Low-Frequency Trading
Using A
Statistical Arbitrage Algorithm

Dayne Fernandes1, Satyam Singh2, Aditya Singh3, and Namit Bhandari4

1. Introduction

The logic behind statistical arbitrage is to find stocks with a significantly high correlation, such that their trend will remain the same in the future. In this project, we aim to apply statistics to find arbitrage opportunities in the stock market using highly correlated stocks in the same industry. The pairs are identified by using a correlation matrix and identifying stocks that surpass the threshold set by us, while trading signals are predicted by using statistics. The end result of our analysis is to prove that the arbitrage mechanism can be applied in the Indian Banking sector stocks.

2. Data & Methods

We decided to limit our universe of assets to Banking sector stocks. We choose equities listed on NSE because they are liquid enough to facilitate decent changes on a day-to-day basis. We chose 9 of the top Banking sector stocks such as 'HDFCBANK.NS', 'SBIN.NS', 'AXISBANK.NS', 'KOTAKBANK.NS', 'INDUSINDBK.NS', 'RBLBANK.NS', 'FEDERALBNK.NS', 'CANBK.NS', 'BANKBARODA.NS'. The Code and simulations are run on a Jupyter notebook. We utilized Python packages such as Pandas, NumPy, yfinance, and SciPy for our analysis.

3. Strategy

3.1 Background

If we determine that there's a strong relationship between two companies, if the stock price of one of them moves in a certain direction, then the stock price of entity 2 is also expected to make a similar move. If not, then there could be a trading opportunity.

We can determine a fair value for either of the stock prices to determine if one of them is overvalued or undervalued on a certain day. The fair value can be determined using statistical techniques like regression, given that they fulfill the other tests.

In terms of arbitrage, our goal would be to buy undervalued stocks & sell overvalued stocks and make a profit from the difference.

3.2 Algorithm

We obtained all our data from Yahoo Finance using an API to obtain 6 years of historic data on a day-to-day basis from 25th November 2016 to 25th November 2022. On obtaining this data we used statistics to determine the correlation between the pairs of stocks, and plotted it in the form of a correlation matrix.

Table 1: Correlation Matrix Of Various Assets

	HDFCBANK.NS	SBIN.NS	AXISBANK.NS	KOTAKBANK.NS	INDUSINDBK.NS	RBLBANK.NS	FEDERALBNK.NS	CANBK.NS	BANKBARODA.NS
HDFCBANK.NS	1.000	0.315	0.692	0.564	0.681	0.515	0.326	0.312	0.305
SBIN.NS	0.315	1.000	0.681	0.304	0.572	-0.485	0.915	0.943	0.901
AXISBANK.NS	0.692	0.681	1.000	0.274	0.816	0.098	0.762	0.699	0.760
KOTAKBANK.NS	0.564	0.304	0.274	1.000	0.469	0.229	0.222	0.248	0.145
INDUSINDBK.NS	0.681	0.572	0.816	0.469	1.000	0.186	0.680	0.505	0.626
RBLBANK.NS	0.515	-0.485	0.098	0.229	0.186	1.000	-0.371	-0.427	-0.386
FEDERALBNK.NS	0.326	0.915	0.762	0.222	0.680	-0.371	1.000	0.922	0.966
CANBK.NS	0.312	0.943	0.699	0.248	0.505	-0.427	0.922	1.000	0.947
BANKBARODA.N	0.305	0.901	0.760	0.145	0.626	-0.386	0.966	0.947	1.000

For our analysis, we decided to only include pairs with correlations greater than or equal to 0.875

Table 2: Pairs Taken For Further Analysis

Stock1	Stock2
SBIN.NS	FEDERALBNK.NS
SBIN.NS	CANBK.NS
FEDERALBNK.NS	CANBK.NS
SBIN.NS	BANKBARODA.NS
FEDERALBNK.NS	BANKBARODA.NS
CANBK.NS	BANKBARODA.NS

Our algorithm selects a single pair at a time and uses the data over one year to determine the slope and intercept. In addition to conducting our regression, we also needed to test if our data is stationary.

Stationarity means that the statistical properties of a time series, i.e., mean, variance, and covariance, do not change over time, implying that the time series has no unit root.

In the case of stationarity, trading signals are generated assuming both stocks' prices will eventually revert to the mean. Hence, we can take the advantage of the prices that deviate from the mean for a short period of time.

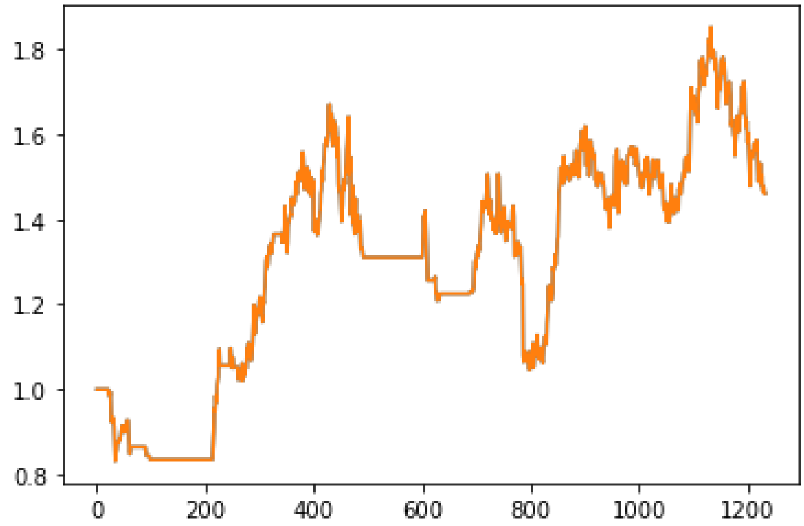
We used the t-statistic value for 1% i.e., -2.5, as a threshold for placing a trade order on a particular day. Any value less than -2.5 was acceptable for our algorithm to take a long and short position on the assets.

Since such a trader would generally be trading at very large volumes, we assumed that the fee charged on the trade would be nearly 0. Our analysis was plotted in order to see the overall profit made in terms of percentage.

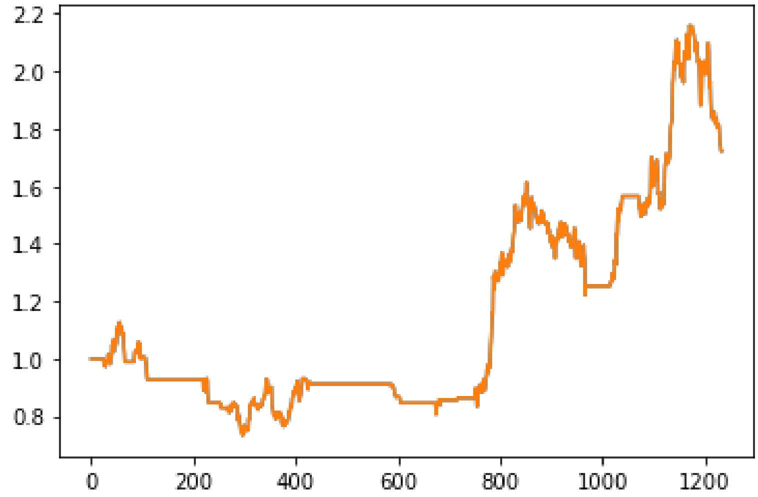
Since our aim was to determine long-term trading strategies, we had to compute the 5-year CAGR. The 5-year CAGR was obtained by dividing the cumulative profit by 5 and generating the average of all the pairs.

3.3 Results

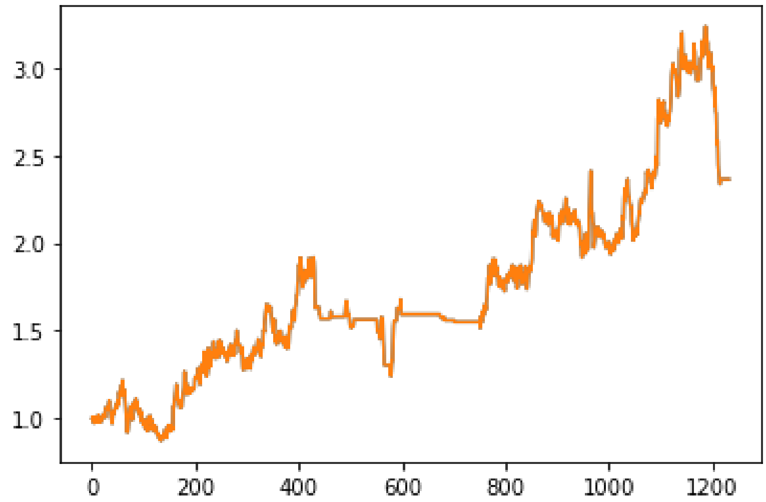
Graph 1: Returns on SBI & Federal Bank



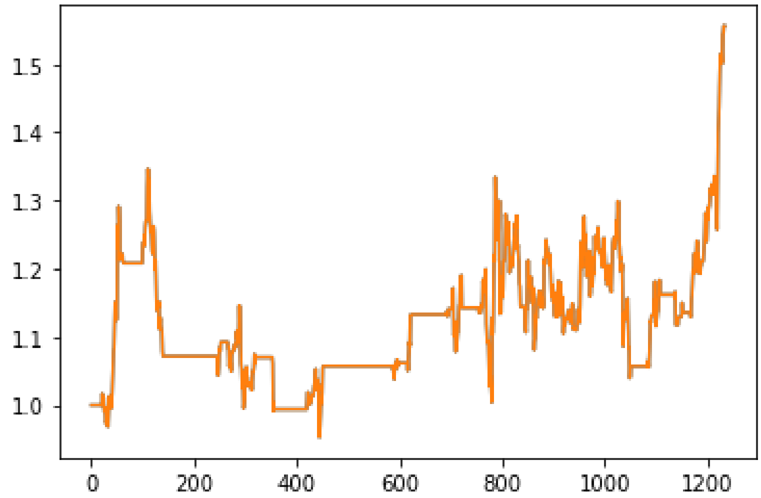
Graph 2: Returns on SBI and Canara Bank



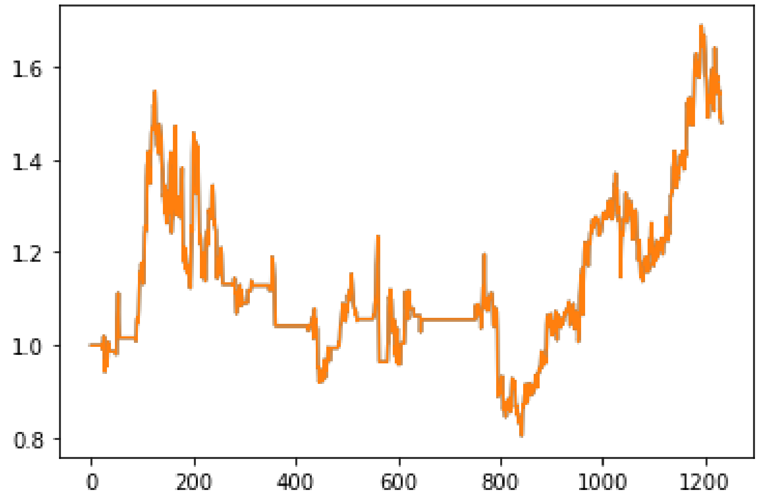
Graph 3: Returns on Federal Bank & Canara Bank



Graph 4: Returns on SBI and Bank Of Baroda



Graph 5: Returns on Federal Bank and Bank Of Baroda



Graph 6: Returns on Canara Bank and Bank Of Baroda

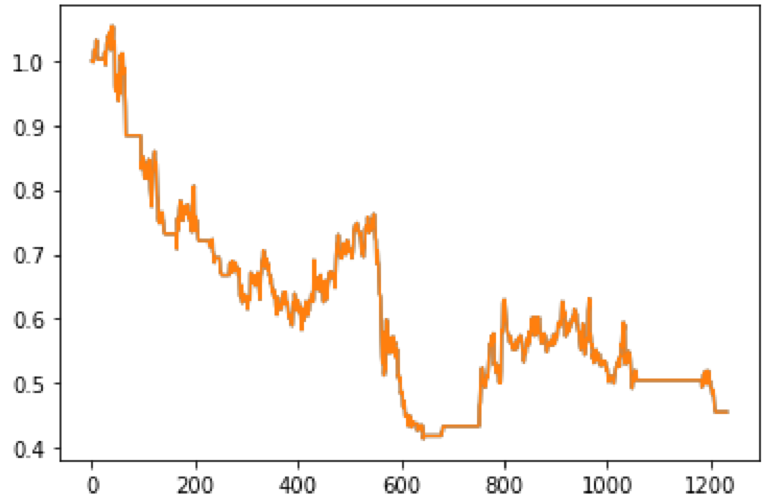


Table 3: 5 Year CAGR

Company	SBIN and FEDERALBNK	SBIN and CANBK	FEDERALBNK and CANBK	SBIN and BANKBARODA	FEDERALBNK and BANKBARODA	CANBK and BANKBARODA	5 Year CAGR
Returns	9.19%	14.43%	27.27%	11.13%	9.57%	-10.89%	10.12%

We finally computed the risk profile of the portfolio of the 6 asset pairs to determine how risky was it to incorporate this as a dedicated fund.

Table 4: Annualised Standard Deviation Of The Asset Pairs

Pairs	Annual Standard deviation
SBIN.NS+FEDERALBNK.NS	0.241
SBIN.NS+CANBK.NS	0.219
FEDERALBNK.NS+CANBK.NS	0.328
SBIN.NS+BANKBARODA.NS	0.216
FEDERALBNK.NS+BANKBARODA.NS	0.314
CANBK.NS+BANKBARODA.NS.csv	0.261

Table 5: Correlation matrix of the portfolio assets

	SBIN.NS + FEDERALBNK.NS	SBIN.NS + CANBK.NS	FEDERALBNK.NS + CANBK.NS	SBIN.NS + BANKBARODA.NS	FEDERALBNK.NS + BANKBARODA.NS	CANBK.NS + BANKBARODA.NS
SBIN.NS+FEDERALBNK.NS	1.0000	-0.0615	0.0954	-0.0284	0.0930	0.0021
SBIN.NS+CANBK.NS	-0.0615	1.0000	0.2120	0.2122	-0.0105	0.1449
FEDERALBNK.NS+CANBK.NS	0.0954	0.2120	1.0000	0.0107	0.0797	0.1166
SBIN.NS+BANKBARODA.NS	-0.0284	0.2122	0.0107	1.0000	0.1621	-0.0455
FEDERALBNK.NS+BANKBARODA.NS	0.0930	-0.0105	0.0797	0.1621	1.0000	-0.0057
CANBK.NS+BANKBARODA.NS	0.0021	0.1449	0.1166	-0.0455	-0.0057	1.0000

The overall risk profile is calculated using the formula:

$$\sigma_P^2 = \sum_{i=1}^{n} \sum_{j=1}^{n} w_i w_j \sigma_{i,j}$$

Portfolio variance

The risk of the portfolio computed was 12.55%

3.4 Discussion

We believe the trading strategy can make money if the right pairs are found over a longer duration of time. By attempting to buy undervalued stocks & sell over priced stocks, we can see that we generate a profit 5 out of 6 times. The overall profit generated by trading all 6 with equal amounts will give us a return of 10.12% CAGR.

Just to give some perspective, the following is a table with the 5-year CAGR of Banking sector focussed mutual funds in India as obtained from moneycontrol.com.

Table 6: Banking Sector focussed mutual funds

Scheme Name	AuM (Cr)	5Y CAGR
Tata Banking And Financial Services Fund	1,305.46	12.17%
SBI Banking & Financial Services Fund	3,990.54	11.92%
Taurus Banking and Financial Services Fund	9.62	11.15%
Nippon India Banking & Financial Services Fund	4,019.67	8.82%
Baroda BNP Paribas Banking and Financial Services Fund	74.14	8.40%
ICICI Prudential Banking and Financial Services Fund	5,518.62	8.28%
Aditya Birla Sun Life Banking and Financial Services Fund	2,531.77	8.21%
UTI Banking and Financial Services Fund	862.65	5.39%
LIC MF BANKING & FINANCIAL SERVICES FUND	111.85	4.99%

Source: moneycontrol.com

As can be seen, the average return from the above mutual funds is 8.81%. Our algorithm applied to this sector can achieve 1.31% higher returns, and effectively puts us in the 4th spot among the above funds.

The overall risk profile of the 6 assets combined was 12.55%, thus effectively bringing down the risk considerably instead of investing in a single pair alone, this was largely due to a negative correlation coefficient due to multiple pairs.

4. References

[1] Fama, Eugene F. and Kenneth R. French, 1996, Multifactor explanations of asset pricing anomalies, Journal of Finance 51, 131-155.

[2] Jegadeesh, Narasimhan, and Sheridan Titman, 1993, Returns to buying winners and Selling Losers: Implications for stock market efficiency, Journal of Finance 48, 65-91.

[3] Bakshi, Gurdip and Zhiwu Chen, 1997, "Stock Valuation in Dynamic Economies," working paper, Ohio State University.

[4] Ready, Mark, 1997, Profits from technical trading rules, working paper, University of Wisconsin, Madison