Applications of random matrix theory

Use the provided data set which contains the daily (close-to-close) returns of 472 stocks in SP500 for a period of 561 days (during 2012-01-04 and 2014-04-30). There is also a bigger data set of stocks in the SP1500, in the interval 2013-07-29 and 2019-07-15.

(a) Compute and plot a histogram of the eigenvalues of the empirical covariace matrix. What do you observe? Plot the same spectrum, but leave out the largest eigenvalues.

(b) Randomly shuffle the entries in the matrix of returns (RETS), and compute the eigenvalues. Repeat this experiment 50 times, and record the average value of the obtained eigenvalues. (You can use the sample function to randomly permute entries RETSRAND = matrix(sample(c(RETS)) , nrow = dim(RETS)[1], ncol = dim(RETS)[2])). Plot a histogram of the resulting averaged eigenvalues (You could set the "breaks" parameter to 50 in your histogram, for better visualization).

(c) Compare the largest eigenvalue obtained in (a) with the largest eigenvalue obtained in (b)

(d) Leaving out the largest eigenvalue in (a), compute a Q-Q plot of the two distributions from (a) and (b). In other words, compute the quantiles of the eigenvalues obtained in (b) (You could use the R command quantile(eigvals, probs = seq(0, 1, 0.05), na.rm = TRUE);) and then do similarly for the eigenvalues from (a) but ignoring the largest eigenvalue, and plot the two quantiles against each other. What can you conclude? Are the two distributions the same?

(e) Compute a scatter plot of the top 20 largest eigenvalues in (a) excluding the largest eigenvalue (on the x-axis), versus the top 20 largest eigenvalues in (b) (on the y-axis). Overlap on this plot the line $x = y$, i.e. use *lines(eigvalsActual[2:21], eigvalsActual[2:21], type='b')*, if eigvalsActual denotes the vector of eigenvalues from part (a). How do the the two sets of eigenvalues compare. What could you conclude?