Assignment #7

Nate Bitting

Introduction

The purpose of this assignment is to perform various methods of explanatory factor analysis (EFA) leveraging the stock data included in a fund managed by Vanguard (VV). We have removed several stocks from the fund for this assignment so that we are left with stocks from various sectors including banking, oil field services, and oil refining, and industrial — chemical. Thus, our initial hypothesis is that four factors would exist in the dataset based on sector. The EFA methods we will be exploring with this dataset include Principal Factor Analysis and the Maximum Likelihood Estimation method. We will also explore how the VARIMAX rotation can aid in the interpretation of the factor analysis results. The ultimate objective for conducting this factor analysis is to compare each method to determine the appropriate number of factors for the model. We will leverage various tools such as the SCREE plot and the PROPORTION criterion.

Results

Principal Factor Analysis without a Factor Rotation

The first method we will conduct for this analysis is to perform principal factor analysis (PFA) without a factor rotation. After running the SAS PROC FACTOR procedure, SAS selected two factors leveraging the proportion criterion and is also supported in the scree plot in Figure 1.1 below.

Figure 1.2 – Scree Plot from PFA

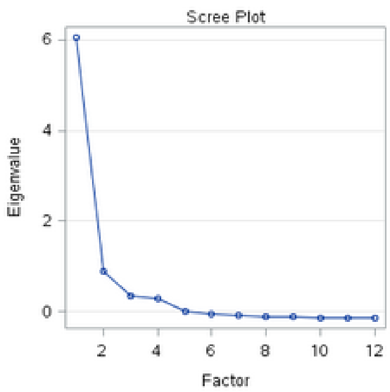


Figure 1.2 – Initial Factor Pattern

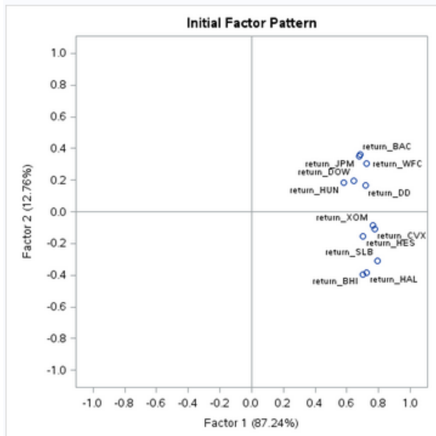
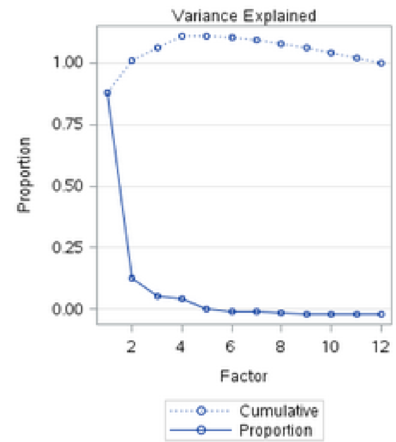


Figure 1.3 – Variance Explained

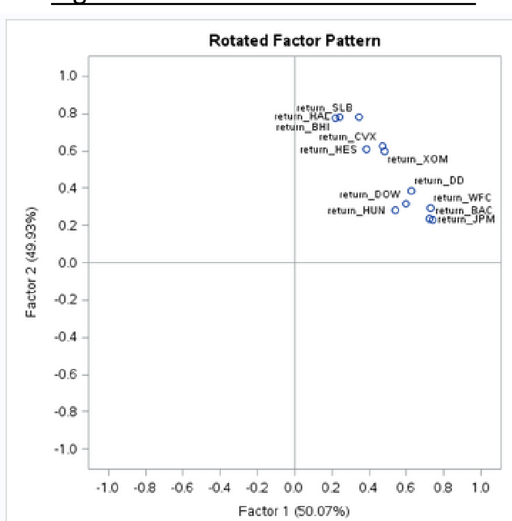


Interestingly enough, our original hypothesis expected four factors, one for each sector, but the initial factor pattern, as shown in Figure 1.2, tells us otherwise. It appears that Factor 2 is comprised of the Banking and Industrial – Chemical sectors and Factor 1 is comprised of the Oil Field Services and Oil Refining sectors. One interesting observation from the SAS output is that the variance explained goes above 1.0 after the 2nd component, and then comes back down to 1.0 for the final component (see Figure 1.3).

Principal Factor Analysis with a Varimax Factor Rotation

After running the PFA with a Varimax factor rotation, SAS retained two factors as well. The primary output that changed was the Initial Factor Pattern plot (Figure 1.2). SAS provides a rotated factor pattern so that interpretability of what falls into each factor much better.

Figure 1.4 – Rotated Factor Pattern



As you can see in Figure 1.4, the rotated factor pattern is much easier to analyze in order to understand what predictors fall into each factor.

Maximum Likelihood Estimation Method with a Varimax Factor Rotation

We will now explore the use of the Maximum Likelihood Estimation method for Factor Analysis with a Varimax factor rotation. After running the ML factor analysis method, SAS also retains two factors as the PFA method did. The primary difference is that ML provides the likelihood ratio χ^2 test statistic difference test between each model which is used to determine the appropriate number of factors. As you can see in Figure 1.5, the p-value for the likelihood ratio χ^2 test statistic is <.0001 with 43 degrees of freedom.

Figure 1.5 – Significance Tests

Significance Tests Based on 501 Observations			
Test	DF	Chi-Square	Pr > Chi Sq
H0: No common factors	66	3656.2617	<.0001
HA: At least one common factor			
H0: 2 Factors are sufficient	43	319.3192	<.0001
HA: More factors are needed			

Figure 1.6 – Convergence Criterion

Iteration	Criterion	Ridge	Change	Communalities											
1	0.6510876	0.0000	0.0766	0.61329	0.67722	0.59574	0.52016	0.43403	0.68819	0.52149	0.34962	0.59394	0.74783	0.65036	0.58326
2	0.6468795	0.0000	0.0233	0.62141	0.68257	0.58380	0.50908	0.42409	0.71149	0.51403	0.34494	0.60820	0.75240	0.65204	0.57178
3	0.6466363	0.0000	0.0044	0.62579	0.68570	0.57998	0.50612	0.42109	0.71292	0.51264	0.34336	0.61064	0.75401	0.65312	0.56870
4	0.6466159	0.0000	0.0017	0.62697	0.68582	0.57927	0.50513	0.42017	0.71351	0.51229	0.34293	0.61230	0.75408	0.65322	0.56796
5	0.6466133	0.0000	0.0006	0.62759	0.68586	0.57907	0.50478	0.41980	0.71345	0.51224	0.34275	0.61278	0.75412	0.65329	0.56779

Another difference between PFA and ML is that the SAS output for ML provides the convergence criterion for each iteration. As you can see in Figure 1.6, the change in the criterion after adding the third component (highlighted in yellow) is negligible, indication two factors is the appropriate number of factors.

Maximum Likelihood Estimation Method with a Varimax Factor Rotation and PRIORS set to MAX

The last method we will employ for this factor analysis is to perform the same procedure as the previous method, but change the PRIORS option by setting it to MAX. This method actually suggests we retain five common factors as opposed to just two. The reason for this is due to the estimates of prior communalities from the MAX method. This is not a valid factor analysis as the only instance in which you should employ the MAX method for prior communality estimation is when your correlation matrix is singular. Therefore, it is critically important you understand the assumptions of prior communalities as it can have a drastic impact on the results of your factor analysis.

Conclusion

Overall, this exercise has shown the immense value of factor analysis and the various methods one can leverage to determine the appropriate number of components. It is good to leverage all available options in order to compare how many factors are appropriate for the dataset being used in the analysis. Given I am new to explanatory factor analysis, I would most likely employ the most commonly used method of maximum likelihood estimation and the Varimax factor rotation for determining the appropriate number of factors. Lastly, it is critical that one understand the assumptions regarding the prior communalities as it can have a dramatic impact on the results of the factory analysis.

SAS Code Output

```
* Code used to get the data into my library;
ods graphics on;
libname mydata '/courses/d6fc9ae5ba27fe300/c_3505/SAS_Data/' access=readonly;
proc datasets library=mydata; run; quit;

data temp;
set mydata.stock_portfolio_data;
drop AA HON MMM DPS KO PEP MPC GS ;
run;

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

data temp;
set temp;
return_BAC = log(BAC/lag1(BAC));
return_BHI = log(BHI/lag1(BHI));
return_CVX = log(CVX/lag1(CVX));
return_DD = log(DD/lag1(DD));
return_DOW = log(DOW/lag1(DOW));
return_HAL = log(HAL/lag1(HAL));
return_HES = log(HES/lag1(HES));
return_HUN = log(HUN/lag1(HUN));
return_JPM = log(JPM/lag1(JPM));
return_SLB = log(SLB/lag1(SLB));
return_WFC = log(WFC/lag1(WFC));
return_XOM = log(XOM/lag1(XOM));
response_VV = log(VV/lag1(VV));
run;

*create a list of all the predictor variables;
%let xlist =return_BAC return_BHI return_CVX return_DD return_DOW
                    return_HAL return_HES return_HUN return_JPM return_SLB
                    return_WFC return_XOM;

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

* Just keep the log return variables;
data return_data;
set temp (keep= return_:);
run;

proc print data=return_data(obs=10); run;

* Principal Factor analysis without rotation;
ods graphics on;
proc factor data=return_data method=principal priors=smc rotate=none
plots=(all);
run; quit;
ods graphics off;

* Principal Factor analysis with varimax rotation;
ods graphics on;
proc factor data=return_data method=principal priors=smc rotate=varimax
plots=(all);
run; quit;
ods graphics off;

* Maximum Likelihood Estimation method with varimax rotation;
ods graphics on;
proc factor data=return_data method=ML priors=smc rotate=varimax
plots=(loadings);
run; quit;
```

```
ods graphics off;

* Maximum Likelihood Estimation method with varimax rotation and using the MAX PRIORS option;
ods graphics on;
proc factor data=return_data method=ML priors=max rotate=varimax
plots=(loadings);
run; quit;
```