MGMTMFE 407: Empirical Methods in Finance
Homework 1

Prof. Lars A. Lochstoer
TA: Danyu Zhang

January 8, 2020

Please use 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. Use the electronic drop box to submit your answers. Submit the R file and the file with a short write-up of your answers separately.

[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 be clear and precise.]

Problem 1: Building a simple autocorrelation-based forecasting model

Fama and French (2015) propose a five-factor model for expected stock returns. One of the factor is based on cross-sectional sorts on firm profitability. In particular, the factor portfolio is long firms with high profitability (high earnings divided by book equity; high ROE) and short firms with low profitability (low earnings divided by book equity; low ROE). This factor is called RMW – Robust Minus Weak.

1. Go to Ken French's Data Library (google it) and download the Fama/French 5 Factors (2x3) in CSV format. Denote the time series of value-weighted monthly factor returns for the RMW factor from 196307-201911 as "rmw." Plot the time-series, give the annualized mean and standard deviation of this return series.

2. Plot the 1st to 60th order autocorrelations of rmw. Also plot the cumulative sum of these autocorrelations (that is, the 5th observation is the sum of the first 5 autocorrelations, the 11th observation is the sum of the first 11 autocorrelations, etc.). Describe these plots. In particular, do the plots hint at predictabilty of the factor returns? What are the salient patterns, if any?

3. Perform a Ljung-Box test that the first 6 autocorrelations jointly are zero. Write out the form of the test and report the p-value. What do you conclude from this test?

4. Based on your observations in (2) and (3), propose a parsimonious forecasting model for rmw. That is, for the prediction model

$$rmw_{t+1} = \alpha + \beta' x_t + \varepsilon_{t+1}, \tag{1}$$

choose the variables in x_t – it could be only one or a $K \times 1$ vector. While this analysis is in-sample, I do want you to argue for your variables by attaching a "story" to your model that makes it more ex ante believeable. (PS: This question is purposefully a little vague. There is not a single correct answer here, just grades of more to less reasonable as in the real world).

5. Estimate the proposed model. Report Robust (White) standard errors for $\hat{\beta}$, as well as the regular OLS standard errors. In particular, from the lecture notes we have that

$$\begin{aligned} Var^{White}\left(\hat{\beta}\right) &= \frac{1}{T}\left(\frac{1}{T}\sum_{t=1}^{T} x_t x_t'\right)^{-1} \frac{1}{T}\sum_{t=1}^{T} x_t x_t' \hat{\varepsilon}_t^2 \left(\frac{1}{T}\sum_{t=1}^{T} x_t x_t'\right)^{-1}, & (2) \\ Var^{OLS}\left(\hat{\beta}\right) &= \frac{1}{T}\left(\frac{1}{T}\sum_{t=1}^{T} x_t x_t'\right)^{-1} \left(\frac{1}{T}\sum_{t=1}^{T} x_t x_t'\right)\left(\frac{1}{T}\sum_{t=1}^{T} \hat{\varepsilon}_t^2\right)\left(\frac{1}{T}\sum_{t=1}^{T} x_t x_t'\right)^{-1}. & (3) \end{aligned}$$

(In asymptotic standard errors, we do not adjust for degrees of freedom which is why we simply divide by T).

Problem 2: Nonstationarity and regression models

1. Simulate T time series observations each of of the following two return series N times:

$$\begin{aligned} r_{1,t} &= \mu + \sigma \varepsilon_{1,t}, \\ r_{2,t} &= \mu + \sigma \varepsilon_{2,t}, \end{aligned} \tag{4}$$

where $\mu = 0.5\%$, $\sigma = 4\%$, and the residuals are uncorrelated standard Normals. Let $T = 600$ and $N = 10,000$. For each of the N time-series, regress:

$$r_{1,t} = \alpha + \beta r_{2,t} + \varepsilon_t, \tag{5}$$

and save the slope coefficient as $\beta^{(n)}$, where $n = 1, ..., N$. Give the mean and standard deviation of β across samples n and plot the histogram of the $10,000$ β's. Does this correspond to the null hypothesis $\beta = 0$? Do the regress standard errors look ok?

2. Next, construct N price sample of length T based on each return using:

$$\begin{aligned} p_{1,t} &= p_{1,t-1} + r_{1,t}, \\ p_{2,t} &= p_{2,t-1} + r_{2,t}, \end{aligned} \tag{6}$$

using $p_{1,0} = p_{2,0} = 0$ as the initial condition. Now, repeat the regression exercise using the regression:

$$p_{1,t} = \alpha + \beta p_{2,t} + \varepsilon_t. \tag{7}$$

Again report the mean and standard deviation of the N estimated β's and plot the histogram. Does this correspond to the null hypothesis $\beta = 0$? Do the regression standard errors look ok? Explain what is going on here that is different from the previous return-based regressions.