**Cross-Currency Strategy Based on Historical Data**

**Cyril de Lavergne**

**Abstract**

We compare the performance of each currency pair predictions to the daily expected returns at t+1 in the forex market of G7 currencies. The model also shows the combined performance of the forex strategy. The methods used in this paper has been properly designed to use parallel programming to compute diverse time series technics such as Arima Garch and Arima Monte Carlo. These time series predictions in addition to technical factors are essentially fed to machine learning algorithms to predict the expected returns at t+1. The combined machine learning algorithms predictions will eventually make a forex trade when reaching a certain threshold. The results per currency differs greatly. In average, the total strategy appears to be profitable.

**Keywords:** Forex Market, Financial time series analysis, Arima Garch, Arima Monte Carlo, Data Mining, Machine Learning, Extra Trees, Random Forest, Trading Strategy

**1 Previous Work**

Mike Halls-Moore from Quantstart team [(Halls-Moore M., 2015)](#HallsMoore) has used Arima Garch time series to the S&P500 index with skew generalized error distribution. He shows that significant profit can be made compared to a simple buy and hold strategy especially when higher volatility is present in the market. Another study is Arima Monte Carlo [(Perez I., 2016)](#perez) carried out on a pool of US stocks that generated significant profit. The study however does not incorporate the fact that some stocks cannot be short sell. In Forex markets, we do not have this inconvenience as one can short any currency pair. In Kaggle’s data science popular website, during a competition, 1st runner-up [(Shahbaz N. and Mohammed C., 2017)](#shahbaz) states ‘We found that Extra Trees and Ridge models were the best fit for this dataset due to the nature of the data and the time constraint. Financial data are highly noisy and unstructured, and we believe for super noisy datasets, using solid basic model to capture the super weak signal are more applicable’.

**2. Characteristics of the dataset**

Data was taken from Interactive Brokers, a popular broker in the financial industry. Since Forex markets is opened 24 hours, equal interval timestamp should be created to use daily data. As such 8pm Hong Kong time was our reference in this paper.

Both Arima garch and monte carlo used 500 historical data point. For Arima Garch model, student-t error distribution was used. In addition, two features were used for arima garch model. The average from t+1 to t+5 predictions and the predictions of t+1. For Arima Monte Carlo, a bootstrap of the error was added 1000 times to the Arima prediction and we eventually use the probability describing how many times the profit was up out of these 1000 predictions.

We then proceed to have our target data as the next period return (time t+1), and get the 5 most recent lags of each currency pair directly related to the target. For example, if EUR.USD was taken, each of currency pair containing either EUR or USD would have their lags taken. A similar operation is carried out for Arima Garch sample and Arima Monte Carlo.

**3. Overview of the system**

We first divide the training set and testing set at 60% of the total number of rows of the dataset. We follow by training the Random forest on the training dataset with 4nodes with an importance based on impurity. At the same time, an extra tree cross validation of 5samples is carried out on the training dataset. We retain the 30 most important variables for Random forest model whereas we keep only the variables in 4 out of 5 samples of cross validation for Extra Trees with a filter based on rank correlation of attributes. This allows us to capture the most important features for both models.

Then, we retrain both models with the features selected in the previous section. Note that Random forest is repeated 10 times to average predictions.

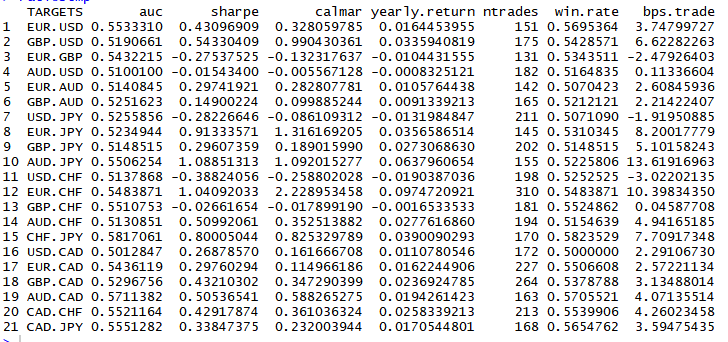
To minimize risks, we prefer to keep only the predictions that are far away from small returns. Hence, we need to extrapolate the predictions that supposes to have higher returns. To achieve so, we compute both the in-sample predictions of Extra Trees and Random Forest. This will be used in the next section to minimize risks.

After testing both models on the testing dataset, the remaining 40% of the dataset, we compare the testing predictions with in-sample predictions. Based on quantiles of these predictions, we will keep the predictions that have returns higher than 55% of the predictions. Similarly, if predictions of returns are negative, they must be lower than 55% of the predictions.

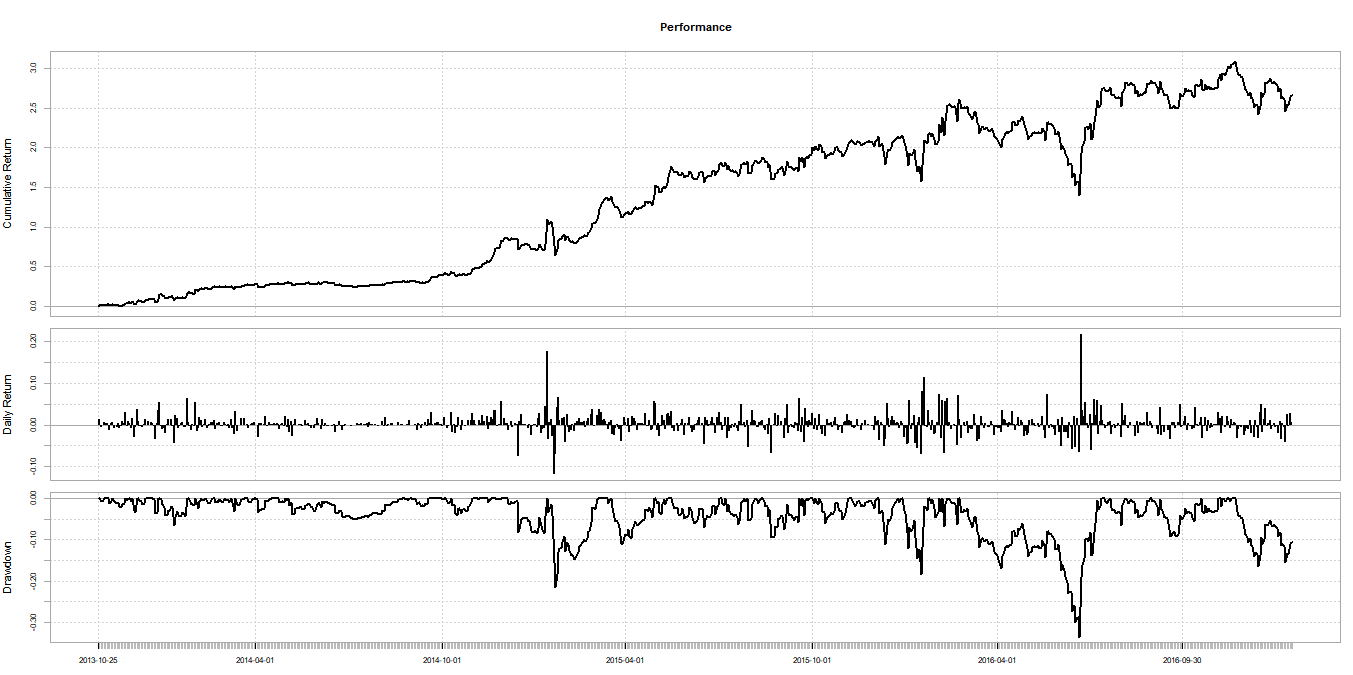
We finally compute the individual performance of each of the 21 currency pairs in addition to the combined performance of the strategy. The performance is based on diverse criteria. Area Under ROC is mainly used to describe how accurate or predictions are. Sharpe Ratio is defined as annualized returns divided by the volatility of the strategy. Calmar Ratio is defined as annualized returns divided by the maximum drawdown of the strategy. Finally, % of winning trades are shown in addition to the number of pips made per trade. Pips per trade is very important to know whether one strategy or each trade in average can be above transaction fees of the forex markets.

**4 Results**

The result per currency pair is as follows:



Note that for every currency pair, the winning rate percentage is above 50%. Only 5 of 21 currency pairs has a negative profit.

The total performance of the strategy is as follows:



The profit is indeed significant at 47% return per year. The Sharpe Ratio is above 1 meaning that the strategy is considered as acceptable to good by investors [(Maverick J. B., n.d.)](#maverick). Note that FX market in Hong Kong has a leverage of 20. Meaning that we could make as much as 940% in one year before transaction costs.

**5 Conclusion**

To conclude, we have shown that a strategy based on historical data can be used in the forex markets. The strategy remains to be improved to obtain a Sharpe Ratio above 2 to be rated as very good by investors.

**6 References**

1. Halls-Moore M., (October 7th, 2015). ARIMA+GARCH Trading Strategy on the S&P500 Stock Market Index Using R. Quantstart. Retrieved from https://www.quantstart.com/articles/ARIMA-GARCH-Trading-Strategy-on-the-SP500-Stock-Market-Index-Using-R
2. Pérez I, (18th June, 2016). MonteCarlo and Arima for stock selection. Github. Retrieved from <https://github.com/imanolperez/MonteCarlo-and-Arima-for-stock-selection/blob/master/trading.r>
3. Shahbaz N. and Mohammed C. (25th May, 2017). Two Sigma Financial Modeling Challenge, Winner's Interview. Kaggle. Retrieved from <http://blog.kaggle.com/2017/05/25/two-sigma-financial-modeling-challenge-winners-interview-2nd-place-nima-shahbazi-chahhou-mohamed/>
4. Maverick J. B. (n.d.). What is a good Sharpe ratio? Investopedia. Retrieved from <https://www.investopedia.com/ask/answers/010815/what-good-sharpe-ratio.asp>