

# Chapitre 1

## Présentation des données et étude de normalité

Les données utilisées dans ce projet sont relativement accessibles, notamment sur les plateformes Boursorama, Yahoo Finance ou encore directement via des API spécialisées :

- <https://fr.finance.yahoo.com/quote/AIR.PA/>
- <https://fr.finance.yahoo.com/quote/SAF.PA/>

Étant donné que l'on souhaite étudier la dépendance des actifs, notamment sur des plages de distributions particulières, j'ai décidé de choisir une période de forte instabilité bancaire pour donner davantage d'intérêt à mon projet, en rendant mes observations les plus volatiles possibles.

J'ai choisi comme période l'année 2020, le début d'une période de pandémie, économiquement très compliquée pour les entreprises évoluant dans le domaine du transport.

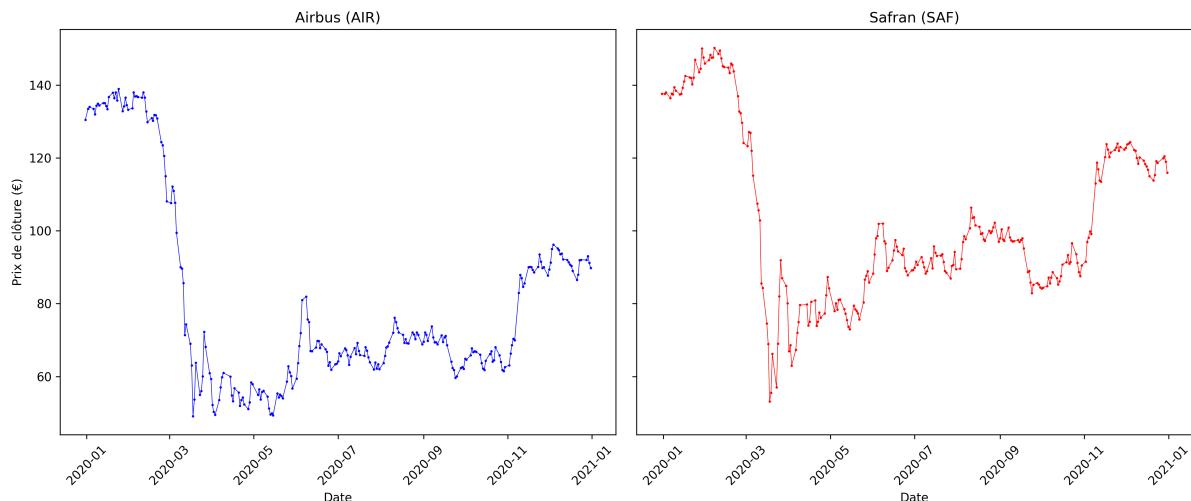


FIG. 1.1 : Evolution des cours

**Remarque :** Étant donné que ce projet a pour but de mettre en lumière certaines limites de modèles financiers connus, il se concentrera sur l'étude de la distribution des rendements logarithmiques (souvent estimés comme normalement distribués), ainsi que sur leurs corrélations.

**Rappel :** Le rendement logarithmique  $R_t$  est défini comme :

$$R_t = \ln \left( \frac{S_t}{S_{t-1}} \right)$$

où  $S_t$  représente le prix de l'actif à l'instant  $t$ .

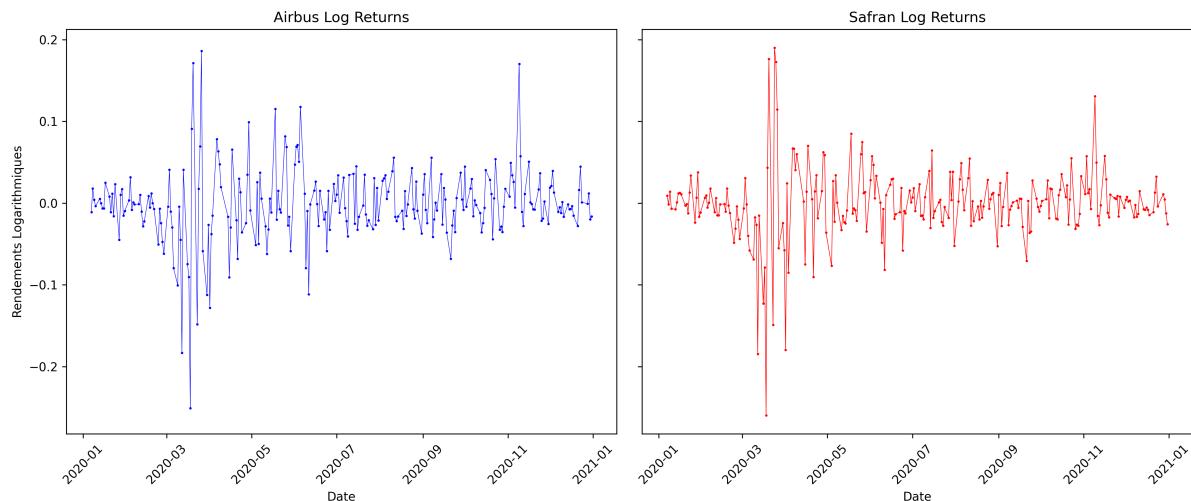


FIG. 1.2 : Rendements logarithmiques

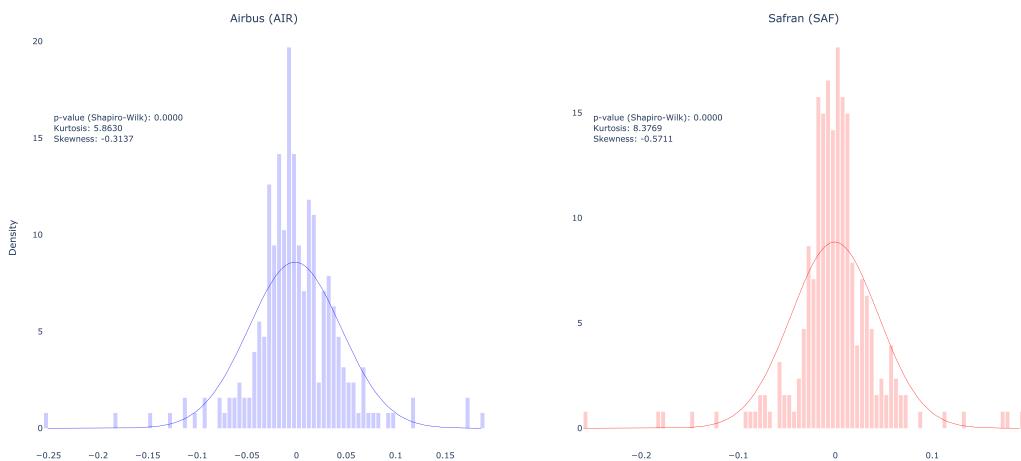


FIG. 1.3 : Distribution des rendements logarithmiques

**Remarque :** Comme attendu, l'ajustement de mes lois normales sur mes rendements logarithmiques est vraiment très mauvais. On remarque que le test de Shapiro-Wilk rejette totalement l'hypothèse de normalité. Le coefficient d'asymétrie (skewness) est très négatif en raison des valeurs extrêmes liées à la crise. La valeur du coefficient d'aplatissement est également très élevée (bien supérieure à 3), ce qui confirme que les queues de ma distribution sont bien trop épaisses pour faire l'hypothèse d'une distribution normale.

Pour confirmer notre intuition à ce sujet, on peut également observer les QQ-plot de mes observations.

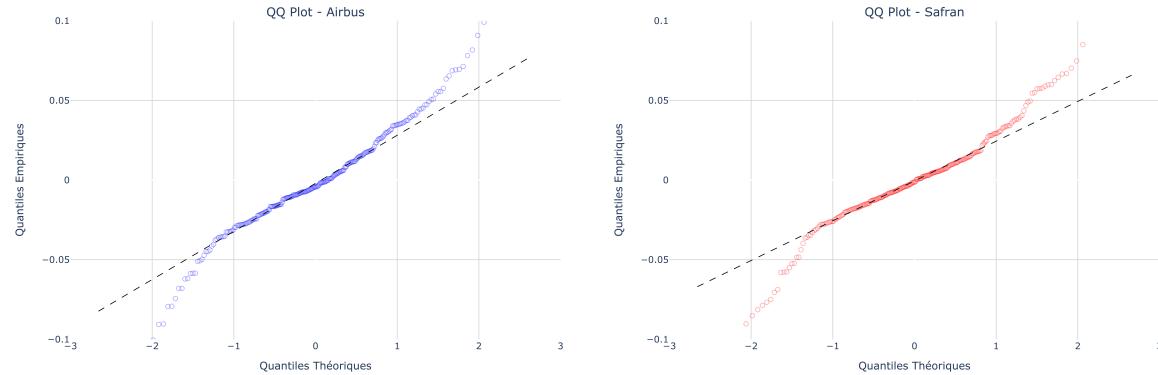


FIG. 1.4 : QQ Plot

**Conclusion :** On remarque que mes quantiles empiriques se détachent de mes quantiles théoriques normaux bien avant la barrière des 5%, ce qui confirme notre intuition. Mes rendements logarithmiques ne peuvent pas être considérés comme normalement distribués dans mon étude.

# Chapitre 2

## Etude de la dépendance

### 2.1 Corrélation entre les variables

Un premier aperçu de la dépendance de mes variables peut être donné par les différents coefficients de corrélation. Les coefficients à estimer sont ceux de Pearson, Kendall et Spearman.

La différence majeure entre ces coefficients est que le coefficient de Pearson mesure la corrélation linéaire. Le tau de Kendall se base plutôt sur la notion de paires concordantes et le rho de Spearman se basent sur le rang de mes observations.

	Pearson	Kendall	Spearman
Coefficient	85.39361%	62.02607%	79.73751%

**Interprétation :** De façon générale, les trois coefficients traduisent une dépendance forte. Étant donné que le tau de Kendall et le rho de Spearman raisonnent sur le rang et les paires concordantes, ils sont donc moins sensibles aux valeurs extrêmes et sont donc ici naturellement moins élevés que le coefficient de Pearson.

## 2.2 Etude de la dépendance des observations

### 2.2.1 Diagramme de dispersion

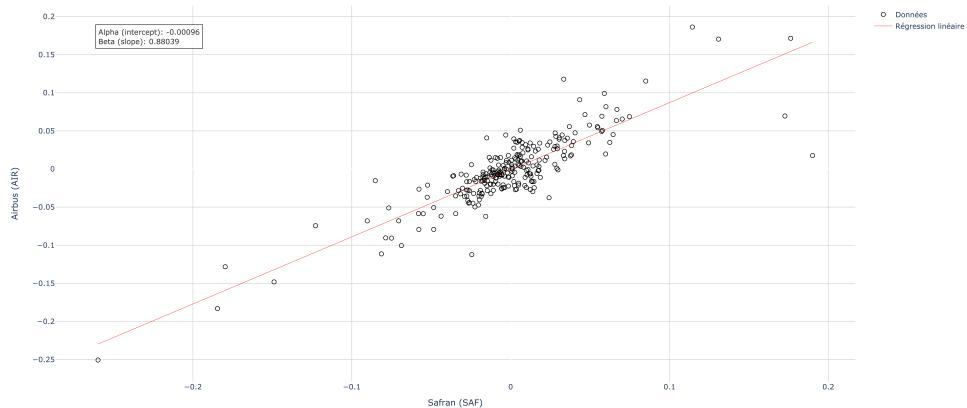


FIG. 2.1 : Diagramme dispersion

**Interprétation :** La dispersion de mes points n'est pas du tout aléatoire. On remarque que mes observations se trouvent relativement proches de la droite d'équation  $y = x$ .

### 2.2.2 Khi plot

Le Khi-plot est une représentation graphique qui permet de détecter la présence d'associations entre deux variables aléatoires. Il s'appuie sur les rangs des observations ainsi que sur la statistique du  $\chi^2$  pour tester l'indépendance.

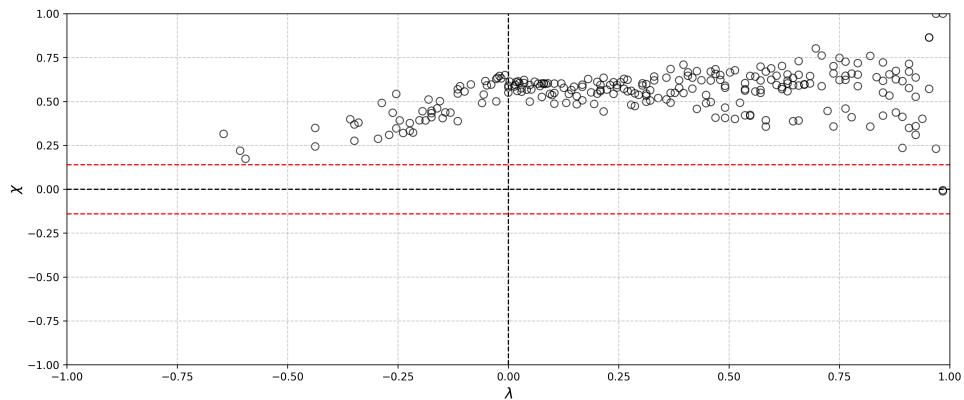


FIG. 2.2 : Khi Plot

**Interprétation :** Toutes les données se situent en dehors de la zone d'acceptation d'indépendance. On peut également rajouter que les points se situent au dessus de cette zone ce qui traduit une dépendance positive.

### 2.2.3 K-Plot (Diagramme de Kendall)

Le K-Plot est une représentation graphique qui évalue la dépendance monotone. Il est également fondé sur le rang des observations ainsi que sur le  $\tau$  de Kendall.

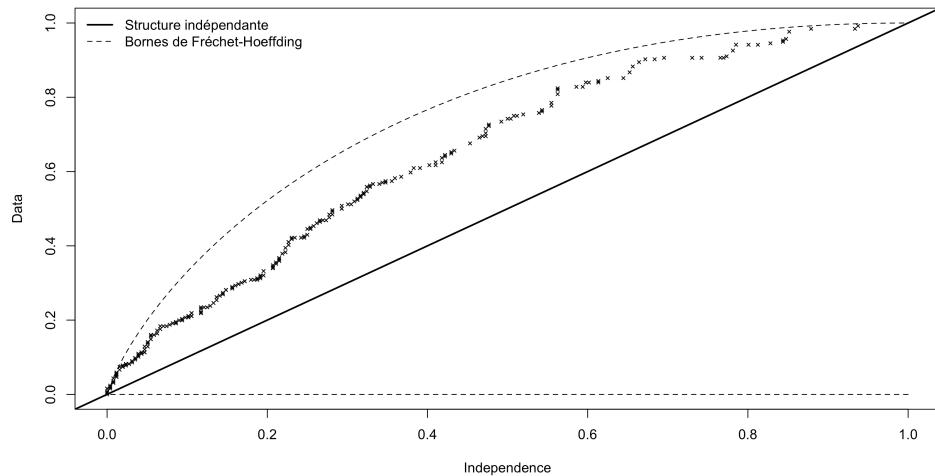


FIG. 2.3 : K-Plot

**Interprétation :** On retrouve une fois de plus la notion de corrélation positive car mes points se situent au-dessus de la droite d'équation  $y = x$  et se rapprochent de la borne supérieure de Fréchet.

# Chapitre 3

## Etude de la dépendance des marginales

### 3.1 Répartitions empiriques

#### 3.1.1 Fonctions de répartition empiriques marginales

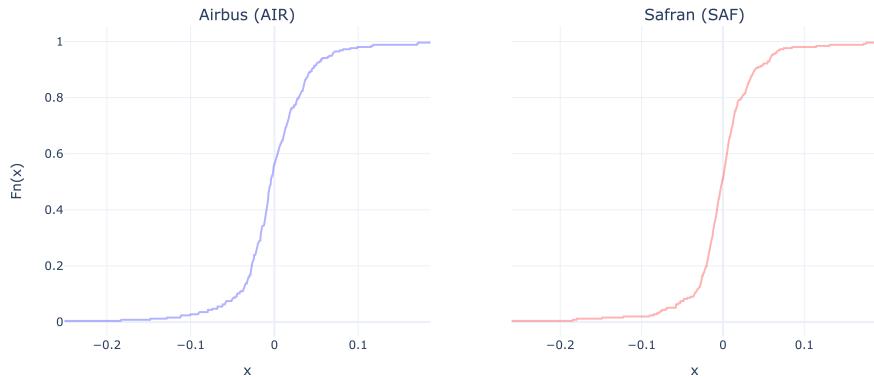


FIG. 3.1 : Répartitions marginales empiriques

#### 3.1.2 Densité empirique multivariée / Kernel density estimation (méthode des noyaux)

L'estimation de la densité empirique peut se faire avec un noyau gaussien.

La forme paramétrique de la densité jointe empirique à noyau gaussien d'un vecteur de variables aléatoires  $X = (X_1, X_2)$  estimée sur un historique  $(x_1^t, x_2^t)$  de taille  $T$  est donnée par la formule suivante :

$$\hat{f}(x) = \frac{1}{Th^2} \sum_{t=1}^T K\left(\frac{1}{h}(x - x^t)\right)$$

avec pour tout  $z \in \mathbb{R}^2$ ,

$$K(z) = (2\pi)^{-1} \exp\left(-\frac{1}{2}z'z\right)$$

Le paramètre  $h$  représente le paramètre de lissage (largeur bande) qui caractérise la densité. On peut l'estimer par le maximum de vraisemblance.

L'estimation de  $h$  est très importante, si  $h$  est trop petit, notre densité sera bruité mais si le paramètre est trop grand, notre densité va perdre sa qualité d'ajustement.

Etant donné que nous sommes en deux dimensions, je décide d'utiliser la fonction **MASS ::kde2d** de la librairie **MASS** pour l'évaluation de ma densité.

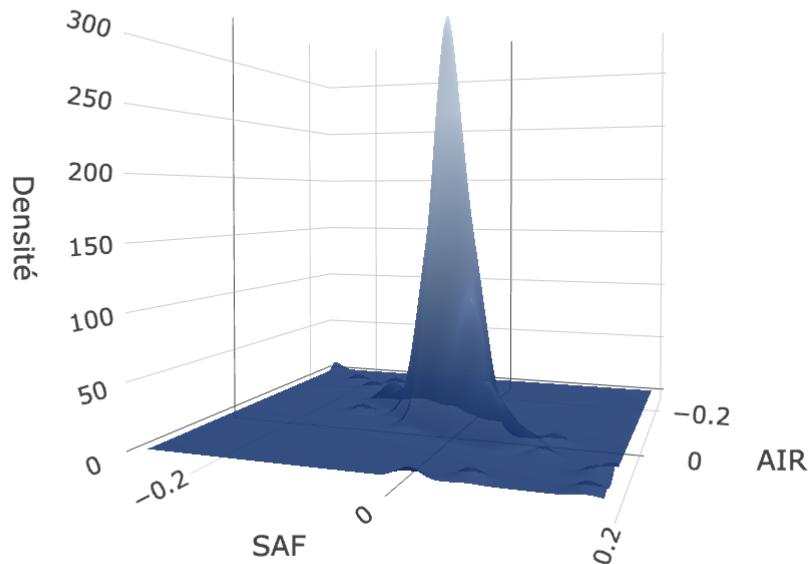


FIG. 3.2 : Densité empirique multivariée

### 3.1.3 Rank-Rank Plot

Le Rank-Rank plot est un outil important pour définir la structure de dépendance, et donc potentiellement la copule idéale à ajuster à nos données. On peut également facilement observer la dépendance des valeurs extrêmes, ce qui pourrait par exemple rapidement nous orienter vers des copules de Clayton, Gumbel ou encore de Student.

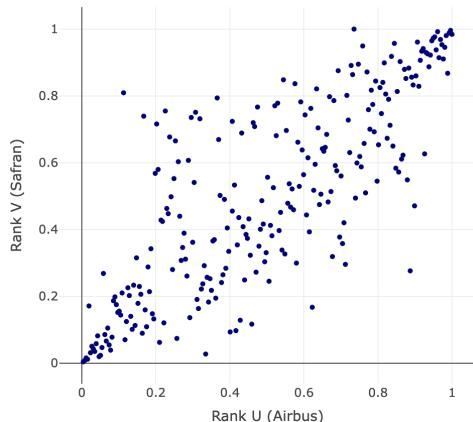


FIG. 3.3 : Rank Rank plot 2D

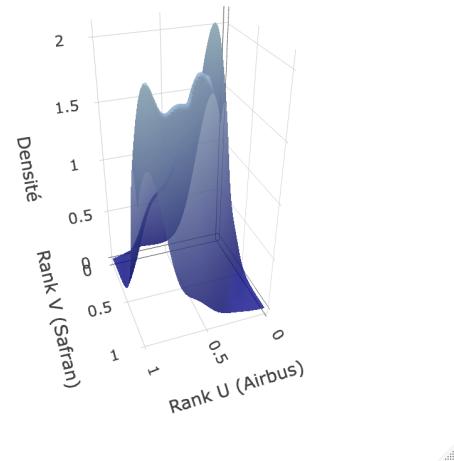


FIG. 3.4 : Rank Rank plot 3D

**Remarque :** En effet, on remarque que les pseudo-observations sont bien alignées sur la droite d'équation  $y = x$ , ce qui confirme une fois de plus la corrélation positive. Il y a également une forte densité d'observations aux deux extrémités de cette droite, ce qui correspond à une forte corrélation des extrêmes.

## 3.2 Copule empirique

### 3.2.1 Répartition du copule empirique

Definition mathématique de la copule empirique :

$$C_K \left( \frac{k_1}{K}, \dots, \frac{k_n}{K} \right) = \frac{1}{K} \sum_{k=1}^K \mathbb{1} \left\{ x_1^{(k_1)} \leq x_1^k, \dots, x_n^{(k_n)} \leq x_n^k \right\}$$

Etant donné que la copule est invariante par toute transformation croissante des marginales, il est plus facile de l'exprimer informatiquement en fonction du rang des observations :

$$C_K \left( \frac{k_1}{K}, \dots, \frac{k_n}{K} \right) = \frac{1}{K} \sum_{k=1}^K \mathbb{1} \left\{ r_1^k \leq k_1, \dots, r_n^k \leq k_n \right\}$$

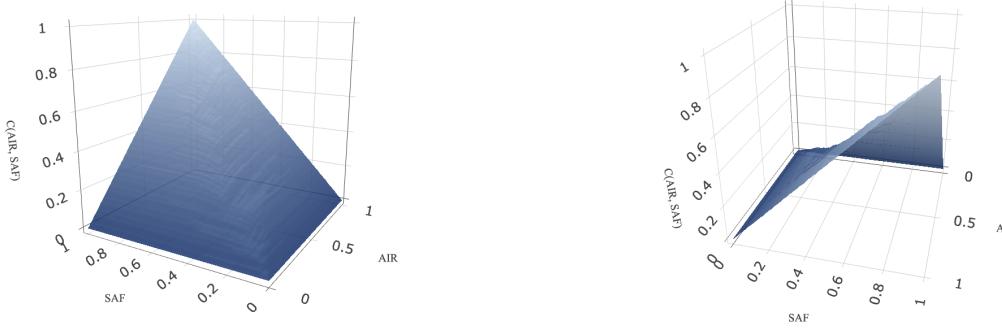


FIG. 3.5 : Copule empirique (fonction de répartition)

**Remarque :** On peut simplement se ramener à la définition mathématique :

$$C(u, v) = F(F_1^{-1}(u), F_2^{-1}(v))$$

et estimer chaque fonction de répartition facilement de façon empirique.

### 3.2.2 Densité du copule empirique

Cette estimation est plus importante que la précédente et va nous permettre de comparer notre densité empirique au copule paramétrique choisi afin d'effectuer une validation graphique de notre modèle.

Selon le théorème de Sklar, elle peut être exprimée à l'aide d'une copule  $C$  :

$$F(x_1, x_2) = C(F_1(x_1), F_2(x_2))$$

où  $F_1(x_1)$  et  $F_2(x_2)$  sont les fonctions de répartition marginales de  $X_1$  et  $X_2$ .

En remplaçant  $x_1 = F_1^{-1}(u)$  et  $x_2 = F_2^{-1}(v)$ , on obtient :

$$C(u, v) = F(F_1^{-1}(u), F_2^{-1}(v))$$

Pour passer de la fonction de répartition d'une copule  $C(u, v) = F(F_1^{-1}(u), F_2^{-1}(v))$  à la densité de copule  $c(u, v)$ , on dérive l'expression par rapport à  $u$  et  $v$ .

$$c(u, v) = \frac{\partial^2 C(u, v)}{\partial u \partial v} = \frac{f(F_1^{-1}(u), F_2^{-1}(v))}{f_1(F_1^{-1}(u)) f_2(F_2^{-1}(v))}$$

On peut donc construire l'estimateur :

$$\hat{c}(u, v) = \frac{\hat{f}(\hat{F}_1^{-1}(u), \hat{F}_2^{-1}(v))}{\hat{f}_1(\hat{F}_1^{-1}(u)) \cdot \hat{f}_2(\hat{F}_2^{-1}(v))}$$

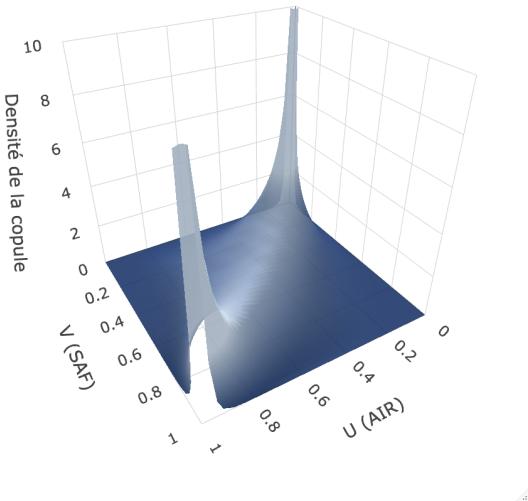


FIG. 3.6 : Densité empirique copule

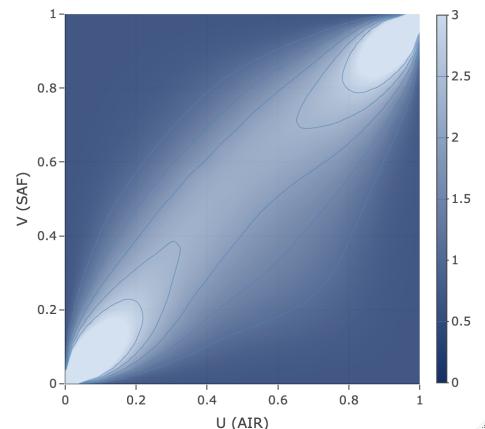


FIG. 3.7 : Contour Plot densité

**Remarque :** Premièrement, étant donné que l'estimateur empirique implique beaucoup d'opérations manuelles (inversion de fonction, etc.), les résultats informatiques manquent de précision. La densité a donc été estimée à l'aide de la librairie **kdecopula**.

Au-delà de ça, la forme de notre densité semble assez cohérente avec les résultats précédents. Cependant, il reste difficile d'avoir une intuition précise sur la copule théorique qui se rapproche le mieux de notre copule empirique. Pour cette raison, nous allons effectuer des tests.

# Chapitre 4

## Ajustement copule théorique

Ajuster nos données à une copule connue est très important dans mon projet. Premièrement, il est difficile de générer des observations à partir de ma copule empirique, car elle est discrète et non continue, elle est uniquement générée sur la marche  $\{0, \frac{1}{K}, \dots, \frac{K-1}{K}\}$ . ( $K = 100$  dans mon code)

L'utilisation de copules connues permet de faire une généralisation et grandement faciliter les différentes simulations.

Deuxièmement, connaître la structure à laquelle se rattachent mes données apporte beaucoup d'interprétabilité à mon raisonnement. (grâce aux propriétés des copules étudiées en cours)

### 4.1 Selection copule

La librairie **gofCopule** (Goodness-of-fit tests), basée sur le processus empirique, compare la copule empirique avec une estimation paramétrique de la copule dérivée sous l'hypothèse nulle.

La statistique de test est le Cramer-von Mises functional  $S_n$  défini par : (C.f. Genest, Remillard et Beaudoin (2009), p 201) (se rapproche beaucoup d'une norme  $L^2$ )

$$S_n = \int_{[0,1]^d} (C_n(\mathbf{u}) - C_{\theta_n}(\mathbf{u}))^2 dC_n(\mathbf{u})$$

et sert à tester l'hypothèse :  $H_0 : C = C_{\theta_n}$

Des p-values pour l'hypothèse de test  $H_0$  peuvent être obtenues soit en utilisant le bootstrap paramétrique, soit à l'aide d'une approche rapide par multiplicateur.

	Copule	Statistic	P_value
statistique 1	gaussian	0.0297334	0.0394605
statistique 2	t_copula	0.0247978	0.0914086
statistique 3	clayton	0.1101302	0.0004995
statistique 4	gumbel	0.0380906	0.0034965
statistique 5	frank	0.0670963	0.0004995

TAB. 4.1 : Résultats statistiques

**Interprétation :** On retient uniquement la copule de student car c'est la seule qui rejette pas l'hypothèse  $H_0$ , ce qui est également cohérent avec la valeur de la statistique car c'est la plus faible.

Ce résultat est peu surprenant : la copule de Student est bien adaptée aux dépendances des valeurs extrêmes. Elle possède également un paramètre contrôlant les queues de distribution, ce qui permet plus de flexibilité dans la modélisation de la dépendance.

## 4.2 Ajustement copule de Student

Rappelons que la fonction de densité de la copule de Student  $d$ -variée, s'écrit pour tout  $(u_1, \dots, u_d) \in [0, 1]^d$  :

$$c(u_1, \dots, u_d) = \frac{f_{\nu, \Sigma}(t_\nu^{-1}(u_1), \dots, t_\nu^{-1}(u_d))}{\prod_{i=1}^d f_\nu(t_\nu^{-1}(u_i))}$$

La fonction de distribution  $t_\nu^{-1}$  est l'inverse de la distribution de Student centrée réduite univariée à  $\nu$  degrés de liberté. La fonction  $f_{\nu, \Sigma}$  est la densité de probabilité de la loi de Student centrée réduite,  $\Sigma$  sa matrice de corrélation, et  $f_\nu$  est la densité univariée de la loi de Student centrée réduite ( $\Sigma = 1$ ).

Pour réaliser ce calibrage, j'utilise une approche paramétrique par maximum de vraisemblance. (Où  $\Sigma = \rho$  car dimension 2)

Paramètre	Valeur estimée
$\rho$	0.8249422
$\nu$	2.650739

TAB. 4.2 : Paramètres estimés de la copule t de Student

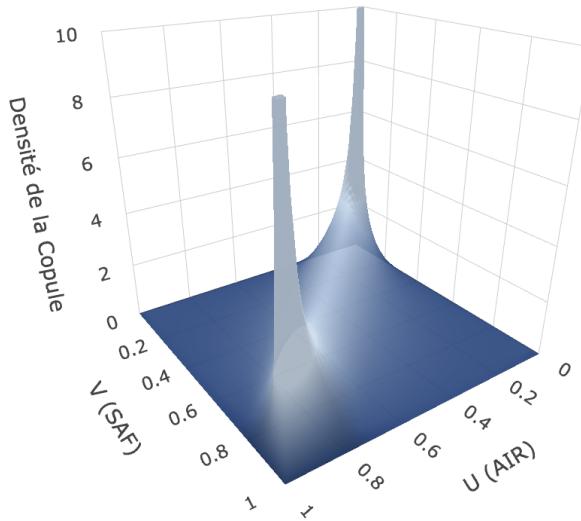


FIG. 4.1 : Densité copule Student ajustée

### 4.3 Fonction Bicopselect

Pour estimer la copule optimale ainsi que ses paramètres associés, il existe également la fonction **Bicopselec** de la librairie VineCopula.

Cette fonction propose un large choix de famille, également un large choix de méthodes d'estimation.

**Les méthodes d'estimation :**

- La méthode d'estimation classique de maximum de vraisemblance appelée Loglik.
- Étant donné que chaque copule théorique ne possède pas le même nombre de paramètres d'ajustement, on peut également intégrer le nombre de paramètres d'ajustement avec le critère **AIC** (**Critère d'information d'Akaike**).
- Étant donné que la taille de mon échantillon est relativement faible, c'est un critère que l'on peut également prendre en compte dans notre estimation avec le critère **BIC** (**Critère d'information bayésien**).

Critère	Copule_Info
logLik	BB7(par = 2.46, par2 = 2.15, tau = 0.62)
AIC	BB7(par = 2.46, par2 = 2.15, tau = 0.62)
BIC	BB7(par = 2.46, par2 = 2.15, tau = 0.62)

TAB. 4.3 : Résultats fonction Bicopselect

**Interprétation :** J'espérai que la fonction Bicopselect confirme mon premier résultat en me renvoyant une copule de Student, cependant la copule trouvée semble être une copule appelée **BB7**. (Copule de Joe Clayton) La copule BB7 est une copule de la famille archimédienne, elle semble être utilisée lorsque la relation entre les variables présentent une asymétrie et une dépendance aux extrêmes .

Etant donné qu'il est difficile d'obtenir plus d'informations sur la copule BB7, je décide de retenir la copule de student pour la suite du projet.

# Chapitre 5

## Application rainbow option (option arc en ciel) : Pricing option rainbow par Monte-Carlo

Une option rainbow est un instrument financier exposé à au moins deux sources d'incertitude provenant de différents actifs sous-jacents, et sa valeur dépend également du prix de ces actifs risqués.

### 5.1 Présentation de l'option étudiée

Symbole	Signification
$S_{it}$	Prix au comptant de l'action $i$ , $i = 1, 2, \dots, n$
$X_t$	Prix au comptant de l'obligation
$K_i$	Prix d'exercice de l'option rainbow / strike $i$ , $i = 1, 2, \dots, n$
$\tau$	Date d'expiration de l'option
$r$	Taux d'intérêt sans risque
$F_t$	Filtration associée à l'ensemble des actifs du marché

TAB. 5.1 : Table notations

Etant donné que l'on se concentre sur les valeurs extrêmes négatives, je décide de choisir l'étude d'un put min rainbow, défini par le payoff suivant :

$$\text{Payoff} = \min_{i \in [1, n] \cap \mathbb{N}} [(K_1 - S_{1\tau})_+; \dots; (K_n - S_{n\tau})_+]$$

Pour que cette option se ramène à mon exemple, je choisis de poser  $n = 2$ . De plus, étant donné que j'ai choisi des actifs ayant des cours relativement proches, je peux également faire la simplification  $K_1 = K_2 = K$ .

On obtient alors :  $\text{Payoff} = \min[(K - S_{1\tau})_+; (K - S_{2\tau})_+] = [K - \max(S_{1\tau}, S_{2\tau})]_+$

On peut donc exprimer le prix de l'option à  $t$  comme :

$$X_t = e^{(-r(\tau-t))} \mathbb{E}[K - \max(S_{1\tau}, S_{2\tau})]_+ | F_t$$

On va donc générer les trajectoires aléatoire du processus  $Y_t = [K - \max(S_{1t}, S_{2t})]_+$  de façon traditionnelle (loi normale multivariée) ainsi qu'avec le copule construit dans la partie précédente afin de comparer les résultats.

On pose également  $K = 100$ .

## 5.2 Approche classique : loi normale multivariée

Etant donné que ce n'est pas l'objet du sujet je ne rentrerai pas dans les détails mathématiques de cette partie. L'intégration de la corrélation de pearson se fait par décomposition de Cholesky.

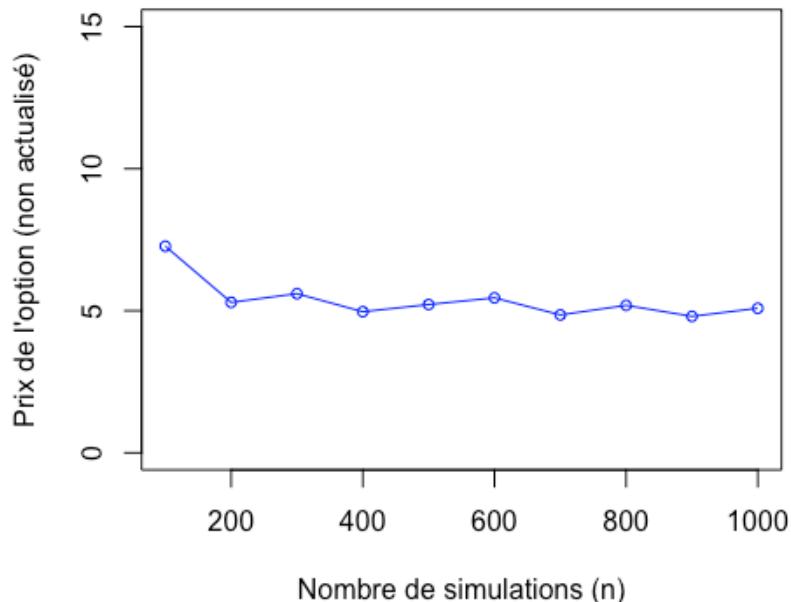


FIG. 5.1 : Convergence du prix par simulation loi normale

### 5.3 Simulation avec copule de Student

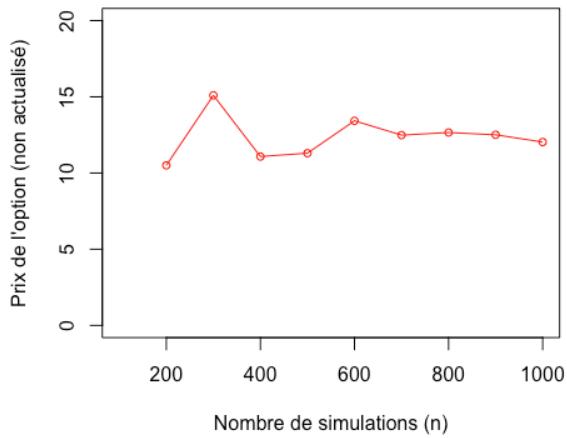


FIG. 5.2 : Convergence du prix par simulation copule de Student

### 5.4 Comparaison des méthodes

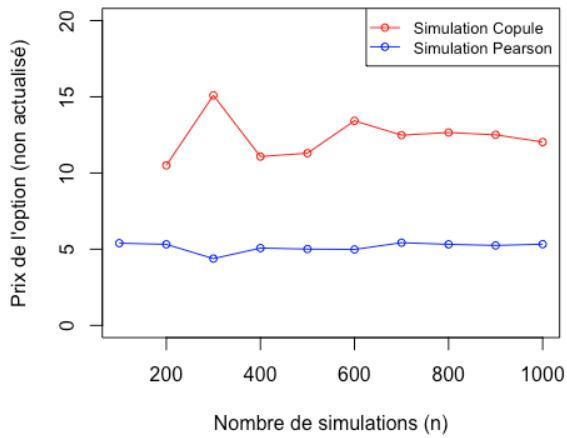


FIG. 5.3 : Convergence-plot loi normale / Copule Student

# Conclusion

Ce dernier résultat est très intéressant et il représente parfaitement l'intérêt des copules. J'ai choisi de prendre un produit financier relativement sensible aux valeurs extrêmes négatives (car put) et corrélé (car min) avec une observation prise au cours d'une crise boursière. Les deux actifs en question, Airbus et Safran, sont particulièrement pertinents dans ce contexte, car leur interdépendance, amplifiée lors des périodes de turbulence économique, n'est pas capturée de manière adéquate par les méthodes traditionnelles de corrélation, telles que la corrélation de Pearson.

En effet, lors d'une crise boursière, les mouvements de marché tendent à être plus synchronisés, exacerbant les baisses conjointes des actions fortement corrélées, surtout dans des secteurs similaires. En utilisant les copules, nous sommes capables de modéliser cette dépendance structurelle, qui est non linéaire et accentuée dans les queues de distribution. Les copules permettent donc de capturer ces événements extrêmes de manière plus précise, modélisant non seulement la corrélation en moyenne, mais aussi les comportements synchronisés dans les situations extrêmes.

En somme, lorsque l'on modélise correctement ces corrélations à l'aide de copules, qui ne sont pas perçues par la corrélation de Pearson, on observe des crashes corrélés beaucoup plus fréquemment au cours des simulations, ce qui se traduit par une valorisation beaucoup plus réaliste et souvent bien supérieure du produit dérivé par rapport à une estimation basée sur une méthode classique. Cette approche offre ainsi une meilleure gestion des risques et une évaluation plus précise des produits financiers sensibles aux conditions de marché extrêmes, renforçant l'importance de la modélisation avancée dans la finance moderne.