

# Sentiment Analysis Trading Strategy vs. MV Portfolio with Shrinkage Estimator

**Zhiyu Cao, Beichen Zhang, Zhiyang Deng**

FE670 Final Project Presentation  
Financial Engineering Department, School of Business  
Stevens Institute of Technology

April 9, 2022

# Overview

- 1 Introduction
- 2 Description of Data Set
- 3 Methodology
- 4 Performance Comparison
- 5 Conclusion

## Motivation

Efficient-market hypothesis (EMH) is a hypothesis in classical financial economics stating that asset prices reflect all available information. If you are a strong EMH believer, you may think that beating benchmark (such as S&P 500 Index) should be extremely difficult, especially when asset price is the only source of your "Alpha". Therefore, many professional investors seek help from "**alternative data**" (e.g. news data, tweets data) to generate new "Alpha", which is the motivation of **sentiment analysis**.

## Sentiment Analysis

The goal of sentiment analysis is, generally, to take large quantities of unstructured data (such as blog posts, newspaper articles, research reports, tweets, video, images etc) and use NLP techniques to quantify positive or negative sentiment about certain assets.

For equities in particular this often amounts to a statistical machine learning analysis of the language utilised and whether it contains **bullish** or **bearish** phrasing. This phrasing can be quantified in terms of strength of sentiment, which translates into numerical values.

**P.S. Positive values reflects bullish sentiment and negative values represents bearish sentiment.**

# Description of Data

Sentdex [<http://sentdex.com/>] provides an API that allows download of their sentiment data for a wide variety of financial instruments. In our project, we utilize a sample data file provided by Sentdex, which contains almost five years (10.2012-01.2016) worth of sentiment signals, at daily resolution, for 10 stocks of the S&P500. It can be seen in next-page slides that each row contains a date, a ticker symbol and then an integer representing strength of sentiment, between +6 and -3.

# Description of Data Set

sentdex\_sample

date	symbol	sentiment_signal
10/15/2012	AAPL	6
10/16/2012	AAPL	2
10/17/2012	AAPL	6
10/18/2012	AAPL	6
10/19/2012	AAPL	6
10/20/2012	AAPL	6
10/21/2012	AAPL	1
10/22/2012	MSFT	6
10/22/2012	GOOG	6
10/22/2012	AAPL	-1
10/23/2012	AAPL	-3
10/23/2012	GOOG	-3
10/23/2012	MSFT	6
10/24/2012	GOOG	-1
10/24/2012	MSFT	-3
10/24/2012	AAPL	-1
10/25/2012	MSFT	6
10/25/2012	GOOG	1
10/25/2012	AAPL	-3

Figure: Sentedex Sentiment Sample File

10 Stocks: MSFT, AMZN, GOOGL, IBM, AAPL, XOM, CVX, SLB, OXY, COP

## Sentiment Analysis Trading Strategy

For simplicity, we present a daily-based Long-only trading strategy for those 10 stocks, in which our trading rules are defined by that: **Long a ticker if its sentiment value reaches +6** and **Close the ticker position if its sentiment value reaches -1**. Moreover, the quantity of share for each ticker is fixed.

## MV portfolio with Shrinkage Estimator

We apply MV portfolio optimization techniques to the same 10 stocks, adding **shrinkage estimators** into our optimization procedure, since we want to decrease the instability of our portfolio caused by high-correlated assets. The shrinkage estimators of expected return and co-variance matrix (Jorion, 1986 & Ledoit and Wolf, 2004) is given by:

$$\mu_g = \frac{1' \hat{\Sigma}^{-1} \hat{\mu}}{1' \hat{\Sigma}^{-1} 1} \text{ and } \hat{\Sigma}_{LW} = (1 - w) \hat{\Sigma} + w \hat{\Sigma}_{CC}$$

Details of shrinkage estimator can be found in FE670 slides (Steve Yang, 2021)

```
Running Backtest...  
LONG AAPL at 2012-10-15 00:00:00  
LONG MSFT at 2012-10-22 00:00:00  
LONG GOOG at 2012-10-22 00:00:00  
CLOSING LONG AAPL at 2012-10-22 00:00:00  
CLOSING LONG GOOG at 2012-10-23 00:00:00  
CLOSING LONG MSFT at 2012-10-24 00:00:00  
LONG MSFT at 2012-10-25 00:00:00  
LONG AAPL at 2012-11-02 00:00:00  
CLOSING LONG MSFT at 2012-11-05 00:00:00  
LONG MSFT at 2012-11-08 00:00:00  
CLOSING LONG AAPL at 2012-11-12 00:00:00  
CLOSING LONG MSFT at 2012-11-13 00:00:00  
LONG GOOG at 2012-11-13 00:00:00  
LONG XOM at 2012-11-14 00:00:00  
LONG IBM at 2012-11-14 00:00:00  
CLOSING LONG GOOG at 2012-11-15 00:00:00  
LONG GOOG at 2012-11-19 00:00:00  
CLOSING LONG XOM at 2012-11-21 00:00:00  
CLOSING LONG GOOG at 2012-11-21 00:00:00
```

Figure: A Snippet of Long-Close Action Generated by Sentiment Trading Strategy



# Performance Comparison

**Table 1** Sentiment Analysis Trading Strategy vs. MV-Shrinkage Portfolio  
from 2012:10 to 2016:01

Portfolio	Sharpe Ratio	Max Daily Drawdown	Annual Volatility
Sentiment	1.41	11.18%	13.02%
MV-Shrinkage	0.71	21.43%	23.33%
S&P 500	0.75	13.04%	13.21%

# Performance Comparison: Cumulative Return

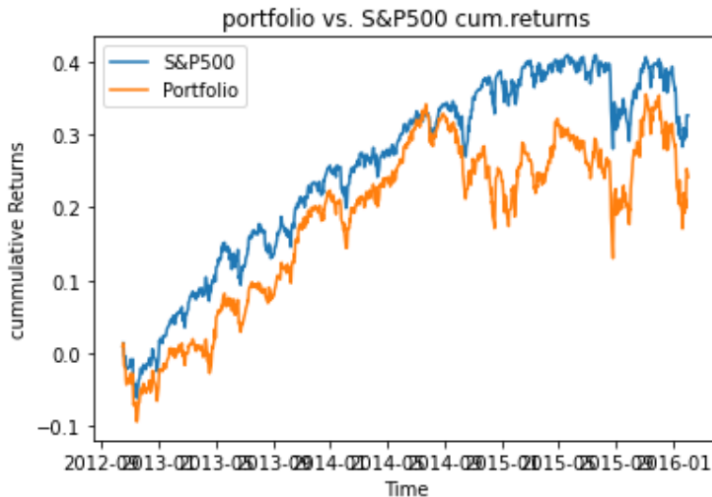


Figure: Cumulative Return: MV-Shrinkage vs. S&P500

# Performance Comparison: Cumulative Return



Figure: Cumulative Return: Sentiment Trading Strategy vs. S&P500

# Conclusion and Future Research

- Compared with classical portfolio theory, sentiment analysis trading strategy is able to generate statistically significant "Alpha" if we can have appropriate sentiment signal data (based on solid data analysis or reliable API)
- Beating the benchmark can, to some extent, rely on "alternative data", but our trading strategies will be extremely sensitive to the quality of data analysis procedure. The theoretical reasoning behind "alternative data Alpha" should be exploited.
- In the future, a research on the combination of classical portfolio theory and other new theory can become the mainstream of academic interest (such as Mean Field Control, Deep Hedging)

# Thank You For The Watching !