**Programming Case Study**

**Introduction:**

The focus of our business is algorithmic trading of large institutional orders for equities, FX, and futures. Our algorithms and our research depend on market data. This case study includes calculations made from market data.

**Data:**

The data files contain every single trade that happened for the stock referenced in the file name. There are 10 individual stock symbols named in the 10 files, and each file should be considered in your analysis.

Timestamp: Milliseconds since Epoch in UTC

Bid : Highest bid for the stock at the time of the trade

Ask : Lowest offer for the stock at the time of the trade

Nbb\_agg\_size : Total Bid size for the stock at the time of the trade

Nbo\_agg\_size: best Offer size for the stock at the time of the trade

Trade\_price: price the trade occurred at

Trade\_size: Size of the trade

Volume: Cumulative volume since the beginning of day at the time of the trade

Vwap: Volume weighted average price since the beginning of the day at the time of the trade.

The data is sorted by timestamp and irregularly spaced in time. Assume that between two data points for any timestamp, the price-level data (bid, ask, Nbb\_agg\_size, Nbo\_agg\_size etc) are the same as the previous timestamp. (A new entry appears at any time there is a change to the bid, ask, or bid size or ask size.)

For all of the questions below, throw away observations before 9:30 and after 16:00 eastern standard time, since the US standard trading hours are from 9:30am to 4:00pm.

The midpoint price (“mid price”) is calculated as the average of bid and ask at any point in time.

**Task:**

Please do your best to answer the questions below using R or Python. We expect this to take no more than 1-2 hours for a strong candidate. Our goal is to understand how you think through programming exercises.

1. Create a visualization to show the following in 15-minute intervals for each stock in the included data -
   1. Average spread (Ask - Bid),
   2. Average total bin volume as a percent of the full day’s volume
   3. Volume-weighted average price (VWAP)

Generate the data first, then create the best visualization to illustrate it. Be sure your code can generate the data (though we do not need to see the data, only the code and the visualization).

1. Write the most efficient code you can to build a cross sectional linear regression model. You will need to start by creating the following variables.
   1. The response variable will be:

* Normalized mid price return from the time of trade to the midpoint price at 10 seconds after the time of a trade.
  + Formula: ( (Mid (t+10) – Mid (t)) / Mid(t) ) / (Ask (t) – Bid(t))
  + Hint: Mid (t+10) is the prevailing midpoint price at time t+10. This would be the current price at the time 10 seconds after the trade.
  + Throw away observations where spread (ask price – bid price) is zero or negative.
  1. Predictors are:
* Bid Offer Imbalance, calculated as

(Nbb\_agg\_size – Nbo\_agg\_size )/(Nbb\_agg\_size + Nbo\_agg\_size)

* Trade sign:
  + 1 if the last trade price is above the average of bid and ask price
  + -1 if the last trade price is below the average of bid and ask price
* Trade Size imbalance over the last 10 seconds: Trade size imbalance for any period can be calculated as:

Sum ( Trade Sign x Trade Size )

Write your code as efficiently as you can and explain it as well as you can using comments.

1. Create the model above from the data you created. Write your comments on the model results.

**Bonus:** How can the model be improved?

1. Calculate the 15-second volatility (standard deviation of mid price returns) for SHOO in each 30 minute interval. Present the volatility in annualized terms.

**Submission:**

Submit your answers to the questions above via email to [kathryn@bestexresearch.com](mailto:kathryn@bestexresearch.com) with the subject “Quant Strategist R/Python Assessment”, formatted in an attached PDF document (so it can be opened without your unique coding software). Keep your document as brief as possible while still conveying the answers to each question. Include code and any necessary discussion as comments.

You will be asked to present and discuss your results during formal interviews following this assessment.