Energy vs. Tech: Quantifying Divergent Risk Drivers in S&P 500 Through Independent Component Analysis

Final Project

XiLong, Jinhao Guo, Haoran Li  
  
Humboldt-Universität zu Berlin  
  
Digital Economy & Data Analysis  
  
Professor WolfGang  
  
July 20th, 2025

# Abstract

This paper applies Independent Component Analysis (ICA) to the financial characteristics of S&P 500 component stocks to identify latent risk factors that drive variations in stock performance. Using FastICA, we extract independent signals from financial indicators such as annual return, volatility, skewness, and kurtosis. We find that three components—IC1, IC2, and IC3—capture key systemic, sectoral, and unique risk features. Case studies of energy companies suggest practical investment strategies based on tail risk exposure and industry volatility. Our findings support the integration of statistical decomposition techniques in portfolio construction and risk modeling.

# 1. Introduction

Financial markets are characterized by complex, multidimensional risks. Traditional risk indicators, such as volatility and skewness, offer partial insights into stock behavior. This paper investigates whether ICA can uncover hidden risk patterns in the S&P 500, particularly focusing on sectoral divergence between energy and technology companies.

# 2. Data and Methodology

We use the dataset SDA\_2020\_St\_Gallen.csv, which includes annualized return, standard deviation, skewness, and kurtosis for S&P 500 companies. The FastICA algorithm is applied to extract statistically independent components from these variables. The dataset includes the following features:  
 annual return log: Log-transformed annual returns  
 Std: Standard deviation of returns (volatility)  
 Skewness: Distribution asymmetry  
 Kurtosis: Tail thickness or 'peakedness'

Unlike PCA

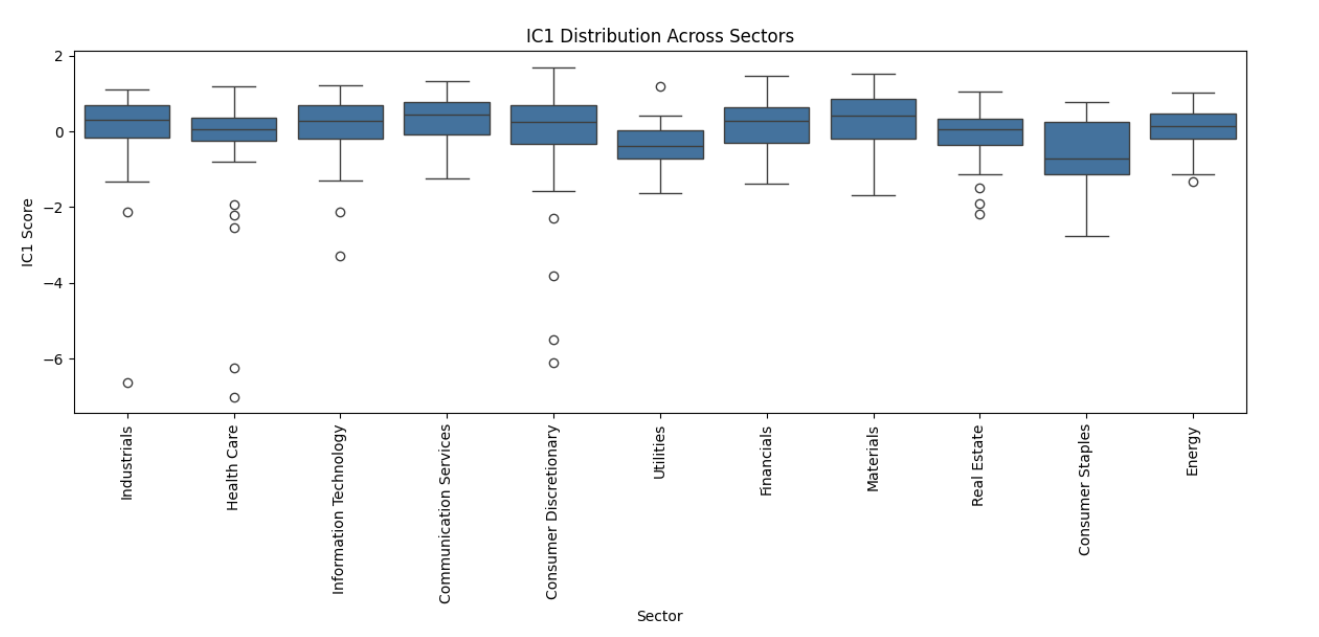
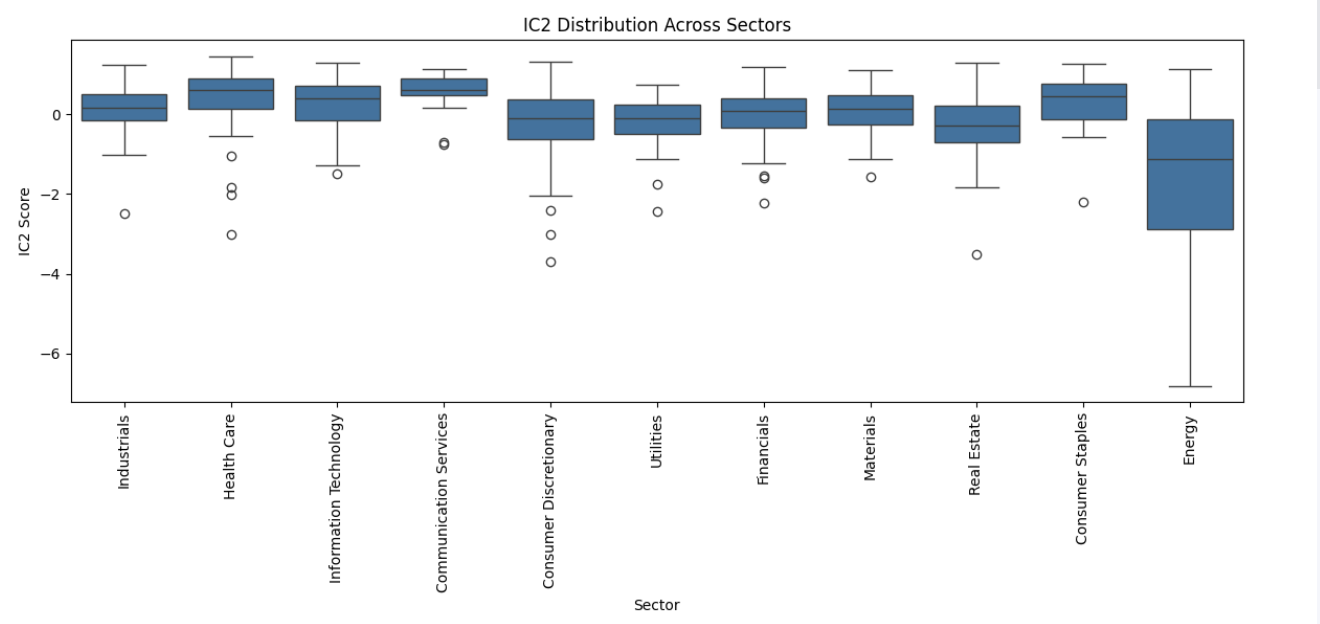
, which identifies directions of maximum variance, ICA aims to find components that are statistically independent. This is more appropriate for revealing latent drivers of financial behavior, particularly when analyzing stock risk and sectoral characteristics.

# ICA Results

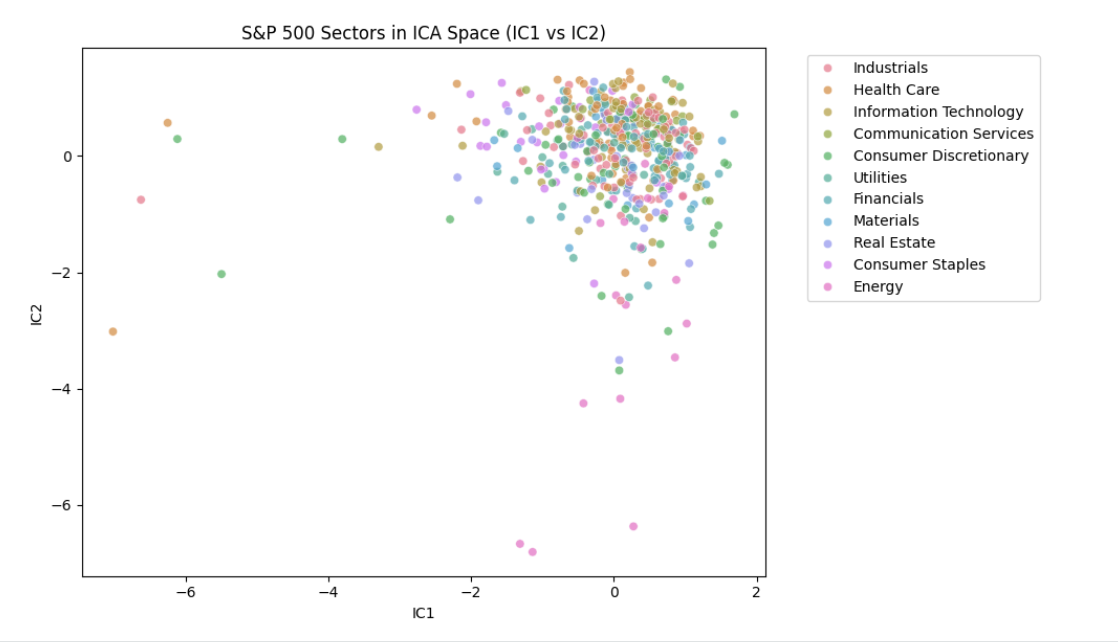
Using ICA algorithm, (assume that all component are independent)

The ICA loadings on financial metrics are as follows:

|  |  |  |  |
| --- | --- | --- | --- |
| Metric | IC1 | IC2 | IC3 |
| Annual Return (log) | 0.0306 | 0.3900 | 0.8003 |
| Std Deviation | 0.2557 | -0.4954 | -0.6507 |
| Skewness | -0.5570 | 0.7808 | -0.2157 |
| Kurtosis | -0.5521 | -0.8176 | 0.1433 |

The interpretation of components is as follows:  
 IC1: Positively correlated with volatility and negatively with skewness/kurtosis, likely reflecting systemic market risk.  
 IC2: Highly positive skewness and low kurtosis — represents tail risk.  
 IC3: High on return and low on volatility — indicating return-risk trade-off or risk-adjusted growth.

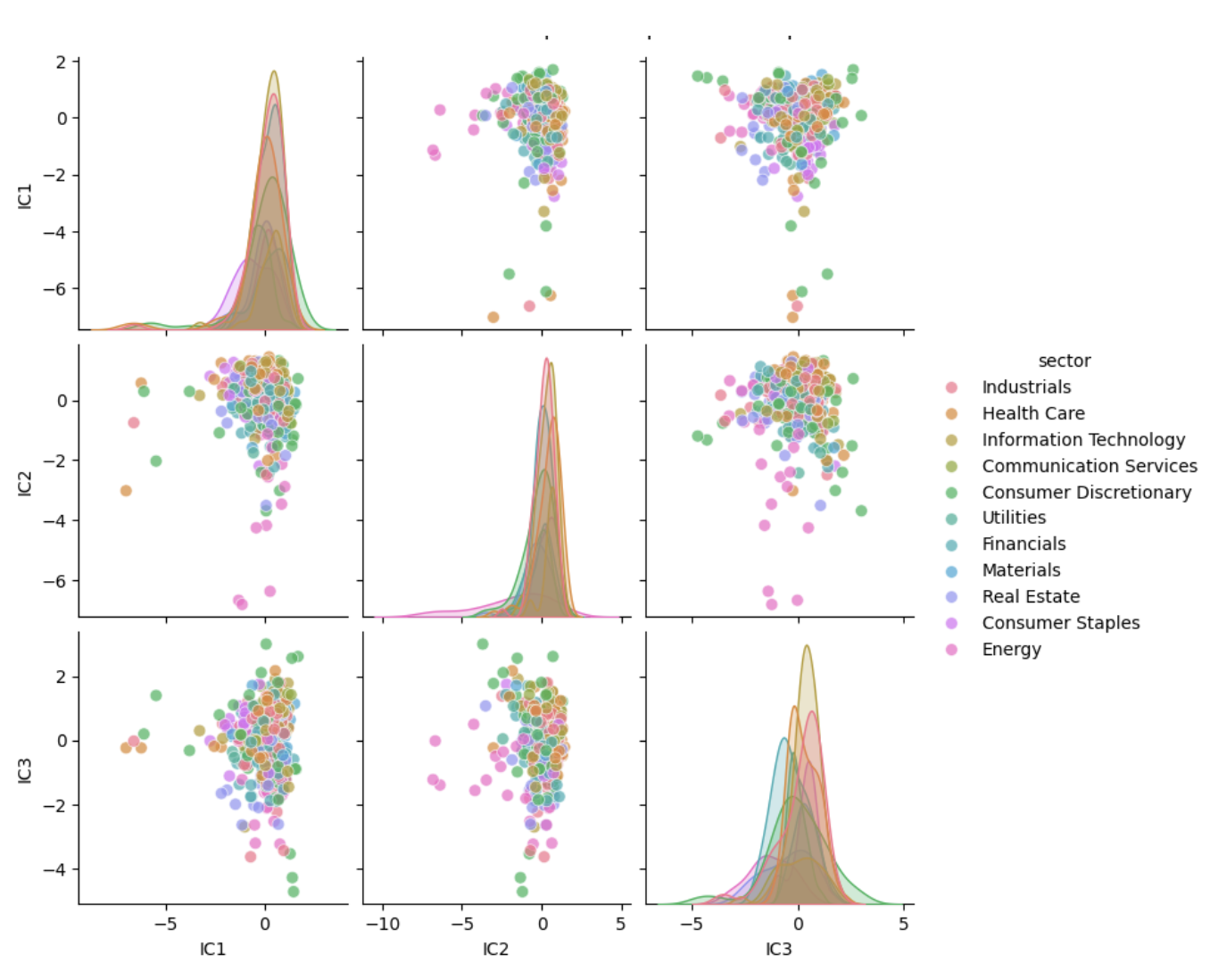
# 4. Sector-Level Interpretation

In the ICA space, stocks from the same sector tend to cluster together, suggesting they are driven by similar latent factors. For example, Information Technology and Health Care companies often overlap in IC1-IC2 space, likely due to shared exposure to innovation cycles or macroeconomic policies. Energy companies, by contrast, show wider dispersion across components, reflecting their exposure to commodity markets and geopolitical factors.

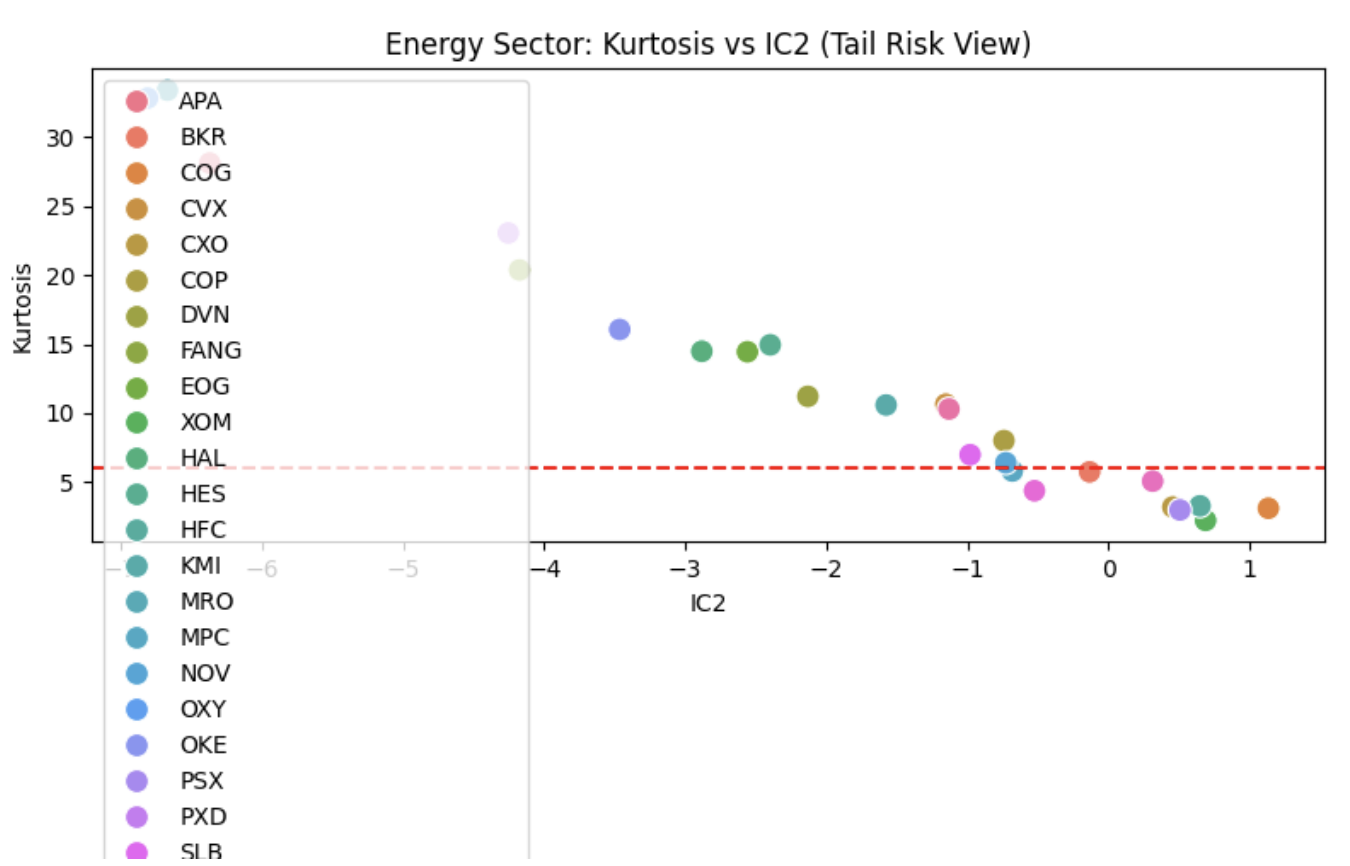
To further understand how the independent components interact and how sectoral patterns emerge, we generate a pairwise plot of IC1, IC2, and IC3. This visualization helps assess:

The independence between components (no strong correlations),

The clustering behavior of sectors in different IC projections,

**** The spread and density of scores across sectors.

# 5. Case Study: Energy Companies

To explore practical investment implications, we analyze several energy companies with different IC2 and kurtosis scores to infer tail risk and volatility characteristics.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Company | Std Dev | Kurtosis | IC2 | Interpretation |
| SLB | 0.77 | 6.99 | 0.77 | Stable; low tail risk |
| KMI | 0.75 | 10.60 | 0.75 | Some extreme risk events |
| FTI | 0.75 | 4.38 | 0.75 | Very low tail risk |
| OKE | 0.74 | 16.07 | 0.74 | High tail risk; caution |
| MPC | 0.72 | 5.81 | 0.72 | Balanced profile |

# 6. Investment Implications

Based on IC2 and kurtosis metrics, we derive general investment strategies for energy companies:

High IC2 + Low Kurtosis (e.g., FTI, SLB): Safe, stable, suitable for risk-averse portfolios.  
 High IC2 + High Kurtosis (e.g., KMI, OKE): Good for yield, but require diversification.  
 Low IC2: More cyclical, suitable during commodity upswings.

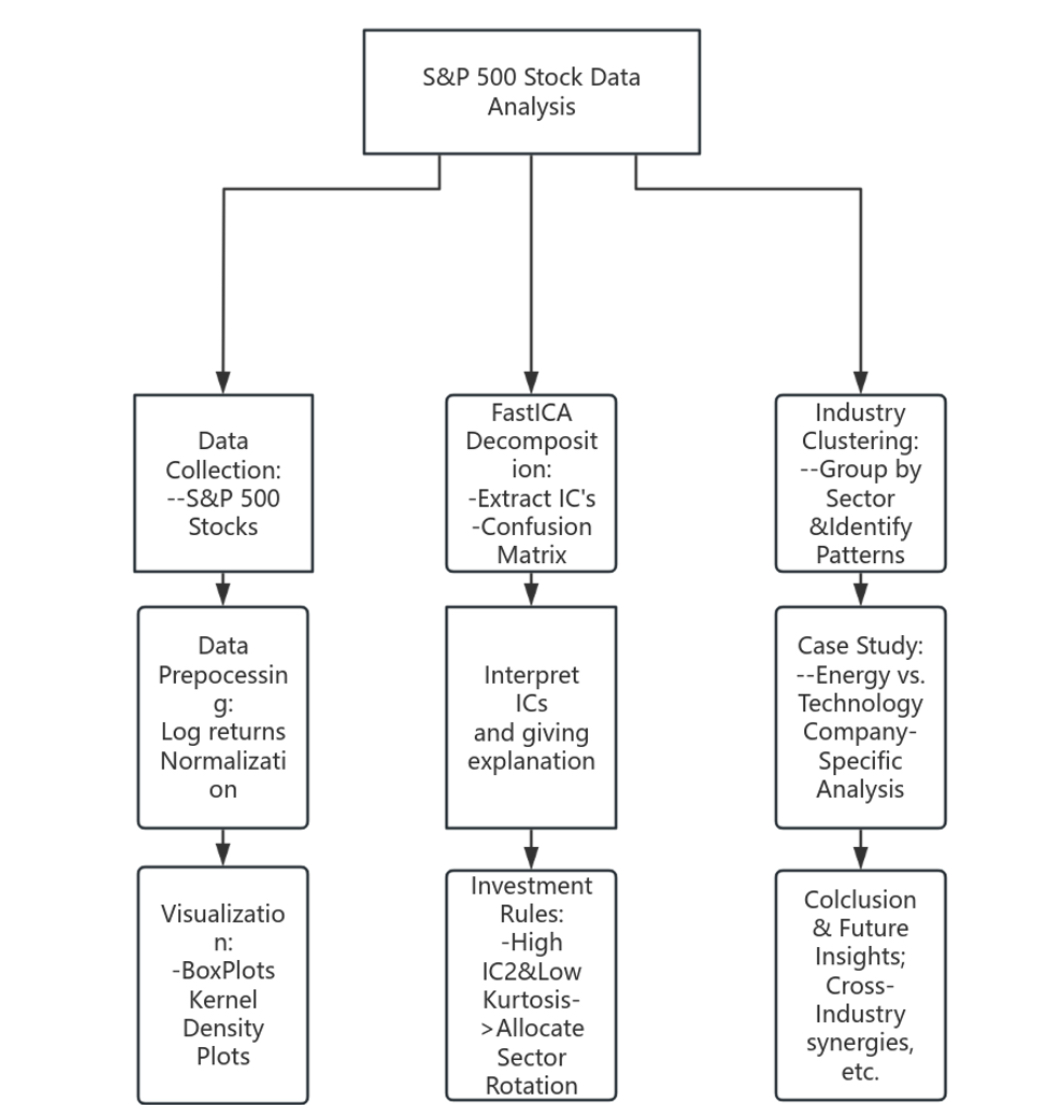
# 7. Conclusion and Future Directions

This study applies Independent Component Analysis to financial indicators of S&P 500 stocks to uncover hidden risk factors. We find three interpretable components that relate to systemic risk, tail risk, and growth-risk trade-offs. Our case studies demonstrate how these components can guide portfolio decisions, especially in volatile sectors like energy.

Future work could include time-series ICA, dynamic factor models, or integrating valuation metrics (e.g., P/E ratio) and ESG scores to improve explanatory power.

**8.Our work**

In this project, we analyzed S&P 500 stock data using Fast Independent Component Analysis (FastICA) to uncover latent risk factors. After collecting financial indicators—log returns, standard deviation, skewness, and kurtosis—we standardized the data and extracted three independent components. These components were interpreted as systemic risk (IC1), tail risk (IC2), and return-volatility tradeoff (IC3). We visualized sector distributions in ICA space and observed clustering behaviors, especially in technology and healthcare. A focused case study on energy companies revealed how IC2 and kurtosis together reflect tail risk. Based on these insights, we proposed investment strategies linking ICA profiles to risk-adjusted allocation.



# References

Hyvärinen, A., & Oja, E. (2000). Independent Component Analysis: Algorithms and Applications. Neural Networks.  
 Dataset: SDA\_2020\_St\_Gallen/sp500\_dataset.csv  
 Quantinar Platform: https://quantinar.com