International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. 10, No.1 (Feb-2021)
http://dx.doi.org/10.12785/ijcds/100124

Portfolio Management: A Financial Application of Unsupervised Shape-based Clustering-Driven Machine Learning Method

Tristan, Lim*[1] and Chin Sin, Ong[2]

[1] *School of Business Management, Nanyang Polytechnic, Singapore*
[2]*DBS Bank, Singapore*

Received 29 May 2020, Revised 13 Jul. 2020, Accepted 31 Jul. 2020, Published 8 Feb. 2021

Abstract: To diversify, investors should avoid adding assets to their portfolio when their prices exhibit high correlation. Industry diversification is a common portfolio diversification method. It is likely that the notion of investing across different industries can help achieve portfolio diversification, as companies in different industries are likely to have different revenue and cost drivers. However, results across various studies have been mixed. This study seeks to identify a novel application to diversify portfolios to overcome the mixed results of industry diversification, through the use of unsupervised time series clustering-based machine learning technique. There are various ways of clustering time-series, namely shape-based, model-based and feature-based. Feature-based approach faces limitation for the need of equal-length feature vectors, and model-based approach faces limitation in terms of scalability. In this research, a shape-based clustering approach, which overcomes the aforementioned limitations, and specifically agglomerative hierarchical clustering algorithm (AHC-DTW), with dynamic time warping as the distance measure, is utilized to perform diversification. AHC-DTW allows clustering to be conducted across different temporal lengths, many-to-one point comparison to measure distances rather than one-to-one point comparison for euclidean distance. Further, AHC-DTW remains robust with scaling and shifting, unlike for instance, euclidean approach which requires clustering of the same time length, and is highly sensitive to outliers, noise, and transformations. The shape-based clustering approach implemented seeks to match the shapes of time series data as closely as possible. Since shape-based clustering technique groups together cumulative stock returns that trends closely across time, it will be intuitive that investors investing in more than one stock in the same cluster will not be better off, in contrast to diversifying investments across different clusters. Research found clear outperformance of shaped-based cluster diversification against industry diversification. Annualized mean return improved by 598 basis points, and Sharpe performance measure improved by 337%. Further, research found that AHC-DTW clustering exhibited time persistency. These robust results suggest promise for industry practitioners in the utilization of shape-based cluster diversification for enhanced investment portfolio performance.

Keywords: Shape-based Clustering; Agglomerative Hierarchical Clustering; Dynamic Time Warping; Invest; Portfolio Management; Portfolio Allocation and Rebalancing; Diversification.

1. INTRODUCTION

Diversification of investments occurs when investors do not want to put all their eggs in one basket, and henceforth, seek to spread the risk of their investment portfolio across different investment assets. In this way, even as certain assets in the portfolio perform poorly, the overall portfolio would generally perform better over time.

In order to best achieve this, Markowitz [1] proposed to avoid investing in securities that tend to move together over time. Investors should avoid adding assets to their portfolio when their prices exhibit high correlation.

Companies in different industries are likely to have different revenue and cost drivers. Those in the same industry would typically face similar industry risk and exhibit similar financial metrics, such as profit margins, debt ratios, and so forth. It is likely that the notion of investing across different industries can help achieve portfolio diversification.

While industry diversification may be useful to generate better portfolio returns [2] [3] [4], several papers have provided mix performances of the applicability of the industry diversification approach [5] [6]. Business models of companies within the same industry can differ and change over time [7]. In addition, businesses can and

*E-mail: tristan_lim@nyp.edu.sg, tris02@gmail.com

are increasingly operating across geographical regions. Changes in business models or geographical coverages are typically not reflected as a change in industry classification. In practice, these result in estimation risks due to risk exposures that span beyond the typical industry group [8] [9].

In this study, we examine stocks listed in the Singapore Stock Exchange (SGX). In Figure 1, we plot the correlation of stocks listed in Singapore against the market benchmark index, or the Straits Times Index (STI), segmented by industry classifications. For box plot visualization, industries with less than five stocks are excluded.

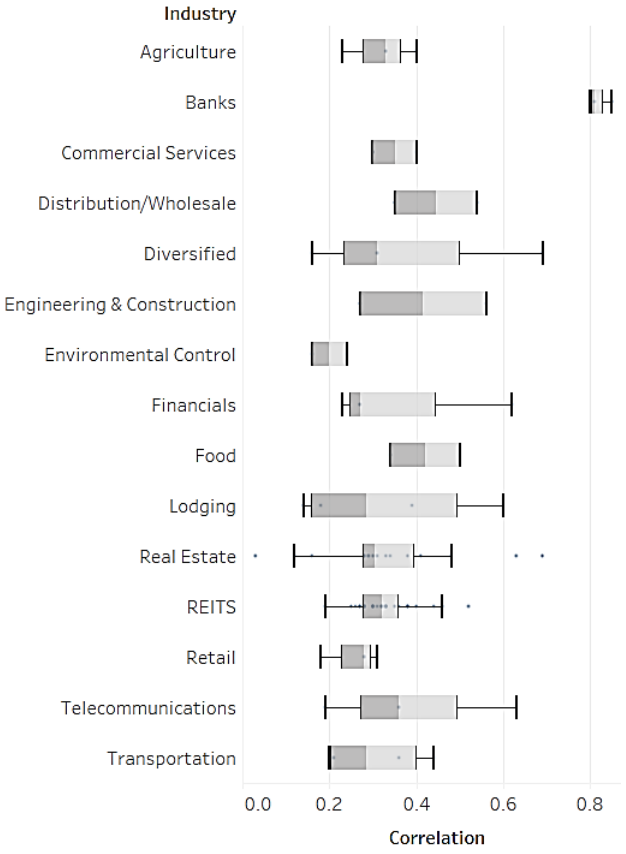

Figure 1. Intra-Industry Correlation Spread

It is interesting to note that within each industry classification, the spread of correlations against the market index are wide, to the extent there exist overlaps of correlation ranges of stocks across industries.

This appears to give insights towards the potential lack of diversification when industry classification is to be used as a diversification tool. Narrow and diversified industry correlation spread ranges should likely lead to increased portfolio diversification when industry diversification is utilized. Conversely the observed wider spreads will likely reduce the usefulness of using industry grouping as a diversification criteria.

With mixed results across various studies on the performance of industry diversification, and the poor spreads of correlation across industries (as observed in Figure 1), this research asks the following question: Can the use of machine learning overcome the shortcoming of industry diversification, to achieve better portfolio performance?

In particular, the study seeks to identify a novel financial application of unsupervised time series clustering-based machine learning method, for portfolio diversification purposes.

2. LITERATURE REVIEW

Clustering is a data mining technique where related or homogeneous sets of similar data are grouped together without the prior knowledge of each group's members [10].

There are various ways of clustering time-series, namely shape-based, model-based and feature-based. In shape-based approach, the shapes of temporal sequenced data are as closely matched as possible, using conventional clustering methods such as k-means clustering or agglomerative hierarchical clustering [11] [12], with appropriate distance measures, such as dynamic time warping, euclidean distance, or triangle similarity measure [13] [14] [15], among others. In model-based approach, raw temporal sequenced data are transformed into model parameters, typically a parametric model for each temporal sequence, and clustering analysis is achieved through conventional clustering algorithms and appropriate distance measures [16]. In feature-based approach, feature vectors, converted from raw temporal sequenced data, are clustered via conventional clustering algorithms and appropriate distance measures.

This study utilizes the shape-based clustering approach, as the latter two approaches suffer from limitations. Feature-based approach typically requires euclidean distance measure as equal length feature vectors are computed from temporal data sequences [17], whereas stock prices may have differing temporal lengths, which may be due to the launch of companies' initial public offerings in the midst of the period of selection. Model-based approaches have limitations in terms of scalability issues [18], and its performance declines when clusters are close to one another [19].

The study specifically utilizes agglomerative hierarchical clustering algorithm, with dynamic time warping (DTW) as the distance measure (collectively hereon referred to as AHC-DTW). DTW allows clustering to be conducted across different temporal lengths, many-to-one point comparison to measure distances rather than

one-to-one point comparison for euclidean distance. Further, DTW remains robust with scaling and shifting [20], unlike for instance, euclidean approach which requires clustering of the same time length, and is highly sensitive to outliers, noise, and transformations [21].

Since shape-based clustering approach groups together stock price series that trend closely across time, it will be intuitive that investors investing in stocks in the same shape-based cluster will not be better off, as compared to diversifying their investments across different shaped-based clusters.

Shaped-based clustering have been applied in financial markets in various contexts. Huang et. al. [22] used a shape-based clustering algorithm known as extended visualization-induced self-organizing map algorithm to find and train the detection of repeatable temporal patterns, and showed outperformance against momentum-based trading strategy [23]. Dose and Cincotti [24] applied shaped-based clustering in index tracking and found robust forecasting applications. Posch et. al. [25] found usefulness in the application of shaped-based clustering in identifying structural changes in financial markets or portfolios. Other papers utilized shape-based clustering and found success at pattern recognition and predictive trading strategies [26] [27] [28] [29] [30]. Literature reviews have found that application of shape-based clustering in financial markets have focused on performing pattern recognition and predictive trading. However, there are a general lack of literature from the context of asset diversification and portfolio construction.

This study aims to study if shape-based clustering technique, and specifically AHC-DTW, is useful to achieve diversification benefits for portfolios, as compared to industry diversification. In particular, the paper looks at the Singapore equity market, and studies how clusters of similar trending stocks may be formed. Research also studies if there are persistence of shape-based clustering observed over time. The findings will be a useful addition to the literature in portfolio management and data science, by providing a data mining portfolio diversification alternative to portfolio management.

3. METHODOLOGY

A. AHC-DTW Clustering

AHC-DTW clustering technique was applied to examine time-series clustering of stocks, for the period 2015-2017, and sub-periods 2015, 2016 and 2017. We mirrored the clustering processes used in [31] [32] [33], which describe how AHC-DTW can be performed.

We performed separate sub-period tests to check for anomalies in similarity results trending across the three-year period. Research excludes cluster sizes with one or two stocks; focus will be on clusters with three stocks and above to achieve observationally and relationally meaningful discussions for portfolio construction.

B. Simulation of Back-test Performance

To test the performance of AHC-DTW cluster-diversified portfolio, we performed a portfolio simulation of 10,000 3-asset portfolio for the period under research. We tested the performance of AHC-DTW cluster-diversified portfolio, against industry-diversified portfolio.

Industry diversification: To compute the performance of industry-diversified portfolios, we randomly simulated 10,000 portfolios of single assets. For each portfolio, we performed a search for the remaining two assets, such that the final three assets in the portfolio are to be of different industry classifications, and the overall portfolio's Sharpe performance measure is the most optimal.

Shape-based cluster diversification: To compute the performance of AHC-DTW-diversified portfolios, we similarly randomly simulated 10,000 portfolios of single assets. For each portfolio, we performed a search for the remaining two assets, such that the final three assets in the portfolio are to be of different AHC-DTW clusters, and the overall portfolio's Sharpe performance measure is the most optimal.

After the performance of portfolio simulation, statistical tests were applied. We performed an assessment of normality using D'Agostino-Pearson omnibus test, followed by one-way ANOVA or Kruskal–Wallis tests to test for statistical significance of the difference of the distribution means, depending on the normality results. For parametric distributions, Tukey Honestly Significant Difference (HSD) test was applied to identify differences in mean returns; Dunn-Bonferroni test was applied for non-parametric distributions [34] [35] [36].

C. Data and Analytical Tool

This research uses the top 82 stocks listed in SGX, with STI as a benchmark market index for the Singapore equity market.

From Bloomberg, we extract the daily closing stock prices between 2015 to 2017. Cumulative geometric return was computed for the full period under review, and the sub-periods 2015, 2016 and 2017. Industry classifications of each stock under investigation were also extracted from Bloomberg.

The following formulae were used to compute the geometric daily return (r_t) (1) and annualized variance (σ_t^2) (2) where d represents the number of trading days, ω represents the weights of asset allocation. Covariance is represented by cov or σ in (3) between the returns of asset m and n. For a random variable X,

$$r_t \triangleq ln\, R_t = ln \frac{p_t}{p_{t-1}} \in \mathbb{R} \quad (1)$$

$$Var[X] \triangleq \sigma_t^2 = \frac{1}{T}\Sigma_{t=1}^{T}(r_t - \bar{r})^2 \cdot d \in \mathbb{R}, \text{ and } Var[X] \triangleq \sigma_t^2 = \omega_t^T\, \Sigma\, \omega_t \in \mathbb{R},$$

$$\text{where } \Sigma = \begin{bmatrix} \sigma_{1,1} & \cdots & \sigma_{1,N} \\ \vdots & \ddots & \vdots \\ \sigma_{N,1} & \cdots & \sigma_{N,N} \end{bmatrix} \quad (2)$$

$$Cov[r_m, r_n] \triangleq \sigma_{mn} = \mathrm{E}[(r_m - \mathrm{E}[r_m])(r_n - \mathrm{E}[r_n])] \in \mathbb{R} \quad (3)$$

Sharpe ratio, an evaluative return performance metric, is used to compute the excess return over risk free rate, for each percentage of risk borne by the investor [37]. Sharpe metric is computed using (4). An annualized risk-free rate (r_f) of 2% is used, approximating the short-term annual interest returns of Singapore Government Securities.

$$\Delta S_{(t-1):t} \triangleq \left(\frac{S_t}{S_{t-1}} - 1\right) \forall\, S_t = \left(\sqrt{T}\frac{r_t - r_f}{\sqrt{\mathrm{Var}[r_{(t-1):t}]}}\right) \in \mathbb{R} \quad (4)$$

The analytical tool and scripting languages used in this research are SAS® Enterprise Miner™ version 14.1 (EM), Python and R.

4. RESULTS AND ANALYSIS

A. DTW Cluster Constellation Plot

Relationships between the 20 clusters generated are visualized using the cluster constellation plot in Figure 2. The plot shows two clear constellation splits, which we rename North-East (NE) and South-West (SW) clusters. Breakage between NE and SW clusters signifies dissimilarities in stock price trends between the two sets of clusters. The positions of each cluster in the constellation plot also provide clues to the dissimilar price movements between each cluster across time.

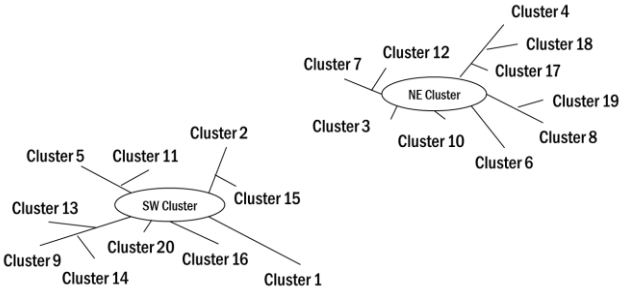

Figure 2. AHC-DTW cluster constellation plot

Of the 20 clusters created, there exist 13 clusters each with three stocks and above: Clusters 3, 4, 6, 7, 8, 10 and 12 exist in the NE cluster; and clusters 1, 2, 5, 9, 11 and 13 exist in the SW cluster. In this research, we exclude Clusters 14 to 20 from our analyses as they each have two or less stocks, which will not add further value to our research objective to establish meaningful stock clusters.

B. Cluster Similarity Score

To find out if trend similarities hold during sub-periods 2015, 2016, and 2017, cluster similarity scores are computed, as shown in Table I.

These similarity scores are computed based on the likelihood of whether stocks within the same 2015 to 2017 supercluster, will be clustering in the same cluster for the sub-periods (i) 2015 (1-year period), (ii) 2015 to 2016 (2-year period), and (iii) 2015 to 2017 (3-year period), weighted by the number of stocks in each cluster. Further, in circumstances where clear segmentations are possible within each 2015 to 2017 supercluster with two or more stocks, each cluster is sub-divided into sub-clusters.

Results positively show trend continuation within sub-periods. For DTW clusters generated for the period 2015-2017, 91% of the clusters are similar in 2015, followed by 83% and 67% for 2016 and 2017 respectively.

Robust similarity scores across the three years of observation indicates strong industry and research application value. There exists a lack of significant similarity decay of clusters across time. In approximation, 9 of 10, 8 of 10, and 7 of 10 stocks display similar clustering trends across the one, two and three-year sub-periods under investigation respectively.

TABLE I. CLUSTER SIMILARITY SCORE

2015 - 2017			**2015**	**2016**	**2017**
Cluster No.	***Sub-cluster No.***	***No. of Stocks***	***% Similarity – 1 year***	***% Similarity – 2 years***	***% Similarity – 3 years***
1	1A	6	100%	100%	83%
	1B	6	83%	83%	67%
2	2A	6	100%	83%	67%
	2B	4	100%	50%	50%
3	3A	4	100%	100%	100%
	3B	4	100%	100%	50%
	3C	4	75%	75%	75%
4	4A	6	100%	83%	50%
	4B	4	100%	100%	75%
5	5A	2	100%	100%	100%
	5B	6	83%	83%	67%
6	-	5	80%	80%	60%
7	-	6	100%	100%	67%
8	-	3	100%	100%	100%
9	-	4	75%	75%	50%
10	-	3	100%	67%	67%
11	-	3	67%	33%	33%
12	-	3	67%	33%	33%
13	-	3	100%	100%	100%
Weighted Average:			**91%**	**83%**	**67%**

C. Clustering Analysis

Cluster summary in Table II shows (i) the cluster-dominant industry, and (ii) the stocks that are present in

each cluster but not explained by the dominant industry (despite holding similar price movements).

TABLE II. CLUSTER SUMMARY

Cluster	Cluster-Dominant Industry (Bloomberg Ticker Code)	Stocks present in cluster but not explained by dominant industry (Bloomberg Ticker Code)
NE 4A	*Telecommunications, Media & Technology* 1. SPH SP Equity [+0.5] 2. ST SP Equity [+0.63] 3. STH SP Equity [+0.36]	*Unexplained Random Industry Stock(s)* 1. SCI SP Equity [+0.65] 2. HPL SP Equity [+0.14] 3. SIE SP Equity [+0.3]
NE 4B	None (*Diversified Mid-Large Capitalization (or Cap) Stocks*)	*Unexplained Random Industry Stock(s)* 1. WIL SP Equity [+0.5] 2. STE SP Equity [+0.56] 3. CAPL SP Equity [+0.69] 4. UIC SP Equity [+0.29]
NE 8	*Financial Services* 1. UOB SP Equity [+0.81] 2. DBS SP Equity [+0.8] 3. GE SP Equity [+0.29]	
NE 6	None (*Diversified Mid-Large Cap Stocks*)	*Unexplained Random Industry Stock(s)* 1. OCBC SP Equity [+0.85] 2. CIT SP Equity [+0.6] 3. HPAR SP Equity [+0.31] 4. VMS SP Equity [+0.26] 5. SIA SP Equity [+0.43]
NE 3A	*Commodity* 1. FR SP Equity [+0.33] 2. OLAM SP Equity [+0.34]	*Property* 1. GUOL SP Equity [+0.28] 2. OUE SP Equity [+0.41]
NE 3B	*Property* 1. HOBEE SP Equity [+0.3] 2. FCT SP Equity [+0.27] 3. CT SP Equity [+0.4]	*Transportation* 1. SBUS SP Equity [+0.2]
NE 3C	*Property* 1. MINT SP Equity [+0.28] 2. WINGT SP Equity [+0.48] 3. FPL SP Equity [+0.33] 4. SUN SP Equity [+0.52]	
NE 7	*Property* 1. STRTR SP Equity [+0.17] 2. UEM SP Equity [+0.38] 3. AREIT SP Equity [+0.44] 4. PREIT SP Equity [+0.19]	*Unexplained Random Industry Stock(s)* 1. HLF SP Equity [+0.27] 2. FNN SP Equity [+0.22]
NE 10	None (*Diversified Mid Cap Stocks*)	*Unexplained Random Industry Stock(s)* 1. SGX SP Equity [+0.62] 2. UOL SP Equity [+0.63] 3. KEP SP Equity [+0.69]
NE 12	None (*Diversified Mid Cap Stocks*)	*Unexplained Random Industry Stock(s)* 1. CD SP Equity [+0.44] 2. M1 SP Equity [+0.19] 3. SMM SP Equity [+0.51]
SW 2A	*Property* 1. AIT SP Equity [+0.16] 2. KDCREIT SP Equity [+0.3] 3. MAGIC SP Equity [+0.32] 4. SPHREIT SP Equity [+0.32]	*Unexplained Random Industry Stock(s)* 1. GENS SP Equity [+0.5] 2. METRO SP Equity [+0.31]
SW 2B	*Property* 1. ART SP Equity [+0.36] 2. MLT SP Equity [+0.35] 3. KREIT SP Equity [+0.38] 4. FSG SP Equity [+0.12]	
SW 1A	*Property* 1. FCOT SP Equity [+0.33] 2. FIRT SP Equity [+0.28] 3. GRAN SP Equity [+0.18] 4. CDREIT SP Equity [+0.33] 5. AAREIT SP Equity [+0.25]	*Financial Services* 1. UOBK SP Equity [+0.23]
SW 1B	*Property* 1. CCT SP Equity [+0.46] 2. MCT SP Equity [+0.33] 3. CRCT SP Equity [+0.3]	*Unexplained Random Industry Stock(s)* 1. RFMD SP Equity [+0.23] 2. KPTT SP Equity [+0.21] 3. SPOST SP Equity [+0.36]
SW 9	*Utilities & Infrastructure* 1. HPHT SP Equity [+0.4] 2. KIT SP Equity [+0.16]	*Property* 1. EREIT SP Equity [+0.28] 2. SML SP Equity [+0.28]
SW 13	None *(Diversified Small Cap Stocks)*	*Unexplained Random Industry Stock(s)* 1. GGR SP Equity [+0.4] 2. OHL SP Equity [+0.31] 3. PAC SP Equity [+0.16]
SW 5A	*Property* 1. PREH SP Equity [+0.29] 2. GLL SP Equity [+0.39]	
SW 5B	*Property* 1. FEHT SP Equity [+0.26] 2. FHT SP Equity [+0.31] 3. OUEHT SP Equity [+0.3] 4. SGREIT SP Equity [+0.38]	*Unexplained Random Industry Stock(s)* 1. CEL SP Equity [+0.16] 2. SILV SP Equity [+0.25]
SW 11	None *(Diversified Small Cap Stocks)*	*Unexplained Random Industry Stock(s)* 1. SSG SP Equity [+0.28] 2. RSTON SP Equity [+0.16] 3. ASCHT SP Equity [+0.27]

[ ] represents pairwise correlation against market index (STI Index).

Further insights can be gleaned from the cluster stock analysis:

(A) Clustering Between Non-Industry Related Stocks

Our results show that stocks from the same cluster may not come from the same industry.

Cluster 4A (Figure 3) comprises Technology, Media and Telecommunications (TMT) cluster companies Singapore Press Holdings Ltd (SPH SP Equity), Singapore Telecommunications Ltd (ST SP Equity) and Starhub Ltd (STH SP Equity). Interestingly, Sembcorp Industries Ltd (SCI SP Equity) of diversified industries, Hotel Properties Ltd (HPL SP Equity) of the hospitality industry, and SIA Engineering Co Ltd (SIE SP Equity) of the aviation industry share similar trends with the TMT

sector stocks. It is likely the idiosyncratic risk and performance profile of the latter three stocks coincided with the TMT sectorial risk and performance profile. Investing in the latter 3 stocks alongside the former TMT stocks will likely not improve the diversification effects of a stock portfolio.

Within cluster 4A, the intra-cluster correlation spread is wide. Against STI market index, stocks in cluster 4A (N = 6) exhibit average correlation of 0.43 and standard deviation of 0.20, with observations ranging from 0.14 to 0.65. AHC-DTW cluster diagram visually exhibits a general downward trend.

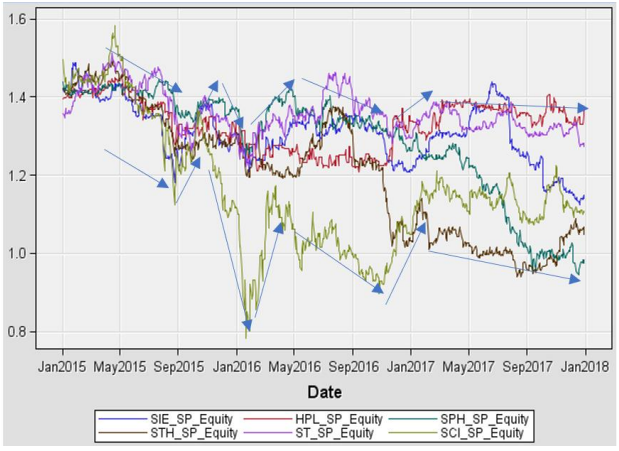

Figure 3. Time-Series Plot - Cluster 4A

Cluster 4B (Figure 4) comprises mid-to-large capitalization stocks Wilmar International Ltd (WIL SP Equity) from the commodity industry, Singapore Technologies Engineering Ltd (STE SP Equity) from the engineering industry, Capitaland Ltd (CAPL SP Equity), a property developer, and United Industrial Corp Ltd (UIC SP Equity), a relatively smaller capitalization property developer.

Similarly, it is likely the idiosyncratic risk and performance profile of the four stocks coincided during the time period under observation. Investing in more than one stock in the seemingly unrelated, albeit mid-to-large capitalization 4B cluster will likely not improve the diversification effects of a stock portfolio.

Within cluster 4B, the intra-cluster correlation spread is again wide. Against STI market index, stocks in cluster 4B (N = 4) exhibit average correlation of 0.51 and standard deviation of 0.17, with observations ranging from 0.29 to 0.69. AHC-DTW cluster diagram visually exhibits a general sideways trend.

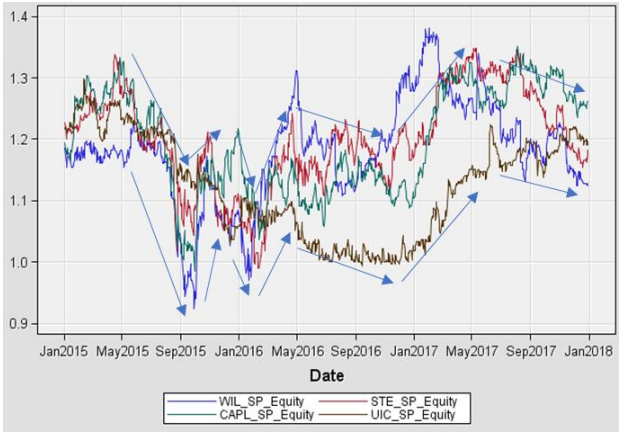

Figure 4. Time-Series Plot - Cluster 4B

(B) Clustering Between Industry Related Stocks

Cluster 8 (Figure 5) comprises financial services stocks, including United Overseas Bank Ltd (UOB SP Equity), DBS Bank Group Holdings Ltd (DBS SP Equity) and insurer Great Eastern Holdings Ltd (GE SP Equity). Good time-series correlations are expected for these stocks, as they faced similar industry dynamics.

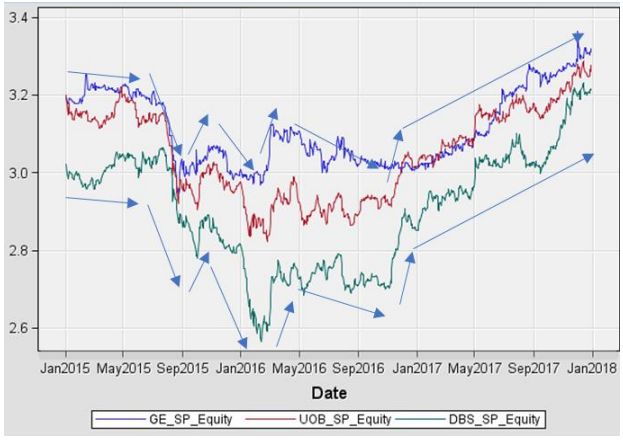

Figure 5. Time-Series Plot - Cluster 8

Within cluster 8, the intra-cluster correlation spread is wide. Against STI market index, stocks in cluster 8 (N = 3) exhibit average correlation of 0.63 and standard deviation of 0.30, with observations ranging from 0.29 to 0.81. AHC-DTW cluster diagram visually demonstrates a general slight-upward trend.

(C) Identification of Exceptional Performers

Strong and weak performers in certain industries can stand out to cluster in separate clusters, allowing star performers or fallen angels to be easily identified.

In our result, the hospitality real estate investment trust (REIT) stocks are mainly located in Cluster 5B (Figure 6). However, a strong REIT performer Ascendas Hospitality Trust (ASCHT SP Equity) is present in Cluster 11 (Figure

7), distinctly separated from its hospitality REIT counterparts.

Similarly, an outlier weak performer, M1 Ltd (M1 SP Equity), a telecommunications company was found in Cluster 12 (Figure 8), distinctly separate from its TMT counterparts in Cluster 4A (Figure 3).

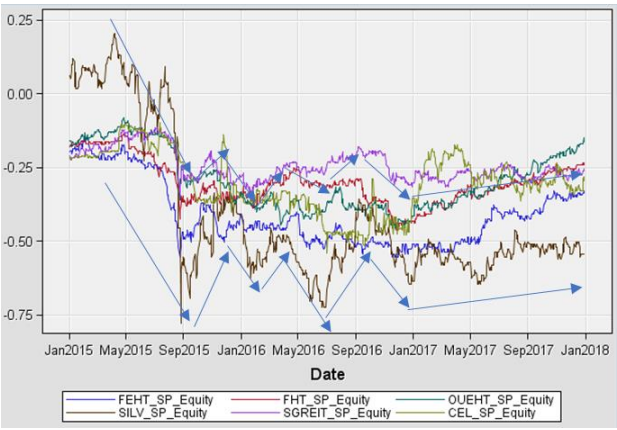

Figure 6. Time-Series Plot - Cluster 5B

Figure 7. Time-Series Plot - Cluster 11

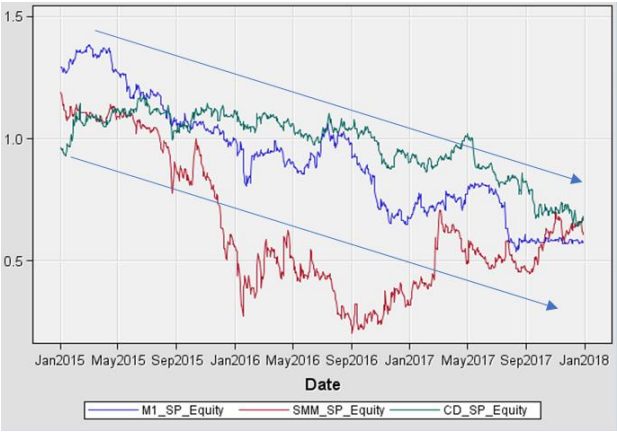

Figure 8. Time-Series Plot - Cluster 12

D. Back-test Performance Analysis

Results from Table III demonstrated that both mean return and Sharpe statistic of the two 10,000 portfolio distributions exhibited non-normality. We applied Kruskal-Wallis non-parametric test, and observed in Table IV that statistically significant difference exists in mean return and Sharpe statistic between the industry- and AHC-DTW-diversified portfolios. In Table V, Dunn-Bonferroni ad-hoc test results confirmed the difference between all test groups at 1% level of significance.

TABLE III. D'AGOSTINO-PERSON OMNIBUS NORMALITY TEST

p-value	**Return**	**Sharpe**
Industry-diversified	0*	3.86E-108*
AHC-DTW-diversified	0*	0*

* represents statistical significance at 1% level.

TABLE IV. KRUSKAL-WALLIS NON-PARAMETRIC TEST

	Test statistic	**p-value**
Return	1874.02	0*
Sharpe	1818.11	0*

* represents statistical significance at 1% level.

TABLE V. DUNN-BONFERRONI NON-PARAMETRIC AD-HOC TEST

p-value	**Industry-diversified**	**AHC-DTW-diversified**
Industry-diversified	-1	4.05E-21*
AHC-DTW-diversified	4.05E-21*	-1

* represents statistical significance at 1% level.

TABLE VI. RETURN AND SHARPE PERFORMANCE

%	**Industry-diversified**	**AHC-DTW-diversified**
Mean Return	3.28 *182**	9.24
Mean Sharpe	7.78 *337**	33.98

* represents percentage improvement (%) of AHC-DTW diversification.

Analysis yielded interesting insights. On an annualized basis, mean return of AHC-DTW-diversified portfolios outperformed industry-diversified portfolios by 596 basis points. This was a performance improvement of 1.82 times.

In addition, DTW-diversified portfolios outperformed industry-diversified portfolio by 337% on a risk-adjusted basis based on Sharpe ratio, providing 337% higher return as compared to industry-diversified portfolios for each percentage of portfolio risk borne by investors.

Robust results demonstrate the usefulness of AHC-DTW-diversification.

5. Conclusion and Future Work

In this study, we performed shape-based clustering, or specifically AHC-DTW clustering, and found clear outperformance against industry-diversification in portfolio performance of 10,000 simulated portfolios.

On an annualized basis, the mean return of AHC-DTW diversified portfolio outperformed industry diversified portfolio by 596 basis points, or a 1.82 times improvement. Even with a haircut of 50% [37], outperformance was 298 basis points – a significant improvement of 0.91 times.

Aside from identifying outlier stocks performing well above or below industry groups, performing shape-based clustering also allowed the identification of similar trending stocks that are not typically seen as closely related in nature, but do somehow cluster together. This may likely be due to coincidence, or structural changes occurring in the stocks' underlying business models, resulting in similar stock return profiles during time periods under observation. As investors, shape-based clustering can highlight these factors above, and guide investment due diligence towards, for instance, outlier performers and allocation of stocks across different shape-based cluster groups to reduce portfolio risk and optimize portfolio return.

Another interesting find is how stocks that underwent shape-based clustering remained in similar cluster groups across time. Robust similarity scores across the three years of observation indicates strong industry and research application value. There exists a lack of significant similarity decay of clusters across time. In approximation, 9 of 10, 8 of 10, and 7 of 10 stocks display similar clustering trends across the one, two and three-year sub-periods under investigation respectively. This persistency points towards the possibility of creating predictive portfolio construction or rebalancing via shaped-based clustering diversification.

For future work, it will also be interesting to perform this study across different financial markets and asset classes, to validate the promise that this study provides.

References

[1] H. Markowitz, "Portfolio Selection," The Journal of Finance, vol. 7, no. 1, p. 77, 1952.

[2] F. Balli, S. A. Basher, and R. J. Louis, "Sectoral equity returns and portfolio diversification opportunities across the GCC region," Journal of International Financial Markets, Institutions and Money, vol. 25, pp. 33–48, 2013.

[3] S. Cavaglia, C. Brightman, and M. Aked, "The Increasing Importance of Industry Factors," Financial Analysts Journal, vol. 56, no. 5, pp. 41–54, 2000.

[4] S. Varotto, "The Causes of International Diversification in the Stock and Eurobond Markets," SSRN Electronic Journal, 2006.

[5] B. S. Kopp, "Conglomerates in Portfolio Management," Financial Analysts Journal, vol. 24, no. 2, pp. 145–148, 1968.

[6] H. Benjelloun, "Portfolio diversification: evidence from Qatar and the UAE," International Journal of Economic Issues, vol. 2, pp. 9–15, 2009.

[7] M. Emiris, "Sectoral vs. Country Diversification Benefits and Downside Risk," SSRN Electronic Journal, 2004.

[8] Fidelity Learning Center, "The limitations of sector classification systems," Markets & Sectors. [Online]. Available: https://www.fidelity.com/learning-center/tradinginvesting/markets-sectors/limitations-sector-classification-systems.

[9] Franklin Templeton Investments, "Concentrated Yet Diversified: A Distinctive Approach to Global Investing," Oct. 2016.

[10] P. Rai, S. Singh. A survey of clustering techniques. Int. J. Comput. Appl., vol. 7, no. 12, pp. 1-5, 2010.

[11] M. Vlachos, J. Lin, E. Keogh. A wavelet-based anytime algorithm for k-means clustering of time series. Proc. Work. Clust, pp. 23-30, 2003.

[12] R.H.R. Shumway. Time-frequency clustering and discriminant analysis. Stat. Probab. Lett, vol. 63, no. 3, pp. 307-314, 2003.

[13] H. Sakoe, S. Chiba. A dynamic programming approach to continuous speech recognition, Proceedings of the Seventh International Congress on Acoustics, vol. 3, pp. 65-69, 1971.

[14] C. Faloutsos, M. Ranganathan, Y. ManolopoulosFast subsequence matching in time-series databases. ACM SIGMOD Rec., vol. 23, no. 2, pp. 419-429, 1994.

[15] X. Zhang, J. Wu, X. Yang, H. Ou, T. LvA notime series classificationvel pattern extraction method for Optim. Eng., vol. 10, no. 2, pp. 253-271, 2009.

[16] T. Warrenliao. Clustering of time series data—a survey. Pattern Recognit., vol. 38, no. 11, pp. 1857-1874, 2005.

[17] V. Hautamaki, P. Nykanen, P. Franti, Time-series clustering by approximate prototypes, in: Proceedings of 19th International Conference on Pattern Recognition, 2008, ICPR 2008, no. D, pp. 1–4, 2008.

[18] M. Vlachos, D. Gunopulos, G. Das"Indexing time-series under conditions of noise," M. Last, A. Kandel, H. Bunke (Eds.), Data Mining in Time Series Databases, World Scientific, Singapore, p. 67, 2004.

[19] T. Mitsa. Temporal Data Mining, vol. 33, Chapman & Hall/CRC Taylor and Francis Group, Boca Raton, FL, 2009.

[20] R. Staub, "Asset Allocation vs. Security Selection—Baseball with Pitchers Only?" The Journal of Investing, vol. 15 no. 3, p.35-42, 2006.

[21] G. P. Brinson, B. D. Singer, and G. L. Beebower, "Determinants of portfolio performance II: An update", Financial Analysts Journal, vol. 47, no. 3, pp. 40-48, 1991.

[22] Y. P. Huang, C. Hsu, and S.. Wang. "Pattern recognition in time series database: A case study on financial database." Expert Systems with Applications, vol. 33, no. 1, pp. 199-205, 2007.

[23] J. Narasimhan, and S. Titman. "Returns to buying winners and selling losers: Implications for stock market efficiency." The Journal of Finance, vol. 48, no. 1, pp. 65-91, 1993.

[24] C. Dose and S. Cincotti. "Clustering of financial time series with application to index and enhanced index tracking portfolio." Physica A: Statistical Mechanics and its Applications 355.1 (2005): 145-151.

[25] P. N. Posch, D. Ullmann, and D. Wied. "Detecting structural changes in large portfolios." Empirical Economics, vol. 56, no. 4, pp. 1341-1357, 2019.

[26] P. Tsinaslanidis and D. Kugiumtzis, "A prediction scheme using perceptually important points and dynamic time warping," Expert Systems with Applications, vol. 41, no. 15, pp. 6848–6860, 2014.

[27] K. Lee, S. Jun, Jeong, and S. Jae, "Trading Strategies based on Pattern Recognition in Stock Futures Market using Dynamic Time Warping Algorithm," Journal of Convergence Information Technology, vol. 7, no. 10, pp. 185–196, 2012.

[28] G.-J. Wang, C. Xie, F. Han, and B. Sun, "Similarity measure and topology evolution of foreign exchange markets using dynamic time warping method: Evidence from minimal spanning tree," Physica A: Statistical Mechanics and its Applications, vol. 391, no. 16, pp. 4136–4146, 2012.

[29] S. Śmiech, "Co-movement of Commodity Prices - Results from Dynamic Time Warping Classification," Zeszyty Naukowe Uniwersytetu Ekonomicznego w Krakowie, no. 4(940), pp. 117–130, 2015.

[30] S. H. Kim, H. S. Lee, H. Ko, S. H. Jeong, H. W. Byun, and K. J. Oh, "Pattern Matching Trading System Based on the Dynamic Time Warping Algorithm," 2018.

[31] R. K.-W. Lee and T. S. Kam, "Time-Series Data Mining in Transportation: A Case Study on Singapore Public Train Commuter Travel Patterns," International Journal of Engineering and Technology, vol. 6, no. 5, pp. 431–438, 2014.

[32] M. Leonard, J. Lee, T. Y. Lee, and B. Elsheimer, "An introduction to similarity analysis using SAS," in Proc. International Symposium of Forecasting, p. 302, 2008.

[33] M. Leonard and B. Wolfe, "Mining transactional and time-series data," in Proc. International Symposium of Forecasting, p. 1-26, 2002.

[34] D. Oztuna, A. H. Elhan, E. Tuccar. "Investigation of four different normality tests in terms of type 1 error rate and power under different distributions." Turkish Journal of Medical Sciences, vol. 36, no. 3, p.171–6, 2006.

[35] M. Blanca, R. Alarcón, J. Arnau, R. Bono, R. Bendayan. "Non-normal data: Is ANOVA still a valid option?" Psicothema, vol. 29, no. 4, p. 552-7, 2017.

[36] A. Dinno, "Nonparametric pairwise multiple comparisons in independent groups using Dunn's test." The Stata Journal, vol.15, no. 1, p. 292-300, 2015.

[37] C. Harvey and Y. Liu, "Backtesting". Journal of Portfolio Management, vol. 42, no. 1, p. 01, 2015.

Tristan Lim is a faculty at the Nanyang Polytechnic, School of Business Management, in Singapore. He is a graduate of the National University of Singapore and The University of Sydney. He specializes in cross-domain research on business, information systems and industrial engineering.

Chin Sin Ong is an Assistant Vice President at DBS Bank in Singapore. He graduated from Nanyang Technological University.