**DS-670-Capstone: Big Data & Business Analytics**

**Assignment 11: Discussion and Conclusion**

**April 8th 2017, Madhumita D**

**Discussion:**

Traditional statistical methods for prediction have their limitations in the applications with data sets which show nonlinearities. Many forecasting techniques are capable only of selecting general trends like positive and negative trends, and show complexity in modelling cycles which by no means are repetitive in period, amplitude, or shape. Even though such inadequacies are present, modelling techniques such as multiple linear regression is used routinely in the Capital Markets and also have proved to be a very useful tool.

In this study we have showed that using a simple neural learning technique such as the back propagation algorithm proves to outperform the routine practice in typical stock ranking applications with the framework of a pricing model. With their smooth interpolation properties neural networks allow models to fit better to the data and generalize significantly better.

We have used a hybrid approach of neural networks and time series analysis to calculate the return predictability. Based on the overall forecasted values we can say that the expected return of stocks in last quarter of 2014 has a negative trend while they show a positive trend in the first two quarters of 2015. However, later in third and fourth quarters in 2015 we see a similar negative trend as shown in last quarter of 2014. On the whole the averages depict a huge volatility in stock price movements. In comparison to our approach the traditional linear regression analysis was performed on stock market data and based on the regression metrics used they find evidence that return predictability fluctuates over time and is highly governed by changing market conditions. It is also found that during a market crash, no return predictability is prominently evident, mostly because of the extreme degree of associated uncertainty. They also state that inflation, risk-free rates and stock market volatility influence stock market predictability over a period of time.

We have assumed that the price for each stock is highly influenced by the fundamental, technical and macro-economic factors while the regression analysis was done under the assumption that the market follows a weak-form of efficient market hypothesis (EMH). This assumption was made to test for evidence of AMH (adaptive market hypothesis) in return predictability. The hybrid approach followed in this study helped to test for semi-strong form of efficient market hypothesis (EMH) based on return predictability. The returns were computed from the pricing information for each stock and predicted for the future time using the neural networks and time series analysis.

Regression analysis uses Dow Jones Industrial Average (DJIA) as a technical indicator. The index is a price-weighted average of 30 blue-chip stocks, accounting for 25–30% of the total value of U.S. stocks. Index created for 30 blue chip stocks (large cap stocks) for century-long U.S. stock market data from 1900 to 2009. We have used daily closing prices of exchange traded fund (ETF) of both small and large cap stocks from the year 2000 to 2017 for U.S. stock market. In our model Relative Strength Index (RSI) is used as a technical indicator. Relative Strength Index compares the magnitude of recent gains and losses over a period of time based on the change of price movements.

Neural Networks and Factor Models were implemented to forecast the returns of a stock in the market. We use a multivariate time series to study the behaviour of stocks. To weigh the use of neural networks we compare the results and metrics to Regression Analysis, which was conducted on the measures of return predictability of stocks. Many different metrics were used to compare and compute the results. In Neural Networks sequential testing of market efficiency plays a major role in computing t the returns accurately based on all the historical and present data available while sequential testing of market efficiency was not taken into consideration in linear regression.

Neural Network Weights from 15 training data sets (normalized data) are used to predict the weights in the test data sets to obtain return predictability. Time series analysis was performed on the weights from the neural networks to analyze if there are any deviations in the predictions. Ranked the test data sets based on predicted returns and grouped these data sets into 5 buckets. These groups are used to study the behaviour of stocks. The metrics used in our algorithm consider the non-linearity of the data. Neural network models are best suited for this kind of data where the model uses a learning algorithm and modelling is based on the data alone. The metrics used in regression analysis are AVR, AQ and generalized spectral test. While automatic variance ratio (AVR) is used in testing weak-form of efficiency of financial markets, automatic portmanteau test (AQ) is used for testing the auto-correlation of returns which are subject to unknown forms of heteroskedasticity. These two tests are used for the linear dependence only. Hence, a generalized spectral test is done where both linear and non-linear dependence is tested. All these metrics are used as inputs to regression analysis which does not completely cover the non-linearity of the financial markets.

The return predictability from regression analysis depends on market conditions which contradicts their initial assumption that the market is weak form of efficient market. Their analysis addresses the fact that return predictability has been smaller during economic bubbles than in normal times. Also, return predictability is associated with stock market volatility and economic fundamentals. From adaptive market hypothesis they have deduced that U.S. Stock markets are susceptible to market conditions such as inflation, crashes and are weak form of efficient markets.

Our analysis addresses that the factors chosen drive the performance of the model but also the Investability of our model with real-time data. This entails simulation of a portfolio controlling for risk and liquidity of the assets traded. We showed that predicted returns are dependent on the publically available information or market conditions which prove that markets are semi-strong form efficient and the U.S. Stock Markets are highly governed by changing market condition but are semi-strong form of efficient markets.

**Conclusion:**

The hybrid approach combining Neural Networks and Time-Series used in this study was able to categorize between the stocks that performed well and stocks which performed poorly in the market. This approach also outperformed the multiple linear regression analysis approach overall. Investors who use this modelling technology should be able to improve their analysis choosing the best investment alternatives based on better predictions. However, further research is required in using this hybrid approach in stock prediction analysis. To further validate the proposed technique in investment analysis, we need to perform qualitative and quantitative tests on our data, considering different time periods and different industries or sectors. With the many advantages, investors and researchers should seriously consider the application of neural network and time-series techniques to investment analysis.