

Anticiper les Mouvements du WTI : L'Apport des Modèles à Facteurs Face aux Méthodes Classiques

Mohamed Sinan, Coraline Journaux, Sébastien Georges, Maiwenn Etienne

March 2025

1 Introduction

1.1 Contexte et Importance de l'Étude

Le prix du pétrole joue un rôle central dans l'économie mondiale, influençant aussi bien les marchés financiers que la politique énergétique des nations. Comprendre les mécanismes sous-jacents qui déterminent ses fluctuations est essentiel pour les investisseurs, les décideurs économiques et les entreprises du secteur énergétique.

1.2 Objectif de l'Étude

Cette étude vise à analyser la dynamique du prix du pétrole en testant trois approches de prévision :

- Un modèle ARMA pour capturer les tendances et la dynamique temporelle du WTI.
- Un modèle ARMA-GARCH pour prendre en compte la volatilité des prix du pétrole.
- Un modèle à facteurs intégrant des indicateurs financiers avancés, qui est supposé offrir la meilleure précision dans la prévision.

Nous comparerons ces modèles pour déterminer lequel offre les meilleures performances prédictives, en mettant un accent particulier sur le modèle à facteurs.

1.3 Problématique

Quel est le modèle le plus pertinent pour prévoir les variations hebdomadaires du prix du pétrole WTI ? Si c'est le modèle à facteurs, comment les interactions des indicateurs de marché peuvent-il améliorer la précision des prévisions ?

1.4 Hypothèse de Recherche

Nous faisons l'hypothèse que le modèle à facteurs intégrant l'Open Interest, le Ratio Futures/Spot, les Positions Nettes des Hedge Funds et la Volatilité OVX offre une meilleure capacité prédictive que les modèles ARMA et GARCH.

2 Revue de Littérature

L'objectif de cette section est de passer en revue les principales recherches académiques qui ont étudié la prévision du prix du pétrole en utilisant des modèles économétriques et des indicateurs financiers.

2.1 Indicateurs de Marché et Leur Influence sur le Pétrole

2.1.1 Open Interest WTI Futures (Futures ouverts)

Définition

C'est le nombre total de contrats à terme (contrat CRUDE OIL, LIGHT SWEET - NEW YORK MERCANTILE EXCHANGE) en circulation qui n'ont pas encore été réglés ou clôturés.

Interprétation

Cette variable indique l'activité spéculative sur le marché des futures. Un Open Interest en hausse signifie qu'il y a de nouvelles entrées sur le marché, soit pour spéculer, soit pour se couvrir. Un Open Interest en baisse signifie que des positions sont clôturées, ce qui peut réduire la volatilité et annoncer une phase d'incertitude ou de consolidation.

2.1.2 CBOE Crude Oil Volatility Index (OVX)

Définition

L'OVX, ou CBOE Crude Oil Volatility Index, est un indicateur qui mesure la volatilité implicite attendue des prix du pétrole brut sur une période de 30 jours. Il est souvent surnommé l'indice de la peur du marché pétrolier, car il reflète les anticipations des investisseurs quant aux fluctuations futures des prix du pétrole. L'OVX est calculé par le Chicago Board Options Exchange (CBOE) en se basant sur les prix des options de l'ETF United States Oil Fund (USO), qui suit les contrats à terme sur le pétrole brut WTI. En analysant les primes de ces options, l'OVX estime la volatilité implicite, c'est-à-dire les attentes du marché concernant les variations futures des prix du pétrole.

Interprétation

- Valeur Élevée de l'OVX : Indique que les investisseurs anticipent une forte volatilité des prix du pétrole dans les 30 prochains jours, souvent en raison d'incertitudes ou de risques accrus sur le marché pétrolier.
- Valeur Faible de l'OVX : Suggère que le marché s'attend à une stabilité relative des prix du pétrole, avec moins de fluctuations prévues.

2.1.3 Ratio Futures/Spot

Le ratio Futures/Spot, souvent analysé à travers les concepts de contango et de backwardation, décrit la relation entre le prix actuel d'une matière première

(prix spot) et le prix de ses contrats à terme (futures). Cette relation est essentielle pour comprendre les anticipations du marché concernant l'évolution des prix du pétrole.

2.1.4 Contango et Backwardation : Définitions

- **Contango** : Cette situation survient lorsque le prix des contrats à terme est supérieur au prix spot actuel. Elle indique généralement une anticipation d'une hausse des prix du pétrole dans le futur, souvent en raison de coûts de stockage ou d'une offre excédentaire.
- **Backwardation** : À l'inverse, la backwardation se produit lorsque le prix des contrats à terme est inférieur au prix spot. Cela suggère une anticipation d'une baisse des prix du pétrole, potentiellement due à une demande élevée ou à une offre restreinte.

2.1.5 Calcul du Ratio Futures/Spot

$$\text{Ratio Futures/Spot} = \frac{\text{Prix du contrat à terme}}{\text{Prix spot}} \quad (1)$$

Interprétation

- Marché en Contango (Futures > Spot ou ratio supérieur à 1) : offre est supérieure à la demande ; Surplus de stock; Anticipation haussière des prix du pétrole.
- Marché en Backwardation (Futures < Spot ou ratio inférieur à 1) : demande excédentaire ; pénurie; Anticipation baissière des prix du pétrole.

2.1.6 Positions Nettes des Hedge Funds

Définition

La position nette des Hedges Funds reflète le sentiment spéculatif des grands investisseurs. La CFTC publie chaque semaine le COT Report, qui détaille les positions détenues par différents types de traders sur les marchés à terme. Dans le COT Report, les hedges funds sont généralement classés sous la catégorie "Managed Money". Cette catégorie inclut les gestionnaires de fonds, les conseillers en trading de matières premières et les opérateurs de fonds spéculatifs.

Les recherches de Kilian et Zhou (2020) ont montré que la combinaison de ces indicateurs dans un modèle à facteurs dynamique (FAVAR) améliore la précision des prévisions par rapport aux modèles traditionnels.

3 Collecte et Préparation des données

3.1 Récupération et préparation des données

Nous avons récupéré les données pour cette étude à partir du site *Commodity Future Trading Commition*. Le fichier Excel dans lequel nous avons sauvegarder nos données contient les colonnes suivantes : **la Date, le prix du WTI, le prix à terme du WTI, le ratio Futures/Spot, les Futures ouverts, l'OVX, les positions longues des Hedge Funds, les positions courtes des Hedge Funds et les positions nettes des Hedge Funds.**

La période de couverture s'étend du 1er Janvier 2010 au 28 Février 2025, sur une base hebdomadaire.

Pour l'analyse des prévisions, nous allons nous concentrer sur les prix du WTI, pour les modèles ARMA et GARCH. Pour le modèle à facteur, nous allons nous concentrer sur l'Open Interest, le Ratio Futures/Spot, les positions nettes des Hedge Funds ainsi que la volatilité OVX qui ont été expliqués précédemment.

3.2 Vérification de la stationnarité

Afin de vérifier la stationnarité des prix du WTI, nous pouvons effectuer le test de Dickey-Fuller avec les hypothèses suivantes :

- H_0 (hypothèse nulle) : La série est non stationnaire .
- H_1 (hypothèse alternative) : La série est stationnaire.

En effectuant le test, on obtient une $p - value = 0,4589 > 0,05$. Ainsi, pour un niveau de confiance à 95%, on ne rejette pas l'hypothèse nulle. Alors la série des prix n'est pas stationnaire.

Afin de la rendre stationnaire, nous utiliserons les rendements du prix du WTI. Pour confirmer que notre choix est bon, nous effectuons le test de Dickey-Fuller avec les mêmes hypothèses que ci-dessus. Nous obtenons alors un $p - value < 0,01$. Celui-ci étant inférieur à 0,05, pour un niveau de confiance de 95%, on rejette l'hypothèse nulle. Ainsi la série des rendements des prix du WTI est stationnaire.

Nous conserverons cette transformation pour la suite de notre étude.

3.3 Vérification de la normalité

Vérifions à présent la normalité des rendements du prix des WTI. La normalité des données est importante pour la suite de notre étude. En effet, pour les prévisions notamment ARMA et GARCH, on suppose que nos erreurs suivent une distribution normale. Afin de vérifier cela, effectuons différents tests :

1. Le test de Jarque-Bera avec les hypothèses suivantes :

- H_0 : Les données suivent une distribution normale.
- H_1 : Les données ne suivent pas une distribution normale.

Nous obtenons une $p - value < 2.2e - 16$. Cette $p - value$ étant inférieur à 0,05, pour un niveau de confiance de 95%, on rejette l'hypothèse nulle. Les rendements des prix du WTI ne suivent donc pas une distribution normale.

2. Le test de Shapiro-Wilk avec les hypothèses suivantes :

- H_0 : Les données suivent une distribution normale.
- H_1 : Les données ne suivent pas une distribution normale.

Nous obtenons un $p - value < 2.2e - 16$. Cette $p - value$ étant inférieur à 0,05, pour un niveau de confiance de 95%, on rejette l'hypothèse nulle. Cela appuie donc notre première analyse de la normalité.

3. L'analyse du graphique du QQplot :

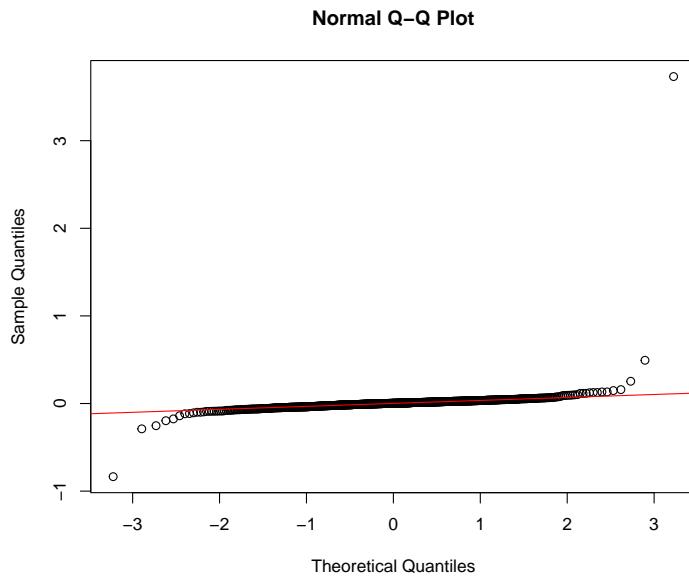


Figure 1: Graphique QQPlot

On constate directement que les extrémités des rendements du prix du WTI ne suivent pas la droite de normalité. De plus, on remarque un point très éloigné qui semble être un point aberrant.

Ainsi, nous pouvons conclure que les rendements du prix de WTI ne suivent pas une loi normale. Le fait que nos données ne soient pas normales renforce l'utilisation des modèles ARMA et GARCH pour les prévisions.

3.4 Analyse des ACF et PACF des prix

Afin de poursuivre correctement notre étude, vérifions à l'aide des graphiques de la fonction d'autocorrélation (ACF) et la fonction d'autocorrélation partielle (PACF), la présence ou non de délai significatifs.

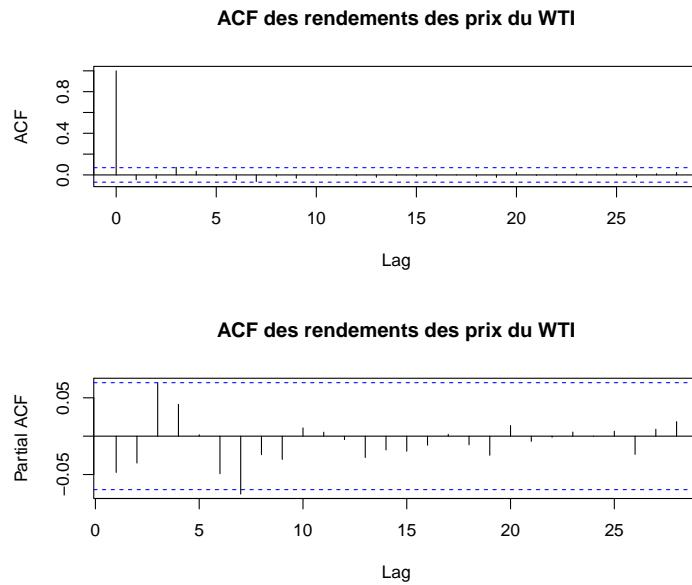


Figure 2: Graphiques des ACF et PACF

On peut observer pour l'ACF au niveau du délai 3 un pic légèrement significatif. Cela pourrait nous mettre sur la piste d'un modèle à moyenne mobile (MA). De même, en observant le PACF, on remarque au niveau des délai 2 et 7 des pics significatifs, ce qui pourrait indiquer la présence d'un modèle autorégressif (AR). Nous reviendrons plus en détail sur ces modèles dans la section suivante.

Dans cette prochaine section, nous allons explorer des modèles adaptés afin de capturer la tendance temporelle présente dans nos données.

4 Etude de différents modèles Econométrique

4.1 Le modèle ARMA

Définition : Lorsque l'on parle de modèle autorégressifs intégrés à moyennes mobiles on peut parler de ARMA. Afin de comprendre l'ensemble du modèle ARMA, nous allons détailler ce qu'est un modèle autorégressifs (AR) et également un modèle à moyennes mobiles (MA) qui compose le modèle ARMA.

1. Un modèle autorégressif d'ordre p s'écrit comme :

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + w_t$$

Où les ϕ_1, \dots, ϕ_p sont des paramètres et w_t est un bruit blanc.

2. Un modèle à moyenne mobile d'ordre q s'écrit comme :

$$X_t = w_t + \theta_1 w_{t-1} + \theta_2 w_{t-2} + \dots + \theta_q w_{t-q}$$

Où les $\theta_1, \dots, \theta_q$ sont des paramètres et w_t sont des bruits blancs.

Maintenant que nous avons développé les différents modèles tel que AR(p) et MA(q), nous pouvons développer le modèle ARMA(p, q).

Un modèle autorégressif à moyennes mobiles d'ordre (p, q) s'écrit comme :

$$X_t = \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + w_t + \theta_1 w_{t-1} + \dots + \theta_q w_{t-q}$$

Où les ϕ_1, \dots, ϕ_p et $\theta_1, \dots, \theta_q$ sont des paramètres et $\{w_t\}$ sont des bruits blancs.

Dans notre cas, pour le calcul de prévisions peut être facilement abordé dans le contexte des modèles ARMA.

Pour choisir le modèle ARMA approprié, plusieurs étapes seront suivies. Tout d'abord, nous identifierons les modèles AR et MA en nous basant sur les fonctions PACF et ACF respectivement, afin d'avoir une première idée du modèle. Ensuite, pour déterminer le meilleur modèle, nous comparerons les valeurs de l'AIC et du BIC, et sélectionnerons celui présentant les valeurs les plus faibles. Nous réaliserons également le test de Jarque-Bera pour vérifier la normalité du modèle. Enfin, nous analyserons les graphiques des fonctions PACF et ACF pour le modèle ARMA sélectionné.

4.2 Le modèle ARMA-GARCH

Dans le modèle ARMA précédent :

$$\phi(B)X_t = \theta(B)w_t$$

Où $\phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$ et $\theta(B) = 1 + \theta_1 B + \dots + \theta_q B^q$

Nous avons présumer que $\{w_t\}$ est un bruit blanc fort. C'est à dire que :

$$Var(w_t | w_{t-1}, w_{t-2}, \dots) = Var(w_t) = \sigma_w^2$$

En effet, la distribution conditionnelle de $\{w_t\}$ conditionnellement à est la même que la distribution inconditionnelle. Ainsi la variance conditionnelle est une constante.

Dans les applications financières, un concept important est la volatilité. Typiquement, la volatilité d'une séries chronologiques de rendements affiche des périodes d'agitation et de calme, que l'on appelle grappes de volatilité. On considère une série de rendements $\{X_t\}$. L'idée générale est de proposer des modèles $\{X_t\}$ qui présument que : $Var(X_t) = \text{une constante en } t$ et que $Var(X_t | X_{t-1}, X_{t-2}, \dots) = \text{une fonction en } t$.

Ainsi la variance conditionnelle n'est pas constante.

L'idée est donc d'exploiter des modèles capable de gérer cette volatilité. Pour cela on va aborder le modèle GARCH.

Définition : Un modèle GARCH d'ordre m , notée GARCH(m) s'écrit comme :

$$X_t = \sigma_t \epsilon_t$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + \dots + \alpha_m X_{t-m}^2 + \beta_1 \sigma_{t-1}^2 + \dots + \beta_s \sigma_{t-s}^2$$

On peut aussi écrire :

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^m \alpha_i X_{t-i}^2 + \sum_{j=1}^s \beta_j \sigma_{t-j}^2$$

Avec $\epsilon \sim \mathcal{N}(0, \sigma^2 I)$ un bruit blanc Gaussien iid

4.3 Le modèle à facteurs

Contexte général

Les modèles à facteurs sont des modèles statistiques ou économétriques permettant de représenter la dynamique d'une variable d'intérêt (ex. rendement, volatilité, état de l'économie, etc.) à l'aide de facteurs latents ou observés.

Dans ce contexte, on cherche à prédire une **variable d'état** (le rendement du WTI) à partir de plusieurs facteurs explicatifs.

Formulation générale

Soit r_{it} une variable observée pour l'unité i au temps t , modélisée par :

$$r_{it} = \alpha_i + \beta_{i1} f_{1t} + \beta_{i2} f_{2t} + \dots + \beta_{im} f_{mt} + \varepsilon_{it}$$

ou, en notation vectorielle :

$$r_{it} = \alpha_i + \boldsymbol{\beta}_i' \mathbf{f}_t + \varepsilon_{it}$$

avec :

- α_i : constante (intercept),
- $\beta_i = (\beta_{i1}, \dots, \beta_{im})'$: vecteur des coefficients,
- $\mathbf{f}_t = (f_{1t}, \dots, f_{mt})'$: vecteur des facteurs,
- ε_{it} : bruit ou erreur spécifique.

Facteurs considérés

En effet, nous avons constaté par un test de stationnarité sur le ratio Futures/Spot, et le nombre de Open Futures que ces deux facteurs ne sont pas stationnaires. Nous avons donc décidé d'utiliser leur rendement qui est stationnaire

- Le rendement du ratio Futures/Spot,
- Le rendement du nombre de Open Futures,
- L'index de la volatilité du CBOE Crude Oil,
- La position nette des Hedge Funds sur les Futures du WTI,

Hypothèses du modèle

$$\begin{aligned} E(f_t) &= \mu_f, & Cov(f_t) &= \Sigma_f \\ E(\varepsilon_{it}) &= 0, & Cov(f_t, \varepsilon_{it}) &= 0 \\ Cov(\varepsilon_{it}, \varepsilon_{js}) &= \begin{cases} \sigma_i^2 & \text{si } i = j \text{ et } t = s \\ 0 & \text{sinon} \end{cases} \end{aligned}$$

Forme matricielle

Pour k variables et T périodes :

$$\mathbf{r}_t = \boldsymbol{\alpha} + \mathbf{B}\mathbf{f}_t + \boldsymbol{\varepsilon}_t$$

avec :

- $\mathbf{r}_t \in R^k$,
- $\boldsymbol{\alpha} \in R^k$,
- $\mathbf{B} \in R^{k \times m}$,
- $\mathbf{f}_t \in R^m$,
- $\boldsymbol{\varepsilon}_t \in R^k$.

Covariance des variables observées

$$Cov(\mathbf{r}_t) = \mathbf{B}\Sigma_f\mathbf{B}' + D$$

où D est une matrice diagonale des variances spécifiques.

Utilisation en prédition

Ce modèle peut être estimé par :

- Régression linéaire multiple (si les facteurs sont observés),
- ACP ou analyse factorielle (si les facteurs sont latents),
- Régression pénalisée (si m est grand),
- Modèles dynamiques (pour les données temporelles).

Nous nous focaliserons sur le model de Regression lineaire multiple, et ferons une ACP qui nous permettra d'extraire des facteurs non observés (latents) à partir des données, en réduisant la dimension tout en capturant le maximum de variance commune entre les variables pour vérifier si la prediction ne serait pas meilleure avec moins de facteurs à considérer.

Avantages du modèle multifactoriel

- Réduction de dimension,
- Interprétation économique claire,
- Filtrage de signaux pertinents,
- Base pour des modèles plus avancés.

5 Estimation et Validation

5.1 Le modèle ARMA

Pour le modèle ARMA, nous avons déterminé qu'un ARMA(2,2) correspond bien à nos données à l'aide de la fonction *auto.arima* sur *R*. Afin d'être sûr de choisir le bon modèle, nous allons essayer trois modèles différents qui sont proche de la proposition de la fonction *auto.arima*.

Nous allons comparer les résultats des différents *AIC* et *BIC*, puis nous sélectionnerons le modèle avec les plus faibles résultats de ces paramètres. Nous obtenons les résultats suivants :

	ARMA(1,2)	ARMA(2,1)	ARMA(2,2)
AIC	-805,73	-803,64	-809,78
BIC	-796,40	-780,34	-786,55

Table 1: Choix du modèle ARMA

On remarque que le plus faible *BIC* est lorsque l'on sélectionne le modèle ARMA(1,2). Ainsi, que le plus faible *AIC* est lorsque l'on sélectionne le modèle ARMA(2,2). Afin de sélectionner le meilleur modèle parmi les modèles ARMA(1,2) et ARMA(2,2), nous pouvons comparer leur ACF respectifs.

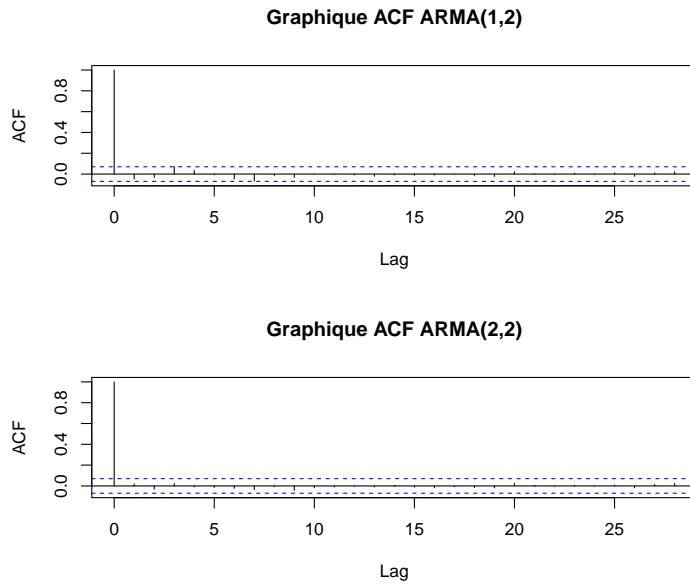


Figure 3: Graphiques ACF des modèles ARMA(1,2) et ARMA(2,2)

Nous observons un léger pic significatif pour le ARMA(1,2). Il serait donc

plus intéressant de sélectionner le modèle ARMA(2,2).

Pour confirmer notre choix nous pouvons faire un test de Ljung Box pour vérifier la présence ou non d'autocorrélation. Pour ce test, nous avons les hypothèses suivantes :

- H_0 : Il n'y a pas d'autocorrélation.
- H_1 : Il y a la présence d'autocorrélation.

Nous obtenons un $p-value = 0.5998$. Cette $p-value$ étant supérieur à 0,05, pour un niveau de confiance à 95%, on ne rejette pas l'hypothèse nulle. Ainsi, il n'y a pas de présence d'autocorrélation. Cela veut dire que notre modèle *ARMA*(2, 2) capture bien la structure temporelle de nos données.

5.2 Le modèle ARMA-GARCH

Afin de sélectionner le meilleur modèle ARMA-GARCH, nous allons tester plusieurs modèles et choisir celui qui aura le plus petit *AIC* et le plus faible *BIC*. Nous avons testé les modèle suivants (chaque modèle est additionné avec un *ARMA*(2, 2)):

	GARCH(1,0)	GARCH(1,1)	GARCH(2,1)	GARCH(2,2)
AIC	-3,52	-3,54	-3,53	-3,55
BIC	-3,49	-3,51	-3,50	-3,50

Table 2: Choix du modèle ARMA-GARCH

On remarque que l'*AIC* le plus faible est lorsque nous avons utilisé le modèle ARMA-GARCH(2,2). Cependant, le *BIC*, le plus faible est celui du modèle ARMA-GARCH(1,1). Comme les deux *AIC* sont très proches, nous allons sélectionner le modèle ARMA(2,2)-GARCH(1,1). En effet, on préfère toujours privilégier un modèle moins surchargé mais tout aussi performant.

Nous pouvons effectuer un test sur les résidus standardisés avec le test de Ljung-Box et les hypothèses suivantes :

- H_0 : Il n'y a pas d'autocorrélation.
- H_1 : Il y a la présence d'autocorrélation.

Nous obtenons un $p-value = 0.2941$. Cette $p-value$ étant supérieure à 0.05, pour un niveau de confiance à 95%, on ne rejette pas l'hypothèse nulle. Ainsi, le choix du modèle GARCH(1,1), sélectionné capture bien la volatilité de nos données.

5.3 Estimation du modèle à facteurs

5.3.1 Extraction des facteurs par ACP

Pour estimer notre modèle à facteurs, nous avons appliqué une **analyse en composantes principales (ACP)** sur quatre variables clés reflétant l'activité, la volatilité et le positionnement des acteurs sur les marchés à terme pétroliers :

- Le rendement du ratio Futures/Spot (`Log_Delta_Ratio`), mesurant la structure des anticipations de prix,
- Le rendement de l'Open Interest (`Log_Delta_OI`), indicateur d'engagement du marché,
- L'indice de volatilité du pétrole (`OVX`), représentant le niveau de risque perçu,
- La position nette des Hedge Funds (`Net_Position`), reflet du positionnement spéculatif.

Après standardisation, l'ACP a permis d'extraire quatre composantes principales (PC1 à PC4) expliquant l'intégralité de la variance (100%) du jeu de données initial :

- PC1 : 30,2% (stress de marché et volatilité),
- PC2 : 25,2% (engagement sur les marchés à terme),
- PC3 : 23,9% (structure du marché via le ratio Futures/Spot),
- PC4 : 20,7% (composante mixte Net Position + OVX).

Les poids (*loadings*) des variables dans les composantes sont résumés dans le tableau suivant :

	Log_Delta_OI	Log_Delta_Ratio	Net_Position	OVX
PC1	-0.1405	0.3857	-0.6390	0.6505
PC2	0.8852	-0.3502	-0.2801	0.1237
PC3	0.4231	0.8532	0.1542	-0.2631
PC4	0.1328	0.0241	0.6996	0.7017

Table 3: Poids des variables dans chaque composante principale

5.3.2 Interprétation des composantes principales

PC1 – Stress spéculatif et nervosité du marché Cette composante est fortement influencée par une hausse de l'OVX (+0.6505) et une baisse de la position nette des hedge funds (-0.6390), ce qui traduit un climat d'incertitude où les investisseurs réduisent leur exposition.

PC2 – Activité et engagement du marché Dominée par Log_Delta_OI (+0.8852), elle reflète directement l'intensité de l'activité sur les marchés à terme du pétrole.

PC3 – Structure des anticipations de prix Fortement corrélée au ratio Futures/Spot (+0.8532), cette composante illustre la perception des anticipations de prix et la structure des contrats (contango vs backwardation).

PC4 – Positionnement et perception du risque Cette dernière composante mêle OVX (+0.7017) et Net Position (+0.6996), traduisant un facteur combiné de positionnement spéculatif dans un environnement volatile.

5.3.3 Visualisation factorielle et lien avec les rendements du WTI

Les projections des rendements du WTI sur ces composantes révèlent des tendances claires :

- PC1 : relation négative forte avec les rendements (stress de marché \Rightarrow baisse du WTI),
- PC2 : relation positive modérée (activité accrue \Rightarrow soutien au prix du WTI),
- PC3 : relation négative marquée (structure anticipative défavorable \Rightarrow rendements négatifs),
- PC4 : impact plus nuancé mais significatif.

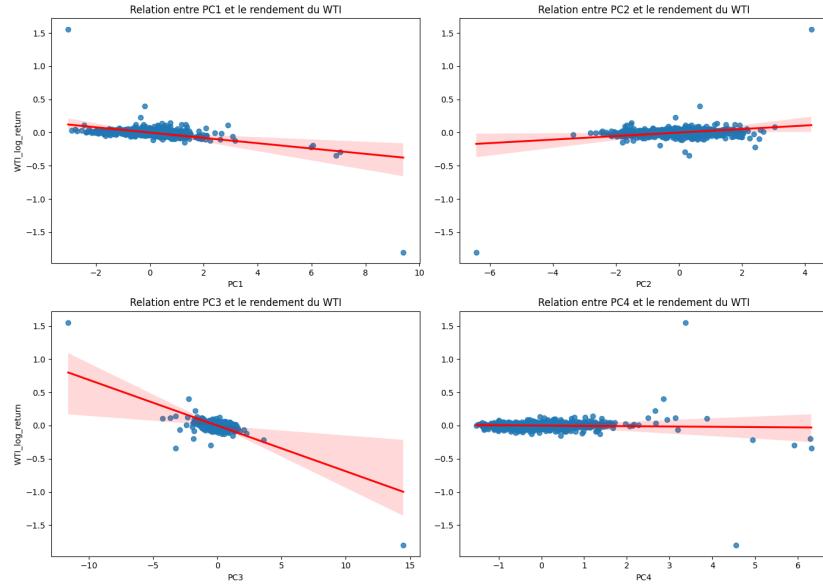


Figure 4: Relations entre composantes principales (ACP) et rendements hebdomadaires du WTI

5.3.4 Estimation du modèle linéaire

Nous avons estimé le modèle économétrique suivant :

$$\text{RWTH}_t = \alpha + \beta_1 \cdot \text{PC1}_t + \beta_2 \cdot \text{PC2}_t + \beta_3 \cdot \text{PC3}_t + \beta_4 \cdot \text{PC4}_t + \varepsilon_t \quad (2)$$

Les résultats du modèle sont les suivants :

- **R² = 0.781** : 78% de la variance expliquée,
- Tous les coefficients β_i sont significatifs au seuil de 1%,
- $\beta_1 < 0$ (stress), $\beta_2 > 0$ (activité), $\beta_3 \ll 0$ (structure du marché), $\beta_4 < 0$ (risque combiné).

```

OLS Regression Results
=====
Dep. Variable: WTI_log_return R-squared:      0.781
Model:          OLS   Adj. R-squared:      0.780
Method:         Least Squares F-statistic:    701.8
Date:        Tue, 08 Apr 2025 Prob (F-statistic): 1.76e-257
Time:           00:11:47 Log-Likelihood:     1330.5
No. Observations: 790   AIC:             -2651.
Df Residuals:    785   BIC:             -2628.
Df Model:        4
Covariance Type: nonrobust
=====
            coef    std err      t      P>|t|      [0.025      0.975]
-----
const      -0.0001    0.002    -0.067    0.947    -0.003     0.003
PC1       -0.0401    0.001   -27.545    0.000    -0.043    -0.037
PC2        0.0266    0.002    16.593    0.000     0.023     0.030
PC3       -0.0689    0.002   -42.039    0.000    -0.072    -0.066
PC4       -0.0045    0.002    -2.580    0.010    -0.008    -0.001
=====
Omnibus:        493.071 Durbin-Watson:      1.624
Prob(Omnibus):    0.000  Jarque-Bera (JB): 29255.312
Skew:           2.072  Prob(JB):            0.00
Kurtosis:        32.523 Cond. No.          1.21
=====

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

```

Figure 5: Sortie du modèle OLS appliqué aux quatre composantes principales

6 Prévision et interprétation

Pour effectuer nos prévisions avec les modèles sélectionnés, nous avons tronqué les 9 dernières lignes de notre jeu de données, soit les données du 3 Janvier 2025 au 28 Février 2025. Pour sélectionner le meilleur modèle, nous comparerons nos prévisions avec les valeurs réelle , en utilisant l' MSE . l' MSE est calculé de la manière suivante :

$$MSE = \frac{1}{9} \sum_{i=1}^9 (\text{valeur réelle}_i - \text{valeur prédite}_i)^2$$

6.1 Le modèle ARMA(2,2)

Une fois nos prévisions faites pour les neuf dernières semaines, nous obtenons les résultats suivants :

Date	Valeur réelle	Valeur prédite	Erreur
Janv 03, 2025	73,15	70,35	2,8
Janv 10, 2025	75,14	70,37	4,77
Janv 17, 2025	79,28	70,51	8,77
Janv 24, 2025	75,73	70,62	5,11
Janv 31, 2025	73,31	70,63	2,68
Fev 7, 2025	72,05	70,56	1,49
Fev 14, 2025	72,17	70,49	1,68
Fev 21, 2025	72,1	70,47	1,63
Fev 28, 2025	69,93	70,50	-0,57

Table 4: Analyse des prévisions du modèle $ARMA(2, 2)$

Nous avons calculé les erreurs de la façon suivante : $erreur = \text{Valeur réelle} - \text{Valeur prédite}$.

Après un rapide calcul, nous obtenons un $MSE = 16,54$. Passons au modèle $ARMA - GARCH$, afin de comparer d'abord entre ces deux modèles, lequel prévoit le mieux.

6.2 Le modèle ARMA(2,2)-GARCH(1,1)

Nous effectuons également nos prévisions sur les 9 dernières semaines, et obtenons les résultats suivants :

Date	Valeur réelle	Valeur prédictive	Erreur
Janv 03, 2025	73,15	70,81	2,34
Janv 10, 2025	75,14	71,11	4,03
Janv 17, 2025	79,28	71,42	7,86
Janv 24, 2025	75,73	71,74	4,49
Janv 31, 2025	73,31	72,08	1,23
Fev 7, 2025	72,05	72,41	-0,36
Fev 14, 2025	72,17	72,75	-0,58
Fev 21, 2025	72,1	73,09	-0,99
Fev 28, 2025	69,93	73,42	-3,49

Table 5: Analyse des prévisions du modèle $ARMA(2,2) - GARCH(1,1)$

Nous avons calculé les erreurs par le même calcul de pour le modèle ARMA. Après un rapide calcul, nous obtenons que l' $MSE = 13,20$. On obtient donc un MSE inférieur pour le modèle AMRA-GARCH comparer au MSE du modèle ARMA. Malgré que nos prévisions ne soient pas exact pour les deux modèles. Entre ces deux là, nous allons préférer utiliser le modèle ARMA-GARCH pour effectuer nos prévisions.

Passons aux prévisions réalisé avec le modèle à facteurs et comparons les résultats afin de sélectionner le meilleur modèle de prévision.

6.3 Le modèle à facteurs

Une prédiction a été réalisée sur 20% des observations pour tester le pouvoir prédictif hors échantillon du modèle.

- Le modèle capture bien la tendance globale des rendements du WTI, en effet, l'erreur moyenne quadratique sur le jeu de test est de 0.0009 et le R^2 est de 0.1751. Ce qui indique que les predictions du model sont assez proches des valeurs réels mais le model n'explique qu'une petite partie de la dynamique sous-jacente des données.
- Des écarts apparaissent lors de fortes variations ponctuelles (typique en finance),
- Le graphe suivant illustre cette adéquation temporelle.

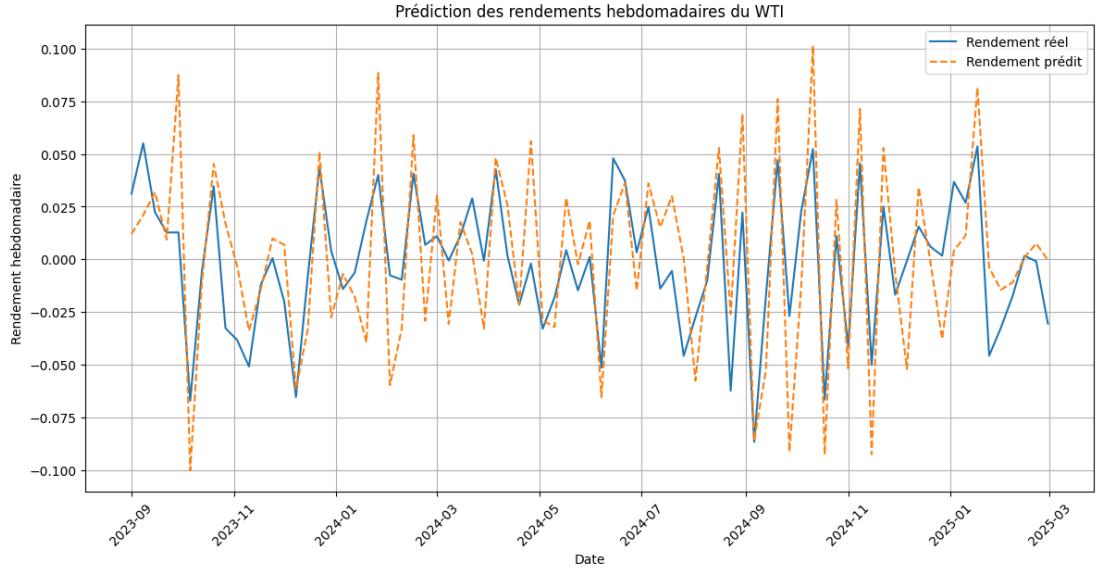


Figure 6: Comparaison des rendements réels et prédits du WTI (hors échantillon)

Conclusion intermédiaire (Model a facteurs)

L'analyse factorielle via ACP permet d'extraire des variables synthétiques portant une interprétation économique forte. Leur utilisation dans une régression linéaire permet de modéliser efficacement les rendements du WTI, tout en conservant une capacité prédictive satisfaisante.

Ces résultats encouragent l'intégration de ces facteurs dans des modèles économétriques dynamiques plus avancés (ARMA, GARCH, etc.), et dans des approches hybrides mêlant machine learning et séries temporelles.

7 Conclusion

Dans ce projet, on cherchait le modèles le plus pertinent pour la prévision hebdomadaire du pétrole WTI. Nous avons comparer trois approches : le modèle ARMA(2,2), le modèle ARMA(2,2)-GARCH(1,1) et un modèle à facteurs fondé sur des indicateurs de marché. Chacune des approches ont été validés à l'aide de tests statistiques et de mesures d'erreur prédictive hors échantillon.

En comparant le modèle ARMA-GARCH et ARMA, on remarque une amélioration des prévisions en ajoutant de la volatilité conditionnelle.

Quant au modèle à facteur, il se distingue par sa bonne précision prédictive, mais surtout par sa capacité à donner un sens économique aux mouvement du WTI. Grâce à une analyse des composantes principales (ACP), on peut rélevé quatre facteurs :

- Le PC1, correspond au **stress spéculatif**. Elle permet de capter la nervosité du marché. Elle est corrélée positivement à OVX et négativement aux positions nettes des hedges funds. De plus, elle est fortement corrélé négativement aux rendements du WTI ce qui implique une baisse de prix lorsqu'il y a de l'incertitude sur le marché.
- Le PC2, correspond à l'**engagement du marché**. Il dépend de l'évolution de l'Open Interest, qui reflète l'activités des spéculateurs. Quand cet indicateur augmente, cela montre que le marché est plus actif, ce qui a tendance à faire augmenter le prix.
- Le PC3, correspond à la **structure des anticipations**. Elle est corrélée au ratio Futures/Spot. Ce rapport montre comment les investisseurs s'attendent à ce que le prix évoluent. Si le prix des Futures est plus élevé que le prix Spot, on parle de contango, cela veut dire que l'on s'attend à une baisse des prix, c'est donc associé à des rendements négatifs.
- Le PC4 correspond au **risque combiné**. Elle combine les deux estimateurs OVX et Net Position. Son impact est assez complexe à interpréter mais elle importante dans notre modèle d'un point de vu statistique.

Pour synthétiser notre évaluation, une analyse SWOT a été conduite :

- Forces (internes) : Le projet repose sur une méthodologie rigoureuse, intègre des outils économétriques standards et avancés, et s'appuie sur des données financières concrètes. L'utilisation conjointe de modèles linéaires et de techniques de réduction dimensionnelle (ACP) permet à la fois interprétabilité et efficacité.
- Faiblesses (internes) : Malgré la pertinence des modèles, les erreurs de prévision demeurent non négligeables, notamment lors de variations extrêmes. Le modèle reste sensible à l'absence de variables exogènes (ex. chocs géopolitiques, décisions de l'OPEP).

- Opportunités (externes) : Ce type de modèle peut s'intégrer à des outils décisionnels pour les acteurs de l'énergie et de la finance, notamment en gestion de portefeuille, en couverture contre le risque de prix, ou en algorithmie de trading. Il est également transposable à d'autres commodités (gaz naturel, métaux précieux).
- Menaces (externes) : Le développement de modèles fondés sur l'intelligence artificielle (LSTM, XGBoost, etc.) pourrait concurrencer les approches économétriques traditionnelles. Par ailleurs, l'imprévisibilité accrue des marchés de l'énergie, exacerbée par les crises politiques ou environnementales, constitue un défi pour tout modèle fondé sur des variables historiques.

En conclusion, ce projet constitue une base solide pour la prévision du WTI, mais suggère aussi que l'avenir réside probablement dans une hybridation entre l'interprétabilité des modèles économétriques et la capacité adaptative des méthodes de machine learning.