# FACTOR ANALYSIS

Vitastata Sharma

# Table of Contents

## Problem Description:

Analysis in this report is done on a commodity data to identify the themes (i.e. factors) in the data. The given data has 26 variables that will categorized into different themes using Factor Analysis.
We use Factor Analysis technique to

- Reduce the number of variables
- To analyze and find the relationship structure among the variables
- Determine the minimum number of factors in the global market that can adequately account for the variance in the data
- Find simpler, easier to interpret factors through rotating the factors
- Determine a meaningful interpretation of the factors that provides insight into the data for use with further analysis.

## Executive Summary:

As per my analysis I have identified 4 factors- Metals, Raw Material, Utilities and Household to classify the given 26 variables. In the process the data was first imported to SAS Studio, missing values in the data were identified using NMISS and then the data was standardized to have a Mean = 0 and variance of one, each observed variable contributes one unit of variance to the "total variance" in the data set. The initial Factor Analysis report shows Eigenvalues and the Scree Plot. The Eigenvalues are obtained from a principal components analysis. The Scree Plot graphs these eigenvalues. The number of factors equals the number of eigenvalues that exceed 1.0.

Per our analyses, we have identified 4 main factors/themes with eigenvalues >1.0 without rotation.

| Variance Explained by Each Factor | | | |
|---|---|---|---|
| **Factor1** | **Factor2** | **Factor3** | **Factor4** |
| **13.14725** | **4.547741** | **1.523902** | **1.119675** |

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 four factors. These associations with the factors can vary from -100% to 100%. Numbers having larger absolute values indicate a stronger association with that particular factor.

Three rotation methods were applied on the same data-set-

***Orthogonal Rotation*** [which retain uncorrelated factors]***:*** VARIMAX, EQUAMAX

***Oblique Rotation*** [which create correlated factors]***:*** PROMAX

Based on the factor parameters and variance explained by each factor table EQUAMAX was the best rotation method.

| Variance Explained by Each Factor | | | |
|---|---|---|---|
| **Factor1** | **Factor2** | **Factor3** | **Factor4** |
| 5.350281 | 5.173427 | 5.119436 | 4.695427 |

## Factor1: Metals

Below are the commodities with their variance that come under Metals.

| COMMODITIES | DESCRIPTION (HOUSEHOLDS) | Factor Loadings |
|---|---|---|
| PZINC_USD | Zinc, high grade 98% pure, US$ per metric ton | 0.89472 |
| PNICK_USD | Nickel, melting grade, LME spot price, CIF European ports, US$ per metric ton | 0.87307 |
| PURAN_USD | Uranium, NUEXCO, Restricted Price, Nuexco exchange spot, US$ per pound | 0.75712 |
| PCOPP_USD | Copper, grade A cathode, LME spot price, CIF European ports, US$ per metric ton | 0.68367 |
| PLEAD_USD | Lead, 99.97% pure, LME spot price, CIF European Ports, US$ per metric ton | 0.65761 |

## Factor2: Raw Material

Below are the commodities with their variance that come under Raw Material.

| COMMODITIES | DESCRIPTION (UTILITIES) | Factor Loadings |
|---|---|---|
| PCOTTIND_USD | Cotton, Cotton Outlook 'A Index', Middling 1-3/32 inch staple, CIF Liverpool, US cents per pound | 0.81408 |
| PCOFFOTM_USD | Coffee, Other Mild Arabicas, International Coffee Organization New York cash price, ex-dock New York, US cents per pound | 0.75385 |
| PIORECR_USD | China import Iron Ore Fines 62% FE spot (CFR Tianjin port), US dollars per metric ton | 0.73854 |
| PSMEA_USD | Soybean Meal, Chicago Soybean Meal Futures (first contract forward) Minimum 48 percent protein, US$ per metric ton | 0.66693 |
| PWHEAMT_USD | Wheat, No.1 Hard Red Winter, ordinary protein, FOB Gulf of Mexico, US$ per metric ton | 0.59169 |
| PROIL_USD | Rapeseed oil, crude, fob Rotterdam, US$ per metric ton | 0.57476 |
| PSALM_USD | Fish (salmon), Farm Bred Norwegian Salmon, export price, US$ per kilogram | 0.23423 |

## Factor3: Utilities

Below are the commodities with their variance that come under Utilities.

| COMMODITIES | DESCRIPTION (RAW MATERIALS) | Factor Loadings |
|---|---|---|
| PSAWORE_USD | Soft Sawnwood, average export price of Douglas Fir, U.S. Price, US$ per cubic meter | 0.91273 |
| PSAWMAL_USD | Hard Sawnwood, Dark Red Meranti, select and better quality, C&F U.K port, US$ per cubic meter | 0.74174 |
| PPOULT_USD | Poultry (chicken), Whole bird spot price, Ready-to-cook, whole, iced, Georgia docks, US cents per pound | 0.74118 |
| POLVOIL_USD | Olive Oil, extra virgin less than 1% free fatty acid, ex-tanker price U.K., US$ per metric ton | 0.70558 |
| PLAMB_USD | Lamb, frozen carcass Smithfield London, US cents per pound | 0.62552 |

| | | |
|---|---|---|
| PHIDE_USD | Hides, Heavy native steers, over 53 pounds, wholesale dealer's price, US, Chicago, fob Shipping Point, US cents per pound | 0.44373 |
| PSHRI_USD | Shrimp, No.1 shell-on headless, 26-30 count per pound, Mexican origin, New York port, US cents per pound | -0.02417 |

## Factor4: Household

Below are the commodities with their variance that come under Household.

| COMMODITIES | DESCRIPTION (RAW MATERIALS) | Factor Loadings |
|---|---|---|
| PPORK_USD | Swine (pork), 51-52% lean Hogs, U.S. price, US cents per pound. | 0.08219 |
| PCOCO_USD | Cocoa beans, International Cocoa Organization cash price, CIF US and European ports, US$ per metric ton | 0.68908 |
| POILDUB_USD | Oil; Dubai, medium, Fateh 32 API, fob Dubai Crude Oil (petroleum), Dubai Fateh Fateh 32 API, US$ per barrel | 0.6238 |
| POILAPSP_USD | Crude Oil (petroleum), Price index, 2005 = 100, simple average of three spot prices; Dated Brent, West Texas Intermediate, and the Dubai Fateh | 0.62167 |
| PFISH_USD | Fishmeal, Peru Fish meal/pellets 65% protein, CIF, US$ per metric ton | 0.60842 |
| PORANG_USD | Oranges, miscellaneous oranges CIF French import price, US$ per metric ton | 0.52859 |

# Advantages of Factor Analysis:

1. It simplifies the data into categories.

2. It is a multivariate statistical technique that is used to summarize the information contained in many variables into a smaller number of subsets or factors.

3. With Factor Analysis there is no distinction between dependent and Independent variable; rather all variables under investigation are analyzed together to identify underlying factors.

# 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:**

There were 26 Variables and 397 Rows of data. NMISS was used to determine the missing values in the data. And no missing values were observed in the data-set as shown in the figure below:

**Missing Values Report**

**The MEANS Procedure**

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

Each of the 25 variables represented the US dollars of the commodity. The Proc CONTENTS shows the variable Type and Length as shown below:

| Alphabetic List of Variables and Attributes | | | |
|---|---|---|---|
| # | Variable | Type | Len |
| 1 | PCOCO_USD | Num | 8 |
| 2 | PCOFFOTM_USD | Num | 8 |
| 4 | PCOPP_USD | Num | 8 |
| 5 | PCOTTIND_USD | Num | 8 |
| 6 | PFISH_USD | Num | 8 |
| 7 | PHIDE_USD | Num | 8 |
| 8 | PIORECR_USD | Num | 8 |
| 9 | PLAMB_USD | Num | 8 |
| 10 | PLEAD_USD | Num | 8 |
| 11 | PNICK_USD | Num | 8 |

| | | | |
|---|---|---|---|
| **12** | POILAPSP_USD | Num | 8 |
| **13** | POILDUB_USD | Num | 8 |
| **14** | POLVOIL_USD | Num | 8 |
| **15** | PORANG_USD | Num | 8 |
| **16** | PPORK_USD | Num | 8 |
| **17** | PPOULT_USD | Num | 8 |
| **3** | PROIL_USD | Num | 8 |
| **18** | PSALM_USD | Num | 8 |
| **19** | PSAWMAL_USD | Num | 8 |
| **20** | PSAWORE_USD | Num | 8 |
| **21** | PSHRI_USD | Num | 8 |
| **22** | PSMEA_USD | Num | 8 |
| **23** | PURAN_USD | Num | 8 |
| **24** | PWHEAMT_USD | Num | 8 |
| **25** | PZINC_USD | Num | 8 |

## Methods Adopted

The factor analysis was performed with assigning variance as one to each variable. The number of factors was initially selected by retaining only those factors with an eigenvalue greater than one. Among the various methods –list here Principal Component Analysis, Principal Factor Analysis, Harris Component Analysis, unweighted least squares factor analysis etc. *Principal Component Analysis* was use for the data analysis here. The Principal Components Analysis (PCA) method (PRINCIPAL) was chosen for the initial factor extraction. The PCA method simply transforms the set of variables into another set of variables; that is, the data is summarized by means of a linear combination of the observed data. This transformation is performed because of the objective mentioned earlier to account for as much variation as possible in the data.

***Scree plot of Eigen values:*** The SCREE PLOT simply displays the eigenvalues for each of the factors in a plot, from the first eigenvalue (the one that explains the most variance) to the last eigenvalue. Also, where the lines levels off indicates the point of diminishing returns

***FACTOR PATTERN***: The elements of the Factor Pattern reflect the unique variance each factor contributes to the variance of an observed variable.

***VARIANCE EXPLAINED BY EACH FACTOR:*** This output simply summarizes the amount of variance in the data that is explained by each factor

In the Factor Pattern, you can see that most of the variables have high loadings on Factor 3 followed by Factor 1, Factor 2 and Factor 4. The interpretation will become clearer after the rotation. The various rotation methods used for this factor analysis were- VARIMAX , EQUAMAX [ they are orthogonal rotation methods] and PROMAX [ Oblique Rotation].

Factor analysis which uses an orthogonal rotation creates a solution that is easier to grasp and interpret than a solution obtained from an oblique rotation. Of the three Rotation methods used *EQUAMAX* rotation (Orthogonal Rotation) outperformed to the others as it was able to categorize the data into relevant themes as listed earlier.

After the rotation, you can see that each variable has a high loading on one factor and a low loading on the other. Therefore, it is easier to interpret the factors.

**EQUAMAX Rotation Output:**

| Rotated Factor Pattern | | | | |
|---|---|---|---|---|
| | Factor1 | Factor2 | Factor3 | Factor4 |
| PZINC_USD | 0.89472 | 0.20019 | 0.22392 | 0.11401 |
| PNICK_USD | 0.87307 | 0.20102 | 0.24666 | 0.24049 |
| PURAN_USD | 0.75712 | 0.23529 | 0.04693 | 0.49629 |
| PCOPP_USD | 0.68367 | 0.54438 | 0.25143 | 0.37731 |
| PLEAD_USD | 0.65761 | 0.48841 | 0.20056 | 0.46034 |
| PCOTTIND_USD | 0.12118 | 0.81408 | -0.19533 | 0.00311 |
| PCOFFOTM_USD | 0.14128 | 0.75385 | -0.25531 | 0.16338 |
| PIORECR_USD | 0.28672 | 0.73854 | 0.17007 | 0.46422 |
| PSMEA_USD | 0.23159 | 0.66693 | 0.24034 | 0.45298 |
| PWHEAMT_USD | 0.43155 | 0.59169 | 0.1944 | 0.46868 |
| PROIL_USD | 0.45346 | 0.57476 | 0.2909 | 0.51169 |
| PSAWORE_USD | 0.2832 | 0.05651 | 0.91273 | -0.01978 |
| PSAWMAL_USD | 0.38222 | 0.42828 | 0.74174 | 0.04277 |
| PPOULT_USD | 0.33142 | 0.36052 | 0.74118 | 0.37264 |
| POLVOIL_USD | 0.413 | -0.06674 | 0.70558 | -0.0708 |
| PLAMB_USD | 0.38386 | -0.05311 | 0.62552 | 0.2853 |
| PPORK_USD | -0.04132 | 0.07468 | -0.74235 | 0.08219 |
| PSALM_USD | 0.18697 | 0.23423 | -0.81347 | 0.00902 |
| PCOCO_USD | 0.14205 | 0.48062 | -0.26549 | 0.68908 |
| POILDUB_USD | 0.47932 | 0.52072 | 0.1772 | 0.6238 |
| POILAPSP_USD | 0.50058 | 0.49551 | 0.19603 | 0.62167 |
| PFISH_USD | 0.45047 | 0.54058 | -0.0102 | 0.60842 |
| PORANG_USD | 0.51375 | 0.20373 | 0.46683 | 0.52859 |
| PHIDE_USD | 0.12073 | 0.31427 | 0.44373 | - |

| | | | | |
|---|---|---|---|---|
| | | | | 0.67973 |
| **PSHRI_USD** | -0.28182 | -0.19105 | -0.02417 | -0.70624 |

| **Orthogonal Transformation Matrix** | | | | |
|---|---|---|---|---|
| | **1** | **2** | **3** | **4** |
| **1** | 0.58982 | 0.53321 | 0.33729 | 0.50402 |
| **2** | -0.08971 | 0.34249 | -0.8756 | 0.32859 |
| **3** | -0.21557 | 0.75771 | 0.08977 | -0.60939 |
| **4** | 0.77304 | -0.15579 | -0.33392 | -0.51637 |

# PROGRAMS

```
/* Project 1 */

/* Vitastata Sharma */

/*************************************************************************************/

/* FACTOR ANALYSIS*/

/*************************************************************************************/

libname mylib "/home/vitastasharma110/";

filename project1 "/home/vitastasharma110/project1_data_25.csv";

/*Import file*/

data mylib.project2;

   drop PBEVE_Index;

   infile project1 dlm=','  Firstobs= 9;

   input PBEVE_Index   PCOCO_USD        PCOFFOTM_USD   PROIL_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 print data= mylib.project2 ;

run;

/*************************************************************************************/

/*Determining Missing Values In The Dataset*/

proc means data= mylib.project2 n nmiss;

        title "Missing Values Report";

run;

title;

/*Result: No missing values in the data-set*/

/*************************************************************************************/

/* Stardardizing the Data */

proc standard data=mylib.project2 out=mylib.standard mean=0 std=1;

        var _numeric_;

run;

proc means data=mylib.standard nmiss mean std maxdec= 0;

        title "Report After Data Standardization";

run;

title;

/*************************************************************************************/

/* Without rotation*/

ods graphics on;

PROC FACTOR DATA= mylib.standard

            METHOD=PRINCIPAL
```

```
        Reorder
        Corr
        plot= all;
        VAR PCOCO_USD--PZINC_USD;
  TITLE "Factor Analysis Without Rotation";
RUN;
ods graphics off;
/*Analysis- Number of Factors= 4*/
/*****************************************************************************************/
/* With Rotation*/
/*****************************************************************************************/
/*Orthogonal Rotation*/
/* Rotation- VARIMAX*/
ods graphics on;
PROC FACTOR DATA= mylib.standard
      METHOD=PRINCIPAL
      ROTATE= VARIMAX reorder
      Corr
      plot = SCREE
      PRIORS=ONE
      NFACTORS= 4 OUT=mylib.rotated;
      VAR PCOCO_USD--PZINC_USD;
  TITLE "Factor Analysis With Rotation = VARIMAX";
RUN;
ods graphics off;
```

```
/* Rotation- 2*/
ods graphics on;
PROC FACTOR DATA= mylib.standard
        METHOD=PRINCIPAL
           ROTATE=EQUAMAX reorder
           Corr
        plots=SCREE
        PRIORS=ONE
        NFACTORS= 4 OUT=mylib.rotated2;
        VAR PCOCO_USD--PZINC_USD;
  TITLE "Factor Analysis With Rotation = EQUAMAX";
RUN;
ods graphics off;
/*****************************************************************************************/
/* Rotation- 3 */
/* Oblique Rotations-In oblique rotations the new axes are free to take any position in the factor
space, but the degree of correlation allowed among factors is, in general, small because two highly
correlated factors are better interpreted as only one factor. Oblique rotations, therefore, relax the
orthogonality constraint in order to gain simplicity in the interpretation*/
ods graphics on;
PROC FACTOR DATA= mylib.standard
           METHOD=PRINCIPAL
        rotate= promax reorder
        Corr
        plots=SCREE
        PRIORS=ONE
        NFACTORS= 4 OUT=mylib.rotated3;
        VAR PCOCO_USD--PZINC_USD;
```

```
   TITLE "Factor Analysis With Rotation = PROMAX ";
RUN;
ods graphics off;
```