

# Pattern Recognition and Prediction in Equity Market

Lang Lang, Kai Wang

## 1. Introduction

In finance, technical analysis is a security analysis discipline used for forecasting the direction of prices through the study of past market data. The technical analysis of the past market data would usually be focused in the moving average of price, support line, resistance line and charting patterns. This type of technique has been widely among the traders and investors of the market. However, technical analysis has not received the same level of academic scrutiny and acceptance as more traditional approaches such as fundamental analysis. One of the main obstacles is the highly subjective nature of technical analysis—the presence of geometric shapes in historical price charts is often in the eyes of the beholder. So in this project, we propose to build appropriate quantitative models to recognize some tradable signals from technical perspective from past data and test the signals in the equity market. We will discuss generating technical signals from supply and demand perspective, from charting patterns perspective and from softmax classification perspective. We back-tested these signals on thousands of stocks and got fairly good results.

## 2. Signal Type One: Support Line and Resistance Line

Support and resistance lines appear as thresholds to price patterns. They are the respective lines where prices stop moving down or up. A support line is the level that a stock's price generally will not fall below. It marks the price level where there is adequate demand to stop, and possibly to turn a downtrend higher. A resistance line is the level above which a stock's price typically will not rise. It indicates a price level where an adequate supply of stock is available to stop and fend off an uptrend in prices.

Support and resistance levels can be identified by trend lines. Some traders believe in using pivot point calculations. The more often a support/resistance level is "tested" (touched and bounced off by price), the more significance given to that specific level. If a price breaks past a support level, that support level often becomes a new resistance level. The opposite is true as well, if price breaks a resistance level, it will often find support at that level in the future.



Figure 1: Support Line and Resistance Line

Support and resistance levels are useful in evaluating the stock supply and demand, but the choosing of support line and resistance line is highly subjective. To test the signals, a quantitative way to detect support line and resistance line is proposed as follows. First, for any specific stock on a specific day, smooth the past data first to eliminate ‘wrong’ local minimums and maximums brought by market noise and find out ten most recent ‘true’ minimum and maximum points. Second, regress over the past ten local minimums and maximums to get slopes of the two lines. In our hypothesis, these two lines represents the support and resistance levels and give us signals for trading.

We use following principles to trade this signal:

1. When price rises to around resistance line, short the stock
2. When price goes above resistance line, buy the stock and consider the resistance line as the new support line
3. When price drops to around support line, buy the stock
4. When price drops below support line, short the stock and consider the support line as the new resistance line

Following are sample plots of support/resistance line and the plot of its trading performance

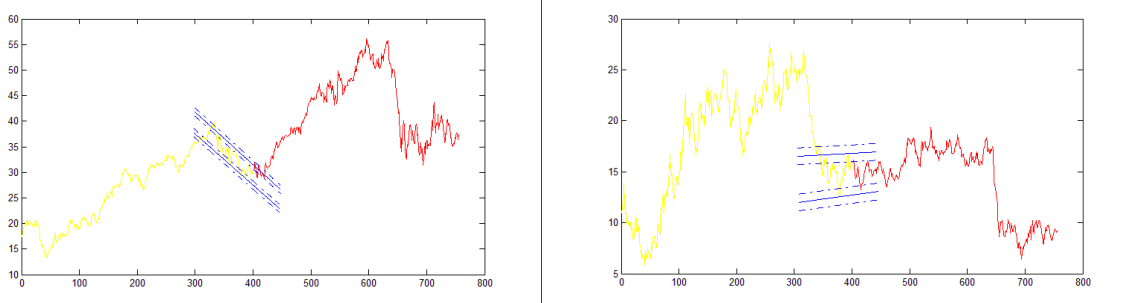


Figure 2: Recognizing Support Line and Resistance Line in Matlab



Figure 3: Performance of Support/Resistance Line Signals

We can see that the return seem to be quite good, with the risk-adjusted return (ir) being 0.036.

## 2. Signal Type Two: Charting Patterns

Candlesticks are usually composed of the body (black or white), and an upper and a lower shadow (wick): the area between the open and the close is called the real body, price excursions above and below the real body are called shadows. The wick illustrates the highest and lowest traded prices of a security during the time interval represented. The body illustrates the opening and closing trades. If the security closed higher than it opened, the body is white or unfilled, with the opening price at the bottom of the body and the closing price at the top. If the security closed lower than it opened, the body is black, with the opening price at the top and the closing price at the bottom. A candlestick need not have either a body or a wick.

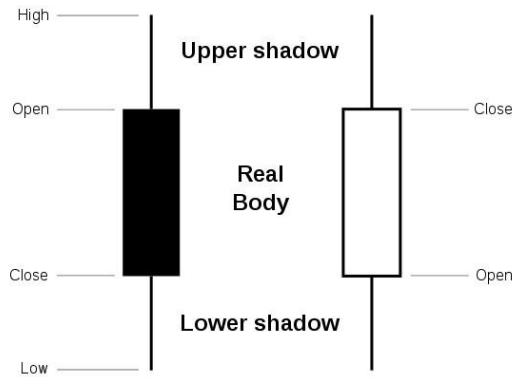


Figure 4: Candlesticks Chart

The second type of trading signals that we want to test comes directly from analyzing the candlesticks chart, which represents a stock's open price, close price, highest price and lowest price of a day. Different shapes of candlesticks typically indicate different distributions of price movement. We can build up programs to keep track of any stock's high, low, open and close and recognize different candlesticks chart patterns and test them. For example, 'hammer' candlesticks are usually a signal of reversal, indicating that the current price trends may not last long any more. The 'hammer' type is characterized as having a very long shadow, typically twice larger than the real body. Using matlab, we are actually able to test many chart patterns to see if they are informative in predicting the future. Following are figures of 'hammer' pattern and its trading performance.

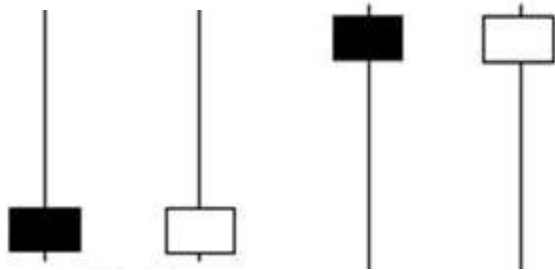


Figure 5: 'Hammer' Type Candlesticks

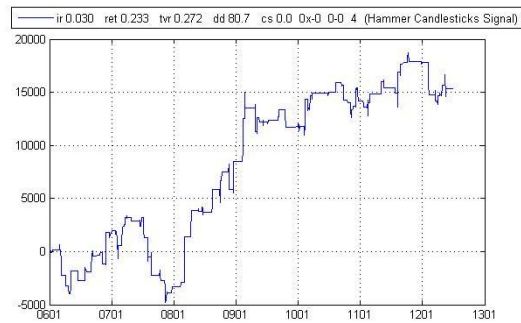


Figure 6: Performance of Charting Pattern Signals

#### 4. Signal Type Three: Softmax Regression and Classification of Future Returns

The first two types of signals that we talked before are traditional technical analysis methods. In general, these methods work fine in the sense that they usually reflect some probability rules. So while we have a very large asset class that we can trade on, we can get positive return. A third type of technical signals we can think of is from the idea of machine learning class. We all know that equity market is not a complete market, which means that past data can contain information about future. Therefore, information like past prices, return, trading volume, volatility, etc, may be useful if we can 'learn' some patterns from them. Here, we choose to use softmax regression to model the relationship between future return and past market data.

The input variables come from a stock's one day return, 5 days return, ratio of intraday high and low, changes in daily high, low, open and close prices in two consecutive days, difference between morning volume and afternoon volume and its daily changes, changes in one-day return, daily morning return, daily afternoon return, change in daily volume, ratio of afternoon return standard deviation and morning return standard deviation, changes in morning return and afternoon return in two days, changes in 5days return in 5 days, and SP&500 return, VIX index, daily top 500 stock return, etc. With above information, we are able to select and construct 24 variables that we consider to be most informative technical input variables.

We want to use these input variables to predict tomorrow's daily return. Future returns are classified into three buckets:

Positive	Neutral	Negative
return > 1.5%	-1.5% ≤ return ≤ 1.5%	return < -1.5%

Our goal is to predict which bucket tomorrow's return will fall into and design a trading strategy using the predicted values. Therefore, we implemented softmax regression for this problem (Part of the program comes from "Exercise: Softmax Regression, Stanford UFLDL Tutorial").

Softmax regression is a method to solve classification problems in which the response variable  $y$  can take on any one of  $k$  values, i.e.  $y \in \{1, 2, \dots, k\}$ .

For softmax regression model, we assume  $p(y = i|x; \theta)$  is linear with  $e^{\theta_i^T x}$ , thus:

$$p(y = i|x; \theta) = \frac{e^{\theta_i^T x}}{\sum_{j=1}^k e^{\theta_j^T x}}$$

By linearity, we only need to consider  $\theta_1, \dots, \theta_{k-1}$  as parameters of the model.

Our hypothesis will output

$$\begin{aligned} h_{\theta}(x) &= E\{(1\{y = 1\}, 1\{y = 2\}, \dots, 1\{y = k - 1\})^T | x; \theta\} \\ &= \left( \frac{e^{\theta_1^T x}}{\sum_{j=1}^k e^{\theta_j^T x}}, \frac{e^{\theta_2^T x}}{\sum_{j=1}^k e^{\theta_j^T x}}, \dots, \frac{e^{\theta_{k-1}^T x}}{\sum_{j=1}^k e^{\theta_j^T x}} \right)^T \end{aligned}$$

In other words, our hypothesis will output the estimated probability that  $p(y = i|x; \theta)$  for every value  $i=1, \dots, k$ .

For parameter fitting, if we have  $m$  training samples  $\{(x^{(i)}, y^{(i)}); i = 1, \dots, m\}$ , the log-likelihood can be written as:

$$\begin{aligned} l(\theta) &= \sum_{i=1}^m \log p(y^{(i)} | x^{(i)}; \theta) \\ &= \sum_{i=1}^m \log \prod_{l=1}^k \left( \frac{e^{\theta_l^T x}}{\sum_{j=1}^k e^{\theta_j^T x}} \right)^{1\{y^{(i)}=l\}} \end{aligned}$$

Thus we can obtain the maximum likelihood estimate of the parameters by maximizing  $l(\theta)$  in terms of  $\theta$  using stochastic gradient ascent.

We trained the model on a data set of size 1000, and applied out-of-sample test on 4706 stocks for 1600 days. The accuracy is 46%, which is slightly higher than random guess. Considering the equity market is extremely hard to predict and highly random, 46% accuracy in choosing one from the three buckets is actually a decent number, indicating a positive R-Square between the input variables and dependent variable.

To test whether this prediction model is useful in trading or not, we first assign a signal value to the three buckets, indicating how much to buy or sell a stock:

Positive	Neutral	Negative
1	0	-1

The performance of this trading strategy is plotted below:



Figure 6: Performance of Softmax Regression Signal

We can see that trading signals from softmax regression perform ok, but the return line is a little less smooth, which means that this method is still not very steady and consistent and can be further improved.

## 5. Conclusion:

Quantitative methods to model three types of technical signals have been discussed and back-tested in this paper. The performance metric of these three types of signals are summarized as follows:

	Information Ratio (Risk-Adjusted Return)	Turnover	Drawdown
Support/Resistance Line	0.038	0.512	48.6
Charting Patterns	0.030	0.272	80.7
Softmax Regression	0.023	1.845	9.1

We can find that quantitative signals generated from traditional technical analysis give fairly good risk-adjusted return and low turnover, but has high drawdown, while the more modern technical signal generated by running a softmax regression model over many technical variables, gives slightly lower but still positive risk-adjusted return with high turnover. The application of machine learning techniques in trading signal construction seems not to be able to beat traditional techniques, but from the performance we can see that this signal has value. Considering that modern methods use totally different principles from traditional ones, they generate signals with very low correlations, and thus add value to one another. We can improve the model by selecting better and more input variables for the model as well as researching to assign better signal values in the trading process.

## Reference

- [1] Andrew Ng. Cs229 lecture notes on Softmax Regression. 2012.
- [2] Wikipedia. "Support and resistance", [http://en.wikipedia.org/wiki/Support\\_and\\_resistance](http://en.wikipedia.org/wiki/Support_and_resistance)
- [3] Wikipedia. "Candlestick chart", [http://en.wikipedia.org/wiki/Candlestick\\_chart](http://en.wikipedia.org/wiki/Candlestick_chart)