



Séries temporelles

Rapport de projet

Francky ANDRIAMAMPIONONA
Ahmed M'SABOU

3 mai 2024

Table des matières

1	Prédiction de la Vente de supercarburant	3
1.1	Modélisation sans variable exogène	3
1.1.1	Une première analyse graphique	3
1.1.2	Test de stationnarité	4
1.1.3	Détermination des paramètres du modèle	5
1.1.4	Out-of-sample testing	6
1.2	Ajout de variables exogènes	7
1.2.1	Choix et modélisation des variables exogènes	7
1.2.2	Modélisation en prenant en compte les variables exogènes	7
1.3	Prévision de la vente de supercarburants auto	9
2	Prédiction de la vente de gazole	10
2.1	Modélisation sans variable exogène	10
2.1.1	Analyse graphique	10
2.1.2	Test de la stationnarité	11
2.1.3	Détermination des paramètres du modèle	12
2.1.4	Out-of-sample testing	13
2.2	Ajout de variables exogènes	14
2.2.1	Choix et modélisation des variables exogènes	14
2.2.2	Modélisation en prenant en compte les variables exogènes	14
2.3	Prévision de la vente de gazole	16
3	Conclusion	17
A	Annexe	18
A.1	Vente de supercarburants auto (en kt)	18
A.1.1	Résumé du modèle SARIMA(5,2,1)x(5,1,1,12)	18
A.1.2	Résumé du modèle final SARIMA(3,2,1)x(5,1,1,12)	19
A.1.3	Test de Causalité de Granger pour les variables exogènes	20
A.2	Vente de gazole (en kt)	20
A.2.1	Résumé du modèle SARIMA(2,1,4)x(5,1,1,12)	20
A.2.2	Test de Causalité de Granger pour les variables exogènes	21

Introduction

L'objectif de ce projet est de déterminer deux modèles de prédiction de deux séries temporelles que sont la Vente de supercarburant auto (en kt) et la vente de gazole (en kt) pour les 12 prochains mois à l'aide de méthodes statistiques. Les tests statistiques et les prédictions sont effectués sous Python. Nous effectuerons dans un premier temps la prédiction de la vente de supercarburant auto (en kt), en comparant l'impact de l'ajout de variables exogènes à notre modèle. Puis, nous réaliserons le même procédé pour la prédiction de la vente de gazole.

1 Prédiction de la Vente de supercarburant

1.1 Modélisation sans variable exogène

1.1.1 Une première analyse graphique

Commençons tout d'abord par visualiser la série temporelle à prédire. Nous remarquons à première vue une saisonnalité, qui semble annuelle. Nous constatons aussi des valeurs extrêmes en 2020, certainement dû à la baisse de production durant la période du Covid 19.

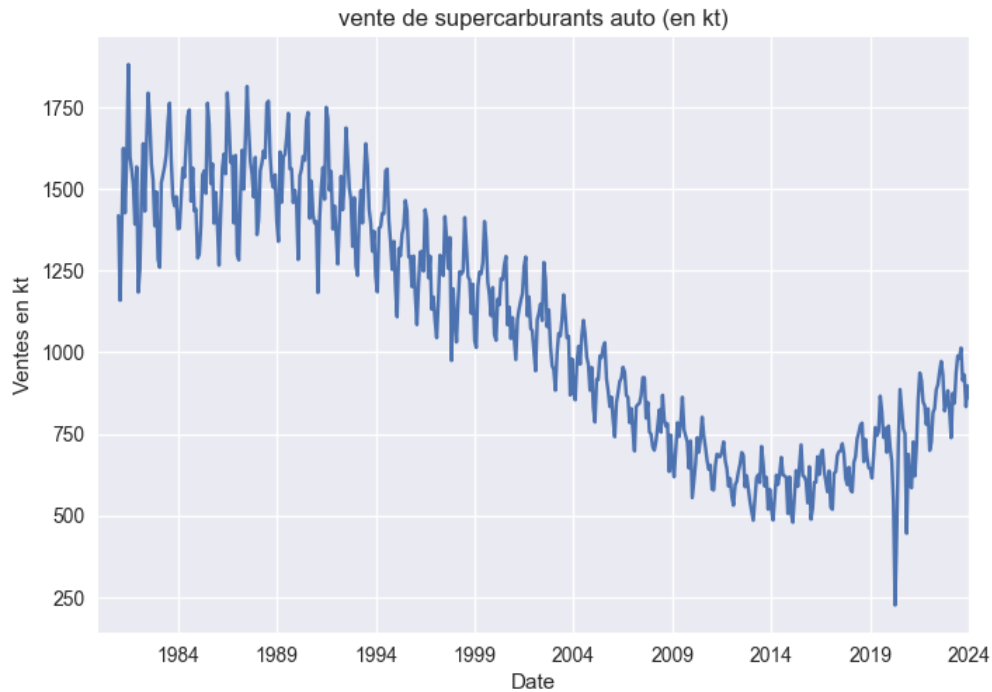


FIGURE 1 – Vente de supercaruburants auto en kt

Pour avoir un meilleur avis sur la nature de notre série, nous décidons de décomposer la série temporelle, pour identifier le trend, la composante saisonnière et les résidus.

Ce que l'on constate, c'est tout d'abord un saisonnalité annuelle clairement identifiable. Cela nous indique qu'il faudra prendre prévoir un modèle tenant compte de la saisonnalité comme un modèle SARIMA. Ensuite, le trend est décroissant avec une remontée en fin de période. Enfin, nous remarquons que les résidus sont centrés autour de 0 avec quelques outliers vers 1900 et 1997. Le modèle semble être additif.

Avant d'estimer un modèle qui conviendrait à notre série temporelle, il nous faut étudier sa stationnarité. Nous décidons également de séparer nos données en train/test. Nous avons choisi de limiter nos données train de janvier 1990 à décembre 2019 et nos données test de janvier 2020 à janvier 2024. Nous expliquerons ce choix de split train/test plus tard. Nos calibrations de modèles se feront donc sur notre base train et le test de pouvoir de prédiction du modèle sur notre base test.

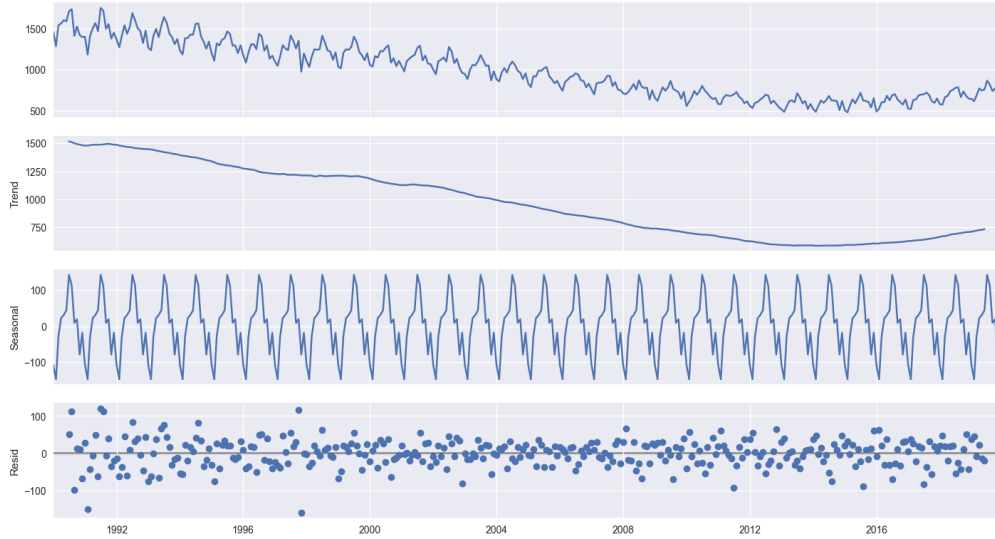


FIGURE 2 – Décomposition de la série temporelle

1.1.2 Test de stationnarité

Pour tester la stationnarité de la série, nous décidons de combiner deux tests statistiques : le test Augmented Dickey-Fuller (ADF) et le test KPSS.

Le test ADF va tester la présence de racine unitaire caractéristique d'un processus où la moyenne et la variance augmentent au cours du temps. L'hypothèse H_0 testée est : " la série temporelle présente une racine unitaire", ce qui implique une non-stationnarité. Le test KPSS est un test contraire au test ADF et va tester la stationnarité de la série temporelle. Ainsi, un p-value inférieure à 0.05 au test ADF et une p-value supérieure à 0.05 permettrait de conclure à la stationnarité de la série temporelle.

	ADF	KPSS
statistic	-1.463577	2.979882
p-value	0.551446	0.010000

FIGURE 3 – Série temporelle originelle

Les tests ADF et KPSS ne sont pas concluant sur la série temporelle originelle. Nous décidons alors de différencier la série temporelle.

	ADF	KPSS
statistic	-2.286811	0.024787
p-value	0.176257	0.100000

FIGURE 4 – Série temporelle différenciée 1 fois

Le test KPSS est validé mais le test ADF ne l'est pas. Cela signifie que la série est stationnaire autour d'un trend déterministe. Nous devons donc encore une fois différencier la série.

	ADF	KPSS
statistic	-1.344204e+01	0.009434
p-value	3.828437e-25	0.100000

FIGURE 5 – Série temporelle différenciée 2 fois

Cette fois-ci, les deux tests indiquent que la série différenciée 2 fois ne présente pas de racine unitaire, ce qui signifie que la série transformée est stationnaire. Nous nous dirigeons donc vers un modèle de type SARIMA avec comme $d = 1$.

1.1.3 Détermination des paramètres du modèle

Afin de déterminer les paramètres, nous nous basons sur les graphiques des fonctions d'autocorrélations et d'autocorrélations partielles.



FIGURE 6 – ACF et PACF de la série stationnaire

En se basant sur le plot ACF, nous remarquons que le premier à rentrer dans l'intervalle de confiance est le 2e donc $q = 1$. Concernant le plot PACF, le premier lag à rentrer dans l'intervalle de confiance est le 6e donc $p = 5$.

Nous remarquons aussi la présence de pics récurrents tous les 12 lags dans l'ACF de la série stationnaire, cela confirme la présence d'une saisonnalité. Nous nous dirigeons vers un modèle SARIMA. Nous allons donc différencier la série de manière à éliminer la composante saisonnière et obtenir une série stationnaire.

	ADF	KPSS
statistic	-9.425274e+00	0.021858
p-value	5.367564e-16	0.100000

FIGURE 7 – Test de stationnarité de la série différenciée de manière saisonnière

Les tests ADF et KPSS indiquent que la série différenciée 1 fois de manière saisonnière est stationnaire. Analysons les graphes ACF et PACF pour déterminer les paramètres P, D et Q du modèle SARIMA.

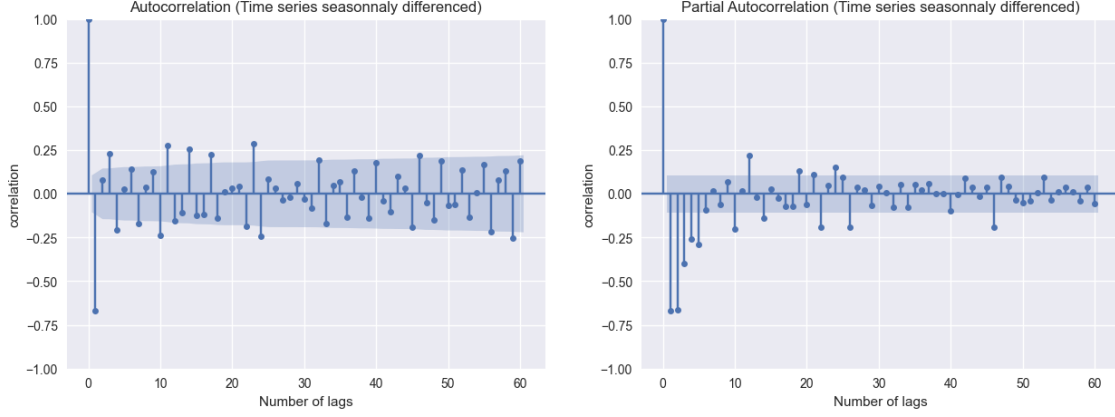


FIGURE 8 – ACF et PACF de la série différenciée de manière saisonnière

De la même manière nous trouvons $P = 5$ et $Q = 1$. Ainsi le modèle que nous utiliserons est un modèle SARIMA(5,2,1)x(5,1,1,12).

Afin de confirmer notre potentiel modèle, il aurait été intéressant de tuner les hyperparamètres par GridSearch afin de trouver un modèle optimal sur la base du critère AIC. Cependant, le modèle donnée par Grid Search est un modèle SARIMA(3,2,0)x(5,1,0,12) ayant un plus grand AIC que le modèle trouvé par analyse des graphiques. Nous décidons donc de conserver notre modèle SARIMA(5,2,1)x(5,1,1,12).

Regardons les caractéristiques de notre modèle sur la base train (cf. Annexe A.1.1). Nous remarquons que les coefficients des composantes AR(4) et AR(5) ne sont pas significatives dans notre modèle. Nous décidons de les supprimer et d'utiliser un modèle SARIMA(3,2,1)x(5,1,1,12) (cf. Annexe A.1.2) A pour lequel nous obtenons un meilleur AIC de **3425**. De plus, la p-value pour le test de Ljung-Box (L1) est de 0.90, ce qui signifie que les résidus ne sont pas auto-corrélés et qu'ils sont donc indépendents. Ainsi, notre modèle ne souffre pas de défauts d'ajustement significatifs.

1.1.4 Out-of-sample testing

Nous décidons de tester notre modèle entraîné sur notre base de test (de janvier 2020 à janvier 2024). Graphiquement, nous constatons que notre modèle semble bien capturer la saisonnalité de la série temporelle. Il y a cependant un écart en 2020, ce qui nous semble normal puisque cela correspond à la période de COVID. Nous regardons le critère MSE. Notre modèle SARIMA(3,2,1)x(5,1,1,12) présente une **MSE de 16592**, ce qui est plus faible que celui trouvé avec le modèle SARIMA(5,2,1)x(5,1,1,12), qui présentait une MSE de 17282.

Nous décidons donc de choisir comme modèle final pour notre variable target le modèle **SARIMA(3,2,1)x(5,1,1,12)**.

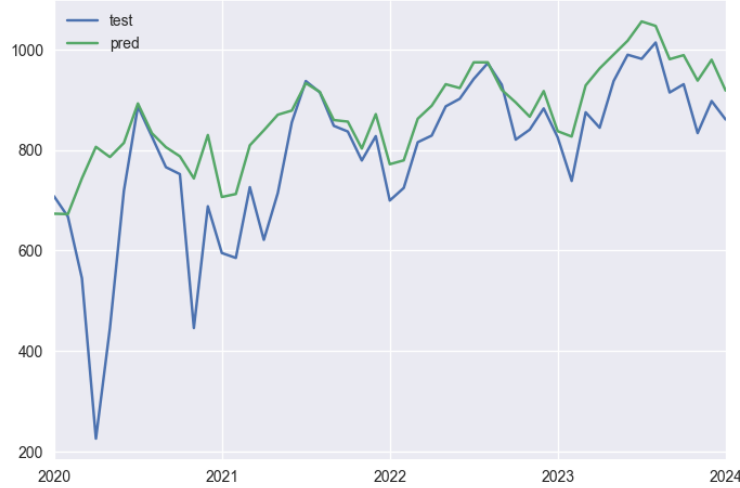


FIGURE 9 – Out-of-sample test du SARIMA(3,2,1)x(5,1,1,12)

1.2 Ajout de variables exogènes

1.2.1 Choix et modélisation des variables exogènes

Nous décidons afin d'améliorer la prédiction de la série temporelle d'inclure 2 variables exogènes. Nous décidons de prendre les variables ayant une corrélation de plus de 0.8 (en valeur absolue). Puis nous effectuons un test de Causalité de Granger pour déterminer si une série a une influence instantanée ou décalée sur la série temporelle target. Toutes les séries avec une corrélation supérieure à 0.8 ont validé le test de Granger. Ces séries sont donc potentiellement utilisables pour améliorer la prédiction de la série temporelle target.

Nous décidons de garder comme variables exogènes les Ventes de supercarburants auto CVS-CVC-CJO (en kt) (7.2) et Production de pétrole (brut et hydrocarbures extraits du gaz naturel en kt) (1) car parmi les 4 séries temporelles exogènes sélectionnées, elles possèdent le plus grand nombre de points de données avec un départ commun en 1990.

Nous effectuons ensuite une analyse de chacune de ces séries temporelles, de la même manière que précédemment et nous trouvons pour la série Ventes de supercarburants auto CVS-CVC-CJO (en kt) (7.2) un modèle SARIMA(2,1,1)x(2,1,1,12) et pour la série Production de pétrole (brut et hydrocarbures extraits du gaz naturel en kt) (1) un modèle SARIMA(2,1,1)x(2,1,2,12). Afin de pouvoir utiliser comme variables exogènes, nous effectuons une prévision de ces deux séries sur les 12 prochains mois.

1.2.2 Modélisation en prenant en compte les variables exogènes

Nous ajoutons tout d'abord la série Ventes de supercarburants auto CVS-CVC-CJO (en kt) (7.2) à notre modèle pour voir l'impact sur les différents critères d'évaluation.

Nous constatons une baisse de l'AIC passant à **3233**, ce qui montre une amélioration de notre ajustement. Concernant le test de Ljung-Box, la p-value est toujours à 0,90 ce qui signifie que notre modèle ne souffre pas de défauts d'ajustement significatifs. Concernant le test Out-of-sample, nous constatons que la prédiction est plus juste et colle mieux à nos données, avec notamment une MSE de **5660**, ce qui est près de 3 fois inférieur au modèle sans variable exogène. Cela montre l'importance

SARIMAX Results			
Dep. Variable:	5.2 Ventes de supercarburants auto (en kt)	No. Observations:	360
Model:	SARIMAX(3, 2, 1)x(5, 1, 1, 12)	Log Likelihood	-1604.815
Date:	Thu, 02 May 2024	AIC	3233.631
Time:	19:36:50	BIC	3279.788
Sample:	01-01-1990	HQIC	3252.011
	- 12-01-2019		
Covariance Type:	opg		

FIGURE 10 – SARIMA(3,2,1)x(5,1,1,12) avec ajout de la première variable exogène

de l'ajout de variables pour améliorer la prédiction.

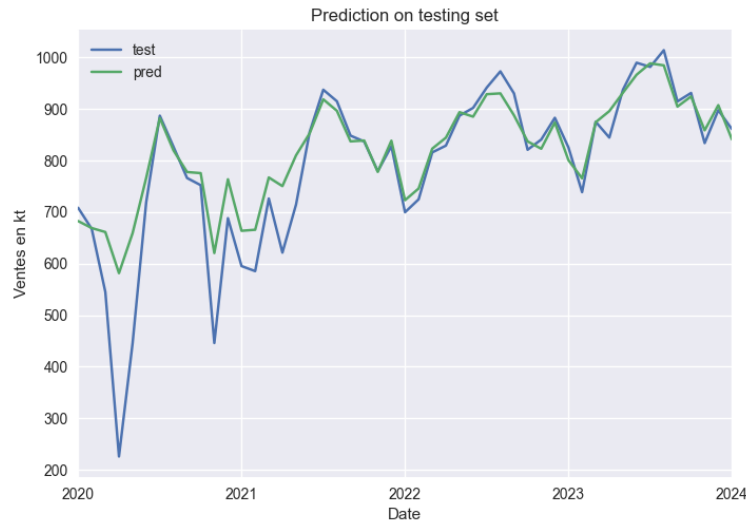


FIGURE 11 – Out-of-sample test avec ajout de la première variable exogène

Nous décidons ensuite de prendre en compte les deux variables exogènes, afin d'améliorer le pouvoir de prédiction. Comme nous le constatons sur le résumé statistique du modèle, le score AIC augmente légèrement par rapport à la modélisation avec uniquement la première variable exogène, passant à **3233**.

SARIMAX Results			
Dep. Variable:	5.2 Ventes de supercarburants auto (en kt)	No. Observations:	360
Model:	SARIMAX(3, 2, 1)x(5, 1, 1, 12)	Log Likelihood	-1603.102
Date:	Thu, 02 May 2024	AIC	3232.205
Time:	19:48:01	BIC	3282.208
Sample:	01-01-1990	HQIC	3252.116
	- 12-01-2019		
Covariance Type:	opg		

FIGURE 12 – SARIMA(3,2,1)x(5,1,1,12) avec ajout des deux variables exogènes

Cependant, regardons comment se comporte notre modèle en terme de prédiction sur la base d'entraînement et surtout en terme de MSE.

Nous ne constatons visuellement pas de différence entre l'ajout d'une variable et l'ajout de deux variables. Cependant, ce modèle est plus précis avec une MSE de **5378**. En se basant sur ce critère,

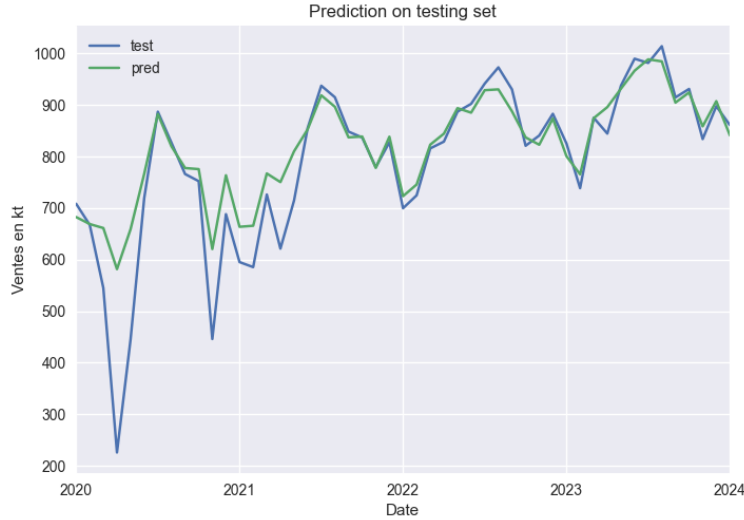


FIGURE 13 – Out-of-sample test avec ajout des deux variables exogènes

nous décidons donc de choisir un modèle $SARIMA(3,2,1) \times (5,1,1,12)$ avec comme variables exogènes les variables Ventes de supercarburants auto CVS-CVC-CJO (en kt) (7.2) et Production de pétrole (brut et hydrocarbures extraits du gaz naturel en kt) (1).

Une dernière validation de modèle est la visualisation des résidus. Si ceux-ci sont décorrélés entre eux alors ce signifie qu'il ne reste que très peu d'information contenu dans ces résidus pouvant être utiles lors du forecast.

Ljung-Box (L1) (Q):	0.03	Jarque-Bera (JB):	125.25
Prob(Q):	0.87	Prob(JB):	0.00
Heteroskedasticity (H):	0.40	Skew:	-0.15
Prob(H) (two-sided):	0.00	Kurtosis:	5.93

FIGURE 14 – Tests des résidus du modèle

Nous avons $Prob(Q) = 0.87 > 0.05$ ce qui signifie par test de Ljung-Box que nos résidus sont décorrélés entre eux. Cependant, le test de Jarque-Bera n'est pas concluant. Nos résidus ne sont donc pas normaux.

1.3 Prévision de la vente de supercarburants auto

Nous effectuons maintenant la prévision des ventes de supercarburants auto à l'aide du modèle final trouvé et des prévisions des séries exogènes modélisées précédemment.

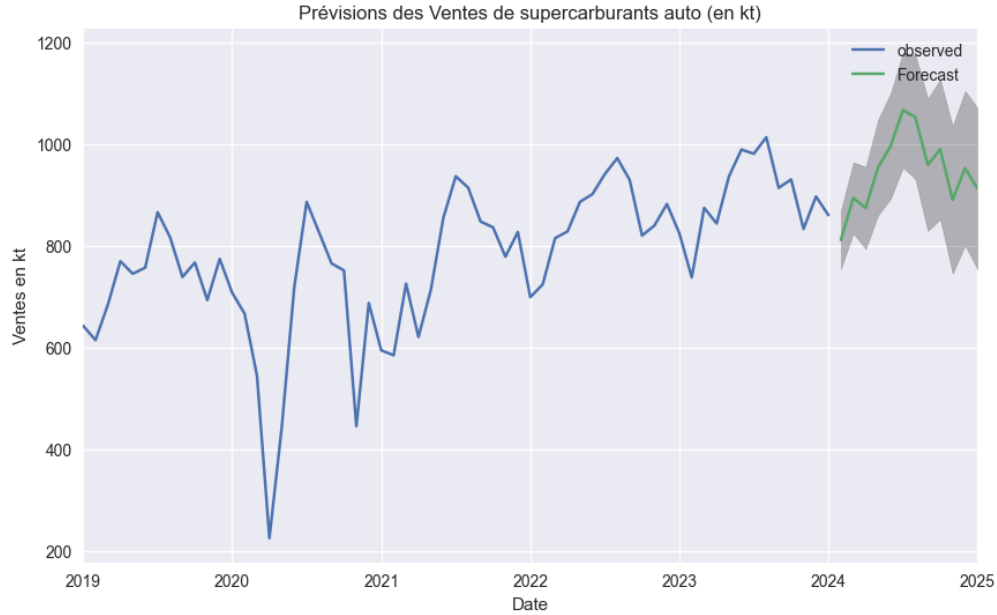


FIGURE 15 – Forecast de la série temporelle target

2 Prédiction de la vente de gazole

2.1 Modélisation sans variable exogène

2.1.1 Analyse graphique

Nous allons tout d'abord visualiser la série temporelle target. la saisonnalité est plus difficile à observer que la série précédente. Nous constatons également un énorme pic en 2020, correspondant également à la période du COVID.

Avant d'effectuer toute analyse de notre série temporelle, nous décidons comme précédemment de séparer nos données en base de train et bases de test. Pour la base train, les données commencent en janvier 1990 et vont jusqu'à janvier 2020. Afin de mieux visualiser les composantes de notre série, notamment le trend et la saisonnalité, nous effectuons une décomposition saisonnière. Ce que nous observons sur le graphe de décomposition, c'est la présence d'une composante saisonnière annuelle ainsi qu'un trend croissant. Enfin, les résidus sont centrés autour de 0, pouvant nous faire penser à du bruit blanc. Le modèle semble être additif.

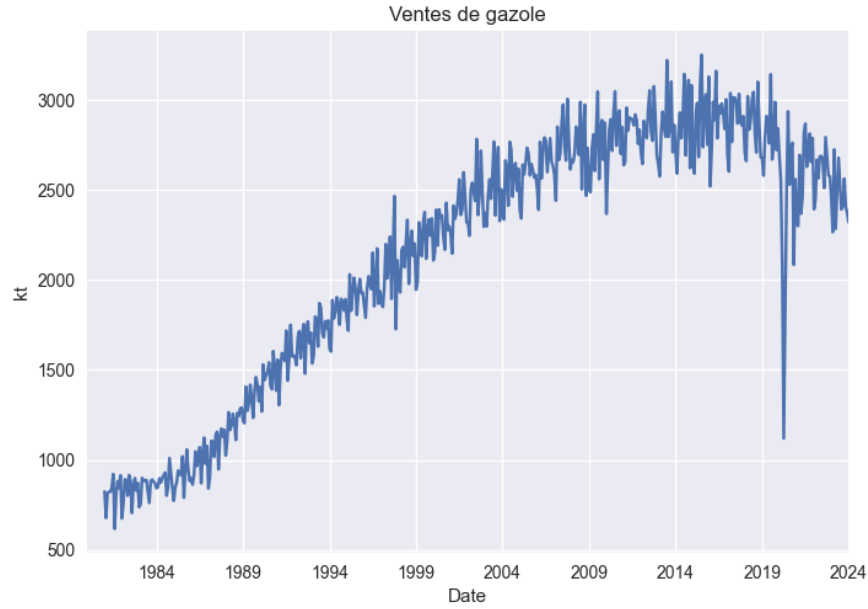


FIGURE 16 – Vente de gazole en kt

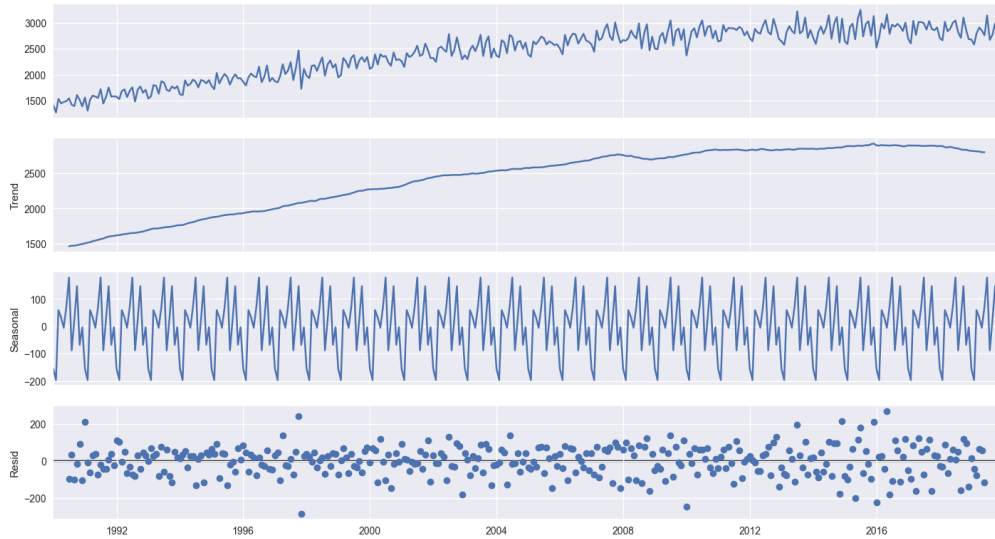


FIGURE 17 – Décomposition de la série temporelle

2.1.2 Test de la stationnarité

Effectuons maintenant des tests de stationnarité à la série originelle. On constate que le test ADF nous indique que la série est stationnaire tandis que le test KPSS rejette l'hypothèse de stationnarité. Ainsi il faut différencier la série pour obtenir une série stationnaire.

Nous constatons dans le tableau de résultat des tests que les tests ADF et KPSS indiquent que la série différenciée est stationnaire. Cela nous pousse donc vers un modèle du type ARIMA (et même SARIMA au vu de la forte composante saisonnière) avec un paramètre $d = 1$.

	ADF	KPSS
statistic	-3.237316	2.895483
p-value	0.017924	0.010000

FIGURE 18 – Tests de stationnarité sur la série originelle

	ADF	KPSS
statistic	-3.237316	2.895483
p-value	0.017924	0.010000

FIGURE 19 – Tests de stationnarité sur la série différenciée 1 fois

2.1.3 Détermination des paramètres du modèle

Afin de déterminer les paramètres, nous nous basons sur les graphiques des fonctions d'autocorrélations et d'autocorrélations partielles.

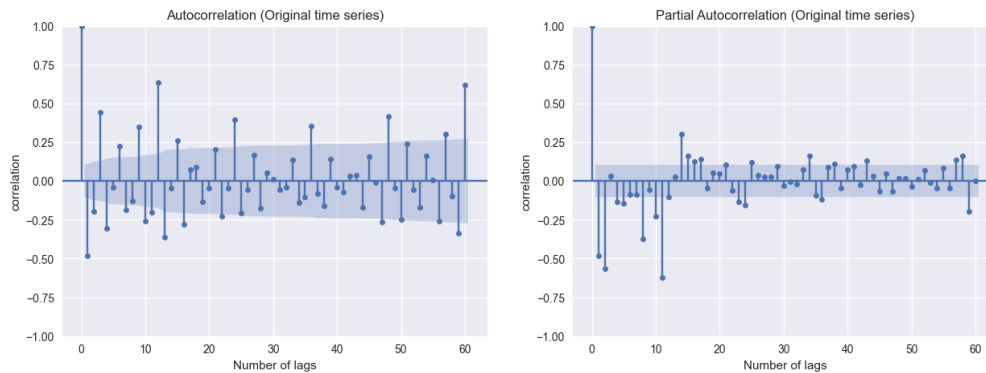


FIGURE 20 – ACF et PACF de la série stationnaire

On trouve sur le graphe PACF $p = 2$ et sur le graphe ACF $q = 4$. On remarque aussi une saisonnalité annuelle avec des pics importants tous les 12 lags, ce qui confirme bien la présence de saisonnalité annuelle dans la série temporelle. Différencions donc la série de manière saisonnière afin d'éliminer la composante saisonnière et testons sa stationnarité.

	ADF	KPSS
statistic	-7.385592e+00	0.029359
p-value	8.253327e-11	0.100000

FIGURE 21 – Tests de stationnarité sur la série différenciée 1 fois de manière saisonnière

Les tests ADF et KPSS indiquent que la série différenciée de manière saisonnière est stationnaire. Analysons maintenant les graphes ACF et PACF pour déterminer les paramètres P, D et Q du modèle SARIMA.

On trouve grâce à ces graphes $P = 5$, $Q = 1$ et $D = 1$ (car différenciée d1 fois de manière saisonnière). On aurait donc un modèle SARIMA(2,1,4)x(5,1,1)

Regardons maintenant les caractéristiques de notre modèle sur la base train. (cf. Annexe A.2.1).



FIGURE 22 – ACF et PACF de la série différenciée de manière saisonnière

Nous constatons un score AIC de **3900**. De plus, la p-value pour le test de Ljung-Box est de 0.91, ce qui indique que les résidus ne sont pas auto-corrélés entre-eux. Ainsi, notre modèle ne souffre pas d'erreurs d'ajustements importantes.

2.1.4 Out-of-sample testing

Testons à présent notre modèle de base sur nos données test (i.e. de janvier 2020 à janvier 2024).

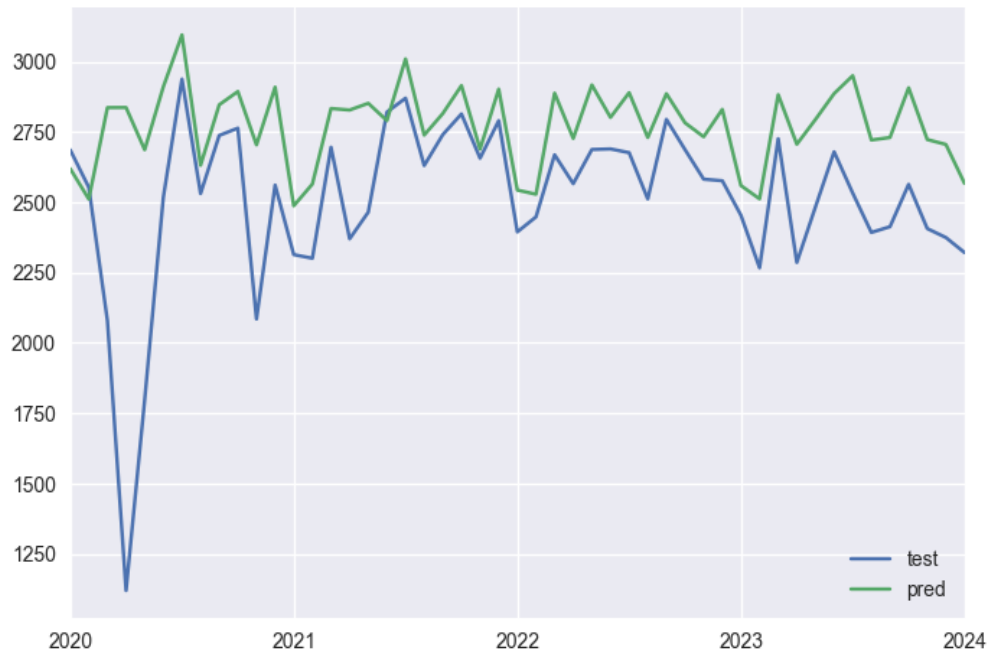


FIGURE 23 – Out-of-sample test du SARIMA(2,1,4)x(5,1,1,12)

Nous obtenons une MSE de **145585**, ce qui semble énorme. Nous décidons donc d'améliorer notre modèle en ajoutant des variables exogènes comme précédemment.

2.2 Ajout de variables exogènes

2.2.1 Choix et modélisation des variables exogènes

Afin d'améliorer notre modèle, nous procédons comme en section 1.2.1. Nous choisissons les variables les plus corrélées à notre série target et possédant un historique suffisant (i.e. commençant en 1990 comme pour la base train de notre variable target). Ensuite nous effectuons un test de causalité de granger pour voir si chacune de ces variables a une influence de façon instantanée ou décalée sur notre série temporelle à prédire. D'après le test de Granger, tous les variables ont un(cf. Annexe A.2.2).

Nous décidons de garder comme variable exogènes les séries Ventes de gazole CVS-CVC-CJO (6.3) et Consommation de carburants routiers (9). Nous effectuons ensuite une analyse graphique de ces séries temporelles. Nous trouvons pour la série Ventes de gazole CVS-CVC-CJO (6.3) un modèle ARIMA(4,1,1) et pour la série Consommation de carburants routiers (9) un modèle SARIMA(3, 1, 4)x(5, 1, 1, 12). Nous effectuons ensuite une prévision de ces deux séries sur les 12 prochains mois.

2.2.2 Modélisation en prenant en compte les variables exogènes

Nous décidons tout d'abord d'ajouter au modèle de base (i.e. le modèle SARIMA(2,1,4)x(5,1,1) trouvé en section 2.1) la variable Ventes de gazole CVS-CVC-CJO (6.3).

SARIMAX Results			
Dep. Variable:	5.3 Ventes de gazole (en kt)	No. Observations:	360
Model:	SARIMAX(2, 1, 4)x(5, 1, [1], 12)	Log Likelihood	-1794.084
Date:	Fri, 03 May 2024	AIC	3616.167
Time:	21:24:48	BIC	3670.058
Sample:	01-01-1990	HQIC	3637.624
	- 12-01-2019		
Covariance Type:	opg		

FIGURE 24 – SARIMA(2,1,4)x(5,1,1) avec ajout de la première variable exogène

Nous constatons tout d'abord une baisse du score AIC passant de **3900** à **3616**, ce qui montre une amélioration du modèle. En ce qui concerne le test out-of-sample de notre modèle, nous observons également une amélioration du pouvoir de prédiction du modèle passant d'une MSE de **145585** à **2592**.

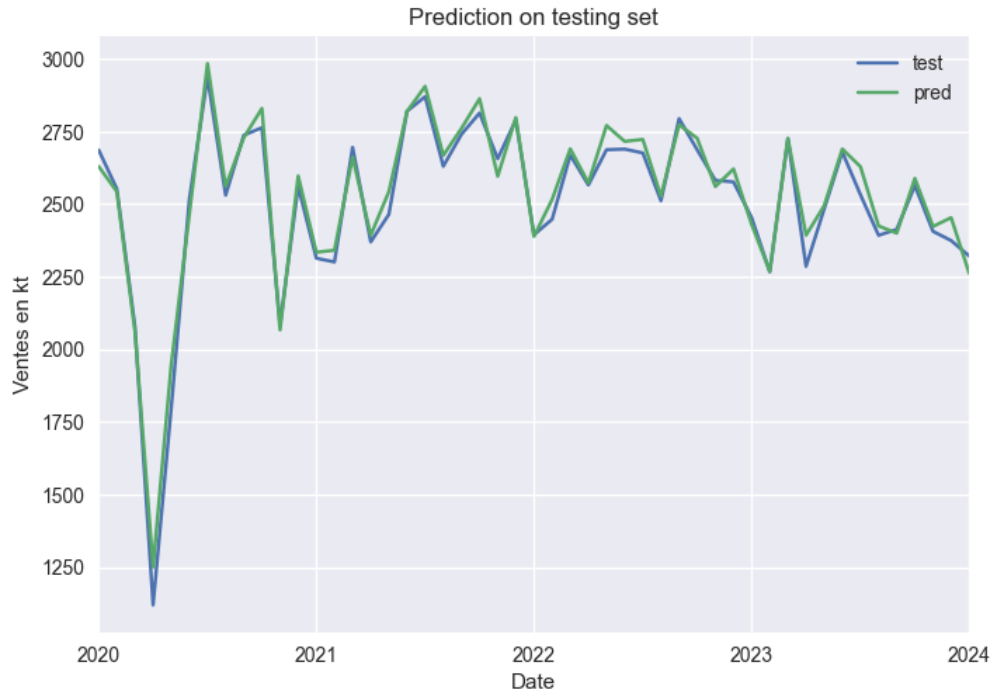


FIGURE 25 – Out-of-sample test du SARIMA(2,1,4)x(5,1,1) avec ajout de la première variable exogène

Ajoutons maintenant la seconde variable à notre modèle et regardons les critères AIC et MSE.

SARIMAX Results			
Dep. Variable:	5.3 Ventes de gazole (en kt)	No. Observations:	360
Model:	SARIMAX(2, 1, 4)x(5, 1, [1], 12)	Log Likelihood	-1470.640
Date:	Fri, 03 May 2024	AIC	2971.280
Time:	21:42:09	BIC	3029.019
Sample:	01-01-1990	HQIC	2994.269
	- 12-01-2019		
Covariance Type:	opg		

FIGURE 26 – SARIMA(2,1,4)x(5,1,1) avec ajout des deux variables exogène

Nous observons également une baisse de l'AIC à **2971**, ce qui montre encore une fois une amélioration du modèle avec l'ajout des deux variables exogènes. Effectuons un test Out-of-sample de ce modèle pour confirmer l'amélioration de l'ajustement.

Nous obtenons une MSE de **1066**. Nous avons donc une MSE quasiment 140 fois inférieure à notre modèle de base ajusté sans variables exogènes. Cela montre l'intérêt de l'utilisation de variables exogènes corrélées pour prédire notre série temporelle.

Avant de passer à la prévision finale de notre série temporelle, nous regardons les résidus du modèle. La p-value pour le test de Ljung Box est supérieure à 0.05 ce qui indique donc que les résidus ne sont pas auto-corrélés. Dans le cas contraire, cela aurait indiqué que certaines informations utiles à la prédiction sont encore contenues dans les résidus.

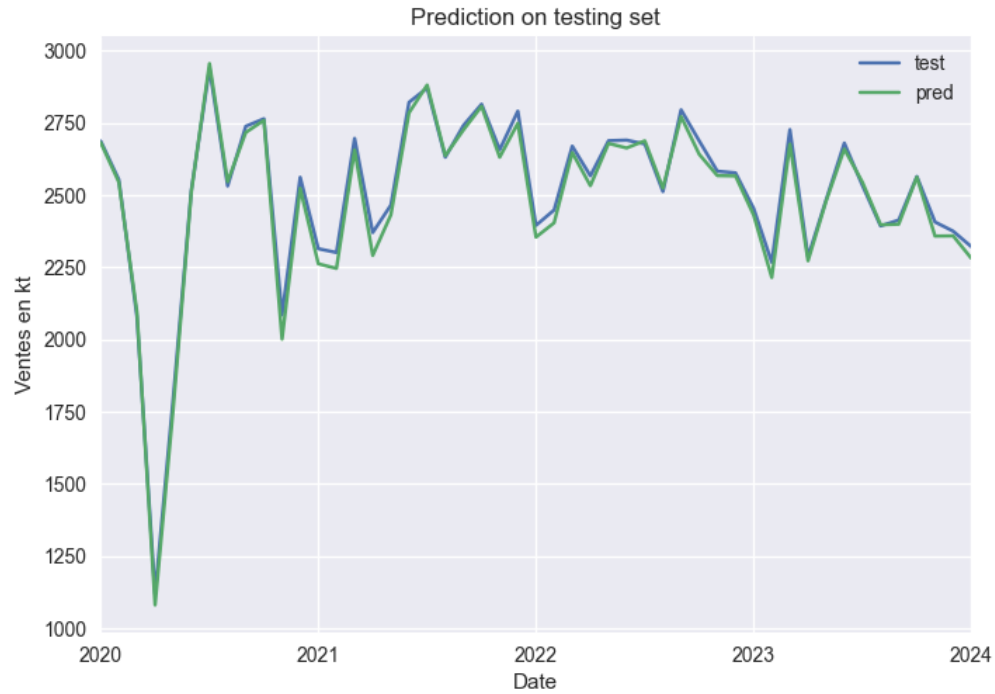


FIGURE 27 – Out-of-sample test du SARIMA(2,1,4)x(5,1,1) avec ajout des deux variables exogènes

Ljung-Box (L1) (Q):	0.07	Jarque-Bera (JB):	14.80
Prob(Q):	0.80	Prob(JB):	0.00
Heteroskedasticity (H):	0.60	Skew:	-0.24
Prob(H) (two-sided):	0.01	Kurtosis:	3.89

FIGURE 28 – Tests des résidus du modèle final

2.3 Prédiction de la vente de gazole

Nous effectuons maintenant la prédiction des ventes de gazole à l'aide du modèle final trouvé et des prévisions des séries exogènes modélisées précédemment.

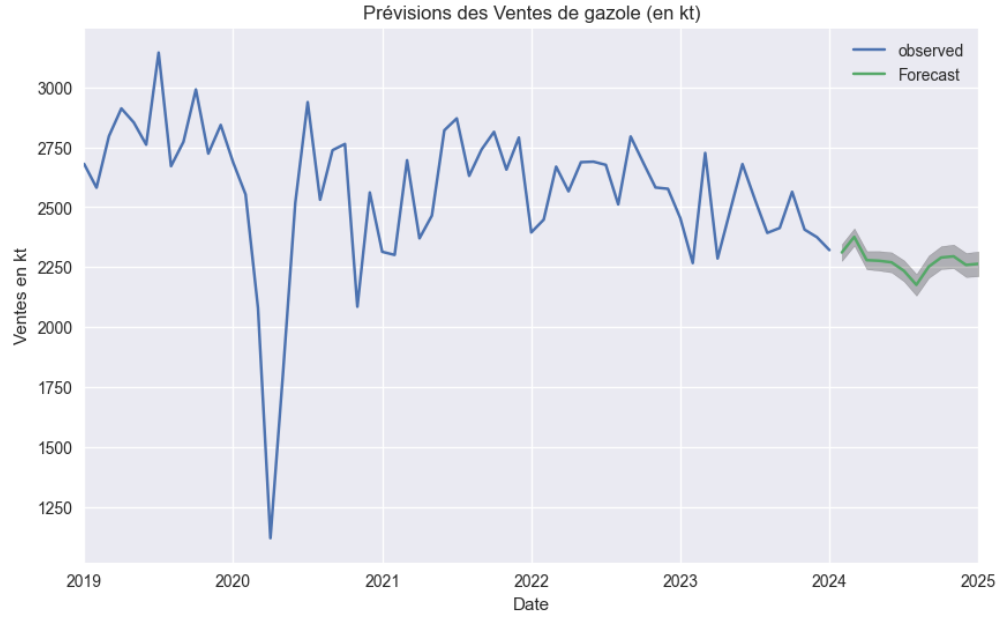


FIGURE 29 – Tests des résidus du modèle final

3 Conclusion

Nous avons estimé deux modèles pour prédire la Vente de supercarburants auto (en kt) et Vente de gazole (en kt). Nous avons pu améliorer le pouvoir prédictif de nos deux modèles en utilisant des variables exogènes fortement corrélées et ayant une influence instantanée ou décalée sur nos séries temporelles target. Une piste d'amélioration possible pour obtenir de meilleures prédictions serait d'utiliser plus de variables exogènes corrélées afin de capturer tous les mouvements des séries temporelles target.

A Annexe

A.1 Vente de supercarburants auto (en kt)

A.1.1 Résumé du modèle SARIMA(5,2,1)x(5,1,1,12)

SARIMAX Results						
Dep. Variable:	5.2 Ventes de supercarburants auto (en kt)			No. Observations:	360	
Model:	SARIMAX(5, 2, 1)x(5, 1, 1, 12)			Log Likelihood	-1700.419	
Date:	Thu, 02 May 2024			AIC	3426.838	
Time:	09:18:18			BIC	3476.841	
Sample:	01-01-1990			HQIC	3446.749	
- 12-01-2019						
Covariance Type:				opg		
	coef	std err	z	P> z	[0.025	0.975]
ar.L1	-0.9723	0.053	-18.473	0.000	-1.076	-0.869
ar.L2	-0.7036	0.069	-10.164	0.000	-0.839	-0.568
ar.L3	-0.2238	0.080	-2.787	0.005	-0.381	-0.066
ar.L4	-0.0871	0.075	-1.158	0.247	-0.234	0.060
ar.L5	-0.0071	0.056	-0.125	0.900	-0.118	0.104
ma.L1	-1.0094	0.019	-53.583	0.000	-1.046	-0.972
ar.S.L12	-1.4847	0.068	-21.933	0.000	-1.617	-1.352
ar.S.L24	-1.1465	0.086	-13.381	0.000	-1.314	-0.979
ar.S.L36	-1.0098	0.094	-10.703	0.000	-1.195	-0.825
ar.S.L48	-0.8010	0.082	-9.823	0.000	-0.961	-0.641
ar.S.L60	-0.4478	0.049	-9.228	0.000	-0.543	-0.353
ma.S.L12	0.9665	0.122	7.908	0.000	0.727	1.206
sigma2	965.8575	90.706	10.648	0.000	788.076	1143.639
Ljung-Box (L1) (Q):	0.01	Jarque-Bera (JB):	30.93			
Prob(Q):	0.94	Prob(JB):	0.00			
Heteroskedasticity (H):	0.34	Skew:	-0.32			
Prob(H) (two-sided):	0.00	Kurtosis:	4.32			

FIGURE 30 – Résumé du SARIMA(5,2,1)x(5,1,1,12)

A.1.2 Résumé du modèle final SARIMA(3,2,1)x(5,1,1,12)

SARIMAX Results						
Dep. Variable:	5.2 Ventes de supercarburants auto (en kt)				No. Observations:	360
Model:	SARIMAX(3, 2, 1)x(5, 1, 1, 12)				Log Likelihood	-1701.554
Date:	Thu, 02 May 2024				AIC	3425.108
Time:	09:19:31				BIC	3467.419
Sample:	01-01-1990				HQIC	3441.956
- 12-01-2019						
Covariance Type:			opg			
	coef	std err	z	P> z	[0.025	0.975]
ar.L1	-0.9572	0.050	-19.073	0.000	-1.056	-0.859
ar.L2	-0.6459	0.060	-10.686	0.000	-0.764	-0.527
ar.L3	-0.1405	0.051	-2.778	0.005	-0.240	-0.041
ma.L1	-0.9926	0.020	-50.540	0.000	-1.031	-0.954
ar.S.L12	-1.4880	0.066	-22.422	0.000	-1.618	-1.358
ar.S.L24	-1.1543	0.083	-13.828	0.000	-1.318	-0.991
ar.S.L36	-1.0203	0.093	-10.925	0.000	-1.203	-0.837
ar.S.L48	-0.8183	0.080	-10.226	0.000	-0.975	-0.661
ar.S.L60	-0.4607	0.047	-9.773	0.000	-0.553	-0.368
ma.S.L12	0.9719	0.138	7.057	0.000	0.702	1.242
sigma2	983.7062	103.626	9.493	0.000	780.602	1186.810
Ljung-Box (L1) (Q):	0.02	Jarque-Bera (JB):		27.60		
Prob(Q):	0.90	Prob(JB):		0.00		
Heteroskedasticity (H):	0.35	Skew:		-0.27		
Prob(H) (two-sided):	0.00	Kurtosis:		4.27		

FIGURE 31 – Résumé du SARIMA(3,2,1)x(5,1,1,12)

A.1.3 Test de Causalité de Granger pour les variables exogènes

	6.2 Ventes de supercarburants auto CVC-CJO (en kt)	5.2.3 Livraisons de SP98 (en kt)	7.2 Ventes de supercarburants auto CVS-CVC-CJO (en kt)	1. Production de pétrole (brut et hydrocarbures extraits du gaz naturel en kt)
1	0.000333	8.330842e-09	1.641575e-07	1.803768e-06
2	0.000068	1.499876e-05	3.382914e-07	1.236957e-14
3	0.000243	7.126380e-05	1.002749e-05	1.063772e-14
4	0.000730	1.564609e-07	2.246608e-09	1.375877e-18
5	0.003687	2.893592e-07	3.500842e-08	8.828239e-18
6	0.005426	1.573166e-06	1.213156e-07	2.905475e-18
7	0.000542	6.314880e-03	7.731677e-06	6.926665e-11
8	0.000622	7.453710e-03	4.320008e-06	1.353731e-13
9	0.000722	2.079372e-02	6.559179e-07	5.293464e-10
10	0.000346	1.342519e-02	7.747539e-07	5.032100e-13
11	0.000146	8.622942e-03	4.564184e-05	2.994613e-16
12	0.000118	3.069097e-08	2.001271e-05	2.188970e-05

FIGURE 32 – Test de Causalité de Granger

A.2 Vente de gazole (en kt)

A.2.1 Résumé du modèle SARIMA(2,1,4)x(5,1,1,12)

SARIMAX Results						
Dep. Variable:	5.3 Ventes de gazole (en kt)			No. Observations:	360	
Model:	SARIMAX(2, 1, 4)x(5, 1, [1], 12)			Log Likelihood	-1982.228	
Date:	Fri, 03 May 2024			AIC	3990.457	
Time:	19:58:48			BIC	4040.498	
Sample:	01-01-1990			HQIC	4010.381	
	- 12-01-2019					
Covariance Type:	opg					
	coef	std err	z	P> z	[0.025	0.975]
ar.L1	-1.3707	0.022	-62.208	0.000	-1.414	-1.328
ar.L2	-0.9900	0.024	-42.056	0.000	-1.036	-0.944
ma.L1	0.1834	0.052	3.542	0.000	0.082	0.285
ma.L2	-0.2907	0.045	-6.516	0.000	-0.378	-0.203
ma.L3	-0.6420	0.050	-12.747	0.000	-0.741	-0.543
ma.L4	0.3461	0.059	5.880	0.000	0.231	0.461
ar.S.L12	-1.5019	0.053	-28.458	0.000	-1.605	-1.398
ar.S.L24	-1.1803	0.074	-15.899	0.000	-1.326	-1.035
ar.S.L36	-1.0945	0.087	-12.633	0.000	-1.264	-0.925
ar.S.L48	-1.0377	0.080	-12.964	0.000	-1.195	-0.881
ar.S.L60	-0.6175	0.045	-13.715	0.000	-0.706	-0.529
ma.S.L12	0.9675	0.088	10.961	0.000	0.794	1.140
sigma2	4778.9481	409.254	11.677	0.000	3976.825	5581.071
Ljung-Box (L1) (Q):	0.01	Jarque-Bera (JB):	9.70			
Prob(Q):	0.91	Prob(JB):	0.01			
Heteroskedasticity (H):	1.47	Skew:	-0.06			
Prob(H) (two-sided):	0.04	Kurtosis:	3.81			

FIGURE 33 – Résumé du SARIMA(2,1,4)x(5,1,1,12)

A.2.2 Test de Causalité de Granger pour les variables exogènes

	6.3 Ventes de gazole CVC-CJO (en kt)	7.3 Ventes de gazole CVS-CVC-CJO (en kt)	9. Consommation de carburants routiers (en kt)	5.6 Ventes de carburéacteur (en kt)
1	0.000223	2.445946e-51	5.206970e-24	9.388961e-02
2	0.000063	1.569777e-34	6.818243e-16	5.135333e-03
3	0.001119	4.509937e-19	7.360717e-08	4.302929e-04
4	0.001966	1.930761e-33	8.292859e-12	2.944368e-03
5	0.003735	1.123827e-30	2.195982e-12	1.469074e-05
6	0.009969	5.038842e-34	1.623807e-14	4.345525e-10
7	0.002566	4.053477e-36	3.335135e-16	9.547894e-13
8	0.002925	2.716349e-48	5.007637e-15	8.478315e-13
9	0.001017	1.075458e-36	2.867179e-14	3.576265e-15
10	0.000302	9.224572e-60	5.949502e-15	4.223402e-18
11	0.000005	1.081786e-68	2.447871e-13	3.547282e-14
12	0.000043	9.156030e-48	1.597643e-07	4.167151e-07

FIGURE 34 – Test de Causalité de Granger