Factor Analysis Report

Global Commodity Market Data

Kousalya Dwarapudi

CONTENTS

Problem Description

The given dataset comprises of various primary commodities prices traded over the global markets. The analysis of this data is mainly done to categorize the variables in to various themes or factors. The problem with this dataset is mainly due to heavy correlation amongst the variables.

The analysis technique has been used to

- Understand the nature and type of the variables
- Study the distribution and relationship amongst the variables
- Analyze the various themes they can be classified into
- Reduce the number of variables by factor analysis

Executive Summary

According to my analysis, I have identified 4 main themes or factors to group the similar variables under.

The four main categories are:

1. Oils, Grains, Beverages– Household Items
2. Metals
3. Livestock
4. Raw Materials

I have clubbed the variables that have larger coefficients to be more associated to that factor than the others. As you observe below, through this factor analysis technique we are now able to show how strongly our dataset variables are associated with each of the three factors. These associations with the factors can vary from -100% to 100%. Numbers having larger absolute values indicate a stronger association with that factor.

1. **Factor 1: Household items**

   Below are the commodities with their percentages that come under heavy duty goods. From the description we have identified that the import and export is evenly spread across heavy duty items.

Commodities	Description	Percent
PROIL_USD	Rapeseed oil, crude, fob Rotterdam, US$ per metric ton	70%
PBEVE_Index	Beverage Price Index, 2005 = 100, includes Coffee, Tea, and Cocoa	82%
PFISH_USD	Fishmeal, Peru Fish meal/pellets 65% protein, CIF, US$ per metric ton	70%
PIORECR_USD	China import Iron Ore Fines 62% FE spot (CFR Tianjin port), US dollars per metric ton	82%
PCOCO_USD	Cocoa beans, International Cocoa Organization cash price, CIF US and European ports, US$ per metric ton	73%
POILAPSP_USD	Crude Oil (petroleum), Price index, 2005 = 100, simple average of three spot prices; Dated Brent, West Texas Intermediate, and the Dubai Fateh	65%
POILDUB_USD	Oil; Dubai, medium, Fateh 32 API, fob Dubai Crude Oil (petroleum), Dubai Fateh 32 API, US$ per barrel	67%
PSMEA_USD	Soybean Meal, Chicago Soybean Meal Futures (first contract forward) Minimum 48 percent protein, US$ per metric ton	74%
PWHEAMT_USD	Wheat, No.1 Hard Red Winter, ordinary protein, Kansas City, US$ per metric ton	69%
PCOFFOTM_USD	Coffee, Other Mild Arabicas, International Coffee Organization New York cash price, ex-dock New York, US cents per pound	82%
PCOTTIND_USD	Cotton, Cotton Outlook 'A Index', Middling 1-3/32 inch staple, CIF Liverpool, US cents per pound	74%

2. **Factor 2: Metals**

   Below are the commodities with their percentages that come under Metals and Ores. From the description we have identified that the export dominates imports in metals and their ores in global market

Commodities	Description	Percent
PCOPP_USD	Copper, grade A cathode, LME spot price, CIF European ports, US$ per metric ton	73%
PLEAD_USD	Lead, 99.97% pure, LME spot price, CIF European Ports, US$ per metric ton	72%
PNICK_USD	Nickel, melting grade, LME spot price, CIF European ports, US$ per metric ton	90%
PORANG_USD	Oranges, miscellaneous oranges CIF French import price, US$ per metric ton	61%
PURAN_USD	Uranium, NUEXCO, Restricted Price, Nuexco exchange spot, US$ per pound	82%
PZINC_USD	Zinc, high grade 98% pure, US$ per metric ton	89%

3. Factor 3: Livestock

Below are the commodities with their percentages that come under Raw Materials. From the description we have identified that imports hold the majority in Raw Materials in global market

Commodities	Description	Percent
PLAMB_USD	Lamb, frozen carcass Smithfield London, US cents per pound	63%
POLVOIL_USD	Olive Oil, extra virgin less than 1% free fatty acid, ex-tanker price U.K., US$ per metric ton	68%
PPORK_USD	Swine (pork), 51-52% lean Hogs, U.S. price, US cents per pound.	73%
PPOULT_USD	Poultry (chicken), Whole bird spot price, Ready-to-cook, whole, iced, Georgia docks, US cents per pound	75%
PSALM_USD	Fish (salmon), Farm Bred Norwegian Salmon, export price, US$ per kilogram	81%
PSAWMAL_USD	Hard Sawn wood, Dark Red Meranti, select and better quality, C&F U.K port, US$ per cubic meter	73%
PSAWORE_USD	Soft Sawn wood, average export price of Douglas Fir, U.S. Price, US$ per cubic meter	89%

4. Factor 4: Raw Materials

Commodities	Description	Percent
PHIDE_USD	Hides, Heavy native steers, over 53 pounds, wholesale dealer's price, US, Chicago, fob Shipping Point, US cents per pound	79%
PSHRI_USD	Shrimp, No.1 shell-on headless, 26-30 count per pound, Mexican origin, New York port, US cents per pound	51%

TECHNICAL APPENDIX

This appendix introduces the various technical issues that were encountered during the factor analysis of the commodities data which is represented above. The analysis was based on a set of 397 observations that were collected from the global commodity market prices.

Data Description

The given dataset consists of 26 variables and all the variables are in the numeric format. Below is the description of the data variables

Alphabetic List of Variables and Attributes			
#	Variable	Type	Len
1	PBEVE_Index	Num	8
2	PCOCO_USD	Num	8
3	PCOFFOTM_USD	Num	8
5	PCOPP_USD	Num	8
6	PCOTTIND_USD	Num	8
7	PFISH_USD	Num	8
8	PHIDE_USD	Num	8
9	PIORECR_USD	Num	8
10	PLAMB_USD	Num	8
11	PLEAD_USD	Num	8
12	PNICK_USD	Num	8
13	POILAPSP_USD	Num	8
14	POILDUB_USD	Num	8
15	POLVOIL_USD	Num	8
16	PORANG_USD	Num	8
17	PPORK_USD	Num	8
18	PPOULT_USD	Num	8
4	PROIL_USD	Num	8
19	PSALM_USD	Num	8
20	PSAWMAL_USD	Num	8
21	PSAWORE_USD	Num	8
22	PSHRI_USD	Num	8
23	PSMEA_USD	Num	8
24	PURAN_USD	Num	8
25	PWHEAMT_USD	Num	8
26	PZINC_USD	Num	8

Looking at the contents, it's clear that each of these variables represent the prices for various commodities that are usually imported and exported in the global marketplace.

Missing Values

I now wanted to check if my data has any missing values. Running the proc means to get the missing values along with the total records for each data variable.

Variable	N Miss	N
PBEVE_Index	0	397
PCOCO_USD	0	397
PCOFFOTM_USD	0	397
PROIL_USD	0	397
PCOPP_USD	0	397
PCOTTIND_USD	0	397
PFISH_USD	0	397
PHIDE_USD	0	397
PIORECR_USD	0	397
PLAMB_USD	0	397
PLEAD_USD	0	397
PNICK_USD	0	397
POILAPSP_USD	0	397
POILDUB_USD	0	397
POLVOIL_USD	0	397
PORANG_USD	0	397
PPORK_USD	0	397
PPOULT_USD	0	397
PSALM_USD	0	397
PSAWMAL_USD	0	397
PSAWORE_USD	0	397
PSHRI_USD	0	397
PSMEA_USD	0	397
PURAN_USD	0	397
PWHEAMT_USD	0	397
PZINC_USD	0	397

From the table, it's clear that there are no missing values for any of my data variables.

Correlation

On computing the correlation matrix amongst our variables, we see that our variables are highly correlated. Hence, this means we need to find the factor analysis in order to classify similar items together.

Standardization

Before we apply the factor analysis techniques, we should ensure that all our variables in the dataset have the same units. But my dataset has different units for each variable, the mainly comprising of these three metric systems:

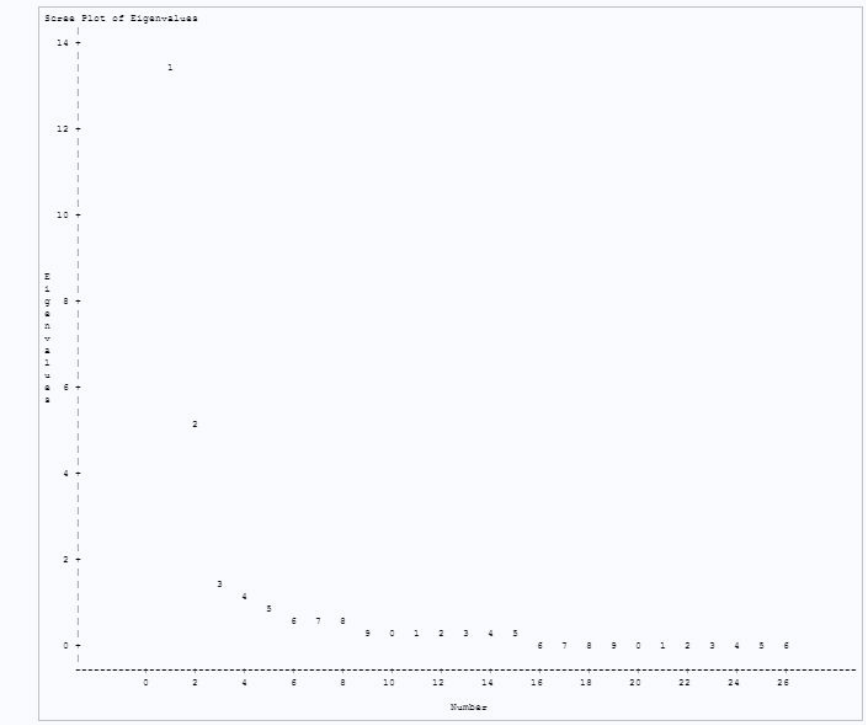
1. US Dollars per Metric Ton
2. US Dollars per Cubic Meter
3. US Cents per Pound
4. US Dollars per Kilogram

We standardize the dataset so that all the variables will have the same units and hence it will then transform it to have zero mean and unit variance.

Factor Analysis: Methods Adopted

The purpose of **factor analysis** is to discover simple **patterns** in the **pattern** of relationships among the variables. It seeks to discover if the observed variables can be explained largely or entirely in terms of a much smaller number of variables called **factors.**

Because we conducted our **factor analysis** on the correlation matrix, the variables are standardized, which **means** that each variable has a variance of 1, and the total variance is equal to the number of variables used in the **analysis.**



Scree plot produces the plot of eigenvalues and proportion variance explained. By default, both **plots** are output in a panel.

After looking at the plot and eigen values, it is found that 4 factors are produced by the factor analysis technique.

It is generally considered that using a rotation in factor analysis will produce more interpretable results. Rotations are helpful as our factor analysis is being performed specifically to gain an explanation of what factors or groups exist in the data. Factor patterns can be rotated through two different ways:

- Orthogonal rotations which retain uncorrelated factors (Varimax Rotation)
- Oblique rotations which create correlated factors (Promax Rotation)

Among the rotations, **VARIMAX rotation has outperformed the others.**

Below are the technical decisions related to the various rotation methods adopted.

I. Factor Analysis with **no rotations**

Variance Explained by Each Factor			
Factor1	Factor2	Factor3	Factor4
13.424779	5.032038	1.558845	1.130599

Looking at the distribution, it looks like the first factor explains much more variance compared to the other factors. Hence, it needs redistribution

II. Factor Analysis with **VARIMAX rotation**

Variance Explained by Each Factor			
Factor1	Factor2	Factor3	Factor4
7.6574310	6.4902382	5.1842426	1.8143493

This rotation nicely distributes the variance among all the factors.

III. Factor Analysis with **PROMAX rotation**

Variance Explained by Each Factor Ignoring Other Factors			
Factor1	Factor2	Factor3	Factor4
10.792687	11.219801	6.982555	3.229708

Reference Axis Correlations				
	Factor1	Factor2	Factor3	Factor4
Factor1	1.00000	-0.55394	0.17729	0.20625
Factor2	-0.55394	1.00000	-0.52038	0.10821
Factor3	0.17729	-0.52038	1.00000	-0.08109
Factor4	0.20625	0.10821	-0.08109	1.00000

Looking at the correlations of the factors, none of them are above 0.32 hence, these don't signify any correlation in the variance explained by the factors.Hence, looking at the latent factors with the variables based on factor loadings, I find the **VARIMAX method rotation as the most appropriate rotation for this dataset**.

PROGRAM CODE

```
libname mydata '/folders/myfolders/';
data mydata.marketdata;
        infile '/folders/myfolders/data15.csv' dlm=',' firstobs=9;
        input PBEVE_Index        PCOCO_USD      PCOFFOTM_USDPROIL_USD        PCOPP_USD
PCOTTIND_USD  PFISH_USD        PHIDE_USD
        PIORECR_USD     PLAMB_USD      PLEAD_USD       PNICK_USD       POILAPSP_USD   POILDUB_USD
POLVOIL_USD     PORANG_USD     PPORK_USD
        PPOULT_USD     PSALM_USD      PSAWMAL_USD  PSAWORE_USD  PSHRI_USD       PSMEA_USD
PURAN_USD       PWHEAMT_USD PZINC_USD;
run;
 Proc contents data=mydata.marketdata;
run;
/* Missing Values*/
proc means data=mydata.marketdata NMISS N; run;
/* Checking the file using the Proc freq, None of my columns have any missing values and all my columns are
numeric */
*Correlation to find if factor analysis is necessary;
proc corr data=mydata.marketdata rank;
run;
 /* Looking at the correlation distribution, we see that some of the variables are highly correlated */
proc standard data=mydata.marketdata mean=0 std=1 out=mydata.marketplace_std;
run;
/*no rotations */
PROC FACTOR DATA=mydata.marketplace_std
         SCREE;
        TITLE "Factor Analysis of Principal dataset";
 RUN;
/*varimax*/
PROC FACTOR DATA=mydata.marketplace_std
                                rotate=varimax
                                nfactors=4     SCREE;
      TITLE "Factor Analysis of Principal dataset VARIMAX";
 RUN;
/*promax*/
ods trace on/listing;
PROC FACTOR DATA=mydata.marketplace_std rotate=promax nfactors=4     SCREE;
ods output OrthRotFactPat=Rotated;     TITLE "Factor Analysis of Principal dataset PROMAX";
RUN;ods trace off;
```