SAPIENZA
UNIVERSITÀ DI ROMA

A Unified Framework for Distance Based Pairs Trading via Graph Laplacians

Professors:

Juri Marcucci

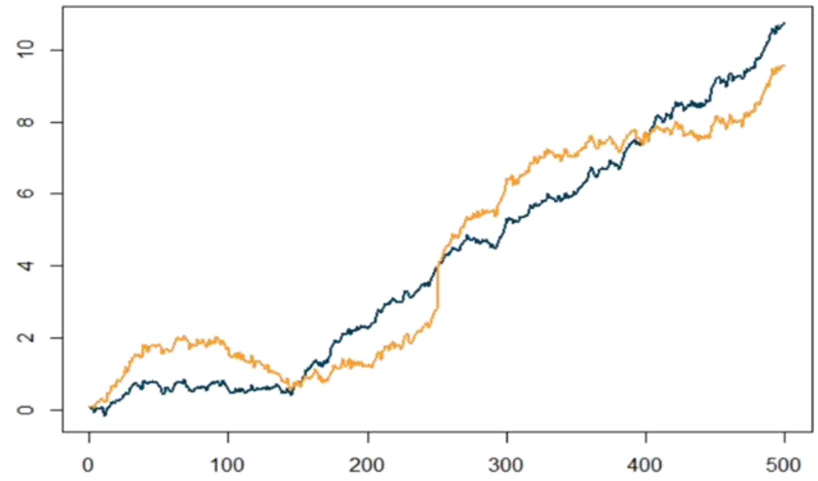
Daniele Bianchi

Student:

Michelangelo Saveriano

Introduction (pairs trading)

- Pairs trading idea:
  1. first find two securities whose prices move together
  2. when they diverge, buy the undervalued asset and sell the overvalued one



- In the quasi-multivariate framework, a security is traded against a basket of securities.

- In the distance-based approach various distance metrics are leveraged to identify comoving assets.

Our contribution

- We presented a new unified framework where relationships between stocks are represented via a graphical structure.
- We introduced two novel approaches for the estimation of the comoving portfolios.
- We back-tested a total of 780 approaches to show the differences between the various techniques.
- We analyzed the evolution of the excess returns over time and how liquidity and market capitalization affect performance.

SAPIENZA
UNIVERSITÀ DI ROMA

Framework

Our framework is spit in 3 phases:

1. Data preprocessing
2. Spreads evaluation
3. Trading

Framework – Data Preprocessing

- Remove the effect of common risk factors when evaluating the idiosyncratic component of a stock
- We assume the following linear factor model:

$$r = \sum_i \beta_i f_i + \epsilon$$

- We compute the spreads using the residuals ϵ rather than the returns r

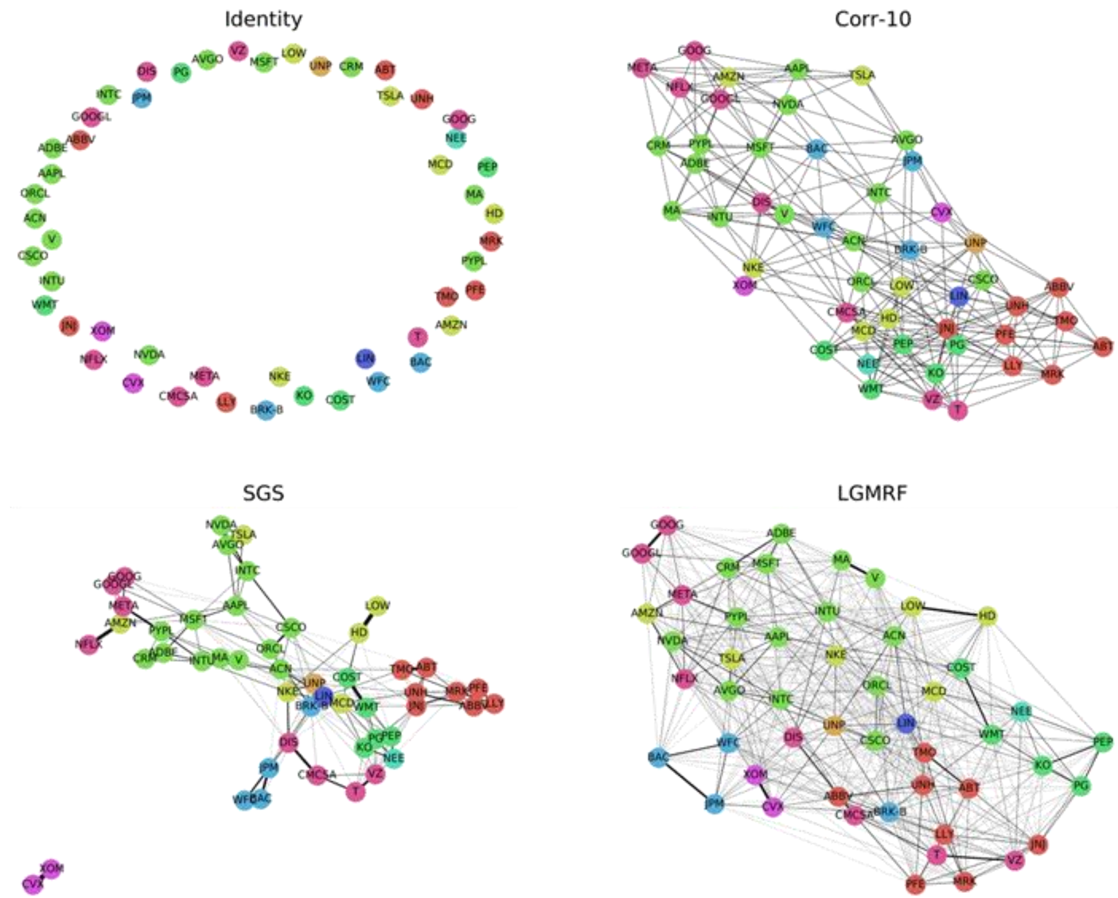
Framework – Spreads Evaluation

- The idiosyncratic component of a stock is represented as the spread between the returns of the stock and the ones of its comoving portfolio

$$s_t^T = r_t^T - r_t^T A = r_t^T L$$

1. Correlation-based approaches - Corr-K
2. Approaches based on the sparse inverse covariance:
   - Laplacian constrained Gaussian Markov Random Field (LGMRF)
   - Smooth Graph Signal (SGS) representation
3. Identity - $L = I$

Graph Visualization



Framework - Trading

- We define the rolling spreads with window w:

$$s_t^{(w)} = \sum_{\tau=t-w}^{t} s_\tau$$

Non-Parametric Threshold

- Given the rolling spreads standard deviation $s_{std}^{(w)}$
- We long asset i if $s_{t,i}^{(w)} \leq -\gamma \cdot s_{std,i}^{(w)}$
- We short asset i if $s_{t,i}^{(w)} \geq \gamma \cdot s_{std,i}^{(w)}$

Quantiles-based

- Given the amount of assets traded q
- We long the $q/2$ assets with the lowest spreads
- We short the $q/2$ assets with the highest spreads

SAPIENZA
UNIVERSITÀ DI ROMA

Framework - Trading

Quantiles-Std

- We long asset i if $s_{t,i} \leq -\gamma \cdot s_{std,i}^{(w)}$ and it belongs to the subset of the $q/2$ assets with the lowest spreads.
- We long asset i if $s_{t,i} \geq \gamma \cdot s_{std,i}^{(w)}$ and it belongs to the subset of the $q/2$ assets with the highest spreads.

- When $\gamma = None$ we neglect the non-parametric threshold trading rule;
- When $q = None$ we neglect the quantiles-based trading rule

Results

- We utilized two datasets:
  - S&P500, 427 stocks from 1995 to 2022;
  - CRSP, 3218 stocks from 1961 to 2022.
- We tested a total of 780 configurations on the S&P500
- Metrics of interest:
  - Monthly Excess Returns;
  - Sharpe Ratio;
  - Sortino Ratio.
- Empirical distributions and 95% confidence intervals are computed using non-parametric bootstrap.

Results – Preprocessing method

- **Returns** vs **Residuals**: The excess returns generated using residuals are less risky.

Method	Excess Returns	Sharpe Ratio	Sortino Ratio
residuals	0.82% (0.78%-0.86%)	**0.47** (0.45-0.49)	**0.76** (0.72-0.81)
returns	**0.86%** (0.8%-0.91%)	0.36 (0.34-0.39)	0.58 (0.54-0.62)

Results – Spreads method

- **Identity** vs **Corr-50** vs **LGMRF** vs **SGS**: Corr-K and LGMRF can considerably reduce the risk of the strategy.

Method	Excess Returns	Sharpe Ratio	Sortino Ratio
Corr-50	0.87%	**0.47**	**0.79**
	(0.76%-0.99%)	(0.42-0.53)	(0.68-0.91)
Identity	**0.99%**	0.35	0.65
	(0.83%-1.17%)	(0.31-0.4)	(0.53-0.77)
LGMRF	0.84%	0.45	0.73
	(0.73%-0.95%)	(0.39-0.5)	(0.62-0.84)
SGS	0.91%	0.38	0.6
	(0.77%-1.06%)	(0.33-0.44)	(0.5-0.71)

Results – Preprocessing method

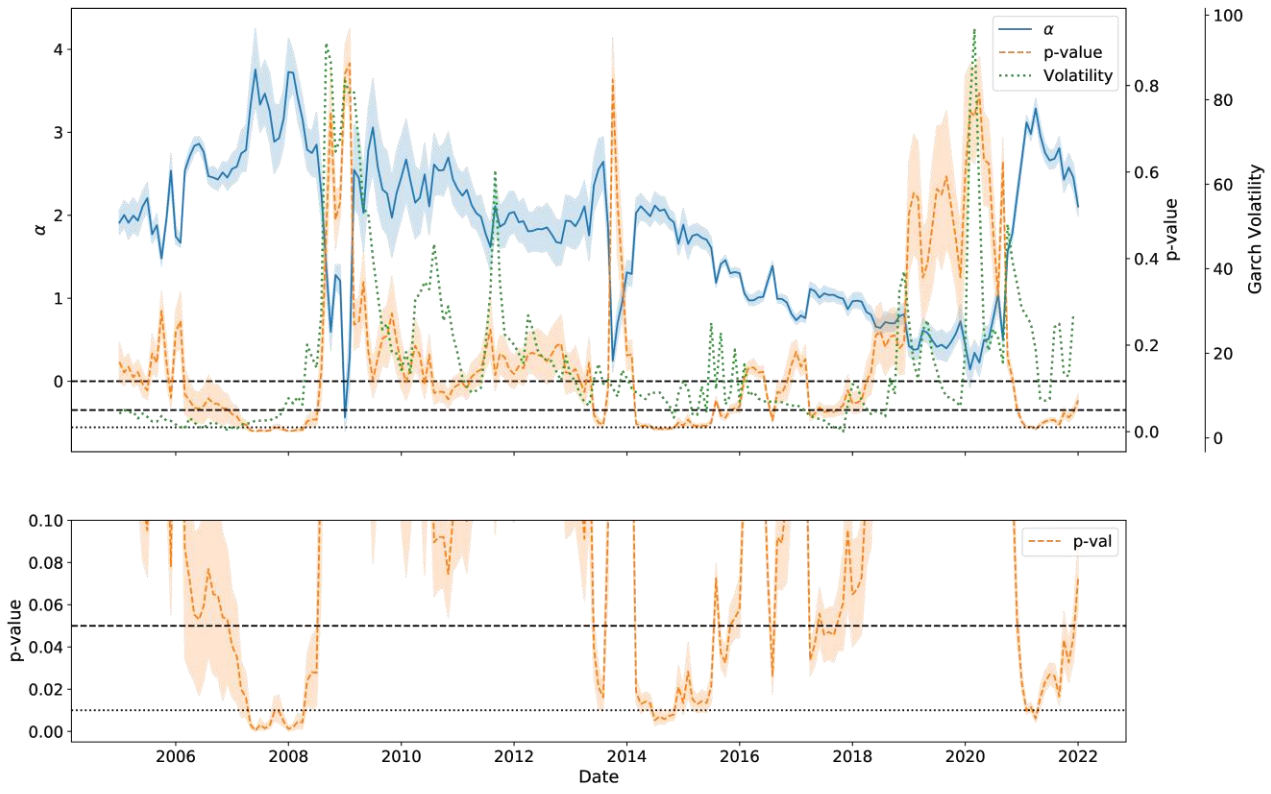
- A small rolling window and the union of the two approaches give the best performance.

Configuration	Excess Returns	Sharpe Ratio	Sortino Ratio
w=1 - γ=None - q=0.1	0.96% (0.85%-1.09%)	0.65 (0.58-0.73)	1.0 (0.85-1.17)
w=1 - γ=None - q=0.2	0.8% (0.71%-0.89%)	**0.72** (0.65-0.8)	1.17 (1.02-1.34)
w=2 - γ=2 - q=0.1	**1.43%** (1.25%-1.61%)	0.64 (0.56-0.73)	**1.18** (1.01-1.37)
w=2 - γ=2 - q=0.2	1.3% (1.13%-1.47%)	0.63 (0.54-0.73)	1.14 (0.97-1.33)
w=2 - γ=2 - q=None	1.25% (1.09%-1.42%)	0.63 (0.54-0.72)	1.12 (0.95-1.3)

Sapienza
Università di Roma

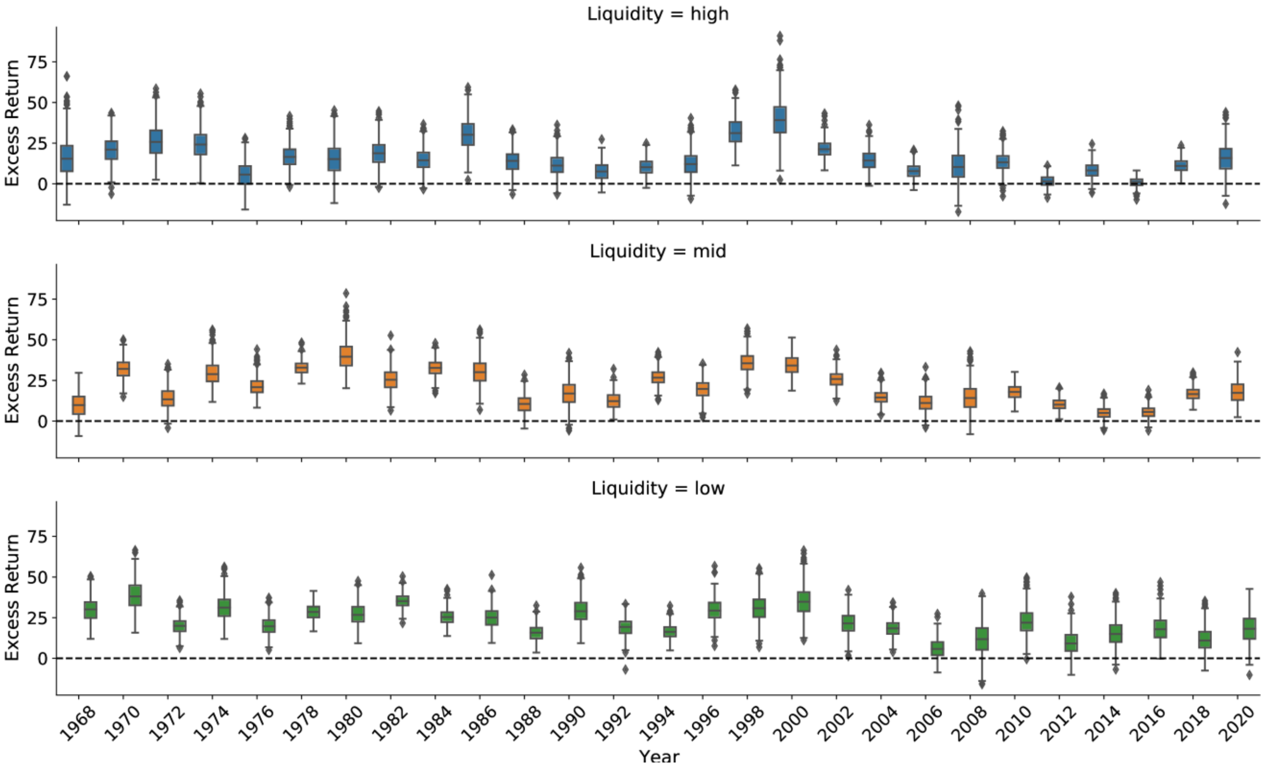
Performance over time – Rolling α

$$r = \alpha + \sum_i \beta_i f_i + \epsilon, \qquad \epsilon \sim \mathcal{N}(0, \sigma^2)$$



Market Cap. and Liquidity

- We split the U.S. CRSP dataset into 3 tertiles based on the average monthly volumes: **Low**, **Mid**, and **High** liquidity.



Conclusion

- We suggest using residuals and information from other stocks to reduce the risk of the trading strategies.
- We found that a low liquidity environment is beneficial to pairs trading.
- We formalized the pairs trading problem, and we presented a statistical arbitrage approach that casts some doubts on the robustness of the idea behind pairs trading.

Thanks for your attention!

SAPIENZA
UNIVERSITÀ DI ROMA