



GESTION QUANTITATIVE DES RISQUES

ÉVALUATION D'UNE STRATÉGIE PAIRS TRADING CONSTRUITE À
PARTIR DE LA COPULE

Flora GUILLET

César ROALDES

I. Introduction

Le *Pairs trading* est une stratégie d'investissement financier développée dans les années 1980 par Gerry Bamberger au sein de la banque d'investissement américaine Morgan Stanley. Il s'agit d'une méthode d'arbitrage statistique permettant de tirer profit des déviations temporaires de prix entre deux actifs interdépendants. Le but de cette stratégie est de sélectionner des paires d'actifs dont les cours évoluent de manière similaire et d'identifier l'instant où ceux-ci divergent. En achetant l'action sous-évaluée et en vendant celle sur-évaluée, le spéculateur fait le pari que les cours convergeront à nouveau. Les rendements de cette stratégie reposent sur deux facteurs, d'une part la bonne identification des paires, et d'autre part la modélisation optimale de la relation entre les deux actifs les constituant. Notre devoir a pour sujet l'optimisation de ce second facteur par la copule.

La stratégie de Pairs Trading est utilisée de manière intensive par les fonds d'investissement spéculatifs. Son application a longtemps reposé sur la méthode des distances, privilégiée pour sa simplicité d'implémentation. Cependant, comme nous le verrons, cette méthode repose sur l'hypothèse restrictive de rendements normaux multivariés, une hypothèse de la théorie moderne du portefeuille qui est rarement vérifiée en pratique.

En 2010, Do et Faff ont testé la performance de la stratégie des distances de 1962 à 2009 sur le marché américain. Le résultat principal tiré de cette étude est l'observation d'une chute de sa rentabilité. Le rendement mensuel moyen de 0.86% de 1962 à 1988 chute à 0.24% en fin de période. Une des explications avancée par les auteurs serait que l'usage très répandu de la stratégie renforcerait l'efficacité du marché financier. Les prix s'ajustant plus rapidement, il en devient plus complexe de capter les opportunités de trade.

Nous présentons dans ce devoir une approche alternative modélisant la distribution jointe des rendements journaliers de nos actifs grâce à l'utilisation de la copule. La méthode des copules nous permettra alors de relâcher la contrainte de normalité et de capter de nouvelles opportunités d'investissement.

Pour réaliser notre stratégie d'investissement Pairs Trading par la méthode des distances, nous avons suivi la procédure décrite par Gatev et al. (2006). Ce papier fait figure de référence académique sur le sujet et présente une analyse poussée des rendements dégagés par cette stratégie. Nous suivons la méthode de Xie et al. (2014) pour modéliser la distribution jointe des rendements par la copule. Pionniers de cette approche dans le domaine académique, ils ont inspiré un grand nombre d'articles sur le sujet ces dernières années¹. Xie et al. (2014) reprennent en grande partie la méthodologie décrite par Gatev dans le but de réaliser une analyse comparative des deux approches et mettre en évidence les atouts apportés par la copule dans la stratégie Pairs Trading.

1. Rad H., Yew L.R., Faff R. (2016); Krauss C., Stübinger J. (2015); Salvatierra, I.D.L., Patton, A. J. (2015).

Notre devoir présentera en premier lieu une version de Pairs Trading basée sur la méthode des distances, puis détaillera la méthodologie d'intégration de la copule. Nous terminerons ce projet en appliquant ces deux stratégies sur une large base d'actifs extraite du NYSE afin d'analyser les rendements dégagés par chacune et les comparer.

II. Construction et Critique de la Méthode des Distances

De manière générale, la stratégie Pairs Trading surveille les mouvements des cours de deux actifs historiquement corrélés. Lorsque la corrélation s'affaiblit, par exemple lorsque l'un des cours monte tandis que l'autre diminue, les positions s'ouvrent. L'actif sur-performant est vendu à découvert et finance l'achat de l'autre actif considéré comme sous-performant. Le résultat de cette opération repose sur la convergence des cours des deux actifs qui entraîne la fermeture des positions.

Différentes mises en place de la stratégie Pairs Trading sont possibles. Chacune d'entre elles implique une période de formation durant laquelle les paires d'actifs sont formées, et une période de trading durant laquelle sont prises les positions long - short. Nous reprendrons ici le cycle de 18 mois utilisé par Gatev et al. (2006) définissant la période de formation à 12 mois pour entamer une période de trading de 6 mois. Nous analyserons dans un premier temps la méthode la plus répandue, celle des distances, avant de présenter une version du Pairs Trading exploitant les avantages procurés par la copule.

Pour réaliser le Pairs Trading, la méthode des distances est la plus souvent utilisée pour modéliser la relation qu'ont les deux actifs entre eux. Aucune hypothèse n'est explicitement faite concernant la distribution de leurs rendements, et cette méthode propose de prendre comme indicateur la distance entre les prix normalisés des deux actifs pour déterminer l'erreur d'évaluation des prix.

Nous reprenons ici méthode utilisée par Gatev et al. (2006) pour implémenter cette stratégie. Pour chaque actif de notre univers, nous trouvons l'actif associé qui minimise la somme des écarts au carré de leurs prix normalisés sur la période de formation. Gatev et al. (2006) rapportent que ce critère de distance minimale est une bonne approximation de la manière dont les traders eux-même choisissent leurs paires afin de former des couples d'actifs ayant des prix qui *bougent ensemble*.

Soit deux d'actifs X et Y dont les prix de clôture au temps t sont respectivement P_t^X et P_t^Y . Nous commençons par normaliser leur prix² durant la période de trading. Les cours normalisés de nos deux actifs sont notés NP_t^X et NP_t^Y et valent 1 en début de période. Le spread, noté Δ_t , correspond à la différence entre les prix normalisés des deux actifs.

$$\Delta_t = NP_t^X - NP_t^Y$$

Mathématiquement, l'actif Y_i formera une paire avec l'actif X s'il vérifie, parmi les $K - 1$ actifs possibles présents dans notre univers, la condition suivante,

$$\underset{Y_j \in \{Y_1, \dots, Y_{K-1}\}}{\operatorname{argmin}} \sum_{t=1}^T (NP_t^X - NP_t^{Y_j})^2 = Y_i.$$

2. La méthode de normalisation utilisée est la suivante : $NP_t^X = \frac{P_t^X}{P_0^X}$

Xie et al. notent que l'hypothèse fondamentale faite par la stratégie Pairs Trading est que les deux actifs sélectionnés partagent la même exposition aux risques. En considérant leurs prix normalisés, les deux actifs devraient générer des flux de trésorerie identiques. Par extension de la loi du prix unique, ils devraient donc se négocier au même prix. En conséquence, l'espérance du spread est nulle.

$$\mathbb{E}(\Delta_t) = \mathbb{E}(NP_t^X) - \mathbb{E}(NP_t^Y) = 0.$$

Durant la période de formation, nous calculons également l'écart-type empirique du spread σ_{Δ_t} , cette mesure nous permettant ensuite de déclencher la stratégie sur la période de trading. L'ouverture des positions s'effectue lorsque les prix des deux actifs, re-normalisés sur la période de trading, s'écartent de plus de deux écarts-types. La prise de position est fondée sur le principe de la *Mean Reversion*. Autrement dit, la stratégie mise sur la convergence des prix des deux actifs historiquement très corrélés. L'actif au prix le plus bas est acheté tandis que l'actif au prix le plus haut, considéré comme sur-performant, est vendu à découvert.

$$\begin{aligned} X \text{ long et } Y \text{ short si : } & \begin{cases} |NP_t^X - NP_t^Y| > 2 \cdot \sigma_{\Delta_t} \\ NP_t^X - NP_t^Y < 0 \end{cases} \\ Y \text{ long et } X \text{ short si : } & \begin{cases} |NP_t^X - NP_t^Y| > 2 \cdot \sigma_{\Delta_t} \\ NP_t^X - NP_t^Y > 0 \end{cases} \end{aligned}$$

D'un point de vue théorique, le critère de la distance peut être suffisant si l'on se place sous l'hypothèse classique d'une distribution normale multivariée des rendements. En effet, sous cette hypothèse, la mesure de corrélation de Pearson décrit parfaitement la dépendance entre les deux variables. La distance, à l'image de la corrélation de Pearson, capture la dépendance linéaire des deux actifs mais n'est pas capable de décrire les dépendances non-linéaires existantes.

Statistiquement, sous l'hypothèse que NP_t^X et NP_t^Y suivent une distribution normale multivariée, Δ_t suit une loi normale centrée, d'écart-type σ_{Δ_t} . Ainsi, le seuil de $2 \cdot \sigma_{\Delta_t}$ à partir duquel s'enclenche la stratégie correspond à la sortie de l'intervalle de confiance à 95% de Δ_t . Seulement, cette hypothèse, nécessaire au bon fonctionnement de la méthode des distances, n'est pas toujours vérifiée. L'hypothèse de normalité des rendements s'est vue massivement rejetée par le domaine académique depuis l'identification d'asymétrie et de leptokurticité dans la distribution des rendements³. Les distributions marginales des rendements d'actifs sont rarement considérées comme normales et la structure de dépendance les liant peut présenter des associations non-linéaires qui sont négligées par une simple mesure de distances. Par exemple, la dépendance en queue de distribution.

En 2014, Xie et al. affirment dans leur article "*Pairs Trading with Copulas*" que la méthode des distances repose implicitement sur l'hypothèse de rendements normaux multivariés. En effet cette méthode considère que la distribution des spreads entre les prix normalisés est symétrique. D'autre part, les associations linéaires sont mesurées par une distance, simple à implémenter mais conduisant à une perte d'information si la distribution des actifs ne suit pas une loi normale bivariée.

3. Cont, R. 2001. *Empirical properties of asset returns : stylized facts and statistical issues*. Quantitative Finance

D'après eux, le rejet de cette hypothèse conduit à l'asymétrie de la variable aléatoire Δ_t autour de zéro. Cela implique que pour une même valeur, Δ_t peut représenter différents degrés d'erreur dans l'évaluation des prix des actifs, et donc que la distribution conditionnelle du spread diffère pour une même valeur de Δ_t .

$$\mathbb{P}(\Delta_t = \bar{\delta} \mid NP_t^X = x^1, NP_t^Y = y^1) \neq \mathbb{P}(\Delta_t = \bar{\delta} \mid NP_t^X = x^2, NP_t^Y = y^2)$$

La distribution de Δ_t devient de fait inconnue. Or, la méthode des distances enclenche l'ouverture des positions lorsque le spread excède un seuil constant, fixé ici à $2 \cdot \sigma_{\Delta_t}$. La méthode n'apparaît donc pas optimale.

A l'inverse, la copule représente une technique de modélisation de la distribution jointe de deux variables aléatoires considérant les associations non-linéaires, et permettra à la stratégie Pairs Trading de capter des opportunités de trade ignorées ou mal évaluées jusqu'alors. Son intégration est présentée dans la section suivante.

III. Pairs Trading par la Méthode de la Copule

La copule est un outil permettant, à l'aide du théorème de Sklar (1959), d'établir un lien entre les distributions marginales et leur distribution jointe. Son utilisation s'est intensifiée ces dernières années dans le domaine de l'assurance et de la finance pour la souplesse et la diversité des modélisations qu'elle offre. L'approche par la méthode des copules résout les difficultés rencontrées par la méthode des distances. Elle admet l'estimation individuelle des distributions marginales des rendements des actifs d'une paire, et permet d'estimer différentes structures de dépendance pouvant décrire les relations en queue de distribution. De cette manière, nous pouvons nous affranchir de l'hypothèse de rendements normaux multivariés. La modélisation de la distribution jointe permise par la copule généralise la méthode des distances.

Xie et al. (2014) utilisent la copule pour modéliser la distribution jointe des rendements engendrés par une paire d'actifs. Cette distribution jointe permet d'évaluer les probabilités conditionnelles et prendre une position adéquate sur le marché. Reprenons nos deux actifs X et Y candidats pour une stratégie de Pair Trading. Nous notons R_t^X et R_t^Y leurs rendements journaliers respectifs⁴. D'après le théorème de Sklar (1959), si nous considérons X_1, \dots, X_d des variables aléatoires dont les fonctions de répartition sont respectivement F_1, \dots, F_d , alors il existe une copule \mathcal{C} définie sur $[0, 1]^d$, tel que,

$$\forall \{x_1, \dots, x_n\} \in \mathbb{R}^d, \quad F(x_1, \dots, x_n) = \mathcal{C}(F_1(x_1), \dots, F_d(x_d)).$$

Nous considérons R_t^X et R_t^Y comme étant nos variables aléatoires et r_t^X et r_t^Y leurs réalisations au temps t . Notons F_X et F_Y les fonctions de répartition marginales respectives de R_t^X et R_t^Y et \mathcal{H} leur distribution jointe. Ainsi, par le théorème de Sklar (1959), il existe une copule \mathcal{C} telle que,

$$\mathcal{H}(r_t^X, r_t^Y) = \mathcal{C}(F_X(r_t^X), F_Y(r_t^Y)).$$

4. Les rendements journaliers sont calculés de la manière suivante : $R_t^X = \frac{P_t^X - P_{t-1}^X}{P_{t-1}^X}$

Cette estimation de la distribution jointe des rendements nous permet de définir de nouvelles mesures d'erreur d'évaluation des prix, les *Mispricing Indexes*, notés $MI_t^{X|Y}$ et $MI_t^{Y|X}$. Ces nouveaux indicateurs sont calculés à partir des probabilités conditionnelles.

$$MI_t^{X|Y} = \mathbb{P}(R_t^X \leq r_t^X \mid R_t^Y = r_t^Y)$$

$$MI_t^{Y|X} = \mathbb{P}(R_t^Y \leq r_t^Y \mid R_t^X = r_t^X)$$

Une fois que nous avons déterminé les fonctions de répartition marginales de R_t^X et R_t^Y et que nous avons choisi la classe de copule optimale \mathcal{C} , l'observation de $R_t^X = r_t^X$ nous permet d'estimer la probabilité conditionnelle que $R_t^Y \leq r_t^Y$ et inversement. Cette distribution conditionnelle est obtenue grâce à la dérivée première de la copule par rapport à la fonction de répartition de X . En notant $u = F_X(r_t^X)$ et $v = F_Y(r_t^Y)$, alors cette probabilité conditionnelle est définie par,

$$\mathbb{P}(R_t^Y \leq r_t^Y \mid R_t^X = r_t^X) = \frac{\partial \mathcal{C}(u, v)}{\partial u}.$$

Nous avons donc une mesure de l'erreur d'évaluation du prix ordonnable, prenant en compte le niveau des prix des actifs, définis au temps t par,

$$MI_t^{X|Y} = \frac{\partial \mathcal{C}(u, v)}{\partial v} \quad \text{et} \quad MI_t^{Y|X} = \frac{\partial \mathcal{C}(u, v)}{\partial u}.$$

L'interprétation des Mispricing Indexes est la suivante :

- si $MI_t^{X|Y} > 0.5$, alors l'actif X est relativement surévalué par rapport à l'actif Y
- si $MI_t^{X|Y} = 0.5$, alors l'actif X est correctement évalué par rapport à l'actif Y
- si $MI_t^{X|Y} < 0.5$, alors l'actif X est relativement sous-évalué par rapport à l'actif Y
- si $MI_t^{Y|X} > 0.5$, alors l'actif Y est relativement surévalué par rapport à l'actif X
- si $MI_t^{Y|X} = 0.5$, alors l'actif Y est correctement évalué par rapport à l'actif X
- si $MI_t^{Y|X} < 0.5$, alors l'actif Y est relativement sous-évalué par rapport à l'actif X

La probabilité conditionnelle $MI_t^{X|Y}$ indique si le rendement de l'actif X peut être considéré comme trop haut ou trop bas au temps t , étant donné le rendement de Y le même jour et la relation historique entre les rendements des deux actifs. Si $MI_t^{X|Y} = 0.5$, alors il y a un même nombre d'observations de rendements de X supérieurs ou inférieurs à r_t^X sachant que le rendement de Y vaut r_t^Y . Dans ce cas, nous pouvons faire l'hypothèse que l'actif X est relativement bien évalué par rapport à Y .

La stratégie Pairs Trading présentée par Xie et al. (2014) est réalisée de la manière suivante. La première étape consiste à estimer les distributions marginales des rendements de notre paire les plus adaptées. Nous avons choisi d'utiliser une estimation non-paramétrique de ces distributions en calculant la fonction de répartition empirique. À partir des valeurs des rendements de X observées durant notre période de formation, $\{r_1^X, \dots, r_T^X\}$, nous définissons \hat{F}_X comme étant la fraction des observations inférieures ou égales à x .

$$\hat{F}_X(x) = \frac{1}{T} \sum_{t=1}^T \mathbb{1}_{\{r_t^X \leq x\}}$$

Vient ensuite la question du choix de la structure de dépendance optimale. Il existent un large éventail de copules bivariées dont nous estimons les paramètres par *Maximum de Vraisemblance*. Une fois les paramètres ajustés, le *critère d'information d'Akaike* (AIC) et le *critère d'information Bayésien* (BIC) sont calculés, la famille de copules minimisant leurs valeurs est ainsi choisie.

Durant la période de trading, les rendements journaliers r_t^X et r_t^Y sont calculés pour ensuite évaluer le niveau des erreurs d'évaluation des prix, $MI_t^{X|Y}$ et $MI_t^{Y|X}$, à l'aide de la copule estimée. Les indicateurs d'ouverture de position sont notés $Flag^X$ et $Flag^Y$. Initialisés à zéro en début de période de trading, ces indicateurs s'incrémentent chaque jour de respectivement $(MI_t^{X|Y} - 0.5)$ et $(MI_t^{Y|X} - 0.5)$. Finalement, nous définissons D comme étant le seuil à partir duquel les positions s'ouvrent et S le point de *stop-loss*. Ainsi, lorsque notre paire fermée, les quatre scénarios suivants sont possibles :

- Quand $Flag^X$ atteint D , nous vendons X à découvert pour financer l'achat de Y
- Quand $Flag^X$ atteint $-D$, nous vendons Y à découvert pour financer l'achat de X
- Quand $Flag^Y$ atteint D , nous vendons Y à découvert pour financer l'achat de X
- Quand $Flag^Y$ atteint $-D$, nous vendons X à découvert pour financer l'achat de Y

Si la position s'ouvre en fonction de la valeur atteinte par $Flag^X$, alors la position sera fermée quand $Flag^X$ retourne à zéro ou atteint la valeur de stop-loss S ou $-S$. Lors de la fermeture des positions, $Flag^X$ et $Flag^Y$ sont réinitialisés à zéro. Toute les paires ouvertes à la fin de la période de trading sont closes, indépendamment de la valeur des deux indicateurs.

Xie et al. (2014) fixent les valeurs de D et S à respectivement 0.6 et 2, soulignant au passage l'infinité des combinaisons possibles. Nous reprendrons ici ces deux valeurs afin de suivre au mieux le protocole décrit. Néanmoins, si l'on cherche à améliorer les performances de la stratégie Pairs Trading par la méthode des copules, l'optimisation de ces deux valeurs est une étape cruciale.

Xie et al. (2014) expliquent les mécanismes statistiques sur lesquelles repose la stratégie. La fonction de répartition d'une variable aléatoire peut-être considérée comme une variable aléatoire uniforme définie sur $[0, 1]$. $MI_t^{X|Y}$ et $MI_t^{Y|X}$ sont des fonctions de distribution conditionnelles suivant également une loi uniforme sur $[0, 1]$, ainsi, $(MI_t^{X|Y} - 0.5)$ et $(MI_t^{Y|X} - 0.5)$ suivent à leur tour une fonction de distribution uniforme sur l'intervalle $[-0.5, 0.5]$. $Flag^X$ et $Flag^Y$, les cumuls journaliers de ces variables aléatoires, sont alors comparables à une marche aléatoire. Si nous définissons mathématiquement $Flag_t^\alpha = Flag_{t-1}^\alpha + \epsilon_t$, et faisons l'hypothèse que les ϵ_t sont indépendants, identiquement distribués, et suivent une distribution uniforme allant de -0.5 à 0.5, alors cela revient à suggérer que $R_{t_i}^\alpha$ est indépendant de $R_{t_j}^\alpha$, pour tout $i \neq j$. De cette manière, $Flag_t^\alpha$ représente une mesure cumulée du degré d'erreur d'évaluation du prix de l'actif α et $Flag_t^\alpha$ devient une pure marche aléatoire.

Xie et al. (2014) soulignent que même si l'espérance de $Flag_t^\alpha$ est toujours nulle, cette série temporelle est non-stationnaire et peut connaître des périodes de stricte croissance et décroissance, ou fluctuer autour d'une valeur différente de zéro. Toutefois, la stratégie Pairs Trading est basée sur la propriété de *Mean Reversion*, signifiant que $Flag_t^\alpha$ a tendance à re-converger vers zéro lorsqu'il s'en écarte, ce qui réfute l'hypothèse d'indépendance de ϵ_t vis-à-vis de ses

réalisations passées.

Trois mécanismes de la série temporelle $Flag_t^\alpha$ ont été identifiés par Xie et al. (2014) :

- *Le mécanisme de base*, la marche aléatoire, sans aucune autocorrélation entre ϵ_t et $Flag_{t-1}^\alpha$. Aucun arbitrage n'est possible dans ce cas et la stratégie Pairs trading ne réalise aucun profit, qu'elle soit réalisée à partir de la méthode des distances ou celle des copules.
- *Le mécanisme de convergence*, la mean reversion, qui suggère une autocorrélation négative entre ϵ_t et $Flag_{t-1}^\alpha$, et assure la convergence de $Flag_t^\alpha$ vers zéro lorsque celui-ci s'en écarte, permettant au Pairs Trading de générer des profits.
- *Le mécanisme de divergence*, où cette fois-ci l'existence d'une autocorrélation positive implique que $Flag_t^\alpha$ diverge de zéro. La stratégie réalise alors des pertes.

Sur le marché, $Flag_t^\alpha$ alterne entre ces trois mécanismes. Tant que le mécanisme de convergence domine celui de divergence, la stratégie Pairs Trading réalise des profits, mais la condition de *stop-loss* assure de limiter les pertes lorsque qu'un mécanisme de divergence se met en place.

IV. Analyse empirique

Dans cette section nous analyserons les rendements obtenus par la méthode des distances et par la méthode des copules. Nous commencerons par présenter la méthode de calcul des rendements dégagés par la stratégie Pairs Trading avant d'effectuer une première étude comparative des deux approches sur la paire d'actifs SIRI - KEM. S'en suivra une application plus large des deux approches sur des portefeuilles de paires sélectionnées en accord avec la méthodologie expliquée plus tôt.

Pour comparer les résultats des deux approches, il est essentiel de définir clairement la méthode de calcul des rendements. Au début de la période de trading, le poids du portefeuille vaut 1. L'achat du premier actif est financé par la vente du second. Le capital considéré d'un euro sur un portefeuille, que nous voulons équipondéré, signifie que nous vendons le premier actifs à 0.5€ et achetons le second à 0.5€. Les positions des actifs constituant la paire s'ouvrent puis se ferment en suivant les règles établies par les deux méthodes. Le résultat financier de l'opération est intégré dans le capital de la prochaine ouverture. En notant ω_i le poids lors de la $(i + 1)^{ième}$ ouverture, après n ouvertures, le rendement r_n^p de notre paire vaut :

$$r_n^p = \frac{\sum_{i=1}^n r_i^{long} - r_i^{short}}{\sum_{i=1}^n \omega_{i-1}} ,$$

Avec :

$$r_i^{long / short} = \frac{P_{Fermeture} - P_{Ouverture}}{P_{Ouverture}} ,$$

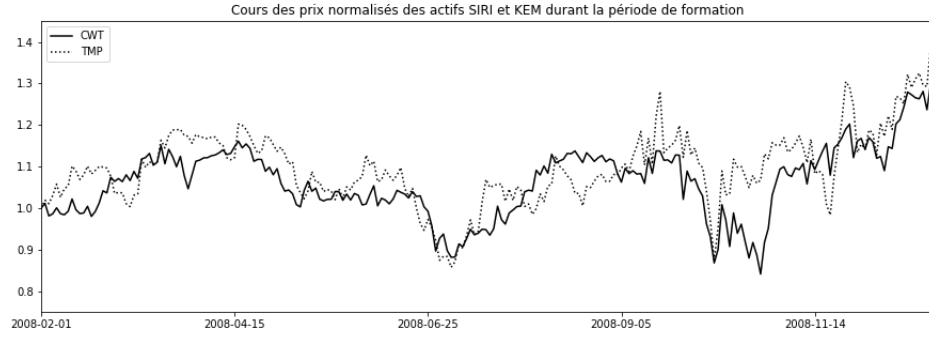
et :

$$\omega_i = \prod_{j=1}^{j=i} \omega_{j-1} \cdot (1 + r_j^p) , \quad \omega_0 = 1 .$$

Le rendement d'un portefeuille constitué de M paires correspond à la moyenne des rendements de ses paires, $r_n^{PF} = \frac{1}{M} \sum_{m=1}^M r_n^m$.

Pour illustrer nos propos, nous proposons d'appliquer les deux méthodes sur la paire SIRI - KEM. La période de formation démarre le 2 janvier 2003 et s'achève le 31 décembre de la même année. La période de trading correspond aux 6 mois suivants, soit du 1 janvier au 30 juin 2004. La Figure 1 présente le cours des prix normalisés des deux actifs durant la période de formation. Les deux actifs bougent de manière coordonnée tout au long de la période.

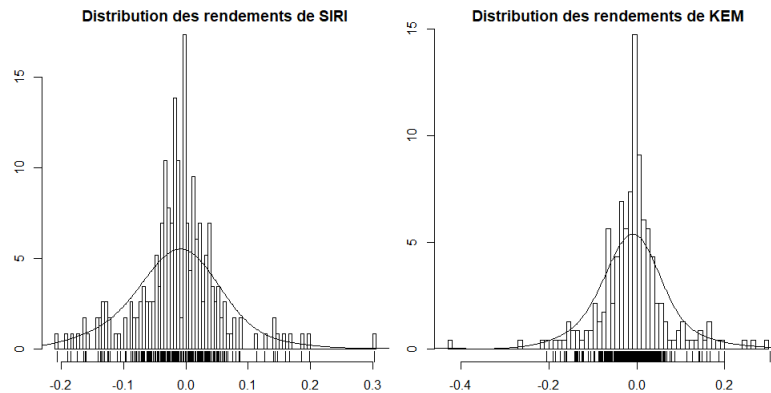
FIGURE 1 : Période de formation



La méthode des distances est mise en place en suivant le protocole expliqué plus tôt. La somme des écarts au carré entre les prix normalisés des deux actifs vaut 1.09 et l'écart-type σ_{Δ_t} de ces spreads est estimé à 0.064 sur la période de formation. Le coefficient de corrélation de Pearson des prix normalisés de la paire est de 98%. Plus les actifs d'une paire sont historiquement corrélés, plus l'écart-type est faible, le spread s'écartant peu de sa moyenne, zéro. Les positions s'ouvriront lorsque Δ_t sera supérieur à 0.128 sur la période de trading.

La mise en place de la méthode des copules est plus complexe. Nous reprenons la méthodologie décrite lors de la section précédente et commençons par estimer les distributions empiriques des rendements des deux actifs. La Figure 2 montre les distributions des rendements observés lors de la période de formation ainsi que leurs fonctions de distribution, estimées de manière non-paramétrique.

FIGURE 2 : Distributions marginales estimées



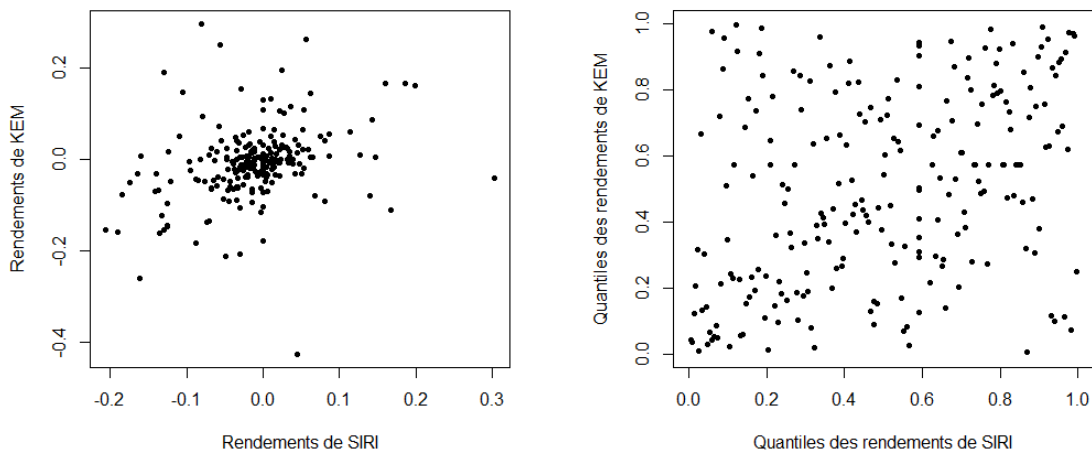
Les distributions des rendements des deux actifs présentent une asymétrie avec des queues épaisses à gauche, remettant en question l'hypothèse de normalité des rendements sur laquelle repose la méthode des distances. La probabilité de réaliser des pertes extrêmes est plus élevée que celle de faire des gains importants. Le test de Kolmogorov-Smirnov vient confirmer le rejet de l'hypothèse de normalité des rendements sur nos deux actifs, mais accepte l'hypothèse d'une distribution identique pour les rendements de SIRI et de KEM. Le résultat de ces tests est présenté dans la Table 1. Le rejet de l'hypothèse de normalité des rendements des deux actifs invalide les fondements statistiques sur lesquels repose la méthode des distances.

TABLE 1 : Test de Kolmogorov-Smirnov comparant les distributions de SIRI et KEM

data: returns_SIRI and pnorm	data: returns_KEM and pnorm
D = 0.42236, p-value < 2.2e-16	D = 0.41217, p-value < 2.2e-16
alternative hypothesis: two-sided	alternative hypothesis: two-sided
data: returns_SIRI and returns_KEM	
D = 0.073593, p-value = 0.559	
alternative hypothesis: two-sided	

La distribution jointe des rendements est présentée ci-dessous, accompagnée de la distribution jointe de leurs pseudo-observations, que nous appellerons *transformation quantile* des rendements. Cette distribution jointe présente une concentration de pertes ou de gains simultanés sur les deux actifs. La tendance se confirme avec l'observation de la distribution jointe des quantiles des rendements par l'accumulation d'observations dans le coin en bas à gauche, où se situent les pertes simultanées, et dans le coin en haut à droite pour les gains simultanés.

FIGURE 3 : Distribution jointe des rendements et de leurs transformation quantile



Suite au calcul des pseudo-observations, nous pouvons désormais définir la copule bivariée appropriée. Dans notre cas, la structure de dépendance choisie est la *copule de Student*. La copule de Student est très utilisée pour modéliser des rendements financiers multivariés⁵. Sa sélection est cohérente avec nos observations précédentes, car la copule de Student présente des masses de probabilité importantes aux extrêmes. Le paramètre de corrélation de Pearson ρ est estimé à 0.41, le nombre de degrés de liberté ν à 2.72. Le coefficient de corrélation de

5. Demarta, S. and McNeil, A. 2004. *The t Copula and Related Copulas*. International Statistical Review

Kendall τ associé à cette copule est de 0.27. La Figure 4 modélise cette copule et la Table 2 présente les résultats de l'estimation de ses paramètres.

FIGURE 4 : Structure de dépendance de la paire SIRI - KEM

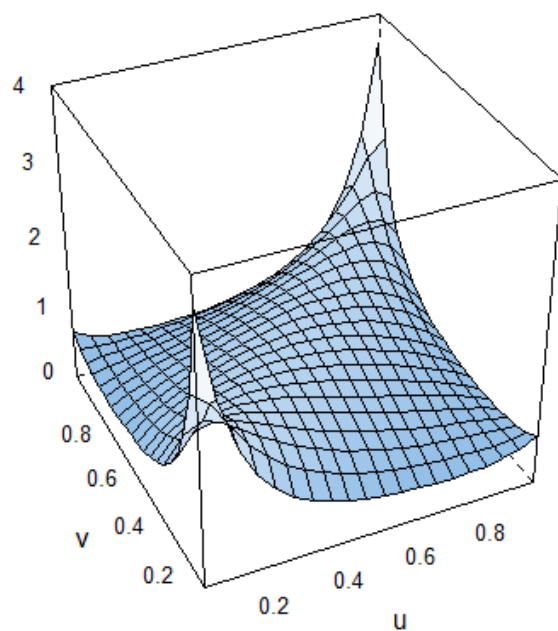


TABLE 2 : Résumé des caractéristiques de la copule sélectionnée

```

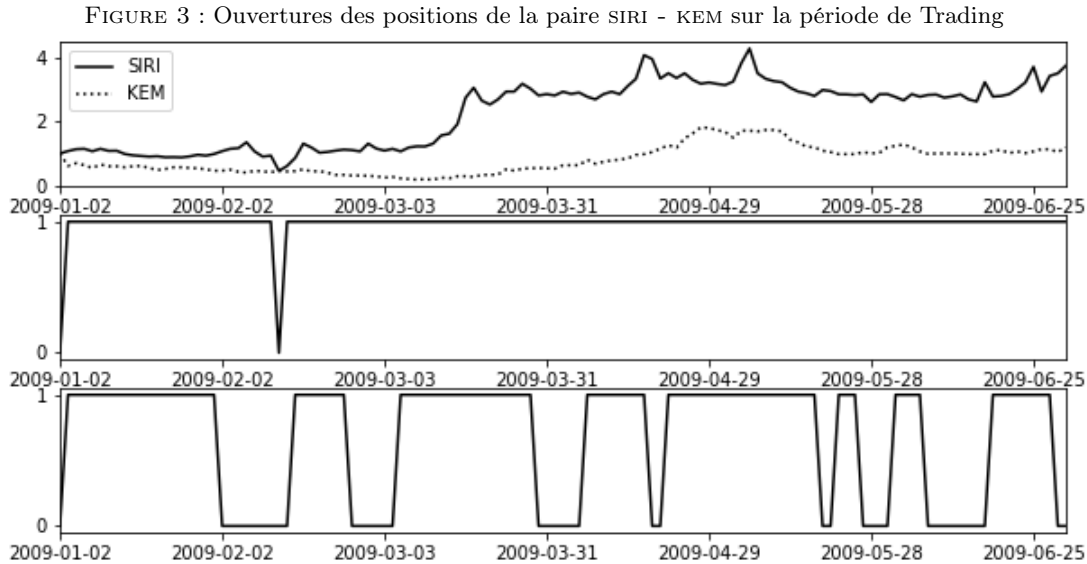
Family
-----
No:      2
Name:    t

Parameter(s)
-----
par:     0.41
par2:    2.72
Dependence measures
-----
Kendall's tau:  0.27 (empirical = 0.26, p value < 0.01)
Upper TD:      0.29
Lower TD:      0.29

Fit statistics
-----
logLik:  28.27
AIC:     -52.54
BIC:     -45.65

```

Nos deux stratégies sont maintenant prêtes à être déployées sur la période de trading. Chaque stratégie dirige les ouvertures de positions en fonction de la trajectoire prise par les prix normalisés et des conditions que nous avons établies. La Figure 5 présente, de haut en bas, les trajectoires des prix normalisés sur la période de formation et les moments d'ouverture des positions décidés par la méthode des distances et par la méthode des copules.



La méthode des distances déclenche seulement deux fois des ouvertures de positions sur la période contre 8 fois pour la méthode des copules. La fermeture de la première position par la méthode des distances intervient lorsque les deux cours se croisent pendant le mois de février, puis s'ouvre une nouvelle fois lorsque le spread franchit le seuil de deux écarts-types. En revanche, le spread ne retourne pas à sa moyenne et étant donné l'absence d'un seuil de stop-loss, la fermeture de cette position n'intervient qu'en fin de période. La méthode des copules, plus dynamique, saisit des opportunités non-captées par la méthode des distances, et limite les pertes lorsqu'un mécanisme de divergence se met en place. Les rendements des deux méthodes sont présentés dans la Table 3.

TABLE 3 : Résultats financiers des deux approches du Pairs Trading sur la paire SIRI - KEM

Méthodes	Ouvertures	Rendements
Distances	2	-173%
Copules	8	31%

Sur la paire d'actif SIRI - KEM, la méthode des distance réalise une perte élevée de -173%. Cette perte, comme nous l'avons vu, est expliquée par l'absence de stop-loss et le rejet des hypothèses implicites de la méthode. L'approche du Pairs Trading par les copules permet en revanche de réaliser des profits significatifs, affichant un rendement de 31% sur 6 mois. La méthode des copules prend ses positions à partir d'une mesure d'erreur d'évaluation des prix plus fine que celle proposée par la méthode des distances.

Ce premier exemple illustre l'intérêt de l'approche du Pairs Trading par les copules. Pour confirmer les avantages apportés par la méthode des copules, nous proposons d'appliquer nos

stratégies sur trois portefeuilles. Nos portefeuilles sont appelés TOP₅, TOP₂₀, et TOP₁₀₀. Ils représentent les 5, 20, et 100 paires d’actifs dont la somme des spreads au carré sur la période de formation sont les plus basses, en accord avec la méthode de sélection définie par Gatev et al. (2006). Les paramètres nécessaires à l’application des deux méthodes sont estimés et les stratégies appliqués sur la période de trading. Le cycle période de formation / période de trading reste le même que celui considéré dans l’exemple précédent. Les rendements réalisés par les deux méthodes entre le 1 janvier et le 30 juin 2004 sur les trois portefeuilles sont présentés dans la Table 4.

TABLE 4 : Rendements des 3 portefeuilles par les deux approches

Portefeuille	Distances	Copules
TOP ₅	-1.08%	0.20%
TOP ₂₀	-2.17%	0.64%
TOP ₁₀₀	-9.35%	0.11%

Les rendements dégagés par la méthode des copules sont relativement faibles, mais positifs sur les 3 portefeuilles, contrairement à la méthode des distances qui réalise des pertes. Il est important de souligner ici que nos portefeuilles ne sont pas indépendants. Le portefeuille TOP₂₀ est compris dans le TOP₁₀₀ et contient lui-même le TOP₅. Malgré cela, l’augmentation de la taille du portefeuille fait chuter les rendements de la méthode des distances.

V. Conclusion

Dans ce devoir, nous avons présenté la stratégie Pairs Trading. Cette stratégie est le plus souvent élaborée par le biais de la méthode des distances. Nous avons vu que cette méthode repose sur l’hypothèse de la distribution normale bivariée des rendements des actifs constituant une paire. Cette hypothèse est rarement vérifiée et conduit à une sous-performance de la stratégie.

Notre sujet a donc été d’élaborer une méthode de réalisation du Pairs Trading s’affranchissant de cette hypothèse. L’utilisation de la copule répond à ce besoin en généralisant la méthode des distances. L’estimation des fonctions de densité marginales de nos rendements et le choix de la spécification de leur structure de dépendance assouplissent les conditions d’application du Pairs Trading. La copule permet la construction d’une mesure d’erreur d’évaluation des prix robuste, par une meilleure estimation de la dépendance résidant au sein d’une paire.

L’application des deux méthodes sur trois portefeuilles a su démontrer la profitabilité supérieure dégagée par la méthode des copules. Cette méthode saisit des opportunités d’ouverture des positions négligées par la simple mesure de distances, rendant la stratégie plus dynamique. Il est important de souligner que ce dynamisme s’accompagne également d’une hausse des coûts de transaction que nous n’avons pas pris en compte dans les résultats présentés.

Plusieurs pistes d’amélioration de la stratégies sont envisageables. La principale est l’amélioration de la sélection de nos paires, en constituant par exemple des paires d’actifs d’entreprises opérant au sein d’une même industrie⁶, ou en modifiant le critère de sélection des paires par une mesure prenant en considération les dépendances linéaires et non-linéaires. La sélection des paires est une étape promordiale si nous souhaitons maximiser les gains du Pairs Trading.

6. Do et Faff (2010) ont souligné que le fait de choisir des paires issues d’un même secteur augmente la rentabilité du Pairs Trading

L'optimisation des paramètres D et S est également envisageable pour maximiser les rendements de la méthode des copules.

Il convient de vérifier que la méthode des copules est plus profitable que la méthode des distances sur plusieurs cycles, en décalant les périodes de 6 mois d'un cycle à l'autre. Cette étape n'a pas pu être réalisée en raison d'une puissance de calcul limitée des outils à notre disposition. Cependant, les travaux académiques sur lesquels nous nous sommes appuyés confirment le gain de performance réalisé par l'utilisation de la copule.

Bibliographie

- [1] Cont, R. 2001. *Empirical Properties of Asset Returns : Stylized Facts and Statistical Issues*. Quantitative Finance.
- [2] Demarta, S. and McNeil, A. 2004. *The t Copula and Related Copulas*. International Statistical Review
- [3] Do B., Faff R. 2010. *Are Simple Pairs Trading still work ?*. Financial Analysts Journal.
- [4] Do B., Faff R. 2012. *Are Pairs Trading Profits Robust to Trading Costs ?*. Journal of Financial Research.
- [5] Gatev E. G., Goetzmann W. N., Rouwenhorst K G. 2006. *Pairs Trading : Performance of a Relative-Value Arbitrage Rule*. The Review of Financial Studies.
- [6] Krauss C., Stübinger J. 2015. *Nonlinear Dependence Modeling with Bivariate Copulas : Statistical Arbitrage Pairs Trading on the S&P100*.
- [7] Rad H., Yew L.R., Faff R. 2016. *The Profitability of Pairs Trading Strategies : Distance, Cointegration and Copula Methods*. Quantitative Finance.
- [8] Salvatierra, I.D.L., Patton, A. J. 2015. *Dynamic Copula Models and High Frequency Data*. Journal of Empirical Finance.
- [9] Sklar, A. 1959. *Fonctions de Répartition à n Dimensions et Leurs Marges*. Publications de l'Institut Statistique de l'Université de Paris.
- [10] Xie W., Liew R. Q., Wu Y., Zou X. 2016a. *Pairs Trading with Copulas*. Journal of Trading.
- [11] Xie et al. 2017. *Dynamic Copula Framework for Pairs Trading*.