MIT Sloan School of Management

Analytics of Finance | 15.457
Hui Chen | Spring 2021

Problem Set 2

(Due: 7:30 AM, Tuesday, March 16)

Instructions:

- Please submit your homework on Canvas. List the names of all of your team members and their IDs in your writeup. Each team only needs to submit once.
- Submit the following files: (1) a PDF file for the writeup of your answers to all the questions; (2) for each question that involves coding, a separate zip file with the code and data for that question.
- For question(s) involving coding, you will receive credit only when the code can be successfully executed.

1. **Interview questions.**

   (a) Return of stock X has mean 20% and volatility 50%. Market return has mean 8% and volatility 20%. The correlation between stock X and market is 50%. If we run regression of stock X return on market return (with intercept), what is the regression coefficient on market return (i.e. beta)?

   (b) True or false? If X and Y are uncorrelated, then they must be independent. Please explain.

   (c) You want to predict Walmart's sales using retail traffic activities. You just learned that one can count the number of cars in Walmart's parking lots using satellite data. Clearly, the car counts are a very noisy measure of the actual level of retail traffic. If you run a linear regression of sales on car counts, how would the noise affect the results?

2. **R^2 of OLS.**

   (a) Consider a simple linear regression $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$. Show that the R^2 statistic of this regression equals to the squared value of the correlation between x and y,

   $$R^2 = \widehat{\mathrm{Corr}}^2(x, y),$$

   where

   $$\widehat{\mathrm{Corr}}(x, y) = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}.$$

(b) Now let's extend the result to multiple linear regression. Show that

$$R^2 = \widehat{\mathrm{Corr}}^2(\widehat{y}, y),$$

where $\widehat{y}_i = \mathbf{x}_i'\widehat{\beta}$ is the predicted value for y_i.

3. **Irrelevant features.**

(a) Prove that the following claim from the lecture notes is true: "*When irrelevant variables are included, the LS estimator is still unbiased.*" Specifically, suppose the true DGP is:

$$Y = X_1\beta_1 + \varepsilon,$$

but we are instead estimating:

$$Y = X_1\beta_1 + X_2\beta_2 + \varepsilon.$$

Show that $E[\hat{\beta}_1|\mathbf{X}] = \beta_1$.

(b) Explain why the inclusion of irrelevant feature X_2 could reduce the precision of LS estimate $\hat{\beta}_1$.

4. **Market timing through sector rotation.**

The MSCI Cyclical and Defensive Sectors Indexes are designed to track the performance of global cyclical and defensive companies across various Global Industry Classification Standard (GICS) sectors.[1] Currently, it classifies 7 GICS sectors as "Cyclical sectors" and 4 as "Defensive sectors" (see https://tinyurl.com/rqxncj3 for details). The indexes go through quarterly reviews (without clearly defined criteria), which can lead to modifications of the components.

In class, we discussed how the market model can be used to measure the cyclicality of different industry portfolios. Our goal in this exercise is to (i) devise a transparent and data-driven methodology to construct our "cyclical" and "defensive" sector indices; (ii) build a sector rotation trading strategy.

(a) From Ken French's website (https://goo.gl/TrqEYX), download (i) monthly market excess returns and one-month Treasury bill rates, and (ii) monthly returns of the 30 industry portfolios. We will define the "full sample" as the period 01/1971 to 12/2020.

(b) Use the full sample, fit the market model to each of the 30 industry portfolios. Plot the data and fitted regression line (as in the lecture notes) for the top 5 and bottom 5 industries ranked by their market betas. In addition, plot the estimated $\hat{\alpha}$ against $\hat{\beta}$ for the 30 industries. Comment on your findings.

[1] For more on GICS, see https://tinyurl.com/yyh7ofkd

(c) We would like to form a "cyclical sector index" as an equal-weight portfolio of the top 5 industries by market beta, and a "defensive sector index" as an equal-weight portfolio of the bottom 5 industries. The realtime index construction cannot use any data from the future. Here is what we need to do to avoid any "look-ahead" bias:

  i. Starting in 01/1981, at the beginning of each month, use data from the previous 5 years (this is called "rolling-window") to estimate the industry betas and then update the two indices according to the new market beta ranking. (If the components of the top and bottom industries change, you would be rebalancing the portfolio by selling the existing positions on some industries and buying stocks of new industries.)

  ii. Report the average excess return, market beta, alpha, Sharpe ratio, information ratio (using market portfolio as benchmark), and maximum drawdown for the two portfolios. Comment on your findings.

(d) Recall the market timing strategy we built in Q4 of Pset1. Assume that we have a mandate to be fully invested in stocks (no cash holding) at all times. Use the model in Q4(e), design a portfolio that is invested in the two sector indices constructed above and rebalanced at the end of each month, where the weights on the cyclical and defensive indices depend on the expected market excess return in each month. Compare your portfolio's performance with the one from Pset1 for the period of 01/2014 to 12/2019.