

SMF PROBLEMS 3. 2.2.2017

Q1. This question is a continuation of Problem 2 in Practical 3.

1. Use the *princomp* function to run a PCA on the `UStreasury.csv` (in R, although you're welcome to use whichever language you wish). (*Hint: the **summary** function gives the components weights, the loadings functions gives the components...*) and comment on which components to retain.
2. Interpret the meaning of the significant components.

Q2. If the variables in the dataset on which you plan to perform PCA are compatible (e.g. they are all daily returns on stocks) consideration of units is not usually required. But, if the components are incompatible (say, unemployment rate vs. tax receipts), some care needs to be taken.

Consider two variables, with a correlation coefficient $r = 0.8$.

1. Write down the correlation matrix. What are the eigenvalues? Which is the principal component? In which direction is most of the variation?
2. Now suppose that variable 1 has variance 1,000,000, and variable 2 variance 1. Write down the covariance matrix. Which variable dominates the PCA?
3. Variable 1 was measured in pounds. I now decide to convert it into millions of pounds. What is the new variance of Variable 1 and what happens to the eigenvectors of the covariance matrix? Does it matter?

NHB/TLS