

1 Introduction

Measuring the association of two assets is a fundamental concept in finance and economics. Portfolio managers have to think about it when they open or close a position [13, 17]. Regulators and large Bank Holding Companies (BHC) have to consider it when designing stress tests. The most common method of measuring association is the Pearson correlation coefficient:

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

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

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

with

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

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

For this paper, I want to investigate other approaches for measuring the association of different time series and compare the results of those approaches to the standard correlation approach used in finance. My main point of reference has been an article by [3] that explores topics in time series data mining. The authors provide a detailed survey of time series representations, various distance metrics for measures of association, and proper ways to index time series data for efficient querying and information retrieval. The discussions around various distance metrics are very interesting. [3] categorize these distance measures into four categories collected in figure 1. A very similar paper was published in Computational Economics very recently by [5].

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

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

2 Relevant Literature

3 Data

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

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

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

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

Figure 1: List of distance measures included in [3]

good reference point to any latent clustering that is discovered. Three different time periods will be considered. The financial crisis of 2007-2008, the moderate period

between 2014 and 2016, and the dramatic shut down economy of 2019-2020.

I've used the S&P's Wikipedia page¹ to collect the stock ticker symbol for each company and the extra industry and sector information. Using an API made freely available by Alpha Advantage², daily historical series of the current companies in the index have been obtained. Dates range from Nov 1999 to Feb 2021. See Figure 6 for a visual summary. Analyzing the entire S&P adds a level computational complexity to this project, but I think the extra scale helps give the article some heft considering the task at hand isn't very complicated.

4 Distance Measures

4.1 ARMA-GARCH Benchmark

To benchmark the shape based measures, we will design a procedure 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.

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)

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

²<https://www.alphavantage.co/>

- 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 a divide and conquer approach. 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. A more detail set of instructions on the modeling strategy can be found in section 5.1 of the appendix.

The second challenge

For model validation, I will follow the same process as in section 4.3 of [2], checking that each model is well specified and the model residuals are tested to see if they are independently and identically distributed

5 Appendix

5.1 Model Estimation Strategy

This section summarizes the modeling strategy for the historical spot series in the S&P 500. Define the return as $X_t = \log(S_t) - \log(S_{t-1})$, where S_t is the spot price at time t . The conditional mean for the log-returns can be formulated by the following ARMA process:

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

Correlation: 2019-02-01 to 2020-02-14

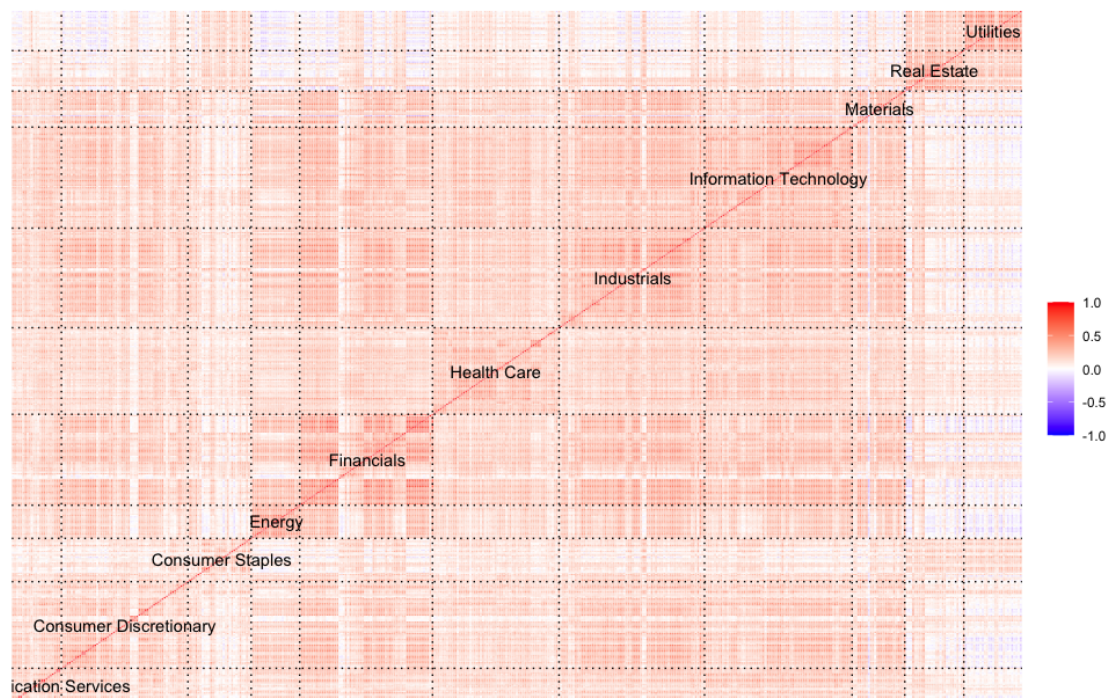


Figure 2:

Correlation: 2020-03-23 to 2021-02-19

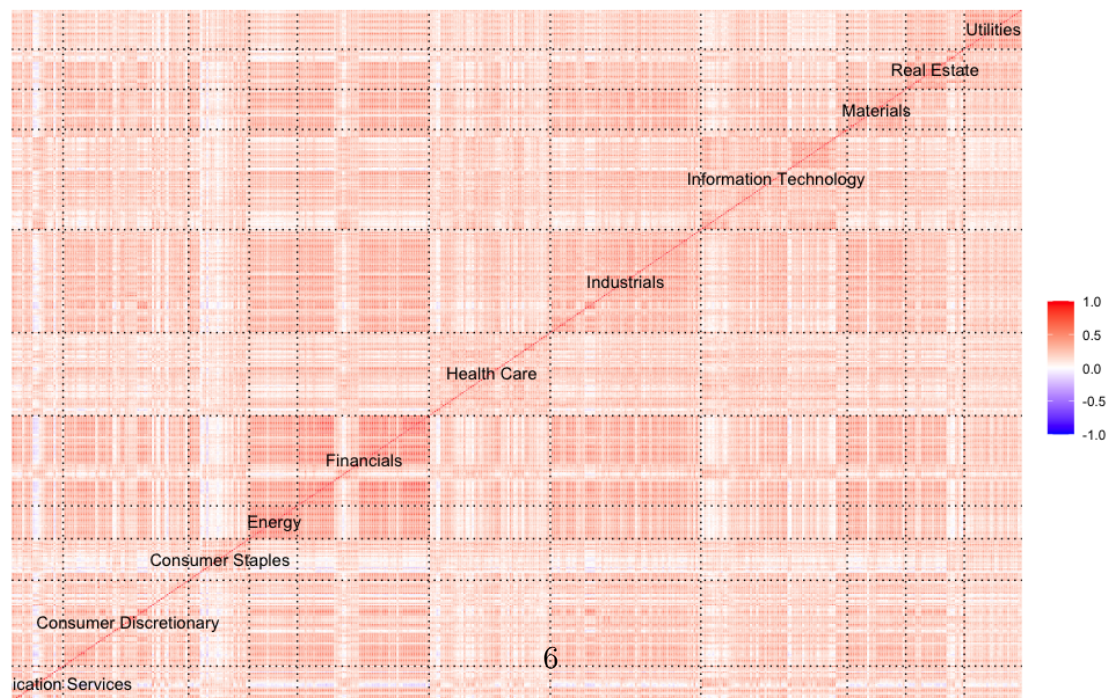


Figure 3:

DTW: 2019-02-01 to 2020-02-14

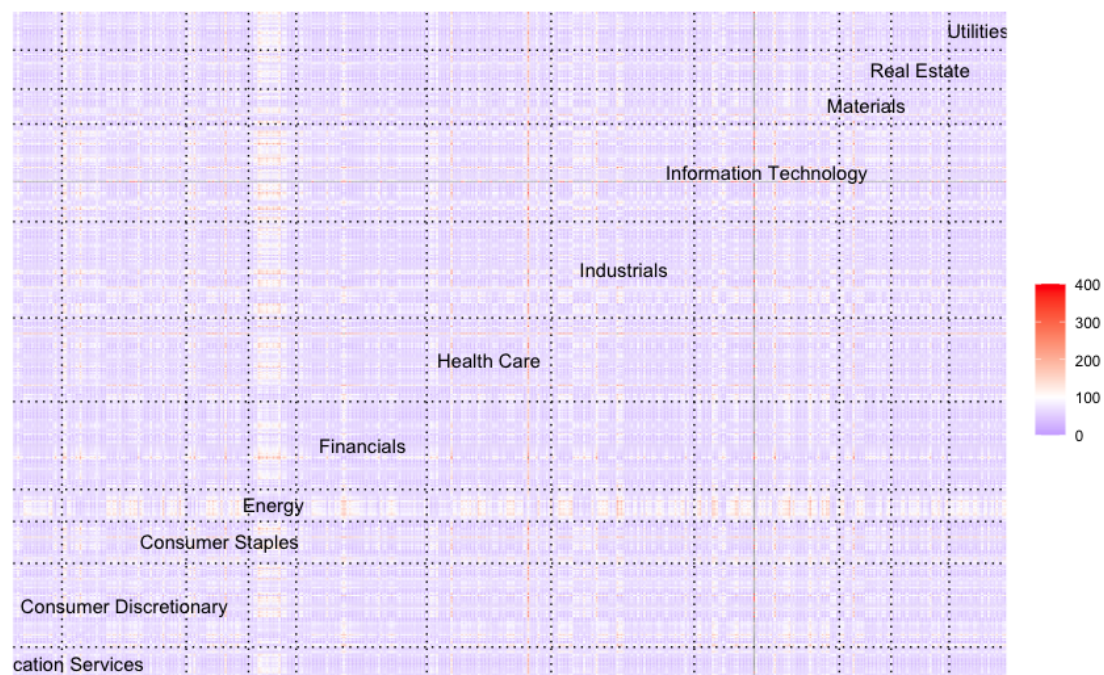


Figure 4:

DTW: 2020-03-23 to 2021-02-19

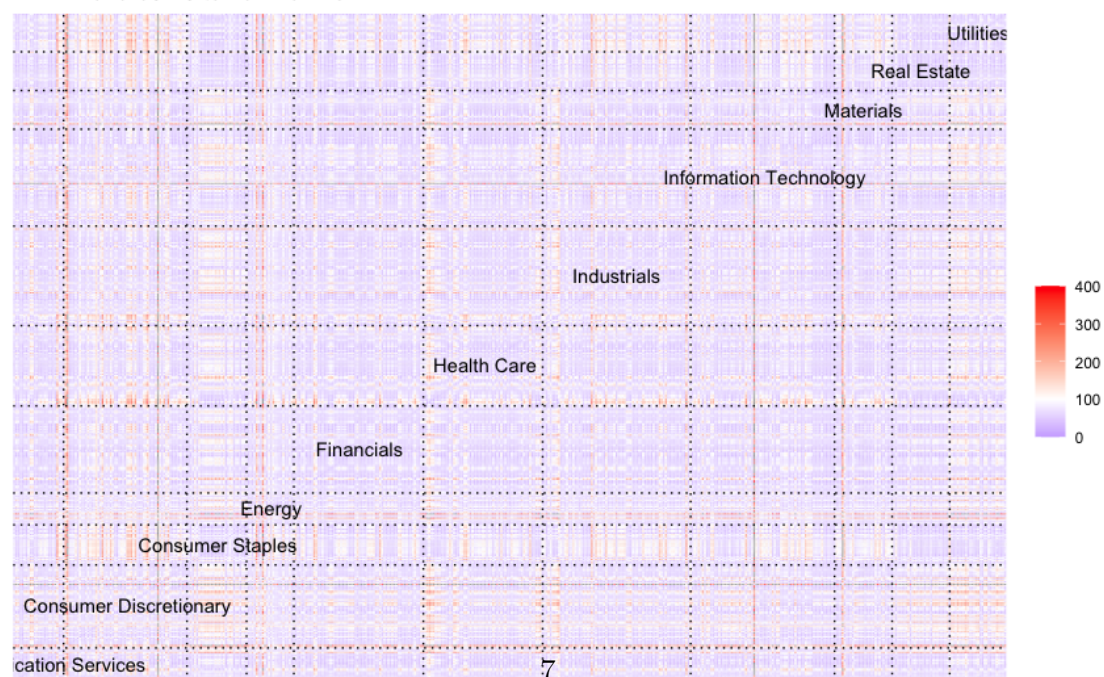


Figure 5:

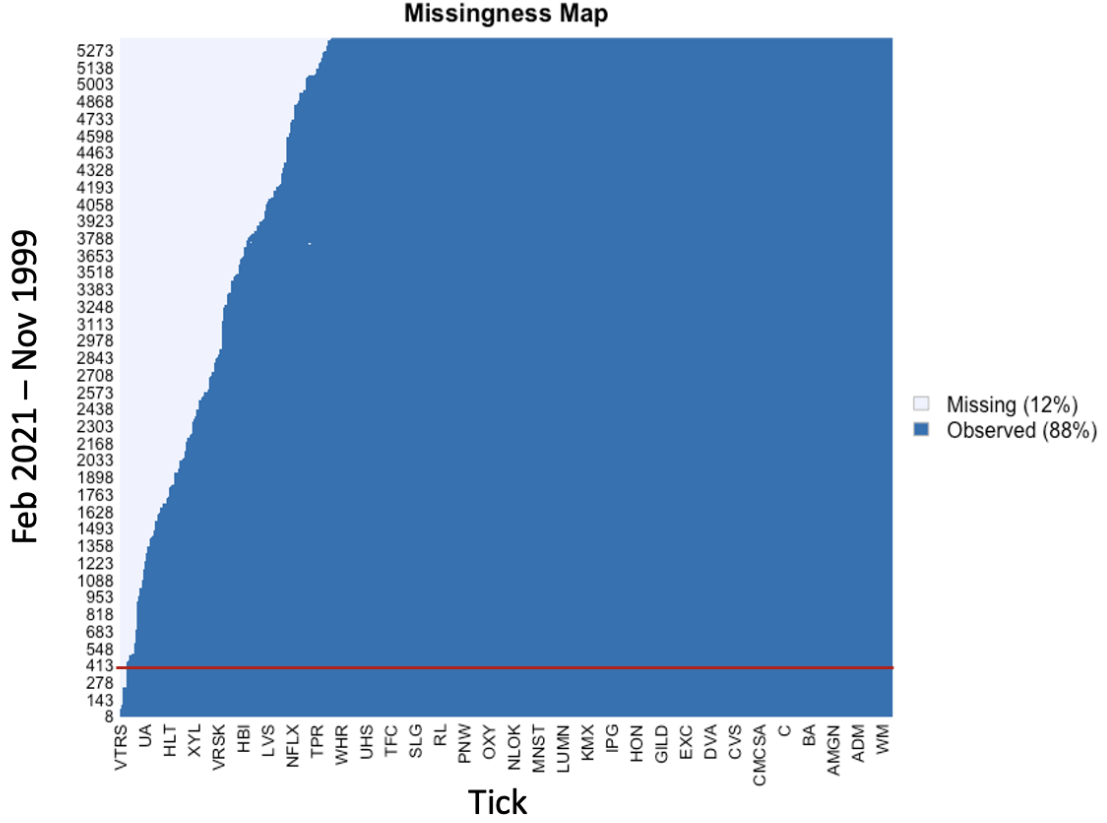


Figure 6: Missing map of the S&P data set. The red horizontal bar demarks the minimum amount of history needed to be included in the study. X of the 505 tickers have enough history

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

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

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

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

$$z_t(\phi, \theta, \alpha, \beta) = \frac{X_t - \mu(X_{t-1}, \phi, \theta)}{\sigma(X_{t-1}, \alpha, \beta)} \quad (7)$$

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

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

With the stage set, the sequence of steps used to automate the model estimation process for the S&P log-returns can be summarized as a three-step process:

1. **Estimate the conditional mean independently of the variance model.** Leveraging the work done by Hyndman and Khandakar [9], 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 [10] running in the R statistical language [15]. 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 [6]. The model set that is searched through considers the following dimensions:
 - Always include a constant term

- ARCH specification: $m = \{1, 2\}$
- GARCH specification: $k = \{1, 2\}$
- Distributions: $g = \{\text{Normal, Student-t, Skewed Student-t [4]}\}$
- Model Specification: $A, B, C = \{\text{Standard GARCH [1], gjr-GARCH [7], Component GARCH [12]}\}$

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 [8] non-parametric density test jointly for i.i.d and correct distribution specification, (d) Shapiro-Wilks [16] test for normality, and (e) Jarque-Bera [11] 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.

5.2 Dynamic Time Warping

Dynamic time warping (DTW) is an alternative method for comparing the association between two discrete time series. It differs from Pearson's correlation measure in that the time indices between the two series at moments of comparison are not constrained to equal each other – like in Equation (1). For the following definitions, this article borrows heavily from [14]. Suppose there are two time series: X_t for $t \in [1 : T]$ and Y_s for $s \in [1 : S]$. A *warping path* is a sequence $\mathbf{p} = [p_1, \dots, p_L]$ where each element is a mapping from the indices of one series to the

other: $p_l = (t_l, w(t_l)) \in [1 : T] \times [1 : S]$ for $l \in [1 : L]$. All warping paths are subject to the following two conditions:

- Boundary Condition: $w(t=1) = 1$ and $w(t=T) = S$
- Step-Size Condition: $\rho_l - \rho_{l-1} \in \{(1, 0), (0, 1), (1, 1)\}$

The boundary condition ensures that the first and last indices of the two series are mapped to each other while the step-size condition ensures a non-decreasing monotonicity in the indices of *both* series. With these conditions, DTW is designed to compare the global shape of the two series rather than the co-movements of the values that share the same time index. The aim of the DTW algorithm is to find the optimal warping path \mathbf{p}^* that minimizes the sum of local cost function, c .

$$\mathbf{p}^* = \arg \min_w \sum_l^L c(X_{t_l}, Y_{w(t_l)}) \quad (9)$$

5.3 Kullback-Leibler Divergence

The model estimation strategy this article takes is ultimately a maximum likelihood one. As discussed in Section 5.1, every ARIMA-GARCH model is estimated by searching for the parameter vector Δ that maximizes the log likelihood for a given distribution g with parameter vector λ

$$\hat{\Delta}, \hat{\lambda} = \arg \max_{\Delta, \lambda} L(\Delta, \lambda | X) \quad (10)$$

with

$$L(\Delta, \lambda | Z) = \prod_{t=1}^T g(z_t | \lambda) \quad (11)$$

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

$$D_{KL}(P||Q) = \int_{-\infty}^{\infty} p(x) \log \left(\frac{p(x)}{q(x)} \right) dx \quad (12)$$

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

$$SKL(P, Q) = \frac{1}{2} (D_{KL}(P||Q) + D_{KL}(Q||P)) \quad (13)$$

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

References

- Bollerslev, Tim (1986). “Generalized autoregressive conditional heteroskedasticity”. In: *Journal of Econometrics* 31.3, pp. 307–327. ISSN: 0304-4076. DOI: [https://doi.org/10.1016/0304-4076\(86\)90063-1](https://doi.org/10.1016/0304-4076(86)90063-1). URL: <https://www.sciencedirect.com/science/article/pii/0304407686900631>.
- Dowiak, L. C. (2018). “A Comparison of Time-Varying Copula-GARCH Models”. In: *Unpublished* 00, p. 00.
- Esling, P. and C. Agon (2012). “Time-Series Data Mining”. In: *ACM Comput. Surv.* 45, pp. 79–82.
- Fernandez, Carmen and Mark Steel (1998). “On Bayesian Modeling of Fat Tails and Skewness”. In: *Journal of The American Statistical Association* 93, pp. 359–371. DOI: 10.1080/01621459.1998.10474117.
- Franses, P. H. and T. Wiemann (2020). “Intertemporal Similarity of Economic Time Series: An Application of Dynamic Time Warping”. In: *Computational Economics* 56, pp. 59–75.
- Ghalanos, Alexios (2020). *rugarch: Univariate GARCH models*. R package version 1.4-4.

- Glosten, Lawrence R, Ravi Jagannathan, and David E Runkle (1993). “On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks”. In: *Journal of Finance* 48.5, pp. 1779–1801. URL: <https://EconPapers.repec.org/RePEc:bla:jfinan:v:48:y:1993:i:5:p:1779-1801>.
- Hong, Yongmiao and Haitao Li (2005). “Nonparametric Specification Testing for Continuous-Time Models with Applications to Term Structure of Interest Rates”. In: *The Review of Financial Studies* 18.1, pp. 37–84. ISSN: 08939454, 14657368. URL: <http://www.jstor.org/stable/3598067>.
- Hyndman, Rob J. and Yeasmin Khandakar (2008). “Automatic Time Series Forecasting: The forecast Package for R”. In: *Journal of Statistical Software, Articles* 27.3, pp. 1–22. ISSN: 1548-7660. DOI: 10.18637/jss.v027.i03. URL: <https://www.jstatsoft.org/v027/i03>.
- Hyndman, Rob et al. (2020). *forecast: Forecasting functions for time series and linear models*. R package version 8.13. URL: <https://pkg.robjhyndman.com/forecast/>.
- Jarque, Carlos M. and Anil K. Bera (1980). “Efficient tests for normality, homoscedasticity and serial independence of regression residuals”. In: *Economics Letters* 6.3, pp. 255–259. ISSN: 0165-1765. DOI: [https://doi.org/10.1016/0165-1765\(80\)90024-5](https://doi.org/10.1016/0165-1765(80)90024-5). URL: <https://www.sciencedirect.com/science/article/pii/0165176580900245>.
- Lee, Gary G. J. and R. Engle (1993). “A Permanent and Transitory Component Model of Stock Return Volatility”. In:
- Markowitz, Harry (1952). “Portfolio Selection”. In: *The Journal of Finance* 7.1, pp. 77–91. ISSN: 00221082, 15406261. URL: <http://www.jstor.org/stable/2975974>.
- Müller, Meinard (2007). “Information Retrieval for Music and Motion”. In: Springer, Berlin, Heidelberg. Chap. 4, pp. 69–84.
- R Core Team (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. URL: <https://www.R-project.org/>.
- SHAPIRO, S. S. and M. B. WILK (1965). “An analysis of variance test for normality (complete samples)”. In: *Biometrika* 52.3-4, pp. 591–611. ISSN: 0006-3444. DOI: 10.1093/biomet/52.3-4.591. eprint: <https://academic.oup.com/biomet/article-pdf/52/3-4/591/962907/52-3-4-591.pdf>. URL: <https://doi.org/10.1093/biomet/52.3-4.591>.
- Sharpe, William F (1963). “A simplified model for portfolio analysis”. In: *Management science* 9.2, pp. 277–293.