

Gordon Lu

STAT 1223

Professor Nelson

21 February 2021

STAT 1223: Assignment 1

In the fast-paced, ever evolving world, buzzwords like “cryptocurrency”, “investing”, “money” are attractive. Typically, with 9-5 jobs occupying most days, finding a leisure that brings wealth with minimal effort. Thus, the notion of investing in the stock market. In high-frequency trading firms, many complex algorithms are designed to optimally buy and sell stocks. For an average Joe, time is a sparse resource and spending hours upon hours on end trying to make profit from investing in the stock market while working is stressful. However, blindly investing into big name stocks is also dangerous. Holistically, at the base level, the response variable that I wish to learn more about is “buy or sell stock”, and how this is correlated with a stock’s previous price and other factors. What this paper seeks to do is not to describe trends in the stock market, but rather, given a certain stock, what determines how well a stock performs, and when should someone sell a stock versus buying the stock.

The literature surrounding stock price prediction is quite dense, and typically do not follow a typical regression model. Rather, the general consensus regarding stock price prediction is to formulate a linear time series. In an article written by Salim Lahmiri (2018), Lahmiri’s objective was to combine two popular approaches, being Singular Spectrum Analysis forecasting, which is used heavily in finance and Support Vector Regression (SVR), which is commonly used in time series analysis (Lahmiri, 2018). With both being deeply rooted in

statistics, Lahmiri furthermore performed optimization techniques to find the optimal SVR initial parameters, and applied the hybrid forecasting model to a set of stocks, and compared the results to benchmark models, which included: “FFN trained with WT coefficients (WT-FFNN), polynomial regression (PolyReg), naïve model, and the classical ARMA process” (Lahmiri, p. 445). What Lahmiri found through comparing the four models to the hybrid model was that it “achieved the lowest MAE, MAPE, and RMSE for all time series used in the study” (Lahmiri, p. 450). This article is fascinating in the sense that it provides intuition that employing a typical time series model may yield sub-optimal results in comparison to using a hybrid approach. It is important to note that there may be problems with overfitting and potential points that may pull the regression up. However, in developing an appropriate model to use to forecast stock prices, this article is a good steppingstone, as it not only provides insight on how to compare and benchmark models, but also approaches and decisions to use in deciding a hybrid approach.

On the other hand, in an article jointly written by Seungwoo Jeon, Bonghee Hong, and Victor Chang, the approach is vastly different. In the article, the authors employ a Dynamic Time Warping algorithm to find patterns between two temporal sequences, then use Stepwise Regression to select the determinants most impacted by stock price and lastly generate an artificial neural network (Jeon, Hong, Chang, 2018). What Jeon, Hong, and Chang concluded was for short-term prediction to use a combination of dynamic time warping, stepwise regression and artificial neural network to help gather similar datasets for each stock and predict daily stock prices (Jeon, Hong, Chang, p.185). The article in itself is very dense, it dives deep into theory, and is ambitious enough to dive into the realm of deep learning. This article is useful in the sense that it employs common regression techniques such as Stepwise Regression, and also utilizes an algorithm to find similar datasets. With this, in developing an optimal model so as to yield

feasible VIF values and finding good predictors, using Stepwise Regression, and perhaps even using a neural network to train the model and generate good predictions aid in developing an optimal model for forecasting stocks.

With the articles written so far, there are already predictors in mind, however Bruce James Vanstone, Adrian Gepp and Geoff Harris jointly worked on the idea of considering a commonly overlooked confounding variable in determining stock prices, news. What Vanstone, Gepp and Harris proposed was to build two Neural Network Autoregressive (NNAR) models, one being the base line, and the other being built on the counts of news articles and twitter posts, and to compare the predictive accuracy of the two models (Vanstone, Gepp, Harris, p. 3819). What Vanstone, Gepp and Harris found is that news and twitter sentiment counts do indeed play a role in stock price modelling. While the article is deeply rooted in financial econometric theory and machine learning, there is a lot of useful information. The trend that news and twitter sentiment counts do have a significant relationship to stock pricing, add to the long list of predictors in the model to develop to determine when to buy or sell a given stock.

Among all three articles presented, it is apparent that some degree of time-series was applied. In particular, in Vanstone, Gepp and Harris' article, a NNAR model, based on the marriage of a neural network and autoregressive model from time series analysis was used. With Jeon, Hong and Chang's article, a Dynamic Time Warping algorithm, which involved searching and identifying similar temporal datasets, and is commonly used in time series analysis. With the article written by Lahmiri, a hybrid model of SSA and SVR was used, both of which are deeply rooted in time series analysis. However, where the articles diverge is the models. It is clear that each of the goals of the models presented is to determine the efficiency and accuracy of predicting the prices of stocks. With each model presented the overall consensus is that the

individual models do provide some degree of leverage over typical rudimentary regression models. What can be conclusively said is that in formulating a model that utilizes a binary response, consolidating the predictors such as a variable reflecting how the price of a given stock changes over time, and considering exogenous variables such as the news and twitter sentiment counts are necessary as to avoid introducing confounding variables into the models. All the articles jointly provide evidence that perhaps forming a regression by considering models rooted in time series and utilizing methods such as Stepwise Selection to decide on relevant predictors are necessary. Additionally, a step to further elevate the regression would be to utilize machine learning and generate a neural network to predict whether a stock would be worth buying or selling.

References

Jeon, S., Hong, B., & Chang, V. (2018). Pattern graph tracking-based stock price prediction using big data. *Future Generation Computer Systems*, 80, 171–187.

<https://doi.org/10.1016/j.future.2017.02.010>

Lahmiri, S. (2018). Minute-ahead stock price forecasting based on singular spectrum analysis and support vector regression. *Applied Mathematics and Computation*, 320, 444–451.

<https://doi.org/10.1016/j.amc.2017.09.049>

Vanstone, B., Gepp, A., & Harris, G. (2019). Do news and sentiment play a role in stock price prediction? *Applied Intelligence (Dordrecht, Netherlands)*, 49(11), 3815–3820.

<https://doi.org/10.1007/s10489-019-01458-9>