Statistical Arbitrage Model

- ADWAYT PRADEEP NADKARNI

Problem Statement

Statistical Arbitrage: For a family of stocks, generally belonging to the same sector or industry, there exists a correlation between prices of each of the stocks. There, though, exist anomalous times when for a small period of time, the correlation is broken. But the market self corrects in some time and the correlation is re-established. During this small window of time when correlation is anomalous, there exists a money-making opportunity for quantitative traders.

** First, you have many types of data The stock market is like candy-land for any data scientists who are even remotely interested in finance. hat you can choose from. You can find prices, fundamentals, global macroeconomic indicators, volatility indices, etc. the list goes on and on. Second, the data can be very granular. You can easily get time series data by day (or even minute) for each company, which allows you think creatively about trading strategies. Finally, the financial markets generally have short feedback cycles. Therefore, you can quickly validate your predictions on new data. Here is the equity data of stocks listed on NSE over 2016 and 2017:

Dataset (https://drive.google.com/file/d/1kyNXxSM-_MSW4kSUJ90HlPQaxPCVti5L/view)

Problem Statement: Develop Machine Learning Algorithm to predict statistical arbitrage opportunities in NSE based on the 2016 data. Test this algorithm on 2017 data.

Submission Guidelines The project must be made completely on command line. You should share the complete source code via GitHub only. And also share the screenshots of the complete application. Also submit a report consisting of graphs providing visualisation of results and conclusions.

Algorithm

- Download the dataset and extract the essential data.
- Use year of record (2016 and 2017) as index and split data into training and testing.
- Perform Linear regression function and use 2016 observations as labels and features.
- Fit the 2017 values to get the predicted label
- Check the accuracy by comparing with actual data.
- Plot the graphs.

Out[8]:

	Unnamed: 0	SYMBOL	SERIES	OPEN	HIGH	LOW	CLOSE	LAST	PREVCLOSE	TOTTRDQTY	TOTTRDVAL	TIMESTAMP	TOTALTRADES	ISIN
0	0	20MICRONS	EQ	37.80	37.80	36.15	36.85	37.40	37.05	27130	9.946579e+05	2017-06-28	202	INE144J01027
1	1	20MICRONS	EQ	52.90	61.70	52.90	59.40	59.65	52.80	2489935	1.481265e+08	2017-11-28	14196	INE144J01027
2	2	20MICRONS	EQ	36.15	36.30	35.60	36.05	36.00	35.65	40594	1.461975e+06	2017-02-28	193	INE144J01027
3	3	20MICRONS	EQ	44.35	44.80	43.55	43.90	43.60	44.80	90170	3.984352e+06	2017-04-28	597	INE144J01027
4	4	20MICRONS	EQ	36.00	36.75	34.20	34.90	34.90	35.40	77310	2.709788e+06	2017-03-20	413	INE144J01027
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
490	490	20MICRONS	EQ	28.45	28.45	27.70	28.30	28.20	28.05	8045	2.263841e+05	2016-06-08	100	INE144J01027
491	491	20MICRONS	EQ	28.45	28.45	26.90	27.60	27.65	27.60	17151	4.718833e+05	2016-03-15	179	INE144J01027
492	492	20MICRONS	EQ	32.40	33.50	31.60	32.40	32.15	32.75	17422	5.712948e+05	2016-04-08	203	INE144J01027
493	493	20MICRONS	EQ	38.10	39.60	37.85	38.30	38.50	38.95	135931	5.229043e+06	2016-09-23	877	INE144J01027
494	494	20MICRONS	EQ	60.10	60.80	57.90	58.30	58.30	59.60	174410	1.031882e+07	2017-12-29	1391	INE144J01027

495 rows × 14 columns

Entire dataset record for Microns

```
stocks.index = stocks['Date']
stocks
```

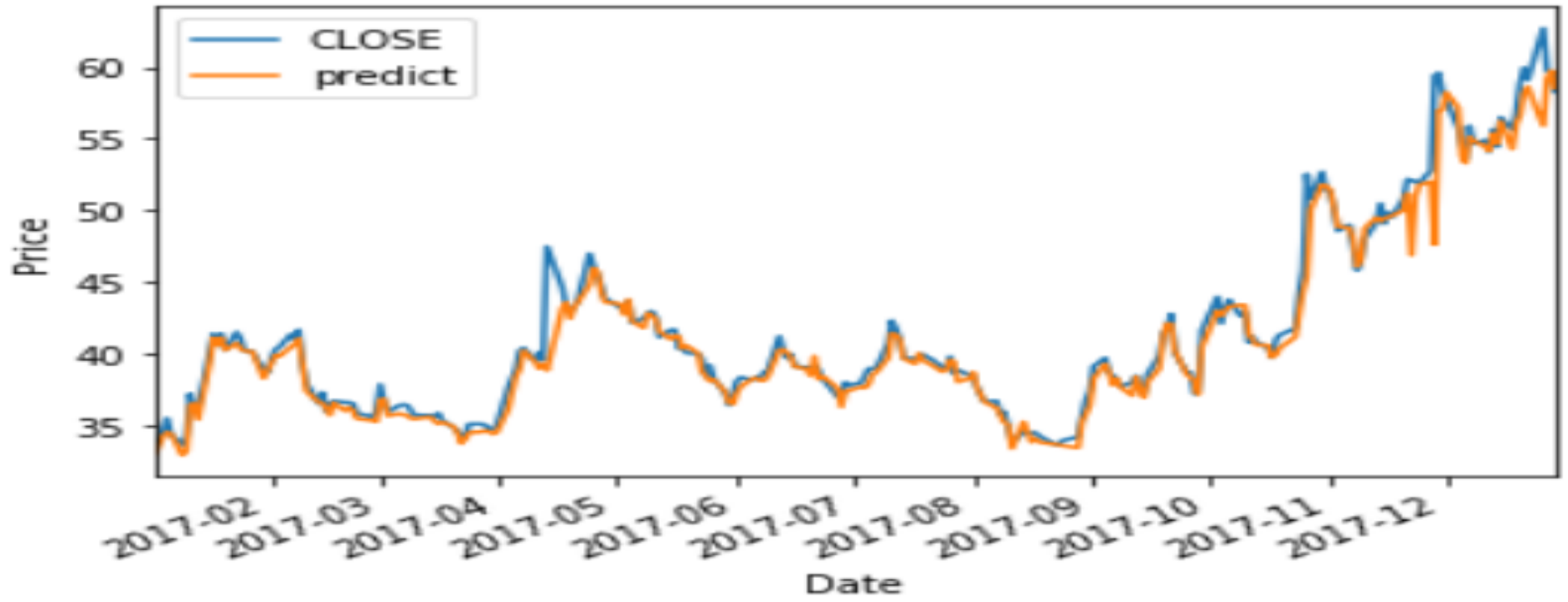
Date	OPEN	HIGH	LOW	CLOSE	TOTTRDQTY	Date	PREVCLOSE	TOTTRDVAL	TOTALTRADES
2017-06-28	37.80	37.80	36.15	36.85	27130	2017-06-28	37.05	9.946579e+05	202
2017-11-28	52.90	61.70	52.90	59.40	2489935	2017-11-28	52.80	1.481265e+08	14196
2017-02-28	36.15	36.30	35.60	36.05	40594	2017-02-28	35.65	1.461975e+06	193
2017-04-28	44.35	44.80	43.55	43.90	90170	2017-04-28	44.80	3.984352e+06	597
2017-03-20	36.00	36.75	34.20	34.90	77310	2017-03-20	35.40	2.709788e+06	413
...	...	...	...	...	...	...	...	...	...
2016-06-08	28.45	28.45	27.70	28.30	8045	2016-06-08	28.05	2.263841e+05	100
2016-03-15	28.45	28.45	26.90	27.60	17151	2016-03-15	27.60	4.718833e+05	179
2016-04-08	32.40	33.50	31.60	32.40	17422	2016-04-08	32.75	5.712948e+05	203
2016-09-23	38.10	39.60	37.85	38.30	135931	2016-09-23	38.95	5.229043e+06	877
2017-12-29	60.10	60.80	57.90	58.30	174410	2017-12-29	59.60	1.031882e+07	1391

495 rows × 9 columns

Filtering out excess data

```
plt.show()
```

```
Accuracy of model is:   0.9628088246357632
```



Result

THANK YOU