Instrument Similarity Analysis
Stage 1 – Proposal (CMP223)

Enrico Dal Pizzol
UFRGS / INF / PPGC

September 3, 2025

Computational Object

Object: Methods for measuring similarity and correlation between financial time series.

Goal: Evaluate computational performance of different algorithms (execution time, CPU, memory).

Methods: Pearson, Spearman, Kendall, DTW (and optionally MI).

Motivation

- Correlation influences risk, diversification, and trading decisions.
- Different methods $\Rightarrow$ different assumptions & costs
- Pearson: linear, very fast
- Spearman/Kendall: monotonic, robust to outliers
- DTW: handles misalignment, but much heavier

Computational Object — Characteristics

- Depending on the chosen period and frequency, time series can become very large.
- Each algorithm has its own computational cost and statistical properties.
- The task is the comparison of two financial time series.
- Each correlation/similarity method has specific strengths:
  - Good for capturing linear trends (Pearson)
  - Robust to outliers and non-linearity (Spearman/Kendall)
  - Suitable for comparing series that are distorted or shifted in time (DTW)

Analysis Method (Measurement)

- Apply correlation methods on real or replayed financial time series
- Environment & tools: Python (NumPy, Pandas, SciPy) and Docker
- Metrics collected: correlation output, CPU usage (avg/peak), memory, and execution time

Parameters

- Period: 1d, 7d, 1m, 6m, 1y
- Frequency: 1s, 1min, 1h, 1d
- Method: Pearson, Spearman, Kendall, DTW
- Instruments: X, Y, Z (pairs)
- Purpose: compare trade-offs: accuracy $\times$ computational cost

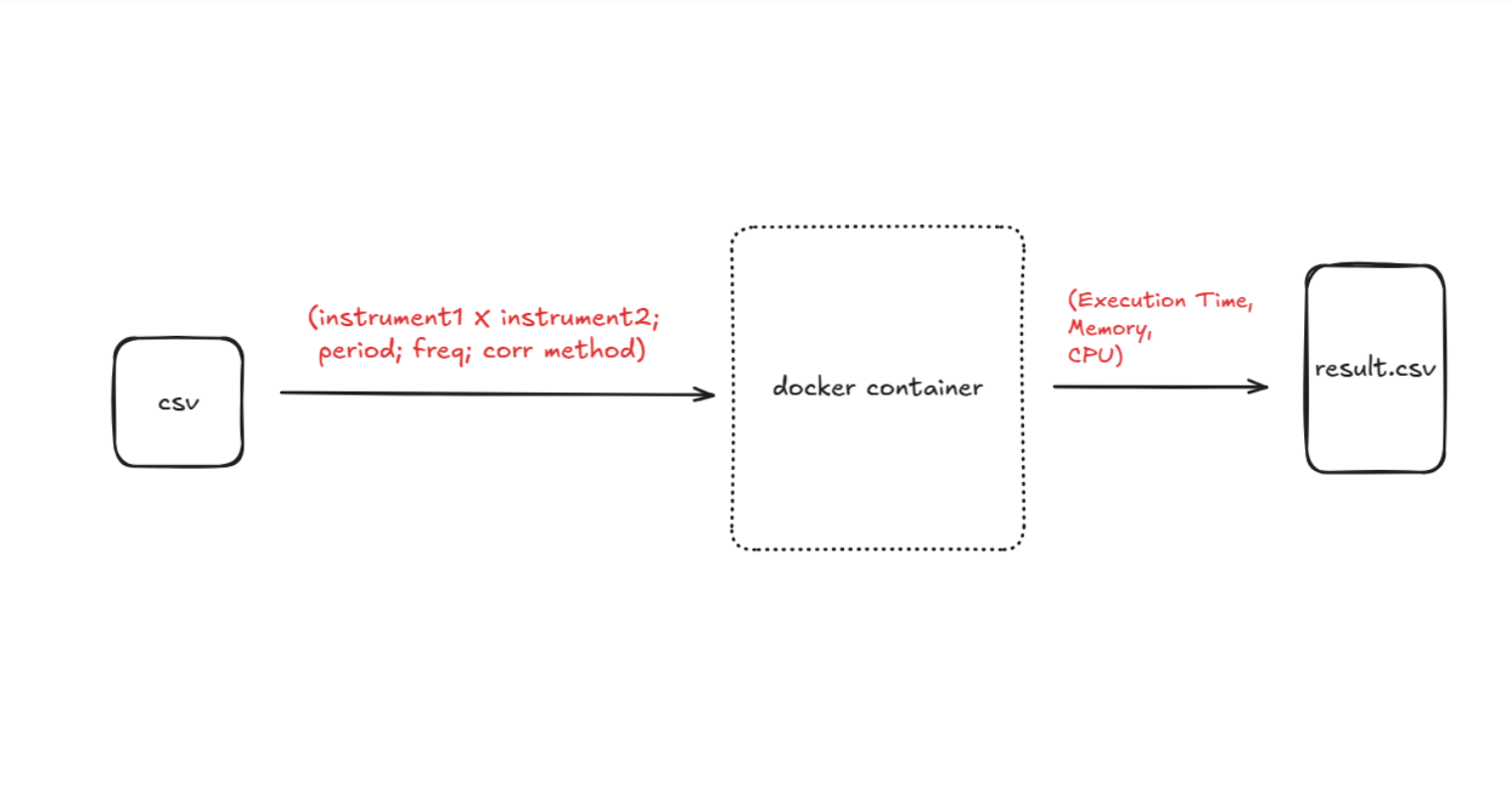
Metrics (What will be measured)

- Execution time
- CPU usage (%)
- Avg/Peak memory
- **Correlation variance**

Experiment Orchestration

- Input: CSV with all parameter combinations
- Runner: one container per row
- Inside container: calculate correlation
- Collector: monitor CPU & memory
- Aggregator: merge into `results.csv`

Experiment Pipeline



Example Result (Mock)

Instruments	Period	Freq	Method	Time	Mem	CPU
X×Y	1d	1min	Pearson	0.01s	5MB	2%
X×Y	1d	1min	Spearman	0.03s	6MB	3%
X×Y	7d	Daily	Kendall	0.02s	5MB	2%
X×Z	6m	1s	DTW	2h	700MB	45%

Next-Steps Schedule (8 Weeks)

- W1: Collect dataset and preprocess
- W2-3: Build test suite (Docker, scripts, ...)
- W4–6: Run experiments; record metrics
- W7: Analyze results; generate plots/tables
- W8: Write and submit final report

Q&A

Thank you!

 `instrument-similarity-analysis`

References

- Alexander Bader. *How can we quantify similarity between time series?*, Gorilla Tech Blog, May 27, 2021. Available at: `https://tech.gorilla.co/how-can-we-quantify-similarity-between-time-series-ed1d0b633`
- Ishan Jain. *Choosing the Right Correlation: Pearson vs Spearman vs Kendall's Tau*. Available at: `https://ishanjainoffical.medium.com/choosing-the-right-correlation-pearson-vs-spearman-vs-kendal`
- P. E. Puspita and Zulkarnain, *A Practical Evaluation of Dynamic Time Warping in Financial Time Series Clustering*, 2020 International Conference on Advanced Computer Science and Information Systems (ICACSIS), Depok, Indonesia, 2020, pp. 61–68. doi: 10.1109/ICACSIS51025.2020.9263123
- M. Winistörfer and I. Zhdankin, *Measure of Dependence for Financial Time-Series*, arXiv preprint 2023. Available at: `https://arxiv.org/abs/2311.12129`