Empirical Methods in Finance
Homework 7

Prof. Lars A. Lochstoer

February 26, 2020

Please use Matlab/R to solve these problems. You can just hand in one set of solutions that has all the names of the contributing students on it in each group.

[The quality of the write-up matters for your grade. Please imagine that you're writing a report for your boss at Goldman when drafting answers these questions. Try to be clear and precise.]

Principal Component Analysis

Download the 48 industry portfolio data (monthly) from Kenneth French's web site. Use the data from 1960 through 2015. Use the value-weighted returns. You may drop the industries that have missing values and are reported as -99.99. Also, download the 3 Fama-French factors from his web site. Use the monthly risk-free rate series provided by French in the same FF factor dataset to compute excess returns on these 48 portfolios.

1. Get the eigenvalues for the sample variance-covariance matrix of the excess returns to the 48 industries. Plot the fraction of variance explained by each eigenvalue in a bar plot.

2. Choose the 3 first (largest) principal components.

   (a) How much of the total variance do these 3 factors explain?

   (b) Give the mean sample return to these 3 factor portfolios, their standard deviation, and correlation.

(c) Consider a multi-factor model of returns using these three factors as pricing factors. Plot the predicted return from this model for all the industries versus the realized average industry returns over the sample. That is, estimate the betas of each industry with respect to these factors, get the expected excess returns as

$$\hat{E}\left[R_{it}^{e}\right] = \hat{\beta}_{i}^{\prime}\hat{E}\left[F_{t}\right], \tag{1}$$

where F_t are the factor returns and $\hat{E}[F_t]$ is estimated as the sample average of each factor. This expected return will be on your x-axis. The sample average of each industry's excess return will be on the y-axis. Add a 45 degree line to this plot.

[You can get factor loadings (betas) from the eigenvectors. Or, if you like, you can run the time-series regression of each industry's return on the 3 factors. The result is the same.]

(d) Give the implied cross-sectional R^2 of the plot in c). That is, calculate:

$$R^{2}_{\text{cross-section}} = 1 - \frac{var\left(\bar{R}^{act} - \hat{R}^{pred}\right)}{Var\left(\bar{R}^{act}\right)}$$

where $\bar{R}^{act}$ is the $N \times 1$ vector of average industry excess returns. $\hat{R}^{pred}$ is the $N \times 1$ vector of predicted industry excess returns from the 3-factor model.

3. Now, download the 25 FF portfolios sorted on size and book-to-market, same sample period.

(a) Get the eigenvalues for the sample variance-covariance matrix of the excess returns to these 25 F-F portfolios. Plot the fraction of variance explained by each eigenvalue in a bar plot.

(b) Given (a), how many factors does do you reckon you need to explain average returns to the 25 F-F portfolios?