Evaluation of copula based pair-trading with bootstrap simulation

Pan, Jiacheng

2014

Pan, J. (2014, March). Evaluation of copula based pair-trading with bootstrap simulation. Presented at Discover URECA @ NTU poster exhibition and competition, Nanyang Technological University, Singapore.

https://hdl.handle.net/10356/102291

NANYANG TECHNOLOGICAL UNIVERSITY

URECA
Undergraduate Research Experience on CAmpus

Category: 1
Nanyang Business School

Student: Pan Jiacheng

Project ID: NBS13019

Evaluation of Copula based Pair-Trading with Bootstrap Simulation

Introduction

Pair trading involves trading two securities with similar trend by different trading positions when their prices diverge. The mean-reverting property thus guarantees profits for the investors. Distance method uses correlation coefficient to evaluate the dependency between securities, which makes sense only when the data are normally distributed. However, most financial assets are not normally distributed but skewed with heavier tails.

Liew, Wu (2013) proposed Copula method, modeling the dependency structure without the assumption of normality, to address this limitation, and found Copula method can identify more trading opportunities and generate more profits than distance method.

This project uses bootstrap simulation to model the data from U.S Utility market. The rationale for bootstrap simulation is that the reality is only one realization of the underlying distribution and may not turn out to be what happens if it were to be observed again. Using bootstrap simulation, we are taking the sample structure as an approximate of population structure. Data of similar patterns are then reconstructed using the modeled structure and the profitability of trading strategies on these data is examined accordingly.

Methodology

There are two periods in the pair trading strategy, formation period and trading period. Minimum sums of squared daily stock price differences in the formation period are adopted to select the top 5, 20 pairs. The pairs will be updated at a predetermined rate.

In Distance method, if distance exceeds threshold d, here d=two standard deviation, we long the undervalued stock and short the overvalued one. The position is closed whenever the distances fall below threshold, the pairs have been updated, or we have reached the end of the data set.

In Copula method, marginal distributions, $F_1(x)$ *and* $F_2(x)$, are modeled for every stock in selected pairs, and the best copula with the best BIC and the best joint distribution by Sklar's theorem, $H(x,y) = C(F(x),G(y))$, are fitted to model the dependency structure between stocks. Conditional probabilities are given by:

$$P(U \le u|V = v) = \frac{\partial C(u,v)}{\partial v}$$

$$P(V \le v|U = u) = \frac{\partial C(u,v)}{\partial u}$$

The stock whose conditional probability of the price being lower than the current price greater than 0.5 is overvalued, and less than 0.5, undervalued. Choosing 0.95 and 0.05 as the upper and lower probability criteria, positions are to be executed on each stocks.

Bootstrap simulation uses sample parameters to generate new samples. Therefore in a manner similar to copula method, we first use the data in formation period to model the marginal distribution of each stock. Log returns $ln\left(\frac{P_{i,t}}{P_{i,t-1}}\right)$ in formation period are modeled and the best possible copula is fitted to marginal distribution by examining its Bayesian Information Criterion (BIC). By Sklar's Theorem, $H(x,y) = C(F(x),G(y))$, the estimated joint distribution $\hat{H}(x,y)$ is then obtained from the fitted copula and pseudo random samples are to be generated by the MATLAB command *copularnd*. The inverse cumulative distribution function converts the generated number back to log returns $ln\left(\frac{P_{i,t}}{P_{i,t-1}}\right)$ that have been modeled. Prices are then generated in the following way $P_{i,t} = P_{i,t-1} \times exp(ln\left(\frac{P_{i,t}}{P_{i,t-1}}\right))$. Modeling log returns avoids the generation of negative prices. The whole process is done dynamically as formation period are constantly updated in a manner similar to the previous case. Both distance method and copula method are then implemented on the simulated data to test the consistency of the strategies.

	Copula Method		Distance Method	
	Top 5	Top 20	Top 5	Top 10
Average excess returns	0.004442602	0.000576336	0.000838113	0.000362064
t-Statistic	3.045277652	0.663988201	0.587720784	0.582210841
Median	0.004545806	0.000361836	0.000000000	-0.000617543
Standard Deviation	0.014733658	0.008766285	0.014402288	0.006280652
Skewness	-0.386026499	-0.222113159	2.0470628	0.266401873
Kurtosis	6.500988529	5.293057547	12.85516673	3.152836424
Minimum	-0.059887987	-0.026432975	-0.034218096	-0.015595493
Maximum	0.054219817	0.033919452	0.081379738	0.018739327
Observations with excess return<0	34.31372549	47.05882353	48.03921569	51.96078431

Results and Conclusion

The U.S Utility market contains 89 stocks, and 10 stocks out of the 89 are chosen as an proxy for the whole market. Evaluation is conducted for the simulated data and the result is displayed in the above tables.

Comparison between two average excess returns shows Copula method outperforms Distance method with significant t-statistics, thus we conclude Copula method can indicate opening and closing point more accurately thus generating more profits than Distance method.

Reference

Gatev, E., Goetzmann, W. N., and Rouwenhorst, K. G. (2006) Pairs Trading: Performance of a Relative-Value Arbitrage Rule. The Review of Financial Studies 19(3): 797–827.

Liew, R. Q. and Wu, Y. (2013) Pairs trading: A copula approach. Journal of Derivatives and Hedge Funds, Vol 19(1), pp12-30.

Project Title: Copula based Pairs-Trading Strategy

Supervisor: Assoc Prof Wu Yuan

Collaborator: Mr Hong Shengkai