Pairs Trading: Optimizing via Mixed Copula versus Distance Method for S&P 500 Assets

Fernando A. B. Sabino da Silva[1], Flavio A. Ziegelmann[1,2], Joao F. Caldeira[2]

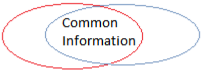
[1]Department of Statistics/UFRGS, [2]PPGE/UFRGS

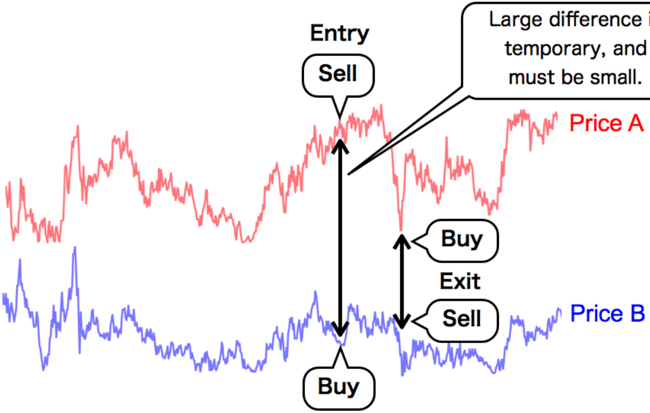
May 17, 2018

History of the Arbitrage Strategy

- The strategy was pioneered by Gerry Bamberger and later led by renowned trader Nunzio Tartaglia's quantitative group at Morgan Stanley in the early 1980s.
  - Morgan Stanley's black box strategy earned the firm a lot of money, and of course, bolster its reputation on Wall Street;
  - One of the members of that team was a computer scientist from Stanford named David Shaw, who later founded one of the most successfull stat arb hedge funds to this day (D. E. Shaw & Co).
- Pairs Trading is a contrarian strategy designed to generate abnormal profits from the mean-reverting behavior between a pair of stocks.
  - It is well-planned assault on the Efficient Market Hypothesis.
- The strategy became the foundation of statistical arbitrage and consequently, algo trading.

Two-Dimensional Pairs Trading





What is Pairs Trading?

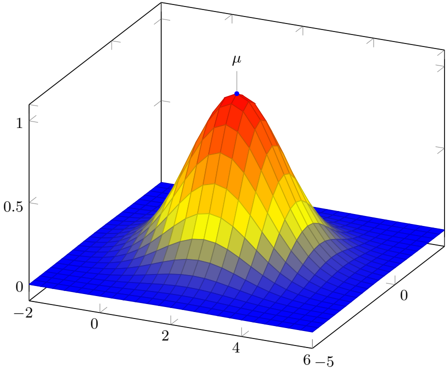
- Pairs Trading is statistical arbitrage that involves the simultaneous long and short positions of two relatively mispriced stocks which have strong historical co-movements - in most cases to create a market-neutral strategy.
- The strategy is self-financing. This means that the investors do not need to invest their own money.
- The idea is to exploit the mean reverting behaviour of co-integrated pairs.
  - If there is a spread between the prices and the further it deviates from its long-term mean value, the greater the probability of a reversal.
- Note, however, that a pairs trading strategy is by no means risk free. The strategy could perform poorly on stock pairs that pick up a trend rather than mean-reverting.

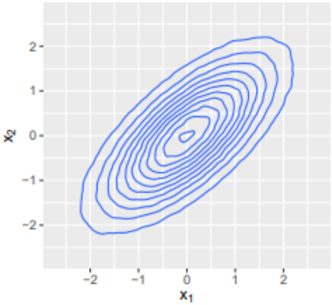
- The most popular strategy is known as distance method (see Gatev *et al.*, 2006).

- The distance strategy (Gatev *et al.*, 2006) uses the distance between normalized prices to capture the degree of mispricing stocks.
  - According to Xie *et al.* (2014) the distance method has a multivariate normal nature since it assumes a symmetric distribution of the spread between normalized prices of the stocks within a pair and it uses a single distance measure, which can be seen as an alternative measurement of the linear association, to describe the relationship between two stocks.

Bivariate Normal Distribution

$$f(x,y) = \frac{\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 - 2\rho\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right) + \left(\frac{y-\mu_y}{\sigma_y}\right)^2\right]\right\}}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}$$





Motivation

- Linear correlation (ρ) fully describes the dependence between securities if the series have joint normal distribution.

- However, a main feature of joint distributions characterized by tail dependence is the presence of heavy and possibly asymmetric tails.
  - Therefore, a single distance measure may fail to catch the dynamics of the spread between a pair of securities, and thus initiate and close the trades at non-optimal positions.

- Lie and Wu (2013) proposed a pairs trading strategy based on 2-dimensional copulas to overcome the limitations of the distance method.

Theorem 1

Let $X_1, ..., X_d$ be random variables with distribution functions $F_1, ..., F_d$, respectively. Then, there exists an d-copula C such that,

$$F(x_1, ..., x_d) = C(F_1(x_1), ..., F_d(x_d)), \tag{1}$$

for all $\mathbf{x} = (x_1, ..., x_d) \in \mathbb{R}^d$. If $F_1, ..., F_d$ are all continuous, then the function C is unique; otherwise C is determined only on $\operatorname{Im} F_1 \times ... \times \operatorname{Im} F_d$.

- Conversely, if C is an n-copula and $F_1, ..., F_d$ are distribution functions, then the function F defined above is an d-dimensional distribution function with marginals $F_1, ..., F_d$.

Why should we care about copulas?

- Assuming that $F(\cdot)$ and $C(\cdot)$ are differentiable, by (1) we have

$$
\begin{aligned}
\frac{\partial^d F(x_1, ..., x_d)}{\partial x_1 ... \partial x_d} &\equiv f(x_1, ..., x_d) = \frac{\partial^d C(F_1(x_1), ..., F_d(x_d))}{\partial x_1 ... \partial x_d} && (2) \\
&= c(u_1, ..., u_d) \prod_{i=1}^{d} f_i(x_i), && (3)
\end{aligned}
$$

where $u_i = F_i(x_i)$, $i = 1, ..., d$.

Copula

- Sklar's theorem states that any multivariate joint distribution can be written in terms of their univariate marginal distribution function and the dependence structure (represented in C) between the variables.

- Copulas accommodate various forms of dependence through suitable choice of the copula "correlation matrix" since they conveniently se parate marginals from dependence component.

Strategy	Associations Captured	Required Marginal Distributions
Distance	Linear	Gaussian
Copula	Linear and Nonlinear	No assumption

- From a modelling perspective, Sklar's Theorem allow us to estimate the multivariate distribution in two parts: (i) finding the marginal distributions; (ii) finding the dependency between the filtered data from (i).

Data

- Daily data of adjusted closing prices of all shares that belongs to the S&P500 market index from July 2nd, 1990 to December 31st, 2015.
  - We obtain the adjusted closing prices from Bloomberg and the returns on the Fama and French factors from French's website.
- The data set sample period is made up of 6,426 days and includes a total of 1100 stocks over all periods.
  - With the objective of eliminating survivorship bias from our database, only stocks that are listed during the formation period are included in the analysis, *i.e.*, around 500 stocks in each trading period.

Distance

- Pairs are sorted based on the sum of squared between their normalized prices during the next 12 months (formation period) adjusting them by dividends, stock splits and other corporate actions.
  - Specifically, the pairs are formed using data from January to December or from July to June.

- Prices are scaled to \$1 at the beginning of each formation period and then evolve using the return series. We then select the top 5, 10,. . . ,35 of those combinations that have the smallest sum of squared spreads, allowing re-selection of a specific pair, during the formation period. These pairs are then traded in the next six months (trading period).

- In Gatev *et al.* (2006) , when the spread diverges by two or more standard deviations (which is calculated in the formation period) from the mean, the stocks are assumed to be mispriced in terms of their relative value to each other and thus we open a short position in the outperforming stock and a long in the underperforming one.

- The price divergence is expected to be temporary, i.e., the prices are expected to converge to its long-term mean value of 0 (mean-reverting behavior).
- Trades that do not converge can result in a loss if they are still open at the end of the trading period when they are automatically closed.
  - This results in fat left tails.
- Since the conditional variance is empirically higher for large negative returns and smaller for positive returns, it may be inappropriate to use constant trigger points because the volatility differs at different price levels.

- To calculate the daily percentage excess returns for a pair, we compute

$$r_{pt} = w_{1t} r_t^L - w_{2t} r_t^S , \tag{4}$$

where L and S stands for long and short, respectively.

- The weights w_{1t} and w_{2t} are initially assumed to be one. After that, they change according to the changes in the value of the stocks, *i.e.*, $w_{it} = w_{it-1}(1 + r_{it-1})$.

- Following Gatev *et al.* (2006), we calculate returns using two weighting schemes: the return on committed capital and on fully invested capital.

- By using the fact that the partial derivative of the copula function gives the conditional distribution function, i.e.,

$$
\begin{aligned}
P\left(U_1 \leq u_1 \left| U_2 = u_2\right.\right) &= \frac{\partial C\left(u_1, u_2\right)}{\partial u_2} = P\left(X_1 \leq x_1 \left| X_2 = x_2\right.\right), \\
P\left(U_2 \leq u_2 \left| U_2 = u_1\right.\right) &= \frac{\partial C\left(u_1, u_2\right)}{\partial u_1} = P\left(X_2 \leq x_2 \left| X_1 = x_1\right.\right),
\end{aligned}
$$

Xie *et al*. (2014) define a measure to denote the degree of mispricing.

Definition 2

- Let R_t^X and R_t^Y represent the random variables of the daily returns of stocks X and Y on time t, and the realizations of those returns on time t are r_t^X and r_t^Y, we have

$$
\begin{aligned}
MI_{X|Y}^t &= P(R_t^X < r_t^X \mid R_t^Y = r_t^Y) \\
&\text{and} \\
MI_{Y|X}^t &= P(R_t^Y < r_t^Y \mid R_t^X = r_t^X),
\end{aligned}
$$

- Given current realizations r_t^X and r_t^Y , if F_X and F_Y are the marginal distribution functions of R_t^X and R_t^Y and C is the copula connecting F_X and F_Y , we define $u_1 = F_X\left(r_t^X\right)$ and $u_2 = F_Y\left(r_t^Y\right)$, and have

$$\begin{aligned} MI^t_{X|Y} = &\quad \frac{\partial C(u_1, u_2)}{\partial u_2} = \quad P(R_t^X < r_t^X \mid R_t^Y = r_t^Y) \\ \text{and} & \\ MI^t_{Y|X} = &\quad \frac{\partial C(u_1, u_2)}{\partial u_1} = \quad P(R_t^X < r_t^X \mid R_t^Y = r_t^Y). \end{aligned} \tag{5}$$

- A conditional value of 0.5 means that the two underlying stocks are considered fairly-valued.

- The conditional probabilities, $M_t^{X|Y}$ and $M_t^{Y|X}$, only measure the degrees of relative mispricing for a single day.
  - To determine an overall degree of relative mispricing we follow Rad *et al*. (2016).

- Let $m_{1,t}$ and $m_{2,t}$ be the overall mispricing indexes of stocks X_1 and X_2, defined by $\left(M_t^{X|Y} - 0.5\right)$ and $\left(M_t^{Y|X} - 0.5\right)$, respectively. At beggining of each trading period two cumulative mispriced indexed M_1 and M_2 are set to zero and then evolve for each day through

$$
\begin{aligned}
M_{1,t} &= M_{1,t-1} + m_{1,t} \\
M_{2,t} &= M_{2,t-1} + m_{2,t}
\end{aligned}
$$

- Positive (negative) M_1 and negative (positive) M_2 are interpreted as stock 1 (stock 2) being overvalued relative to stock 2 (stock 1).

- We perform a sensitivity analysis to open a long-short position once one of the cumulative indexes is above 0.05, 0.10,. . . , 0.55 and the other one is below -0.05, -0.10,. . . ,-0.55 at the same time for Top 5, 10,. . . ,35 pairs.

- The positions are unwound when both cumulative mispriced indexes return to zero. The pairs are then monitored for other possible trades throughout the remainder of the trading period.

Methodology: Copula

1. Calculate daily returns for each stock during the formation period and estimate the marginal distributions of returns separately.

   - We fit an appropriate ARMA(p,q)-GARCH(1,1) model to each univariate time series.

2. Estimate the two-dimensional copula model to data that has been transformed to [0,1] margins, i.e.,

$$H\left(r_t^X, r_t^Y\right) = C\left(F_X\left(r_t^X\right), F_Y\left(r_t^Y\right)\right),$$

where H is the joint distribution, r_t^X e r_t^Y are stock returns and C is the copula.

   - Copulas that are tested in this step are Gaussian, t, Clayton, Frank, Gumbel, one Archimedean mixture copula consisting of the optimal linear combination of Clayton, Frank and Gumbel copulas and one mixture copula consisting of the optimal linear combination of Clayton, t and Gumbel copulas.

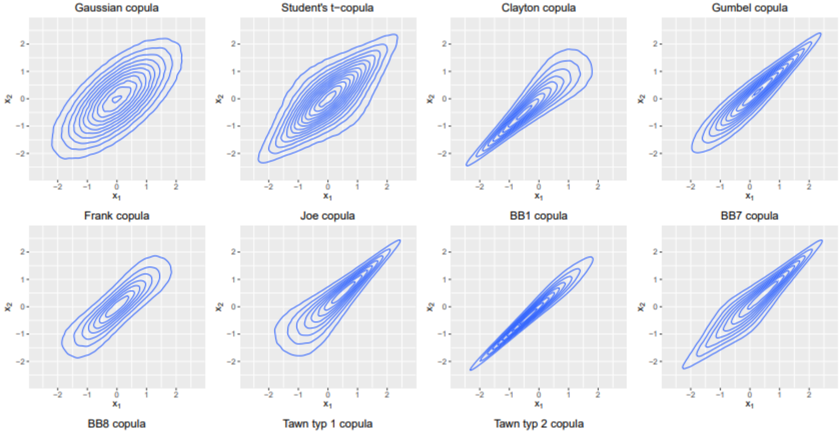
$$\mathcal{C}_{\theta}^{CFG}\left(u_1,u_2\right)=\pi_1\mathcal{C}_{\alpha}^{C}\left(u_1,u_2\right)+\pi_2\mathcal{C}_{\beta}^{F}\left(u_1,u_2\right)+\left(1-\pi_1-\pi_2\right)\mathcal{C}_{\delta}^{G}\left(u_1,u_2\right),$$

and

$$\mathcal{C}_{\xi}^{CtG}\left(u_1,u_2\right)=\pi_1\mathcal{C}_{\alpha}^{C}\left(u_1,u_2\right)+\pi_2\mathcal{C}_{\Sigma,\nu}^{t}\left(u_1,u_2\right)+\left(1-\pi_1-\pi_2\right)\mathcal{C}_{\delta}^{G}\left(u_1,u_2\right),$$

where $\theta=(\alpha,\beta,\delta)'$ are the Clayton, Frank and Gumbel copula (dependence) parameters and $\xi=(\alpha,(\Sigma,\nu),\delta)'$ are the Clayton, t and Gumbel copula parameters, respectively, and π_1, $\pi_2\in[0,1]$.

Tail Dependence



Methodology: Mixed Copula

(3 Take the first derivative of the copula function to compute conditional probabilities and measure mispricing degrees $MI_{X|Y}$ and $MI_{Y|X}$ for each day in the trading period using the copula and estimated parameters.

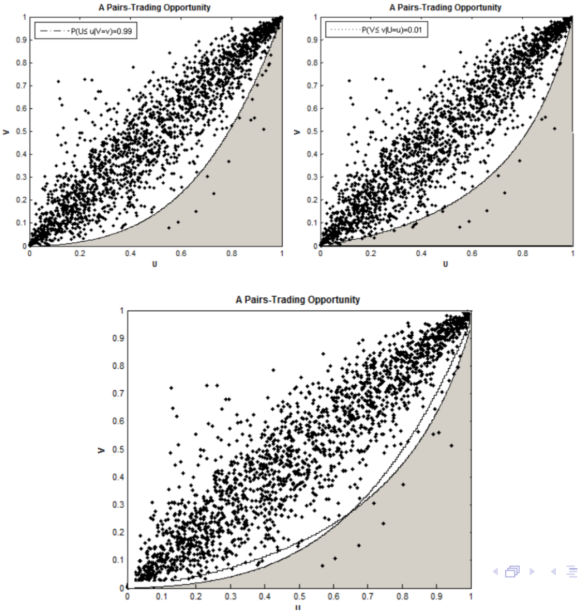
(4 Build long and short positions of Y and X on the days that $M_{1,t} > \Delta_1$ and $M_{2,t} < \Delta_2$ if there is no positions in X or Y.

  - Conversely, build positions long/short of X and Y on the day that $M_{1,t} < \Delta_2$ and $M_{2,t} > \Delta_1$ if there is no positions in X or Y.

- All positions are closed if $M_{1,t}$ reaches Δ_3 or $M_{2,t}$ reaches Δ_4, where $\Delta_1, \Delta_2, \Delta_3$ and Δ_4 are predetermined thresholds or are automatically closed out on the last day of the trading period if they do not reach the thresholds.

- Here we use $\Delta_1 = 0.2, \Delta_2 = -0.2$ and $\Delta_3 = \Delta_4 = 0$.

Illustration



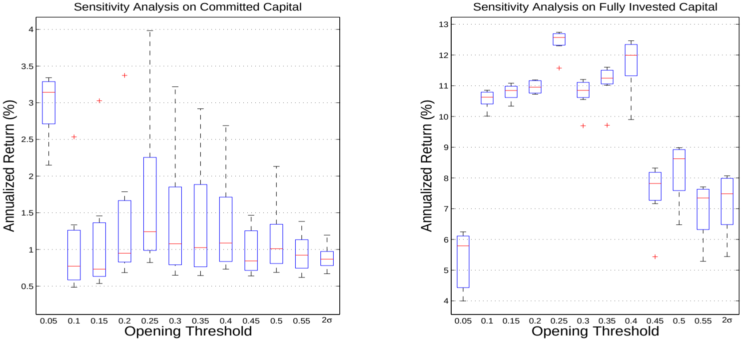


Figure 1: **Annualized returns of pairs trading strategies after costs on committed and fully invested capital**

These boxplots show annualized returns on committed (left) and fully invested (right) capital after transaction cost to different opening thresholds from July 1991 to December 2015 for Top 5 to Top 35 pairs.

Risk-Return characteristics

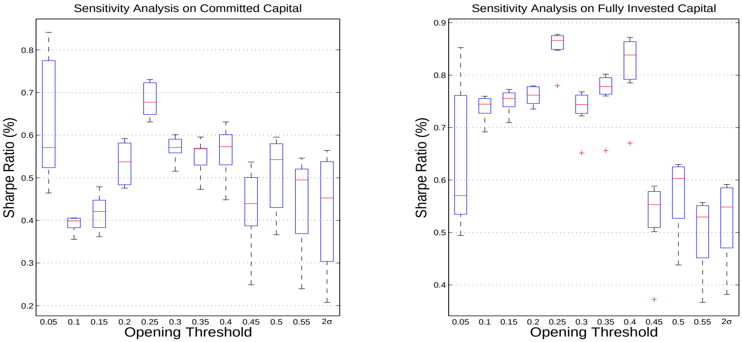


Figure 2: **Sharpe ratio of pairs trading strategies after costs on committed and fully invested capital**

Table 1: Excess returns on committed capital of pairs trading strategies on portfolios of Top 5, 20 and 35 pairs after costs.

Strategy	Mean Return (%)	Sharpe ratio	Sortino ratio	t-stat	% of negative trades	MDD1	MDD2
Return on Committed Capital							
Panel A - Top 5 pairs							
Distance	2.60	0.31	0.58	1.86*	46.98	6.73	19.62
Mixed Copula	3.98	0.63	1.08	3.49***	41.79	4.36	9.29
Panel B - Top 20 pairs							
Distance	3.14	0.65	1.13	3.32***	48.02	3.88	9.69
Mixed Copula	1.24	0.64	1.04	3.52***	41.33	2.07	3.43
Panel C - Top 35 pairs							
Distance	3.12	0.77	1.36	3.92***	47.97	2.70	7.52
Mixed Copula	0.82	0.73	1.19	3.95***	41.31	1.18	1.98

Note: Summary statistics of the annualized excess returns, annualized Sharpe and Sortino ratios on portfolios of top 5, 20 and 35 pairs between July 1991 and December 2015 (6,173 observations). The t-statistics are computed using Newey-West standard errors with a six-lag correction. The columns labeled MDD1 and MDD2 compute the largest drawdown in terms of maximum percentage drop between two consecutive days and between two days within a period of maximum six months, respectively.

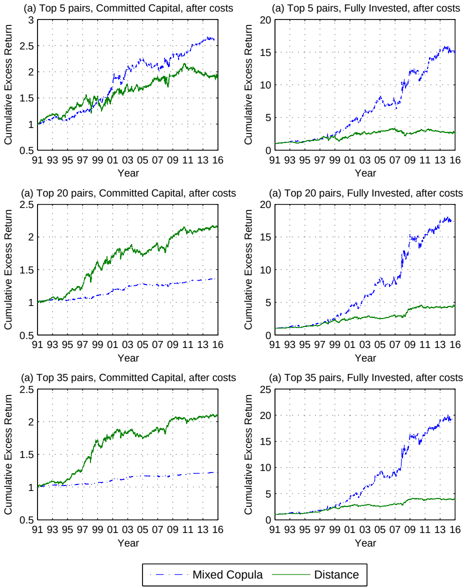
***, **, * significant at 1%, 5% and 10% levels, respectively.

Table 2: Excess returns on fully invested capital of pairs trading strategies on portfolios of Top 5, 20 and 35 pairs after costs.

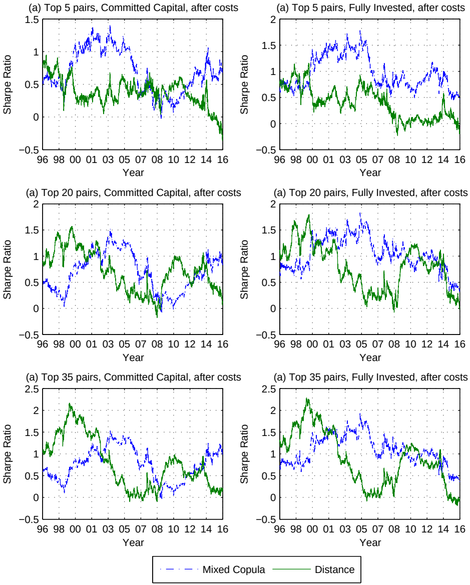
Strategy	Mean Return (%)	Sharpe ratio	Sortino ratio	t-stat	% of negative trades	MDD1	MDD2
			Return on Fully Invested Capital				
			Panel A - Top 5 pairs				
Distance	4.01	0.28	0.57	1.81*	46.98	8.70	38.36
Mixed Copula	11.58	0.78	1.43	4.26***	41.79	9.00	25.68
			Panel B - Top 20 pairs				
Distance	6.07	0.66	1.19	3.55***	48.06	5.43	20.03
Mixed Copula	12.30	0.85	1.54	4.60***	41.31	9.00	25.68
			Panel C - Top 35 pairs				
Distance	5.76	0.76	1.38	4.05***	47.97	4.24	15.07
Mixed Copula	12.73	0.88	1.59	4.73***	41.28	9.00	25.68

***, **, * significant at 1%, 5% and 10% levels, respectively.

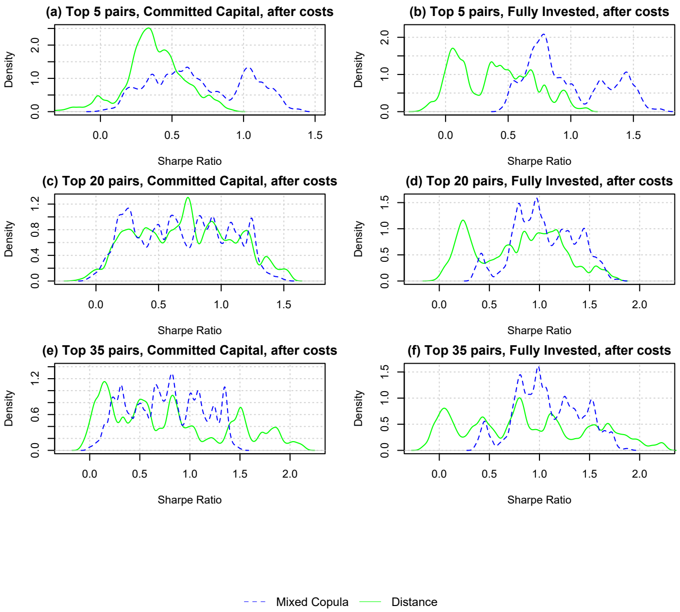
Cumulative excess returns of pairs trading strategies after costs



Five-year rolling window Sharpe ratio after costs



Kernel density estimation of 5-year rolling window Sharpe ratio after costs



Trading Statistics

Table 3: Trading statistics.

Strategy	Distance	Mixed Copula
	Panel A: Top 5	
Average price deviation trigger for opening pairs	0.0594	0.0665
Total number of pairs opened	352	348
Average number of pairs traded per six-month period	7.18	7.10
Average number of round-trip trades per pair	1.44	1.42
Standard Deviation	1.0128	1.33
Average time pairs are open in days	50.70	37.70
Standard Deviation	39.24	38.93
Median time pairs are open in days	38.5	19
	Panel B: Top20	
Average price deviation trigger for opening pairs	0.0681	0.0821
Total number of pairs opened	1312	749
Average number of pairs traded per six-month period	26.78	15.29
Average number of round-trip trades per pair	1.34	0.76
Standard Deviation	0.99	0.99
Average time pairs are open in days	51.65	23.60
Standard Deviation	39.62	32.90
Median time pairs are open in days	41	9

Table 4: Trading statistics.

Strategy	Distance	Mixed Copula
	Panel C: Top 35	
Average price deviation trigger for opening pairs	0.0729	0.0893
Total number of pairs opened	2238	941
Average number of pairs traded per six-month period	45.68	19.20
Average number of round-trip trades per pair	1.30	0.55
Standard Deviation	1.02	0.84
Average time pairs are open in days	52.72	19.35
Standard Deviation	40.48	30.56
Median time pairs are open in days	42	6

Note: Trading statistics for portfolio of top 5, 20 and 35 pairs between July 1991 and December 2015 (49 periods). Pairs are formed over a 12-month period according to a minimum-distance (sum of squared deviations) criterion and then traded over the subsequent 6-month period. Average price deviation trigger for opening a pair is calculated as the price difference divided by the average of the prices.

Table 5: Monthly risk profile of Top 5 pairs: Fama and French (2016)'s five factors plus Momentum and Long-Term Reversal.

Strategy	Intercept	Rm-Rf	SMB	HML	RMW	CMA	Mom	LRev	R^2	R^2_{adj}
Section 1: Return on Committed Capital										
Distance	0.0025	0.0091	-0.0032	0.0113	0.0003	-0.0029	-0.0107	-0.0084	0.028	0.027
	(1.89)*	(4.22)***	(-0.71)	(2.05)**	(0.25)	(-0.18)	(−4.80)***	(−1.96)**		
Mixed Copula	0.0035	0.0052	-0.0043	0.0039	-0.0035	0.0027	-0.0054	-0.0057	0.015	0.014
	(3.55)***	(3.68)***	(−1.83)*	(1.20)	(-0.99)	(0.63)	(−2.99)***	(-1.57)		
Section 2: Return on Fully Invested Capital										
Distance	0.0040	0.0170	-0.0031	0.0185	0.0049	-0.0018	-0.0161	-0.0150	0.025	0.024
	(1.75)*	(4.88)***	(-0.45)	(2.22)**	(0.76)	(0.05)	(−4.30)***	(−1.97)**		
Mixed Copula	0.0098	0.0148	-0.0084	0.0152	-0.0053	0.0087	-0.0082	-0.0222	0.018	0.017
	(4.17)***	(3.51)***	-1.45	1.6355	-0.60	0.75	(−2.19)**	(−2.08)**		

***, **, * significant at 1%, 5% and 10% levels, respectively.

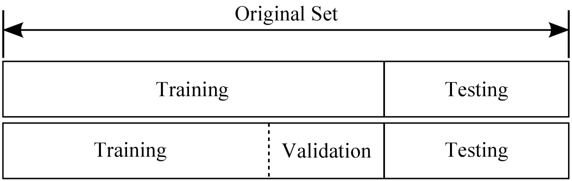
- We find that the alphas of the regressions are significantly positive and higher than the raw excess returns by about 2-7 bps per month, indicating that only a small part of the excess returns can be attributed to their exposures to the seven risk determinants.

Future Research

- Varying Threshold
- Create similar copula-based arbitrage for triplets to increase information dependency information and measure relative pricing more comprehensively.

- Vine Copulas (Pair-Copula Constructions)
  - Superior flexibility

Future Research



- Machine Learning
  - Explore non-linear approaches
- News Sentiment Factor
  - Can provide increased risk-adjusted returns

Conclusions

1. By capturing linear/nonlinear associations and covering a wider range of possible dependencies structures, the mixed copula strategy outperforms the distance method when the number of trading signals is equiparable, especially after the subprime mortgage crisis.

2. The alphas of all strategies remain large and significant even after several asset pricing factors such as market portfolio (beta), size (SMB), value (HML), investment(CMA), profitability (RMW), momentum (Mom) and reversal (LRev) are taken into account.

Thanks! Any questions?