**MTH9898 Assignment A**

Due Feb 29th.2016

**STUDENT: Yue Wang**

Overall Summary:

This document outlines the methodology and toolbox being used to tackle the assignment. Please note many avenues being used in this implementation is a collaborative effort from discussion with classmates.

Initial data understanding:

I used python to visualize the data pattern for both price and volume of the small data input, (apply a non-parallel fashion) to statistically test some potential methodologies of recognizing the pattern and separating the noise (given the smaller testing input set represents the larger testing data set). The traditional three sigma deviation from price and volume applied, the underlying assumption is to leverage the central limit theorem, the i.i.d. price/volume series’ means will converge to a normal distribution, which implies the possibility of tail events happens should diminish as it moves far away from mean. By applying this methodology, the 10K, 100K, 1000K both applied, it turned out a terrible filter as almost 50% data input being identified as noise for 100k and 1000k. By plotting things out, we found the above mentioned techniques serve terrible purpose, so the assumption of normality is forgone. From the plot, we do see price jump frequently and spike in a short period of time, which testify the poor performance of central limit theorem. Intuitively, the volume could vary a lot from time to time, so it doesn’t really make sense from a practical perspective to apply central limit theorem. For volume, a

Methodology: