Submitted By: Murli Manohar Singh (msingh40)

Project Report for IDS 462

**TABLE OF CONTENTS**

[Introduction 2](#_Toc445861109)

Factor1: Households Items and Beverages 2

Factor2: Raw Materials and Energy items 3

Factor3: Meat 4

Factor4: Agriculture and Energy Items 4

[Technical Appendix 5](#_Toc445861110)

Programs………………………………….……………………………………………………………………..7

[References 9](#_Toc445861111)

**INTRODUCTION**

Analysis in this report is done on a commodity data to identify the themes (i.e. factors) in the data. We used Factor Analysis technique to

* Reduce the number of variables
* To analyze and find the relationship structure among the variables
* To describe [variability](https://en.wikipedia.org/wiki/Variance) among observed, correlated [variables](https://en.wikipedia.org/wiki/Variable_%28mathematics%29) in terms of a potentially lower number of unobserved variables

Per our analyses, we have identified four main factors/themes which are Households Items and Beverages, Raw Materials and Energy items, Meat and Agriculture and Energy Items.

As we will explain below, using Factor Analysis technique we are now able to show how strongly our dataset variables are associated with each of the four factors. These associations with the factors can vary from -100% to 100%. Numbers having larger absolute values indicate a more strong association with that particular factor.

After analyzing dataset and performing Factor analysis, we come up with four below mentioned factors with highly correlated variables.

**Factor1: Households Items and Beverages**

Below are the commodities with their percentages that come under household’s items and beverages which we use for our day today work. Beverages include coffee and cocoa beans. In addition, this factor contains items such as fish meal, Soybean Meal, Maize and Wheat which we use for eating purposes. Moreover, the factor also contains oils we use for day to day cooking purposes. From the description we have identified that the data contains both import and export items across households items.

|  |  |  |
| --- | --- | --- |
| COMMODITIES | DESCRIPTION (HOUSEHOLDS ITEMS AND BEVERAGES) | % |
| PCOFFOTM\_USD | Coffee, Other Mild Arabicas, International Coffee Organization New York cash price, ex-dock New York, US cents per pound | 91.6 |
| PCOFFROB\_USD | Coffee, Robusta, International Coffee Organization New York cash price, ex-dock New York, US cents per pound | 88.5 |
| PCOTTIND\_USD | Cotton, Cotton Outlook 'A Index', Middling 1-3/32 inch staple, CIF Liverpool, US cents per pound | 77.3 |
| PFISH\_USD | Fishmeal, Peru Fish meal/pellets 65% protein, CIF, US$ per metric ton | 77.1 |
| PMAIZMT\_USD | Maize (corn), U.S. No.2 Yellow, FOB Gulf of Mexico, U.S. price, US$ per metric ton | 93.6 |
| PSMEA\_USD | Soybean Meal, Chicago Soybean Meal Futures (first contract forward) Minimum 48 percent protein, US$ per metric ton | 80.6 |
| PSUNO\_USD | Sunflower oil, Sunflower Oil, US export price from Gulf of Mexico, US$ per metric ton | 87.1 |
| PWHEAMT\_USD | Wheat, No.1 Hard Red Winter, ordinary protein, FOB Gulf of Mexico, US$ per metric ton | 79.7 |
| PCOCO\_USD | Cocoa beans, International Cocoa Organization cash price, CIF US and European ports, US$ per metric ton | 75.5 |

From the above mentioned correlation matrix or factor loadings, we can see that factor1 is able to explain 91.6% variability in PCOFFOTM\_USD; 88.5% variability in PCOFFROB\_USD; 77.3% variability in PCOTTIND\_USD; 77.1% variability in PFISH\_USD; 93.6% variability in PMAIZMT\_USD; 80.6% variability in PSMEA\_USD; 87.1% variability in PSUNO\_USD; 79.7% variability in PWHEAMT\_USD; 75.5% variability in PCOCO\_USD.

So, it is very clear that factor1 is able to explain all the 9 variables quiet strongly.

**Factor2: Raw Materials and Energy items**

Below are the commodities with their percentages that come under Raw Materials and Energy items which is used by industries to produce products and to create energy such as electricity. It is very clear from the table that Lead and Uranium are very precious mineral used as Raw material and energy items. Hard logs can is the raw material to produce bigger products. Natural gas is used to generate electricity and used by various industries. From the description we have identified that the import and export spread pretty well in the global market.

|  |  |  |
| --- | --- | --- |
| COMMODITIES | DESCRIPTION (RAW MATERIALS AND ENERGY ITEMS) | % |
| PLEAD\_USD | Lead, 99.97% pure, LME spot price, CIF European Ports, US$ per metric ton | 89.6 |
| PLOGSK\_USD | Hard Logs, Best quality Malaysian meranti, import price Japan, US$ per cubic meter | 76.9 |
| PNGASEU\_USD | Natural Gas, Russian Natural Gas border price in Germany, US$ per thousands of cubic meters of gas | 80.4 |
| PNGASJP\_USD | Natural Gas, Indonesian Liquefied Natural Gas in Japan, US$ per cubic meter of liquid | 69.2 |
| PURAN\_USD | Uranium, NUEXCO, Restricted Price, Nuexco exchange spot, US$ per pound | 88.2 |

From the above mentioned correlation matrix or factor loadings, we can see that factor2 is able to explain 89.6% variability in PLEAD\_USD; 76.9% variability in PLOGSK\_USD; 80.4% variability in PNGASEU\_USD; 69.2% variability in PNGASJP\_USD; 88.2% variability in PURAN\_USD.

So, it is very clear that factor2 is able to explain all the 5 variables quiet strongly.

**Factor3: Meat**

Below are the commodities with their percentages that come under meat. Lamb carcass and other parts used for eating purposes. It is heavily exported and imported across the world. In addition, from the description we have identified that exports holds the majority in meat in global market.

|  |  |  |
| --- | --- | --- |
| COMMODITIES | DESCRIPTION (MEAT) | % |
| PLAMB\_USD | Lamb, frozen carcass Smithfield London, US cents per pound | 87.6 |

From the above mentioned correlation matrix or factor loadings, we can see that factor3 is only contains one variable and is able to explain 87.6% variability in PLAMB\_USD.

So, it is very clear that factor3 is able to explain PLAMB\_USD variable very strongly.

**Factor4: Agriculture and Energy Items**

Below are the commodities with their percentages that come under Agriculture and Energy Items. Soft logs have a wide range of uses such as it can be used in agriculture, industries and in some countries to produce coal and hence satisfy the energy needs. Natural Gas can be used in industries and for energy generation purposes. From the description we have identified that exports holds the majority in Agriculture and Energy Items in global market.

|  |  |  |
| --- | --- | --- |
| COMMODITIES | DESCRIPTION (AGRICULTURE AND ENERGY ITEMS) | % |
| PLOGORE\_USD | Soft Logs, Average Export price from the U.S. for Douglas Fir, US$ per cubic meter | 71.1 |
| PNGASUS\_USD | Natural Gas, Natural Gas spot price at the Henry Hub terminal in Louisiana, US$ per thousands of cubic meters of gas | 52.3 |

From the above mentioned correlation matrix or factor loadings, we can see that factor4 is able to explain 71.1% variability in PLOGORE\_USD; 52.3% variability in PNGASUS\_USD. One more thins to mention here is that factor4 is negatively correlated to PLOGORE\_USD which means if factor4 increases PLOGORE\_USD decreases but this relationship is strong.

So, it is very clear that factor4 is able to explain both the variables quiet strongly.

**TECHNICAL APPENDIX**

The analysis was based on a set of 133 observations and 20 variables that were collected from the global commodity market prices.

Below are the steps/ technical decisions taken to perform the Factor Analysis on the dataset:

* Before performing Factor analysis, we inspected the data set variables. By this initial analysis following things can be explained:
  + There are no missing data values
  + All the variables are numeric or quantitative
  + Data needs standardization as variables are in the different measures such as few are indexes, few are prices, and few has different magnitude of the values.
* Then we inspected the correlation between the variables in the data set. As per the analysis the data is pretty correlated and that’s what it should logically be as the dataset represents the commodity data as a whole.

* Initially after standardizing the data, we executed simple factor analysis without any rotation and with equal importance given to each variable in the data set. The number of factors was initially selected by retaining only those factors with an eigenvalue greater than one. This means that we selected the factors those explained at least one complete variable (eigenvalue greater than 1).
* After inspecting the factors and eigenvalues we decided to go with 4 factors as the 4th factor has Eigen value just below 1.0 value i.e. 0.98 and by taking this factor into our analysis the variability explained on the whole data set got better i.e. 85% and was increased by 6% over when we took 3 factors.
* We also observed that removing indexes variables from the dataset will be more productive as keeping only commodity prices makes sense here the indexes are just combined representation of the prices and performance of the commodities in the market. So, all our further analysis has been performed by removing the indexes.
* We performed Varimax rotation with and without prior SMC and with 4 factors. By analyzing the results we can say that the factors were not that impressive lot of overlap of variables those should have been in other factors as per their logical functioning/use.
* Next we performed Promax rotation with and without Prior SMC. By analyzing the data we focused our analysis more on Promax rotation with Prior SMC because it made a perfect sense and the variability explained was also impressive i.e. 83%.
* Among the various methods listed above Promax rotation with Prior SMC outperformed the others because of the following reasons:

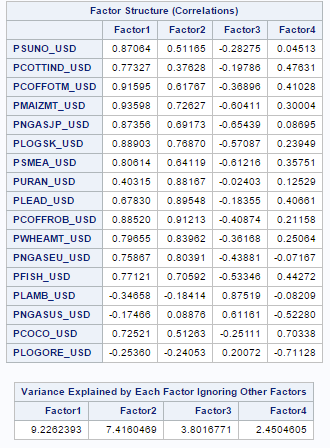


Figure 1. Result of Promax rotation with Prior SMC

* + It explained 83% variability in the dataset.
  + It is very unrealistic to assume that the variables will be un-related. As we know that it is after all a commodity data set and all the variables are connected to each other some way or the other. We can infer the correlation from our correlation analysis which we performed before applying Factor analysis. So, it makes more sense to apply Promax rotation which considers oblique relation (means variables are connected with each other) into account.
  + SMC method takes into account the importance of variables individually as compared to no SMC which unrealistically consider every variable equally important.

**PROGRAMS**

**/\* Read data csv and load it into sas dataset and name it projfact\*/**

proc import

out=projfact

datafile="/folders/myfolders/project1\_data\_16.csv"

dbms= csv replace;

getnames=yes;

datarow=2;

run;

**/\* Analyse contents for the data set: all are quantitative values so no need to drop any variable\*/**

proc contents data=projfact;

title 'content analyzation for the dataset';

run;

**/\* Examine if we have a correlated variables in the dataset\*/**

proc corr data=projfact;

title 'initial varables correlation matrix';

run;

**/\* We need standardization in the dataset values as lot many values are weighed differently \*/**

proc standard data=projfact out=Stdprojfact(drop= PALLFNF\_Index PNFUEL\_Index PFANDB\_Index) mean=0 std=1;

title 'Standardisation of the dataset';

run;

**/\* Analyze data set after standardisation \*/**

proc means data=Stdprojfact;

title 'Summary Statistics for the standardised dataset';

run;

**/\* Factor analysis without any rotation to get the top factors\*/**

proc factor data=Stdprojfact scree reorder;

title 'Initial Factor analysis without any rotation';

run;

**/\* Using Varimax rotation without Prior SMC\*/**

proc factor data=Stdprojfact rotate=varimax nfactors=4 scree reorder out=scoreVar;

title 'Factor analysis using varimax with 4 best factors';

run;

**/\* Using Varimax rotation with Prior SMC\*/**

proc factor data=Stdprojfact rotate=varimax nfactors=4 plot scree reorder;

title 'Factor analysis using varimax with 4 best factors with prior smc';

priors smc;

run;

**/\* Using Promax rotation without Prior SMC\*/**

proc factor data=Stdprojfact rotate=promax nfactors=4 scree reorder out=scorePro;

title 'Factor analysis using promax with 4 best factors';

run;

**/\* Using Promax rotation with Prior SMC\*/**

proc factor data=Stdprojfact rotate=promax nfactors=4 scree reorder;

title 'Factor analysis using promax with 4 best factors with prior smc';

priors smc;

run;

**REFERENCES**

Data Source:

<http://www.imf.org/external/np/res/commod/index.aspx>

Technical help:

<https://support.sas.com/documentation/cdl/en/statug/63347/HTML/default/viewer.htm#factor_toc.htm>