Problem Set #1
Business 35902
Winter 2019

The due date for this assignment is January 24 (Thursday)
(upload on canvas)

Write up the solution clearly, using tables and graphs where appropriate, preferably using LaTeX. In addition to showing the results, you must also explain and discuss your results in words, as you would do if you wrote a paper. Carry out your analysis in MATLAB and attach to your solution the MATLAB program(s) that you wrote (if you would like to use other software, please consult with me first). Structure your program(s) carefully and insert some explanatory comments so that it is transparent to an outsider what your program does.

For this problem, use the following two files that you already used for Problem Set #1:

- `StockPortfolios.csv` contains monthly returns for 30 different portfolios based on sorts of stocks according to 15 different stock characteristic. For each characteristic there is one portfolio composed of stocks with very low values of this characteristic and one of stocks with very high values of the characteristic. (These returns are from the website of Robert Novy-Marx at U of Rochester).
- `FFMktFct.csv` contains a series of monthly Treasury Bill yields and a series of monthly returns on the CRSP value-weighted index (= the value-weighted return on a portfolio including virtually all stocks traded in the U.S.)

Construct monthly excess returns on the market factor F (i.e., CRSP value-weighted index in excess of the T-bill return) and monthly excess returns on the 30 stock characteristics portfolios ($\boldsymbol{z}$), both for the sample period 1964m1 to 2013m12.

1. Estimate an SDF in which F (demeaned) is the only factor, $M_t = 1 - b(F_t - \mu_F)$. Follow the procedure we discussed in class, in two variants:

   (a) Imposing Assumption IID

   (b) imposing only no serial correlation in the estimation of the $\boldsymbol{S}$ matrix, but not independence of factor model residuals from factor innovations. Use the two-step optimal GMM estimator with identity weighting matrix in the first stage. De-mean $\hat{\boldsymbol{g}}_t$ when estimating the $\boldsymbol{S}$ matrix.

   For both cases, report the SDF and factor mean parameter estimates and their standard errors, the pricing errors for the $N = 30$ portfolios, the χ^2-statistic for the test of the null hypothesis of zero pricing errors for the N portfolios, and the estimates of the HJ-distance (unexplained squared Sharpe Ratio) $\boldsymbol{\alpha}\boldsymbol{\Sigma}^{-1}\boldsymbol{\alpha}$. Comment on the similarity/differences in the estimates under Assumption IID, first-stage GMM, and second-stage GMM.

2. Now construct the first five principal component factors (i.e., the five associated with the highest eigenvalues) from the excess returns of the 30 portfolios. Redo your estimation with an SDF based on these five (demeaned) factors, in three ways: (i) imposing Assumption IID; (ii) first-stage GMM with identity weighting matrix; (iii) second-stage GMM.

   (a) Compare the estimates from approach (i) and (ii). Why are they similar? Can you prove that they should be equal? (*Hint*: Consider the $\boldsymbol{d}$ matrix and the fact that the factors are PCs of the N test asset returns)

   (b) What happens when you try to implement the second-stage estimation (iii)? Why is $\boldsymbol{S}$ not invertible? Try two alternative approaches. First, use the pseudo-inverse instead of the inverse of $\boldsymbol{S}$. Second, re-run first and second-stage using only the moment conditions for the N assets and K factor means, but not the K pricing conditions for the factors. Compare the results of the two alternative approaches. Use the result to discuss, intuitively, what the pseudo-inverse does in terms of the weighting of the data.