Problem Set 7

Instructions: Work in a group of 1-3 people. Each group hands in one electronic copy of their answers. Try to make your answers readable. Be brief and to the point, but be sure to explain your logic. Do not print data, entire spreadsheets, or programs – instead, copy the relevant statistics to a table. All tables and charts should have legends and explanations. Answer text (excluding tables and figures) should be typed and maximum six pages long. Exceeding these limits will draw a penalty.

This problem set relates to initial tests of market efficiency that focuses largely on the relation between past returns and expected returns, seasonality in returns, and the dangers and cautions of data mining. In order to proceed you need Microsoft Excel and file "Problem_Set7.xls". This file contains returns on 75 portfolios sorted on size and various lengths of past returns (1-month, 12-month, and 60-month past returns). In addition, the file contains monthly returns to the Fama and French factors, including a proxy for the excess return on the market portfolio, RMRF, plus the T-bill returns.

a) Estimate the January mean return for each of the 75 size-and-past-return based portfolios. Compare January mean return to mean return February through December. The easiest way to do this is to run a regression of the excess (of the T-bill return) portfolio return on an intercept and a January dummy variable (e.g., a variable which equals 1 if the month is January and zero otherwise). The dummy variable coefficient gives the January specific "extra" mean return. Is there a strong January seasonal that is different than other months? For which portfolios is this largest or smallest? Why might this be?

b) Do the same as above for December by adding a December dummy variable as well to the regression (and still include the January dummy) and then compare the mean return in December to the mean return February through November (which is the intercept) to the mean return in January for the portfolios. Is there a strong December and January seasonal that is different than other months and different from each other? For which portfolios are these seasonalities largest or smallest? Can you conjecture why these differences exist?

c) Are these findings evidence *for* or *against* market efficiency? How would you try to reconcile the above findings with the efficient markets hypothesis? How are the findings perhaps a violation of efficient markets? (hint: consider statistical explanations, transaction costs, the model we have assumed, etc.)

d) Now, let's look at the relation between past returns and expected returns. Run a series of cross-sectional regressions in the style of Fama and MacBeth (1973) where the left-hand-side of the regression is the return of each of the 75 portfolios in month t in excess of the risk-free rate in month t on the following right-hand-side variables:

i) The past one-month return of each portfolio (e.g., return on the portfolio in month *t-1*)

ii) The past 12-month return of each portfolio, skipping the most recent month (e.g., cumulative return on the portfolio from month *t-2* to *t-12*)

iii) The past 60-month return of each portfolio, skipping the most recent year (e.g., cumulative return on the portfolio from month *t-13* to *t-60*)

Run this cross-section regression every month, save the coefficients, and then report the time-series average of the coefficients over all months as well as the time-series *t*-statistics on those coefficients in the style of Fama and MacBeth (1973). Comment on the coefficients on the lagged returns of each portfolio. What pattern of predictability do they indicate? What is the relation between past returns and average returns? Does this violate some form of market efficiency? Why might it? What else could explain these patterns?

e) Now, let's put together all of the information. Create a strategy that best exploits all of the facts found above. For example, this might entail going long small stocks and past 12-month losing stocks in January, and short large stocks and past 12-month winning stocks in January, doing the reverse in Dec., doing the opposite for 60-month winners and losers, etc. Try to maximize the Sharpe ratio of your portfolio based on the information you learned from the regressions above. Once you have created this strategy, compute its time-series mean, standard deviation, and Sharpe ratio.

f) Compute the confidence interval around the Shape ratio you calculated in e). The Sharpe ratio is approximately distributed under a normal distribution:

$$SR \sim N\left(\frac{\mu}{\sigma}, \frac{1}{T-1}\left(1+\frac{\mu^2}{2\sigma^2}\right)\right)$$

The Sharpe ratio is approximately distributed under more general conditions:

$$SR \sim N\left(SR, \frac{1}{T-1}\left(1+\frac{SR^2}{4}\left[\frac{\mu_4}{\sigma^4}-1\right]-SR\frac{\mu_3}{\sigma^3}\right)\right)$$

$$\mu_k = E[(r_t-\overline{r})^k]$$

$$SR \sim N\left(SR, \frac{1}{T-1}\left(1+\frac{SR^2}{4}\left[kurt-1\right]-SR * skew\right)\right)$$

Please compute the confidence intervals of your Sharpe ratio under both the normal and more general distributions.

g) Now let's generate some random strategies. We have 75 size-past-return-based portfolios. Each month, randomly select a portfolio from the 75 size-past-return portfolios, and use its return as our random strategy return. (To do this, generate a random integer between 1 and 75, and use the portfolio corresponding to the random integer.) Do this each month and you have generated a time-series of returns to a random strategy.

- Then, repeat this process of producing a random strategy 5,000 times so that you get 5,000 time-series of different random strategies.
- A useful function in Matlab is `R = ceil(75.*rand(N,M))` where R equals an NxM matrix of random numbers between 1 and 75.
- Compute the mean, standard deviation, and Sharpe ratio for each of these random strategies.

- What fraction of these random strategies exhibit returns that are statistically different from zero?
- Plot the 5,000 *t*-stats of the mean returns to these strategies in a histogram and comment on the number of *t*-stats greater than 2 or less than –2. For the strategies that produced *t*-stats greater than 2 or less than –2, does this mean we have found a strategy that delivers performance?

- What fraction of these random strategies exhibit Sharpe ratios that are larger than the strategy you created in part e)?
- What fraction of these random strategies exhibit Sharpe ratios that are larger than the upper part of the confidence intervals you calculated in f)?

- **Bonus question:** Under the normal distribution, how probable are the max and min returns to your strategy in part e) (e.g., what is the probability you would see a return at least as high as the max or as low as the min) assuming
  a. the backtested Sharpe ratio was the true Sharpe ratio?
  b. the full range of confidence interval around the Sharpe ratio is possible?

What danger does your answer to these last set of questions highlight in empirical work?