Designing Efficient Pair-Trading Strategies Using Cointegration for the Indian Stock Market

Kartik Pandey
University Teaching Department
Chhattisgarh Swami Vivekanand Technical University
Bhilai 491001, India
Email: kartik0pandey00@gmail.com

Abstract

A pair-trading strategy is an approach that utilizes the fluctuations between prices of a pair of stocks in a short-term time frame, while in the long-term the pair may exhibit a strong association and co-movement pattern. This paper presents a cointegration-based approach that identifies stocks listed in five sectors of the National Stock Exchange (NSE) of India for designing efficient pair-trading portfolios. Based on stock prices from Jan 1, 2018, to Dec 31, 2020, the cointegrated stocks are identified, and the pairs are formed. The pair-trading portfolios are evaluated on their annual returns for the year 2021. The results show that pairs of stocks from the auto and the realty sectors, in general, yielded the highest returns among the five sectors studied. However, two pairs from the IT sector resulted in negative returns.

1 Introduction

Pair trading is a market-neutral strategy involving the simultaneous buying and selling of two correlated stocks to capitalize on their price divergence. When prices of the two stocks exhibit a significant deviation from their historical spread, the over-performing stock is shorted, and the underperforming stock is bought. Once the prices revert to their historical equilibrium, the positions are reversed, thus generating profits.

This strategy assumes that the prices of the two stocks are cointegrated, meaning they share a long-term equilibrium relationship despite short-term deviations. This paper explores the application of a cointegration-based pair-trading strategy in the Indian stock market, focusing on five sectors: Auto, Banking, IT, Pharma, and Realty.

2 Research Methodology

The methodology consists of nine steps that are followed sequentially to design the pair-trading portfolios. These steps are explained in detail in the following subsections.

2.1 Choice of Sectors for Analysis

From the list of sectors listed on the National Stock Exchange (NSE) of India, five sectors were chosen based on their diversity: Auto, Banking, Information Technology (IT), Pharma, and Realty. For each sector, the top ten stocks were selected based on their influence on the respective sectors. The NSE's report published on Dec 31, 2021, was used for selecting the stocks.

2.2 Historical Stock Prices

The historical stock prices were extracted using the `pandas_datareader` module from Yahoo Finance for the period from Jan 1, 2018, to Dec 31, 2020. The close values of the stock prices were used for building the pair-trading models. The evaluation is done based on the annual returns of the models on the test data from Jan 1, 2021, to Dec 31, 2021. In this study, pairs from specific sectors were considered rather than from diversified indices like NIFTY 50.

2.3 Correlation Matrix

At this step, a correlation matrix is generated for each sector using the `heatmap` function of the `seaborn` library to visualize the correlation coefficients among the pairs of stocks in that sector. Since there are 10 stocks in a sector, Pearson's product-moment correlation coefficients will be available for 45 distinct pairs. The correlation coefficients are computed based on the return values of the close prices. The `pct_change` function is used to compute the return series for each stock.

2.4 Identifying Cointegrated Pairs of Stocks

A pair of time series variables are said to be cointegrated if they exhibit significant correlation in the long-term, even though no relationship may be apparent in the short term. To identify the pairs that are cointegrated, the `coint` function from Python's `statsmodels` module is used. The null hypothesis assumes that a given pair is not cointegrated. Pairs with p-values below 0.05 are assumed to be cointegrated and are considered for pair-trading portfolios.

2.5 Building OLS Regression Models

For the pairs that passed the cointegration test, an Ordinary Least Squares (OLS) regression model is built. The stock with the higher mean close price is chosen as the predictor variable, and the OLS model is fitted using the `OLS` function from the `statsmodels` library. The model output includes several components, including:

- Hedge ratio and the p-value of its t-statistics.
- P-value of the F-statistics.
- Durbin-Watson test statistic.
- Prob. of the Jarque-Bera test statistics.

For the regression model to be accurate, the p-values of the F-statistics and Omnibus test should be significant (below 0.05), and the Durbin-Watson test statistic should be close to 2, indicating no autocorrelation among the residuals.

2.6 Stationarity of Residuals

To validate the cointegration of the pair, the residuals of the OLS model are tested for stationarity using the Augmented Dickey-Fuller (ADF) test. The `adfuller` function is used for this purpose. If the test statistic is negative and smaller than the critical value at the 1% significance level, or if it is positive and greater than the critical value, the null hypothesis of a unit root (non-stationarity) is rejected, confirming the stationarity of the residuals.

2.7 Generating Trading Signals

Once cointegration is validated, the trading signals for the pair are generated. These signals are based on the spread between the two stocks, and when the spread deviates from its mean by a predefined threshold, positions are taken. The long position is taken on the stock whose price falls, while a short position is taken on the stock whose price rises.

3 Results

The pair-trading portfolios were evaluated for the test period of Jan 1, 2021, to Dec 31, 2021. The results show that pairs from the auto and realty sectors yielded the highest returns, while two pairs from the IT sector and one pair each from the banking and pharma sectors yielded negative returns. The portfolios for the remaining pairs showed moderate to high positive returns.

The top-performing sectors, the auto and realty sectors, showed consistent profits from the cointegrated pairs. The success of the cointegration-based approach in these sectors suggests its potential for practical trading strategies. However, negative returns in some sectors, particularly the IT sector, highlight the need for caution when selecting pairs based on cointegration.

4 Conclusion and Future Work

This paper proposed a cointegration-based approach for pair-trading in the Indian stock market, focusing on five diverse sectors. The results demonstrate that cointegrated pairs, particularly from the auto and realty sectors, can yield significant returns. However, the approach has limitations, as negative returns were observed in some sectors, such as IT and pharma.

Future work can explore alternative methods for pair-trading, such as clustering-based approaches, to compare with the cointegration method. Additionally, refining the trigger points for trading signals and incorporating more robust risk management strategies could enhance the effectiveness of pair-trading models in different sectors.