Data 698

Spring 2022

Jimmy Ng

2022-02-15

**Final Project Proposal (version 2)**

**Research Objective**

This research examines the efficacy of technical indicators and predictive modeling in optimizing a set of investment strategy for the S&P 500. Specifically, this research makes comparison of the S&P 500 2-year prior to and 2-year after the beginning of the global pandemic.

Below analogy drawing from a hypothetical experiment can illustrate the goal and design of this research study.

A/B testing: Imagine we are conducting A/B testing for treating a disease (or enhancing physical performance). We have 500 patients. We have two treatment plans, i.e., RSI, MACD. This is a longitudinal study, and we have two study trials. First, we follow each patient for 2-year prior to the start of pandemic (Y2018, Y2019). Second, we follow them for 2-year after the pandemic began (Y2020, Y2021). At the end of each trial, we measure their performance (return of investment, ROI) based on the treatment plans. In addition, we can build classification model based on features generating from these treatment plans. We will evaluate the treatment plans by drawing conclusion from the data (based on ROI).

**Hypotheses**

1) Null hypothesis (H01): all indicators make no significant difference in terms of return of investment (ROI), i.e., ROI is not significantly different from 0

2) Null hypothesis (H02): all indicators are statistically the same in terms of ROI

3) Null hypothesis (H03): there is no significant difference in terms of ROI (generated by the indicators) among sectors

4) Null hypothesis (H04): there is no significant difference in terms of ROI (generated by the indicators) before and after the pandemic

5) Null hypothesis (H05): a predictive model built from using the features generated by the technical indicators do not perform significantly better than traditional technical analysis, strategy

**Methodology**

We can conduct this hypothetical study using simulation on historical data of the S&P 500. In the first trial, we begin the simulation of investing $10,000 in January 2018 for each stock. The trading decision (“entry”, “exit”) is based on the indicators (and different set of parameters). We will measure the ROI by the end of December 2019. Similarly for the second trial, we start investing $10,000 in January 2020 until December 2021, and then we will measure the ROI at the end.

We will evaluate the efficacy of these technical indicators based on the performance of these 500 “subjects” in these two trials.  
  
**Sample Data**

This is open-source data, e.g., quantmod. Below is a mock-up of a data frame for the experimental simulation.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **symbol** | **company\_name** | **sector** | **trial** | **RSI** | **MACD** |
| A | Agilent Technologies Inc | Health Care | pre-pandemic | $3,000 | $2,000 |
| AAL | American Airlines Group | Industrials | pre-pandemic | $1,000 | $3,300 |
| AAP | Advance Auto Parts | Consumer Discretionary | pre-pandemic | ($500) | ($700) |
| AAPL | Apple Inc. | Information Technology | pre-pandemic | $2,500 | $2,000 |
| ABBV | AbbVie Inc. | Health Care | pre-pandemic | ($900) | $200 |

GENERAL COMMENTS

Jim,

I like the fact that you’ve re-scoped your stock index trading study idea, but I need a little more clarity and specificity in the following areas of your proposal:

**Timeframe**: As you note, you’re comparing the returns of two common technical trading strategies (RSI and MACD on the S&P 500 2-year prior to and 2-year after the beginning of the global pandemic.

But why this timeframe?  What is your research question: whether the pandemic has changed the efficiency/performance of technical buy/sell indicators? You state that in H4, but I don’t understand why that is/would be. Why do you believe the pandemic has effected the use of technical indicators?     
  
The global pandemic crisis created a significant crash and then a dramatic rebound within 60 days (counting from mid-Feb) in the market in 2020. Instead of crashing, we have seen a dramatic and huge come back for majority of stocks from the S&P 500 throughout the rest of 2020. Did people just buy in without paying attention to (or simply ignoring) the indicators? In other words, people might have paid more attention to the stimulus plan(s) instead of the technical indicators. If so, that means the indicators or technical analysis is meaningless in terms of predicting or affecting the significant change in direction of the stock market. We can find out if that is the case by comparing the before- and after-pandemic performances.

**Hypotheses:** in your first (null) hypothesis, you state that “all indicators make no significant difference in terms of return of investment (ROI), i.e., ROI is not significantly different from 0.”

By “0” do you mean beta (market return)? And by “ROI” do you mean capital appreciation (I assume you’re not capturing dividend returns)?

Not that complicated. I decide to write a function that will do the "investing" based on the signal ("buy" or "sell") generated by the indicators. We will start investing $10,000. We "buy" whenever there's a signal of buying, and vice versa when the "sell" signal show up. At the end of each trial, we can calculate the ROI by subtracting the total value of each security by $10,000. The difference is the ROI returned based on the strategies according to the indicators. The key is to compare which indicator(s) or strategies (based on a predictive model that I intend to build using the features generated from the indicators) outperform the others. The idea is similar to the research paper that I had previously attached. 

**Methodology**: in the first two-year timeframe, you’re proposing to buy in January 2018 and measure ROI at the end of December 2019.

That sounds like a traditional “buy and hold” strategy, but you also state that “The trading decision (“entry”, “exit”) is based on the indicators (and different set of parameters).” What parameters or threshold will you use to trigger a buy or sell based on RSI and MACD strategies? How many trades are you envisioning? If you envision frequent trades, how can are factoring in trading commissions in your return calculations?

By parameters, for example, RSI, any value less than 30 is considered to be "oversold", whereas anything above 70 is considered to be "overbought". Similarly, for MACD, we can look for "cross-over" and "divergence" of the MACD line and the single line. We can tune the parameters to trigger the "buy" and "sell" signal.  
  
For example,  
dplyr::mutate(signal = case\_when(  
((RSI\_today >30) & (RSI\_today > RSI\_yesterday) & (RSI\_yesterday <30)) ~ "buy",  
((RSI\_today >70) & (RSI\_today > RSI\_yesterday) & (RSI\_yesterday >70)) ~ "sell",  
TRUE ~ "hold"  
)  
  
Here's just a quick example, but I will do more research to fine tune the parameters for generating more accurate signals.

I intend to write a function to take care of all these. And then I will use the "split-apply-combine" strategy to the whole S&P 500 df. The function will input a df (or ts object for each stock), and then it will output a single value for the ROI. I try to simplify the calculation of ROI by ignoring all the commission, dividend calculation. The function will apply to "daily" calculation only, so it's not for "intraday" trading.

I’ve attached your draft proposal, and uploaded it to the “Draft Proposals with Track Changes” folder in the “Feedback on Drafts” section of the course site.

Please let me know if you would care to discuss.

A