How to choose a profitable firm:
In The View of Investors

20175404 Rocku Oh
20176027 Seonu Jang

Ulsan National Institute of Science and Technology

Business Issue

- **ROE** is one of the most important factors that **Warren Buffett** looks at when deciding whether to make a deal or not.

- ROE (Returns on Equity) = Net Income / Shareholder's Equity

    - The factor represents profitability of firms.

Background

The capital structure

- The choice of capital structure is fundamentally a essential problem (Brealey and Myers, 2013)

- Firms can use either debt or equity capital to finance their assets

- Pecking order theory (first preferring internal financing, and then debt, lastly raising equity as a "last resort")

- It is crucial for business organizations to maximize returns of the firms on the firm's ability to deal with its competitive environment

The relationship between capital structure and profitability

- An improvement in the profitability is necessary for the long-term survivability of the firm.

Problem Statement and Hypothesis

Problem statement

- It is difficult for investors to understand the nature of firms due to too many financial factors with some amount multicollinearity.

    - Investors should know which capital structure is the most ideal to improve ROE (profitability)

    - Investors need to know how a firm handles their debt properly

Hypothesis

- Capital structure of firms reflects the current state and capability of firms.

- Therefore, highly profitable firms and poorly profitable firms will have different capital structure, and we can find good companies by analyzing financial structure of firms.

Research Question

✓ **RQ1) To what extends are the capital structure and expenses related to the firm profitability?**

- *Regression analysis*

✓ **RQ2) How can investors estimate or predict the future profitability of firm on the basis of the current capital structure and expenses?**

- *Classification analysis*

✓ **RQ3) Is high leverage always bad for firms? : In the view of investors**

- *Clustering analysis*

Data Preprocessing

✓ **We gathered capital structure data from "Dataguide" subscribed by UNIST**

- We remove data where equity is less than 10 Mio.

- We use box plot to check the data distribution, and remove data which seems to be erroneous, or extreme outliers

    - ROE(Return on Equity) between -2 and 2

    - Debt Ratio (Total Debt / Equity) less than 1

    - Cost-related Ratio between 0 and 2

    - Income-related Ratio between -5 and 1

- 1419 firms listed in KOSPI and KOSDAQ stock exchange during 2001 to 2017.

- Total number of observations is 13714 after preprocessing stage

RQ1

Effects of Capital Structure and Expense on Firm Profitability

Previous Work

A number of theories have been advanced in explaining the capital structure of firms

- Wald (1999) found a negative correlation between leverage and profitability.

- Abor (2005) found i) a positive relationship between the ratio of short-term debt to total assets and return on equity, ii) a negative relationship between the ratio of long-term debt to total assets and return on equity

- Mendell, Sydor, and Mishra found (2006) negative relationship between profitability and debt.

Limitation of previous studies

- Most studies have been conducted on a single industry.

- Does not consider fixed effects (e.g. firm effect and year effect)

  - *The residuals may be correlated across firms or across time, and OLS standard errors can be biased. (Peterson, 2009)*

Variables

- To remain consistent with previous studies, measures pertaining to capital structure and profitability were taken from Abor's (2005, p. 442) study.

- Two control variables (firm size, sales growth) were also included as standard determinants of corporate profitability.

Proxy variables definition

Proxy Variables	Definition
Short-Term Debt (STD)	Short-term debt divided by the total assets
Long-Term Debt (LTD)	Long-term debt divided by the total assets
Firm Size (SIZE)	Natural Logarithm of Firm's Sales, lagged one-year period
Sales-Growth (SG)	Current year's Sales minus previous year's sales divided by previous year's sales
Cost of Goods-Sold Ratio (COGSR)	Cost of goods sold divided by the total sales
Labor Cost Ratio (LCR)	Labor cost divided by the total sales
ADVCR	Advertising cost divided by the total sales
SCR	Selling cost divided by the total sales
ADMCR	Administrative cost divided by the total sales

Normality

Outlier Detection

- After removing outliers by utilizing Cook Distance measure, the dependent variable (ROE) are followed as a Normal Distribution

- with some relatively large kurtosis value
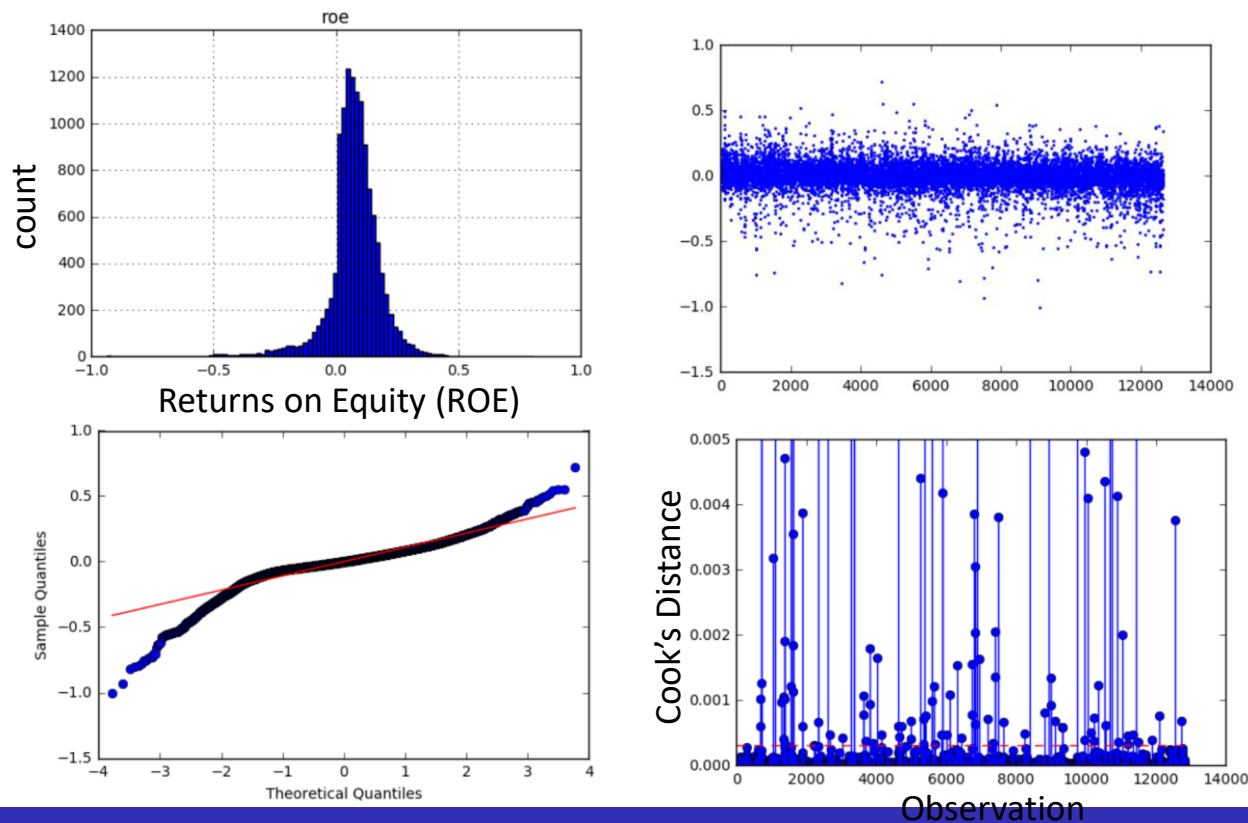
Summary Stats.

Descriptive statistics of the collected dataset.

- All variables are calculated based on the book value on the yearly financial statements.

- Korean companies tend to depend on short-term debt rather than long-term debt. Short-term debt to total assets is 29.5% and long-term to total assets is 10.8%

- In terms of expense, cost of goods sold takes the biggest portion followed by labor cost, administrative costs, and advertisement costs.

Descriptive statistics of the variables

	ROE	SG	STD	LTD	SCR	COGSR	LCR	ADVCR	ADMCR
Obs.	12638	12638	12638	12638	12638	12638	12638	12638	12638
Mean	0.073	0.151	0.295	0.108	0.030	0.765	0.058	0.010	0.045
Std.	0.110	1.276	0.150	0.096	0.045	0.180	0.055	0.023	0.057
Min	-0.935	-0.999	0.000	0.000	-0.062	0.000	0.000	0.000	-0.009
Max	0.807	90.635	0.849	0.680	0.542	5.412	2.344	0.384	0.638

Multicollinearity

- Debt ratio components are less correlated with the cost ratio components

- Within cost category, there are relatively large correlation

   - COGSR and other costs, ADMCR and other costs

- We take all the variables as regressors because individual factors are has economical meaning even though relatively high Variance Inflation Factor(VIF) values

Correlation matrix of the variables

		1	2	3	4	5	6	7	8	VIF
1	STD	1.000								5.5
2	LTD	0.105	1.000							2.3
3	SG	0.041	0.008	1.000						1
4	COGSR	0.274	0.117	-0.006	1.000					6.5
5	LCR	-0.206	-0.085	-0.035	-0.594	1.000				3
6	ADVCR	-0.095	-0.045	0.003	-0.574	0.408	1.000			1.7
7	SCR	-0.016	-0.029	-0.013	-0.369	0.173	0.265	1.000		1.5
8	ADMCR	-0.123	-0.028	-0.025	-0.636	0.556	0.476	0.161	1.000	2.6

OLS and regression result of capital structure and expense on firm profitability (ROE)

	OLS (1)	OLS (2)	OLS (3)	OLS (4)	OLS (5)
Dep. Var.	ROE	ROE	ROE	ROE	ROE
Intercept	0.001 (0.1055)	-0.044*** (-3.6071)	-0.057*** (-4.6707)	0.327*** (24.0990)	0.257*** (6.3637)
SIZE	0.005*** (7.8209)	0.007*** (11.1050)	0.009*** (13.6850)	0.007*** (10.8740)	0.011*** (4.9483)
STD	-0.082*** (-12.281)		-0.080*** (-12.070)	-0.012* (-1.9122)	**-0.047*** (-4.7478)**
LTD		-0.204*** (-19.144)	-0.201*** (-19.006)	-0.156 (-16.122)	**-0.161*** (-11.386)**
SG	0.006*** (7.8201)	0.006*** (7.7246)	0.006*** (8.4126)	0.004*** (5.9128)	0.004*** (5.7201)

Cont.					
SCR				-0.536*** (-25.645)	**-0.778*** (-18.822)**
COGSR				-0.390*** (-50.034)	-0.376*** (-37.436)
LCR				-0.484*** (-22.823)	-0.212*** (-7.0671)
ADMCR				-0.445*** (-21.185)	-0.518*** (-14.198)
ADVCR				-0.279*** (-5.9416)	-0.431*** (-4.6269)
Year Effect	NO	NO	NO	NO	Yes
Firm Effect	NO	NO	NO	NO	Yes
Observations	12638	12638	12638	12638	12638
R-squared	0.034	0.017	0.045	0.217	0.163
F-statistic	146.4	74.115	147.48	388.46	242.4

Note: *t*-statistics are in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

✓ **RQ1) To what extends are the capital structure and expenses related to the firm profitability?**

- In consistent with other studies, We find that there is **negative relationship of both short-term debt ratio and long-term debt ratio on firm profitability(ROE)**.

- **Many companies cannot take the advantages of leverage effect** enough. Long-term debt has more negative impact on ROE than short-term debt has.

- Selling costs has the most negative effect, followed by administrative costs, and advertisement costs. These costs are considered relatively easy to reduce.

- Without considering fixed effects, short-term debt and long-term debt effects are underestimated. In addition, selling costs, administrative costs and advertisement costs also are underestimated.

OLS and regression result of capital structure and expense on firm profitability (ROE) with different industries

	IT OLS (1)	경기소비재 OLS (2)	금융 OLS (3)	산업재 OLS (4)	소재 OLS (5)	Cont.					
Dep. Var.	ROE	ROE	ROE	ROE	ROE	SCR	-1.163*** (-7.4682)	-0.551*** (-8.5636)	17.361 (0.8197)	-1.120*** (-11.103)	-1.161*** (-7.4352)
Intercept	0.430*** (3.9793)	0.349*** (4.3830)	-31.491 (-0.6876)	0.740*** (7.0979)	0.666*** (6.6157)	COGSR	-0.214*** (-10.092)	-0.398*** (-18.299)	-0.648 (-0.3246)	-0.806*** (-25.008)	-0.810*** (-24.736)
SIZE	-0.005 (-0.8715)	0.007 (1.5993)	1.532 (0.6833)	0.007 (1.3143)	0.012** (2.3403)	LCR	-0.300*** (-2.6586)	-0.618*** (-8.2816)	9.129 (0.7660)	-0.967*** (-6.1911)	-1.388*** (-6.7670)
SDR	**-0.097*** (-3.2933)**	0.005 (0.2787)	0.542 (0.7402)	-0.024 (-1.1366)	-0.100*** (-4.6498)	ADMCR	-0.361 (-0.8527)	-0.439*** (-3.0746)	-9.929 (-0.5551)	-1.041 (-1.5937)	-0.329 (-0.3400)
LTR	**-0.262*** (-5.4109)**	-0.128*** (-4.6873)	-0.687 (-0.6769)	-0.141*** (-4.7440)	-0.068** (-2.2524)	ADCR	-0.449*** (-4.3518)	-0.378*** (-7.2220)	17.547 (0.5818)	-1.439*** (-8.0844)	-1.434*** (-6.5216)
SG	0.008*** (2.8964)	0.001 (0.4985)	0.531 (0.6730)	0.009*** (3.9593)	0.003*** (3.0624)	Year Effect	Yes	Yes	Yes	Yes	Yes
						Firm Effect	Yes	Yes	Yes	Yes	Yes
						Observations	2264	2894	30	2605	2181
						R-squared	0.178	0.172	0.869	0.259	0.282
						F-statistic	45.556	58.872	0.738	89.512	85.278

*Note: t-statistics are in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.*

OLS and regression result of capital structure and expense on firm profitability (ROE) with different industries

	에너지 OLS (1)	유틸리티 OLS (2)	의료 OLS (3)	통신서비스 OLS (4)	필수소비재 OLS (5)	Cont.					
Dep. Var.	ROE	ROE	ROE	ROE	ROE	SCR	-0.568 (-1.4734)	0.038 (0.0252)	-0.794*** (-10.913)	-2.996 (-0.2739)	-0.751*** (-7.0983)
Intercept	-0.183 (-0.6300)	-0.588 (-1.4943)	0.018 (0.1886)	3.231 (0.1994)	0.497*** (3.7499)	COGSR	-0.546*** (-6.1350)	-0.197*** (-3.2999)	-0.484*** (-18.936)	0.087 (0.0554)	-0.567*** (-9.8235)
SIZE	**0.038** (**2.4709**)	**0.044** (**2.3683**)	0.030*** (5.6687)	-0.131 (-0.1812)	0.004 (0.5578)	LCR	0.595 (1.4904)	-1.608** (-1.9975)	-0.715*** (-9.6476)	-1.826 (-0.9032)	-0.121*** (-3.0198)
SDR	-0.050 (-0.6497)	-0.050 (-0.7615)	-0.026 (-1.0183)	-0.239 (-0.2383)	0.054 (1.4802)	ADMCR	2.866 (1.2859)	-5.316 (-0.9706)	-0.497*** (-4.1174)	-3.171 (-0.5903)	0.080 (0.4306)
LTR	-0.151 (-1.4471)	-0.180*** (-3.1457)	-0.077*** (-2.5695)	0.353 (0.3006)	-0.288*** (-6.6626)	ADCR	-0.731* (-1.9493)	0.161 (0.2808)	-0.756*** (-10.130)	-1.508 (-0.6075)	-0.489*** (-4.2071)
SG	0.011 (1.6334)	-0.029 (-1.2316)	0.029*** (6.5738)	0.082 (0.2600)	0.001 (0.4981)	Year Effect	Yes	Yes	Yes	Yes	Yes
						Firm Effect	Yes	Yes	Yes	Yes	Yes
						Observations	270	206	1132	31	1025
						R-squared	0.176	0.351	0.387	0.765	0.171
						F-statistic	5.092	10.024	68.975	1.083	20.673

Note: t-statistics are in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

✓ **RQ1) To what extends are the capital structure and expenses related to the firm profitability?**

- **Firm size positively affects energy and utility firms** requiring economies of scale.

  - they can efficiently reduce variable costs when the firm size is large.

- **Long-term debt ratio negatively affects IT firms at most** among the 12 sectors

  - Debt itself cannot contribute to the profitability of technology due to the pace of technology change (shift).

RQ2)

***How can investors estimate or predict
the future profitability?***

It is important to predict how ROE will change in the future.

- It matters for the investors who considering investment to the firm in financial markets as well as for the managerial decision

Many financial factors in the financial statements has high collinearity.

- it is necessary to use adequate subset selection methods and pick up significant factors in order to mitigate the multicollinearity problem

Hypothesis

- Financial structure in the current year will be important factor to predict firm profitability in the next year.

- Firms with negative income, and positive income in the past has different capital structure respectively, so it requires to use different prediction model for each case.

Variables

✓ **Variables**

- We first start with 19 predictors from financial factors related with debt and expense costs.

- We use the best subset selection method based on AIC and BIC to reduce features

✓ **Dependent Variable**

- We use a variable PI(Profitability indicator): "ROE Increased" and "ROE Decreased" compared to ROE of previous fiscal year for all firms

$$PI_t = \begin{cases} 1, where\ ROE_t \leq ROE_{t+1} \\ 0, otherwise \end{cases}$$

Subset Selection

We first start with 19 predictors.

- We remove predictors with high collinearity

Variable 1	Variable 2	Correlation
~~net_income_to_sales~~	roe	66.49%
~~gross_profit_to_sales~~	cost_of_goods_sold_to_sales	-100.00%
~~selling_and_administrative_expenses_to_sales~~	cost_of_goods_sold_to_sales	-51.48%
selling_and_administrative_expenses_to_sales	~~gross_profit_to_sales~~	51.48%
gross_profit_to_sales	~~selling_and_administrative_expenses_to_sales~~	51.48%
labor_costs_to_sales	~~selling_and_administrative_expenses_to_sales~~	84.77%
rnd_costs_to_sales	~~selling_and_administrative_expenses_to_sales~~	55.62%
selling_costs_to_sales	~~selling_and_administrative_expenses_to_sales~~	52.57%
administrative_costs_to_sales	~~selling_and_administrative_expenses_to_sales~~	80.04%
rent_to_sales	~~selling_and_administrative_expenses_to_sales~~	53.19%
operating_income_to_sales	~~selling_and_administrative_expenses_to_sales~~	-52.90%
~~administrative_costs_to_sales~~	labor_costs_to_sales	68.99%
~~administrative_costs_to_sales~~	rent_to_sales	59.67%
roe	~~operating_income_to_sales~~	58.01%
~~net_income_to_sales~~	~~operating_income_to_sales~~	79.58%
roe	~~net_income_to_sales~~	66.49%
cost_of_goods_sold_growth	~~sales_growth~~	92.75%
sna_expenses_growth	~~sales_growth~~	51.02%

Subset Selection

**We use the best subset selection based on AIC and BIC to reduce features.**

- 9 and 3 variables are selected according to the AIC and BIC criteria respectively.

- We use 6 (=(9+3)/2) variables in this model
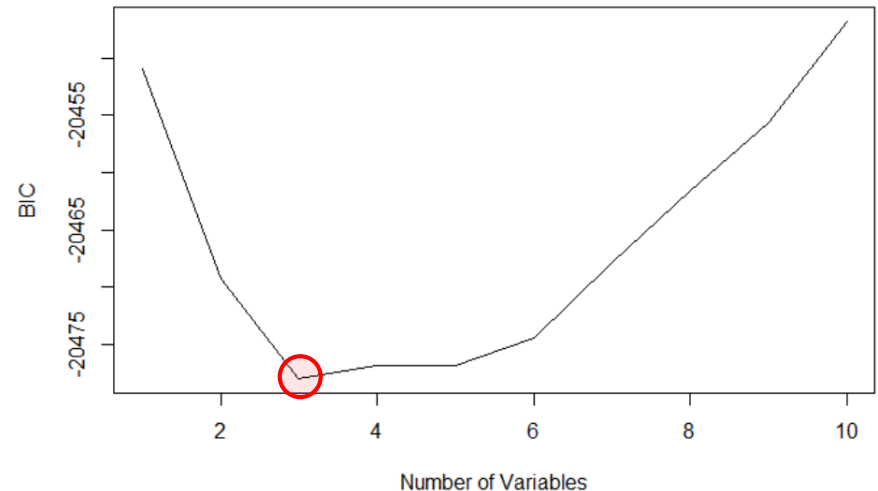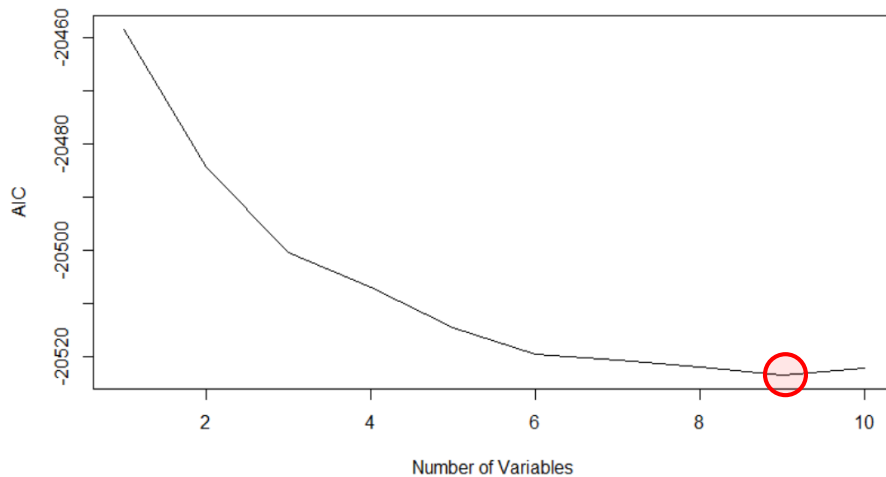
Subset Selection

Visualization of the most important two factors which is result of subset selection.

- High ROE has tendency to decrease in the future, and low ROE has tendency to increase in the future – "Mean reversion theory"

- High short term ratio seems to increase the future ROE, but not statistically significant
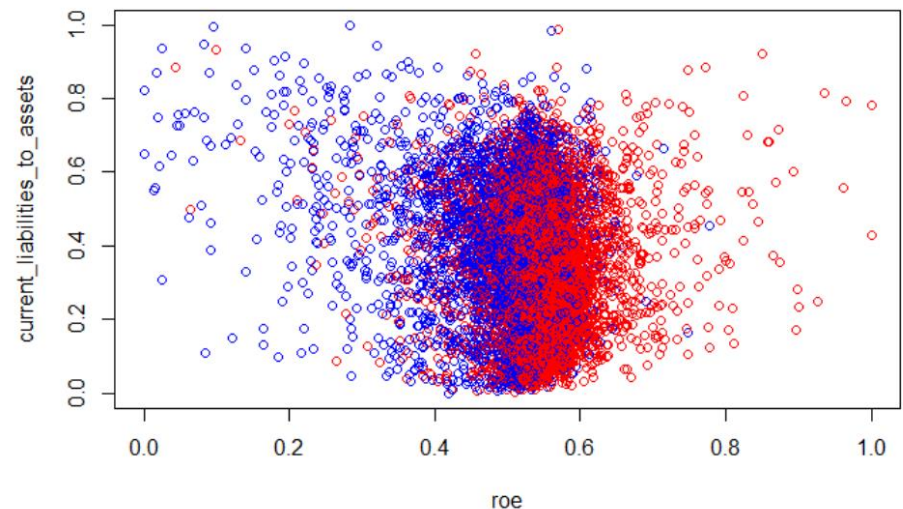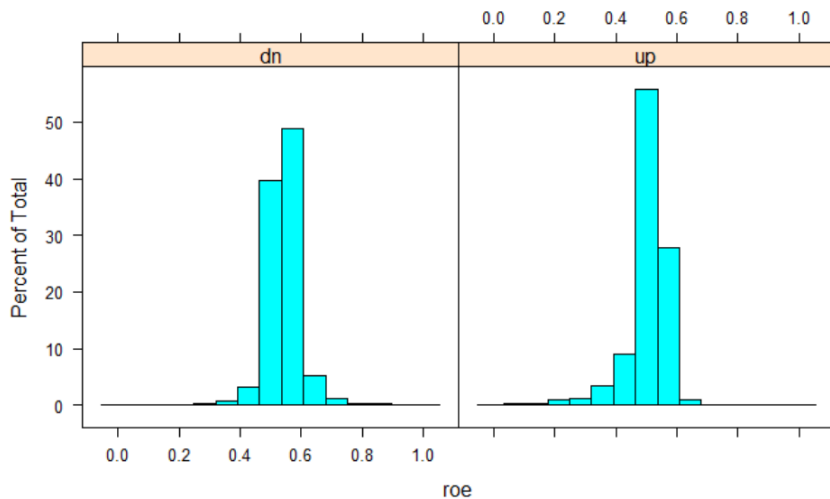
Final Variables

Dependent Variable

- We use a variable PI(Profitability indicator): "ROE Increased" and "ROE Decreased" compared to ROE of previous fiscal year for all firms

$$PI_t = \begin{cases} 1, where\ ROE_t \leq ROE_{t+1} \\ 0, otherwise \end{cases}$$

Predictor variables definition

Predictor Variables	Definition
Short-Term Debt (STD)	Short-term debt divided by the total assets
Long-Term Debt (LTD)	Long-term debt divided by the total assets
ROE (ROE)	Returns on equity
SNA Expense Growth(SEG)	
Labor Cost Ratio (LCR)	Labor cost divided by the total sales
Rent Cost Ratio(RR)	Rent cost divided by the total sales
Sectors (SECTOR)	Advertising cost divided by the total sales

Summary Stats.

Descriptive statistics of the collected dataset.

- All variables are calculated based on the book value on the yearly financial statements.

- Korean companies tend to depend on short-term debt rather than long-term debt. Short-term debt to total assets is 32.25% and long-term to total assets is 11.85% on average.

- Selling and Expenses grows by 27.64% on average every year.

Descriptive statistics of the variables

	ROE	STD	LTD	SEG	LCR	RR
Obs.	13655	13655	13655	13655	13655	13655
Mean	0.0573	0.3225	0.1185	0.0971	0.0506	0.0018
Std.	0.2501	0.1642	0.1055	0.2764	0.0703	0.0114
Min	-1.9715	0.0009	0.0000	-0.9620	0.0020	0.0006
Max	1.8766	0.9613	0.9025	1.9581	0.8850	0.1918

Methodology

Classification 1

- Dataset is divided into training dataset and test dataset as 7:3 ratio.

- We apply logistic regression (LR), linear discriminant analysis (LDA), quadratic discriminant analysis (QDA), and Support Vector Machine (SVM) models for stepwise features.
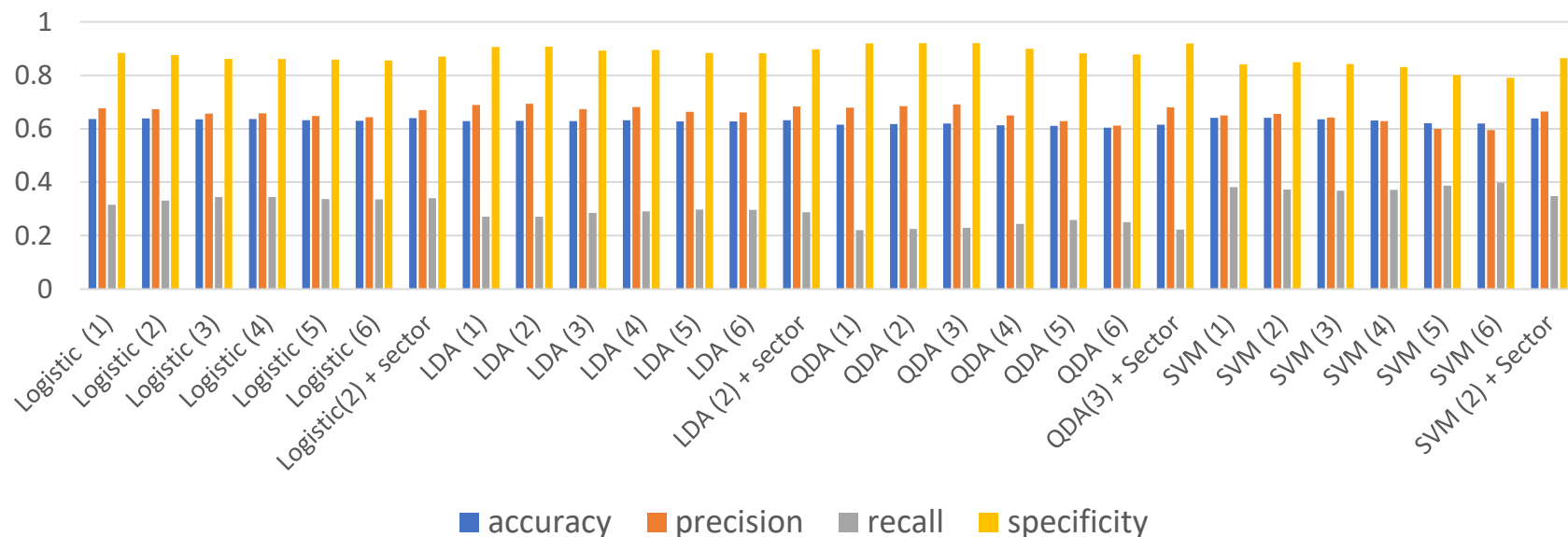
Classification 2 and 3

- We split the dataset into two groups: negative ROE_t companies and positive ROE_t companies at $t's$ fiscal period

- We classify dependent variables(PI_t) as "up" and "down" respectively.

- Datasets are divided into training datasets and test datasets as 7:3 ratio for each dataset.

- We apply logistic regression(LR), linear discriminant analysis(LDA), quadratic discriminant analysis(QDA) models for stepwise features.

Classification 1 (10-fold cross validation result)

- SVM model using the first two variables has the most high accuracy and recall compared to other models.

- LDA models have high average precision compared to the other methods. The model can effectively predict the future ROE as "increasing" in terms of precision

- QDA model using the first three variables has the most high specificity compared to other models.
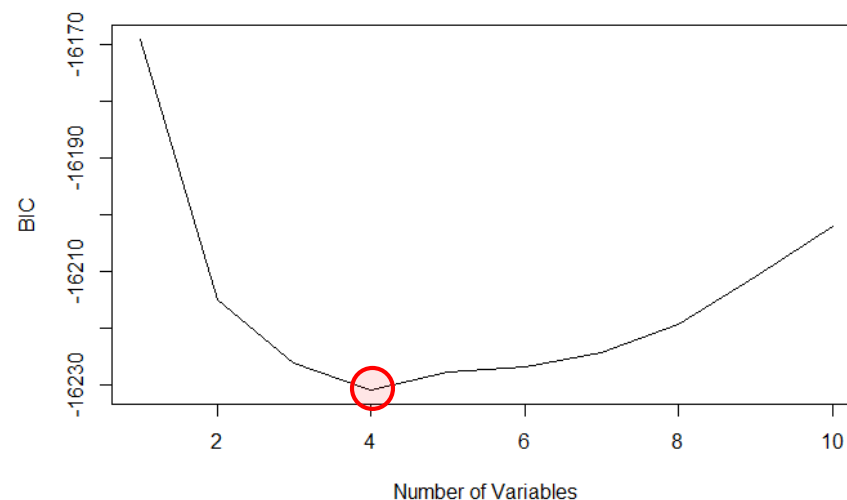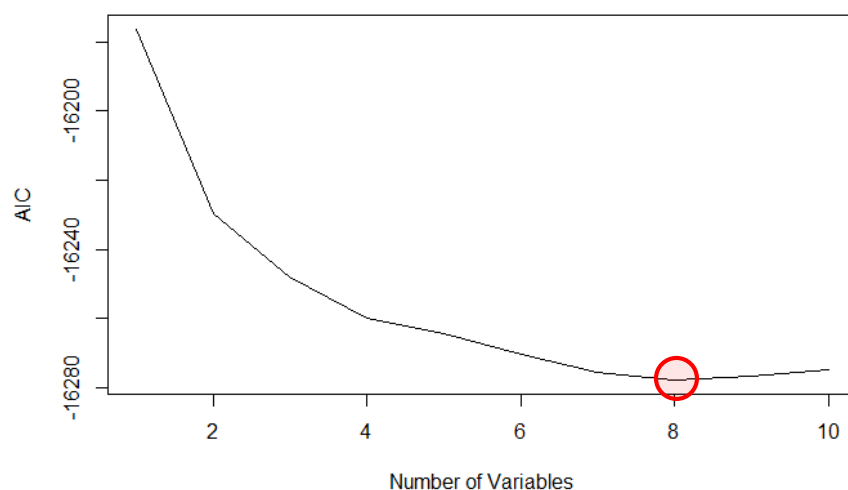
To sum up, We select the SVM (2) to be the best model for the prediction of the future ROE.

Classification 2 – 흑자기업(Positively Profitable firms at year t)

- 8 and 4 variables are selected according to the AIC and BIC criteria respectively.

- We use 4 variables in this model for the better interpretation

    - ROE, Short-term Debt Ratio, Long-term Debt Ratio, RND Cost Ratio

Classification 2 – Positively Profitable firms at year t (10-fold cross validation result)

| Method	Variables	Accuracy	Precision	Recall	Specificity
LR	ROE	0.632	0.542	0.191	0.901
	ROE, STD	0.627	0.518	0.213	0.879
	ROE, STD, RND	0.627	0.519	0.217	0.877
	roe, STD, RND, LTD	0.632	0.531	0.237	0.872
LDA	ROE				
	ROE, STD	0.623	0.519	0.088	0.950
	ROE, STD, RND	0.624	0.521	0.101	0.943
	ROE, STD, RND, LTD	0.633	0.566	0.129	0.939
QDA	ROE	0.571	0.462	0.805	0.428
	ROE, STD	0.574	0.462	0.739	0.475
	ROE, STD, RND	0.584	0.460	0.565	0.595
	ROE, STD, RND, LTD	0.585	0.461	0.571	0.594

Note COGSR : RNDR: R&D Ratio, STD: Short-term Debt Ratio, LTD: Long-term Debt Ratio

RQ2) How can investors estimate or predict the future profitability?

Classification 2 – Positively Profitable firms at year t (10-fold cross validation result)

- LDA models with three variables (ROE, STD, and RNDR) have the most high average accuracy and specificity compared to the other methods.

  - The model can effectively predict the decreasing future ROE as "decrease" in terms of specificity

- QDA model using the two variables(ROE and STD) has the most high recall compared to other models.

- We focus on the case that positively profitable firms goes less performance than that of previous year.

 - We give more weight to the specificity for the purpose when we select the best model

To sum up, We select the LDA (3) to be the best model for the prediction of the future ROE.
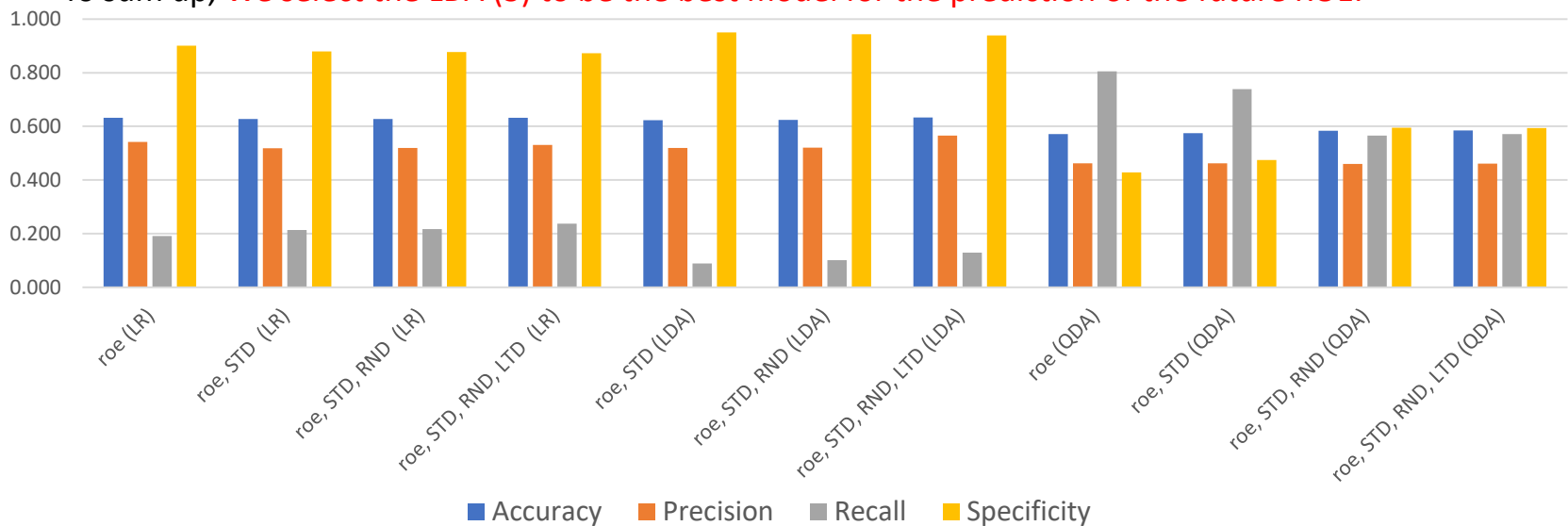
Classification 3 – 적자기업(Negatively Profitable firms at year t)

- 6 and 2 variables are selected according to the AIC and BIC criteria respectively.

- We use 4 variables in this model for the better interpretation

    - Sales, COGS Ratio, ROE, Rend Cost Ratio
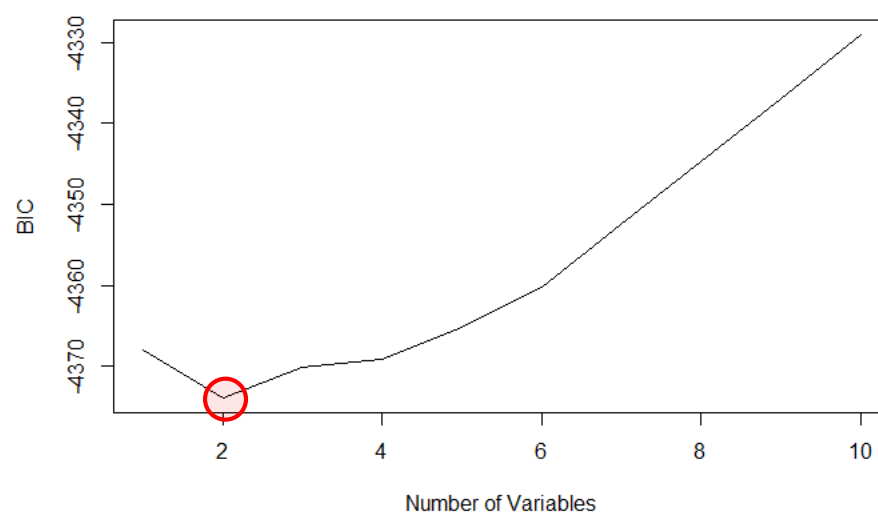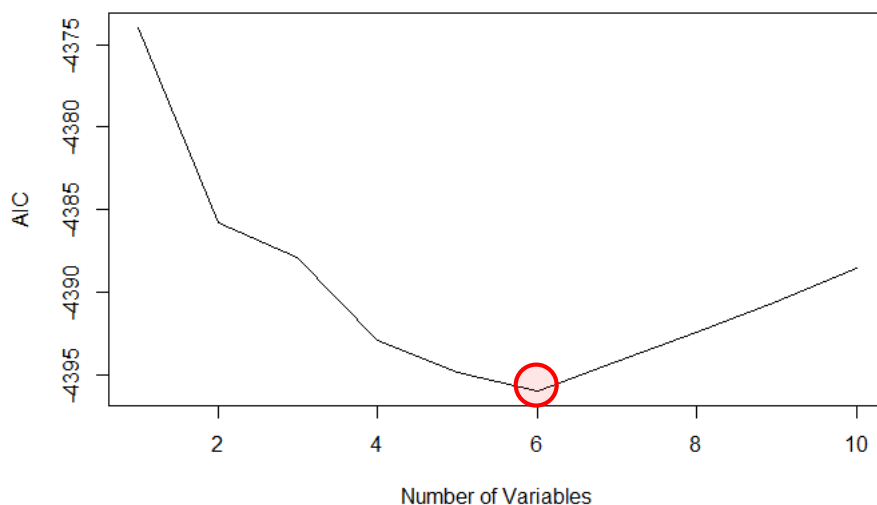
Classification 3 – Negatively Profitable firms at year t (10-fold cross validation result)

Method	Variables	Accuracy	Precision	Recall	Specificity
Logistic Regression	ROE	0.530	0.708	0.521	0.521
	ROE, Sales	0.552	0.725	0.543	0.569
	ROE, Sales, COGSR	0.555	0.731	0.542	0.583
	ROE, Sales, COGSR, RCR	0.569	0.739	0.562	0.583
LDA	ROE	0.677	0.677	1.000	0.000
	ROE, Sales	0.677	0.677	1.000	0.000
	ROE, Sales, COGSR	0.677	0.676	1.000	0.000
	ROE, Sales, COGSR, RCR	0.677	0.678	1.000	0.000
QDA	ROE,	0.677	0.676	1.000	0.000
	ROE, Sales	0.660	0.675	0.960	0.033
	ROE, Sales, COGSR	0.629	0.668	0.896	0.069
	ROE, Sales, COGSR, RCR	0.633	0.671	0.900	0.076

Note COGSR : Cost-of-goods-sold Ratio, RCR : Rent Cost Ratio

Classification 3 – Negatively Profitable firms at year t (10-fold cross validation result)

- Logistic regression models with four variables (ROE, Sales, COGSR, and RCR) have the most high precision and specificity compared to the other methods.

- LDA and QDA models have similarly high accuracy. However, the performance of specificity is too low to select the model.

- We focus on the case that negatively profitable firms goes better performance than that of previous year.

  - We give more weight to the recall for the purpose when we select the best model

To sum up, We select the LR (4) to be the best model for the prediction of the future ROE.
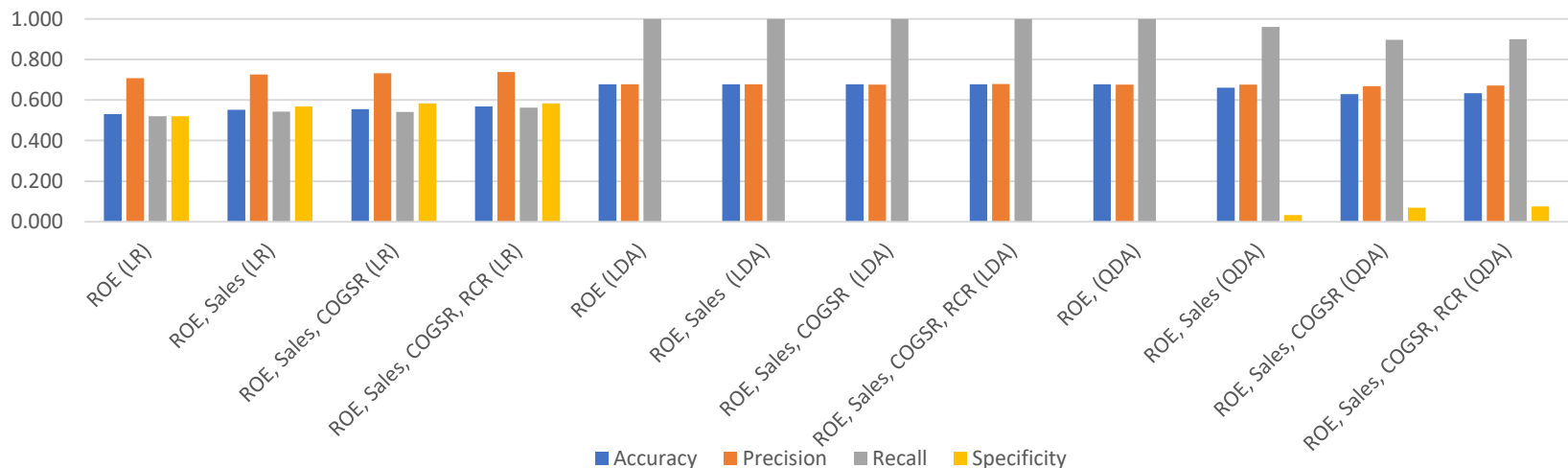
Classification 1 (10-fold cross validation result)

- SVM model using the two variables(ROE and STD) has the most high accuracy and recall compared to other models.

Classification 2 – Positively Profitable firms at year t (10-fold cross validation result)

- LDA models with three variables (ROE, STD, and RNDR) have the most high average accuracy and specificity compared to the other methods.

  - *The model can effectively predict the decreasing future ROE as "decrease" in terms of specificity*

Classification 3 – Negatively Profitable firms at year t (10-fold cross validation result)

- Logistic regression models with four variables (ROE, Sales, COGSR, and RCR) have the most high precision and specificity compared to the other methods.

  - *We focus on the case that negatively profitable firms goes better performance than that of previous year.*

✓ **RQ2) How can investors estimate or predict the future profitability?**

- In Regression results show that **high leverage negatively** affects firm's profitability.

- Classification results show that the firms with high ROE will likely have low ROE in the following year(Mean-reversion), and **high leverage** will likely to affect **positively** ROE in the following year.

- Therefore, firms which can **use leverage efficiently to increase profitability** will likely maintain consistent high profitability.

We cluster firms according to debt-handling ability in the next stage.

RQ3)

Is high leverage always bad for firms?

RQ3) *Is high leverage always bad for firms?*

It is difficult for investors to understand the nature of firms due to too many financial factors with high collinearity.

Hypothesis

- Firms will have different strategies and ability in terms of how to deal with leverage.

- There are firms which take advantage of leverage to improve there profitability, and firms which are not. These firms will be differently clustered.

Variable Selection

We select the initial variables based on the previous project (Classification)

- We remove variables with high correlation coefficient to remove multicollinearity issue

- We standardize values of each variables, which result in low variance inflation factors

Variables definition

Variables	Definition	VIF
Short-Term Debt (STD)	Short-term debt divided by the total assets	1.11
Long-Term Debt (LTD)	Long-term debt divided by the total assets	1.02
ROE (ROE)	Returns on equity	1.38
Administrative Cost Ratio(ADMCR)	Administrative cost divided by total sales	1.23
RND Cost Ratio(RNDR)	RND cost divided by total sales	1.17
Sales Growth (SG)	Sales Growth	1.07
Net Income Ratio(NIR)	Net Income divided by total sales	1.48

Summary Stats.

Descriptive statistics of the collected dataset.

- All variables are calculated based on the book value on the yearly financial statements.

- Korean companies tend to depend on short-term debt rather than long-term debt. Short-term debt to total assets is 32.25% and long-term to total assets is 11.85% on average.

- Sales growth by 33.50% on average every year.

- Net income margin is about 1% on average.

Descriptive statistics of the variables

	ROE	STD	LTD	SG	ADMCR	RNDR	NIR
Obs.	9525	9525	9525	9525	9525	9525	9525
Mean	0.0573	0.3225	0.1185	0.1241	0.0529	0.0239	0.0099
Std.	0.2501	0.1642	0.1055	0.3350	0.0610	0.0515	0.2515
Min	-1.9715	0.0009	0.0000	-1.4449	0.0007	0.0000	-4.8795
Max	1.8766	0.9613	0.9025	3.7923	1.3774	1.4240	0.8470

Methodology

Factor analysis - PCA

- Based on the PCA result, we analyze variables.

- We extract three PCs from the original 7 variables.

Clustering 1 – Hierarchical clustering

- We use hierarchical clustering using the principal components.

- We use Ward's minimum variance method as a distance measure for the clustering

Clustering 2 – K-means clustering

- We clusters the principal components using the K-means clustering

- we select the number of clusters as four for a consistency comparison

Principal Component Analysis

- We cannot find distinctive elbow from the scree plot.

- We can explain more than 84% of variance using PC1 to 5.

- Thus, we select the number of principal component to five based on the cumulative explained variance.

# of PCs
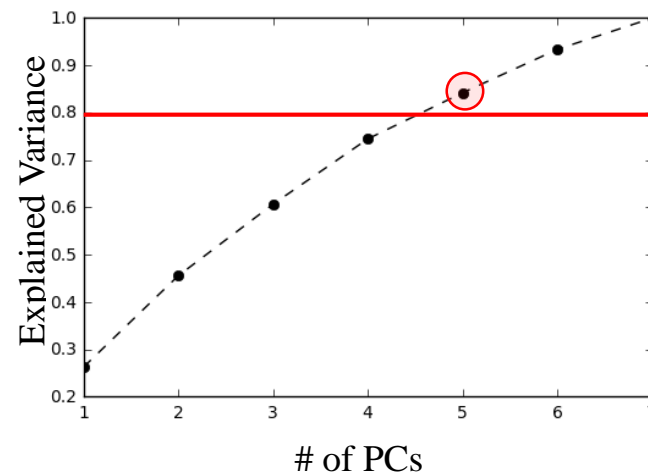
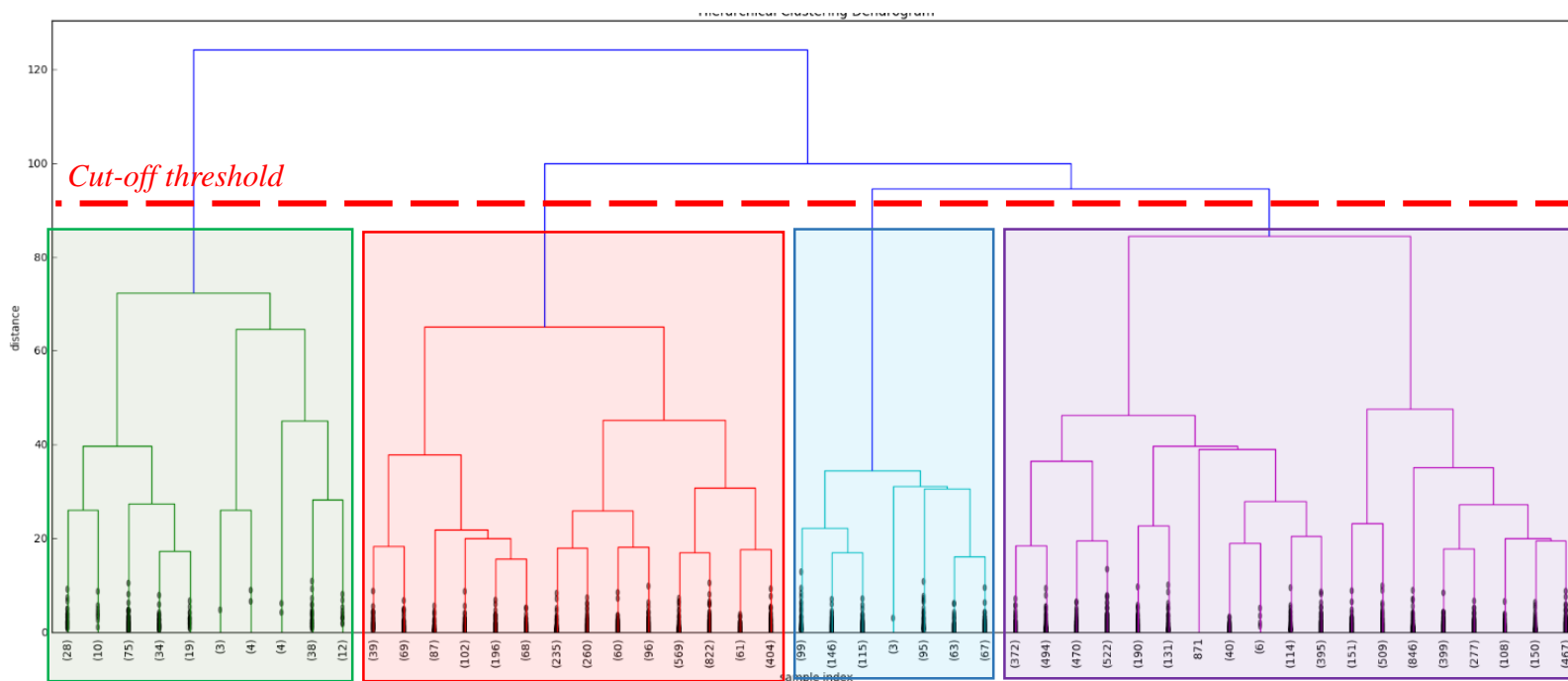|	Factor Loadings							Lambda	Explained variance	Cumulative Explained variance
	STD	LTD	ROE	RNDR	ADMCR	NIR	SG			
PC1	-0.063	0.119	-0.513	0.355	0.470	-0.583	-0.177	1.839	0.263	0.263
PC2	-0.601	-0.273	0.365	0.502	0.330	0.213	0.156	1.351	0.193	0.456
PC3	0.418	0.056	0.105	0.174	0.133	-0.173	0.856	1.058	0.151	0.607
PC4	-0.278	0.942	0.159	0.025	-0.015	0.067	0.065	0.966	0.138	0.745
PC5	0.035	-0.003	0.271	-0.620	0.734	-0.016	-0.041	0.682	0.097	0.842
PC6	-0.615	-0.127	-0.351	-0.454	-0.205	-0.224	0.430	0.642	0.092	0.934
PC7	-0.062	-0.063	0.609	-0.025	-0.266	-0.728	-0.141	0.462	0.066	1.000

Principal Component Analysis

- The scatter plot for the PC 1 and 2 indicates that we can group the factors as three main features;

  - Profitability factors (ROE, Net-income ratio, and Sales growth)

  - Cost-related factors (RND cost ratio and administrative cost ratio)

  - Debt-related factors (Short-term debt and Long-term Debt)
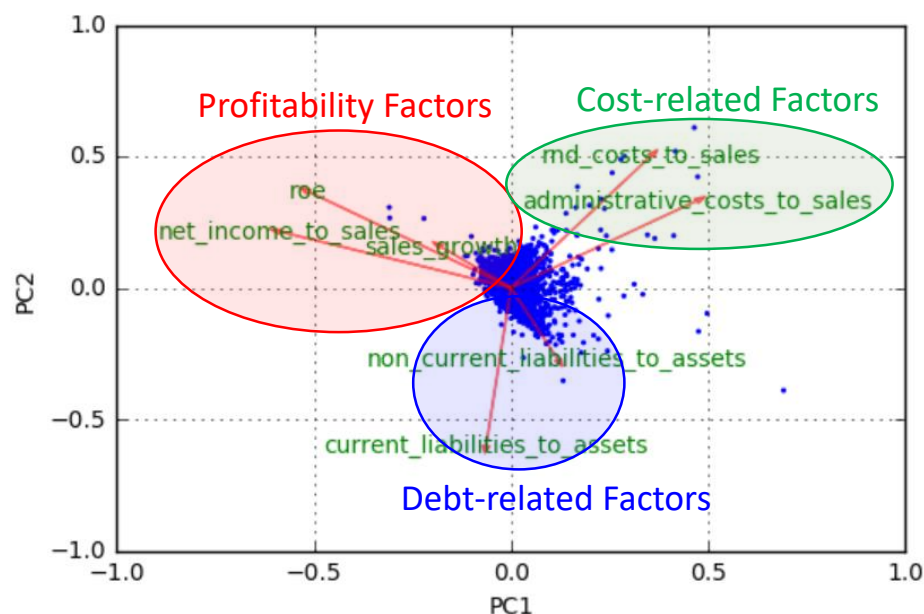
Hierarchical clustering

- Hierarchical clustering is conducted based on Ward's minimum variance method

- We select four clusters for the business interpretation with the cut-off threshold from the dendrogram analysis
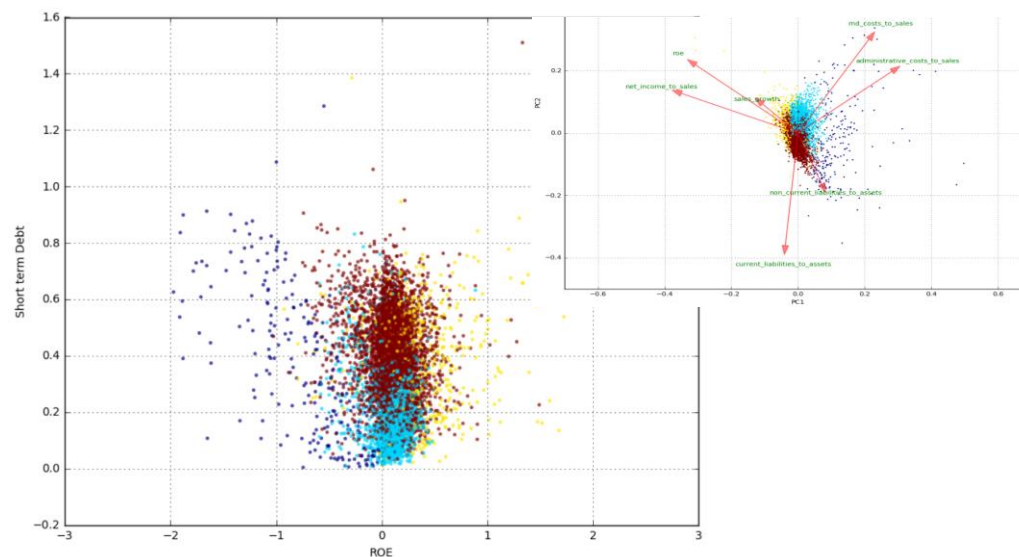
Hierarchical clustering

- Cluster 1 (Long-term-oriented Firms) - High RND cost and Low current profitability (Future investment)

- Cluster 2 (Stable Firms) -Low debt(STD and LTD) and Relatively High ROE

- Cluster 3 (High Growth Firms) - High Sales growth and Profitability (ROE and Net-income)

- Cluster 4 (High Risky Firms) - High Debt and low Cost (Improper debt handing)

| Cluster	1	2	3	4
STD	0.360	0.206	0.390	0.383
LTD	0.174	0.068	0.097	0.147
ROE	-0.720	0.095	0.335	0.094
RNDR	0.134	0.032	0.023	0.013
ADMCR	0.173	0.074	0.047	0.034
NIR	-0.906	0.074	0.131	0.036
SG	0.163	0.140	0.972	0.211

*Note Highest value, lowest value

K-means clustering

- Cluster 1 (Stable Firms) -Low debt(STD and LTD) and Relatively High ROE

- Cluster 2 (High Growth Firms) - High Sales growth and Profitability (ROE and Net-income)

- Cluster 3 (Long-term-oriented Firms) – High RND cost and Low current profitability (Future investment)

- Cluster 4 (High Risky Firms) - High Debt and low Cost (Improper debt handing)

Clusters	1	2	3	4
STD	0.207	0.385	0.250	0.426
LTD	0.076	0.118	0.155	0.159
ROE	0.125	0.298	-0.585	0.026
RNDR	0.030	0.023	0.186	0.009
ADMCR	0.062	0.044	0.244	0.035
NIR	0.091	0.102	-1.182	0.001
SG	0.154	0.791	0.199	0.164

*Note Highest value, lowest value
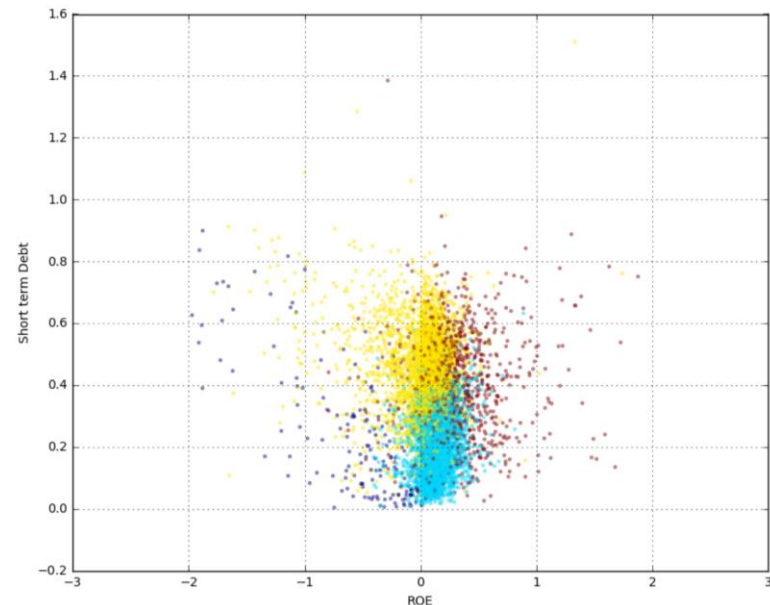
Discussion and Conclusion

✓ **There are four kinds of firms in terms of the Debt-handing.**

- Firms are well clustered into "Stable Firms", "High Growth Firms", "Long-term-oriented Firms", and "High Risky Firms".

- In general, firms with high debt has tendency of low profitability, but **there are firms which efficiently use leverage to improve profitability**.

- The clustering result shows how firms use leverage properly to improve their profitability and sales growth rate.

✓ **The clustering result is helpful for investors to easily understand the nature and characteristics of firms.**

- **High Growth Firms are much attractive to investors** because they use their leverage properly.

- Investors need to consider firms' future growth when they are willing to invest the Long-term-oriented firms.

Limitation and Action Plan

Limitation of this study

- Firms will be differently clustered other than the debt-handling ability such as sensitivity to economic condition.

- Financial statements are updated quarterly, but this study only use yearly data.

    - The time span between the announcements is more narrow than yearly updated data

    - We can improve the accuracy of the model by considering some factors that include time series information

Further studies

    - We can consider macro-economic effects.

    - We can predict firm productivity on quarterly basis.

    - We can check the relationship between stock price and the future firm productivity.

References

1. Azhagaiah, R., & Gavoury, C. (2011). The impact of capital structure on profitability with special reference to IT industry in India vs. Domestic products. *Managing Global Transitions*, *9*(4), 371.

2. Gill, A., Biger, N., & Mathur, N. (2011). The effect of capital structure on profitability: Evidence from the United States. International Journal of Management, 28(4), 3.

3. Nasimi, A. N. (2016). Effect of Capital Structure on Firm Profitability (An Empirical Evidence from London, UK). *Global Journal of Management And Business Research*.

4. Abor, J. (2005). The effect of capital structure on profitability: an empirical analysis of listed firms in Ghana. *The journal of risk finance*, *6*(5), 438-445.

5. Petersen, M. A. (2009). Estimating standard errors in finance panel data sets: Comparing approaches. The *Review of Financial Studies*, 22(1), 435-480.

6. Rajan, R. G., & Zingales, L. (1995). What do we know about capital structure? Some evidence from international data. *The journal of Finance*, *50*(5), 1421-1460.

7. Wald, J. K. (1999). How firm characteristics affect capital structure: an international comparison. *Journal of Financial research*, *22*(2), 161-187.

8. Berger, A. N., & Di Patti, E. B. (2006). Capital structure and firm performance: A new approach to testing agency theory and an application to the banking industry. *Journal of Banking & Finance*, *30*(4), 1065-1102.

9. 백원선, 조현우, & 호용익. (2004). 자기자본이익률 구성요소와 미래수익성 및 가치관련성. *경영학연구*, *33*(5), 1329-1354.

10. 윤봉한. (2005). 한국 상장기업의 자본구조 결정요인에 대한 장기분석: 정태적 절충모델과 자본조달순위모델간의 비교. *경영학연구*, 34(4), 973-1000

Thank you

Regression - Literature review (Previous study)

- *The effect of capital structure on profitability: Evidence from the United States. Gill, A., Biger, N., & Mathur, N. (2011).*

Table 1: Proxy variables definition and predicted relationship

Proxy Variables	Definitions	Predicted sign
Short-Term Debt (SDA)	Short-term debt divided by the total assets	+/-
Long-Term Debt (LDA)	Long-term debt divided by the total assets	+/-
Total Debt (DA)	Total debt divided by the total assets	+/-
Firm Size (SIZE)	Natural Logarithm of Firm's Sales, lagged one year period	+/-
Sales Growth (SG)	Current year's Sales minus previous year's sales divided by previous year's sales	+/-
Industry	Firm is assigned value one if firm is a manufacturing firm and zero otherwise	+/-

$$\text{Profitability}_{i,t} = b_0 + b_1*SDA + b_2*SIZE + b_3*SG + \mu_{i,t}$$

Profitability $_{i,t}$ - profitability for firm i between 2005-2007 measured by ROE.

SDA $_{i,t}$ - short-term debt/total assets for firm i in time t.

LDA $_{i,t}$ - long-term debt/total assets for firm i in time t.

DA $_{i,t}$ - total debt/total assets for firm i in time t.

SIZE $_{i,t}$ - Natural logarithm of firm's sales, lagged one year period.

SG $_{i,t}$ - Current year's sales minus previous year's sales divided by previous year's sales.

$\mu_{i,t}$ = the error term.

Service Industry: Short-Term Debt to Total Assets

$[R^2 = 0.050;$ SEE = 0.306; F = 2.212]

Regression Equation (A): ROE = 0.378 + 0.428 SDA - 0.021 SG - 0.013 SIZE

	Unstandardized Coefficients		Standardized Coefficients c	t	Sig.	Collinearity Statistics	
	B	Std. Error	Beta			Tolerance	VIF
(Constant)	0.378	0.225		1.678	0.096		
SDA	0.428	0.175	0.213	2.451	0.016	0.995	1.005
SG	-0.021	0.192	-0.010	-0.109	0.914	0.982	1.018
SIZE	-0.013	0.014	-0.081	-0.923	0.358	0.982	1.018

Service Industry: Long-Term Debt to Total Assets

$[R^2 = 0.011;$ SEE = 0.312; F = 0.448]

Regression Equation (B): ROE = 0.418 + 0.114 LDA - 0.001 SG - 0.012 SIZE

	Unstandardized Coefficients		Standardized Coefficients c	t	Sig.	Collinearity Statistics	
	B	Std. Error	Beta			Tolerance	VIF
(Constant)	0.418	0.230		1.817	.072		
LDA	0.114	0.133	0.076	0.858	.392	0.998	1.002
SG	-0.001	0.196	0.000	-0.005	.996	0.984	1.016
SIZE	-0.012	0.015	-0.072	-0.801	.425	0.984	1.016