

Master 1 Économétrie et Statistiques

Étude des séries temporelles de l'indice de prix à la
consommation (IPC), le prix du pétrole et la variation des
températures en France, à l'aide de la méthodologie ARIMA

Une analyse des tendances et des fluctuations de 2010 jusqu'à
2022

Séries temporelles univariées - D.Girard

Gloria Isabel PALACIO

Avril 2024

Sommaire

1	INTRODUCTION	3
2	ANALYSE ECONOMIQUE DU SUJET	4
3	ANALYSE DES TROIS VARIABLES	5
4	PRESENTATION DE LA METHODOLOGIE ARIMA.....	10
5	CHOIX DU MODELE ARIMA POUR YT	17
6	ANALYSE DES VARIABLES PETROLE ET TEMPERATURE	28
7	TEST DE COINTEGRATION SELON ENGLE-GRANGER.....	33
8	CONCLUSION	40
9	DISCUSSION.....	42
10	ANNEXES	43
11	BIBLIOGRAPHIE	50
12	TABLE DES FIGURES	52
	TABLE DES MATIERES	53

1 Introduction

L'indice des prix à la consommation (IPC) des produits alimentaires constitue un baromètre crucial de notre économie, reflétant non seulement les fluctuations des coûts de la vie mais aussi leur impact direct sur le quotidien des individus. Dans un contexte mondial marqué par des fluctuations significatives des prix du pétrole et des variations climatiques de plus en plus prononcées, comprendre l'interaction entre ces variables et l'IPC devient impératif.

Ce projet vise à analyser l'influence du prix du pétrole et des variations de températures sur l'IPC des produits alimentaires en France, de 2010 à 2022, en utilisant la méthodologie ARIMA (AutoRegressive Integrated Moving Average), une approche reconnue pour sa précision dans la prévision des séries temporelles.

L'objectif de cette analyse est double : d'une part, nous explorons les tendances et fluctuations au sein de ces séries temporelles, et d'autre part, nous examinons leurs interrelations à travers l'analyse de cointégration.

Pour ce faire, nous procéderons à une analyse économique approfondie en présentant les différentes variables pertinentes à notre étude.

Ensuite, nous examinerons les données en détail, afin d'identifier les tendances et les variations.

Par la suite, une partie théorique sera consacrée à la présentation et à l'application de la méthodologie ARIMA, qui nous permettra d'effectuer des prévisions quant à l'évolution des séries temporelles.

Enfin, l'utilisation de tests tels que celui de Dickey-Fuller augmenté (ADF), et de cointégration d'Engle-Granger, nous aidera à établir la validité de ces relations à long terme.

En révélant comment les fluctuations économiques et environnementales peuvent influencer l'inflation alimentaire, ce travail ambitionne de contribuer à une meilleure anticipation des défis futurs liés à l'économie et au climat.

2 Analyse économique du sujet

L'augmentation des prix alimentaires pèse lourdement sur le budget de nombreux Français. Selon l'Observatoire des Vulnérabilités Alimentaires de la Fondation Nestlé France, 37 % des Français se trouvent en situation d'insécurité alimentaire, c'est-à-dire qu'ils n'ont pas un accès régulier à une alimentation à la fois saine et suffisante. Ce chiffre a triplé par rapport à il y a huit ans, passant de 11 % en 2015 à 37 %¹. L'inflation est le principal facteur explicatif, notamment l'année 2022 qui a connu une forte inflation, particulièrement marquée dans le secteur des produits alimentaires.

L'objectif est d'explorer comment l'indice des prix à la consommation des aliments, qui offre une perspective sur l'évolution de l'inflation, est influencé par le prix du pétrole et les variations de température. Le choix de ces deux variables explicatives est motivé par le fait que notre première variable joue un rôle important sur les coûts de production et de transport, tandis que la deuxième impacte directement la production agricole.

La relation entre le prix du pétrole et l'IPC est bien connue et documentée ; la Banque de France note par exemple que chaque augmentation de 10 euros du prix du pétrole se traduit par une hausse de 0,4% de l'IPC tant dans la zone euro qu'en France².

Concernant la variation de température, un nouveau terme est apparu, "heatflation", qui combine "chaleur" et "inflation" pour illustrer comment les températures élevées peuvent réduire la production agricole et conduire à une augmentation de l'inflation. Selon une étude publiée dans Communications Earth & Environment, si la tendance à la hausse des températures se poursuit, nous pourrions observer une augmentation moyenne des prix de l'alimentation de 1,49% par an d'ici à 2035, même dans le meilleur des scénarios³. Ce phénomène révèle l'impact direct et significatif du changement climatique sur notre

¹ (Bleu, 2023)

² (Ouvrard, 2018)

³ (AFP, 2024)

économie, en particulier dans le secteur de l'alimentation, avec des implications qui varient selon les régions et les saisons.

3 Analyse des trois variables

Dans cette section, nous allons analyser les caractéristiques générales de nos 3 séries temporelles. Un point particulier sera mis sur l'évolution des séries à travers une analyse graphique, qui vise à détecter les moments clés de variation. Pour réaliser les graphiques, ainsi que pour l'ensemble du document, nous avons utilisé le logiciel Gretl.

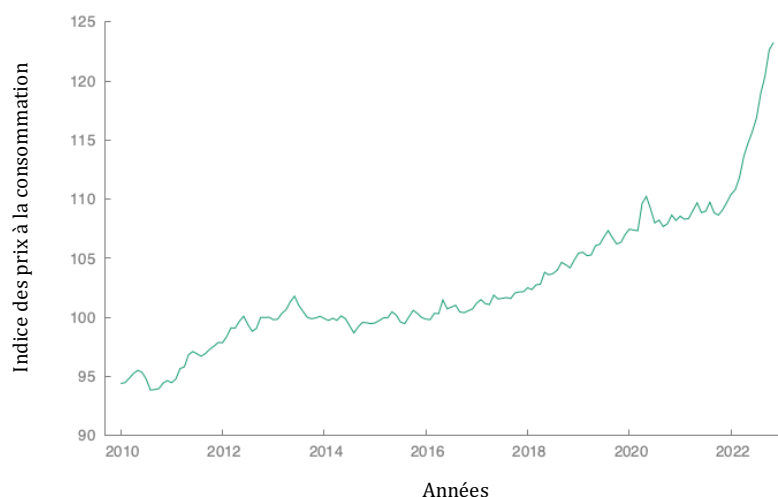
3.1 Variable à expliquer

Notre variable dépendante est l'indice des prix à la consommation (IPC), un indicateur économique clé qui mesure l'évolution des prix d'un "panier moyen" de biens et services consommés par les ménages.

Nous nous intéresserons spécifiquement au « Prix à la Consommation, Indices d'alimentation », pour la période allant de janvier 2010 à novembre 2022 en France (figure 1). Cet indice fonctionne comme un indicateur des variations des prix des produits alimentaires, signalant une inflation lorsque les prix augmentent (indice supérieur à 100) et une déflation lorsqu'ils diminuent (indice inférieur à 100). En s'appuyant sur les données de la FAOSTAT⁴, notre étude permet de comprendre un élément clé qui influence directement le pouvoir d'achat et la qualité de vie des consommateurs français.

⁴ FAOSTAT, est une division de la FAO (Organisation des Nations Unies pour l'alimentation et l'agriculture)

Figure 1 - Évolution de l'IPC alimentaire en France



Le graphique ci-dessus nous permet d'observer l'évolution depuis 2010, une période marquée par l'application de politiques de relance économique post-crise de 2008. Initialement, en 2010, l'IPC dans le secteur alimentaire a enregistré des légères hausses, passant par exemple 94,42 en janvier à 94,66 en décembre. Cette faible tendance à la hausse s'est maintenue jusqu'en 2012, où l'IPC franchi le seuil des 100 pour la première fois. Entre 2013 et 2015, nous observons une certaine stabilité relative et c'est à partir de 2015 qu'il a augmenté progressivement (période de référence 2015 = 100) et connaît une accélération marquée.

A partir de 2021, la France fait face à un épisode d'inflation exceptionnel, touchant particulièrement le secteur alimentaire. Initialement considérée comme une conséquence temporaire de la reprise économique post-COVID (2020) et du rebond des prix du pétrole⁵, cette inflation s'est avérée plus complexe et persistante. En effet, elle s'est étendue à un large éventail de biens et services, avec des répercussions importantes sur les coûts alimentaires. L'escalade du conflit en Ukraine en 2022 a exacerbé la situation, entraînant une hausse supplémentaire des prix de l'énergie ainsi que des perturbations dans les marchés des matières premières agricoles, l'Ukraine étant un fournisseur majeur de céréales et d'huiles végétales. Le secteur de l'alimentation demeure parmi le plus touché par l'inflation⁶.

⁵ (Ulgazi, 2021)

⁶ (Haut Conseil de la famille, de l'enfance et de l'âge, 2023)

3.2 Variable explicative 1 : Prix du pétrole

Notre première variable explicative est le prix du pétrole, car il est essentiel à l'économie française. Nous nous focalisons sur le prix au comptant du Brent, exprimé en euros par baril selon les données de l'INSEE⁷. Ce prix spécifique du « Brent », provenant d'un champ pétrolifère en mer du Nord, mais qui est coté à l'Intercontinental Exchange (ICE) de Londres, sert de référence mondiale pour le pétrole brut. Sa qualité et sa disponibilité font du Brent le standard international pour le prix du pétrole brut, influençant les coûts énergétiques et, par conséquence, les économies et politiques énergétiques.

L'impact du prix du pétrole sur l'économie est considérable, car il affecte directement les coûts de production et de transport de nombreux biens, y compris les produits alimentaires. Mais au-delà de son influence sur les coûts de transport, le prix du pétrole affecte la production agricole. Cela s'observe notamment à travers le coût des engrais et de l'énergie nécessaires non seulement aux processus de production, mais aussi de conservation des aliments. Ainsi, le prix du pétrole joue un rôle essentiel dans l'évolution de l'inflation alimentaire.

Figure 2 - Pétrole brut Brent (Londres) - Prix au comptant en euros par baril



⁷ (INSEE, 2024)

Sur la figure 2, nous pouvons observer l'évolution du prix en euros durant la période janvier 2010 à novembre 2022.

Au début de la décennie, les prix augmentent progressivement, passant de 53,5 € le baril en janvier 2010 à 82,3 € en juin 2014. Cette période est caractérisée par une reprise économique post-crise financière de 2008, entraînant une augmentation de la demande de pétrole.

Fin 2014 marque le début d'une chute significative des prix, avec une descente à 52,2 € en décembre 2014, cette chute est principalement due à une surproduction mondiale de pétrole et au ralentissement économique⁸.

Cette tendance à la baisse s'est poursuivie jusqu'en janvier 2016, où le prix a atteint un plancher de 28,3 €, soulevant des inquiétudes sur les marchés. Car bien que cette diminution puisse bénéficier au pouvoir d'achat des Français, elle représente un risque pour l'économie. En effet, un prix trop bas peut déstabiliser les pays exportateurs de pétrole. Ces derniers, confrontés à des difficultés financières, pourraient réduire leurs importations de produits manufacturés en provenance des pays développés.⁹

Finalement après 2016, les prix se redressent légèrement, stabilisant autour de 50-60 € jusqu'en 2019. Cette reprise est freinée par la pandémie de COVID-19 en 2020, provoquant une chute historique des prix en avril 2020 à seulement 16,9 €, sous l'effet d'une demande mondiale anéantie par les confinements et les restrictions de voyage.

À partir de 2021, les prix connaissent une remontée soutenue, reflétant la reprise économique mondiale et une demande en augmentation.

En 2022, les prix atteignent des sommets historiques, avec un pic à 115,5 € en juin, en raison de l'invasion de l'Ukraine par la Russie et des craintes subséquentes concernant l'approvisionnement en énergie.

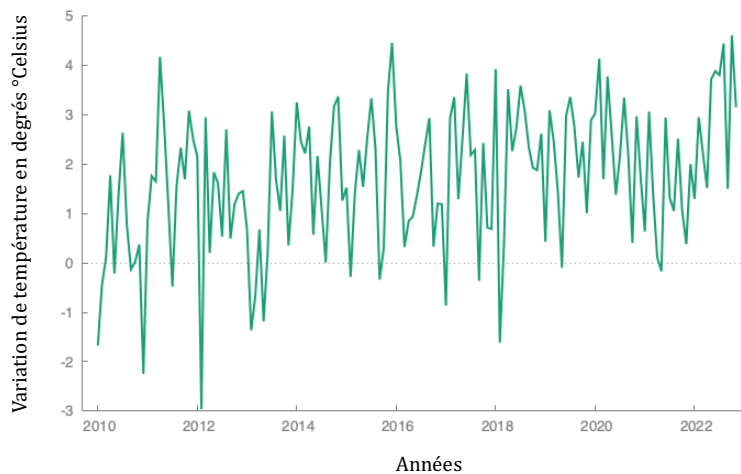
⁸ (Gazzane, 2014)

⁹ (Calignon, 2016)

3.3 Variable explicative 2 : La température

Pour notre deuxième variable explicative, il était intéressant de choisir un indicateur permettant de saisir ou du moins d'approfondir notre compréhension des problématiques climatiques actuelles. Ainsi, nous examinerons les variations de température, exprimées en degrés Celsius (°C), également disponibles via le site de la FAO¹⁰, qui publie des données sur l'évolution moyenne de la température de surface par pays, avec des mises à jour annuelles. Notre période d'analyse couvrira de janvier 2010 à novembre 2022 pour la France, et se concentrera sur les anomalies de la température moyenne mensuelle, c'est-à-dire les variations de température par rapport à une norme climatologique de référence définie sur la période 1951-1980.

Figure 3 - Variation de température en France.
Période 01/2010 au 11/2022



Le graphique des variations de température nous montre des tendances et des anomalies saisonnières qui pourraient avoir un impact significatif sur divers aspects économiques, notamment la consommation d'énergie et l'agriculture.

Au début de notre série, les années 2010 à 2012 se caractérisent par des températures nettement inférieures aux normales saisonnières, avec des anomalies de -1.673 en janvier 2010, de -2.245 en décembre 2010, et un point marquant en février 2012 avec une

¹⁰ (Organisation des Nations Unies pour l'alimentation et l'agriculture, 2024)

température exceptionnellement basse de -2.960, soulignant des hivers particulièrement rigoureux. En contraste, mars 2012 connaît une hausse significative, avec une anomalie de 2.953.

À partir de l'année 2014, les variations négatives deviennent rares, et nous observons des hivers anormalement doux, comme le montre décembre 2015, qui se distingue par une anomalie très élevée de 4.459.

Par la suite, nous avons observé plusieurs étés particulièrement chauds et des variations plus marquées. Notamment, janvier 2018 a connu une augmentation significative de la température, avec une variation de 3.928, rapidement suivie en février par une baisse marquée à -1.613, illustrant ainsi les écarts extrêmes de température d'un mois à l'autre. Les tendances chaudes se poursuivent, et février 2020 se distingue par une variation exceptionnelle de 4.141.

La fin de notre série, l'année 2022, est marquée par des records de température, notamment en août avec une variation de 4.446 et en octobre avec 4.611, soulignant une année exceptionnellement chaude.

4 Présentation de la méthodologie ARIMA

La méthode ARIMA (AutoRegressive Integrated Moving Average) est une technique d'analyse et de prévision des séries temporelles, qui a pour objectif de construire un modèle qui peut être utilisé pour faire des prédictions futures basées sur les données historiques. Elle intègre trois composantes : la partie autorégressive (AR) qui reflète l'influence des observations passées sur les actuelles, la partie de moyenne mobile (MA) qui représente l'impact des erreurs antérieures sur les prédictions futures, et l'intégration (I) qui implique la différenciation des données pour obtenir une série stationnaire, c'est-à-dire jusqu'à ce que les propriétés telles que la moyenne et la variance de la série ne varient pas avec le temps.

Cependant, l'intégration n'est pas toujours nécessaire. Si la série est déjà stationnaire, le composant d'intégration (I) peut être omis.

Ainsi, un modèle ARIMA est caractérisé par trois paramètres : (p, d, q), qui dénotent respectivement, l'ordre de la partie autorégressive, le degré de différenciation nécessaire pour atteindre la stationnarité, et l'ordre de la partie de moyenne mobile. En ajustant ces paramètres, le modèle peut être calibré pour capturer les spécificités de la série temporelle concernée, permettant des prévisions plus précises.

4.1 Stationnarité

La stationnarité d'une série temporelle est une condition essentielle dans l'analyse de données chronologiques. Elle implique que les propriétés statistiques de la série, comme la moyenne et la variance, restent constantes dans le temps. Une série non stationnaire peut présenter des tendances ou des saisons, affectant ainsi sa moyenne et sa variance au fil du temps. Par exemple, les températures quotidiennes subissent une saisonnalité tandis que les tendances économiques à long terme peuvent modifier la moyenne d'une série liée au PIB. Ces caractéristiques soulignent la nécessité d'identifier et de corriger la non-stationnarité.

Pour qu'une série soit stationnaire, elle doit satisfaire les conditions suivantes :

Espérance constante :

- $E[Y_t] = \mu$, où μ est une constante, pour tout t

Variance constante :

- $\text{Var}[Y_t] = \sigma^2$ où σ^2 est une constante, pour tout t

Covariance qui ne dépend que du décalage :

- $\text{Cov}[Y_t, Y_{t+h}] = \gamma(h)$, où $\gamma(h)$ dépend seulement de h et non de t

Si ces conditions ne sont pas respectées, la non-stationnarité peut induire en erreur quant à la persistance et à l'impact des chocs économiques ou d'autres événements.

Pour rendre la série stationnaire, nous pouvons utiliser des transformations telles que le logarithme ou la différenciation, où cette dernière consiste à calculer les différences entre les observations consécutives afin de stabiliser la moyenne. Ces transformations aident à éliminer ou à réduire les effets de la tendance et de la saisonnalité.

Si nous utilisons Y_t pour représenter la série brute, où $t = 1, 2, \dots, N$ et N , est le nombre total d'observations dans la série. Nous pouvons exprimer la première différenciation comme :

$$Z_t = Y_t - Y_{t-1}$$

$$Z_t = \Delta Y_t$$

Cette opération calcule la différence entre chaque observation et sa précédente, résultant en une nouvelle série où les tendances linéaires peuvent être atténuées.

Un indicateur fréquemment utilisé pour tester la stationnarité est le test Augmented Dickey-Fuller (ADF), qui évalue la présence d'une racine unitaire, un signe de non-stationnarité.

Ainsi, s'assurer de la stationnarité d'une série est une étape préliminaire avant d'appliquer des modèles comme ARIMA, car leurs prédictions se basent sur l'hypothèse que les caractéristiques sous-jacentes des séries ne changent pas avec le temps. Ignorer la non-stationnarité peut conduire à des résultats fallacieux.

4.2 Corrélogramme

Le corrélogramme est un graphique qui nous permet d'observer l'autocorrélation pour une série temporelle à différents retards (décalages), permettant d'identifier les modèles et de déterminer les ordres appropriés pour les composantes AR (p) et MA (q) du modèle ARIMA.

Dans le corrélogramme, deux fonctions principales sont représentées : la fonction d'autocorrélation (FAC ou ACF en anglais) et la fonction d'autocorrélation partielle (PACF en anglais ou FAC partielle). La FAC mesure l'autocorrélation entre les observations d'une série

séparées par k intervalles, donnant une vue d'ensemble des corrélations sur l'ensemble des décalages. La FAC partielle, quant à elle, mesure l'autocorrélation entre deux observations avec un décalage k tout en contrôlant l'influence des autres décalages, ce qui aide à identifier le nombre de termes AR dans un modèle ARIMA.

4.2.1 Identification du processus AR(p)

Pour reconnaître visuellement un modèle autorégressif d'ordre p à partir de la FAC partielle (PACF), nous nous attendons à qu'elle affiche des valeurs significatives (au-dessus du seuil de confiance) pour les premiers " p " décalages, et qu'elle diminue brusquement ou coupe après le " p " décalage, indiquant le niveau d'autorégression dans la série.

La forme générale d'un processus AR(p) est donnée par la relation :

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \epsilon_t$$

Où :

- Y_t est la valeur de la série au temps t
- c est une constante (intercept)
- $\phi_1, \phi_2, \dots, \phi_p$ sont les paramètres du modèle (coefficients autorégressifs)
- ϵ_t est un terme d'erreur qui est généralement supposé être du bruit blanc.

Les conditions de stationnarité à vérifier pour un modèle autorégressif d'ordre p sont :

Si le modèle est AR(1), c'est-à-dire $Y_t = c + \phi Y_{t-1} + \epsilon_t$ alors la condition est que la valeur absolue du coefficient autorégressif ϕ doit être inférieure à 1 : $|\phi| < 1$

Dans un modèle AR(2), avec $Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \epsilon_t$ les racines du polynôme caractéristique associé à l'équation doivent être en dehors du cercle unité du plan complexe.

Ces conditions se traduisent par :

- $\phi_1 + \phi_2 < 1$
- $\phi_2 - \phi_1 < 1$
- $|\phi_2| < 1$

4.2.2 Identification du processus MA(q)

Pour déterminer l'ordre q d'un processus de moyenne mobile (MA(q)), nous nous concentrons sur la fonction d'autocorrélation (FAC), elle présente des pics significatifs jusqu'au q -ième décalage et tombe rapidement en dessous du seuil de confiance pour les décalages supérieurs à q . Cela indique que les termes d'erreur sont corrélés avec leurs valeurs passées jusqu'à q périodes précédentes, après quoi ils ne présentent plus de corrélation significative.

La forme générale d'un modèle MA(q) est donnée par :

$$Y_t = \mu + \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - \dots - \theta_q \epsilon_{t-q}$$

Où :

- Y_t est la valeur la série au temps t
- μ la moyenne de la série
- $\theta_1, \theta_2, \dots, \theta_q$ sont les coefficients de la moyenne mobile qui mesurent l'impact des termes d'erreur des périodes précédentes sur la valeur actuelle
- ϵ_t représente le bruit blanc au temps t , c'est-à-dire une séquence de variables aléatoires indépendantes et identiquement distribuées avec une moyenne de zéro et une variance constante.
- q est l'ordre de la moyenne mobile, indiquant le nombre de termes d'erreur passés inclus dans le modèle

Les conditions d'inversibilité à respecter sont :

Si $q=1$ $\theta_1 < 1$

Si $q=2$ $|\theta_2| < 1; \theta_1 + \theta_2 < 1; \theta_2 - \theta_1 < 1$

4.3 Construction du modèle de Box Jenkins

Le modèle de Box-Jenkins est une technique pour prévoir les séries temporelles en ajustant des modèles ARIMA, en se basant sur les étapes suivantes :

Identification : La première étape consiste à déterminer si la série temporelle est stationnaire. Si ce n'est pas le cas, nous procédons à la différenciation de la série jusqu'à ce qu'elle le devienne. Ensuite, à l'aide des fonctions FAC et FAC partielle, nous identifions les ordres potentiels pour les composantes AR et MA. Ce sont les termes p et q dans la notation ARIMA(p,d,q), où d est le nombre de différenciations appliqué.

Estimation : Cet étape implique l'application du modèle précédemment identifié pour calculer et évaluer la significativité de ses paramètres. Au cours de cette phase, une attention particulière est accordée à la stationnarité des coefficients autorégressifs (AR) et à l'inversibilité des paramètres de moyenne mobile (MA), garantissant que le modèle est conforme aux hypothèses. Si ces critères ne sont pas remplis, nous devons réitérer le processus d'identification afin de revoir et ajuster les paramètres du modèle.

Vérification : Les résidus du modèle ajusté sont ensuite analysés pour vérifier qu'ils ressemblent à du bruit blanc. Cela signifie qu'ils doivent être indépendants les uns des autres et normalement distribués avec une moyenne de zéro. Des tests comme celui de Ljung-Box peuvent être utilisés pour tester l'autocorrélation des résidus.

Prévision : Si le modèle est adéquat, nous pouvons l'utiliser pour faire des prédictions sur les valeurs futures de la série. Si les prédictions ne sont pas satisfaisantes ou si de nouvelles

données sont disponibles, nous pouvons retourner aux étapes précédentes pour affiner le modèle. D'ailleurs, chaque étape de la méthode Box-Jenkins est itérative.

Cette approche souligne l'importance de la parcimonie : un modèle avec moins de paramètres est préférable tant qu'il capture la dynamique sous-jacente de la série temporelle. Cela permet de réduire le risque de surajustement et de fournir des prévisions plus robustes.

4.4 Modèles Multiplicatifs

Un modèle multiplicatif ARIMA, souvent désigné sous le terme SARIMA (Seasonal ARIMA), est une extension du modèle ARIMA classique conçu pour mieux gérer les séries temporelles présentant des motifs saisonniers en plus des tendances non saisonnières. Alors que le modèle ARIMA classique est exprimé en termes de composantes autorégressives (AR), intégrées (I) et de moyenne mobile (MA), le modèle multiplicatif introduit des composantes saisonnières qui fonctionnent de manière similaire mais sur des cycles saisonniers.

La notation est $\text{ARIMA}(p,d,q)(P,D,Q)_s$, où :

- p, d, q sont les ordres du modèle ARIMA non saisonnier (comme dans ARIMA classique)
- P, D, Q sont les ordres des composantes saisonnières AR, I, et MA respectivement
- s est la périodicité de la saisonnalité (par exemple, 12 pour des données mensuelles avec une saisonnalité annuelle)

Permettant de modéliser à la fois la structure de dépendance dans les données (à travers les composantes AR et MA) et la différenciation nécessaire pour rendre la série stationnaire (composante I), tout en intégrant ces mêmes aspects dans un contexte saisonnier.

Les composantes saisonnières sont multiplicatives dans le sens où elles interagissent avec les composantes non saisonnières pour modéliser les effets saisonniers qui varient en fonction du niveau de la série temporelle.

Cette approche est particulièrement utile pour les séries où l'effet saisonnier s'intensifie ou diminue en fonction de la tendance de la série, offrant une modélisation plus flexible et précise des phénomènes saisonniers.

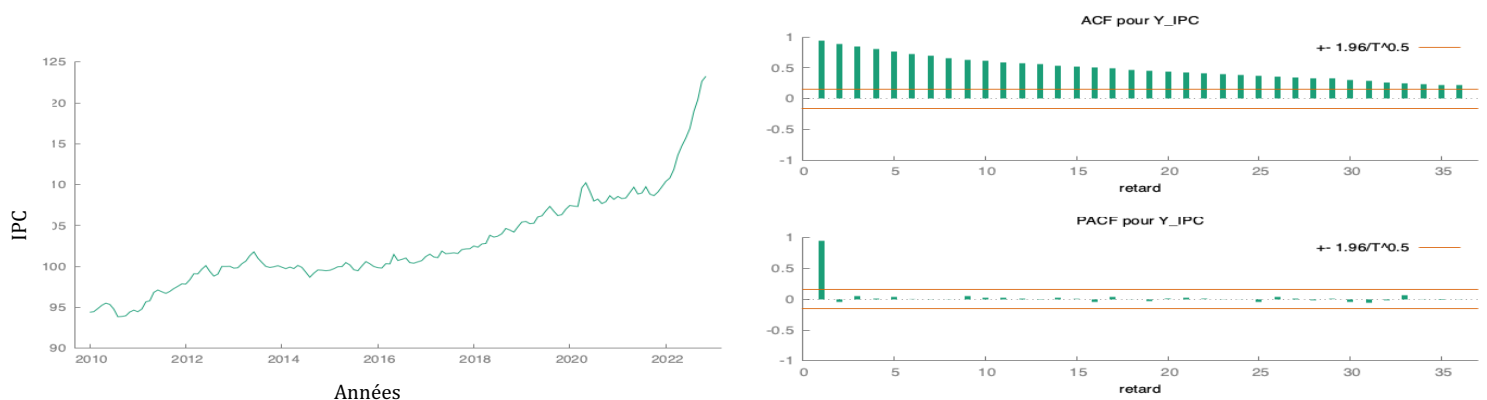
5 Choix du modèle ARIMA pour Y_t

5.1 Stationnarité

Avant d'appliquer l'approche ARIMA, nous devons d'abord, comme nous l'avons mentionné précédemment, confirmer la stationnarité de la série temporelle analysée. Cette vérification est déterminante pour la validité des conclusions que nous tirons de ce modèle. Il a été précédemment établi qu'il y a deux stratégies pour stabiliser une série temporelle : ajuster la variance par des méthodes de transformation telles que l'application de logarithmes, ou réguler la moyenne en procédant à une différenciation.

Pour initier notre analyse, il est nécessaire d'observer graphiquement la série brute afin de déterminer sa stationnarité. La figure 4 nous permet d'évaluer les différentes caractéristiques de la série, notamment les tendances et la variabilité.

Figure 4 - Série brute Y : IPC et son corrélogramme



Pour qu'une série soit considérée comme stationnaire, elle doit présenter une moyenne et une variance constantes. Cependant nous pouvons observer que la série montre une tendance claire à la hausse, ce qui suggère que la moyenne n'est pas constante dans le temps ; cela constitue un problème de stationnarité en moyenne.

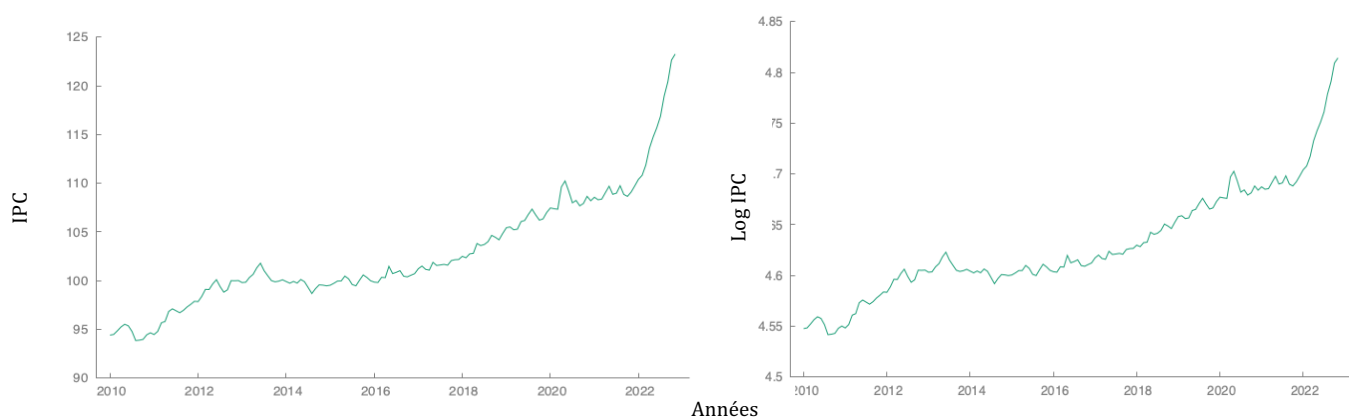
De plus, bien qu'il soit moins évident d'examiner la variance, l'augmentation significative observée vers la fin de la période 2022 suggère également une variance non constante, indiquant ainsi un potentiel problème de non-stationnarité en termes de variance.

Poursuivant l'analyse, l'examen du corrélogramme à 36 retards nous permet de confirmer la non-stationnarité de la série brute. En effet, nous observons un déclin lent et régulier des retards de la fonction d'autocorrélation (la FAC ne tombe pas rapidement à zéro). Et la 1ère valeur de la PACF proche de 1 indique la présence potentielle d'une racine unitaire et donc une non-stationnarité.

Pour rendre la série stationnaire, deux approches peuvent être envisagées : la transformation logarithmique et la différenciation.

La Transformation logarithmique : D'abord, il faut traiter le problème potentiel de variance que nous avons observé précédemment, pour cela, nous allons procéder à une transformation logarithmique de la série temporelle brute. Cette transformation est efficace pour atténuer les effets de la variance, car elle réduit l'ampleur des fluctuations des valeurs élevées de la série.

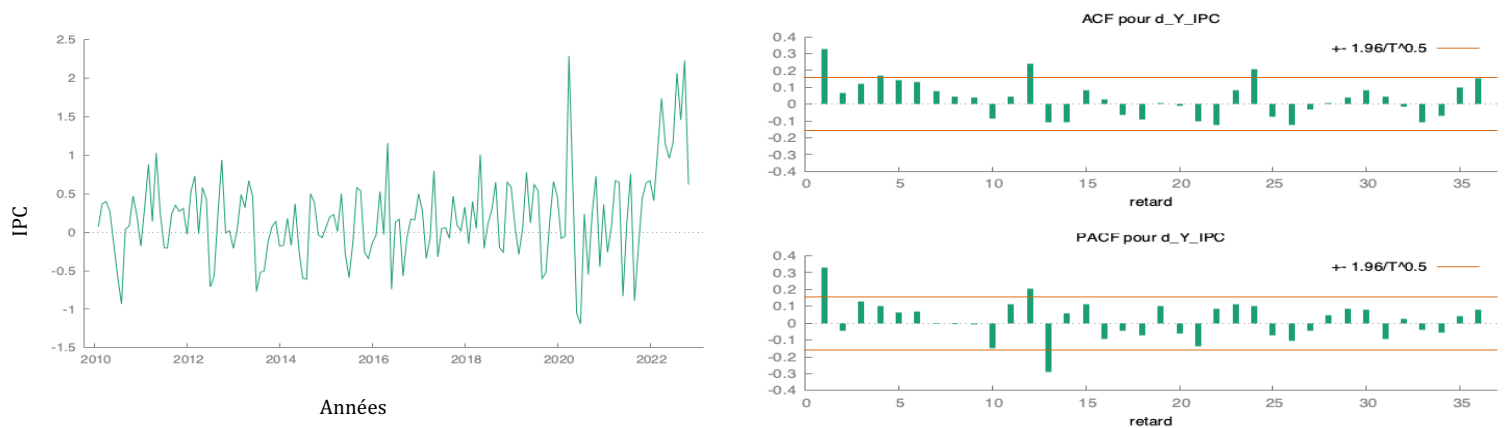
Figure 5 - Comparaison série brute et logarithmique



Cependant, le comparatif présenté dans la figure 5 révèle que la transformation logarithmique n'induit pas de changement significatif pour la série. Par conséquent, nous continuerons l'analyse sans appliquer cette transformation.

Différenciation : Deux types de différenciation sont envisageables : la simple et la saisonnière. Nous commencerons par la première, qui a pour objectif d'éliminer les tendances linéaires en calculant les différences entre chaque valeur de la série et sa précédente, à l'exception de la toute première valeur. Ainsi, la série modifiée aura une valeur de moins que la série initiale. Cette démarche simplifie l'identification et la correction des fluctuations saisonnières qui pourraient subsister, et nous aidera à évaluer si cette technique seule peut garantir la stationnarité de la série.

Figure 6 -Série Y différenciation d=1 et son corrélogramme



Visuellement, la série différenciée apparaît comme étant plus stable, les valeurs d'autocorrélation diminuent rapidement après le premier décalage et se situent majoritairement à l'intérieur de la bande de confiance.

5.2 Identification ARIMA (p, d, q), (P, D, Q) et estimation

Dans la phase d'identification, nous commencerons avec un modèle initial établi en fonction du corrélogramme de la série différenciée figure 5. Ensuite, nous entamerons un processus itératif en tenant compte de divers paramètres, tels que l'analyse des corrélogrammes des résidus, la significativité des paramètres ϕ (phi) et θ (thêta), ainsi que les statistique Q. Ces observations seront essentielles pour élaborer un tableau récapitulatif des modèles, facilitant ainsi la sélection du modèle le plus adéquat.

Premier modèle :

L'analyse de la FAC partielle révèle un premier retard significatif, suggérant un modèle AR (1). De même, la FAC indique un premier retard significatif, ce qui implique un terme MA (1). Dès lors, un modèle ARIMA (1,1,1) devient un choix initial pertinent pour notre série temporelle.

Analyse 1^{er} modèle :

Les paramètres ϕ (phi) et θ (thêta) ne sont pas significatifs, ce modèle ne sera pas retenu.

Dans cette situation, il serait judicieux de reconsidérer la complexité du modèle et de le simplifier en testant les termes AR et MA indépendamment.

Analyse des modèles 2 et 3 MA (1) et AR (1) :

Une fois testés les modèles MA (1) et AR (1), bien qu'on obtienne le paramètre θ dans MA (1) soit statistiquement significatif, l'analyse du corrélogramme des résidus révèle une statistique Q inférieure à 0,05, indiquant une autocorrélation résiduelle significative. Cette observation est également vraie pour le modèle AR (1). Par conséquent, ces modèles sont insuffisants, car ils violent l'hypothèse fondamentale de l'indépendance des résidus, qui doivent être aléatoires et ne présenter aucune structure autocorrélative (bruit blanc). Cette autocorrélation des résidus suggère que ces modèles ne captent pas entièrement la

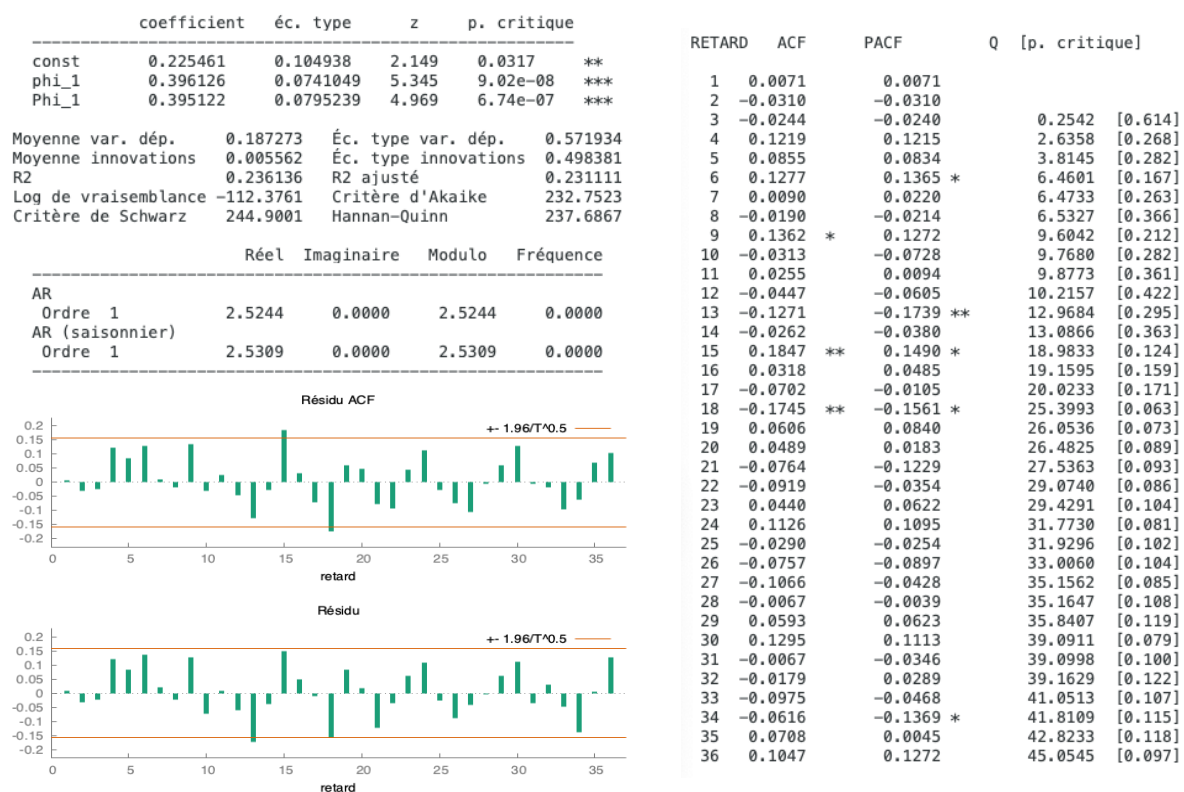
dépendance dans les données, impliquant potentiellement l'inclusion d'autres termes AR ou MA.

Si nous examinons de nouveau le corrélogramme de la série différenciée, nous détectons également des pics notables dans la FAC et FAC partielle au douzième retard, révélant ainsi une saisonnalité annuelle. Il est donc possible que la série temporelle présente une saisonnalité qui n'a pas été prise en compte, pour vérifier nous allons tester le modèle ARIMA avec une partie classique et une partie saisonnière (1, 1, 0) (1, 0, 0).

Résultat :

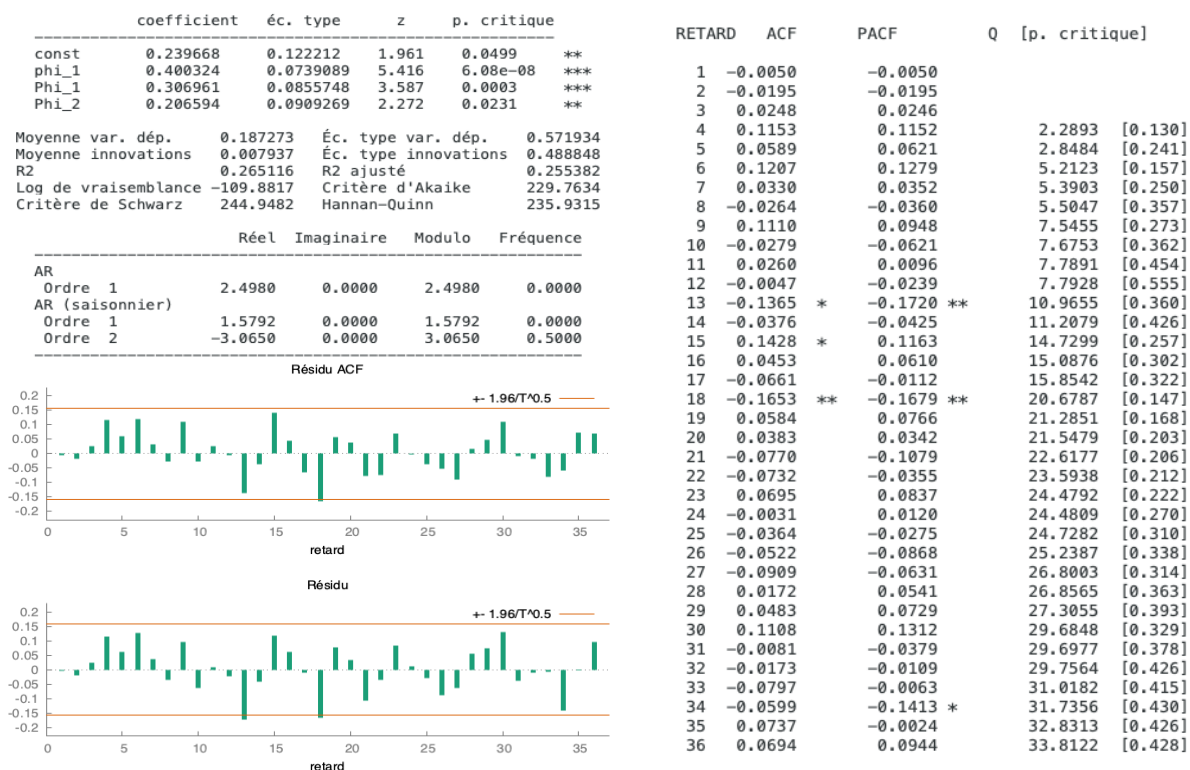
Les deux paramètres phi_1 et Phi_1 sont significatifs. De plus, l'examen des résidus indique que la p critique est supérieure à 0,05, comme illustré dans la figure 7. Ainsi, nous disposons d'un premier modèle qui répond aux attentes fixées.

Figure 7 -Modèle (1, 1, 0) (1,0,0)



Dans le corrélogramme de résidus nous observons encore la présence d'une autocorrélation partielle significative au retard saisonnier de 13, si bien ce retard n'est pas immédiatement adjacent au retard saisonnier principal de 12, il est proche, ce qui peut indiquer une interaction saisonnière plus complexe qui ne serait pas entièrement capturée par un seul terme SAR (1). En ajoutant un SAR (2), le modèle peut potentiellement fournir des prévisions plus précises, en tenant compte de plus de la structure saisonnière inhérente aux données (voir figure 8).

Figure 8 - Modèle (1, 1, 0) (2, 0, 0)



Étant donné que notre série est composée d'une partie classique et une saisonnière, il est également pertinent de tester les modèles ARIMA(0,1,1)(0,0,1) et ARIMA(0,1,1)(0,0,2), nous permettent d'explorer l'effet des termes moyens mobiles (MA) sur la série, en complément des termes autorégressifs (AR) déjà explorés. (Voir figure 9 et 10)

Figure 9 Modèle (0,1,1) (0,0,1)

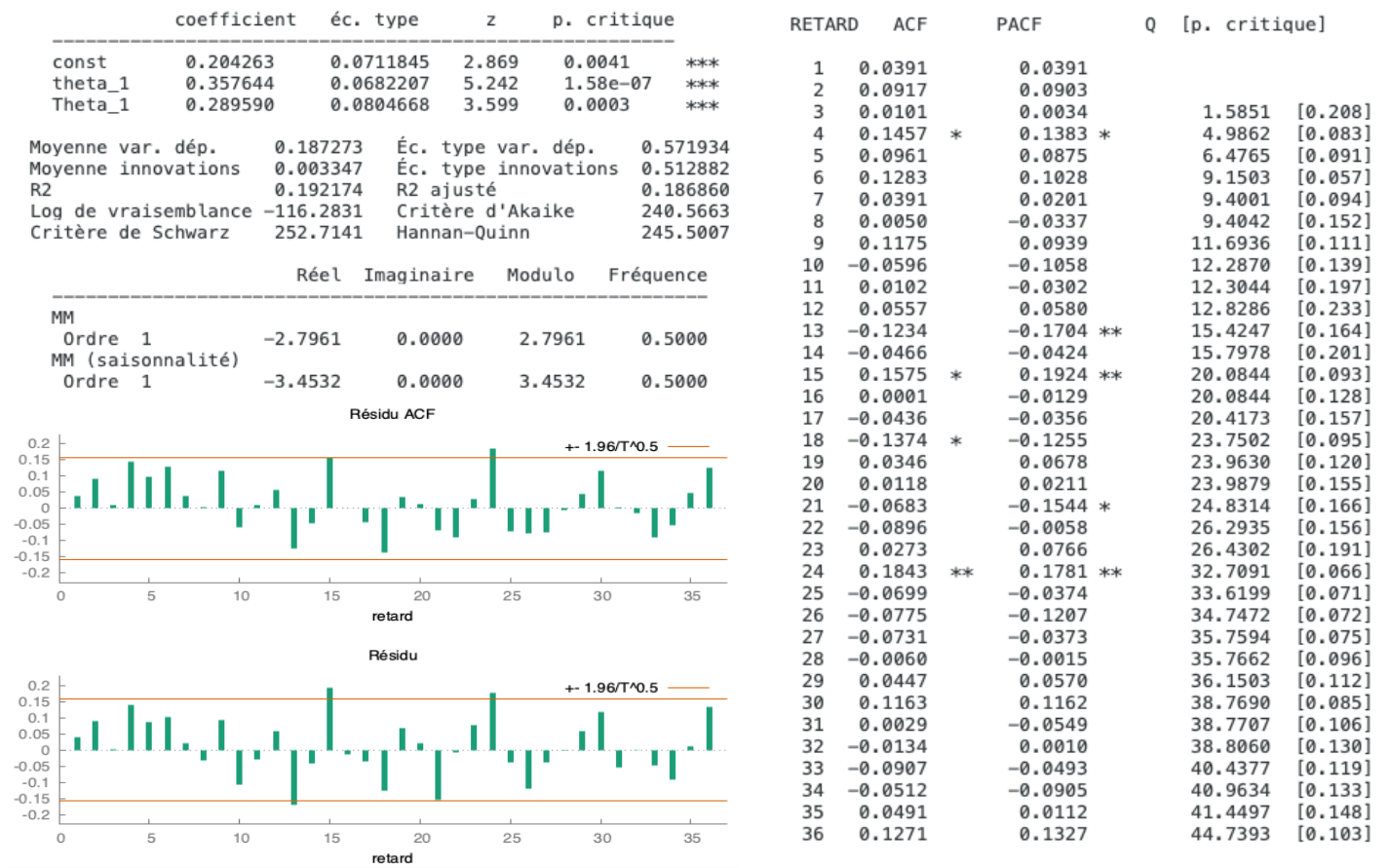
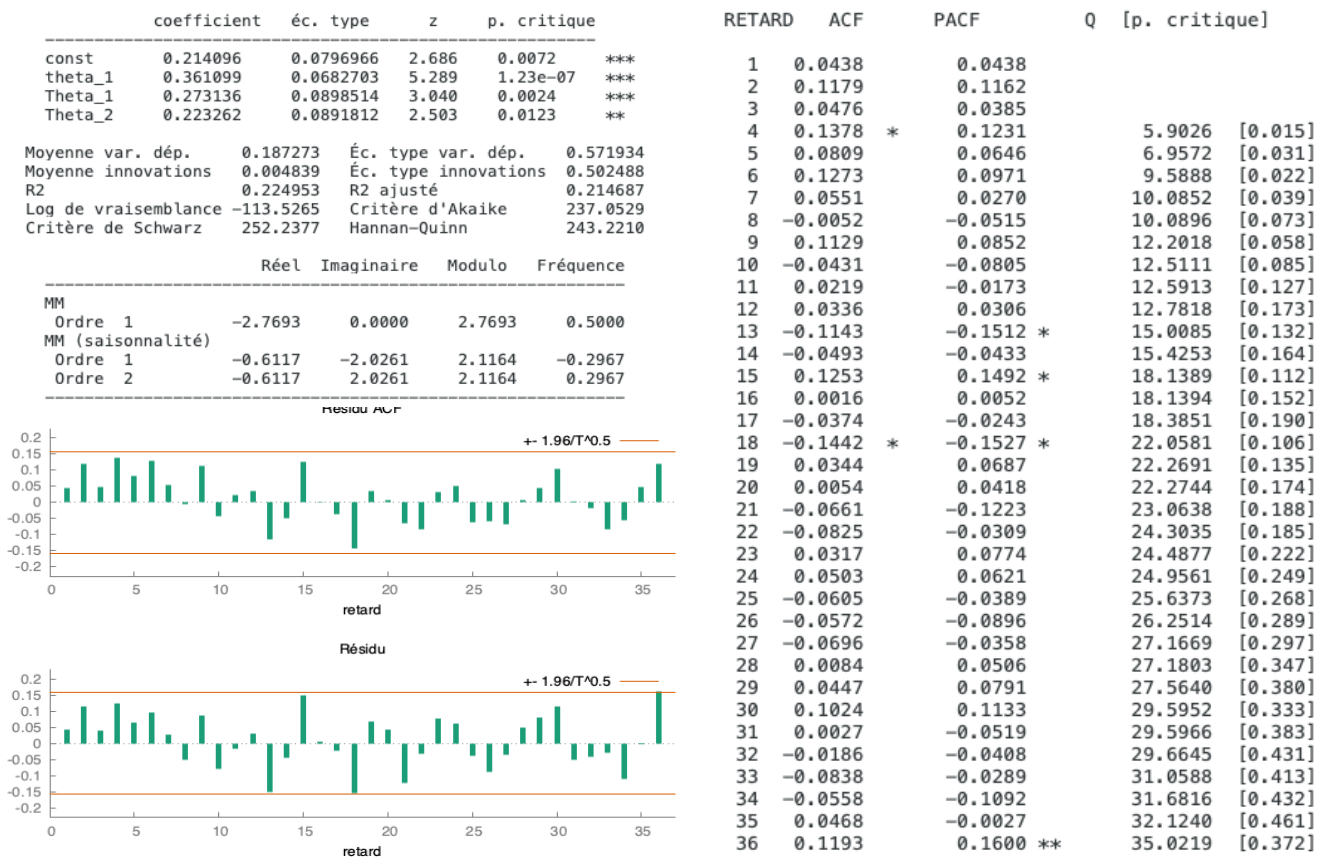


Figure 10 - Modèle (0,1,1) (1,0,2)



5.3 Vérification et choix du modèle

Le tableau 1 présente les différents modèles évalués en termes de leur qualité selon différents critères tels que la statistique Q, l'information de critère d'Akaike (AIC), l'écart-type, le coefficient de détermination (R2) et les prévisions.

Tableau 1 Comparaison de modèles

Modèle	Paramètres	Q stat (k=36)	AIC	Écart-type	R2	Zt	% d'erreur
(1,1,0) (1, 0,0)	2	0,09	232,75	0,49	0,23	0,51	0,07%
(1,1,0) (2, 0,0)	3	0,42	229,76	0,48	0,26	0,31	0,23%
(0,1,1) (0,0,1)	2	0,10	240,56	0,51	0,19	0,30	0,24%
(0,1,1) (0,0,2)	3	0,37	237,05	0,50	0,22	0,06	0,43%

Comme nous pouvons le constater dans le tableau de comparaison 1, les 4 modèles présentent des valeurs p associées à la statistique Q de Ljung-Box (souvent décrite comme un type de test 'portemanteau'¹¹) qui dépassent 0,05, ce qui indique l'absence d'autocorrélation significative dans les résidus du modèle ajusté, jusqu'à un certain nombre de décalages, ici $k=36$.

L'hypothèse nulle (H_0) de ce test est que les résidus sont indépendants, c'est-à-dire qu'il n'y a pas d'autocorrélation significative à aucun des décalages jusqu'à k . Cela suggère que nos modèles ont efficacement capturé la structure de dépendance dans les données, et que les résidus se comportent comme du 'bruit blanc'. En tenant compte de ce critère, et en évaluant également d'autres indicateurs tels que l'AIC, le modèle $(1,1,0)(2,0,0)$ qui présente une statistique Q de 0,42, pourrait être considéré comme le meilleur modèle.

Cependant, d'après les prévisions réalisées avec le logiciel gretl, il apparaît que le modèle affichant le taux d'erreur le plus faible n'est pas celui initialement décrit, mais plutôt le modèle $(1,1,0)(1,0,0)$.

Pour calculer le pourcentage d'erreur de nos prévisions, nous utilisons la vraie valeur de 123,86, correspondant au mois de décembre 2022, que nous avons réservée à des fins de comparaison.

Puisque notre modèle implique une différenciation d'ordre 1, la première étape consiste à réintégrer cette différence pour obtenir la prévision au niveau original de la série. Cela se fait en ajoutant la dernière observation 123,26 (novembre 2022) de la série originale, qui précède immédiatement la période de prévision, à notre prévision différenciée. Cette étape nous permet de transformer la prévision différenciée en une valeur prévue au même niveau que les données originales.

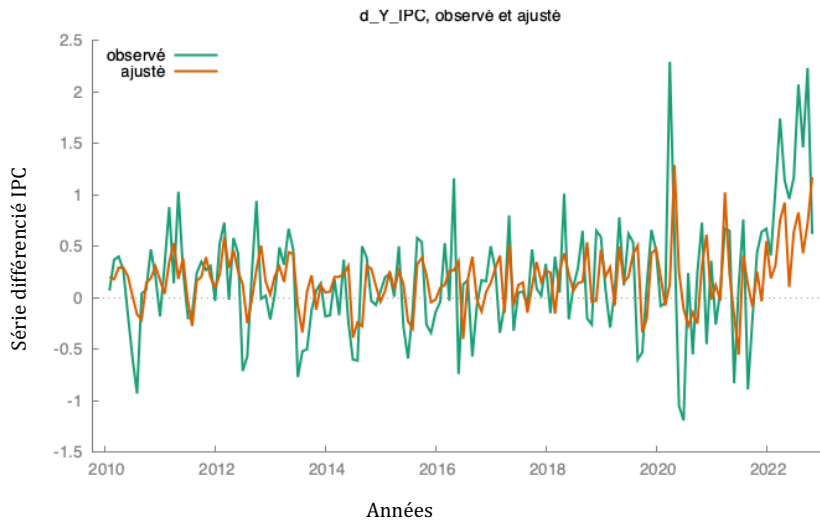
Ensuite, nous calculons l'erreur absolue, qui est la valeur absolue de la différence entre notre prévision réintégrée et la vraie valeur observée de 123,86. Pour normaliser cette erreur et la convertir en pourcentage, nous divisons l'erreur absolue par la vraie valeur, puis multiplions

¹¹ (Souktani, 2003)

le résultat par 100%. Ce processus nous donne le pourcentage d'erreur de la prévision, nous permettant ainsi d'évaluer l'exactitude de notre modèle par rapport à l'échelle de notre série.

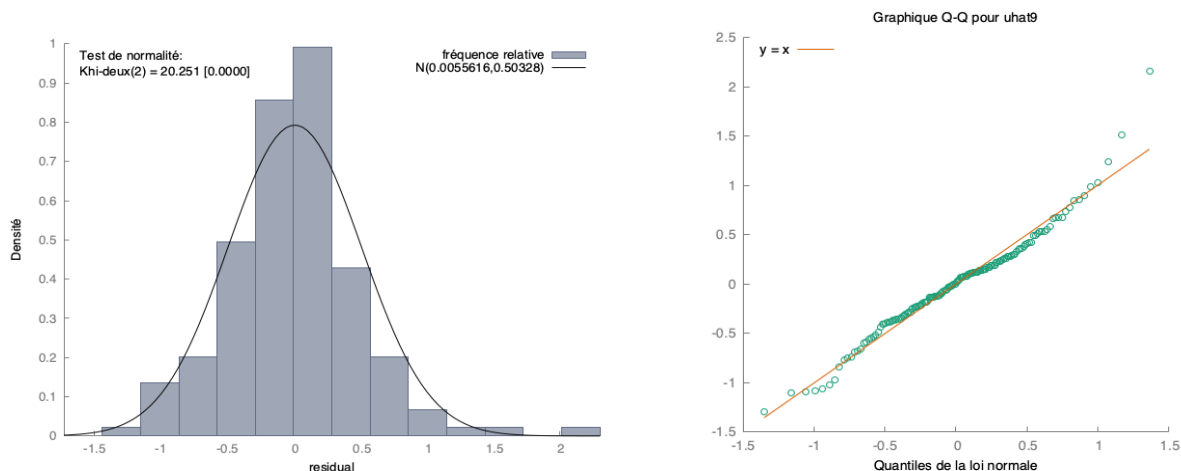
Nous pouvons observer la pertinence du modèle sur la figure 11 :

Figure 11 - Valeurs observées et prédites



Comme nous pouvons l'observer dans l'histogramme (figure 12) la distribution semble asymétrique, suggérant que les résidus de notre série ne suivent pas une distribution normale. Le test de normalité basé sur le test du Khi-2, nous indique une p value très faible de 0.00004 (voir annexe 1), nous conduisant à rejeter l'hypothèse que les résidus suivent une distribution normale. Cela est également le cas pour les quatre modèles proposés.

Figure 12 - Normalité des résidus



5.4 Prévision à une période

Pour notre modèle ARIMA (1,1,0) (1, 0, 0) la forme générale en termes de composantes serait :

$$(1-\phi_1B)(1-B)Y_t = (1-\Phi_1B^{12})\epsilon_t$$

Où :

- B est l'opérateur de retard
- ϕ_1 est le coefficient du terme autorégressif non saisonnier
- Φ_1 est le coefficient du terme autorégressif saisonnier
- ϵ_t est le terme d'erreur au temps t

Pour calculer la prévision en différence ΔY_{t+1} nous avons besoin des valeurs historiques et aussi des coefficients du modèles :

Pour le terme autorégressif non saisonnier (Y_t, Y_{t-1})

- $Y_t = 123,26$ (la valeur la plus récente, pour novembre 2022)
- $Y_{t-1} = 122.64$ (la valeur du mois précédent, octobre 2022)

Pour le terme autorégressif saisonnier (Y_{t-12}, Y_{t-13})

- $Y_{t-12} = 109.10$ (la valeur d'il y a un an à la même période)
- $Y_{t-13} = 108.66$ (la valeur du mois précédent l'année dernière)

Les coefficients du Modèle

- ϕ_1 (AR non saisonnier) = 0.396126
- Φ_1 (AR saisonnier) = 0.395122
- Constante = 0.225461

Les différences premières pour le terme autorégressif non saisonnier :

$$\Delta Y_t = Y_{2022:11} - Y_{2022:10} = 0,62$$

Et saisonnier : 0,44

Et on applique les coefficients sur la formule :

$$\Delta Y_{t+1} = \phi_1 \cdot \Delta Y_t + \Phi_1 \cdot \Delta Y_{t-12} + c$$

$$= 0,396126 \cdot 0,62 + 0,395122 \cdot 0,44 + 0,225461$$

$$\Delta Y_{t+1} = 0,645$$

Soit une prévision de 123.91 pour décembre 2022

Pour rappel : $Y_{t+1} = 123,86$ (la vraie valeur pour décembre 2022)

Et donc l'erreur est de 0,04%

Comme nous l'avons vu, les erreurs de prévision sont très faibles, tant pour le calcul « manuel » que pour celui effectué avec gretl, ce qui témoigne de l'efficacité de la méthodologie utilisée pour ajuster le modèle. Cette précision dans la prévision des valeurs futures suggère que le modèle sélectionné saisit correctement la dynamique de la série temporelle.

6 Analyse des variables pétrole et température

Poursuivant notre analyse, nous aborderons dans cette section les variables explicatives en utilisant la méthodologie de Box-Jenkins, qui a déjà été mise en œuvre dans la section précédente. Afin de minimiser les répétitions, l'analyse dans cette partie sera plus directe.

6.1 Stationnarité

Les corrélogrammes des séries brutes des variables explicatives, présentés dans les annexes 3 et 4, permettent d'établir rapidement que la variable pétrole n'est pas stationnaire. En effet, nous constatons que les autocorrélations représentées par la FAC demeurent significatives et décroissent lentement, ce qui est indicatif d'une non-stationnarité. Pour cette série

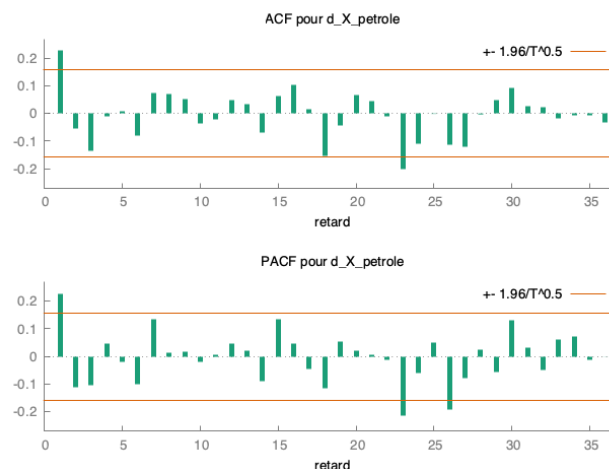
l'application de logs n'est pas pertinente, et la série devient stationnaire grâce à une différenciation.

En examinant la série des températures, notre interprétation graphique se complique. Le corrélogramme pourrait suggérer une stationnarité à première vue ; toutefois, la présence répétée de significativités à des retards plus conséquents (notamment aux retards 1, 3, 7, 12, 14, 28, 29, 30, et 36) dans la fonction d'autocorrélation (FAC) et la tendance haussière mise en évidence dans la section « Analyse des variables » (figure 3) nous indiquent une probable non-stationnarité. Pour ces raisons, nous optons pour une différenciation, qui nous permettra d'éliminer cette tendance et potentiellement de stabiliser la variance, améliorant ainsi la fiabilité de nos analyses.

6.2 Pétrole

Le corrélogramme présenté dans la figure 13, après application d'une différenciation, confirme la transformation de la série en une série stationnaire. On remarque que les autocorrélations chutent rapidement au sein de la bande de confiance juste après le premier décalage dans la FAC, suggérant l'absence de corrélations à long terme au sein de la série différenciée. Quant à la fonction d'autocorrélation partielle FACP, un pic notable apparaît au premier décalage, puis, pour les décalages suivants, les valeurs deviennent majoritairement non significatives, indiquant une influence limitée des termes passés sur les valeurs futures au-delà du premier décalage.

Figure 13 Corrélogramme de la série pétrole différencié



Sur la base du corrélogramme, nous avons initialement envisagé un modèle ARIMA (1,1,1), mais comme aucun paramètre n'est significatif (comme il est indiqué dans l'annexe 4), et que l'exclusion de la constante ne change pas les résultats, nous avons décidé de proposer un modèle ARIMA (1,1,0) à la place. Ce modèle comprend un terme autorégressif de premier ordre avec une différenciation d'ordre un. Avec ce modèle, le paramètre phi s'avère significatif. Compte tenu de l'insignifiance de la constante dans notre modèle, nous avons procédé à un nouvel ajustement en l'omettant, et nous obtenons les résultats suivants :

Figure 14 - Résultats d_Pétrole modèle ARIMA(1,1,0)



Nous avons donc un modèle qui respecte la significativité de paramètres et une statistique $Q > 0,05$

Tableau 2- Modèle ARIMA (1,1,0) pour Pétrole

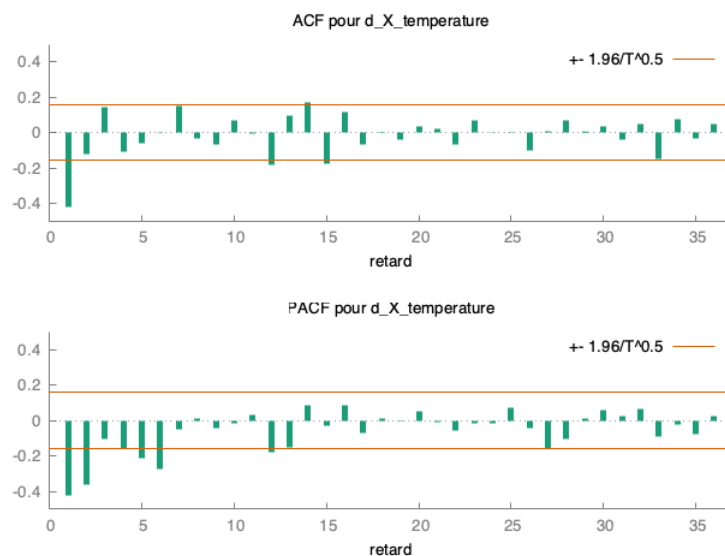
Modèle	Paramètres	Q stat (k= 36)	AIC	Écart-type	R2	Zt Prévision	% d'erreur
(1, 1, 0)	1	0.481	945.837	5.148	0.051	-1.2	15,68%

Concernant le calcul du pourcentage d'erreur, nous procédons comme suit : la dernière observation enregistrée pour le mois de novembre 2022 est de 89,7. Après différenciation, cela nous donne 89,7 moins 1,2, soit 88,5. Étant donné que la valeur réelle pour le mois de décembre 2022 est de 76,5, nous obtenons un pourcentage d'erreur de 15,68%.

6.3 Température

Comme mentionné précédemment, nous avons procédé à une différenciation de la série des températures dans le but de stabiliser sa tendance croissante et d'assurer la stationnarité de notre série, sans recourir à des tests formels.

Figure 15 - Corrélogramme série d_température



Sur la base du corrélogramme et des graphiques vu lors du cours de séries temporelles ¹², un modèle initial à envisager pourrait être un modèle ARIMA combinant à la fois de composantes AR et MA, mais aussi la forme observée suggère un modèle mixte, c'est-à-dire avec une partie saisonnière.

Nous avons une valeur significative à la fois dans la FAC et la FAC partielle au premier retard (-0.4206, significatif à 1%), suggérant comme nous l'avons dit une partie AR et une partie MA, mais aussi nous constatons plusieurs retards significatifs à 1% dans FAC partielle.

Ces observations nous conduisent à la proposition d'un modèle mixte ARIMA (0,1,1)(1,0,0) qui intègre à la fois les effets de moyenne mobile immédiats et la saisonnalité, pour une représentation plus complète des dynamiques de la série temporelle. Et on obtient :

Figure 16 - Série d_température, modèle ARIMA(0,1,1)(1,0,0)



¹² (Girard, 2023 -2024)

Tableau 3- Modèle ARIMA (1,1,0) pour *Température*

Modèle	Paramètres	Q stat (k=36)	AIC	Écart-type	R2	Zt	% d'erreur
ARIMA (0,1,1)(1,0,0)	2	0.429	535.166	1.336	0.429	-0.687	43%

Pour calculer le terme d'erreur, nous prenons comme référence la dernière valeur observée pour 2022, celle de novembre, qui est de 3,155. En soustrayant à cette valeur le Z_t calculé par Gretl, qui est de $-0,687$, nous obtenons 2,468. La valeur réelle pour décembre 2022 est cependant de 1,715. Ainsi, nous constatons un pourcentage d'erreur de 43%.

7 Test de cointégration selon Engle-Granger

Le concept de cointégration repose sur l'idée que certaines séries temporelles, bien qu'elles soient individuellement non stationnaires (c'est-à-dire qu'elles présentent des tendances ou des variations saisonnières persistantes), peuvent être combinées d'une manière qui neutralise ces non-stationnarités, révélant ainsi une relation stable à long terme entre elles. Le test de cointégration d'Engle-Granger est une approche pour identifier de telles relations, offrant un moyen de distinguer entre les régressions fallacieuses et les relations significatives.

Lorsque deux séries, comme X_t et Y_t , sont intégrées d'ordre 1 $I(1)$ mais que leur combinaison linéaire, $Y_t - aX_t - b$, est stationnaire $I(0)$, elles sont dites cointégrées, indiquant une relation stable à long terme malgré leur non-stationnarité individuelle.

Deux séries X_t et Y_t sont dites cointégrées si les deux conditions suivantes sont vérifiées¹³:

- Elles sont intégrées d'ordre d
- La combinaison linéaire de ces 2 séries permet de se ramener à une série d'ordre d'intégration inférieur

¹³ (Hamisultane, 2016)

7.1 Test de cointégration

Pour nous assurer que nos régressions ne soient pas fallacieuses, nous procéderons au test de cointégration d'Engle-Granger en suivant les étapes ci-après :

Étape 1 : Tester l'ordre d'intégration des deux variables

Nous nous concentrons sur les séries qui sont I (1), c'est-à-dire celles qui sont non stationnaires mais qui deviennent stationnaires après une première différenciation. Il est crucial de vérifier que les séries concernées deviennent effectivement stationnaires après cette différenciation. Pour cela, nous appliquerons le test de Dickey-Fuller augmenté (ADF), qui confronte l'hypothèse nulle de la présence d'une racine unitaire (indiquant la non-stationnarité) à l'hypothèse alternative selon laquelle la série est stationnaire (donc sans racine unitaire) :

Résultats des Tests ADF

Figure 17 - Test ADF série IPC différenciée

```
Test de Dickey-Fuller augmenté pour d_Y_IPC
test à reculons à partir de 4 retards, suivant le critère AIC
taille de l'échantillon 153
hypothèse nulle de racine unitaire : a = 1

test avec constante
avec 0 retards de (1-L)d_Y_IPC
modèle: (1-L)y = b0 + (a-1)*y(-1) + e
valeur estimée de (a - 1): -0.669296
statistique de test: tau_c(1) = -8.6977
p. critique asymptotique 4.067e-15
Coeff. d'autocorrélation du 1er ordre pour e: 0.014

avec constante et tendance temporelle
avec 0 retards de (1-L)d_Y_IPC
modèle: (1-L)y = b0 + b1*t + (a-1)*y(-1) + e
valeur estimée de (a - 1): -0.711198
statistique de test: tau_ct(1) = -9.09562
p. critique asymptotique 2.712e-16
Coeff. d'autocorrélation du 1er ordre pour e: 0.019
```

Figure 18 - Test ADF série Température différenciée

```
Test de
test à reculons à partir de 6 retards, suivant le critère AIC
taille de l'échantillon 148
hypothèse nulle de racine unitaire : a = 1

test avec constante
avec 5 retards de (1-L)d_X_temperature
modèle: (1-L)y = b0 + (a-1)*y(-1) + ... + e
valeur estimée de (a - 1): -4.07509
statistique de test: tau_c(1) = -10.181
p. critique asymptotique 8.696e-20
Coeff. d'autocorrélation du 1er ordre pour e: -0.029
différences retardées: F(5, 141) = 11.701 [0.0000]

avec constante et tendance temporelle
avec 5 retards de (1-L)d_X_temperature
modèle: (1-L)y = b0 + b1*t + (a-1)*y(-1) + ... + e
valeur estimée de (a - 1): -4.07842
statistique de test: tau_ct(1) = -10.1571
p. critique asymptotique 1.97e-20
Coeff. d'autocorrélation du 1er ordre pour e: -0.030
différences retardées: F(5, 140) = 11.641 [0.0000]
```

Figure 19 - Test ADF série Pétrole différenciée

```
Test de Dickey-Fuller augmenté pour d_X_petrole
test à reculons à partir de 2 retards, suivant le critère AIC
taille de l'échantillon 152
hypothèse nulle de racine unitaire : a = 1

test avec constante
avec un retard de (1-L)d_X_petrole
modèle: (1-L)y = b0 + (a-1)*y(-1) + ... + e
valeur estimée de (a - 1): -0.860875
statistique de test: tau_c(1) = -8.51373
p. critique asymptotique 1.474e-14
Coeff. d'autocorrélation du 1er ordre pour e: -0.016

avec constante et tendance temporelle
avec un retard de (1-L)d_X_petrole
modèle: (1-L)y = b0 + b1*t + (a-1)*y(-1) + ... + e
valeur estimée de (a - 1): -0.863372
statistique de test: tau_ct(1) = -8.50345
p. critique asymptotique 3.744e-14
Coeff. d'autocorrélation du 1er ordre pour e: -0.017
```

Le nombre de retards pour nos tests a été choisi en utilisant la sélection des retards du VAR sur gretl, détail que l'on peut retrouver dans l'annexe 6 « Sélection des retards pour test ADF ». Comme illustré dans les figures 9, 10 et 11, les valeurs des p critiques asymptotiques de nos trois séries sont inférieures à 0,05. Ce résultat confirme ce que nous avons mentionné précédemment : que le test ayant pour hypothèse nulle (H_0) l'existence d'une racine unitaire, nous pouvons rejeter cette hypothèse et confirmer la stationnarité de nos séries.

Étape 2 : établir et à évaluer la relation de long terme

On procède à une régression par les moindres carrés ordinaires (MCO) pour estimer la relation de long terme entre les deux séries $X_t \sim I(1)$ et $Y_t \sim I(1)$.

Cette régression prend la forme $Y_t = aX_t + b + \varepsilon_t$ pour qu'il y ait cointégration, il faut que les résidus e_t issus de la régression, soient stationnaires et $e_t = Y_t - \hat{a}X_t - b \sim I(0)$.

Nous appliquerons à nouveau le test ADF pour vérifier la stationnarité, mais cette fois ci sur les résidus.

Tableau 3 Résultats du test ADF pour les résidus

Séries	p. critique (ADF)
Y (IPC), X (Température)	0.999
Y(IPC), X (Pétrole)	1
X(Température), Y(IPC)	1.945e-22
X(Température), X(Pétrole)	5,098e-19
X(Pétrole), X(Température)	0,267
X(Pétrole), Y(IPC)	0,237

Comme nous pouvons observer dans le tableau 3, les résidus des régressions Y sur X_1 et Y sur X_2 ne sont pas stationnaires, donc il n'existe pas de relation de cointégration entre ces paires de variables selon le test d'Engle et Granger. En conséquence, toute analyse de régression linéaire simple entre ces variables pourrait mener à des conclusions erronées ou à des régressions fallacieuses.

Face à cette situation, nous avons exploré d'autres combinaisons et constaté que les résidus des régressions de X (température) sur Y (IPC), ainsi que de X (température) sur X (pétrole),

sont effectivement stationnaires. Cela indique que ces séries sont cointégrées, ce qui signifie qu'il existe une relation d'équilibre à long terme entre elles.

Concernant les données présentées dans le tableau 3, l'estimation par les Moindres Carrés Ordinaires (MCO) pour chaque paire, la sélection des résidus et les tests de Dickey-Fuller augmenté (ADF) sont disponibles respectivement dans les annexes 7, 8 et 9.

7.2 Modèle à correction d'erreur pour les variables Co intégrées

A partir de nos séries cointégrés $X_1, X_2 \sim CI(1,1)$, nous allons estimer le modèle à correction d'erreur. Pour plus de clarté dans les formules présentées dans cette section, X_1 , qui représente la température, sera considérée comme notre variable dépendante et sera notée Y .

On commence par estimer la relation de long terme

$$Y_t = a X_t + b + \varepsilon_t$$

Ici, le terme ε_t représente l'écart à la relation de long terme au moment t , soit l'erreur de prédiction de Y_t basée sur X_t . En décalant les séries d'une période en arrière, notée (-1) , nous obtenons :

$$Y_{t-1} = a X_{t-1} + b + \varepsilon_{t-1}$$

L'objectif est de calculer l'écart (erreur de cointégration) de la période précédente :

$$\varepsilon_{t-1} = Y_{t-1} - (a X_{t-1} + b)$$

Qui mesure à quel point les variables étaient éloignées de leur équilibre à long terme à ce moment-là.

Le MCE repose donc sur l'idée que, même si les variables peuvent fluctuer à court terme pour diverses raisons, il existe un mécanisme intrinsèque qui les ramène vers une relation

préétablie à long terme, assurant ainsi la stabilité de leur relation malgré ces fluctuations temporelles.

Le modèle intègre les ajustements de court terme via la relation :

$$\Delta Y_t = \gamma \Delta X_t + \delta \varepsilon_{t-1} + \mu_t \text{ avec } \delta < 0$$

(Où Δ représente la différence d'une période à l'autre). Cette intégration permet de saisir non seulement comment les variables reviennent à leur équilibre à long terme après un écart ε_{t-1} , mais aussi comment elles réagissent aux changements immédiats l'une de l'autre ΔX_t .

Le coefficient δ est associé au résidu de long terme décalé d'une période, il doit être négatif et significatif. Si ce n'est pas le cas, le MCE ne serait pas approprié. L'importance du signe négatif pour δ réside dans le fait qu'il symbolise une force de rappel, qui agit pour ramener la variable dépendante vers son équilibre en cas de déséquilibres.

Figure 20 - MCE

Variable dépendante: d_X_temperature				
	coefficient	éc. type	t de Student	p. critique
const	0.0478100	0.110906	0.4311	0.6670
d_Y_IPC	-0.0768293	0.185476	-0.4142	0.6793
residus_temp_I~_1	-0.879579	0.0797714	-11.03	4.14e-21 ***
Moyenne var. dép.	0.031351	Éc. type var. dép.	1.745529	
Somme carrés résidus	257.9449	Éc. type régression	1.306998	
R2	0.446673	R2 ajusté	0.439345	
F(2, 151)	60.94745	P. critique (F)	3.94e-20	
Log de vraisemblance	-258.2326	Critère d'Akaike	522.4653	
Critère de Schwarz	531.5761	Hannan-Quinn	526.1661	
rho	-0.001614	Durbin-Watson	1.997364	

Dans la figure 20, nous nous intéressons uniquement au résidu -1 (δ), c'est-à-dire à notre coefficient associé au terme de correction d'erreur qui est de -0.879. Et il est non seulement significatif, mais aussi négatif, ce qui nous indique que nous sommes face à une régression non fallacieuse et que nous pouvons interpréter nos résultats.

Concernant l'ajustement à court terme, δ nous indique que lorsque notre variable dépendante, ici la température, s'écarte de sa valeur prévue par la relation de cointégration, nous pouvons nous attendre à ce que quasiment 88 % de cet écart soit corrigé en un an, ramenant progressivement la température vers la trajectoire à long terme établie par la cointégration.

Lorsque nous examinons la relation de cointégration à long terme figure 21, nous constatons que la p-value associée à l'IPC est extrêmement faible, soit 5.72e-06, ce qui indique que le coefficient de l'IPC est statistiquement significatif au niveau de confiance de 1%. Par conséquent, nous sommes en mesure de rejeter l'hypothèse nulle selon laquelle l'IPC n'a pas d'effet sur la température.

Figure 21 - MCO Température sur IPC

Modèle 1: MCO, utilisant les observations 2010:01-2022:11 (T = 155)
Variable dépendante: X_temperature

	coefficient	éc. type	t de Student	p. critique	
const	-7.35335	1.92945	-3.811	0.0002	***
Y_IPC	0.0881958	0.0187596	4.701	5.72e-06	***
Moyenne var. dép.	1.703761	Éc. type var. dép.	1.418335		
Somme carrés résidus	270.6929	Éc. type régression	1.330126		
R2	0.126227	R2 ajusté	0.120517		
F(1, 153)	22.10278	P. critique (F)	5.72e-06		
Log de vraisemblance	-263.1464	Critère d'Akaike	530.2927		
Critère de Schwarz	536.3796	Hannan-Quinn	532.7651		
rho	0.126720	Durbin-Watson	1.720312		

Le coefficient pour l'IPC est de 0.088, ce qui implique qu'une augmentation d'une unité de l'IPC entraîne une augmentation de 0.088 unités de la température, toutes choses étant égales par ailleurs. Ce qui signifie que l'augmentation de l'IPC est associée à une hausse de la température.

8 Conclusion

Dans le cadre de ce projet d'analyse des séries temporelles, l'objectif était de comprendre les facteurs influençant les fluctuations de l'Indice des Prix à la Consommation (IPC) des produits alimentaires, notamment en étudiant les interactions entre le prix du pétrole et les variations de température. À cette fin, la méthodologie ARIMA de Box-Jenkins a été adoptée pour prédire les tendances futures.

L'application de la méthode ARIMA à notre série Y_t nous a guidé vers le choix du modèle $ARIMA(1,1,0)(1,0,0)$. Bien que ce modèle n'ait pas montré la valeur la plus proche de 1 selon la statistique Q, il a fourni les meilleures prévisions selon Gretl, avec un pourcentage d'erreur remarquablement faible.

A l'aide du corrélogramme des résidus, nous avons pu confirmer que les résidus se comportent comme un bruit blanc. Toutes les barres du corrélogramme se situent à l'intérieur de la bande de confiance, ce qui indique que les résidus pourraient être indépendants, ce qui peut indiquer une distribution normale. Cependant, le test de normalité des résidus réalisé avec Gretl n'a pas validé cette hypothèse.

Nous avons procédé à la prévision pour une période ultérieure, aboutissant à une estimation pour l'IPC de décembre 2022 à 123,91. La valeur réelle, préalablement mise de côté pour comparaison, était de 123,86, validant ainsi la proximité de nos résultats avec la réalité.

Concernant les variables explicatives, notre démarche n'a pas été exhaustive. Nous avons privilégié l'identification d'un modèle qui satisfait à la fois aux critères de significativité des paramètres et à la qualité de l'ajustement, cette dernière étant évaluée par la statistique Q. Ainsi, pour la série temporelle liée à la température, nous avons opté pour le modèle $ARIMA(0,1,1)(1,0,0)$, tandis que pour la série concernant le prix du pétrole, le choix s'est porté sur le modèle $ARIMA(1,1,0)(0,0,0)$.

L'analyse de trois séries temporelles a révélé que celles-ci devenaient stationnaires après une différenciation d'ordre 1, ce qui a été confirmé par le test de Dickey-Fuller augmenté (ADF).

Néanmoins, la tentative de cointégration entre Y et X1 ou X2 n'a pas abouti comme initialement envisagé, en raison de la non-stationnarité des résidus issus de la régression.

En examinant d'autres options, nous avons découvert que la série de la température par rapport à l'IPC remplissait toutes les conditions nécessaires pour établir une cointégration, indiquant une relation significative et fiable entre l'Indice des Prix à la Consommation (IPC) et la température.

Cela met en évidence l'existence d'un équilibre à long terme entre ces deux variables. Le modèle d'ajustement à court terme révèle que presque 88 % de toute divergence par rapport à cet équilibre est corrigée au cours de l'année suivante, ce qui témoigne d'une forte capacité des variables à retrouver leur relation de cointégration à long terme.

Cette relation s'avère statistiquement significative, avec un coefficient de 0,088 pour l'IPC, ce qui implique qu'une hausse d'une unité de l'IPC est liée à une augmentation de 0,088 unités de la température.

Cela nous amène à considérer l'influence potentielle des fluctuations économiques sur des facteurs environnementaux tels que la température. Par exemple, un lien indirect pourrait exister par le biais de la croissance économique (que l'IPC peut illustrer), susceptible de provoquer une augmentation des émissions de gaz à effet de serre et, en conséquence, d'affecter le climat à long terme.

9 Discussion

Dans cette dernière section, nous abordons les divers défis et limitations rencontrés durant le projet, ainsi que mes réflexions personnelles.

Dans la section consacrée à l'analyse de la stationnarité, où nous avons tenté de déterminer la stationnarité uniquement à partir d'observations graphiques sans avoir recours à des tests formels, nous avons certes stimulé une réflexion critique intéressante. Toutefois, cette approche peut aussi introduire une certaine incertitude dans le choix des modèles. C'est le cas, par exemple, pour la température : dans ce projet, les données s'étendant de 2010 à 2022 révèlent une tendance croissante dans le graphique de la série brute, me conduisant à considérer la série comme non stationnaire. Néanmoins, si j'avais utilisé un échantillon de données plus restreint, je ne suis pas certaine que j'aurais fait le même choix.

J'aurais pu, explorer d'autres améliorations et options de modélisation, notamment pour les variables température et pétrole. Alors que l'erreur liée au modèle du pétrole aurait pu être considérée comme acceptable, celle associée à la température s'est révélée bien plus conséquente, mettant en lumière le besoin d'une réévaluation approfondie pour cette variable.

En ce qui concerne l'analyse de cointégration, les contraintes inhérentes au processus d'établissement d'une relation de cointégration ont restreint nos options. Et la méthode d'Engle et Granger, par exemple, ne permet d'identifier qu'une seule relation de cointégration, soulignant ainsi les limites de notre approche.

Pour finir, il est important de manier avec prudence toute association suggérée entre l'IPC et la température. Car l'interprétation des liens entre des variables économiques et environnementales est complexe et peut souvent dépasser ce que suggèrent les résultats à première vue.

10 Annexes

Annexe 1. Normalité de résidus, série d Y modèle AR(1) SAR(1)

Distribution des fréquences pour residual, obs 2-155
nombre de classes = 13, moyenne 0.00556161, éc. type = 0.503276

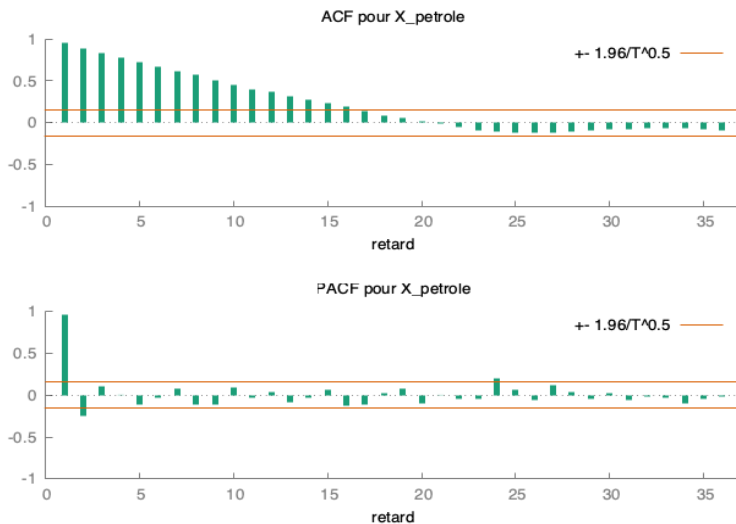
intervalle	centre	fréquence	rel.	cum.
< -1.1553	-1.2993	1	0.65%	0.65%
-1.1553 - -0.86711	-1.0112	6	3.90%	4.55% *
-0.86711 - -0.57897	-0.72304	9	5.84%	10.39% **
-0.57897 - -0.29083	-0.43490	22	14.29%	24.68% *****
-0.29083 - -0.0026932	-0.14676	38	24.68%	49.35% *****
-0.0026932 - 0.28545	0.14138	44	28.57%	77.92% *****
0.28545 - 0.57359	0.42952	19	12.34%	90.26% ****
0.57359 - 0.86173	0.71766	9	5.84%	96.10% **
0.86173 - 1.1499	1.0058	3	1.95%	98.05%
1.1499 - 1.4380	1.2939	1	0.65%	98.70%
1.4380 - 1.7261	1.5821	1	0.65%	99.35%
1.7261 - 2.0143	1.8702	0	0.00%	99.35%
>= 2.0143	2.1584	1	0.65%	100.00%

Test de l'hypothèse nulle de normalité de la distribution:
Khi-deux(2) = 20.251 avec p. critique 0.00004

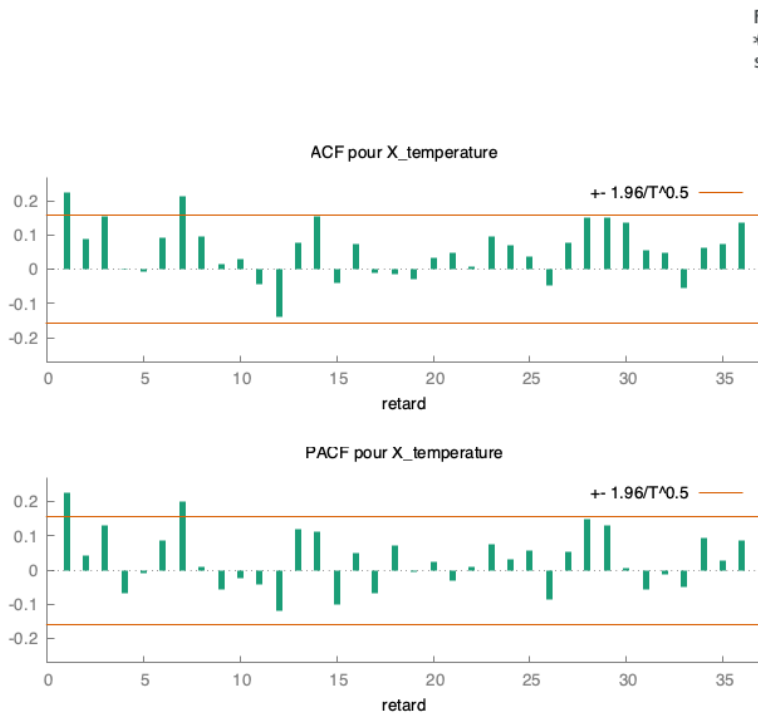
Annexe 2. Corrélogramme séries brutes Pétrole

Fonction d'autocorrélation de X_petrole
***, **, * indiquent une significativité au seuil de 1%,
sur la base d'écart-type $1/T^{0.5}$

RETARD	ACF	PACF	Q	[p. critique]
1	0.9549 ***	0.9549 ***	144.0930	[0.000]
2	0.8900 ***	-0.2480 ***	270.0815	[0.000]
3	0.8321 ***	0.0993	380.9201	[0.000]
4	0.7812 ***	0.0043	479.2732	[0.000]
5	0.7255 ***	-0.1068	564.6592	[0.000]
6	0.6664 ***	-0.0271	637.1962	[0.000]
7	0.6170 ***	0.0832	699.7920	[0.000]
8	0.5667 ***	-0.1126	752.9578	[0.000]
9	0.5063 ***	-0.1159	795.6782	[0.000]
10	0.4513 ***	0.0976	829.8604	[0.000]
11	0.4035 ***	-0.0362	857.3733	[0.000]
12	0.3636 ***	0.0334	879.8653	[0.000]
13	0.3181 ***	-0.0856	897.2046	[0.000]
14	0.2700 ***	-0.0297	909.7832	[0.000]
15	0.2323 ***	0.0672	919.1625	[0.000]
16	0.1920 **	-0.1222	925.6160	[0.000]
17	0.1413 *	-0.1091	929.1380	[0.000]
18	0.0905	0.0238	930.5934	[0.000]
19	0.0541	0.0746	931.1167	[0.000]
20	0.0211	-0.1054	931.1972	[0.000]
21	-0.0167	0.0037	931.2479	[0.000]
22	-0.0565	-0.0443	931.8314	[0.000]
23	-0.0918	-0.0497	933.3847	[0.000]
24	-0.1098	0.2015 **	935.6234	[0.000]
25	-0.1160	0.0590	938.1440	[0.000]
26	-0.1205	-0.0593	940.8842	[0.000]
27	-0.1155	0.1160	943.4194	[0.000]
28	-0.1030	0.0383	945.4516	[0.000]
29	-0.0909	-0.0478	947.0467	[0.000]
30	-0.0832	0.0282	948.3939	[0.000]
31	-0.0773	-0.0612	949.5659	[0.000]
32	-0.0669	-0.0146	950.4507	[0.000]
33	-0.0638	-0.0376	951.2629	[0.000]
34	-0.0700	-0.1037	952.2491	[0.000]
35	-0.0788	-0.0393	953.5069	[0.000]



Annexe 3. Corrélogramme séries brutes Température



Fonction d'autocorrélation de X_temperature
 ***, **, * indiquent une significativité au seuil de 1%, 5%
 sur la base d'écart-type $1/T^{0.5}$

RETARD	ACF	PACF	Q	[p. critique]
1	0.2256 ***	0.2256 ***	8.0406	[0.005]
2	0.0901	0.0413	9.3316	[0.009]
3	0.1536 *	0.1317	13.1093	[0.004]
4	0.0017	-0.0661	13.1098	[0.011]
5	-0.0080	-0.0094	13.1202	[0.022]
6	0.0933	0.0869	14.5424	[0.024]
7	0.2130 ***	0.2016 **	22.0018	[0.003]
8	0.0976	0.0103	23.5795	[0.003]
9	0.0164	-0.0564	23.6245	[0.005]
10	0.0307	-0.0224	23.7830	[0.008]
11	-0.0444	-0.0415	24.1161	[0.012]
12	-0.1396 *	-0.1168	27.4304	[0.007]
13	0.0764	0.1195	28.4308	[0.008]
14	0.1542 *	0.1130	32.5349	[0.003]
15	-0.0380	-0.0999	32.7852	[0.005]
16	0.0725	0.0511	33.7052	[0.006]
17	-0.0102	-0.0675	33.7235	[0.009]
18	-0.0156	0.0708	33.7667	[0.013]
19	-0.0300	-0.0047	33.9276	[0.019]
20	0.0346	0.0244	34.1430	[0.025]
21	0.0486	-0.0298	34.5721	[0.031]
22	0.0097	0.0105	34.5893	[0.043]
23	0.0959	0.0752	36.2866	[0.039]
24	0.0691	0.0320	37.1740	[0.042]
25	0.0358	0.0558	37.4139	[0.053]
26	-0.0456	-0.0853	37.8069	[0.063]
27	0.0771	0.0526	38.9368	[0.064]
28	0.1516 *	0.1507 *	43.3387	[0.032]
29	0.1494 *	0.1294	47.6497	[0.016]
30	0.1382 *	0.0076	51.3692	[0.009]
31	0.0546	-0.0565	51.9541	[0.011]
32	0.0472	-0.0111	52.3945	[0.013]
33	-0.0528	-0.0498	52.9508	[0.015]
34	0.0623	0.0929	53.7319	[0.017]
35	0.0727	0.0293	54.8050	[0.018]
36	0.1381 *	0.0855	58.7025	[0.010]

Annexe 4. Pétrole ARIMA (1, 1, 1)

Variable dépendante: d_X_petrole
 Écart-types basés sur la matrice hessienne

	coefficient	éc. type	z	p. critique
const	0.222830	0.516282	0.4316	0.6660
phi_1	0.0109943	0.239074	0.04599	0.9633
theta_1	0.236987	0.226408	1.047	0.2952

Moyenne var. dép.	0.235065	Éc. type var. dép.	5.301764
Moyenne innovations	0.000763	Éc. type innovations	5.128992
R2	0.057999	R2 ajusté	0.051801
Log de vraisemblance	-470.3241	Critère d'Akaike	948.6482
Critère de Schwarz	960.7960	Hannan-Quinn	953.5826

	Réel	Imaginaire	Modulo	Fréquence
AR				
Ordre 1	90.9560	0.0000	90.9560	0.0000
MM				
Ordre 1	-4.2196	0.0000	4.2196	0.5000

Estimation par AS 197 (MV exact)
 Variable dépendante: d_X_petrole
 Écart-types basés sur la matrice hessienne

	coefficient	éc. type	z	p. critique
phi_1	0.0143185	0.238261	0.06010	0.9521
theta_1	0.235204	0.225739	1.042	0.2974

Moyenne var. dép.	0.235065	Éc. type var. dép.	5.301764
Moyenne innovations	0.178717	Éc. type innovations	5.132075
R2	0.058007	R2 ajusté	0.051810
Log de vraisemblance	-470.4170	Critère d'Akaike	946.8340
Critère de Schwarz	955.9448	Hannan-Quinn	950.5348

	Réel	Imaginaire	Modulo	Fréquence
AR				
Ordre 1	69.8395	0.0000	69.8395	0.0000
MM				
Ordre 1	-4.2516	0.0000	4.2516	0.5000

Annexe 5. Série température, modèle SARIMA(0,1,1)(0,0,1)

Variable dépendante: d_X_temperature
Écart-types basés sur la matrice hessienne

	coefficient	éc. type	z	p. critique	
const	0.0101717	0.00193268	5.263	1.42e-07	***
theta_1	-1.00000	0.0209277	-47.78	0.0000	***
Theta_1	-0.226786	0.0767557	-2.955	0.0031	***
Moyenne var. dép.	0.031351	Éc. type var. dép.	1.745529		
Moyenne innovations	0.105043	Éc. type innovations	1.294919		
R2	0.454489	R2 ajusté	0.450900		
Log de vraisemblance	-261.3900	Critère d'Akaike	530.7801		
Critère de Schwarz	542.9279	Hannan-Quinn	535.7145		

	Réel	Imaginaire	Modulo	Fréquence
MM				
Ordre 1	1.0000	0.0000	1.0000	0.0000
MM (saisonnalité)				
Ordre 1	4.4095	0.0000	4.4095	0.0000

Annexe 6. Sélection de retards pour le test ADF

- Série d IPC

Les astérisques indiquent les meilleures (donc les plus faibles) valeurs des critères d'information suivants, AIC = critère d'Akaike, BIC = critère bayésien de Schwartz et HQC = critère d'Hannan-Quinn.

retards	log-vrais.	p(LR)	AIC	BIC	HQC
1	-117.34740		1.657603	1.698850*	1.674363*
2	-117.23042	0.62860	1.669867	1.731738	1.695008
3	-115.43342	0.05799	1.658797	1.741292	1.692319
4	-113.93551	0.08348	1.651882*	1.755001	1.693784
5	-113.11914	0.20133	1.654433	1.778175	1.704714
6	-112.21600	0.17895	1.655778	1.800144	1.714440
7	-112.10104	0.63159	1.668070	1.833060	1.735112
8	-111.96048	0.59596	1.680007	1.865620	1.755429
9	-111.81021	0.58355	1.691808	1.898045	1.775612
10	-110.49434	0.10475	1.687421	1.914282	1.779605

- Série d Température

Les astérisques indiquent les meilleures (donc les plus faibles) va des critères d'information suivants, AIC = critère d'Akaike, BIC = critère bayésien de Schwartz et HQC = critère d'Hannan-Quinn.

retards	log-vrais.	p(LR)	AIC	BIC	HQC
1	-244.70652		3.795485	3.839601	3.813411
2	-234.21509	0.00000	3.649463	3.715637	3.676352
3	-233.00774	0.12020	3.646273	3.734505	3.682124
4	-230.76153	0.03405	3.627100	3.737390	3.671915
5	-227.43934	0.00995	3.591375	3.723722	3.645152
6	-222.65372	0.00198	3.533134*	3.687540*	3.595874*
7	-222.17512	0.32789	3.541156	3.717619	3.612859
8	-222.06720	0.64223	3.554880	3.753402	3.635546
9	-222.03534	0.80071	3.569774	3.790354	3.659403
10	-222.02509	0.88614	3.585001	3.827639	3.683593
11	-221.88194	0.59260	3.598184	3.862879	3.705738
12	-218.82522	0.01342	3.566542	3.853295	3.683059
13	-215.66611	0.01195	3.533325	3.842136	3.658805
14	-215.66152	0.92370	3.548639	3.879508	3.683082
15	-214.91708	0.22239	3.552570	3.905498	3.695977
16	-214.67306	0.48500	3.564706	3.930101	3.715575

- Série d Pétrole

Les astérisques indiquent les meilleures (donc les plus faibles) valeurs des critères d'information suivants, AIC = critère d'Akaike, BIC = critère bayésien de Schwartz et HQC = critère d'Hannan-Quinn.

retards	log-vrais.	p(LR)	AIC	BIC	HQC
1	-403.92592		6.245014	6.289130*	6.262940*
2	-402.73728	0.12311	6.242112*	6.308286	6.269001
3	-401.97000	0.21543	6.245692	6.333924	6.281544
4	-401.77342	0.53064	6.258053	6.368342	6.302867
5	-401.77063	0.94050	6.273394	6.405742	6.327172
6	-400.83859	0.17215	6.274440	6.428846	6.337180
7	-399.37169	0.08674	6.267257	6.443720	6.338960
8	-399.32105	0.75031	6.281862	6.480384	6.362528
9	-399.30705	0.86711	6.297032	6.517611	6.386660
10	-399.24838	0.73192	6.311513	6.554151	6.410105
11	-399.24070	0.90137	6.326780	6.591475	6.434335
12	-399.11429	0.61510	6.340220	6.626973	6.456737
13	-399.10765	0.90829	6.355502	6.664314	6.480983
14	-398.41460	0.23906	6.360225	6.691094	6.494668
15	-396.87795	0.07959	6.351968	6.704896	6.495375

Annexe 7. Recherche de séries cointégrées. MCO pour les séries :

- Y(IPC), X(température)
- Y(IPC), X(pétrole)

Variable dépendante: Y_IPC						Variable dépendante: Y_IPC					
	coefficient	éc. type	t de Student	p. critique			coefficient	éc. type	t de Student	p. critique	
const	100.255	0.673980	148.8	2.11e-167	***	const	101.249	1.60044	63.26	1.33e-111	***
X_temperature	1.43122	0.304427	4.701	5.72e-06	***	X_petrole	0.0226599	0.0240598	0.9418	0.3478	
Moyenne var. dép.	102.6933	Éc. type var. dép.	5.713576			Moyenne var. dép.	102.6933	Éc. type var. dép.	5.713576		
Somme carrés résidus	4392.736	Éc. type régression	5.358236			Somme carrés résidus	4998.344	Éc. type régression	5.715673		
R2	0.126227	R2 ajusté	0.120517			R2	0.005764	R2 ajusté	-0.000734		
F(1, 153)	22.10278	P. critique (F)	5.72e-06			F(1, 153)	0.887017	P. critique (F)	0.347772		
Log de vraisemblance	-479.1174	Critère d'Akaike	962.2347			Log de vraisemblance	-489.1268	Critère d'Akaike	982.2536		
Critère de Schwarz	968.3216	Hannan-Quinn	964.7070			Critère de Schwarz	988.3405	Hannan-Quinn	984.7260		
rho	0.918148	Durbin-Watson	0.226094			rho	0.994985	Durbin-Watson	0.011765		

- X(température), Y(IPC)
- X(température), X(pétrole)

Variable dépendante: X_temperature					Variable dépendante: X_temperature				
	coefficient	éc. type	t de Student	p. critique		coefficient	éc. type	t de Student	p. critique
const	-7.35335	1.92945	-3.811	0.0002	***	const	1.66323	0.398428	4.174
Y_IPC	0.0881958	0.0187596	4.701	5.72e-06	***	X_petrole	0.000636089	0.00598965	0.1062
Moyenne var. dép.	1.703761	Éc. type var. dép.	1.418335			Moyenne var. dép.	1.703761	Éc. type var. dép.	1.418335
Somme carrés résidus	270.6929	Éc. type régression	1.330126			Somme carrés résidus	309.7751	Éc. type régression	1.422910
R2	0.126227	R2 ajusté	0.120517			R2	0.000074	R2 ajusté	-0.006462
F(1, 153)	22.10278	P. critique (F)	5.72e-06			F(1, 153)	0.011278	P. critique (F)	0.915564
Log de vraisemblance	-263.1464	Critère d'Akaike	530.2927			Log de vraisemblance	-273.5981	Critère d'Akaike	551.1963
Critère de Schwarz	536.3796	Hannan-Quinn	532.7651			Critère de Schwarz	557.2831	Hannan-Quinn	553.6686
rho	0.126720	Durbin-Watson	1.720312			rho	0.227328	Durbin-Watson	1.505053

- X(pétrole), Y(IPC)
- X(pétrole), X(température)

Variable dépendante: X_petrole					Variable dépendante: X_petrole					
	coefficient	éc. type	t de Student	p. critique		coefficient	éc. type	t de Student	p. critique	
const	37.6014	27.7789	1.354	0.1779	const	63.5264	2.41568	26.30	1.24e-58	***
Y_IPC	0.254374	0.270089	0.9418	0.3478	X_temperature	0.115875	1.09113	0.1062	0.9156	
Moyenne var. dép.	63.72387	Éc. type var. dép.	19.14325	Moyenne var. dép.	63.72387	Éc. type var. dép.	19.14325			
Somme carrés résidus	56110.16	Éc. type régression	19.15028	Somme carrés résidus	56431.30	Éc. type régression	19.20500			
R2	0.005764	R2 ajusté	-0.000734	R2	0.000074	R2 ajusté	-0.006462			
F(1, 153)	0.887017	P. critique (F)	0.347772	F(1, 153)	0.011278	P. critique (F)	0.915564			
Log de vraisemblance	-676.5381	Critère d'Akaike	1357.076	Log de vraisemblance	-676.9804	Critère d'Akaike	1357.961			
Critère de Schwarz	1363.163	Hannan-Quinn	1359.549	Critère de Schwarz	1364.048	Hannan-Quinn	1360.433			
rho	0.964429	Durbin-Watson	0.077091	rho	0.966570	Durbin-Watson	0.076161			

Annexe 8. Sélection du nombre de retards pour les résidus de :

- Y(IPC), X(température)
- Y(IPC), X(pétrole)

Les astérisques indiquent les meilleures (donc les plus faibles) valeurs des critères d'information suivants, AIC = critère d'Akaike, BIC = critère bayésien de Schwartz et HQC = critère d'Hannan-Quinn.

retards	log-vrais.	p(LR)	AIC	BIC	HQC
1	-307.30169		4.722163	4.766059	4.740000
2	-297.56874	0.00001	4.588836	4.654680	4.615591
3	-289.31334	0.00005	4.478066	4.565859*	4.513740*
4	-288.68025	0.26048	4.483668	4.593408	4.528260
5	-287.74710	0.17190	4.484688	4.616377	4.538199
6	-286.51398	0.11632	4.481129	4.634766	4.543559
7	-283.48778	0.01389	4.450195*	4.625780	4.521543
8	-283.35429	0.60536	4.463424	4.660957	4.543691
9	-282.78042	0.28402	4.469930	4.689411	4.559115
10	-282.74807	0.79924	4.484703	4.726132	4.582807
11	-282.74107	0.90583	4.499864	4.763241	4.606886
12	-282.69034	0.75008	4.514356	4.799681	4.630297

Les astérisques indiquent les meilleures (donc les plus faibles) valeurs des critères d'information suivants, AIC = critère d'Akaike, BIC = critère bayésien de Schwartz et HQC = critère d'Hannan-Quinn.

retards	log-vrais.	p(LR)	AIC	BIC	HQC
1	-104.31415		1.786792	1.833500	1.805759
2	-101.48499	0.01737	1.756050	1.826112*	1.784500*
3	-100.38141	0.13737	1.754309	1.847725	1.792243
4	-100.16788	0.51344	1.767527	1.884297	1.814944
5	-99.82012	0.40429	1.778489	1.918613	1.835389
6	-99.81265	0.90273	1.795171	1.958648	1.861554
7	-99.46595	0.40501	1.806150	1.992982	1.882017
8	-99.46317	0.94048	1.822910	2.033096	1.908260
9	-98.90701	0.29158	1.830370	2.063910	1.925203
10	-98.78858	0.62649	1.845186	2.102080	1.949503
11	-97.02269	0.06020	1.832314	2.112562	1.946114
12	-96.60012	0.35793	1.842019	2.145621	1.965302
13	-94.46295	0.03869	1.822907	2.149862	1.955673
14	-89.30380	0.00132	1.753005	2.103315	1.895255
15	-89.30235	0.95712	1.769787	2.143451	1.921520
16	-88.95351	0.40356	1.780731	2.177749	1.941948
17	-88.66685	0.44895	1.792720	2.213092	1.963420
18	-88.21169	0.34003	1.801877	2.245603	1.982060
19	-85.12282	0.01294	1.766770	2.233850	1.956436
20	-84.95164	0.55847	1.780700	2.271133	1.979849
21	-84.33866	0.26819	1.787204	2.300992	1.995837
22	-78.87376	0.00095	1.712164*	2.249306	1.930280
23	-78.84811	0.82081	1.728540	2.289035	1.956139
24	-78.76353	0.68087	1.743035	2.327774	1.981088

- X(température), Y(IPC)
- X(température), X(pétrole)

Système VAR, ordre des retards max 24

Les astérisques indiquent les meilleures (donc les plus faibles) valeurs des critères d'information suivants, AIC = critère d'Akaike, BIC = critère bayésien de Schwartz et HQC = critère d'Hannan-Quinn.

retards	log-vrais.	p(LR)	AIC	BIC	HQC
1	-219.69955		3.384726*	3.428622*	3.402563*
2	-219.65958	0.77738	3.399383	3.465227	3.426138
3	-219.08512	0.28378	3.405880	3.493672	3.441554
4	-218.55857	0.30479	3.413108	3.522848	3.457700
5	-218.52939	0.80912	3.427930	3.559618	3.481441
6	-218.35523	0.55506	3.440538	3.594174	3.502967
7	-216.66515	0.06599	3.430002	3.605587	3.501350
8	-216.64101	0.82611	3.444901	3.642434	3.525167
9	-215.94083	0.23666	3.449478	3.668959	3.538663
10	-215.71936	0.50571	3.461364	3.702793	3.559468
11	-215.43237	0.44868	3.472250	3.735627	3.579272
12	-214.14706	0.10887	3.467894	3.753219	3.583834

Système VAR, ordre des retards max 24

Les astérisques indiquent les meilleures (donc les plus faibles) valeurs des critères d'information suivants, AIC = critère d'Akaike, BIC = critère bayésien de Schwartz et HQC = critère d'Hannan-Quinn.

retards	log-vrais.	p(LR)	AIC	BIC	HQC
1	-225.10474		3.467248*	3.511144*	3.485085*
2	-225.02013	0.68082	3.481223	3.547068	3.507979
3	-223.53822	0.08515	3.473866	3.561658	3.509540
4	-223.36223	0.55299	3.486446	3.596187	3.531039
5	-223.33785	0.82523	3.501341	3.633030	3.554852
6	-222.82882	0.31298	3.508837	3.662473	3.571266
7	-220.78622	0.04326	3.492919	3.668504	3.564267
8	-220.74869	0.78411	3.507614	3.705146	3.587880
9	-220.11020	0.25846	3.513133	3.732614	3.602318
10	-219.93879	0.55820	3.525783	3.767212	3.623886
11	-219.71240	0.50101	3.537594	3.800971	3.644616
12	-218.65934	0.14671	3.536784	3.822109	3.652724
13	-217.37606	0.10914	3.532459	3.839732	3.657318

- X(pétrole), X(température)
- X(pétrole), Y(IPC)

systeme VAR, ordre des retards max 24

Les astérisques indiquent les meilleures (donc les plus faibles) valeurs des critères d'information suivants, AIC = critère d'Akaike, BIC = critère bayésien de Schwartz et HQC = critère d'Hannan-Quinn.

retards	log-vrais.	p(LR)	AIC	BIC	HQC
1	-409.37885		6.280593	6.324489	6.298430
2	-404.75250	0.00235	6.225229*	6.291073*	6.251984*
3	-404.05466	0.23745	6.229842	6.317634	6.265516
4	-403.64939	0.36796	6.238922	6.348662	6.283514
5	-403.17823	0.33168	6.246996	6.378684	6.300507
6	-403.16145	0.85464	6.262007	6.415643	6.324436
7	-402.76262	0.37179	6.271185	6.446770	6.342533
8	-400.74366	0.04449	6.255628	6.453161	6.335895
9	-400.52800	0.51135	6.267603	6.487084	6.356788
10	-400.48670	0.77379	6.282240	6.523668	6.380343
11	-400.46142	0.82210	6.297121	6.560498	6.404143

systeme VAR, ordre des retards max 24

Les astérisques indiquent les meilleures (donc les plus faibles) valeurs des critères d'information suivants, AIC = critère d'Akaike, BIC = critère bayésien de Schwartz et HQC = critère d'Hannan-Quinn.

retards	log-vrais.	p(LR)	AIC	BIC	HQC
1	-409.63330		6.284478	6.328374	6.302315
2	-404.93831	0.00218	6.228066*	6.293910*	6.254821*
3	-404.20295	0.22523	6.232106	6.319898	6.267780
4	-403.74074	0.33632	6.240317	6.350057	6.284909
5	-403.23355	0.31385	6.247840	6.379529	6.301351
6	-403.22386	0.88933	6.262960	6.416596	6.325389
7	-402.75366	0.33217	6.271048	6.446633	6.342396
8	-400.83299	0.05000	6.256992	6.454525	6.337259
9	-400.61341	0.50753	6.268907	6.488388	6.358092
10	-400.58029	0.79689	6.283669	6.525097	6.381772
11	-400.57917	0.96230	6.298919	6.562296	6.405940
12	-400.49471	0.68107	6.312896	6.598221	6.428837

Annexe 9. Test de stationnarité ADF pour les résidus des séries :

- Résidus : Y(IPC), X(température)
- Résidus : Y(IPC), X(pétrole)

Test de Dickey-Fuller augmenté pour res_Y_temp
test à reculons à partir de 7 retards, suivant le critère AIC
taille de l'échantillon 148
hypothèse nulle de racine unitaire : $\alpha = 1$

test avec constante
avec 6 retards de (1-L)res_Y_temp
modèle: $(1-L)y = b_0 + (a-1)*y(-1) + \dots + e$
valeur estimée de $(a - 1)$: 0.0799983
statistique de test: $\tau_c(1) = 1.94246$
p. critique asymptotique 0.9999
Coeff. d'autocorrélation du 1er ordre pour e: -0.015
différences retardées: $F(6, 140) = 10.894 [0.0000]$

avec constante et tendance temporelle
avec 6 retards de (1-L)res_Y_temp
modèle: $(1-L)y = b_0 + b_1*t + (a-1)*y(-1) + \dots + e$
valeur estimée de $(a - 1)$: 0.0401848
statistique de test: $\tau_{ct}(1) = 0.396871$
p. critique asymptotique 0.999
Coeff. d'autocorrélation du 1er ordre pour e: -0.013
différences retardées: $F(6, 139) = 6.012 [0.0000]$

Test de Dickey-Fuller augmenté pour r2_Y_pet
test à reculons à partir de 22 retards, suivant le critère AIC
taille de l'échantillon 133
hypothèse nulle de racine unitaire : $\alpha = 1$

test avec constante
avec 21 retards de (1-L)r2_Y_pet
modèle: $(1-L)y = b_0 + (a-1)*y(-1) + \dots + e$
valeur estimée de $(a - 1)$: 0.0326975
statistique de test: $\tau_c(1) = 2.73702$
p. critique asymptotique 1
Coeff. d'autocorrélation du 1er ordre pour e: -0.009
différences retardées: $F(21, 110) = 2.932 [0.0001]$

avec constante et tendance temporelle
avec 21 retards de (1-L)r2_Y_pet
modèle: $(1-L)y = b_0 + b_1*t + (a-1)*y(-1) + \dots + e$
valeur estimée de $(a - 1)$: 0.0726802
statistique de test: $\tau_{ct}(1) = 1.60151$
p. critique asymptotique 1
Coeff. d'autocorrélation du 1er ordre pour e: -0.016
différences retardées: $F(21, 109) = 2.884 [0.0002]$

- X(température), Y(IPC)
- X(température), X(pétrole)

Test de Dickey-Fuller augmenté pour residus_temp_IPC
test à reculons à partir de 1 retards, suivant le critère AIC
taille de l'échantillon 154
hypothèse nulle de racine unitaire : $a = 1$

test avec constante
avec 0 retards de $(1-L)\text{residus_temp_IPC}$
modèle: $(1-L)y = b_0 + (a-1)y(-1) + e$
valeur estimée de $(a - 1)$: -0.873303
statistique de test: $\tau_c(1) = -10.9982$
p. critique asymptotique $1.945e-22$
Coeff. d'autocorrélation du 1er ordre pour e: -0.008

avec constante et tendance temporelle
avec 0 retards de $(1-L)\text{residus_temp_IPC}$
modèle: $(1-L)y = b_0 + b_1t + (a-1)y(-1) + e$
valeur estimée de $(a - 1)$: -0.873183
statistique de test: $\tau_{ct}(1) = -10.9575$
p. critique asymptotique $8.739e-24$
Coeff. d'autocorrélation du 1er ordre pour e: -0.008

test de Dickey-Fuller augmenté pour res_temp_Petro
test à reculons à partir de 1 retards, suivant le critère AIC
taille de l'échantillon 154
hypothèse nulle de racine unitaire : $a = 1$

test avec constante
avec 0 retards de $(1-L)\text{res_Temp_Petro}$
modèle: $(1-L)y = b_0 + (a-1)y(-1) + e$
valeur estimée de $(a - 1)$: -0.77256
statistique de test: $\tau_c(1) = -9.94187$
p. critique asymptotique $5.098e-19$
Coeff. d'autocorrélation du 1er ordre pour e: -0.028

avec constante et tendance temporelle
avec 0 retards de $(1-L)\text{res_Temp_Petro}$
modèle: $(1-L)y = b_0 + b_1t + (a-1)y(-1) + e$
valeur estimée de $(a - 1)$: -0.866276
statistique de test: $\tau_{ct}(1) = -10.8683$
p. critique asymptotique $2.11e-23$
Coeff. d'autocorrélation du 1er ordre pour e: -0.007

- X(pétrole), X(température)
- X(pétrole), Y(IPC)

Test de Dickey-Fuller augmenté pour res_Petro_Tempe
test à reculons à partir de 2 retards, suivant le critère AIC
taille de l'échantillon 153
hypothèse nulle de racine unitaire : $a = 1$

test avec constante
avec un retard de $(1-L)\text{res_Petro_Tempe}$
modèle: $(1-L)y = b_0 + (a-1)y(-1) + \dots + e$
valeur estimée de $(a - 1)$: -0.0451669
statistique de test: $\tau_c(1) = -2.0454$
p. critique asymptotique 0.2674
Coeff. d'autocorrélation du 1er ordre pour e: 0.023

avec constante et tendance temporelle
avec un retard de $(1-L)\text{res_Petro_Tempe}$
modèle: $(1-L)y = b_0 + b_1t + (a-1)y(-1) + \dots + e$
valeur estimée de $(a - 1)$: -0.0464518
statistique de test: $\tau_{ct}(1) = -2.02808$
p. critique asymptotique 0.5853
Coeff. d'autocorrélation du 1er ordre pour e: 0.023

Test de Dickey-Fuller augmenté pour reta_Petro_IPC
test à reculons à partir de 2 retards, suivant le critère AIC
taille de l'échantillon 153
hypothèse nulle de racine unitaire : $a = 1$

test avec constante
avec un retard de $(1-L)\text{reta_Petro_IPC}$
modèle: $(1-L)y = b_0 + (a-1)y(-1) + \dots + e$
valeur estimée de $(a - 1)$: -0.0468713
statistique de test: $\tau_c(1) = -2.11963$
p. critique asymptotique 0.237
Coeff. d'autocorrélation du 1er ordre pour e: 0.023

avec constante et tendance temporelle
avec un retard de $(1-L)\text{reta_Petro_IPC}$
modèle: $(1-L)y = b_0 + b_1t + (a-1)y(-1) + \dots + e$
valeur estimée de $(a - 1)$: -0.0502824
statistique de test: $\tau_{ct}(1) = -2.15069$
p. critique asymptotique 0.5167
Coeff. d'autocorrélation du 1er ordre pour e: 0.023

11 Bibliographie

- AFP, G. a. (2024, Mars). *GEO*. Récupéré sur Climat: pourquoi les prix de l'alimentation devraient augmenter dans les prochaines années: <https://www.geo.fr/environnement/climat-pourquoi-les-prix-de-l-alimentation-devraient-augmenter-dans-les-prochaines-annees-219435>
- Bleu, F. (2023, Novembre). *France Bleu*. Récupéré sur Inflation : le nombre de Français en insécurité alimentaire en France : <https://www.francebleu.fr/infos/economie-social/inflation-le-nombre-de-francais-en-insecurite-alimentaire-en-france-a-triple-en-huit-ans-selon-une-etude-3268320>
- Calignon, G. d. (2016, Janvier). *Les Echos*. Récupéré sur Pétrole : la baisse des prix va continuer de soutenir le pouvoir d'achat en 2016: <https://www.lesechos.fr/2016/01/petrole-la-baisse-des-prix-va-continuer-de-soutenir-le-pouvoir-dachat-en-2016-194140>
- Gazzane, H. (2014, Décembre). *Le figaro*. Récupéré sur Comprendre la chute vertigineuse des prix du pétrole: <https://www.lefigaro.fr/conjoncture/2014/12/17/20002-20141217ARTFIG00244-comprendre-la-guerre-du-petrole.php>
- Girard, D. (2023 -2024, Février). Séries temporelles univariées. *Graphiques ARIMA* . Nantes.
- Hamisultane, H. (2016, Janvier). *MODELE A CORRECTION D'ERREUR (MCE) ET APPLICATIONS*. Récupéré sur HAL open science: <https://shs.hal.science/cel-01261167/document>
- Haut Conseil de la famille, de l'enfance et de l'âge*. (2023, Décembre). Récupéré sur Le pouvoir d'achat des familles face au choc d'inflation: https://www.hcfea.fr/IMG/pdf/pouvoir_d_achat_des_familles_face_au_choc_d_inflation-2.pdf
- INSEE*. (2024, 03). Récupéré sur Cours des matières premières importées - Pétrole brut Brent (Londres) - Prix au comptant en euros par baril : <https://www.insee.fr/fr/statistiques/serie/010002078#Documentation>
- Organisation des Nations Unies pour l'alimentation et l'agriculture*. (2024, 24). Récupéré sur Variation de température sur la superficie des terres: <https://www.fao.org/faostat/fr/#data/ET>

Organisation des Nations unies pour l'alimentation et l'agriculture. (2024, janvier). Récupéré sur Indices de Prix à la Consommation: <https://www.fao.org/faostat/fr/#data/CP>

Ouvrard, Y. K.-F. (2018, Mars). *Banque de France*. Récupéré sur L'impact du prix du pétrole sur l'inflation en France et en zone euro: https://publications.banque-france.fr/sites/default/files/billet_50_vf_rec.pdf

Souktani, L. (2003, décembre). Consulté le mars 2024, sur Université de Montréal: https://papyrus.bib.umontreal.ca/xmlui/bitstream/handle/1866/14584/Souktani_Lamya_2003_memoire.pdf?sequence=1&isAllowed=y

Ulgazi, Y. (2021, Octobre). *Banque de France* . Récupéré sur La hausse de l'inflation en France est-elle temporaire ?: <https://www.banque-france.fr/fr/publications-et-statistiques/publications/la-hausse-de-l-inflation-en-france-est-elle-temporaire>

12 Table des figures

FIGURE 1 - ÉVOLUTION DE L'IPC ALIMENTAIRE EN FRANCE	6
FIGURE 2 - PETROLE BRUT BRENT (LONDRES) - PRIX AU COMPTANT EN EUROS PAR BARIL	7
FIGURE 3 - VARIATION DE TEMPERATURE EN FRANCE. PERIODE 01/2010 AU 11/2022.....	9
FIGURE 4 - SERIE BRUTE Y : IPC ET SON CORRELOGRAMME	17
FIGURE 5 - COMPARAISON SERIE BRUTE ET LOGARITHMIQUE	18
FIGURE 6 -SERIE Y DIFFERENCIATION $D=1$ ET SON CORRELOGRAMME	19
FIGURE 7 -MODELE $(1, 1, 0) (1,0,0)$	21
FIGURE 8 - MODELE $(1, 1, 0) (2, 0,0)$	22
FIGURE 9 MODELE $(0,1,1) (0,0,1)$	23
FIGURE 10 - MODELE $(0,1,1) (1,0,2)$	24
FIGURE 11 - VALEURS OBSERVEES ET PREDITES.....	26
FIGURE 12 - NORMALITE DES RESIDUS	26
FIGURE 13 CORRELOGRAMME DE LA SERIE PETROLE DIFFERENCIE	29
FIGURE 14 - RESULTATS D_PETROLE MODELE ARIMA $(1,1,0)$	30
FIGURE 15 - CORRELOGRAMME SERIE D_TEMPERATUE	31
FIGURE 16 - SERIE D_TEMPERATURE, MODELE ARIMA $(0,1,1)(1,0,0)$	32
FIGURE 17 - TEST ADF SERIE IPC DIFFERENCIEE	34
FIGURE 18 - TEST ADF SERIE TEMPERATURE DIFFERENCIEE	35
FIGURE 19 - TEST ADF SERIE PETROLE DIFFERENCIEE.....	35
FIGURE 20 - MCE	38
FIGURE 21 - MCO TEMPERATURE SUR IPC	39

13 Table des matières

1	INTRODUCTION	3
2	ANALYSE ECONOMIQUE DU SUJET	4
3	ANALYSE DES TROIS VARIABLES	5
3.1	VARIABLE A EXPLIQUER	5
3.2	VARIABLE EXPLICATIVE 1 : PRIX DU PETROLE.....	7
3.3	VARIABLE EXPLICATIVE 2 : LA TEMPERATURE.....	9
4	PRESENTATION DE LA METHODOLOGIE ARIMA.....	10
4.1	STATIONNARITE.....	11
4.2	CORRELOGRAMME.....	12
4.2.1	Identification du processus $AR(p)$	13
4.2.2	Identification du processus $MA(q)$	14
4.3	CONSTRUCTION DU MODELE DE BOX JENKINS	15
4.4	MODELES MULTIPLICATIFS.....	16
5	CHOIX DU MODELE ARIMA POUR YT	17
5.1	STATIONNARITE.....	17
5.2	IDENTIFICATION ARIMA (P, D, Q), (P, D, Q) ET ESTIMATION	20
5.3	VERIFICATION ET CHOIX DU MODELE	24
5.4	PREVISION A UNE PERIODE	27
6	ANALYSE DES VARIABLES PETROLE ET TEMPERATURE	28
6.1	STATIONNARITE.....	28
6.2	PETROLE.....	29
6.3	TEMPERATURE.....	31
7	TEST DE COINTEGRATION SELON ENGLE-GRANGER.....	33
7.1	TEST DE COINTEGRATION.....	34
7.2	MODELE A CORRECTION D'ERREUR POUR LES VARIABLES CO INTEGREES.....	37
8	CONCLUSION	40
9	DISCUSSION.....	42
10	ANNEXES	43
11	BIBLIOGRAPHIE	50

12	TABLE DES FIGURES	52
13	TABLE DES MATIERES.....	53