**Algorithmic Trading – Fall 2017**

**Final Group Project**

**Due: December 11, 2017**

**Group Project Data:**

* Each Group is assigned a different 50 stock sample.
* Data for each stock is in the file: “Final Project Data\_Fall 2017.xlsx.”
* Groups will need to extract the data for their 50 stocks.
* Each problem will require students to map together data from different Excel sheets. A large part of algorithmic trading is working with large data sets and also mapping and merging data from different sources into the format needed for analysis.
* Data collection is often one of the most time consuming tasks of algorithmic trading.
* Each group will have different results based on their set of stocks.
* Stock Data is in different sheets in the Excel file.
* Students will need to extract the data for their group and map the data together.
* Students in each group will need to work together to address each problem below.
* The results from each group will be different, and groups are encouraged to collaborate on the best way to solve each problem.
* Excel VBA Macro files have been posted to BB during the semester. These macros provide examples of how to solve different problems using Excel VBA (e.g., running multiple regressions via a loop, solving a non-linear optimization, sampling with and without replacement, constructing a risk model, and portfolio optimization with TCA).

**Final Project Questions:**

The final project consists of six (6) problems each with multiple parts. Details of each question is provided below.

#1 – Construct a Market Impact Model

#2 - Calculate Risk Model using the Fama French Three Factor Model

#3 – Construct a Daily Volume Forecasting Model

#4 – Perform Portfolio Optimization with Market Impact Cost

#5 – Develop a Pre-Trade Model in Excel

#6 – Perform Post Trade Analysis

**Data Description**

All data for the final project is in the file “Final Project Data\_Fall 2017.xlsx.” The sheets in the file have the following data:

**Project Groups:**

* Students in each group, and the list of 50 stocks for each group.

**Daily Stock Data:**

* Daily Stock Data for all stocks for the period 7/1/2016 through 9/29/2017.
* The daily stock data consists of Date, Stock, Open, Close, Volume.

**Stock Data:**

* Current Stock Data as of 9/29/2017. The data consists of Symbol, Current Price, Shares Outstanding, Est. 1-Year Return.

**Market Impact Data:**

* The data used to calibrate the market impact model. The data consists of trades over the period 9/1/2016 through 8/31/2017, including Date, Stock, Trade Imbalance, and Avg Execution Price. The data includes the Trade Imbalance (e.g., Share Quantity) for each stock for a one-year period. There are 252 data points for each stock. Each group will have 1260 data points to construct the market impact model (50 stocks \* 252 days = 12600 data points).

**TCA Data:**

* TCA Post Trade Data used in the Post-Trade Analysis.
* The Data consists of one month of trades for the period 9/1/2017 through 9/30/2017.
* The data includes Date, Symbol, Side, Shares, Broker, Period Volume, Avg Exec Price, and Period VWAP.
* There are 20 trades to analyze for each symbol. Each group will evaluate performance for 1000 trades (50 stocks \* 20 trades = 1000 data points)

**Fama French Factors:**

* Data used to construct the Risk Model. These data are the factors for the Fama French Three Factors Model are provided for the period 9/1/2016-8/31/2017. These factors are expressed as a decimal (e.g., 0.10=10%) to be consistent with the returns which are to be expressed in decimals. The factors are: Mkt-Rf, SMB, HML, and Rf.

**Pre-Trade Single Stock:**

* Example layout for the Single Stock Pre-Trade Model.

**Pre-Trade Portfolio:**

* Example layout for the Pre-Trade Portfolio Analysis.

**Post-Trade Analysis**:

* Example layout for the Post-Trade Analysis report.

**Final Project Problems:**

**Problem #1 – Construct a Market Impact Model**

Each group is provided with end of day trade imbalances for each of their 50 stocks over a 1-year period. In total, there are 252 data points for each stock, and 12,600 data points in total that will be used to calibrate the market impact model.

The actual market impact parameters for each group will be different.

Data for this problem is provided on the Market Impact Data sheet, Daily Stock Data sheet, and Stock Data sheet. Students will need to map and merge the appropriate data together to solve the problem.

**1.1 Estimate the market impact parameters for the market impact model below. This is a non-linear model with four (4) factors and (6) parameters using the stock data for your group.**

* You will need to estimate the parameters using non-linear regression techniques
* We can use non-linear OLS in this case because the error term of the model is symmetric.
* In general, analysts will need to perform a full and proper evaluation of the model error to determine if non-linear OLS or MLE is the best approach to use to solve for the model parameters.

**Definitions:**

= market impact for stock k and day t

= order size for stock k on day t

= Volatility for stock k on day t

= Natural Log of Market Cap (in millions) for stock k on day t

= POV rate (percentage of volume) for stock k on day t

= stock k

= day t

**Calculations:**

Market Impact Cost

Size:

Order size expressed as a percentage of average daily volume.

We need to use the absolute value of the Trade Imbalance.

ADV (Average Daily Volume)

Moving average of the stock’s average daily historical period of time. We will use 30 days.

The ADV is computed using a simple 30-day moving average. For example, the ADV for Stock A on 9/1/2016 is computed as the average of Stock A’s Daily Volume over the period 7/21/2016 through 8/31/2016. This is the 30-day period prior to 9/1/2016. The ADV will need to be continuously updated.

Volatility

Standard deviation of daily log price returns computed over a 30-day period annualized.

The volatility measure is a moving average and need to be updated each day.

For example, the volatility on 9/1/2016 is computed using prices over the period 7/20/2016 through 8/31/2016.

The natural log of the stock’s market cap expressed in millions of dollars.

The LnMktCap is a moving average using the stock’s current shares outstanding (see Sock Data sheet) and the previous day’s closing price.

The order’s percentage of volume.

For the Market Impact calculation, the percentage of volume is computed using the actual volume on the day.

We need to use the absolute value of the Trade Imbalance.

**1.2 Calculate the standard error of each parameter value using sampling techniques.**

* The standard error of each parameter will help us determine if the factor (variable) used in the model is a significant predictor of market impact.
* When sampling, select about 10% of the data observations, then estimate the MI parameter values.
* Perform about 200-300 simulations using 10% of the data points in each simulation to estimate the parameter values. The standard deviation of these 200-300 parameter estimates are the standard error of the parameter.
* Provide the results of the sampling simulation in an Excel spreadsheet. There should be one column for each parameter value, and a row for each calculated value.
* An Excel VBA spreadsheet was uploaded to BB to provide insight into the Market Impact non-linear sampling technique.
* Are all the factors in the model significant explanatory factors of market impact?

**1.3 If any of the factors were found to be “not significant” then construct a new market impact model without that factor.**

* For example, if the LnMktCap factor was found to be “not significant” based on the sampling technique and the calculated standard error of parameter a5, then formulate a new market impact model, estimate the parameter values, and repeat the sampling technique to determine the standard errors. This model without LnMktCap is:
* Please note that some groups will have factors that are not significant predictor variables for market impact.
* After estimating the parameters of the market impact model, this model and model parameters will be used for the remainder of the problems.

**Problem #2 - Calculate Risk Model using the Fama French Three Factor Model**

* Calculate a Covariance Matrix (Risk Model) using the Fama French Three Factor Model over the period 9/1/2016 through 8/31/2017.
* Price data for each stock is in the sheet titled “Daily Stock Data.”
* Fama-French Factors is in the sheet titled Fama French Factors.
* There are 252 days in the period 9/1/2016 through 8/31/2017.

**2.1 Estimate regression model parameters values for each stock.**

The regression model is as follows:

* Compute log returns for each stock, e.g.,
* Subtract the risk free rate from the calculated returns before running the regression model.
* The risk free rate for each day is included as a column in the Fama French Factors sheet.
* Please note that we can run the regression on actual returns and actual factors, or we can run the regression on excess returns and excess factors as we did in class.
* Excess return subtracts out the mean return – rf. The mean excess R-rf in this case will have mean = 0.
* Excess factor value substracts out the mean factor value. The mean excess factor will have mean = 0.
* Since we are not using the constant term in the derivation of the Risk Matrix, we can use either data set to estimate the Fama French beta’s for each stock.
* Record all regression results in a table.
* The rows will be the stocks (50 rows) and the columns will be the following values/statistics: .
* The regression can be run from Excel or MATALB or Python.
* Please note that Excel’s “=linest()” function returns the Standard Deviation of the Regression and not the MSE. So you will need to square this value to calculation the MSE. Excel’s Regression Add-In returns both values so you will need to determine which value you are taking.
* Matlab’s regression functions regress() and regstat() return the MSE so you will not need to square this value if working in MATLAB.

**Factor Model:**

**Risk Model (derived from Factor Model):**

Where,

Where,

Where,

**Important Notes on Computation of Covariance Matrix:**

* Be sure that you are using the correct B and B’ to compute C. C needs to be a 50 x 50 matrix.
* Be sure that you are using the mean square error for the diagonal matrix above and not the standard deviation of the regression if using Excel. This will give incorrect results throughout the final project.
* Different software packages provide different statistical results, e.g., mse or stdev of regression.
* Excel’s =linest() function provides the standard deviation of the regression error SEy so in this case we need to square the value.
* Excel’s regression analysis in the Regression Add-In provides both the standard deviation of the regression error and also the mean square error of the regression.
* Matlab’s regression function “regress()” and “regstats()” returns the mean square error “mse.”
* You need to be sure you are selecting the correct value for the diagonal term when using this Regression Function. An incorrect diagonal term will dramatically change the covariance matrix and cause all subsequent problems to be incorrect.
* To check that your matrix is correct and you have the correct diagonal term, check to make sure diagonal of the covariance matrix C is equal to the variance of returns for each stock. There may be some slight difference.

**Problem #2 – What to Answer and/or Provide:**

* 1. Calculate and provide the table of Regression Results for each of the 50 stocks.
  2. Provide the Covariance Matrix for your fifty (50) stocks. Be sure to annualize the covariance my multiplying by 250. Refer to this Covariance Matrix as C.
  3. Calculate a second covariance matrix for your 50 stocks using only price returns data, e.g., log returns. This is the sample covariance matrix computed using rice returns, e.g., variance and covariance across all stocks. Be sure to annualize the covariance my multiplying by 250. Refer to this Matrix as C\_Sample.
  4. Provide both the C matrix and C\_Sample matrix in an Excel sheet.
  5. Compare the “C” matrix to “C\_Sample” matrix. How do these matrices differ? Are any of the covariances the opposite sign? E.g., positive covariance in one matrix and a negative covariance in the other matrix?
  6. You will use the C matrix (computed from the factor model) for all other final project questions.

**Problem #3 – Construct a Daily Volume Forecasting Model**

* Construct a daily volume forecasting model for your group of stocks.
* Daily volume data is provided in the sheet titled “Daily Stock Data.”
* Use data from 9/1/2016 through 8/31/2017 for your volume forecast.

**3.1 Construct a daily volume forecasting model for each of your 50 stocks using the daily forecasting technique from our class lecture and text.**

This model has form:

This model is a variation of an ARMA model.

= denotes the number n of historical days used in the mean or median calculation.

The model includes:

* Historical measure which could be the Average or Median
* Day of Week Adjustment factor
* Previous Error Term (there is evidence of serial correlation with market volumes)

Your group will need to:

* Compute the Day of Week Factor for all stocks for the period 9/1/2016 through 8/31/2017.
* Determine the best measure to use for , e.g., average or median.

Avg:

Median:

* The forecast error at each time period t is:
* Compute the value showing the persistence of the volume error.

Provide in an Excel file:

* The Day of Week Factor for each day.
* You can compute an average Day of Week Factor for all stocks. That is, we do not need a day of week factor adjustment for each stock.
* The forecast error parameter value, , for each stock.
  + Does the forecast error parameter value different for each stock?
  + Would you suggest using a stock specific parameter value or using an average forecast error parameter across all stocks? Why?

Hint:

* This problem requires more work than it may initially appear.
* There are statistical techniques that can be used to compute the value
* For simplicity, we can set the very first forecasting error term to be zero, and only look at the previous error term (this is a little different than time series analysis and ARMA modeling and is fine for our purposes).
* That is, set on 8/31/2016 = 0. Then, we can run a regression of the lagged error term to determine the value. E.g.,
* You will need to determine the best methodology to use, such an Average or Median.

**3.2 Forecast Daily Volume for each of your 50 stocks for each day in September 2017.**

Forecast the daily volume for each of your 50 stocks for all 20 days in Sept 2017.

There were twenty (20) trading days in Sept 2017, 9/1/2017 through 9/29/2017.

Sept 4, 2017 was the Labor Day Holiday.

The forecast on September 1, 2017 will incorporate the forecast error on August 31, 2017 (and the estimated value).

The forecast error on September 5, 2017 (the next trading day in Sept 2017) will not include the forecast error term from September 1, 2017. Why? Because when forecasting, we expect this error term to be equal to zero, that is, we expect the forecast to be accurate since we have .

The daily volume forecast equation for 9/1/2017 is:

Where,

is the forecast error on 8/31/2017.

The daily volume forecast equation for 9/5/2017 through 9/29/2017 is:

**3.2 What to provide:**

* Provide the results of your daily volume forecast in an Excel Spreadsheet.
* The spreadsheet will have 50 rows, 1 row for each stock, and 20 columns, one column for each day.

**3.3 Extra Credit :**

* Formulate the Regression Equation that can be used to calculate the value of the forecast error term where the for all 0 stocks is the same.
* Solve the First Order Conditions of Regression.
* This solution can be handwritten and provided as a pdf file or as a picture.

**Problem #4 – Perform Portfolio Optimization with Market Impact Cost**

* Perform Portfolio Optimization to determine the optimal set of weights for your fifty (5) stocks.
* Compute the trading cost for each stock and for the basket if you were to trade these fifty stocks.
* You will perform the analysis for one scenarios:
  + $1 Billion dollar portfolio

The data for this problem is from:

* Market Impact Model – Problem #1
* Covariance Matrix – Problem #2
* Estimated Returns – in sheet titled “Stock Data”
* Price – in sheet titled “Stock Data”
* Volatility – data in sheet “Daily Stock Data.” Compute volatility using Log Returns. Be sure to sure to annualize volatility by multiplying by sqrt(250). Use the Volatility Estimate for each stock as of 9/1/2017. That is, use the 30 days of price returns ended on 8/31/2017.
* ADV (Average Daily Volume) – data in sheet “Daily Stock Data.” Compute a 30 day average daily volume (ADV) for each of your 50 stocks as of 9/1/2017. This calculation will use 30 days of data ended on 8/31/2017.
* LnMktCap – If your market impact model includes the LnMktCap factor, calculate the LnMktCap on 9/1/2017 using Shares Outstanding and Current Price (as of 9/1/2017) in the Stock Data sheet.

**4.1 Traditional QP Optimization with Market Impact Analysis**

Traditional QP mean-variance determines the optimal set of weights that satisfies the objective function and all constraints.

Market impact cost, however, is a function of Shares not weight. For example, a $100 million dollar portfolio will have a different number of shares to buy or sell than a a $500 million dollar portfolio even if the weight of each stock in both portfolios are exactly the same.

Using traditional Mean-Variance optimization, determine the optimal portfolio weights that will minimize portfolio risk with a specified return or 10%.

Use the following constraints:

* Lower Bound = -0.05
* Upper Bound = +0.05
* This portfolio can be two sided and consist of long and short positions.
* This optimization can be solved in Excel using the Solver Add-In or via MATLAB.

The optimization can be formulated as show below:

(Minimize Portfolio Variance Risk)

Subject to:

(Total Return)

(Total Weight)

(Lower Bounds of -5% and Upper Bounds of +5% for each stock)

What is the minimum risk for this portfolio?

We will use this risk value in problem 4.2 below.

For a $100 Million Dollar Portfolio:

Determine the number of Shares to purchase for each of the 50 stocks if you are investing $100 million dollars. You will need the optimal stock weight and price to determine the number of shares.

Compute the Market Impact cost in basis points and total dollars for each stock.

Compute the Market Impact cost in basis points and total dollars for the entire basket.

What is the expected Net Return for this Portfolio? e.g., the net return is the expected return minus the market impact cost.

For a $500 Million Dollar Portfolio:

Determine the number of Shares to purchase for each of the 50 stocks if you are investing $500 million dollars. You will need the optimal stock weight and price to determine the number of shares.

Compute the Market Impact cost in basis points and total dollars for each stock.

Compute the Market Impact cost in basis points and total dollars for the entire basket.

What is the expected Net Return for this Portfolio? e.g., the net return is the expected return minus the market impact cost.

**4.2 Portfolio Optimization with Market Impact Cost**

This question will help use determine if we can do better incorporating market impact cost directly into the portfolio optimization process than we can if we only perform traditional portfolio optimization and consider market impact costs after determining the optimal portfolio weights.

Portfolio Optimization with Market Impact will provide both the optimal weights and the optimal number of shares for each stock based on the investment dollar value.

Market impact cost is dependent upon Shares and Dollars invested, we will need to compute different optimal portfolio for each portfolio value. Thus, we will need to run this optimization twice: $100 million dollar investment and $500 million dollar investment.

To accomplish this, we need to specify the portfolio optimization objective function to maximize the net return of the portfolio, e.g., expected return minus market impact cost.

This is:

(Net Return)

Subject to:

(Variance from QP Optimization above, e.g. )

(Total Weight)

(Bounds)

/1000

We will need to solve the above optimization for two portfolios, first with a Dollar Value of $100 million and second with a Dollar value of $500 million.

The result from this optimization will also provide the portfolio weights and portfolio shares.

What is the portfolio risk for each portfolio?

What is the expected return for each portfolio (e.g., weights \* estimated 1-year return)?

What is the net return for each portfolio

**4.3 Compare results for the two portfolios for both traditional QP portfolio optimization and Portfolio Optimization with Market Impact cost.**

* You will now have four (4) optimal portfolios.
* Two (2) portfolios A1&B1 for the $100 million investment and two (2) portfolios A2&B2 for the $500 million investment.
* Denote these portfolios as:
  + A1: Traditional and $100 million
  + A2: Traditional and $500 million
  + B1: Traditional and $100 million
  + B2: Traditional and $500 million
* The risk should be exactly the same for portfolio A1 and B1 and for A2 and B2. If not, there is an error and you will need to redo the steps above.
* For each of these portfolios, provide a table with Stock Weight, Est. Return, Market Impact Cost, and Net Return.
* For each scenario, provide a table with Est. Return, Portfolio Risk, Est. Portfolio Net Return.

Questions to Answer:

What optimization technique provides higher expected net return?

Which portfolio would you select for the $100 million dollar investment and why?

Which portfolio would you select for the $500 million dollar investment and why?

Is the Net Return for A1 higher or lower than B1?

Is the Net Return for A2 higher or lower than B2?

Compare the Weights across all 50 stocks for the $100M and $500M scenarios.

Do these results contradict traditional financial theory in any way? That is, can we start with a seemingly inefficient portfolio and achieve higher expected net returns?

**4.4 Extra Credit:**

* Is there a single efficient frontier as previously described in portfolio management literature in classes and journals, or is the efficient frontier now dependent upon another variable such as portfolio dollar value?
* This extra credit question is much more difficult that it appears, and is a large research area for portfolio managers and investment funds.

**Problem #5 – Develop a Pre-Trade Model in Excel**

**Develop a Working Pre-Trade model for Single Stock Trades and for a Basket of Trades in EXCEL.**

Example Layouts of these pre-trade analyses are shown in the sheets titled “Pre-Trade Single Stock” and “Pre-Trade Portfolio.”

You will be required to use the results your market impact model and model parameters from Problem #1 and market data as of 9/1/2017. E.g., Price, ADV, Volatility, and LnMktCap (if applicable) as of 9/1/2017.

This problem consists of:

* Develop a Working Single Stock Pre-Trade Model (that will analyze any or your 50 stocks). A sample layout is shown in the sheet titled “Pre-Trade Single Stock.” This analysis will also include four(4) graphs.
* Develop a Working Pre-Trade Model for a Basket of Stock (using your 50 stocks). A sample layout is shown in the sheet titled “Pre-Trade Portfolio.”
* You will need to include all columns and fields shown in the Excel spreadsheets.

These models will compute MI & TR for the user entered order and portfolios.

These models will have to provide updated results and calculations if the user changes the Shares and/or POV rate.

For the Single Stock Pre-Trade:

* The single stock pre-trade model will need to work for all stocks in the group. That is, the user will select the stock to evaluate and all data fields and calculations need to update based on the user specified selection.
* This single stock pre-trade will also need to show the four (4) graphs shown on the sample layout sheet “Pre-Trade Single Stock.”

For the Basket Pre-Trade:

* Use the 50 stocks from your group and use the Share Quantities and Side from the $500 million portfolio.
* Please determine the 1-day liquidation cost (market impact cost) of the entire basket of stock under current market conditions.
* How many days would you have to trade if you wanted to reduce the 1-day market impact cost by 50%.
* Please determine the expected liquidation cost (market impact cost) of the entire basket of stock if the volatility increases 2x times for all stocks and if the ADV decreases by 50% for each stock. This is known as a “sensitivity” or “what-if” analysis.
* How many days would you have to trade to reduce the liquidation cost from the “sensitivity” analysis to be equal to the current liquidation cost, e.g., the liquidation cost under current market conditions.

**Problem #6 – Perform Post Trade Analysis**

**Develop a Working Post-Trade Analysis in EXCEL.**

An example layout of the post trade analysis is provided in the sheet titled “Post-Trade Analysis.”

You will evaluate trades for you group of stock.

There are 20 trades for each stock and 50 stocks for each group. Thus, in total, there are 1000 orders that will be analyzed.

The TCA Post Trade Data is provided in the sheet titled “TCA Post Trade Data.”

You will need to map the TCA Post Trade Data with the Stock Prices on the sheet “Daily Stock Data.”

You will also need to compute updated ADV and Volatility for each stock through September.

Additionally, if your market impact model includes LnMktCap as a factor, you will need to update the LnMktCap factor through September.

**Post-Trade Analysis Variables:**

* Weight = Traded Dollar weight for the group or stock
* AvgExecPrice = Avg Price of the trade or group. For groups, it is computed as a weighted average.
* Size (%ADV) = weighted average Order Size
* Volatility = weighted volatility for the group
* POV = weighted interval POV rate
* Time = weighted Trade Time (expressed as a percentage of the day)
* Arrival Cost = actual cost of the order (weighted average for category summaries)
* Est. Cost = estimated cost of the order using your market impact model from problem #1 (weighted average for category summaries)
* VWAP PnL = VWAP performance measure (weighted average for summary categories)
* Value-Add = broker value add measure (weighted average for summary categories)

**Additional calculations:**

**Actual Cost:**

For this analysis only, we take the arrival price to be the open price on the trade day.

In reality, the arrival price is measured as the mid-point of the bid-ask spread at the time the order was released to the market. If the order begins trading at the open, then the arrival price is the opening price on the day.

**Order Size**

The ADV measure is the average daily volume computed over the previous 30 trading days. This value needs to be updated for each day in September 2017.

**Interval POV**

The interval POV rate is calculated as the actual POV rate of the order based on the actual market volume over the trading period. Periods Volume is computed from tick data.

**VWAP Benchmark Analysis**

The interval VWAP price performance metric. Provides an evaluation of how the order executed in comparison with the interval VWAP price. In this problem, the interval price is provided. In reality, the interval price is computed as a volume weighted average price over the trading horizon from tick data.

**Est. Market Impact**

If your market impact model includes the LnMktCap factor, you will need to include this factor in the equation above. You will use the model function and model parameters from Problem #1.

If your market impact model includes the LnMktCap factor, you will need to include this factor in the equation above.

**Est. Timing Risk**

**Value-Add:**

Measure of the Brokers Performance compared to the estimated trading cost.

**Z-Score:**

Risk Adjusted Measure of the Brokers Performance and Value-Add

**Summary Categories**

Summaries are to be provided Side, Broker, and for all individual trades.

All summary calculations are to be weighted calculations.

**Questions:**

Who is the best performing broker?

How would you rank and rate Broker performance for each broker A, B, C, D, and E?