# Project Problem

By Jae Lee for GA Data Science 43

Can moving day averages predict financial market direction, such as the stock market? By employing market trading strategies based on moving day averages, can one beat market performance of simply buying and holding stocks for the long term?

The aim of my final data science project is to predict whether trading (buy-sell) strategies based on moving day averages can beat the average return of traditional buy-hold investing strategies. To fulfill this aim I will test moving day trading strategies by backtesting it against a benchmark (the S&P 500 index, a comparable stock, or other investment assets[[1]](#footnote-1)).

**Background**

Since the advent of the modern financial markets, including access to readily available long-term financial data, notable pioneers, such Robert D. Edwards, John Magee, and Richard Donchian, have developed and introduced viable methods to successfully trade the financial markets (i.e. equities, commodities, currencies, stock indices, and bonds), which can be employed alone or in addition to traditional fundamental analysis such as corporate financial balance sheets, economics, and other “traditional” supply and demand analytics to assess the merits of a stock or an investment vehicle. This new method, known as “technical analysis,” is a security analysis methodology for forecasting the direction of security prices strictly through the study of “past market data, primarily its price and volume,”[[2]](#footnote-2)\*with trading strategies based off finding *patterns* in this historical price and volume.

One prominent technical analysis method utilized widely for trading by individual traders and professional money managers is “trend following”[[3]](#footnote-3)—buying upward markets (or *selling* downward markets) and betting that that market trend (upward or downward) will continue.

In an April, 2015 abstract entitled, ““Two centuries of trend following,” researchers, Y. Lempérière, C. Deremble, P. Seager, M. Potters, and J. P. Bouchaud, after a comprehensive study, concluded that the “the existence of trends [is] one of the most statistically signiﬁcant anomalies in ﬁnancial markets.”[[4]](#footnote-4)

And “In A Century of Evidence on Trend-Following Investing,” Brian Hurst, Yao Hua Ooi, and Lasse H. Pedersen, Ph.D., studied the performance of trend-following investing across global markets since 1903 (by constructing “a time series momentum strategy”[[5]](#footnote-5) all the way back to 1903) and concluded that “Our analysis provides significant out-of-sample evidence…that trends are *pervasive features of global markets* and…Trend-following investing has performed well consistently over more than a century...”[[6]](#footnote-6)

So with “positive reported performance over long periods, suggesting that the anomaly is to a large extent *universal*, both across epochs [time] and asset classes, [our study] have established the existence of *anomalous excess returns* based on trend following strategies across all asset classes and over very long time scales.”[[7]](#footnote-7)

**Moving Day Averages**

Traders who employ a trend following strategy (buying when the price goes up; selling when it goes down) do not aim to forecast or predict specific price levels; they simply jump on the trend (when they perceived that a trend has been established [identifying it with their own parameters or set of rules]) and “ride it,” betting that the trend will persist.

To confirm, identify, and/or gauge the direction of a trend, traders may employ a simple *moving average* to see if the security price is trending up or down.

A moving average is calculated by taking the arithmetic mean of a given set of values. To calculate a basic 10-day moving average add up the closing prices from the past 10 days and then divide the result by 10.

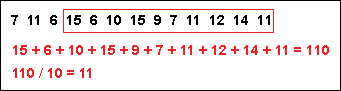


Figure 1. Calculation of a moving day average.

However, in order to continue to calculate the moving average on a daily basis, the oldest number must be replaced with the most recent closing price. See Figure 2 and 3.

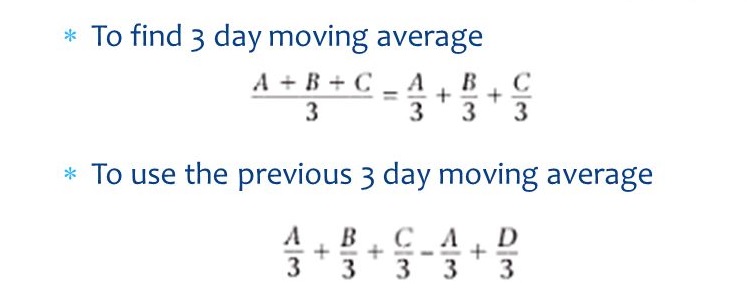


Figure 2. Formula for current and past moving average. [[8]](#footnote-8)

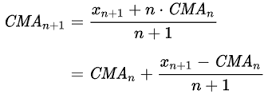


Figure 3. Two calculations of a moving day average: one current, the other lagging (by one day).

Once the values of the moving day average have been calculated, they are plotted onto a chart and then connected to create a moving average line in order to allow traders to look at smoothed data rather than focusing on the day-to-day price fluctuations that are inherent in all financial markets. See Figure 4.

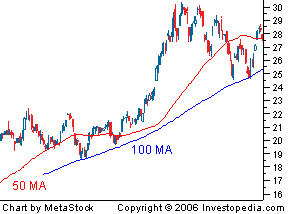


Figure 4. Price chart with a 50 and 100-day moving average.

**Trading Strategy: Moving Day Average Crossovers**

With a smoothed out visual aid, moving average lines can help a trader identify a current trend or spot a possible trend reversal. This trend-defining property makes it possible for moving averages to generate trading signals—simple buys or sells. In its simplest application, traders buy when prices move *above* the moving average and sell when prices cross *below* that line[[9]](#footnote-9), thus utilizing moving averages as a trading *method* and *strategy*.

The goal, then, of individual traders, banks, hedge funds, CTAs, and professional money managers who use moving day average crossovers trading strategies as part of their primary trading strategy are to identify the most optimal parameters (how many moving days to select) in order to satisfy their goal of identifying the investment asset (e.g. a stock) to buy or sell as it crosses from one side of a moving average to another.[[10]](#footnote-10)

**Hypothesis**

Trend following “presupposes that [asset] prices will move in long sweeps like bull [up] and bear [down] markets.”[[11]](#footnote-11) Therefore, moving averages may *not* be an effective tool when the markets are *not* trending (up or down) and may even prove to be perilous when employed during cyclical markets.

My Hypothesis: can trading signals produced from moving averages outperform a baseline of buy-and-hold throughout all market conditions?

My Experience

As a stockbroker for six (6) years, I have employed a simple moving day average crossover for my trading strategy as well, selecting 7, 18, and 21-day moving average crossover system with some success.[[12]](#footnote-12)

For my project, I will backtest the strategy I had employed as a broker, using actual data to evaluate my strategy. The simulator would generate estimated number of trades, the fraction of winning/losing trades, average profit/loss, average holding time, maximum drawdown, and the overall profit/loss with the hopes of further experimenting and refining this strategy (care taken, however, to avoid over-optimization) as well as developing money management/asset allocation algorithms, optimizing where optimization is appropriate.[[13]](#footnote-13)

**Study**

My goal will be to construct trading strategies using moving averages, formulate exit strategies upon entering a position, and evaluate this strategy with backtesting.

**An outline of any potential methods and models**

There are several studies of data science applied to the stock market/trading posted online, among them:

1. *A Simple Trading System* - <http://www.seykota.com/tribe/TSP/EA/> - Exponential Average Crossover - This page shows how an exponential average crossover system works. It also presents results from back-testing a system on an S&P continuous Panama chart.
2. *Stock Market Prediction in Python Intro* - <http://francescopochetti.com/stock-market-prediction-part-introduction/> - “The idea at the base of this project is to build a model to predict financial market’s movements. The forecasting algorithm aims to foresee whether tomorrow’s exchange closing price is going to be lower or higher with respect to today. Next step will be to develop a trading strategy on top of that, based on our predictions, and backtest it against a benchmark.”
3. The First Python Project in Data Science: Stock Price Prediction - <http://lovelearning9.blogspot.com/2015/12/the-first-python-project-in-data.html> - “The objective of this project is to predict stock price.”

However, Curtis Miller, Associate Instructor at University of Utah College of Science, posted a two-part series on applying and backtesting moving day average crossover trading strategy to the stock market using Python, *An Introduction to Stock Market Data Analysis with Python* (<https://ntguardian.wordpress.com/2016/09/19/introduction-stock-market-data-python-1/> and <https://ntguardian.wordpress.com/2016/09/26/introduction-stock-market-data-python-2/>). I will draw heavily from this study, including, with permission, some of his code and methodology.

**Methodology**

Steps

1. Install Dependencies
   1. install csv
   2. install numpy (perform calculations on our data)
   3. install scikit-learn (build a predictive model)
   4. install Matplotlib (plot our data points on a graph)
2. Collect Dataset
3. Write scripts
4. Backtest

Stock data can be obtained from Yahoo! Finance, Google Finance, or a number of other sources. In my study, I will try to obtain my data from Yahoo! Finance. My data will include:

* The average price from the past 7,18, 21 days.
* The average price for the past month.
* The average price for the past year.
* The ratio between the average price for the past 7,18, and 21 days, and the average price for the past year.
* The standard deviation of the price over the past five days.
* The standard deviation of the price over the past year.
* The ratio between the standard deviation for the past five days, and the standard deviation for the past year.
* The average volume over the past five days.
* The average volume over the past year.
* The ratio between the average volume for the past five days, and the average volume for the past year.
* The standard deviation of the average volume over the past five days.
* The standard deviation of the average volume over the past year.
* The ratio between the standard deviation of the average volume for the past five days, and the standard deviation of the average volume for the past year.
* The year component of the date.

**Datasets**

Sample dataset:

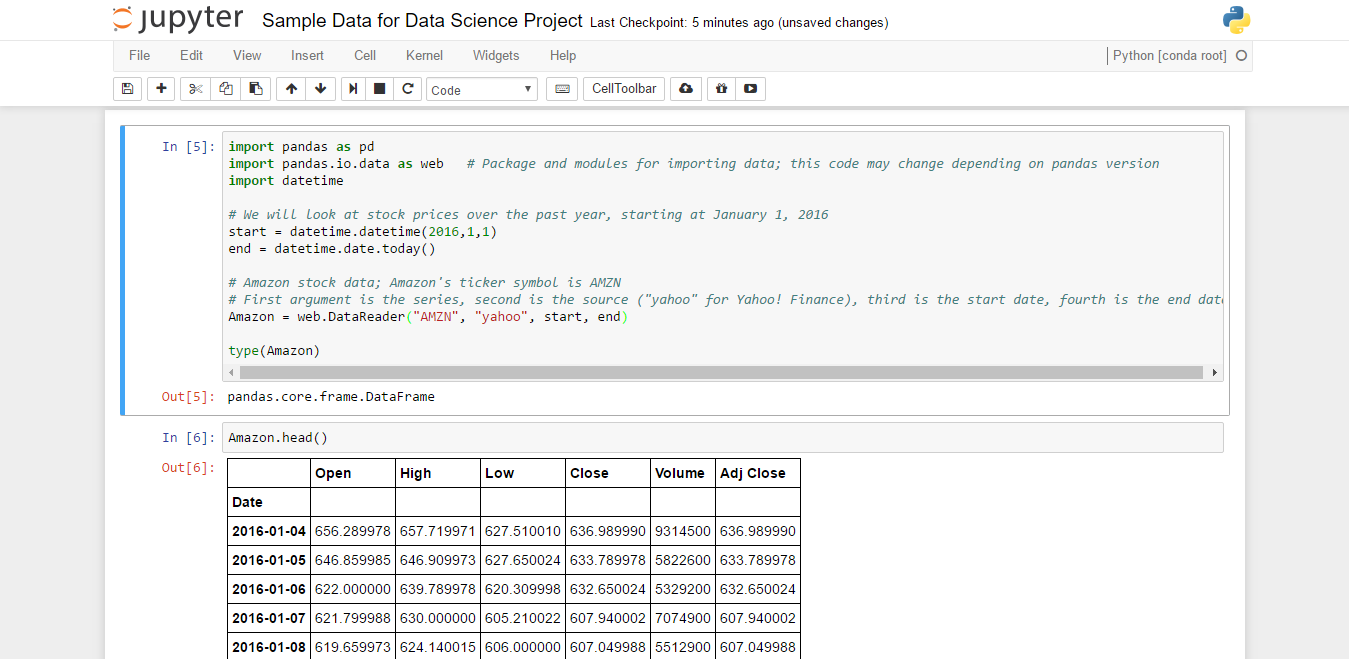
****

Figure 5. Data Sample[[14]](#footnote-14)

**Partial Data Dictionary\***

**“Open** is the price of the stock at the beginning of the trading day (it need not be the closing price of the previous trading day), **high** is the highest price of the stock on that trading day, **low** the lowest price of the stock on that trading day, and close the price of the stock at closing time. **Volume** indicates how many stocks were traded. **Adjusted close** is the closing price of the stock that adjusts the price of the stock for corporate actions.”[[15]](#footnote-15)\*\*

**Conclusion**

Pandas provides functionality for easily computing moving averages. I will employ its use by creating a 7,18, 21 days moving average for each stock data, plot it alongside the stock, and set parameters to stimulate buys/sells (buy entries and exits) based on current and past price crossover over these moving day averages. I will then create a simulation portfolio and see how my buy/sell triggers performed vs. either simply buying and holding the same stock over the same period or comparing it with any other benchmarks such as against the S&P 500 index.

1. The S&P 500 index is a stock market index based on the market capitalizations of 500 large companies having common stock listed on the NYSE (New York Stock Exchange) or NASDAQ. Being such a diversified portfolio, the S&P 500 index is typically used as a market benchmark. [↑](#footnote-ref-1)
2. *Technical analysis*, Wikipedia, <https://en.wikipedia.org/wiki/Technical_analysis>

   \*”The efficacy of…technical…analysis is disputed by the efficient-market hypothesis which states that stock market prices are essentially *unpredictable*.” – Ibid. Italics mine. [↑](#footnote-ref-2)
3. A “trend” is simply a price that is continuing to move in a certain direction. [↑](#footnote-ref-3)
4. *Two centuries of trend following*, Y. Lempérière, C. Deremble, P. Seager, M. Potters, and J. P. Bouchaud. April 12, 2014. <https://www.trendfollowing.com/whitepaper/Two_Centuries_Trend_Following.pdf> [↑](#footnote-ref-4)
5. “…an equal-weighted combination of 1-month, 3-month, and 12-month time series momentum strategies for 59 markets across 4 major asset classes—24 commodities, 11 equity indices, 15 bond markets, and 9 currency pairs—from January 1903 to June 2012.” *A Century of Evidence on Trend-Following Investing* Brian Hurst, Yao Hua Ooi, and Lasse H. Pedersen, Ph.D. <https://www.trendfollowing.com/whitepaper/Century_Evidence_Trend_Following.pdf> [↑](#footnote-ref-5)
6. Ibid. Italics mine. [↑](#footnote-ref-6)
7. *Two centuries of trend following*, Y. Lempérière, C. Deremble, P. Seager, M. Potters, and J. P. Bouchaud. April 12, 2014. <https://www.trendfollowing.com/whitepaper/Two_Centuries_Trend_Following.pdf> [↑](#footnote-ref-7)
8. Thus, a "moving" average differs from a regular mean because as new values becomes available, the oldest data points must be dropped from the set and new data points must replace them—the data set is constantly "moving" to account for new data as it becomes available. In Pandas, there are functions to compute moving (rolling) statistics, such as rolling\_mean and rolling\_std., but a shift in the column is required of that rolling statistics forward by one day precisely due to this difference of a moving average from a regular mean. *Part I – Stock Market Prediction in Python Intro*, Francesco Pochetti, September, 2014, <http://francescopochetti.com/stock-market-prediction-part-introduction/> [↑](#footnote-ref-8)
9. A move below the moving average may suggests that the market will move lower. Conversely, a cross above a moving average may suggests that the market price may move higher. [↑](#footnote-ref-9)
10. Specifically, when a price crosses above or below a certain moving average or even when one moving average (e.g., a 5-day moving average) crosses another moving day average (e.g., a 10-day moving day average). [↑](#footnote-ref-10)
11. *Richard Donchian*, Wikipedia. <https://en.wikipedia.org/wiki/Richard_Donchian> [↑](#footnote-ref-11)
12. Inspired by a similar selection by trader Ed Seykota, who, with this simple system (along with some custom modification) returned 250,000% over a 16-year period. [↑](#footnote-ref-12)
13. Including: Money management--asset allocation, e.g. knowing when or if to reduce/increase trading sizes when conditions warrant (e.g., during cyclical or black swan markets [i.e., during periods of lower or higher market volatility and market crashes]); Risk control—accepting losses by setting a stop loss; diversification of assets and others—all of which can be optimized and statistically validated with traditional data science application and algorithms. [↑](#footnote-ref-13)
14. \* This code and the following data dictionary are taken entirely (with the exception of the stock symbol) from *An Introduction to Stock Market Data Analysis with Python*by Curtis Miller, <https://ntguardian.wordpress.com/2016/09/19/introduction-stock-market-data-python-1/> [↑](#footnote-ref-14)
15. Ibid.

    \*\* This sample dataset is missing some of the data profiles listed under § Methodology, as I am still working on how to retrieve the data I have in mind. [↑](#footnote-ref-15)