International Journal of Finance, Accounting and Economics Studies
Vol.1 / No.1 / Winter 2011

Applying Variable Deletion Strategies in Bankruptcy Studies to Capture Common Information and Increase Their Reality

Receipt date: 89/7/25 Acceptance date: 89/9/10

Hashem Nikoomaram
Professor, Department of Accounting, Science and Research Branch,
Islamic Azad University, Tehran, Iran
nikoomaram@srbiau.ac.ir

Zahra Pourzamani
Assistant Professor, Department of Accounting, Central Tehran Branch,
Islamic Azad University, Tehran, Iran

Abstract

In financial distress studies selection of variable is commonly based on the success of variables in variable sets employed in earlier bankruptcy studies, suggestions in the literature or an accompanying data reduction in a large set of variables. If seemingly different variable sets exhibit a strong relationship then heterogeneous variable sets capture common information. Canonical correlation analysis appropriately examines the relationship between two sets of measured variables.

The main purpose of the present study was to illustrate the value of variable deletion strategies in canonical correlation analysis for more parsimonious to capture common information. In research contents, the law of parsimony states that the fewer variables used to explain a situation, the more probable that the explanation will be closer to reality. Therefore, as variable sets become more parsimonious there are greater probabilities that the results of the analysis will be replicable. To determine the common information between variable sets in financial distress studies, the study selected two specific bankruptcy models: Altman, the most famous model, and Deakin, the biggest model. The results indicated that as the number of variables increase, the probable effect of these sources of error variation on the canonical correlation increases. Therefore, the goal of a variable deletion strategy is to estimate as much variance with the smallest variable set possible. In this study the goal was achieved by removing the three variables in variable sets employed in selected bankruptcy studies.

Keywords: Bankruptcy Studies, Variable Deletion Strategies Canonical Correlation Analysis.

Introduction

Numerous corporate failure prediction models have so far been developed, based on various modeling techniques and financial ratios. In fact interest in the ability to predict financial distress has produced a considerable body of research in accounting and finance over the last years. These methods all have their particular strengths and weaknesses in discrimination of failing and non-failing firms. But the resulting consequence of ad hoc variable selection in financial distress studies is that consensus does not exist on a definitive set of variables that distinguish between distressed and non-distressed firms.

To determine the common information between variable sets, The study selects two bankruptcy studies and employs canonical correlation analysis to examine the relationships that exist between two variable sets and then to illustrate the value of applying the law of parsimony to canonical correlation analysis (CCA) solutions.

2. Litrature review

As suggested by Ball and Foster (1982), ad hoc variable selection limits the financial distress models that are consistent predictors of financial distress. Thorndike (1978) stated that “as the number of variables increase, the probable effect of these sources of error variation on the canonical correlation increases”

Thompson (1991) showed that CCA subsumes all other parametric methods including t-tests, ANOVA, regression, MANOVA and discriminate analysis. Also according to Henson (2000) and Knapp (1978) purport was that this connection helped to reinforce the concept that all parametric techniques are subsumed under CCA as the classical form of the general linear model.

As Pedhazur (1997) has demonstrated, canonical correlation matrix computation can become “prohibitive” and “complex”. Besides Knapp’s demonstration modern statistical packages almost eliminate the need to create these matrixes.

Cantrell (1999) indicated as variable sets become more parsimonious, there are greater probabilities that the results of the analysis will be replicable.

Rim (1975) suggested that more parsimonious models are not only more stable and replicable but also more able to be generalized. According to Thompson (1982), reducing the number of variables lessen Type II error probability since degrees of freedom model are also lessened. In an analysis with three criterion variables and six predictor variables, the 18 degrees of freedom would be reduced by nine if three predictor variables were deleted from the final model.

Thompson (1984) also suggested that in multiple regressions dropping of variables in CCA would be synonymous with “backward elimination” stepwise procedures

According to Humphries-Wadsworth (1998), canonical correlation analysis is a “rich tool for examining the multiple dimensions of the synthetic variable relationships”

Capraro and Capraro (2001)**,** stated that “the goal of deletion strategies in canonical correlation analysis is a more parsimonious solution. Therefore, choosing the smaller variable set when the same amount of variance can be accounted for is achieved”. They showed that “bigger is not better”, at least in reference to the number of variables, when using canonical correlation analysis.

Leclere (2006) also suggested that if one variable set is redundant to another variable set, it is because the redundant variable set, is much smaller than the predictor variable set.

3. Statistical method

Canonical correlation analysis determines the extent of the relationship between two variable sets with redundancy coefficients. Redundancy coefficients indicate the degree of overlap between two sets of variables; more specifically, they are an index of the average proportion of variance in one variable set that is predictable from or shared with the canonical variates in the other set (Stewart and Love, 1968; Lambert et al.1988). Employing one set of variables to predict a second set of variables implies the second set is ‘‘redundant’’ upon knowing the first set. The examination of redundancy coefficients is either individually or pooled across canonical functions.

Canonical correlation analysis examines the independent statistical relationships that exist between two variable sets by analyzing the sets simultaneously and identifying and quantifying the elements of one variable set most highly related to the elements of the other variable set (Kotz and Johnson, 1982; Thompson, 1984). This statistical technique can treat the two variable sets symmetrically or it can treat one variable set as the predictor set (independent or exploratory measures) and the other set as the criterion set (dependent measures). Furthermore multiple regression analysis could do the job if there were only one dependent variable; however, canonical analysis goes a step farther by allowing multiple dependent variables.

4. Sample method

The research limited to manufacture firms listed in the Tehran Stock Exchange from 1998 to 2007. The research relies on a sample of 30 failed and 30 non-failed manufacturing firms. A sample of 30 manufacturing companies which had become bankrupt between 1998 and 2007 were identified from The Article No.141 of Commercial Law of Iran and matched to 30 non-failed companies on the basis simple Q-tubin.

5. Research hypothesis

To determine the common information between two variable sets the research hypothesis "Much of the information contained in the Altman variable set is presented in the Deakin variable set" was developed.

6. An overview of the selected bankruptcy studies

Altman (1968) employed discriminate analysis to classify firms as failed or no failed. The five ratios employed in the model were earnings before interest and taxes/total assets, market value of equity/book value of debt, retained earnings/total assets, sales/total assets, and working capital/total assets. In the year prior to bankruptcy, the best model was 95 per cent effective in classifying the firms.

Deakin (1972) utilized the ratios of Beaver (1966; 1968a) to build a discriminate model for predicting business failure. The ratios employed were cash/current debts, cash/sales, cash/total assets, cash flow/total debts, current assets/current debts, current assets/sales, current assets/total assets, net income/total assets, quick assets/current debts, quick assets/sales, quick assets/total assets, total debts/total assets, working capital/sales, and working capital/total assets. The model was 97 per cent effective in classifying the firms in the year prior to failure.

7. Results

Looking at table 1 in function 3 the pooled redundancy coefficient of the Deakin set with respect to the Altman set is 0.29.The variable set employed by Deakin is not similar to the variable set employed by Altman. The Altman variable set is not a good predictor of the Deakin variable set. On the other hand, the pooled redundancy coefficient of the Altman variable set with respect to the Deakin variable set is still moderately high at 0.52. A large part of the Altman variable set is redundant to the Deakin variable set after dropping common variables.

Table 1: canonical correlation analysis to examine the relationships that exist between two variable sets

Variable statistic	FUNCTION1	FUNCTION2	FUNCTION3
	rs2	rs2	rs2
Altman model			
adequacy	28.79%	24.41%	22.80%
RD	26.20%	16.55%	9.53%
∑Rd	26.20%	42.75%	52.28%
Rc2	91.00%	67.80%	41.80%
RD	15.97%	5.29%	7.32%
∑Rd	15.97%	21.26%	28.57%
adequacy	17.55%	7.80%	17.51%
Deakin model			

r_s^2 - squared canonical structure coefficient - how much variance a variable linearly shares with a canonical variant (Thompson, 1980).
Rc^2 - squared canonical coefficient– how much each function is contributing to the overall canonical solution (Thompson, 1991).

The results of canonical analysis in table 1 indicate that the pooled redundancy coefficient of the Altman set with respect to the Deakin variants is 0.52, the Deakin canonical variants account for 52 per cent of the variability among the Altman variables. One conclusion is that a large part of the variable set used by Altman is redundant to the variable set used by Deakin and much of the information in the Altman variable set is common to the Deakin variable set. To determine the common information between these two variable sets, one of the different deletion methods are delineated in the paper.

To illustrate the deletion process, the results of full canonical analysis are compiled in Table 2. The “Func” (canonical function coefficient), the “rs” (canonical structure coefficient) along with the Rc2 (squared canonical correlation coefficient) for each function was obtained directly from the SPSS printout. The rs2 (squared canonical structure coefficient) was calculated by squaring the canonical structure coefficients for each variable and converting them into percentage format. The h2 (communality coefficient) for each variable was obtained by summing all the rs2s. The adequacy coefficient, “how will a canonical variant represents the variance of the original variables in a domain” (Thompson. 1980, p.10), was an average of all the squared structure coefficients for the variables in one set with respect to one function. The adequacy coefficient for the criterion variable set was calculated by adding all the structure coefficients in the criterion set and dividing by the number of variables in the set and converting it into percentage format. The adequacy coefficient for the predictor set was determined by the same method. The redundancy coefficient, the redundancy of C (criterion variable set) given P (predictor variable set), was calculated by multiplying the adequacy coefficient by the Rc2 for each function (Roberts, 1999).

In Table 2, in initial solution the predictor variables with the lowest h2s were cash/sales (7.49%). This variable was quite a bit lower than the other twelve-predictor variables that ranged from 17.68% to 93.55%. Through this variable deletion strategy, the variable with the lowest h2, cash/sales, was dropped first.

In iteration #2 the remaining canonical solution still contained current assets/total assets with an h2 of 17.55%. That variable was considerably lower than the other variables, therefore current assets/total assets was dropped

In iteration #3 the remaining canonical solution still contained cash/current debts with an h2 of 24.96%. That variable was lower than the other variables, therefore cash/current debts was dropped

Table 2: Canonical Solution After Based on Canonical Communality Coefficients Deletion Strategy

Variable statistic	Initial Solution		Iteration #2		Iteration #3		Iteration #4	
	FUNCTION	h2	FUNCTION	h2	FUNCTION	h2	FUNCTION	h2
	rs2		rs2		rs2		rs2	
Sales / total assets	2.16%	100.0%	43.56%	99.21%	46.65%	97.75%	46.10%	99.08%
Market value of equity / book value of debt	3.42%	4.50%	4.71%	7.00%	6.54%	7.93%	6.71%	10.23%
Earnings before interest and taxes / total assets	0.41%	99.84%	0.00%	99.82%	0.00%	99.90%	0.00%	99.59%
Retained earnings / total assets	85.19%	99.66%	42.51%	98.96%	38.19%	98.98%	38.56%	97.50%
Working capital / total assets	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
adequacy	22.80%		22.70%		22.82%		22.84%	
RD	9.53%		6.88%		6.82%		6.83%	
∑Rd	52.28%		45.98%		45.01%		44.42%	
Rc2	41.80%		30.30%		29.90%		29.90%	
RD	7.32%		3.58%		3.40%		3.37%	
∑Rd	28.57%		28.31%		27.53%		26.65%	
adequacy	17.51%		11.82%		11.38%		11.29%	
Quick assets/total assets	4.49%	93.55%	0.66%	93.60%	.72%	93.82%	0.74%	94.00%
Quick assets/current debts	29.81%	31.70%	23.04%	36.93%	22.85%	37.61%	23.14%	40.27%
Net income/total assets	32.95%	73.59%	6.92%	73.59%	5.06%	75.53%	5.24%	78.90%
Working capital/total assets	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Total debt / total assets	31.81%	36.43%	40.83%	48.71%	40.58%	49.39%	40.83%	50.44%
Cash flow/total debts	34.81%	70.78%	12.25%	71.23%	10.18%	72.51%	10.43%	75.71%
Working capital/ sales	29.92%	33.01%	15.44%	34.12%	13.62%	34.19%	13.76%	35.02%
Quick assets/sales	5.15%	78.90%	0.17%	80.68%	0.19%	80.61%	0.18%	81.55%
Current assets/total assets	11.97%	17.68%	1.37%	17.55%	0.00%	0.00%	0.00%	0.00%

Current assets/sales	1.35%	27.42%	20.70%	39.84%	21.53%	39.56%	21.16%	41.87%
Current assets/current debts	32.04%	32.64%	23.81%	36.29%	22.28%	36.24%	22.47%	36.90%
Cash/total assets	2.82%	33.50%	7.18%	43.83%	9.00%	46.10%	8.76%	48.53%
Cash/sales	6.71%	7.49%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Cash/current debts	3.76%	20.40%	1.35%	24.14%	1.90%	24.96%	0.00%	0.00%

h^2 – canonical communality coefficients - sum of all rs^2; how much of the variance in a given observed variable is reproduced by the complete canonical solution (Thompson, 1991).

8. Conclusion

Variable selection in financial distress studies is commonly based on ad hoc variable selection. Therefore, the main purpose of the present study was to employ canonical correlation analysis in order to illustrate the value of applying the law of parsimony to canonical correlation analysis solutions.

As variable sets become more parsimonious there are greater probabilities that the results of the analysis will be replicable. Therefore the research hypothesis "Much of the information contained in the Altman variable set is present in the Deakin variable set" was developed and variable deletion strategies are delineated in the paper to determine the common information between two variable sets, Altman and Deakin. The goal of parsimony was achieved by removing the three variables, Cash/Sales, Current Assets/Total Assets and Cash/Current Debts. In each function a very small change was noted in the communality coefficients. In other words, the Deakin variable set would be a good predictor of the Altman variable set

References

1) Altman, E. 1968, 'Financial ratios, discriminate analysis, and the prediction of corporate bankruptcy', *Journal of Finance*, vol. 23, no. 4, pp. 587-690.
2) Ball, R. & Foster, G. 1982, 'Corporate financial reporting: a methodological review of empirical research', *Journal of Accounting Research*, vol. 20, supplement, pp. 161-234.
3) Beaver, W. (1966), ''Financial ratios as predictors of failure'', *Journal of Accounting Research*, Vol. 4, Supplement, pp. 71-111.

4) Beaver, W. (1968a), ''Alternative accounting measures as predictors of failure'', *The Accounting Review*, Vol. 43 No. 1, pp. 113-22.
5) Cantrell, C. 1999, 'Variable deletion', In B. Thompson (Ed.), *Advances in social science methodology*, vol. 5, pp. 321-333, Stamford, CT: J.A.I. Press.
6) Capraro, M. & Capraro, R. (2001), ''Bigger is Not Better: Seeking Parsimony in Canonical Correlation Analysis via Variable Deletion Strategies '', *Multiple Linear Regression Viewpoints,* Vol. 27(2) , pp. 24-33.
7) Deakin, E.B. 1972, 'A discriminate analysis of predictors of business failure', *Journal of Accounting Research*, vol. 10, no. 1, pp. 167-179.
8) Foster, G. 1986, *Financial Statement Analysis*, Prentice-Hall, New York.
9) Henson, R. K. 2000, 'Demystifying parametric analyses: illustrating canonical correlation analysis as the multivariate general linear model', *Multiple Linear Regression Viewpoints*, vol. 26, no. 1, pp. 11-19.
10) Humphries-Wadsworth, T. 1998, *Features of published analyses of canonical results,* Paper presented at annual meeting of the American Educational Research Association, San Diego, CA, (ERIC Document Reproduction Service No. ED 418 125).
11) Knapp, T.R. 1978, 'Canonical correlation analysis: A general parametric significance-testing system', *Psychological Bulletin*, vol. 85, pp. 410-416.
12) Kotz, S. and Johnson, N.L, 1982, Encyclopedia of Statistical Sciences, John Wiley and Sons, New York, NY.
13) Lambert, Z.V., Wildt, A.R. & Durand, R.M. 1988, 'Redundancy analysis: an alternative to canonical correlation and multivariate multiple regression in exploring interest associations', *Psychological Bulletin*, vol. 104, pp. 282-289.
14) Leclere, M. (2006), ''Bankruptcy studies and ad hoc variable selection: a canonical correlation analysis'', `.
15) Pedhazur, E. J. 1997, 'Canonical and discriminate analysis, and multivariate analysis of variance', *Multiple regression in behavior research*, 3rd ed., pp. 924-979, Fort Worth, TX: Harcourt Brace.
16) Rim, E. 1975, *A stepwise canonical procedure and the shrinkage of canonical correlations*, Paper presented at annual meeting of the

American Educational Research Association, Washington, DC, (ERIC Document Reproduction Service No. ED 109 237).

17) Roberts, J.K. 1999, 'Canonical redundancy (Rd) coefficients, they should (almost never) be computed and interpreted', In B. Thompson (Ed.), *Advances in social science methodology*, vol. 5, pp. 331-333, Stamford, CT: J.A.I. Press.
18) Stewart, D. & Love, W. 1968, 'A general canonical correlation index', *Psychological Bulletin*, vol. 70, no. 3, pp. 160-163.
19) Thompson, B. 1980, April, *Canonical correlation: Recent extensions for modeling educational processes.* Paper presented at annual meeting of the American Educational Research Association, Boston, MA. (ERIC Document Reproduction Service No. ED 199 269)
20) Thompson, B. 1982, February, *Stepwise canonical correlation analysis: A new research technique*. Paper presented at annual meeting of the Southwest Educational Research Association, Austin, TX. (ERIC Document Reproduction Service No. ED 222 546)
21) Thompson, B. 1984, *Canonical correlation analysis*, Sage Publications, Thousand Oaks, CA.
22) Thompson, B. 1991, Methods, plainly speaking: A primer on the logic and use of canonical correlation analysis. *Measurement and Evaluation in Counseling and Development, 24* (2), 80-93.
23) Thorndike, R. 1978, *Correlation procedures for research.* New York: Gardner