1. **Principal Component Analysis and Risk Factors.**
   A. Consider the 15 eigenportfolios corresponding to the top 15 eigenvalues in the PCA conducted in the previous assignment (constituents of S&P 500, 1 year lookback window). Let $F_{kt}$ denote the daily return of the eigenportfolio $k$ on date $t$. For each $k$, estimate the tails of the distribution of 1-day returns $F_k$ by performing a Q-Q plot versus a Student-t distribution with $d$ degrees of freedom and finding the best fit for $d$.

   B. The above method suggests using the following factor model for stock returns

   $R_s = \sigma_s \left( \sum_{k=1}^{15} \beta_{sk} F_k \right) + \sigma_s \left( \sqrt{1 - \sum_{k=1}^{15} \beta_{sk}^2} \right) G_s$  

   where $F_k, k = 1, \ldots, 15$ are uncorrelated Student-t variables. We assume that $G_s$ is also a Student-t variable, which is uncorrelated with $F_k$ and with $G_{s'}$ if $s'$ is a different stock. Calculate the correlation matrix for stock returns corresponding to the model. What are its eigenvalues?

2. **Stock Returns regressed on ETF returns.**
   This exercise uses de data of Exercise 1 of the previous section as well as additional data on ETFs for the same period. Consider the following 16 etfs:

   - XLF Financial SPDR
   - XLE Energy SPDR
   - XLU Utilities SPDR
   - IYR Real Estate IShare
   - IYE Energy IShare
   - IBB Biotechnology IShare
   - RTH Retail HOLDR
   - RKH Regional Banks HOLDR
IYT Transportation IShare

XLY Consumer Discretionary SPDR

A. Perform a multiple regression (least-squares fit) of the returns of each of the constituents of the S&P 500 on the 16 ETFs. List, for each ticker (i) the 16 regression coefficients, (ii) the ETF corresponding to the largest regression coefficient (iii) the ETF corresponding to the largest explained variance (iv) the R-squared of the regression. Rank the stocks by decreasing order in R-squared and explain your findings.

B. Write a “Matching Pursuit” algorithm along the lines explained in class. Perform an analysis similar to A. on the 500 equities using Matching Pursuit instead of Multiple Regression.

C. Using the ETF with largest explained variance in B, perform a regression of each stock on the corresponding ETF and consider the corresponding R-squared for each stock. For each stock, compare the latter with the R-squared in A and with the R-squared in Assignment 1.