Index Tracking in the Structure of Fund of Funds based on Cointegration

# Abstract

Index tracking funds have grown significantly in previous decade and attracted more and more investors as an outperforming passive investment vehicle. There are two main different ways to track indices. One is called full replication, funds can take long position on all the constituents as the same weights of an index. The other tracking method is known as sample replication, funds only buy part of the stocks from a family of index stocks using different analytics tools like correlation, mixed integer programming and cointegration. In this paper, our goal is to construct a portfolio to track S&P 500 in a structure of fund of funds (FoF) using cointegration analysis. In contrast with traditional index funds, we do not buy constituent stocks directly to mimic index, we buy sector ETFs. S&P 500 consists of 11 different sectors and industries, there are numerous sector ETFs on the market. We can construct an index fund by purchasing sector funds to track S&P 500 deploying cointegration analysis to make sure long run equilibrium. In FoF structure, we can cut transaction cost enormously and reduce turn over rate, which are essential for an index fund. Index funds hold stocks directly for both full and sample replication methodologies.

Contents

[Abstract 1](#_Toc515966266)

[Introduction 3](#_Toc515966267)

[Statement of Problem 4](#_Toc515966268)

[Literature Review 5](#_Toc515966269)

[Data 5](#_Toc515966270)

[Methodology and Research Design 5](#_Toc515966271)

[Empirical Results 5](#_Toc515966272)

[Conclusion 5](#_Toc515966273)

[Limitations and Extensions 5](#_Toc515966274)

[References 5](#_Toc515966275)

# Introduction

In investment industry there are two opposite equity portfolio management philosophies, active portfolio management and passive management. Active investment is aiming to beat the benchmark index based on professional analytics and fund manager’s judgment in picking securities and determining the right moment to long and short them. Hedge fund is a typical active management portfolio. However, it is not as easy as planed in theory for active funds to outperform respective benchmarks. Based on SPIVA® U.S. Scorecard, around 83.18% of all domestic funds underperformed their benchmark for the 10-year period by middle year of 2015 in the US. Besides us stock market, many active funds failed to beat the targeted indices over the 10-year period in other major capital markets. On the contrary, passive investment pursues the same performance as targeted index over a long period of time instead of trying to beat the benchmark. A tracking fund is a classic passive product whose mission is to mimic a specified benchmark passively with buy-and-hold strategy, the benchmark can be a stock index, a commodity, bonds, even bitcoin. As opposed to underperformed active funds, index tracking funds attracted more and more investors and grown significantly through nearly a decade bull market since the worst situation in early 2009. Plenty of capital flow into passive funds rapidly. For 2017, investors poured more than $692 billion into index funds across all asset classes. For the same period, actively managed funds experienced $7 billion in outflows. Now the total asset in index funds including index mutual funds and index ETFs is about 1112 trillion in the US.

There are two main conventional methods to track indices.

One is called full replication, the fund can take long position on all the constituents of an index in the respective weights with buy-and-hold strategy, which is straightforward to implement and can achieve the precise tracking performance as long as fund managers rebalance the weights once a while. Even though full replication can closely track indices in theory, it has a few nonnegligible flaws. Full replication funds need to rebalance quite often with high volatility stock weights, which could lead to inflated costs. Liquidity is another issue, especially for small capitalization stocks, this may affect fund construction and increase the transaction costs. Low cost is a signature characteristic of passive management funds, but full replications funds cannot bring out this feature.

The other traditional tracking method is known as sample replication. Some indexes may contain large number of constituents, such as S&P Global 1200, Russell 2000. In those cases, full replication approach is not efficient to conduct, however, sample replication methodology can be appropriate. Sample replication funds need to long part of total stocks that could represent the underlying index based on correlations, risks and returns. As the funds trade a relative fewer constituents, which could significantly reduce the costs, but this may potentially cause higher tracking errors. <此处应有引用谁是最早提出了各种tracking 方法>

In additional to traditional physical funds holding a portfolio of assets, there is another alternative approach so called synthetic portfolio to replicate the performance of an index by using corresponding derivative and swaps instead of holding stocks directly. Proponents claim that synthetic funds are a better financial instrument than traditional tracking funds to track illiquid indexes at a low cost and small tracking error. However, the synthetic portfolios are born with a few risks involving counterparty risk, liquidity risk, and collateral risk. Synthetic funds are not popular in US markets due to regulation by US Securities and Exchange Commission.

In this paper, our goal is to construct a portfolio to track S&P 500 Total Return Index (SPTR) on the fund of funds(FoF) structure by using cointegration analysis that is a powerful econometrics tool that could ensure the long run equilibrium relationship between portfolio and the underlying index. As an approach of sample replication, we buy sector and industry ETFs to mimic SPTR index which consists of 11 different sectors and industries. There are numerous sector ETFs on the market, many of them have over a decade history, large asset size and high liquidity. We are going to select about 5 adequate ETFs from each sector to form a ETFs pool, then we design our portfolio based on LASSO regression to select ETFs and find corresponding weights in the co-integration system.

This paper is organized as follows. Section 2 reviews other literatures and compares different tracking methodologies. Section 3 describes targeted index S&P 500 Total Return and sector ETFs on the market. Section 4 introduces our tracking methodology cointegration and variable selection asset allocation method LASSO regression. Section 5 shows strategy implementation and empirical results for our tracking portfolios with varying number of ETFs and different rebalance strategies. Finally, section 6 makes the conclusion for our tracking strategy. In the end, we will discuss the limitations and potential research extensions based on this paper.

# Literature Review

Alexander (2005) is a pioneer who applied cointegration to passive portfolio management field. This paper deployed cointegration analysis to track a stock index, then built the long short market neutral strategy based on index tracking. The purpose of using cointegration is to identify any common stochastic trends in stock prices, and then achieve stationary tracking errors between a portfolio of stocks and the stock index over the long run. The author divided the process of constructing index tracking portfolio into two parts, selection and allocation. This paper took ‘brute force’ approach to select stocks. Firstly, pick the number of stocks to form the portfolio, then use all the combinations of stocks as possible portfolios. Next step is to optimize the weights of each stocks from every possible combination by using Engle-Granger cointegration methodology. This paper amplified the ordinary index tracking to long short market neutral strategy, which consists of a long portfolio tracking index plus, and a short portfolio tracking index minus. This long short strategy, as one of statistical arbitrage strategies, could provide double alpha opportunities in stock markets. Vast back testing results confirmed that Engle-Granger cointegration is a sound methodology to build index tracking portfolios with relative few stocks and less turnover rates.

Glova, Pastor and Sabol(2015) studied cointegration as a time series model and discovered its application in passive portfolio management. They discussed the statistical characteristics of cointegration and compared it with correlation from asset management perspective. They noted that cointegration and correlation are related, both describe the relationship between assets. Cointegration is a long-term relationship among time series. If cointegration existing, then it could ensure long run equilibrium between stock prices. Correlation is a short time statistic based on assets’ returns, that is not appropriate for constructing a long term buy and hold strategy. This literature tracked Dow Jones Industrial Average Index and Dow Jones Composite Average Index by exploiting the mean reverting property of cointegration. They used daily closed prices of indices and daily closed prices of component stocks adjusted for splits and dividends from 2000 to 2013. This paper conducted a lot of portfolios from different selection process and compared each return and risk metrics. At the end, they approved that cointegration is a right apply in passive portfolio management, which can create a comparable low volatility and low-cost tracking portfolio.

Sant’Anna, Filomena and Caldeira (2017) compared cointegration and correlation methodologies in index tracking and enhanced index tracking on Brazil Ibovespa index and U.S. S&P 100 index. This paper pointed out that both methodologies are outperforming for index tracking portfolios, but no significant advantages turn towards neither method for enhanced index tracking. The authors constructed a series of portfolios consists of at most 10 stocks by different combinations between in sample and out of sample data intervals through both approaches. Then they found different patterns between Brazilian and U.S. stock markets. There is a trade-off between tracking performance and costs in Brazilian market, which is correlation based portfolios have larger average tracking errors, but smaller turnover values, on the other hand, cointegration based portfolios have smaller tracking errors, but higher turnover rates accompanied higher cost. However, no empirical evidences revealed the similar features on S&P 100 index, tracking results did no favor neither correlation nor cointegration.

Overall, this paper failed to find robust evidences to demonstrate different characterises of cointegration and correlation in passive portfolio management area. It is worth noting that all portfolios have only 10 assets, which may be a potential reason why this paper did not generate strong findings. We will build portfolios with relatively more stocks and compare between numbers of stocks.

Numerous studies proved that cointegration is a sound and robust methodology to track index. However, there are another one crucial problems affect the tracking performance, construction costs, rebalance costs: asset selection. Asset selection is a picking art for fund managers. For our index tracking funds, assets selection helps to selection appropriate subset of assets out of total assets pool to represent the index, moreover, we can allocate different proportion of total capital to each asset in the portfolio. Alexander (2001)(cointegration and asset allocation) proposed ‘brute force’ approach to select assets. The author tested all possible combinations of a fixed number of stocks in a portfolio.

Therefore, brute force method requires huge computing power. When we track some other indexes contain large number of constitutes, it may cause explosive growth of computing, so that it is not applicable. For instance, the Russell 2000 index has 2000 stocks, if we want to pick 1200 stocks to construct a tracking fund, we may have N!/k!(N-k)!, the number of combinations is incredible giant, then the brute force becomes computationally infeasible even using modern supercomputers.

There are countless number of approaches to the asset selection problem. For linear regression, penalty methods are widely used as an effective statistical modelling technique. Regression with L1 penalty term is known as least absolute shrinkage and selection operator (Tibshirani, 1996), LASSO for short. For more details, we will discuss it in methodology section. Many academyes tested the tracking funds by using nonnegative-lasso method such as Wu et al. (2014) and Wu and Yang (2014). Yuehang Yang, Lan Wu (2016) proposed a two-stage nonnegative adaptive lasso method to do asset selection in ultra-high dimensional regression models based on adaptive lasso algorithm proposed by Zou (2006), which can deal with hundreds even thousands of stocks. They tracked CSI 300 Index that is a major index in Chinese stock market by using long and hold sample replication strategy. They did not use cointegration to ensure the long run equilibrium between tracking fund and CSI 300 Index. First stage solved the asset selection problem, they used nonnegative adaptive lasso method to select 30 stocks out of total 300 stocks, the number 30 is a predetermined number. Once they determined the stocks, second stage solved the asset allocation problem. They applied nonnegative OLS method to estimate the weights of the 30 stocks in the tracking fund. The authors did not show long time tracking performance, the results for short time were satisfactory.

# Data

In this paper we constructed a fund to track S&P 500 Total Return Index(SPTR) by using sector ETFs. We used daily close prices of SPTR and these ETFs adjusted for paying dividends and stock splits, and the adjusted closing prices are easy for us to perform analysis of historical returns. We took 10-year data from beginning of 2008 to end of 2017, this period contains 2518 trading days, downloaded data from Yahoo! Finance.

S&P 500 Total Return Index is a very similar index to standard S&P 500 index. They have the same components that comprises 500 large capitalization companies listed on NYSE and NASDAQ, and these constituents are categorized into 11 sectors: consumer discretionary, consumer staples, utilities, technology, health care, financial, energy, telecom, industrials, material, and real estate. Both SPTR and SPX are calculated in capitalization-weighted method in which the constituents are weighted based on the market value of their outstanding shares, however, the only difference is SPTR tracks the stocks with dividends are reinvested back into underlying stocks instead of tracking stock price movements only.

The total market capitalization of SPTR is about 23 trillion dollars at the end of 2017. Real Estate was separated from the Financials sector at the end of August 2016.

# Methodology and Research Design

# Empirical Results

# Conclusion

# Limitations and Extensions

# References