Analysis of Stock Market with Unsupervised Learning
- An Analysis of ETFs In China Stock Market using Dynamic Time Warping

Chen Li, Lim Yong Kai, A/P Kam Tin Seong, Singapore Management University

ABSTRACT

Economic theories are supposed to interpret why and how the economy behaves and what are the best solutions to influence or solve the economic phenomena. However, they are full of assumptions, hypotheses and contexts in terms of moral values and politics. Price movements of stock market have never been well explained by any economic theories. Hence, our question as investors comes: Is there a holistic way to understand the price movements in the market without application of any economic theory? An option would be to use unsupervised learning to detect objective patterns of the subject without the requirement of any domain knowledge. We believe one approach to understand real world complexity is to get the pattern first, followed by forming and studying the theories.

In this paper, we explore over one hundred ETFs in China stock market without prior domain knowledge of each ETF using dynamic time warping clustering (DTW) and agglomerative hierarchical clustering method to sense the similarities of their price movements. Our results show that clusters from DTW method largely coincide with the type of industries that ETFs involve. Analysis of the clusters' price movement also revealed that certain industries performed better after 2019 when compared to 2018 in light of China's new self-reliance economic direction.

INTRODUCTION

Exchange Traded funds (ETF) is a financial security that can track an index, sector, commodity or other asset and it is available for trading on the stock market. ETF is a type of fund that holds a basket of assets rather than only one stock. There are different types of ETFs such as Bond ETF which include government bonds or corporate bonds, Industry ETFs which track a particular industry such as technology or banking, Commodity ETFs such as gold and Currency ETFs that invest in foreign currencies. ETF allows investors to gain access to many stocks across various industries and have better risk management through the diversification in their portfolio. ETFs have become increasingly popular among investors and nowadays they are in the investing mainstream. This resulted in many new ETFs created over the years. It is important for investors to pick the correct ETF to invest in, amidst the abundance of ETFs available in the stock market.

China stock exchange operator consisting of the Shanghai Stock Exchange and Shenzhen Stock Exchange is the world second largest stock exchange operator by market capitalization at 11.4 trillion U.S. dollars behind the United States. Chinese stock presents exceptional and unique future growth in the twenty-first century but at the same time considered to be with high risks of corruption, murky corporate financial statements, shady corporate governance, and complicated opaque government bureaucracy. These risks are thought of as main contributors to make the Chinese stock market extremely volatile. Burton and Taylor (2008, p. 282-284) propose to minimize risks of investing in Chinese stock by investing in funds rather than individual stocks for diversification.

In this paper, we performed cluster analysis on the ETFs listed in the Shanghai and Shenzhen Stock Exchange followed by evaluating the clusters. We perform two Clustering techniques, Hierarchical Clustering and Dynamic Time Warping to aggregate the ETFs into clusters with similar price movement patterns. We seek to aid new investors without extensive knowledge to build their investment portfolio using China ETFs.

LITERATURE REVIEW

Several studies involving Cluster Analysis for ETFs have been published before. One study focused on the performance of ETFs among themselves by using three multivariate methods: K-means Cluster Analysis, Factor Analysis and Chernoff-Faces using Intraday, Year-to-Date, 3-Month, 1-Year and 3-Year returns as the performance measures. (Solis, 2011) Another study performed agglomerative hierarchical clustering using the dissimilarity measure of clusters based on Pearson correlation coefficient using weekly returns in a two-year period. (Isakov, 2019) Both studies utilized traditional clustering algorithm for their cluster analysis. As the stock market is directly influenced by time series changes, time-series clustering technique might produce better clustering results. A study which tested 8 popular clustering methods which includes hierarchical clustering algorithms using Euclidean and Dynamic Time Warping (DTW) methods concluded that there is no one clustering method that performs better than others for all dataset and highlighted the importance of creating a pool of clustering methods to determine which is the most suitable for the respective dataset. (Javed, Lee, & Rizzo, 2020)

METHODOLOGY

DATASET

The dataset use in this project is the daily closing price of 255 ETFs listed on the Shanghai Stock Exchange and 111 ETFs from the Shenzhen Stock Exchange from 2nd January 2018 to 31st December 2020. The dataset is sourced from Choice Financial Terminal.

DATA PREPARATION

The three years daily prices of all 366 ETFs were merged using JMP Pro 16.0.0. Each row represents one ETF and each column represents the daily closing price from 2nd January 2018 to 31st December 2020. As there are new ETFs being created constantly, we excluded 223 ETFs that did not have complete closing price in 2018. Hence, the remaining 143 ETFs with complete daily closing price during the three-year period were used for our study.

New columns were created to calculate the monthly returns of each ETFs. The formula for monthly return calculation is:

$$monthly\ return = \frac{P_1 - P_0}{P_0}$$

where P_1 is the price of an individual ETF at the last day of the month and P_0 is the price at the start of the month. The formula computes the percentage or proportion of the EFT monthly return. If the return yields a positive figure, it signifies that a positive increase in the price of the ETF in percentage terms for the month. Likewise, if the return yields a negative figure, it means that the ETF saw a negative fall in the price in percentage for the month. The resulting dataset consists of 143 ETFs in the rows and 36 monthly returns expressed in percentage in the columns.

Exploratory analysis is performed on the 143 ETFs monthly return data and there is an outlier for ETF 512030.SH where the return for August 2020 is -79.62%. This is because the ETF split their shares into 5 portions, resulting in a negative return of 79% due to correction of the ETF price. The ETF is removed from our analysis as clustering algorithm does not provide good results from extreme outliers. Figure 1 shows the data preparation workflow.

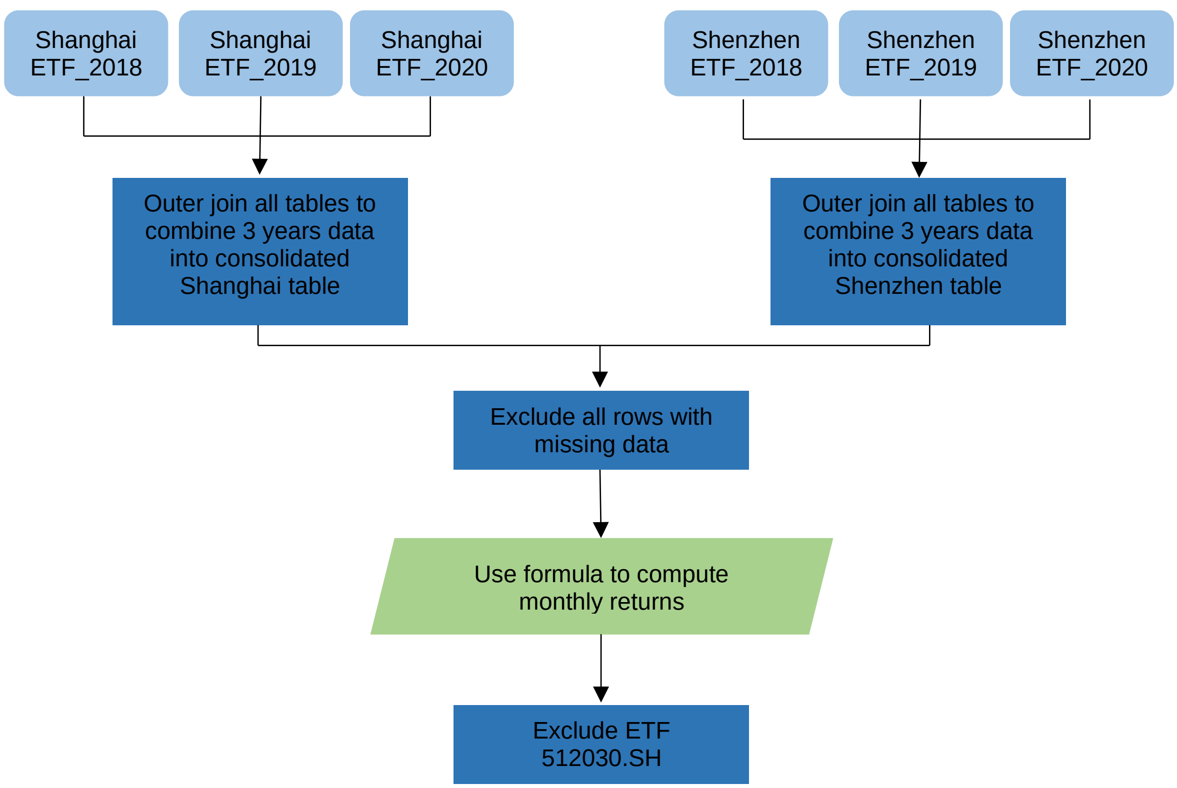


Figure 1. Data Preparation Process

CLUSTERING ANALYSIS

Dynamic Time Warping and hierarchical clustering is performed to investigate the differences in cluster results.

Dynamic Time Warping

Dynamic Time Warping (DTW) technique is used to compare similarity between time series. Instead of directly inferring similarity based on the distance between two data points with the same time reference, DTW algorithm accounts for the time factor when comparing different time-series. Hence, it could factor in propagation delays, and detect time-series that share similar patterns but are out of phase (Kam & Lee, 2014).

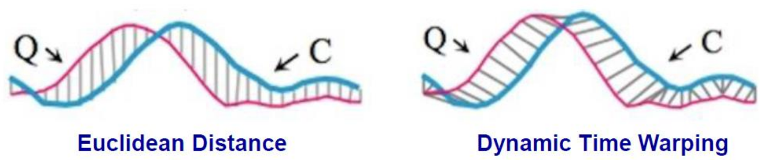


Figure 2. Comparison between Euclidean Distance and DTW Method

SAS Enterprise Miner 14.1 is used as the data mining tool for this study. DTW analysis is performed using the "*TS Similarity*" node which offers a means of identifying similar trends across a set of sequenced parameters which are the monthly returns. Figure 3 shows the SAS Enterprise Miner Workflow, adopting similar process as in Lee & Kam (2014), which covered the Similarity analysis process in detail. The dataset is transposed into time series format where the months are in the row and each ETF is in the columns before importing to accommodate the DTW workflow in SAS Enterprise Miner. The "*Time Interval*" is set to "*Month*" and no normalization of the variables is applied because each data is in the same scale of percentage return.

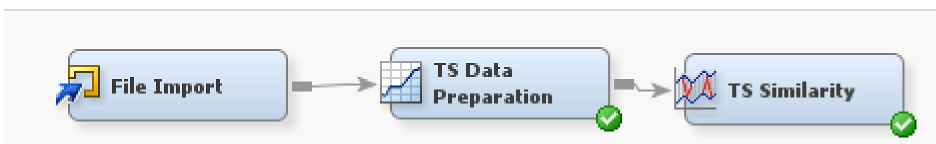
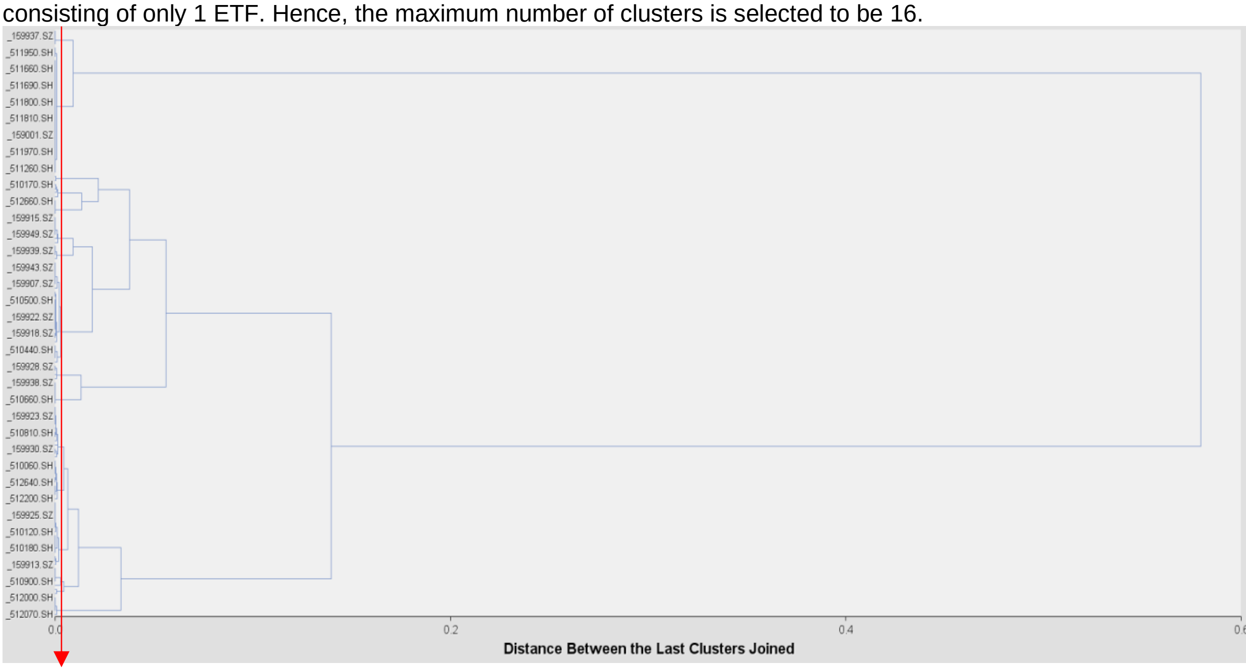


Figure 3. DTW method workflow in SAS Enterprise Miner

The dendrogram is observed and the plausible numbers of clusters are 14, 16 and 18 by drawing a cutoff line based on the dendrogram legs length. It is observed that splitting beyond 16 clusters results in the formation of clusters consisting of only 1 ETF. Hence, the maximum number of clusters is selected to be 16.



Cut-off of 16 clusters

Figure 4. Dendrogram of DTW

Figure 5 shows the multiple time series comparison plot for all ETFs and Table 1 shows the multiple time series comparison plot for each individual clusters.

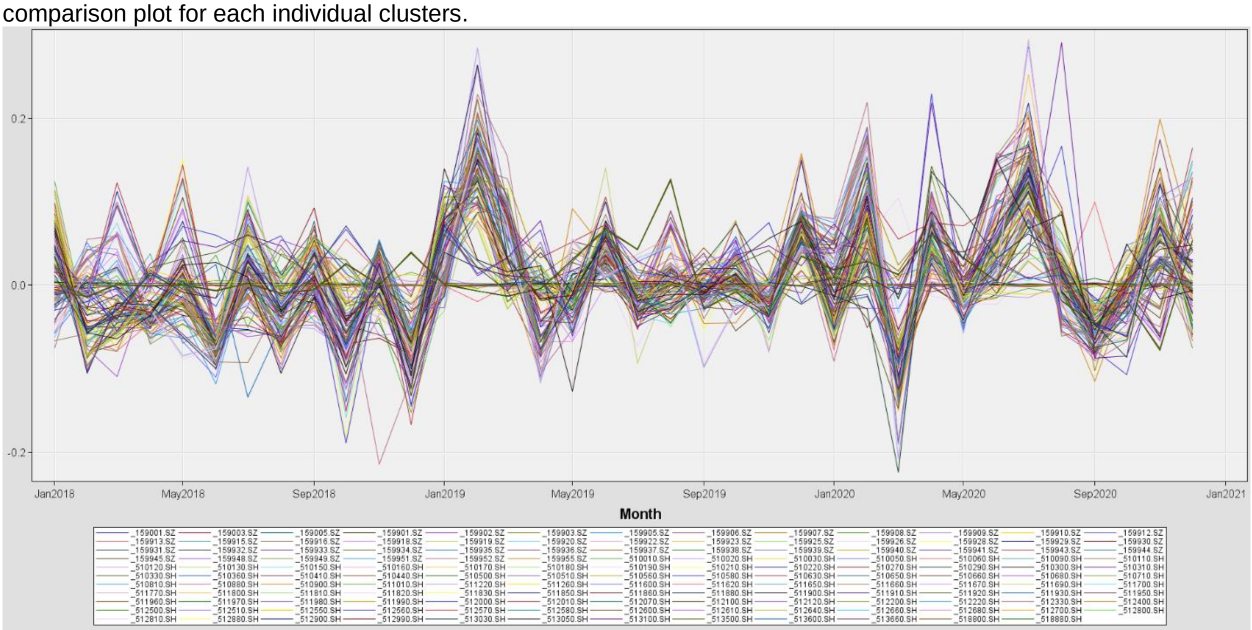
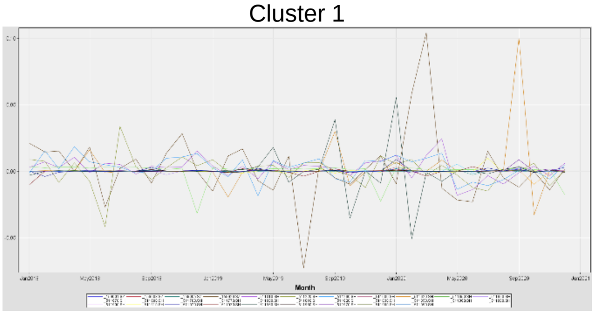
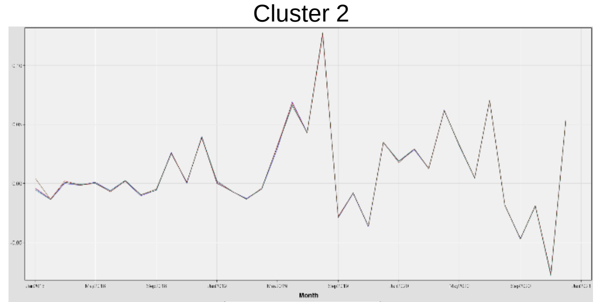
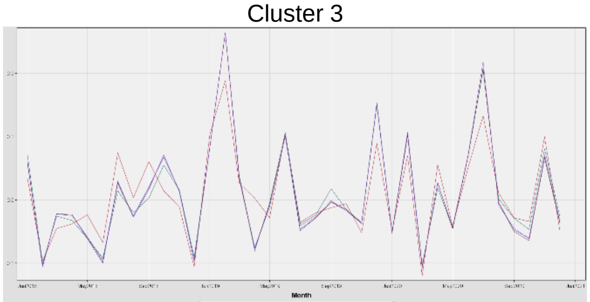
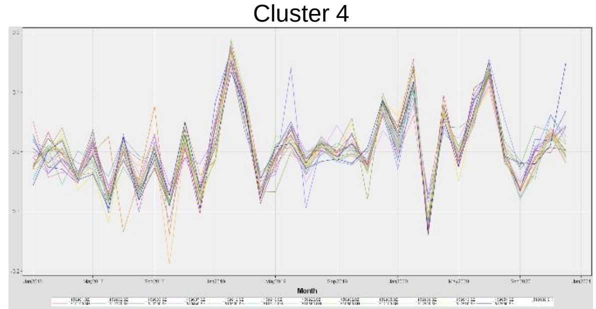
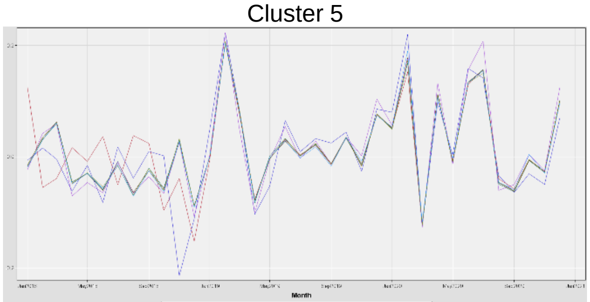
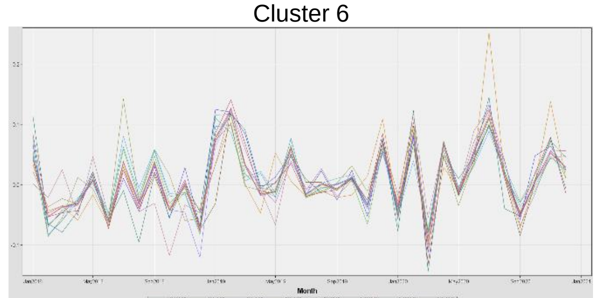


Figure 5. Multiple Time Series Comparison Plot

Cluster 1	Cluster 2
	
Cluster 3	Cluster 4
	
Cluster 5	Cluster 6
	

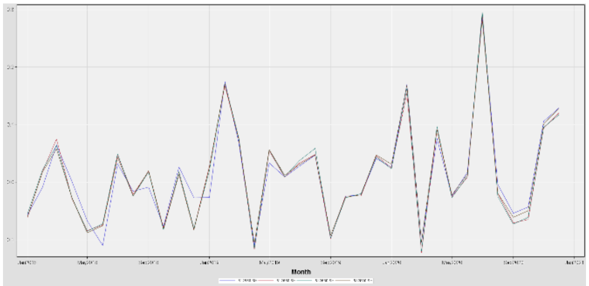
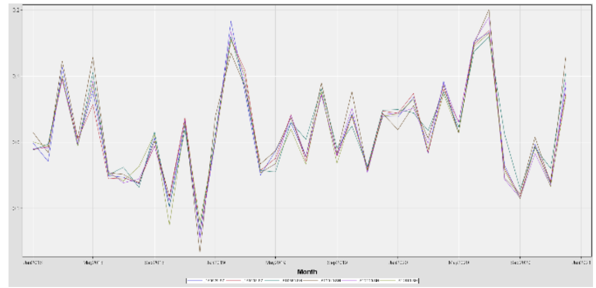
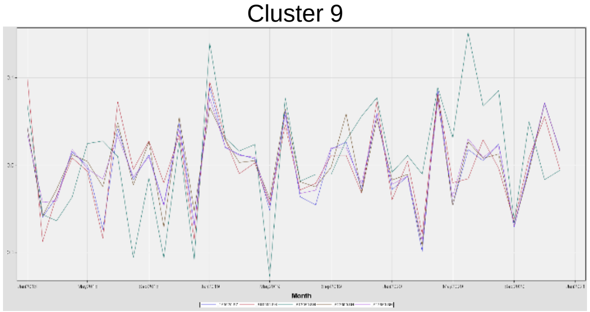
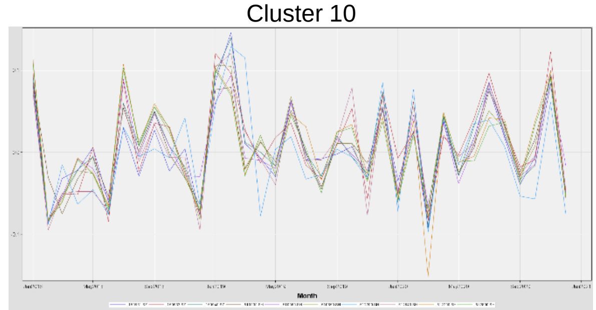
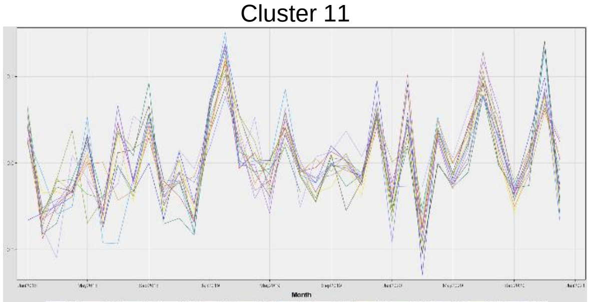
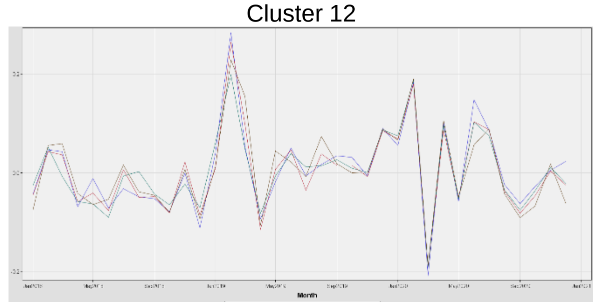
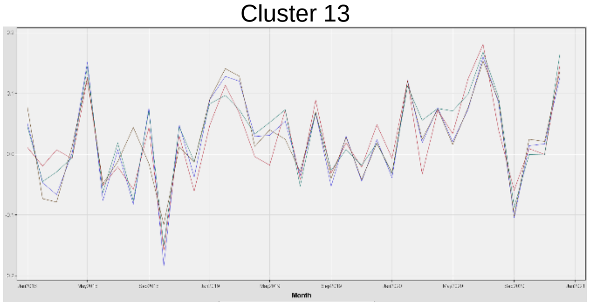
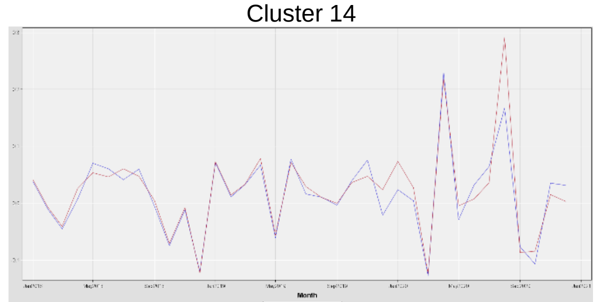
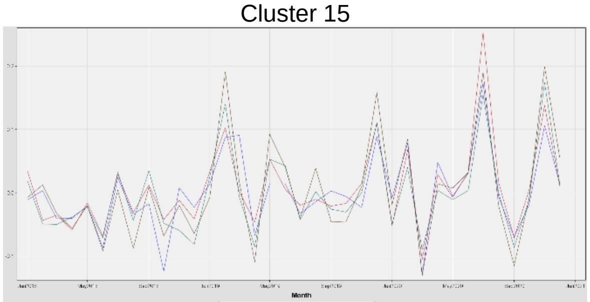
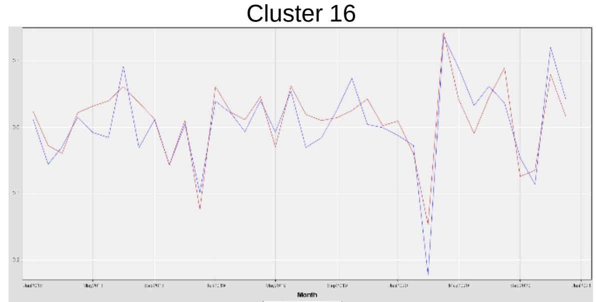
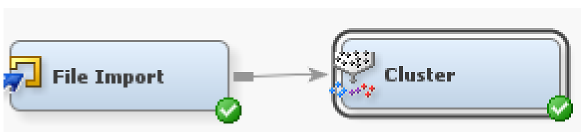
Cluster 7 	Cluster 8 
Cluster 9 	Cluster 10 
Cluster 11 	Cluster 12 
Cluster 13 	Cluster 14 
Cluster 15 	Cluster 16 

Table 1. Multiple Time Series Comparison Plot for each cluster

Hierarchical Clustering

SAS Enterprise Miner 14.1 is used to perform agglomerative hierarchical algorithms for cluster analysis. Agglomerative hierarchical clustering uses a proximity matrix to determine all the pairwise dissimilarities or similarities between each monthly return using the Euclidean distance. No standardization is performed on the variables before clustering. The clustering method uses Ward method which calculates the distance between two clusters using the ANOVA sum of squares between the two clusters summed over all the variables. At each generation, the within-cluster sum of squares is minimized over all partitions obtainable by merging two clusters from a previous generation. The "*Final Maximum*" number of clusters is set to 16 to match the DTW analysis. The "*Ordinal Encoding*" method is set to "*Bathtub*" because the dataset consist of more than one variable and their values are of both positive and negative float. Figure 6 shows the workflow and parameters to perform hierarchical clustering in SAS Enterprise Miner.



Property	Value
General	
Node ID	Clus2
Imported Data	...
Exported Data	...
Notes	...
Train	
Variables	...
Cluster Variable Role	Input
Internal Standardization	None
Number of Clusters	
Specification Method	Automatic
Maximum Number of Clusters	16
Selection Criterion	
Clustering Method	Ward
Preliminary Maximum	50
Minimum	2
Final Maximum	16
CCC Cutoff	3
Encoding of Class Variables	
Ordinal Encoding	Bathtub
Nominal Encoding	GLM
Initial Cluster Seeds	
Seed Initialization Method	Default
Minimum Radius	0.0
Drift During Training	No

Figure 6. Hierarchical Cluster Workflow and Parameters in SAS Enterprise Miner

Clustering Results and Analysis

Dynamic Time Warping Results		
Cluster	**Description**	**ETFs**
1	Government Bond & Monetary Fund ETFs	159001.SZ, 159003.SZ, 159005.SZ, 159926.SZ, 511010.SH, 511220.SH, 511260.SH, 511600.SH, 511620.SH, 511650.SH, 511660.SH, 511670.SH, 511690.SH, 511700.SH, 511770.SH, 511800.SH, 511810.SH, 511820.SH, 511830.SH, 511850.SH, 511860.SH, 511880.SH, 511900.SH, 511910.SH, 511920.SH, 511930.SH, 511950.SH, 511960.SH, 511970.SH, 511980.SH, 511990.SH
2	Gold spot contracts ETFs	159934.SZ, 159937.SZ, 518800.SH, 518880.SH
3	Securities ETFs	512000.SH, 512070.SH, 512570.SH, 512880.SH, 512900.SH
4	ETFs with Manufacturing and Finance companies with varying proportions	510130.SH, 510190.SH, 510220.SH, 159918.SZ, 159936.SZ, 510290.SH, 510440.SH, 510580.SH, 159935.SZ, 512580.SH, 159901.SZ, 159912.SZ, 510560.SH, 512500.SH, 159902.SZ, 159903.SZ, 159907.SZ, 159922.SZ, 159943.SZ, 510500.SH, 510510.SH, 512100.SH, 512510.SH, 159932.SZ, 159951.SZ
5	Technology companies in the Growth Enterprise Market ETFs	159906.SZ, 159908.SZ, 159915.SZ, 159948.SZ, 159949.SZ, 159952.SZ, 159955.SZ
6	ETFs with Manufacturing and Finance companies with varying proportions	159913.SZ, 159919.SZ, 159905.SZ, 159910.SZ, 159925.SZ, 51030 0.SH, 510310.SH, 510330.SH, 510360.SH, 512990.SH, 510020.SH, 510050.SH, 510180.SH, 510710.SH, 512550.SH, 510120.SH
7	Military Manufacturers ETFs	512560.SH, 512660.SH, 512680.SH, 512810.SH
8	Medicine & Medicare Companies ETFs	159929.SZ, 159938.SZ, 510660.SH, 512010.SH, 512120.SH, 512610.SH
9	Chinese companies listed in HK and US Exchange ETFs	159920.SZ, 510900.SH, 513600.SH, 513660.SH, 513050.SH
10	Finance and Real Estate ETFs	159931.SZ, 159933.SZ, 159940.SZ, 510030.SH, 510060.SH, 510650.SH, 512200.SH, 512640.SH, 512700.SH, 512800.SH
11	ETFs with Manufacturing and Finance companies with varying proportions	159930.SZ, 159945.SZ, 510110.SH, 510160.SH, 159916.SZ, 510210.SH, 159923.SZ, 510010.SH, 510090.SH, 510270.SH, 510680.SH, 510810.SH, 510880.SH
12	Technology companies in Mature Enterprise Market ETFs	159909.SZ, 159939.SZ, 512220.SH, 512330.SH
13	Consumer ETFs	159928.SZ, 510150.SH, 510630.SH, 512600.SH
14	NASDAQ ETFs	159941.SZ, 513100.SH
15	Mineral resource companies ETFs	159944.SZ, 510170.SH, 510410.SH, 512400.SH
16	Overseas companies in Standard & Poor and Germany index	513030.SH, 513500.SH

Table 2. Cluster results from DTW

Hierarchical Clustering Results		
Clusters	**Description**	**ETFs**
1	Mineral resource companies ETFs with oil and coal resources companies ETFs	159930.SZ, 159945.SZ, 510110.SH, 510160.SH, 159944.SZ, 510170.SH, 510410.SH, 512400.SH
2	Overseas companies in Standard & Poor and Germany index	513030.SH, 513500.SH
3	Military Manufacturers ETFs	512560.SH, 512660.SH, 512680.SH, 512810.SH
4	Medicine & Medicare Companies ETFs	159929.SZ, 159938.SZ, 510660.SH, 512010.SH, 512120.SH, 512610.SH
5	Cluster with 1 ETF	159906.SZ
6	Cluster with 1 ETF	513050.SH
7	Cluster with 1 ETF	159908.SZ
8	ETFs with Manufacturing and Finance companies with varying proportions	159901.SZ, 159905.SZ, 159910.SZ, 159912.SZ, 159913.SZ, 159916.SZ, 159918.SZ, 159919.SZ, 159925.SZ, 159936.SZ, 510120.SH, 510130.SH, 510190.SH, 510210.SH, 510220.SH, 510290.SH, 510300.SH, 510310.SH, 510330.SH, 510360.SH, 510440.SH, 510560.SH, 510580.SH, 512500.SH, 512990.SH
9	Technology companies in the Growth Enterprise Market ETFs and Mix cluster of ETFs	159902.SZ, 159903.SZ, 159907.SZ, 159915.SZ, 159922.SZ, 159943.SZ, 159948.SZ, 159949.SZ, 159952.SZ, 159955.SZ, 510500.SH, 510510.SH, 512100.SH, 512510.SH, 512580.SH
10	Cluster with 1 ETF	159935.SZ
11	Finance, Real Estate ETFs and mixed cluster of ETFs	159920.SZ, 159923.SZ, 159931.SZ, 159933.SZ, 159940.SZ, 510010.SH, 510020.SH, 510030.SH, 510050.SH, 510060.SH, 510090.SH, 510180.SH, 510270.SH, 510650.SH, 510680.SH, 510710.SH, 510810.SH, 510880.SH, 510900.SH, 512200.SH, 512550.SH, 512640.SH, 512700.SH, 512800.SH, 513600.SH, 513660.SH
12	Securities ETFs	512000.SH, 512070.SH, 512570.SH, 512880.SH, 512900.SH
13	Technology companies in Mature Enterprise Market ETFs	159909.SZ, 159932.SZ, 159939.SZ, 159951.SZ, 512220.SH, 512330.SH
14	Government Bond & Monetary Fund ETFs and Gold spot contract ETFs	159001.SZ, 159003.SZ, 159005.SZ, 159926.SZ, 159934.SZ, 159937.SZ, 511010.SH, 511220.SH, 511260.SH, 511600.SH, 511620.SH, 511650.SH, 511660.SH, 511670.SH, 511690.SH, 511700.SH, 511770.SH, 511800.SH, 511810.SH, 511820.SH, 511830.SH, 511850.SH, 511860.SH, 511880.SH, 511900.SH, 511910.SH, 511920.SH, 511930.SH, 511950.SH, 511960.SH, 511970.SH
15	Consumer ETFs	159928.SZ, 510150.SH, 510630.SH, 512600.SH
16	NASDAQ ETFs	159941.SZ, 513100.SH

Table 3. Cluster results from Hierarchical Clustering

Comparison of Clusters: Case Study 1

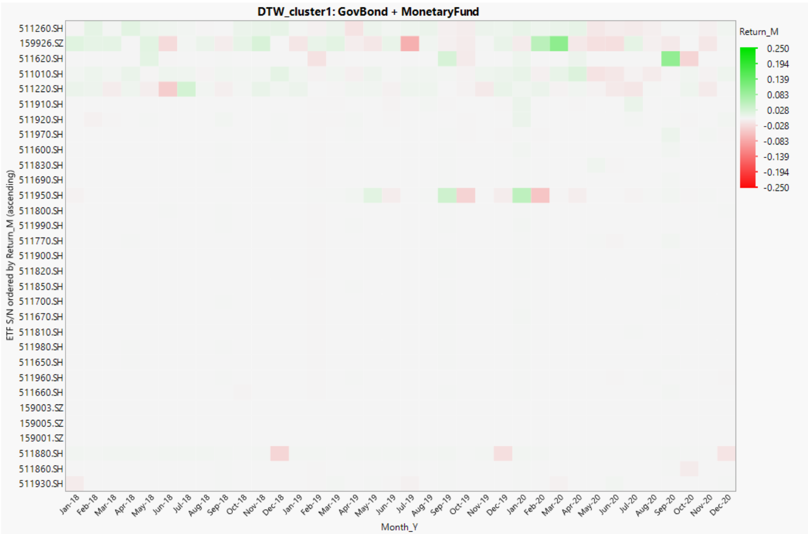


Figure 7. DTW Cluster 1

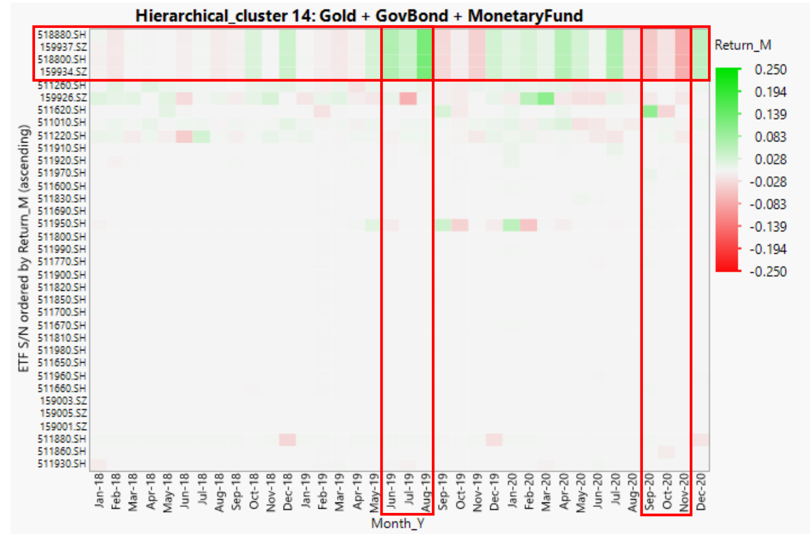


Figure 8. Hierarchical Cluster 14

- DTW algorithm clusters ETFs that are Government Bonds with Monetary Fund into cluster 1 (Figure 7) and Gold spot contact into cluster 2, whereas Hierarchical clustering algorithm clustered all three types of ETFs together into cluster 14 (Figure 8).
- Observing the heat map for Hierarchical cluster 14 in figure 8, the intensity of the first four rows of ETFs (Gold spot contract ETFs) is visually unique in comparison to the other ETFs in the cluster. In June to August 2019 and September to November 2020 (boxed in red), the intensity of returns is distinctly different from the remaining ETFs in the cluster. These show the advantages of DTW algorithm in comparison to hierarchical clustering algorithm in such datasets.

Comparison of Clusters: Case Study 2

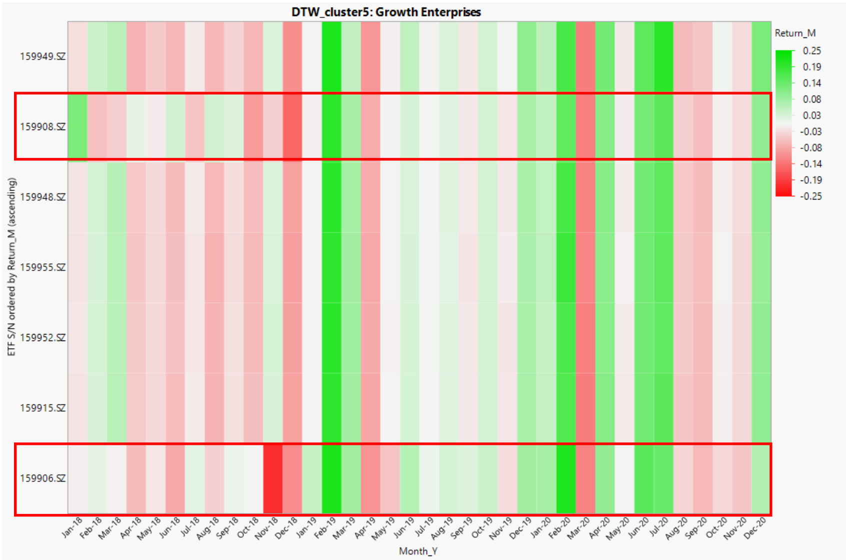


Figure 9. DTW Cluster 5

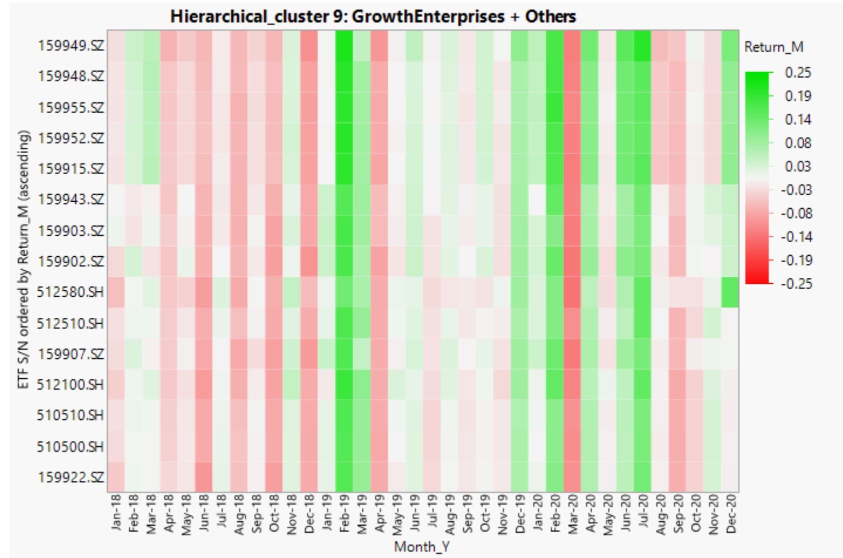


Figure 10. Hierarchical Cluster 9

- DTW algorithm clusters seven Growth enterprises ETFs into cluster 5 (Figure 9) while hierarchical cluster 9 (Figure 10) is a combination of some Growth enterprises ETFs and other ETFs.
- However, hierarchical cluster 9 did not include ETF 159908.SZ and 159906.SZ which is included in DTW cluster 5 (boxed in red). Checking on the two ETF portfolio details, both ETF also belongs to the Growth Enterprises. This exhibits merit of the DTW algorithm in comparison to hierarchical clustering algorithm.

Comparison of Clusters: Case Study 3

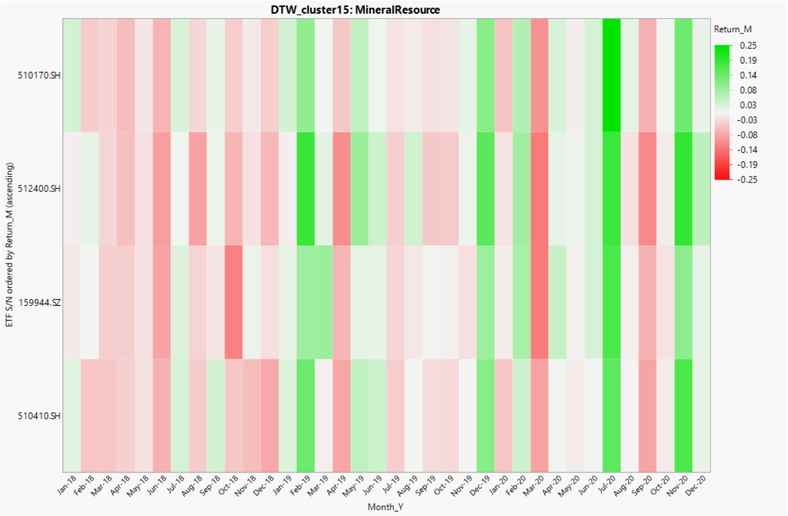


Figure 11. DTW Cluster 15

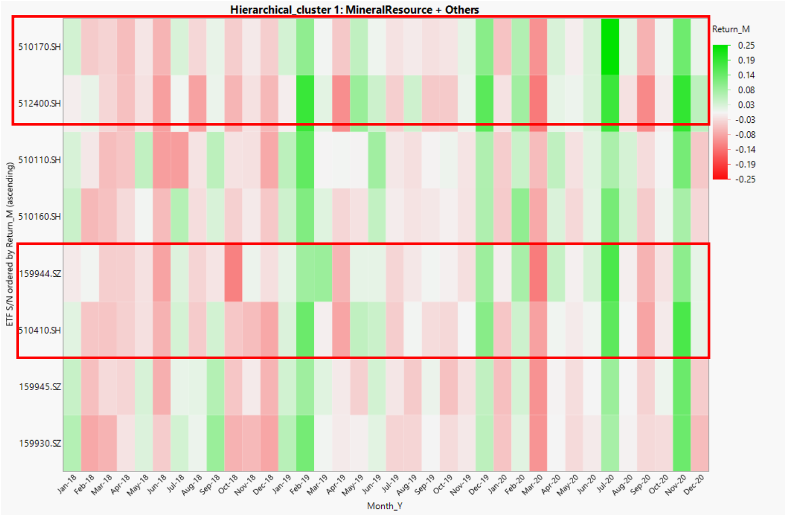


Figure 12. Hierarchical Cluster 1

- DTW cluster 15 (Figure 11) consists of four ETFs and Hierarchical cluster 1 (Figure 12) consists of the same four ETFs with an additional four other ETFs.
- Checking on the ETF portfolio details, the four ETFs in DTW cluster 15 are companies in the mineral resources industry and hierarchical cluster 1 included some other ETFs from fossil resource industry such as oil and coal. This shows the differences in ETFs selection from both clustering algorithm.

Comparison of Clusters: Case Study 4

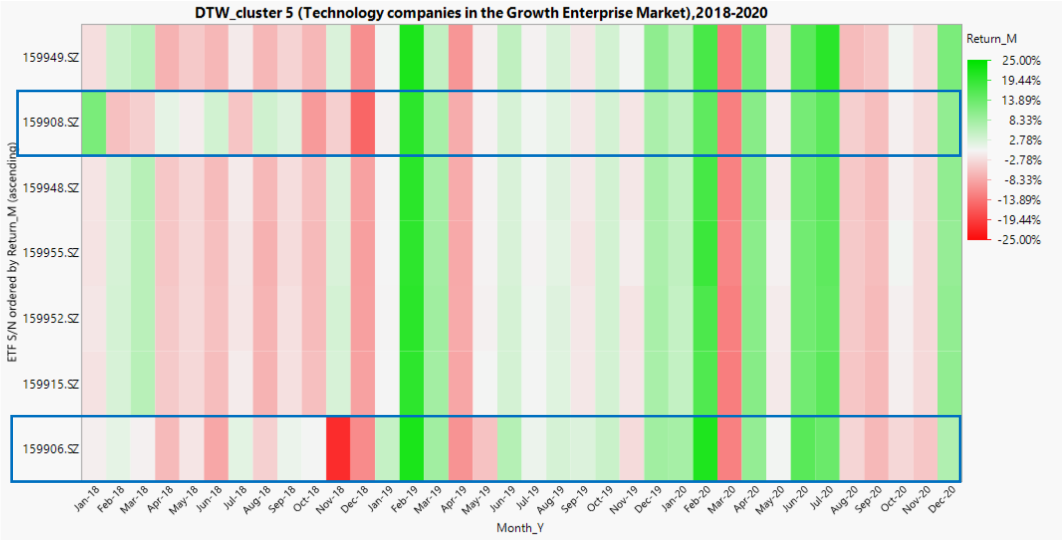


Figure 13. DTW Cluster 5

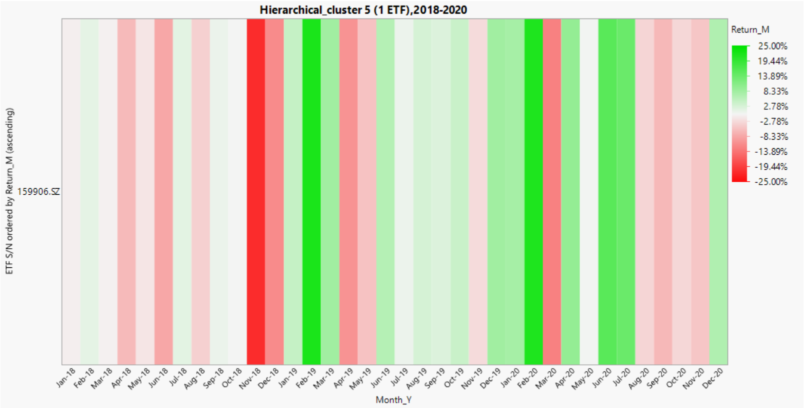


Figure 14. Hierarchical Cluster 5

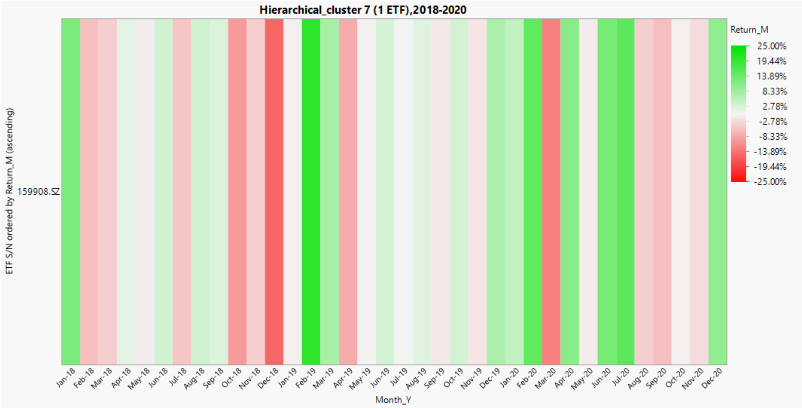


Figure 15. Hierarchical Cluster 7

- DTW clustered ETF 159906.SZ and 159908.SZ into cluster 5 (Figure 13) whereas hierachical singled out both ETF to form individual cluster with only one ETF.
- Although the intensity of returns for both ETFs in DTW cluster 5 seem to be different from the others in 2018 (Figure 13), those two ETFs focus on technology companies in growth enterprise market. The ETFs are succesfully grouped together by DTW method.

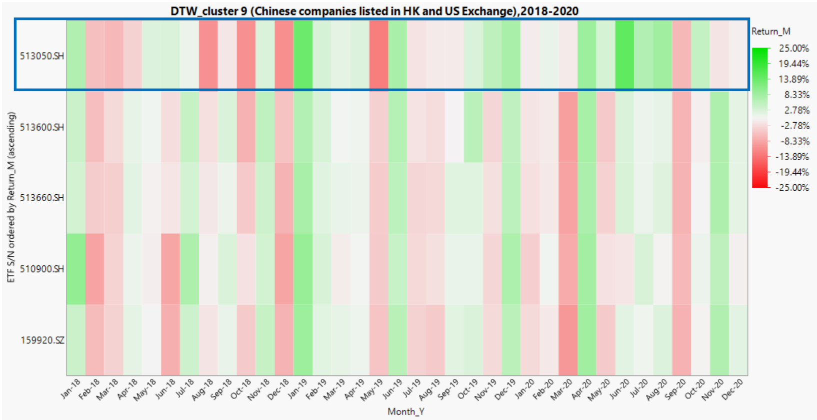


Figure 16. DTW Cluster 9

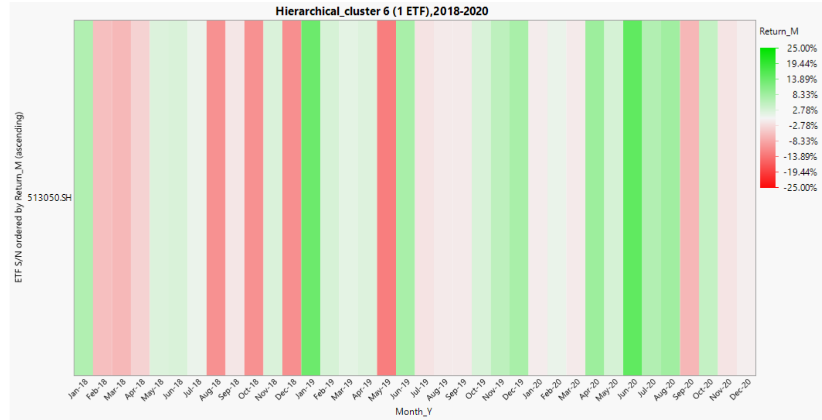


Figure 17. Hierarchical Cluster 6

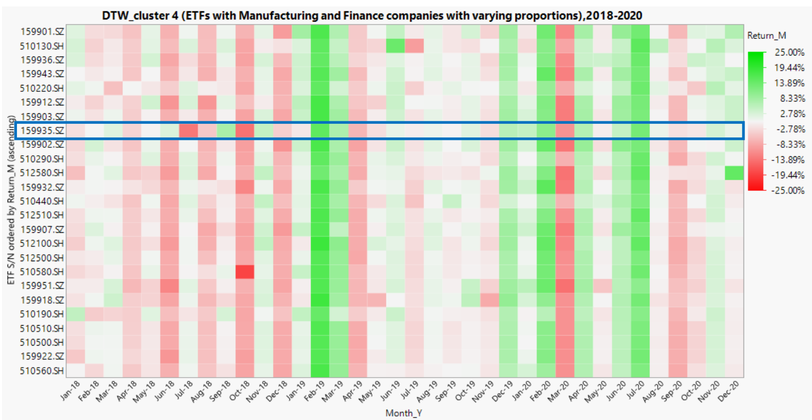


Figure 18. DTW Cluster 4

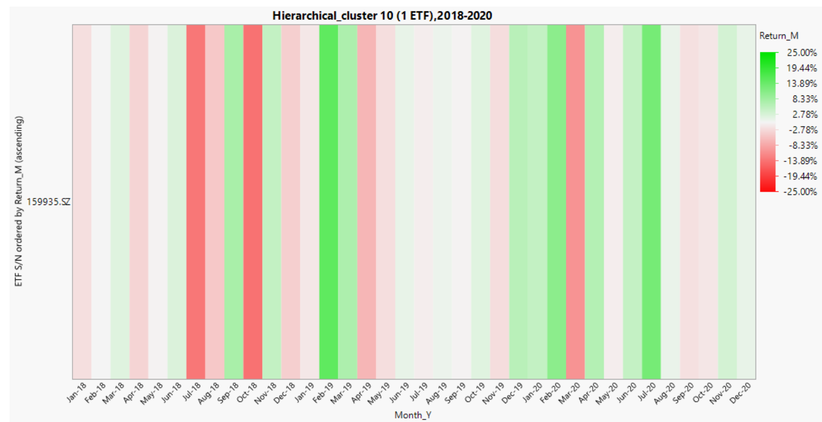


Figure 19. Hierarchical Cluster 10

- Similarily, ETF 513050.SH(Chinese internet companies listed in HK and US in Figure 17) and 159935.SZ (Manufacturing and Finance companies in Figure 19) are clustered individually by hierarchical clustering.
- On the other hand, DTW manages to cluster ETF 513050.SH together with other 4 ETFs focusing companies listed in HK (Figure 16). DTW also group ETF 159935.SZ with other ETFs focusing on the Manufacuting and Finance industries (Figure 18).

In conclusion, most of the cluster selections from both DTW (Table 2) and hierarchical clustering algorithm (Table 3) are similar with the exception of a few ETFs. Comparing the clustering results with the ETF portfolio composition, DTW algorithm manages to cluster ETFs of similar industry or portfolio more accurately. This may be due to the fact that instead of only calculating Euclidean distance between same time period as per hierarchical clustering algorthim, DTW can slide along the time axis to calculate the shortest distance between two time series and detect patterns that are out of phase.

ANALYSIS OF ETF CLUSTER PERFORMANCE

Best and Worst Performance month and cluster

The study over the 2018 to 2020 period identifies clusters with similar monthly returns trends. To further investigate the cluster performance in the three-year period, the graph in figure 20 is plotted with the boxplot and the median line is added for all 142 ETFs. JMP Pro 16.0.0 is used for the analysis.

From the graph, we observe that the range of fluctuation for monthly returns is larger in 2019 to 2020 compared to 2018. The interquartile range of months with volatile returns are also larger for all 142 ETFs. This signifies that for those months, the variance for monthly returns among the ETFs are larger.

The median for monthly returns is used to determine the 3 best performing and 3 worst performing months during the three-year period. Next, we also determine the 3 best performing and 3 worst performing clusters for the months and the results is documented in table 4.

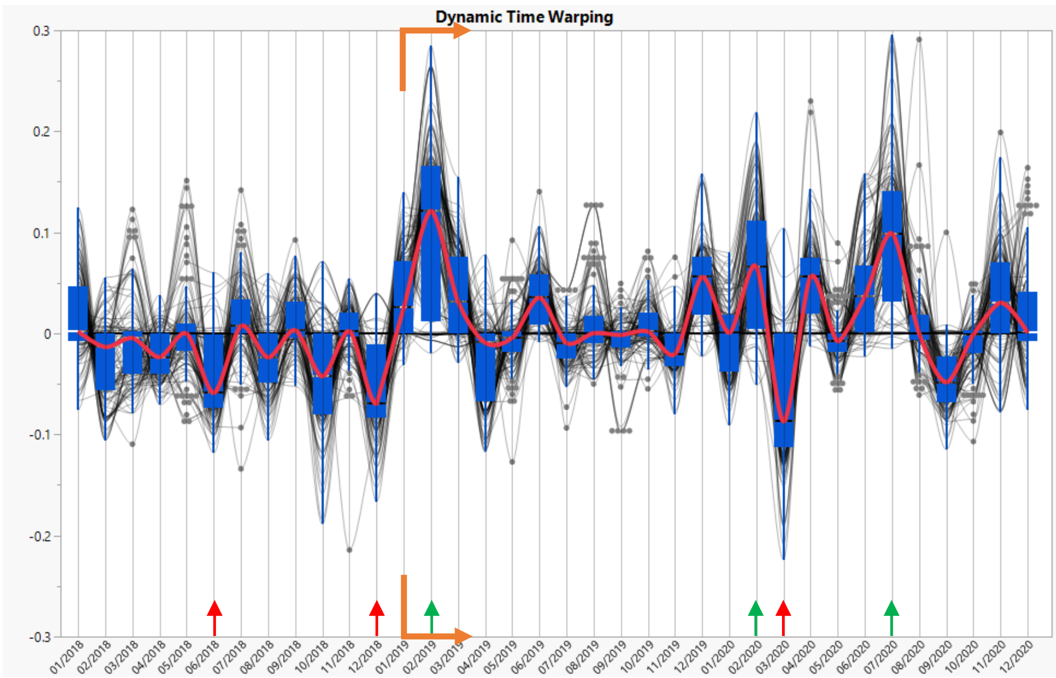


Figure 20. Monthly Return Plot with Boxplot and Median Line over Time

Rank	Best performing month	Worst performing month
1	February 2019 (Median return of 12.1%) 1) Cluster **3** (Median return of 26.4%) 2) Cluster **12** (Median return of 24.6%) 3) Cluster **5** (Median return of 20.6%)	March 2020 (Median return of -8.6%) 1) Cluster 12 (Median return of -19.0%) 2) Cluster **16** (Median return of -18.5%) 3) Cluster **14** (Median return of -12.6%)
2	July 2020 (Median return of 9.9%) 1) Cluster **7** (Median return of 28.9%) 2) Cluster **3** (Median return of 20.7%) 3) Cluster 15 (Median return of 18.1%)	December 2018 (Median return of -6.9%) 1) Cluster 8 (Median return of -13.7%) 2) Cluster **14** (Median return of -12.1%) 3) Cluster **16** (Median return of -11.2%)
3	February 2020 (Median return of 6.5%) 1) Cluster **12** (Median return of 18.5%) 2) Cluster **5** (Median return of 17.6%) 3) Cluster **7** (Median return of 16.5%)	June 2018 (Median return of -5.8%) 1) Cluster 3 (Median return of -9.4%) 2) Cluster 15 (Median return of -7.9%) 3) Cluster 4 (Median return of -7.7%)

Table 4. Best and Worst Performing Months

From the results in table 4, we observe that during 2018 to 2020, cluster 3(Securities ETFs), 5(Technology companies in the Growth Enterprise Market ETFs), 7(Military Manufacturers ETFs) and 12(Technology companies in Mature Enterprise Market ETFs) generally perform better during the best performing months and cluster 14(NASDAQ ETFs) and 16(S&P500 and Germany ETFs) generally perform worse during the worst performing months.

Interestingly, the worst performing clusters, 14 (Nasdaq) and 16 (S&P500 and Germany) are both overseas segments. The two clusters and their ETFs can be further explored to understand why these catergories performs the worst during a bad perfoming month for the ETFs market in China.

Figure 21 plots the monthly returns of cluster 7 and 12. For most months in 2019 to 2020, ETFs in cluster 7 (Military Manufacturers ETFs) have smaller interquartile range than cluster 12 (Technology Companies in Mature Enterprise Market). This indicates that picking any of the ETFs from cluster 7 will produce more consistent monthly returns when compared to ETFs in cluster 12 in these two years.

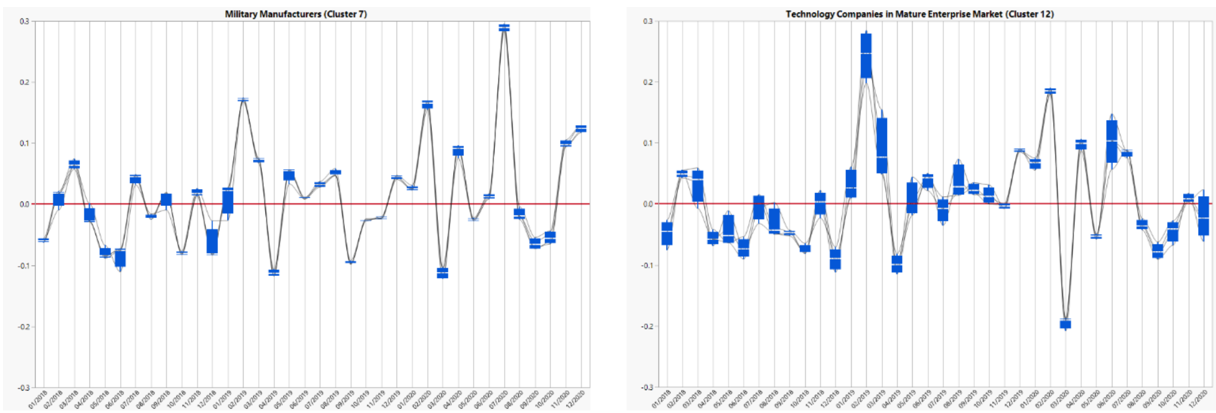


Figure 21. Monthly Return Plot for Cluster 7 and 12 over Time

Strong monthly return growth in 2019 and 2020

From the previous analysis, the results shows that the monthly return results in 2019 to 2020 performs generally better than 2018. Hence, we further explored the monthly returns for certain clusters. Figure 22 shows the monthly return heatmap for cluster 5, 7, 12 and 13. The heatmap plot saw more high intensity positive returns (bright green region) from the 2019 to 2020 period as compared to 2018.

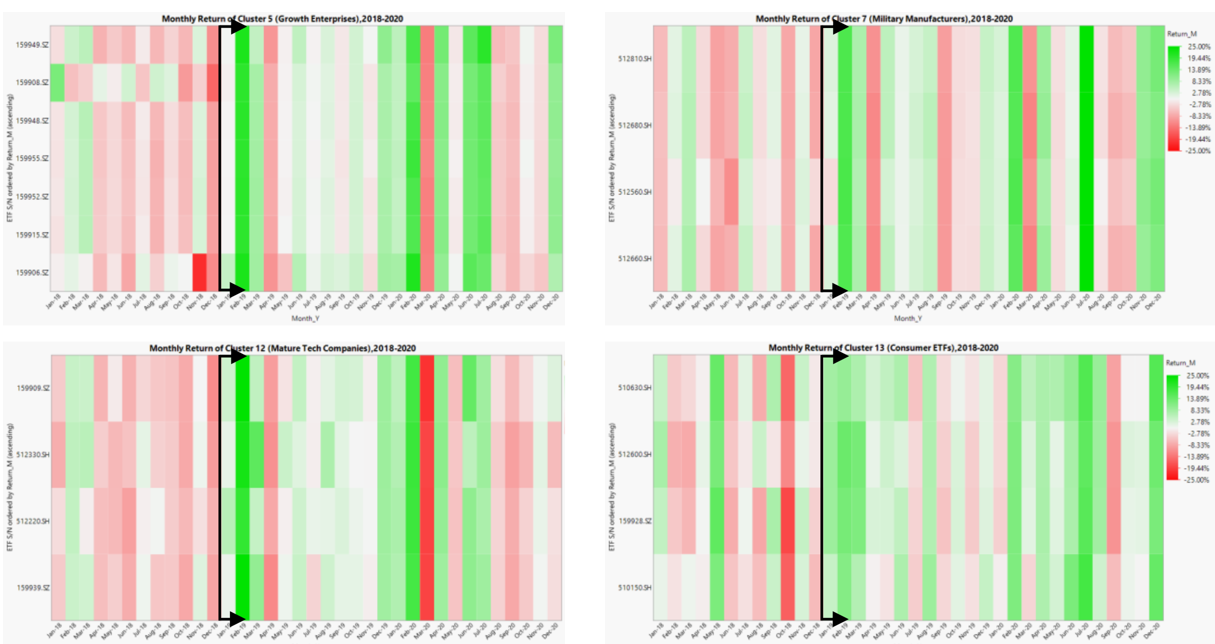


Figure 22. Heatmap for Cluster 5, 7, 12 and 13

We assume that it has to do with the China and United States rising trade war tension that started in 2017 which disrupted China's economy and trade market significantly. In 2019's New Year Address, Chinese President Xi called for China's self-reliance. According to statistics, China's high-tech manufacturing makes up a larger portion of the country's industrial growth in the first half of 2019 and they are shifting away from dependence on foreign technology and other products. Furthermore, China invested heavily in industries like AI and integrated circuits (IC) to achieve their goals towards self-reliance. These economic driven goals might have led the few industries to grow positively in 2019 to 2020.

Therefore, the ETFs from these industries can be further explored to evaluate whether they will continue to yield positive monthly returns in 2021 and so forth.

CONCLUSION AND FUTURE WORKS

In this paper, we employed dynamic time warping techniques to cluster China's ETFs. Clusters were formed based on similar price movement patterns from the monthly returns of ETFs.

It was observed that although traditional hierarchical clustering method also produces similar clusters as the DTW, there are merits of DTW to better cluster the ETFs. As ETFs prices are in a time series, DTW produced better clustering results as compared to hierarchical clustering.

We explored the ETFs monthly return and observed that there is larger fluctuation in 2019 to 2020 as compared to 2018. Upon investigation of the better performing clusters, we notice that there is some correlation of the clusters industry and China's economic goals announced in 2019. We assume that China's push for self-reliance in industries like AI and integrated circuits (IC) might have affected directly or indirectly the ETFs prices and monthly return.

Future work to explore the correlation of the better performing industries ETFs and China's economic goals can be carried out to test our assumption. Other macro factors can also be explored to determine if it affects the ETFs cluster positively or negatively.

Future research could broaden the scope by applying the same techniques to study different financial instruments in the market such as stocks prices, commodity prices, currencies or derivatives. This could unveil objective information for investors.

Finally, the analysis could be performed using different time interval such as daily returns, weekly returns, quarterly returns or yearly returns with different time periods. Our paper only focused on the data in 2018 to 2020 in China which had major global events occurring such as the China-United States trade war and COVID-19 pandemic in 2020. A research over a longer time period might uncover other useful information or seasonal trends in the ETF market.

REFERENCES

Degen, R. J. (2009). *Understanding China's historical development: The profit and the risk that China's stock market provides investors.* Portugal: globADVANTAGE Center of Research in International Business & Strategy.

Investopedia. (2021, March 3). *Exchanged Traded Fund (ETF)*. Retrieved from Investopedia: https://www.investopedia.com/terms/e/etf.asp

Isakov, V. (2019). *Performance Appraisal of Exchange-Traded Funds Using Clustering and Data Envelopment Analysis (XETRA, Germany).* Finland: LUT School of Business and Management.

Javed, A., Lee, B. S., & Rizzo, D. M. (2020). *Machine Learning with Applications.* ELSEVIER.

Lee, R. K.-W., & Kam, T. S. (2014). *Time-Series Data Mining in Transportation: A Case Study on Singapore Public Train Commuter Travel Patterns.* Singapore: Singapore Management University.

Liu, L. (2019, April 14). *China and its Pursuit for Chip Self-Sufficiency (Part 1)*. Retrieved from EETimes China: https://www.eetasia.com/china-and-its-pursuit-for-chip-self-sufficiency-part-1/

Solis, R. (2011). *Exchange-Traded Funds: a Multidimensional Analysis.* Fresno: ResearchGate.

TAKAHASHI, T. (2019, January 1). *Xi calls for 'self-reliant' China in New Year address*. Retrieved from Nikkei Asia: https://asia.nikkei.com/Politics/Xi-calls-for-self-reliant-China-in-New-Year-address

WorldAtlas; WFE. (2021, March). *Largest stock exchange operators in 2021, listed by market cap of listed companies.* Retrieved from Statista: https://www-statista-com.libproxy.smu.edu.sg/statistics/270126/largest-stock-exchange-operators-by-market-capitalization-of-listed-companies/

Zheng, H. (2019, July 15). *The China-U.S. tech rivalry fuels Chinese self-reliance*. Retrieved from CGTN: https://news.cgtn.com/news/2019-07-15/The-China-U-S-tech-rivalry-fuels-Chinese-self-reliance-IljJeLHN3G/index.html

国泰基金管理有限公司. (2020, August 10). 关于上证 180 金融交易型开放式指数证券投资基金实施基金份额拆分及相关业务安排的第一次提示性公告. *国泰基金管理有限公司*, pp. 1-5.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Name: Chen Li
E-mail: lichen.2020@mitb.smu.edu.sg

Name: Lim Yong Kai
E-mail: yongkai.lim.2020@mitb.smu.edu.sg

Name: Associate Professor Kam Tin Seong
E-mail: tskam@smu.edu.sg