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 [15, 19]. 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:

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

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 assessment 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:

$$\boldsymbol{Z_t} \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}) \tag{2}$$

with

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix} \tag{3}$$

and I estimated the sample variance-covariance matrix, we should not be surprised to find that the sample statistics are close to the population values: $\sigma_1 \approx \hat{s}_1$, $\sigma_2 \approx \hat{s}_2$, and $\rho \approx \hat{\rho}$. Now, if we create a second multivariate sample $\boldsymbol{Z}' = [\boldsymbol{Z}_{(-T),1}, \boldsymbol{Z}_{(-1),2}]$ where the first column is the all but the last value of $\boldsymbol{Z}_1$ while the second column is all but the first value of $\boldsymbol{Z}_2$. In this case we still have $\sigma_1' \approx \hat{s}_1'$ and $\sigma_2' \approx \hat{s}_2'$ 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 unrelated, 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 [3] 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. [3] categorize the these distance measures into four categories collected in figure 1. A very similar paper was published in Computational Economics very recently by [5].

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
- Parameter value clustering (or by other metrics e.g. persistance)
- Autocorrelation
- Kullback-Leibler

2 Relevant Literature

[3] view the analysis of time series data from a data mining perspective. The authors provide a detailed survey of time series representations, various distance metrics for measuring association, and proper ways to index time series data for efficient querying and information retrieval. The distance measure are classified into four categories: shape, edit, feature, and structure-based.

[6] cluster time series into K groups based on the fitted parameters of AR(p) and dynamic regression models.

[16] provides an excellent summary of the principles of DTW and discuss several extensions with respect to the local and global parameters of the technique.

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			√	√	√	$O(nlogn)$	0
Vector quantization		√	√	√	√	$O(n^2)$	2
Threshold Queries (TQuest)		√	√	√		$O(n^2logn)$	1
Random Vectors		√	√	√		$O(n)$	1
Histogram			√	√	√	$O(n)$	0
WARP	√	√	√		√	$O(n^2)$	0
Structure-based							
Model-based							
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
Compression-based							
Compression Dissimilarity (CDM)		√	√	√		$O(n)$	0
Parsing-based		√	√	√		$O(n)$	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 [3]

[5] apply dynamic time warping to quarterly real GDP of the 50 US states during the 2007 recession. The authors employ a novel distance metric and utlilize a K-Means

algorithm proposed by [17] and adapted for time series to group the US states into seven distinct business cycle clusters.

[21] apply DTW to measure the pair-wise similarity between 35 foreign currencies. The authors then utilize a minimum spanning trees (MST) – a novel graph algorithm – to document the structural changes in the dependencies of the FX market over three separate years.

[13] are also motivitated from a data mining perspective and the ability to return relevant results from query to a large number of time series. The authors compare the an example-base query using DTW with a model-based approach using Hidden Markov Models (HMM).

3 Data

Studying the companies that make up the Standard and Poor's 500 will provide a rich sample to look for comparisons. Three different time periods will be considered. The financial crisis of 2007-2008, the moderate economic period between 2015 and 2016, and the dramatic shut-down economy of 2019-2020. With a focus on the these three time periods, the market upheaval surrounding the financial crisis and Covid-19 serve as a natural experiment with the moderate-but-steady growth of 2015-2016 acting as a control. This is especially true of the Covid-19 timeframe. Some industries bounced back strongly (Technology) while others 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.

The companies comprising the S&P's Wikipedia page[1] as of February 2021 are included in this essay's analysis. Using an API made freely available by Alpha Advantage[2], 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 of the price history for the individual companies in the index.

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

[2] https://www.alphavantage.co/

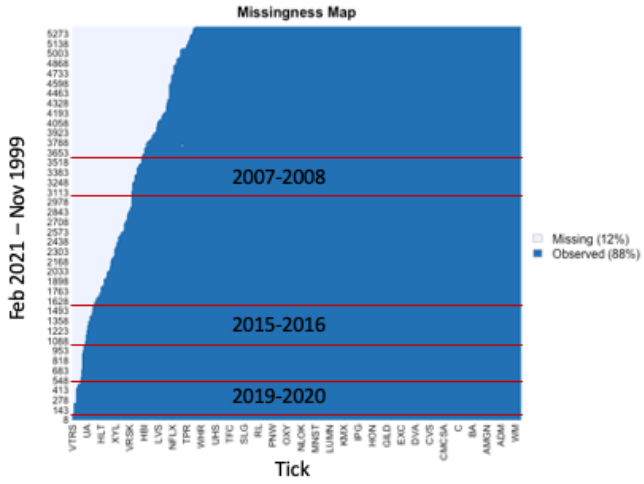

Figure 2: Missing map of the S&P data set. The red horizontal bars demark the timeframes under consideration. Of the 505 tickers in the dataset, there are 423 with a full series stock values from 2007 onward, 480 from 2015 onward, and 494 from 2019 onward.

4 Distance Measures

4.1 ARMA-GARCH Benchmark

To benchmark the competing measures, a procedure is designed to estimate a valid ARMA-GARCH model for the log returns of each member of the S&P 500. After controlling for the conditional mean (ARMA) and the conditional variance (GARCH) of the stock's DGP, pair-wise correlation values are calculated on the standardized residuals for each stock. For model validation, I will follow the same process as in section 4.3 of [2]. Each model is checked to see if it is well specified and that the model residuals are abiding by the independent and identical distribution assumptions.

Define the return as $X_t = \log(S_t) - \log(S_{t-1})$, where S_t is the spot price at time t. The conditional mean for the log-returns can be formulated by the following ARMA process:

$$x_t = \mu_t + \epsilon_t \tag{4}$$

$$\mu_t = \mu(\phi, \theta, x_{\{s:\, s<t\}}, \varepsilon_{\{s:\, s<t\}}) = \phi_0 + \sum_{i=1}^{p} \phi_i x_{t-i} + \sum_{j=1}^{q} \theta_j \, \varepsilon_{t-j} + \varepsilon_t \tag{5}$$

where ε_s is an innovation term that satisfies $E[\varepsilon_s] = 0$ and $E[\varepsilon_s^2] = \sigma_s^2$. The conditional volatility process for the models under consideration can be generalized with the following formula. Functions A, B, C are generic stand-ins that will differ across the various GARCH specifications.

$$\sigma_t^2 = C(\varepsilon_{t-1}^2,\, \sigma_{t-1}^2) + \sum_{j=1}^{m} \alpha_j A(\varepsilon_{t-j}^2) + \sum_{i=1}^{k} \beta_i B(\sigma_{t-i}^2) \tag{6}$$

The conditional mean and variance are used to center and scale the innovation terms.

$$z_t(\phi, \theta, \alpha, \beta) = \frac{x_t - \mu(x_{t-1}, \phi, \theta)}{\sigma(x_{t-1}, \alpha, \beta)} \tag{7}$$

By collecting all the parameters in the conditional mean and variance equations into one vector $\Delta = [\phi, \theta, \alpha, \beta]$, the functional form for the error terms f can be written as a product of Δ, the chosen error distribution g, and any necessary shape parameters λ of the distribution:

$$f(x_t|\mu_t, \sigma_t^2, \lambda, \Delta) = \frac{1}{\sigma_t} g(z_t|\lambda, \Delta) \tag{8}$$

The challenge for automating this process is two-fold. First, the size of the full model space is quite large. To list out all the dimensions that must be considered:

- ARIMA model
  - Constant term
  - Autoregressive order (p)
  - Moving average order (q)
  - Order of integration (d)
  - Seasonal autoregressive order (P)
  - Seasonal moving average order (Q)
  - Order of seasonal integration (D)
- GARCH model
  - Constant term
  - Autoregressive order (m)
  - Moving average order (k)
  - GARCH specification
  - Error distribution

A full grid-search over these dimensions is time-consuming and impractical. To reduce the number of specifications in the model set, the fitting procedure takes the following divide-and-conquer approach.

1. **Estimate the conditional mean independently of the variance model.** Leveraging the work done by Hyndman and Khandakar [10], a step-wise strategy is used to search through the ARIMA model space for the best fit, which

is evaluated via the Akaike information criterion (AIC). This is accomplished by using the Forecast package [11] running in the R statistical language [20]. The following decisions are made:

- Always include a constant term
- Set the integration terms to zero: $d = D = 0$

2. **Estimate a set of GARCH models.** Set the conditional mean model to the specification found in step 1. Then iterate over every GARCH specification in the model set, re-estimating the combined ARMA and GARCH parameters at the same time for each conditional variance model. This is accomplished by using the rugarch package [7]. The model set that is searched through considers the following dimensions:

   - Always include a constant term
   - ARCH specification: m = {1, 2}
   - GARCH specification: k = {1, 2}
   - Distributions: g = {Normal, Student-t, Skewed Student-t [4]}
   - Model Specification: A, B, C = {Standard GARCH [1], gjr-GARCH [8], Component GARCH [14] }

   With these dimensions, a total of 36 volatility models are available to choose from.

3. **Select the best model specification.** The fitted residuals of a model are checked against a battery of tests to confirm the independent and identical assumptions as well as to verify the correct distribution has been selected. The tests used include: (a) Moment LM tests to check for any remaining autocorrelation in the first four moments, (b) Kolmogorov-Smirnov test to check the residuals against the chosen theoretical distribution, (c) Hong and Li [9] non-parametric density test jointly for i.i.d and correct distribution specification, (d) Shapiro-Wilks [18] test for normality, and (e) Jarque-Bera [12] test for joint normality for skew and kurtosis.

   With these tests in hand, finding the best model reduces to selecting the GARCH specification that:

   - Passes all five distributional tests
   - Minimizes the Bayesian information criterion (BIC)

If no specifications pass all five tests, then the one that minimizes the BIC across the 36 candidate models is selected.

The specification for the ARMA process is found first independently of the GARCH process. Once the AR(p) and MA(q) orders have been found, this specification is set and remains the same as different volatility models are estimated. Note that the parameter estimates are not held constant, just the specification. For each new GARCH fit, the ARMA parameters are all re-estimated.

4.2 Dynamic Time Warping

Dynamic time warping (DTW) is an alternative method for comparing the association between two discrete time series. It differs from Pearson's correlation measure in that the time indices between the two series at moments of comparison are not constrained to equal each other – like in Equation (1). Time is allowed to stretch and compress before a local cost[3] c function is applied to the pairs of values from the two time series. This article will adhere to the classic definition of dynamic time warping. For notation this article this article borrows heavily from [16].

4.2.1 Algorithm

Suppose there are two time series: x_t for $t \in [1:T]$ and y_s for $s \in [1:S]$. A *warping path* is a sequence $\boldsymbol{p} = [p_1, ..., p_L]$ where each element is a mapping from the time index of one series to the other: $p_l = (t_l, s_l) \in [1:T] \times [1:S]$ for $l \in [1:L]$. For each point in the warping path p_l there is a cost function quantifying the distance between the values of the series.

$$c : x_{t_l} \times y_{s_l} \to \mathbb{R}_{\geq 0} \tag{9}$$

The behaviour of the warping paths do follow some conditions, which are listed below:

- *Boundary Condition*: $p_1 = (1, 1)$ and $p_L = (T, S)$
- *Step-Size Condition*: $p_l - p_{l-1} \in \{(1, 0), (0, 1), (1, 1)\}$

[3]This is the languaged used in [16] but other authors use "association", "dissimilarity", or "distance" functions.

The boundary condition requires that the first and last indices of the two series are mapped to each other. The step-size condition governs the evolution of the warping path. It ensures a non-decreasing monotonicity in the indices of *both* series such that $t_i \leq t_j$ and $s_i \leq s_j$ if $i \leq j$. A $T \times S$ cost matrix can be created that stores the associated cost between all values in the two series:

$$\mathbf{C}(\boldsymbol{X}, \boldsymbol{Y}) = \begin{bmatrix} c(x_T, y_1) & c(x_T, y_2) & \cdots & c(x_T, y_S) \\ \vdots & \vdots & \vdots & \vdots \\ c(x_2, y_1) & c(x_2, y_2) & \cdots & c(x_2, y_S) \\ c(x_1, y_1) & c(x_1, y_2) & \cdots & c(x_1, y_S) \end{bmatrix} \tag{10}$$

A warping path's total cost is the sum of the local costs it incurs as it travels from the start of the series (bottom left of $\boldsymbol{C}$) to their end (top right of $\boldsymbol{C}$):

$$\mathbb{c}_{\boldsymbol{p}}(\boldsymbol{X}, \boldsymbol{Y}) = \sum_{l=1}^{L} c(x_{t_l},\, y_{s_l}) \tag{11}$$

There are many permissable[4] warping paths between two series. In fact the number of permissable paths has exponential growth in the values of T and S. The aim during optimization is to find the warping path that minimizes the total cost. If the set of all warping paths are denoted $\mathbb{P}(\boldsymbol{X}, \boldsymbol{Y})$ then the value of the optimal warping path has the property

$$\mathbb{c}_{\boldsymbol{p}^*}(\boldsymbol{X}, \boldsymbol{Y}) \leq \mathbb{c}_{\boldsymbol{p}}(\boldsymbol{X}, \boldsymbol{Y}) \text{ for all } \boldsymbol{p} \in \mathbb{P}(\boldsymbol{X}, \boldsymbol{Y}) \tag{12}$$

and the value of the DTW measure between $\boldsymbol{X}$ and $\boldsymbol{Y}$ is set to $\mathbb{c}_{\boldsymbol{p}^*}(\boldsymbol{X}, \boldsymbol{Y})$. An interested party could solve this optimization problem by estimating the total cost of all warping paths and select the one that minimizes this value. The challenge though is efficient computation. Since the number of warping paths grows exponentially the compuational time needed to check every warping path becomes problematic for large series. The proposed solution is to leverage *dynamic programming* to reduce the computational time needed to find the optimal solution. Instead of dealing with exponential growth of permissable warping paths the DTW algorithm can find the optimal warping path in $\mathcal{O}(TS)$ calculations. To do so an *accumulated cost matrix* $\boldsymbol{A}$ needs to be defined. The accumulated cost matrix has the same dimension as the

[4]Permissable as governed by the boundary and step conditions

cost matrix. Defining $A_{t,s}$ as the value of $\boldsymbol{A}$ at the t^{th} row and the s^{th} column of the accumulated cost matrix, the matrix has the following three identities:

$$A_{t,1} = \sum_{k=1}^{t} c(x_k, y_1) \text{ for } t \in [1:T] \tag{13}$$

$$A_{1,s} = \sum_{k=1}^{s} c(x_1, y_k) \text{ for } s \in [1:S] \tag{14}$$

$$A_{t,s} = c(x_t, y_s) + \min\left(A_{t-1,s-1},\, A_{t-1,s},\, A_{t,s-1}\right) \tag{15}$$

With this definition of the accumulative cost matrix the optimal warping path can be found by the following recursive procedure:

1. Set $p_L = (T, S)$
2. Given $p_l = (t, s)$ select $p_{l-1} = \begin{cases} (1, s-1) & \text{if } t = 1 \\ (t-1, 1) & \text{if } s = 1 \\ \arg\min\left[A_{t-1,s-1},\, A_{t,s-1},\, A_{t-1,s}\right] & \text{otherwise} \end{cases}$

4.2.2 Example

To put these concepts into practice the dynamic time warping algorithm will be demonstrated on the stock prices for Agilent Technologies (A) and General Electric (GE) from the first half of 2019. Instead of the original price level the series will be standardized by setting the price on January 2, 2019 equal to one. Figure 3 summarizes the price movement of these two stocks and shows which price values are matched together by their optimal warping path. Figure 4 displays the full optimal warping paths and discusses the influence of boundary and step-size conditions on it's shape. Figure 5 displays the cost and accumulated cost matrices for the first ten days of the two series. The optimal warping path is annotated in both.

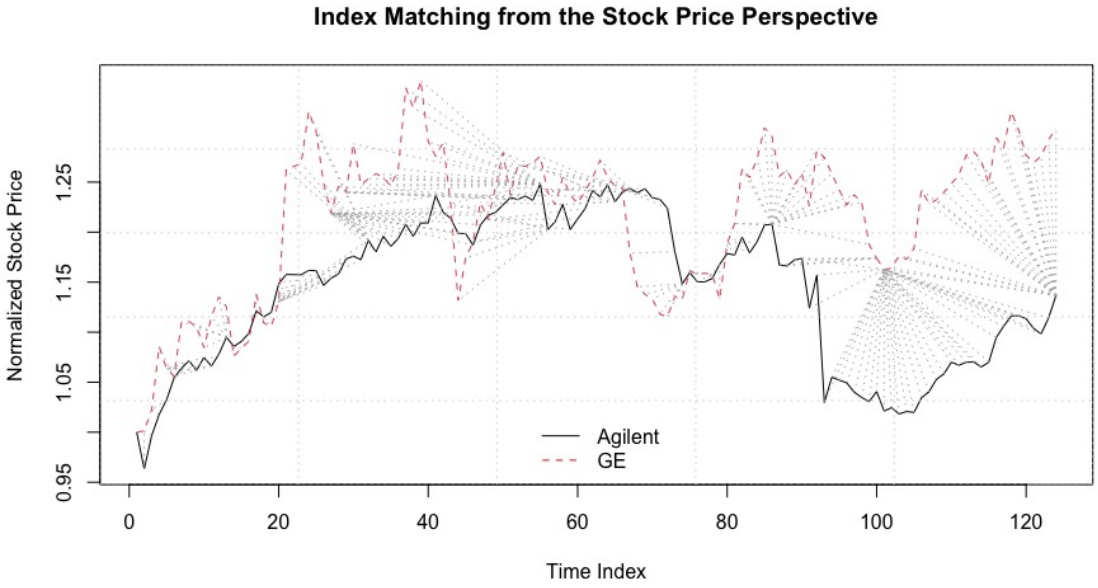

Figure 3: This figures shows a normalized view of the Aligent and GE stock price for the first half of 2019. The normalization divides the stock price by it's value on the first market day of the year, January 2nd, 2019. Dynamic time warping is applied on the standardized series. The total length is 124 days ending on June 28th. The dashed gray lines connect elements of the index pair in the optimized warping path. The DTW finds the warping path that has the least total cost.

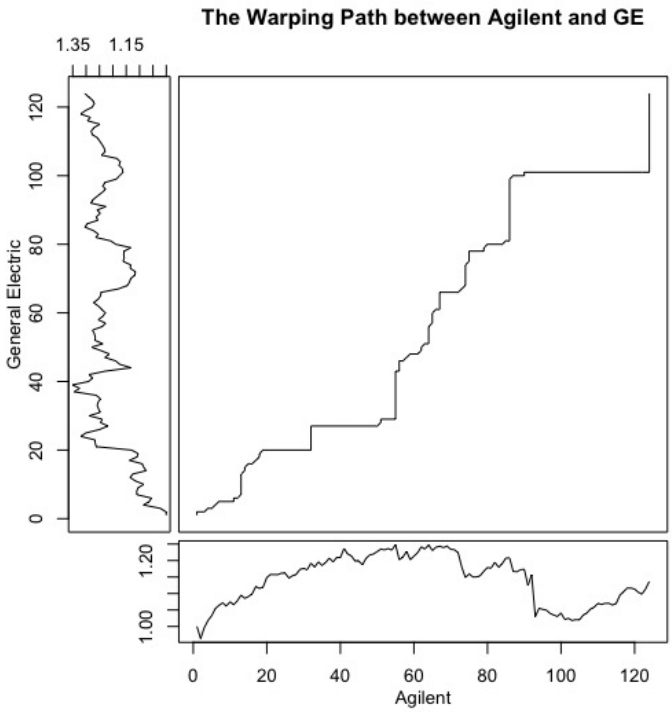

Figure 4: This figure displays the full warping path between Aligent and GE stock prices in the first half of 2019. The two peripheral images show their respective standardized stock prices while the central image traces the warping path as it connects the indicies of the two series. The influence of the step-condition is seen immediately in the marginal increments of the path. They are non-decreasing. This allows time to "stretch" but not be put into reverse. At the end of the warping path (top right part of the image) the boundary condition becomes binding–seen by the vertical line at the tail end–as the last index in GE's series is reached before the last index of Aligent's.

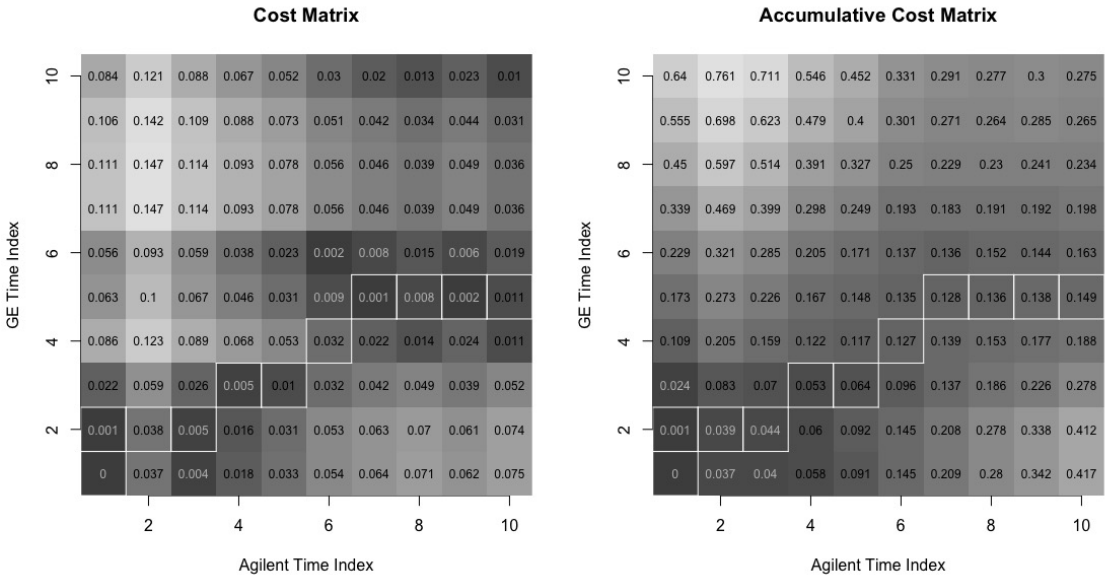

Figure 5: The left image is the cost matrix for the first ten values of the Aligent and GE stock prices after standardization. The local cost function is the euclidean distance. The right image is the accumulative cost matrix for the same sequence of prices. The optimal warping path is annotated with white outlines. The gray shading tracks the range of values in each matrix. Darker shades of gray represent smaller values while lighter shades of gray represents larger values. Note that the scale of shading is not uniform across images.

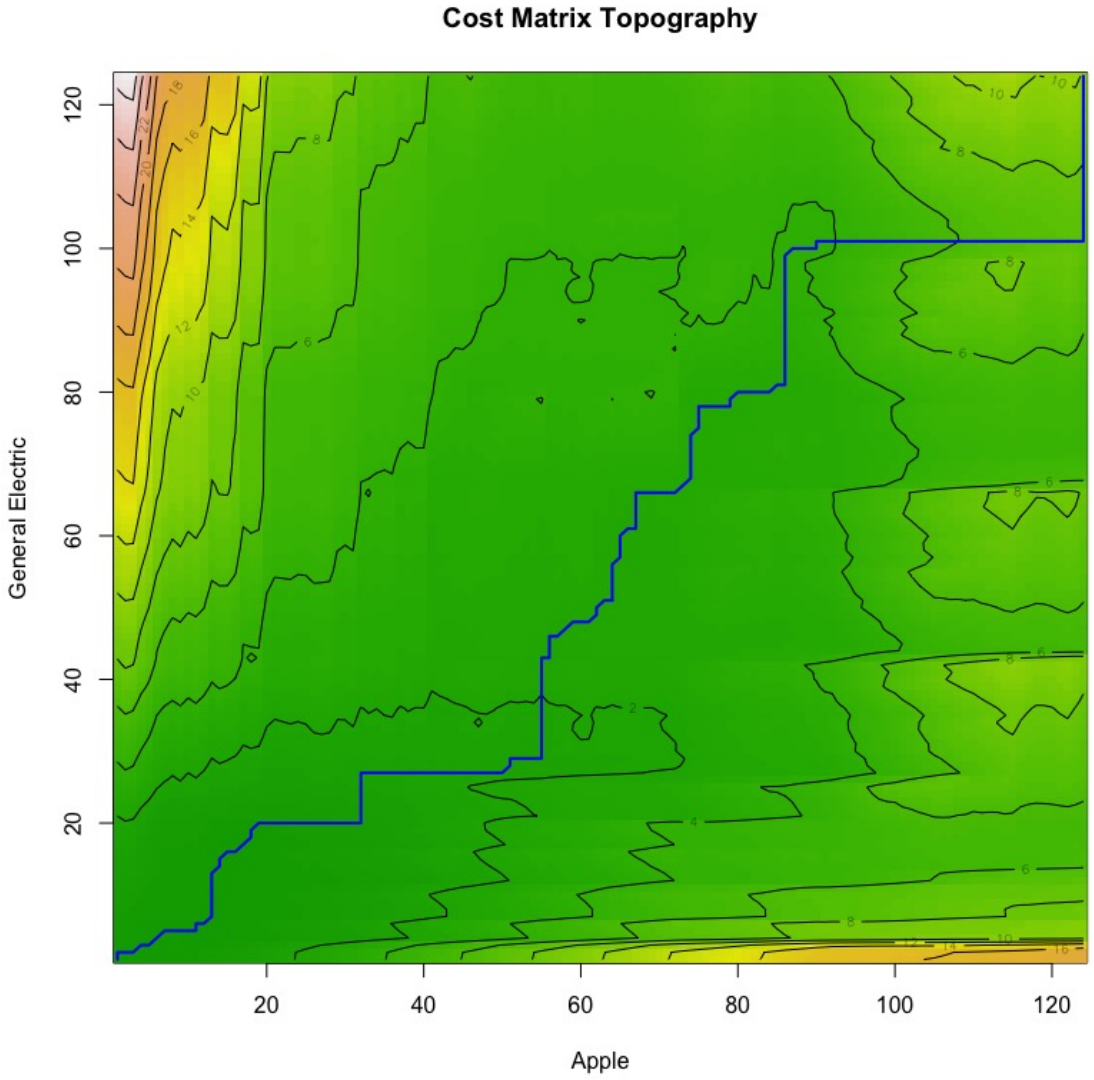

Figure 6: This figure displays the optimal alignment between Aligent and GE stock return series. The topology lines are based on the cost matrix. From this view the warping path (in blue) is a "valley of least cost" traversing the local cost landscape.

4.3 Kullback-Leibler Divergence

The model estimation strategy this article takes is ultimately a maximum likelihood one. As discussed in Section 4.1, every ARIMA-GARCH model is estimated by searching for the parameter vector Δ that maximizes the log likelihood for a given distribution g with parameter vector λ

$$\hat{\Delta},\ \hat{\lambda} = \arg\max_{\Delta,\,\lambda} L(\Delta, \lambda|X) \tag{16}$$

with

$$L(\Delta, \lambda|Z) = \prod_{t=1}^{T} g(z_t|\lambda) \tag{17}$$

The Kullback-Leibler divergence is a measure of relative entropy between two distributions and provides another avenue to measure the association between two time series. For two stock tickers (XOM vs AMD), the distributions estimated from the data are used to compare the relative entropy between the log returns of those stocks (e.g: $g_{xom}(\cdot|\hat{\lambda})$). For continuous distributions Q,P the relative entropy between the reference distribution Q and a second distribution P is written:

$$D_{KL}(P||Q) = \int_{-\infty}^{\infty} p(x) \log\left(\frac{p(x)}{q(x)}\right) dx \tag{18}$$

where p and q are the density functions of P and Q, respectively (e.g: $p = \frac{d}{dx}P(x)$). Since KL divergence is not symmetric, this article will use a symmetrized version of KL divergence

$$SKL(P, Q) = \frac{1}{2}\left(D_{KL}(P||Q) + D_{KL}(Q||P)\right) \tag{19}$$

Unlike DTW, there is a strong connection in this article between the correlation measure and the symmetric KL divergence. The distributions used here are derived during the estimation of the ARIMA-GARCH model and due to that, the SKL measure is subject to the same dangers of model misspecification as the correlation analysis.

5 Simulation Results

In order to gauge the relationship between the DGP process of two time series and their resulting DTW value we turn to simulation. For the first simulation a stationary process is developed with only two factors: an intercept and an error term.

$$X_t = \omega + \epsilon_t \qquad \epsilon_t \sim \mathcal{N}(0, \sigma^2) \tag{20}$$

This stationary process represents a simplified ARIMA-GARCH model, replacing the conditional mean and variance their unconditional equivalents. After fixing parameter values (ω_A, σ_A, ω_B, σ_B) we can sample from the normal distribution to produce stationary time series of arbitrary length. By sampling a series each from specifications A and B, then calculating the dynamic time-warping distance between the two, we can trace the effect of the parameters on the value of the DTW metric. By repeating this process N times we can build up a distribution of these values. A similar apprach can be used to capture the null distribution of the dynamic time-warping metric for two unit root processes that obey the following specification.

$$X_t = \omega + X_{t-1} + \epsilon_t \qquad \epsilon_t \sim \mathcal{N}(0, \sigma^2) \tag{21}$$

From equation (21) at any time t we know both the expectation and the variance of both random walk processes:

$$X_{i,t} \sim \mathcal{N}(t\omega_i,\, t\sigma_i^2) \tag{22}$$

ω_A	ω_B	σ_A^2	σ_B^2	Min	1st Qu.	Median	Mean	3rd Qu.	Max	IQR
0	0	1	1	0.72	0.82	0.85	0.85	0.89	0.97	0.06
0	0	1	2	0.86	1.04	1.08	1.09	1.13	1.25	0.09
0	0	1	4	1.19	1.44	1.52	1.52	1.59	1.83	0.15
0	0	1	8	1.72	2.07	2.18	2.19	2.29	2.70	0.22
0	0	1	16	2.55	2.99	3.16	3.17	3.32	4.05	0.33
0	0	2	2	0.99	1.16	1.22	1.21	1.26	1.42	0.10
0	0	4	4	1.39	1.64	1.71	1.71	1.77	2.02	0.13
0	0	8	8	2.06	2.32	2.42	2.42	2.51	2.93	0.18
0	0	16	16	2.95	3.31	3.44	3.44	3.56	4.14	0.25
2	0	1	1	1.23	1.41	1.48	1.48	1.53	1.80	0.12
4	0	1	1	2.32	3.32	3.58	3.58	3.86	4.76	0.54
8	0	1	1	9.47	11.06	11.41	11.39	11.75	12.86	0.69
16	0	1	1	25.16	27.06	27.41	27.38	27.74	28.68	0.68

Table 1: Stationary White Noise

N = 500; T = 100

Note: The DTW value is divided by the series length (T).

Note: The distributions of the DTW samples are really symmetric. The median and mean remain close in value over all configurations.

Note: For the stationary series the deterministic component (distance between the mean values) plays an influencial role in the location of the center of the DTW distribution.

References

Bollerslev, Tim (1986). "Generalized autoregressive conditional heteroskedasticity". In: *Journal of Econometrics* 31.3, pp. 307–327. ISSN: 0304-4076. DOI: https://doi.org/10.1016/0304-4076(86)90063-1. URL: https://www.sciencedirect.com/science/article/pii/0304407686900631.

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.

Fernandez, Carmen and Mark Steel (1998). "On Bayesian Modeling of Fat Tails and Skewness". In: *Journal of The American Statistical Association* 93, pp. 359–371. DOI: 10.1080/01621459.1998.10474117.

ω_A	ω_B	σ_A^2	σ_B^2	Min	1st Qu.	Median	Mean	3rd Qu.	Max	IQR
0	0	1	1	0.84	2.36	4.42	7.17	9.31	46.87	6.94
0	0	1	2	1.39	3.57	7.05	10.09	13.84	50.89	10.27
0	0	1	4	1.70	4.87	9.01	12.88	17.17	74.00	12.29
0	0	1	8	2.88	6.81	12.30	16.34	22.30	83.36	15.48
0	0	1	16	3.97	10.93	18.41	23.23	32.04	101.38	21.11
0	0	2	2	1.43	3.44	6.48	10.41	13.85	69.20	10.41
0	0	4	4	1.73	4.83	8.98	14.41	19.41	91.95	14.58
0	0	8	8	2.75	7.38	15.66	23.12	31.83	152.90	24.45
0	0	16	16	4.26	10.05	19.77	31.23	40.61	251.30	30.56

Table 2: Random Walk without Drift

N = 500; T = 100

Note: The DTW value is divided by the series length (T).

Note:

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.

Frühwirth-Schnatter, Sylvia and Sylvia Kaufmann (2004). "Model-Based Clustering of Multiple Time Series". In: *Journal of Business and Economic Statistics* 26. DOI: 10.1198/073500107000000106.

Ghalanos, Alexios (2020). *rugarch: Univariate GARCH models*. R package version 1.4-4.

Glosten, Lawrence R, Ravi Jagannathan, and David E Runkle (1993). "On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks". In: *Journal of Finance* 48.5, pp. 1779–1801. URL: https://EconPapers.repec.org/RePEc:bla:jfinan:v:48:y:1993:i:5:p:1779-1801.

Hong, Yongmiao and Haitao Li (2005). "Nonparametric Specification Testing for Continuous-Time Models with Applications to Term Structure of Interest Rates". In: *The Review of Financial Studies* 18.1, pp. 37–84. ISSN: 08939454, 14657368. URL: http://www.jstor.org/stable/3598067.

Hyndman, Rob J. and Yeasmin Khandakar (2008). "Automatic Time Series Forecasting: The forecast Package for R". In: *Journal of Statistical Software, Articles* 27.3, pp. 1–22. ISSN: 1548-7660. DOI: 10.18637/jss.v027.i03. URL: https://www.jstatsoft.org/v027/i03.

Hyndman, Rob et al. (2020). *forecast: Forecasting functions for time series and linear models*. R package version 8.13. URL: https://pkg.robjhyndman.com/forecast/.

ω_A	ω_B	σ_A^2	σ_B^2	Min	1st Qu.	Median	Mean	3rd Qu.	Max	IQR
1	1	1	1	0.85	1.12	1.48	1.97	2.36	8.62	1.24
1	1	1	2	1.00	1.42	2.02	2.75	3.48	12.82	2.05
1	1	1	4	1.31	1.97	2.79	4.01	4.62	19.91	2.64
1	1	1	8	1.68	3.08	4.55	6.49	8.08	40.11	5.00
1	1	1	16	2.88	5.19	7.64	11.02	13.26	68.12	8.07
1	1	2	2	1.23	1.63	2.38	3.23	3.76	22.63	2.13
1	1	4	4	1.73	2.43	3.66	5.62	7.60	36.24	5.17
1	1	8	8	2.60	4.06	6.49	9.38	11.70	48.24	7.64
1	1	16	16	3.93	6.57	10.87	17.00	20.03	108.67	13.46
2	1	1	1	15.18	27.33	32.43	32.74	37.32	59.31	9.99
4	1	1	1	115.79	136.75	143.33	143.48	150.31	181.75	13.56
8	1	1	1	355.60	382.08	389.51	389.48	396.75	422.52	14.67
16	1	1	1	854.08	883.86	891.20	891.33	899.22	927.54	15.36

Table 3: Random Walk with Drift

N = 500; T = 100

Note: The DTW value is divided by the series length (T).

Note: The distributions are skewed towards larger values of the DTW metric (mean is greater than the median).

Note: Compare the last row of each section where variance or the distance between intercepts is set to 16. For the white noise process the strongest factor contributing to the mean DTW value is the difference between the mean values. In comparison, the response to an increase in the variance of either series is muted.

Jarque, Carlos M. and Anil K. Bera (1980). "Efficient tests for normality, homoscedasticity and serial independence of regression residuals". In: *Economics Letters* 6.3, pp. 255–259. ISSN: 0165-1765. DOI: https://doi.org/10.1016/0165-1765(80)90024-5. URL: https://www.sciencedirect.com/science/article/pii/0165176580900245.

Kotsifakos, Alexios et al. (2011). "Model-Based Search in Large Time Series Databases". In: vol. 36, p. 36. DOI: 10.1145/2141622.2141666.

Lee, Gary G. J. and R. Engle (1993). "A Permanent and Transitory Component Model of Stock Return Volatility". In:

Markowitz, Harry (1952). "Portfolio Selection". In: *The Journal of Finance* 7.1, pp. 77–91. ISSN: 00221082, 15406261. URL: http://www.jstor.org/stable/2975974.

Müller, Meinard (2007). "Information Retrieval for Music and Motion". In: Springer, Berlin, Heidelberg. Chap. 4, pp. 69–84.

Petitjean, François, Alain Ketterlin, and Pierre Gançarski (2011). "A global averaging method for dynamic time warping, with applications to clustering". In: *Pattern Recognition* 44.3, pp. 678–693. ISSN: 0031-3203. DOI: https://doi.org/10.1016/j.patcog.2010.09.013. URL: https://www.sciencedirect.com/science/article/pii/S003132031000453X.

Shapiro, S. S. and M. B. Wilk (1965). "An analysis of variance test for normality (complete samples)". In: *Biometrika* 52.3-4, pp. 591–611. ISSN: 0006-3444. DOI: 10.1093/biomet/52.3-4.591. eprint: https://academic.oup.com/biomet/article-pdf/52/3-4/591/962907/52-3-4-591.pdf. URL: https://doi.org/10.1093/biomet/52.3-4.591.

Sharpe, William F (1963). "A simplified model for portfolio analysis". In: *Management science* 9.2, pp. 277–293.

Team, R Core (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. URL: https://www.R-project.org/.

Wang, Gang-Jin et al. (2012). "Similarity measure and topology evolution of foreign exchange markets using dynamic time warping method: Evidence from minimal spanning tree". In: *Physica A: Statistical Mechanics and its Applications* 391.16, pp. 4136–4146. ISSN: 0378-4371. DOI: https://doi.org/10.1016/j.physa.2012.03.036. URL: https://www.sciencedirect.com/science/article/pii/S0378437112002889.