corresponding DA estimates and test statistic were computed using FORTRAN software written by the author.[12] Due to MLOGIT constraints, the dependent variable indicating the status of solvency was defined to be 1 if bankrupt and 2 if solvent and solvency was the normalized alternative. The logit and DA estimation results are reported in Table 4 with asymptotic standard errors enclosed in parentheses. Table 4 indicates that the same three variables are significant at the 5% level or better for both logit and DA; OLTA, NITA, and BANK. In contrast to Ohlson's (1980) estimated SIZE coefficient, which is significant at the 1% level, the SIZE coefficient estimated here is insignificant. This may be due to the fact that larger firms are often acquired or reorganized when faced with financial distress and do not file for bankruptcy whereas smaller firms are not included in the COMPUSTAT database, leaving the sample of bankrupt and solvent firms with little systematic variation in SIZE.[13] That the ratio of current assets to total assets CATA is insignificant does not contradict the model presented in Lo (1984) since in that framework the probability of default is determined by the ratio of cash plus intangibles to current liabilities and is unaffected by the cash to intangibles plus cash ratio. The fact that CLTA is also insignificant may seem inconsistent with our theoretical framework since current liabilities directly affects the default trigger. However, because the data were collected one to three years prior to bankruptcy, this result may be reasonable. This is also supported by the fact that the ratio of other debt liabilities to total assets OLTA is significant since "other liabilities" includes debt obligations maturing in the actual year of bankruptcy.

TABLE 4. LOGIT AND DA ESTIMATION RESULTS.

VARIABLE	LOGIT ESTIMATE (STD. ERROR)	DA ESTIMATE (STD. ERROR)
CONSTANT	1.2140 (2.5162)	- (-)
SIZE	-0.0441 (0.2977)	-0.0418 (0.2811)
CLTA	0.1074 (2.4177)	-2.3803 (2.0231)
OLTA	-3.3258 (1.8103)	-2.6131 (1.5159)
CATA	-1.2321 (1.9827)	-0.6597 (1.7085)
NITA	10.7971 (4.8249)	5.1616 (2.2561)
BANK	4.1642 (2.2755)	1.0686 (0.6245)

CONVERGENCE CRITERION ON EACH PARAMETER = 0.001000

CONVERGENCE CRITERION ON SUMS OF SQUARES = 0.000100

VALUE OF LOG-LIKELIHOOD AT CONVERGENCE = -30.662786

Since the solvent alternative was normalized, positive coefficients imply that higher values for the associated variable correspond to larger probabilities of solvency. In terms of the significant variables, we infer that:

(i) A larger non-current debt to total assets ratio increases the probability of default.

(ii) A larger net income to total assets ratio decreases the probability of default.

(iii) A larger cash plus intangibles to current debt ratio decreases the probability of default.

Implication (i) has been discussed above. Implications (ii) and (iii) seem to support Lo's (1984) multi-period model of bankruptcy.

Of course, all interpretations of these results in terms of the theoretical model should be seen as suggestive at best since no structural stochastic specification of bankruptcy has been made. In some cases, it may be possible to derive a logistic specification from economic behavior but this has not been considered in this study.[14] In addition, the accounting variables used may only loosely correspond to their theoretical quantities if at all. There is also the timing problem mentioned above.

Given the estimated parameters from the DA and logit analysis estimation, the Hausman test statistic is easily computed to be 6.2105 which is χ^2 with 7 degrees of freedom and has a corresponding p-value of 0.515. Since H_0 cannot be rejected at any level better than 49% it seems that the data support the normality of $X|y$ against other distributions of the exponential class. This suggests that DA is the preferred method of estimation since it is asymptotically more efficient. In fact, Efron's (1975) calculations relate the asymptotic relative efficiency (ARE) of the two procedures to the square root of the Mahalanobis distance $\Delta \equiv \left[(\mu_1 - \mu_0)\Sigma^{-1}(\mu_1 - \mu_0)\right]^{1/2}$ implying that a value of 4 for Δ yields an ARE of 0.343. The estimated value of Δ for the data set used here is 1.5248 which, according to (1.12) in Efron (1975), corresponds roughly to an ARE of 0.968. Although the Hausman test supports the appropriateness of DA, the estimated loss in efficiency of logit under H_0 is less than 4%.

5. Finite Sample Properties of the Specification Test.

In this section, we present the results of several sampling experiments which explore the finite-sample properties of the specification test statistic J proposed in Section 3 under the null hypothesis that $X|y$ is normally distributed. Since these simulations involve only a single X (in addition to the constant term) and consider only one of several interesting alternative hypotheses, these simulation results are meant only to be suggestive.

Under the null hypothesis H_0, $X|y$ is normal with mean u_y and variance σ^2, where y takes on the values 0 or 1. The unconditional probabilities π_0 and π_1 are assumed to be $1/2$ throughout the simulations. Given numerical values for μ_0, μ_1, and σ^2, a random sample of observations (y_1, X_1, ..., y_T, X_T) of size T is generated in the following manner: y_1 is first generated as an outcome of a Bernoulli random variate with $p = 1/2$, then X_1 is generated as an outcome of a normal variate with mean μ_y and variance σ^2, and so on for observations 2 to T. The logit and DA estimators $\hat{\beta}_L$ and $\hat{\beta}_{DA}$, their estimated asymptotic variances $\hat{V}_L$ and $\hat{V}_{DA}$, and the test statistic J are then computed for the sample. The following parameter values were assumed and held constant throughout all experiments:

$$\mu_0 = 0.20 \qquad \sigma^2 = 0.01 .$$

Experiments were then performed for sample sizes of 50, 100, 200, and 300, and for various values of the parameter μ_1. More specifically, because Efron (1975) has shown that the asymptotic relative efficiency of logit versus DA is related to the square root of the Mahalanobis distance, successive values for μ_1 were chosen to vary the Mahalanobis distance from 5.0 to 9.0 in unit increments. Asymptotic relative efficiency is particularly relevant for the specification test of Section 3 because as the ARE approaches unity, the matrix difference $[\hat{V}_L - \hat{V}_{DA}]$ is more likely to be non-positive-definite in finite samples even though the logit estimator is asymptotically less efficient than DA. Although performing the Hausman test in such situations is problematic, Newey (1983) has developed a general method of constraining the difference of the two covariance matrices to be positive definite. However, the use of Newey's approach involves deriving both the consistent and efficient estimators as generalized method of moments estimators and, unfortunately in the case of logit and DA, this would sacrifice much of the computational simplicity of the proposed test. In these experiments, samples which yield negative variance differences were simply discarded and replications continued until 1000 samples with positive variance differences were obtained. This procedure obviously introduces serious biases into our simulations when the logit estimator is "close" to the DA estimator in efficiency and when the sample size is small. Indeed, the fraction of the 1000 replications with a non-positive variance difference often exceeded 20 percent for experiments with Mahalanobis distances less than 5.0 and are not reported here due to their unreliability.[15] However, for larger values of the Mahalanobis distance and for sample sizes above 100, the simulation results are more reliable. As an indication of the seriousness of the nonpositive variance difference problem in each experiment, the number of samples with

TABLE 5a

Performance of specification test for sample size of 50 observations.

T (%-Bad Draws)	Δ^2	β	$\bar{\hat{\beta}}_{DA}$ (SE)	$\bar{\hat{\beta}}_L$ (SE)	V_{DA}	$\bar{V}_{DA}$ (SE)	$\bar{\tau} \times 10$ (t-stat)	$\overline{SE}$ (τ) (t-stat)[a]	0.01 tail (t-stat)[b]	0.05 tail (t-stat)	0.10 tail (t-stat)
50 (12.9)	5.0	22.36	24.62 (5.93)	28.42 (14.66)	1400.0	1717 (730.2)	0.3496 (1.11)	0.9265 (-3.29)	0.015 (1.59)	0.025 (-3.63)	0.031 (-7.27)
50 (10.8)	6.0	24.49	26.67 (6.07)	33.88 (48.34)	1600.0	1928 (786.9)	0.4004 (1.27)	0.7810 (-9.80)	0.015 (1.59)	0.029 (-3.05)	0.039 (-6.43)
50 (10.2)	7.0	26.46	28.51 (6.22)	34.57 (21.83)	1800.0	2131 (845.6)	-0.3281 (-1.04)	1.443 (19.80)	0.0170 (2.22)	0.030 (-2.90)	0.046 (-5.69)
50 (10.4)	8.0	28.28	30.18 (6.48)	38.98 (33.50)	2000.0	2331 (928.4)	-0.6544 (-2.07)	1.262 (11.71)	0.018 (2.54)	0.030 (-2.90)	0.039 (-6.43)
50 (13.7)	9.0	30.00	31.49 (6.59)	45.30 (121.4)	2200.0	2491 (982.5)	-0.9058 (-2.86)	1.068 (3.06)	0.021 (3.50)	0.034 (-2.32)	0.043 (-6.01)

[a]Asymptotic t-statistic for the hypothesis that the true standard deviation is unity, calculated as $[(\overline{SE} - 1)(2N)]^{1/2}$ where N = 1000 is the number of replications.

[b]Asymptotic t-statistic for the hypothesis that the true proportion is 0.01, calculated as $(\hat{p} - p)/[p(1 - p)/N]^{1/2}$ where $\hat{p}$ is the estimated proportion, p is the true proportion 0.01 and N = 1000 is the number of replications. The t-statistics for the 0.05 and 0.10 tails are calculated similarly.

TABLE 5b

Performance of specification test for sample size of 100 observations.

T (%-Bad Draws)	Δ^2	β	$\hat{\beta}_{DA}$ (SE)	$\hat{\beta}_L$ (SE)	V_{DA}	$\hat{V}_{DA}$ (SE)	$\tau \times 10$ (t-stat)	SE (τ) (t-stat)[a]	0.01 tail (t-stat)[b]	0.05 tail (t-stat)	0.10 tail (t-stat)
100 (13.2)	5.0	22.36	23.65 (4.10)	24.76 (5.58)	1400.0	1572 (463.6)	0.0860 (0.272)	1.006 (0.262)	0.028 (5.72)	0.045 (-0.73)	0.072 (-2.95)
100 (9.0)	6.0	24.49	25.85 (4.33)	27.39 (6.69)	1600.0	1793 (526.4)	-0.5453 (-1.72)	1.435 (19.47)	0.029 (6.04)	0.049 (-0.15)	0.066 (-3.58)
100 (6.4)	7.0	26.46	27.84 (4.62)	30.05 (8.85)	1800.0	2011 (597.7)	-0.9019 (-2.85)	1.961 (42.96)	0.032 (6.99)	0.044 (-0.871)	0.065 (-3.69)
100 (4.7)	8.0	28.28	29.68 (4.86)	32.65 (10.09)	2000.0	2228 (664.0)	-0.3854 (-1.22)	1.094 (4.21)	0.028 (5.72)	0.051 (0.145)	0.065 (-3.69)
100 (3.9)	9.0	30.00	31.47 (5.07)	36.00 (19.47)	2200.0	2450 (729.2)	-0.2202 (-0.696)	1.029 (1.32)	0.027 (5.40)	0.043 (-1.02)	0.066 (-3.58)

[a] Asymptotic t-statistic for the hypothesis that the true standard deviation is unity, calculated as $[(\overline{SE} - 1)(2N)]^{1/2}$ where N = 1000 is the number of replications.

[b] Asymptotic t-statistic for the hypothesis that the true proportion is 0.01, calculated as $(\hat{p} - p)/[p(1 - p)/N]^{1/2}$ where $\hat{p}$ is the estimated proportion, p is the true proportion 0.01 and N = 1000 is the number of replications. The t-statistics for the 0.05 and 0.10 tails are calculated similarly.

TABLE 5c

Performance of specification test for sample size of 200 observations.

T (%-Bad Draws)	Δ^2	β	$\hat{\beta}_{DA}$ (SE)	$\hat{\beta}_L$ (SE)	V_{DA}	$\hat{V}_{DA}$ (SE)	$\tau \times 10$ (t-stat)	SE (τ) (t-stat)[a]	0.01 tail (t-stat)[b]	0.05 tail (t-stat)	0.10 tail (t-stat)
200 (6.2)	5.0	22.36	22.96 (2.75)	23.42 (3.43)	1400.0	1479 (295.8)	-0.8180 (-2.59)	1.409 (18.30)	0.030 (6.36)	0.055 (0.726)	0.092 (-0.843)
200 (4.2)	6.0	24.49	25.11 (2.94)	25.74 (3.96)	1600.0	1687 (339.9)	-0.447 (-1.31)	1.238 (10.65)	0.032 (6.99)	0.056 (0.871)	0.085 (-1.58)
200 (2.5)	7.0	26.46	27.09 (3.11)	27.91 (4.51)	1800.0	1896 (382.9)	0.6564 (-2.08)	1.297 (13.30)	0.032 (6.99)	0.053 (0.435)	0.084 (-1.69)
200 (1.7)	8.0	28.28	28.95 (3.28)	30.04 (5.17)	2000.0	2107 (427.3)	-0.4017 (-1.27)	1.286 (12.80)	0.031 (6.67)	0.055 (0.726)	0.082 (-1.90)
200 (0.9)	9.0	30.00	30.70 (3.44)	32.15 (5.96)	2200.0	2317 (470.6)	-0.5474 (-1.73)	1.556 (24.87)	0.032 (6.99)	0.055 (0.726)	0.081 (-2.00)

[a]Asymptotic t-statistic for the hypothesis that the true standard deviation is unity, calculated as $[(\widetilde{SE} - 1)(2N)]^{1/2}$ where N = 1000 is the number of replications.

[b]Asymptotic t-statistic for the hypothesis that the true proportion is 0.01, calculated as $(\hat{p} - p)/[p(1 - p)/N]^{1/2}$ where $\hat{p}$ is the estimated proportion, p is the true proportion 0.01 and N = 1000 is the number of replications. The t-statistics for the 0.05 and 0.10 tails are calculated similarly.

TABLE 5d

Performance of specification test for sample size of 300 observations.

T (%-Bad Draws)	Δ^2	β	$\hat{\beta}_{DA}$ (SE)	$\hat{\beta}_L$ (SE)	V_{DA}	$\hat{V}_{DA}$ (SE)	$\tau \times 10$ (t-stat)	SE (τ) (t-stat)[a]	0.01 tail (t-stat)[b]	0.05 tail (t-stat)	0.10 tail
300 (3.2)	5.0	22.36	22.73 (2.21)	23.02 (2.75)	1400.0	1450 (233.1)	-0.8941 (-2.83)	1.196 (8.75)	0.029 (6.04)	0.057 (1.02)	0.089 (-1.16)
300 (1.1)	6.0	24.49	24.89 (2.35)	25.28 (3.15)	1600.0	1657 (267.1)	-0.8254 (-2.61)	1.193 (8.65)	0.033 (7.33)	0.053 (0.435)	0.083 (-1.79)
300 (0.4)	7.0	26.46	26.87 (2.50)	27.40 (3.61)	1800.0	1862 (302.3)	-0.7345 (-2.32)	1.264 (11.82)	0.029 (6.04)	0.052 (0.290)	0.089 (-1.16)
300 (0.1)	8.0	28.28	28.71 (2.64)	29.42 (4.13)	2000.0	2069 (337.8)	-0.3095 (-0.979)	1.081 (3.62)	0.025 (4.77)	0.051 (0.145)	0.085 (-1.58)
300 (0.0)	9.0	30.00	30.46 (2.77)	31.42 (4.74)	2200.0	2277 (372.6)	-0.3154 (-0.998×10^{-4})	1.014 (0.646)	0.022 (3.81)	0.049 (0.145)	0.076 (-2.53)

[a]Asymptotic t-statistic for the hypothesis that the true standard deviation is unity, calculated as $[(\overline{SE} - 1)(2N)]^{1/2}$ where N = 1000 is the number of replications.

[b]Asymptotic t-statistic for the hypothesis that the true proportion is 0.01, calculated as $(\hat{p} - p)/[p(1 - p)/N]^{1/2}$ where $\hat{p}$ is the estimated proportion, p is the true proportion 0.01 and N = 1000 is the number of replications. The t-statistics for the 0.05 and 0.10 tails are calculated similarly.

negative differences as a percentage of the total number of replications performed is reported. Tables 5a-d summarize the results of the simulations. Each row in Tables 5 corresponds to a separate independent experiment. The first column indicates the sample size T and, in parentheses, the percentage of the replications which were discarded (and replaced) because of a non-positive variance difference in the J-statistic. The second and third columns indicate the theoretical values for the Mahalanobis distance and β respectively. The fourth and fifth columns report the mean and standard errors of the DA and logit estimates of β respectively. Column six gives the theoretical value for the asymptotic variance of the DA estimator and column seven reports the mean and standard error of the estimated asymptotic variance. Since the J-statistic is χ^2 with one degree of freedom under the null, it may be transformed into a N(0, 1) random variable τ. The mean of the τ-statistic across the 1000 replications is reported in column eight with asymptotic t-statistics for the hypothesis that the true mean is zero given in parentheses. In column nine, the standard error of τ over the replications is given with asymptotic t-statistics for the hypothesis that the true standard deviation is unity given in parentheses. The last three columns report estimated 1, 5, and 10 percent tail probabilities respectively. Asymptotic t-statistics for the hypotheses that the true tail probabilities are 1, 5, and 10 percent respectively are reported in parentheses.

Tables 5a-d are largely self-explanatory. It is clear from the tables that the difference of the variances in the J-statistic is nonpositive for a significant fraction the replications when the Mahalanobis distance is small. For example, Table 5a reports that for sample sizes of 50 and a Mahalanobis distance of 5.0, 1148 replications were required, with close to 13 percent of the replications having negative variance differences. However, Tables 5c and d indicate that with sample sizes of 200 or more the number of replications with negative variance differences declines considerably, and declines monotonically as the Mahalanobis distance increases.

The majority of the experiments yielded means of the τ-statistic which were insignificantly different from zero. However, for almost all the experiments the hypothesis that the true standard deviation is unity can be rejected. Nevertheless, the estimated size of a 5%-test for sample sizes of 100 or more are not statistically different from 0.05. Although the 5%-test seems well-behaved, tests at the 1% and 10% level have estimated sizes which do differ significantly from 0.01 and 0.10 respectively, the 1%-test rejecting too frequently and the 10%-test rejecting less often than it should.

6. Power of the Specification Test.

In this section we investigate the power of the proposed specification test against the alternative hypothesis that the distribution of $X|y$ is gamma with parameters (η, λ_y) where the density function is given by:

$$f(X|y) = \frac{X^{\eta-1}}{\lambda_y^{\eta}\Gamma(\eta)} \exp\left[-\frac{X}{\lambda_y}\right] . \tag{14}$$

Since the gamma distribution approaches the normal distribution as η approaches infinity, we may examine the power of the test as the alternative hypothesis becomes "closer" to the null by simply increasing η.

Note that, given the density function in (14), the parameter in (7c) is now given by:

$$\beta = \frac{\lambda_1 - \lambda_0}{\lambda_0\lambda_1} . \tag{15}$$

In these simulations, it is also assumed throughout that the unconditional

Table 6a

Power of specification test for sample size of 50 observations.[a]

T (%-Bad Draws)	η	β	$\bar{\hat{\beta}}_{DA}$ (SE)	$\bar{\hat{\beta}}_{L}$ (SE)	$\bar{\tau}$ (SE)	Power-10% (SE)	Power-5% (SE)	Power-1% (SE)
50 (11.7)	2.0	0.[illegible]	0.222 (0.08)	0.284 (0.12)	0.670 (0.72)	0.054 (0.007)	0.026 (0.005)	0.011 (0.003)
50 (11.4)	4.0	0.250	0.220 (0.07)	0.278 (0.10)	0.749 (0.78)	0.071 (0.008)	0.036 (0.006)	0.012 (0.003)
50 (8.3)	6.0	0.250	0.221 (0.07)	0.285 (0.10)	0.760 (0.85)	0.098 (0.009)	0.042 (0.006)	0.012 (0.003)
50 (7.1)	8.0	0.250	0.221 (0.06)	0.296 (0.16)	0.764 (0.82)	0.089 (0.009)	0.039 (0.006)	0.008 (0.003)
50 (3.6)	10.0	0.250	0.223 (0.07)	0.308 (0.15)	0.695 (0.83)	0.074 (0.008)	0.035 (0.006)	0.007 (0.003)
50 (4.2)	12.0	0.250	0.221 (0.06)	0.319 (0.29)	0.635 (0.66)	0.039 (0.006)	0.018 (0.004)	0.006 (0.002)
50 (4.7)	14.0	0.250	0.221 (0.06)	0.325 (0.20)	0.531 (0.81)	0.042 (0.006)	0.017 (0.004)	0.010 (0.003)
50 (5.3)	16.0	0.250	0.218 (0.06)	0.329 (0.23)	0.434 (1.25)	0.023 (0.005)	0.011 (0.003)	0.007 (0.003)

[a]Standard errors for power estimates p are calculated as $[p(1 - p)/1000]^{1/2}$.

Table 6b

Power of specification test for sample size of 100 observations.[a]

T (%-Bad Draws)	η	β	$\bar{\hat{\beta}}_{DA}$ (SE)	$\bar{\hat{\beta}}_{L}$ (SE)	$\bar{\tau}$ (SE)	Power-10% (SE)	Power-5% (SE)	Power-1% (SE)
100 (6.2)	2.0	0.250	0.211 (0.06)	0.266 (0.08)	1.063 (0.81)	0.160 (0.012)	0.093 (0.009)	0.032 (0.006)
100 (4.9)	4.0	0.250	0.212 (0.05)	0.265 (0.06)	1.237 (0.85)	0.254 (0.014)	0.130 (0.011)	0.038 (0.006)
100 (4.7)	6.0	0.250	0.210 (0.04)	0.262 (0.06)	1.263 (0.91)	0.296 (0.014)	0.162 (0.012)	0.044 (0.007)
100 (3.0)	8.0	0.250	0.209 (0.04)	0.264 (0.06)	1.196 (1.08)	0.290 (0.014)	0.153 (0.011)	0.035 (0.006)
100 (2.3)	10.0	0.250	0.212 (0.04)	0.272 (0.07)	1.148 (0.89)	0.251 (0.014)	0.133 (0.011)	0.034 (0.006)
100 (1.6)	12.0	0.250	0.212 (0.04)	0.277 (0.07)	1.113 (0.85)	0.231 (0.013)	0.111 (0.010)	0.027 (0.005)
100 (0.8)	14.0	0.250	0.211 (0.04)	0.278 (0.08)	1.003 (0.85)	0.168 (0.012)	0.079 (0.009)	0.016 (0.004)
100 (0.6)	16.0	0.250	0.211 (0.04)	0.282 (0.09)	0.878 (1.69)	0.140 (0.011)	0.066 (0.008)	0.013 (0.004)

[a]Standard errors for power estimates p are calculated as $[p(1 - p)/1000]^{1/2}$.

Table 6c

Power of specification test for sample size of 200 observations.[a]

T (%-Bad Draws)	η	β	$\overline{\hat{\beta}}_{DA}$ (SE)	$\overline{\hat{\beta}}_{L}$ (SE)	$\overline{\tau}$ (SE)	Power-10% (SE)	Power-5% (SE)	Power-1% (SE)
200 (2.0)	2.0	0.250	0.206 (0.04)	0.256 (0.05)	1.599 (0.75)	0.422 (0.016)	0.250 (0.014)	0.087 (0.009)
200 (2.1)	4.0	0.250	0.207 (0.03)	0.259 (0.04)	1.953 (0.90)	0.651 (0.015)	0.453 (0.016)	0.172 (0.012)
200 (1.3)	6.0	0.250	0.204 (0.03)	0.256 (0.04)	1.991 (0.85)	0.667 (0.015)	0.498 (0.016)	0.211 (0.013)
200 (0.6)	8.0	0.250	0.207 (0.03)	0.260 (0.04)	1.935 (0.82)	0.648 (0.015)	0.471 (0.016)	0.193 (0.013)
200 (0.5)	10.0	0.250	0.205 (0.03)	0.257 (0.04)	1.785 (0.84)	0.587 (0.016)	0.426 (0.016)	0.151 (0.011)
200 (0.3)	12.0	0.250	0.205 (0.03)	0.258 (0.04)	1.663 (0.88)	0.524 (0.016)	0.370 (0.015)	0.127 (0.011)
200 (0.1)	14.0	0.250	0.204 (0.03)	0.261 (0.04)	1.622 (0.83)	0.507 (0.016)	0.327 (0.015)	0.101 (0.010)
200 (0.0)	16.0	0.250	0.205 (0.03)	0.262 (0.05)	1.457 (0.82)	0.430 (0.016)	0.260 (0.014)	0.066 (0.008)

[a]Standard errors for power estimates p are calculated as $[p(1 - p)/1000]^{1/2}$.

Table 6d

Power of specification test for sample size of 300 observations.[a]

T (%-Bad Draws)	η	β	$\hat{\beta}_{DA}$ (SE)	$\hat{\beta}_L$ (SE)	$\bar{\tau}$ (SE)	Power-10% (SE)	Power-5% (SE)	Power-1% (SE)
300 (0.6)	2.0	0.250	0.206 (0.03)	0.257 (0.04)	2.009 (0.75)	0.689 (0.015)	0.481 (0.016)	0.173 (0.012)
300 (0.5)	4.0	0.250	0.204 (0.03)	0.253 (0.03)	2.407 (0.85)	0.839 (0.012)	0.709 (0.014)	0.392 (0.015)
300 (0.5)	6.0	0.250	0.204 (0.03)	0.254 (0.03)	2.470 (0.95)	0.849 (0.011)	0.740 (0.014)	0.434 (0.016)
300 (0.2)	8.0	0.250	0.203 (0.02)	0.254 (0.03)	2.469 (0.92)	0.862 (0.011)	0.731 (0.014)	0.418 (0.016)
300 (0.3)	10.0	0.250	0.204 (0.02)	0.255 (0.03)	2.248 (0.80)	0.794 (0.013)	0.649 (0.015)	0.335 (0.015)
300 (0.1)	12.0	0.250	0.205 (0.02)	0.258 (0.03)	2.148 (0.79)	0.747 (0.014)	0.598 (0.016)	0.298 (0.015)
300 (0.0)	14.0	0.250	0.204 (0.02)	0.258 (0.03)	1.983 (0.81)	0.695 (0.015)	0.558 (0.016)	0.212 (0.013)
300 (0.2)	16.0	0.250	0.204 (0.02)	0.258 (0.04)	1.827 (0.82)	0.603 (0.016)	0.443 (0.016)	0.189 (0.012)

[a]Standard errors for power estimates p are calculated as $[p(1 - p)/1000]^{1/2}$.

probabilities π_0 :d π_1 of the dichotomous population indicator y are both $^1/_2$. As before, the construction of the random samples i͏r ͏olve first an outcome of a Bernoulli trial with $p = {}^1/_2$ for y, and then a draw from a gamma distribution with parameters (η, λ_y). The parameters λ_0 and λ_1 are held at 2.0 and 4.0 respectively for all experiments. Experiments were conducted for sample sizes of 50, 100, 200, and 300, and for values of η from 1.0 to 16.0 in unit increments. Tables 6a-d summarize the results of these simulations.

The first column of Tables 6 indicate the sample size T and, in parentheses, the percentage of negative-variance difference draws. The second and third columns display the theoretical values of η and β respectively. Columns four and five report the means and standard errors of the DA and logit estimates of β respectively. In column six the mean and standard error of the τ-statistic estimates are reported and the last three columns display respectively the estimated power and associated standard errors of 10, 5, and 1 percent tests.

As in the null simulations for sample sizes of 50, the results for the alternative simulations in Table 6a may be unreliable due to the large proportion of negative variance-difference draws. However, for sample sizes of 100, the largest fraction of bad draws is only 6.2 percent and only 0.6 percent for 300-observation samples. By and large, the tests seem to perform well for sample sizes of 100 or larger. For example, Table 6d shows that the power of the 5 percent test at $\eta = 2.0$ is 0.48 and reaches a peak of 0.74 for $\eta = 6.0$. As η increases beyond 6.0, the power declines as expected since the alternative distribution is moving closer to ͏rmality. This pattern is characteristic of all the simulations; the test power increases monotonically as η increases to 6.0 and generally declines monotonically with increases in η thereafter.

The results of Sections 5 and 6 seem to indicate that for sample sizes larger than 100, the proposed specification test performs well and has power against a gamma alternative in the univariate case. Of course, simulations for multivariate null and alternative distributions are required before any general conclusions concerning the test's performance may be drawn. However, it does seem that for sample sizes less than 100, the negative variance-difference problem is significant and the proposed test may not be viable in such cases.

7. Conclusion.

In this paper, we have presented a specification test for the conditional normality of the attributes X and hence a test for the appropriateness of applying normal discriminant analysis under the maintained hypothesis of logistic conditional response probabilities. The specification test was performed for the analysis of corporate failures and it was concluded that the null hypothesis that DA and logit are equivalent may not be rejected.

In view of the distinction between causal and conjoint probability models which McFadden (1976) points out, the above test may be particularly useful when the estimated parameters have "structural" interpretations or if they are to be used to forecast impacts of policy changes. For example DA is often used to forecast future bankruptcies conditional on various macroeconomic scenarios. The result of the specification test for the data set used in this paper seems to support the use of DA. The standard hypothesis tests of the structural parameters may then be performed since the data do not reject normality in favor of some other member of the exponential family.

For purposes of classification, nonnormality may be less problematic as Amemiya and Powell's (1983) study suggests. Their calculations indicate that the use of normal DA when the X's are in fact binary does not appreciably increase the rate of misclassification. In this case, the proposed specification test may not have much power. However, the simulations in Section 6 indicate that the test does have power against a gamma alternative. Other alternative simulations should be performed on a case-by-case basis.

FOOTNOTES

[1]Discrimination analysis may, in principle, be performed for any distribution. Since the most common distribution employed is the multivariate normal, this will be the only case considered here.

[2]In this context, efficiency is in terms of ARE.

[3]In this case, efficiency is defined as minimum-variance in the class of CUAN estimators.

[4]See also Amemiya (1981) and Amemiya and Powell (1983).

[5]See McFadden (1976), Martin (1977), and Zavgren (1980).

[6]The Industrial Research File is composed of data for firms deleted from the Industrial Annual File because of

Code 01 - Acquisition or Merger

Code 02 - Bankruptcy

Code 03 - Liquidation

Code 04 - Other (no longer files with S.E.C., etc.).

[7]Most empirical studies of bankruptcy use data one year prior to default. The COMPUSTAT database did not, however, always have data for firms one year prior to failure. In such cases, the firm was included in our sample if data were available within three years of default and rejected otherwise.

[8]The GNP price deflator was taken from the 1984 Economic Report of the President (Table B-3) and normalized to 1.00 in 1972.

[9]Linear combinations of these variables are also often included in other empirical studies. For example, Ohlson (1980) includes the ratio of working capital to total assets in his estimation, but this is simply CATA - CLTA. We exclude them to preserve degrees of freedom and to avoid problems of multicollinearity.

[10]Note that the correspondence of this ratio to the ratio of accounting data used is almost certainly inexact. Therefore the results of our estimation should not be interpreted as conclusively supporting or rejecting any structural model of default.

[11]All software was implemented on a Digital VAX 11-780 in single-precision due to MLOGIT software constraints.

[12]Available from author upon request.

[13]Since the COMPUSTAT database only contains data for firms which are listed on the major exchanges, smaller privately held firms are excluded.

[14]See Palepu (1983) for an example.

[15]The complete set of simulation results are available from the author upon request.

Appendix 1 - The Asymptotic Distribution of $\hat{\beta}_{DA}$

Let $\mu \equiv \mu_1 - \mu_0$ and $\hat{\mu} \equiv \hat{\mu}_1 - \hat{\mu}_0$ where $\hat{\mu}_0$ and $\hat{\mu}_1$ are given by (9a). We seek the asymptotic distribution of $\sqrt{T}(\hat{\mu} - \mu)$ first. Recall that

$$\hat{\mu}_0 = \frac{1}{T_0} \sum_{i \varepsilon I_0} X_i \Rightarrow \sqrt{T_0}(\hat{\mu}_0 - \mu_0) \overset{A}{\sim} N(0, \Sigma) \tag{A1a}$$

$$\hat{\mu}_1 = \frac{1}{T_1} \sum_{i \varepsilon I_1} X_i \Rightarrow \sqrt{T_1}(\hat{\mu}_1 - \mu_1) \overset{A}{\sim} N(0, \Sigma) \tag{A1b}$$

and $\hat{\mu}_0$ and $\hat{\mu}_1$ are independent. By definition, we have

$$\hat{\mu} - \mu = (\hat{\mu}_1 - \mu_1) - (\hat{\mu}_0 - \mu_0) \tag{A2}$$

and thus

$$\sqrt{T}(\hat{\mu} - \mu) = \frac{\sqrt{T}}{\sqrt{T_1}} \sqrt{T_1}(\hat{\mu}_1 - \mu_1) - \frac{\sqrt{T}}{\sqrt{T_0}} \sqrt{T_0}(\hat{\mu}_0 - \mu_0) \tag{A3}$$

where $T = T_0 + T_1$. Since $\frac{\sqrt{T}}{\sqrt{T_i}}$ converges in probability to $\frac{1}{\sqrt{\pi_i}}$ $i = 0, 1,$ we conclude that

$$\sqrt{T}(\hat{\mu} - \mu) \overset{A}{\sim} N\left(0, \left(\frac{1}{\pi_0} + \frac{1}{\pi_1}\right)\Sigma\right) \tag{A4a}$$

or

$$\sqrt{T}(\hat{\mu} - \mu) \overset{A}{\sim} N(0, \delta\Sigma) \tag{A4b}$$

where $\delta = \frac{1}{\pi_0 \pi_1}$. Following Richard's (1975) notation, let σ denote the $\frac{1}{2}k(k + 1)$ vector of distinct elements of Σ, i.e., $\sigma = (\sigma_{11} \cdots \sigma_{k1}, \sigma_{22} \cdots \sigma_{k2}, \cdots, \sigma_{kk})'$ and let R be the $\frac{1}{2}k(k + 1) \times k^2$ selection matrix such that $R'\sigma = \text{vec}(\Sigma)$. Define Q' as the Moore-Penrose inverse of R, so that RQ' = I. By Richard (1975) we have:

$$\sqrt{T} \left(\begin{vmatrix} \hat{\mu} \\ \hat{\sigma} \end{vmatrix} - \begin{vmatrix} \mu \\ \sigma \end{vmatrix} \right) \overset{A}{\sim} N\left(0, \begin{vmatrix} \delta\Sigma & 0 \\ 0 & 2Q(\Sigma \otimes \Sigma)Q' \end{vmatrix}\right). \tag{A5}$$

Pre-multiplying $(\hat{\mu}' \quad \hat{\sigma}')'$ by $\mathrm{diag}(I_k, R')$, where I_k is the k^{th} order identity, then yields:

$$\sqrt{T} \left(\begin{vmatrix} \hat{\mu} \\ \mathrm{vec}\ (\hat{\Sigma}) \end{vmatrix} - \begin{vmatrix} \mu \\ \mathrm{vec}\ (\Sigma) \end{vmatrix} \right) \overset{A}{\sim} N(0,\ V_\gamma) \tag{A6a}$$

where

$$V_\gamma = \begin{vmatrix} \delta\Sigma & 0 \\ 0 & 2R'Q(\Sigma \otimes \Sigma)Q'R \end{vmatrix} . \tag{A6b}$$

Let $\gamma \equiv (\mu',\ \mathrm{vec}(\Sigma)')'$ and define $\hat{\gamma}$ similarly. Define the function f as

$$f(\gamma) \equiv \Sigma^{-1}\mu = \beta \tag{A7}$$

and let $J_f \equiv \frac{\partial f}{\partial \gamma}$ denote the Jacobian matrix of f. Applying the delta-method to f then yields:

$$\sqrt{T}\ (\hat{\beta}_{DA} - \beta) \overset{A}{\sim} N(0,\ J_f V_\gamma J_f') . \tag{A8}$$

To evaluate J_f explicitly, observe that

$$J_f = \left[\frac{\partial f}{\partial \mu} \quad \frac{\partial f}{\partial \mathrm{vec}(\Sigma)}\right] . \tag{A9}$$

But $\frac{\partial f}{\partial \mu} = \Sigma^{-1}$ and

$$\frac{\partial f}{\partial \mathrm{vec}(\Sigma)} = \frac{\partial \Sigma^{-1}\mu}{\partial \mathrm{vec}(\Sigma)} = \frac{\partial \mathrm{vec}(\Sigma^{-1}\mu)}{\partial \mathrm{vec}(\Sigma)} \tag{A10a}$$

$$= \frac{\partial (\mu' \otimes I_k)\mathrm{vec}(\Sigma^{-1})}{\partial \mathrm{vec}(\Sigma)} \tag{A10b}$$

$$= (\mu' \otimes I_k)\frac{\partial \mathrm{vec}(\Sigma^{-1})}{\partial \mathrm{vec}(\Sigma)} \tag{A10c}$$

$$\frac{\partial f}{\partial \mathrm{vec}(\Sigma)} = -(\mu' \otimes I_k)(\Sigma^{-1} \otimes \Sigma^{-1}) . \tag{A10d}$$

Substituting (A10) into (A9) results in the relation

$$J_f = \left[\Sigma^{-1} \quad -(\mu' \otimes I)(\Sigma^{-1} \otimes \Sigma^{-1})\right] . \tag{A11}$$

Using (A11) and simplifying yields the desired result

$$\sqrt{T}\ (\hat{\beta}_{DA} - \beta) \overset{A}{\sim} N(0,\ V_{DA}) \tag{A12}$$

$$V_{DA} = J_f V_\gamma J_f' = \delta\Sigma^{-1} + 2(\mu' \otimes I_k)(\Sigma^{-1} \otimes \Sigma^{-1})R'Q(\Sigma \otimes \Sigma)Q'R(\Sigma^{-1} \otimes \Sigma^{-1})(\mu \otimes I_k) .$$

Appendix 2 - Summary Statistics of Data Set

	MEANS MATRIX			STANDARD DEVIATION MATRIX		
VARIABLE	BANKRUPT	SOLVENT	COMBINED	BANKRUPT	SOLVENT	COMBINED
CONSTANT	1.0000	1.0000	1.0000	0.0000	0.0000	0.0000
SIZE	2.4569	2.6604	2.5586	1.1995	1.3219	1.2747
CLTA	0.5706	0.3394	0.4550	0.2114	0.2070	0.2406
OLTA	0.2974	0.2117	0.2545	0.1646	0.2581	0.2221
CATA	0.6533	0.6613	0.6573	0.2054	0.2091	0.2087
NITA	-0.1548	0.0313	-0.0618	0.2108	0.0883	0.1877
BANK	0.1086	0.6734	0.3910	0.1284	0.7954	0.6401

CORRELATION MATRIX (COMBINED SAMPLE)

	SIZE	CLTA	OLTA	CATA	NITA	BANK
SIZE	1.0000					
CLTA	-0.2855	1.0000				
OLTA	0.1556	-0.0975	1.0000			
CATA	-0.4561	0.1007	-0.3601	1.0000		
NITA	0.2155	-0.6498	-0.0618	0.0524	1.0000	
BANK	-0.0525	-0.5780	-0.1099	-0.0401	0.3370	1.0000

Summary Statistics of Data Set (continued)

CORRELATION MATRIX (BANKRUPT SAMPLE)

	SIZE	CLTA	OLTA	CATA	NITA	BANK
SIZE	1.0000					
CLTA	-0.3515	1.0000				
OLTA	0.2682	-0.4729	1.0000			
CATA	-0.3397	0.0081	-0.4535	1.0000		
NITA	0.3041	-0.5370	0.0486	0.0233	1.0000	
BANK	-0.1073	-0.3976	0.3679	-0.0556	0.2139	1.0000

CORRELATION MATRIX (SOLVENT SAMPLE)

	SIZE	CLTA	OLTA	CATA	NITA	BANK
SIZE	1.0000					
CLTA	-0.2200	1.0000				
OLTA	0.1253	-0.0700	1.0000			
CATA	0.5664	0.2438	-0.3201	1.0000		
NITA	0.0511	-0.6827	0.0586	0.1246	1.0000	
BANK	-0.1193	-0.6051	-0.0701	-0.0684	0.3072	1.0000

REFERENCES

Amemiya, T., "Qualitative Response Models: A Survey," Journal of Economic Literature, 19 (1981), 1483-1536.

Amemiya, T. and J. Powell, 1983, "A Comparison of the Logit Model and Normal Discriminant Analysis when the Independent Variables are Binary," in: Samuel Karlin, Takeshi Amemiya, and Leo A. Goodman, Studies in Econometrics, Time Series, and Multivariate Statistics (Academic Press, New York).

Directory of Obsolete Securities, Financial Information Inc., Jersey City, NJ.

Efron, B., "The Efficiency of Logistic Regression Compared to Normal Discriminant Analysis," Journal of the American Statistical Association, 70 (1975), pp. 892-898.

Hausman, J., "Specification Tests in Econometrics," Econometrica, 46 (1978), 1251-1271.

Lo, A., Essays in Financial and Quantitative Economics, unpublished Ph.D. dissertation, Harvard University, 1984.

McFadden, D., "A Comment on Discriminant Analysis 'Versus' Logit Analysis," Annals of Economic and Social Measurement, 5/4 (1976), pp. 511-523.

Martin, D., "Early Warning of Bank Failure: A Logit Regression Approach," Journal of Banking and Finance, 1 (1977), pp. 249-276.

Newey, W. K., Specification Testing and Estimation Using a Generalized Method of Moments, unpublished Ph.D. dissertation, M.I.T., 1983.

Ohlson, J. A., "Financial Ratios and the Probabilistic Prediction of Bankruptcy," Journal of Accounting Research, 18 (1980), pp. 109-131.

Palepu, K., "The Determinants of Acquisition Likelihood," Harvard Business School Working Paper # 9-783-056, February 1983.

Richard, J. F., "A Note on the Information Matrix of the Multivariate Normal Distribution," Journal of Econometrics, 3 (1975), 57-60.

Scott, J. H., Jr., "The Probability of Bankruptcy: A Comparison of Empirical Predictions and Theoretical Models," Journal of Banking and Finance, 5 (1981), pp. 317-344.

Wall Street Journal Index, Dow Jones and Company, New York, NY.

Zavgren, Christine, A Probabilistic Model of Financial Distress, unpublished Ph. D. dissertation, University of Nebraska, 1980.