The Canonical Correlation Analysis - An Application to Bank Performance and Consumers Satisfaction

M. Roungu Ahmmad[*], Md. Matair Rahman[**] , Md. Sanwar Hossain[*], Azizur Rahman[*]

[*]Department of Statistics, Jagannath University,Dhaka, Bangladesh
[***]Department of Statistics, Islamic University, Kustia, Bangladesh.

Abstract

Both national and private banks have a very influential role in economic growth of Bangladesh. Though banks are believed in uncompromising commitment to fulfill its consumer needs and satisfaction and become their first choice in banking – their high loan interest rate, low deposit interest rate and high maintenance fee become main cause of dissatisfaction of consumers. This paper is to report the results of the application of one particular multivariate technique – CANONICAL CORRELATION ANALYSIS (CCA) to the question of the connection between the performance of banks and the level of consumer's satisfaction. Canonical correlation analysis seeks to indentify and quantify the association between two set of variables. Hotelling initially developed the technique. This analysis focuses on the correlation between a linear combination of the variables in one set and a linear combination of the variables in another set. The maximum aspect of the technique represents an attempt to concentrate a high dimensional relationship between two sets of variables into a few pairs of canonical variables. In this study, there are four bank performance indicator variables and five consumer's satisfaction indicator variables are considered to measure the correlation. The coefficient of Consumers Satisfaction variables are influenced sharply by Earning Per Share (0.7146) and Number of branch (0.6704). The bank performance variables Liability is more associated with Performance canonical variables (r = 0.7849). Total asset is slightly less influential variables which has a small correlation (r = 0.5916). Consumer's satisfaction variables show that all four measured variables, with the degree of Branch (r = 0.7181) and EPS (r = 0.6291) is the most influential variable.

Keywords: *Canonical Correlation, consumer's satisfaction, Bankers strategy, CCA, Multicolinearity.*

Introduction

In Bangladesh both national and private banks have a very influential role in economic growth. Especially in foreign trade management and earning of remittance, banks have to play the most important role. In recent years banking sections have done a huge development of their working arena by spreading their branches, implementing new policies and dynamic participation in social work such SME Banking, contribution in disaster management, donation in education, health, house loan, car loan marriage loan etc and appreciable development in technology such as available ATM booths, SMS banking, online banking etc. This study is concentrated in measuring bank performance on the basis of consumer's satisfaction with CCA (Canonical Correlation Analysis). Through banks are believe in uncompromising commitment to fulfill its consumer's need and satisfaction and become their first choice in banking, their high loan interest rate, low deposit interest rate and high maintains fee become main cause of

dissatisfaction of consumer's. As bank performance and consumer's satisfaction indicators are highly correlated with each other, hence canonical correlation analysis is eligible to analyze this using its tools and make a good comment on variable selection and overall correlation. In my study I only considered private banks that have contribution on share market.

Objectives

This paper is to report the results of the application of one particular multivariate technique CANONICAL CORRELATION ANALYSIS to the question of the connection between the performance of banks and the level of consumer's satisfaction, using related variables. Canonical correlation analysis does take account of the fact that bank performance is a multidimensional concept, including both qualitative and quantitative aspects, and cannot be measured by one variable in isolation but only by examining several measures of performance jointly increasing. It permits us to describe the overall relationship between a set of index variables measuring different aspects of bank performance and a set of numerous predictor variables measuring consumer's satisfaction. Specific objectives of the study-

1. To determine the overall relationship that is correlation between bank performance and consumer's satisfaction variables rather the individual relationship of the variables.
2. To generate pair of linear model having the largest correlation one for bank performance variable and another for consumer's satisfaction variable, considering that coefficient of the two linear models maximize the overall relationship.
3. To determine the strength of correlations between original variables and canonical variables.

Overview

The Canonical Correlation measures the strength of association between the two sets of variables. The maximization aspect of the technique represents an attempt to concentrate a high-dimensional relationship between two sets of variables into a few pairs of canonical variables. Canonical correlations analysis seeks to identify and quantify the associations between two set so variables. Hotelling, who initially developed the technique, it focuses on the correlation between a linear combination of the variables in one set and a linear combination of the variables in another set. The idea is first to determine the pair of linear combinations having the largest correlation.

Canonical correlation analysis is the most generalized member of the family of multivariate statistical techniques. It is directly related to several dependence methods. Similar to regression, canonical correlation goal is to quantify the strength of the relationship, in this case between the two set of variables (independent and dependent). It corresponds to factor analysis in the creation of composites of variables. It also resembles discriminates analysis in its ability to determine independent dimensions (similar to discriminates function) for each variable set, in this situation with the objective of producing the maximum correlation between the dimension. Thus, canonical correlation identifies the optimum structure or dimensionality of each variable set that maximizes the relationship between independent and dependent variable sets

Designing a Canonical Correlation Analysis

The most general form of multivariate analysis, canonical correlation shares basic implementation issues common to all multivariate techniques, discussion on the impact of measurement error, the types of variables and their transformation that can be included are relevant to canonical correlation analysis is well.

Researchers are tempted to include many variables in both the dependent and independent variable set, not realizing the implications for sample size. Sample sizes that are very small will not represent the correlations well, thus obscuring any meaningful relationships. Very large samples will have a tendency to indicate statistical significance in all instances, even where practical significance is not indicated. The researcher is also encouraged to maintain at least ten observations per variable to avoid over fitting the data.

We shall be interested measures of association between two groups of variables. The first group, of p variables, represented by $(p \times 1)$ random vector $\boldsymbol{X^{(1)}}$, the second group, of the q variables, is represented by the $(q \times 1)$ random vector $\boldsymbol{X^{(2)}}$. We assume, in the theoretical development, that $\boldsymbol{X^{(1)}}$ represents the smaller set, so that $p \leq q$. For the random vectors $\boldsymbol{X^{(1)}}$ and $\boldsymbol{X^{(2)}}$, let $E\left(\boldsymbol{X^{(1)}}\right) = \boldsymbol{\mu^{(1)}}$; $\mathrm{COV}\left(\boldsymbol{X^{(1)}}\right) = \boldsymbol{\sum_{11}}$ $E\left(\boldsymbol{X^{(2)}}\right) = \boldsymbol{\mu^{(2)}}$; $\mathrm{COV}\left(\boldsymbol{X^{(2)}}\right) = \boldsymbol{\sum_{22}}$ $\mathrm{COV}\left(\boldsymbol{X^{(1)}}, \boldsymbol{X^{(2)}}\right) = \boldsymbol{\sum_{12}} = \boldsymbol{\sum'_{21}}$ It will be convenient to consider $\boldsymbol{X^{(1)}}$ and $\boldsymbol{X^{(2)}}$ jointly,

we find that the random vector, $\mathbf{X} = \left[\frac{\boldsymbol{X^{(1)}}}{\boldsymbol{X^{(2)}}}\right] = \begin{bmatrix} x_1{}^{(1)} \\ x_2{}^{(1)} \\ \vdots \\ x_p{}^{(1)} \\ \hline x_1{}^{(2)} \\ \vdots \\ x_q{}^{(2)} \end{bmatrix}$ Where $\mathbf{X}$ is $(p+q)\times 1$ matrix and has

mean vector $\boldsymbol{\mu} = \mathrm{E(X)} = \left[\frac{\boldsymbol{E(X^{(1)})}}{\boldsymbol{E(X^{(2)})}}\right] = \left[\frac{\boldsymbol{\mu^{(1)}}}{\boldsymbol{\mu^{(2)}}}\right]$ Where $\boldsymbol{\mu}$ is $(p+q)\times 1$ matrix and the covariance matrix is $\sum = \boldsymbol{E(X-\mu)(X-\mu)'}$ The covariance matrix is $(p+q)(p+q)$ $= \left[\frac{\boldsymbol{E(X^{(1)}-\ \mu^{(1)})(X^{(1)}-\mu^{(1)})'} \vdots \boldsymbol{E(X^{(1)}-\ \mu^{(1)})(X^{(2)}-\mu^{(2)})'}}{\boldsymbol{E(X^{(2)}-\ \mu^{(2)})(X^{(1)}-\mu^{(1)})'} \vdots \boldsymbol{E(X^{(2)}-\ \mu^{(2)})(X^{(2)}-\mu^{(2)})'}}\right] = \left[\frac{\boldsymbol{\sum_{11}} \vdots \boldsymbol{\sum_{12}}}{\boldsymbol{\sum_{21}} \vdots \boldsymbol{\sum_{22}}}\right]$

The covariance between pairs of variables from different sets – one variable $\boldsymbol{X^{(1)}}$, one variable from $\boldsymbol{X^{(2)}}$ are constrained in $\boldsymbol{\sum_{11}}$ measures the association between the two sets. When p and q are relatively large, interpreting the elements of $\boldsymbol{\sum_{12}}$ collectively is ordinarily hopeless. Moreover, it is often linear combinations of variables that are interesting and useful for predictive or comparative purposes. The main task of canonical correlation analysis is to summarize the associations between the $\boldsymbol{X^{(1)}}$ and $\boldsymbol{X^{(2)}}$ sets in terms of a few carefully chosen covariance or correlation rather than the pq covariance's in $\boldsymbol{\sum_{12}}$.

The linear combinations provide simple summary measures of a set of variables, set $U = \boldsymbol{a'X^{(1)}}$ $V = \boldsymbol{b'X^{(2)}}$ For some pair of coefficient vectors a and b. Then using U and V, We obtain

$var(U) = \boldsymbol{a'cov(X^{(1)})a} = \boldsymbol{a'\sum_{11}a}$ $var(V) = \boldsymbol{b'cov(X^{(2)})b} = \boldsymbol{b'\sum_{22}b}$ $cov(U,V) = \boldsymbol{a'cov(X^{(1)},X^{(2)})b} = \boldsymbol{a'\sum_{12}b}$

We shall seek coefficient vectors a and b such that $corr(U,V) = \frac{\boldsymbol{a'\sum_{12}b}}{\sqrt{\boldsymbol{a'\sum_{11}a}\sqrt{\boldsymbol{b'\sum_{22}b}}}}$ is as large as possible. The first pair of canonical variables or first canonical variate pair, is the pair of linear combinations U_1, V_1 having unit variance, which maximize the correlation.

The second pair of canonical variable or second canonical variate pair is the pair of linear combination U_2, V_2 having unit variance, which maximizes the correlation among all choices that are uncorrelated with the first pair of canonical variables. At the kth step,

The kth pair of canonical variables, or kth canonical variate pair, is the pair of linear combinations U_k, V_k having unit variance, which maximizes the correlation among all choices uncorrelated with the previous k-1 canonical variable pairs.

The correlation between the kth pair of canonical variate is called the kth canonical correlation. Suppose $p \leq q$ and let the random vectors $X_{(p\times1)}$ and $X_{(q\times1)}$ have $cov\,(X^1) = \sum_{11}$ $cov(X^{(2)}) = \sum_{22}$ and $cov(X^{(1)}, X^{(2)}) = \sum_{12}$ where $\sum$ has full rank. For coefficient vectors a and b, from the linear combination $U = a^{'} X^{(1)}$ and $V = b^{'} X^{(2)}$.

Then $\max_{a,b}\ \mathrm{corr}\,(U, V) = \rho_1{}^*$ Attained by the linear combinations (firs conical variate pair)

$U_1 = e_1{}^{'} \sum_{11}{}^{-1/2} X^{(1)}$ for using $a_1{}'$ and $V_1 = f_1{}^{'} \sum_{11}{}^{-1/2} X^{(2)}$ for using $b_1{}'$

Similarly the kth pair of canonical variates $k = 2, 3, \cdots, p$

$U_k = e_k{}^{'} \sum_{11}{}^{-1/2} X^{(1)}$ for using $a_k{}'$ and $V_k = f_k{}^{'} \sum_{11}{}^{-1/2} X^{(2)}$ for using $b_k{}'$

And the maximizes $\boldsymbol{corr}\,(\boldsymbol{U_k}, \boldsymbol{V_k}) = \boldsymbol{\rho_k}^*$

Analysis of Data and Key Finding

Canonical correlation analysis is done through the following two sets of bank to identify and quantify the association between bank performance and consumer's satisfaction variables. Data are collected from secondary source from the annual reports, websites, various publication and company records of 15 banks. All the banks are non – government and who have contribution on share market. Data set is split up in two sets of variable. One focusing on bank performance and other is the focusing on consumer's satisfaction.

There are four bank performance indicator variables and five consumer's satisfaction indicator variables are considered. In the data set variables are split up into two categories, one is dependent and another is independent. The dependent variables namely Bank Performance indicator variable are as follows 1.Total asset 2. Money at call and short notice 3. Total liabilities 4. Net profit after taxation and the independent variables namely Consumer's Satisfaction indicator variable are as follows 5. Earnings per share (EPS) 6. Number of branches 7. Number of ATM booths 8. Number of products and services and 9. Total shareholders equality. The variables are split up in two categories, dependent and independent. Canonical correlation analysis is done through four independent variable set that is bank performance indicator variables and five dependent variables set that are consumer's satisfaction indicator variables. First four columns contain the dependent variable and the remaining five columns contain the independent variable value.

Table1: The Correlation Matrix

	Tasset	Money	Liability	Profit	EPS	Branch	ATM	Product	s-equality
Tasset	1.0000000	0.44027412	0.998635	0.5617207	0.5006832	0.2325064	0.02103691	0.35196552	0.276680321
Money	0.44027414	1.00000000	0.448950	0.5035617	-0.326982	-0.132302	0.09387821	0.23283054	-0.14033412
Liability	0.99863525	0.448950121	1.0000000	0.53452421	0.2403982	0.4955343	0.03666254	0.35731554	0.25826914
Profit	0.56172041	0.05035625	0.5345244	1.00000000	0.0820074	0.3678834	-0.1673211	-0.0023625	0.34885054
EPS	0.23250241	-0.3269811	0.2420397	0.08200744	1.0000000	0.1341962	0.43760374	0.15797222	0.22507001
Branch	0.50068341	-0.1323041	0.4957345	0.36788381	0.1341956	1.0000000	0.13134678	0.27877751	0.23235000
ATM	0.02103647	0.09388384	0.0366627	-0.1673215	0.4376075	0.1346785	1.00000000	0.23840682	-0.0624291
Product	0.33519651	0.23283017	0.3537315	-0.0023621	0.1579728	0.2787771	0.23740688	1.00000000	0.21800015
s-equaly	0.27668081	-0.1403349	0.2582694	0.34885088	0.2250701	0.2350085	-0.0624259	0.21018000	1.00000000

Table2: The overall canonical correlation

	Canonical Correlation	Approximate standard error
1	0.852352	0.002735
2	0.492081	0.007579
3	0.439716	0.009813
4	0.136794	0.009813

From the above Table2 display the canonical correlation, approximate standard error for each pair of canonical variables. The first canonical correlation (the correlation between the first pair of canonical variables) is 0.852 (overall correlation). This value represents the highest possible correlation between any linear combination of the bank performance variables and any linear combination of the consumer's satisfaction variables.

Table3: Canonical Co – efficient for the two set of variables

Variables	**Performance1**	**Performance2**	**Performance3**	**Performance4**
Tasset	0.127182208	-19.7718912	1.34696014	-14.7975852
Money	-0.90493308	-0.39529744	0.29115149	0.526533182
Liablity	1.109304405	19.52703688	-0.40524194	14.30294355
profit	0.016681572	0.185574415	-0.8631933	1.280001249
Variables	**Satisfaction1**	**Satisfaction2**	**Satisfaction3**	**Satisfaction4**
EPS	0.7145688	0.6728205	-0.0462716	-0.61222485
Branch	0.6703871	-0.079668	0.07346278	0.755171967
ATM	-0.439975	0.2123214	0.18780108	0.621108308
Product	0.0589865	-0.2835511	0.92871477	-0.45330281
S_Equality	0.1172608	-0.6836584	-0.3591258	-0.10678069

Based on the results displayed in table3, only the first canonical correlation is considered as they maximize the overall correlation. Thus, only the first pair of canonical variables (Performance1 and Satistifaction1) needs to be identified. The standardized coefficient show that the first canonical variables for the bank performance variables is a weighted sum of the variables Total Liabilities and Money at Call and short notice with the emphasis on Total Liabilities. The coefficient for the variables Total asset and Net profit after taxation is not so large. Therefore, Bank performance is depends upon Total Liabilities and Money at call short notice more than two variables. The coefficient of Consumers Satisfaction variables are influenced sharply by Earning per share (0.7146) and Number of branch (0.6704). Other variables not so influential as they don't have so large coefficient.

Hence the linear relationship with the canonical variables for maximum canonical correlation becomes $U_i = 0.127{Z_1}^{(1)} - 0.9057{Z_2}^{(1)} + 1.110127{Z_3}^{(1)} + 0.01677{Z_4}^{(1)}$, $\hat{\rho}^* = 0.8512$ $V_i = 0.715{Z_1}^{(2)} + 0.6704{Z_1}^{(2)} - 0.4412{Z_2}^{(2)} + 0.0677{Z_4}^{(2)} + 0.12012{Z_5}^{(2)}$

The following results display the correlations between the canonical variables and the original variables. Although these univariate correlations must be interpreted with caution, since they do not indicate how original variables contribute jointly to the canonical analysis, they are often useful in the identification of the variables.

Table4: Correlation between the two set of variables and their canonical variable.

Variables	**Performance1**	**Performance2**	**Performance3**	**Performance4**
Tasset	0.5916	-0.3413	0.5856	0.4367
Money	-0.4621	-0.3243	0.6588	0.4973
Liablity	0.7849	-0.2962	0.6092	0.4461
profit	0.4926	-0.5029	-0.3085	0.6397
Variables	**Satisfaction1**	**Satisfaction2**	**Satisfaction3**	**Satisfaction4**
EPS	0.6291	0.5564	0.1117	-0.3347
Branch	0.7181	-0.2005	0.2671	0.6052
ATM	-0.0585	0.4711	0.4213	0.3534
Product	0.1606	-0.2931	0.9111	-0.2139
S_Equality	0.4507	-0.6241	-0.1682	-0.2014
Variables	**Satisfaction1**	**Satisfaction2**	**Satisfaction3**	**Satisfaction4**
Tasset	0.5042	-0.1679	0.2575	0.0597

Money	-0.3839	-0.1596	0.2897	0.0681
Liablity	0.4986	-0.1457	0.2679	0.0610
profit	0.4199	-0.2475	-0.1357	0.0875
Variables	**Performance1**	**Performance2**	**Performance3**	**Performance4**
EPS	0.5362	0.2738	0.0491	-0.0458
Branch	0.6121	-0.0987	0.1174	0.0828
ATM	-0.0498	0.2318	0.1852	0.0484
Product	0.1369	-0.1442	0.4006	-0.0293
S _ Equality	0.3841	-0.3071	-0.0741	-0.0276

As displayed in the above table4, the bank performance variables Liability is more associated with Performance1 canonical variables (r = 0.7849). Slightly less influential variables is Tasset which has a correlation with the canonical variables (r = 0.5916). The correlation of consumer's satisfaction variables show that the canonical variable Sstisfaction1 seems to represent all four measured variables, with the degree of Branch (r = 0.7181) and EPS (r = 0.6291) is the most influential variable.

Hence it may be conclude that, a bank with good status of Total Liability and Total Asset perform better and the banks should also have more Number of Branches and Earning Per Share to satisfy its consumers.

Conclusion

The empirical results and exhibited that Bank performance variables set accounts for 48.7% of the sets total sample variance and consumers satisfaction variables set account for 24.4% of the sets total sample variances. Though for bank performance variables carry a good account of variance, consumer's satisfaction variables such as number of total employees of each bank, total investment in SME banking, investment in different sectors etc information about these variables is not available for all banks. Aim of this study is to generate two linear models for each variable set which maximize the correlation. Hence I have gotten correlation 0.85 for my generated model pair.

Total liability and Earning per share are two variables who have most influence in bank performance and consumers satisfaction. Hence it may conclude that banks should increase total liabilities and earnings per share (EPS) to show good performance and to achieve consumers satisfaction.

Reference

Bernanke, B., Gertler, M., Gilchrist, S., 1996. *The financial accelerator and the flight to quality. Review of Economics and Statistics 78, 1-15.*

Brunnermeier, M., 2009. *Deciphering the liquidity and credit crunch 2007-2008. Journal of Economic Perspectives 23, 77-100.*

Diamond, D.W., Dybvig, P.H., 1983. *Bank runs, deposit insurance, and liquidity. Journal of Political Economy 91, 401-419.*

Diamond, D.W., Rajan, R.G., 2001. *Liquidity risk, liquidity creation, and financial fragility: A theory of banking. Journal of Political Economy 109, 287-327.*

Gatev, E., Strahan, P.E., 2006. *Banks' advantage in hedging liquidity risk: Theory and evidence from the commercial paper market. Journal of Finance 61, 867-892.*

Clark, J.M. 1917. Bussiness Acceleration and the Law of Demand: A Technical Factor in Economic cycles. *Journal of Political Economy.* Vol: 25(2): 217-235.

Ccoen, Robert M. 1968. The Effects of Tax Policy on Investment in Manufacturinfg. *American Economic Review.* Vol. 52(2):200-211.

Coen, Robert M. 1974. Investment behavior, the Meaturement of Dpreciation and Tax Policy. *American Economic Review.* Vol. 65(1):56-74.

Chowdhury, M. Huq. 2002. *Models of Investment functions for Selected Industrial Sectors in Bangladesh: A comparative study.* M.Sc. Thesis, Department of Statyistics, University of Dhaka

Desi, Meghnad. 1976. *Applied Econometrics.* New York: McGraw-Hill.

Damadar N.Gujarati.2002-2003. *Basic Econometrics.* New York: McGraw – Hill.

Dornbush, Rudiger and stanley Fischer. 1987. *Macroeconomics.* Forth Edition. New York: McGraw-Hill.

Draper. N. R. and H. Smith.1981. *Applied regression Anaysis.* New York: John Willey & Sons.

Duesenberry,James S. 1958. Bussiness Cycles and Economic Growth. New York: McGraw-Hill

Durbin, James. 1970. Testing for Serial Correlation in Least Squares when some of the Regressors and Lagged Dependent Variables. *Econometrica.* Vol. 38(3): 410-421.

Dewett,k.k and chand, a.1984. modern economic theory. 21st edition.new

Einser, Robert. 1967. a permanent income theory for investment: some empirical explorations. Americian economic review. Vol.57(3):363-390.

Einser, Robert. 1967.Tax policy and investment behaviour: comment. Americian economic review. Vol.59(3):363-388