Chapter 1

Introduction

This essay focuses on the application of *dynamic time warping* (DTW) in the fields of finance and economics. Dynamic time warping is a well known method to compare the similarity of real-valued sequential data. It started to receive widespread attention in the mid-to-late 1970's [33, 20] as a means to group together the audio signals of the same word spoken by different speakers. As the name suggests, it treats the time dimension of a series as elastic. Under certain conditions the original temporal alignment between two series can be stretched and pulled so that a new optimally aligned series are closer in value. From an economist's perspective the idea of manipulating the time dimension of their data may seem strange. Whether studying a cohort using longitudinal data or comparing the historical Philip's curve of a country's inflation and unemployment rate, the researcher usually takes as granted the time synchronicity of their data.

To give an idea of how time synchronicity is pre-baked into classic financial analysis consider how correlation would be calculated between two time series, $\boldsymbol{x}$ and $\boldsymbol{y}$.

$$\rho(\boldsymbol{x}, \boldsymbol{y}) = \frac{\sum_{t=1}^{T}(x_t - \bar{x})(y_t - \bar{y})}{\sqrt{\sum_{t=1}^{T}(x_t - \bar{x})^2 \sum_{t=1}^{T}(y_t - \bar{y})^2}} \tag{1.1}$$

When looking at the covariance component in the numerator of equation (1.1), the statistic's assessment of association is based on the synchronous co-movements of the two variables. Suppose, for instance, there is a sample of T draws from the following multivariate normal distribution

$$\boldsymbol{Z_t} \sim N(\mathbf{0}, \boldsymbol{\Sigma}) \tag{1.2}$$

with

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix} \tag{1.3}$$

and the sample is given an integer valued time index from one to T. Consider the variance-covariance matrix estimated from the given sample. No one would be surprised to find that the sample statistics are close to their population values: $\sigma_1 \approx \hat{s}_1$, $\sigma_2 \approx \hat{s}_2$, and $\rho \approx \hat{\rho}$. Now, if a second multivariate sample is created $\boldsymbol{Z}' = [\boldsymbol{Z}_{(-T),1}, \boldsymbol{Z}_{(-1),2}]$ where the first column is all values of $\boldsymbol{Z}_1$ but the last one and the second column is all values of $\boldsymbol{Z}_2$ but the first one. 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 by the one day lag between the two series. Now consider if $\boldsymbol{x}$ and $\boldsymbol{y}$ are two different lengths or have missing values. Without some kind of preliminary data processing or imputation method the correlation value can not be calculated at all.

As documented below, DTW is already present in the finance and economics fields. What is missing from the literature, though, is consideration of how non-stationary behavior impacts the outcome of the dynamic time warping algorithm. This essay aims to fill that gap as well as provide context to using DTW as a complement, or even as a substitute, to correlation in certain situations. The rest of this section discusses some of the existing literature of DTW applied to fields of economics and finance and the data source used in this essay. Section 2 formally defines dynamic time warping, correlation, and euclidean distance. The latter two measures will be used as benchmarks to ground the analysis of DTW. Sector 3 directly compares DTW to Pearson's Correlation in their ability to form profitable portfolios while executing a specific pair-trading strategy. Section 4 conducts two separate simulation exercises to study the impact that non-stationarity has on the DTW algorithm. Section 5 uses regression analysis to analyze the mean relationship between DTW and correlation while Section 6 concludes.

1.1 Relevant Literature

As noted before, [33] and [20] are two early authoritative articles defining the core dynamic time warping algorithm and demonstrate its effectiveness to the field of speech recognition. Its influence has extended well beyond speech recognition and includes applications in the fields of data mining [3, 9, 24], information retrieval [28], agricultural and forestry management [5, 27, 29, 31, 38], and economics and finance [2, 7, 11, 17, 21, 26, 32, 36, 37], amongst many others. The reason for its wide adoption comes from its many attractive properties. First, the algorithm is non-parametric and operates on the data with minimal tuning. It is also flexible, being able to handle data of different lengths as well as data with time stamps that are not chronologically synced. DTW may also be helpful when the data is sparse or has many missing observations. Applications of DTW in the finance and economics do exist in the literature but the number of them remains relatively small. Of those mentioned in this essay, their application usually falls into one of the following three categories.

The most popular use of DTW is to find clusters or other sub-populations in a large set of sequential data without any prior knowledge of the latent underlying multinomial distribution. The approach is similar in spirit to the common K-means classifier 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 the geometric mean, or centroid, between series can lead to a group representation that does not capture important shape features of the underlying set of series it is summarizing. A few different approaches have been suggested. Instead of using the centroid (or geometric mean) at heart of K-means clustering, [7] describes a procedure based on creating groups centered around a shared *medoid*. The advantage here is that unlike the centroid, the medoids is a member of the dataset. Proposing yet another time series clustering approach, [30] leverages the concept of a barycenter from astronomy and develops a heuristic that finds the "barycentric average" of a set of series. [11] leverages the work of [30] to quarterly real GDP growth of the 50 US states during the 2007 recession. The authors group the US states into seven distinct business cycle clusters. [2] combines DTW with graph structures to analyze financial network effects and identify the dominant set of variables within the network.

Another set of use-cases involves the classic information retrieval task where a researcher submits a query to a database, and in return, receives a set of relevant elements from the database, ranked in order of their relevance to the query. In the context of this essay the query and the elements in the database are time series. [36]

uses DTW in support of technical analysis of financial assets. The author constructs a database containing numerous stock pattern "templates" and then submit short histories of contemporary stock prices as queries to the database. The database returns the stock "templates" that are most relevant, ranked by their DTW distance. In a similar vein [32] uses DTW to match current Treasury Spreads against historical spreads to predict recessions in the US economy.

A third DTW application in the literature is to identify the existence of leading or lagging relationships between financial time series. [17] uses DTW to control for any temporal structure in stock returns before estimation of a stock's beta[1]. They find that applying DTW before calculating a stock's beta restores the linear relationship between risk and expected return that is assumed in the capital asset pricing model (CAPM). In support of a high frequency trading strategy, [21] propose a multinomial DTW algorithm to estimate any existing lead/lag relationship between pairs of financial assets. The authors validate their approach on simulated brownian motion. Finally, [26] uses DTW to identify periods where the prices of WTI and Brent crude "decouple" and then "recouple" at a later date.

1.2 Data

This essay's analysis examines the companies comprising the S&P 500 on the date of of October 1st, 2023. 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 macroeconomic 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[2]. The adjusted close takes into account any reinvested dividends as well as stock splits or the introduction of new stock that

[1]From wikipedia: In finance, "beta" is a statistic that measures the expected increase or decrease of an individual stock price in proportion to movements of the stock market as a whole. Beta can be used to indicate the contribution of an individual asset to the market risk of a portfolio when it is added in small quantity. It refers to an asset's non-diversifiable risk, systematic risk, or market risk.

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

occurred 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 1.1 for a visual summary of the price history for the sample of companies considered in this essay.

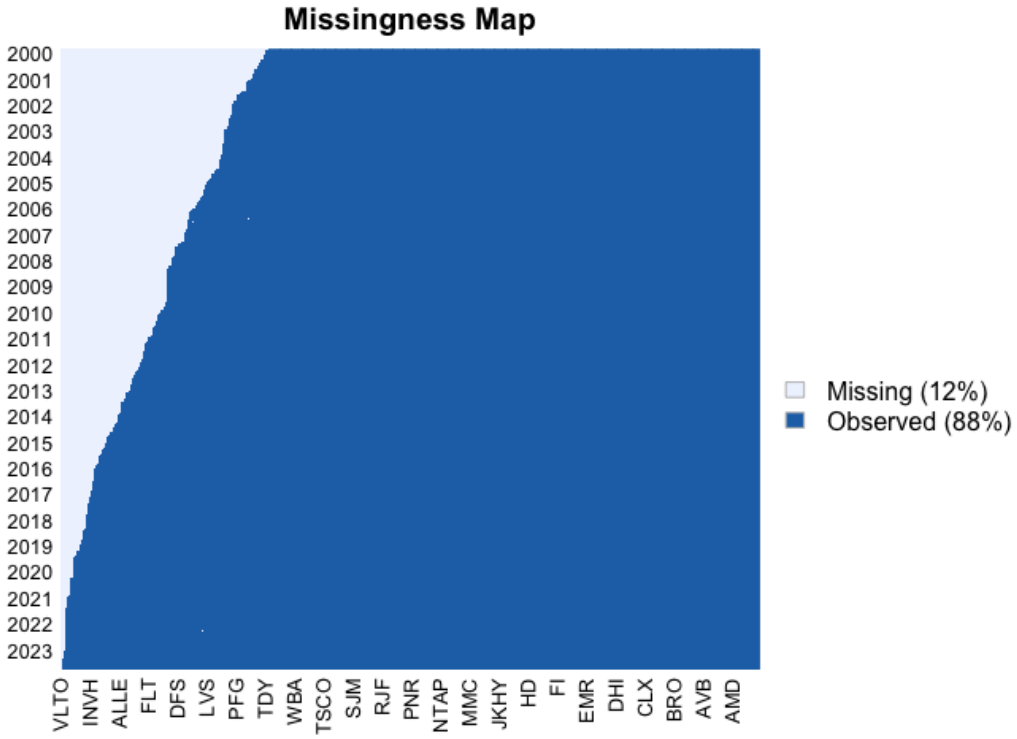

Figure 1.1: This image captures the available history, by stock, of the S&P data used in this essay. Light blue shading indicates timestamps where price data is not available. Dark blue represents available history. Of the 503 stocks in the dataset, there are 349 with a full series of prices throughout the time period under investigation.

Chapter 2

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 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.

$$x_t = \log(S_t) - \log(S_{t-1}) \tag{2.1}$$

Correlation 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 cluster time series usually use the price level or the nominal value of the series. This contrasts with 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 essay 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 essay[1].

[1]Simulation exercises conducted in Section 4.1 are an exception

2.1 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

$$\rho(\boldsymbol{x}, \boldsymbol{y}) = \frac{\sum_{t=1}^{T}(x_t - \bar{x})(y_t - \bar{y})}{\sqrt{\sum_{t=1}^{T}(x_t - \bar{x})^2 \sum_{t=1}^{T}(y_t - \bar{y})^2}} \tag{2.2}$$

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 4, 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 conjunction with a Copula to estimate a bivariate distribution for a pair of stock returns. A t-Copula, which is characterized by its correlation and degrees-of-freedom parameters (ρ, ν), is optimized via maximum likelihood.

$$\begin{aligned} C_{\boldsymbol{t}}(u, w; \rho, \nu) &= \boldsymbol{t}_v\left(t_\nu^{-1}(u), t_\nu^{-1}(w)\right) \\ &= \int_{-\infty}^{t_\nu^{-1}(u)} \int_{-\infty}^{t_\nu^{-1}(w)} q\,(1-\rho)^{-\frac{1}{2}}\left[1 + \nu^{-1}\left(x^2 - 2\rho xy + y^2\right)\right]^{-(\nu+2)/2} dxdy \end{aligned} \tag{2.3}$$

The values u and w are a transformed representation[2] of random variables x and y from equation (2.2). 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 [6] summarizes Copula theory and its application to foreign exchange rates.

[2]See equation (A.5) in the Appendix

2.2 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.

$$||\boldsymbol{X}, \boldsymbol{Y}||_2 = \left[(X_1 - Y_1)^2 + (X_2 - Y_2)^2 + \cdots + (X_T - Y_T)^2\right]^{\frac{1}{2}} \tag{2.4}$$

Euclidean distance, as defined in equation (2.4), serves as benchmark similarity measure where no time elasticity is allowed between the two series.

2.3 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 constrained to equal each other – like in equations (2.2) and (2.4). Time is allowed to stretch and compress before a local cost function is applied to the pairs of values from the two series. This essay will adhere to the classic definition of dynamic time warping. For notation this essay borrows heavily from [28].

2.3.1 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.

$$c : X_{t_l} \times Y_{s_l} \to \mathbb{R}_{\geq 0} \tag{2.5}$$

The behavior of the warping paths obeys two conditions which are listed below.

$$\text{Boundary Condition:} \quad p_1 = (1,1) \text{ and } p_L = (T,S) \tag{2.6}$$
$$\text{Step-Size Condition:} \quad p_l - p_{l-1} \in \{(1,0),(0,1),(1,1)\} \tag{2.7}$$

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:

$$\mathbf{C}(\boldsymbol{X}, \boldsymbol{Y}) = \begin{bmatrix} c(X_T, Y_1) & c(X_T, Y_2) & \cdots & c(X_T, Y_S) \\ \vdots & \vdots & \vdots & \vdots \\ c(X_2, Y_1) & c(X_2, Y_2) & \cdots & c(X_2, Y_S) \\ c(X_1, Y_1) & c(X_1, Y_2) & \cdots & c(X_1, Y_S) \end{bmatrix} \tag{2.8}$$

This article uses squared distance[3] for the local cost function to make it comparable with the euclidean distance defined in Section 2.2. A warping path's total cost is the sum of the local costs it incurs as it travels from the start of the series (bottom left of $\boldsymbol{C}$) to their end (top right of $\boldsymbol{C}$):

$$\mathbb{c}_{\boldsymbol{p}}(\boldsymbol{X}, \boldsymbol{Y}) = \sum_{l=1}^{L} c(X_{t_l}, Y_{s_l}) \tag{2.9}$$

There are many admissible[4] 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

$$\mathbb{c}_{\boldsymbol{p}^*}(\boldsymbol{X}, \boldsymbol{Y}) \leq \mathbb{c}_{\boldsymbol{p}}(\boldsymbol{X}, \boldsymbol{Y}) \text{ for all } \boldsymbol{p} \in \mathbb{P}(\boldsymbol{X}, \boldsymbol{Y}) \tag{2.10}$$

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

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

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

is efficient computation. Since the number of warping paths grows exponentially[5] in T and S, the computation time needed to check every warping path becomes problematic for series of even moderate size.[6]

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 dynamic 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:

$$A_{t,1} = \sum_{k=1}^{t} c(X_k, Y_1) \text{ for } t \in [1:T] \tag{2.12}$$

$$A_{1,s} = \sum_{k=1}^{s} c(X_1, Y_k) \text{ for } s \in [1:S] \tag{2.13}$$

$$A_{t,s} = c(X_t, Y_s) + \min\left(A_{t-1,s-1},\ A_{t-1,s},\ A_{t,s-1}\right) \tag{2.14}$$

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

[5]The total number of paths can be calculated by the *Delannoy Number*:

$$D(T, S) = \sum_{k=0}^{min(T,S)} \binom{T}{k}\binom{S}{k} 2^k \tag{2.11}$$

To give the reader an idea of the rapid growth of the total number of warping paths consider the following Delannoy values: $D(1,1) = 3$, $D(2,2) = 13$, $D(4,4) = 321$, and $D(8,8) = 265,729$. On average there are close to 250 trading days per year. The Delannoy number $D(250, 250)$ is approximately 8.842641×10^{189}.

[6]Reducing the computational burden of the DTW algorithm is a well studied topic in computer science. See [23] for a discussion on reducing the time complexity of the DTW algorithm

One more condition is added to this essays'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 [33] and is frequently used for its simplicity and its performance [13]. 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. The value of new information in markets dissipates quickly over time.

2.3.2 An Example

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.

Figure 2.1 demonstrates the extent of time warping made possible by the DTW algorithm and the impact of using a global constraint versus an unconstrained implementation. All figures on the left plot the prices for Teradyne Inc. (TER) and Lam Research Corp. (LRCX) stock during the 2021 calendar year. Instead of the original stock price, each series is presented in their standard price form by using equation (3.1). The line segments connect the matched time index from the two series together. All figures on the right show the path of the optimal warping path as a solid line. The dotted lines denote a 15-day window bounding the main diagonal, which is the line stemming from the origin at a 45 degree angle. This main diagonal indicates a synchronous time alignment between the two series. Three sets of graphs are presented in Figure 2.1. Figure 2.1.A captures the standard approach of comparing time series with no time dilation allowed. This is why the line segments on the left graph are vertical and why the time pairing on the right never deviates from the main diagonal. In this case, the total number of index pairs, L, is equal to the length of the series, 252, and the sum of squared distances is 11.19. Figure 2.1.B summarizes the optimal warping path of the DTW algorithm when it is subjected to a 15-day Sakoe-Chiba window. The time warping can be viewed in two ways. First, the line-segments on the left connecting the time indicies are allowed to be non-vertical. The two series have the most consistent temporal alignment at the beginning of the series, from time index 1 to 40, and at the end of the series, from time index 205 to 252. As a heuristic, two series have a strong temporal alignment when the optimal warping path is parallel to the main diagonal. If the warping path is parallel to the main diagonal for an extended duration then this impies that the warping distance between the two series is stable as well. This is easiest to see at

the end of the two series when the optimal warping path bumps against the lower bound of the Sakoe-Chiba window. Since this part of the warping path lies below the main diagonal, it can be inferred that LRCX is the leading series. In contrast, the temporal alignment between the two series during the middle part of the year is weaker. The alignment oscillates between TER leading LRCX, LRCX catching up and leading, and then back to TER. This oscillation is easy to see when looking at the optimal warping path. During the middle of the year the optimal warping path looks like an irregular staircase, flat at times when the optimal warping path holds at a single time index for LRCX, but punctuated with vertical line segments when the optimal warping path holds at a single time index for TER. In this case of constrained DTW the total number of index pairs, L, is 462 and the total sum of squared distances is 62.62. The bottom two figures summarize the optimal warping path of the DTW algorithm without constraint. In this special case time is allowed to be infinitely elastic and Figure 2.1.C captures the extreme amount of time warping that occurs between TER and LRCX. The unconstrained optimal warping path is a tale of two parts. The first part is defined by the ever increasing time warping observed over the first 205 days of the year. When the standard stock price of LRCX hits a local minimum on January 29th (time index 19), the optimal warping path holds at this time index for 168 steps in the path. This is easily seen by the figure on the right of Figure 2.1.C with its extended flat portion of the warping path. The second part of this tale is the unwinding[7] of that extreme time warping.

Figure 2.2 presents four graphs demonstrating the impact that using ARMA-GARCH to model univariate stock return series can have on the characterization of a stock pair's bivariate distribution. Again, all figures pertain to the daily stock returns of Teradyne Inc. (TER) and Lam Research Corp. (LRCX) or their joint distribution. Figure 2.2.A shows the contours of a t-Copula estimated on the bivariate distribution of unadjusted returns for TER and LRCX. For each return series a t-distribution was fitted on the unadjusted log-returns, individually. The two fitted t-distributions are used to estimate a t-Copula which characterizes their joint distribution. The stocks are highly correlated with a correlation coefficient of 0.86, and with a degree of freedom parameter of 4.79, the distribution has substantial amounts of excess kurtosis (fat-tails). Figure 2.2.B shows a similar contour plot, but in this case, each marginal return series is model as an ARMA-GARCH process before the estimation of the t-Copula. The result is noticeable. Correlation is reduced slightly from 0.86 to 0.84, but the level of excess kurtosis drops considerably. The degree of freedom parameter raises to 9.01 (from 4.91) after controling

[7]The time warping must be unwound to satisfy boundary condition (2.6)

for spikes in market volatility. Figures 2.2.C and Figure 2.2.D show the fitted return and volatility values estimated from the ARMA-GARCH models, respectively. TER is estimated with a ARMA(0,0)-GARCH(1,1) with intercept parameterization, a standard GARCH volatility model, and a t-distribution. LRCX is estimated with a ARMA(2,0)-GARCH(1,1) with intercept parameterization, a component GARCH volatility model, and a t-distribution.

Finally, Figure 2.3 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. Much like water running down a mountain, the optimal warping path follows the path of least resistance, or lowest cost, from the top-right of the accumulated cost matrix down to the lower-left.

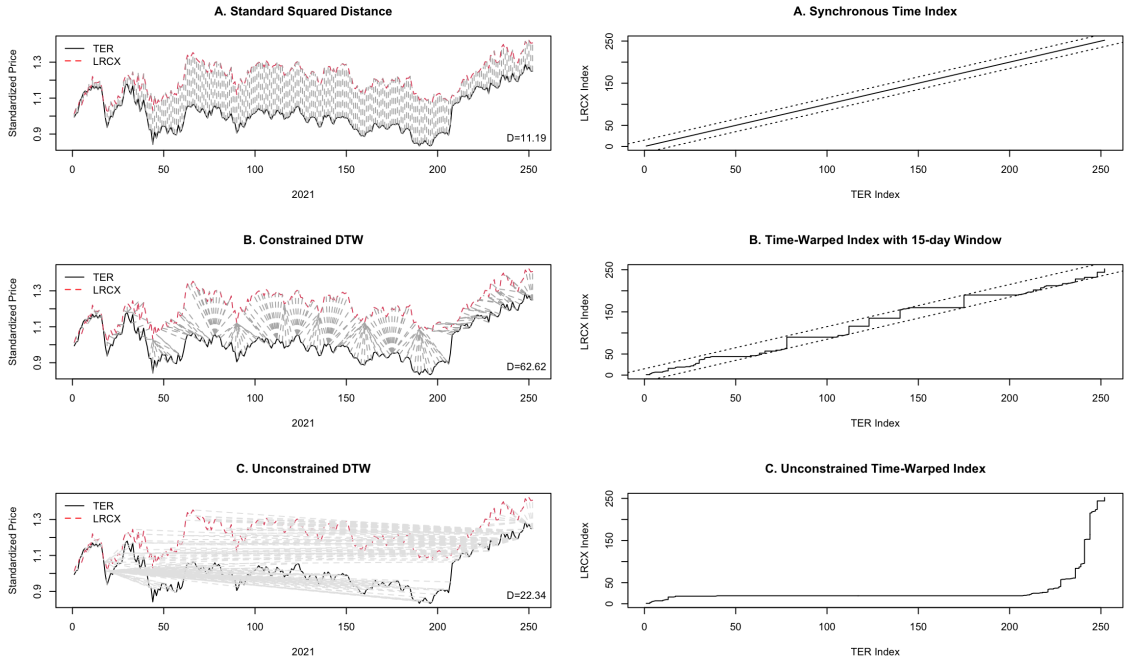

Figure 2.1: These figures demonstrate the extent of time warping made possible by the DTW algorithm and the impact of using a global constraint versus an unconstrained implementation. The figures of the left plot the standard prices for Teradyne Inc. (TER) and Lam Research Corp. (LRCX) stock during the 2021 calendar year. The line segments connect the matched time index from the two series together. The figures on the right show the path of the optimal warping path as a solid line. The dotted lines denote a 15-day window surrounding the main diagonal. **Top**: The top two figures show the natural time-alignment of the two stocks. Without time warping, the time index of the two stocks are paired contemporaneously. This is seen by the vertical lines between the two series. In this case, the total number of index pairs, L, is equal to the length of the series, 252, and the sum of squared distances is 11.19. **Middle**: The middle two figures summarize the optimal warping path of the DTW algorithm with a 15-day Sakoe-Chiba constraint. The total number of index pairs is 462 and their sum of squared distances is 62.62. **Bottom**: The bottom two figures summarize the optimal warping path of the unconstrained DTW algorithm. The total number of index pairs is 474 and their sum of squared distances is 22.34.

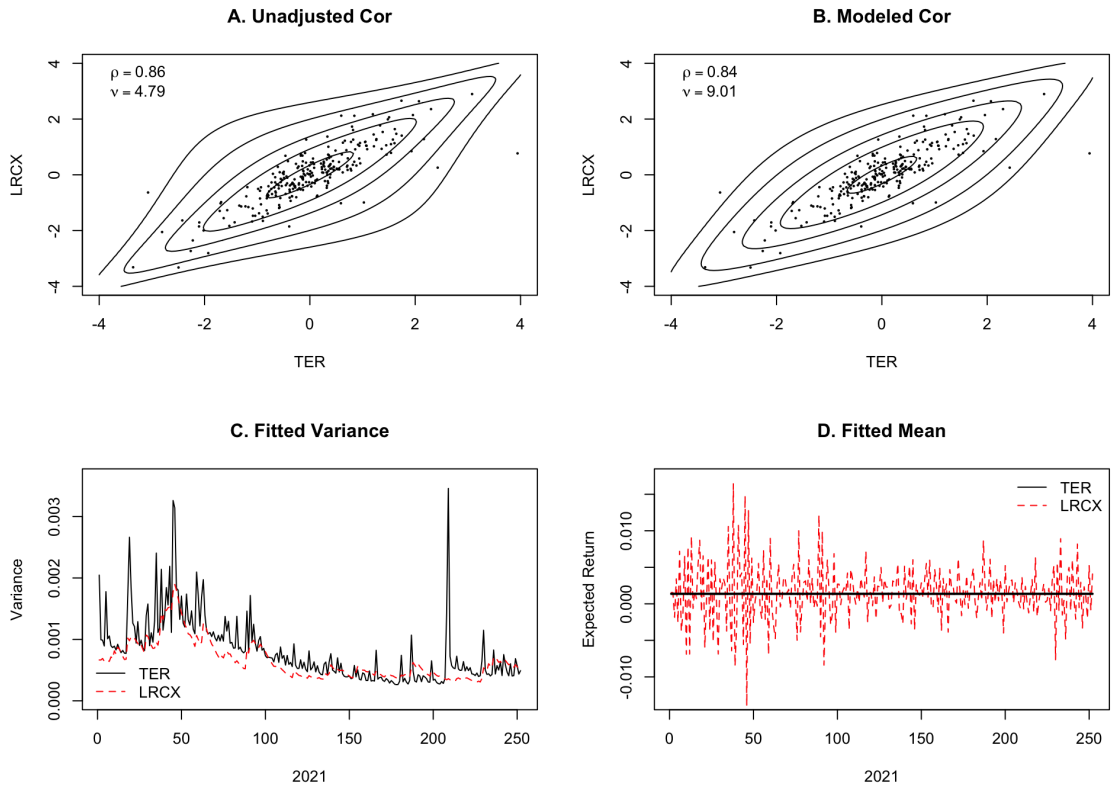

Figure 2.2: These figures demonstrate the impact that using ARMA-GARCH to model univariate stock return series can have on the characterization of a stock pair's bivariate distribution. All figures pertain to the daily stock return distributions of Teradyne Inc. (TER) and Lam Research Corp. (LRCX). To provide clean visuals, the returns have been z-scored. **Top-left**: The contours of a t-Copula is estimated on the bivariate distribution of unadjusted returns for TER and LRCX. The stocks are highly correlated with a correlation coefficient of 0.86, and with a degree of freedom parameter of 4.79, the distribution has substantial amounts of excess kurtosis (fat-tails). **Top-right**: The contours of a t-Copula estimated on the normalized returns for TER and LRCX where each stock's return series is modeled as an ARMA-GARCH process. Correlation is reduced slightly from 0.86 to 0.84. Notably, the level of excess kurtosis drops considerably. The degree of freedom parameter raises from 4.79 to 9.01 after controling for spikes in market volatility. **Bottom-left**: The fitted variance process from the estimated ARMA-GARCH models. **Bottom-right**: The fitted return series from the estimated ARMA-GARCH models. For TER the mean process was estimated with an intercept term only which explains its flat appearance in the graph.

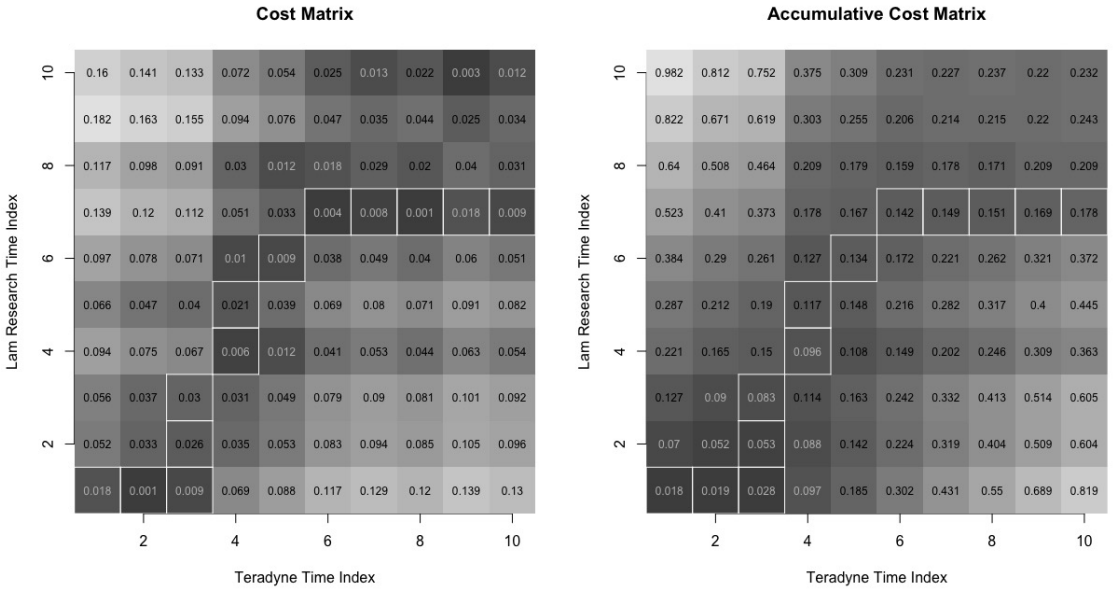

Figure 2.3: **Left**: The left image is the cost matrix for the first ten values of the Teradyne and Lam Research stock prices after standardization. The local cost function is squared distance. **Right**: The right image is the accumulative cost matrix for the same sequence of prices. The *optimal warping 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.

Chapter 3

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 1990s and 2000s 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.

3.1 Trading Strategy

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

3.1.1 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:

$$p_t = \prod_{1}^{t}(1 + x_t) \tag{3.1}$$

Here, x_t is the log return of the stocks price as defined in equation (2.1). 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 two standard 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 2: unadjusted and modeled correlation, euclidean distance, and dynamic time warping. Portfolios are formed by the closest 25, 50, and 100 pairs of stocks.

3.1.2 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 deviates beyond the two standard deviation limit, then a position in a pair will never be opened.

3.2 Post Trade Analysis

Table 3.1 provides summary results of the trading strategy contrasted by distance measure, portfolio size, and definition of "returns" to the portfolios constructed here. Table 3.1.A records committed returns where the total return of the portfolio during the trading period is divided by the number of pairs selected to trade. Table 3.1.B 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 table, 3.1.C, 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 chosen by dynamic time warping. But the return is small in magnitude, especially compared to the alternative of the baseline long-position as seen in 3.1.C. The DTW and euclidean measures perform better than the correlation measures in almost all categories. The only exception is the one percent return of fully invested equity to the top 25 pairs chosen by unadjusted correlation, which beats out the portfolio constructed 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 synchronized time this may give euclidean distance an inherent advantage over correlation.

Looking closely at parts A and B of Table 3.1, note 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 sizable 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 chosen 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 substantial number of inter-sector pairs formed using DTW and euclidean distance (compare Tables 3.2 and 3.3). 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 3.1, several bar-charts 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 fall-out from the dot-com bust in 2001 and 2002, the mortgage backed security crisis in 2008 and 2009, and the shut-down economy of 2020.

Panel A: 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

Panel B: 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

Panel C: 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 3.1: 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	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 3.2: 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 3.3: 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 3.2. 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.

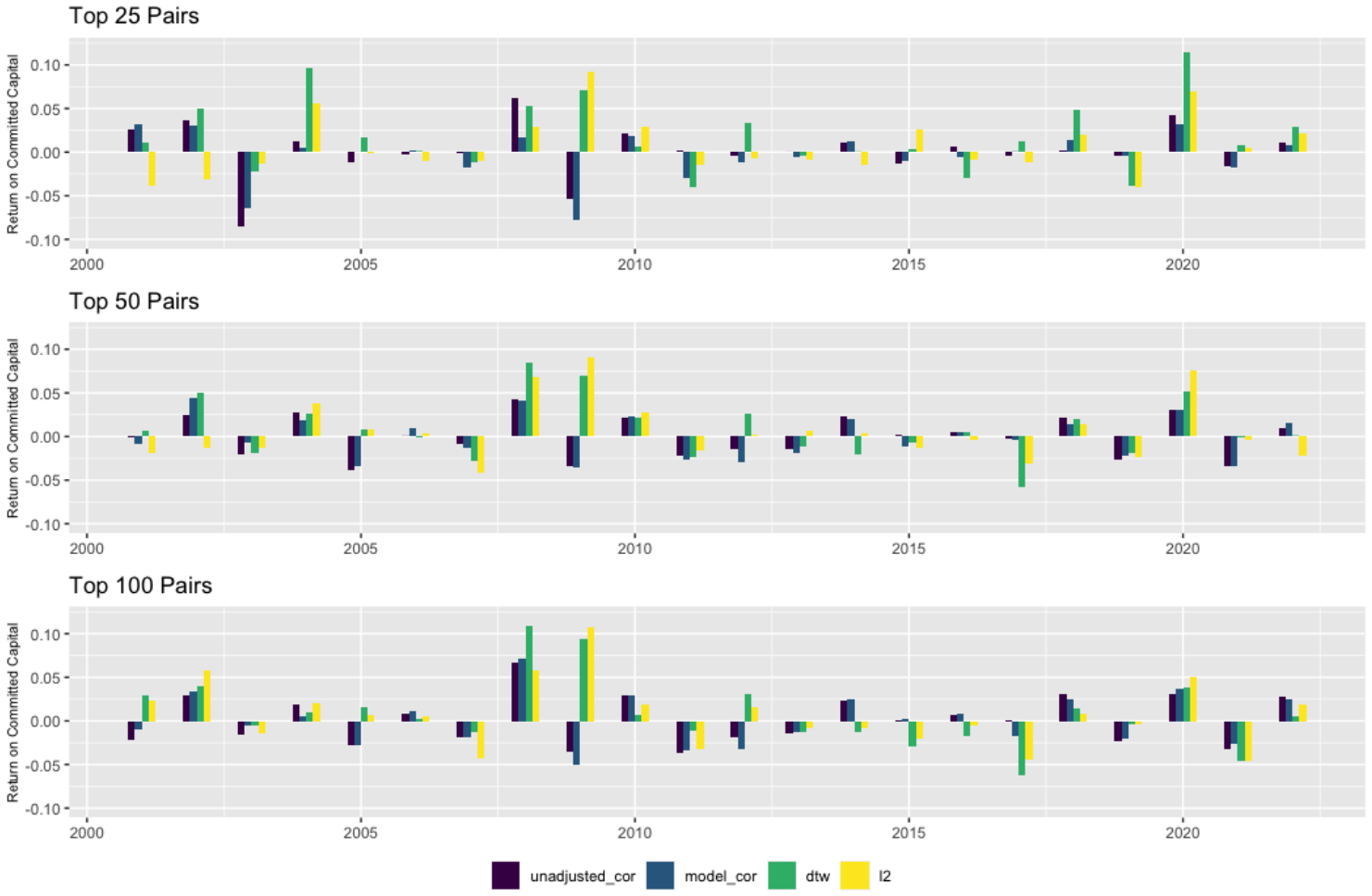

Figure 3.1: Bar-charts summarizing the annual committed returns of the trading strategy under different buy signals and portfolio sizes.

Chapter 4

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[1], there is no guarantee that it satisfies the triangle inequality[2].

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. The first approach investigates the relationship via simulation. Two sets of simulations 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 3 with the copula approach to modeling joint distributions.

In addition to the simulation studies, the second approach 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 3. The sample

[1]Symmetry is not always guaranteed. This essay uses a symmetric cost function and a symmetric step-condition (see equation (2.5) and condition (2.7) in Section 2.3). See one of the early authoritative works [33] for a detailed discussion on global constraints and step-conditions.

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

used for the regressions is the same sample used in the Monte Carlo exercises. Taken together, the analysis in this section provides the contours of the relationship between Pearson's correlation and dynamic time warping when estimated on a set of stock pairs.

4.1 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 (4.1) below summarizes the process.

$$X_t = \omega + \epsilon_t \qquad \epsilon_t \sim \mathcal{N}(0, \sigma^2) \tag{4.1}$$

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 approach 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 (4.1), two series can be constructed as a random walk that obey the following specification:

$$X_t = \omega + X_{t-1} + \epsilon_t \qquad \epsilon_t \sim \mathcal{N}(0, \sigma^2) \tag{4.2}$$

In equation (4.2) 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 (4.1) and (4.2) 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 3 to backtest the pair-trading strategy. To further align the procedure used in Section 3 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 4.1, 4.2, and 4.3.

The results of dynamic time warping applied to a pair of stationary white noise series offers a natural starting point. Considering Table 4.1.A under constrained DTW, a few points are worth noting. First, dynamic time warping values have a positive relationship 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 subpanel (constant $\sigma_A = 1$ with increasing σ_B) with the third subpanel (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 subpanel). This divergence 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, examining 4.1.B, 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 subpanels of Tables 4.1.A and 4.1.A.

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 Table 4.2.A, the mean DTW value from two independent standard normal processes ($\sigma_A = \sigma_B = 1$) is 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 distributions have long right-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 subpanels of Tables 4.2 and 4.3, regardless of whether DTW paths are constrained or unconstrained. Having a

shared drift parameter appears to keep the two simulated series in a closer relative neighborhood than two random walk series with no drift.

4.2 Monte Carlo Simulation based on the ARMA-GARCH Framework

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

$$\mathbb{E}[\epsilon_{A,t}\epsilon_{B,s}] = 0 \qquad \forall t, s \tag{4.3}$$

$$\mathbb{E}[\epsilon_{A,t}\epsilon_{A,s}] = 0 \qquad \forall t, s \tag{4.4}$$

This restriction is convenient 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[3] [4].

In Section 3 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 estimated 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

[3] Evidence from [21] supports our analysis that the variance of DTW increases as the correlation between asset movements decreases. This aligns with the results from the simulation exercises carried out in this section.

[4] [8] is a rare paper dealing directly with statistical inference of the DTW measure.

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 confirming a negative relationship between the two we find that the variance of DTW values is significance at all levels of correlation.

Instead of analyzing the entire 2.2 million pairs from Section 3 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 are 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 [6] 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, conditional 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 (A.2), (A.3), and (A.5) from Appendix A).

2. The probability integral transform is used to reformulate each marginal model's fitted residuals as a sample from the uniform distribution.

$$\hat{U}_A = \hat{F}_A(s) = \int_{-\infty}^{s} \frac{1}{\hat{\sigma}_{A,\,t}} \hat{g}_A\left(z|\hat{\mu}_t\right) dz \tag{4.5}$$

$$\hat{U}_B = \hat{F}_B(s) = \int_{-\infty}^{s} \frac{1}{\hat{\sigma}_{B,\,t}} \hat{g}_B\left(z|\hat{\mu}_t\right) dz \tag{4.6}$$

   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 its 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[5].

$$r^{(i)}_{A,\,t} = \hat{F}^{-1}_A\left(u^{(i)}_{A,\,t}\ ;\ \mu = \hat{\mu}_{A,\,t},\, \sigma^2 = \hat{\sigma}^2_{A,\,t}\right) \tag{4.7}$$

[5]All marginal return models estimated in Section 3 have a time-varying variance equation (see equation (A.3)) but not every return series has a time-varying mean component (see equation (A.2)).

$$r_{B,\,t}^{(i)} = \hat{F}_B^{-1}\left(u_{B,\,t}^{(i)}\ ;\ \mu = \hat{\mu}_{B,\,t},\, \sigma^2 = \hat{\sigma}_{B,\,t}^2\right) \tag{4.8}$$

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 equation (3.1).

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) correlation and tail-dependency observed in the stock pairs is preserved by the copula while the unique idiosyncrasies (see Appendix A) of each stock is retained by the marginal models. This leaves only the variation in 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 process described in Appendix 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 stock's 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 (4.7) and (4.6). 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[6].

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 4.4 represent the proportion of the sampled stock pairs from Section 3 whose observed dynamic warping value falls outside the confidence intervals constructed using the Monte Carlo simulation

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 (4.7) and (4.8): $\mu = \hat{\mu}_i$ for $i \in A, B$.

[6]The vignette [14] summarizing the software implementation used in this essay provides the functional relationship between the conditional mean and variance equations for each GARCH specification with their unconditional analogues.

approach described in this section. The sampled pairs are split into two categories based on the ARMA specification chosen for stock returns modeled in Section 3. 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 (A.2)). 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 (4.7) and (4.8). 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 (4.7) and (4.8) (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 procedure (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 the significance level.

A. 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

B. 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 4.1: Distribution of DTW under Stationary White Noise DGP

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 slightly 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.

A. 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

B. 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 4.2: Distribution of DTW under a Random Walk without Drift DGP

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 4.1 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.

A. 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

B. 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 4.3: Distribution of DTW under Random Walk with Drift DGP
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 4.4: 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 3 whose observed dynamic warping value falls outside the Monte Carlo confidence intervals constructed using the simulation approach described in Section 4.2. 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 component in their mean equation (see equation (A.2)). 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 (4.7) and (4.8).
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 bias observed at the one percent significance level.

Chapter 5

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 was constructed for the simulations in Section 4.2: 7,765 stock pairs formed during the annual formation periods while executing the pair-trading strategy. The results summarized in Tables 4.1, 4.2, and 4.3 help inform this section's choice of explanatory variables. Each pair in the sample has an observed correlation and dynamic time warping value. This essay will focus on the general regression equation described below[1].

$$\log(\widehat{dtw}) = \beta_0 + \beta_1\hat{\rho} + \beta_2\hat{\rho}^2 + \boldsymbol{\beta}_3\boldsymbol{X} + \varepsilon \tag{5.1}$$

The main relationship of interest, between the observed correlation and dynamic time warping values, is written explicitly in equation (5.1). It has two important features. One is the log-transformation that is applied to the observed dynamic time warping values. The second is the quadratic polynomial applied to the correlation parameter obtained by fitting a t-Copula to a stock pairs return series. Since many of the remaining explanatory variables used in this section—captured by $\boldsymbol{X}$ in equation (5.1)—are derived from the optimized ARMA-GARCH models, the estimated correlation value used in the regression is the modeled correlation described in equation (2.3). Informed by Tables 4.1 through 4.3 two variables are constructed to capture

[1]The hats are kept to reinforce the fact that we are regressing on the observed values from Section 3

the difference in expected return and the aggregate volatility of the two return series for each pair in the sample. A third variable is constructed to capture the strength of "mean-reversion" in each stock's return series:

1. Δ **Unconditional Mean**: The absolute difference between the unconditional mean return for each stock in the pair

$$|\mu_A - \mu_B| \tag{5.2}$$

2. **Total Unconditional Variance**: The sum of unconditional volatility model by each stock's ARMA-GARCH model

$$\sigma_A^2 + \sigma_B^2 \tag{5.3}$$

3. **Average Pair Persistence**: The geometric average of each stock return's persistence as estimated by each series' optimized ARMA model. Persistence is a measure of a series' *memory* or how long a small perturbation at a point in time effects the future values of the series. It takes values on the unit interval with a value of one indicating an infinitely long memory.

Additional explanatory variables are added to help isolate the relationship between correlation and dynamic time warping:

1. **Year**: A set of indicator variables for yearly fixed effects
2. **Intra-Sector Pair**: An indicator variable that has a value of one if each stock in the pair comes from the same sector
3. **Volatility Model**: An indicator variable that uniquely identifies what kind of volatility models are used for each stock in the pair. Three choices are available for each marginal model. **std**: the standard GARCH model of [4]. **gjr**: a GARCH model from [15] that allows for asymmetric responses to positive or negative returns. **cs**: The component-GARCH model from [25].
4. **Return Distribution** An indicator variable that uniquely identifies what kind of distribution models are used for each stock in the pair. Two choices are available for each marginal model. **Standard**: The standard t-Distribution. **Skewed**: The skewed t-Distribution of [10].
5. **No. of ARMA Parameters**: The total number of parameters in the conditional mean equations from both marginal models

6. **No. of GARCH Parameters**: The total number of parameters in the conditional variance equations from both marginal models

Results for the applied regressions are summarized in Table 5.1.

Table 5.1: Relationship between correlation and dynamic time warping

	Coef	Std Err	Coef	Std Err	Coef	Std Err	Coef	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**
Fitted Corr. $\hat{\rho}$	-46.914	0.000**	-40.856	0.000**	-48.546	0.000**	-50.500	0.000**
(Fitted Cor. $\hat{\rho}$)^2	-19.836	0.000**	-17.612	0.000**	-17.660	0.000**	-18.263	0.000**
Δ Unconditional Mean[a]	–	–	308.194	0.000**	280.256	0.000**	276.413	0.000**
Total Unconditional Var[b]	–	–	0.428	0.000**	0.239	0.044*	0.254	0.037*
Average Pair Persistence[c]	–	–	0.089	0.188	-0.034	0.697	0.015	0.841
Intra-Sector Pair	–	–	–	–	–	–	0.068	0.023*
No. of ARMA Par[d]	–	–	–	–	–	–	0.001	0.729
No. of GARCH Par[d]	–	–	–	–	–	–	-0.003	0.802
2001[e]	–	–	–	–	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[f]	–	–	–	–	–	–	0.005	0.852
Vol Model: cs-std[f]	–	–	–	–	–	–	0.133	0.000**
Vol Model: cs-gjr[f]	–	–	–	–	–	–	0.014	0.855
Vol Model: cs-cs[f]	–	–	–	–	–	–	0.099	0.704
Vol Model: gjr-gjr[f]	–	–	–	–	–	–	0.125	0.132
T-Dist: Skewed-Standard[g]	–	–	–	–	–	–	-0.061	0.014*
T-Dist: Skewed-Skewed[g]	–	–	–	–	–	–	0.072	0.438

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

[a] The absolute difference between the unconditional mean return for each stock in the pair

[b] The sum of the unconditional variances from each stock's return series in the pair

[c] The geometric average of each stock return's persistence as estimated by each series' optimized ARMA model. Persistence a measure of the a series' *memory* or how long a small perturbation at a point in time effects the future values of the series. It takes values on the unit interval with a value of one indicating an infinitely long memory.

[d] The total number of ARMA or GARCH parameters across both pairs.

[e] A set of indicator variables for annual fixed effects

[f] Indicator variable that uniquely identifies what kind of volatility models are used for each stock in the pair. Three choices are available for each marginal model. **std**: the standard GARCH model of [4]. **gjr**: a GARCH model from [15] that allows for asymmetric responses to positive or negative returns. **cs**: The component-GARCH model from [25].

[g] Indicator variable that uniquely identifies what kind of distributions models are used for each stock in the pair. Two choices are available for each marginal model. **Standard**: The standard t-Distribution. **Skewed**: The skewed t-Distribution of [10].

Chapter 6

Conclusion

This essay offered an overview of the dynamic time warping distance as a measure of time series similarity and its application in the fields of Economics and Finance. Pearson's coefficient of correlation, indispensable to these fields in its own right, provides a foil. Section 3 confirms that daily arbitrage opportunities are very difficult to obtain but it also illustrates key differences between DTW and correlation. Section 4 points out that we can expect that DTW values from two stationary series to be much lower, consistently, than for time series with a non-stationary character. The difference between the average return of the two series and the total variance each impact the distribution of DTW but in different ways.

Appendix A

ARMA-GARCH Modeling

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 [6]. 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:

$$x_t = \mu_t + \epsilon_t \tag{A.1}$$

$$\mu_t = \mu(\phi, \theta, x_{\{s:\, s<t\}}, \varepsilon_{\{s:\, s<t\}}) = \phi_0 + \sum_{i=1}^{p} \phi_i x_{t-i} + \sum_{j=1}^{q} \theta_j \, \varepsilon_{t-j} + \varepsilon_t \tag{A.2}$$

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.

$$\sigma_t^2 = C(\varepsilon_{t-1}^2,\, \sigma_{t-1}^2) + \sum_{j=1}^{m} \alpha_j A(\varepsilon_{t-j}^2) + \sum_{i=1}^{k} \beta_i B(\sigma_{t-i}^2) \tag{A.3}$$

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

$$z_t(\phi, \theta, \alpha, \beta) = \frac{x_t - \mu(x_{t-1}, \phi, \theta)}{\sigma(x_{t-1}, \alpha, \beta)} \tag{A.4}$$

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:

$$f(x_t|\mu_t, \sigma_t^2, \lambda, \Delta) = \frac{1}{\sigma_t} g(z_t|\lambda, \Delta) \tag{A.5}$$

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 [35]. 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 [10]}
   - Model Specification: A, B, C = {Standard GARCH [4], gjr-GARCH [15], Component GARCH [25] }

   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 [16] non-parametric density test jointly for i.i.d and correct distribution specification, (d) Shapiro-Wilks [34] 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.

Bibliography

[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] Donald J Berndt and James Clifford. "Using dynamic time warping to find patterns in time series". In: *Proceedings of the 3rd international conference on knowledge discovery and data mining*. 1994, pp. 359–370.

[4] 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.

[5] 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.

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

[7] 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.

[8] 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].

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

[10] 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.

[11] 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.

[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] 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.

[24] 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.

[25] Gary G. J. Lee and R. Engle. "A Permanent and Transitory Component Model of Stock Return Volatility". In: *Cointegration, Causality and Forecasting: A Festschrift in Honor of Clive W.J. Granger*. New York, NY: Oxford University Press, Dec. 1999.

[26] Loretta Mastroeni et al. "Decoupling and recoupling in the crude oil price benchmarks: An investigation of similarity patterns". In: *Energy Economics* 94 (2021), p. 105036. ISSN: 0140-9883. DOI: https://doi.org/10.1016/j.eneco.2020.105036.

[27] 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.

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

[29] 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.

[30] 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.

[31] 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.

[32] Tasneem Raihan. "Predicting US Recessions: A Dynamic Time Warping Exercise in Economics". Sept. 2017.

[33] 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.

[34] 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.

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

[36] 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.

[37] 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.

[38] 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.