# Concluding Remarks

## Main Findings

Before answering the two central research questions, this paper analyzes crude oil returns and shows that the series of returns only exhibits weak intertemporal correlation. As a result, it is tough to forecast crude oil returns using historical returns only. In contrast, the series of news sentiments shows significant intertemporal correlation and predictable patterns. Altogether the significant relationship between these two time series, it is reasonable to conjecture that one can forecast crude oil returns by using sentiments as a bridge.

In the experiment section, this paper firstly examines the market efficiency hypothesis (MEH) on the crude oil market with respect to different combinations of information sets, predictive models, and searching technologies. In general, the market shows efficiency (i.e., unpredictability) with respect to simple models such as moving average predictors and ARIMA models. Still, returns are predictable using more sophisticated models such as LSTM-RNNs. As for the second research question, it turns out that news sentiments help predict returns only in one single case in which LSTM-RNN is used. In most cases, incorporating news sentiments even worsen models’ test time performances. This is not a surprising result. Firstly, most of news of day $t$ arrive before the market is close, and if the price adjusts based on news quickly and accurately, the close spot price of day $t$ has already reflected a great percentage of news on day $t$. At the beginning of day $t+1$ when we gather news on day $t$ (the information flow of day $t$) and predict $r\_{t+1}$ (equivalently, predict $p\_{t+1}$), only the small portion of news that arrive after market closes on day $t$ is the novel content in the information flow. Consequently, adding daily information flow and news sentiments is too slow for the fast-evolving market and does not help much on forecasting daily returns. Secondly, recall that the models presented in this paper use information from the past 31 trading days to forecast the return on the upcoming trading day. In total, there are 385 sentiment features but only 31 lagged values of returns used to generate one prediction. It could be the case that there are far too many sentiment features than necessary so that these noisy predictors obscure any information from historical returns. These two conjectures explain why adding news sentiments lead to an even worse test-time performance.

## Limitations and Extensions

Even though this paper answers its own research questions, the analysis in this paper can still be extended in many ways. At the beginning of this paper, we define a market to be efficient with respect to an information set, a class of models and a searching technology if no trading strategy based on the optimal model can attain positive (expected) economic profit. Throughout this paper, we claim the market to be inefficient whenever we found the market is predictable using some models. However, accurate prediction can only guarantee profit only if transaction costs are zero, which is a strong assumption. Incorporating a non-negligible transaction cost could flip some of conclusions in this paper: the market could be efficient even if it is predictable. More robust conclusions can be drawn by using back-testing techniques. After choosing an optimal predictive model, we can design some trading strategies based on model’s prediction. Then the back-testing system simulates real world trading based on the designed trading strategy and calculates the profit. Gaining positive profit in multiple rounds of simulation provides stronger evidence on market inefficiency than just showing there exists models can attain accuracy higher than 50%.

Another limitation is that this paper uses intra-day high-frequency trading data and conclude the market can react to news quickly so that adding news sentiments does not produce advantage to forecasting daily returns. Experimental results in this paper only imply that market’s reaction time is faster than 24 hours, further analysis using intraday oil returns can produce more meaningful results. Potentially, one can build a profitable forecasting system for the crude oil market on an intraday frequency.

The characteristic function is an important component in the proposed framework. In this paper, characteristic functions only summarize information flows using simple statistics such as mean and standard deviation of sentiment scores. It turns out that too many features are generated, and these constructed sentiment features are too noisy. As a result, adding these sentiment data does not help improve models’ performances. Future works can include more sophisticated and meaningful technical indicators built on news sentiments such as the moving average of daily average sentiments. Informative content, if any, within news sentiments can be fully captured using only a few technical indicators and help handle the curse of dimensionality.

To analyze the usefulness of news sentiments, this paper looks at two information sets only (historical returns v.s. both historical returns and news sentiments). Including other macroeconomic indicators as additional features improve performances of models implemented in this paper. For example, crude oil and gold are substitute as commodities for investments. As a result, changes in gold prices would affect crude oil prices and hence contain useful information to forecast crude oil returns. It is worthy to devote more time to design models incorporating other exogenous features.

The performance of a time series forecasting model depends on whether it can successfully grasp two key factors: intertemporal correlations and non-linearities. This paper examines ARIMA models which can capture intertemporal correlation but not non-linearities, random forests and support vector machines that can model non-linear patterns but not intertemporal correlation, and LSTM-RNNs capturing both factors. It turns out that only LSTM-RNNs attain a satisfactory and potentially profitable result on the test set, which suggest that both factors are necessary for a successful forecasting system. However, LSTM-RNN is not the only model handle two factors at the same time. Other models such as CNN-RNN can capture these two factors as well and preforms much better on high dimensional problems.