**Strategy – Pair Trading - Copula**

Archimedean Copula

An Archimedean Copula with a strict generator has the form :

Kendall’s Tau

The Relation between Copula generation function and Kendall rank correlation tau  
(Requiring the Data pairs are IID)

Proof:

(1)

Note that

(2)

Where

(3)

By the substitution ,   
Then the Jocobian

Then

For more details, please refer to “The Joy of Copulas : Bivariate Distribution with Uniform Marginals” written by Christian Genest and Jock MacKay

Copula Strategy

Step 1 : Selecting the Paired Stocks

(1)Selecting (Fundamentally) related pairs

(2)Estimating Marginal Distributions of log-return ()  
Using the Empirical distribution (Using statsmodel library : ECDF)

(3)Ljung-Box Test on (Filter out the asset if the the series cannot pass the test (showing not IID))

(4)Filter the trading pair with statistical correlation  
Graphical user interface

Description automatically generated with low confidence

Finding the Kendall Rank Correlation of the Pairs’ log-return (Using the SciPy Stat Functions : kendalltau)

Selecting the pair with highest Kendall’s tau

Step 2 : Estimating Copula Parameters  
-Clayton Copula  
Generator Function :

-Gumbel Copula  
Generator Function :

-Frank Copula  
Generator Function :

Step 3 : Selecting the Best Fitting Copula  
By AIC

|  |  |
| --- | --- |
| Copula | Copula density |
| Clayton |  |
| Gumbel Copula |  |
| Frank Copula |  |

Step 4 : Generating the Trading Signals

Basic Idea  
We execute the pair trade when the following case (similar as u , v swap) since such probability show the high likelihood of reversing motion

In other words

Consider to be the Joint CDF (Copula) and to be the copula density

Long X short Y when   
Short X Long Y when

Background

Training Formation Period : 3 years (Jun 2018 – Apr 2022)  
Rolling Formation Period : (Hyperparameter)  
[30 , 60 , 90 , 120 , 150 , 180 , 210 , 240 , 270 , 300 , 330 , 360] days  
Trading Frequency : Daily or Hour  
Marginal Copula Threshold : 0.01

We would use Rolling Sample Data to estimate   
(1) Empirical CDF -> Converting the log return to uniform variable (u)  
(2) Testing the whether series of u is iid or not by Ljung-Box Test  
(3) -> Selecting the best pair  
(4) selecting the best fitting Copula  
We would update it at the beginning of each month (Step 1 to 3)   
We would fit the daily data every day for the selected pair and check whether the trading signal is generated

Result

(Frequency : Daily)

|  |  |  |  |
| --- | --- | --- | --- |
| Rolling Window (Days) | Annualized Return | Sharpe | Calmar |
| 30 | -0.079 | -0.35 | -0.13 |
| 60 | -0.047 | -0.30 | -0.06 |
| 90 | -0.028 | -0.15 | -0.05 |
| 120 | -0.021 | -0.17 | -0.03 |
| 150 | -0.105 | -0.68 | -0.17 |
| 180 | 0.038 | 0.30 | 0.05 |
| 210 | -0.060 | -0.70 | -0.08 |
| 240 | -0.040 | -0.86 | -0.05 |
| 270 | -0.024 | -0.34 | -0.03 |
| 300 | -0.051 | -0.29 | -0.08 |
| 330 | -0.140 | -0.94 | -0.26 |
| 360 | -0.053 | -0.30 | -0.08 |

(Frequency : Hour)

|  |  |  |  |
| --- | --- | --- | --- |
| Rolling Window (Days) | Annualized Return | Sharpe | Calmar |
| 30 | -0.17 | -1.35 | -0.34 |
| 60 | -0.19 | -1.67 | -0.40 |
| 90 | -0.15 | -2.19 | -0.29 |
| 120 | -0.14 | -1.75 | -0.23 |
| 150 | -0.16 | -1.95 | -0.30 |
| 180 | -0.08 | -1.32 | -0.11 |
| 210 | -0.07 | -1.51 | -0.10 |
| 240 | -0.12 | -1.21 | -0.20 |
| 270 | -0.04 | -0.49 | -0.05 |
| 300 | -0.23 | -1.64 | -0.61 |
| 330 | -0.17 | -1.4 | -0.37 |
| 360 | -0.04 | -0.92 | -0.05 |

Discussion

From the above result, we can see that the strategy does not perform well in general. The possible reason could be related to the IID requirement on the pair data

The data is required to be IID in order to estimate the Kendall’s tau and Copula Parameter. This means that there is no any dependency or correlation between the pairs. In addition, the conditional marginal distribution only measure the current situation. If the distribution function value exceed the threshold, it only imply that the log-return pair are under an extreme situation today. This does not reflect any information for the future given their IID property. Besides, when the log-return pair back to normal for the next time point, the motion may still against our bet like if we are long A and short B, the log return might slightly lower than zero and that of B might slightly higher than zero.