

WQD7005 DATA MINING

MILESTONE 3: CO-VARIANCE OF STOCKS

Group Members:

WQD180063 YONG WAI SHIN

WQD170033 JEREMY CHIN EE WEI

WQD180069 YAP BOON HUI

**1.0 Introduction**

Stock analysis consist constant dynamic changes of variables in terms of opening price, closing price, change percentage, stock volume, etc. Various analysis techniques may be applied to study relationships between stocks to help analysts to have a better understanding to evaluate stocks.

**1.1 Covariance of Stocks**

Predicts how two stocks might perform relative to each other in the future. A positive covariance would indicate the stocks move in the same direction. A negative covariance means that when one stock advances, the other will retreat. Covariance close to zero would indicate the two stocks act mostly independently.

**1.2 Piecewise Aggregated Approximation (PAA) of Time Series**

Dimensionality reduction of time series from *n* dimensions to *w* dimensions whereby the data is divided into *w* dimensions with equal sized frames which provides the important lower bounding property. Each time series is normalized to have mean of zero and a standard deviation of one before converting it to the PAA representation.

**1.3 Symbolic Aggregate Approximation (SAX)**

Takes a PAA transformed times series and maps it into a sequence of symbols. SAX provides lower bounding of Euclidean distance. SAX approximations are trusted to do comparisons between two times-series data sets which aim to still keep the characteristics after PAA times series dimensionality reduction.

References:

<https://www.linkedin.com/pulse/part-i-introduction-time-series-approximation-sax-supreet-oberoi/>

<http://www.cs.ucr.edu/~eamonn/SAX.pdf>

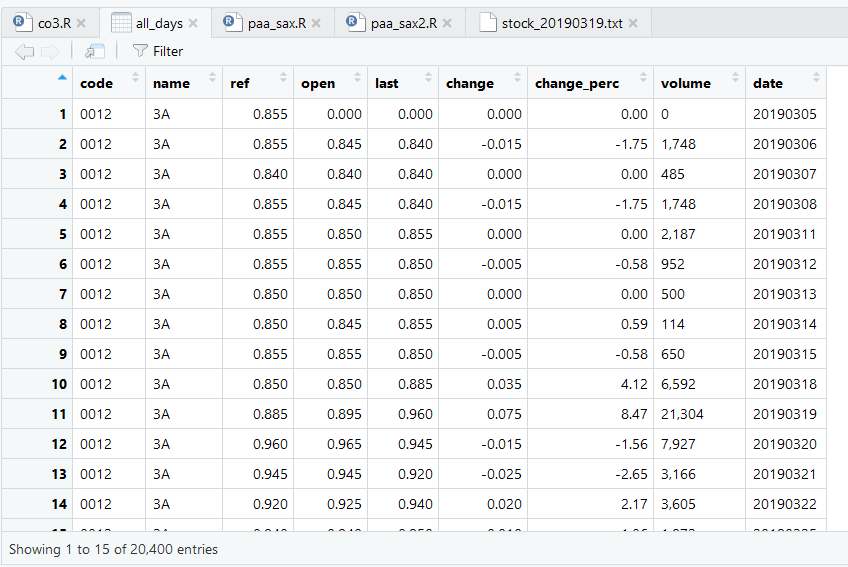
<https://cs.gmu.edu/~jessica/SAX_DAMI_preprint.pdf>

**2.0 Data Preprocessing**

Data consist 20 days of approximately 1800 crawled stocks data with variables:

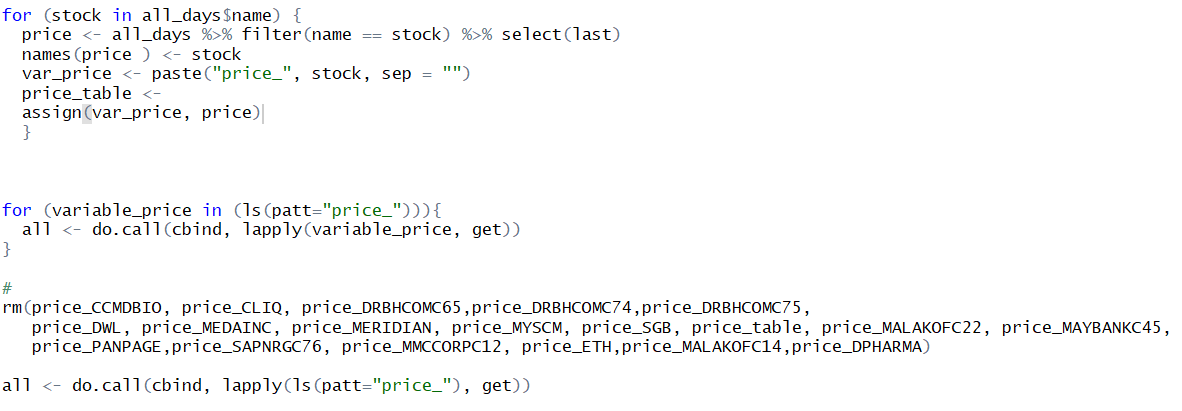
* Stock Code
* Stock Name
* Stock Reference
* Stock Opening
* Stock Closing
* Stock Change
* Stock Change Percentage
* Stock Volume

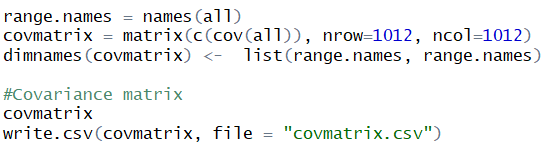
Preprocessing was carried out using the dplyr package in R to bind and sort each stock in ascending date order as shown below. These include naming of columns, mutating a date column, and reading file names from folder directory.

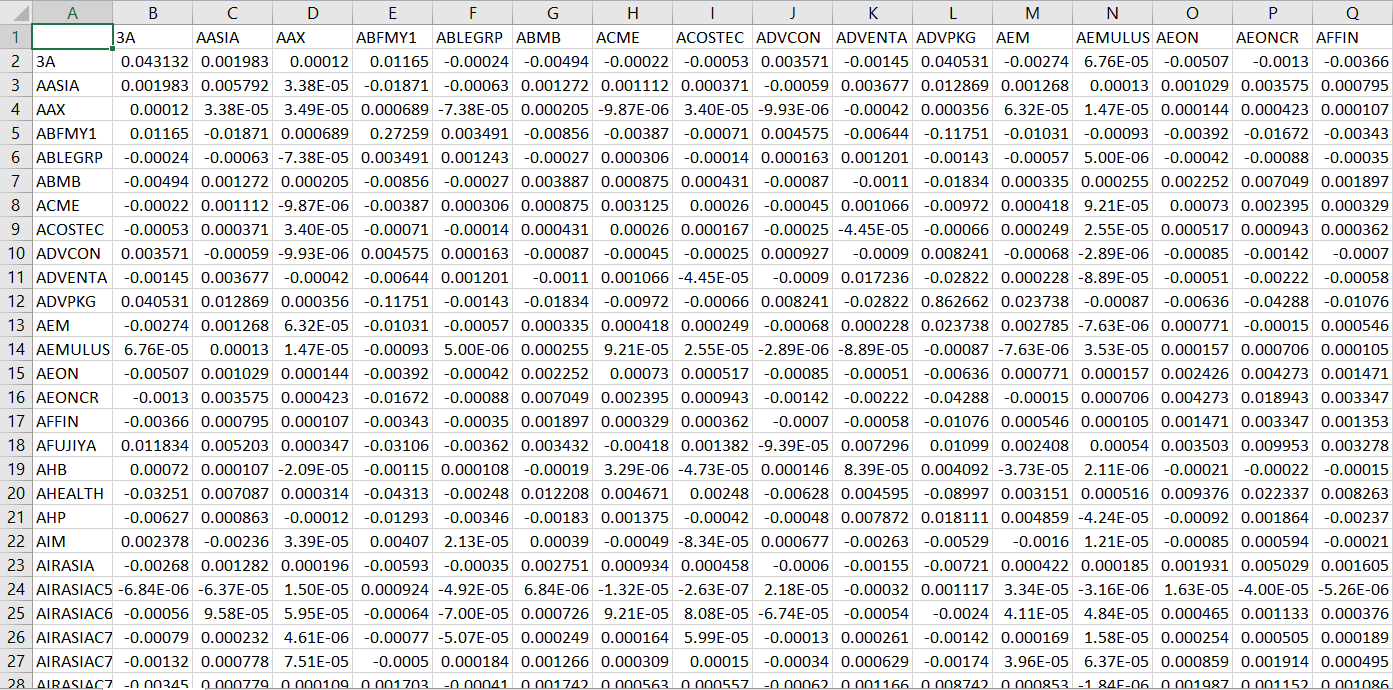


**3.0 Covariance**

For loop was used to separate all stock closing prices into individual list, column bind into a list called “all”and the covariance matrix of closing prices were generated

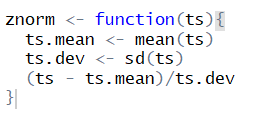




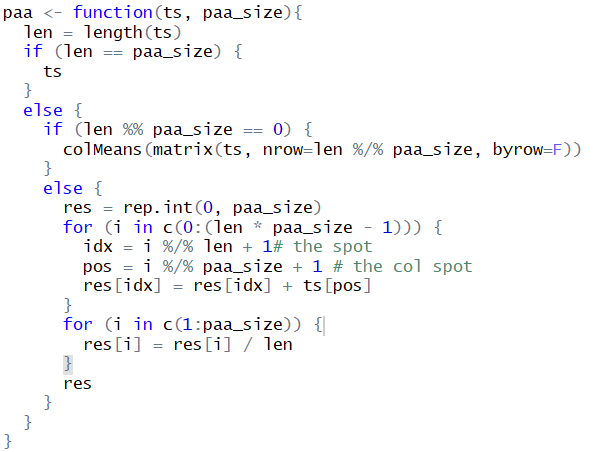


**4.0 PAA-SAX Time Series**

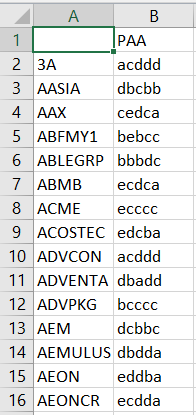
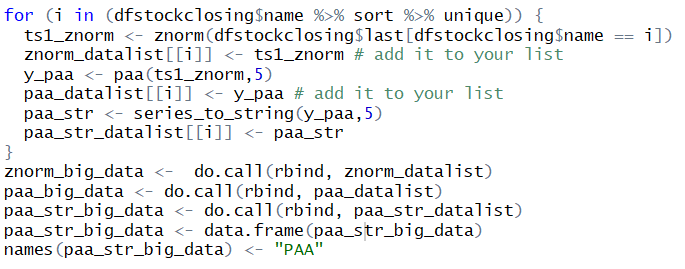
Znorm function was used to normalize the data before transforming the timeseries



PAA transform function



For loop was used to transform each stock’s closing prices reduced to times series of 5 points and output to string values and saved into a csv file.



Data Visualization

