Ablation Study for Correlation Matrix Clustering for Statistical Arbitrage Portfolios

Yifan Gu, Khang Nguyen, Lunji Zhu

Department of Mathematics
University of California, Los Angeles

June 11, 2025

UCLA College | Physical Sciences Mathematics

Introduction

Introduction

- Financial data suffers from the curse of dimensionality and ill-conditioned problems.
- [Jin et al., 2023, Khelifa et al., 2024] proposed constructing portfolio strategies using signed graph clustering methods such as SPONGE $\rightarrow$ over 10% annualized returns and statistically significant Sharpe ratios above one.
- We conduct further ablation studies to see the effect of using different modifications to construct the signed graph-based portfolios

Main contributions

We conduct an ablation study on several modifications of the trading strategy described by [Jin et al., 2023].

- Construct the stocks available in each trading windows individually, thereby increasing the number of stocks in the trading universe by 4-5 times;
- Introduce winsorization to raw and residual returns to reduce outlier effects;
- Only construct portfolios using the best clusters, as measured by a correlation density metric that we introduced;
- Put more weights onto stocks that give the strongest signals in each cluster instead of using a uniform weighting scheme;
- Evaluate our portfolio performance via a new metric of arbitrage success rate.

Methodology

Data preprocessing

We base our data preprocessing on that of [Jin et al., 2023] for the CRSP dataset, with notable differences:

- Our sample period is between January 2000 to December 2021, one year shorter than [Jin et al., 2023]
- For each trading period, our portfolio contains all stocks that have complete close price information, yielding a universe of approximately $2000 - 3000$ stocks. [Jin et al., 2023]'s universe has around 600 stocks

Data preprocessing (continued)

The following steps are replicated from [Jin et al., 2023].

- Market residual return: record the idiosyncratic component:

$$R_{i,t}^{res} = R_{i,t} - \beta_i R_{mkt,t},$$

- Correlation matrix at time $T + 1$:

$$C_{i,j} = \frac{\sum_{t=T-w+1}^{T} (R_{i,t}^{res} - \overline{R}_i^{res})(R_{j,t}^{res} - \overline{R}_j^{res})}{(w-1)\sigma_i \sigma_j},$$

where

- T: the window length for market, w: window length for correlation matrix
- $R_{i,t}$: raw return of stock i at time t
- β_i: beta coefficient of the stock i in the sliding window
- $R_{mkt,t}$: market return at time t (SPY ETF)
- σ_i: standard deviations of returns for the stock i

Data preprocessing modification – winsorization

- We set a threshold ϵ (typically 1% or 5%) so that the smallest and largest ϵ fraction of our data is clipped to the respective values exactly at the ϵ percentile in the cumulative data distribution.
- Because we are more interested in the collective behavior of stocks at each given time than the trend of individual stocks over its time series, we decide to winsorize across stocks.
- We experiment with winsorizing either the raw or residual returns.

Graph clusterization

We want to partition the graph nodes into k clusters such that most edges (correlations) within clusters are positive, and most edges between clusters are negative.

SPONGE [Cucuringu et al., 2019] was used by [Jin et al., 2023] to do this.

We hypothesize that there's a mean reversion phenomenon happening within each cluster independent of the market.

Winners and losers identification

On each day s over the w-day lookback window, define the cluster mean return as

$$\bar{R}_{C,s} = \frac{1}{|C|} \sum_{i=1}^{|C|} R_{i,s},$$

where $R_{i,s}$ represent raw returns.
A threshold p for winners and losers is set so that if

$$\sum_{s=T-w}^{t-1} (R_{i,s} - \bar{R}_{C,s}) > p$$

then stock i is identified as a winner $\rightarrow$ take short position. Conversely, when

$$\sum_{s=T-w}^{t-1} (R_{i,s} - \bar{R}_{C,s}) < -p,$$

stock i is a loser $\rightarrow$ take long position

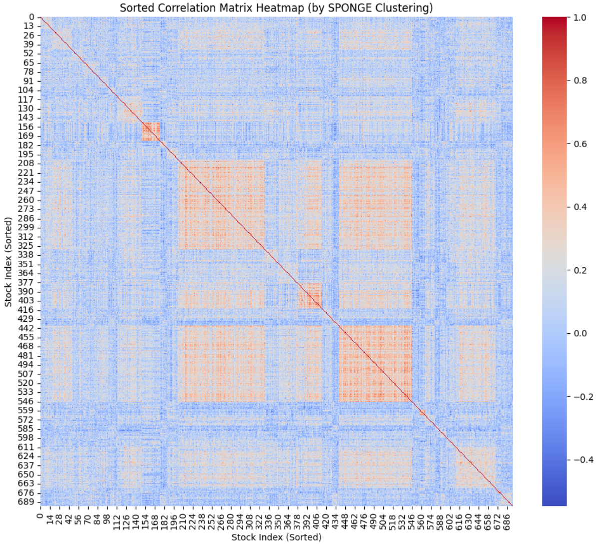

Figure: Heat map of the correlation matrix based on the given return data in our dataset. A clear distinction between strong and weak correlations is observed.

Cluster selection (continued)

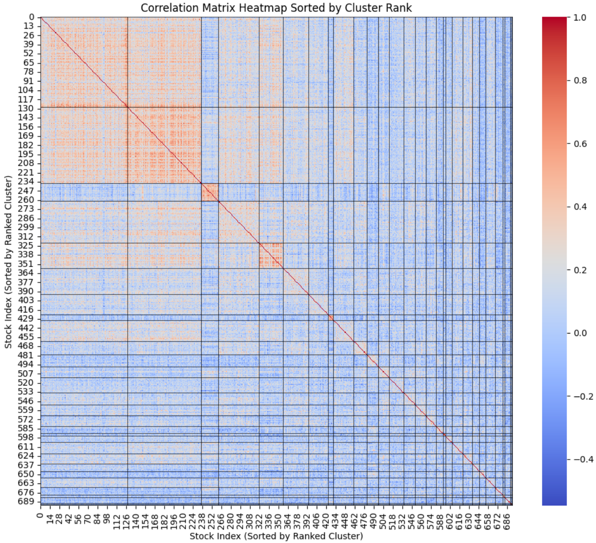

Figure: Correlation matrix with the entries grouped by the cluster and the clusters ranked by their performance with respect to the objective function of SPONGE method. The top left cluster represents the "best cluster" and the bottom right cluster represents the "worst cluster" under the evaluation of SPONGE objective function. We artificially set number of cluster to be $k = 30$ (aligned with [Jin et al., 2023]) and $\tau^+ = \tau^- = 1$ as a default setting.

Cluster selection (continued)

However, we are more interested in the positive intra-correlation of clusters instead of the negative inter-correlation between clusters.

The SPONGE objective function is defined as:

$$\frac{\mathrm{cut}_{G^+}(C,\overline{C}) + \tau^- \mathrm{vol}_{G^-}(C)}{\mathrm{cut}_{G^-}(C,\overline{C}) + \tau^+ \mathrm{vol}_{G^+}(C)}, \quad \text{which summarizes } \begin{cases} \frac{\mathrm{cut}_{G^+}(C,\overline{C})}{\mathrm{vol}_{G^+}(C)} \\ \frac{\mathrm{vol}_{G^-}(C)}{\mathrm{cut}_{G^-}(C,\overline{C})} \end{cases}$$

Cluster selection (continued)

For sake of statistical arbitrage, we care more about the positive intra-correlation. Thus, we consider *Correlation Density*:

$$D_C = \frac{\sum_{i,j \in C, i \neq j} W_{ij}}{|C|^2 - |C|}$$

as a measure of positive intra-correlation. It is the **average of inter-node edges** within a cluster.

Cluster selection (continued)

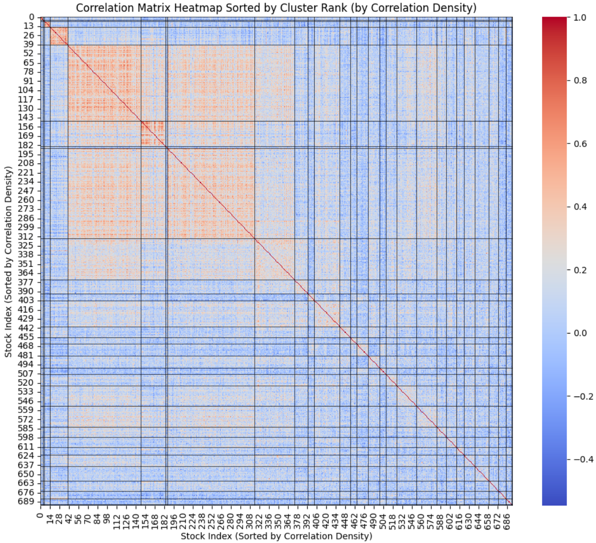

Figure: Correlation matrix with the entries grouped by the cluster and the clusters ranked by their performance with respect to Correlation Density function we just defined. The top left cluster represents the "best cluster" and the bottom right cluster represents the "worst cluster" under the evaluation of correlation density function.

Good clusters are good, bad clusters are bad.

Cluster selection (continued)

Clust.	CD	Clust.	CD	Clust.	CD
27	0.5612	12	0.2270	15	0.1701
6	0.5370	9	0.2227	5	0.1602
20	0.4698	7	0.2220	4	0.1398
8	0.4546	10	0.2130	3	0.1394
18	0.3819	23	0.2045	24	0.1338
14	0.3783	28	0.1945	19	0.1295
26	0.3287	0	0.1924	22	0.1218
11	0.3191	17	0.1914	16	0.0000
25	0.2518	2	0.1804	29	0.0000

Table: Clusters sorted by Correlation Density, by the matrix above as an example.

Cluster selection (continued)

In summary, we perceive more intra-correlated clusters are better for us to perform statistical arbitrage. Thus, we discard some bad clusters and only trade with stocks in good clusters.

Portfolio construction

[Jin et al., 2023] use 1 unit of capital to take short and long positions. This ensures market neutrality, and implies that the gross exposure for each cluster is 2 units of money.

For example, if we have 5 winners and 4 losers in one cluster, then we short sell \$0.20 on each winner and bet \$0.25 on each loser.

This gives rise to a cost-neutral trading strategy.
Downside: this does not take into account the signal strength

Portfolio construction modification – weighting strategies

Let $x_1 \geq x_2 \geq \ldots \geq x_n \geq -y_m \geq \ldots \geq -y_2 \geq -y_1$ be the deviation from the cluster mean of the winners and losers, respectively. We explore several different ways to assign capital to the stocks inside each cluster based on the signal strength, i.e. how far removed from the cluster mean a given stock's current return is:

- Linear: we short the i-th winner by $\frac{x_i}{\sum_{i=1}^{n} x_i}$ and long the j-th loser by $\frac{y_j}{\sum_{j=1}^{m} y_j}$.
- Exponential: we short the i-th winner by $\frac{\exp(x_i)}{\sum_{i=1}^{n} \exp(x_i)}$ and long the j-th loser by $\frac{\exp(y_j)}{\sum_{j=1}^{m} \exp(y_j)}$.

Trading strategy execution

Pseudo-code

count $= 0$
for day in all_trading_days **do**
    count $+= 1$
    if count $> \ell$ **or** cumulative_PnL $> q$ **then**
        record PnL
        get into new position
        count $= 0$
    end if
end for

We hold our position until a lookforward window of ℓ days is reached, or a win threshold q for cumulative PnL is exceeded, whichever is earlier.

Evaluation criteria

- PnL: as already introduced.
- Sharpe ratio: $SR = \frac{\text{mean}(PnL)}{\text{stdev}(PnL)} \times \sqrt{252}$.
- Arbitrage success rate: for each trading period, if the cumulative PnL for that trading period exceeds the winning threshold (either it terminates earlier or not), then we consider this trading period to be "successful." If not, then we say it is not "successful." We compute the success rate of each trading execution.

Empirical results

Winsorization

Figure: Cumulative PnL from experiments with 4 different types of winsorization compared to the baseline.

Winsorization (continued)

Run Name	Final PnL	Sharpe Ratio	Success Rate
baseline	0.983313	0.935421	0.527091
winsorize raw 0.05	0.438600	0.567879	0.465858
winsorize res 0.05	1.058808	0.961939	**0.538955**
winsorize raw 0.01	0.548323	0.627381	0.494918
winsorize res 0.01	**1.086277**	**1.022078**	0.533362

Table: Performance comparison of different winsorization strategies.

Cluster selection

Setting #	Original Cluster #	Trading Cluster #
baseline	40	40
1	40	5
2	40	10
3	40	20
4	40	30
5	5	5
6	10	10
7	20	20
8	30	30

Table: The Experiment Settings for Cluster Selection. The row in blue is the run that we failed to implement since the SPONGE package constantly reports an error that we cannot fix. All other results are reported and discussed.

Cluster selection (continued)

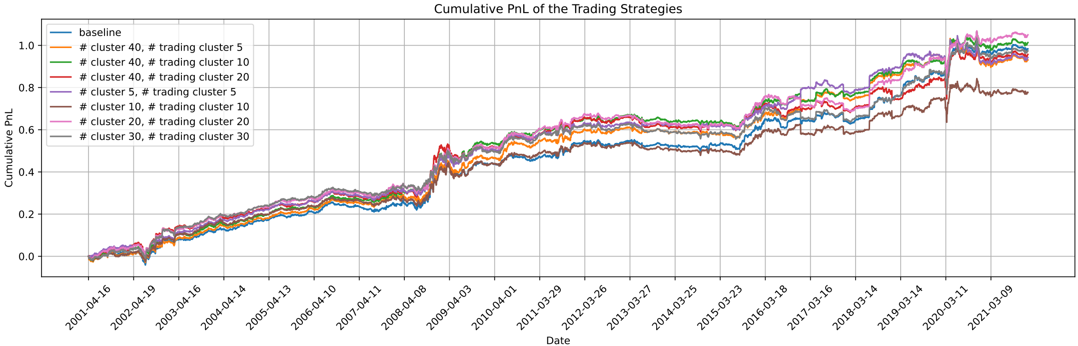

Figure: Overview of Result for Cluster Selection

Cluster selection (continued)

Run Name	Final PnL	Sharpe Ratio	Success Rate
baseline	0.983313	0.935421	0.527091
40 choose 5	0.930600	0.855173	0.541011
40 choose 10	1.013241	0.942249	**0.547154**
40 choose 20	0.957010	0.873160	0.545766
5	0.942771	0.870318	0.543015
10	0.776835	0.716692	0.546968
20	**1.049930**	**0.963611**	0.537377
30	0.973835	0.924952	0.539086

Table: Performance comparison of different clustering strategies

Stock weighting

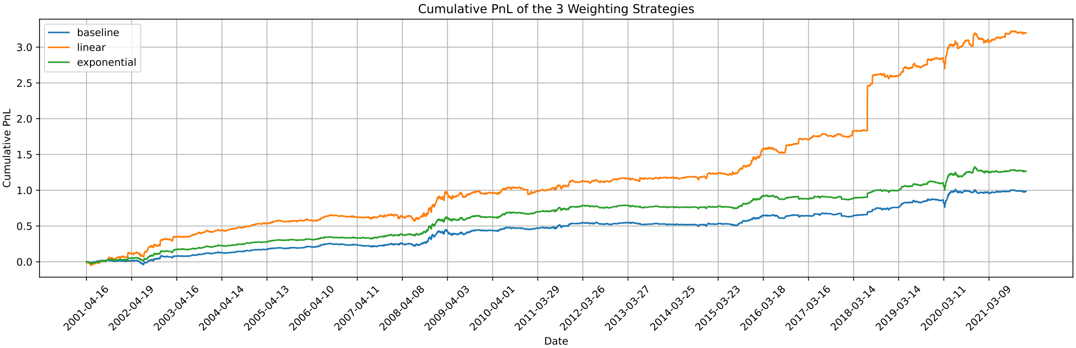

Figure: Methods that put more weight on stocks with stronger signals outperform the baseline

Stock weighting (continued)

Run Name	**Final PnL**	**Sharpe Ratio**	**Success rate**
`uniform`	0.983313	0.935421	0.527091
`linear`	**3.200949**	0.917903	**0.614686**
`exponential`	1.268054	**1.126639**	0.549596

Table: Performance comparison of different weighting strategies

Conclusion

Takeaways

- **Winsorization:** Winsorizing residual returns is a good idea; winsorizing raw returns is a bad idea.
- **Cluster selection:** Merely "selecting good clusters" does not necessarily help.
- **Weighting:** can give really good results both in term of PnL and Sharpe ratio

Future Work

- **Winsorization:** winsorize deviations from cluster means for winner/loser reweighting
- **Cluster Selection:** design better criteria for deciding which clusters to keep/discard; verify the discrepancy of trading performance between good and bad clusters.
- **Weighting**: explore thresholding
- Fix potential scaling issue in the PnL calculation.

Acknowledgments

The presenters gratefully acknowledge Professor Mihai Cucuringu, for leading this seminar and offering valuable advice for our project.

References I

Cucuringu, M., Davies, P., Glielmo, A., and Tyagi, H. (2019).
Sponge: A generalized eigenproblem for clustering signed networks.
In *The 22nd International Conference on Artificial Intelligence and Statistics*, pages 1088–1098. PMLR.

Jin, Q., Cucuringu, M., and Cartea, Á. (2023).
Correlation matrix clustering for statistical arbitrage portfolios.
In *Proceedings of the Fourth ACM International Conference on AI in Finance*, pages 557–564.

Khelifa, N., Allier, J., and Cucuringu, M. (2024).
Cluster-driven hierarchical representation of large asset universes for optimal portfolio construction.
In *Proceedings of the 5th ACM International Conference on AI in Finance*, pages 177–185.