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 [27, 36]. 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 6, 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:

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

with

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

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 [8] 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. [8] categorize the these distance measures into four categories collected in figure 1. A very similar paper was published in Computational Economics very recently by [10].

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. persistance)
- Autocorrelation
- Kullback-Leibler

2 Relevant Literature

[34] provides the original paper defining dynamic time warping with application in the field of speech recognition. It's influence has extended into applications such data mining, agricultural and forestry management, high-frequency trading.

The algorithm has certian properties that are really convienent from the view of an analyst. It can handle times series of different lengths as well as series with observations that are not perfectly synchronous in time like equationn 6. This makes it useful where the timing of observations from inputs are not aligned such as in high frequency trading or streams of satalite sensors [28].

Applications of dynamic time warping in the finance and economic literature usually fall into one of the following categories.

1. The most popular use is to find clusters or other sub-populations in a large set of sequential data. The approach is similar in spirit to the common K-means classifiers but the challenge in sequential data, including economic time series, is to find an appropriate version of the "average" of a group of series. Using the euclidean distance between series can lead to a group representation that does not capture important shape features of the underlying set of series it is summarizing [31]. One approach that prevents Several articles use a K-Medoid approach which selects a member of the set that minimizes the within-group dissimilarity.

2. Another general use for dynamic time warping Template Pattern Matching and Nearest Neighbor Classication. Satelite imagery and stock pattern matching

3. Lead/Lag estimation and Temporal Realignment

There's the simple k-means classification where the researcher is interested in sub-population structures. Class classification and latent clusering

DTW is only one way to index a set of series for clustering applications. [8] provides an overview of various methods and metrics to used to index a broad class of sequential data.

1. Latent cluster identification where no explicit labels are provided or prior knowledge of any underlying multinomial distribution

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

    - [10] 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 [31] and adapted for time series to group the US states into seven distinct business cycle clusters.

    - [31] illustrates a method to find a global average for a set of sequences. The "average" of a set of sequences is defined as the sequence that minimizes the within group sum of sequares distances from each sequence in the set to the "average" sequence. The distance in between sequences is calculated by the dynamic time warping algorithm. The bulk of the paper discusses the process of updating the "average" sequence from one iteration to the next.

2. Template Pattern Matching and Nearest Neighbor Classification
   - [38] uses DTW in support of technical anaylsis of financial assets.
   - [32] demonstrates that [31] can be used to form efficient nearest neighbor classification decisions where the nearest group average determines the classification. Performs better that Euclidean averages or mediods.
   - [25] 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 an example-base query using DTW with a model-based approach using Hidden Markov Models (HMM).
3. Lead/Lag estimation. Temporal alignment. Linking asynchroneous time series. High frequency trading application. Leading economic indicator analysis. Land-use identification from Satelite data.
   - [23] 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.
   - [28] This paper presents a time-weighted version of the dynamic time warping (DTW) method for land-use and land-cover classification using remote sensing image time series. Authors of the dtwSat package. Uses logistic-based time weights to temporally align signals from satelite imagery before classification decision. Weights are *additive* to the DTW cost function. Supervised learning with manually labeled examples. The approach in this paper is copied by [4], [40], [30], [33]
   - [21] Propose a Multinomial DTW (MDTW) to directly estimate any existing lead/lag relationship between pairs of financial assets trading at high-frequency. They compare their approach to existing methods in the literature. Evidence from this paper (See table 8) supports our analysis that the variance of DTW increases as the correlation between asset movements decreases. This paper uses simulates stock pairs evolutions with a

bivariate brownian motion governed by ρ and σ^2. DTW is calculated on the (stationary) log return series similuated by the 2d brownian motion, something that differentiates this work from mine.

4. Counter-argument to using DTW on financial data - Hypothesis: Unit roots cause issues even controlling for different time series scales.

    - [6] Use raw returns instead of modeling conditional mean or variances. Multivariate time series clustering approach based on a trimmed mediod clustering model. Similar to [31] in that both are concerned with finding an efficient "center" of a group of time series.

    - [2] uses graph theory to model the dynamic nature of financial networks

    - [17] Unpublished manuscript attempt that applies a DTW step to a classic CAPM stock return analysis. Concentrates on the high frequency domain. The CAPM model states that an asset's return is governed by the following relationship:

    $$E[r_j] = r_f + \beta_j(E[r_m] - r_f) \tag{4}$$

      Instead of running the above regression using synchronous returns the authors first "align" the sequences using the DTW algorithm and then calculate the betas $\beta_j = Cov(r_{t_l}, r_{s_l})/Var(r_m)$. The authors build in The authors note that after approaching the estimation of β in this way. Additional simulation results are used to show that their approach back out different forms of dynamic lead/lag evolution. Model price movements as a random walk without drift.

[8] 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.

[7] Rare paper dealing directly with statistical inference of the DTW measure.

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

[39] 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.

[24] 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.

[39] TBD

3 Data

The companies comprising the S&P 500, as of October 2023, are included in this essay's analysis. These large-cap U.S. based companies provide a reliable sample with highly liquid markets for their stock and broad diversity across sectors. The date range under consideration begins in January 2000 and extends to October 2023. This two decade timeframe covers a number of macroeconomic crises: the fallout from the bursting of the dot-com bubble from 2000-2002, the financial crisis of 2007-2008, and the dramatic shut-down economy of 2020. Interceding these marcroeconomic shocks are two bull markets: a five year bull-run in the early 2000's and the decade long bull-run throughout the 2010s. The daily adjusted close for each stock is sourced by an API provided by Alpha Advantage[1]. The adjusted close takes into account any reinvested dividends as well as stock splits or the introduction of new stock that occured over the duration of the period.

Over time, the composition of the index changes as new companies are introduced and existing companies fall out. There are many companies in this essay that do not have a full history dating back to 2000. Of the 503 stocks in the sample there are 352 with a history that extends back to the beginning of 2000. See Figure 2 for a visual summary of the price history for the sample of companies considered in this essay.

4 Distance Measures

In this section the distance or dissimilarity measures discussed in this article are formally defined. It's important to distinguish the price of a stock versus its return since the distance measures discussed below do not always apply to the same representation of data. All upper-case variables are assumed to be invoiced in the price of

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

the stock while lower-case variables are assumed to be return values. If S_t is the spot price of a stock at time t then the return on a stock from date $t-1$ to t is defined as the log difference between consecutive spot prices.

$$x_t = \log(S_t) - \log(S_{t-1}) \tag{5}$$

Correlatation is applied to a stock's return series while euclidean distance and dynamic time warping are applied to a pair of stock's spot price. The decision to apply dynamic time warping to the price level is purposeful. The papers reference here that want to cluser time series usually use the price level or the nominal value of the series. This contrasts with the articles that want to estimate the magnitude of a leading or lagging temporal relationship. These articles typically apply dynamic time warping on the stationary log returns. Since the main focus of this article is on studying the relationship between correlation and dynamic time warping the decision is made to calculate dynamic time warping at the price level for the majority of this article[2].

4.1 Correlation

The formal definition of Pearson's correlation was given in the introductory paragraph. It is reproduced here for easy reference. For two series of stock returns $(\boldsymbol{x}, \boldsymbol{y})$ the unadjusted correlation is

$$\rho(\boldsymbol{x}, \boldsymbol{y}) = \frac{\sum_{t=1}^{T}(x_t - \bar{x})(y_t - \bar{y})}{\sqrt{\sum_{t=1}^{T}(x_t - \bar{x})^2 \sum_{t=1}^{T}(y_t - \bar{y})^2}} \tag{6}$$

where $\bar{x}$ is the arithmetic mean $\bar{x} = \frac{1}{T}\sum_{t=1}^{T} x_t$. This correlation measure is an unconditional one and makes no attempt to control for any explainable variance in the returns of either stock series. In order to support a set of simulation exercises in section 6 a second, more complex, process will be used to estimate the correlation between to series. The conditional mean and conditional variance of each stock return series will be estimated by the ARMA-GARCH modeling framework. The models, once properly fit, can be used in conjuction with a Copula to estimate a bivariate distribution for a pair of stock returns. A t-Copula, which is characterized by

[2]Simulation exercises conducted in section 6.1 are an exception

it's correlation and degrees-of-freedom parameters (ρ, ν), is optimized via maximum likelihood.

$$\begin{aligned} C_{\boldsymbol{t}}\left(u, w; \rho, \nu\right) &= \boldsymbol{t}_{v}\left(t_{\nu}^{-1}\left(u\right), t_{\nu}^{-1}\left(w\right)\right) \\ &= \int_{-\infty}^{t_{\nu}^{-1}(u)} \int_{-\infty}^{t_{\nu}^{-1}(w)} q\left(1-\rho\right)^{-\frac{1}{2}} \left[1 + \nu^{-1}\left(x^2 - 2\rho xy + y^2\right)\right]^{-(\nu+2)/2} dxdy \end{aligned} \tag{7}$$

The values u and w are a transformed representation[3] of random variables x and y from equation 6. The modeled distribution (g), conditional mean (μ_t), and conditional variance (σ_t^2) produced by the ARMA-GARCH framework facilitates this transform. The fitted correlation parameter $(\hat{\rho}_{x,y})$ from this estimation process is used as the second measure of correlation. Greater detail about the ARMA-GARCH modeling strategy used in this article is contained in the Appendix while [5] summarizes Copula theory and it's application to foreign exchange rates.

4.2 Euclidean Distance

An important baseline to compare against dynamic time warping will be euclidean distance. Also referred to as the L^2-norm, euclidean distance is heavily used in non-finance fields to compare the similarity (or dissimilarity) between a pair of sequential data. It is defined on a pair of series with synchronous time stamps.

$$||\boldsymbol{X}, \boldsymbol{Y}||_2 = \left[(X_1 - Y_1)^2 + (X_2 - Y_2)^2 + \cdots + (X_T - Y_T)^2\right]^{\frac{1}{2}} \tag{8}$$

Euclidean distance, as defined in equation 8, serves as benchmark similarity measure where no time elasticity is allowed beween the two series.

4.3 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

[3]See equation 32 in the Appendix

constrained to equal each other – like in equations 6 and 8. Time is allowed to stretch and compress before a local cost function is applied to the pairs of values from the two series. This article will adhere to the classic definition of dynamic time warping. For notation this article borrows heavily from [29].

4.3.1 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 behavior of the warping paths obey two conditions which are listed below.

$$\text{Boundary Condition:} \quad p_1 = (1,1) \text{ and } p_L = (T,S) \tag{10}$$

$$\text{Step-Size Condition:} \quad p_l - p_{l-1} \in \{(1,0),(0,1),(1,1)\} \tag{11}$$

The boundary condition requires that the first and last indices of the two series are to be 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:

$$\mathbf{C}(\boldsymbol{X}, \boldsymbol{Y}) = \begin{bmatrix} c(X_T, Y_1) & c(X_T, Y_2) & \cdots & c(X_T, Y_S) \\ \vdots & \vdots & \vdots & \vdots \\ c(X_2, Y_1) & c(X_2, Y_2) & \cdots & c(X_2, Y_S) \\ c(X_1, Y_1) & c(X_1, Y_2) & \cdots & c(X_1, Y_S) \end{bmatrix} \tag{12}$$

This article uses squared distance[4] for the local cost function to make it comparable with the euclidean distance defined in section 4.2. A warping path's total cost is the

[4] $c(X_i, Y_j) = (X_i - Y_j)^2$

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 $\boldsymbol{C}$):

$$\mathbb{c}_{\boldsymbol{p}}(\boldsymbol{X}, \boldsymbol{Y}) = \sum_{l=1}^{L} c(X_{t_l}, Y_{s_l}) \tag{13}$$

There are many admissable[5] warping paths between the two series. 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

$$\mathbb{c}_{\boldsymbol{p}^*}(\boldsymbol{X}, \boldsymbol{Y}) \leq \mathbb{c}_{\boldsymbol{p}}(\boldsymbol{X}, \boldsymbol{Y}) \text{ for all } \boldsymbol{p} \in \mathbb{P}(\boldsymbol{X}, \boldsymbol{Y}) \tag{14}$$

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 in T and S, the computation time needed to check every warping path becomes problematic for large series.

The proposed solution is to leverage *dynamic programming* to reduce the computational needed to find the optimal solution. Instead of dealing with exponential growth of warping paths the dynamict time warping algorithm is designed to 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 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:

$$A_{t,1} = \sum_{k=1}^{t} c(X_k, Y_1) \text{ for } t \in [1 : T] \tag{15}$$

$$A_{1,s} = \sum_{k=1}^{s} c(X_1, Y_k) \text{ for } s \in [1 : S] \tag{16}$$

$$A_{t,s} = c(X_t, Y_s) + \min\left(A_{t-1,s-1},\, A_{t-1,s},\, A_{t,s-1}\right) \tag{17}$$

[5] Admissable as governed by the boundary and step conditions

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

One more condition is added to this article's application of dynamic time warping. The maximum warping distance between the two return series no greater than 15 trading days. This type of global constraint is called a Sakoe-Chiba band [34] and is frequently used for it's simplicity and it's performance [13]. Most applications of dynamic time-warping constrain the extent of time distoration in some way. From a finance and macroeconomic perspective there are reasonable arguments for a limit as well. The value of new information in markets dissipates quickly over time.

4.3.2 Extended Discussion

The details in section 4.3.1 describe the classic dynamic time warping used in this article. There are numerous variations and extensions that have been discussed that touch on every facet of the algorithm. The boundary condition can be relaxed and the step-condition can be altered to change the local constraints at each step of the dynamic programming procedure [34]. Weighting the local cost value in equations 13 - 17 as a function of the time difference between the two return values has been a popular approach in financial applications, especially when trying to estimate a constant temporal relationship[6]. Curiously, additive weights are the prefered scheme when looking to optimize cluster performance [31, 32]. Aside from local constraints, global restrictions, like the Sakoe-Chiba band discussed at the end of section 4.3.1, can be used to restrict the total amount of warping allowed between two series [20, 34, 29].

[6] The warping path from equation 13 with multiplicative weights could have the following definition: $\mathbb{c}_{\boldsymbol{p}}(\boldsymbol{X}, \boldsymbol{Y}) = \sum_{l=1}^{L} c(X_{t_l}, Y_{s_l}) \cdot \omega(|t_l - s_l|)$

4.4 Examples

To put these concepts into practice the dynamic time warping algorithm will be demonstrated on the stock prices for Teradyne Inc. (TER) and Lam Research Corp. (LRCX) for the 2021 calendar year. Instead of the original stock price each series is normalized by calculating the cumulative return of their log-returns using equation 18. Figure 3 displays the cost matrix and the accumulated cost matrix for the first ten trading days of the two series. The optimal warping path is annotated by the sequence of boxes outlined in white. Figure 4 provides four graphs summarizing topics covered in this section. The top-left graph shows the cumulative return of Teradyne Inc. and Lam Research Corp. stock with line segements connecting the indicies from the optimal warping path. The algorithm is run on the pair's cumulative return series (see equation 18). A 15-day Sakoe-Chiba window provides a global constraint. The bottom-left graph is the same as the top-left but contains vertical line segments between synchronous points in time. The euclidean distance between the pair's normalized prices is 3.345. On the top right is a scatter-plot of un-modeled log-returns. The sample pearson correlation is 0.84. And on the bottom right is the fitted conditional variance from each series' esitmated ARMA-GARCH model.

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 (normalized) prices are eventually unwound before the end of the trading period. For this trade strategy the critical decision is how to select pairs of stocks and what criterion to use for similarity. In this section we directly compare the returns of portfolios formed using correlation against portfolio returns that use dynamic time warping and euclidean distance to measure similarity. 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 this metric can provide meaningful information to use in the analysis of stock returns.

The specifics of this pair trading strategy are the same as the approach taken by [12]. The execution of this strategy takes place in two parts: a *formation period* where a trader will create a portfolio of stock pairs and the *trading period* where the trader will buy and sell those pairs according to a defined set of market signals.

This 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 microseconds[1]. This article studies stock returns captured at a daily frequency and is at a natural disadvantage in comparison. In addition [12] was published close to two decades ago in a widely read finance journal making its recipe part of the established 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.

5.1 Trading Strategy

The following subsections provide more detail about the formation and trading periods and how the different measures of association will be compared.

5.1.1 Formation Period

During the formation period the trader compares historical stock returns and uses a measure of similarity to form a portfolio of N pairs of the closest related stocks. In this article a one year time span serves as the duration of the formation period. Before stock comparisons are made prices are first transformed from their nominal price to a measure of their cumulative return. For any date t a stock's standard price is found by using the following calculation:

$$p_t = \prod_{1}^{t} (1 + r_t) \tag{18}$$

After pair selection the trader again uses the price history in the formation period to calculate a critical value for the price differential experienced by each pair. This article follows [12] and uses a two standard deviation threshold as a signal to open a position on the pair during the trading period. That is if p_t and q_t are the two standardize prices for a pair and $\boldsymbol{d}$ is the vector of absolute price differentials with elements $d_t = |p_t - q_t|$, then over the entire formation period the threshold that acts as a signal to open a trading position is set to $\bar{\boldsymbol{d}} + 2\sigma_{\boldsymbol{d}}$. After price normalization four different portfolio collections are created using the distance measures discussed in section 4: unadjusted and modeled correlation, euclidean distance, and dynamic time warping. Portfolios are formed by the closest 25, 50, and 100 pairs of stocks.

5.1.2 Trading Period

Given a single stock pair in a trader's portfolio a position is opened if the price deviation between the two stocks exceeds the two standard deviation threshold. Once triggered the trader goes long in the lower priced stock and takes a short position 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 fresh buy-signal. That is a single pair can be opened and closed multiple times during a single trading period. If this happens this leads to multiple revenue streams over the duration of the trading period. The total return of this pair is calculated as the compounded return of these multiple revenue streams. Take note that the inclusion of a stock pair in the trader's portfolio does not mean that it must be traded. If the differential in the standard prices of the pair never deviate beyond the two standard deviation limit then a position in a pair will never be opened.

5.2 Post Trade Analysis

Table 1 provides summary results of the trading strategy contrasted by distance measure, portfolio size, and definition of "returns" to the portfolios constructed here. Section I 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 total number of pairs selected in the portfolio. This is a more conservative estimate than committed returns. The third section records the returns on the portfolio if the trader had chosen to buy-and-hold the portfolio for the duration of the trading period.

As expected there does not appear to be significant arbitrage opportunities from this strategy. There is a some evidence that dynamic time warping is a metric with practical use, though. The only statistically positive return on equity from the trading strategy comes from the smaller portfolio of top 25 pairs choosen by dynamic time warping. But the return is small in magnitude, especially compared to the alternative of the baseline long-postion as seen in section III. The DTW and euclidean measures perform better than the correlation measures in amost all categories. The only exception is the the one percent return of fully invested equity to the top 25 pairs chosen by unadjusted correlation, which beats out the portfolio construced by euclidean distance but not the portfolio constructed by dynamic time warping. This could be attributed to how the market signals are constructed during the formation period. Because we are forced to trade in syncronized time this may give euclidean distance an inherent advantage over correlation.

Lookly closely at sections I and II in table 1 the reader will notice that for DTW and Euclidean portfolios, returns on fully invested equity are close to the returns on committed equity across portfolio size. This stands in contrast to portfolios created with either correlation measure. The reason for this is simple. A sizeable proportion of the pairs formed with correlation are never traded because the differential in the cumulative return never exceeds the two standard deviation threshold estimated during the formation period. In fact for the portfolio of top 100 pairs choosen with correlation only 60% of pairs are traded. This compares to 98% of pairs formed using DTW or euclidean distance.

Another contrast between the portfolios constructed with correlation versus DTW and euclidean distance is the sector-level diversification achieved by the latter. Pairs formed in the correlation based portfolios almost always come from the same sector but there is a substatial number of inter-sector pairs formed using DTW and euclidean distance (compare tables 2 and 3). This diversification effect of the DTW and Euclidean portfolios may explain their positive and statistically significant baseline long returns even as the comparable correlation portfolios see positive but non-statistically significant return values.

In figure 5, several barcharts are presented summarizing the annual committed returns to the trading strategy under different buy signals and portfolio sizes. The return volatility to this pair trading strategy is cyclical in nature. The absolute value of returns are largest during and immediately following the three large financial crisis experienced over the time period studied: the follow-out from the dot-com bust in 2001 and 2002, the mortgage back security crisis in 2008 and 2009, and the shut-down economy of 2000.

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	0.0063	0.030	0.0011	-0.873	5.039	-0.085	0.061
Top 25	Model cor	-0.0033	0.0057	0.027	0.0003	-1.199	4.417	-0.078	0.032
Top 25	DTW	0.0185	0.0085*	0.041	0.0096	0.695	2.992	-0.041	0.114
Top 25	Euclidean	0.0064	0.0071	0.034	-0.0074	0.945	3.359	-0.040	0.092
Top 50	Unadjusted cor	-0.0003	0.0049	0.023	-0.0001	0.028	1.918	-0.038	0.043
Top 50	Model cor	-0.0011	0.0052	0.025	-0.0054	0.226	1.935	-0.036	0.044
Top 50	DTW	0.0084	0.0070	0.034	0.0032	0.512	2.999	-0.057	0.084
Top 50	Euclidean	0.0064	0.0072	0.035	-0.0009	1.136	3.505	-0.042	0.091
Top 100	Unadjusted cor	0.0013	0.0058	0.028	0.0004	0.430	2.357	-0.037	0.067
Top 100	Model cor	0.0007	0.0061	0.029	-0.0017	0.412	2.735	-0.051	0.071
Top 100	DTW	0.0083	0.0082	0.039	0.0034	0.897	4.070	-0.062	0.109
Top 100	Euclidean	0.0073	0.0078	0.037	0.0057	0.763	3.619	-0.045	0.107

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	0.0106	0.051	0.0021	0.057	4.071	-0.112	0.132
Top 25	Model cor	0.0034	0.0104	0.050	0.0005	0.434	3.994	-0.093	0.133
Top 25	DTW	0.0189	0.0086*	0.042	0.0096	0.721	3.062	-0.041	0.119
Top 25	Euclidean	0.0067	0.0072	0.035	-0.0075	0.919	3.245	-0.040	0.092
Top 50	Unadjusted cor	0.0015	0.0078	0.038	-0.0006	0.041	1.701	-0.060	0.064
Top 50	Model cor	0.0010	0.0089	0.043	-0.0086	0.343	2.234	-0.065	0.097
Top 50	DTW	0.0084	0.0071	0.034	0.0033	0.457	2.973	-0.060	0.084
Top 50	Euclidean	0.0064	0.0073	0.035	-0.0009	1.116	3.435	-0.042	0.091
Top 100	Unadjusted cor	0.0032	0.0095	0.046	0.0007	0.134	1.556	-0.062	0.073
Top 100	Model cor	0.0020	0.0096	0.046	-0.0022	0.127	1.758	-0.072	0.077
Top 100	DTW	0.0082	0.0082	0.040	0.0035	0.845	4.017	-0.064	0.109
Top 100	Euclidean	0.0073	0.0078	0.037	0.0057	0.740	3.556	-0.046	0.107

III. Baseline Long Position

# Pairs	Pair Strategy	Mean	Std Err[1]	Std Dev	Median	Skew	Kurtosis	Min	Max
Top 25	Unadjusted cor	0.0557	0.0483	0.232	0.1049	-0.234	2.753	-0.463	0.513
Top 25	Model cor	0.0487	0.0490	0.235	0.0815	-0.003	3.256	-0.472	0.597
Top 25	DTW	0.0899	0.0310**	0.149	0.1371	-1.196	4.021	-0.333	0.266
Top 25	Euclidean	0.0970	0.0315**	0.151	0.1418	-0.892	3.847	-0.314	0.364
Top 50	Unadjusted cor	0.0648	0.0472	0.227	0.1161	-0.633	2.837	-0.490	0.413
Top 50	Model cor	0.0596	0.0472	0.227	0.0947	-0.466	2.674	-0.470	0.431
Top 50	DTW	0.0711	0.0311**	0.149	0.1266	-1.079	3.986	-0.339	0.274
Top 50	Euclidean	0.0895	0.0302**	0.145	0.1238	-0.947	3.559	-0.302	0.304
Top 100	Unadjusted cor	0.0700	0.0450	0.216	0.1292	-0.809	3.100	-0.488	0.358
Top 100	Model cor	0.0715	0.0458	0.220	0.1253	-0.648	2.952	-0.472	0.403
Top 100	DTW	0.0875	0.0312**	0.150	0.1335	-1.027	3.882	-0.334	0.288
Top 100	Euclidean	0.0912	0.0307**	0.148	0.1300	-0.943	3.600	-0.312	0.287

Table 1: Average Annual Return Distribution

Three different measures of portfolio returns are summarized in this table. Section I 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 total number of pairs in the portfolio. Section III records the annual return of the portfolio if the trader simply goes long in each pair at the beginning of the trading period.

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

[1] Standard errors are Newey-West estimates and significant values are for normally distributed tails.

	Communication Services	Consumer Discretionary	Consumer Staples	Energy	Financials	Health Care	Industrials	Information Technology	Materials	Real Estate	Utilities	Count	%
Communication Services	17	.	.	.	.	.	.	.	.	.	.	34	3.1
Consumer Discretionary	.	30	.	.	.	.	.	.	.	.	.	60	5.5
Consumer Staples	.	.	.	.	.	.	.	.	.	.	.	.	.
Energy	.	.	.	47	.	.	.	.	.	.	.	94	8.5
Financials	.	.	.	.	147	.	.	.	.	.	.	294	26.7
Health Care	.	.	.	.	.	.	.	.	.	.	.	.	.
Industrials	.	.	.	.	.	.	10	.	.	.	.	20	1.8
Information Technology	.	.	.	.	.	.	.	36	.	.	.	72	6.5
Materials	.	.	.	.	.	.	.	.	5	.	.	12	1.1
Real Estate	.	.	.	.	.	.	.	.	2	161	.	324	29.5
Utilities	.	.	.	.	.	.	.	.	.	.	95	190	17.3

Unadjusted Correlation

	Communication Services	Consumer Discretionary	Consumer Staples	Energy	Financials	Health Care	Industrials	Information Technology	Materials	Real Estate	Utilities	Count	%
Communication Services	17	.	.	.	.	.	.	.	.	.	.	34	3.1
Consumer Discretionary	.	36	.	.	.	.	.	.	.	.	.	72	6.5
Consumer Staples	.	.	.	.	.	.	.	.	.	.	.	.	.
Energy	.	.	.	52	.	.	.	.	.	.	.	104	9.5
Financials	.	.	.	.	117	.	.	.	.	.	.	234	21.3
Health Care	.	.	.	.	.	1	.	.	.	.	.	2	0.2
Industrials	.	.	.	.	.	.	13	.	.	.	.	26	2.4
Information Technology	.	.	.	.	.	.	.	44	.	.	.	88	8.0
Materials	.	.	.	.	.	.	.	.	10	.	.	23	2.1
Real Estate	.	.	.	.	.	.	.	.	3	150	.	303	27.5
Utilities	.	.	.	.	.	.	.	.	.	.	107	214	19.5

Model Correlation

Table 2: Sector Distribution of Portfolio Pairs: Top 25 by Correlation

This table tabulates the sector distribution of all pairs in the annual portfolios constructed from 2011 to 2022. Two views are presented. On the left is the cross sector count of *pairs*. On the right is the (marginal) sector count for all individual *stocks* in the set of portfolios. Intra-industry pairs are the rule using either correlation based pairing strategy. The only few exceptions that are pairs between Real Estate and Materials stocks. Taken together the Financial, Real Estate, and Utility sectors account for 73.5 and 68.3 percent out of all stocks chosen in the unadjusted and modeled portfolios, respectively.

	Communication Services	Consumer Discretionary	Consumer Staples	Energy	Financials	Health Care	Industrials	Information Technology	Materials	Real Estate	Utilities	Count	%
Communication Services	14	.	.	.	.	.	.	.	.	.	.	44	4.0
Consumer Discretionary	.	3	.	.	.	.	.	.	.	.	.	24	2.2
Consumer Staples	2	3	13	.	.	.	.	.	.	.	.	92	8.4
Energy	.	.	1	5	.	.	.	.	.	.	.	21	1.9
Financials	2	2	13	.	50	.	.	.	.	.	.	163	14.8
Health Care	3	.	12	1	12	6	.	.	.	.	.	59	5.4
Industrials	3	4	7	3	8	6	20	.	.	.	.	91	8.3
Information Technology	2	1	1	2	9	3	6	2	.	.	.	31	2.8
Materials	.	3	1	2	4	1	6	2	2	.	.	28	2.6
Real Estate	.	1	1	.	2	.	2	.	2	49	.	115	10.5
Utilities	4	4	25	2	11	9	6	1	3	9	178	430	39.2

Dynamic Time Warping

Communication Services	14	.	.	.	.	.	.	.	.	.	.	38	3.5
Consumer Discretionary	.	.	.	.	.	.	.	.	.	.	.	8	0.7
Consumer Staples	1	3	19	.	.	.	.	.	.	.	.	86	7.8
Energy	.	.	.	7	.	.	.	.	.	.	.	18	1.6
Financials	2	1	13	.	71	.	.	.	.	.	.	191	17.4
Health Care	1	.	8	1	7	4	.	.	.	.	.	34	3.1
Industrials	1	1	2	.	10	3	17	.	.	.	.	70	6.4
Information Technology	1	.	1	2	3	1	3	2	.	.	.	18	1.6
Materials	.	1	.	1	3	.	9	2	3	.	.	26	2.4
Real Estate	.	1	.	.	2	.	2	.	1	72	.	156	14.2
Utilities	4	1	20	.	8	5	5	1	3	6	200	453	41.3

Euclidean Distance

Table 3: Sector Distribution of Portfolio Pairs: Top 25 by L^2 Norm

This table tabulates the sector distribution of all pairs in the annual portfolios constructed from 2011 to 2022. Two views are presented. On the left is the cross sector count of *pairs*. On the right is the (marginal) sector count for all individual *stocks* in the set of portfolios. Diversification away from intra-sector pairs is the main differentiator between the portfolios summarized here versus those using correlation as captured in table 2. Taken together the Financial, Real Estate, and Utility sectors account for 64.8 and 72.9 percent out of all stocks chosen in the time-warped and euclidean portfolios, respectively.

Time-Series Data Mining 12:21

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 [8]

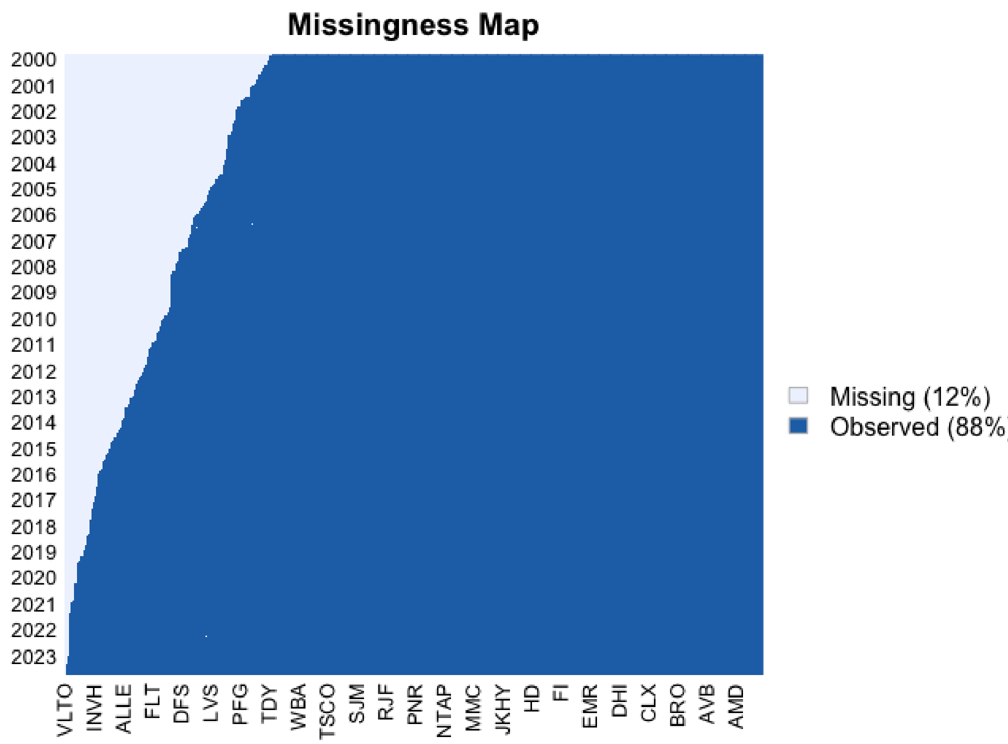

Figure 2: This image captures the available history, by stock, of the S&P data used in this essay. Light blue shading indicates timestamps where price data is not available. Dark blue represents available history. Of the 503 stocks in the dataset, there are 349 with a full series of prices throughout the time period undner investigation.

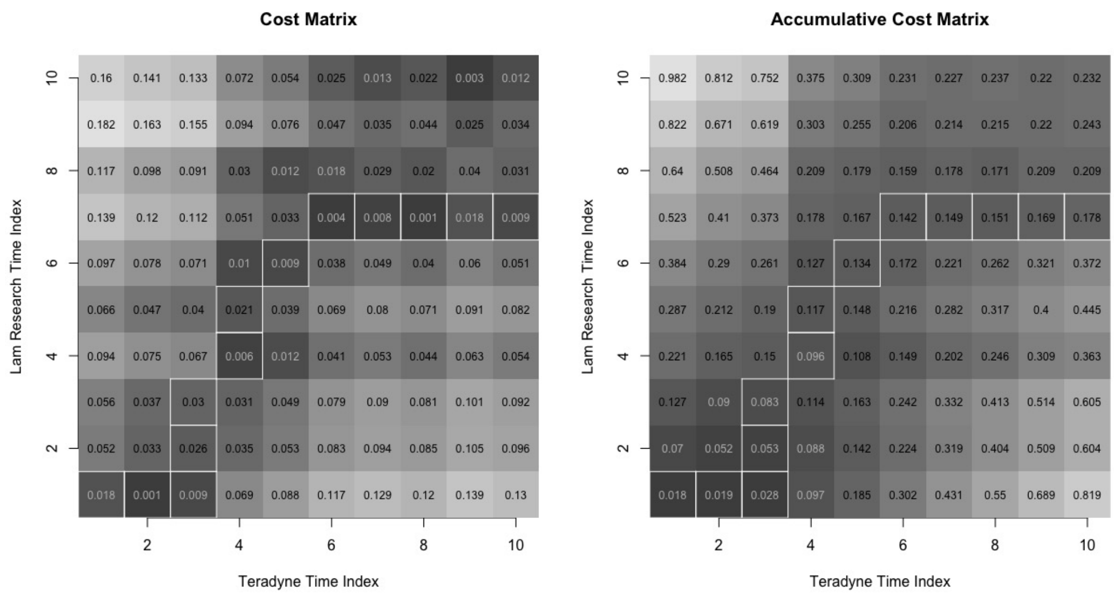

Figure 3: The left image is the cost matrix for the first ten values of the Teradyne and Lam Research stock prices after standardization. The local cost function is the euclidean distance. The right image is the accumulative cost matrix for the same sequence of prices. The *optimal warping path* is annotated 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.

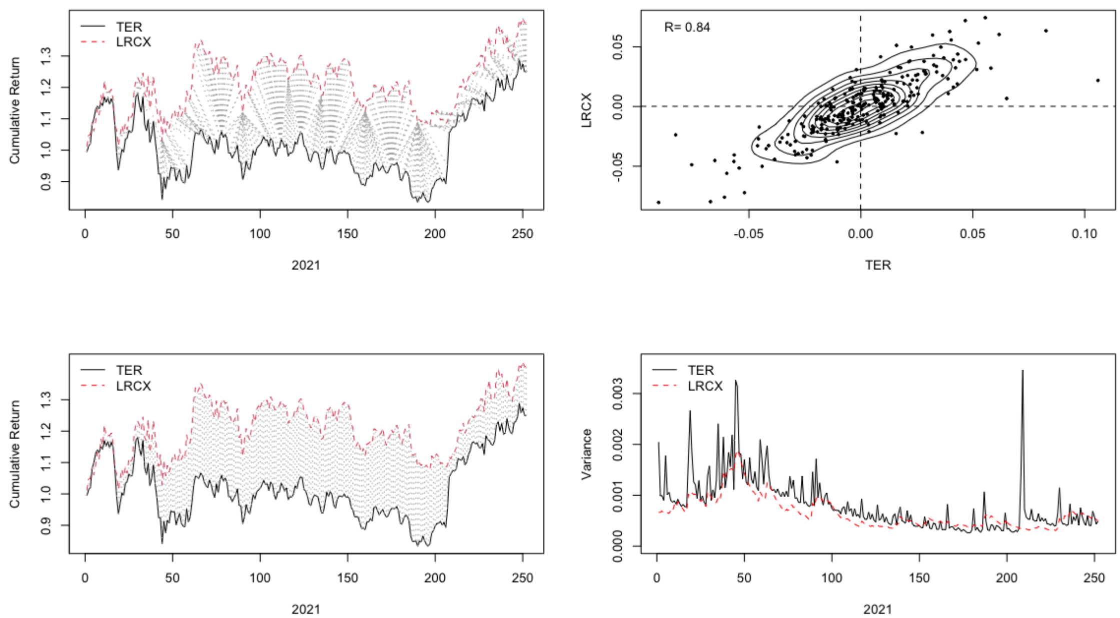

Figure 4: These figures record an unusual case of a stock pair with very strong positive correlation *and* a large dynamic time warping distance. All figures summarize data on Teradyne Inc. (TER) and Lam Research Corp. (LRCX) stock for the 2021 calendar year. **Top-Left** A graph showing the cumulative return of Teradyne Inc. and Lam Research Corp. stock with line segements connecting the indicies from the optimal warping path. The algorithm is run on the pair's cumulative return series (see equation 18). A 15-day Sakoe-Chiba window provides a global constraint. **Bottom-Left** The same cumulative price chart as the top-left but with Euclidean line segments. The line segments are vertical because there is no time dilation. The euclidean distance between the pair's normalized prices is 3.345. **Top-Right** A scatter-plot of unmodeled log-returns. The sample pearson correlation is 0.84. **Bottom-Right**. The fitted conditional variance from each series' esitmated ARMA-GARCH model.

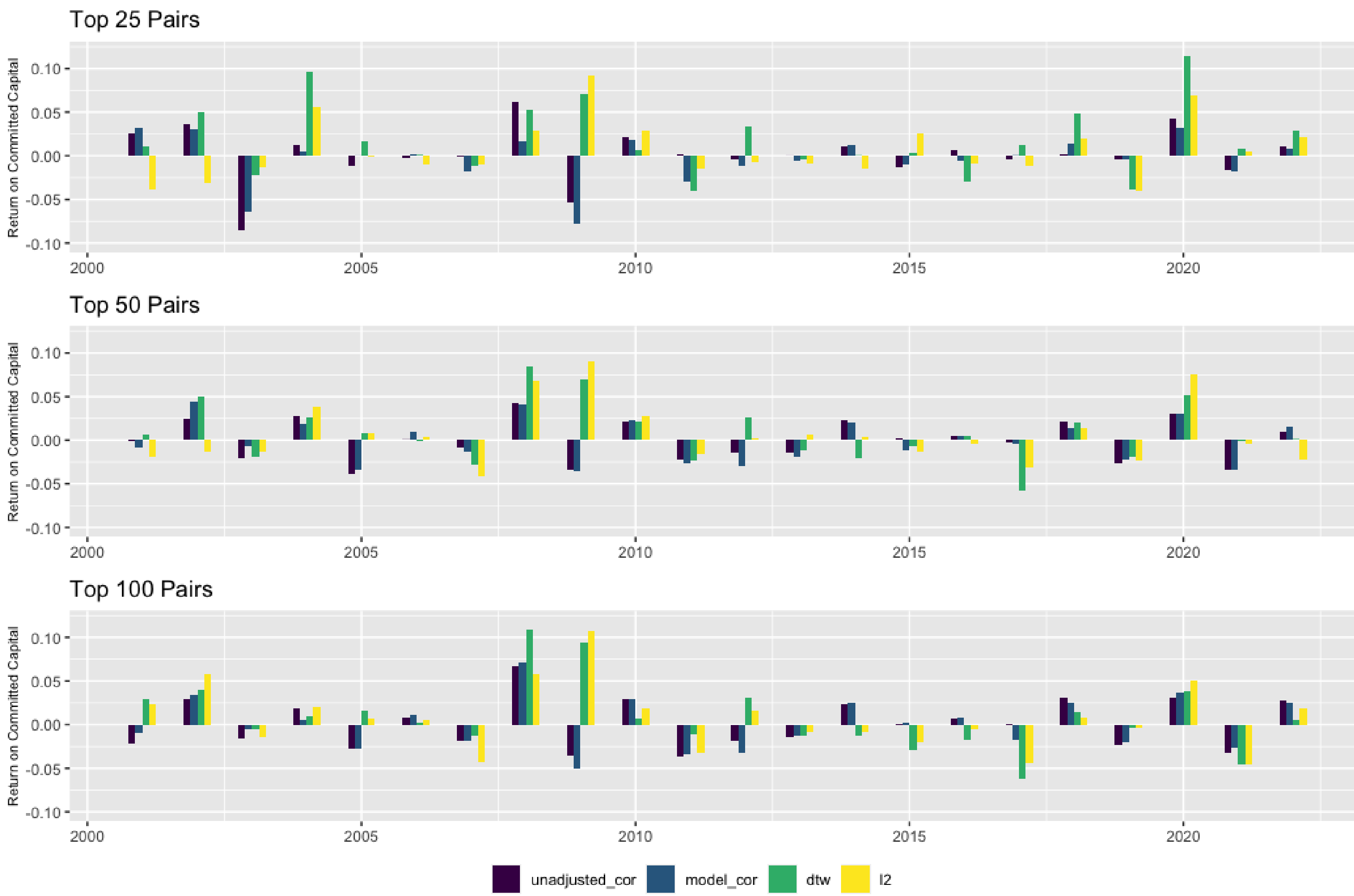

Figure 5: Barcharts summarizing the annual committed returns of the trading strategy under different buy signals and portfolio sizes.

6 Testing for a Relationship Between Correlation and DTW

At the time of writing the authors have found no existing theory on the convergence of dynamic time warping when applied to time series that exhibit brownian motion, random walk, or other non-stationary behavior. This may be due to the fact that dynamic time warping is not a valid metric. Although DTW takes non-negative values and is symmetric[7], there is no gurantee that it satisfies the triangle inequality[8].

In order to gauge the relationship between the bivariate DGP process of two financial time series and their resulting DTW value this section pursues two separate approaches. First, the relationship is investigated via simulation. Two sets of simulation exercises are carried out. The first simulation uses iid normal variates to produce time series of white noise, random walks without drift, and random walks with drift. A second, more sophisticated, simulation exercise uses a monte carlo approach to inference by combining the optimized ARMA-GARCH models from section 5 with the copula approach to modeling joint distributions.

In addition to the simulation studies, a second analysis uses regression to study the mean relationship between correlation and dynamic time warping on a sample of stock pairs created during the pair trading strategy of section 5. The sample used for the regressions is the same sample used in the monte carlo exercises. Taken together the analysis in this section provide the contours of the realtionship between Pearson's correlation and dynamic time warping when estimated on a set of stock pairs.

6.1 Uncorrelated White Noise and Random Walks

For the first simulation a stationary process is developed with only two factors: an intercept and an error term. The errors are independent and identical draws from the normal distribution. Equation 19 below summarizes the process.

$$X_t = \omega + \epsilon_t \qquad \epsilon_t \sim \mathcal{N}(0, \sigma^2) \tag{19}$$

[7]Symmetry is not always guranteed. This article uses a symmetric cost function and a symmetric step-condition (see equation 9 and condition 11 in section 4.3). See one of the early authoritative works [34] for a detailed discussion on global constraints and step-conditions.

[8]$d(x, z) \leq d(x, y) + d(y, z)$

After fixing parameter values (ω_A, σ_A, ω_B, σ_B), synthetic stationary series of arbitrary length can be produced by sampling from the normal distribution. By sampling a series from two separate specifications, A and B, and then calculating the dynamic time-warping distance between the two, a connection between the parameters and the estimated DTW metric can be established. By repeating this process N times a distribution of DTW values can be built up.

A similar apprach can be used to capture the distribution of the dynamic time warping distance for two random walk processes. Instead of the stationary process defined in equation 19, two series can be constructed as a random walk that obey the following specification:

$$X_t = \omega + X_{t-1} + \epsilon_t \qquad \epsilon_t \sim \mathcal{N}(0, \sigma^2) \tag{20}$$

In equation 20 the intercept term, ω, captures the underlying drift of the random walk. By setting ω equal to zero a random walk processes absent of any drift can be simulated. If ω is set to a non-zero value this results in a random walk process with drift. Using equations 19 and 20 three different kinds of simulation exercises are conducted: white noise, random walk without drift, and random walk with drift. The length of the simulated series in this section is set to 252 to match the number of trading days in an average year. Choosing this length aligns these simulations with the annual formation and trading periods used in section 5 to backtest the pair-trading strategy. To further align the proceedure used in section 5 to these simulation exercises, this essay compares the results of unconstrained dynamic time warping to dynamic time warping using a 15-day Sakoe-Chiba global constraint. Results are contained in tables 4, 5, and 6.

The results of dynamic time warping applied to a pair of stationary white noise series offers a natural starting point. Looking at the unconstrained section of table 4 a few items are worth noting. First, dynamic time warping values have a positive realtionship with the variance of the two series. As the variance increases so too does the range and central location of the DTW distributions. Contrasting the second section (constant $\sigma_A = 1$ with increasing σ_B) with the third section (jointly increasing σ_A and σ_B), the total variance of the two series matters less than the max value of the two variances. Second, dynamic time warping values increase as the intercept terms diverge in size (see the fourth section). This divergance in intercept between the two series has a greater impact on the size of the DTW values than increasing the the variance of one or both series. Third, the distribution of DTW values appears symmetric. The mean and median are close in value and when tested for normality using the Anderson-Darling test, only the simulation with $\omega_A = 16$ and $\omega_B = 0$ rejects the null of normality at the 0.05 significance level. Fourth, the constrained dynamic time warping values are slightly larger, at the mean, than their unconstrained analogue. This is unsurprising given the distance minimizing principle of the algorithm. Any constraints on the amount of time warping can only increase the value of the distance measure. Interesting enough, the Sakoe-Chiba constraints produces larger IQRs than the unconstrained version for the first three groups of simulations but acts in a constraining fashion when the intercept terms diverge and σ_A is greater than two. See the last section of table 4.

When moving from a white-noise process to a random walk process the resulting dynamic time warping values are not only materially larger, the range of values explodes. As recorded in unconstrained part of table 5, the benchmark case ($\sigma_A = \sigma_B = 1$) records a mean DTW value of 10.93, compared to 0.84 for its white-noise equivalent, and an interquartile range of 11.49, compared to 0.04. The distributions are no longer symmetric and all specifications reject the null hypothesis of a normal distribution when subjected to Anderson-Darling tests for normality. The disptributions have long tails extending into larger values of the distance measure. Counterintuitively, the random walk process with drift results in smaller dynamic time warping values than the random walk without drift.This can be seen by comparing the same specification in the first three sections of tables 5 and 6. Having a shared drift paramter appears to keep the two simulated series in a closer relative neighborhood than two random walk series with no drift.

6.2 Monte Carlo Simulation based on the ARMA-GARCH Framework

One characteristic the white noise and random walk specifications in the preceeding section both share is independence between the two series. The error terms are drawn independently from each other and from all previous draws:

$$\mathbb{E}[\epsilon_{A,t}\epsilon_{B,s}] = 0 \qquad \forall t, s \tag{21}$$

$$\mathbb{E}[\epsilon_{A,t}\epsilon_{A,s}] = 0 \qquad \forall t, s \tag{22}$$

This restriction is convienent for simulation but does not accurately represent returns to many financial assets, especially stock returns which tend to be positively correlated with each other. In this section a suitable monte carlo framework is outlined to test if a dynamic time warping value observed between two stock prices is statistically rare for the observed correlation between the pair's returns. The results show that taking into account the time-varying mean and variance is necessary for proper inference.

In section 5 a total of 10,018 ARMA-GARCH models were estimated during the formation periods from 2000 to 2021 where each model was trained on a single year's worth of trading data. From these ten thousand marginal models a total of 2,199,276 pairs were formed and estimates of the pair's correlation and dynamic time warping distance were calculated over the duration of the study. These observed stock returns and the models optimized on top of them provide a diverse sample to study what, if any, relationship exists between correlation and dynamic time warping when estimtated on the same pair of stock price returns. Common sense would suggest that dynamic time warping values should be at their lowest when the correlation of the stock returns of the pair are at their highest, implying a negative relationship between the two. In the special case where the log price returns are perfectly correlated the dynamic time warping value should be equal to zero. While this section does lay out evidence comfirming a negative relationship between the two we find that the variance of DTW values is significance at all levels of correlation.

Instead of analysing the entire 2.2 million pairs from section 5 a smaller sample is created using stratified sampling. The idea is to stratify across the correlation parameters observed correlation parameter. The range of correlation values is segmented into adjacent bins with a uniform width set to 0.05, e.g: [-0.65, -0.6), ..., [0.9, 0.95), [0.95, 1). For each bin 300 hundred stock pairs or randomly sampled. If a bin has less than 300 pairs then all pairs in that bin are included. In total there are 7,765 pairs selected for this simulation study. The optimized marginal ARMA-GARCH models for these pairs, combined with Copula theory, can be leveraged to study the range of dynamic time warping values that can be expected for a pair of stocks with a specific level of correlation (See section 2 of [5] for more details on copulas and their applications). The simulation exercise has the following steps.

1. For two arbitrary stocks, A and B, the conditional mean, conditionl variance, and conditional density of the returns are recorded for each asset: $\hat{\mu}_{A,t}$, $\hat{\sigma}^2_{A,t}$, $\hat{g}_A$, and $\hat{\mu}_{B,t}$, $\hat{\sigma}^2_{B,t}$, $\hat{g}_B$ (see equations 29, 30, and 32 from section A).

2. The probability integral transform is used to reformulate each marginal model's

fitted residuals as a sample from the uniform distribution.

$$\hat{U}_A = \hat{F}_A(s) = \int_{-\infty}^{s} \frac{1}{\hat{\sigma}_{A,\,t}} \hat{g}_A\left(z|\hat{\mu}_t\right) dz \tag{23}$$

$$\hat{U}_B = \hat{F}_B(s) = \int_{-\infty}^{s} \frac{1}{\hat{\sigma}_{B,\,t}} \hat{g}_B\left(z|\hat{\mu}_t\right) dz \tag{24}$$

The residuals, in their uniform representation $[(\hat{u}_{A,\,1},\, \hat{u}_{B,\,1}),\, \cdots,\, (\hat{u}_{A,\,T},\, \hat{u}_{B,\,T})]$, are used to estimate a bivariate t-Copula which is defined by it's correlation $\rho_{A,B}$ and degrees-of-freedom $\nu_{A,B}$ parameters.

3. The estimated t-Copula is used as a data generating process to produce new sets of bivariate uniform distributions $[(u^{(i)}_{A,\,1},\, u^{(i)}_{B,\,1}),\, \cdots,\, (u^{(i)}_{A,\,T},\, u^{(i)}_{B,\,T})]$. Each observation t in a new sample is centered and scaled according to the conditional mean and variance from the estimated ARIMA-GARCH models[9].

$$r^{(i)}_{A,\,t} = \hat{F}_A^{-1}\left(u^{(i)}_{A,\,t}\ ;\ \mu = \hat{\mu}_{A,\,t},\, \sigma^2 = \hat{\sigma}^2_{A,\,t}\right) \tag{25}$$

$$r^{(i)}_{B,\,t} = \hat{F}_B^{-1}\left(u^{(i)}_{B,\,t}\ ;\ \mu = \hat{\mu}_{B,\,t},\, \sigma^2 = \hat{\sigma}^2_{B,\,t}\right) \tag{26}$$

4. After a new set of return series are generated they are transformed from returns to price level using the standard price equation defined in 18.

5. Repeat steps one through four to build up a distribution of dynamic time warping values for a pair of stock prices with a specific correlation value

By marrying the conditional return models with a copula-based resampling approach many of the important characteristics of stock returns can be included in the monte carlo samples. These include auto-correlation, asymmetric returns, fat-tails, and volatility clustering observed in individual stocks as well as high levels of correlation in price movements between stocks over time. The original (Pearson's) correaltion and tail-dependency observed in the stock pairs is preserved by the copula while the unique idiosyncracies (See section A) of each stock is retained by the marginal models. This leaves only the variance of the distributions (g_A, g_B) to drive the variation in the dynamic time warping values calculated in step 4. Assuming multivariate t-distributions for this simulation exercise is well founded. In the model selection

[9]All marginal return models estimated in section 5 have a time-varying variance equation (see equation 30) but not every return series has a time-varying mean component (see equation 29). The automated modeling process allows for a constant term only for the mean equation. This can impact the dynamics of the sampling process in equations 25 and 26: $\mu = \hat{\mu}_i$ for $i \in A, B$.

process described in section A the t-distribution is chosen for all 10,018 marginal models in this study. The symmetric t-distribution is chosen 9,324 times while the remaining 694 models are modeled with a skewed t-distribution.

In the context of inference between correlation and dynamic time warping, the importance of capturing a stocks time-varying dynamics via the ARMA-GARCH model is assessed by running four different simulation exercises that alternate between using conditional and unconditional mean and variance specifications of equations 25 and 24. For all ARMA-GARCH specifications used in this essay, the unconditional mean and unconditional variance of a stock's return series can be estimated by the the fitted conditional models[10].

This essay finds evidence that modeling the conditional mean, and especially the conditional variance, is materially important if the researcher wants to perform bootstrapped confidence intervals. The values in the table 7 represent the proportion of the sampled stock pairs from section 5 whose observed dynamic warping value falls outside the confidence intervals constructed using the Monte Carlo simulation approach described in this section. The sampled pairs are split into two categories based on the ARMA specification choosen for stock returns modeled in section 5. On the left is a set of 5,087 pairs where one or more of the stocks in the pair were chosen to have an autoregressive or moving average component in their mean equation (see equation 29). The set of columns on the right are for pairs where both stocks were modeled with intercept terms only. In the latter case the conditional mean and the unconditional mean of the series are the same. Rows correspond to different model assumptions when simulating new pairs of return series described in equations 25 and 26. It is clear that not controlling for time-varying variance (i.e. volatility clustering) can lead to improper inference. The simulations that use unconditional variances in equations 25 and 26 (first and second rows) create confidence intervals that are far too narrow. For pairs with at least one stock fitted with an AR or MA component (*Time-Varying Mean Pair*), the proportion of pairs whose observed dynamic time warping value falls outside the monte carlo confidence intervals far exceeds the expected significance level. This bias is reduce somewhat for pairs where both stocks are estimated with a constant mean process (*Constant Mean Pair*) but the bias remains significant. When adding the fitted conditional variance values to the simulation proceedure (third and fourth rows) the observed bias narrows considerably, resulting in confidence intervals where the proportion of observed dynamic time warping values falling outside of them aligns closely given

[10] The vignette [14] summarizing the software implementation used in this essay provides the relationship between the conditional mean and variance equations with their unconditional analogues.

the significance level.

6.3 Regression Analysis

In this section regression analysis is used to identify the important features that can help explain the relationship between a pair of stock return's observed correlation and the dynamic time warping value of their standard prices. The regression sample is the same set of pairs that were created for the simulations in section 6.2: 7,765 stock pairs formed during the annual formation periods while executing the pair-trading strategy. The results summarized in tables 4, 5, and 6 help inform this sections choice of explanatory variables.

Each pair in the sample has an observed correlation and dynamic time warping value.

$$\log(\widehat{dtw}) = \beta_0 + \beta_1\hat{\rho} + \beta_2\hat{\rho}^2 + \boldsymbol{\beta}_3\boldsymbol{X} + \epsilon \tag{27}$$

A ARMA-GARCH Modeling

Discuss 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 [5]. 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.

The conditional mean for the log-returns can be formulated by the following ARMA process:

$$x_t = \mu_t + \epsilon_t \tag{28}$$

$$\mu_t = \mu(\phi, \theta, x_{\{s:\, s<t\}}, \varepsilon_{\{s:\, s<t\}}) = \phi_0 + \sum_{i=1}^{p} \phi_i x_{t-i} + \sum_{j=1}^{q} \theta_j\, \varepsilon_{t-j} + \varepsilon_t \tag{29}$$

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.

$$\sigma_t^2 = C(\varepsilon_{t-1}^2,\, \sigma_{t-1}^2) + \sum_{j=1}^{m} \alpha_j A(\varepsilon_{t-j}^2) + \sum_{i=1}^{k} \beta_i B(\sigma_{t-i}^2) \tag{30}$$

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

$$z_t(\phi, \theta, \alpha, \beta) = \frac{x_t - \mu(x_{t-1}, \phi, \theta)}{\sigma(x_{t-1}, \alpha, \beta)} \tag{31}$$

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:

$$f(x_t|\mu_t, \sigma_t^2, \lambda, \Delta) = \frac{1}{\sigma_t} g(z_t|\lambda, \Delta) \tag{32}$$

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 [18], 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 [19] running in the R statistical language [37]. 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 [14]. 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 [9]}
   - Model Specification: A, B, C = {Standard GARCH [3], gjr-GARCH [15], Component GARCH [26] }

   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 auto-correlation in the first four moments, (b) Kolmogorov-Smirnov test to check the residuals against the chosen theoretical distribution, (c) Hong and Li [16] non-parametric density test jointly for i.i.d and correct distribution specification, (d) Shapiro-Wilks [35] test for normality, and (e) Jarque-Bera [22] 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.

References

[1] Matteo Aquilina, Eric Budish, and Peter O'Neill. "Quantifying the High-Frequency Trading Arms Race". In: *The Quarterly Journal of Economics* 137.1 (Feb. 2022), 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.

[2] Lu Bai et al. "Entropic Dynamic Time Warping Kernels for Co-Evolving Financial Time Series Analysis". In: *IEEE Transactions on Neural Networks and Learning Systems* 34.4 (2023), pp. 1808–1822. DOI: 10.1109/TNNLS.2020.3006738.

[3] Tim Bollerslev. “Generalized autoregressive conditional heteroskedasticity”. In: *Journal of Econometrics* 31.3 (1986), pp. 307–327. ISSN: 0304-4076. DOI: https://doi.org/10.1016/0304-4076(86)90063-1.

[4] Michel E.D. Chaves et al. “Time-weighted dynamic time warping analysis for mapping interannual cropping practices changes in large-scale agro-industrial farms in Brazilian Cerrado”. In: *Science of Remote Sensing* 3 (2021), p. 100021. ISSN: 2666-0172. DOI: https://doi.org/10.1016/j.srs.2021.100021.

[5] L. C. Dowiak. “A Comparison of Time-Varying Copula-GARCH Models”. In: *Unpublished* 00 (2018), p. 00.

[6] Pierpaolo D’Urso, Livia De Giovanni, and Riccardo Massari. “Trimmed fuzzy clustering of financial time series based on dynamic time warping”. In: *Annals of Operations Research* 299.1 (), pp. 1379–1395. DOI: 10.1007/s10479-019-03284-1.

[7] Vo Nguyen Le Duy and Ichiro Takeuchi. *Statistical Inference for the Dynamic Time Warping Distance, with Application to Abnormal Time-Series Detection*. 2023. arXiv: 2202.06593 [stat.ML].

[8] P. Esling and C. Agon. “Time-Series Data Mining”. In: *ACM Comput. Surv.* 45 (2012), pp. 79–82.

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

[10] P. H. Franses and T. Wiemann. “Intertemporal Similarity of Economic Time Series: An Application of Dynamic Time Warping”. In: *Computational Economics* 56 (2020), pp. 59–75.

[11] Sylvia Frühwirth-Schnatter and Sylvia Kaufmann. “Model-Based Clustering of Multiple Time Series”. In: *Journal of Business and Economic Statistics* 26 (Oct. 2004). DOI: 10.1198/073500107000000106.

[12] Evan Gatev, William N. Goetzmann, and K. Geert Rouwenhorst. “Pairs Trading: Performance of a Relative-Value Arbitrage Rule”. In: *Review of Financial Studies* 19.3 (Feb. 2006), 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.

[13] Zoltan Geler et al. “Dynamic time warping: Itakura vs sakoe-chiba”. In: *2019 IEEE International Symposium on INnovations in Intelligent SysTems and Applications (INISTA)*. IEEE. 2019, pp. 1–6.

[14] Alexios Ghalanos. *rugarch: Univariate GARCH models.* R package version 1.4-4. 2020.

[15] Lawrence R Glosten, Ravi Jagannathan, and David E Runkle. "On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks". In: *Journal of Finance* 48.5 (1993), pp. 1779–1801.

[16] Yongmiao Hong and Haitao Li. "Nonparametric Specification Testing for Continuous-Time Models with Applications to Term Structure of Interest Rates". In: *The Review of Financial Studies* 18.1 (2005), pp. 37–84. ISSN: 08939454, 14657368.

[17] Clint Howard, Tālis Putniņš, and Vitali Alexeev. "To lead or to lag? Measuring asynchronicity in financial time-series using dynamic time warping". Oct. 2020.

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

[19] Rob Hyndman et al. *forecast: Forecasting functions for time series and linear models*. R package version 8.13. 2020.

[20] F. Itakura. "Minimum prediction residual principle applied to speech recognition". In: *IEEE Transactions on Acoustics, Speech, and Signal Processing* 23.1 (1975), pp. 67–72. DOI: 10.1109/TASSP.1975.1162641.

[21] Katsuya Ito and Ryuta Sakemoto. "Direct Estimation of Lead–Lag Relationships Using Multinomial Dynamic Time Warping". In: *Asia-Pacific Financial Markets* 27.3 (2020), pp. 325–342. DOI: 10.1007/s10690-019-09295-z.

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

[23] Young-Seon Jeong, Myong K. Jeong, and Olufemi A. Omitaomu. "Weighted dynamic time warping for time series classification". In: *Pattern Recognition* 44.9 (2011). Computer Analysis of Images and Patterns, pp. 2231–2240. ISSN: 0031-3203. DOI: https://doi.org/10.1016/j.patcog.2010.09.022.

[24] Eamonn Keogh and Chotirat Ann Ratanamahatana. "Exact indexing of dynamic time warping". In: *Knowledge and Information Systems* 7.3 (2005), pp. 358–386. DOI: 10.1007/s10115-004-0154-9.

[25] Alexios Kotsifakos et al. "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, 2011. ISBN: 9781450307727. DOI: 10.1145/2141622.2141666.

[26] Gary G. J. Lee and R. Engle. "A Permanent and Transitory Component Model of Stock Return Volatility". In: 1993.

[27] Harry Markowitz. “Portfolio Selection”. In: *The Journal of Finance* 7.1 (1952), pp. 77–91. ISSN: 00221082, 15406261.

[28] Victor Maus et al. “A Time-Weighted Dynamic Time Warping Method for Land-Use and Land-Cover Mapping”. In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 9.8 (2016), pp. 3729–3739. DOI: 10.1109/JSTARS.2016.2517118.

[29] Meinard Müller. “Information Retrieval for Music and Motion”. In: Springer, Berlin, Heidelberg, 2007. Chap. 4, pp. 69–84.

[30] O. G. Narin et al. “Using RVI and NDVI Time Series for Cropland Mapping with Time-Weighted Dynamic Time Warping”. eng. In: *International archives of the photogrammetry, remote sensing and spatial information sciences.* XLVIII-4/W3-2022 (2022), pp. 97–101. ISSN: 2194-9034.

[31] François Petitjean, Alain Ketterlin, and Pierre Gançarski. “A global averaging method for dynamic time warping, with applications to clustering”. In: *Pattern Recognition* 44.3 (2011), pp. 678–693. ISSN: 0031-3203. DOI: https://doi.org/10.1016/j.patcog.2010.09.013.

[32] François Petitjean et al. “Faster and more accurate classification of time series by exploiting a novel dynamic time warping averaging algorithm”. In: *Knowledge and Information Systems* 47.1 (), pp. 1–26. DOI: 10.1007/s10115-015-0878-8.

[33] Xuzhou Qu et al. “Monitoring maize lodging severity based on multi-temporal Sentinel-1 images using Time-weighted Dynamic time Warping”. eng. In: *Computers and electronics in agriculture* 215 (2023), pp. 108365–. ISSN: 0168-1699.

[34] H. Sakoe and S. Chiba. “Dynamic programming algorithm optimization for spoken word recognition”. In: *IEEE Transactions on Acoustics, Speech, and Signal Processing* 26.1 (1978), pp. 43–49. DOI: 10.1109/TASSP.1978.1163055.

[35] S. S. Shapiro and M. B. Wilk. “An analysis of variance test for normality (complete samples)”. In: *Biometrika* 52.3-4 (Dec. 1965), 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.

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

[37] R Core Team. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing. Vienna, Austria, 2021.

[38] Yuqing Wan and Yain-Whar Si. “A formal approach to chart patterns classification in financial time series”. In: *Information Sciences* 411 (2017), pp. 151–175. ISSN: 0020-0255. DOI: https://doi.org/10.1016/j.ins.2017.05.028.

[39] Gang-Jin Wang et al. "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 (2012), pp. 4136–4146. ISSN: 0378-4371. DOI: https://doi.org/10.1016/j.physa.2012.03.036.

[40] Xingyuan Xiao et al. "Limited-Samples-Based Crop Classification Using a Time-Weighted Dynamic Time Warping Method, Sentinel-1 Imagery, and Google Earth Engine". eng. In: *Remote sensing (Basel, Switzerland)* 15.4 (2023), pp. 1112–. ISSN: 2072-4292.

Unconstrained DTW

ω_A	ω_B	σ_A^2	σ_B^2	Min	1st Qu.	Median	Mean	3rd Qu.	Max	IQR
0	0	1	1	0.76	0.82	0.84	0.84	0.86	0.94	0.04
0	0	1	2	0.92	1.03	1.06	1.07	1.10	1.25	0.06
0	0	1	4	1.29	1.44	1.49	1.49	1.55	1.78	0.11
0	0	1	8	1.82	2.08	2.16	2.16	2.24	2.55	0.16
0	0	1	16	2.60	3.04	3.14	3.15	3.27	3.74	0.23
0	0	2	2	1.03	1.15	1.18	1.18	1.21	1.35	0.06
0	0	4	4	1.49	1.63	1.68	1.68	1.72	1.88	0.10
0	0	8	8	2.07	2.30	2.37	2.37	2.44	2.71	0.13
0	0	16	16	2.97	3.26	3.36	3.36	3.44	3.74	0.18
2	0	1	1	1.26	1.39	1.44	1.44	1.47	1.67	0.08
4	0	1	1	1.95	2.90	3.14	3.14	3.39	4.21	0.49
8	0	1	1	9.43	10.61	10.93	10.90	11.22	12.20	0.61
16	0	1	1	25.10	26.58	26.92	26.88	27.22	28.30	0.64

Constrained DTW with a 15-day Sakoe-Chiba Window

ω_A	ω_B	σ_A^2	σ_B^2	Min	1st Qu.	Median	Mean	3rd Qu.	Max	IQR
0	0	1	1	0.76	0.83	0.85	0.86	0.88	0.98	0.05
0	0	1	2	0.94	1.05	1.08	1.09	1.12	1.27	0.07
0	0	1	4	1.30	1.46	1.51	1.52	1.57	1.83	0.11
0	0	1	8	1.85	2.11	2.19	2.20	2.28	2.57	0.16
0	0	1	16	2.63	3.09	3.20	3.20	3.32	3.76	0.23
0	0	2	2	1.03	1.18	1.21	1.21	1.24	1.38	0.07
0	0	4	4	1.50	1.66	1.71	1.71	1.76	1.94	0.10
0	0	8	8	2.07	2.35	2.42	2.42	2.49	2.79	0.14
0	0	16	16	3.06	3.33	3.43	3.43	3.52	3.87	0.20
2	0	1	1	1.36	1.53	1.59	1.59	1.65	1.96	0.11
4	0	1	1	3.53	4.22	4.39	4.39	4.56	5.19	0.34
8	0	1	1	11.42	12.14	12.31	12.31	12.49	13.21	0.34
16	0	1	1	27.49	28.11	28.29	28.28	28.45	29.05	0.35

Table 4: Stationary White Noise

This table summarizes the distribution of DTW values estimated on pairs of stationary white noise processes. The length of the simulated series is set to 252. A total of 1000 simulations are run for each set of parameter values. The DTW values in table are divided by the series length.
Note: The distributions of the DTW samples are symmetric. The median and mean remain close in value over all configurations. With one exception, the simulated groups fail to reject the null hypothesis of a normal distribution when subjected to Anderson-Darling test. The sole exception is the last simulated group with $\omega_A = 16$.
Note: For stationary series the deterministic component (distance between the mean values) plays the definitive role in the location of the center of the DTW distribution. This factor outweighs the role variance plays in the location of the DTW distribution. A pair with one relatively large variance has a similar DTW distribution than the DTW distribution where both series have large variances.
Note: All other factors equal, constrained DTW leads to slighly larger values of DTW than their unconstrained analogue. Another consequence of using constrained DTW is the fact that it restricts the dispersion (as measured by the inter-quartile range) of the DTW distributions as the absolute difference between ω_A and ω_B grows. This can seen by comparing the IQRs the last section of each table.

Unconstrained DTW

ω_A	ω_B	σ_A^2	σ_B^2	Min	1st Qu.	Median	Mean	3rd Qu.	Max	IQR
0	0	1	1	1.15	3.35	6.97	10.93	14.84	71.27	11.49
0	0	1	2	1.41	4.59	9.70	14.91	19.96	108.66	15.36
0	0	1	4	2.27	7.07	14.46	21.91	30.33	105.10	23.26
0	0	1	8	3.15	9.25	18.26	29.15	40.56	217.00	31.31
0	0	1	16	4.48	13.90	28.54	43.11	59.89	242.73	45.99
0	0	2	2	1.26	4.27	8.64	13.22	17.66	101.12	13.39
0	0	4	4	2.00	6.57	12.33	16.99	23.33	77.76	16.76
0	0	8	8	3.13	9.67	16.79	23.33	31.23	141.76	21.57
0	0	16	16	5.06	15.34	26.64	33.73	46.19	131.63	30.84

Constrained DTW with a 15-day Sakoe-Chiba Window

ω_A	ω_B	σ_A^2	σ_B^2	Min	1st Qu.	Median	Mean	3rd Qu.	Max	IQR
0	0	1	1	1.73	8.67	14.68	18.86	25.66	89.12	16.99
0	0	1	2	2.20	11.78	21.16	26.02	34.83	133.52	23.05
0	0	1	4	3.49	17.99	31.58	38.64	52.47	135.20	34.48
0	0	1	8	4.49	23.33	40.22	51.43	69.51	265.81	46.18
0	0	1	16	4.88	33.55	60.11	75.34	102.81	319.32	69.26
0	0	2	2	1.63	10.33	18.44	22.76	31.39	122.94	21.07
0	0	4	4	2.72	14.11	24.49	29.51	41.02	113.10	26.90
0	0	8	8	3.70	19.03	31.46	39.69	52.57	173.66	33.54
0	0	16	16	5.96	26.79	45.56	56.23	77.23	193.73	50.45

Table 5: Random Walk without Drift

This table summarizes the distribution of DTW values estimated on pairs of a random walk process without drift. The length of the simulated series is set to 252. A total of 1000 simulations are run for each set of parameter values. The DTW values in table are divided by the series length.

Note: The symmetry of the DTW distribution found in table 4 is lost when estimated on a pair of series evolving with random walk behavior. All distributions have much larger mean and median values than DTW estimated on white noise processes. The distribution of DTW values in this table are skewed toward higher values of DTW.

Note: All other factors equal, constrained DTW leads to significantly larger values of DTW than their unconstrained analogue. This a natural consequence of the unbounded evolution of the two simulated series.

Unconstrained DTW

ω_A	ω_B	σ_A^2	σ_B^2	Min	1st Qu.	Median	Mean	3rd Qu.	Max	IQR
1	1	1	1	0.89	1.12	1.45	1.92	2.22	11.92	1.10
1	1	1	2	1.10	1.38	1.92	2.71	3.13	15.72	1.74
1	1	1	4	1.45	1.95	2.71	3.90	4.76	23.09	2.81
1	1	1	8	2.08	3.04	4.42	6.73	7.88	66.83	4.84
1	1	1	16	3.08	5.09	7.71	11.96	15.20	100.90	10.12
1	1	2	2	1.23	1.61	2.27	3.29	4.00	21.91	2.38
1	1	4	4	1.76	2.49	3.74	5.60	6.41	47.59	3.92
1	1	8	8	2.53	4.09	6.40	10.49	14.32	65.27	10.23
1	1	16	16	3.73	6.60	11.57	17.91	21.89	129.09	15.29
2	1	1	1	37.67	57.90	64.23	64.48	71.01	93.85	13.11
4	1	1	1	235.92	276.81	286.75	286.26	295.34	342.13	18.53
8	1	1	1	724.32	766.67	778.44	778.24	790.03	828.62	23.36
16	1	1	1	1732.09	1772.40	1783.98	1784.40	1796.35	1841.19	23.95

Constrained DTW with a 15-day Sakoe-Chiba Window

ω_A	ω_B	σ_A^2	σ_B^2	Min	1st Qu.	Median	Mean	3rd Qu.	Max	IQR
1	1	1	1	0.89	1.33	2.86	6.43	8.02	53.37	6.69
1	1	1	2	1.14	2.18	5.56	10.82	13.96	87.11	11.78
1	1	1	4	1.45	4.55	10.93	17.21	24.25	99.93	19.70
1	1	1	8	2.14	9.29	19.63	27.99	39.91	200.91	30.61
1	1	1	16	3.65	17.01	32.24	44.84	61.00	226.39	43.99
1	1	2	2	1.27	3.40	8.82	14.63	20.51	110.96	17.11
1	1	4	4	1.86	7.91	15.96	24.39	32.21	158.55	24.31
1	1	8	8	3.17	16.01	32.29	44.41	63.89	236.21	47.88
1	1	16	16	4.72	25.33	49.39	63.81	86.45	369.86	61.11
2	1	1	1	142.79	192.71	211.14	210.31	225.75	285.64	33.04
4	1	1	1	591.19	670.46	687.05	686.34	703.69	775.79	33.24
8	1	1	1	1545.19	1624.21	1641.39	1640.70	1658.21	1713.89	34.00
16	1	1	1	3460.67	3530.21	3547.09	3547.86	3565.62	3629.28	35.40

Table 6: Random Walk with Drift

This table summarizes the distribution of DTW values estimated on pairs of a random walk process with drift. The length of the simulated series is set to 252. A total of 1000 simulations are run for each set of parameter values. The DTW values in table are divided by the series length.

Note: All other factors equal, constrained DTW leads to significantly larger values of DTW than their unconstrained analogue. This a natural consequence of the unbounded evolution of the two simulated series.

Table 7: Violations of Monte Carlo Confidence Intervals

	Time-Varying Mean Pair				Constant Mean Pair			
	1%	5%	10%	N	1%	5% d	10%	N
Unconditional Model	0.108	0.151	0.190	5,087	0.075	0.114	0.155	2,333
Conditional Mean	0.114	0.161	0.203	5,087	–	–	–	–
Conditional Variance	0.020	0.052	0.087	5,087	0.022	0.055	0.090	2,333
Conditional Mean and Var	0.026	0.061	0.106	5,087	0.023	0.051	0.091	2,333

The values in the table represent the proportion of randomly sampled stock pairs from section 5 whose observed dynamic warping value falls outside the Monte Carlo confidence intervals constructed using the simulation approach described in section 6.2. The left set of columns summarizes stock pairs where one or more of the stocks in the pair were chosen to have an autoregressive or moving average componennt in their mean equation (see equation 29). The set of columns on the right are for pairs where both stocks were modeled with intercept terms only. In this case the conditional mean and the unconditional mean are the same. Rows correspond to different model assumptions when simulating new pairs of return series described in equations 25 and 26.
Note: The number of simulations is set to 250. The small number is due to the high computational cost of running the simulations. This relatively small number of simulations per pair may contribute to the consistent bais observed at the one percent significance level.

Table 8: Relationship between correlation and dynamic time warping

	Estimate	Std Err	Estimate	Std Err	Estimate	Std Err	Estimate	Std Err
Dependent Var.	log(dtw)		log(dtw)		log(dtw)		log(dtw)	
Intercept	3.842	0.000**	3.413	0.000**	3.589	0.000**	3.523	0.000**
Pearson's ρ	-46.914	0.000**	-40.856	0.000**	-48.546	0.000**	-50.500	0.000**
(Pearson's $\rho)^2$	-19.836	0.000**	-17.612	0.000**	-17.660	0.000**	-18.263	0.000**
Δ Unconditional Mean	–	–	308.194	0.000**	280.256	0.000**	276.413	0.000**
Total Unconditional Var	–	–	0.428	0.000**	0.239	0.044*	0.254	0.037*
Average Pair Persistance	–	–	0.089	0.188	-0.034	0.697	0.015	0.841
Intra-Sector Pair	–	–	–	–	–	–	0.068	0.023*
No. of ARMA Par	–	–	–	–	–	–	0.001	0.729
No. of GARCH Par	–	–	–	–	–	–	-0.003	0.802
2001	–	–	–	–	0.071	0.194	0.043	0.449
2002	–	–	–	–	-0.266	0.000**	-0.247	0.001**
2003	–	–	–	–	0.059	0.426	0.048	0.519
2004	–	–	–	–	-0.276	0.000**	-0.284	0.000**
2005	–	–	–	–	-0.191	0.011*	-0.215	0.006**
2006	–	–	–	–	-0.114	0.096+	-0.116	0.095+
2007	–	–	–	–	-0.033	0.660	-0.026	0.740
2008	–	–	–	–	0.142	0.017*	0.160	0.009**
2009	–	–	–	–	0.604	0.000**	0.628	0.000**
2010	–	–	–	–	0.082	0.181	0.109	0.087+
2011	–	–	–	–	0.049	0.373	0.057	0.316
2012	–	–	–	–	-0.203	0.002**	-0.202	0.003**
2013	–	–	–	–	0.008	0.902	-0.015	0.823
2014	–	–	–	–	-0.313	0.000**	-0.329	0.000**
2015	–	–	–	–	-0.162	0.006**	-0.159	0.010**
2016	–	–	–	–	-0.007	0.887	-0.008	0.873
2017	–	–	–	–	-0.403	0.000**	-0.423	0.000**
2018	–	–	–	–	-0.158	0.005**	-0.185	0.002**
2019	–	–	–	–	-0.135	0.005**	-0.150	0.002**
2020	–	–	–	–	0.262	0.000**	0.263	0.000**
2021	–	–	–	–	-0.020	0.659	-0.033	0.482
2022	–	–	–	–	0.163	0.002**	0.172	0.001**
Vol Model: gjr-std	–	–	–	–	–	–	0.005	0.852
Vol Model: cs-std	–	–	–	–	–	–	0.133	0.000**
Vol Model: cs-gjr	–	–	–	–	–	–	0.014	0.855
Vol Model: cs-cs	–	–	–	–	–	–	0.099	0.704
Vol Model: gjr-gjr	–	–	–	–	–	–	0.125	0.132
T-Dist: Skewed-Standard	–	–	–	–	–	–	-0.061	0.014*
T-Dist: Skewed-Skewed	–	–	–	–	–	–	0.072	0.438

Signif. Codes: 0 '**' 0.01 '*' 0.05 '+' 0.1 ' ' 1
Regression analysis ran on the correlation between observed correlation and dynamic time warping values for a stratified sample of stock pairs modeled in section 5