Primordial Ooze

Andrew Tillmann

In the field of finance, many traders use different technical analysis tools to help them forecast the market. One of these tools is Japanese candlestick patterns. A candlestick pattern is formed using four price attributes of a given trading period, they are: the open, close, high and low, in a more visual depiction of price action during a single time period or series of time periods. The idea behind the Japanese candlestick trading is that there are recurring patterns related to the candlestick shape or a group of candlesticks shape that will help illuminate future price action. More specifically, there are 12 major featured patterns, including doji, gravestone doji, long-legged doji, piercing pattern, hammer, etc... Different patterns have been discovered to indicate different market situations. For example, a doji pattern might convey a sense of indecision or tug-of-war between buyers and sellers. While the morning star pattern foretells the rise of prices. Our group is proposing to use clustering techniques to investigate groups of candlesticks. With these groups we would like to see if there is any relationship between historical and future candlestick groupings. While we are not exactly using the given patterns that traders use, we will investigate the fundamental idea of recurring patterns within the candlesticks.

This research could be very helpful to traders that use candlestick trading to help validate the underlying theory of Japanese candlestick trading by discovering features of patterns. Traders may be able to exploit patterns within the markets to increase returns for themselves and their clients. In addition to trading returns, this pattern discovery can also give insights into the psychology of financial markets.

There is much literature on the herd mentality and animal spirits (emotion) that seems to have a great deal of influence on market prices. The psychology of the financial markets sometimes causes speculative malinvestment which can raise or lower prices of assets to levels that do not reflect the intrinsic value of those assets. Incorrect valuations lead to asset price bubbles that can damage retirement investment and the economy at large. If through this research, investors can have a greater knowledge of the psychology of the financial markets, the agents of the markets and the markets themselves may become more efficient in valuing assets, leading to less volatility and greater price stability.

Datasets

Candlestick trading can really be applied to any financial market. Thus, for our research we could study the effects on any of these markets. With that in mind we have chosen to look at several stock and future markets.

The explanatory variables would be at minimum, the open, high, low and close of the asset being traded. We could decide to add others or some sort of interaction later. Using these explanatory variables we will be clustering them into different groups. With these different groups we can use various regression techniques to find which group the next candlestick will most likely belong to. Thus, these groups will be our response variable.

Current Research

Study done by Stanford University engineers along the line of our idea of exploring the interconnectedness different nations economies

http://cs229.stanford.edu/proj2012/ShenJiangZhang-StockMarketForecastingusingMachineLearningAlgorithms.pdf

NYU study on using machine learning techniques to predict market prices

http://www.vatsals.com/Essays/MachineLearningTechniquesforStockPrediction.pdf

Norwegian University of Science and Technology study on machine learning and stock price prediction

http://www.diva-portal.org/smash/get/diva2:354463/FULLTEXT01.pdf

The fundamental purpose of our project is to use data analysis techniques in order to discover patterns within the historical price action of major worldwide stock market indices, then leveraging these discovered patterns to predict future price action. This could be done in an isolated fashion using only the historical data from a single index on itself. With increasing globalization and interconnectedness of the economies of different nations, we believe it may also be relevant and interesting to explore how the price action of one nation's market affects the price action of another's yet to come.

Another layer of complexity we could add to our project is exploring the relationship between the derivative and spot markets with respect to various indices. The futures market consists of contracts that lock in the current price of an asset for future delivery of the asset. Investors and companies typically use these contracts to hedge their price exposure. For example, an airline might believe the price of fuel will increase in the next 6 months. In order to alleviate or hedge this price exposure the airline will purchase a 6 month futures contract for jet fuel, locking in the current price of jet fuel and paying this price in 6 months when they require delivery of the fuel. This concept also applies to stock market indices and the futures markets for indices are typically thought to be a gauge of investors' expectations of future price movements.

For this project we have decided to all focus on different ideas relating to the financial markets and candlesticks. A candlestick is comprised of the open, high low and close of the price in a market session. We are looking to see if when these candlesticks are grouped together if they have a bias or recurring pattern. Each of us are going to be looking into separate ideas and datasets but follow this theme.

- Foreign Exchange Markets and Datasets(FOREX):

The dataset FOREX will contain these six pairs EURUSD,EURJPY,EURGBP, USDJPY, USDGBP and GBPJPY. Each of these pairs data will be looked at separately and not be mix together in an analysis. The data will consist of prices of the open, high, low, close from the Dukascopy price feed(bid side only) for each pair. The data was off the daily time period from the dates Jan 1st 2004 to Dec 31 2013 thus 10 years of data.

This data was then changed to only show the change in price that day. Thus, besides the weekend gaps the open prices will all be 0s and then the high, low and close will be adjusted off of that. This way the candlesticks are not dependent off previous candlesticks.

After performing some rudimentary histograms and scatter plots I see only noise in the data and no clear patterns or underlying themes.

The application that will be used with the data will first be clustering. The goal is to cluster the candlesticks into groups similar how traders group candlesticks together that have certain predefined characteristics. However, with the cluster techniques to be used no predefined characteristics are going to be needed. Once they are grouped together it will be investigated to see if after each group the next group that follows is distributed evenly or if there is a bias towards some groups.

- Global Financial Interconnectedness

We would like to explore the interconnectedness of the modern financial markets by analyzing the stock price and bond yield action of different global indices. The data we will be using are the S&P 500, Shanghai Composite Index(China), Nikkei 225(Japan), CAC 40(France), DAX(Germany), FTSE 100(U.K.), RTSI(Russian), ASX(Australia) and the historical sovereign bond yields of these respective countries. We would like to look at these over a 32 year period going back to 1980. The data we will use comes from two sources, YahooFinance(stocks) and Economagic.com(bonds). The bond yield data is on a monthly basis and the stock price data is daily, but can be transformed into monthly open, close, intra-day high and intra-day low attributes, giving 384 data points on the monthly scale. The first problem with the data in this form is that that theoretically interest rates and prices are both on a $[0,\infty)$ bound. In practice stock prices have an upward trend, whereas bond yields stay within a range of 0%-10% regardless of the index price. This difference may lead to earlier yields having a greater effect when the stock indices where at lower values. In order to resolve this problem we would like to look at bond yield spreads between the 10 year government bond and the 2 year government bond. Looking at this spread will capture that effects of the steepening or flattening of the yield curve, We do still need to attain the 2 year government bond data, but we are actively working with a connection at the Federal Reserve Bank of Chicago to resolve this issue.

The two approaches that we would like to take in order to analyze the data is either LDA or Multinomial Logistic Regression. We must first gather this last portion of the data and examine to see whether the data seems to ascertain the distribution of the data. Once this is known then we will make a distributional drive choice on which method would be most appropriate. They may both be useful to examine the data using LDA to look at up-month and down-month classes, as well as separating those classes further by country. The Multinomial Logistic regression could be done by separating the data into three classes up-months, down-months and unchanged months as the categorical dependent variables and the previous day's values from the other countries as the dependent variables. One issue these strategies will be limited by is that there are only 384 data points, but many attributes. This may cause us to extend the timeframe some.

http://www.economagic.com/em-cgi/data.exe/fedstl/irltlt01dem156n

http://www.economagic.com/em-cgi/data.exe/fedstl/irltlt01aum156n

- Spot Prices and Future Prices

In this part, I would compare different characters of candlesticks shown spot prices and the futures price of the underlying asset to see somehow they may have some relationships. The data set would contain two main categories - spot price and the future price of underlying asset. The data set contains four pairs S&P Stock Index and its index futures, Hang Seng Index and its index futures, Nikkei Exchange Index and its futures and DAX Index and its futures. The data will consist of prices of the open, high, low and close within period from the dates April 25, 2010 to April 25, 2014. The research would focus more on the price trends after 2008 financial crisis.

The spot price of indexes are collected from Yahoo Finance.

S&P 500 Index Historical Spot Prices (in total): Daily - Jan 3rd 1950 - Current

http://finance.yahoo.com/q/hp?s=^GSPC+Historical+Prices

Hang Seng Historical Spot Prices: Daily - Dec 31st, 1986 - Current

http://finance.yahoo.com/q/hp?s=^HSI+Historical+Prices

The Nikkei Exchange Historical Spot Prices: Daily - Jan 4th, 1984 - Current

http://finance.yahoo.com/q/hp?s=^N225&a=00&b=1&c=2998&d=03&e=13&f=2014&g=d

DAX Historical Spot Prices: Daily - Nov 26th, 1990 - Current

http://finance.yahoo.com/q/hp?s=^GDAXI+Historical+Prices

The future price of indexes are collected from Investing.com

S&P 500 Historical Futures: Daily - Oct 15th, 2009 - Current

http://www.investing.com/indices/us-spx-500-futures-historical-data

Hang Seng Historical Futures: Daily - Oct 15th, 2009 - Current

http://www.investing.com/indices/hong-kong-40-futures-historical-data

Nikkei 225 Historical Futures: Daily - Oct 15th, 2009 - Current

http://www.investing.com/indices/japan-225-futures-historical-data

DAX Historical Futures: Daily - Oct 15th, 2009 - Current

http://www.investing.com/indices/germany-30-futures-historical-data

The technique used include cluster analysis to group different signs of candlesticks (bearish signs and bullish signs, etc). The research will try to adapt three methods of cluster analysis into analysis, including k-means clustering, map-self clustering and grey clustering. They are classic ways of studying candlesticks patterns based on the research of Hus et al. (2009).
(Reference: Y,-T, Hsu et al., Profit refiner of futures trading using clustering algorithm, Expert Systems with Applications 36 (2009) pp. 6192-6198). CDA will also be discovered to use in analysis to explore the relationship between spot prices and future prices.

- Alternative vs. Legacy Energy Stocks

This analysis will explore the relationships between the alternative energy sector (solar, wind, geothermal, biofuel) and the legacy energy sector (coal, oil and gas). The time scale of the data it from 2004 to 2014 and consists of a set of stocks and indices from both energy sectors. Each dataset will consist of the open, close, high and low prices recorded on a daily basis. Data from the larger global indices (FTSE, DAX, S&P500) will also be included in the study to understand the relationship between each energy sector and the overall markets. The idea being to predict candlestick categories for each energy sector based on the daily performance of the other energy sector and the performance of the overall markets.

In order to analyze this data, clustering would be used in order to determine relationships and categorizations between the different energy types and sectors. This may help to determine whether the energy sectors candlesticks and therefore categorizations change over time. LDA could be used to determine which factors are causing the most effect in the relationships between the energy sectors.

- Dataset Name - Renewable Energy (re)
- The data was collected from the Yahoo Finance Web Site.
    - http://finance.yahoo.com/q/hp?a=00&b=01&c=2010&d=03&e=27&f=2014&g=d&s=scty&ql=1
    - http://finance.yahoo.com/q/hp?a=00&b=01&c=2010&d=03&e=27&f=2014&g=d&s=my&ql=1
    - http://finance.yahoo.com/q/hp?a=00&b=01&c=2010&d=03&e=27&f=2014&g=d&s=optt&ql=1
    - http://finance.yahoo.com/q/hp?a=00&b=01&c=2010&d=03&e=27&f=2014&g=d&s=vrnm&ql=1
    - http://finance.yahoo.com/q/hp?a=00&b=01&c=2010&d=03&e=27&f=2014&g=d&s=csiq&ql=1
    - http://finance.yahoo.com/q/hp?s=REW&a=00&b=01&c=2010&d=03&e=28&f=2014&g=d
- To the best of my knowledge, this data was collected from the Nasdaq where the stock is traded by Yahoo. It is collected on at least a daily basis for public

viewing and analysis. I do not see, nor would I expect any bias or experimental limitations.

- Dataset Name - Legacy Energy (le)
- The data was collected from the Yahoo Finance Web Site.
    - http://finance.yahoo.com/q/hp?a=&b=&c=&d=3&e=28&f=2014&g=d&s=aci&ql=1
    - http://finance.yahoo.com/q/hp?a=&b=&c=&d=3&e=28&f=2014&g=d&s=btu&ql=1
    - http://finance.yahoo.com/q/hp?a=&b=&c=&d=3&e=28&f=2014&g=d&s=bp&ql=1
    - http://finance.yahoo.com/q/hp?a=&b=&c=&d=3&e=28&f=2014&g=d&s=oxy&ql=1
    - http://finance.yahoo.com/q/hp?a=&b=&c=&d=3&e=28&f=2014&g=d&s=sto&ql=1
    - http://finance.yahoo.com/q/hp?s=VDE&a=08&b=29&c=2000&d=03&e=28&f=2014&g=d
- To the best of my knowledge, this data was collected from the Nasdaq where the stock is traded by Yahoo. It is collected on at least a daily basis for public viewing and analysis. I do not see, nor would I expect any bias or experimental limitations.

Hang Seng Historical Spot Prices

http://finance.yahoo.com/q/hp?s=^HSI+Historical+Prices

S&P 500 Index Historical Spot Prices (in total)http://finance.yahoo.com/q/hp?s=^GSPC+Historical+Prices

The Nikkei Exchange Historical Spot Prices

http://finance.yahoo.com/q/hp?s=^N225&a=00&b=1&c=2998&d=03&e=13&f=2014&g=d

FTSE 100 Historical Spot Prices

http://finance.yahoo.com/q/hp?s=^FTSE+Historical+Prices

DAX Historical Spot Prices

http://finance.yahoo.com/q/hp?s=^GDAXI+Historical+Prices

The rest of this paper will just focus on the subtopic of the FOREX dataset. The general question that was being answered by the research done from the model was if Japanese candlesticks can be used to help predict the outlook on future prices. Furthermore, the use of other related pairs and cross pairs candlesticks are used to see if they also have prediction power. The techniques being used that relates to data mining are k-means and association rules.

Given the dataset of FOREX pairs EURUSD,EURJPY,EURGBP, USDJPY, USDGBP and GBPJPY on

the daily timeframe . With the last 10 years of data were used with 625 observations used for

testing. Firstly, the data points needed to be made price independent of each other. To make a candlestick price independent each Open, High, Low and Close was offset by the close of previous market close. These independent data points were then clustered into 6 groups of candlesticks.

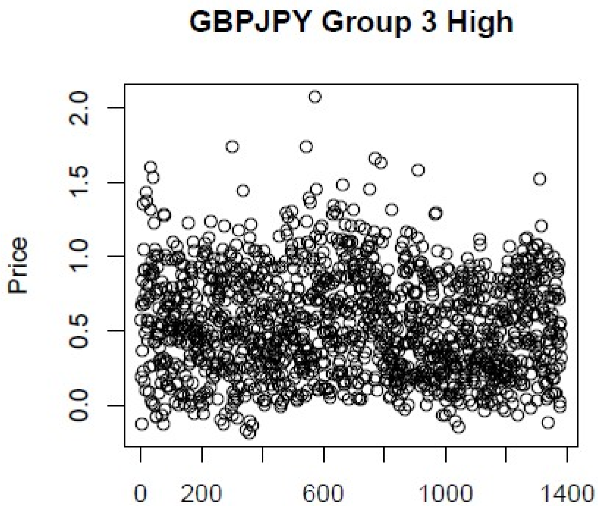
Using the offset values(current values minus yesterday's close) on open, high, low and
close as inputs into k-Means of 6 groups. Feeding these given k-Means group for the past 3 days
across all pairs at once into algorithm to build association rules to predict the next day's group
for all pairs.

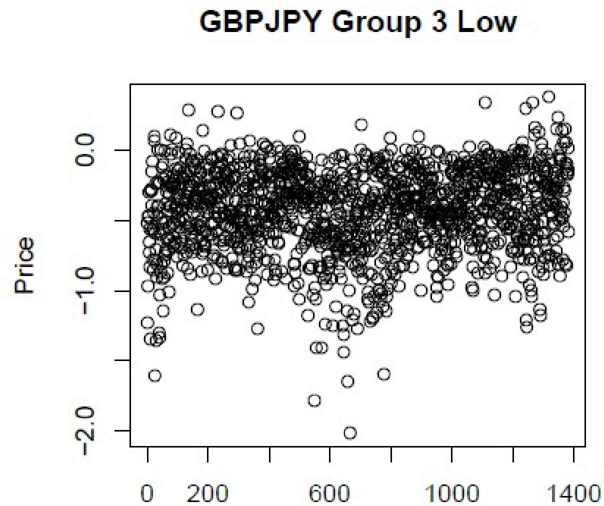
With each of these of these pairs the last three full candlesticks(N-2,N-1,N) and the next candlestick to occur (N+1) were feed into the association rules.

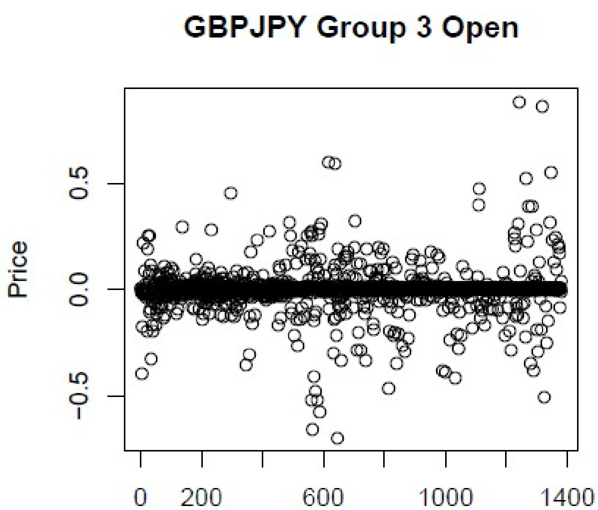
The performance metrics for the association rules would be support, confidence and lift. The support
relates to the number of times of occurrences in the data. The confidence is the percent of likelihood
the association rules feels confident of occurrence. The lift is the a ration of the targeted outcome
divided by average response. The higher the lift value the stronger the relationship of a given occurrence
over a random chosen value. The model uses the settings of 6 groups for the k-means and support =.02 and confidence=0.80 .

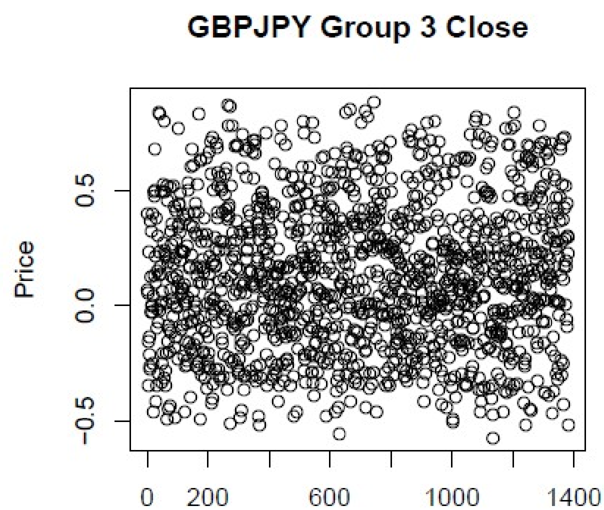
The discovery was that a few hundred rules appeared to develop within the data. This is very interesting. I was half expecting to run the program and have no rules appear on the output. However, the result only built rules for 3 of the 36 overall groups. 36 overall groups come from 6 different FOREX pairs and 6 different groups thus 6*6=36. With the association rules established it appears that candlesticks might have a least a little predicting processing is some cases. However, looking at the confusion matrix out of the 625 on the random samples there were only 53 prediction for GBPJPY, 27 prediction for USDJPY, 52 for the EURGBP, thus the prediction sample is somewhat small. This smaller sample size is due to the fact that the model didn't make predictions often.

With GBPJPY Group 3 it made 53 predictions receiving an accuracy rate of 58.49 percent. This is 14.45 percent better than random. GBPJPY seems to be a neutral group with most highs and lows within +- 100 pips and a close within -+50 pips(a pip is 1/100 of a penny).


GBPJPY Group 3 High
Price
0.0
0.5
1.0
1.5
2.0
0
200
600
1000
1400


GBPJPY Group 3 Low
Price
−2.0
−1.5
−1.0
−0.5
0.0
0
200
600
1000
1400


GBPJPY Group 3 Open
Price
−0.5
0.0
0.5
0
200
600
1000
1400

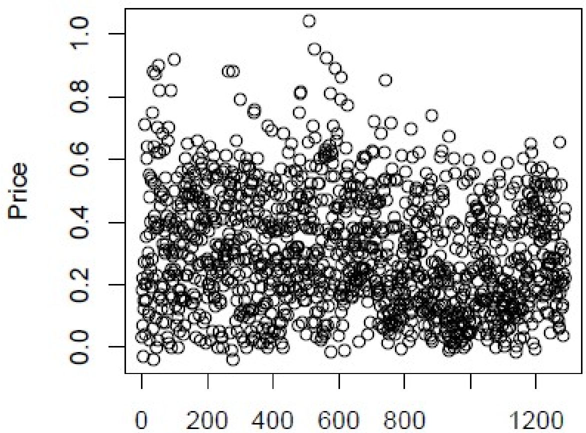

GBPJPY Group 3 Close
Price
−0.5
0.0
0.5
0
200
600
1000
1400

With USDJPY Group 5 it make 27 predictions receiving an accuracy rate of 48.14 percent. This is

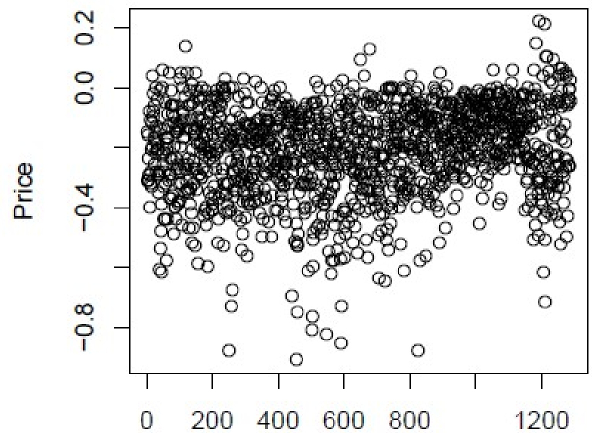
7.02 percent better than random. The USD JPY group seems to be neutral to a high. With most

highs below +60 pips and most lows above -40 pips following with a close between -20 to +40 pips.

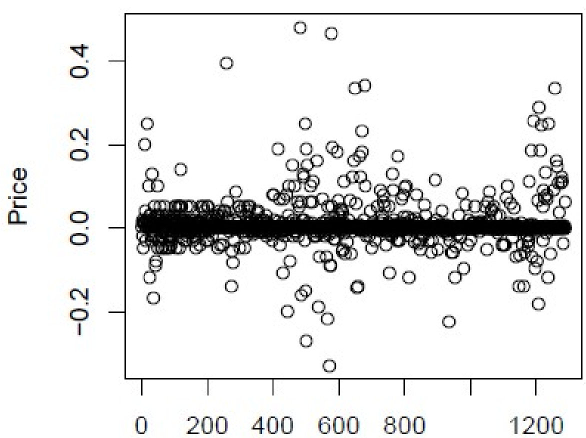
USDJPY Group 5 High



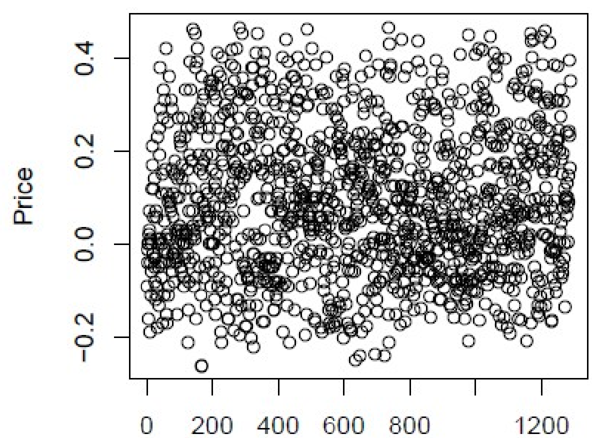
USDJPY Group 5 Low



USDJPY Group 5 Open



USDJPY Group 5 Close



With EURGBP Group 5 it make 52 predictions receiving an accuracy rate of 55.76

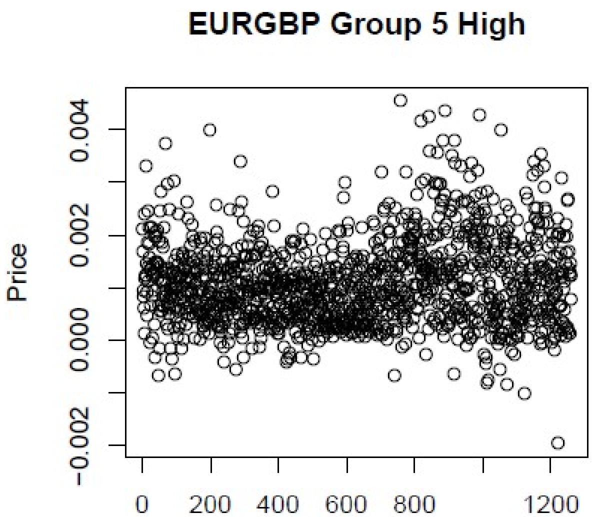
percent. This is

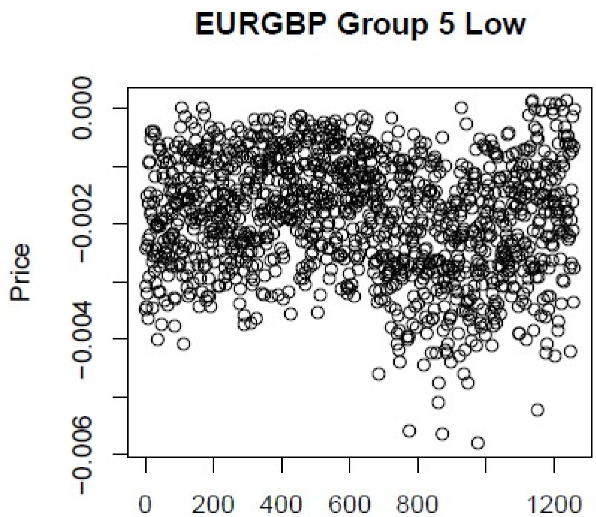
15.56 percent better than random. The EURGBP group was neutral to bearish. With

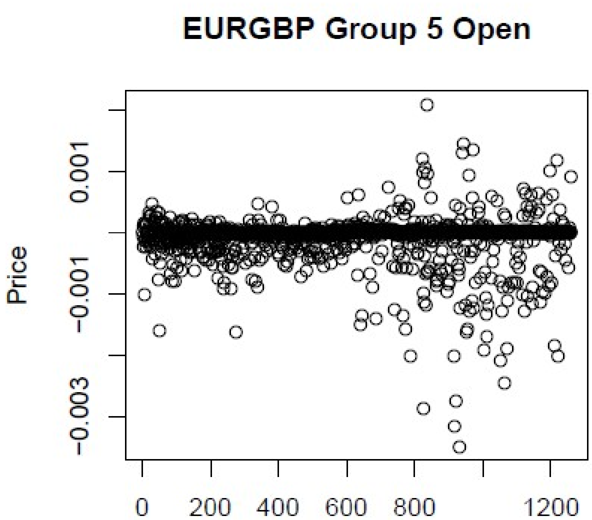
most highs

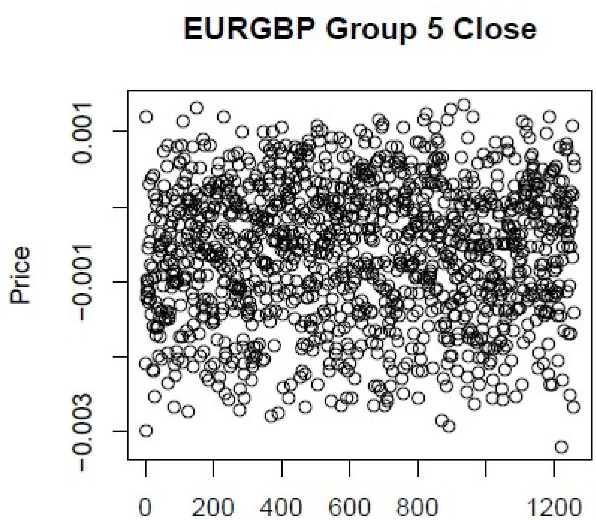
below +20pips and most lows above -40 pips with a closing between +10 and -20

pips.









In finance building useful prediction models that are based solely off price that work on the daily time
frame I believe to be universally difficult. That said this model in particular was not that difficult to
create. It is however, not that useful since the group that it predicts are those that range most of the
time and not the groups that have a very clear one direction. However, given this data I would use it to help confirm a trade direction or trend for the next day. This success indicates that there is some validity in using the shape of the candlestick two predict the future price. Furthermore, when looking at the association rules they show how it takes the cross pairs uses it in determining the future price. However, it shouldn't be used on its own as a standalone trading strategy. What needs to be looked into is when the predictions were wrong and what those open, high, low, close distributions look like when misclassified. Say if the model predicted neutral to bearish but when wrong it is mostly very bearish that would be much better than an incorrect grouping as say very bullish. Thus, not all incorrect grouping may be hurtful to the overall use of the model. To look into ways of improvement of the model would be to figure out a better way to judge the misclassification. In order to do this first one would need to know the use of the end user. Another thought of improvement would be to investigate the use of a different supervise learning technique applied together with k-Means. Also, looking into trying the model on different / multiple time frames. In general candlesticks do seem to have a predictive power, however, a lot of ideas could be investigated into based on the specified needs and goals of the end user.

R Code: You just need to have the dataset files and then set the variable fileLocation correctly to these files(See the six line in the R code). With this done you can run the code and go through the final model. However, it should be noted that there is some randomness in the model. This is because it divides the data set randomly for training and testing.

```
install.packages("arulse");

install.packages("caret");

library(caret);

library(arules);

set.seed(3456);



fileLocation="C:/Users/Main/Desktop/424/424_project/";

pairs=c("EURJPY","GBPJPY","USDJPY","EURUSD","GBPUSD","EURGBP");

minLength=2;

numberOfGroups=6;

support=.01;

confidence=0.65;




graphs=c("High","Low","Open","Close");



for( i in 1:length(pairs))

{

	##load dataFile and put into variable data
```

```
    eval(parse(text =paste("data<-
read.csv(\"",fileLocation,pairs[i],"_Candlestick_1_D_BID_01.01.2004-
31.12.2013.csv\")",sep="")));

    ##make database of each offset candlestick
    eval(parse(text=paste(pairs[i],"_candlesticks<- data.frame(
##Open offset
(data[2:length(data$Open),2]-data[1:(length(data$Close)-1),5])
##High offset
,(data[2:length(data$Open),3]-data[1:(length(data$Close)-1),5])
##Low offset
,(data[2:length(data$Open),4]-data[1:(length(data$Close)-1),5])
##Close offset
,(data[2:length(data$Open),5]-data[1:(length(data$Close)-1),5])

)",sep="")));

##change dataset column names to Open,High,Low,Close
eval(parse(text=paste("names(",pairs[i],"_candlesticks)[1]<-\"Open\"",sep="")));
eval(parse(text=paste("names(",pairs[i],"_candlesticks)[2]<-\"High\"",sep="")));
eval(parse(text=paste("names(",pairs[i],"_candlesticks)[3]<-\"Low\"",sep="")));
eval(parse(text=paste("names(",pairs[i],"_candlesticks)[4]<-\"Close\"",sep="")));

##remove weekend days from data use pair one to determine weekend
if(i==1){
```

```
str="";
        for( j in 1:length(graphs)){
                str<-paste(str,pairs[i],"_candlesticks$",graphs[j],"[]==00",sep="");
                if(j<length(graphs)){str<-paste(str," & ",sep="")}
        }
        eval(parse(text=paste("x<-",str,sep="")));
}
eval(parse(text=paste(pairs[i],"_candlesticks<-",pairs[i],"_candlesticks[x==FALSE,]"
,sep="")));
eval(parse(text=paste("rownames(",pairs[i],"_candlesticks) <- NULL",sep="")));

##use data rows of the offsetted candlestick data with K-means to make
numberOfGroups groups
eval(parse(text=paste("Group<-
kmeans(",pairs[i],"_candlesticks,numberOfGroups,iter.max=1000)",sep="")));



##add group number to current dataset
eval(parse(text=paste(pairs[i],"_candlesticks[\"Group\"]<-Group$cluster",sep="")));

##build new dataset that is the data of the k-means group N-2,N-1,N,N+1
eval(parse(text=paste("data","<-data.frame("
,pairs[i],"_candlesticks[1:(length(",pairs[i],"_candlesticks$Open)-3),5],"
,pairs[i],"_candlesticks[2:(length(",pairs[i],"_candlesticks$Open)-2),5],"
,pairs[i],"_candlesticks[3:(length(",pairs[i],"_candlesticks$Open)-1),5],"
```

```
,pairs[i],"_candlesticks[4:(length(",pairs[i],"_candlesticks$Open)),5]"
,")",sep="")));



##change dataset column names to N_Minus2, N_Minus1, N ,N_Plus1


eval(parse(text=paste("names(","data",")[1]<-\"N_Minus2\"",sep="")));
eval(parse(text=paste("names(","data",")[2]<-\"N_Minus1\"",sep="")));
eval(parse(text=paste("names(","data",")[3]<-\"N\"",sep="")));
eval(parse(text=paste("names(","data",")[4]<-\"N_Plus1\"",sep="")));




##make the groups Binany
Ns=c("N_Minus2","N_Minus1","N","N_Plus1");
for( k in 1:length(Ns)){
	for( j in 1:numberOfGroups){
		string=paste(Ns[k],"_",as.character(j),sep="");
		eval(parse(text=paste(string,"<-
data.frame(data$",Ns[k],"==",j,")",sep="")));


		if(k==1 && j==1){
			eval(parse(text=paste(pairs[i],"<-
data.frame(",string,")",sep="")));
			eval(parse(text=paste("names(",pairs[i],")
[1]<-\"",pairs[i],"_N_Minus2_1\"",sep="")));
```

```
			}
			else{

		eval(parse(text=paste(pairs[i],"[\"",pairs[i],"_",string ,"\"]<-",string,sep="")));
			}
		}
}

##Save the data
eval(parse(text=paste("save(",pairs[i],",file=\"",fileLocation,pairs[i],".Rda\")",sep=""
)));
eval(parse(text=paste("save(",pairs[i],"_candlesticks,
file=\"",fileLocation,pairs[i],"_candlesticks.Rda\")",sep="")));

##eval(parse(text=paste(,sep="")));



}

##build master dataset and save it
string=pairs[1];
for( i in 2:length(pairs))
{
	string<-paste(string,",",pairs[i],sep="");
}
```

```
eval(parse(text=paste("masterData<-data.frame(",string,")",sep="")));
save(masterData,file=paste(fileLocation,"mastersData.Rda",sep=""));

##build the test and training dataset
trainIndex<-createDataPartition(masterData$EURUSD_N_Minus2_1,
p=.8,list=FALSE, times=1);
head(trainIndex);
train<-masterData[trainIndex,];
test<- masterData[-trainIndex,];

##build the LSH and RSH
LHS<-c()
RHS<-c()
names=c("_N_Minus2_","_N_Minus1_","_N_","_N_Plus1_");
for( k in 1:length(pairs)){
	for( i in 1:length(names)){
		for ( j in 1:numberOfGroups){
			if(i<4){
				LHS<-
cbind(LHS,paste(pairs[k],names[i],as.character(j),"=TRUE",sep=""));
			}else{
				RHS<-
cbind(RHS,paste(pairs[k],names[i],as.character(j),"=TRUE",sep=""));
			}
		}
```

```
        }
}


rules <- apriori(train,parameter = list(minlen=minLength, supp=support,
conf=confidence),appearance = list(lhs=LHS, rhs=RHS,default="none"),control =
list(verbose=T));

# find redundant rules
rules.sorted <- sort(rules, by="lift");
subset.matrix <- is.subset(rules.sorted, rules.sorted);
subset.matrix[lower.tri(subset.matrix, diag=T)] <- NA;
redundant <- colSums(subset.matrix, na.rm=T) >= 1;

# remove redundant rules
rules.pruned <- rules.sorted[!redundant];
inspect(rules.pruned);
```

```
##save the rules
eval(parse(text=paste("save(rules.sorted",",file=\"",fileLocation,"rules",".Rda\")",sep
="")));

##build confusion matrix to show end results
confusionMat<-
array(0,dim=c(length(pairs)*numberOfGroups,length(pairs)*numberOfGroups));



##runs through each row on the test group of data
for( row in 1:length(test[,1])){

##build output count double array
outputCount <- array(0,dim=c(length(pairs),numberOfGroups));
##set all values to 0

##run through all rules
for( i in 1:length(rules.pruned)){

##create a string of the condishions needed for the rules as array of strings
```

```
rightside=array(unlist(strsplit(labels(rules.pruned[i]),"=>")))[1];
rightside=rightside[1];
rightside=substr(rightside,2,nchar(rightside)-2);
rightside=array(unlist(strsplit(rightside,",")));
for( j in 1:length(rightside)){
	rightside[j]=sub("=TRUE","",rightside[j]);
}
for( j in 1:length(rightside)){

if(eval(parse(text=paste("test$",rightside[j],"[",row,"]","==TRUE",sep="")))==FALS
E){
	break;
	}
	##all condishion rules were met add to count
	if(j==length(rightside)){
		##see condishion rules output as string
		leftside=array(unlist(strsplit(labels(rules.pruned[i]),"=>")))[2];
		leftside=leftside[1];
		leftside=substr(leftside,3,nchar(leftside)-6);
		group=as.integer(substr(leftside,nchar(leftside),nchar(leftside)));
		for( k in 1:length(pairs)){
			if(grepl(substr(leftside,1,6),pairs[k])){
				outputPair=k;
			}
		}
```

```
                    ##adds to total count
                    outputCount[outputPair,group]= outputCount[outputPair,group] +1;
          }

}
}##ends for loop on all rules

##add into confusion matrix
for( i in 1:length(pairs)){
          ##if max is greating than 0 prediction is made
          if(max(outputCount[i,])>0){

                    for( j in 1:numberOfGroups){
                              ##look for max index on prediction
                              if(outputCount[i,j]==max(outputCount[i,])){
                                        predicted_value=(((i-1)*numberOfGroups)+j);
                              }
                              ##look for actural group


if(eval(parse(text=paste("test$",pairs[i],"_N_Plus1_",as.character(j),"[row]==TRUE",
sep="")))){
                                        actual_value=(((i-1)*numberOfGroups)+j);
                              }
                    }
                    ##update confusion matrix
```

```
confusionMat[actual_value,predicted_value]=confusionMat[actual_value,predicted_v
alue]+1;
          }
}##ends confusion matrix update
print(paste("Test case ",row, " of ", length(test[,1]),sep=""));

}##ends testing each row in test data group





##build a outputSuccess and performance net plot and plot each candlestick data
of predicted group
pdf(paste(fileLocation,"output.pdf",sep=""));
outputSuccess <- array(0,dim=c(length(pairs),numberOfGroups));
outputPerformance <- array(0,dim=c(length(pairs),numberOfGroups));
for( i in 1:length(pairs)){
          for( j in 1:numberOfGroups){
                    if(sum(confusionMat[,(((i-1)*numberOfGroups)+j)])>0){

                              outputPerformance[i,j]<-((confusionMat[(((i-1)*numberOfGroups)
+j),(((i-1)*numberOfGroups)+j)]/sum(confusionMat[,(((i-1)*numberOfGroups)+j)]))-
eval(parse(text=paste("sum(masterData$",pairs[i],"_N_Plus1_",as.character(j),")/len
gth(masterData$",pairs[i],"_N_Plus1_",as.character(j),")",sep=""))));
```

```
                              outputSuccess[i,j]<-(confusionMat[(((i-1)*numberOfGroups)+j),
(((i-1)*numberOfGroups)+j)]/sum(confusionMat[,(((i-1)*numberOfGroups)+j)]));

                              ##build a normal curve of open,high,close,low
                              str=paste(pairs[i],"_candlesticks$Group[]==",j,sep="");
                              eval(parse(text=paste("x<-",str,sep="")));

          eval(parse(text=paste("x<-",pairs[i],"_candlesticks[x,]",sep="")));

                              par(mfrow=c(2,2))
                              for( k in 1:length(graphs)){
                                        eval(parse(text=paste("plotData=x$",graphs[k],sep="")));
                                        plot(plotData,main=paste(pairs[i]," Group ",j,"
",graphs[k],sep=""),type="p",xlab="",ylab="Price");
                              }
                    }
          }
}
dev.off();
sink(paste(fileLocation,"output.txt",sep=""), append=FALSE, split=FALSE);
print(outputPerformance);
print(outputSuccess);
sink();
```