

# SAS<sup>®</sup> GLOBAL FORUM 2018

USERS PROGRAM

## Analysis of Profitability-Bank Systems in South Korea Using SAS Base

April 8 - 11 | Denver, CO

**#SASGF**

# Analysis of Profitability-Bank Systems in South Korea Using SAS Base

Jinwoo Cho, Dongwoo Kim

Sungkyunkwan University

## ABSTRACT

Banks have the most important role in the finance market. The major role of banks is to link fund of persons and firms. In other words, banks are lubricants to allow domestic economy to grow smoothly in that they maximize the fluidity of finance in each person or company. However, in the aspect of customers who want to deposit their own money, selecting banks must be deliberated. Customers should be conservative in their choice since the credibility of banks is directly connected to their assets. Additionally, it means that customers should consider profitability of banks when saving their money. In this regard, our project mainly focused on the index which can discriminate the earning with SAS base. Above all, our project's object also lies on the possibility which you can analyze the system approximately with only SAS base. Actually, it is not an easy way to make a project with bank data, so that you might be looking forward to using more complicated or sophisticated SAS programs. But, we can suggest a significant solution using SAS base.

## METHODS

### • Data description

The data is collected from 'Financial Statistical Information System (FISIS)', which can access from 'Financial Supervisory Service (FSS)' in South Korea. The data table contains nine variables including return on equity, proportion of allowance for bad debts, selling and administrative expenses, and proportion of sub-standard. 97 banks are included in this data, that is the number of observation is 97.

### • Objectives

To analyze the structure of bank system, Return on Equity(ROE) is used as the main index. ROE is the main revenue of one's bank, so that the profitability structure is reflected to ROE. Through this response variable, the main objective is to figure out the structures of 97 banks with discriminate analysis method. To concrete, two division were generated by ROE order, that is the first group has more profit and the second is not. Given two groups, discriminant analysis help us to select influence variables among 9 variables. Therefore, chosen variables would be crucial to explain the profitability of bank in Korea.

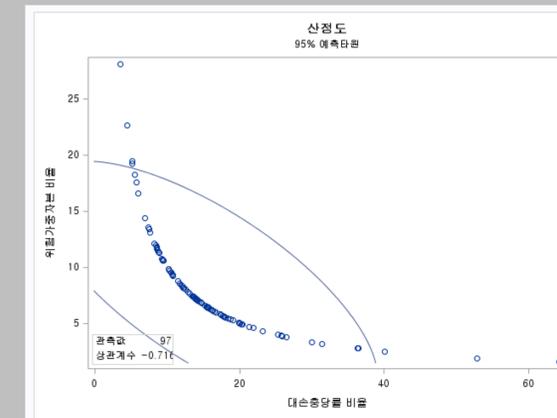
### • Methods

At the beginning, our project fulfilled correlation analysis to select significant variables. Then, we explored the data by principal analysis and factor analysis. Finally, by utilizing new variables resulted from the previous analyses, we made some clusters, and finally, conducted discrimination analysis related to profitability of banks.

## RESULTS

### • Correlation analysis

Before main analyses, data exploration is important step. If ignore this step, it is possible to fail to analyze the real structure of data. Similarly, before factor analysis or principal component analysis, correlation has the role of preprocessing. Even though factor or principal component analysis have significant effect to control high correlation structure among variables, it is useful to have another filter to eliminate high correlated variables. Unlike other statistic package, SAS base gives several configurations to explore the correlation structure. The given figure is the one of them explaining correlation between proportion of risk capital and proportion of bad debts



Form of the ellipse shows brief correlation, so that you can check the high correlation structure without statistics. Conclusively, from this step, we removed 3 variables including proportion of bad debts, loans and own capital conforming to BIS.

### • Principal Component Analysis and Factor Analysis

Principal Component Analysis (PCA) and Factor Analysis (FA) have similar proposition, but there is a serious contrast point in the intention of analyses. PCA concentrates on shrinkage of dimension, so that a set assumption is variables. On the other hand, FA focuses on a latent variable which is contributed to given ordinary variables. Thus, FA and PCA has opposite direction in planning of analysis. First of all, we conducted PCA to figure out how interpret principal components and their significance. With this result, assuming 3 factors that represent principal components, FA has been fulfilled to check whether same interpretation could be realized.

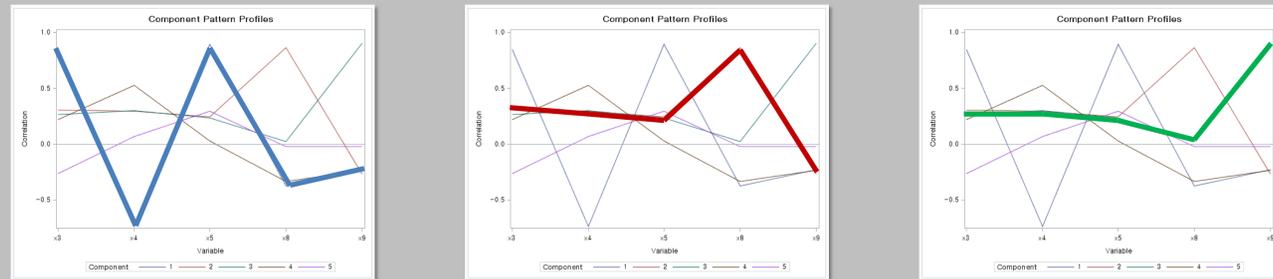
# Analysis of Profitability-Bank Systems in South Korea Using SAS Base

Jinwoo Cho, Dongwoo Kim

Sungkyunkwan University

## RESULTS CONTINUED

- SAS base solution gives useful figures to have convenient interpretation in PCA.

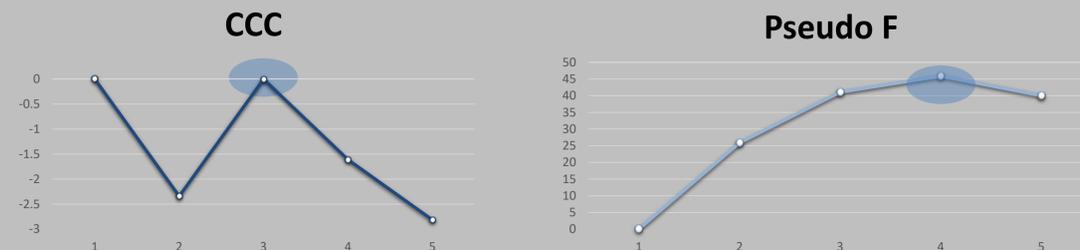


The first blue principal component has high coefficient on deposit and low coefficient on loan, so that we can interpret the first component as scale or quality of asset. The second red line has a significant coefficient on proportion of risk weight asset. Thus, the second component means risk of bank. The last only shows meaningful coefficient on fluidity. So the last component could be interpreted as fluidity of bank asset.

Factor analysis can do similarly. There are 3 methods to set the factor analysis. Among them, we choose the principal component methods. Also, you can rotate principal component to maximize the variance explanation, which called varimax rotation. When you do the varimax in this case, however, loadings of factor is not commensurate with the PCA components. Therefore, we choose the result of FA without rotation.

## Cluster Analysis

Clustering has two approach; hierarchical and non-hierarchical method. In this study, we only used non-hierarchical method, since the hierarchical method has difficulty in choosing the number of clusters and its visualization. There are several solutions to select the number of clusters. SAS base gives two approach Cubic Clustering Criterion (CCC) and Pseudo F statistic. If both statistics are high, then it means that number of clusters is significant. The graph is following



## RESULTS CONTINUED

- Finally, we can conclude that the suitable number of clusters is three. Then we can analyze each cluster with their cluster means in SAS base. Following table shows those means as a table

Cluster Means			
Cluster	Factor1	Factor2	Factor3
1	-0.354380682	-0.100112559	1.629347262
2	-0.399943977	0.114876135	-0.418528193
3	1.976305383	-0.353134948	-0.057068694

You can analyze the cluster, for example, cluster 1 has negative values on factor 1 and factor 2. In contrasts, the cluster mean of factor 3 is positive. Form this interpretation, cluster 1 contains banks that has little scale, but low risk and high fluidity. Moreover, when aligning the observation within cluster 1, the characteristic of cluster 1 is obviously shown as a group of low risk saving banks. Likewise, cluster 2 involves comparably high risk saving banks. And cluster 3 includes 16 major banks in South Korea.

## Discriminant Analysis

Unlike cluster analysis, discriminant analysis has response variable, which is called supervising analysis, otherwise unsupervising analysis including clustering. The final objective of our study is to find out a meaningful factor relevant to profitability or ROE. So, before beginning, we divide two group by ROE index's median. That is, one group has higher profitability than the other.

With that information, discriminant analysis has done for each cluster that already divided, and shows the following result for the cluster 1, low risk saving banks. In this case, factor 2 is influential to classify banks by profitability.

Total-Sample Standardized Canonical Coefficients	
Variable	Can1
Factor1	-0.279057578
Factor2	1.070091580
Factor3	-0.420848994

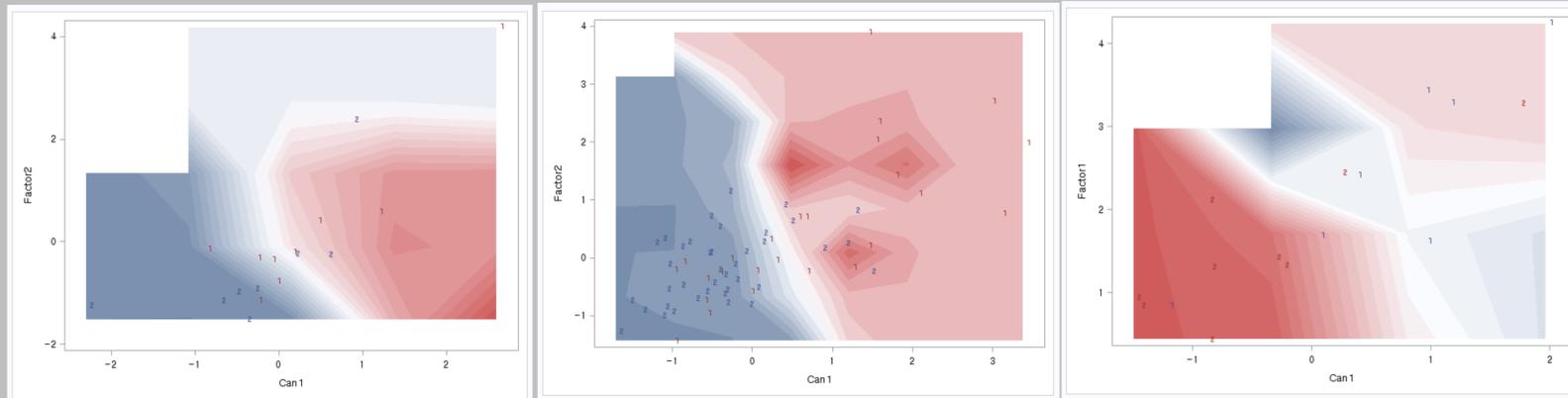
# Analysis of Profitability-Bank Systems in South Korea Using SAS Base

Jinwoo Cho, Dongwoo Kim

Sungkyunkwan University

## RESULTS CONTINUED

- Moreover, with proc template, SAS gives a beautiful plot for a discriminant analysis. These three figures explain each cluster 1, 2 and 3 from left. Vertical axis are the crucial factor of each cluster. In cluster 1 and 2, factor 2 which represents risk of asset occupies a important role to divide profitability. However, cluster 3 including 16 major banks is not the case of factor 2. Instead, factor 3 is crucial to classify the profitability. Conclusively, in the major banks which have comparably high quality asset, scale of asset is important to measure its profitability.



## CONCLUSIONS

To summarize, first we conduct correlation analysis. Correlation analysis is a sort of preprocessing before main analysis. Through this first step, we can remove the duplicate variables and make more accurate statistical result. Next, PCA and FA were fulfilled. Frankly, the asset index is not easy to access to out-field people. To reduce this difficulty, we did PCA and FA and shrank those remain variables. Lastly, for more precise result, clusters were created through the non-hierarchical approach. Then, discriminant analysis was conducted for each cluster.

Until now, our study shows the several steps to find out the structure of bank market. Although there is no advanced program, only SAS base program was used, the conclusion and the interpretation of this analysis was quite realistic. Thus, the objective that we set has completed through this several statistical methods.

## REFERENCES

- Financial Supervisory Service in South Korea, <http://english.fss.or.kr/fss/eng/main.jsp>
- C.S.Hong, SAS and Statistical data analysis 4<sup>th</sup> edition, Tamjin



# SAS<sup>®</sup> GLOBAL FORUM 2018

April 8 - 11 | Denver, CO  
Colorado Convention Center

#SASGF

```

%macro rename;
data final; set data;
rename F1=bank;
label bank='À°Çà, í';
%do i=1 %to 9;
rename _col&i=x&i;
%end;
run;
%mend;
%rename;
data final; set final;
label bank='À°Çà, í';
run;
data final; set final;
if bank= '¿;½°ñ¾ÆÀÌÀúÃàÀ°Çà' then delete;
if bank= '¿;½°ñ¾ÆÀÌ2ÀúÃàÀ°Çà' then delete;
if bank= '¿;½°ñ¾ÆÀÌ3ÀúÃàÀ°Çà' then delete;
if bank= '¿;½°ñ¾ÆÀÌ4ÀúÃàÀ°Çà' then delete;
if bank= 'À°ÀúÃàÀ°Çà' then delete;
if bank= 'ÇØ¼ÖÀúÃàÀ°Çà' then delete;
if bank= 'ÇÑ½ÀúÃàÀ°Çà' then delete;
if bank= '¿;®ÀúÃàÀ°Çà' then delete;
if x1<0 then class=1;
else class=2;
run;

proc corr data=final plot=all;
var x1-x9;
run;

proc sort data=final;
by x8;
run;

proc princomp data=final plot=all std;
var x3 x4 x5 x8 x9;
run;

proc factor data=final plot=pathdiagram fuzz=0.5 nfactors=3 out=factor;
var x3 x4 x5 x8 x9;
run;
proc factor data=final rotate=varimax plot=pathdiagram fuzz=0.5 nfactors=3 out=factor;
var x3 x4 x5 x8 x9;
run;

proc cluster data=factor method=com ccc;
var factor1-factor3;
id bank;
run;
proc tree;run;

proc standard data=factor;
var factor1-factor3;

```

```

run;

proc fastclus data=factor maxclusters=3 list out=clust;
var factor1-factor3;
id bank;
run;

data clus1 clus2 clus3; set clust;
if cluster=1 then output clus1;
else if cluster=2 then output clus2;
else output clus3;
run;

proc means data=clus3 median;
var x1;
run;

data clus3;set clus3;
if x1 < 4.34 then class=1;
else class=2;
run;

proc discrim data=clus3 out=can3 can crossvalidate ;
var factor1-factor3;
class class;
run;
proc template;
define statgraph classify;
begingraph;
layout overlay;
contourplotparm x=Can1 y=factor1 z=_into_ / contourtype=fill
nhint = 30 gridded = false;
scatterplot x=Can1 y=factor1 / group=class includemissinggroup=false
markercharactergroup = class;
endlayout;
endgraph;
end;
run;
proc sgrender data = can3 template = classify;
run;

proc discrim data=clus2 out=can2 can pool=test;
var factor1-factor3;
class class;
run;
data can2;set can2;
into=-_into_;
run;
proc template;
define statgraph classify;
begingraph;
layout overlay;
contourplotparm x=Can1 y=factor2 z=into / contourtype=fill
nhint = 30 gridded = false;

```

```

scatterplot x=Can1 y=factor2 / group=class includemissinggroup=false
            markercharactergroup = class;;
endlayout;
endgraph;
end;
run;
proc sgrender data = can2 template = classify;
run;

```

```

proc discrim data=clus1 out=can1 can crossvalidate;
var factor1-factor3;
class class;
run;
data can1;set can1;
into=-_into_;
run;
proc template;
define statgraph classify;
begingraph;
layout overlay;
contourplotparm x=Can1 y=factor2 z=into / contourtype=fill
                nhint = 30 gridded = false;
scatterplot x=Can1 y=factor2 / group=class includemissinggroup=false
            markercharactergroup = class;;
endlayout;
endgraph;
end;
run;
proc sgrender data = can1 template = classify;
run;

```