

1 Introduction

Measuring the association of two assets is a fundamental concept in finance and economics. Portfolio managers have to think about it when they open or close a position. Regulators and large Bank Holding Companies (BHC) have to consider it when designing stress tests. The most common method of measuring association is the pearson correlation coefficient:

$$\rho(\mathbf{x}, \mathbf{y}) = \frac{\sum_{t=1}^T (x_t - \bar{x})(y_t - \bar{y})}{\sqrt{\sum_{t=1}^T (x_t - \bar{x})^2 \sum_{t=1}^T (y_t - \bar{y})^2}} \quad (1)$$

There's certainly nothing wrong about this statistic as the measure of association. I use it all the time. And even though the economic literature is clear-eyed about it's shortcomings (asymmetry, fat-tails) when trying to capture true return correlations between different financial time series, it's still the first measure researches estimate. When looking at the covariance expression on the numerator of equation (1), the correlation's assesment of association is based on the co-movements of the two variables for each time unit. This tight pairing is completely logical but it can also be restrictive. For instance, if I were to sample T draws from the following multivariate normal distribution:

$$\mathbf{Z}_t \sim N(\mathbf{0}, \Sigma) \quad (2)$$

with

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix} \quad (3)$$

and I estimated the sample variance-covariance matrix, we should not be suprised to find that the sample statistics are close to the population values: $\sigma_1 \approx \hat{\sigma}_1$, $\sigma_2 \approx \hat{\sigma}_2$, and $\rho \approx \hat{\rho}$. Now, if we create a second multivariate sample $\mathbf{Z}' = [\mathbf{Z}_{(-T),1}, \mathbf{Z}_{(-1),2}]$ where the first column is the all but the last value of \mathbf{Z}_1 while the second column is all but the first value of \mathbf{Z}_2 . In this case we still have $\sigma'_1 \approx \hat{\sigma}'_1$ and $\sigma'_2 \approx \hat{\sigma}'_2$ but $\hat{\rho}'$ will be a completely untrustworthy estimate of the original correlation because of the misalignment of the time index. This does not mean that the two variables are unreleated, they still are, but the relationship has been masked.

For this paper, I want to investigate other approaches for measuring the association of different time series and compare the results of those approaches to the standard correlation approach used in finance. My main point of reference has been an article by Esling and Agon 2012 that explores topics in time series data mining. The authors provide a detailed survey of time series representations, various distance metrics for measures of association, and proper ways to index time series data for efficient querying and information retrieval. The discussions around various distance metrics are very interesting. Esling and Agon 2012 categorize these distance measures into four categories collected in figure 1. A very similar paper was published in Computational Economics very recently by Franses and Wiemann 2020.

Instead of focusing on exact synchronous co-movements, many of the distance measures discussed in the data mining literature focus on comparing the shape or global structure of two different series. Below are a handful of distance measures that are mentioned in figure 1 that I believe can be easily implemented.

- Pearson Correlation
- L_p norms
- Dynamic Time Warping
- Time Warp Edit distance
- Autocorrelation
- Kullback-Leibler

2 Data

Studying the companies that make up the Standard and Poors 500 will provide a rich sample to look for comparisons. If we focus on the most recent two years, we can use the market upheaval surrounding Covid-19 as a natural experiment. Some industries have bounced back strongly (Technology) while others have struggled to recover (Travel, Hospitality). These alternate reactions will ensure that there is enough heterogeneity in the "shape" of the stock values of these companies to make meaningful comparisons. Natural industry classifications will also provide a good reference point to any latent clustering that is discovered.

To benchmark the shape based measures, we will design a process to estimate a valid ARMA-GARCH model for each member of the S&P 500. For model validation, I will

follow the same process as in section 4.3 of Dowiak 2018, checking that each model is well specified and the model residuals are tested to see if they are independently and identically distributed.

I’ve used the S&P’s Wikipedia page¹ to collect the stock ticker symbol for each company and the extra industry and sector information. Using an API made freely available by Alpha Advantage², daily historical series of the current companies in the index have been obtained. Dates range from Nov 1999 to Feb 2021. See figure 2 for a visual summary. Analyzing the entire S&P adds a level computational complexity to this project, but I think the extra scale helps give the article some heft considering the task at hand isn’t very complicated.

References

- Dowiak, L. C. (2018). “A Comparison of Time-Varying Copula-GARCH Models”. In: *Unpublished* 00, p. 00.
- Esling, P. and C. Agon (2012). “Time-Series Data Mining”. In: *ACM Comput. Surv.* 45, pp. 79–82.
- Franses, P. H. and T. Wiemann (2020). “Intertemporal Similarity of Economic Time Series: An Application of Dynamic Time Warping”. In: *Computational Economics* 56, pp. 59–75.

¹https://en.wikipedia.org/wiki/List_of_S&P_500_companies

²<https://www.alphavantage.co/>

Table I. Comparison of the Distance Measures surveyed in This Article with the Four Properties of Robustness

| Distance measure | Scale | Warp | Noise | Outliers | Metric | Cost | Param |
|-----------------------------------|-------|------|-------|----------|--------|-----------------|-------|
| Shape-based | | | | | | | |
| L_p norms | | | | | ✓ | $O(n)$ | 0 |
| Dynamic Time Warping (DTW) | | ✓ | | | | $O(n^2)$ | 1 |
| LB-Keogh (DTW) | | ✓ | ✓ | | ✓ | $O(n)$ | 1 |
| Spatial Assembling (SpADe) | ✓ | ✓ | ✓ | | | $O(n^2)$ | 4 |
| Optimal Bijection (OSB) | | ✓ | ✓ | ✓ | | $O(n^2)$ | 2 |
| DISSIM | | ✓ | ✓ | | ✓ | $O(n^2)$ | 0 |
| Edit-based | | | | | | | |
| Levenshtein | | | | ✓ | ✓ | $O(n^2)$ | 0 |
| Weighted Levenshtein | | | | ✓ | ✓ | $O(n^2)$ | 3 |
| Edit with Real Penalty (ERP) | ✓ | | | ✓ | ✓ | $O(n^2)$ | 2 |
| Time Warp Edit Distance (TWED) | ✓ | | | ✓ | ✓ | $O(n^2)$ | 2 |
| Longest Common SubSeq (LCSS) | ✓ | | ✓ | ✓ | | $O(n)$ | 2 |
| Sequence Weighted Align (Swale) | ✓ | | ✓ | ✓ | | $O(n)$ | 3 |
| Edit Distance on Real (EDR) | ✓ | | ✓ | ✓ | ✓ | $O(n^2)$ | 2 |
| Extended Edit Distance (EED) | ✓ | | ✓ | ✓ | ✓ | $O(n^2)$ | 1 |
| Constraint Continuous Edit (CCED) | ✓ | | ✓ | ✓ | | $O(n)$ | 1 |
| Feature-based | | | | | | | |
| Likelihood | | | ✓ | ✓ | ✓ | $O(n)$ | 0 |
| Autocorrelation | | | ✓ | ✓ | ✓ | $O(n \log n)$ | 0 |
| Vector quantization | ✓ | | ✓ | ✓ | ✓ | $O(n^2)$ | 2 |
| Threshold Queries (TQuest) | ✓ | | ✓ | ✓ | | $O(n^2 \log n)$ | 1 |
| Random Vectors | ✓ | | ✓ | ✓ | | $O(n)$ | 1 |
| Histogram | | | ✓ | ✓ | ✓ | $O(n)$ | 0 |
| WARP | ✓ | ✓ | ✓ | | ✓ | $O(n^2)$ | 0 |
| Structure-based | | | | | | | |
| <i>Model-based</i> | | | | | | | |
| Markov Chain (MC) | | | ✓ | ✓ | | $O(n)$ | 0 |
| Hidden Markov Models (HMM) | ✓ | ✓ | ✓ | ✓ | | $O(n^2)$ | 1 |
| Auto-Regressive (ARMA) | | | ✓ | ✓ | | $O(n^2)$ | 2 |
| Kullback-Leibler | | | ✓ | ✓ | ✓ | $O(n)$ | 0 |
| <i>Compression-based</i> | | | | | | | |
| Compression Dissimilarity (CDM) | ✓ | ✓ | ✓ | | | $O(n)$ | 0 |
| <i>Parsing-based</i> | | | | | | | |
| | ✓ | ✓ | ✓ | | | $O(n)$ | 0 |

Each distance measure is thus distinguished as *scale* (amplitude), *warp* (time), *noise* or *outliers* robust. The next column shows whether the proposed distance is a metric. The cost is given as a simplified factor of computational complexity. The last column gives the minimum number of parameters setting required by the distance measure.

Figure 1: List of distance measures included in Esling and Agon 2012

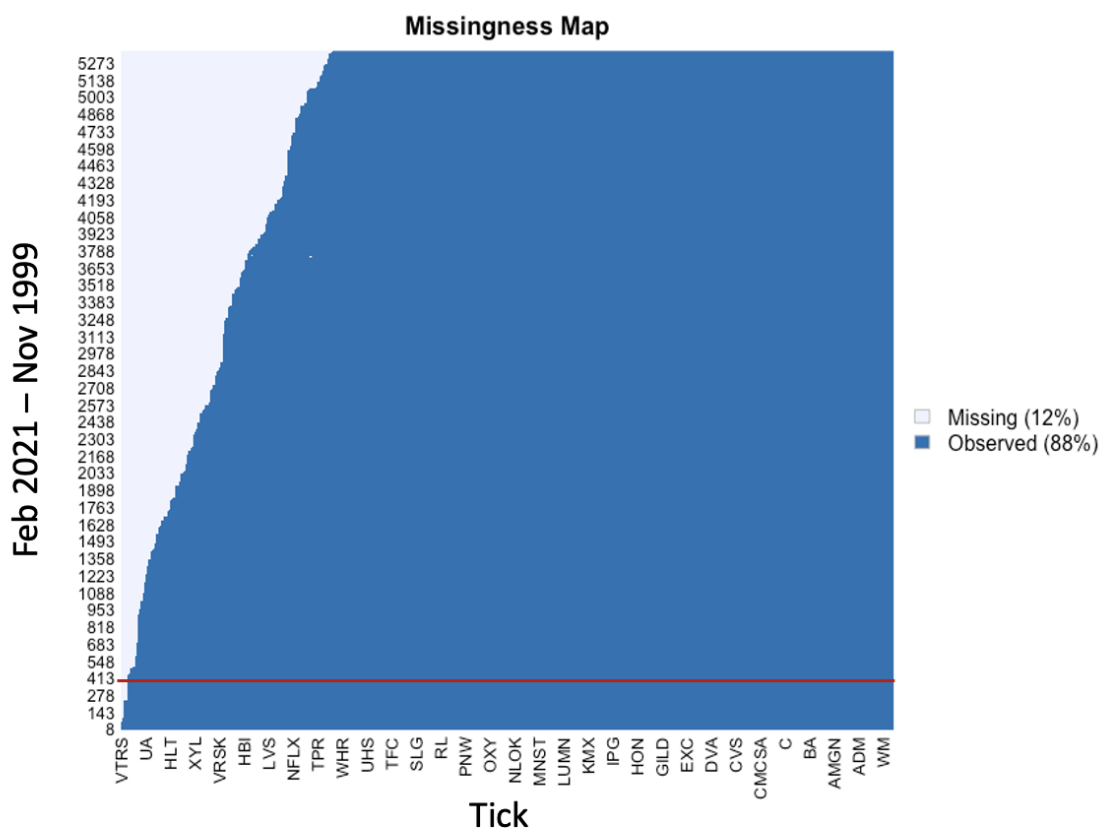


Figure 2: Missing map of the S&P data set. The red horizontal bar demarks the minimum amount of history needed to be included in the study. X of the 505 tickers have enough history