

INSTITUT DE SCIENCE FINANCIÈRE ET D'ASSURANCES

M2 ACTUARIAT
ANNÉES 2023/2024

ESTIMATION DE COPULES

Projet 1 : Étude de la structure de dépendance entre le CAC 40 et le FTSE 100

Léo ALLERS
Théo JALABERT
Lou SIMONEAU-FRIGGI

Table des matières

Introduction	2
I Présentation des données	3
I.1 Description quantitative des données	3
I.2 Dispersion	5
I.3 Coefficients de corrélation	6
I.4 K-plots	7
I.5 Khi-plots	7
II Analyse de la copule empirique	9
III Estimation paramétrique	13
IV Estimation non paramétrique	15
IV.1 La copule empirique	15
IV.2 Méthodes non paramétriques	15
IV.3 Qualité de l'ajustement non paramétrique	17
Conclusion	19
Bibliographie	20
Annexes	21
A Copule BB1	21

Introduction

L'entrée en vigueur de la directive européenne Solvabilité 2 en janvier 2016 a modifié en profondeur le cadre réglementaire s'appliquant aux assurances. L'un des impacts les plus significatifs concerne les fonds propres des assureurs qui doivent désormais satisfaire au Capital Requis de Solvabilité (SCR), ce qui augmente significativement les montants immobilisés. Son calcul repose sur différents facteurs de risque dont notamment celui de marché qui jouent désormais un rôle prépondérant – notamment pour les assurances vie du fait de l'évaluation en valeur de marché (Mark to Market) des actifs de la compagnie.

Dans ce contexte il est particulièrement intéressant pour les assureurs, les asset managers et les Banques de Financement et d'Investissement de pouvoir quantifier l'impact d'un actif sur le SCR, et de mettre en place des stratégies de couverture sur cet actifs.

Les stratégies de couverture consistent à acheter des options sur les actions à couvrir, qui protègent de la baisse via un payoff positif pour une baisse du sous-jacent. L'exemple le plus classique est donc l'option de vente *put*.

Les deux stratégies les plus connues sont :

- la couverture par un put simple ;
- la couverture par un put spread.

Ces deux stratégies sont, chez les assureurs européens, souvent basées sur l'indice ESTX 50, *Euro Stoxx 50*, qui regroupent les 50 plus grosses capitalisation boursières européennes. En effet, la plus part des assureurs possèdent des actions de ces entreprises et, du fait de la diversification, considèrent que cet indice représente une majeure partie de leurs portefeuilles.

Nous nous posons la question de la faisabilité d'une stratégie de couverture avec une option **worst-off put**. Ce put fonctionne exactement comme un put classique mais repose sur un panier d'indice. Le sous-jacent est la moyenne des indices. Une telle stratégie pourrait permettre à un assureur de se spécialiser dans certains secteurs géographique, et ne plus considérer l'Europe entière.

Nous avons décidé de comparer la performance du **FTSE 100**, l'indice britannique, avec l'indice français, le **CAC 40**. Cette comparaison nous est venue du fait des différences dans les évolutions du marché entre ces deux indices. Le Brexit est venu rajouter une autre dimension à l'évolution de ces deux indices. Nous nous posons la question sur ce qu'il en est réellement.

I Présentation des données

Les données financières sont récupérées via le package R *quantmod*. Les données sont récupérées du 1er janvier 2010 au 31 décembre 2023. Nous récupérons ainsi pour chaque jour de cotation :

- le cours d'ouverture ;
- le point le plus haut atteint par l'indice au cours de la séance ;
- le point le plus bas atteint par l'indice au cours de la séance ;
- le cours de fermeture ;
- le volume échangé ;
- le cours ajusté à la fin de la séance.

Nous pouvons calculer, à partir de ces données, les rendements journaliers des deux indices qui nous permettent d'obtenir notre jeu de données sous la forme suivante :

Date	r_{CAC}	r_{FTSE}
2010-01-04	0.0016	0.016
...
2023-12-31	0.0014	0.0011

TABLE 1 – Données utilisées dans notre étude

I.1 Description quantitative des données

Avant de nous lancer dans toutes analyses, nous avons voulu observer la distribution des rendements de nos deux indices.

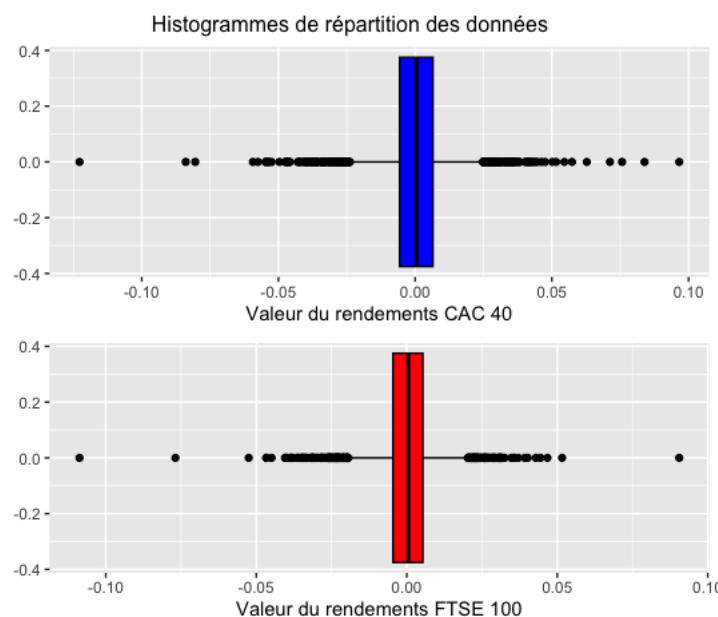


FIGURE 1 – Boxplot des rendements journaliers

On observe que les rendements du **CAC 40** semblent légèrement supérieurs en moyenne aux rendements du **FTSE 100**. Ils semblent également plus dissipés. Pour les deux indices, les rendements sont centrés en 0.

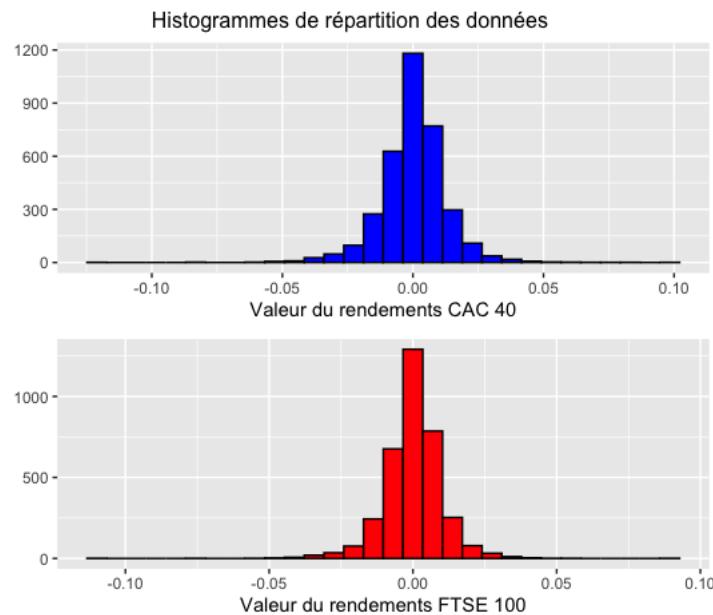


FIGURE 2 – Histogramme des rendements journaliers

Les histogrammes semblent indiquer que nos rendements ont une tendance normale. Cela est important pour modéliser leur dépendance grâce aux copules.

De plus, on observe que la plupart des rendements journaliers de ces indices sont compris dans l'intervalle $\pm 10\%$.

Nous pouvons analyser les **QQ-plot** pour nous fournir une première idée des possibles marginales à utiliser pour la suite :

Q Plot of Sample Data versus Ne Data versus Student-t with :

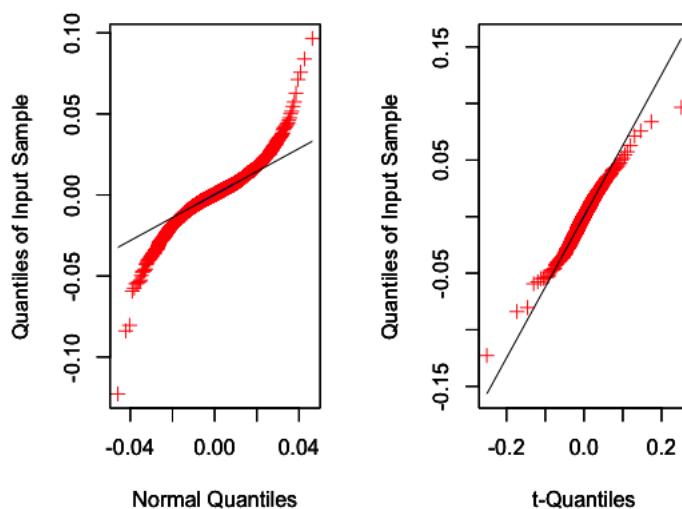


FIGURE 3 – QQ-plot des lois normale et de Student pour le CAC 40

Les lois gaussiennes et de Student semblent adaptées à l'étude de nos rendements. Toutefois, c'est la loi de Student qui paraît être la plus adaptée.

Q Plot of Sample Data versus Ne Data versus Student-t with :

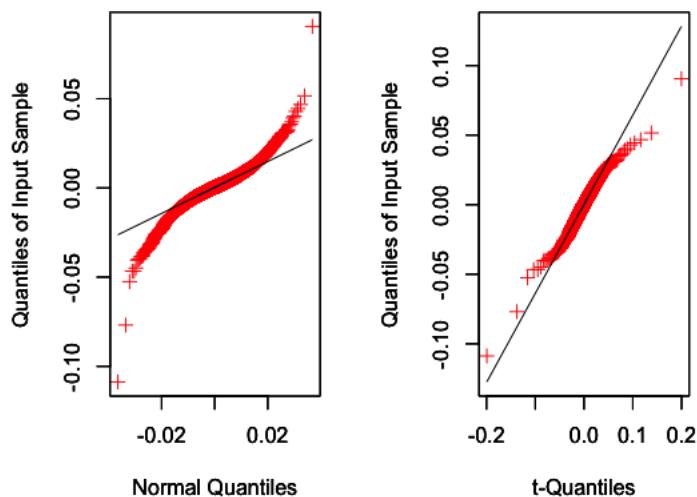


FIGURE 4 – QQ-plot des lois normale et de Student pour le FTSE 100

Les mêmes observations que pour le **CAC 40** peuvent être faites pour le **FTSE 100**.

I.2 Dispersion

Le graphique ci-dessous fournit une représentation visuelle des données, permettant d'observer les patterns de corrélation entre les deux variables. Cependant, il est important de souligner que le diagramme de dispersion présente certaines limites dans la détection de la structure de dépendance spécifique entre les variables.

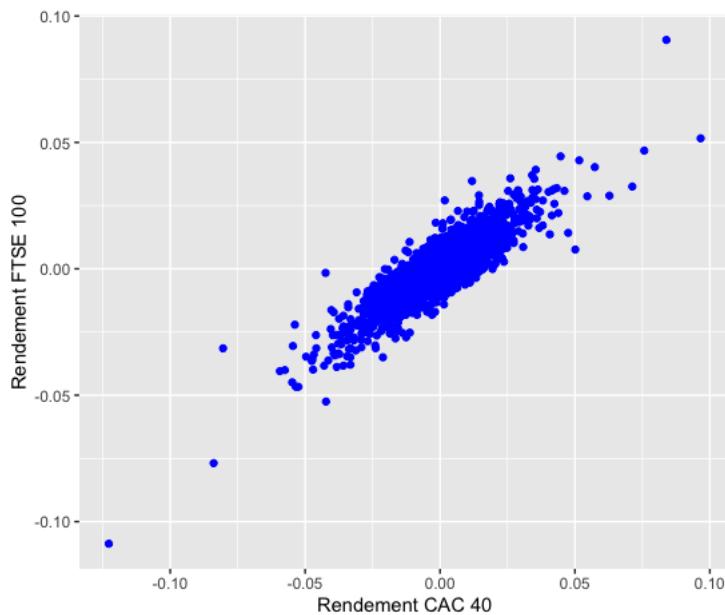


FIGURE 5 – Diagramme de dispersion

On observe un nuage de point regroupé dans une forme ovale. L'orientation de la forme ovale indique une faible corrélation positive. Il est à noter que les valeurs extrêmes présentent également une corrélation positive.

Il faut cependant rester vigilant car la forte concentration de points au centre du nuage peut masquer des corrélations.

Il est donc encore difficile de conclure quant à la structure de dépendance entre les rendements de nos indices.

I.3 Coefficients de corrélation

La corrélation est une analyse bivariée qui mesure la force de l'association entre deux variables et le sens de la relation. En terme de force de la relation, la valeur du coefficient de corrélation varie entre $+1$ et -1 . Une valeur de ± 1 indique un degré parfait d'association entre les deux variables. Plus la valeur du coefficient de corrélation se rapproche de 0, plus la relation entre les deux variables est faible. Le sens de la relation est indiqué par le signe du coefficient :

- un signe + indique une relation positive ;
- un signe - une relation négative.

Dans cette section nous étudierons la corrélation entre les rendements sous 3 formes :

- le coefficient de corrélation de Pearson ;
- le coefficient de rang de Spearman ;
- le τ de Kendall.

On constate que les coefficients sont tous les trois relativement élevés ce qui indique une forte corrélation. Les coefficients étant tous positifs, cela indique une dépendance positive forte.

Pearson	τ de Kendall	ρ de Spearman
0,8686582	0,6338505	0,8189616

TABLE 2 – Critères de sélection de la copule non paramétrique

I.4 K-plots

Un outil d'analyse supplémentaire est le K-plot.

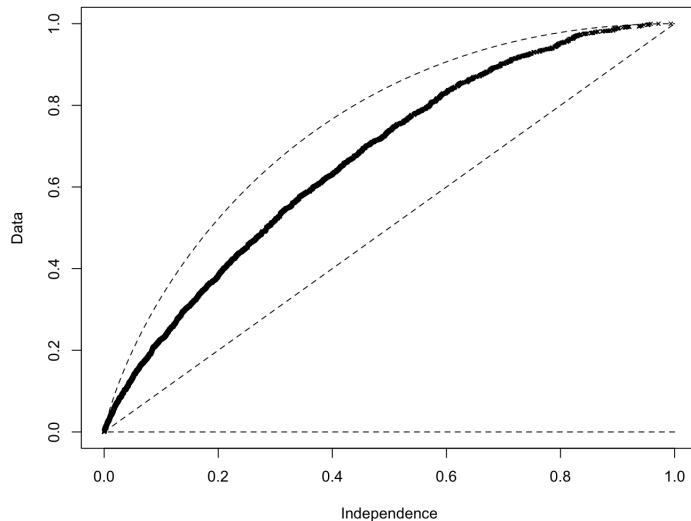


FIGURE 6 – K-plot associé à l'échantillon

L'ensemble des points se situent au dessus de la diagonale ce qui implique une corrélation positive. Ils sont également proche de la courbe supérieure en pointillé, ce qui démontre une forte corrélation.

I.5 Khi-plots

Pour observer, tester et mieux comprendre la relation de dépendance entre nos rendements, nous réalisons un χ -plot en traçant (χ_i, λ_i) . Avec χ_i et λ_i définis grâce aux notations vues dans le cours :

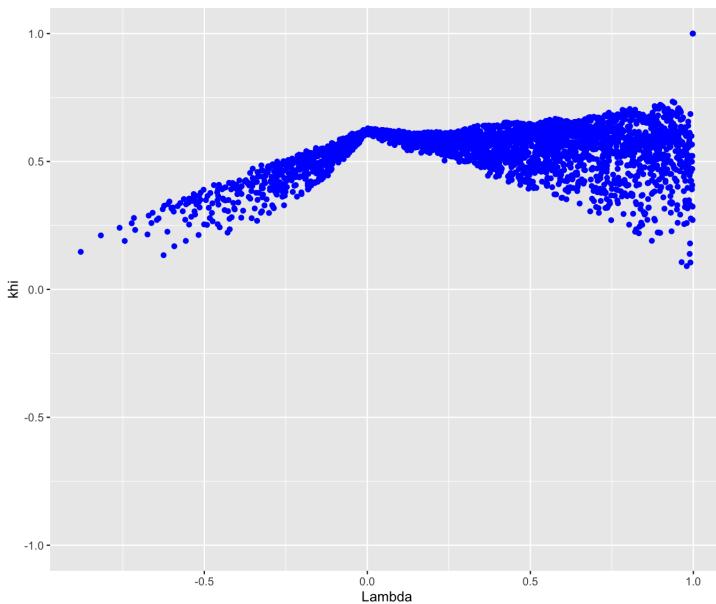


FIGURE 7 – Khi-plot associé à l'échantillon

Des valeurs de χ proches de 0 correspondent au cas d'indépendance c'est à dire $H_i = G_i F_i$. Les points (χ_i, λ_i) auront tendances à être situés au-dessus de zéro pour les marges positivement dépendantes et vice versa pour les marges négativement dépendantes. Les bornes de contrôle, autour de zéro, indiquent si une dépendance significative est présente.

On observe pour notre jeu de données que les rendements des indices ne sont clairement pas indépendants. On peut même ajouter qu'ils sont corrélés positivement. Les estimations du χ -plot réalisées ci-dessous sont des estimations réalisées avec une marge non paramétrique.

Ainsi, toutes les analyses effectuées jusqu'à maintenant permettent d'affirmer plusieurs choses sur nos données :

- il apparaît une forte corrélation positive entre les rendements journaliers de nos deux indices ;
- la loi normale et la loi de Student semblent être de bonne qualité pour l'estimation de nos données, ce qui nous sera utile pour l'estimation paramétrique.

Pour la suite, nous allons nous intéresser à la copule empirique, puis nous examinerons les estimations paramétriques et non paramétriques de la copule associée à nos données.

II Analyse de la copule empirique

D'après le théorème de Sklar, la copule est fonction de marges uniformes. Par conséquent, nous transformons les observations de notre jeu de données en marges uniformes standard :

$$(U_i^{CAC}, V_i^{FTSE}) = \left(\frac{R_i^{CAC}}{n}, \frac{S_i^{FTSE}}{n} \right)$$

Avec

- R_i^{CAC} et R_i^{FTSE} les rangs des rendements journaliers de nos deux indices ;
- n : Le nombre de jours de notre période d'étude

Cette transformation nous permet d'obtenir le rank-rank plot suivant :

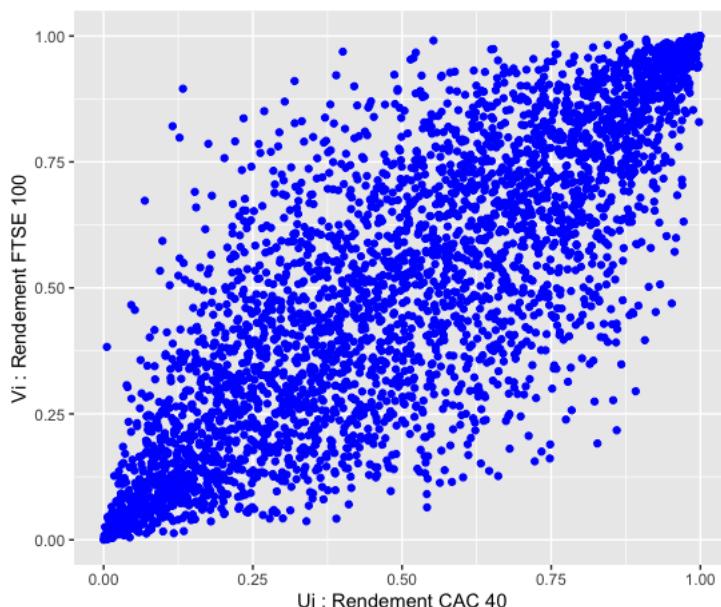


FIGURE 8 – Rank-rank plot des marges empiriques

On observe une concentration de points autour des angles en bas à gauche et en haut à droite, ce qui indique une dépendance forte des extrêmes. De plus, il semble ne pas y avoir de concentration autour des deux autres angles de ces graphes. Cette répartition de points nous fait donc penser à la copule de Student. Ensuite, on constate que les points semblent répartis autour de la première diagonale, ce qui implique une corrélation positive entre nos données.

On peut maintenant tracer la copule empirique :

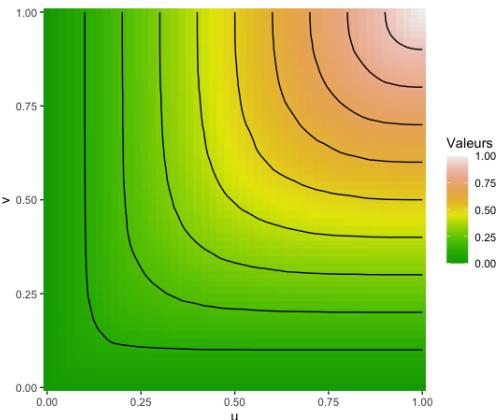


FIGURE 9 – Tracé en 2D de la copule empirique

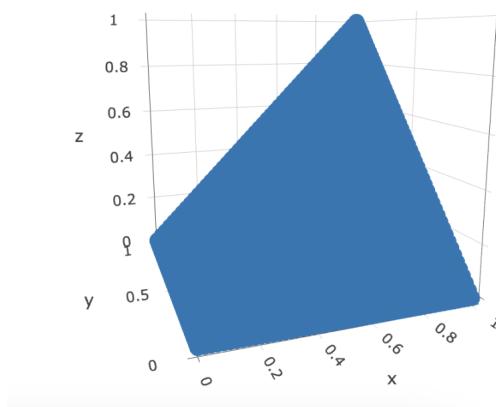


FIGURE 10 – Tracé en 3D de la copule empirique

Nous avons voulu comparer nos données à trois copules classiques que nous connaissons bien :

- la copule indépendante ;
- la copule comonotone ;
- la copule antimonotone.

Pour comparer la copule empirique à ces trois copules, nous regarderons la représentation en deux dimensions.

Copule indépendante : Pour rappel, la copule indépendante est telle que $c(u, v) = uv$.

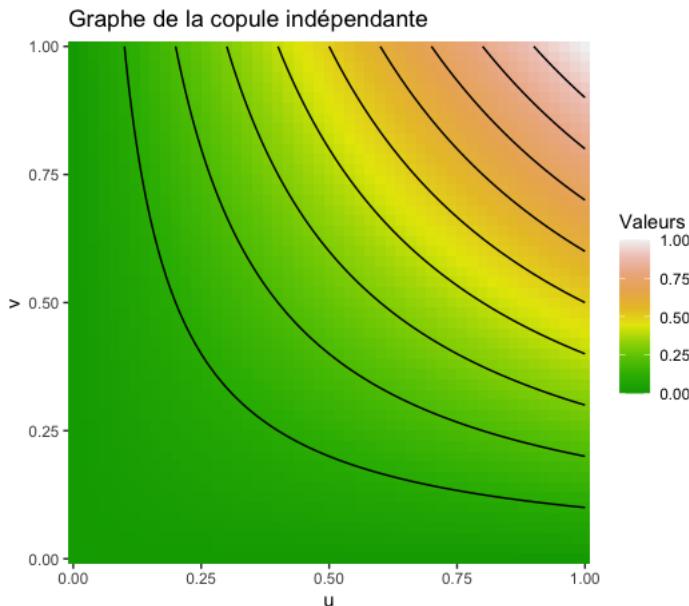


FIGURE 11 – Copule indépendante

Elle ne présente que très peu de lignes de niveaux mais sa forme est triangulaire car la valeur de la copule pour u est linéaire en v et inversement.

Copule comonotone : Pour rappel, la copule comonotone est telle que $c(u, v) = \min(u; v)$.

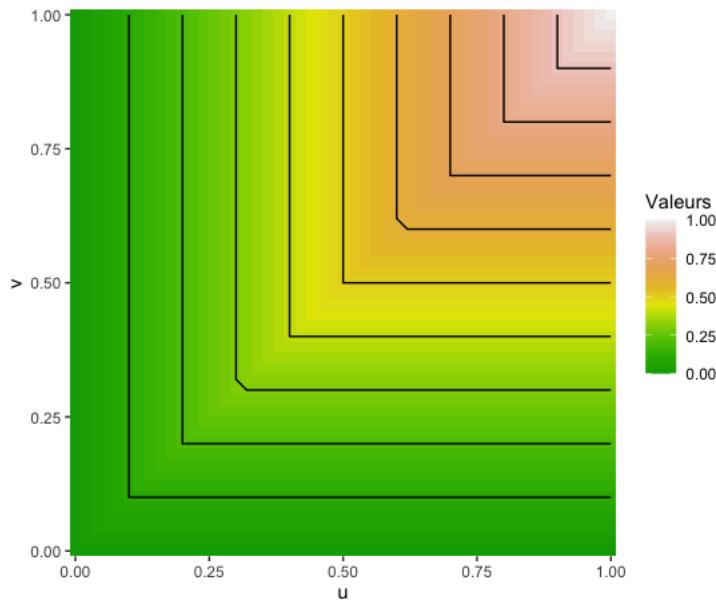


FIGURE 12 – Copule comonotone

On ne distingue pas de lignes de niveau, c'est le cas extrême de dépendance positive.

Copule antimonotone : Pour rappel, la copule antimonotone est telle que $c(u, v) = \max(u + v - 1; 0)$.

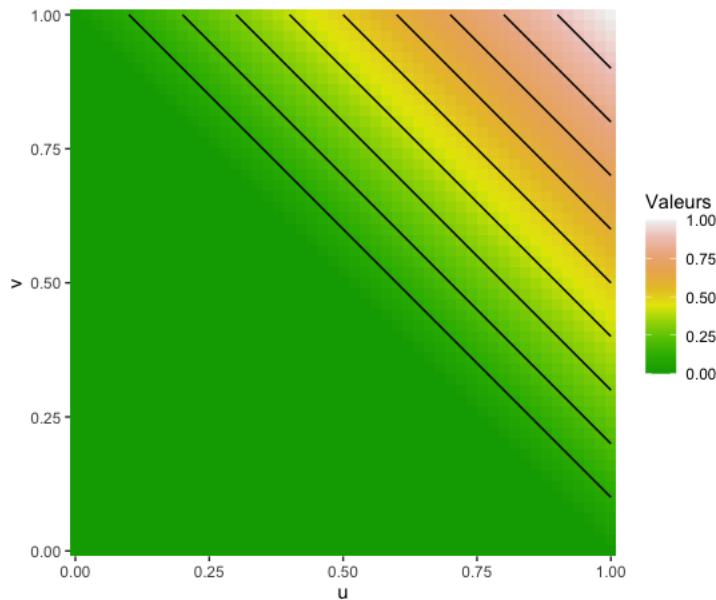


FIGURE 13 – Copule antimonotone

On observe dans notre cas que notre copule présente une forme plus elliptique que les deux cas extrêmes, avec des lignes de niveaux qui sont présentent et arrondies. Elle

se rapproche cependant plus de la copule avec une dépendance positive. Cela confirme notre intuition lors de l'observation du diagramme de dispersion. La forme générale de la copule empirique semble ressembler à une copule de Student ou Gaussienne car elle paraît elliptique. Il faudra donc être vigilant quant à nos résultats concernant la corrélation linéaire de nos rendements car - Paul EMBRECHTS, 2001 - à montrer que en présence de copules elliptiques les résultats pouvaient s'avérer contre-intuitifs. Cependant il faudra étudier la dépendance aux extrêmes pour confirmer cette idée.

Ainsi, on observe clairement que la copule empirique ne ressemble pas à ces copules connus.

Cette estimation de la copule empirique est une première estimation non paramétrique réalisée à l'aide des pseudos-observations

III Estimation paramétrique

Dans cette phase de l'étude, une hypothèse a été formulée concernant les distributions marginales des variables analysées.

D'après ce qui a été clairement expliqué par la *Figure 3* et par la *Figure 4*, nous allons étudier les distributions marginales normales et de Student.

Nous avons ainsi :

$$(U_i^{\text{CAC}}, V_i^{\text{FTSE}}) = (F(X_i^{\text{CAC}}), G(X_i^{\text{FTSE}}))$$

Avec :

- F : La fonction de répartition d'une loi centrée en μ_{CAC} , d'écart type σ_{CAC} ;
- G : La fonction de répartition d'une loi centrée en μ_{FTSE} , d'écart type σ_{FTSE} .

Les paramètres de ces lois ont été estimés par la méthode du maximum de vraisemblance.

Rank-Rank plot :

Nous avons tout d'abord représenté les Rank-Rank plot avec les deux lois :

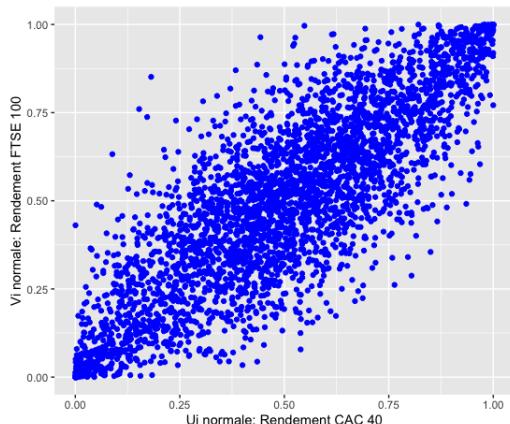


FIGURE 14 – Rank-rank plot avec marginales gaussiennes

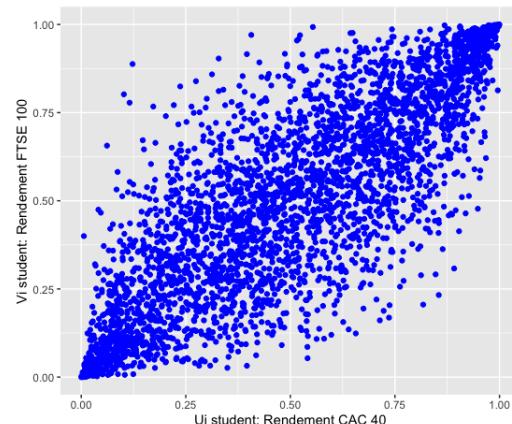


FIGURE 15 – Rank-rank plot avec marginales Student

On observe une concentration de points autour des angles en bas à gauche et en haut à droite, ce qui correspond à ce que nous disions de la copule empirique, à savoir une dépendance forte des extrêmes. Cependant, les marginales gaussiennes présentent un nuage de points aux extrêmes qui sont plus dispersés et moins denses.

D'après ce qui découle de cette analyse, nous aurions tendance à privilégier les marginales de Student. Cependant, nous testerons également les marginales gaussiennes qui représente bien les données.

K-plots :

Nous avons également voulu représenter les K-plot pour les différentes copules.

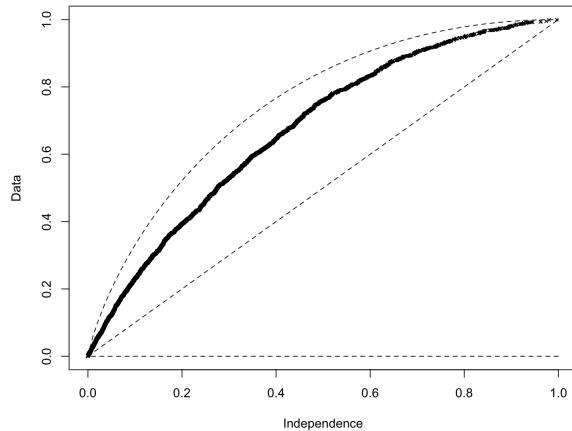


FIGURE 16 – K-plot avec copule gaussienne

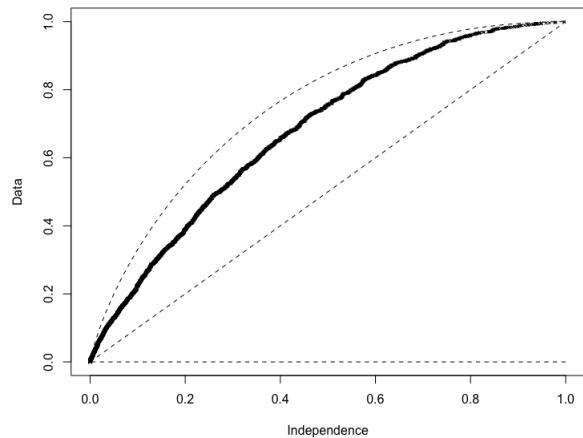


FIGURE 17 – K-plot avec copule de Student

La forme de notre copule donnée en *Figure 6* semble bien correspondre aux deux K-plot ci-dessus. Les deux graphiques sont très semblables. Il serait difficile de les départager à l’œil nu. On va donc tester les deux familles de copule.

Comparaison des estimations

Hésitant entre les copules gaussiennes et les copules de Student, nous avons utilisés des tests statistiques :

- Les test de Cramer-Von-Misses et de Kolmogorov-Smirnov pour la copule gaussienne ;
- Le test de White pour la copule de Student.

On utilise la fonction **BiCopSelect** du package *VineCopula* qui sélectionne la meilleure famille bivariée en utilisant le critère de l’AIC. Les copules sont ajustées sur les données, puis en utilisant le critère de l’AIC, la fonction détermine la meilleure famille de copule selon ce critère.

Marge	Gaussienne	Student
Copule sélectionnée	Student	Student
Dimensions	2	2
AIC	-4576,06	-4950,44
BIC	-4563,72	-4938,1
Log vraisemblance	2290,03	2477,22

TABLE 3 – Critères de sélection de la copule paramétrique

IV Estimation non paramétrique

IV.1 La copule empirique

L'estimation de la copule empirique que nous avons réalisé est une estimation naïve mais qui peut être considérée comme non paramétrique. Ainsi, nous réaliserons les tests similaires à ceux que nous avons réalisés pour les méthodes paramétriques.

Marge	Empirique
Copule sélectionnée	BB1
Dimensions	7
AIC	-4578,05
BIC	-4565,71
Log vraisemblance	2291,02

TABLE 4 – Critères de sélection de la copule empirique

La définition de la copule BB1 est donné en *Annexe A*.

IV.2 Méthodes non paramétriques

Grâce au package **kdecopula** disponible dans R, développé par Nagler, nous avons accès à des outils puissants pour la modélisation non paramétrique des copules. Ce package fournit des fonctionnalités pour estimer les copules à l'aide d'estimateurs à noyau, une méthode non paramétrique populaire pour estimer les densités de copules à partir de données échantillonnées.

Comme abordé en cours, il existe une variété d'estimateurs à noyau pour estimer les copules d'un échantillon. Ces estimateurs diffèrent dans leur construction et leurs propriétés. La performance de chaque estimateur est différente, notamment en termes de gestion des problèmes liés aux bords des distributions. Dans notre analyse, nous avons opté pour l'utilisation d'un estimateur à noyau particulier fourni par kdecopula. Ce choix est motivé par le désir d'éviter les problèmes communs rencontrés aux bords de la distribution, qui peuvent fausser les estimations de la copule, surtout dans les zones de faible densité de données ou près des limites de l'échelle des variables.

L'utilisation de cet estimateur à noyau spécifique nous permet d'obtenir une estimation plus fiable et précise de la densité de la copule pour notre échantillon. Cela nous aide à mieux comprendre la structure de dépendance sous-jacente entre les données, sans les contraintes imposées par les modèles paramétriques. Cette approche complète nos résultats paramétriques et offre une perspective plus large sur la relation entre ces deux variables importantes dans les données de musique.

Les quatre méthodes que nous avons utilisées pour l'estimation non paramétrique des densités de copule dans notre étude sont les suivantes :

- **TLL**¹ : Cette méthode, expliquée par G. Geenens et al. en 2014², repose sur une transformation locale des données pour maximiser la vraisemblance. La méthode TLL a plusieurs variantes, offrant une flexibilité dans la modélisation de la densité de la copule.
- **T³** : Présenté par A. Charpentier et al. en 2006⁴, cet estimateur utilise une transformation des données pour estimer la densité de la copule. L'approche se concentre sur la modification des marges pour obtenir une meilleure estimation de la copule.
- **Bêta** : Cette méthode utilise le Bêta Kernel, comme décrit par A. Charpentier en 2006. L'approche du Bêta Kernel est particulièrement utile pour estimer les densités de copule dans les zones proches des bords, où d'autres méthodes peuvent rencontrer des difficultés.
- **MR Mirror-Reflection Estimator** : Développé par I. Gijbels en 1990⁵, le mirror-reflection estimator est une technique qui améliore l'estimation de la densité de la copule en réfléchissant les données autour des bords. Cette méthode aide à réduire les biais et les problèmes liés aux estimations près des limites de la distribution.

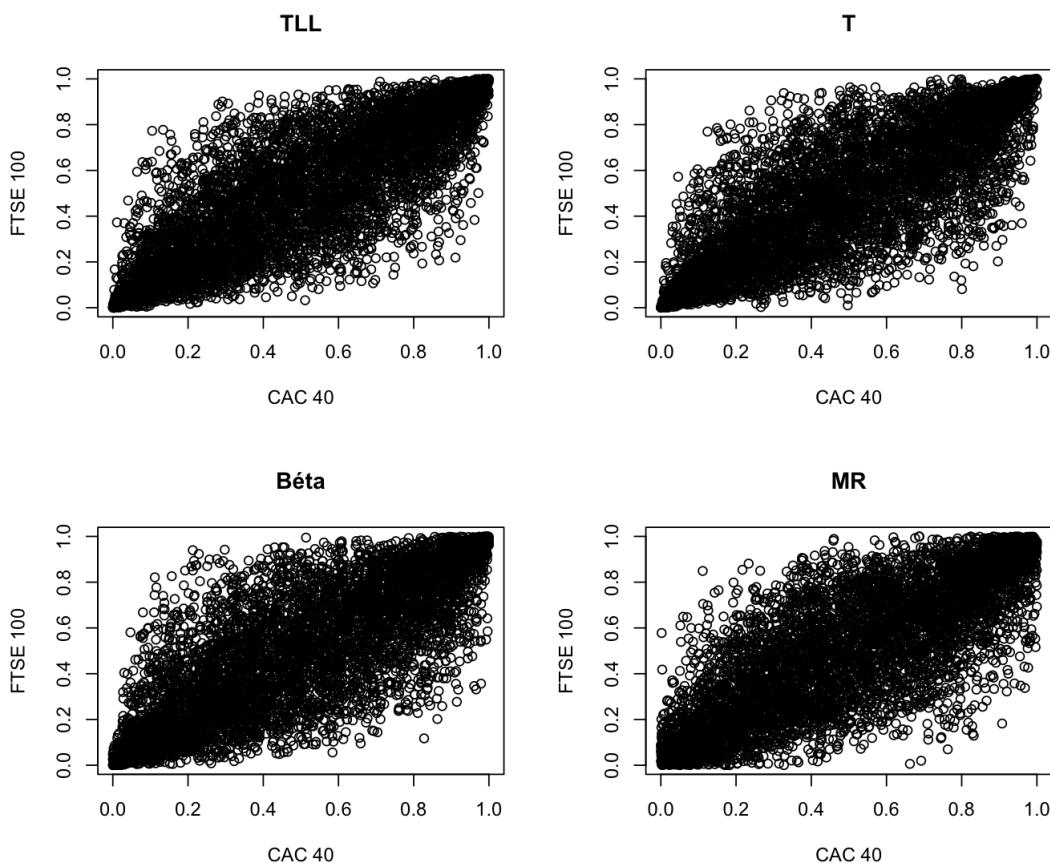


FIGURE 18 – Rank-rank plot par méthodes non paramétriques

1. Transformation Locale du Maximum de Vraisemblance
2. Probit Transformation for Non-Parametric Kernel Estimation of the Copula Density
3. Estimateur de Transformation
4. The Estimation of Copulas : Theory and Practice
5. Estimating the Density of a Copula Function

Les densités estimées par les quatre méthodes (TLL, T, Bêta et MR) montrent une forte dépendance dans les extrêmes à gauche et à droite. De plus, les formes de densité sont similaires entre les méthodes, suggérant une cohérence dans la capture de la structure de dépendance. Cependant, les méthodes Bêta et MR semblent moins précises aux extrêmes avec une densité plus faible au niveau des points de par et d'autre de la distribution. Les méthodes TLL et T sont très « pointu » à ces endroits là ce qui implique une plus grandes précisions aux extrêmes. C'est pour cela que nous conserverons l'une de ces deux méthodes seulement.

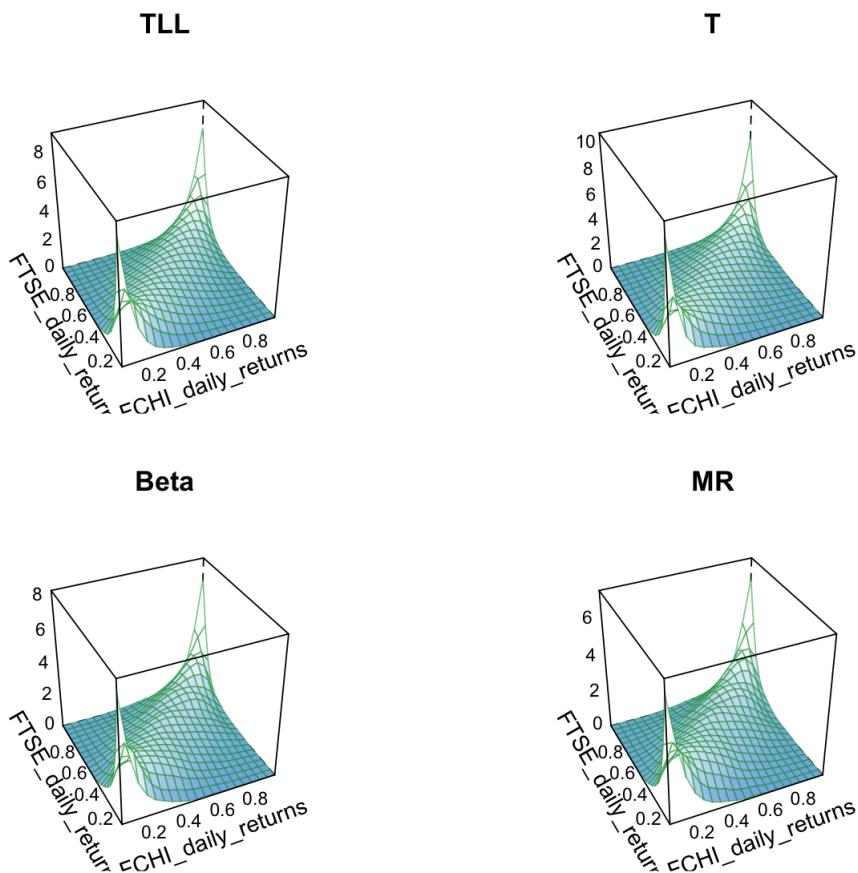


FIGURE 19 – Densité selon les méthodes non paramétriques

Les densités ressemblent très fortement à ce que nous attendions.

En conclusion, ces analyses non paramétriques complètent et confirment les résultats obtenus par les méthodes paramétriques, tout en offrant des perspectives supplémentaires sur la nature de la dépendance entre nos variables d'intérêt.

IV.3 Qualité de l'ajustement non paramétrique

Afin de choisir la copule qui nous paraît la plus adaptée à notre échantillon de données, nous avons décidé d'utiliser les critères présentés dans le tableau ci-dessous.

Méthode	TLL	T	Beta	MR
AIC	-4577,84	-4716,43	-4393,98	-4215,7
BIC	-4488,28	-4697,99	-3894,2	-4117,91
Log vraisemblance	2303,44	2361,21	2278,03	2123,71

TABLE 5 – Critères de sélection de la copule non paramétrique

La **méthode T**, qui utilise l'estimateur de transformation, se distingue nettement par sa capacité à minimiser les critères AIC (Akaike Information Criterion) et BIC (Bayesian Information Criterion), tout en maximisant la log-vraisemblance.

Cette analyse reste cohérente avec les observations que nous avons faites précédemment en nous basant sur les densités et les rank-rank plots.

Ainsi, la meilleure copule non-paramétrique pour notre analyse est celle estimée par l'estimateur de transformation. Cette conclusion est soutenue par des critères statistiques rigoureux et par une analyse visuelle des données, confirmant que cette approche fournit une représentation fidèle et précise de la structure de dépendance entre nos variables d'intérêt.

Conclusion

En comparant les résultats obtenus à travers les approches paramétrique et non paramétrique, notre objectif est de déterminer la copule qui modélise le mieux la dépendance entre les rendements de nos deux indices.

Les analyses paramétriques ont montré que la copule de Student est la plus adaptée à nos données. C'est la copule minimisant l'AIC et le BIC, tout en maximisant la log vraisemblance. D'après les exposés réalisés en ERM, ce résultat nous semble cohérent. Dans l'approche non paramétrique, selon les mêmes critères, c'est la méthode T que nous conserverons.

Bien que les deux approches fournissent des résultats convaincants, si nous devons faire un choix entre elles, nous privilégierions l'approche paramétrique du fait des critères plus discriminants.

En conclusion, notre étude nous conduit à sélectionner la copule de Student pour modéliser la relation entre les rendements du **CAC 40** et du **FTSE 100**. Cette copule offre une compréhension approfondie et nuancée de la structure de dépendance, alignée sur notre problématique initiale.

Dans l'idée d'utiliser une stratégie de couverture basée sur un *put worst-off*, il faudrait ensuite réaliser des simulations de trajectoires à l'aide de la copule sélectionnée. Une fois les simulations effectuées, une méthode de pricing classique via *Monte-Carlo* permet d'obtenir le prix d'une telle option. L'utilisation des copules est beaucoup plus flexible par rapport aux méthodes classiques pour le pricing d'option, et ceux notamment pour le pricing d'option sur plusieurs sous-jacent.

Bibliographie

BiCopSelect. (s.d.). *Selection and Maximum Likelihood Estimation of Bivariate Copula Families.* <https://search.r-project.org/CRAN/refmans/VineCopula/html/BiCopSelect.html>

cran.r, P. (s.d.). *Copula Modeling.* <https://cran.r-project.org/web/packages/univariateML/vignettes/copula.html>

fitCopula, R. (s.d.). *Fitting Copulas To Data – Copula Parameter Estimation.* <https://www.rdocumentation.org/packages/copula/versions/1.0-1/topics/fitCopula>

Nagler, T. (s.d.). *kdecopula : An R Package for the Kernel Estimation of Copula Densities.* <https://cran.microsoft.com/snapshot/2017-04-02/web/packages/kdecopula/vignettes/kdecopula.pdf>

Annexes

A Copule BB1

La copule BB1, également connue sous le nom de copule de Clayton-Gumbel ou copule de Gumbel-Hougaard inversée, fait partie de la famille des copules archimédiennes. Elle est particulièrement utile pour modéliser des dépendances asymétriques entre les marges, où la dépendance est plus forte dans la queue supérieure qu'inférieure, ou vice versa.

Les copules BB1 sont définies par une fonction générateur de copule archimédienne et son inverse. La copule BB1 est paramétrée par deux paramètres, habituellement notés θ et δ , qui contrôlent respectivement la dépendance globale et l'asymétrie de la copule. Le paramètre θ doit être positif ($\theta > 0$) et δ doit être supérieur ou égal à 1 ($\delta \geq 1$).

La fonction générateur de la copule BB1 est donnée par :

$$\phi(t) = (1 + t^{-\theta})^{-\delta}$$

Et sa dérivée, qui est utilisée pour construire la fonction de copule elle-même, est :

$$\phi'(t) = -\delta\theta t^{-\theta-1} (1 + t^{-\theta})^{-\delta-1}$$

La copule BB1 est capable de capturer une large gamme de structures de dépendance, y compris des dépendances plus fortes dans les queues supérieures, ce qui la rend utile dans divers domaines, notamment en finance et en assurance pour modéliser des risques conjoints et des événements extrêmes.

En R, la copule BB1 peut être utilisée avec le package `VineCopula`. Vous pouvez créer un objet copule BB1 et l'ajuster aux données pour étudier la structure de dépendance entre les variables.



Fin du Rapport