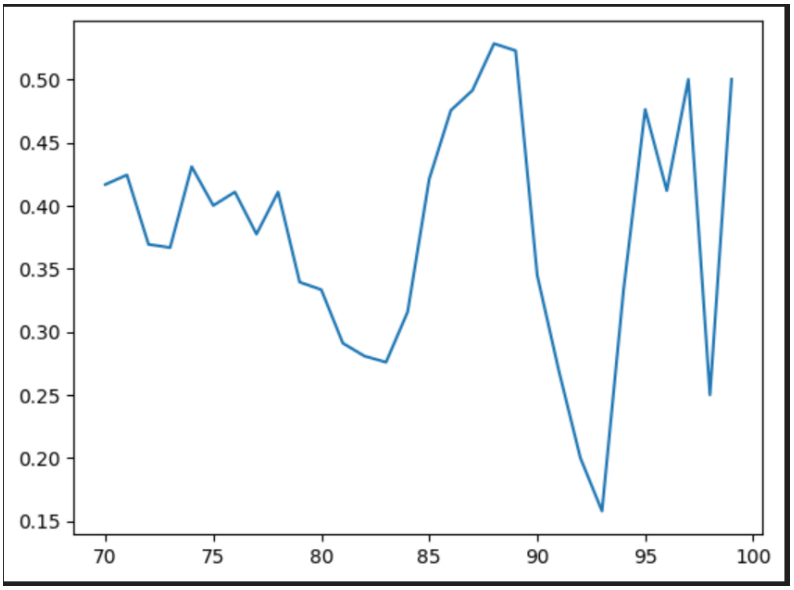
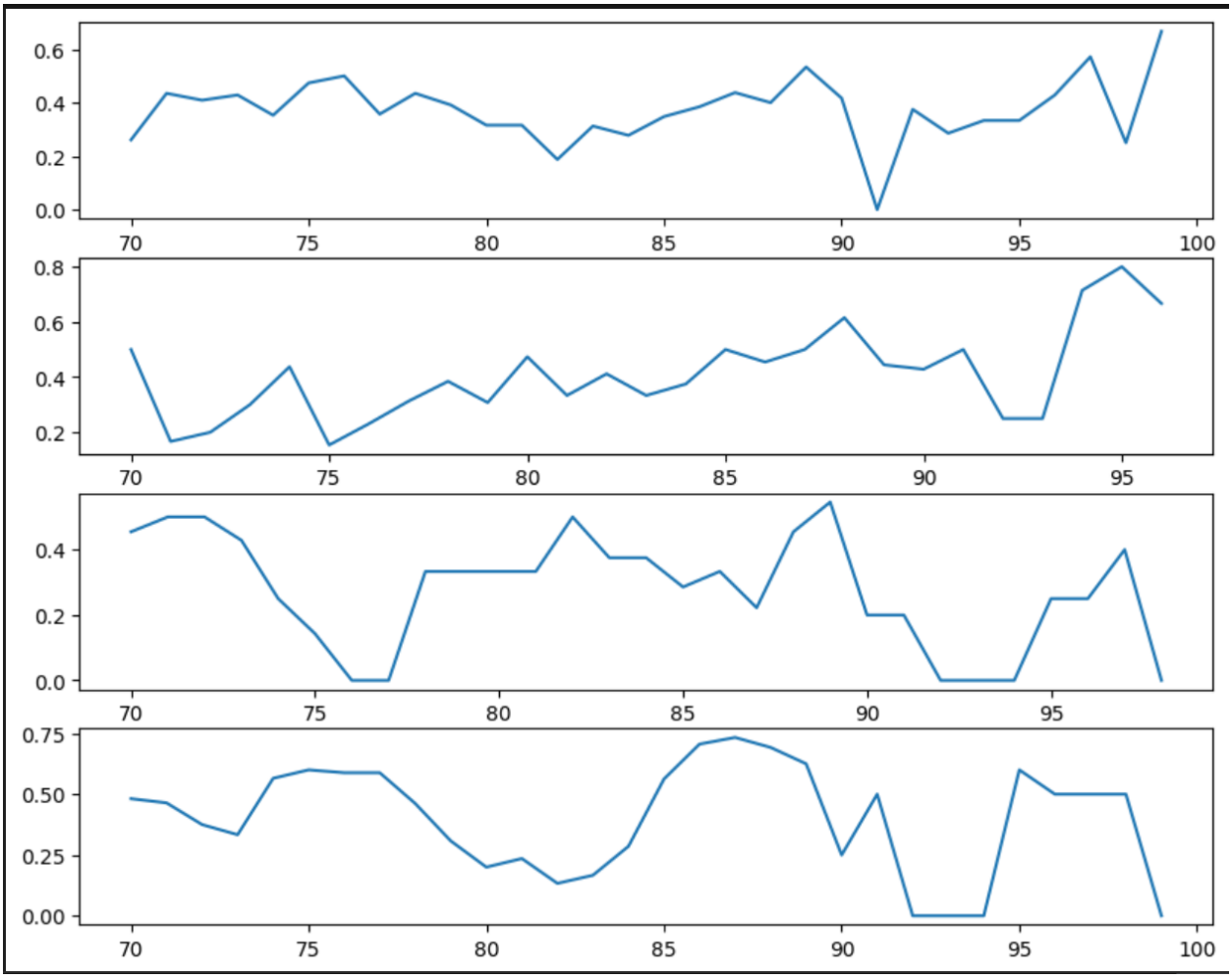
Baseline Alarm Testing:

Testing was done using the month of data pulled to test the baseline with the initial 95th quantile threshold value. We define an alarm as a data point whose rolling standard deviation is greater than the threshold value. Flags were then defined as the first alarm in a sequence of alarms. False alarms were then defined as occurrences of flags which had the vol drop back below 90% of the threshold value within the next 30 minutes. The purpose of the testing was to classify false alarms using a reasonable metric and get a sense for how many false alarms can be expected using the baseline detector, as well as to choose a suitable threshold value which tends to have minimal false alarms.

For each quantile in the range 70-99%, the false alarm rate over the month was measured. i.e. classified the months data into alarms and not alarms and calculated the proportion of flags which were false alarms. The results were plotted and can be seen below



The x-axis is the threshold value used, represented as a quantile of the original rolling standard deviations. The y-axis is the false alarm rate that was measured over the month using the given threshold value. The data shows the false alarm rate tends to stay between 25 and 50 percent but can drop as low as 17 percent for thresholds in the 90-95 percentile range.



Above is the same type of graph but the False Alarm Rate (y axis) is calibrated over each of the four weeks of the month represented in the initial data. There are some thresholds where the baseline is fairly predictive even at lower values, but those periods are not as significant of volatility spikes and are thus less important events. There is clearly a pattern in the data that for the 92-93% range very few false alarms are measured.

Conclusion:

Based on the testing done and using the under 90% of threshold in 30 min definition of false alarm, I think a threshold value of about 0.000775 is where the baseline is most predictive. Note: this value corresponds to the 93rd percentile of the data used but should be kept as a constant for classification of future data. Note based on the results, we should expect a false alarm rate of about 15-20% using the baseline detector.