WQD 7005 Data Mining

Milestone 3: Co-variance of stocks

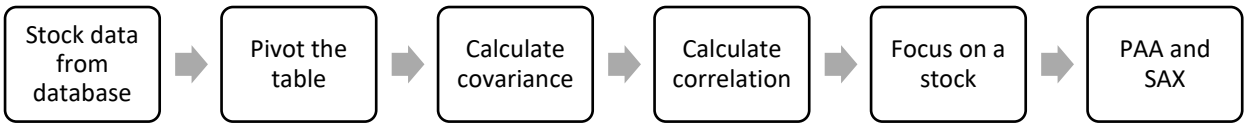
Ng Kang Wei

WQD170068

Video presentation: https://youtu.be/9VzZzI56TxE (Please turn on subtitle)

Problem statement: Do the movement of stock in plantation sector influences the movement of stock in consumer products sector and industrial product and services sector?

In milestone 3, a correlation between the stocks need to be discovered. The correlation is discovered by using the covariance between the stocks. In my opinion, the movement of plantation stocks might influence other stocks, such as consumer products stock and industrial product and services stock.



In order to discover the correlation between stocks, the stock data scraped since milestone 1 would be analyzed. The stock data need to be loaded into a data frame in Python. Then the data frame needs to be pivoted so that the stock code become the columns rather than the rows. At this stage, we can just call the pandas built in function to calculate the covariance and the correlation between the stocks. Both covariance and correlation measures if the stocks are positively or inversely related to each other. Covariance is used to measures variables with different units of measurement while correlation standardizes the unit of measurement.

Covariance dataframe

stock_code	0002	0008	0012	...	9962	9997	9997WB
stock_code				...			
0002	0.278592	0.017101	-0.284168	...	0.113734	0.518133	0.236916
0008	0.017101	2.072311	0.145112	...	0.071739	1.868690	0.159174
0012	-0.284168	0.145112	5.232152	...	0.033372	0.348701	0.421476
0021	-0.204587	-0.725278	0.283484	...	-0.149589	0.413093	2.552913
0029	-0.722014	-0.043122	-1.687284	...	1.856576	5.399028	-1.950342
0029WB	-0.208884	-0.580740	3.109682	...	-0.318878	1.487728	1.687796
0037	0.073184	-0.552317	-0.651969	...	0.724048	-1.005560	1.529444
0041	-0.058488	0.165079	0.399187	...	0.208333	0.416564	0.472583
0043	0.345341	-1.420428	1.113631	...	-2.716999	6.112844	-4.899607
0047	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.000000
0049	0.060286	1.091319	-0.522624	...	0.027921	-0.700848	0.637858
0051	-0.102440	0.600453	-2.420098	...	-0.626308	-0.320473	0.235064
0054	0.179587	0.129181	-1.778442	...	-0.182267	0.747873	0.051399
0056	-0.144876	0.301504	-0.288862	...	-0.533629	0.274681	-0.578575
0058	-0.104536	0.027866	0.175870	...	-0.736479	0.146291	0.350058
0059	-0.033168	0.444229	-1.089039	...	-1.265445	1.632062	1.952482
0064	-0.693034	-0.699115	-0.078287	...	1.138279	-0.918749	-0.059369
0065	-0.110962	0.577952	0.102943	...	0.236321	0.902539	1.134927
0065WA	-0.431104	-0.162690	-1.373358	...	-1.571780	-0.463335	1.592981

Figure 1 Covariance matrix for all stocks

Correlation dataframe

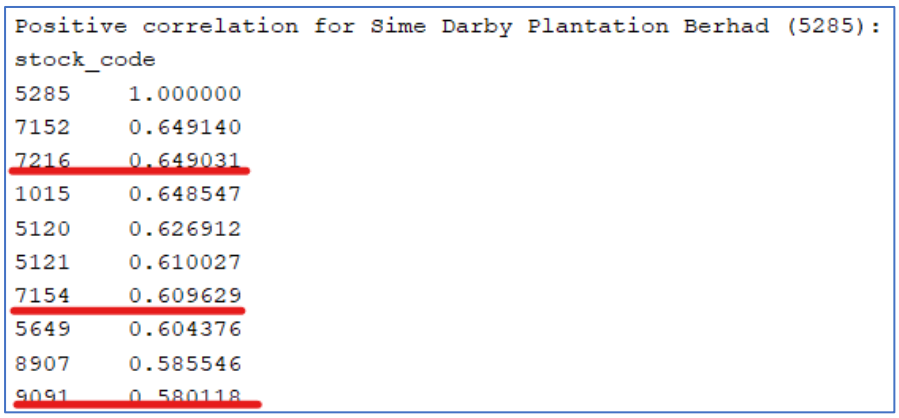
stock_code	0002	0008	0012	...	9962	9997	9997WB
stock_code				...			
0002	1.000000	0.022506	-0.235370	...	0.113588	0.217549	0.147006
0008	0.022506	1.000000	0.044069	...	0.026269	0.287680	0.036214
0012	-0.235370	0.044069	1.000000	...	0.007691	0.033784	0.060347
0021	-0.235301	-0.305850	0.075235	...	-0.047869	0.055575	0.507567
0029	-0.332732	-0.007286	-0.179424	...	0.238051	0.291037	-0.155371
0029WB	-0.072542	-0.073948	0.249199	...	-0.030812	0.060436	0.101325
0037	0.080328	-0.222277	-0.165128	...	0.221118	-0.129105	0.290197
0041	-0.070696	0.073161	0.111340	...	0.070065	0.058898	0.098746
0043	0.083770	-0.126333	0.062334	...	-0.183374	0.173448	-0.205453
0047	NaN	NaN	NaN	...	NaN	NaN	NaN
0049	0.042376	0.281262	-0.084769	...	0.005461	-0.057625	0.077506
0051	-0.087215	0.187438	-0.475443	...	-0.148360	-0.031915	0.034595
0054	0.193248	0.050967	-0.441593	...	-0.054570	0.094135	0.009561
0056	-0.260626	0.198871	-0.119910	...	-0.267097	0.057801	-0.179925
0058	-0.299056	0.029229	0.116098	...	-0.586215	0.048954	0.173116
0059	-0.019624	0.096367	-0.148679	...	-0.208312	0.112949	0.199692
0064	-0.350898	-0.129787	-0.009147	...	0.160355	-0.054414	-0.005196
0065	-0.084650	0.161661	0.018122	...	0.050161	0.080539	0.149669
0065WA	-0.173993	-0.024075	-0.127902	...	-0.176502	-0.021874	0.111140
0078	0.241206	0.258207	0.319923	...	0.350764	-0.021358	-0.141073
0078WB	-0.052297	0.201954	-0.076568	...	-0.284754	-0.217793	-0.076088

Figure 2Correlation matrix for all stocks

A correlation coefficient of value one, it means that the variables have a perfect positive correlation. When one moves so does the other in the same direction, proportionally. A positive correlation coefficient that is less than 1, indicates a weaker positive correlation. If the correlation efficient is 0, that means that the variables has no identifiable relationship between them. If the correlation coefficient is less than 0, it indicates that the variables are inversely related to one another.

The covariance and correlation are calculated with the percentage change of the stock in a single day. (Close – Open) / Open * 100.

In order to illustrate the problem statement stated above, one of the prominent plantation stocks is chosen. The stock chosen is Sime Darby Plantation Berhad (5285). When the price of the stock 5285 goes up, the consumer product stock, Kawan Food Bhd (7216) would rise as well. These 2 stocks have a positive correlation of 0.649. This is probably because Kawan Food Bhd is in the food business which would use a large quantity of palm oil in its product, an increase in demand of palm oil drive the rise of the palm oil or plantation stock. Caely Holdings Bhd (7154) and Emico Holdings Bhd (9091) are consumer product and services sector and have a positive correlation with Sime Darby Plantation Bhd stock movement.



```
Positive correlation for Sime Darby Plantation Berhad (5285):
stock_code
5285    1.000000
7152    0.649140
7216    0.649031
1015    0.648547
5120    0.626912
5121    0.610027
7154    0.609629
5649    0.604376
8907    0.585546
9091    0.580118
```

Figure 3 Positive correlation with Sime Darby Plantation Berhad

Besides positive correlation, there is negative correlation to the movement of Sime Darby Berhad stock prices. When the plantations stock (Sime Darby) rises, the industrial product and services stock has a tumble. This might be due to a rise in the prices of the raw materials from plantation causing an increase of cost in the production of the industrial product. For

instance, Hock Heng Stone Industries Bhd (5165) from the industrial product and services sector has a negative correlation coefficient of -0.707 with Sime Darby Plantation Bhd. Daibochi Berhad (8125), FACB Industries Incorporated (2984), Ablegroup Berhad (7086) are from the industrial product and services sector, they all have a negative correlation with Sime Darby Berhad, a plantation stock.

```
Negative correlation for Sime Darby Plantation Berhad (5285):
stock_code
5165   -0.707354
8125   -0.605335
2984   -0.602856
5008   -0.585498
7086   -0.575729
7053   -0.575729
7544   -0.575729
7199   -0.575729
9539   -0.555750
7062   -0.520808
```

Figure 4 Negative correlation with Sime Darby Plantation Berhad

Finally, SAA and PAX methods are applied on the time series. These algorithms would smoothen the time series data and convert the time series data to a string. With the positive correlation between the stocks, the graphical output from SAA and PAX would be similar. Stocks with the negative correlation would have a very dissimilar graphs.

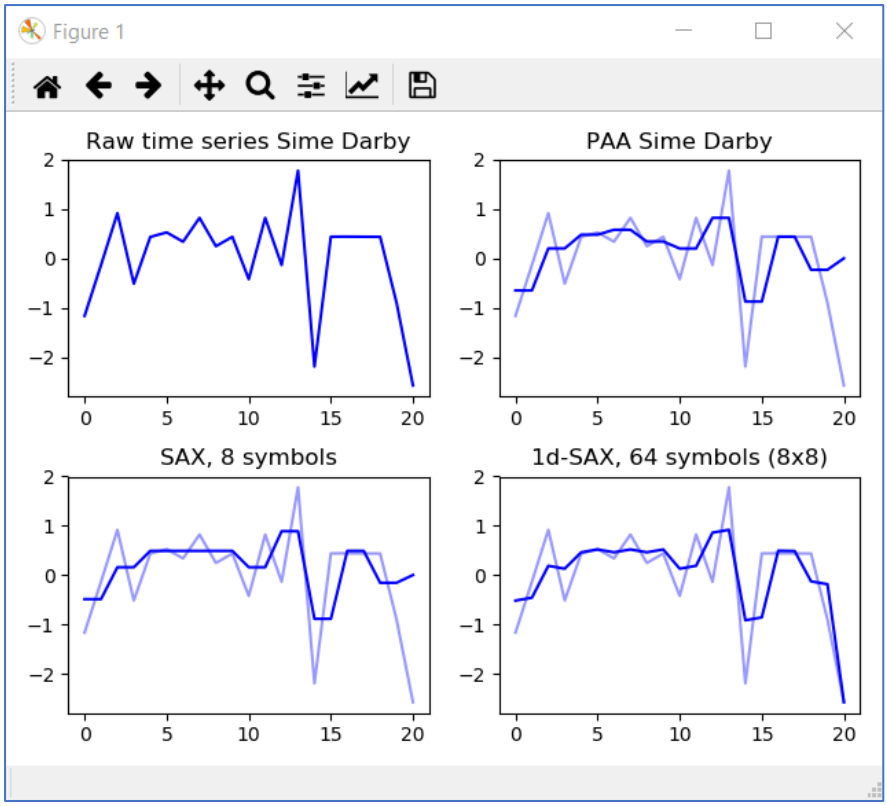


Figure 5 PAA and SAX of Sime Darby Plantation Berhad

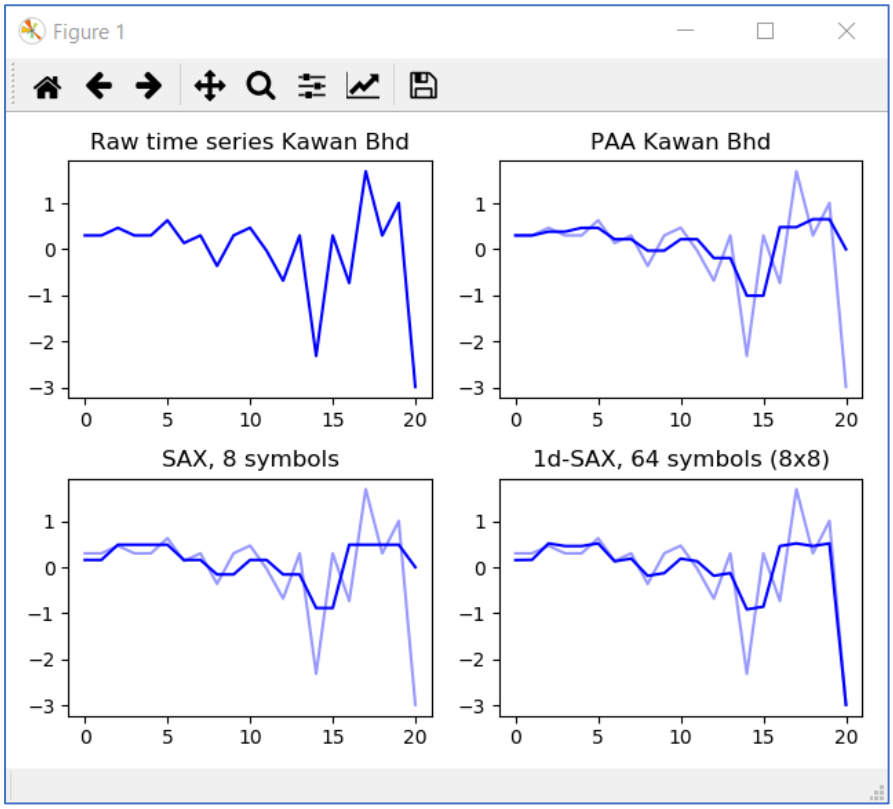


Figure 6 PAA and SAX of Kawan Berhad

Figure 7 PAA and SAX of Caely Holdings

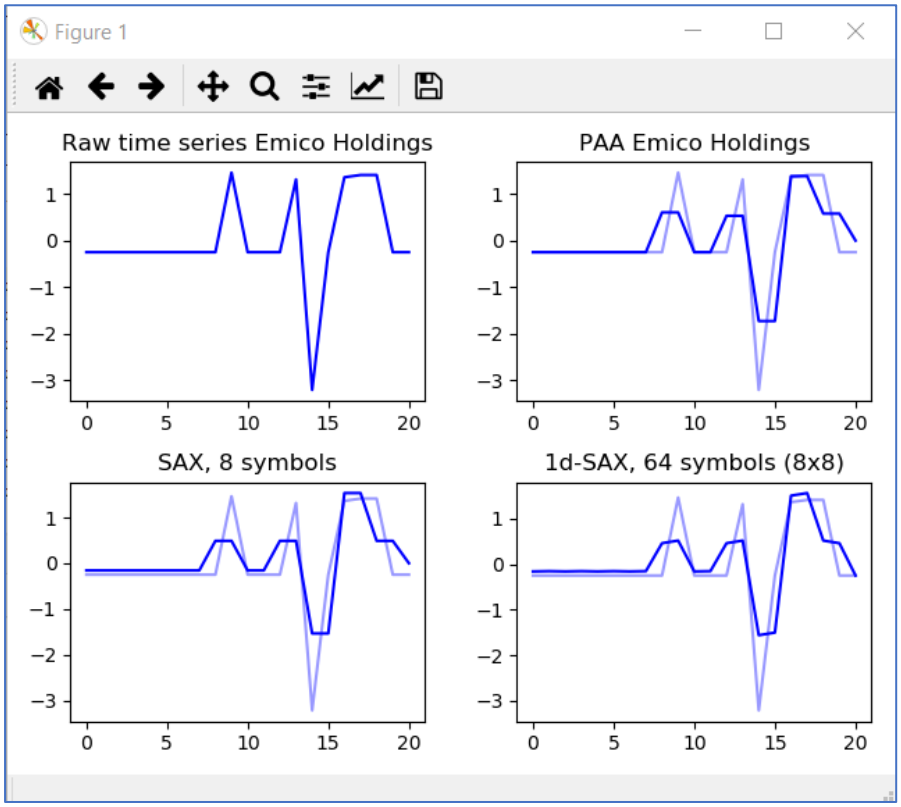


Figure 8 PAA and SAX of Emico Holdings

The few stock time-series that follows are from stocks that have high negative correlation with Sime Darby Plantation Berhad. The transformed PAA and SAX graphs would be in a distinctly different pattern from Sime Darby Plantation Berhad stock time-series.

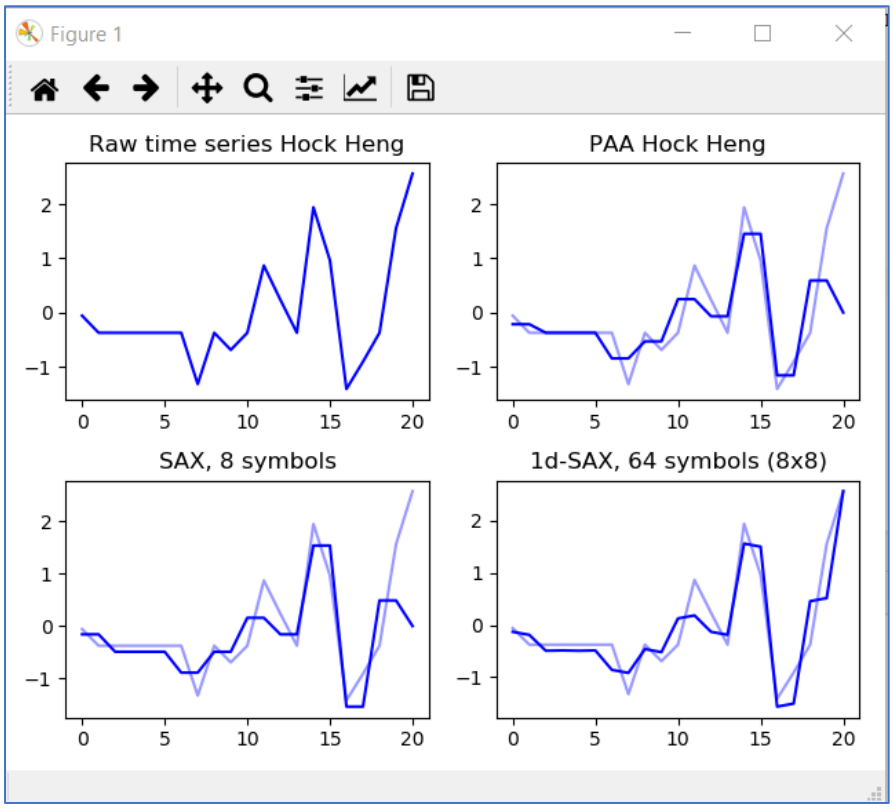


Figure 9 SAA and PAX of Hock Heng Stone Industries Berhad

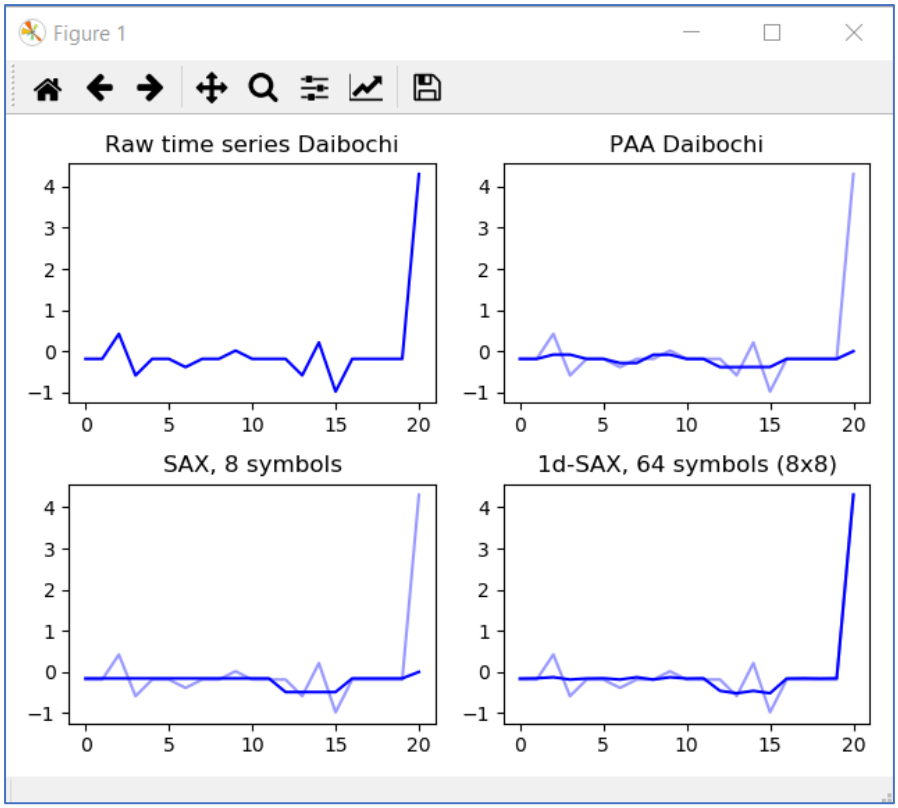


Figure 10 PAA and SAX of Daibochi Berhad

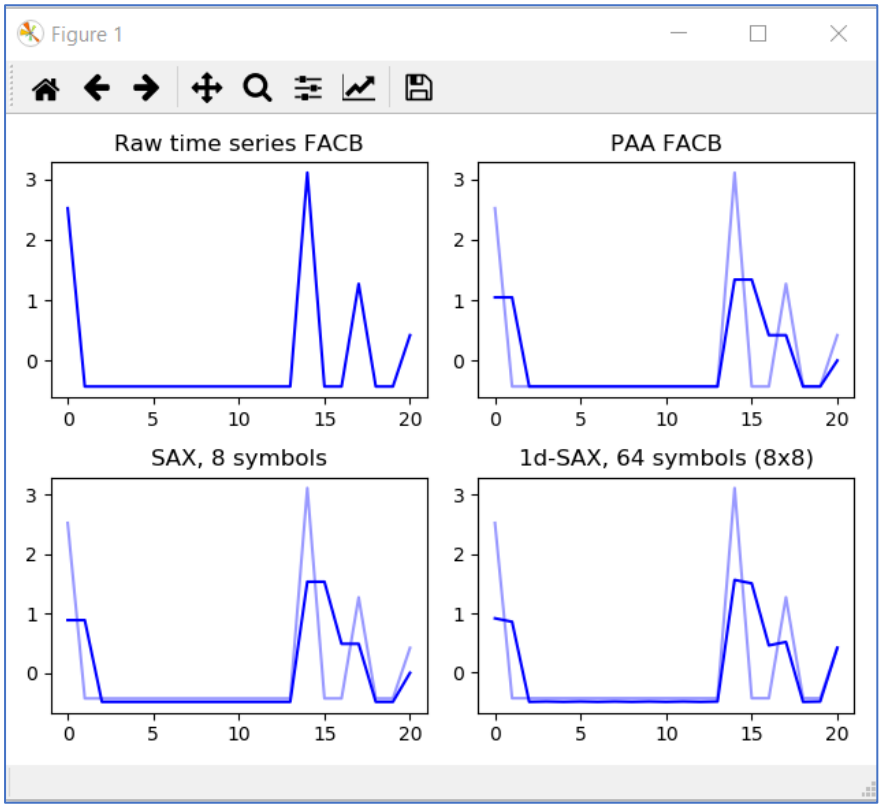


Figure 11 PAA and SAX of FACB Industries Incorporated

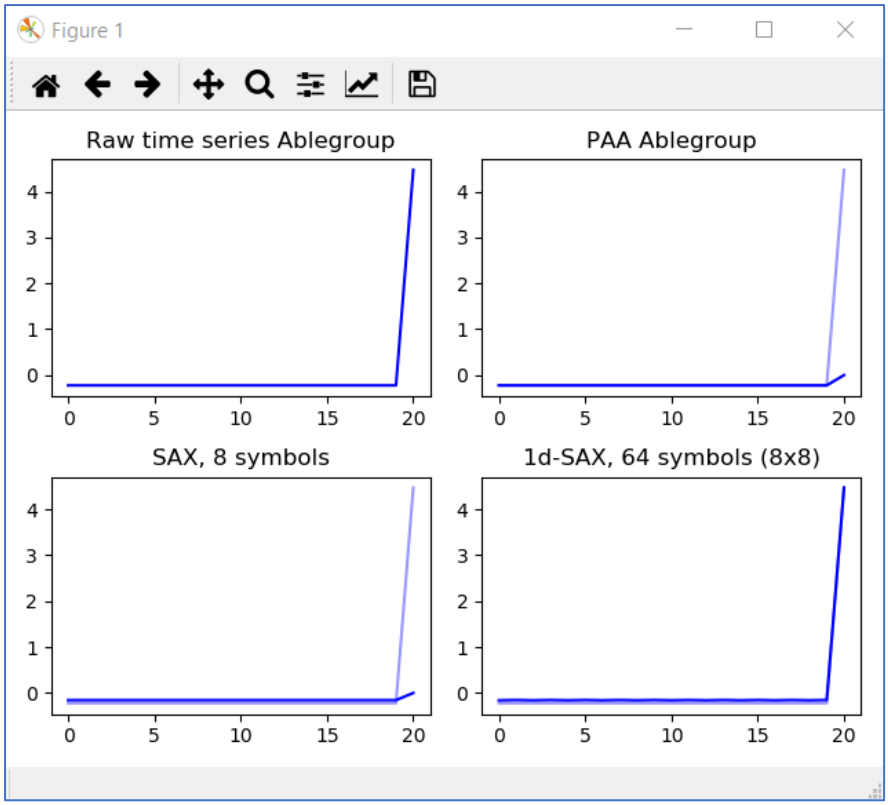


Figure 12 PAA and SAX of Ablegroup Berhad

Conclusion

From the correlation analysis and PAA and SAX transformed time series it is shown that the problem statement does hold true. Movement of plantation stock or Sime Darby Plantation Berhad in the past month has a positive correlation with some stocks in consumer products sector and a negative correlation with some stocks in industrial product and services sector.