Financial Markets - 15.433 – Assignment #3
Professor Charles Hadlock

You should complete this assignment in a group of 3-6 students. The group should email a single copy of the assignment to (sloan15433ta@gmail.com) by **Tuesday, October 6th at 9:30 a.m.**. These exercises all require computer programming and you are free to use any language of your choice (Python, R, Matlab, Stata, or other). Your assignment should include a single pdf document that contains a summary of your tabulations from the computer/statistics work related to all of the exercises. You should then also include a document that includes your programming work and associated output. All questions related to the assignment should be directed to the course TAs. Please list all student names as they appear on Canvas along with your MIT ID numbers on the first page of your PDF submission.

All t-statistics, p-values and statistical tests should use Newey-West standard errors with a lag of six to account for possible serial correlation in the time-series.

Exercise #1

The dataset assignment3a.xls contains data on the returns to the three Fama-French factors at the monthly level in percentage units (i.e., 2.36 represents 2.36%).

(a) Estimate at the monthly level the relation between last month's market excess return and this month's excess return (MKTRF is market excess return during the indicated month). What does the coefficient tell you about the economic size of this relation? What does the t-statistic tell you about the statistical significance of this relation? Does it look like past returns explain a large fraction of the variation of future returns at the monthly level?

(b) Characterize market returns each month as either "high" or "low" by whether MKTRF in the indicated month is above or below the historical average over all sample months. Estimate at the monthly level whether high returns one month predict high returns the next month. Present this evidence using both (i) a regression-based approach, and (ii) a simpler test based on the binomial distribution.

(c) Estimate at the annual level the relation between last year's market excess return and this year's excess return. Estimate this relation first using only calendar year windows (i.e., do annual returns from January 1 to December 31 of one year predict returns in the same Jan. 1 to Dec. 31 window in the subsequent year) and also using rolling annual windows (i.e., every month consider whether the annual returns from the past 12 month predict the annual return in the next 12 months).

Exercise #2

The dataset assignment3a.xls also contains a column with Prof. Robert Shiller's start-of-month Price-Earnings ratio (shillerpe) for the market as a whole (called the cyclically-adjusted total return PE or CAPE). Use the reciprocal of this variable as an indicator of the earnings-yield for the market (E/P).

(a) Estimate at the monthly level the relation between a given month's market excess return and the market's start of month earnings-yield. What is the p-value on this relation?

(b) Estimate at the annual level the relation between a given year's excess return and the start-of-year earnings-yield. Use only calendar year windows (i.e., January to December) in predicting excess returns.

(c) Returning to the monthly level relation in part (a) of this problem, calculate the out-of-sample predictability of earnings-yield for monthly returns using the out of sample R_{os}^2 outlined in the lecture7b slides. Does it appear that the earnings-yield has a high level of out-of-sample predictability?

(d) At the monthly level, calculate the percentage of months in which the regression predicted return using only past data and earnings yield is closer to the actual return than prediction just based on the past historical mean. Does the evidence here agree with your evidence in part (c)? [Note: your calculations in (d) should use the same figures as in part (c). Instead of calculating the out-of-sample R_{os}^2 using a formula, you are being asked to look at the fraction of the times that the regression predicted return is closer in magnitude to the realized return than the distance between the simple historical average predicted return and the realized return.]

Exercise #3

The dataset assignment3b.xls contains data on the returns in decimal format of the value-weighted CRSP index (a value-weighted measure of the overall market return) at the daily level starting in late 2010.

(a) For each trading day from June 1, 2011 onwards, calculate the SMA of daily return volatility using the past 50 trading days in your calculation.

(b) For each trading day from June 1, 2011 onwards, calculate the EWMA of daily return volatility using a lambda value of 0.8.

(c) For each trading day from June 1, 2011 onwards, calculate the EWMA of daily return volatility using a lambda value of 0.94.

(d) Graph all three data series from parts a/b/c of this problem on a single graph and calculate the simple correlation between each pair of measures (again, only for period June 1, 2011 and onwards).

(e) Calculate the RMSE of all three volatility measures/estimates from parts a/b/c. Based on this evidence, which estimator seems to best predict volatility?

Exercise #4

The dataset assignment3c.xls contains data on the bid ask-spread for Amazon stock and trading prices for every transaction that occurred on October 1 of 2010. The data items are as follows: "smbl" is the stock trading symbol, "time" is the time of the quote or transaction measured in nanoseconds from 12:00 a.m. on October 1, 2010, "bid" is the national best bid price (NBB) at the indicated time, "ask" is the national best offer price (NBO), "shares" is how many shares are traded in the transaction (if a transaction took place), "buysell" is whether the reported trade is initiated by the buyer or seller, and "price" is the transaction price per share in dollars (again, if a transaction tookplace). The file reflects the fact that quotes are continually updated over time and whenever a transaction takes place the most recent quote is entered as the best quote that was available at the time the trade took place.

(a) For buyer-initiated trades (i.e., when buysell is coded as "B"), calculate the volume-weighted average transaction cost for buyers. Specifically, calculate the transaction cost for each buyer-initiated trade, and then compute a weighted average of these transaction costs for each stock, using the numbers of shares traded as weights. The transaction cost is computed as a fraction of the benchmark price, where the benchmark price is the midpoint of the national best bid and offer prices at the time of the transaction. For example, if the national best bid and offer prices are $10.00 and $10.20, respectively, and a transaction of 200 shares occurs at $10.20, the transaction cost for this trade as a proportion of the benchmark price is (10.20 – 10.10)/10.10 = 0.99%, or 99 bps.

(b) For seller-initiated trades, calculate the volume-weighted average transaction cost.

(c) Does the data indicate any relation between trade size and transactions cost? Justify your answer with relevant calculations.

(d) The graph below represents the relation between buyer-initiated volume-weighted transactions costs and the natural logarithm of a firm's start-of-day market capitalization for 117 stocks on October 1 of 2010. Amazon is one data point out of 117 data points on this graph. Does it appear that transactions costs vary by market capitalization?