

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 Markowitz, 1952; Sharpe, 1963. 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 its shortcomings (asymmetry, fat-tails) when trying to capture true return correlations between different financial time series, it's still the first measure researchers 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:

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

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 surprised 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 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 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
- Pearson Correlation
- L_p norms
- Dynamic Time Warping
- Parameter value clustering (or by other metrics e.g. persistence)
- Autocorrelation
- Kullback-Leibler

2 Relevant Literature

(Esling & Agon, 2012) 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 measures are classified into four categories: shape, edit, feature, and structure-based.

(Frühwirth-Schnatter & Kaufmann, 2004) cluster time series into K groups based on the fitted parameters of AR(p) and dynamic regression models.

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 (TQues)	✓	✓	✓			$O(n^2 \log n)$	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

(Müller, 2007) provides an excellent summary of the principles of DTW and discuss several extensions with respect to the local and global parameters of the tech-

nique.

(Franses & Wiemann, 2020) 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 utilize a K-Means algorithm proposed by Petitjean, Ketterlin, and Gançarski, 2011 and adapted for time series to group the US states into seven distinct business cycle clusters.

(Wang, Xie, Han, & Sun, 2012) 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.

(Kotsifakos, Athitsos, Papapetrou, Hollmén, & Gunopulos, 2011) are also motivated 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).

(Keogh & Ratanamahatana, 2005) Provide a theory for the most efficient lower bound on the dtw value. The authors adapt Piecewise Aggregate Approximation (PAA) for comparing series in a time-warping context.

(Jeong, Jeong, & Omitaomu, 2011) Instead of providing a universal bound or step condition the authors come up with a temporal weighting function and apply it to the cost matrix before calculating the accumulated cost matrix. For temporal weights the authors use a generalized logistic function that uses two parameters to center and scale the output of the function. The scale parameter is optimized on a validation set while the Performance is evaluated on a test set.

(Duy & Takeuchi, 2023) Rare paper dealing directly with statistical inference of the DTW measure.

(Wang, Xie, Han, & Sun, 2012) TBD

(Howard, Putniņš, & Alexeev, 2020) Unpublished manuscript attempt that applies a DTW step to a classic CAPM stock return analysis. The CAPM model states that an asset’s return is governed by the following relationship:

$$E[r_j] = r_f + \beta_j(E[r_m] - r_f) \quad (4)$$

Instead of running the above regression using synchronous returns the authors first

”align” the sequences using the DTW algorithm and then calculate the betas $\beta_j = Cov(r_{t_i}, r_{s_i})/Var(r_m)$. The authors note that after approaching the estimation of β in this way

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 page¹ as of February 2021 are included in this essay’s analysis. 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 of the price history for the individual companies in the index.

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

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

²<https://www.alphavantage.co/>

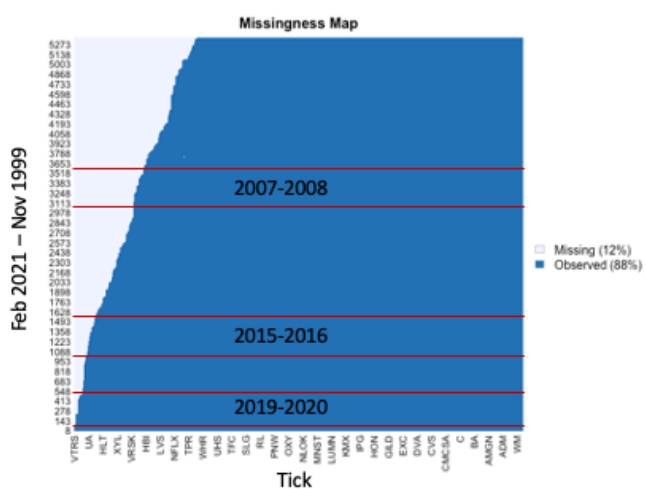


Figure 2: Missing map of the S&P data set. The red horizontal bars demarcate the timeframes under consideration. Of the 505 tickers in the dataset, there are 423 with a full series stock values from 2007 onward, 480 from 2015 onward, and 494 from 2019 onward.

residuals for each stock. For model validation, I will follow the same process as in section 4.3 of Dowiak, 2018. 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:

$$x_t = \mu_t + \epsilon_t \quad (5)$$

$$\mu_t = \mu(\phi, \theta, x_{\{s: s < t\}}, \varepsilon_{\{s: s < t\}}) = \phi_0 + \sum_{i=1}^p \phi_i x_{t-i} + \sum_{j=1}^q \theta_j \varepsilon_{t-j} + \varepsilon_t \quad (6)$$

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.

$$\sigma_t^2 = C(\varepsilon_{t-1}^2, \sigma_{t-1}^2) + \sum_{j=1}^m \alpha_j A(\varepsilon_{t-j}^2) + \sum_{i=1}^k \beta_i B(\sigma_{t-i}^2) \quad (7)$$

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

$$z_t(\phi, \theta, \alpha, \beta) = \frac{x_t - \mu(x_{t-1}, \phi, \theta)}{\sigma(x_{t-1}, \alpha, \beta)} \quad (8)$$

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:

$$f(x_t | \mu_t, \sigma_t^2, \lambda, \Delta) = \frac{1}{\sigma_t} g(z_t | \lambda, \Delta) \quad (9)$$

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 Hyndman and Khandakar, 2008, 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 Hyndman et al., 2020 running in the R statistical language Team, 2021. 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 Ghalanos, 2020. 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: $g = \{\text{Normal, Student-t, Skewed Student-t Fernandez and Steel, 1998}\}$
- Model Specification: $A, B, C = \{\text{Standard GARCH Bollerslev, 1986, gjr-GARCH Glosten, Jagannathan, and Runkle, 1993, Component GARCH Lee and Engle, 1993 }\}$

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 Hong and Li, 2005 non-parametric density test jointly for i.i.d and correct distribution specification, (d) Shapiro-Wilks Shapiro and Wilk, 1965 test for normality, and (e) Jarque-Bera Jarque and Bera, 1980 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 cost³ 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 Müller, 2007.

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 $\mathbf{p} = [p_1, \dots, p_L]$ 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.

$$c : x_{t_l} \times y_{s_l} \rightarrow \mathbb{R}_{\geq 0} \quad (10)$$

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*: $p_l - p_{l-1} \in \{(1, 0), (0, 1), (1, 1)\}$

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

³This is the language used in Müller, 2007 but other authors use "association", "dissimilarity", or "distance" functions.

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:

$$\mathbf{C}(\mathbf{X}, \mathbf{Y}) = \begin{bmatrix} c(x_T, y_1) & c(x_T, y_2) & \cdots & c(x_T, y_S) \\ \vdots & \vdots & \vdots & \vdots \\ c(x_2, y_1) & c(x_2, y_2) & \cdots & c(x_2, y_S) \\ c(x_1, y_1) & c(x_1, y_2) & \cdots & c(x_1, y_S) \end{bmatrix} \quad (11)$$

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 \mathbf{C}) to their end (top right of \mathbf{C}):

$$\mathfrak{c}_{\mathbf{p}}(\mathbf{X}, \mathbf{Y}) = \sum_{l=1}^L c(x_{t_l}, y_{s_l}) \quad (12)$$

There are many permissable⁴ 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 $\mathbb{P}(\mathbf{X}, \mathbf{Y})$ then the value of the optimal warping path has the property

$$\mathfrak{c}_{\mathbf{p}^*}(\mathbf{X}, \mathbf{Y}) \leq \mathfrak{c}_{\mathbf{p}}(\mathbf{X}, \mathbf{Y}) \text{ for all } \mathbf{p} \in \mathbb{P}(\mathbf{X}, \mathbf{Y}) \quad (13)$$

and the value of the DTW measure between \mathbf{X} and \mathbf{Y} is set to $\mathfrak{c}_{\mathbf{p}^*}(\mathbf{X}, \mathbf{Y})$. 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 computational 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* \mathbf{A} needs to be defined. The accumulated cost matrix has the same dimension as the cost matrix. Defining $A_{t,s}$ as the value of \mathbf{A} at the t^{th} row and the s^{th} column of the accumulated cost matrix, the matrix has the following three identities:

⁴Permissable as governed by the boundary and step conditions

$$A_{t,1} = \sum_{k=1}^t c(x_k, y_1) \text{ for } t \in [1 : T] \quad (14)$$

$$A_{1,s} = \sum_{k=1}^s c(x_1, y_k) \text{ for } s \in [1 : S] \quad (15)$$

$$A_{t,s} = c(x_t, y_s) + \min(A_{t-1,s-1}, A_{t-1,s}, A_{t,s-1}) \quad (16)$$

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 $p_{l-1} = \begin{cases} (1, s-1) & \text{if } t = 1 \\ (t-1, 1) & \text{if } s = 1 \\ \arg \min [A_{t-1,s-1}, A_{t,s-1}, A_{t-1,s}] & \text{otherwise} \end{cases}$

4.2.2 Common Extensions

The definition of the DTW in section 4.2.1 is an unconstrained one without restrictions beyond the boundary and step-size conditions. Most applications of dynamic time-warping constrain the extent of time distortion in some way. From a finance and macroeconomic perspective there are reasonable arguments for a limit as well. New information in markets dissipates over time. There are numerous modifications of the original DTW diverse set of industries and domains but applications to financial or economics are relatively sparse. Unit root behavior is a common challenge in the finance and macro-economic fields.

- Temporal weight approaches
 - (Jeong, Jeong, & Omitaomu, 2011)
 - This article if we choose a weighting function suitable for
- Step-Size Constraints
 - Sakoe and Chiba (1978)
- Global Bounds

4.2.3 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.

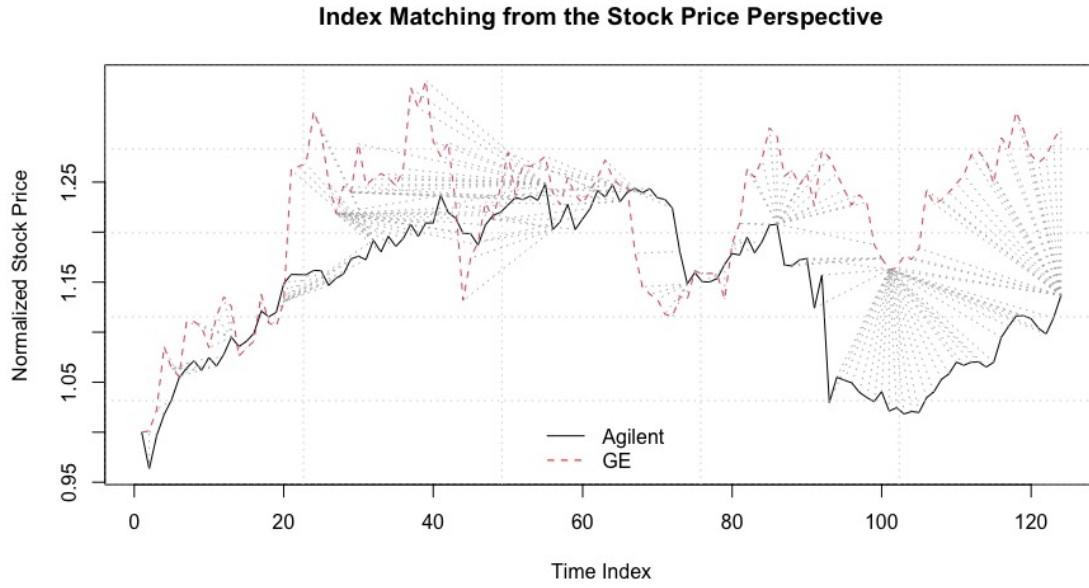


Figure 3: This figure shows a normalized view of the Agilent and GE stock price for the first half of 2019. The normalization divides the stock price by its value on the first market day of the year, January 2nd, 2019. Dynamic time warping is applied on the standardized series. The total length is 124 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.

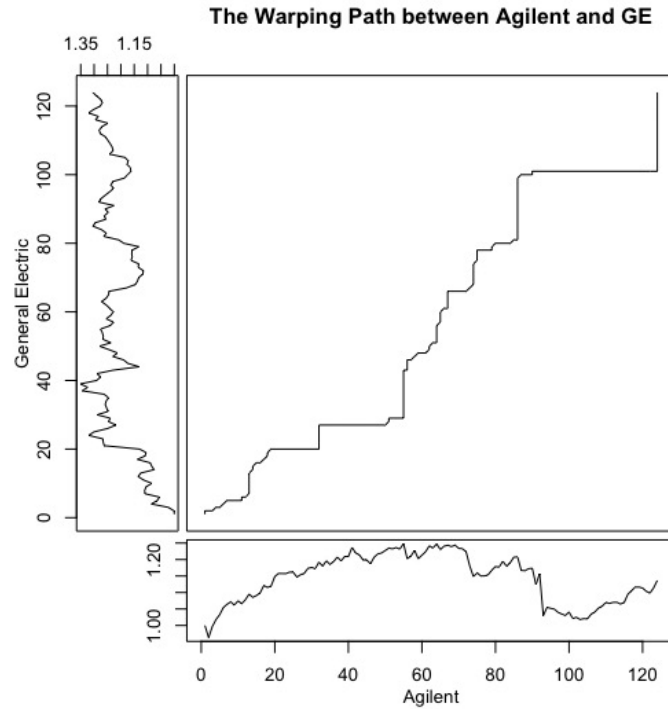


Figure 4: This figure displays the full warping path between Agilent and GE stock prices in the first half of 2019. The two peripheral images show their respective standardized stock prices while the central image traces the warping path as it connects the indices of the two series. The influence of 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 Agilent's.

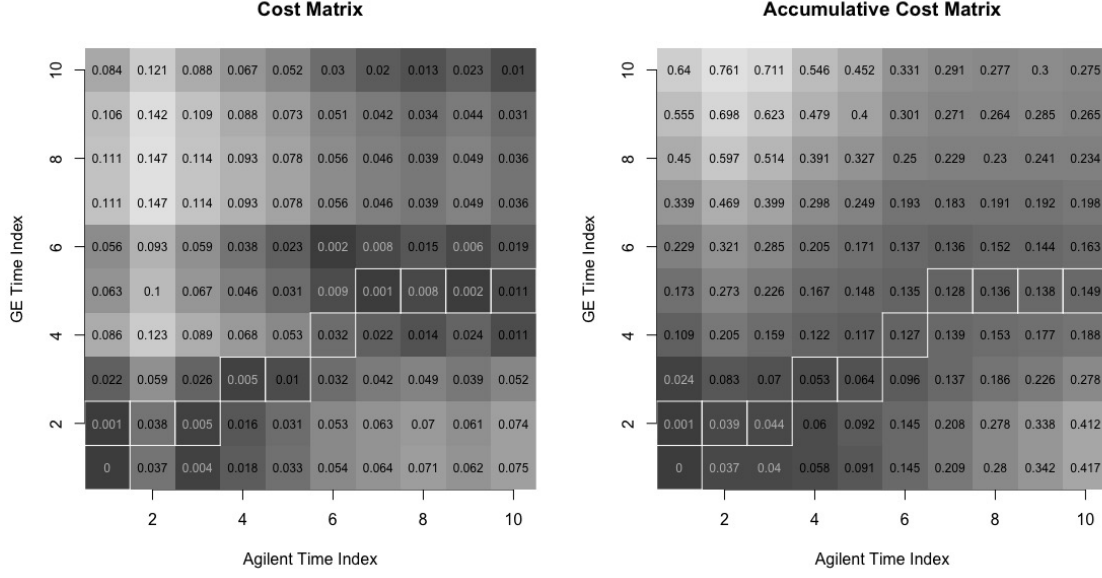


Figure 5: The left image is the cost matrix for the first ten values of the Agilent and GE stock prices after standardization. The local cost function is the euclidean distance. The right image is the accumulative cost matrix for the same sequence of prices. The optimal warping path is annotated with 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.

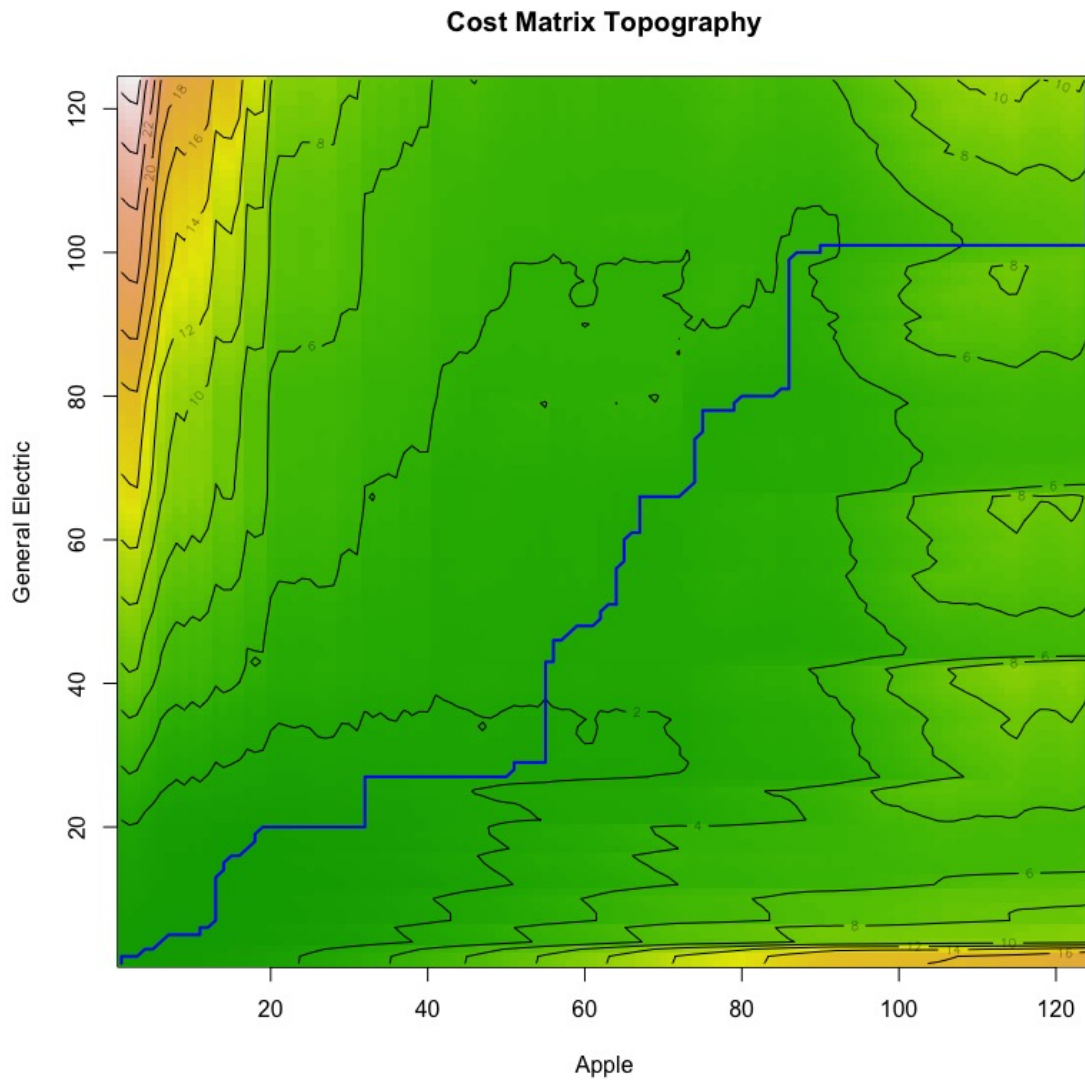


Figure 6: This figure displays the optimal alignment between Aligent and GE stock return series. The topology lines are based on the cost matrix. From this view the warping path (in blue) is a "valley of least cost" traversing the local cost landscape.

5 Pair Trading

One way to compare these measures is putting them to practical use. In this section a simple trading strategy is described that centers around finding "similar" pairs of stocks and trading off the expectation that any major short term deviations in their prices are eventually unwound over a longer time horizon. If the returns to using dynamic time warping to form trading pairs are materially different than returns from a correlation based approach it could help us understand whether dtw can provide complimentary information to use in combination with correlation. The specifics of this pair trading strategy are the same as the approach taken by Gatev, Goetzmann, and Rouwenhorst, 2006 with a few adjustments made to take into account the unique nature of dynamic time warping and its difference to correlation as a measure of similarity. This backtesting experiment is not carried out because of high expected returns, especially in the context of our modern trading environment. The rise of high frequency trading in the 1990's and 2000's has created the financial infrastructure that sees arbitrage opportunities resolved in microseconds⁵ while this article studies asset returns at a daily frequency. In addition Gatev, Goetzmann, and Rouwenhorst, 2006 was published close to two decades ago in a widely read journal making its recipe part of the established financial canon. The benefits to using this particular trading strategy is that it provides a suitable context to directly compare the impact of using dynamic time warping – instead of correlation – to identify the "similar" pairs to trade.

Given a single stock-pair in a trader's portfolio a buy-signal is triggered if the price deviation between the two stocks exceeds a certain threshold. Once triggered the trader goes long in the lower priced stock and short in the higher priced stock. The positions are unwound when the prices come back to parity, profits are recorded, and the trader stands ready to open another position given a new buy-signal. At the end of the trading period all open positions are closed at the prevailing price. To summarize, the general strategy has the following sequence of steps:

1. Set length of time to use for pair selection and cross-validation for optimal buy-signal thresholds. Set length of time for the execution of the trading strategy. In this article a one year duration is used for both pair selection and trade execution periods.
2. Standardize all stock prices at the beginning of the pair selection and trading period intervals.

⁵See Aquilina, Budish, and O'Neill, 2022, for a more detailed.

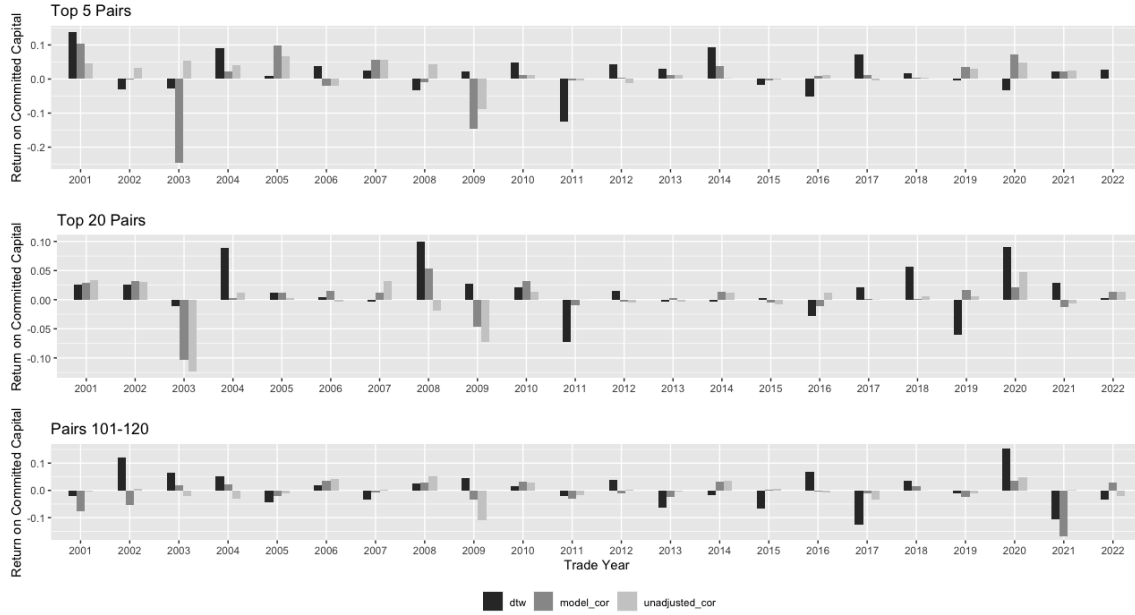


Figure 7: Barcharts indicating the annual committed returns of the trading strategy under different buy signals and portfolio sizes.

3. Pair Selection: Use recent past price movements to identify the top N pairs of stocks that are the most *similar*. Use the same period to find the optimal buy-signal threshold via cross-validation.
4. Trading Period: Execute the strategy for each of the N stock pair in the portfolio. Close out any open positions at the end of the trading period.

There are three different proposed methods for choosing the association measure:

1. Use the un-treated returns. Measure closeness via correlation. Use static (unconditional) distributions and a two standard deviation threshold on the cumulative return differential as a signal to open positions.
2. Filter the returns through an ARIMA-GARCH model. Measure closeness on the fitted residuals. Use rolling mean/variance values to calculate adjusted threshold during trading period
3. Calculate the DTW metric between the cumulative returns of the pairs. Use synchronous L1 value and the two standard deviation signal to open positions.

TODO:

1. Need to compare pairs formation between the three buy strategies
2. Compare Top 5 and Top 20 correlation pairs. Returns between the two strategies are highly correlated. How often does the model make different pair recommendations?

6 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.

$$X_t = \omega + \epsilon_t \quad \epsilon_t \sim \mathcal{N}(0, \sigma^2) \quad (17)$$

This stationary process represents a simplified ARIMA-GARCH model, replacing the conditional mean and variance with their unconditional equivalents. After fixing parameter values $(\omega_A, \sigma_A, \omega_B, \sigma_B)$ 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 approach can be used to capture the null distribution of the dynamic time-warping metric for two unit root processes that obey the following specification.

$$X_t = \omega + X_{t-1} + \epsilon_t \quad \epsilon_t \sim \mathcal{N}(0, \sigma^2) \quad (18)$$

From equation (18) at any time t we know both the expectation and the variance of both random walk processes:

$$X_{i,t} \sim \mathcal{N}(t\omega_i, t\sigma_i^2) \quad (19)$$

A random walk temporal weighting function could have the following form:

$$\omega_{t,s} = |t - s| \sqrt{\sigma_{A,t} \sigma_{B,s}} \quad (20)$$

$$\omega_{t,s} = \rho |t - s| \sigma_{A,t} \sigma_{B,s} \quad (21)$$

$$\omega(t, s) = \sum_{k=1}^{|t-s|} \beta^{|t-s|} (\sigma_{A,t}^2 + \sigma_{B,s}^2) \quad (22)$$

References

- Aquilina, Matteo, Eric Budish, and Peter O'Neill (2022). "Quantifying the High-Frequency Trading Arms Race". In: *The Quarterly Journal of Economics* 137.1, pp. 493–564. ISSN: 0033-5533. DOI: 10.1093/qje/qjab032. eprint: <https://academic.oup.com/qje/article-pdf/137/1/493/41929626/qjab032.pdf>.
- 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](https://doi.org/10.1016/0304-4076(86)90063-1).
- Dowiak, L. C. (2018). "A Comparison of Time-Varying Copula-GARCH Models". In: *Unpublished* 00, p. 00.
- Duy, Vo Nguyen Le and Ichiro Takeuchi (2023). *Statistical Inference for the Dynamic Time Warping Distance, with Application to Abnormal Time-Series Detection*. arXiv: 2202.06593 [stat.ML].
- 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.
- 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.
- Gatev, Evan, William N. Goetzmann, and K. Geert Rouwenhorst (2006). "Pairs Trading: Performance of a Relative-Value Arbitrage Rule". In: *Review of Financial Studies* 19.3, pp. 797–827. ISSN: 0893-9454. DOI: 10.1093/rfs/hhj020. eprint: <https://academic.oup.com/rfs/article-pdf/19/3/797/24421685/hhj020.pdf>.
- 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.
- 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.

- Howard, Clint, Tālis Putniņš, and Vitali Alexeev (2020). “To lead or to lag? Measuring asynchronicity in financial time-series using dynamic time warping”.
- 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](https://doi.org/10.18637/jss.v027.i03).
- Hyndman, Rob et al. (2020). *forecast: Forecasting functions for time series and linear models*. R package version 8.13.
- 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](https://doi.org/10.1016/0165-1765(80)90024-5).
- Jeong, Young-Seon, Myong K. Jeong, and Olufemi A. Omitaomu (2011). “Weighted dynamic time warping for time series classification”. In: *Pattern Recognition* 44.9. Computer Analysis of Images and Patterns, pp. 2231–2240. ISSN: 0031-3203. DOI: <https://doi.org/10.1016/j.patcog.2010.09.022>.
- Keogh, Eamonn and Chotirat Ann Ratanamahatana (2005). “Exact indexing of dynamic time warping”. In: *Knowledge and Information Systems* 7.3, pp. 358–386. DOI: [10.1007/s10115-004-0154-9](https://doi.org/10.1007/s10115-004-0154-9).
- Kotsifakos, Alexios et al. (2011). “Model-Based Search in Large Time Series Databases”. In: *Proceedings of the 4th International Conference on Pervasive Technologies Related to Assistive Environments*. PETRA ’11. Heraklion, Crete, Greece: Association for Computing Machinery. ISBN: 9781450307727. DOI: [10.1145/2141622.2141666](https://doi.org/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.
- 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>.
- 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](https://doi.org/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>.

- 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.
- 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>.

I. Returns on Committed Equity									
# Pairs	Pair Strategy	Mean	Std Err ¹	Std Dev	Median	Skew	Kurtosis	Min	Max
Top 25	unadjusted cor	0.0018	0.0024	0.030	0.0011	-0.873	5.039	-0.085	0.061
Top 25	model cor	-0.0033	0.0040	0.027	0.0003	-1.199	4.417	-0.078	0.032
Top 25	dtw	0.0185	0.0050**	0.041	0.0096	0.695	2.992	-0.041	0.114
Top 50	unadjusted cor	-0.0003	0.0017	0.023	-0.0001	0.028	1.918	-0.038	0.043
Top 50	model cor	-0.0011	0.0026	0.025	-0.0054	0.226	1.935	-0.036	0.044
Top 50	dtw	0.0084	0.0054	0.034	0.0032	0.512	2.999	-0.057	0.084
101-200	unadjusted cor	0.0047	0.0015**	0.019	0.0058	0.205	2.592	-0.029	0.044
101-200	model cor	0.0015	0.0020	0.022	0.0073	-1.242	5.263	-0.066	0.034
101-200	dtw	-0.0044	0.0057	0.042	-0.0012	-1.043	3.218	-0.110	0.040
1001-1100	unadjusted cor	-0.0074	0.0031*	0.033	-0.0018	-0.337	2.962	-0.083	0.060
1001-1100	model cor	0.0005	0.0039	0.026	-0.0014	1.276	6.203	-0.044	0.085
1001-1100	dtw	-0.0027	0.0071	0.042	0.0057	-0.085	2.294	-0.086	0.065

II. Returns on Fully Invested Equity									
# Pairs	Pair Strategy	Mean	Std Err ¹	Std Dev	Median	Skew	Kurtosis	Min	Max
Top 25	unadjusted cor	0.0100	0.0063	0.051	0.0021	0.057	4.071	-0.112	0.132
Top 25	model cor	0.0034	0.0089	0.050	0.0005	0.434	3.994	-0.093	0.133
Top 25	dtw	0.0189	0.0050**	0.042	0.0096	0.721	3.062	-0.041	0.119
Top 50	unadjusted cor	0.0015	0.0032	0.038	-0.0006	0.041	1.701	-0.060	0.064
Top 50	model cor	0.0010	0.0055	0.043	-0.0086	0.343	2.234	-0.065	0.097
Top 50	dtw	0.0084	0.0055	0.034	0.0033	0.457	2.973	-0.060	0.084
101-200	unadjusted cor	0.0097	0.0023**	0.031	0.0088	0.139	2.066	-0.037	0.065
101-200	model cor	0.0046	0.0024+	0.032	0.0128	-0.600	2.913	-0.076	0.047
101-200	dtw	-0.0045	0.0058	0.042	-0.0012	-1.046	3.222	-0.111	0.041
1001-1100	unadjusted cor	-0.0118	0.0060*	0.050	-0.0023	-0.318	2.235	-0.113	0.075
1001-1100	model cor	-0.0028	0.0058	0.045	-0.0020	-0.293	4.727	-0.125	0.107
1001-1100	dtw	-0.0031	0.0071	0.043	0.0058	-0.080	2.253	-0.087	0.065

III. Baseline Long Position									
# Pairs	Pair Strategy	Mean	Std Err ¹	Std Dev	Median	Skew	Kurtosis	Min	Max
Top 25	unadjusted cor	0.0557	0.0228*	0.232	0.1049	-0.234	2.753	-0.463	0.513
Top 25	model cor	0.0487	0.0225*	0.235	0.0815	-0.003	3.256	-0.472	0.597
Top 25	dtw	0.0899	0.0140**	0.149	0.1371	-1.196	4.021	-0.333	0.266
Top 50	unadjusted cor	0.0648	0.0237**	0.227	0.1161	-0.633	2.837	-0.490	0.413
Top 50	model cor	0.0596	0.0241*	0.227	0.0947	-0.466	2.674	-0.470	0.431
Top 50	dtw	0.0711	0.0148**	0.149	0.1266	-1.079	3.986	-0.339	0.274
101-200	unadjusted cor	0.0739	0.0219**	0.210	0.1129	-0.890	3.521	-0.496	0.362
101-200	model cor	0.0907	0.0255**	0.216	0.1432	-0.719	3.599	-0.505	0.427
101-200	dtw	0.0840	0.0181**	0.165	0.1309	-1.073	4.193	-0.388	0.293
1001-1100	unadjusted cor	0.0756	0.0281**	0.214	0.1423	-1.033	3.905	-0.522	0.362
1001-1100	model cor	0.0655	0.0280*	0.205	0.0941	-0.982	3.869	-0.500	0.333
1001-1100	dtw	0.0954	0.0170**	0.171	0.1381	-1.014	4.276	-0.402	0.326

Table 1: Average Annual Return Distribution

Three different measures of portfolio returns are summarized in this table. The top section records committed returns where the total return of the portfolio during the trading period is divided by the number of pairs selected to trade. Section II of the table records fully invested returns where the total return of the portfolio during the trading period is divided by the number of pairs selected to trade. Section III records the annual return of the portfolio if the trader simply goes long in at the beginning of the trading period.

Signif. Codes: 0 '***', 0.01 '**', 0.05 '+', 0.1 ' ', 1

¹Standard errors are Newey-West estimates and significant values are

		Communication Services	Consumer Discretionary	Consumer Staples	Energy	Financials	Health Care	Industrials	Information Technology	Materials	Real Estate	Utilities
Dynamic Time Warping												
Communication Services	14
Consumer Discretionary	.	3
Consumer Staples	2	3	13
Energy	.	.	1	5
Financials	2	2	13	.	50
Health Care	3	.	12	1	12	6
Industrials	3	4	7	3	8	6	20
Information Technology	2	1	1	2	9	3	6	2
Materials	.	3	1	2	4	1	6	2	2	.	.	.
Real Estate	.	1	1	.	2	.	2	.	2	49	.	.
Utilities	4	4	25	2	11	9	6	1	3	9	178	.
Model Correlation												
Communication Services	17
Consumer Discretionary	.	36
Consumer Staples
Energy	.	.	.	52
Financials	117
Health Care	1
Industrials	13
Information Technology	44
Materials	10	.	.	.
Real Estate	3	150	.	.
Utilities	107	.
Unadjusted Correlation												
Communication Services	17
Consumer Discretionary	.	30
Consumer Staples
Energy	.	.	.	47
Financials	147
Health Care
Industrials	10
Information Technology	36
Materials	5	.	.	.
Real Estate	2	161	.	.
Utilities	95	.

Table 2: Sector Distribution of Top 25 Portfolio Pairs
Footnote size table note

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 3: 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 influential role in the location of the center of the DTW distribution.

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 4: Random Walk without Drift

$N = 500$; $T = 100$

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

Note:

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 5: 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.