

“Seeking Thunder” Executive Summary

Members: Joshua Pfeffer, Samson Johnson, Bahareh Baharinezhad, Mike LaCroix, Atharva Patil

Mentor: Rongqing Ye

Github link: [Seeking Thunder Github repo](#)

Introduction

Exchange Traded Funds (ETFs) are investment vehicles that pool a variety of securities, typically aligned around a specific theme. This thematic grouping often leads to increased correlations between the prices of the constituent stocks, even if some of these stocks—referred to as "outsider" stocks—have minimal or negative exposure to the theme. This misalignment creates potential trading opportunities, as the prices of these outsider stocks may revert to their nominal values, offering a chance for profit.

Lynch et al. (2019) explored this phenomenon, finding that high-volume sell-off days for ETFs lead to predictable price movements in outsider stocks. By identifying these stocks, they were able to construct portfolios that showed significant returns over a 40-day period post-sell-off. This project builds on their work to further investigate the dynamics and potential trading strategies around ETF-induced mispricings.

Motivation

The rapid growth of ETFs, now accounting for around 30% of all trading volume on US exchanges, has introduced new dynamics in the financial markets. When ETFs experience significant trading volumes, correlations among their constituent stocks increase, even for those with little thematic relevance. This phenomenon presents trading opportunities, as highlighted by the 2015 case when the Health Care Select Sector SPDR ETF (XLV) saw a significant sell-off following a tweet by Hillary Clinton about pharmaceutical price gouging. Stocks within the ETF that were unrelated to human pharmaceuticals, such as animal medicine producers and medical equipment companies, were sold off as well, presenting an opportunity for investors to capitalize on the subsequent price correction.

By systematically identifying and investing in these outsider stocks during high-volume ETF sell-offs, investors can capture significant alpha. This project aims to develop a data-driven approach to exploit these opportunities.

Data

In our study, we analyzed eleven ETFs from the Select Sector SPDR Trust, encompassing themes from health care to entertainment. Data was collected from NPORT-P filings (Public) on the SEC EDGAR database, which provide quarterly reports on ETF compositions - data can be found in the data directory (e.g. XLV_holdings.csv). Daily stock prices within each ETF were obtained using the Python API [yfinance](#), and we measured returns based on the relative change in closing prices.

Method

We can describe our strategy in three parts.

1. When to invest?

After statistically significant high ETF trading volume and negative return, generally causing abnormally high correlations.

2. Which stocks to invest in?

We measure the correlation between individual stock returns and ETF returns, we calculate each stock's "ETF beta", defined as the coefficient of the exponentially weighted linear regression of that stock's daily returns against the daily returns of the ETF - giving greater weight to more recent returns. We define "outsider" stocks of the ETF (on a given day) as stocks whose ETF beta values fall in the bottom 10% among all stocks in the ETF.

3. How do we measure the performance of a selected outsider stock relative to the ETF?

We compare their returns with those of the ETF over the 40 days following the ETF trading volume spike. (The choice of 40 days was based on the empirical findings in Lynch, et al.)

Findings

Our analysis confirmed that average cross-correlations among constituent stocks in the ETF are abnormally high during ETF trading volume spikes. We identified the outsider stocks and implemented the strategy proposed in Lynch et al. (2019). Our results find that, from October 2019 to March 2024, this strategy does not generate returns that consistently outperform the corresponding ETFs. This difference may indicate that the strategy proposed by Lynch et al. (2019) may not have persisted since it was proposed.

Conclusion

Our analysis provides mixed results regarding the strategy proposed by Lynch et al. (2019). While we confirmed that high trading volume spikes in ETFs lead to increased correlations among constituent stocks, including those with little exposure to the ETF's theme, we did not consistently find significant returns from investing in these outsider stocks. It is possible that the strategy's effectiveness may have diminished over time.

Acknowledgement

We would like to express our gratitude to the Erdos Institute for their support and resources. Special thanks to Steven Gubkin, the Lead Instructor, and to Professor Thomas Polstra from the Quantitative Research course for his insightful contributions, and our group mentor Rongqing Ye.