Statistical Significance Analysis of the CHAOS-BASED PREDICTION MODEL

In this paper we study a measure of statistical significance of the forecasting models employed at Trading Pro
There have been in the market several prediction systems that offer market price forecasts. But how good are they? In general they all fail to prove the statistical significance of their predictions. In this article we will try to prove the accuracy of Trading Pro' forecasting models by analyzing its predictive capabilities against chance.

Abstract

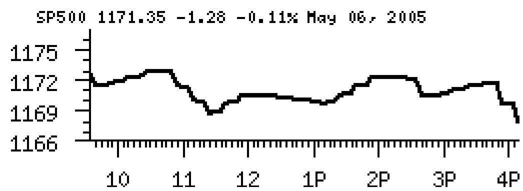
Trading Pro distributes the signal generated by a market prediction system, based on Dynamical Systems Theory and Artificial Intelligence. The system makes intraday predictions every 15 minutes, for the rest of the day. We believe the predictor has an edge that can be exploited to obtain significant returns by trading any cash or futures market.
The objective of this paper is to analyze the statistical significance of Trading Pro's chaos based forecasting model accuracy.
The analysis will be focused on proving that there is a low probability that the daily forecast could be the the result of a random process.

Introduction

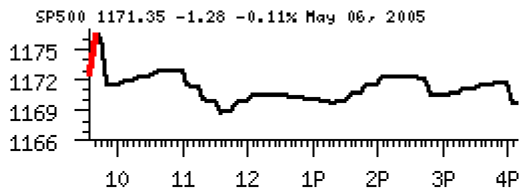
To run the necessary tests we have chosen the DJIA Index for the period 12/1/2003 to 12/15/2005. All time values are Eastern Time. The predictor generates 20 predictions per day, one every 15 minutes. Each prediction contains a dataset, sampled every 5 minutes, starting at the time of the computed prediction, and ending at the end of the day.

The first prediction of the day is generated at market opening 9:30A . The next prediction is at 9:45A, 10:00A, 10:15A, and so on. The last 15 minutes of the market are computed on each new prediction, by the system. As an example lets take a look at the S&P500 May 06, 2005 prediction.

First Prediction
At 9:30A the predictor showed the following chart for the day.

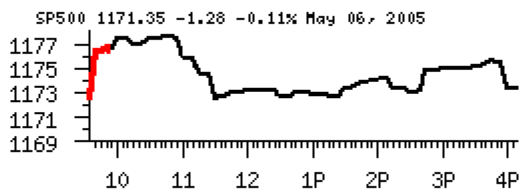


Second Prediction
The second prediction is executed at 9:45A Market opened with a gap to the LONG side. The system computed the new 15 minutes market data, and adjusted the prediction as follows.



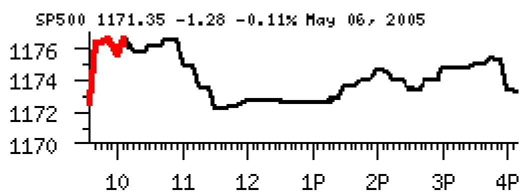
Third Prediction

The third prediction of the day is computed at 10:00A. The system add to it's historical data, the new 15 minutes market data and recomputed the prediction as follows.



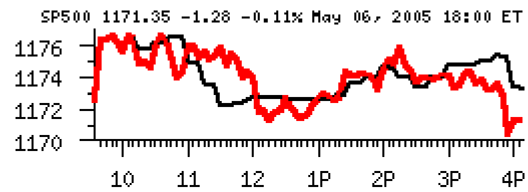
Fourth Prediction

The fourth prediction is computed at 10:15A. The system recomputed the prediction for the rest of the day as follows.

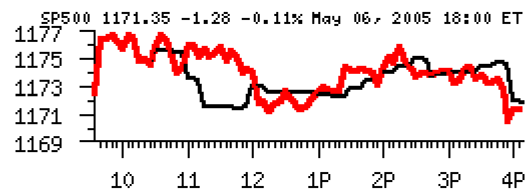


As the system needs more data to adjust the prediction, and generally there are opening gaps that can affect the first predictions of the day, we do not consider the first 3 forecasts.

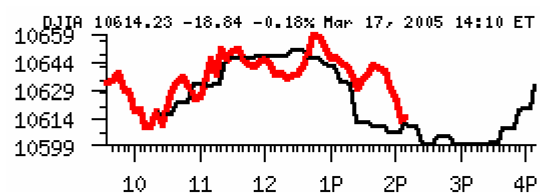
If we superpose the fourth prediction of the day (the black line), with what the market finally did (the red line) the chart would look as follows.



The system goes on making prediction every 15 minutes.
For example the fifth prediction of the day, with the market on top of it looked as follows. .



The following is another example of the fourth prediction of the day for the DJIA for March 17th, 2005. In this case, the market is plotted up to 14:10.



For the purpose of this paper only the fourth prediction of the day will be used. Two different prediction scopes will be analyzed.

Forecast	Start Time	End Time
A	10:20	11:20
B	10:20	12:20
C	10:20	13:20
D	10:20	14:20
E	10:20	15:20
F	10:20	16:00

Null Hypothesis

We define that the null hypothesis is that the model has no predictive capabilities (there is no difference between a random prediction and the model prediction or a random market and the prediction).

A random market is a simulated market created with some random process that will generate market data.

The alternative hypothesis is that the prediction model does predict.

We have chosen the Root Mean Square Error, [RMSE], as the metric to be measured. We will compute the RMSE of the prediction with respects to the index, and the RMSE of the prediction with respects to a random market. We will reject the null hypothesis, if the distributions obtained from the RMSEs of the prediction are not one more occurrence of the distribution obtained from the random market RMSEs.

Method

We compute the RMSE as follows.

$$RMSE_{Pj} = \frac{\sqrt{\sum_{i=0}^{n}\left(F_i^{\,j} - I_i^{\,j}\right)^2}}{i}$$

$RMSE_{Pj}$ is the Root Mean Square Error for the day *j* of the predicted time series against the index.

Where:
n = is the number of 5 minutes samples between the time where the prediction is computed, and the horizon of such prediction within a certain *j* day. n has the same value for any *j* day.
$F_i^{\,j}$ = is the forecasted value for the 5 minute *i* sample element *j* day.
$I_i^{\,j}$ =is the index value for the 5 minute *i* sample element *j* day.

We defined the critical alpha level [α], to be used in the tests, as follows.

$$\alpha = \left\{1 - \frac{RMSE_P}{\mu\left(RMSE_{RND}\right)}\right\} * 100$$

Where $RMSE_P$ is the $RMSE_{Pj}$ averaged over the approximately 500 trading days of the testing set:

$$RMSE_P = \frac{\sum_{j=1}^{m} RMSE_{Pj}}{m}$$

And *m* is the number of days considered for the test.
$\mu\left(RMSE_{RND}\right)$ is the average over 250 different random runs, each representing one year of random data.
$RMSE_{RND}$ is different for each of the two tests considered.

Test 1 computes the random RMSE by comparing the forecast against a random market:

$$RMSE_{RND} = \frac{\sum_{j=1}^{m} RMSE_{RNDj}}{m}$$

With:

$$RMSE_{RNDj} = \frac{\sqrt{\sum_{i=0}^{n}\left(F_i^{\,j} - I_{RNDi}^{\,j}\right)^2}}{i}$$

$I_{RNDi}^{\,j}$ is a random market, constructed by picking returns obtained from the real market data set $I_i^{\,j}$, but randomly.

The random market creation was done with the following procedure.

1. We took the returns from the 5 minutes DJIA Index sampled data on the analyzed timeframe. The 5 minutes real volatility.
2. We created a vector with the computed returns.
3. We then recreated the market by picking from the real returns vector a random element and recalculated the market.
4. Each day, the market is recomputed beginning at the starting time of the forecast, taking the real index value at that time and adding the returns, positives or negatives.

Test2 computes the random RMSE by comparing a random prediction against the real market

$$RMSE_{RNDj} = \frac{\sqrt{\sum_{i=0}^{n}\left(F_{RNDi}^{\,j} - I_i^{\,j}\right)^2}}{i}$$

$F_{RNDi}^{\,j}$ is a random prediction, constructed the same way as the random market: by picking returns obtained from the real market data set I_i^j, but randomly.

We define

$$\alpha_{\min} = 8\%.$$

We will reject the null hypothesis if

$$\alpha_{test} > \alpha_{\min}$$

Tests - Implementation

For each test, six prediction lengths have been tested as follows.

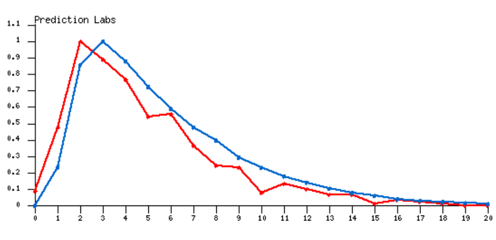
Forecast	Start Time	End Time
A	10:20	11:20
B	10:20	12:20
C	10:20	13:20
D	10:20	14:20
E	10:20	15:20
F	10:20	16:00

For each forecast, an α_{Test} has been computed and the corresponding RMSE distribution charted.

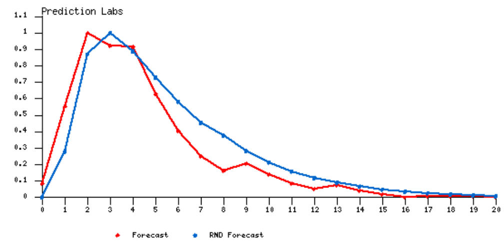
Test 1 – Implementation

Forecast	Start Time	End Time	RMSE Forecast	RMSE RND	α_{Test}
A	10:20	11:20	5.37	6.30	14.75%
B	10:20	12:20	4.96	6.05	18.15%
C	10:20	13:20	4.69	5.93	20.95%
D	10:20	14:20	4.51	5.87	23.15%
E	10:20	15:20	4.45	5.91	24.76%
F	10:20	16:00	4.47	6.00	25.42%
Average					**21.20%**

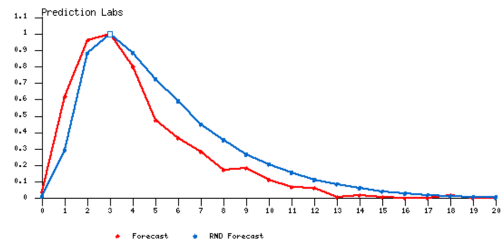
Distribution A – T1



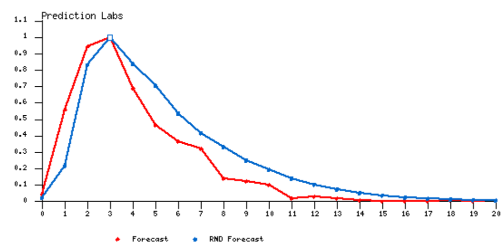
Distribution B – T1



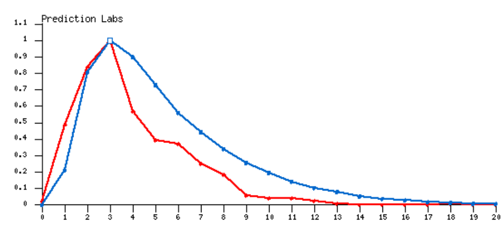
Distribution C – T1



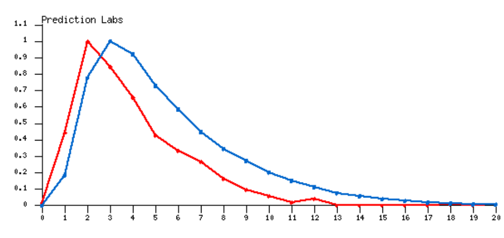
Distribution D – T1



Distribution E – T1



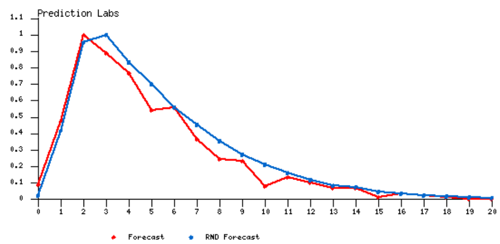
Distribution F – T1



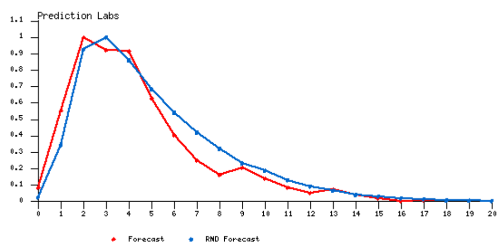
Test 2 – Implementation

Forecast	Start Time	End Time	RMSE Forecast	RMSE RND	α_{Test}
A	10:20	11:20	5.37	5.90	8.96%
B	10:20	12:20	4.96	5.61	11.75%
C	10:20	13:20	4.69	5.50	14.74%
D	10:20	14:20	4.51	5.39	16.29%
E	10:20	15:20	4.45	5.30	16.02%
F	10:20	16:00	4.47	5.23	14.55%
Average					**13.72%**

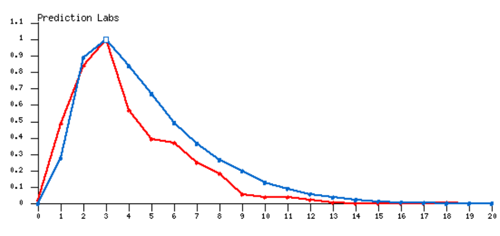
Distribution A – T2



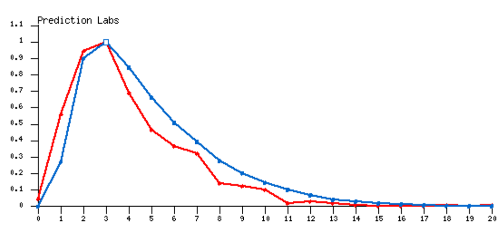
Distribution B – T2



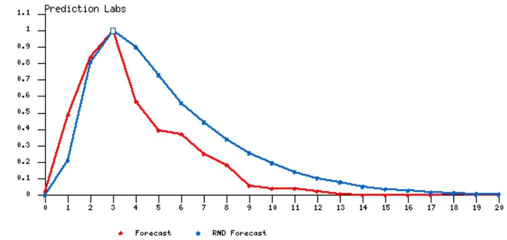
Distribution C – T2



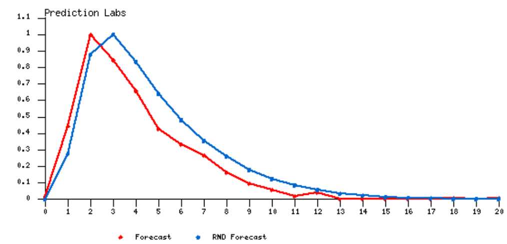
Distribution D – T2



Distribution E – T2



Distribution F – T2



Conclusions and perspectives

After running the tests with 1 hour to 6 hours forecasts, we reject in all cases the null hypothesis.

We can conclude that on average considering Test 1 we get an edge of 21.20% above the 50% chance.

Considering Test 2, the edge is 13.72% above the chance.

We can also conclude that the longer the prediction within the day, the better the odds for the predictor.