



République algérienne démocratique et populaire



Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

AKLI MOHAND OULHADJ UNIVERSITY -BOUIRA-

Faculté des Sciences et Sciences Appliquées

Département de mathématiques

Mémoire Master

En mathématiques

Spécialité: Recherche opérationnelle

Thème

**Application d'Approche statistique pour la Préviation
des Ventes : Cas de l'Entreprise Cevital - Béjaia**

Réalisé par:

- GUILLAL Ahmed
- CHABANE CHAOUCHE Ghiles

Devant le jury composé de:

- HAMID Karim MAA U.A.M.O.BOUIRA **PRESIDENT**
- IFTISSEN El-ghani MAA U.A.M.O.BOUIRA **PROMOTEUR**
- BOUGHANI L'hadi MAA U.A.M.O.BOUIRA **EXAMINTATEUR**

2022/2023

Remerciements

Nous remercions avant tout le bon DIEU tout puissant, qui nous a donné la force, la volonté et le courage pour terminer ce travail. Nous tenons à exprimer notre profond respect à Mr IFTISSEN EL-Ghani pour l'aide, les orientations et les encouragements qui nous a donné tout le long de notre travail, nous lui sommes reconnaissants pour la pertinence de ces remarques et pour ces précieux conseils. Nous remercions aussi les membres du jury d'avoir accepté de nous honorer par leurs présences. Nous tenons à exprimer notre gratitude absolue à tous ceux qui nous ont aidés de près ou de loin à l'élaboration de ce modeste mémoire.

Dedicace

Je tiens à exprimer ma plus grande reconnaissance à dieu le tout puissant de m'avoir accordé la vie, la santé et la patience afin d'arriver là où je suis maintenant. Je dédie ce travail à mes chers parents pour leur encouragement leurs soutiens et leur confiance et leur immense sacrifice, à mon frère et mes sœurs à qui je souhaite beaucoup de réussite, à mes grands-parents à qui je souhaite une longue vie, à mes tantes et mes oncles pour leur encouragement, à toute la Famille Guillal et à tous mes amis et ceux qui m'ont soutenu de près ou de loin par un mot d'encouragement Et à mon cher binôme pour tous les moments de joie et de peine qu'on a partagé ensemble tout le long de notre vie universitaire

GUILLAL Ahmed.

Dedicace

Je dédie ce modeste travail: A ma mère qui a donnée sens à ma vie, symbole de tendresse, qui m'a toujours aidé dans ma vie et qui n'a cessé de m'encourager et me soutenir tout au long de mes études. à mon père qui peut être fier et trouver ici le résultat de longues années de sacrifice et de privations pour m'aider à avancer dans ma carrière académique. Puisse Dieu faire en sorte que ce travail porte fruit; merci pour les valeurs nobles, l'éducation et le soutien permanent, que Dieu vous gardes et vous protèges, à mes chère frères Koceila et Aymen, à ma chère belle-sœur Lyliya ainsi que mon neveu et petit prince adoré Syphax, et à mon cher binôme Ahmed qui m'a aidé, soutenu tout au long de notre parcours et il a été à côté de moi dans mes moments difficiles.

CHABANE Chaouche ghiles.

Abstract

In this work we aim at improving the costs of production and storage of products in a large scale enterprise through the modeling of its sales and the application of mathematical models. We are interested in the Box and Jenkins method with the models AR, MA and ARMA, covering sales trend detection of vegetable oil using monthly and emphasizing the importance of time series in econometrics for the optimization of management and the importance of sales forecast to the adaptation to market fluctuation.

Keywords: econometry, ARMA, Time series, box and Jenkins method, Forecasts.

ملخص

نهدف في هذا العمل إلى تحسين تكاليف إنتاج وتخزين المنتجات في مؤسسة كبيرة الحجم من خلال نمذجة مبيعاتها وتطبيق النماذج الرياضية. نحن مهتمون بطريقة Box et Jenkins مع نماذج AR و MA و ARMA ، التي تغطي اكتشاف اتجاه مبيعات الزيوت النباتية شهرياً وتؤكد على أهمية السلاسل الزمنية في الاقتصاد القياسي لتحسين الإدارة وأهمية توقعات المبيعات للتكيف لتقلبات السوق.

الكلمات المفتاحية : الاقتصاد القياسي ، ARMA ، السلاسل الزمنية ، Box et Jenkins ، التنبؤات.

Résumé

Dans ce travail, nous visons à améliorer les coûts de production et de stockage des produits dans une entreprise à grande échelle grâce à la modélisation de ses ventes et à l'application de modèles mathématiques. Nous nous intéressons à la méthode de Box et Jenkins avec les modèles AR, MA et ARMA, couvrant la détection des tendances des ventes d'huile végétale en utilisant des données mensuelles et en soulignant l'importance des séries temporelles en économétrie pour l'optimisation de la gestion et l'importance de la prévision des ventes pour l'adaptation à la fluctuation du marché.

Mots clés: économétrie, ARMA, Série temporelle, Méthode de box and Jenkins, Prévision.

Table des matières

Table des matières	i
Table des figures	iv
Liste des tableaux	vi
Notations et Abréviations	vii
Introduction générale	1
1 Présentation de l'organisme d'accueil	3
1.1 Historique et évolution de Cevital	4
1.2 Situation géographique	6
1.3 Missions et activités	7
1.3.1 Activités	7
1.3.2 Objectifs principaux d'entreprise	8
1.4 Gammes de produit	8
1.5 Organisation de l'entreprise	9
1.5.1 La structure et l'organisation de l'entreprise Cevital	10
1.6 Conclusion	13
2 Généralités sur les Séries temporelles	14
2.1 Introduction	14
2.2 Définition d'une série temporelles	15
2.3 Domaines d'application	16

2.4	Objectifs principaux	17
2.5	Les composantes d'une série chronologique	17
2.6	Les modèles d'une série temporelle	18
2.6.1	Modèle d'ajustement	18
2.6.2	Modèle auto-projectif	20
2.6.3	Modèle explicatif	20
2.7	Notion d'opérateur de retard L et de différenciation Δ	21
2.7.1	L'opérateur retard	21
2.7.2	L'opérateur de différenciation	21
2.8	Processus bruit blanc	22
2.9	Stationnarité	23
2.9.1	La stationnarité stricte	23
2.9.2	La stationnarité du second ordre	23
2.10	Non Stationnarité	24
2.10.1	Non stationnarité déterministe	24
2.10.2	Non stationnarité stochastique	25
2.11	Autocovariance et autocorrélation	25
2.11.1	La fonction d'autocovariance	25
2.11.2	La fonction d'autocorrélation	25
2.12	Les modèles d'une série chronologique	26
2.12.1	Modèle autorégressif	26
2.12.2	Modèle moyenne mobile	28
2.12.3	Les modèles ARMA	29
2.12.4	Les modèles ARIMA	30
2.13	Prévision par la méthode de Box-Jenkins	30
2.13.1	Identification	31
2.13.2	Estimation des paramètres	32
2.13.3	Validation	34
2.13.4	Choix du modèle	36
2.13.5	Prévision	37
2.14	Conclusion	38

3	Modélisation statistique des ventes d’huile par la technique de Box-jenkins	39
3.1	Présentation des données	39
3.2	Présentation du logiciel EViews	40
3.2.1	Généralités	40
3.2.2	Champs d’application	41
3.3	Analyse de la série vente huile	41
3.3.1	Représentation graphique	41
3.3.2	Analyse du Corrélogramme	42
3.4	Le processus de dessaisonnalisation des ventes d’huile	45
3.4.1	Estimation des coefficients saisonniers	45
3.4.2	Série corrigée des variations saisonnières (CVS)	45
3.5	Analyse de la Stationnarité de la Série Désaisonnalisée ventesa	45
3.5.1	Test de dickey-fuller augmented	47
3.6	Modélisation ARMA	52
3.6.1	Identification des ordres p et q de ARMA	52
3.6.2	Évaluation des modèles	53
3.7	Validation du modèle	55
3.7.1	Test de Akaike et Schwarz	55
3.7.2	Test de Ljung-Box	55
3.7.3	Test de Jarque-Bera	56
3.8	Prédiction des ventes d’huile pour l’année 2023	56
Conclusion générale		59
Bibliographie		60

Table des figures

1.1	Situation géographique du complexe de production de Cevital[1].	7
1.2	Structure du complexe Cevital. (Source: document interne à l'entreprise).	10
1.3	La distribution des produits du complexe Cevital vers les plateformes puis les CLR[2].	13
2.1	représentent la température mensuelle moyenne à Nottingham entre 1920-1939 [3].	15
2.2	représente L'évolution du nombre mensuel des passagers aériens par mois entre 1949-1960 [3].	16
2.3	Composantes d'une série temporelle[3].	18
2.4	Représentation graphique d'une modèle additif[3].	19
2.5	Représentation graphique d'une modèle multiplicatif[3].	20
2.6	Graphe représentant un processus bruit blanc [4].	22
2.7	Stratégie simplifiée des tests de racine unitaire [5].	33
3.1	L'icône de logiciel Eviews.	40
3.2	l'allure générale de la série vente huile.	42
3.3	la représentation du corrélogramme de la série vente huile.	43
3.4	Modèle avec tendance et constante.	43
3.5	Modèle sans tendance et avec constante.	44
3.6	Modèle sans tendance et sans constante.	44
3.7	Table des coefficients saisonniers de la série huile. (Source: Résultat obtenu à partir logiciel Eviews 12).)	45

3.8 Graphe de la série désaisonné VENTESA.	46
3.9 Les valeurs fournies par le logiciel Eviews12.	47
3.10 Evaluation de modèle 3 de la série VENTESA.	48
3.11 Evaluation de modèle 2 de la série VENTESA.	49
3.12 Evaluation de modèle 1 de la série VENTESA.	50
3.13 Test de première différenciation de la série DVENTESA.	51
3.14 la représentation graphique de la série DVENTESA autour de sa moyenne.	51
3.15 Graphique de l'autocorrélation de la série DVENTESA.	52
3.17 Évaluation du modèle MA(1) sans constante.	53
3.18 Évaluation du modèle ARMA (1,1) sans constante.	54
3.19 l'examen graphique des autocorrélations résiduelles pour effectuer le test de ljung-Box.	55
3.20 Test de normalité des résidus.	56
3.21 Tableau des prévisions pour l'année 2023.	57
3.22 un graphique qui représente à la fois les données de vente d'huile réelles et les valeur prévues à partir d'une méthode de prévision.	58

Liste des tableaux

- 1.1 Les CLR de Cevital.(Source: selon les informations recueillies au sein du service logistique.) 12
- 3.1 évolution des ventes en palettes d’huile. 39
- 3.2 les données de la Série désaisonnalisée. 46
- 3.3 Les valeurs du test de Akaike et Schwarz. 55

Notations et Abréviations

SPA	Société Par Action.
PLF	plate-forme.
CLR	Centre de Livraison Régional.
ACF	Fonction d'autocorrélation simple.
PACF	Fonction d'autocorrélation simple partielle.
AR	Autorégressif.
MA	Moyenne mobile.
ARMA	Autorégressif moyenne mobile.
ARIMA	Autorégressif moyenne mobile intégré.
ADF	Test de Dickey-Fuller augmenté.
DF	Test de Dickey-Fuller simple.
BB	Bruit blanc.
AIC	Critère d'information d'Akaike.
BIC	Critère d'information bayésien.
JB	Statistique de Jarque-Bera.
β_1	Skewness.
β_2	Kurtosis.
$\tilde{\chi}^2$	Loi de Khi-deux.
MAE	Erreur absolue moyenne.
MSE	Erreur quadratique moyenne.
RMSE	Racine carrée de l'erreur quadratique moyenne.
Q_{LB}	Statistique de Ljung et Box.

L	Opérateur de retard.
Δ	Opérateur de différentiation.
TS	Non stationnarité déterministe.
DS	Non stationnarité stochastique.
iid	Indépendante et identiquement distribuée.
E(.)	Espérance.
var(.)	Variance.
corr(.)	Corrélation.
Cov(.)	Covariance.
ϵ_t	Résidu(Erreur).
S_t	Saisonnalité.
C_t	Tendance.
\bar{X}	Moyenne empirique.
R_p	Matrice des corrélations.

Introduction générale

Les statistiques sont une discipline scientifique qui permet de tirer parti d'un volume considérable de données provenant de l'observation d'un phénomène. Elles incluent un ensemble de techniques mathématiques essentielles pour analyser relativement une grande quantité de données et visent à fournir les composantes nécessaires pour évaluer une situation et faciliter la prise de décision. À l'heure actuelle, il existe un éventail de méthodes rigoureuses permettant d'analyser et de traiter des données liées à un phénomène, qui permet de mieux comprendre l'évolution d'un phénomène donné et d'améliorer les prévisions, ce qui est généralement le cas pour les séries temporelles[6].

Une série temporelle est une séquence de points de données ou d'observations collectés ou enregistrés à des moments successifs, à des intervalles réguliers ou irréguliers. Ces données peuvent représenter divers phénomènes, tels que les ventes d'un produit, les températures climatiques, les fluctuations du marché boursier, les niveaux de pollution, les taux de chômage, et bien plus encore.

L'analyse des séries chronologiques permet de dégager des informations précieuses sur le comportement et l'évolution des phénomènes au fil du temps. Grâce à cette analyse, il est possible de détecter des tendances, des saisons, des cycles, ainsi que des modèles récurrents ou des ruptures. Ces informations sont cruciales pour comprendre le passé, éclairer le présent et anticiper l'avenir. Ce mémoire s'organise en trois parties :

Dans le premier chapitre, nous présenterons l'entreprise Cevital, où nous avons effectué notre stage, nous parlerons de son histoire, son activité, ses gammes de produits, ses capacités de production, les missions et les objectifs de l'entreprise et les différentes directions du groupe Cevital, ainsi que la position du problème.

Le deuxième chapitre développera la taxonomie d'une série temporelle, expliquant en détail les aspects théoriques nécessaires pour l'analyse et la prévision, ainsi que les outils statistiques qui nous aideront à réaliser notre étude empirique, en particulier, nous présenterons la méthode de Box-Jenkins.

Finalement, nous avons élaboré une application de la méthodologie de Box-Jenkins sur des données réelles pour obtenir le meilleur modèle qui représente cette série afin de faire la prévision.

Parmi les nombreux indicateurs clés du bien-être économique d'une entreprise, l'optimisation de la production vient tout de suite à l'esprit. La production est un processus complexe avec de nombreux sous-processus à améliorer, tels que l'ordonnancement et le contrôle et les adaptations en temps réel, en particulier dans le contexte d'un marché hautement compétitif avec des produits de consommation quotidienne. En réponse à ces exigences, de nombreux modèles ont été développés pour étudier ce processus d'optimisation pour des besoins commerciaux, visant à trouver un équilibre entre la gestion des coûts de production et la fourniture de la meilleure qualité aux consommateurs. L'objectif principal de notre travail est de s'approcher de la réponse à la question :

"Quelle est la stratégie et les méthodes de prévision efficaces pour estimer les ventes totales des huiles de l'entreprise Cevital a fin de réalisé une bonne gestion ? "

Présentation de l'organisme d'accueil



Le groupe Cevital est une société par action (SPA) dont les actionnaires principaux sont M.ISSAD Rebrab et fils créée en 1998 avec un capital privé de 68 ,760 milliards de DA. Cevital est un Groupe familial qui s'est bâti sur une histoire, un parcours et des valeurs qui ont fait sa réussite et sa renommée.

Première entreprise privée algérienne à avoir investi dans des secteurs d'activités diversifiées, elle a traversé d'importantes étapes historiques pour atteindre sa taille et sa notoriété actuelle. Porté par 18 000 employés répartis sur 3 continents, il représente le fleuron de l'économie algérienne, et œuvre continuellement dans la création d'emplois et de richesse [7].

1.1 Historique et évolution de Cevital

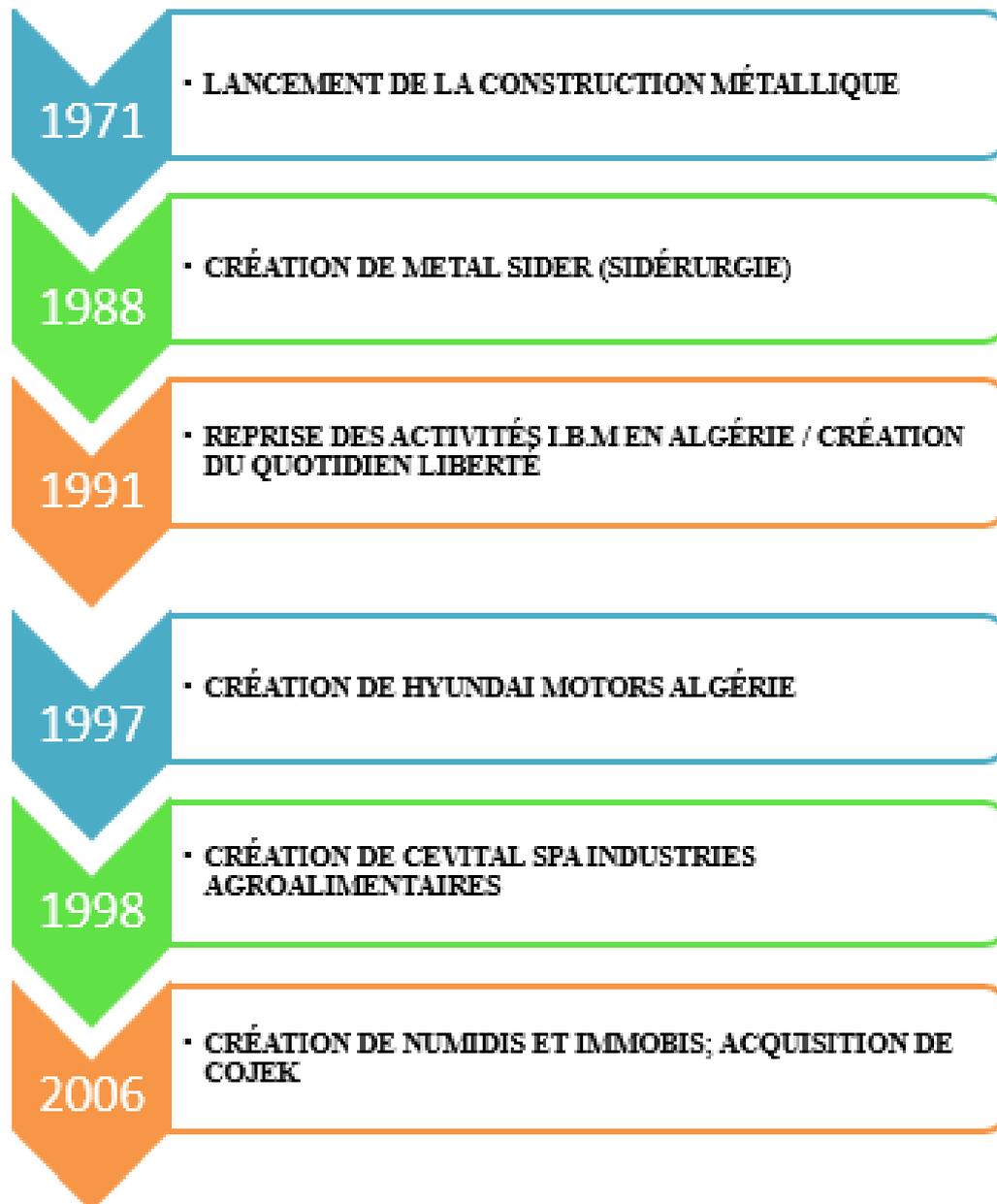
ISSAD Rebrab est né en 1944. Il crée son cabinet d'expert-comptable en 1968, puis se lance dès 1971 dans l'entrepreneuriat, en créant des sociétés dans la métallurgie et en 1991 dans la sidérurgie. À la tête du Groupe Cevital, il n'a cessé de le faire grandir en diversifiant ses activités, pour compter aujourd'hui 26 filiales dans l'industrie, l'agro-alimentaire, la grande distribution ou encore l'automobile.

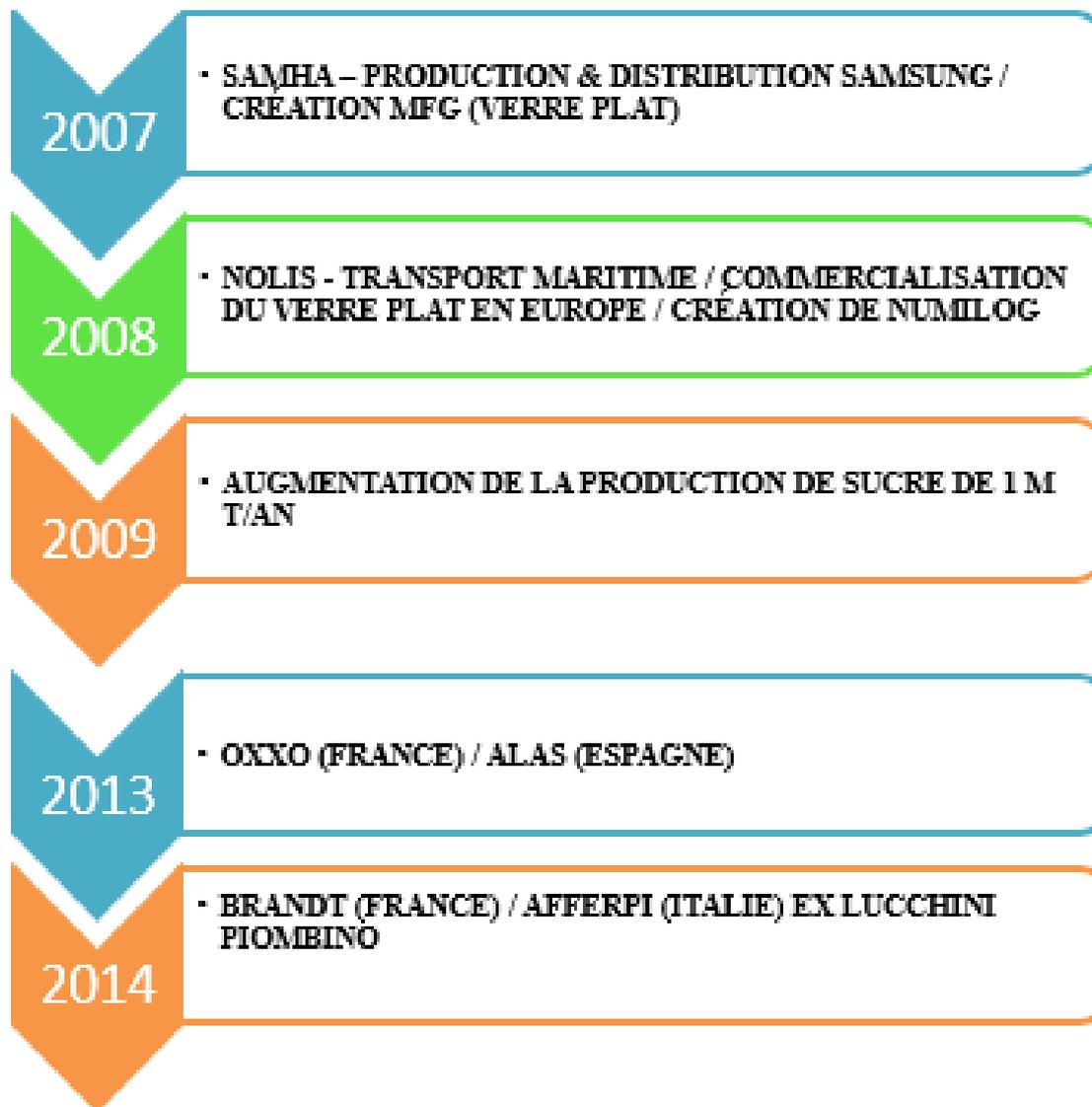
Depuis quelques années, Il développe ses activités à l'international, notamment en Europe (France, Italie, Espagne), en Tunisie, au Maroc et au Brésil. En France, Issad Rebrab a racheté en 2013 la PME Oxxo (fabricant de fenêtres haute-performance), en 2014 Brandt France, pour relancer ce leader de l'électro-ménager, et en 2015 un Luccuni un complexe sidérurgique d'acier spéciaux. Issad Rebrab est, depuis 2008, Président du Conseil d'Administration du Groupe Cevital.

Père de 5 enfants, ces derniers sont tous impliqués au sein du management du Groupe.

En 2009, Cevital a élargi sa gouvernance en décidant d'ouvrir son conseil d'administration à des membres indépendants. Une décision motivée par la volonté de soutenir la croissance de l'entreprise et d'assurer sa pérennité, à l'image des grandes entreprises internationales.

Issad Rebrab a été élu CEO de l'année à l'Africa CEO Forum 2015 et Personnalité de l'Année en 2016 par la région de la Toscane (Italie), suite à la reprise des activités sidérurgiques de Piombino [7].





(Source: élaboré par nos soins à partir des données de l'entreprise.)

1.2 Situation géographique

Le principal complexe de production de Cevital est situé dans le port de Béjaïa, plus précisément sur le quai principal, à environ 3 km de la ville, du côté sud-ouest, juste à côté des routes nationales 26 et 9. La proximité du port et de l'aéroport de Bejaïa est un avantage économique pour l'entreprise. Le complexe s'étend sur une superficie de 45 000 m², avec une capacité de stockage de 182 000 tonnes/an et un terminal de déchargement portuaire de 200 000 tonnes/heure pour la réception de matières premières. L'entreprise

possède également un réseau de distribution de plus de 52 000 points de vente sur tout le territoire national .



Figure 1.1: Situation géographique du complexe de production de Cevital[1].

1.3 Missions et activités

1.3.1 Activités

Le complexe de Cevital a été lancé en mai 1998 et sa première opération a été le conditionnement d'huile végétale en décembre de la même année. La construction de la raffinerie d'huile s'est déroulée en 1999, débutant en février par l'établissement de la plate-forme de base et devenant pleinement opérationnelle en août.

Comme nous l'avons mentionné, la principale activité de Cevital concerne les opérations liées à l'huile végétale, de la production à la commercialisation, au fil du temps d'autres produits ont été ajoutés, principalement le sucre et la margarine, les quantités journalières estimées sont les suivantes [8]:

- ✓ **Raffinage d'huile végétale:** 1800 tonnes par jour

- ✓ **Conditionnement d'huile végétale:** 1400 tonnes par heure
- ✓ **Margarine:** 600 tonnes par jour
- ✓ **Sucre:** 1600 tonnes par jour

1.3.2 Objectifs principaux d'entreprise

La mission principale de l'entreprise est de développer une production à des prix plus compétitifs et d'assurer la qualité et le conditionnement des huiles, margarines et sucres dans le but de satisfaire et de fidéliser la clientèle [8].

tels sont les principaux objectifs :

- Ses produits s'étendent dans toutes les régions du pays.
- Les graines oléagineuses sont cultivées pour l'extraction directe du pétrole brut.
- Optimiser les opportunités d'emploi sur le marché du travail.
- Les agriculteurs sont encouragés à produire des graines oléagineuses locales grâce à une aide financière.
- Moderniser ses installations en termes de machinerie et de technologie pour augmenter la production.
- Exporter ses produits sur le marché international.

1.4 Gammes de produit

Cevital se spécialise dans la production et la vente d'une large gamme de produits agroalimentaires diversifiés, principalement du sucre, de la margarine et des huiles végétales[9].

- **Les types et marques d'huiles végétales**

Il existe 3 types d'huile, chacun ayant son propre Logo et sa qualité spécifique : :

- ✓ **Fleurial** : à base de tournesol, sans cholestérol, enrichi en vitamines A, D, E.
- ✓ **Fridor et Elio** : bases diversifiées, à base de céréales, enrichies en vitamine E, sans cholestérol.

Les principales céréales utilisées dans la production sont le soja et le tournesol, ainsi que l'huile de palme. Ces produits occupent 70% de la part du marché national, avec une capacité de production qui dépasse les 570000 tonnes par an, ce qui permet une capacité d'exportation, ces produits sont exportés au Maroc, au Moyen-Orient et dans certaines parties de l'Europe.

- **Graisses végétales et margarinerie:**

Cevital produit différentes margarines telles que fleurial, matina et beurre gourmet qui sont destinées à la consommation directe, tandis que d'autres, telles que Smen Medina, sont destinées à la production, notamment aux pâtisseries. Ces margarines représentent 30% de la part du marché national et sont conditionnées en boîtes de 250g, 500g et 1,8kg. Elles sont exportées vers les pays mentionnés ci-dessus, avec une capacité de production de 180000 tonnes par an.

- **Sucre Blanc:**

La marque de sucre raffiné "SKOR" est un sucre de haute qualité à usages multiples, vendu à la fois sous forme de poudre et de liquide, dans différents conditionnements allant de 1 à 50 kilogrammes. Elle occupe 85% de la part du marché national du sucre et sa capacité annuelle varie de 650 000 à 1 800 000 tonnes.

- **Boissons:**

Cevital produit une variété de boissons, de l'eau minérale régulière ou gazeuse, des jus naturels et même des sodas, ces produits sont commercialisés sous des marques célèbres telles que "Lalla Khedjidja" et Djurdjura, les quantités de production sont attendues énormes atteignant, par exemple, 3 millions de bouteilles de sodas par jour.

1.5 Organisation de l'entreprise

Nous proposons la figure suivante pour la représenter :

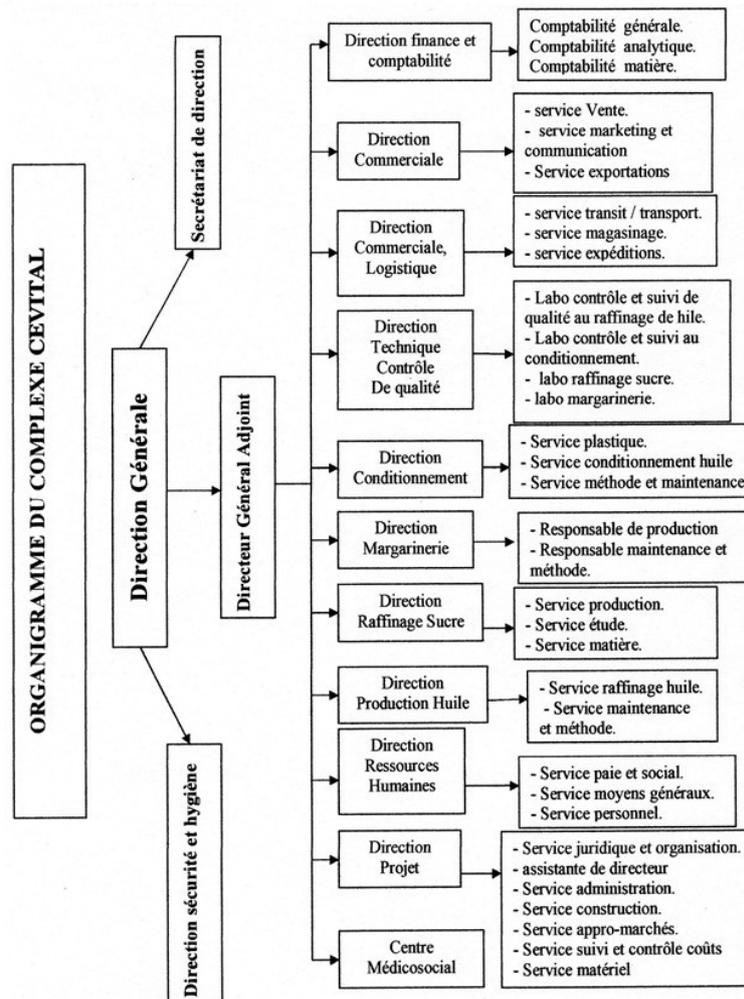


Figure 1.2: Structure du complexe Cevital. (Source: document interne à l'entreprise).

1.5.1 La structure et l'organisation de l'entreprise Cevital

Le schéma organisationnel de la direction générale est composé de différentes directions qui se présentent comme suit [10]

- ❖ la direction générale:
- ❖ DRH:
- ❖ Direction commerciale:

- Service marketing : son rôle est d'attirer la clientèle.
- Service gestion de stock : une gestion appropriée des stocks .
- Service des ventes: son rôle est de:
 - * mettre à jour les dossiers de la clientèle;
 - * établir les programmes de distributions périodiques et quotidiennes en fonction des demandes de la clientèle;
 - * établir les conventions de livraison;
 - * établir les demandes de chaque client;

❖ **Direction comptabilité et finances:**

❖ **Direction logistique:**

- -Service transit et transport.
- -Service expéditions.
- -Service magasinage.
- -Service approvisionnement.

❖ **La Direction Technique contrôle de Qualité:**

❖ **La Direction de Conditionnement:**

❖ **La Direction de Production:**

Composée de trois services de raffinage :

- **Raffinerie d'Huile:**
- **La Margarinerie:**
- **Raffinerie de sucre:**

Elle gère le processus de raffinage du sucre roux afin de le convertir en sucre blanc prêt à être emballé.

❖ **La Direction Projet:**

- **Les CLR (Centres de Livraison Régionaux)**

	EST	CENTRE	OUEST
Les CLR	25-Constantine	15-Tizi Ouzou	31-Oran
	04- Oum Al bouaghi	16-Alger	13-Tlemcen
	05-Batna	26-Médea	22-Sidi bel abbas
	23- Annaba	09- Blida	27-Mostaganem
	19-Sétif	35-Boumerdes	29-Mascara
		06-Béjaia	14-Tiaret
			48-Relizane

Table 1.1: **Les CLR de Cevital.**(Source: selon les informations recueillies au sein du service logistique.)

Au sein du service logistique, on trouve cinq correspondants CLR qui sont chargés du suivi de tous les PLF existant sur le territoire national. Le principe des CLR consiste à travailler avec zéro stock, mais vu plusieurs aléas, ils disposent toujours d'un stock de couverture suivant le programme des ventes $j+2$.

- **Les plateformes:**

Ce sont des zones de stockage externes qui sont propres à l'entreprise Cevital, il existe deux plateformes : une au centre, qui est celle de Bouira dont sa capacité de stockage est de 50000 palettes (dont 9000 palettes des produits agroalimentaires).

Et une autre à l'ouest, celle de HASSI AMER à Oran, qui a une capacité de stockage de 25000 palettes (dont 12000 palettes des produits agroalimentaires), et la dernière plateforme qui se trouve à l'est (El kharoub) avec une capacité de stockage de 4.000 emplacements. Le choix de ces plateformes n'est pas venu au hasard, mais après une étude approfondie.

La preuve est le positionnement de ces plateformes (Est, Centre, Ouest), qui permet d'alimenter la plupart des marchés du pays [7].

Dans les figures qui suivent on va voir comment se fait la distribution usine plateforme CLR:

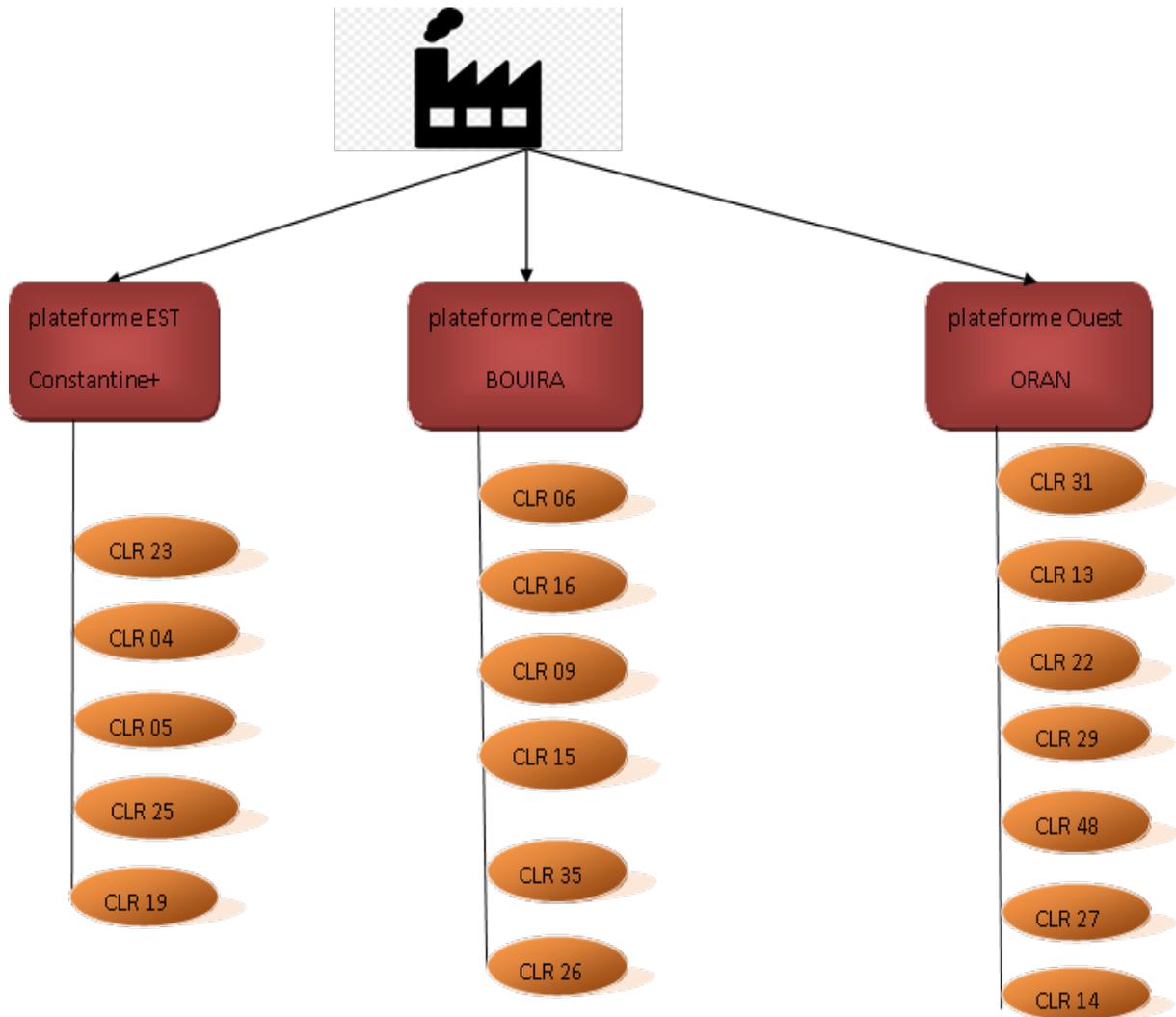


Figure 1.3: La distribution des produits du complexe Cevital vers les plateformes puis les CLR[2].

1.6 Conclusion

Cevital agro-industrie est le leader du secteur agro-alimentaire en Algérie. Sa mission principale est le développement de sa production afin d'assurer la qualité et les conditions de ses différents produits (huile, margarine, sucre, eau minérale, boissons fruitées) et de satisfaire ses différents clients par la couverture du marché national. Nous avons pu constater que l'entreprise Cevital se caractérise par des moyens efficaces (Capital) et outils modernes tel que leur logiciels, et ainsi d'un bon savoir-faire, qui l'aide à mieux gérer ses différentes fonctions.

Généralités sur les Séries temporelles

2.1 Introduction

Les séries temporelles représentent l'étude des données dans le temps, elles visent à identifier les tendances, leur stabilité et leurs différentes caractéristiques afin de mieux les comprendre et de les utiliser.

L'idée est de prendre un échantillon de données et de construire le meilleur modèle qui ajuste ces données. Ce modèle nous permet de tirer certaines conclusions sur la série. Par exemple, on peut établir une formule pour la prédiction de données, détecter certains pics ou modéliser la tendance (orientation) de la série. Un autre aspect important de la série est la composante saisonnière, c'est-à-dire la présence de cycles. Un autre concept intéressant serait le phénomène de causalité, c'est-à-dire l'influence d'une série sur une autre. Cependant il n'est pas toujours évident de choisir le bon modèle, car normalement plusieurs peuvent être de bons candidats [11].

Tout au long de ce chapitre, nous essaierons d'établir le contexte théorique des modèles que nous allons explorer, nous traiterons du modèle autorégressif AR (p), de la moyenne mobile MA (q) et du modèle mixte ARMA (p,q). Nous mettrons ces modèles en œuvre avec le modèle de Box-Jenkins.

2.2 Définition d'une série temporelles

Definition 2.1 On appelle série temporelles, ou bien encore chronique ou série temporelle, une suite réelle finie $(X_t, 1 \leq t \leq n)$ tel que $n \in \mathbb{N}^*$ de données quantitatives indexée par le temps. L'indice t représente une unité de temps qui peut être selon les cas, la seconde, la minute, l'heure, le jour, le mois, l'année, ...etc [12].

On distingue deux grands types de séries temporelles:

- **Série continue:** L'observation est prise de manière continue.
- **Série discrète:** L'observation est prise à un intervalle de temps fixe et prédéterminé.

Exemples de séries chronologiques

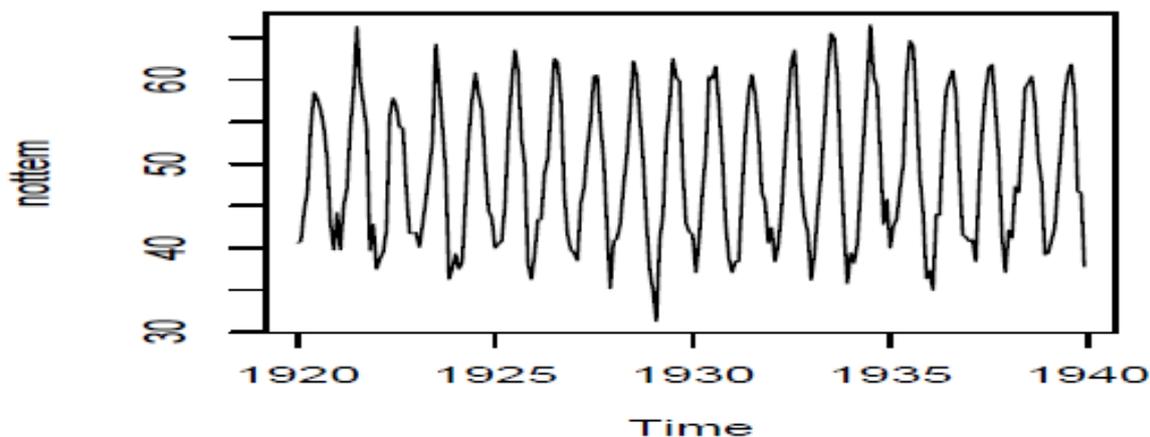


Figure 2.1: représentent la température mensuelle moyenne à Nottingham entre 1920-1939 [3].

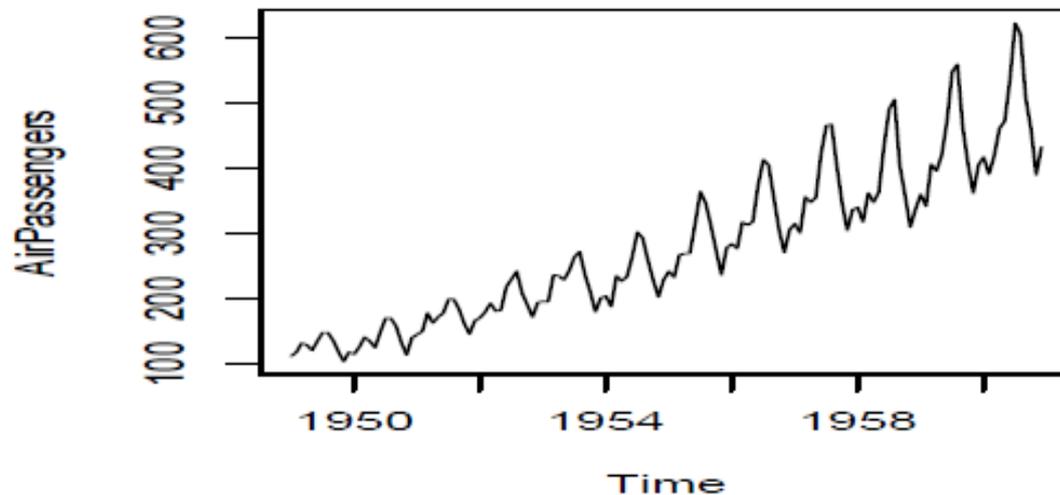


Figure 2.2: représente L'évolution du nombre mensuel des passagers aériens par mois entre 1949-1960 [3].

2.3 Domaines d'application

Il y a nombreux domaines d'application des séries chronologiques [12]:

- Finance: Cours de bourse quotidiens.
- Epidémiologie: Décès mensuels dus à des certaines maladies (COVID 19).
- Econométrie.
- Economie.
- Météorologie, pollution.
- Médecine.
- Biologie.
- Traitement du signal.
- Imagerie médicale.

2.4 Objectifs principaux

L'objectif de l'étude des séries temporelles est de faire des prédictions sur l'évolution de la série [12].

- **Analyser un phénomène:** Décrire-comprendre-juger l'évolution de la série.
- **Prévision:** faire des prédictions sur l'évolution de la série et prévoir les valeurs futures de X_t .

2.5 Les composantes d'une série chronologique

Une série X_t est composée de trois parties comme suit [13]:

1. Tendances ou trend (C_t):

Il s'agit d'un terme de la série qui exprime la façon dont le phénomène évolue à moyen terme. Cette évolution sera estimée en utilisant une forme paramétrique telle que linéaire, polynomiale, logarithmique, exponentielle, etc.

- (a) **Tendance linéaire:** La tendance la plus simple est linéaire on peut estimer les paramètres au moyen de la méthode des moindres carrés c'est une régression simple:

$$C_t = a + bt$$

- (b) **Tendance polynomiale d'ordre q:** on peut ajuster la série par un polynôme d'ordre q. les paramètres peuvent être estimés au moyen de la méthode des moindres carrés c'est une régression avec q variables explicatives :

$$C_t = b_0 + b_1t + b_2t^2 + \dots + b_qt^q$$

- (c) **Tendance logistique:** la fonction logistique permet de modéliser des processus ne pouvant dépasser une certaine valeur c:

$$C_t = \frac{c}{1 + be^{-at}} \quad \text{Ou} \quad a, b, c \in \mathbb{R}^+.$$

2. Saisonnalité (S_t):

Elle représente des effets périodiques de période connue p qui se reproduisent de façon plus ou moins identique d'une période à l'autre elle est notée par S_t $t=1, \dots, T$. Elle est généralement supposée périodique : $S_{t+P} = S_t$ d'une période p .

3. Résidus (ϵ_t):

Les résidus notée ϵ_t est la partie non structurée du phénomène. Elle est modélisée par une suite de variables aléatoire ϵ_t , centrées, non corrélées et de même variance, on parle donc de Bruit Blanc.

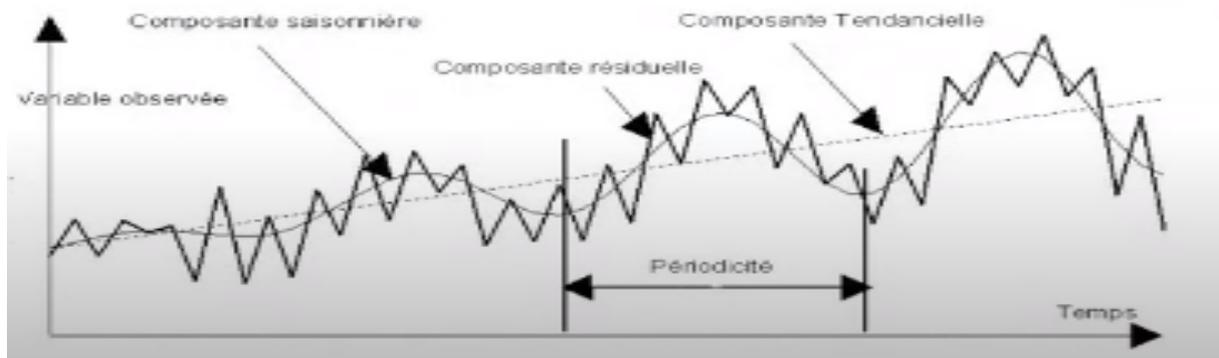


Figure 2.3: Composantes d'une série temporelle[3].

2.6 Les modèles d'une série temporelle

Il existe trois types de modèles pour décrire la façon dont la série évolue. [4]:

1. Les modèles d'ajustement.
2. Les modèles auto-projectifs.
3. Les modèles explicatifs.

2.6.1 Modèle d'ajustement

Ce type de modèle comporte deux sous-types:

* **Modèles additif:**

L'amplitude du mouvement saisonnier est constante, de type :

$$X_t = C_t + S_t + \epsilon_t$$

Où C_t représente la tendance (déterministe), S_t la saisonnalité (déterministe aussi) et ϵ_t les composantes ("erreurs au modèle") aléatoires iid.

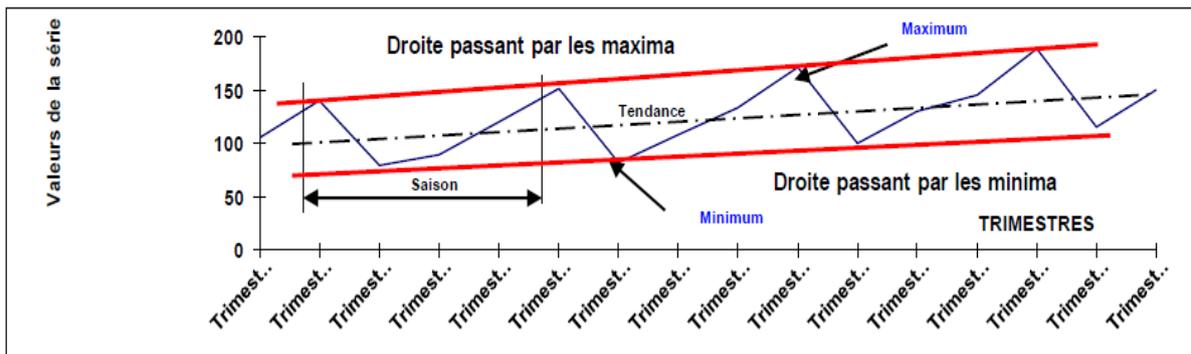


Figure 2.4: Représentation graphique d'un modèle additif[3].

Remarque 2.1 En remarque que les deux droites tracées sont parallèles entre elles

* **Modèles multiplicatif:**

L'amplitude du mouvement saisonnier est croissante ou décroissante, de type :

$$X_t = C_t * S_t * \epsilon_t$$

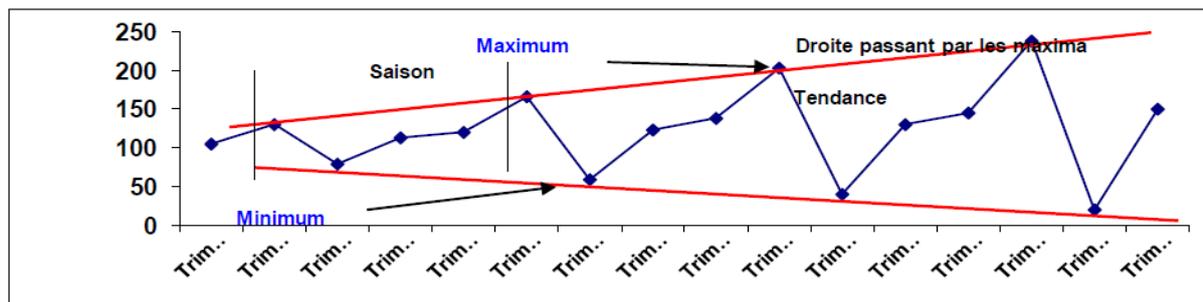


Figure 2.5: Représentation graphique d'un modèle multiplicatif[3].

Remarque 2.2 *En remarque les deux droites tracées ne sont pas parallèles.*

2.6.2 Modèle auto-projectif

Dans ce cas X_t s'écrit à partir d'une composante irrationnelle et de ses valeurs passées, d'où le "auto", on distingue deux modèles autoprojectifs :

Modèle de Box-Jenkins.

Modèle de lissage exponentiel.

2.6.3 Modèle explicatif

Ce modèle est représenté comme :

$$X_t = f(C_t) + \epsilon_t$$

- X_t : La variable aléatoire
- C_t : vecteur de variables aléatoires observables exogènes
- ϵ_t : perturbation aléatoire

f est une fonction déterministe, soit aléatoire, dans ce dernier cas, le processus C_t et ϵ_t ont certaines propriétés ou de non corrélation.

2.7 Notion d'opérateur de retard L et de différenciation

Δ

Le traitement des processus chronologiques est souvent simplifié par l'intermédiaire des deux opérateurs que nous allons présenter [14].

2.7.1 L'opérateur retard

Definition 2.2 On appelle « opérateur retard » (et l'on note L) l'opérateur qui à une donnée chronologique associe sa valeur précédente dans le temps. Il peut s'agir de quantités déterministes. Par exemple, on a

$$Lt = t - 1 \quad \text{et} \quad Lf(t) = f(t - 1)$$

Si t représente le temps sur l'espace discret Z , et f une fonction déterministe du temps. L'opérateur s'applique aussi sur des quantités aléatoires. En particulier, si (X_t) est une série chronologique définie sur Z , on a

$$LX_t = X_{t-1}.$$

La notation L provient de l'anglais *lag*, mais il est fréquent de rencontrer aussi la notation B (comme *backward*) ou, de manière plus francisée, la notation R (comme *retard*).

2.7.2 L'opérateur de différenciation

Definition 2.3 On appelle « opérateur différenciation » (et l'on note Δ) l'opérateur qui à une donnée chronologique associe la valeur de son dernier incrément dans le temps [14]. Reprenant les exemples précédents, on a

$$\Delta t = (1 - L)t = t - (t - 1) = 1 \quad \text{et} \quad \Delta f(t) = (1 - L)f(t) = f(t) - f(t - 1).$$

De même, pour une série chronologique (X_t) définie sur Z , on a

$$\Delta X_t = (1 - L)X_t = X_t - X_{t-1}.$$

On vérifie facilement que l'on a

$$X_t = LX_t + \Delta X_t$$

et donc que l'on a l'équivalence $\Delta = 1 - L$ en termes d'opérateurs.

2.8 Processus bruit blanc

Definition 2.4 On appelle processus stochastique toute famille de variables aléatoires (v.a.) $(X_t, t \in T)$ d'un espace probabilisé (Ω, A, P) vers un espace probabilisable (E, ϵ) , i.e [15].

$\forall t \in T : X_t$ est une (v.a) de (Ω, A, P) vers (E, ϵ) .

L'ensemble T est appelé espace des temps et E espace des états. Chacun de ces espaces peut être discret ou continu.

Pour $\omega \in \Omega$, on appelle trajectoire du processus la fonction (déterministe):

$$t \rightarrow X_t(\omega).$$

Definition 2.5 Un processus bruit blanc est une suite de variables aléatoires indépendantes ϵ_t telles que:

- $E(\epsilon_t) = 0 \forall t$; (hypothèse de centralité)
- $V(\epsilon_t) = \sigma^2$, minimale $\forall t$; (hypothèse d'homoscédasticité)
- $COV(\epsilon_t, \epsilon_s) = E(\epsilon_t, \epsilon_s) = 0 \forall t, s \quad t \neq s$. (absence d'autocorrélation des erreurs)

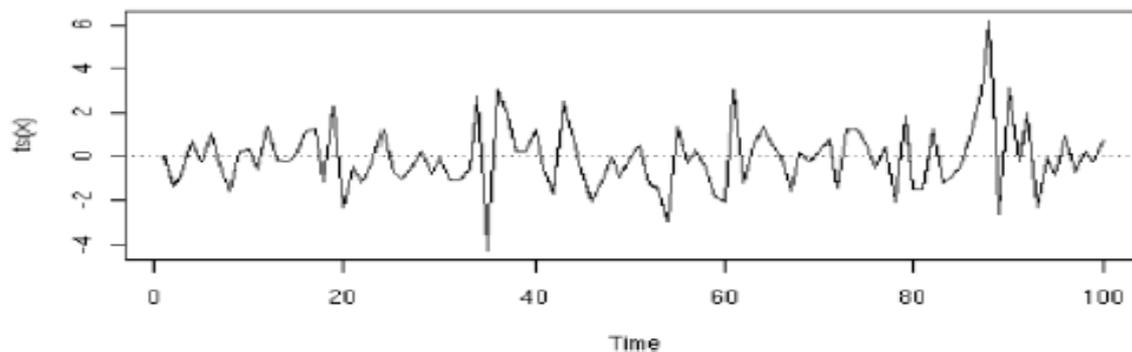


Figure 2.6: Graphe représentant un processus bruit blanc [4].

Definition 2.6 Une bruit blanc gaussien $(X_t, t \in Z)$ est une suite de variables aléatoire i.i.d de loi normale.

Definition 2.7 Un processus gaussien à temps discret $(X_t, t \in Z)$ est une série temporelle telle que la loi de n'importe quel vecteur extrait est gaussien, i.e [15].

$\forall n \in N^*, \forall (t_1, \dots, t_n) \in Z^n : (X_{t_1}, \dots, X_{t_n})$ est un vecteur gaussien.

2.9 Stationnarité

La notion de stationnarité caractérise la capacité d'un processus à se décorrélérer totalement de l'indice temporel. Ainsi la loi du processus, bien que restant souvent inconnue, sera bien plus aisée à manipuler par l'intermédiaire de ses propriétés d'espérance et de covariance [14].

2.9.1 La stationnarité stricte

Definition 2.8 Une série chronologique $(X_t, t \in T)$ est dite « strictement stationnaire » (ou « fortement stationnaire ») si, pour tout $n \geq 1$ et tout vecteur $(t_1, \dots, t_n) \in T^n$ les vecteurs

$$(X_{t_1}, \dots, X_{t_n}) \quad \text{et} \quad (X_{t_1+k}, \dots, X_{t_n+k})$$

ont la même loi, et ce pour tout décalage temporel $k \in Z$. Cette définition illustre bien le fait que l'indice temporel ne joue plus aucun rôle dans le comportement de la série chronologique puisque, en considérant tous les vecteurs de même taille translatés dans le temps, la loi du processus reste inchangée. La loi d'un processus strictement stationnaire ne dépend donc pas du temps[14].

2.9.2 La stationnarité du second ordre

Definition 2.9 Une série chronologique $(X_t, t \in T)$ est dite « stationnaire au second ordre » (ou « faiblement stationnaire », ou même simplement « stationnaire ») si sa moyenne $m(\cdot)$ et sa covariance $\tau(\cdot, \cdot)$ sont invariantes par translation dans le temps.

Autrement dit, pour tout $t, s \in T$ et tout décalage temporel $k \in Z$, on a

$$\checkmark E[X_t] = m(t) = m,$$

$$\checkmark cov(X_t, X_s) = \tau(t, s) = \tau(t + k, s + k).$$

Cette définition implique que $T = Z$, mais il n'est pas difficile de la généraliser à $T=N$. Nous pouvons en outre en déduire que la variance du processus est constante. En effet,

$$\text{var}(X_t) = \tau(t, t) = \tau(0, 0)$$

En choisissant $k=-t$ comme décalage temporel. Ainsi, l'espérance et la variance d'un processus stationnaire au second ordre sont constantes tandis que sa covariance ne dépend que du décalage temporel entre les deux valeurs considérées. Comme corollaire immédiat, nous en déduisons que la covariance, initialement fonction de deux variables t et s , est, pour un processus stationnaire au second ordre, fonction que d'une seule variable. En effet, avec $k = -s$,

$$\tau(t, s) = \tau(t - s, 0) = \tau(h, 0) = \gamma(h),$$

En posant $h=t-s$ [14].

Proposition 2.1 *Une série chronologique $(X_t, t \in T)$ strictement stationnaire est stationnaire au second ordre. Cependant, la réciproque est généralement fausse.*

2.10 Non Stationnarité

La plupart des séries sont non stationnaires, c'est-à-dire que le processus qui les décrit ne vérifie pas au moins une des conditions de la définition d'un processus stationnaire du second ordre. Ceci nous conduit à définir deux types de non stationnarité : non stationnarité déterministe et non stationnarité stochastique [16].

2.10.1 Non stationnarité déterministe

On dit que le processus Y_t est caractérisé par une non stationnarité déterministe, ou encore que le processus Y_t est TS (Trend stationary) s'il peut s'écrire:

$$Y_t = f(t) + X_t$$

Où $f(t)$ est une fonction qui dépend du temps et X_t est un processus stationnaire.

Ainsi, ce processus est rendu stationnaire en lui enlevant sa tendance déterministe:

$$Y_t - f(t) = X_t$$

2.10.2 Non stationnarité stochastique

On dit que le processus Y_t est caractérisé par une non stationnarité stochastique, ou encore que le processus Y_t est DS (Différence stationnary) si le processus différencié une fois $(1-L) Y_t$ est stationnaire. On parle aussi de processus intégré d'ordre 1, on note $Y_t \sim I(d)$:

De manière générale, on dit que le processus Y_t est un processus intégré d'ordre d , avec d le degré d'intégration, si le processus différencié d fois $(1-L)^d Y_t$ est stationnaire. On note $Y_t \sim I(d)$:

$$(1-L)^d Y_t = X_t.$$

2.11 Autocovariance et autocorrélation

2.11.1 La fonction d'autocovariance

Definition 2.10 Soit $(X_t, t \in T)$ une série chronologique stationnaire. Sa « fonction d'autocovariance » $\gamma(\cdot)$ est définie, pour tout $h \in Z$, par

$$\gamma(h) = \tau(h, 0) = \text{cov}(X_h, X_0).$$

Son abréviation usuelle est « CVF: Covariance Fonction ». Puisque (X_t) est stationnaire, il est clair que pour tout, $t \in T$, on a

$$\text{cov}(X_{t+h}, X_t) = \dots = \text{cov}(X_{h+1}, X_1) = \text{cov}(X_h, X_0) = \gamma(h),$$

Ce qui montre bien que l'autocovariance d'un processus stationnaire ne dépend que du décalage temporel h entre deux observations, et non de l'instant t de la mesure.

Comme nous l'avons montré, la fonction d'autocovariance du bruit blanc est caractérisée par:

$$\gamma(h) = \begin{cases} \sigma^2 & \text{si } h = 0 \\ 0 & \text{si } h \neq 0 \end{cases}$$

2.11.2 La fonction d'autocorrélation

Definition 2.11 Soit $(X_t, t \in T)$ une série chronologique stationnaire. Sa « fonction d'autocorrélation » $\rho(\cdot)$ est définie, pour tout $h \in Z$, par

$$\rho(h) = \text{coor}(X_h, X_0) = \frac{\gamma(h)}{\gamma(0)}.$$

Son abréviation usuelle est «ACF: Autocorrelation Function». Puisque (X_t) est stationnaire, il est clair que pour tout $t \in T$, on a

$$\text{coor}(X_{t+h}, X_t) = \dots = \text{coor}(X_{h+1}, X_1) = \text{corr}(X_h, X_0) = \rho(h).$$

Ce qui montre également que l'autocorrélation d'un processus stationnaire ne dépend que du décalage temporel h entre deux observations, et non de l'instant t de la mesure. De même, la fonction d'autocorrélation du bruit blanc est caractérisée par

$$\rho(h) = \begin{cases} 1 & \text{si } h = 0 \\ 0 & \text{si } |h| \neq 0 \end{cases}$$

Proposition 2.2 Soit $(X_t, t \in T)$ une série chronologique stationnaire. Alors, sa fonction d'autocovariance et sa fonction d'autocorrélation sont paires.

2.12 Les modèles d'une série chronologique

2.12.1 Modèle autorégressif

Definition 2.12 On dira que le processus X_t est un modèle autorégressif d'ordre p , s'il existe un Bruit Blanc $\{\epsilon_t\}$, et s'il peut s'écrire sous la forme suivant [13]:

$$X_t = \sum_{j=1}^p \theta_j X_{t-j} + \epsilon_t.$$

θ_j : ($j=1, \dots, p$) les termes du modèle

$$X_t - \sum_{j=1}^p \theta_j X_{t-j} = \epsilon_t \iff (1 - \theta_1 L - \dots - \theta_p L^p) X_t = \epsilon_t \iff \phi(L) X_t = \epsilon_t.$$

✓ **Représentation stationnaire:**

Si le polynôme ϕ a toutes ses racines de module différent de 1, on peut inverser l'opérateur $\phi(L)$.

On en déduit que l'équation admet une solution unique, avec une écriture MA(∞):

$$X_t = \phi(L^{-1})\epsilon_t = \sum_{j=-\infty}^{\infty} f_t \epsilon_{t-j}.$$

On peut alors montrer que l'on a $\sum_{j=-\infty}^{\infty} |f_t| < \infty$. et donc la représentation est stationnaire.

✓ Estimation des paramètres:

On dispose d'une observation $\{x_0, \dots, x_T\}$ d'un processus stationnaire X_t suivre un modèle AR(p), c'est-à-dire

$$X_t = \sum_{j=1}^p \theta_j X_{t-j} + \epsilon_t.$$

$t \in Z, X_0 = x_0$ et $\theta_1, \dots, \theta_p$ des paramètres inconnus.

- * **Maximum vraisemblance:** l'estimation d'un modèle AR(p) par la méthode du maximum de vraisemblance est délicate car la fonction de vraisemblance est très complexe et n'a pas de dérivée analytique. Cette difficulté provient de l'interdépendance des valeurs, ainsi que du fait que les observations antérieures ne sont pas toutes disponibles pour les p premières valeurs.
- * **Méthode des moments:** utilisation des équations de Yule-Walker.
- * **Moindres carrés.**

✓ Autocorrélation d'un AR(p)

Sous l'hypothèse de stationnarité, on a $E(X_t) = 0$, il est clair que la fonction d'autocovariance d'un modèle AR(p) s'écrit:

$$\begin{aligned} \gamma_x(h) &= \text{cov}(X_t, X_{t-h}) = E(X_t X_{t-h}) \\ E(X_t X_{t-h}) &= \theta_1 E(X_t X_{t-h}) + \dots + \theta_p E(X_{t-p} X_{t-h}) E(\epsilon_t X_{t-h}) \end{aligned}$$

$$E(\epsilon_t X_{t-h}) = \begin{cases} \delta_\epsilon^2 & \text{si } h = 0 \\ 0 & \text{sinon} \end{cases}$$

$$\gamma_x(0) = \sum_{j=1}^p \theta_j \gamma_x(j) + \delta_\epsilon^2 \text{ et } \gamma_x(h) = \sum_{j=1}^p \theta_j \gamma_x(h-j), \quad h \geq 1 \quad (2.1)$$

On divise (2.1) par $\gamma_x(0)$, nous obtenons la fonction d'autocorrélation :

$$\varphi_x(h) = \frac{\gamma_x(h)}{\gamma_x(0)} = \frac{\sum_{j=1}^p \theta_j \gamma_x(h-j)}{\gamma_x(0)} = \sum_{j=1}^p \theta_j \frac{\gamma_x(h-j)}{\gamma_x(0)} = \sum_{j=1}^p \theta_j \varphi_x(h-j), \quad h \geq 1 \quad (2.2)$$

(2.1) et (2.2) s'appelle le système de Youle-Walker :

$$\begin{cases} \gamma_x(h) = \sum_{j=1}^p \theta_j \gamma_x(h-j) \\ \varphi_x(h) = \sum_{j=1}^p \theta_j \varphi_x(h-j) \end{cases}$$

2.12.2 Modèle moyenne mobile

Definition 2.13 On appelle modèle moyenne mobile d'ordre q , un modèle stationnaire X_t sous forme [13]:

$$X_t = c + \epsilon_t + \sum_{j=1}^q \theta_j \epsilon_{t-j}, \quad t \in Z$$

θ_j : Réels non nul et ϵ_t est $BB(0, \delta^2)$.

✓ Autocorrélation d'un MA(q):

La fonction d'autocovariance d'un MA(q) est:

$$\gamma_x(h) = cov(X_t, X_{t-h})$$

$$= E((\epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \dots + \theta_q \epsilon_{t-q})(\epsilon_{t-h} + \theta_1 \epsilon_{t-h-1} + \dots + \theta_q \epsilon_{t-h-q}))$$

$$= \begin{cases} (1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2) \delta_\epsilon^2 & h = 0 \\ (-\theta_k + \theta_1 \theta_{h+1} + \dots + \theta_{q-h} \theta_q) \delta_\epsilon^2 & 0 < h \leq q \\ 0 & h < q \end{cases}$$

D'où la fonction d'autocorrélation est

$$\rho_h = \frac{\gamma_h}{\gamma_0} \begin{cases} 1 & h = 0 \\ \frac{-\theta_h + \theta_1\theta_{h+1} + \dots + \theta_{q-h}\theta_q}{1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2} & 0 < h \leq q \\ 0 & h < q \end{cases}$$

Remarque 2.3 La fonction d'autocorrélation d'un modèle $MA(q)$ s'annule au-delà du rang

($h < q$):

Propriété 2.1 Pour un $MA(\infty)$, on a les propriétés suivantes

1. $E(X_t) = 0, \forall t$
2. $\text{var}(X_t) = \delta_\epsilon^2 \sum_{j=-\infty}^{+\infty} \theta_j^2$
3. $\text{cov}(X_t, X_{t-h}) = \gamma_x(h) = \gamma_\epsilon^2 \sum_{j=-\infty}^{+\infty} \theta_j \theta_{j+|h|}$

2.12.3 Les modèles ARMA

Definition 2.14 Un processus stationnaire X_t admet une représentation ARMA (p, q) minimale [13]:

$$X_t - \sum_{j=1}^p \theta_j \epsilon_{t-j} = \epsilon_t - \sum_{i=1}^q \alpha_i \epsilon_{t-i} \iff \phi(L)X_t = \Theta(L)\epsilon_t$$

S'il satisfait les conditions suivantes :

1. $\theta_j \neq 0, \alpha_i \neq 0$.
2. les polynômes ϕ et Θ toutes leurs racines de module strictement supérieur à 1.
3. ϕ et Θ n'ont de racine commune.
4. ϵ_t et un bruit blanc, de variance $\delta_\epsilon^2 \neq 0$.

✓ Autocorrélation d'un ARMA (p, q)

Proposition 2.3 Soit (X_t) un processus ARMA (p, q), alors les autocorrélations $\gamma(h)$ satisfont

$$\gamma(h) - \sum_{i=1}^p \phi_i \gamma(h-i) = \delta^2 [\theta_h + h_1 \theta_{h+1} + \dots + h_{q-h} \theta_q] \quad \text{pour} \quad 0 \leq h \leq q$$

Propriété 2.2 Soit (X_t) un processus ARMA (p, q) , alors les autocorrélations $\gamma(h)$ satisfont

$$\gamma(h) - \sum_{i=1}^p \phi_i \gamma(h-i) = 0 \quad \text{pour} \quad h \geq q+1$$

Où les h_i correspondent aux coefficients de la forme MA(∞) de (X_t) :

$$X_t = \sum_{j=0}^{\infty} h_j \epsilon_{t-j}.$$

2.12.4 Les modèles ARIMA

Definition 2.15 Pour $d \geq 1$, le processus X_t est un ARIMA (p, d, q) si le processus $Y_t = (1-L)^d X_t$ est un processus ARMA (p, q) de moyenne nulle. X_t satisfait donc une équation de la forme : [10]

$$\Phi(L)(1-L)^d X_t = \Theta(L)\epsilon_t.$$

Φ et Θ : des polynômes de degrés respectifs p et q

L : l'opérateur de retard

$$\Phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p$$

$$\Theta(z) = 1 + \theta_1 z + \dots + \theta_q z^q$$

Remarque 2.4 On dit aussi que le processus ARIMA (p, d, q) est un processus ARMA (p, q) de moyenne nulle "intégré" d fois.

2.13 Prédiction par la méthode de Box-Jenkins

La méthodologie de Box et Jenkins permet de déterminer le modèle ARIMA adéquat pour la modélisation d'une série chronologique, donc il s'agit de construire un modèle restituant le mieux possible le comportement d'une série temporelle. Cette méthodologie de Box-Jenkins utilise une approche itérative en cinq étapes

- ☞ Identification à priori.
- ☞ Estimation des modèles retenus.
- ☞ Validation.

☞ Choix du modèle.

☞ Prévision.

2.13.1 Identification

L'identification consiste à spécifier les trois paramètres p , d et q du modèle ARIMA (p , d , q). La stationnarité du modèle est testée par étude graphique, de corrélogramme et test de dickey fuller augmenté. Si la série n'est pas stationnaire, nous devons la transformer en une série stationnaire. L'ordre d'intégration d est le nombre de fois que la série initiale a été différenciée pour obtenir la stationnarité. Les autocorrélations et les autocorrélations partielles permettent d'estimer les ordres p et q pour les modèles AR et MA [5].

Test de Dickey Fuller simple : Dickey et Fuller sont les premiers à fournir un ensemble d'outils statistiques formels pour détecter la présence d'une racine unitaire dans un processus autorégressif du premier ordre, ce test permet de tester l'hypothèse [5].

$$\begin{cases} H_0 : \text{Le modèle a une racine unitaire.} \\ H_1 : \text{le modèle n a pas de racine unitaire.} \end{cases}$$

Ce test est regroupé en 4 cas :

$$\begin{aligned} y_t &= \rho y_{t-1} + \epsilon_t & H_0 : \rho &= 1, \\ y_t &= \alpha + \rho y_{t-1} + \epsilon_t & H_0 : \alpha &= 0 \quad \text{et} \quad \rho = 1, \\ y_t &= \alpha + \rho y_{t-1} + \epsilon_t & H_0 : \alpha &\neq 0 \quad \text{et} \quad \rho = 1, \\ y_t &= \alpha + \beta_t + \rho y_{t-1} + \epsilon_t & H_0 : \alpha = 0, \beta &= 0 \quad \text{et} \quad \rho = 1, \end{aligned}$$

➤ Test de Dickey Fuller augmenté:

Dickey et Fuller en 1981 étendent ensuite cette procédure de test à des processus autorégressifs d'ordre p , il s'agit alors des tests ADF «Augmented Dickey-Fuller » [5].

Ce test permet de tester

$$\begin{cases} H_0 : \text{Le modèle a une racine unitaire.} \\ H_1 : \text{le modèle n a pas de racine unitaire.} \end{cases}$$

Ces tests peuvent être regroupés en 4 cas :

$$y_t = \rho y_{t-1} + \sum_{i=1}^p \alpha_i y_{t-i} + \epsilon_t \quad H_0 : \rho = 1$$

$$y_t = \alpha + \rho y_{t-1} + \sum_{i=1}^p \alpha_i y_{t-i} + \epsilon_t \quad H_0 : \alpha = 0 \text{ et } \rho = 1$$

$$y_t = \alpha + \rho y_{t-1} + \sum_{i=1}^p \alpha_i y_{t-i} + \epsilon_t \quad H_0 : \alpha \neq 0 \text{ et } \rho = 1$$

$$y_t = \alpha + \beta_t + \rho y_{t-1} + \sum_{i=1}^p \alpha_i y_{t-i} + \epsilon_t \quad H_0 : \alpha = 0, \beta = 0 \text{ et } \rho = 1$$

Pour simplifier, on écrira :

$$\Delta y_t = \phi y_{t-1} + \sum_{i=1}^p \alpha_i \Delta y_{t-i} + \epsilon_t, \text{ avec } \phi = \rho - 1 \quad \text{''Modèle[1]''},$$

$$\Delta y_t = \alpha + \phi y_{t-1} + \sum_{i=1}^p \alpha_i \Delta y_{t-i} + \epsilon_t, \quad \text{Modèle[2]''},$$

$$\Delta y_t = \alpha + \beta_t + \phi y_{t-1} + \sum_{i=1}^p \alpha_i \Delta y_{t-i} + \epsilon_t, \quad \text{''Modèle[3]''},$$

Nous pouvons résumer la stratégie d'ADF à partir de schéma suivant :

2.13.2 Estimation des paramètres

L'estimation des paramètres d'un modèle ARIMA (p, d, q) lorsque p, d, q sont supposés connus peut se réaliser par différentes méthodes dans le domaine temporel, et parmi ces méthodes on a :

1. Maximum de vraisemblance.
2. Dans le cas q = 0, on utilise les équations de Yule Walker [17].

1. Estimation par la méthode de maximum de vraisemblance:

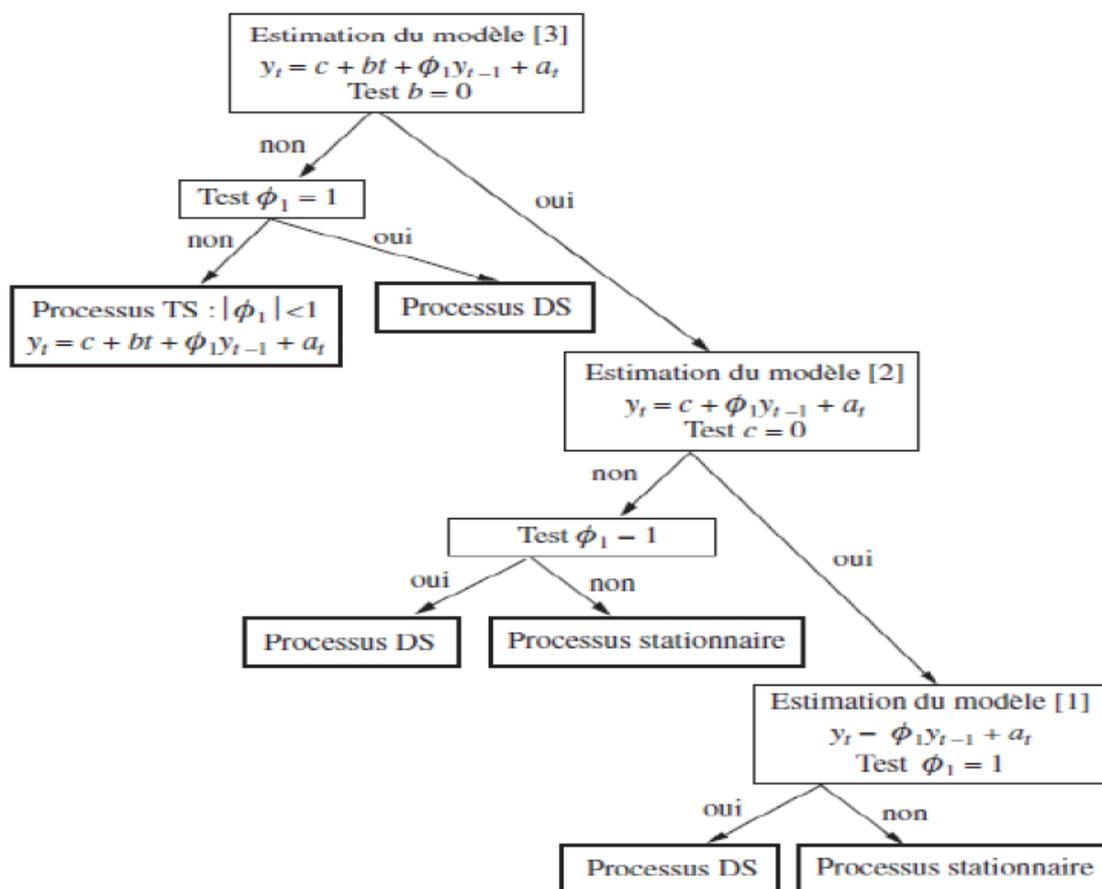


Figure 2.7: Stratégie simplifiée des tests de racine unitaire [5].

Soit $X_t \sim ARMA(p,q)$. Pour appliquer cette méthode, il est nécessaire de supposer que $\epsilon_t \sim N(0, \sigma_\epsilon^2)$ (ϵ_t est par hypothèse un Bruit Blanc).

$$L(X, \phi, \theta, \sigma_\epsilon^2) = \frac{1}{(2\pi\sigma_\epsilon^2)^{\frac{T}{2}}} \frac{1}{\sqrt{\det\Omega}} \exp\left(-\frac{1}{2\sigma_\epsilon^2} X^t \Omega^{-1} X\right).$$

Où : Ω est la matrice $(T \times T)$ de variance-covariance du vecteur, $X = (X_1, \dots, X_T)^t$ et $L(X, \phi, \theta, \sigma_\epsilon^2)$ est la fonction de vraisemblance.

La maximisation et le calcul de cette quantité étaient relativement difficiles à cause du calcul de l'inverse Ω^{-1} et du déterminant de Ω , surtout quand T est assez grand. Newbold a proposé une autre expression plus facile à calculer. On note par H la matrice triangulaire inférieure, à éléments positifs sur la diagonale, telle que $HH^t = \Omega$ (décomposition de Cholesky), et soit e le vecteur où $e = H^{-1}X$. La log-vraisemblance du modèle s'écrit donc :

$$\log L = -\frac{T}{2} \log 2\pi - \frac{1}{2} \log \sigma_\epsilon^2 - \frac{1}{2} |\det\Omega| - \frac{1}{2\sigma_\epsilon^2} X^t \Omega^{-1} X$$

$$\begin{aligned}
&= -\frac{T}{2} \log 2\pi - \frac{T}{2} \log(e^t e) - \log |\det H| \\
&= -\frac{T}{2} \log(|\det H|^{\frac{1}{T}} e^t e |\det H|^{1/T}).
\end{aligned}$$

La méthode de maximum de vraisemblance revient finalement à chercher le minimum de : $L = (|\det H|^{\frac{1}{T}} e^t e |\det H|^{1/T})$.

2. Estimation par les équations de Yule-Walker:

Dans le cas d'un modèle AR(p), on peut utiliser le système de Yule-Walker, pour estimer les paramètres $\phi = (\phi_1, \phi_2, \dots, \phi_p)$, comme suit

$$\hat{\phi} = \hat{R}_p^{-1} \hat{\rho}$$

Où : \hat{R}_p est la matrice des autocorrélations estimées et $\hat{\rho} = (\hat{\rho}_X(1), \hat{\rho}_X(2), \dots, \hat{\rho}_X(p))^t$ avec

$$\hat{\rho}_X(h) = \frac{\hat{\gamma}_X(h)}{\hat{\gamma}_X(0)}$$

De plus, l'estimateur de la variance σ_ϵ^2 est donné par :

$$\sigma_\epsilon^2 = \hat{\gamma}_X(0) \left(1 - \sum_{j=1}^p \hat{\phi}_j \hat{\rho}_X(j)\right).$$

2.13.3 Validation

Après l'étape l'identification, si on dispose de plusieurs modèles, il est nécessaire d'effectuer des tests sur les paramètres et sur les résidus [16].

Test de significativité des coefficients

Les coefficients du modèle doivent être significativement différent de zéro, pour ce faire on utilise le test classique de student [16].

On rejette l'hypothèse nulle $H_0 : \theta_j = 0$, si $|t_c| > |\Gamma_{T-q}^\alpha|$, où $|t_c| = \left| \frac{\hat{\theta}}{\hat{\sigma}_{\hat{\theta}}} \right|$.

Test sur le bruit blanc

Pour que les modèles obtenues soient valides, il convient de vérifier que les résidus estimés, notés $\hat{\epsilon}_t$ suivent un bruit blanc, non autocorrélé et de même variance $\hat{\sigma}_\epsilon^2$ et qu'ils suivent une loi normale. Si ces hypothèses ne sont pas rejetées, on peut alors mener des tests sur les paramètres

1. Test de Box-Pierce et Ljung-Box:

Le test de Box-Pierce permet d'identifier les processus de bruit blanc. Ce test s'écrit:

$$\begin{cases} H_0 : \rho(1) = \rho(2) = \dots = \rho(h) = 0. \\ H_1 : \text{il existe } i \text{ telle que } \rho(i) \neq 0. \end{cases}$$

Pour effectuer ce test, on utilise la statistique de Box et Pierce (1970), donnée par:

$$Q = T \sum_{k=1}^h \hat{\rho}^2(k),$$

Où h est le nombre de retards, T est le nombre d'observations et $\hat{\rho}(k)$ l'autocorrélation empirique d'ordre k . Sous l'hypothèse que les résidus suivent un bruit blanc. Cette statistique, suit une loi du khi-deux à h degré de liberté, notée $\tilde{\chi}_h^2$. Nous rejetons donc H_0 si $Q > \tilde{\chi}_h^2(1 - \alpha)$, où $\tilde{\chi}_h^2(1 - \alpha)$ est le quantile d'ordre $(1 - \alpha)$ de la loi du $\tilde{\chi}_h^2$

Ljung et Box (1978) ont améliorés ce test en considérant la statistique

$$Q_{LB} = T(T + 2) \sum_{k=1}^h \frac{\hat{\rho}^2(k)}{T - j}.$$

2. Test de normalité:

Plusieurs des modèles de séries temporelles supposent que les résidus sont indépendants et distribués selon la loi Normale. Un des tests permettant de vérifier la normalité des résidus est celui de Jarque-Bera. Les hypothèses à confronter sont:

$$\begin{cases} H_0 : \epsilon_t \simeq N(0, 1). \\ H_1 : \epsilon_t \not\simeq N(0, 1). \end{cases}$$

La statistique du test est définie par :

$$JB = \frac{T}{6}\beta_1^2 + \frac{T}{24}(\beta_2 - 3)^2,$$

Où β_1 et β_2 sont respectivement les coefficients d'asymétrie (Skewness) et d'aplatissement (Kurtosis), donné par:

$$\beta_1 = \frac{1}{T} \sum_{t=1}^T \left(\frac{X_t - \bar{X}}{S} \right)^3 \text{ et } \beta_2 = \frac{1}{T} \sum_{t=1}^T \left(\frac{X_t - \bar{X}}{S} \right)^4,$$

Ici, \bar{x} et S sont respectivement la moyenne et l'écart type empiriques. La statistique JB suit asymptotiquement la loi du $\tilde{\chi}_2^2$. On rejette donc l'hypothèse de normalité des résidus H_0 , si $JB > \tilde{\chi}^2$.

2.13.4 Choix du modèle

Une fois quelques modèles choisis, l'erreur de prévision dépend de la variance σ_ϵ^2 , notre but est donc de choisir le meilleur modèle qui minimise cette erreur par l'intermédiaire de l'un des critères suivants : [15]

Critères standard:

- **L'erreur quadratique moyenne (MSE: Mean Square Error):**

$$MSE = \frac{1}{T} \sum_{t=1}^T \epsilon_t^2$$

- La racine de l'erreur quadratique moyenne (RMSE: Root Mean Square Error):

$$RMSE = \sqrt{\frac{1}{T} \sum_{t=1}^T \epsilon_t^2}.$$

- **L'erreur absolue moyenne (MAE: Mean absolute Error):**

$$MAE = \frac{1}{T} \sum_{t=1}^T |\epsilon_t|.$$

Critères d'information:

- Le critère AIC (Akaike Information Criterion):

$$AIC(p, q) = \ln \hat{\sigma}_\epsilon^2 + \frac{2(p+q)}{T},$$

- Le critère BIC (Schwartz Bayesian Information Criterion):

$$BIC(p, q) = \ln \hat{\sigma}_\epsilon^2 + \frac{(p+q) \ln(T)}{T},$$

2.13.5 Prédiction

C'est la dernière étape de la méthodologie de Box and Jenkins. Etant donné une série stationnaire, observée entre 1 et T, on cherche à faire de la prédiction à horizon h, et donc de prévoir X_{T+1}, \dots, X_{T+h} [5].

Il s'agit de calculer les prévisions optimales du modèle ARIMA estimé, à savoir X_T (h) la prédiction de X_{T+h} sachant l'ensemble d'information disponible en T, noté

$$\hat{X}_{T+h} = E\left(\frac{X_{T+h}}{X_t}, X_{T-1}, \dots, X_1\right).$$

Prédiction d'un modèle AR(p) Le modèle s'écrit :

$$X_t = \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + \epsilon_t$$

La prédiction optimale à la date T + 1, faite à la date T est

$$\hat{X}_{T+1} = E\left(\frac{X_{T+1}}{X_T}, X_{T-1}, \dots, X_1\right),$$

Donc

$$\hat{X}_{T+1} = \phi_1 X_T + \dots + \phi_p X_{T+1-p}$$

De façons analogue

$$X_{T+h} = \phi_1 X_{T+h-1} + \dots + \phi_p X_{T+h-p} + \epsilon_{T+h}.$$

Donc

$$\hat{X}_{T+h} = E\left(\frac{X_{T+h}}{X_T}, X_{T-1}, \dots, X_1\right).$$

De façons récurrente

$$\hat{X}_{T+h} = \begin{cases} \phi_1 \hat{X}_{T+h-1} + \dots + \phi_{h-1} \hat{X}_{T+1} + \phi_h X_T + \dots + \phi_p X_{T+h-p} & \text{pour } h \leq p \\ \phi_1 \hat{X}_{T+h-1} + \dots + \phi_p \hat{X}_{T+h-p} & \text{pour } h > p \end{cases}$$

Prédiction d'un modèle MA(q)

Soit le modèle MA(q)

$$X_t = \epsilon_t + \theta_1 \epsilon_{t-1} + \dots + \theta_q \epsilon_{t-q} = \Theta(L)\epsilon_t.$$

La prévision optimale à la date $T + 1$ est

$$\hat{X}_{T+1} = \theta_1 \epsilon_T + \dots + \theta_q \epsilon_{T+1-q}$$

De façons analogue

$$\hat{X}_{T+h} = \begin{cases} \theta_h \hat{X}_T + \dots + \theta_q \hat{X}_{T+h-q} & \text{pour } h \leq q \\ 0 & \text{pour } h > q \end{cases}$$

Prévision d'un modèle ARMA (p, q)

(X_t) est un ARMA (p, q), alors :

$$X_t = \sum_{i=1}^p \phi_i X_{t-i} + \epsilon_t - \sum_{j=1}^q \theta_j \epsilon_{t-j}$$

Donc

$$X_{T+h} = \sum_{i=1}^p \phi_i X_{T+h-i} + \epsilon_t - \sum_{j=1}^q \theta_j \epsilon_{T+h-j}.$$

On a alors

$$\hat{X}_{T+h} = \sum_{i=1}^p \phi_i \hat{X}_{T+h-i} + \epsilon_t - \sum_{j=1}^q \theta_j \hat{\epsilon}_{T+h-j}$$

Où

$$\hat{\epsilon}_{T+h-j} = \begin{cases} 0 & \text{si } h > j \\ \epsilon_{T+h-j} & \text{si } h \leq j \end{cases}$$

2.14 Conclusion

Dans le chapitre suivant, nous allons mettre en pratique les étapes de Box-Jenkins sur le cas réel de vente de produits alimentaires par l'entreprise Cevital.

Modélisation statistique des ventes d'huile par la technique de Box-jenkins

Ce chapitre est dédié à l'étude des ventes d'huile pour l'année 2023. En utilisant la méthode de box-Jenkins, nous examinons la série de données pour identifier le modèle le plus approprié qui servira de base à nos prévisions.

3.1 Présentation des données

Les données que nous utiliserons dans cette étude ont été généreusement fournies par le service de production. Cette source de données nous a offert une opportunité précieuse pour analyser et comprendre les aspects clés du processus de vente d'huile de 2019 à 2022.

annéesmois	2019	2020	2021	2022
Janvier	2230	8019	13382	16615
Février	1941	3214	11649	13897
Mars	2440	13824	14641	14509
Avril	2265	14820	13591	13805
Mai	2307	13161	13847	14440
Juin	2053	11829	12321	14789
Juillet	1973	12642	11843	18566
Août	2268	12861	13609	13160
Septembre	2023	13248	12142	14506
Octobre	2553	14003	15318	14718
Novembre	1872	12333	11232	12940
Décembre	1951	12595	11708	14975

Table 3.1: évolution des ventes en palettes d'huile.

3.2 Présentation du logiciel EViews

3.2.1 Généralités

Eviews est un logiciel qui permet de réaliser une analyse des données, notamment des régressions et des prévisions. C'est donc un outil performant d'analyse des séries temporelles. Aujourd'hui, Eviews connaît une large application dans beaucoup de domaines. Il offre la possibilité d'entrer les données à partir du clavier ou de fichiers, de créer de nouvelles séries à partir de celles existantes, de faire des analyses statistiques des relations entre des séries. Eviews utilise les caractéristiques visuelles des logiciels sous Windows. L'utilisateur peut utiliser la souris pour faire des opérations à l'aide des menus et des boîtes de dialogue standards de Windows. Les résultats apparaissent en Windows et peuvent être manipulés par les techniques standard. Ils peuvent être copiés puis collés dans le Word pour faciliter le processus de rédaction du rapport. Alternativement, Eviews dispose d'un langage de programmation et de commandes assez variées que pourrait utiliser l'utilisateur. Enfin, Eviews contient un menu d'aide en ligne riche en documentation sur la mise en œuvre des procédures d'analyse, des commandes et des opérations qu'il permet de faire[18].

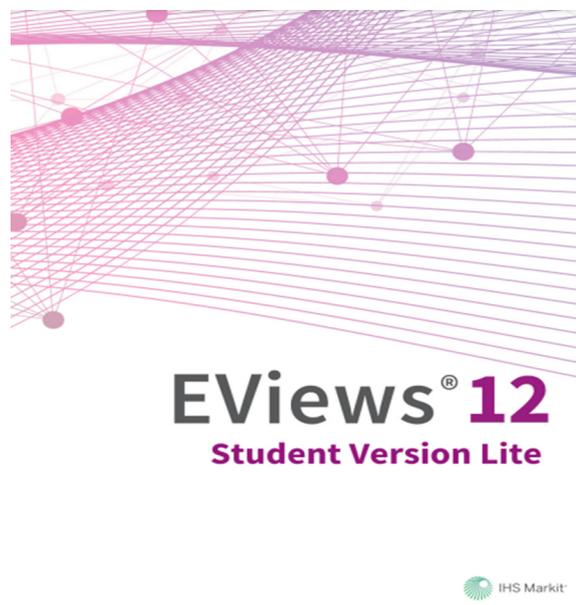


Figure 3.1: L'icône de logiciel EViews.

3.2.2 Champs d'application

Eviews connaît une application dans de nombreux domaines impliquant le traitement et l'analyse des données. En particulier il peut être utile à [18]:

- ✓ La gestion et la prévision des modèles mathématique de domaine économique
 - Prévision de l'inflation;
 - Prévision du taux de croissance;
 - Prévision du taux d'investissement;
 - Prévision des recettes budgétaires etc...;

- ✓ La prévision en entreprise
 - Prévision des ventes;
 - Prévision de la demande;
 - Prévision des prix, etc...;

- ✓ L'analyse financière
 - Caractéristiques des séries financière;
 - Les modèles a effets ARCH, ARCH en moyenne, ARCH généralise;

- ✓ La simulation

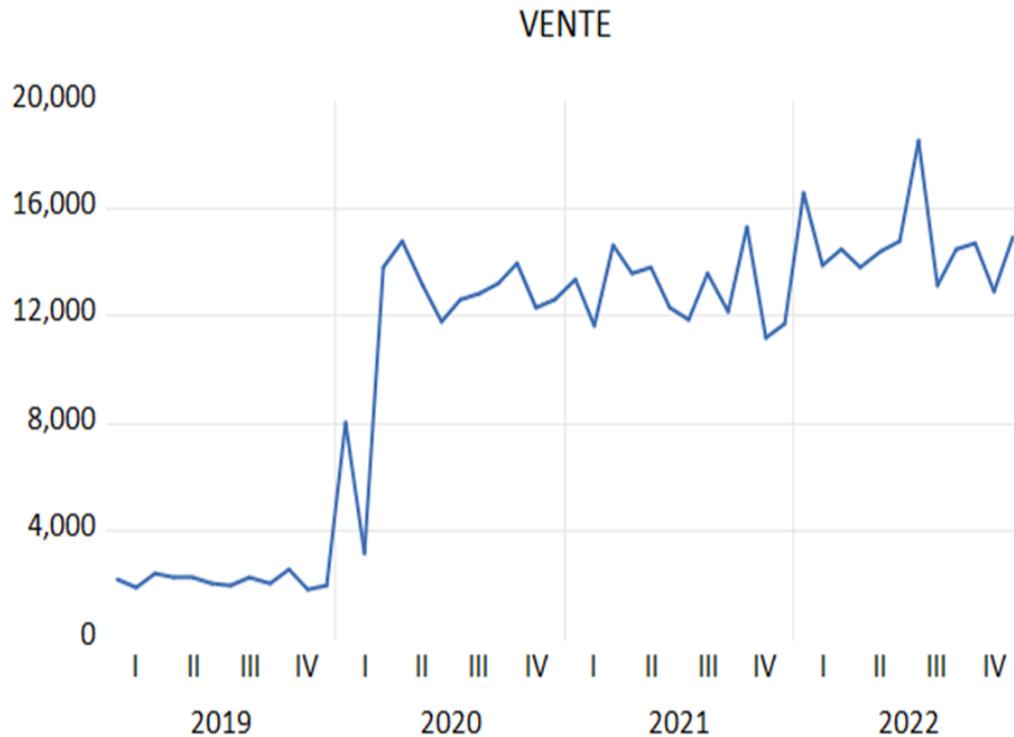
3.3 Analyse de la série vente huile

L'examen de la serie ventes d huile implique l'analyse du graphique et de corrélogramme de la série afin de détecter toute indication de tendances ou de saisonnalités.

3.3.1 Représentation graphique

Le graphe de la figure 3.2 représente les ventes des palettes d'huile de l'entreprise Cevital allant du mois de janvier 2019 au mois de décembre 2022. Le graphe montre qu'en général, la série évolue d'une manière croissante ce qui suggère l'existence d'une tendance. On voit aussi une fluctuation remarquable (fin 2019, début 2020) et cela peut être le résultat de

Figure 3.2: **l'allure générale de la série vente huile.**

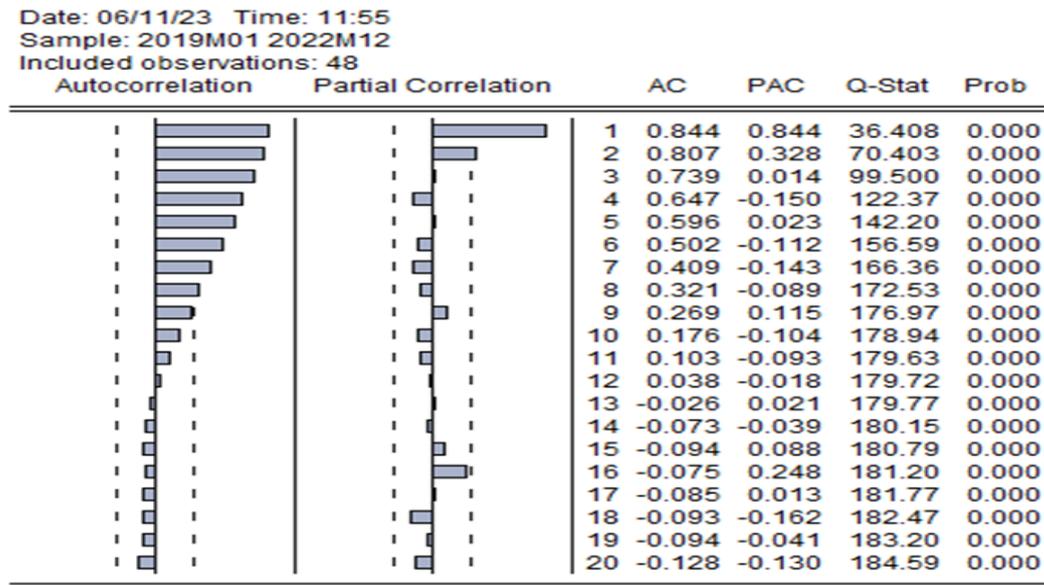


l'apparition de la Pandémie du COVID 19 ainsi les variations saisonnières demeurent relativement constantes, accompagnées de pics périodiques, ce qui signifie que la série n'est pas stationnaire.

3.3.2 Analyse du Corrélogramme

Nous remarquons qu'ils existent des barres de ce corrélogramme dont les auto corrélation sont en dehors de l'intervalle de confiance donc la série comporte une saisonnalité. À travers ce graphe, nous avons remarqué que les probabilités affichées dans le correlograme Sont inférieure à 0.05 .

Figure 3.3: la représentation du corrélogramme de la série vente huile.



Pour confirmer la non stationnarité, on utilise aussi le test de racine unitaire (test ADF) pour confirmer la non stationnarité:

Figure 3.4: Modèle avec tendance et constante.

Null Hypothesis: VENTE has a unit root
 Exogenous: Constant, Linear Trend
 Lag Length: 1 (Automatic - based on SIC, maxlag=9)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-1.720527	0.7260
Test critical values:		
1% level	-4.170583	
5% level	-3.510740	
10% level	-3.185512	

*Mackinnon (1996) one-sided p-values.

dans notre cas, il est pertinent que toutes les probabilités associées à nos modèle sont supérieur à 5%, ce qui nous permet de conclure à l'existence de racines unitaires dans nos données.

D'après ce qu'on a vu précédemment, la série VENTE est non stationnaire.

Nous avons déduire que notre série est saisonnière donc il faut éliminer le facteur de saisonnalité en créant une nouvelle série temporelle corrigée à l'aide de Eviews et ajoutée des coefficients significatifs.

Figure 3.5: **Modèle sans tendance et avec constante.**

Null Hypothesis: VENTE has a unit root
 Exogenous: Constant
 Lag Length: 1 (Automatic - based on SIC, maxlag=9)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-1.589777	0.4796
Test critical values:		
1% level	-3.581152	
5% level	-2.926622	
10% level	-2.601424	

*Mackinnon (1996) one-sided p-values.

Figure 3.6: **Modèle sans tendance et sans constante.**

Null Hypothesis: VENTE has a unit root
 Exogenous: None
 Lag Length: 1 (Automatic - based on SIC, maxlag=9)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	0.341588	0.7797
Test critical values:		
1% level	-2.616203	
5% level	-1.948140	
10% level	-1.612320	

*Mackinnon (1996) one-sided p-values.

3.4 Le processus de dessaisonnalisation des ventes d'huile

Pour progresser dans notre analyse, nous allons appliquer la dessaisonnalisation en utilisant les coefficients saisonniers (CS).

3.4.1 Estimation des coefficients saisonniers

Figure 3.7: Table des coefficients saisonniers de la série huile. (Source: Résultat obtenu à partir logiciel Eviews 12).)

```

Date: 06/11/23 Time: 11:56
Sample: 2019M01 2022M12
Included observations: 48
Difference from Moving Average
Original Series: VENTE
Adjusted Series: VENTESA

```

Scaling Factors:	
1	1411.063
2	-2056.007
3	2357.340
4	1762.340
5	1183.660
6	12.71528
7	-332.6319
8	61.52083
9	-713.8264
10	445.2708
11	-2029.187
12	-2102.257

3.4.2 Série corrigée des variations saisonnières (CVS)

La série ajustée pour corriger les variations saisonnières, notée “ventesa”, représente la manifestation du phénomène sans l’influence des saisons. En examinant attentivement le tableau et le graphique de la série désaisonnalisée ci-dessous, il est apparent que l’effet saisonnier a été éliminés.

3.5 Analyse de la Stationnarité de la Série Désaisonnalisée ventesa

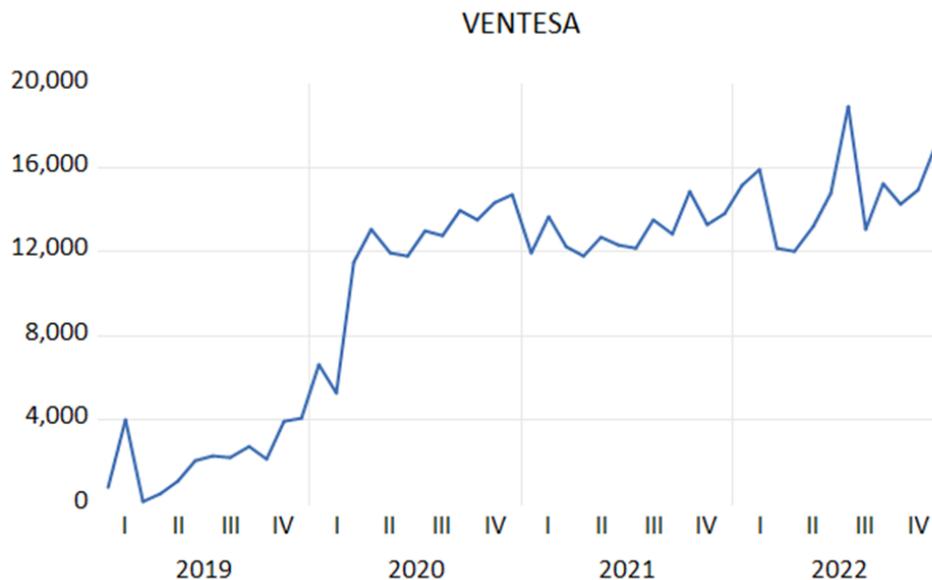
Dans notre analyse, nous débutons par l’élimination des effets saisonniers de notre série de données, car ces variations périodiques peuvent obscurcir la vraie nature des tendances.

Table 3.2: les données de la Série désaisonnalisée.

annéesmois	2019	2020	2021	2022
Janvier	818.9375	6606.938	11970.94	15203.94
Février	3997.007	5270.007	13705.01	15953.01
Mars	82.65972	11466.66	12283.66	12151.66
Avril	502.6597	13057.66	11828.66	12042.66
Mai	1123.340	11977.34	12663.34	13256.34
Juin	2040.285	11816.28	12308.28	14776.28
Juillet	2305.632	12974.63	12175.63	18898.63
Août	2206.479	12799.48	13547.48	13098.48
Septembre	2736.826	13961.83	12855.83	15219.83
Octobre	2107.729	13557.73	14872.73	14272.73
Novembre	3901.187	14362.19	13261.19	14969.19
Décembre	4053.257	14697.26	13810.26	17077.26

Cependant, même après cette désaisonnalisation, il est possible que notre série reste non stationnaire, c'est-à-dire qu'elle continue à présenter des fluctuations imprévisibles dans le

Figure 3.8: Graphe de la série désaisonnalisé VENTESA.



temps. Pour déterminer si notre série est stationnaire, nous utilisons des tests DF (Dickey-Fuller) qui évaluent la présence de racines unitaires. Ces tests sont fondamentaux pour garantir que nos données sont appropriées pour des analyses statistiques plus avancées.

3.5.1 Test de dickey-fuller augmented

Avant d'effectuer le test de Dickey-Fuller, nous entreprenons la sélection du nombre optimal de retards, noté P. Nous recherchons le nombre de retards qui minimise les critères d'Akaike et Schwarz pour les trois modèles, à savoir celui avec tendance et constante, celui avec constante seulement, et celui sans tendance ni constante. Ces valeurs de critères sont obtenues à l'aide du logiciel Eviews et sont récapitulées dans le tableau ci-dessous.

Figure 3.9: Les valeurs fournies par le logiciel Eviews12.

	Modèle 1		Modèle 2		Modèle 3	
	Akaike	Schwarz	Akaike	Schwarz	Akaike	Schwarz
L =0	18.08	18.12	18.04	18.12	17.99	18.11
L =1	17.95	18.03	17.94	18.06	17.94	18.10

Nous optons pour L=1 car il offre la minimisation des critères d'Akaike et Schwarz. Nous poursuivons ensuite avec l'estimation des trois modèles distincts : le modèle 3, modèle 2 et enfin le modèle 1.

❖ Evaluation du modèle 3

Dans ce modèle nous testons les deux hypothèses :

- ✓ H0 : absence de la tendance
- ✓ H1 : l'existence d'une tendance

Figure 3.10: Evaluation de modèle 3 de la série VENTESA.

Null Hypothesis: VENTESA has a unit root
 Exogenous: Constant, Linear Trend
 Lag Length: 1 (Fixed)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-1.657912	0.7537
Test critical values:		
1% level	-4.170583	
5% level	-3.510740	
10% level	-3.185512	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
 Dependent Variable: D(VENTESA)
 Method: Least Squares
 Date: 06/11/23 Time: 12:07
 Sample (adjusted): 2019M03 2022M12
 Included observations: 46 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
VENTESA(-1)	-0.173465	0.104628	-1.657912	0.1048
D(VENTESA(-1))	-0.287061	0.147048	-1.952163	0.0576
C	948.9845	624.5321	1.519513	0.1361
@TREND("2019M01")	50.77720	39.77032	1.276761	0.2087

R-squared	0.196590	Mean dependent var	284.3533
Adjusted R-squared	0.139203	S.D. dependent var	1975.585
S.E. of regression	1832.931	Akaike info criterion	17.94816
Sum squared resid	1.41E+08	Schwarz criterion	18.10717
Log likelihood	-408.8077	Hannan-Quinn criter.	18.00773
F-statistic	3.425715	Durbin-Watson stat	1.810247
Prob(F-statistic)	0.025560		

En premier lieu, nous observons une augmentation de la probabilité associée au test de Dickey-Fuller. Cette probabilité s'élève à 0,75, ce qui dépasse le seuil de signification de 0,05. Cela suggère la présence d'une racine unitaire dans les données. De plus, la probabilité associée à la tendance est de 0,20, également supérieure à 0,05. En conséquence, nous acceptons l'hypothèse nulle, ce qui signifie qu'il n'y a pas de tendance apparente dans les données, Alors on passant a l'autre modèles.

❖ Evaluation du modèle 2

Dans ce modèle nous testons les hypothèses suivantes :

- ✓ H0 : absence de la constante.
- ✓ H1 : l'existence de la constante.

Figure 3.11: Evaluation de modèle 2 de la série VENTESA.

Null Hypothesis: VENTESA has a unit root

Exogenous: Constant

Lag Length: 1 (Fixed)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-1.086799	0.7133
Test critical values:		
1% level	-3.581152	
5% level	-2.926622	
10% level	-2.601424	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation

Dependent Variable: D(VENTESA)

Method: Least Squares

Date: 06/11/23 Time: 12:08

Sample (adjusted): 2019M03 2022M12

Included observations: 46 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
VENTESA(-1)	-0.058883	0.054180	-1.086799	0.2832
D(VENTESA(-1))	-0.354026	0.138379	-2.558382	0.0141
C	1011.181	627.1747	1.612280	0.1142
R-squared	0.165407	Mean dependent var		284.3533
Adjusted R-squared	0.126589	S.D. dependent var		1975.585
S.E. of regression	1846.312	Akaike info criterion		17.94276
Sum squared resid	1.47E+08	Schwarz criterion		18.06202
Log likelihood	-409.6835	Hannan-Quinn criter.		17.98744
F-statistic	4.261071	Durbin-Watson stat		1.837036
Prob(F-statistic)	0.020498			

Nous constatons que la probabilité liée à la constante est de 0,11, ce qui dépasse le seuil de 5 %. Par conséquent, nous acceptons l'hypothèse nulle(H_0), indiquant l'absence de la constante dans notre analyse, Alors on passant au modèle suivant.

❖ Evaluation du modèle 1

Dans ce modèle, nous examinons les deux affirmations suivantes *:

- ✓ H_0 : Le processus n'est pas stationnaire.
- ✓ H_1 : Le processus est de nature stationnaire.

Figure 3.12: Evaluation de modèle 1 de la série VENTESA.

Null Hypothesis: VENTESA has a unit root
 Exogenous: None
 Lag Length: 1 (Fixed)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	0.808706	0.8835
Test critical values:		
1% level	-2.616203	
5% level	-1.948140	
10% level	-1.612320	

*Mackinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
 Dependent Variable: D(VENTESA)
 Method: Least Squares
 Date: 06/11/23 Time: 12:09
 Sample (adjusted): 2019M03 2022M12
 Included observations: 46 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
VENTESA(-1)	0.019592	0.024227	0.808706	0.4230
D(VENTESA(-1))	-0.363305	0.140749	-2.581215	0.0133
R-squared	0.114954	Mean dependent var		284.3533
Adjusted R-squared	0.094840	S.D. dependent var		1975.585
S.E. of regression	1879.570	Akaike info criterion		17.95798
Sum squared resid	1.55E+08	Schwarz criterion		18.03748
Log likelihood	-411.0335	Hannan-Quinn criter.		17.98776
Durbin-Watson stat	1.854938			

Nous observons que la probabilité liée au test ADF est de 0,88, ce qui excède le seuil de 5 %. Par conséquent, nous acceptons l'hypothèse H0 et rejetons l'hypothèse H1. Cela indique que notre série n'est pas stationnaire et qu'elle suit un processus de différenciation sans dérive(DS). Pour déterminer l'ordre d'intégration de notre série, nous devons procéder à la différenciation.

Différenciation pour le modèle 1

Nous notons que la probabilité liée au test de Dickey-Fuller est inférieure à 5 %. Par conséquent, nous en déduisons que l'existence d'une racine unitaire est peu probable. En fin de compte, nous validons l'hypothèse H1 tout en rejetant H0, ce qui indique que notre série, après la première différenciation, est stationnaire. Pour illustrer cela, nous présen-

Figure 3.13: Test de première différenciation de la série DVENTESA.

Null Hypothesis: DVENTESA has a unit root
 Exogenous: None
 Lag Length: 1 (Fixed)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-5.368317	0.0000
Test critical values:		
1% level	-2.617364	
5% level	-1.948313	
10% level	-1.612229	

*Mackinnon (1996) one-sided p-values.

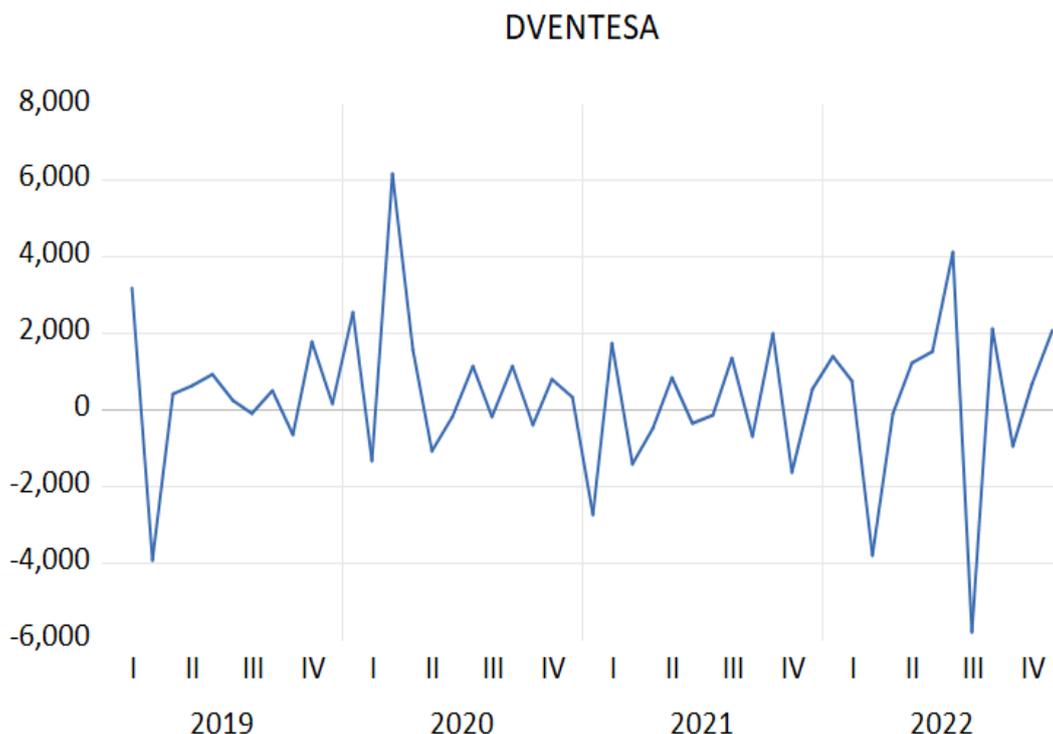
Augmented Dickey-Fuller Test Equation
 Dependent Variable: D(DVENTESA)
 Method: Least Squares
 Date: 06/13/23 Time: 23:02
 Sample (adjusted): 2019M04 2022M12
 Included observations: 45 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
DVENTESA(-1)	-1.319160	0.245731	-5.368317	0.0000
D(DVENTESA(-1))	0.021079	0.146995	0.143402	0.8866

R-squared 0.660616 Mean dependent var 133.8315
 Adjusted R-squared 0.652724 S.D. dependent var 3161.119
 S.E. of regression 1862.852 Akaike info criterion 17.94103
 Sum squared resid 1.49E+08 Schwarz criterion 18.02133
 Log likelihood -401.6732 Hannan-Quinn criter. 17.97097
 Durbin-Watson stat 1.989712

tons ci-dessous le graphique de la série désaisonnalisée après la première différenciation.

Figure 3.14: la représentation graphique de la série DVENTESA autour de sa moyenne.

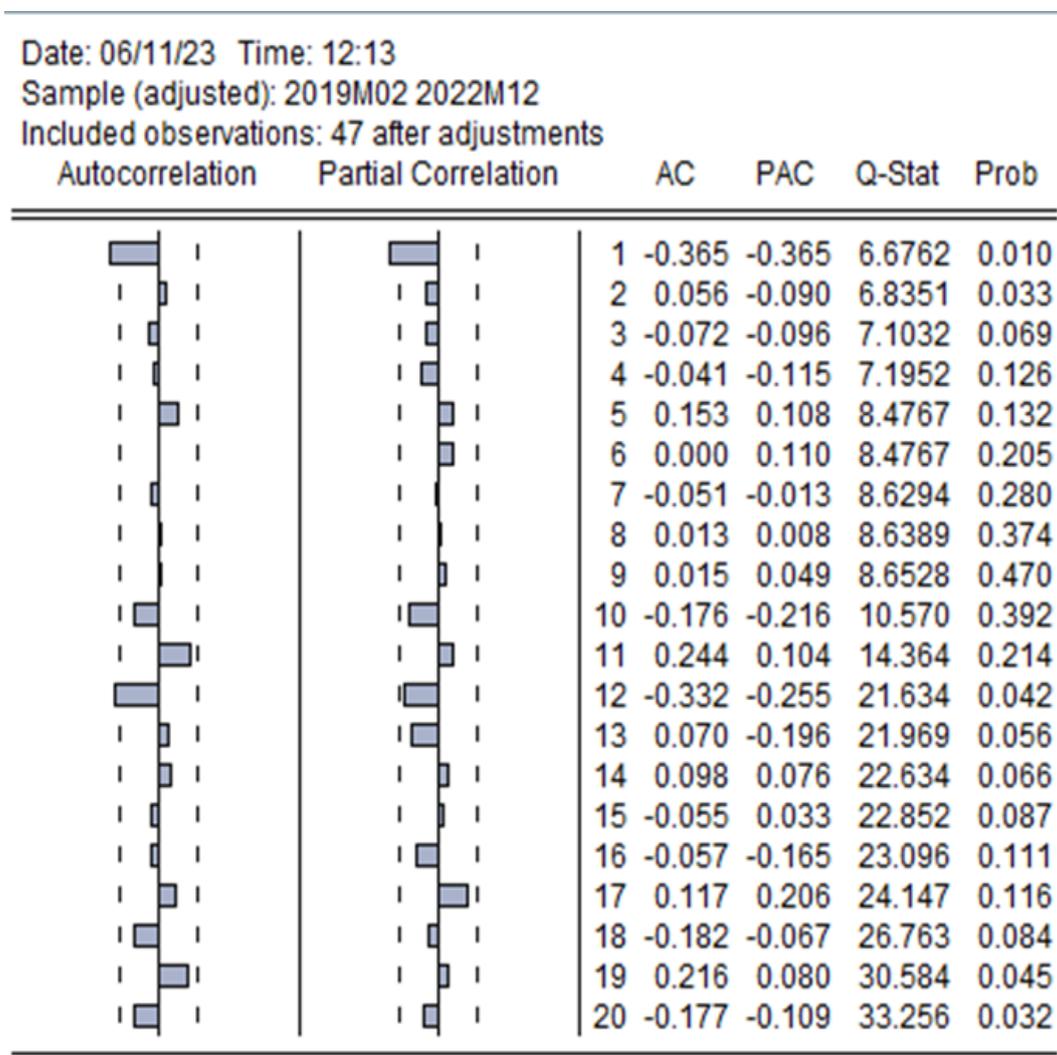


Une série devient stationnaire lorsque ses valeurs fluctuent autour de sa moyenne, ce qui signifie qu'elles ne présentent pas de tendance ou de motifs systématiques au fil du temps.

3.6 Modélisation ARMA

3.6.1 Identification des ordres p et q de ARMA

Figure 3.15: Graphique de l'autocorrélation de la série DVENTESA.



Nous sommes chargés de sélectionner un modèle ARMA (p, q) pour notre série stationnaire, DVENTESA. Dans cette quête, nous examinons le corrélogramme de la série pour déterminer les ordres appropriés de p et q. Le corrélogramme simple nous guide

dans l'identification d'un modèle MA(q), tandis que le corrélogramme partiel nous aide à discerner un modèle AR(p).

En analysant le corrélogramme de la corrélation partielle, nous distinguons la présence du modèle autorégressif (AR), tandis que l'autocorrélation suggère un modèle à moyenne mobile (MA). Notre objectif consiste à identifier les pics situés en dehors de l'intervalle de confiance. Nous remarquons que la première autocorrélation simple est significativement différente de zéro, ce qui nous amène à conclure que $q=1$. En outre, le corrélogramme de la fonction d'autocorrélation partielle révèle que la première autocorrélation partielle est également significativement différente de zéro. Par conséquent, nous déduisons que $p=1$. En conséquence, nous pouvons envisager trois modèles pour la prédiction de la série : un modèle AR(1), un modèle MA(1) ou un modèle ARMA(1,1).

3.6.2 Évaluation des modèles

Étude de la signification des paramètres individuels du modèle

Dans l'évaluation de la signification individuelle des coefficients des variables explicatives, il est évident que ces coefficients diffèrent considérablement de zéro. La t-statistique, évaluée à 2,48, dépasse largement la valeur de 1,96, et les probabilités associées sont toutes inférieures à 5 %. En conséquence, nous conservons ce modèle pour les futures analyses des résidus.

Figure 3.17: **Évaluation du modèle MA(1) sans constante.**

Dependent Variable: DVENTESA
 Method: ARMA Conditional Least Squares (BFGS / Marquardt steps)
 Date: 06/11/23 Time: 12:16
 Sample (adjusted): 2019M02 2022M12
 Included observations: 47 after adjustments
 Convergence achieved after 14 iterations
 Coefficient covariance computed using outer product of gradients
 MA Backcast: 2019M01

Variable	Coefficient	Std. Error	t-Statistic	Prob.
MA(1)	-0.335896	0.140800	-2.385620	0.0212
R-squared	0.088087	Mean dependent var		345.9217
Adjusted R-squared	0.088087	S.D. dependent var		1999.063
S.E. of regression	1908.987	Akaike info criterion		17.96758
Sum squared resid	1.68E+08	Schwarz criterion		18.00695
Log likelihood	-421.2381	Hannan-Quinn criter.		17.98239
Durbin-Watson stat	1.955144			
Inverted MA Roots	.34			

Test de significative:

Nous constatons que les coefficients des variables explicatives sont significativement distincts de zéro, avec une t-statistique de 2,38 largement supérieure à 1,96, et des probabilités associées inférieures à 5 %. Par conséquent, nous conservons ce modèle pour les futures analyses des résidus.

Figure 3.18: **Évaluation du modèle ARMA (1,1) sans constante.**

Dependent Variable: DVENTESA
 Method: ARMA Conditional Least Squares (BFGS / Marquardt steps)
 Date: 06/11/23 Time: 12:19
 Sample (adjusted): 2019M03 2022M12
 Included observations: 46 after adjustments
 Convergence achieved after 17 iterations
 Coefficient covariance computed using outer product of gradients
 MA Backcast: 2019M02

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	-0.509193	0.281280	-1.810274	0.0771
MA(1)	0.236192	0.336393	0.702131	0.4863
R-squared	0.112527	Mean dependent var		284.3533
Adjusted R-squared	0.092357	S.D. dependent var		1975.585
S.E. of regression	1882.146	Akaike info criterion		17.96072
Sum squared resid	1.56E+08	Schwarz criterion		18.04022
Log likelihood	-411.0965	Hannan-Quinn criter.		17.99050
Durbin-Watson stat	1.979792			
Inverted AR Roots	-.51			
Inverted MA Roots	-.24			

Test de significative:

Nous observons que les coefficients des variables explicatives ne se distinguent pas de manière significative de zéro. Cela est corroboré par la t-statistique associée à AR(1) qui atteint 1,81 et à MA(1) qui est de 0,70, toutes deux inférieures à 1,96. De plus, les probabilités liées à MA(1) et AR(1) dépassent 5 %. Par conséquent, nous décidons de ne pas conserver ce modèle pour les analyses ultérieures des résidus.

Après avoir estimé plusieurs modèles et évalué la signification de leurs coefficients, nous retenons les deux premiers : AR (1) et MA (1). Nous progressons maintenant vers la phase de validation du modèle approprié.

3.7 Validation du modèle

3.7.1 Test de Akaike et Schwarz

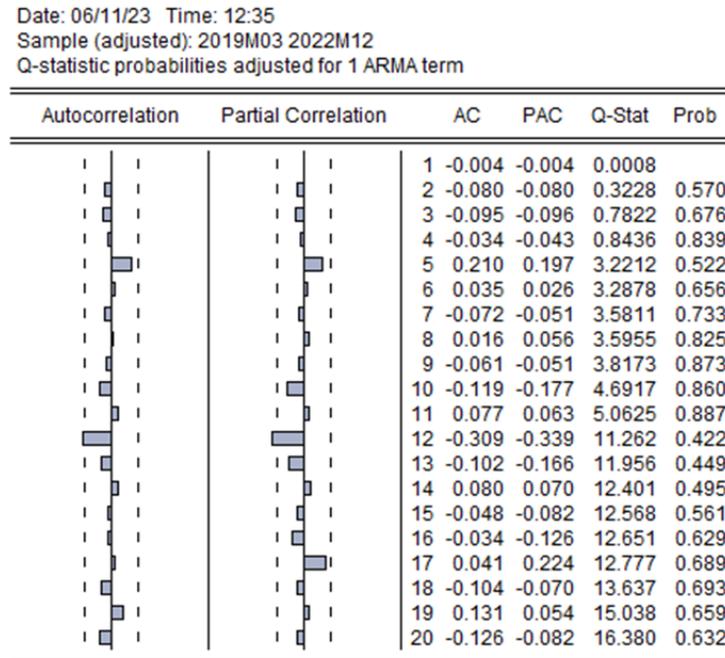
Table 3.3: Les valeurs du test de Akaike et Schwarz.

	AR(1)	MA(1)
AIC	17.92	17.96
SC	17.96	18.00

Les tests d'Akaike et Schwarz indiquent que le modèle AR(1) peut être choisi, car il minimise les deux statistiques.

3.7.2 Test de Ljung-Box

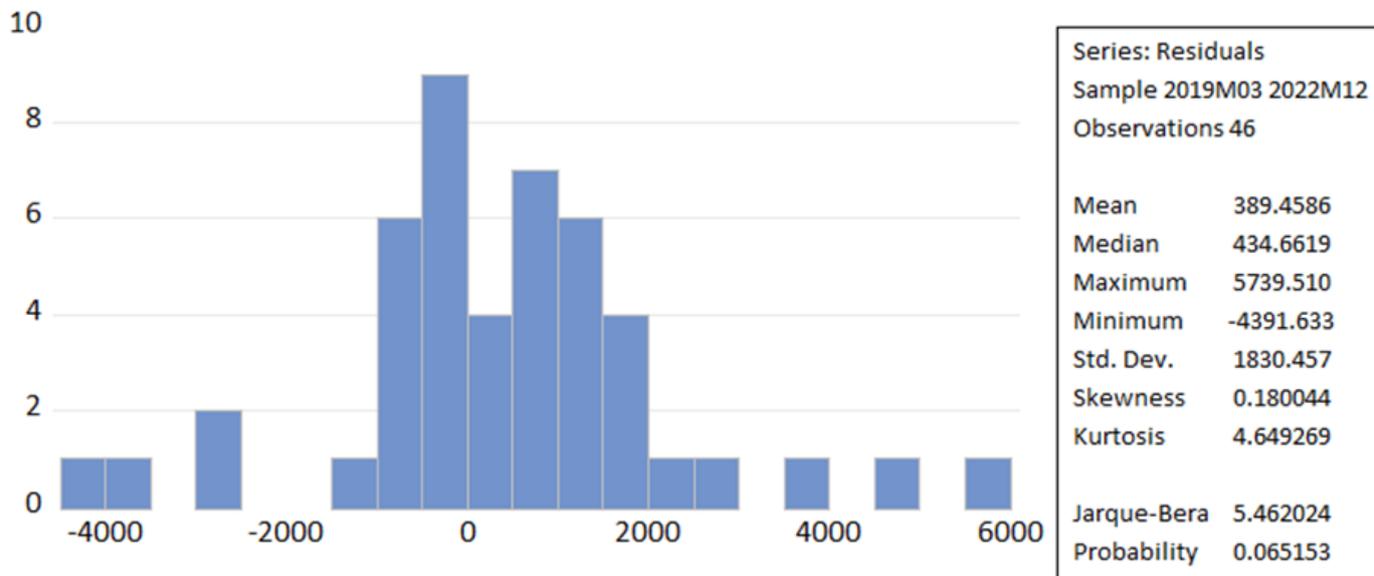
Figure 3.19: l'examen graphique des autocorrélations résiduelles pour effectuer le test de Ljung-Box.



Nous observons que les graphiques des autocorrélations simple et partielle ne présentent aucune pointe en dehors de l'intervalle de confiance, et toutes les probabilités sont inférieures à 5 %. En conséquence, nous en concluons que ce modèle correspond à un processus de bruit blanc.

3.7.3 Test de Jarque-Bera

Figure 3.20: Test de normalité des résidus.



La statistique de Jarque-Bera, qui s'élève à 5,462024, est inférieure à la valeur critique tabulée du chi-deux à 2 degrés de liberté, qui est de 5,991 à un seuil de 5 %. De plus, la probabilité associée à cette statistique, égale à 0,065, dépasse 5 %. Par conséquent, nous pouvons conclure que les résidus du modèle suivent une distribution normale, confirmant ainsi qu'ils correspondent à un bruit blanc gaussien.

En conclusion le modèle AR(1) est validé, ce qui nous permet d'effectuer des prévisions de la série VENTESA en utilisant l'équation suivante :

$$DVENTESA_t = -0.34DVETESA_{t-1}$$

3.8 Prédiction des ventes d'huile pour l'année 2023

En appliquant la méthodologie de Box-Jenkins aux données observées du complexe Cevital pour les années 2019, 2020, 2021 et 2022, nous avons pu générer des prévisions pour les ventes d'huile en 2023. L'équation qui en découle est la suivante :

Nous avons: $DVENTESA_t = -0.34DVETESA_{t-1}$

A partir de cette équation nous allons calculer les prévisions comme suite :

Nous sommes dans le cas d'un modèle additif et nous avons une série dessaisonnalisée, pour revenir à la série d'origine à savoir VETNE, nous allons suivre les étapes suivantes :

$$DVENTESA_t = -0.34DVENTESA_{t-1}$$

Pour revenir à la série vente, nous allons poursuivre ainsi :

$$DVENTESA49 = -0.34 (2108.069)$$

$$DVENTESA49 = -716.7434$$

Nous rajoutons maintenant le coefficient saisonnier.

$$DVENTE49 = -716.7434 + 1411.063$$

$$VENTESA49 - VENTESA48 = 694.3195$$

$$VENTE49 = 694.3195 + VENTE48$$

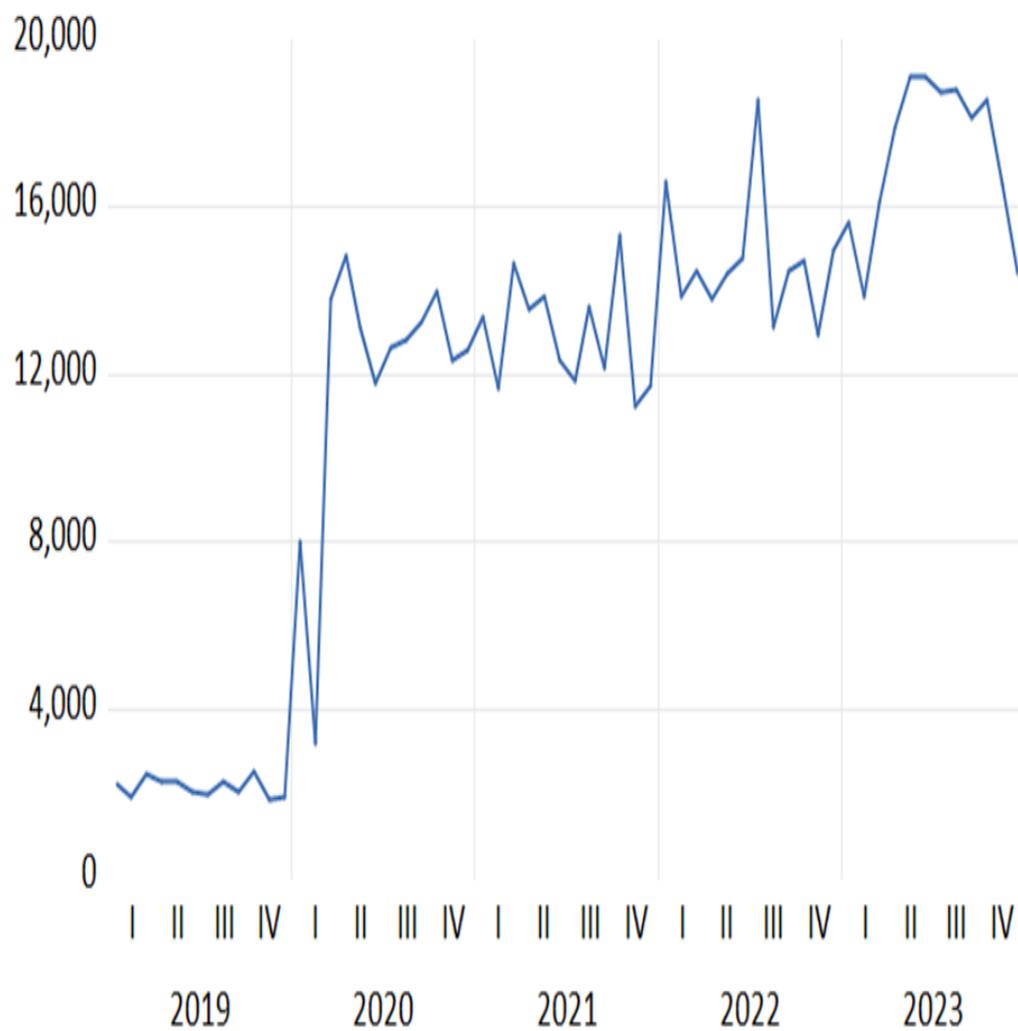
$$VENTE49 = 694.3195 + 14975$$

$$VENTE49 = 15669$$

Figure 3.21: **Tableau des prévisions pour l'année 2023.**

Mois	Prévision 2023 (Unités : palettes)
Janvier	15669
Février	13856
Mars	16130
Avril	17920
Mai	19094
Juin	19109
Juillet	18775
Août	18836
Septembre	18122
Octobre	18567
Novembre	16537
Décembre	14434

Figure 3.22: un graphique qui représente à la fois les données de vente d'huile réelles et les valeur prévues à partir d'une méthode de prévision.



Conclusion générale

Dans cette étude, nous avons réalisé une revue statistique approfondie d'une série chronologique représentant les ventes de palettes sur une période de quatre ans. En parallèle, nous avons développé un modèle ARMA pour les prévisions, dans le but de fournir des estimations pour les périodes futures, en mettant en œuvre la méthodologie de prévision statistique développée par Box et Jenkins, nous avons utilisé le logiciel Eviews 12 pour anticiper les ventes de palettes prévues pour l'année 2023.

En optimisant la gestion de la production et des ventes grâce à l'analyse économétrique des données chronologiques en utilisant l'approche de Box et Jenkins, Cevital pourra répondre plus efficacement aux besoins du marché. En tenant compte des prévisions commerciales et des demandes anticipées, l'entreprise pourra réduire les coûts tout en maintenant la satisfaction des clients. Cela permettra d'éviter les pertes potentielles liées à une capacité de production insuffisante ou à des surcoûts dus à une surcapacité.

L'utilisation des séries temporelles dans la gestion des ventes de Cevital souligne l'importance cruciale de cette approche pour maintenir la compétitivité sur un marché en constante évolution. En exploitant les données temporelles pour élaborer des modèles prédictifs fiables, l'entreprise sera mieux préparée à anticiper les fluctuations du marché et à prendre des décisions stratégiques éclairées. Ainsi, cette méthodologie s'avère essentielle pour atteindre l'objectif ultime de toute entreprise : réaliser un profit maximal et garantir sa durabilité. Les enseignements tirés de cette étude offrent des perspectives prometteuses pour l'avenir de Cevital dans un environnement économique dynamique et exigeant.

Bibliographie

- [1] Situation géographique du complexe CEVITAL. <https://www.cevital.com>. , [Accédé:11/05/2023].
- [2] document interne à l'entreprise.
- [3] F.ABDERRAHMANI. Séries chronologiques. *département des sciences économiques, université de Béjaïa, Année 2017-2018 (Cour)*.
- [4] N.BERRANE S. BELHOUL. Minimisation de la rogne dans la production du carton ondulé au niveau de l'entreprise général emballage. *Mémoire de master en mathématiques, en recherche opérationnelle, Université Abderrahmane MIRA Bejaïa, Année 2018/2019*.
- [5] K. MECHERI. Application de la méthodologie de box-jenkins sur la série du nombre des séances d'hémodialyse. *Mémoire de master en mathématiques, université KASDI MERBAH OUARGLA, Année 2016*.
- [6] G. LYAMINE A. RAMDANI. L'impact de la crise sanitaire sur le trafic portuaire de béjaia (epb). *Mémoire de master en mathématiques, en recherche opérationnelle, Université Abderrahmane MIRA Bejaïa, Année 2021*.
- [7] R. AMOURAT L. MOUZAIA. Analyse et amélioration de stockage des produits agro-alimentaire du groupe cevital. *Mémoire de fin d'étude, école supérieure en sciences appliquées-Tlemcen, Année 2019 /2020*.
- [8] L. MAMMASSE MD A. BOUNCEUR. Gestion optimale des silos au niveau de l'entreprise cevital, rapport de stage. *Université Abderrahmane MIRA Bejaïa*.

-
- [9] S. DJERROUD. Etude de la logistique du transport au sein de cevital. *Mémoire de master en mathématiques, en recherche opérationnelle, Université de Béjaïa, Année 2018.*
- [10] Documentations commerciales et techniques de cevital.
- [11] F.MOKHTARI. Guide pratique des séries temporelles macroéconomique et financières avec evIEWS 9,5. *Note de cours, Université Dr Moulay Tahar Faculté des Sciences Département de Mathématiques Saïda.*
- [12] F. MERAHI. Méthodes chronologiques et prévision. *université Mustapha ben Boulaid de Batna 2 département de mathématique, Première année Master Mathématiques (SAD) Année 2021-2022.*
- [13] S. ROUBA. Modélisation statistique d'une série chronologique. *Mémoire de master en mathématiques, université Mohamed Khider, Biskra, Année 2019.*
- [14] J. BIGOT. Polycopie rédigé à partir des notes de cours de Frédéric Proia. *Université de Bordeaux, Année 2016-2017.*
- [15] J.DAUXOIS. Introduction à l'Étude des séries temporelles, année 2016-2017.
- [16] R. CKEKIMA. Modélisation des séries chronologiques. *Mémoire de master en mathématiques, université KASDI MERBAH OUARGLA, Année 2021-2022.*
- [17] M. CHAOUCHKHOUANE. Méthode de box et Jenkins. *Mémoire de master en mathématiques, université Mohamed Khider, Biskra, Année 2021.*
- [18] Y.KEHO. initiation au logiciel evIEWS. <https://www.scribd.com/document/485113733/SUPPORT-DE-COURS-EVIEWS#>, Année 2007 [Accédé:15/05/2023].