3/16/2016

Harish Visweswaraiya

679799263 / hviswe2

FACTOR ANALYSIS of IMF COMMODITIES

IDS 462 – Statistical Software in Business

Contents

[Background & Objective 2](#_Toc445898006)

[Factor Analysis - Definition 2](#_Toc445898007)

[Data Description 2](#_Toc445898008)

[Results Summary 2](#_Toc445898009)

[Factor1: Crops & Poultry 2](#_Toc445898010)

[Factor2: Industrial Metals 3](#_Toc445898011)

[Factor3: Oil & Natural Gas 3](#_Toc445898012)

[Factor4: Livestock 4](#_Toc445898013)

[Factor5: Other Commodities 4](#_Toc445898014)

[Technical Appendix 4](#_Toc445898015)

[Program 4](#_Toc445898016)

[/\* Loading Input file into SAS \*/ 4](#_Toc445898017)

[/\* Find Missing values in the data \*/ 5](#_Toc445898018)

[/\* Standardizing the dataset \*/ 6](#_Toc445898019)

[/\* 5 Factor loading using Equamax rotation with ‘max’ prior estimates\*/ 6](#_Toc445898020)

[References 8](#_Toc445898021)

## Background & Objective

The International Monetary Fund (IMF) is an organization of 188 countries, working to foster global monetary cooperation, secure financial stability, facilitate international trade, promote high employment and sustainable economic growth, and reduce poverty around the world.

Objective of this project is to identify the hidden themes from observed variables of a commodities data obtained from IMF. This can be achieved through one of the commonly used variable reduction techniques, Factor Analysis.

## Factor Analysis - Definition

Factor Analysis is a method for reducing a large number of variables down to a few underlying unobserved or latent variables called ‘factors’ or ‘themes’ in the data.

## Data Description

Given dataset, ‘project1\_data\_20.csv’, contains 20 variables and 133 observations from IMF commodities dataset.

## Results Summary

Factor Analysis technique has been used on the given dataset to

* Reduce the number of variables and
* To analyze and find the relationship structure among the variables

Five main factors/themes identified using Factor Analysis are as follows:

1. Crops & Poultry
2. Industrial Metals
3. Oil & Natural Gas – Fuels
4. Livestock
5. Other commodities

Results of factor analysis technique indicates how strongly observed variables are associated with each of the five latent 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.

Detailed analysis of each factor is provided below:

### Factor1: Crops & Poultry

Below are the commodities with their percentages that come under Crops & Poultry. From the description we have infer that the import and export is evenly spread across Consumable commodities, crops & poultry

|  |  |  |
| --- | --- | --- |
| **Commodities** | **Description (Crops & Poultry - Consumable Goods)** | **%** |
| PNFUEL\_Index | Non-Fuel Price Index, 2005 = 100, includes Food and Beverages and Industrial Inputs Price Indices | 66.79 |
| PFANDB\_Index | Food and Beverage Price Index, 2005 = 100, includes Food and Beverage Price Indices | 59.32 |
| PGNUTS\_USD | Groundnuts (peanuts), 40/50 (40 to 50 count per ounce), cif Argentina, US$ per metric ton | 68.53 |
| PMAIZMT\_USD | Maize (corn), U.S. No.2 Yellow, FOB Gulf of Mexico, U.S. price, US$ per metric ton | 64.75 |
| PPOULT\_USD | Poultry (chicken), Whole bird spot price, Ready-to-cook, whole, iced, Georgia docks, US cents per pound | 53.83 |
| PRICENPQ\_USD | Rice, 5 percent broken milled white rice, Thailand nominal price quote, US$ per metric ton | 59.67 |
| PSOYB\_USD | Soybeans, U.S. soybeans, Chicago Soybean futures contract (first contract forward) No. 2 yellow and par, US$ per metric ton | 74.39 |

### Factor2: Industrial Metals

Below are the commodities with their percentages that come under Industrial Metals. From the description we notice that the imports dominate the Industrial Metals category.

|  |  |  |
| --- | --- | --- |
| **Commodities** | **Description (Industrial Metals)** | **%** |
| PINDU\_Index | Industrial Inputs Price Index, 2005 = 100, includes Agricultural Raw Materials and Metals Price Indices | 76.828 |
| PCOPP\_USD | Copper, grade A cathode, LME spot price, CIF European ports, US$ per metric ton | 76.871 |
| PIORECR\_USD | China import Iron Ore Fines 62% FE spot (CFR Tianjin port), US dollars per metric ton | 49.436 |
| PNICK\_USD | Nickel, melting grade, LME spot price, CIF European ports, US$ per metric ton | 74.446 |

### Factor3: Oil & Natural Gas

Below are the commodities with their percentages that come under Oil & Natural Gas, which can be otherwise called as fuel resources. From the description we notice that the imports dominate the fuel resource category

|  |  |  |
| --- | --- | --- |
| **Commodities** | **Description (Oil & Natural Gas - Fuel)** | **%** |
| PNGASEU\_USD | Natural Gas, Russian Natural Gas border price in Germany, US$ per thousands of cubic meters of gas | 58.799 |
| PNGASJP\_USD | Natural Gas, Indonesian Liquefied Natural Gas in Japan, US$ per cubic meter of liquid | 46.563 |
| POILAPSP\_USD | Crude Oil (petroleum), Price index, 2005 = 100, simple average of three spot prices; Dated Brent, West Texas Intermediate, and the Dubai Fateh | 52.45 |
| POILBRE\_USD | Crude Oil (petroleum), Dated Brent, light blend 38 API, fob U.K., US$ per barrel | 51.542 |
| PSUNO\_USD | Sunflower oil, Sunflower Oil, US export price from Gulf of Mexico, US$ per metric ton | 76.794 |

### Factor4: Livestock

Below are the commodities with their percentages that are grouped as Livestock, which are animal products. From the description we notice that the exports dominate Livestock

|  |  |  |
| --- | --- | --- |
| **Commodities** | **Description (Livestock)** | **%** |
| PHIDE\_USD | Hides, Heavy native steers, over 53 pounds, wholesale dealer's price, US, Chicago, fob Shipping Point, US cents per pound | 68.783 |
| PPORK\_USD | Swine (pork), 51-52% lean Hogs, U.S. price, US cents per pound. | 56.363 |

### Factor5: Other Commodities

Below are the commodities with their percentages that are grouped as Others, as the variance represented by these observed variables is maximum for this hidden factor. We could see that this doesn’t logically group into a factor so inclusion of more observed variables might contribute to make sense out of this factor.

|  |  |  |
| --- | --- | --- |
| **Commodities** | **Description (Others)** | **%** |
| PNGASUS\_USD | Natural Gas, Natural Gas spot price at the Henry Hub terminal in Louisiana, US$ per thousands of cubic meters of gas | 80.709 |
| PSUGAEEC\_USD | Sugar, European import price, CIF Europe, US cents per pound | 82.732 |

## Technical Appendix

This appendix introduces the various technical aspects that were considered and challenges faced during the factor analysis of commodities data which is represented above. The analysis was based on a set of 133 observations that were collected from the global commodity market prices.

Factor Analysis cannot be performed on observations with missing values. On analyzing, it’s been noted that the dataset is clean with no missing values for all 20 attributes under consideration. Screenshot of missing value check provided in Program section.

Since, each observed variable is scaled differently in the dataset, for example price and index represent two different entities factor loading will be biased. That Standardized values are computed with Mean = 0 & SD = 1.

Identification of Latent Factors included changing different parameters and their inputs to make sense out of the unobserved factors. Parameters given below:

1. Rotation – Several rotation techniques, oblique & non-oblique(orthogonal) have been considered. These include

**Promax & Obequamax** (Oblique)

**Varimax, Equamax, Orthomax, Parsimax, Quartimax** (Orthogonal)

1. Prior Communality Estimates – Two different estimates have been considered.

**SMC** – Square Multiple correlations. This parameter sets the prior communality estimate for each variable to its squared multiple correlation with other variables.

**MAX** – This estimates assigns a maximum correlation value of 1.0 between observed variables and performs rotated factor loadings

1. **Nfactors** – Factor values from 3 – 5 were tried and tested for above parameters to make logical groupings out of the observed variables.
2. **MinEigen** – MinEigen value of 1 was tried against different other parameters. It was noted that number of factors were reduced to 3 after rotation on all occasions. And those factors didn’t make much sense.
3. **Fuzz** – Fuzz value represents the cut-off value of variance that can be displayed in the final ‘Rotated Factor Matrix’. A value of about 0.3 is normally considered so, 0.4 cut-off has been used here to display in the final matrix between observed and latent factors.

Below are the technical decisions relative to the factor analysis:   
Out of the all possible combinations from the parameters listed above, one combination that made more sense out of the grouping for the given dataset is given below:

**Equamax rotation with prior estimates ‘Max’ for nfactors=5. MinEigen value has been omitted because it reduced the number of factors before rotation.** By eliminating the MinEigen value = 1, it could be noted that the final loading after rotation had Eigenvalues more than 1 for all 5 factors.

## Program

### /\* Loading Input file into SAS \*/

**data imf\_commodity\_init;**

**infile '/folders/myfolders/project1\_data\_20.csv' DLM=',' DSD MISSOVER FIRSTOBS=2;**

**input PNFUEL\_Index**

**PFANDB\_Index**

**PINDU\_Index**

**PCOPP\_USD**

**PGNUTS\_USD**

**PHIDE\_USD**

**PIORECR\_USD**

**PMAIZMT\_USD**

**PNGASEU\_USD**

**PNGASJP\_USD**

**PNGASUS\_USD**

**PNICK\_USD**

**POILAPSP\_USD**

**POILBRE\_USD**

**PPORK\_USD**

**PPOULT\_USD**

**PRICENPQ\_USD**

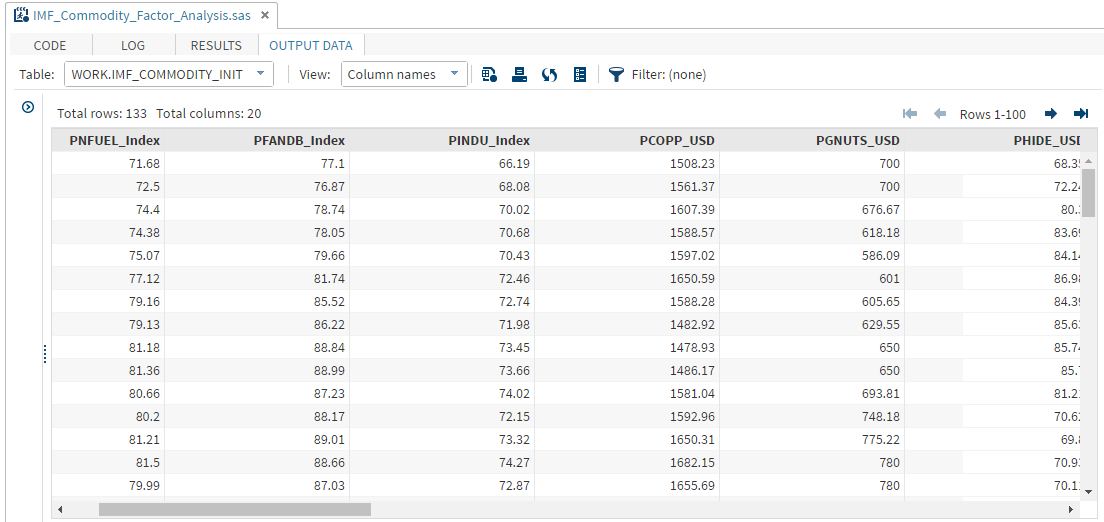
**PSOYB\_USD**

**PSUGAEEC\_USD**

**PSUNO\_USD**

**;**

**run;**

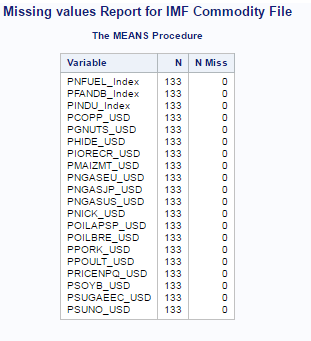


### /\* Find Missing values in the data \*/

**proc means data=imf\_commodity\_init n nmiss maxdec=0;**

**title "Missing values Report for IMF Commodity File";**

**run;**



### /\* Standardizing the dataset \*/

**proc standard data=imf\_commodity\_init out=imf\_commodity mean=0 std=1;**

**var \_numeric\_;**

**run;**



### /\* 5 Factor loading using Equamax rotation with ‘max’ prior estimates\*/

**proc factor data=imf\_commodity**

**corr**

**plot**

**nfactor=5 fuzz=0.4**

**rotate=equamax scree;**

**priors max;**

**run;**

|  |
| --- |
| ***The FACTOR Procedure*** |
| ***Rotation Method: Equamax*** | |

| **Orthogonal Transformation Matrix** | | | | | |
| --- | --- | --- | --- | --- | --- |
|  | **1** | **2** | **3** | **4** | **5** |
| **1** | 0.54771 | 0.49909 | 0.51714 | 0.40610 | -0.13629 |
| **2** | -0.07612 | 0.41127 | -0.07987 | 0.00259 | 0.90481 |
| **3** | -0.35404 | 0.32985 | -0.52396 | 0.66286 | -0.22787 |
| **4** | 0.09770 | -0.68325 | 0.15169 | 0.62566 | 0.33039 |
| **5** | -0.74788 | 0.07824 | 0.65471 | 0.06508 | -0.04087 |

| **Rotated Factor Pattern** | | | | | |
| --- | --- | --- | --- | --- | --- |
|  | **Factor1** | **Factor2** | **Factor3** | **Factor4** | **Factor5** |
| **PNFUEL\_Index** | 0.44787 | 0.66790 | 0.41802 | . | . |
| **PFANDB\_Index** | 0.59323 | 0.50927 | 0.46915 | . | . |
| **PINDU\_Index** | . | 0.76828 | . | 0.43280 | . |
| **PCOPP\_USD** | . | 0.76871 | . | . | . |
| **PGNUTS\_USD** | 0.68526 | . | . | 0.47822 | . |
| **PHIDE\_USD** | . | . | . | 0.68783 | . |
| **PIORECR\_USD** | . | 0.49436 | 0.47441 | . | -0.51581 |
| **PMAIZMT\_USD** | 0.64749 | . | . | 0.47201 | . |
| **PNGASEU\_USD** | 0.54945 | . | 0.58799 | . | . |
| **PNGASJP\_USD** | 0.52182 | . | 0.46563 | 0.59020 | . |
| **PNGASUS\_USD** | . | . | . | . | 0.80709 |
| **PNICK\_USD** | . | 0.74446 | . | . | 0.40677 |
| **POILAPSP\_USD** | 0.52856 | 0.47848 | 0.52450 | 0.42250 | . |
| **POILBRE\_USD** | 0.52542 | 0.46480 | 0.51542 | 0.46293 | . |
| **PPORK\_USD** | . | . | 0.48999 | 0.56363 | . |
| **PPOULT\_USD** | 0.53831 | . | 0.56679 | . | . |
| **PRICENPQ\_USD** | 0.59666 | . | 0.65132 | . | . |
| **PSOYB\_USD** | 0.74396 | . | . | . | . |
| **PSUGAEEC\_USD** | . | . | . | . | 0.82732 |
| **PSUNO\_USD** | 0.44856 | . | 0.76794 | . | . |
| **Values less than 0.4 are not printed.** | | | | | |

| **Variance Explained by Each Factor** | | | | |
| --- | --- | --- | --- | --- |
| **Factor1** | **Factor2** | **Factor3** | **Factor4** | **Factor5** |
| 4.3052614 | 4.0536610 | 4.0208884 | 2.9422771 | 2.0950102 |