

– Étude de cas –

**Estimation d'un modèle d'évaluation des actifs, du bêta des
actions et évaluation des prévisions des rendements**

Préambule.....	1
Base de données.....	1
Étude de nos actions.....	2
Liens entre nos actions.....	5
Recherche de modèles performant contre la crise du COVID-19.....	9
Prévision des rendements.....	12

Préambule

Une entreprise de conseil en gestion nous a sollicités pour les aider dans une étude visant à prédire le comportement de trois actions à l'aide d'un modèle prédictif. Pour cela, nous avons importé les données boursières de ces trois actions françaises depuis Yahoo Finance, en utilisant leurs '**tickers**' (identifiants boursiers) et en les intégrant dans nos outils d'analyse grâce au package '**yfinance**'. Les données couvrent la période de janvier 2019 à aujourd'hui.

Nous avons choisi d'analyser les actions du Crédit Agricole, de Carrefour et de Total Énergie pour étudier l'évolution d'entreprises dans des secteurs variés. Étant donné que notre période d'étude inclut la crise du COVID, nous pourrions extraire des informations précieuses sur la manière dont les cours des actions fluctuent en fonction du secteur d'activité.

Base de données

Pour optimiser notre étude, nous avons d'abord effectué une exploration préliminaire de la base de données afin de corriger d'éventuelles erreurs. En effet, il a été nécessaire de réaffecter les colonnes aux données qu'elles représentent, car un décalage survenu lors de l'importation a rendu la base de données difficilement interprétable sans modification. Les variables ont été converties au format numérique pour garantir leur interprétabilité et une représentation cohérente.

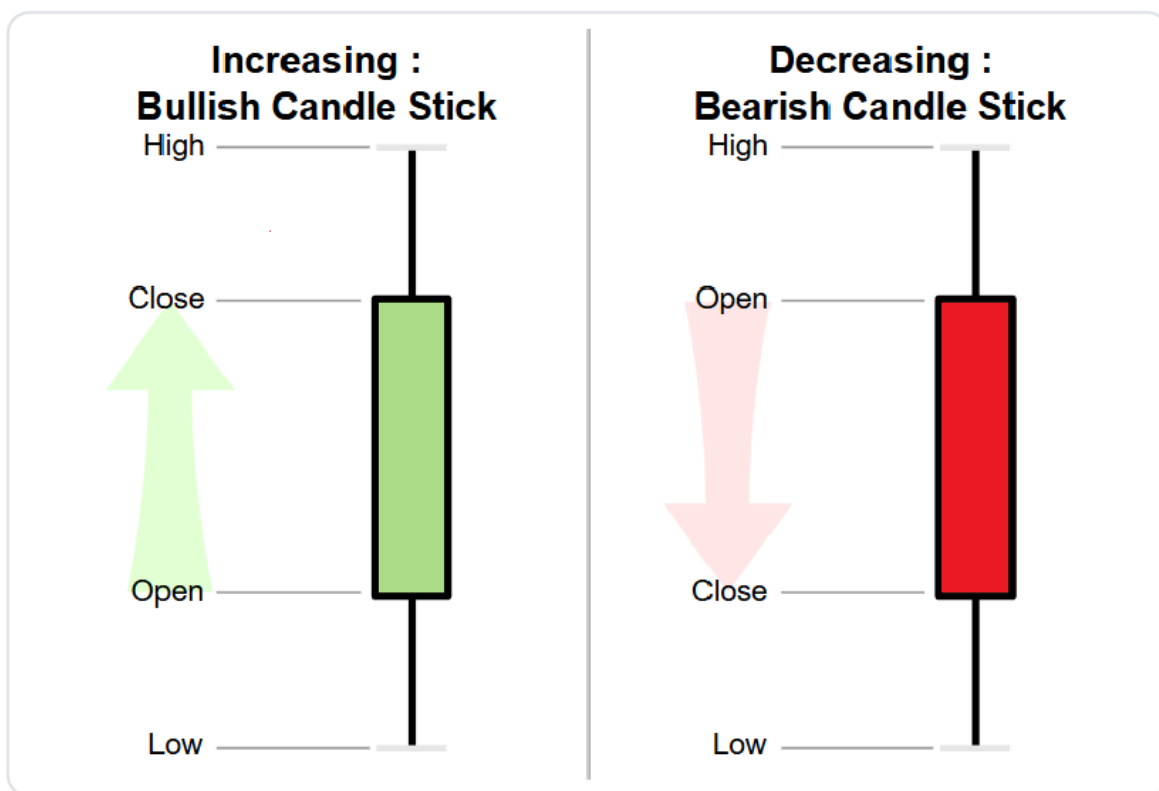
Hormis ces modifications, les données étaient très propres et sans défauts.

Étude de nos actions

Pour mieux visualiser l'évolution de nos actions et comprendre le fonctionnement du marché boursier, nous avons utilisé des graphiques en chandeliers japonais pour représenter nos trois actions. Ce type de graphique permet de visualiser les variations de prix d'une action tout au long d'une journée (lorsqu'il est configuré en intervalle quotidien). Dans ces graphiques, un chandelier est représenté en **vert** pour indiquer une tendance haussière (lorsque le prix de clôture est supérieur au prix d'ouverture) et en **rouge** pour une tendance baissière (lorsque le prix de clôture est inférieur au prix d'ouverture).

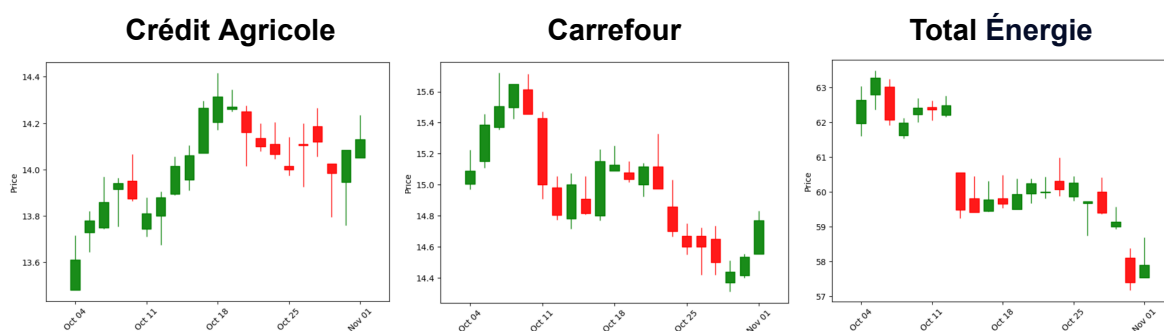
Les variations de prix pendant la période considérée sont représentées par les deux extrémités de la mèche (la ligne verticale). Les extrémités du corps du chandelier correspondent au prix d'ouverture, lorsque le marché commence à échanger, et au prix de clôture, lorsque le marché termine ses échanges.

Le schéma suivant résume ces éléments :



Source : <https://www.investopedia.com>

Graphiquement, l'évolution de nos actions sur les 4 dernières semaines (à partir du 1er Novembre) :



Le graphique permet d'observer les tendances des cours de chaque action, mais sur le long terme, ces investissements sont-ils réellement rentables ? Quels rendements génèrent-ils ?

Pour les calculer simplement, nous utiliserons le prix de clôture du marché le lendemain, divisé par le prix de clôture d'aujourd'hui. Cela nous donnera le rendement de l'action pour le jour suivant par rapport à aujourd'hui, et ainsi de suite...

Nous allons exprimer ces rendements en logarithmes pour en faciliter l'interprétation et pour plus de précision dans nos calculs. Voici les résultats obtenus :

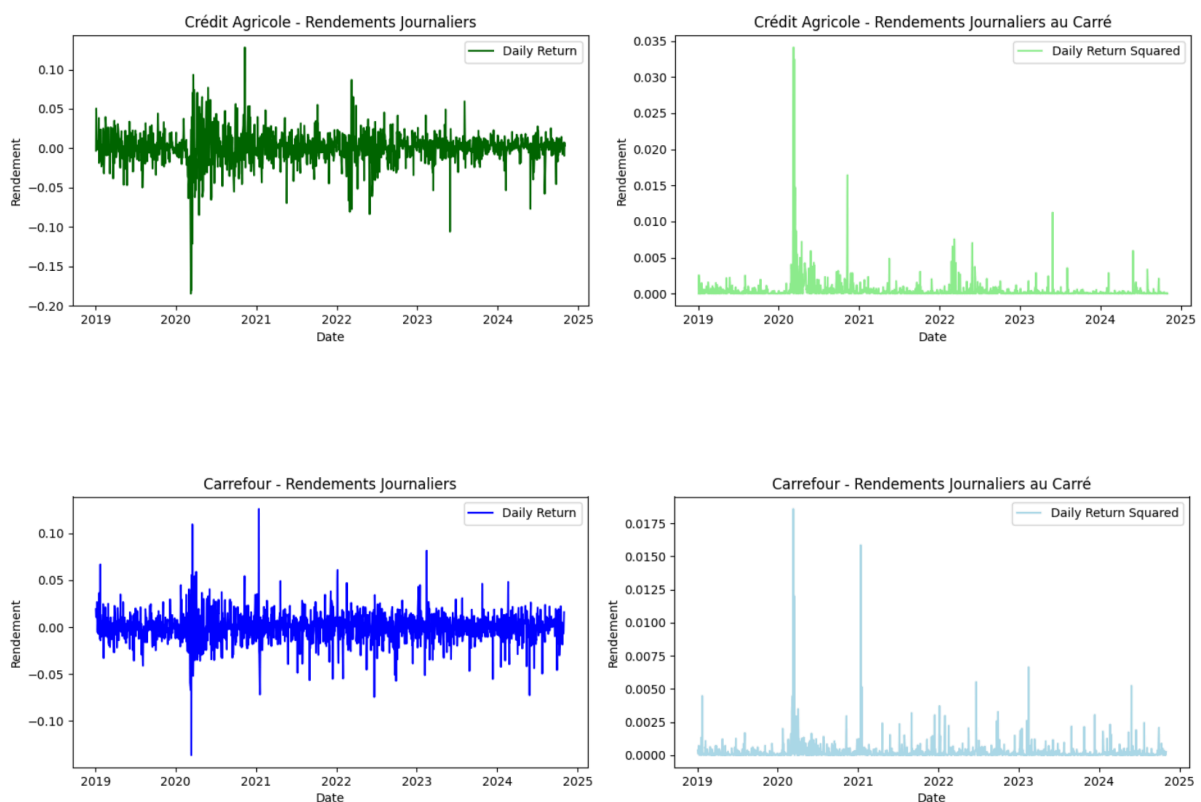
Crédit Agricole - Moyenne des rendements journaliers: 0.0002782970365473789
Crédit Agricole - Volatilité des rendements journaliers: 0.020513992925653018

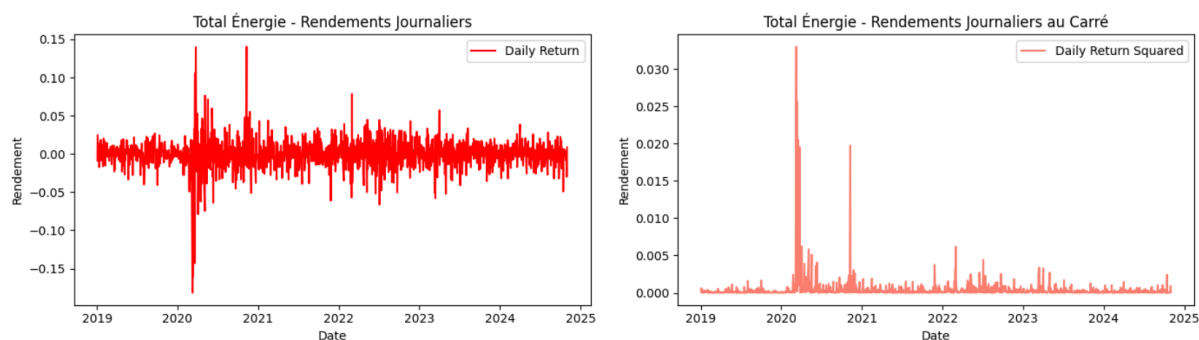
Carrefour - Moyenne des rendements journaliers: 4.080162590907994e-06
Carrefour - Volatilité des rendements journaliers: 0.01651131756997421

Total Énergie - Moyenne des rendements journaliers: 0.0001441504235054229
Total Énergie - Volatilité des rendements journaliers: 0.019183638062128104

Le Crédit Agricole et Total Énergie montrent une légère tendance haussière, tandis que Carrefour reste essentiellement stable. Cependant, la croissance quotidienne moyenne reste faible pour les trois. Cependant, les actions du Crédit Agricole et de Total Énergie présentent une volatilité similaire, plus élevée que celle de Carrefour, ce qui traduit une plus grande incertitude quant à leurs prix quotidiens.

Nous pouvons, en quelque sorte, observer la moyenne des rendements à partir de la série des rendements, que l'on peut représenter sous forme de graphique. Il en va de même pour la volatilité :





La série de rendements journaliers permet de visualiser la **direction et l'ampleur** des variations quotidiennes du prix de l'action. Cela aide à observer les périodes de hausse ou de baisse et peut montrer la volatilité de manière directe. C'est à partir d'ici qu'on peut établir une moyenne.

Le rendement au carré mesure l'amplitude des variations sans tenir compte de leur direction (positif ou négatif). En effet, en prenant le carré, les variations positives et négatives sont toutes deux représentées de manière positive, ce qui met en évidence la **volatilité** en amplifiant les périodes de volatilité.

Pour nous donner une idée, si on calculait les rendements cumulatifs, calculés en multipliant les rendements journaliers pour obtenir le rendement sur toute la période. Cela donnerait une idée de la performance totale d'un investissement initial jusqu'à aujourd'hui. Ici, il est de **10.09%** sur toute la période étudiée pour le Crédit Agricole, **-18%** pour Carrefour et **-6,1%** pour Total Energie.

Comme autre indicateur, on aurait le rendement journalier moyen pour estimer le rendement annualisé en élevant la moyenne journalière à la **puissance 252** (le nombre approximatif de jours de bourse par an en enlevant les week-ends). Ce calcul permet d'évaluer la rentabilité potentielle annuelle si la performance moyenne se poursuit. Ici, il est de **7.26%** par an pour le Crédit Agricole, **0,1%** pour Carrefour et **3,70%** pour Total Energie.

On observe tout de même certaines tendances similaires sur nos graphiques malgré la diversité de nos actions. Nous allons donc tester s'il y a des saisonnalités dans nos données. Une première fois de façon mensuelle, puis trimestrielle. Pour cela, nous avons décomposé la série de rendements journaliers avec une fréquence de **21 jours** pour le mois et **63 jours** pour le trimestre (approximant les jours de bourse en enlevant les week-ends) et appliqué le **test de Ljung-Box** qui nous permet de tester cela statistiquement.

Crédit Agricole et **Total Énergie** présentent une saisonnalité claire, aussi bien mensuelle que trimestrielle, avec des p-valeurs très faibles aux deux échelles de temps. Alors ici, seul **Carrefour** ne présente pas de saisonnalité, suggérant que ses rendements ne suivent pas de cycle spécifique identifiable dans cette analyse.

Alors, elle est stationnaire si elle ne suit pas de cycle ?

Comme visible, sur nos graphiques précédents et nos analyses sur la variance et la moyenne de nos séries, nos conclusions seront cohérentes. Pour cela, nous avons utilisé le **test de Dickey-Fuller (ADF)** qui est un test statistique utilisé pour déterminer si une série

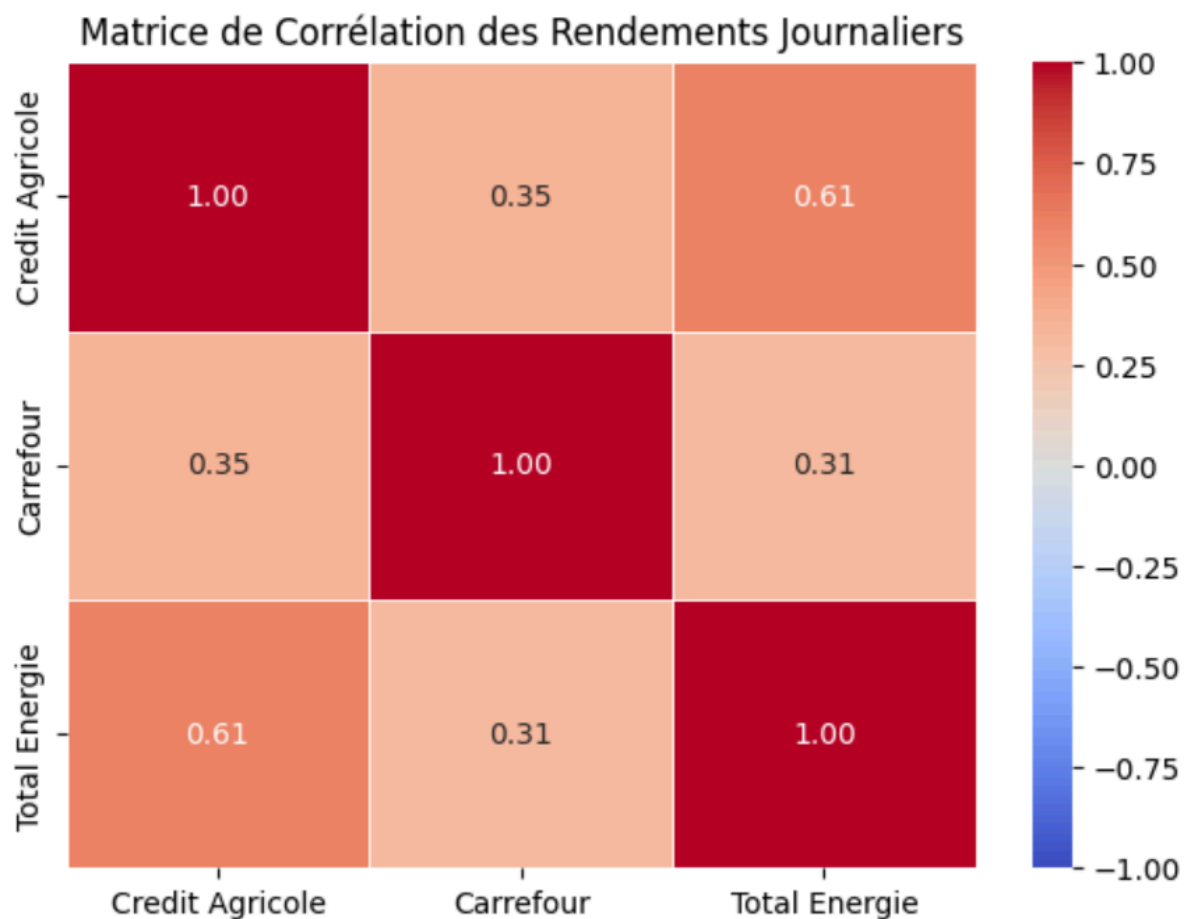
temporelle est stationnaire, c'est-à-dire si ses propriétés statistiques (comme la moyenne et la variance) sont constantes dans le temps. On prendra un **seuil de 1%**, qui signifie que l'on accepte une marge d'erreur de 1 % dans notre décision. Nos résultats traduisent que les prix de clôture des trois actions sont **non stationnaires** au seuil de 1%. Cela signifie qu'ils contiennent probablement une tendance et ne fluctuent pas de manière stable autour d'une moyenne. Cependant, les rendements journaliers des trois actions sont **stationnaires** au seuil de 1% ce qui indique qu'ils fluctuent autour d'une moyenne constante.

Liens entre nos actions

Comme explicité précédemment, nos actions proviennent de différents secteurs, il serait intéressant de calculer la matrice des corrélations des rentabilités.

Pour calculer la **matrice de corrélation des rentabilités** (ou rendements journaliers), nous allons mesurer le lien entre les rendements des différentes actions. La corrélation quantifie l'intensité et la direction de la relation linéaire entre les rendements de chaque paire d'actions. Une corrélation de 1 indique une relation parfaitement positive (les deux actions évoluent de manière identique), alors qu'une corrélation de -1 signifie une relation parfaitement négative (quand une action monte, l'autre descend). Une corrélation proche de 0 indique une relation faible ou inexistante.

Les corrélations entre actifs sont essentielles pour un investisseur, car elles facilitent l'accès à la diversification, optimisent le portefeuille et réduisent le risque global. En combinant des actifs faiblement ou négativement corrélés, l'investisseur peut compenser les pertes sur certains actifs par les gains sur d'autres, minimisant ainsi le risque du portefeuille sans sacrifier le rendement attendu. Par exemple, si les rendements de Crédit Agricole et Carrefour sont peu ou négativement corrélés, cela réduit le risque du portefeuille par rapport à des actifs fortement corrélés. En effet, la corrélation influence la volatilité du portefeuille global : en intégrant des actifs faiblement corrélés, il est possible de limiter la volatilité globale, ce qui est stratégique sur des marchés instables. De plus, les corrélations aident les investisseurs à repérer des relations sectorielles importantes : si **Crédit Agricole (banque)** et **Total Énergie (services)** montrent une forte corrélation, cela peut indiquer une dépendance commune à certaines conditions économiques, ce qui oriente les décisions d'investissement. En somme, des corrélations élevées signalent une interdépendance accrue entre actifs, accroissant le risque en cas de tendance baissière. À l'inverse, des corrélations faibles ou négatives diversifient et protègent mieux le portefeuille contre les fluctuations du marché.



La forte corrélation entre **Crédit Agricole** et **Total Énergie** suggère que leurs performances sont susceptibles d'être similaires, réduisant ainsi les avantages de diversification, tandis que la faible corrélation de **Carrefour** avec les autres actions offre un meilleur potentiel de diversification bien que positivement corrélée (d'autres facteurs peuvent être pris en compte).

L'objectif est d'estimer l'impact des actions par rapport au marché boursier français (**CAC40**), c'est-à-dire d'évaluer la volatilité des actions par rapport au marché. Pour ce faire, nous avons utilisé une régression linéaire simple pour estimer les paramètres du modèle. La **régression linéaire** est une méthode statistique utilisée pour modéliser la relation entre une variable dépendante et une ou plusieurs variables indépendantes. Ici, expliquer le rendement de nos actions (R_t^i) par rapport au rendement du marché à la même période (R_t^M) soit :

$$R_t^i = \alpha + \beta R_t^M + \varepsilon_t$$

Les résultats de notre analyse nous suggèrent que le **bêta** (l'impact qu'à le rendement de marché sur le rendement de nos actions) de **Crédit Agricole** est de **1.27**, ce qui indique que l'action de **Crédit Agricole** est plus volatile que le marché. En d'autres termes, pour chaque 1% de variation du marché, le rendement de **Crédit Agricole** varie de **1,27%**. Le **beta**

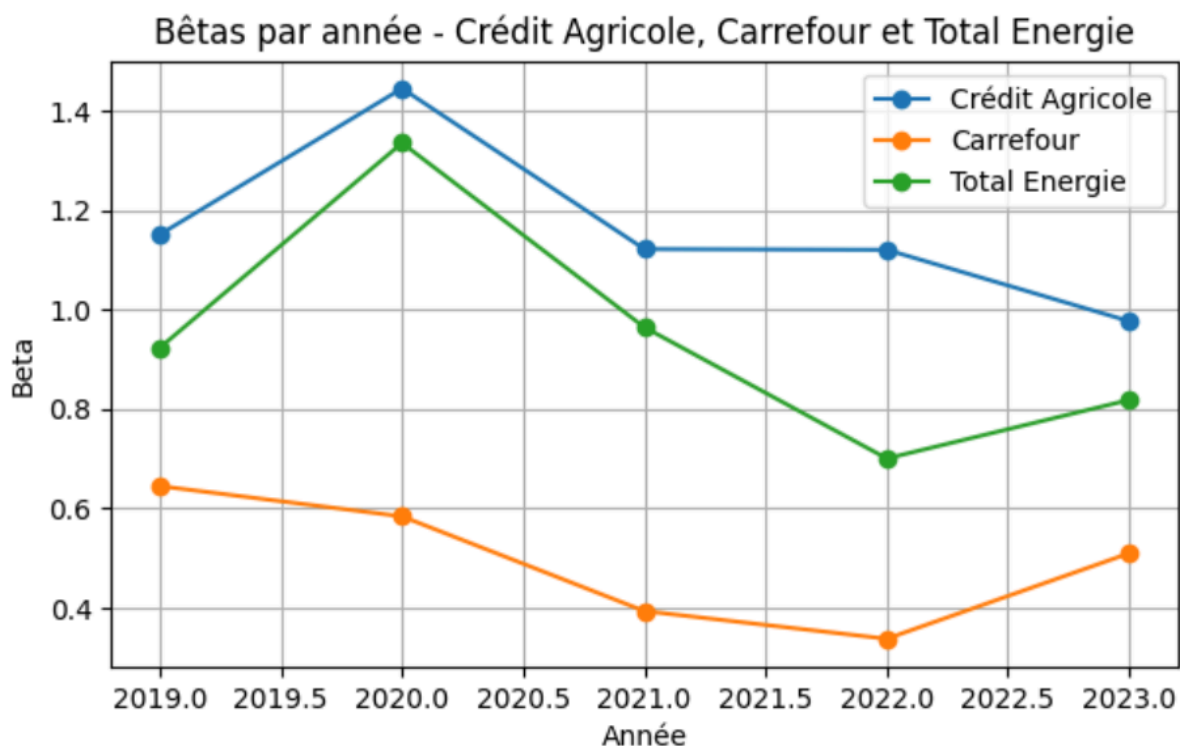
de Carrefour est plus faible, elle serait moins sensible aux variations du marché (une variation de 1% du marché entraîne une variation de **0.505%** dans le rendement de Carrefour). Pour Total Énergie, il est égal à **1.08**, ce qui signifie que l'action de Total Énergie suit de manière quasi parfaite les variations du marché.

Les alphas étant très proches de zéro, on interprète qu'il n'y a pas de surperformance ou sous-performance significative en dehors des fluctuations du marché.

Cependant, les années ne se ressemblent pas et l'impact peut être mal traduit à cause de la longévité de la période. Nous allons donc effectuer le même travail mais par année, pour obtenir des **bêtas** qui varient et s'adaptent aux périodes en cours. Cela permet de mieux comprendre comment chaque action réagit à l'environnement économique au fil du temps et comment cette sensibilité évolue.

Year	Beta_Credit_Agricole	Beta_Carrefour	Beta_Total_Energie
2019	1.150491	0.645002	0.922797
2020	1.444155	0.584086	1.334506
2021	1.121268	0.393788	0.964103
2022	1.119846	0.338114	0.700501
2023	0.976419	0.510019	0.818054

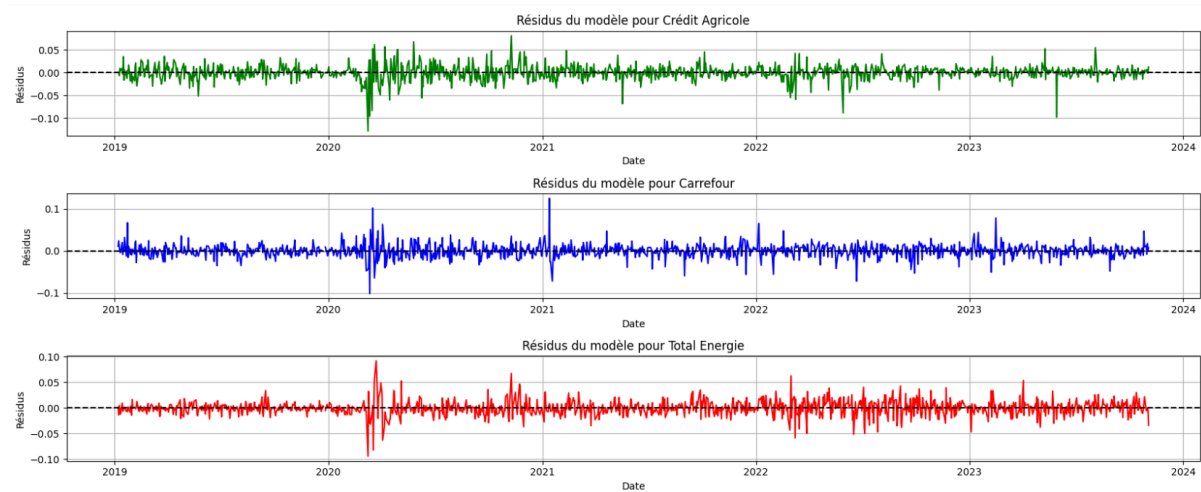
Soit graphiquement pour une meilleure visualisation :



Grâce à ces résultats, nous pouvons calculer les rendements de chaque action à un moment donné grâce aux prédictions. Mais ça ne suffit pas, afin de garantir la performance de notre modèle, nous devons observer les résidus de notre modèle. Les résidus sont la

différence entre les rendements réels des actions et les rendements prédits par le modèle de régression.

Nous obtenons cette représentation des résidus :



Nous observons une variance des résidus, ce qui suggère qu'il y a une **présence d'hétéroscédasticité**. On parle d'hétéroscédasticité lorsque la variance des erreurs stochastiques (résidus) de la régression est la même pour chaque observation. Pour s'en assurer, nous avons effectué des **tests statistiques de White pour tester l'hétéroscédasticité**. Ce test consiste à prendre les **carrés des résidus** du modèle initial comme variable dépendante et à les régresser sur les variables explicatives (régression auxiliaire).

Comme prévu, **les tests de White** montrent que l'hétéroscédasticité est présente pour les trois actions ce qui signifie que cela nous a conduit à des erreurs dans le calcul des **écarts-types** des coefficients estimés, donc de nos prévisions.

Pour corriger cela, on va utiliser un modèle **GARCH(1,1)** classique (Generalized Autoregressive Conditional Heteroskedasticity), qui est utilisé spécifiquement dans les séries temporelles pour modéliser la **volatilité conditionnelle (volatilité d'une série temporelle à un moment donné, en fonction des valeurs passées des rendements et de la volatilité elle-même)**, permettant ainsi de tenir compte de l'hétéroscédasticité qui change dans le temps.

Cette estimation nous a permis d'obtenir plusieurs paramètres, dont **la moyenne des rendements** dans le modèle (μ qui est environ égale à 0), **la volatilité de base** (ω), **la variance résiduelle** (non expliquée par les chocs passés qui est α), ainsi que **l'impact des chocs passés sur la volatilité actuelle** (β).

Action	omega	alpha_1	beta_1	AIC
Crédit Agricole	0.000004	0.050000	0.930000	-7131.726246
Carrefour	0.000024	0.099986	0.800027	-6886.679001
Total Energie	0.000004	0.050000	0.930000	-7182.122308

Crédit Agricole

Carrefour

Total Energie

Écart-types des coefficients:		Écart-types des coefficients:		Écart-types des coefficients:	
mu	4.214636e-04	mu	4.166764e-04	mu	1.464005e-05
omega	1.029703e-12	omega	6.504420e-07	omega	8.566207e-12
alpha[1]	5.563869e-03	alpha[1]	3.835620e-02	alpha[1]	2.172026e-02
beta[1]	3.840550e-04	beta[1]	3.052701e-02	beta[1]	2.159261e-02

Tout étant significatif ici.

Concernant l'**AIC**, il s'agit d'une mesure utilisée pour évaluer la qualité d'un modèle statistique en tenant compte à la fois de l'ajustement du modèle aux données et de la complexité du modèle. Il est souvent utilisé pour comparer plusieurs modèles et, plus il est petit, meilleur est le modèle.

Les résultats des paramètres des modèles GARCH pour chaque action montrent que la volatilité des rendements est persistante et sensible aux événements passés avec des paramètres significatifs.

Toutefois, il est important de noter que la volatilité estimée peut avoir été influencée par certains facteurs externes. En 2020, la crise du COVID-19 a bouleversé les marchés financiers, entraînant une volatilité accrue et de fortes fluctuations des prix des actifs.

Recherche de modèles performant contre la crise du COVID-19

Pour contrôler l'impact de cette crise dans notre modèle et éviter qu'elle ne fausse nos résultats sur des périodes normales, il serait pertinent d'introduire des variables spécifiques dans le modèle afin de capturer cet effet.

On pourrait introduire une **variable binaire (dummy)** qui prend la valeur 1 pendant la période de la crise du Covid et 0 sinon et va interagir avec la variable rendement des marchés. Le coefficient de cette variable d'interaction capturera la relation entre les rendements de l'action et ceux du marché en fonction de la crise.

La **dummy Covid seule** servira à capturer l'**effet global de la crise** en tant qu'événement exogène sur les rendements des actions, sans prendre en compte la relation avec le rendement du marché.

Cela donnerait un modèle comme suit :

$$R_t^i = \alpha + \beta R_t^M + \Psi \cdot EffetCovid_t + \gamma (EffetCovid_t * R_t^M) + \varepsilon_t$$

Nous allons désormais tester la robustesse de notre modèle face à la crise du COVID en intégrant **un modèle AR(1)** dans notre approche sur notre nouveau modèle.

Pour cela, nous devons introduire un terme "autorégressif" dans notre modèle. Le coefficient de **l'AR(1)** mesurera la relation entre le rendement actuel et celui de la période précédente ($\xi \cdot R_{t-1}^i$).

Cela donne :

$$R_t^i = \alpha + \beta R_t^M + \Psi \cdot EffetCovid_t + \gamma \cdot (EffetCovid_t * R_t^M) + \xi \cdot R_{t-1}^i + \varepsilon_t$$

On trouve les estimations suivantes :

Action	Intercept (Const)	Market_Return	EffetCovid	Interaction	AR1
Crédit Agricole	-0.000259	1.098540 ·	0.000095 ·	0.361791	0.037077
Carrefour	0.000069	0.420192 ·	-0.000610 ·	0.178461	0.000469
Total Energie	-0.000003	0.812045 ·	-0.000403 ·	0.530736	0.034754
AIC					
-7101.985502					
-6793.007271					
-7023.873653					

Avec “·” traduisant la significativité du coefficient.

Les coefficients associés à la variable **Market_Return** et à l'interaction avec la crise du COVID sont significatifs, tandis que les autres paramètres ne le sont pas. Cela montre que le marché a un impact significatif sur les rendements de chaque action. En revanche, l'effet direct de la crise du COVID n'est pas significatif, bien que son interaction avec le marché indique que les actions deviennent plus sensibles aux fluctuations du marché en période de crise. L'effet de l'interaction étant significatif et positif pour toutes les entreprises, il indique que l'impact du marché est amplifié durant la période COVID. Total Énergie présente la sensibilité la plus marquée, reflétant sa forte exposition aux chocs systémiques mondiaux, tandis que Crédit Agricole montre une dépendance plus légère mais tout de même importante. Carrefour en revanche, est le moins affecté, ce qui peut paraître questionnable car les enseignes alimentaires comme Carrefour ont été des bénéficiaires indirects de la pandémie grâce à leur statut de secteur essentiel (la performance peut dépendre de beaucoup d'autres caractéristiques comme la région par exemple).

Il pourrait être pertinent de tester un modèle plus performant, comme **un modèle ARMA(p,q)**, pour capturer les dépendances temporelles résiduelles qui ne sont pas expliquées par les variables du modèle de régression. Nous avons sélectionné les paramètres optimaux en analysant **l'ACF et la PACF** de nos modèles. **L'ACF** mesure la corrélation entre les valeurs d'une série temporelle à différents décalages temporels et la **PACF** mesure la corrélation entre deux valeurs de la série temporelle après avoir éliminé l'influence des autres valeurs situées entre elles. Les résultats sont les suivants :

- **Crédit Agricole** : Absence de structure de dépendance significative dans les résidus (**ARMA(4,4)**) avec un **AIC de -6921**.

- **Carrefour** : **Modèle ARMA(5,5)** pour les résidus, suggérant une dépendance significative, avec un **AIC de -6805,99**.
- **Total Énergie** : **Modèle ARMA(4,5)** pour les résidus, indiquant une faible dépendance avec une composante de moyenne mobile, et un **AIC de -7119**.

Pour déterminer de manière plus précise si l'hétéroscédasticité conditionnelle autorégressive doit être prise en compte (en raison de l'hétéroscédasticité détectée dans les résidus précédemment), il est nécessaire d'effectuer un **test d'Engle**, spécifiquement conçu pour les modèles GARCH. En effet, ce test suggère la présence de cette hétéroscédasticité, ce qui justifie l'utilisation d'un modèle **GARCH ou ARCH** pour la modélisation.

Le modèle GARCH, qui intègre la dépendance de la variance actuelle non seulement aux erreurs passées, mais aussi à ses propres valeurs passées, est une généralisation du **modèle ARCH**. Ce dernier se base uniquement sur les erreurs passées pour modéliser la variance conditionnelle d'une série temporelle. Il convient donc d'examiner si ces deux modèles, dans leur forme optimale, peuvent être utilisés. Pour ce faire, nous utiliserons la même fonction qu'auparavant afin de déterminer les meilleurs modèles. Nous limiterons l'estimation des paramètres à un maximum de 5, pour des raisons de pertinence et d'efficacité de calcul.

Les résultats obtenus sont les suivants :

Meilleur modèle ARCH pour Crédit Agricole : ARCH(5) avec AIC = -7126.5084

Meilleur modèle GARCH pour Crédit Agricole : GARCH(5,2) avec AIC = -7139.7406

Meilleur modèle ARCH pour Carrefour : ARCH(3) avec AIC = -6883.6983

Meilleur modèle GARCH pour Carrefour : GARCH(4,3) avec AIC = -6894.5013

Meilleur modèle ARCH pour TotalEnergies : ARCH(5) avec AIC = -7179.4225

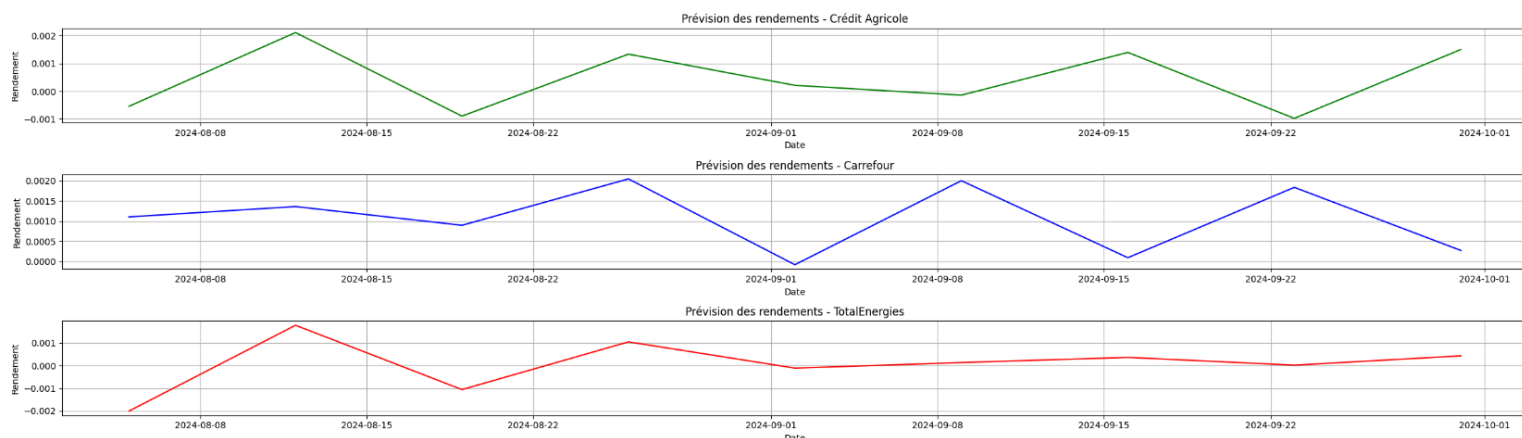
Meilleur modèle GARCH pour TotalEnergies : GARCH(1,3) avec AIC = -7194.3052

En conclusion, nos modèles finaux et optimaux pour les prévisions de rendements futurs seront les modèles GARCH mentionnés ci-dessus.

En conclusion, **les modèles GARCH** mentionnés ci-dessus représentent nos modèles finaux et optimaux. Toutefois, en ce qui concerne la modélisation des rendements, ces modèles ne sont pas les plus performants. **Un ARMA(p,q)** offrirait de meilleures performances pour prédire les rendements, et le modèle GARCH serait plus performant sur la prédiction de la volatilité. **On utilisera donc les modèles ARMA pour prédire nos rendements.**

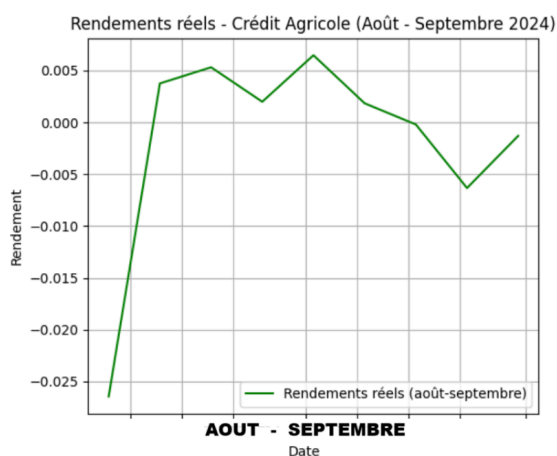
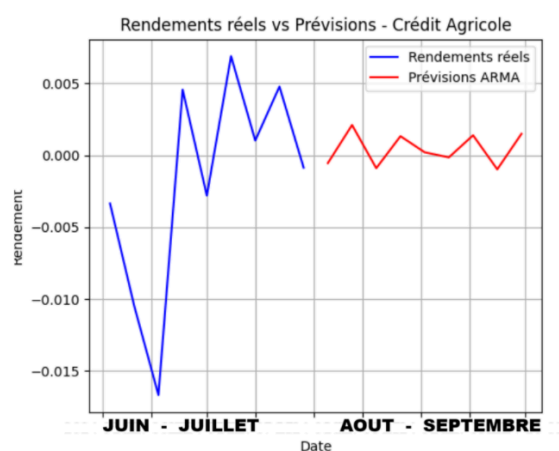
Prévision des rendements

En couplant l'estimation de la volatilité **des modèles GARCH** à l'estimation des rendements **des modèles ARMA (optimaux)**, nous avons pu prédire la valeur des rendements d'une semaine à l'autre pendant 2 mois (début août jusqu'à la fin septembre).



On observe un rendement qui fluctue modérément et autour d'un centre, ce qui est logique, car les séries de rendements sont considérées comme stationnaires évoluant autour d'une moyenne.

Cependant, la performance de notre modèle peut être remise en question de plusieurs manières :



Bien que l'échelle ne soit pas identique, il est possible de comparer les valeurs et la forme de la courbe grâce à l'axe des ordonnées. Les tendances sont bien suivies, et le modèle réussit à prédire la dynamique du marché. Cependant, il a des difficultés à estimer précisément les valeurs des rendements.

Autrement dit, il surestime ou sous-estime parfois les prévisions, mais les résultats restent globalement satisfaisants.

Nous pouvons aussi tester statistiquement la performance de notre modèle avec **la RMSE** qui mesure de l'écart type des erreurs de prévision. Elle mesure la distance entre les valeurs

réelles et les valeurs prédites (ici, les rendements réels contre les rendements prévus par notre modèle).

La **RMSE est de 0,0076** est ce qui suggère que, bien que le modèle soit globalement proche de la tendance des rendements, il y a une certaine variabilité dans les erreurs de prévision. Pour nos 2 autres actions, celle-ci est plus élevée avec des divergences plus importantes comme nous pouvons le voir ci-dessous.

Pour nos 2 autres actions :



La forme des prévisions suit globalement celle des données réelles, même si elle le fait avec un ajustement moins précis que dans les étapes précédentes, et reste sujette à un certain sur-ajustement ou sous-ajustement.

En conclusion, ce projet a permis de construire un guide simplifié de la prévision du cours d'une action (et de ses rendements) , en suivant une approche accessible et progressive. Certains tests et analyses plus complexes ont été omis dans ce rapport pour le rendre plus concis et centré sur l'essentiel.

Pour aller plus loin, il serait possible d'améliorer nos modèles en explorant des méthodes de séries temporelles plus avancées et en optimisant davantage les paramètres (en utilisant un modèle VAR par exemple). L'intégration de variables exogènes comme des indicateurs macroéconomiques ou sectoriels, ou des données de sentiment du marché pourrait enrichir significativement les prévisions et les rendre plus robustes.

Enfin, bien qu'il existe encore de nombreuses façons d'affiner nos estimations, ce rapport offre une base d'introduction à la prévision des cours boursiers et constitue un bon point de départ pour les entreprises qui souhaitent approfondir l'analyse de séries temporelles financières.