Principal Component Analysis

PCA is widely used in multivariate statistics such as factor analysis and it is used to reduce the number of variables. It is used to identify a new set of variables such that each new variable, called a principal component, is a linear combination of original variables.

Our data is composed of 130 times series or variables, each time series x\_i represents the number of posts about one corresponding keyword on trading days. The principal component analysis is conducted in SPSS (Statistical Package for the Social Sciences) version 20 for Windows.

First, we use Z standardization method to standardize data and calculate the correlation matrix of standardized data. After standardization, the average value of the variables is 0 and the standard deviation is 1, thus eliminating the effects of magnitude.

Next, It extracts 32 principal components from the previous 130 variables based on the condition that eigenvalues greater than 1. The result is shown in Figure 1 and Figure 2.

Finally, the component score coefficient matrix is obtained and can be seen in Table 1 . For each principal component, we try to extract some main variables to explain its specific meaning. However, the variables we extract for each principal component have no obvious relationship and there is no unified financial explanation to describe these variables.

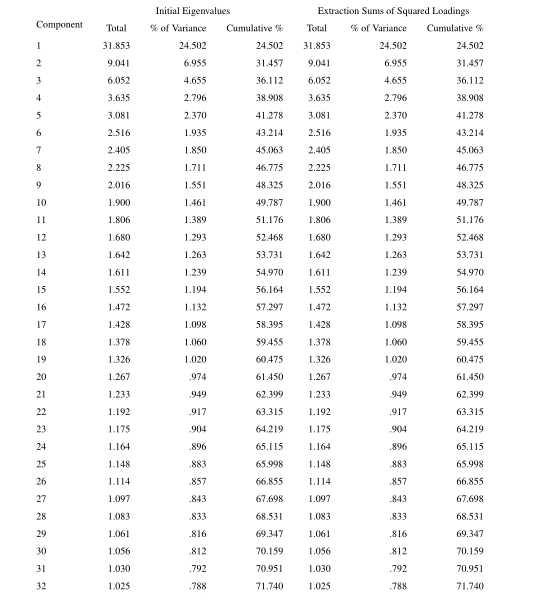


Figure 1: Total Variance Explained

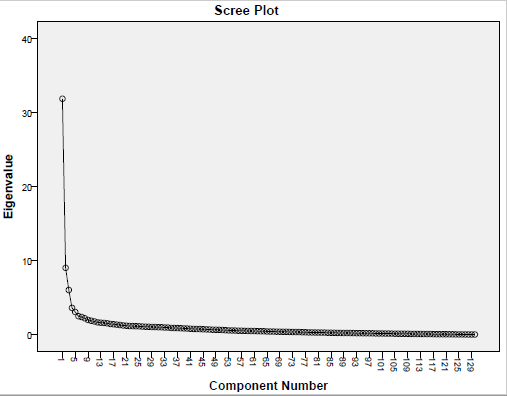


Figure 2: Eigenvalue vs. Component number

On the one hand, the number of extracted principal components is still too large to go on our research; on the other hand, we have no way to give a specific financial meaning for each principal component. Based on the above analysis, we are not going to present the process of principal component analysis in our paper.

Table 1: Component Score Coefficient Matrix

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Zscore(x\_i)  Factor | Zscore(x1) | Zscore(x2) | Zscore(x3) | Zscore(x4) | …… | Zscore(x130) |
| 1 | .024 | .014 | .018 | .021 | …… | .010 |
| 2 | -.015 | .053 | .049 | -.021 | …… | .020 |
| 3 | -.001 | .004 | -.018 | .077 | …… | .060 |
| 4 | .019 | -.042 | -.054 | -.023 | …… | -.047 |
| 5 | -.018 | .041 | -.047 | .010 | …… | .014 |
| 6 | .029 | .052 | -.061 | -.022 | …… | .013 |
| 7 | -.043 | -.051 | .078 | -.059 | …… | .090 |
| 8 | .009 | .052 | -.002 | -.017 | …… | .063 |
| 9 | -.045 | -.031 | -.049 | -.002 | …… | -.041 |
| 10 | -.033 | .010 | .040 | -.019 | …… | .000 |
| 11 | .044 | .002 | -.047 | .066 | …… | .110 |
| 12 | .000 | .025 | .020 | .018 | …… | -.094 |
| 13 | .023 | .039 | -.014 | -.030 | …… | -.036 |
| 14 | -.029 | -.086 | .062 | .002 | …… | -.051 |
| 15 | .000 | -.096 | .050 | -.007 | …… | -.038 |
| 16 | -.043 | .003 | -.026 | -.011 | …… | .092 |
| 17 | -.086 | .003 | -.059 | -.083 | …… | -.042 |
| 18 | -.006 | -.112 | -.004 | .019 | …… | .045 |
| 19 | -.022 | .014 | .025 | -.068 | …… | -.042 |
| 20 | .005 | -.095 | .005 | -.060 | …… | .024 |
| 21 | .080 | -.181 | .002 | -.021 | …… | -.078 |
| 22 | .006 | .031 | -.082 | -.071 | …… | -.009 |
| 23 | .098 | .083 | .037 | -.066 | …… | -.123 |
| 24 | .042 | .059 | -.096 | .052 | …… | -.036 |
| 25 | .006 | .099 | -.008 | .055 | …… | .037 |
| 26 | .067 | -.053 | .036 | .031 | …… | -.017 |
| 27 | -.009 | .024 | .024 | .061 | …… | -.062 |
| 28 | -.017 | -.020 | .024 | -.039 | …… | -.010 |
| 29 | .022 | .087 | -.026 | .007 | …… | .060 |
| 30 | -.031 | .068 | .044 | .001 | …… | .084 |
| 31 | .011 | .040 | .028 | -.042 | …… | -.019 |
| 32 | -.015 | -.003 | .024 | .004 | …… | .138 |