Risk & Portfolio Management, Spring 2010
Homework 1

1. Principal Component Analysis of Stock Market Returns. The goal of this exercise is to understand the correlation of returns of a large group of stocks, such as the Standard and Poor's 500 Index or the Nasdaq-100 (the latter being more focused on tech stocks, thus slightly narrower than the SP500).

A. Collect daily price data for the components of SP500 from Yahoo!Finance for the last year (252 trading days in the past), to generate an approximate 252 rows by $N_{stocks} \approx 500$ columns of prices. Construct the corresponding matrix of standardized returns

$$Y_{it} = \frac{R_{it} - \overline{R}_i}{\overline{\sigma}_i} \quad i = 1, ..., N_{stocks}, \quad t = 1, ..., T, \tag{1}$$

where $T = 251$,

$$\overline{R}_i = \frac{\sum_{t=1}^{T} R_{it}}{T}, \quad \text{and} \quad \overline{\sigma}_i = \sqrt{\frac{\sum_{t=1}^{T} (R_{it} - \overline{R}_i)^2}{T}}.$$

B. Calculate the $N_{stocks} \times N_{stocks}$ sample correlation matrix, defined as

$$C_{ij} = \frac{1}{T} \sum_{t=1}^{T} Y_{it} Y_{jt} \tag{2}$$

and the *regularized correlation matrix* as

$$\tilde{C}_{ij} = \frac{C_{ij} + \epsilon \delta_{ij}}{1 + \epsilon}, \tag{3}$$

where δ_{ij} is the Kronecker delta and $\epsilon = 10^{-9}$. Verify that $\tilde{C}$ corresponds to the correlation matrix of another random vector. Which one?

C. Calculate the eigenvalues and eigenvectors of the matrix $\tilde{C}_{ij}$ and graph the eigenvalues $\lambda_1 > \lambda_2 \geq \lambda_3 ...$ in decreasing order. What percent of the trace is explained by summing the first 5 eigenvalues?

D. Focus on the first eigenvector, and denote it by $(V_1, V_2, ..., V_{N_{stocks}})$. Define the factor

$$F_t \;=\; \frac{1}{\sqrt{\lambda_1}} \sum_{i=1}^{N_{stocks}} \frac{V_i}{\overline{\sigma}_i} R_{it} \quad t = 1, 2, ..., T \tag{4}$$

In other words, F_t is the time-series of returns of a portfolio which invests $\frac{1}{\sqrt{\lambda_1}} \frac{V_i}{\overline{\sigma}_i}$ dollars in the i^{th} stock, $i = 1, ..., N_{stocks}$. Calculate the sample mean and the sample standard deviation of the factor F.

E. Consider a series of returns of the Standard and Poors 500 index (symbol= $\hat{}GSPC$) for the same time period. Perform a linear regression of the returns of F with the standardized returns of $\hat{}GSPC$. Calculate the R-squared of the regression and discuss whether F and the capitalization-weighted market portfolio are good proxies for each other. Discuss the result and argue why F and the market index might be related.

2. Totally optional. Repeat for NASDAQ-100.