How to use Copula for pairs trading

We use the first 3 years of data to choose the best fitting copula and asset pair ("training formation period"). Next, we use a period of 8 years from ("the trading period"), to execute the strategy. During the trading period we use a rolling 12 month window of data to get the copula parameters ("rolling formation period").

Filter the trading pair with statistical correlation

To determine which stock pairs to include in the analysis, correlations between the pre-selected pairs are analysed. Below are three types of correlation measures we usually use in statistics:

Pearson Rank correlation

Kendall rank correlation

Spearman Rank correlation

We can get these coefficients in Python using functions from the stats library in SciPy.

The correlations have been calculated using daily log stock price returns during the training formation period. The Pearson correlation assumes that both variables should be normally distributed. Thus here we can use Kendall rank as the correlation measure and choose the pairs with the highest Kendall rank correlation to implement the pairs trading.

We get the daily historical closing price of our pairs ,compute log returns

Estimating Marginal Distributions of log-return

In order to construct the copula, we need to transform the log-return series to two uniformly distributed values u and v. This can be done by estimating the marginal distribution functions of individual tickers and plugging the return values into a distribution function. As we make no assumptions about the distribution of the two log-return series, here we use the empirical distribution function to approach the marginal distribution. The Python ECDF function from the statsmodel library gives us the Empirical CDF as a step function.

Estimating Copula Parameters

As discussed above, we estimate the copula parameter theta by the relationship between the copula and the dependence measure Kendall’s tau, for each of the Archimedean copulas.

Selecting the Best Fitting Copula

Once we get the parameter estimation for the copula functions, we use the AIC criteria to select the copula that provides the best fit in algorithm initialization.

$$AIC=-2L(\theta)+2kAIC=-2L(\theta)+2k$$

where $L(\theta)=\sum Tt=1logc(ut,vt;\theta)L(\theta)=\sum t=1Tlog c(ut,vt;\theta)$ is the log-likelihood function and k is the number of parameters, here k=1.

Generating the Trading Signals

The copula functions include all the information about the dependence structures of two return series. According to Stander Y, Marais D, Botha I. in Trading strategies with copulas, the fitted copula is used to derive the confidence bands for the conditional marginal distribution function of $C(v|\ u)C(v|\ u)$ and $C(u|\ v)C(u|\ v)$, that is the mispricing indexes. When the market observations fall outside the confidence band, it is an indication that pairs trading opportunity is available. Here we choose 95% as the upper confidence band, 5% as the lower confidence band as indicated in the paper. The confidence level was selected based on a back-test analysis in the paper that shows using 95% seems to lead to appropriate trading opportunities to be identified.

Given current returns $Rx,RyRx,Ry$ of stock X and stock Y, we define the "mis-pricing indexes" are:

$$MIX|Y=P(U\leq u|\ V\leq v)=\partial C(u,v)\partial vMIX|Y=P(U\leq u|\ V\leq v)=\partial C(u,v)\partial v$$

$$MIY|X=P(V\leq v|\ U\leq u)=\partial C(u,v)\partial uMIY|X=P(V\leq v|\ U\leq u)=\partial C(u,v)\partial u$$

For further mathematical proof, please refer to Xie W, Wu Y. Copula-based pairs trading strategy. The conditional probability formulas of bivariate copulas can be derived by taking partial derivatives of copula functions shown in Table 1. The results are as follows:

Gumbel Copula

$$C(v|\ u)=C(u,v;\theta)[(-\ln u)\theta+(-\ln v)\theta]1-\theta\theta(-\ln u)\theta-11uC(v|\ u)=C(u,v;\theta)[(-\ln u)\theta+(-\ln v)\theta]1-\theta\theta(-\ln u)\theta-11u$$

$$C(u|\ v)=C(u,v;\theta)[(-\ln u)\theta+(-\ln v)\theta]1-\theta\theta(-\ln v)\theta-11vC(u|\ v)=C(u,v;\theta)[(-\ln u)\theta+(-\ln v)\theta]1-\theta\theta(-\ln v)\theta-11v$$

Clayton Copula

$$C(v|\ u)=u-\theta-1(u-\theta+v-\theta-1)-1\theta-1C(v|\ u)=u-\theta-1(u-\theta+v-\theta-1)-1\theta-1$$

$$C(u|\ v)=v-\theta-1(u-\theta+v-\theta-1)-1\theta-1C(u|\ v)=v-\theta-1(u-\theta+v-\theta-1)-1\theta-1$$

Frank Copula

$$C(v|\ u)=(\exp(-\theta u)-1)(\exp(-\theta v)-1)+(\exp(-\theta v)-1)(\exp(-\theta u)-1)(\exp(-\theta v)-1)+(\exp(-\theta)-1)C(v|\ u)=(\exp(-\theta u)-1)(\exp(-\theta v)-1)+(\exp(-\theta v)-1)(\exp(-\theta u)-1)(\exp(-\theta v)-1)+(\exp(-\theta)-1)$$

$C(u| \ v)=(\exp(-\theta u)-1)(\exp(-\theta v)-1)+(\exp(-\theta u)-1)(\exp(-\theta u)-1)(\exp(-\theta v)-1)+(\exp(-\theta)-1)C(u| \ v)=(\exp(-\theta u)-1)(\exp(-\theta v)-1)+(\exp(-\theta u)-1)(\exp(-\theta u)-1)(\exp(-\theta v)-1)+(\exp(-\theta)-1)$

After selection of trading pairs and the best-fitted copulas, we take the following steps for trading. Please note we implement the Steps 1, 2, 3 and 4 on the first day of each month using the daily data for the last 12 months, which means our empirical distribution functions and copula parameters theta estimation are updated once a month.

In summary each month:

- During the 12 months' rolling formation period, daily close prices are used to calculate the daily log returns for the pair of tickers identified and then compute Kendall's rank correlation.
- Estimate the marginal distribution functions of log returns of X and Y, which are ecdf_x and ecdf_y separately.
- Plug Kendall's tau into copula parameter estimation functions to get the value of theta.
- Run linear regression over the two price series. The coefficient is used to determine how many shares of stock X and Y to buy and sell. For example, if the coefficient is 2, for every X share that is bought or sold, 2 units of Y are sold or bought.
- Finally during the trading period, each day we convert today's returns to u and v by using empirical distribution functions ecdf_x and ecdf_y. After that, two mispricing indexes are calculated every trading day by using the estimated copula C. The algorithm constructs short positions in X and long positions in Y on the days that $MI_{Y|X}<0.05MI_{Y|X}<0.05$ and $MI_{X|Y}>0.95MI_{X|Y}>0.95$. It constructs short position in Y and long positions in X on the days that $MI_{Y|X}>0.95MI_{Y|X}>0.95$ and $MI_{X|Y}<0.05MI_{X|Y}<0.05$.