Optimizing Statistical Arbitrage Models

Guo Liang

Institute of Statistics and Big Data

Renmin University of China

April 17, 2023

Presentation Overview

Introduction

Background:

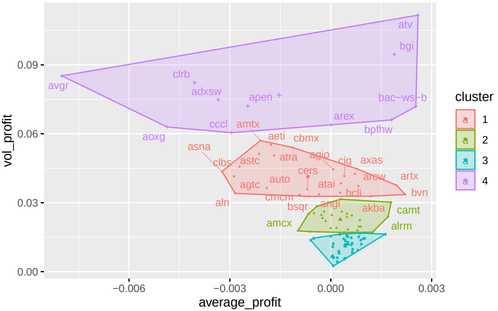
Statistical arbitrage is using statistical analysis of historical data to guide arbitrage trading. It estimates the probability distribution of relevant variables and integrates fundamental data to guide arbitrage trading.

Aims:

Choose stock pairs from the data and create a trading strategy based on the stock portfolio with the goal of achieving a high annualized return, a high Sharpe ratio, and a low maximum drawdown in the backtest.

- Our data is from The U.S. Daily Price/Volume Data for Stocks and ETFs.
- The dataset provide the full historical daily price and volume data for all US-based stocks and ETFs trading on the New York stock exchange (NYSE) and national association of securities dealers automated quotations (NASDAQ).
- The data is presented in CSV format as follows: Date, Open, High, Low, Close, Volume, OpenInt. The prices have been adjusted for dividends and splits.

- *Step 1:* Remove stocks containing little data.
- *Step 2:* Estimate the log-return and volatility for each of the stocks to be used, and use this as a basis for clustering.
- An example:



- Check the stationary of the stock and the stationary after differencing.

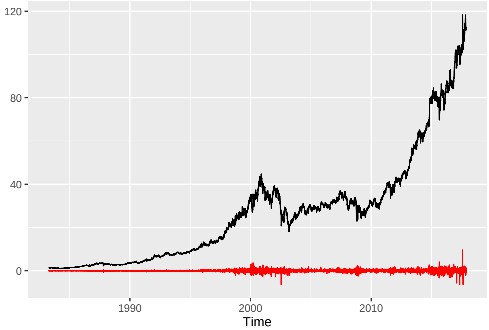


Figure: Stationary test.

- Use the Engle-Granger method to test the co-integration.

- Price of stock pairs with co-integration relationship.

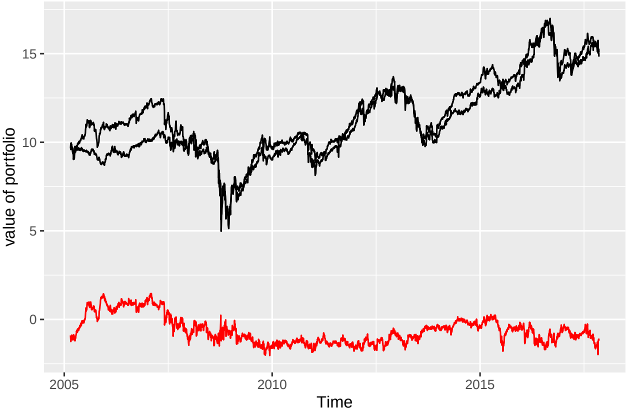
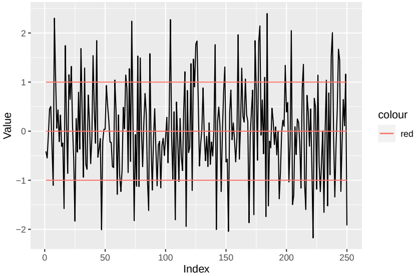


Figure: Co-integration stock pairs.

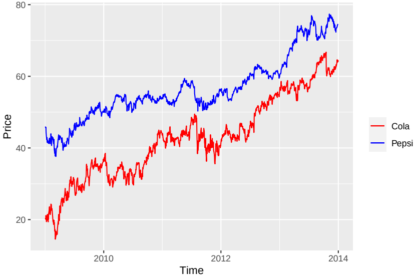
Model

- **Idea:** If the two stocks are fully co-integrated, the portfolio they form is a mean-reverting series. When the portfolio value deviates from the mean, we can then choose the appropriate strategy (long or short). When the price returns to the mean, the position is sold or liquidated to gain a profit.
- A toy example:



Model

- **Stock selection:** We select data for Coca-Cola and Pepsi for the five years from 2009 to 2014. And the data of the first three years are set as the training set and the data of the last two years are set as the test set.



- **Questions:**
  - How to set threshold?
  - What is the mean?
  - When to liquidate a position and stop loss?

Optimal Threshold

- **Interpolation:** In contrast to the general situation, we accept threshold asymmetry. Annualized returns, Sharpe ratios, and maximum drawdown for in-sample data are calculated using several threshold combinations, and the best threshold combination is chosen.

- **Optimization:** We select the strategy with the lowest maximum drawdown and maximize the following equation in the experiment.

$$\frac{1}{4}(\text{Annualized returns} + \text{Sharpe ratios}) - \frac{1}{2}\text{max drawdown}$$

Optimal Threshold

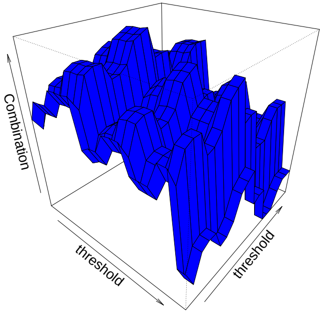


Figure: Change of combination about threshold.

- **Upper threshold:** σ **Lower threshold:** -0.85σ
- **Annualized returns:** 40.33058% **Sharpe ratios:** 2.1316185 **max drawdown:** 11.82071%

Optimal Mean and Volatility

Assumption

- The series after co-integration is uncorrelated.
- The series after co-integration is normally distributed at each point.
- Mean value is constantly updated.

- We use the Bayesian method to estimate the mean and volatility.
- We use a conjugate prior

$$(\mu, \sigma^2) \sim N - Inv - \chi^2(\mu_0, \sigma_0^2/\kappa_0; v_0, \sigma_0^2)$$

- The hyper parameters are estimated with in-sample data.

Profit and Robust

- The value of the portfolio sometimes does not cross the average but is close to crossing, when closing the position will bring gains.
- Establishing a position when the value of the portfolio crosses a threshold and reverts to the mean. This reduces the probability of excessive deviation of the portfolio value from the mean, i.e., reduces the probability of stop loss.

Comparisons

Table: Comparison of the results of in-sample.

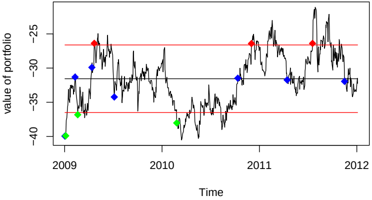
Method	*Ann-Return*	*Sharpe-Ratio*	*Max-Drawdown*
General	30.92%	1.44	14.05%
Opt-Thre	40.33%	2.13	11.82%
Opt-Thre-Post	33.40%	1.30	16.93%
Bayes	34.56%	1.93	11.82%
Bayes-Post	33.40%	1.30	16.93%

Comparisons

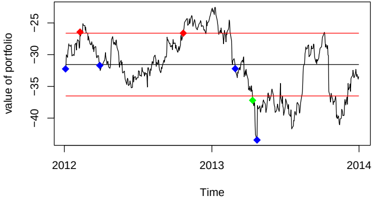
Table: Comparison of the results of out-of-sample.

Method	*Ann-Return*	*Sharpe-Ratio*	*Max-Drawdown*
General	15.76%	0.95	12.82%
Opt-Thre	30.19%	1.87	15.34%
Opt-Thre-Post	25.34%	1.25	15.20%
Bayes	40.15%	1.10	17.23%
Bayes-Post	43.84%	1.18	15.65%

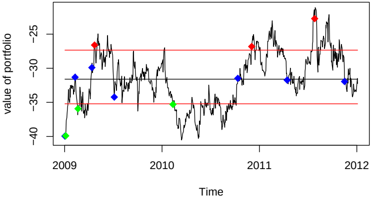
Trading Points



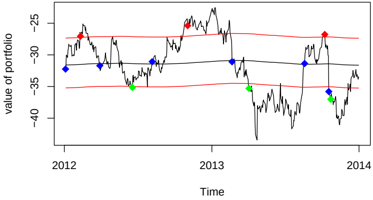
(a) General for in-sample



(b) General for out-of-sample



(c) Bayes-Post for in-sample



(d) Bayes-Post for out-of-sample

Excess Return

- We use the S&P 500 as a benchmark to see the excess returns of our strategy.

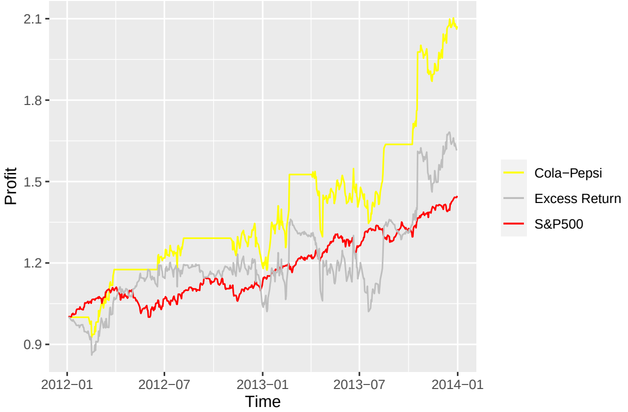


Figure: Portfolio excess return on S&P 500 index.

Discussion

- Bayesian methods can better estimate the mean and variance, thus increasing the probability of profitability.
- Postponing the opening and closing of positions early can increase stability (reduce the probability of stop loss), but since our data is daily, the effect may not be very obvious.
- **Future work:**
    - Finding minute stock data that makes the algorithm tradeable at every moment of the day.
    - Optimize the algorithm and improve the algorithm rate.

References

Vidyamurthy, G. (2004)
Pairs Trading: quantitative methods and analysis.
Vol. 217.

Zeng, Z., & Lee, C. G. (2014)
Pairs trading: optimal thresholds and profitability.
Quantitative Finance 14(11), 1881-1893.

Sarmento, S. M., & Horta, N. (2020)
Enhancing a pairs trading strategy with the application of machine learning.
Expert Systems with Applications 14(11), 158, 113490.