A COMPARISON: TRADITIONAL CANONICAL CORRELATION ANALYSIS VERSUS THE MIMIC/SEM APPROACH USING A MULTI-GROUP STUDY OF CORPORATE SIZE VERSUS PROFITABILITY

David W. Johnk

The University of Texas – Pan American
College of Business Administration
1201 W University Dr., Edinburg, TX 78541
Telephone: 956-457-2453
Email: djohnk@gmail.com

Abstract

This study employs the use of two methods of canonical correlation analysis to investigate the relationship between two latent variables: corporation "size" and "profitability". The variables examined are divided into an independent set of variables representing the latent variable "size", and a dependent set of variables representing the latent variable called "profitability". The first method employed is the traditional canonical correlation analysis (CCA) approach. The second method used is the structural equation modeling (SEM) approach. The SEM approach uses the Multiple Indicators / Multiple Causes (MIMIC) model.

Three variables are used to measure company size; number of employees, total assets, and total income. Three variables are also used to indicate company profitability; return on equity, return on income, and earnings per share.

This study demonstrates that the MIMIC model can be quite useful and in some ways is superior to the traditional canonical approach. The traditional approach finds that the first canonical function is significant and the same is found with the MIMIC model. Furthermore, it was shown that multiple groups which are moderated by categorical variables, in this case industry sector, can be evaluated for differences using the MIMIC approach. The traditional model does not allow this. In the literature review no other study was found which demonstrates the use of the SEM/MIMIC model in the examination of the affects of categorical moderators.

Introduction

This paper is a presentation of the results of an exploratory empirical study of a sample of 316 United States corporations. These corporations are first examined using traditional canonical correlation analysis (CCA). The study is then repeated using the multiple indicators/multiple causes (MIMIC), which is a structural equation modeling (SEM) approach. Fan (1997) and Guarino (2004) have already compared traditional canonical correlation analysis with the MIMIC model. This study extends what they demonstrate by examining group differences using a *multi-group* MIMIC model.

The purpose of canonical correlation analysis is to determine if a significant linear relationship exists between two sets of variables, and if a relationship exists, the magnitude and the direction of that relationship. The variables in this study are divided into an independent set of variables which indicate "size", and a dependent set of variables measure "profitability". The "size" variables used were: number of employees (NumEmploy), total assets (TotAssets), and

total sales (TotInc). The "profitability" variables used were: return on equity (ROE), return on invested capital (ROIC), and earnings per share (EPS).

The size variables are all excellent indicators of firm size; collectively they are representative of the size of a company. In accounting terms, total assets are equal to liabilities plus owner's equity; thus total assets should tend to be larger for bigger companies. Larger companies should have more employees. Total income should similarly tend to be larger for large companies.

Most profitability measures are ratios and are generally defined as indicators of a firm's success at generating profits. These indicators are most useful when combined together with several others for getting a whole picture of how a company is doing. Analysts rarely look at one profitability ratio by itself when determining how successful a company will be at generating profits in the future. When compare they compare between companies to chose which one to invest in, they often combine several measures. One reason for this is that it is possible for firm management to manipulate one or even more than one of the ratios in order to make their company more attractive to stockholders, raising the value of the firm. For example ROE can be artificially raised by using 100% debt to purchase another company which earns less than its cost of capital but more than the cost of debt. (Copeland, Weston, & Shastri, p.474), but it is far more difficult to manipulate all of them, thus it makes sense to examine a set of variables for each, as done in this study.

The ratio ROIC is defined as the net operating profit minus taxes, the result then divided by the total capital. ROIC is a measure of how well a company generates cash (profits) in terms of the money (capital) it has invested in doing business. ROE is defined as net income divided by stockholders equity. The ROE ratio is commonly used by stockholders to determine how well each dollar invested in a firms stock is turned into profits. EPS is defined as net income minus dividends on preferred stock, the result divided by the average outstanding shares. EPS is also considered to be a good indicator of profitability; it measures the portion of the companies profit allocated to each share of outstanding stock.

Methodology

Sample Selection

The sample data was drawn from the Datastream[i] database. In Datastream there are forty United States equity market sectors, which are divided by industry. Examples of some of the sectors are Automobiles and Parts, Banks, Beverages, Chemicals, etcetera. The intent of the study is to select two industries. The selection criteria is that the two sectors must be relatively large (above one-hundred cases) and also be similar in size. During the selection process five sectors were too small and one was too large. The semiconductor and electrical equipment sectors were selected for use. The semiconductor sector has 200 initial cases; 37 cases were removed due to missing data yielding a final sample size of 163. The electrical equipment sector has 275 initial cases; 122 were removed due to missing data resulting final sample of 153 cases. With six total variables being tested this represents about 25 samples per variable; which is an acceptable number and exceeds the rule of thumb of 10 for a minimum (or about 15 to 20 per sample in the case where a multi-group model is being tested). After the sample is drawn, all variables are standardized in order to better interpret the results. Standardization is necessary

because some variables have extremely large values (such as total revenue) and others have comparatively small values (such as the ratio's ROIC, ROE, and EPS).

Previously the same analysis was done with two-digit NAICS codes 31 and 32 (manufacturing codes). The resulting companies were too diverse and insignificant results were obtained during analysis; they are subsequently not shown here. Narrowing the industry by using 3 or 4-digit North American Industry Classification System (NAICS) codes resulted in sample sizes which are too small. Using Standard Industrial Codes (SIC) is considered to have similar the same problem as NAICS codes and subsequently were not used.

Subject

The subject of this study is US companies randomly drawn from two different Datastream equity market sectors, specifically the sectors were "electronic equipment" and "semiconductors". The sample is taken from firms in similar industries (sectors) in order to control for the fact that the linkages between the variables; independent (size) and the dependent (profitability ratios) may be affected by differences between industries. US companies are used because standard data was more readily available for this study.

Procedure

The traditional canonical correlation analysis is done using the program SPSS® [ii]. Canonical correlation is not directly accessible from the menu on SPSS®, but can it can be accessed in several different ways using the syntax commands. The syntax which is used for this study:

```
MANOVA set1 WITH set2
/DISCRIM ALL ALPHA(1)
/PRINT SIG(EIG DIM).
```

(Set 1 is replaced with the dependent variable list and set2 with the independent variable list, with a space between each variable.)

The SEM approach is done using a MIMIC model; all calculations are done with the program AMOS® [iii]. Figure 1 graphically illustrates one group MIMIC model that is used to estimate the first canonical function. After the first function is estimated, its regression weights are used in order to estimate the second function. The change in the χ^2 values and degrees of freedom is calculated in order to determine significance of fit between the two models. If the change is not significant the second canonical function is deemed to be insignificant. Only if the change is significant (the second function is useful) will the same process be repeated in order to analyze the third function.

A multi-group analysis is then conducted in order to analyze whether there are differences between the combined group model and the multi-group model, and also whether there are differences between the groups (see figures 2 and 3)

Canonical Correlation Analysis
MIMIC Model - One
Size vs. Profitability
Both Industries (Semiconductors & Electrical Equip.)
Standardized data

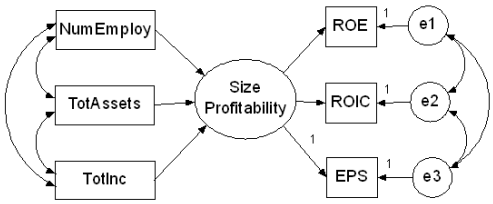

Figure 1

Canonical Correlation Analysis
Multi-group MIMIC Model
Size vs. Profitability
Electrical Equipment Group
Standardized data

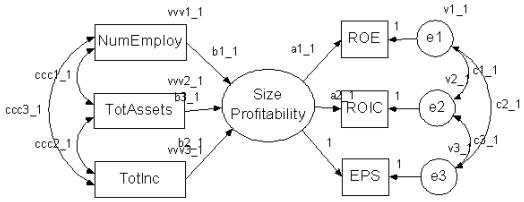

Figure 2

Canonical Correlation Analysis
Multi-group MIMIC Model
Size vs. Profitability
Semiconductors
Standardized data

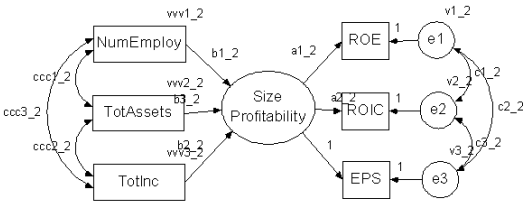

Figure 3

Results

CCA Analysis

Table 1 shows the overall fit of the model using the traditional method. The number of canonical functions which can be derived is equal to the minimum number of variables in either the dependent or independent set of variables. Because both sets of variables have three indicators, the resulting number of canonical functions is three. It can be seen that none of the functions produce significant results. The first function has a canonical correlation of 0.283 and a resulting R-squared value of 0.08. The first function also has an F statistic of 3.162 and its corresponding p-value is 0.001. According to Hair, Black, Babin, Anderson, & Tatham (2006) the level of significance is typically 5%. The other two functions are not significant at a 5%, and also have very low R-square values, and thus should not be interpreted.

Measures of Overall Model Fit for Canonical Correlation Analysis (traditional approach)

Canonical Function	Canonical Correlation	Canonical R-squared	F Statistic	Probability
1	0.283	0.08	3.16225	0.001
2	0.081	0.007	0.51174	0.727
3	0.002	0	0.00184	0.966

Table 1

Multivariate Tests of Significance (traditional approach)

Statistic	Value	Approximate F Statistic	Probability
Pillai's Trace	0.0867	3.095	0.001
Hotelling's Trace	0.0937	3.214	0.001
Wilks' Lambda	0.9138	3.162	0.001
Roys' gcr	0.0801		

Table 2

Multivariate tests of significance were collectively performed for all three functions, with the information presented in Table 2. Wilk's lambda, Pillia's trace, Hotelling's trace and Roy's gcr all are statistically significant at a 5% level (p-value of 0.001). The canonical correlation of 0.283, is quite low, but is significant so interpretation of it is questionable (it may not be practically significant).

MIMIC analysis

Evaluating the first canonical function using the SEM MIMIC model, the χ^2 value for goodness-of-fit (GOF) is 2.072 with degrees of freedom (df) of 4 resulting in a p-value of 0.723 (see Table 3). The CFI index is 1.000 and the RMSEA is 0.000 (0.000 to 0.060 at 90% confidence). The GOF index shows that there is not statistical evidence of a significant difference between the observed sample covariance matrix and the SEM estimated covariance matrix. It has been previously observed that using the GOF index as the sole indicator of fit is problematic due to sensitivity issues with sample size and the number of indicators (Hair et. al 2006). The other indexes (CFI and RMSEA) are also acceptable; therefore the model has acceptable fit.

Canonical Correlation Analysis
MIMIC Model - One
Size vs. Profitability
Both Industries (Semiconductors & Electrical Equip.)
Standardized data

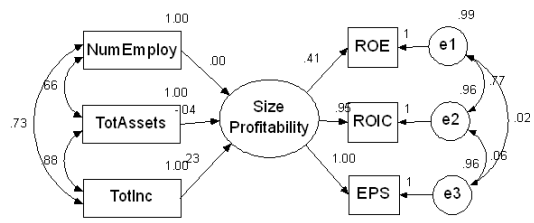

Figure 4

Canonical Correlation Analysis
MIMIC Model - Two
Size vs. Profitability
Both Industries (Semiconductors & Electrical Equip.)
Standardized data

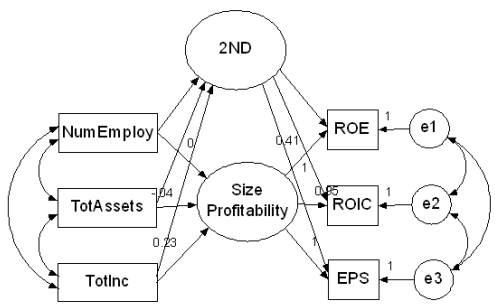

Figure 5

Evaluating the second canonical function, the $\Delta\chi^2$ values between the first and second canonical functions is 2.048 with a Δdf of 5 and a resulting p-value of 0.842 (see table 3). The addition of the second canonical function does not provide significantly better model fit, therefore only the first model is significant. As expected, this result is confirms the results from the traditional method, where the only the first canonical functions is found significant. It is not necessary to evaluate the third function using SEM because if the second is found to be not significant, the third will not be significant. One advantage with the MIMIC approach is that it additionally provides standard errors and p-values for interpretation. Referring to table 4 it can be observed that the only regression weight that is significant is the Total Income variable with positive value of 0.235. Figures 4 and 5 illustrate the graphical AMOS models used to evaluate the first and second canonical functions.

Overall results (SEM/MIMIC)

	SEM diagram	Model	Chi-squared value	df	p-value	CFI	RMSEA
1	MIMIC (one function)	default	2.072	4	0.723	1.000	0.000
2	MIMIC (two functions)	default	0.024	4	1.000	1.000	0.000
difference 1 - 2			2.048	5	0.842		
3	Multi-group	unconstrained	2.252	8	0.972	1.000	0.000
difference 3 - 1			0.180	4	0.996		
4	Multi-group	model 2	4.363	13	0.987	1.000	0.000
difference 4 - 3			2.111	5	0.834		

Table 3

Regression Weights - Independent Variate

	Regression Weights	Standard Error	p-value
NumEmploy --> Profitability	0	0.056	0.995
TotInc --> Profitability	0.235	0.112	0.037
TotAssets --> Profitability	-0.04	0.087	0.646

Table 4

Evaluating the multi-group model the χ^2 value for goodness-of-fit (GOF) is 2.252 with degrees of freedom (df) of 8 resulting in a p-value of 0.972. The CFI index is 1.000 and the RMSEA is 0.000 (0.000 to 0.000 at 90% confidence). Improvement of model fit between the combined model and the multi-group model the $\Delta\chi^2$ is 0.180 with a Δdf of 4 for a resulting p-value of 0.996, meaning there is not significant improvement in fit between the multi-group model and the combined one canonical function model.

Conclusion

This study shows that both the CCA and MIMIC approaches produce similar results, for both methods the first canonical function was found significant and the second and third functions are found to be insignificant.

The MIMIC approach can be quite useful, and depending on the application, may be superior to the traditional CCA approach. Multiple groups can easily be evaluated using the MIMIC approach; the traditional CCA model does not allow this. Another advantage of the MIMIC model is the availability of standard errors and p-values for the regression coefficients, which are not simply calculated using the CCA approach. A disadvantage of the MIMIC model is that the model may not converge; this problem does not happen with the traditional CCA approach.

In the analysis of size versus profitability, it can be seen that the only independent variable which is significant is the total income. It had a positive relationship with profitability. Further research could include the study and of how the MIMIC model would be useful in the understanding of the relationships between company size and profitability using other industry sectors, and using additional size and profitability variables. More research should to be done in order to understand how useful the MIMIC approach is in analysis of size versus profitability, and whether it might be a useful tool for additional insight to that area of financial literature.

---

[i] Thomson Financial DataStream Advance, version 3.5 for Windows®, run on Microsoft Windows XP Professional®, Second Edition.

[ii] SPSS®, version 15 for Windows®, run on Microsoft Windows XP Professional®, Second Edition.

[iii] AMOS®, version 7 for Windows®, run on Microsoft Windows XP Professional®, Second Edition.

References

Copeland, T.E., Weston J.F, & Shastri, K. (2005). *Financial Theory and Corporate Policy* (4th ed.). Boston: Pearson Addison Wesley.

Fan, Xitao (1997). Canonical correlation analysis and structural equation modeling: what do they have in common? *Structural Equation Modeling.* 4(1), 65-79

Guarino, A. J. (2004) A Comparison of First and Second Generation Multivariate Analysis: Canonical Correlation Analysis and Structural Equation Modeling. *Florida Journal of Education Research*. 2004, 42, 22 – 40

Hair, J.F., Black, W.C., Babin, B.J. Anderson, R.E., & Tatham, R.L. (2006). *Canonical Correlation Analysis – supplement to: Multivariate Data Analysis* (6th ed.) [Electronic Version] Retrieved March 2007 from: http://www.mvstats.com/downloads/supplements/canonical_correlation.pdf.