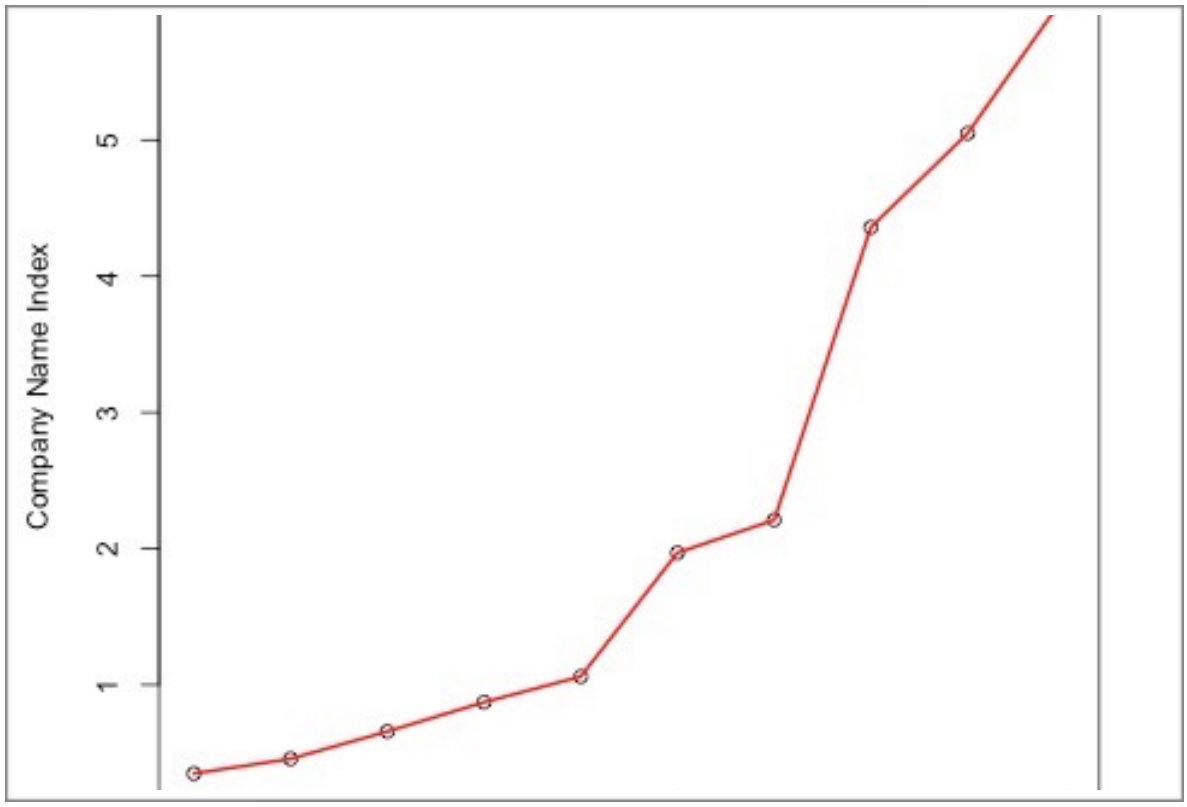
Time Series Forecast

Using R programming Language



CSE587- Data Intensive Computing

Spring 2015

Time Series Forecast

Using R programming language

Problem:

Use R to implement the time-series forecast of stocks in NASDAQ

Description:

In this assignment, you are required to use R language/tools in CCR to do time-series forecast of stock price using the same data in hw#1. There are many approaches to forecasting. In this homework, you will compare three techniques, namely, Linear Regression Model, Holt-Winters Model, and ARIMA model.

Implementation:

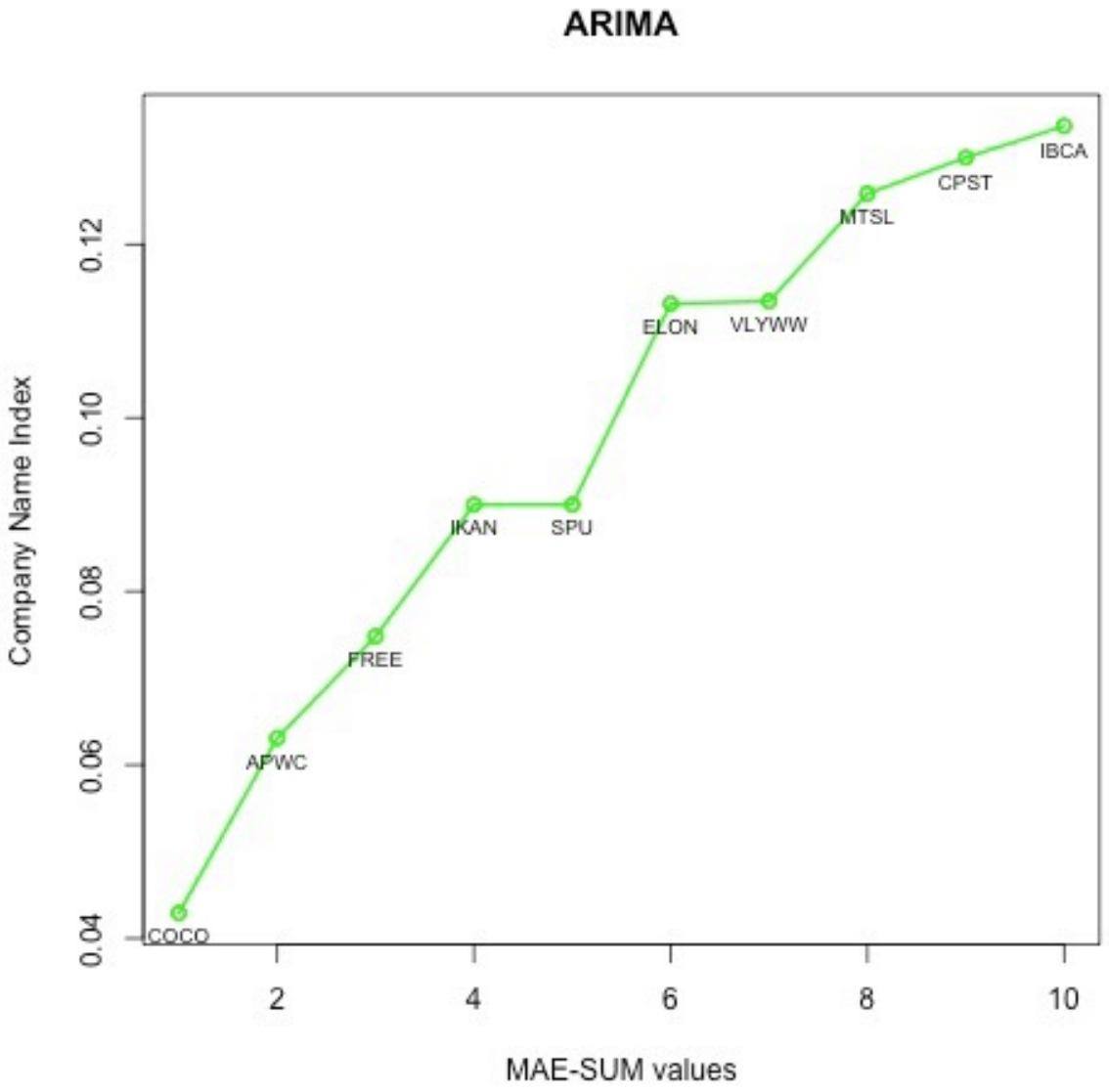
In order to finish this task, the data for each stock will be split into two parts:

- The first part with 744 trading days is used for training
- The second part with 10 trading days is used for testing.

Algorithm:

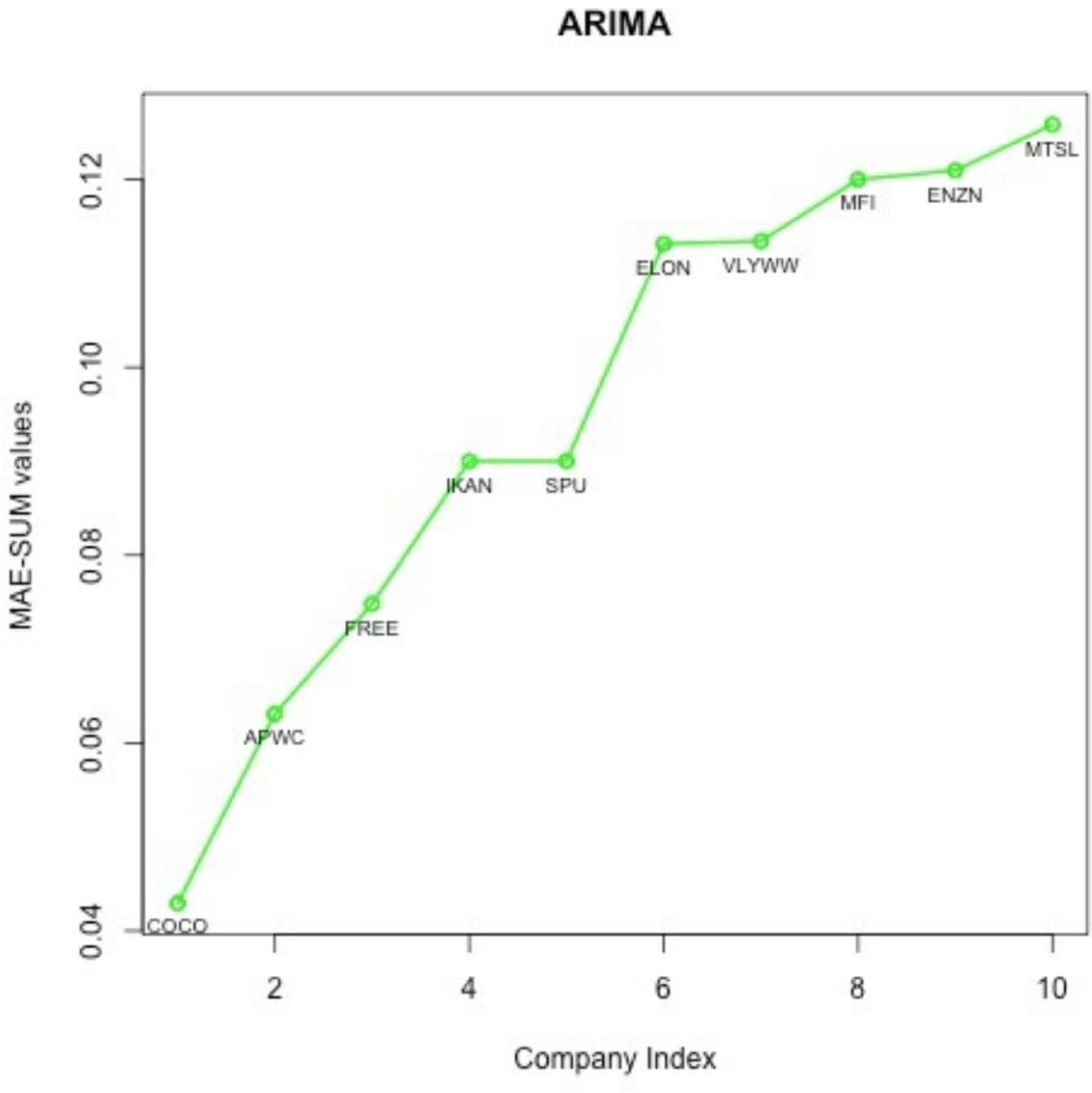
1) get the list of stocks which does not satisfy the condition.
2) Split the data into training and test data.
3) Compute the forecast for test and training data using three models Linear Regression, Holt-winters, Arima.
4) predict the MAE which is the difference between predicted value and the actual value.
5) get the top 10 stocks with minimum MAE for all the models.

Time Series Graph



"finalList"	"arima_vector"	
"469"	"COCO.csv"	0.0429102862857264
"141"	"APWC.csv"	0.0630886637689909
"821"	"FREE.csv"	0.0748033738137716
"1042"	"IKAN.csv"	0.0900000000000001
"1919"	"SPU.csv"	0.0900000000000001
"674"	"ELON.csv"	0.113156343435634
"2156"	"VLYWW.csv"	0.113436790461283
"1353"	"MTSL.csv"	0.125836227508328
"494"	"CPST.csv"	0.13
"1008"	"IBCA.csv"	0.133633916554677

With Approximation, seasonal and lambda set i get the below values:



"finalList" "arima_vector"

"469" "COCO.csv" 0.0429102862857658

"141" "APWC.csv" 0.0630886648410769

"821" "FREE.csv" 0.0748033738137716

"1042" "IKAN.csv" 0.0900000000000001

"1919" "SPU.csv" 0.0900000000000001

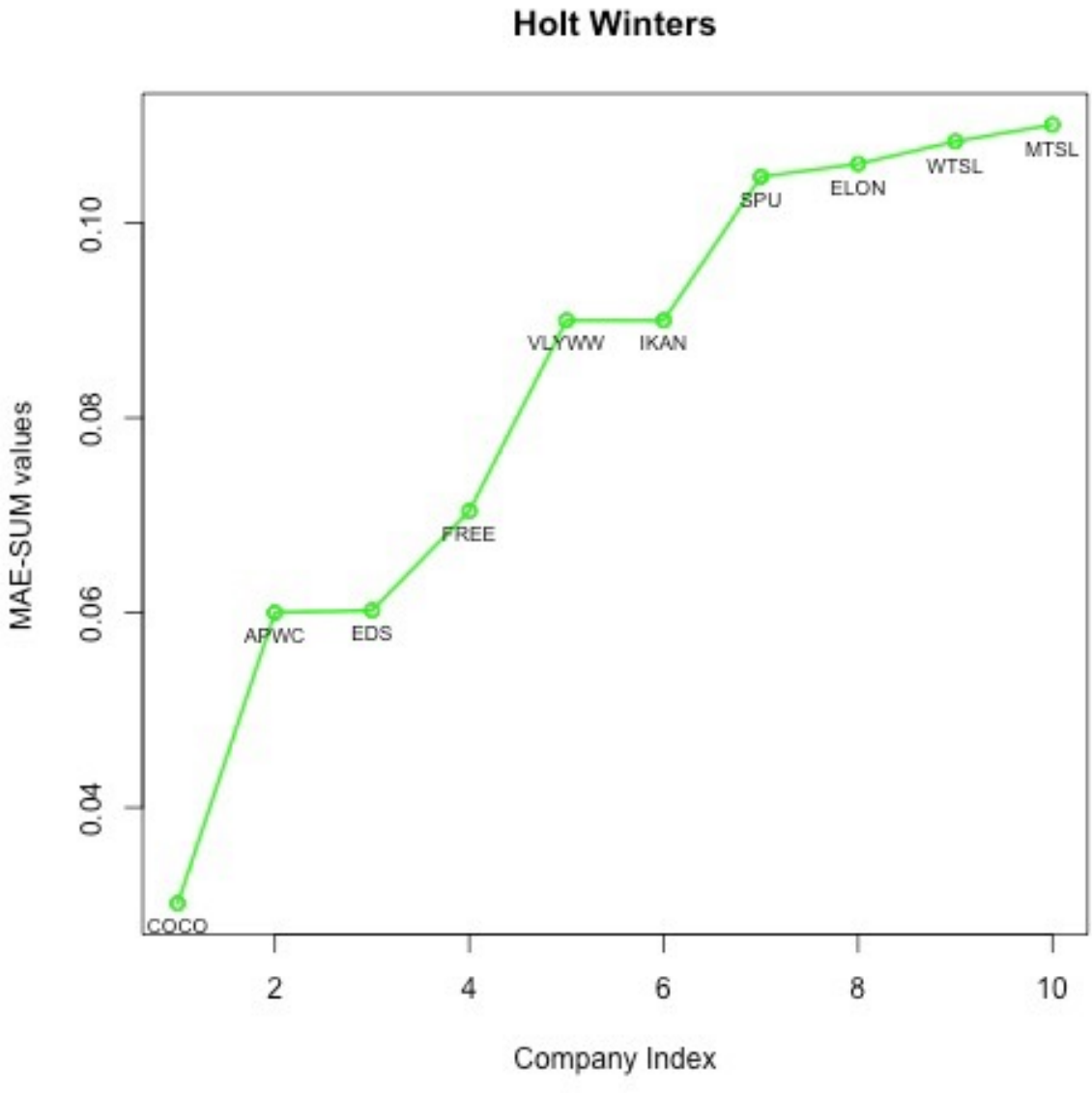
"674" "ELON.csv" 0.11315609717094

"2156" "VLYWW.csv" 0.113436790470689

"1280" "MFI.csv" 0.120000000000006

"694" "ENZN.csv" 0.12096462378173

"1353" "MTSL.csv" 0.12583622799868



"finalList" "hw_vector"

"469" "COCO.csv" 0.030156758080105

"141" "APWC.csv" 0.0600065426668421

"658" "EDS.csv" 0.0602272459239397

"821" "FREE.csv" 0.0704327240953608

"2156" "VLYWW.csv" 0.09

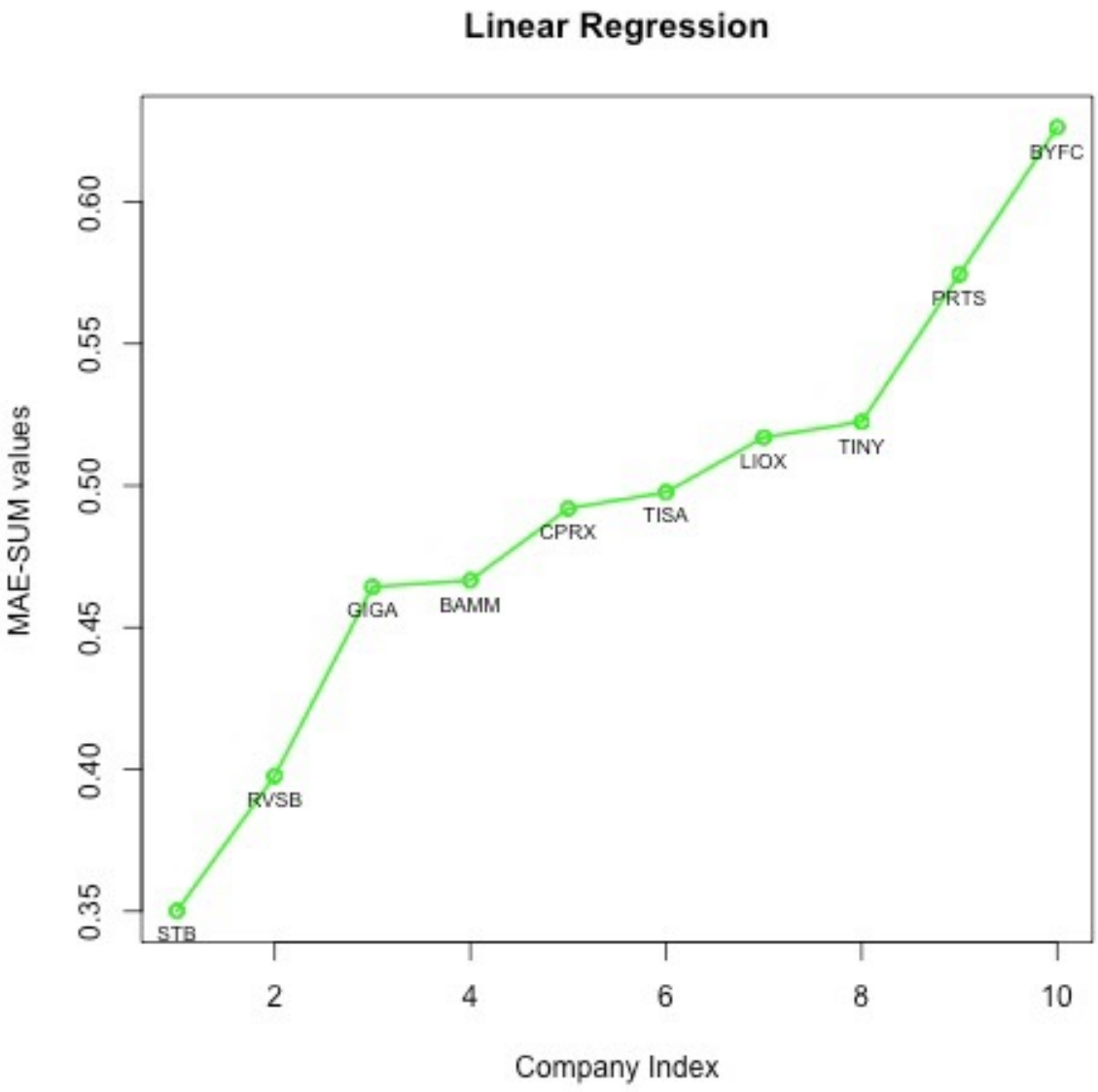
"1042" "IKAN.csv" 0.0900000000000001

"1919" "SPU.csv" 0.104746328560384

"674" "ELON.csv" 0.106063964745531

"2242" "WTSL.csv" 0.108373386847075

"1353" "MTSL.csv" 0.110086738825851



	"finalList"	"lr_vector"
"1940"	"STB.csv"	0.350085995086024
"1766"	"RVSB.csv"	0.397678132678126
"870"	"GIGA.csv"	0.464299754299754
"220"	"BAMM.csv"	0.466609336609332
"490"	"CPRX.csv"	0.491928746928756
"2023"	"TISA.csv"	0.497628992628996
"1188"	"LIOX.csv"	0.517002457002481
"2021"	"TINY.csv"	0.522469287469304
"1623"	"PRTS.csv"	0.574410319410301
"328"	"BYFC.csv"	0.626216216216218

Conclusion

Arima model without utilising the MPI took 30 hours to compute in the 16 node cluster. With MPI it took less than 2 hours. Thus we can come to a conclusion that MPI helps in utilising the maximum benefit from the CPUs.