IAE Nantes
Économie & Management
Pôle Sociétés

Nantes Université

Master 1 – Econometrics and Statistics
Applied Econometrics Track

Executive Summary

Index Tracking and Asset Selection Using Elastic Net Penalized Regression

Florian CROCHET
Supervised by Mr. Olivier DARNÉ

Academic Year 2024–2025

Source code: github.com/floriancrochet/master-year1-thesis

Summary

This study proposes a method to replicate the S&P 500 index with fewer stocks while maintaining high accuracy. Using daily data from 2017 to 2024, an Elastic Net regression selects the most relevant stocks. The approach reduces portfolio size and costs without sacrificing performance. It provides a practical solution for passive investment. Portfolio managers can track the index efficiently with optimized portfolios.

Keywords: S&P 500, index tracking, asset selection, penalized regression, Elastic Net.

Research Question and Study Framework

Passive investment products, such as Exchange-Traded Funds (ETFs), commonly seek to replicate indices like the S&P 500. Fully replicating such indices requires holding all constituents, which can be costly due to transaction fees, frequent rebalancing, and operational complexity. Partial replication, which involves selecting only a subset of the index's stocks, offers a way to reduce these costs while still tracking the index closely. The central question addressed here is: which subset of S&P 500 stocks can accurately replicate the index with fewer holdings?

The analysis uses daily returns from January 2017 to March 2024, covering different market phases including periods of sharp volatility such as the COVID-19 crisis. A long-only constraint is imposed to reflect the reality of most ETFs, which do not engage in short selling. One of the key challenges arises from the strong correlations between many S&P 500 stocks. Traditional selection methods often struggle with such multicollinearity, which can lead to unstable estimations and poor out-of-sample performance. The Elastic Net approach addresses this by blending two penalization techniques: Lasso, which promotes sparsity by excluding irrelevant stocks, and Ridge, which stabilizes estimation when predictors are highly correlated.

Data Used

The returns data were sourced from Yahoo Finance, consisting of 484 individual S&P 500 stocks and the index itself. Logarithmic returns were computed to stabilize variance and make the data suitable for statistical modeling. Non-trading days, such as weekends and public holidays, were removed. Outliers were adjusted to limit the influence of extreme market events. Stationarity tests (ADF, PP, KPSS) confirmed that both index and stock returns could be treated as stationary processes, validating their use in penalized regression frameworks.

Statistical Methodology

The Elastic Net regression models the index return as a weighted sum of selected stock returns, with non-negativity constraints to reflect long-only investment. The optimization problem involves minimizing both the sum of squared residuals and a penalty term that blends Lasso and Ridge penalties. This dual penalization allows the model to balance between variable selection and estimation stability. The penalty terms are controlled by two hyperparameters: α, which determines the balance between Lasso and Ridge, and λ, which controls the overall degree of shrinkage. These hyperparameters were tuned using rolling-origin cross-validation to ensure robust out-of-sample performance.

Formally, the Elastic Net estimator solves the following optimization problem:

$$\hat{\beta} = \underset{\beta \geq 0}{\arg\min} \left\{ \frac{1}{2} \sum_{i=1}^{N} \left(y_i - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^{p} \left(\alpha \beta_j^2 + (1-\alpha)|\beta_j| \right) \right\}$$

where β represents the portfolio weights, α controls the balance between penalties, and λ determines the shrinkage intensity.

Results

The estimation process identified 198 stocks out of the 484 available as sufficient to closely track the S&P 500 index. This represents a substantial reduction in portfolio complexity while preserving replication quality. The hyperparameter values selected through cross-validation were $\alpha = 0.35$ and $\lambda = 0.00000368$, indicating a moderate balance between sparsity and stability.

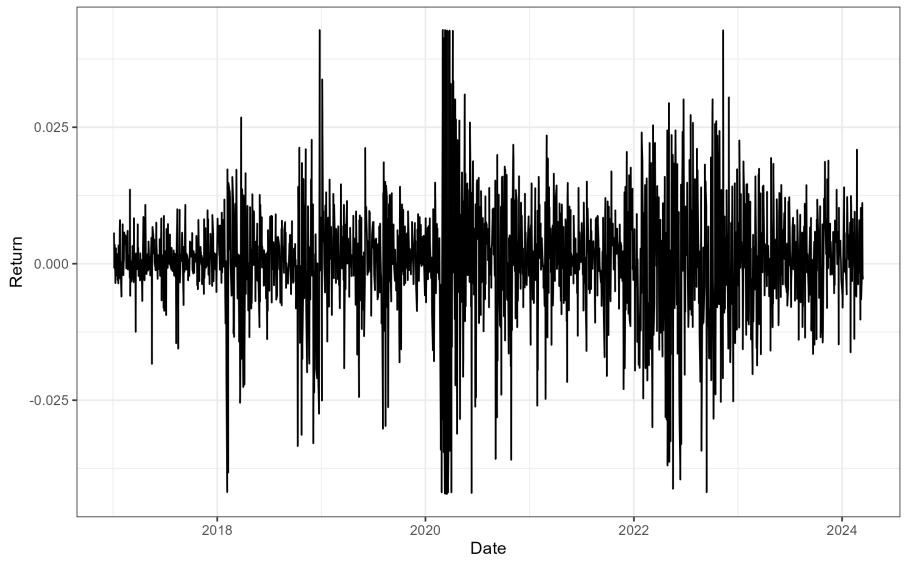
The performance of the selected portfolio demonstrates the effectiveness of this approach. Over the full period, the tracking error stands at 1.98%, indicating a very close replication of the index. The information ratio reaches 1.77, highlighting the model's efficiency in generating excess return relative to risk. Correlation with the S&P 500 is virtually perfect, and the estimated beta is close to one, confirming that the constructed portfolio moves in line with the index. Moreover, Jensen's alpha is positive, suggesting that the portfolio delivers a modest but consistent outperformance after adjusting for systematic risk.

Table 1: Performance Metrics

Metric	**Value**
Tracking Error	0.01981
Active Return	Positive
Information Ratio	1.77
Correlation with S&P 500	≈ 1
Beta	≈ 1
Jensen's Alpha	Positive

Figure 1 below illustrates the daily logarithmic returns of the S&P 500 index over the period studied. Volatility spikes, such as during the COVID-19 crisis in early 2020, are clearly visible, underscoring the challenge of accurately tracking the index through highly volatile markets.

Figure 1: Logarithmic Returns of the S&P 500



The constructed portfolio effectively mirrors these dynamics, maintaining close alignment with the index even during periods of market stress. By selecting fewer than half of the index's constituents, the Elastic Net approach significantly simplifies portfolio management while maintaining excellent replication fidelity. This reduction in portfolio size has important practical implications for ETF managers, as it reduces trading costs, simplifies operational management, and allows for more flexible portfolio construction without materially sacrificing performance.

Conclusion

Elastic Net penalized regression offers a powerful and practical framework for constructing sparse, stable, and efficient index-tracking portfolios. It enables ETF managers to replicate the S&P 500 with fewer stocks while achieving high replication accuracy and reducing operational costs. Beyond its application to the S&P 500, this methodology could be extended to other indices or adapted to incorporate dynamic rebalancing rules, alternative penalty structures, or additional asset classes. For asset managers and financial engineers, such approaches provide valuable tools to design passive investment products that are both cost-effective and robust across market regimes.

Bibliography

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 67(2), 301-320.