**Assignment #7: Factor Analysis (20 points)**

**Data:** The data for this assignment is the stock portfolio data set. This data will be made available by your instructor.

Currently there is no up to date SAS reference for multivariate analysis and the new SAS SG graphical output other than the SAS User’s Manual. Since the SAS User’s Manual can be difficult for a new SAS user, we have provided some SAS information in the course supplements, and we will include more SAS example code and code snippets in the multivariate analysis portion of this course. Some code will be explicit and complete, and some code will only be pseudo code or code snippets. The student is expected to recognize the difference by the time that we reach this section of the class.

**Assignment Instructions:**

For this assignment we will use factor analysis to identify sectors in the stock market.

1. In order to get better factor analysis results let’s drop some of the variables from the data set and create the return data set. By eliminating these companies we will be left with: (1) Banking, (2) Oil Field Services, (3) Oil Refining, and (4) Industrial – Chemical. Within the context of factor analysis we can hypothesize that we have three or four factors (or industry sectors) in this data set. Let’s perform a factor analysis and see what we get for results. Before we begin we will need to create a data set of log-returns, just as we did in our principal components assignment.

**data** temp**;**

set mydata.stock\_portfolio\_data**;**

\* Let's drop some variables to get better factor analysis results;

drop AA HON MMM DPS KO PEP MPC GS **;**

**run;**

**proc** **sort** **data**=temp**;** by date**;** **run;** **quit;**

**data** temp**;**

set temp**;**

\* Compute the log-returs;

\* Note that the data needs to be sorted in the correct

direction in order for us to compute the correct return;

return\_BAC = log**(**BAC/lag1**(**BAC**));**

\* Compute the remainder of the log-returs;

response\_VV = log**(**VV/lag1**(**VV**));**

**run;**

**proc** **print** **data**=temp**(**obs=**10);** **run;** **quit;**

**data** return\_data**;**

set temp **(**keep= return\_:**);**

\* What happens when I put this keep statement in the set statement?;

\* Look it up in The Little SAS Book;

**run;**

**proc** **print** **data**=return\_data**(**obs=**10);** **run;**

1. We will begin our factor analysis by performing a Principal Factor Analysis without a factor rotation. Under this SAS procedure call SAS will automatically select the number of factors to retain. How many factors did SAS retain? Do we have enough factors to support our initial hypothesis? Do are common factors exhibit a simple factor structure?

ods graphics on**;**

**proc** **factor** **data**=return\_data method=principal priors=smc rotate=none

plots=**(**all**);**

**run;** **quit;**

ods graphics off**;**

What criterion did SAS use to select the number of factors to retain? Is there anything strange about the default SAS output?

In addition to estimating the common factors SAS will produce a plot of Factor 1 against Factor 2. Is there anything interesting in this plot?

1. Now let us apply a VARIMAX rotation to the Principal Factor Analysis.

ods graphics on**;**

**proc** **factor** **data**=return\_data method=principal priors=smc rotate=varimax

plots=**(**all**);**

**run;** **quit;**

ods graphics off**;**

Did SAS retain the same number of factors? What components of the PROC FACTOR output did the rotation change? Did we obtain a ‘simple structure’ from our factor rotation? Did we increase the interpretability using the factor rotation?

1. Now let us use Maximum Likelihood Estimation to estimate the common factors (Maximum Likelihood Factor Analysis) with a VARIMAX rotation.

ods graphics on**;**

**proc** **factor** **data**=return\_data method=ML priors=smc rotate=varimax

plots=**(**loadings**);**

**run;** **quit;**

ods graphics off**;**

How many common factors are suggested by ML Factor Analysis? How does ML Factor Analysis arrive at this number of factors, and in general how do we interpret the output from a ML Factor Analysis? From a modeling perspective what does ML Factor Analysis provide that Principal Factor Analysis does not?

1. Every factor analysis procedure requires estimates of the prior communalities. Let’s consider a Maximum Likelihood Factor Analysis with a VARIMAX rotation (the same factor analysis as in (4)), but with the MAX argument for the PRIORS option.

ods graphics on**;**

**proc** **factor** **data**=return\_data method=ML priors=max rotate=varimax

plots=**(**loadings**);**

**run;** **quit;**

ods graphics off**;**

How many common factors are now suggested by ML Factor Analysis? Is this a valid factor analysis? Why or why not? What does this suggest about the sensitivity of common factor estimation to the prior estimates of the communalities?

**Assignment Document:**

All assignment reports should conform to the standards and style of the report template provided to you. Results should be presented and discussed in an organized manner with the discussion in close proximity of the results. The report should not contain unnecessary results or information. The document should be submitted in pdf format. Name your file Assignment7\_LastName.pdf.