A Machine Learning Approach to Combine the Trend-following and Counter-trend Trading Strategies

Guangyu Wu, Lu Guo
{ gw923 , lg2691 }@nyu.edu

Background:
Trend-following strategies (TF) and counter-trend (CT) strategies are popular systematic trading strategies, but as their names suggest, they have vastly different underlying mechanisms and work best under different market conditions. The trend-following strategies are witnessed to work well in trending periods but suffers when markets are turbulent or experiencing financial crisis. In contrast, certain counter-trend strategies take huge benefits in the recovery from a crisis but they tend to lag behind during less volatile, trending periods. **The contrasting natures and features of the two types of strategies prompt us to look into the possibility of constructing a combined strategy on top of them.**

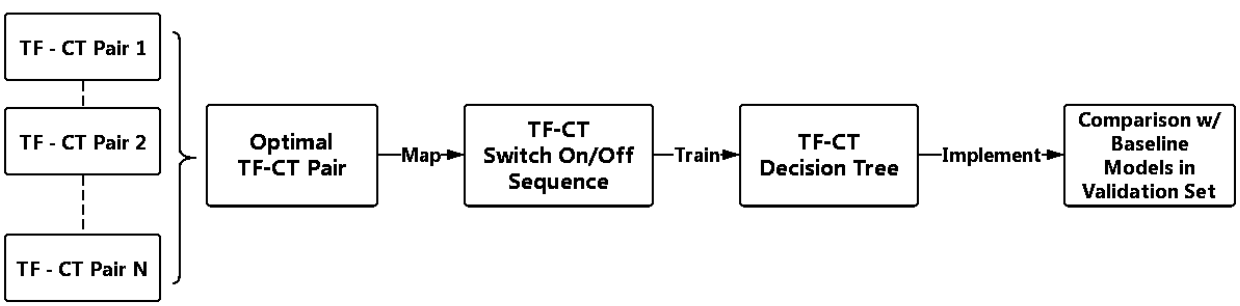
Research Question:

- What are the possible ways to combine trend-following and counter-trend strategies, to leverage the best features of both and avoid their downsides?
- How does the application of machine learning algorithms play a role in selecting strategies to combine, as well as implementing the combined strategy?

Data:
The primary data studied is S&P 500 E-mini Futures contracts price data (including "Date", "Open", "High", "Low", "Close") from 1999 to 2019.

Reason for choosing this data: commonly traded/low transaction costs/relative abundant liquidity.

Methodology:
The trading strategy is to have two base strategies (one trend-following, one counter-trend) and use a machine learning model to predict which base strategy will perform better on the next day, then that strategy is executed on the next day accordingly.

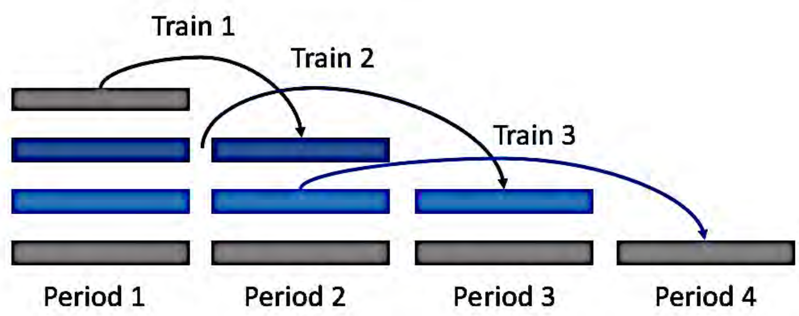


Such a trading system is constructed in two steps:
1) Identify the optimal pair of trend-following strategy and counter-trend strategy to combine;
2) Train a machine learning algorithm to predict which strategy to use in the next day in training set.

These two steps can be repeated to update trading system model as new data comes in. For this project, the time interval we chose to update is every 2 years i.e. train a new model when additional 500 data points become available each time.

To measure the performance of the trading system, we created a standard backtesting process that simulates what would happen if this trading system was implemented since 1999 (the start time of our dataset).

Backtesting Process Based on Accumulative Historical Data:



We retrain a trading model every 2 years, and there are 10 periods (20 years) in total in this project. In this way, we dynamically train and test our model to exploit all the data we have and avoid aliasing at the same time.

Step 1 Base Pair Identification:
We listed several criteria to identify the optimal TF-CF pair. Through testing these hypotheses on real data, we identified limitations of each and constructed an improved criterion to identify the optimal TF-CF pair.

	Methodology	Limitation
Criterion 1: **Highest IR**	Pair the one with highest IR among all TFs and the one with highest IR among all CFs.	Two strategies that have high IR individually are not necessarily complementary with each other if they move closely together.
Criterion 2: **Most Negative Correlation**	Pair the two strategies that have most negative correlation.	By the definition of correlation, it is sensitive to relative move between two strategies. A pair with most negative correlation means that the two strategies move apart most. However, **in the context of our problem, we are interested in not only the magnitude of relative movement between the two, but also what range does the returns of these two moving strategies fall into.** For example, we tend to find two strategies that always have flipped sign of returns rather than two always have returns with the same sign. Since when we try to combine the two, the former situation yields a higher IR mathematically.
Criterion 3: **Top Avg Ranking on Criterion 1&2**	Find the pair that has the top average ranking on highest IR and most negative correlation.	By picking the top average ranking of the two criteria above, we have no clue if it would reinforce the strength of the two or the weakness of the two. The underlying methodology is relatively arbitrary.

Test out the 3 criterion above in real data:

	Highest IR	Lowest Correlation	Average Ranking	Complement Finding
IR	0.365	0.222	0.087	0.610
Accuracy	0.835	0.444	0.383	0.525
F1 Score	0.855	0.335	0.365	0.517

The first three columns verify our worries of the three criteria above. Inspired by the limitations of these strategies, we finally settled down on a new criterion to leverage the strengths of them and at the same time avoid the weaknesses of them.

1) The best criterion in practice among the three is the first one: Highest IR. Therefore, we decided to choose the one strategy that yields highest IR as our base strategy.
2) The most negative correlation only takes care of the magnitude in relative direction, but fails to take care of the range of returns. Therefore, we then go on to search for a complementary strategy with highest returns on the days when our pre-identified base strategy performs worst (yielding bottom 10 percentile returns).

We named the procedures above **Complement Finding process**, which is the fourth column in the table above. Data shows that this methodology beats the other three a lot in terms of IR.

Step 2 Prediction Algorithm Training:
ML task: Predict which strategy in the pre-identified pair will generate higher return tomorrow.
ML algo: Logistic Regression, SVM, Random Forest (supervised learning, classification problem)

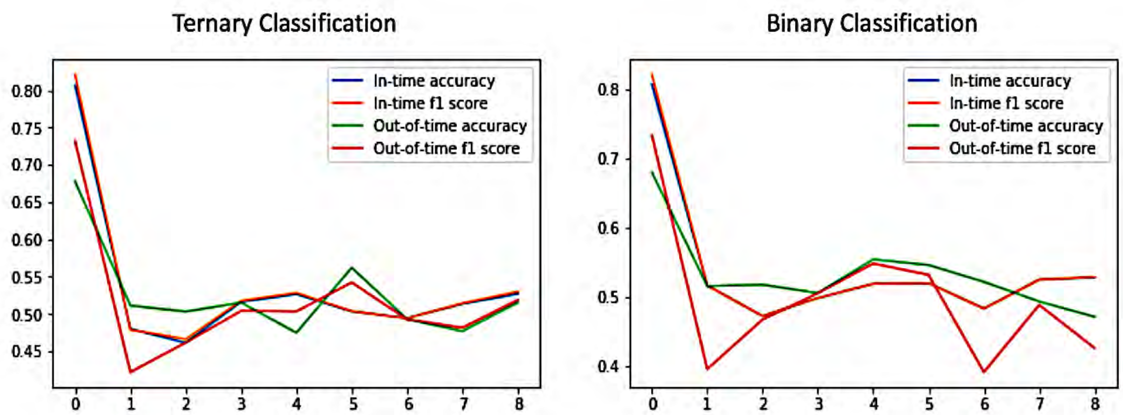
Y variable: Which strategy to use (or we shouldn't trade tomorrow)
X variables:
1) Moving average, moving vol, daily price range (with different time windows)
2) Normalize indicators in (1) and get their z-values (with different time windows)
3) Macro data[1]: CPI, GDP, civilian unemployment rate, FED fund rate, yield spread (T10Y2Y)
However, macro data are much lower in frequency compared to daily S&P data, and they proved to contribute little to prediction accuracy in our trials, so macro data are dropped in the final prediction model.

<u>Subtask 1: Y Variable Choice – Binary/Ternary Classification</u>
We specified two kinds of classification task for ML models:
- Binary classification: TF or CT.
- Ternary classification: TF or CT or OFF.

Acknowledging that there are days where both strategies lose money, the "ternary classification" way seems to be more accurate and scientific. However, due to the **low predictivity** nature of our problem (as indicated by experiments on real data), the predicted result from binary classification model may outperform that from ternary classification model given the ambiguity and complexity in the ML task.

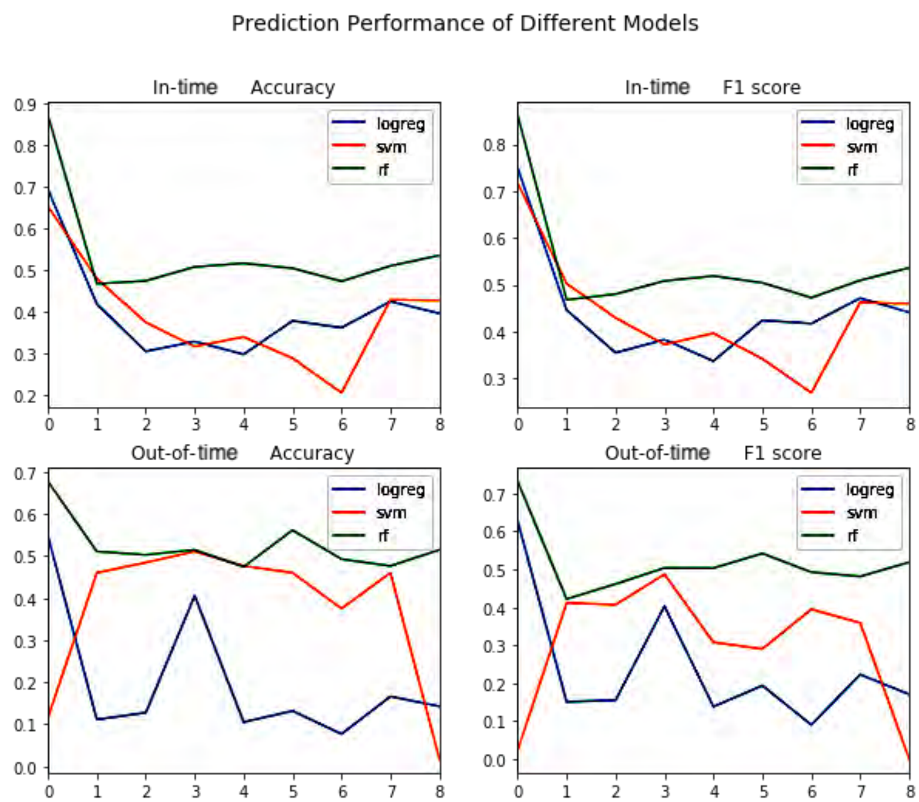
Therefore, to carry out the ML algo training, we first tested on real data whether presenting the ML task as binary classification or ternary classification is a more friendly ML task.

[1] Data obtained from the website of Federal Reserve Bank of St. Louis.



By implementing RF model, we found that **presenting the task as a ternary classification one is more ML-friendly.** (IR: 0.61>0.45)

<u>Subtask 2: ML Algorithm Choice – Logistic/SVM/Random Forest</u>
We compared the performance of prediction models (Logistic Regression/SVM/Random Forest) in terms of in-sample accuracy, out-of-sample accuracy, in-sample F1, and out-of-sample F1.



Result shows that **Random Forest model is consistently better than the other two**. Then we went on to fine-tune the hyper parameters in RF using grid search. However, the result suggests it overfits the data when we conducted aggressive grid search. Therefore, we set limits to the hyper parameters of the RF model to **avoid overfitting**.

Results & Discussions:

Equity Graph of Our Combined Strategy, Two Underlying Base Strategies, and BuyNHold (2002-2019)

```
Comparison of the Information Ratio of Different Strategies (2002 -2019):

Prediction Strategy: 0.61
BuyNHold_Return: 0.383
Long_20_2.2_Return: 0.709
SMA_10_30_Return: 0.069
```

BuyNHold IR: 0.383
Our Combined Strategy IR: 0.61>0.383>0.069
Underlying TF IR: 0.069
Underlying CT IR: 0.709

In the 10-year period after 2008 financial crisis:
Our Combined Strategy IR: 0.772
Underlying CT IR: 0.39
BuyNHold IR: 0.76

Analysis:

1) Although IR of our combined strategy is not as high as the underlying CT, **our combined strategy still captured the upsides much better than the underlying CT during non-crisis periods especially in bull periods which the CT strategy missed out.**
2) If we look at the 10-year period after 2008 financial crisis, the underlying CT strategy only has an IR of 0.39, which means that **CT only shows its relative strength in crisis periods**. Imagine if an investor traded based on this CT strategy after the crisis, he or she would be most likely unsatisfied with the low IR so far, and might only get compensated if another crisis occurs in the near future.
3) Similarly on the other side, during the last 10-year period, BuyNHold has a superior IR of 0.76, but it failed miserably during the crisis around 2008. **It's too much a risk to adopt BuyNHold strategy hoping that there won't be a similar crisis coming in the future.**

To further verify the idea above, we find the top two strategies during a crisis period and those during a non-crisis period, and test them in another period.

<u>Equity Graph of Our Prediction-Based Combined Strategy, BuyNHold, and a Strategy that Combines the Two Base Strategies with Equal Weighting</u>
(First Graph: 2001-2010; Second Graph: 2011-2019)

Test the top two strategies from a normal ten-year in a ten-year period with financial crisis

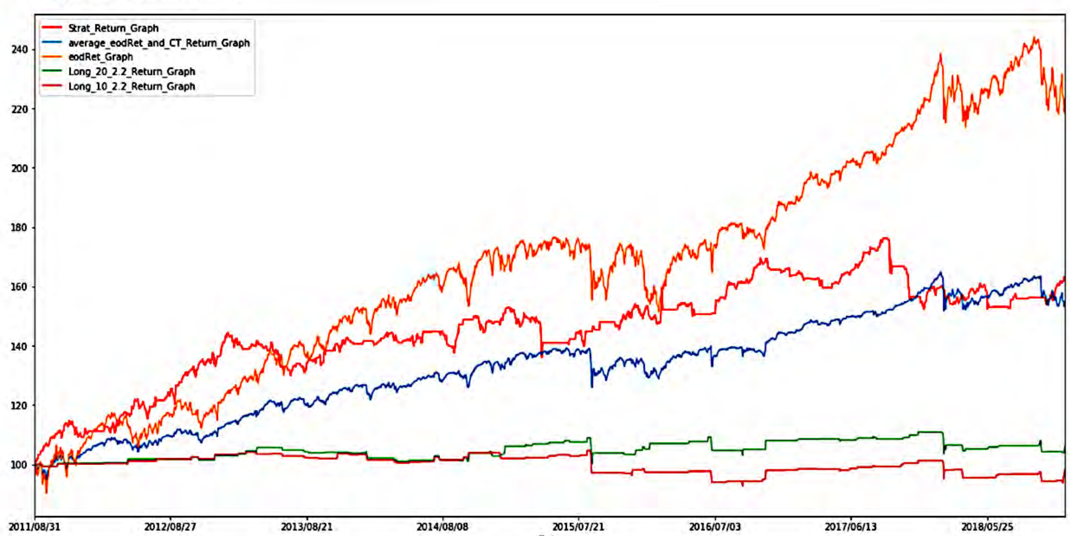
Prediction Strategy, Buy & Hold, Simple average of bases strategies are graphed as well

```
Strat_Return 0.752
average_eodRet_and_CT_Return 0.404
eodRet 0.173
Short_30_0.6_Return 0.059
Short_20_0.6_Return -0.062
```

Test the top two strategies from a ten-year that has a crisis in a ten-year period without one

Prediction Strategy, Buy & Hold, Simple average of bases strategies are graphed as well

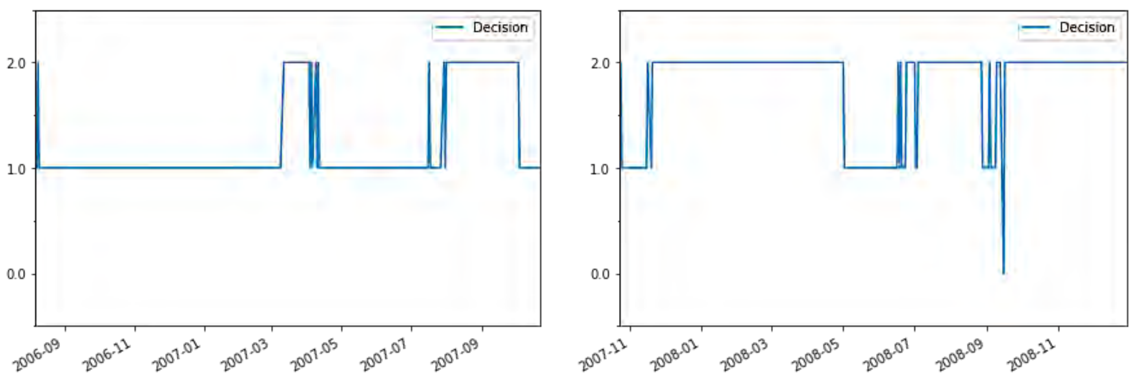
```
Strat_Return 0.772
average_eodRet_and_CT_Return 0.789
eodRet 0.848
Long_20_2.2_Return 0.187
Long_10_2.2_Return -0.042
```



Observations:

1) Top strategies selected from 2001-2010 (a period with crisis) only beats our combined strategy in 2001-2010, but they perform poorly in other market conditions (2011-2019 a period without crisis). While **our combined strategy is relatively consistent in its performance throughout the time.**
2) **Our combined strategy outperforms the equal-weighting strategy in a crisis period while doing equally well in non-crisis periods, which demonstrates the value that machine learning prediction added.**

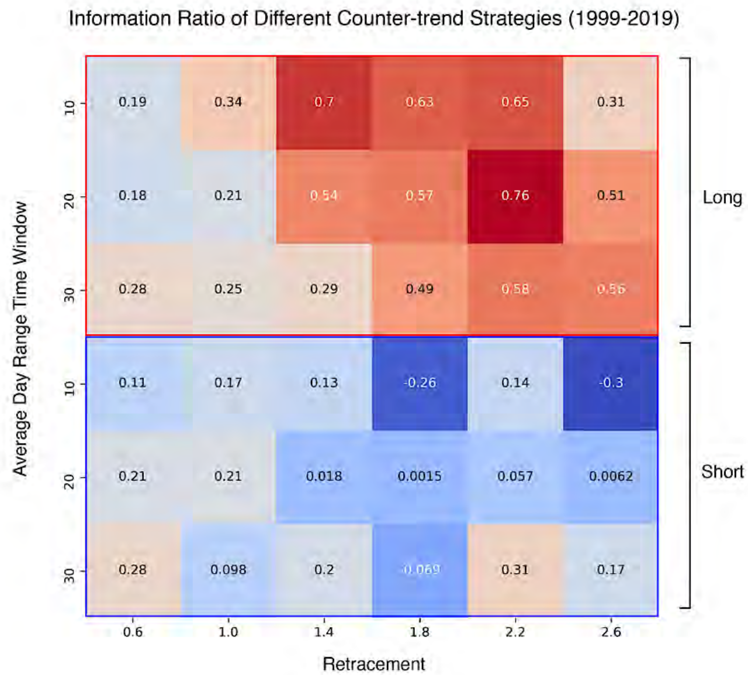
A Deeper Look into Our Combined Strategy

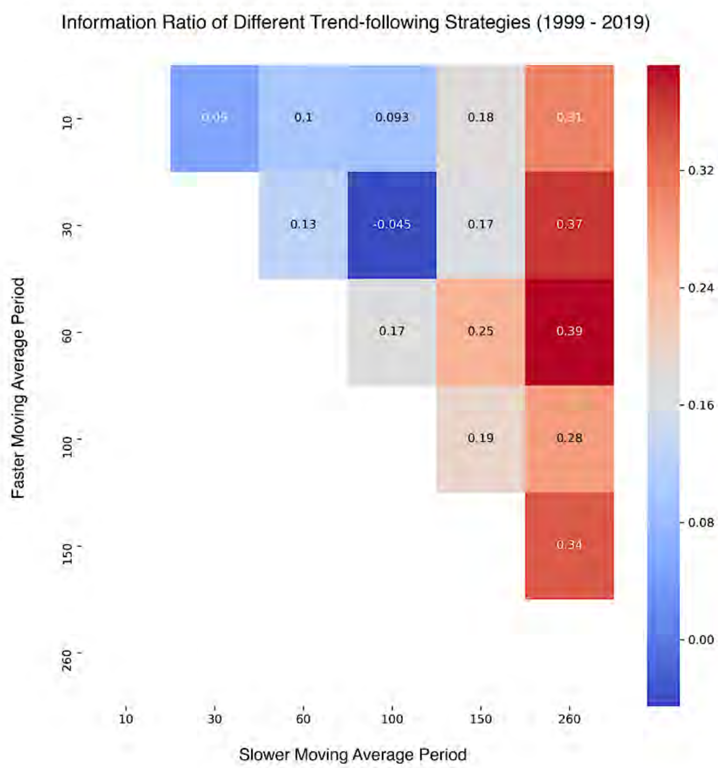


Indeed, when digging into the strategy decision based on our prediction model, we can observe that the model is mostly using trend-following strategy in 2006 - 2007 period which is a trending period leading up to the crisis. And during the crisis in 2008, the model decides to use mostly counter-trend strategy or not trading.

In summary, our combined strategy with machine learning switch is successful in striking a balance between capturing market upsides and avoid losses during crisis, which means it has partially achieved the goal of combining the best features of both trend-following and counter-trend strategies by dynamically reacting to the market and making sensible decisions.

Appendix:


Information Ratio of Different Counter-trend Strategies (1999-2019)
Average Day Range Time Window
Retracement
Long
Short
0.6
1.0
1.4
1.8
2.2
2.6
10
20
30
0.19
0.34
0.7
0.63
0.65
0.31
0.18
0.21
0.54
0.57
0.76
0.51
0.28
0.25
0.29
0.49
0.58
0.56
0.11
0.17
0.13
-0.26
0.14
-0.3
0.21
0.21
0.018
0.0015
0.057
0.0062
0.28
0.098
0.2
-0.069
0.31
0.17


Information Ratio of Different Trend-following Strategies (1999 - 2019)
Faster Moving Average Period
Slower Moving Average Period
10
30
60
100
150
260
0.05
0.1
0.093
0.18
0.31
0.13
-0.045
0.17
0.37
0.17
0.25
0.39
0.19
0.28
0.34
0.32
0.24
0.16
0.08
0.00

Correlation between Different Strategies (1999-2019)