The dataset consists a few different documents, including one dataset on the stock prices of SP500 and a few datasets of factors of the financial market. Denote N as the total number of stocks and T as the length of longitude.

1. For each stock i, consider the linear factor model

$$r_{it} = \alpha_i + f_t'\beta_i + \epsilon_{it}, \tag{1}$$

where r_{it} is the return of the stock from t to $t+1$, ϵ_{it} is a residual with mean 0, f_t is a vector of all the available factors. For each stock, the volatility of the idiosyncratic risk is

$\sigma_i^2 := \mathbb{E}[\epsilon_{it}^2]$.

How to construct a unbiased estimate of σ_i^2?

2. Filter the data from Jan. 1990-Dec. 2012. Estimate the model for all the stocks using the filtered data and report the average estimated volatility of them.

3. You have got residuals $\hat{\epsilon}_{it} = r_{it} - \hat{\alpha}_i - f_t\hat{\beta}_i$, where $\hat{\alpha}_i$ and $\hat{\beta}_i$ are the estimates you get from the previous exercise. Consider the covariance matrix of the idiosyncratic risk $V = Var(\epsilon_t)$, where $\epsilon_t = (\epsilon_{1t}, ..., \epsilon_{Nt})$. You can construct an estimated covariance matrix $\hat{V}$ as the following:

$$\hat{V} := \frac{1}{T}\sum_{t=1}^{T} \hat{\epsilon}_t\hat{\epsilon}_t', where \tag{2}$$

$\hat{\epsilon}_t = (\epsilon_{1t}, ..., \epsilon_{Nt})'$ is a $N \times 1$ vector.

Perform a principle component analysis (PCA) on $\hat{V}$ and report the first 3 largest eigenvalues of it. How much percentage of idiosyncratic risk does the first principle component explain in an equally weighted portfolio?