



Ministère de l'Enseignement Supérieur et
De La Recherche Scientifique
Université M'hmed Bougara-Boumerdès
Faculté Des Sciences
Département de mathématiques



Mémoire présenté pour l'obtention du diplôme de Master
en Mathématiques

Spécialité : Mathématique Financière

Thème

Prévision budgétaire en sein de la CNR

Présenté par :

- LAICHAOUI Lila
- RATNI Zahia

Devant le jury :

- Présidente : Mme MEDAHI Samia (UMBB)
- Promoteur : Mr TAZEROUTI Moussa (UMBB)
- Examinatrice: Mr BENMANSOUR (UMBB)
- Encadreur : Mr ZEMMOURI Yacine (CNR)

2018-2019

Introduction

Dans sa définition la plus simple le travail est l'activité au moyen de laquelle nous produisons de quoi satisfaire nos besoins.

« Il est travailleur, c'est un courageux », « il ne doit sa réussite qu'à lui-même » : ces expressions montrent l'estime qu'on porte à ceux qui réussissent à force de travail et à chaque réussite une récompense s'impose, matérielle soit-elle ou moral peut être juste une satisfaction personnel, une chose est sûre après l'effort le réconfort.

Dans notre cas on peut aussi dire qu'après des années de travail viens en fin la retraite ; une situation sociale et financière d'un individu qui, ayant atteint l'âge minimum requis, cesse son activité professionnelle, en général de manière définitive, et perçoit régulièrement un revenu sous forme de pension versée par la CNR (caisse nationale de retraite) ex : "assurance vieillesse" ou de ressources provenant de ses cotisations qui ont été versées à un régime d'assurance vieillesse.

Pour la CNR, il est important de garder le contrôle sur ses finances et pour cela le budget est l'outil par excellence à utiliser pour planifier, évaluer et contrôler les dépenses et d'identifier les problèmes. La gestion de ce dernier permis d'établir des prévisions et le dégageant des écarts entre réalisations et prévisions afin de prévoir les dépenses à venir, pour démontrer que sa stratégie commerciale est bien documenté et qu'elle est au courant de tous les pièges qui peuvent la rencontrer et comment les dépasser.

L'objectif de notre travail est de faire des prévisions budgétaires par des méthodes statistiques et les comparées avec les prévisions faites par une méthode financière proposée par la CNR, la modélisation et par la suite les prévisions sont faites sur la base de données mensuelles sur les différentes séries pour une durée de (5 ans et 10 ans) qui ont pour unité de mesure de dinars.

Dans le cadre de notre étude, nous allons essayer de répondre à la question principale suivante :

- ✓ **Quelle est la méthode statistique de prévision à court terme la plus adéquate pour faire les prévisions budgétaires au sein de la CNR?**

Pour l'objectif assigné, il existe un axe important dans les séries chronologiques ce qui nous amène à organiser notre étude comme suit, nous commençons par une étude univariée des séries chronologiques ou nous utilisons la méthode de box et Jenkins, la méthode de Holt et winters. Pour l'approche multivariée nous allons utilisés la modélisation VAR et l'approche de cointégration bivariée.

La dernière partie sera consacrée a la comparaison entre les méthodes univariées et multivariée pour tirer le meilleur modèle en se basant sur le critère de l'erreur quadratique moyenne des erreurs (RMSE :root mean square error) . Nous terminons notre travail par une étude comparative.

Dédicace

Je dédie ce modeste travail

A ma très cher Mère et mon très cher Père

A ceux qui ont veillé pour mon bien être

A ceux qui m'ont toujours encouragé pour que je réussisse dans mes études

A tout ce qui est m'encouragé dans la réalisation de ce modeste travail

A mes chères sœurs

A tous mes frères

A toutes les familles : LAICHAOUI, RATNI

N'oublier pas de dédier à mes très chère amies intime

Tous mes camarades surtout la promotion 2018/2019

Spécialité : Mathématiques Financières

Tous qui m'on aider de proche ou de loin.

Remerciements

Tout d'abord, nous remercions ALLAH Le tout Puissant qui m'a accordée la volonté et le courage pour réaliser ce mémoire.

Mes plus vifs remerciements et ma profonde gratitude vont à Mr. TAZEROUTI Moussa notre promoteur de mémoire pour sa grande patience, pour sa disponibilité, pour ses nombreux conseils, pour ses corrections et son appréciation au cours de L'élaboration de ce travail.

J'adresse mes remerciement à Mr. ZAMOURI Yacine financier de l'agence CNR de la wilaya de Boumerdes pour son soutien, disponibilité, ces conseils et pour son aide durant la réalisation de notre travail, et à tous les travailleurs de service finance de la CNR.

Je remercie sincèrement, les membres de jury d'avoir bien voulu accepter de faire partie de la commission d'examineur.

Nos remerciements les plus sincères vont au Mme FILICI Fatiha et tous les enseignants du département de mathématiques.

Je remercier aussi l'ensemble de mes camarades de master, mes amis, mes proches et mes familles (au sens large) qui m'ont soutenue durant ce travail, et spécialement à celles ou ceux, elles ou ils reconnaîtront, qui m'ont encouragée à finir ce travail et qui m'ont accompagnée dans tous les moments de joie et de tristesse.

Enfin, je ne saurais terminer cette partie sans exprimer notre gratitude à mes parents, mes grands parents, mes frères et mes sœurs qui j'ai toujours soutenue, encouragé et stimulée pendant mes études.

Merci à tous et à toutes.

Notations et abréviations

| | |
|----------------|--|
| ADF | Dickey-Fuller Augmenté |
| AR | Autorégressifs, (Autorégressive) |
| ARMA(p; q) | Autoregressive Moving Average |
| ARIMA(p; d; q) | Autoregressive Integrated Moving Average |
| B | Opérateur retard |
| BB | Bruit blanc |
| CNAS | Caisse Nationale des Assurances Sociale |
| CNR | Caisse Nationale des Retraites |
| DS | Différence Stationary |
| DW | Durbin Watson |
| F | Opérateur d'avance |
| FAC | Fonction d'Autocorrélation |
| FAP | Fonction d'Autocorrélation Partielle |
| JB | Test de Jarque & Bira |
| Ku | Kurtosis |
| MA | Moyennes Mobiles (Moving Average) |
| MCO | Moindres carrés ordinaires |
| SC | Critère de Schwarz |
| Sk | Skewness |
| SSR | Sum Squared Residual |
| R^2 | Coefficient de détermination |
| TS | Trend Stationary |
| VCM | Valeur Critique de Mackinnon |

Sommaire

| | |
|--|----|
| Dédicace | |
| Remerciements | |
| Introduction | 1 |
| 1 La caisse nationale de retraite da la wilaya de Boumerdes..... | 2 |
| 1.1 Historique du système de retraite algérien | 2 |
| 1.2 Présentation de la CNR de Boumerdes..... | 3 |
| 1.3 Organisation..... | 3 |
| 1.4 Financement..... | 3 |
| 1.5 Notion sur le budget..... | 5 |
| 1.5.1 Notion sur le budget | 5 |
| 1.5.2 Budget prévisionnel | 6 |
| 1.5.3 Un budget prévisionnel : pour quoi faire ? | 6 |
| 1.5.4 Le budget de la CNR | 6 |
| 2 Généralité sur les séries chronologiques..... | 8 |
| 2.1 Définition d'une série chronologique | 8 |
| 2.2 Représentation graphique d'une série chronologique..... | 8 |
| 2.3 Composantes d'une série temporelle..... | 8 |
| 2.3.1 La composante tendancielle (Tt)..... | 8 |
| 2.3.2 La composante cyclique (Ct) | 8 |
| 2.3.3 La composante saisonnière (S_t) | 8 |
| 2.3.4 La composante résiduelle ou aléatoire | 8 |
| 2.4 Modèle de décomposition d'une série chronologique..... | 9 |
| 2.4.1 Modèle additif | 9 |
| 2.4.2 Modèle multiplicatif..... | 9 |
| 2.5 Identification des composantes de la série chronologique | 9 |
| 2.5.1 Test graphique | 9 |
| 2.5.2 Test par l'analyse de la variance | 9 |
| 2.6 Prévision..... | 11 |
| 2.7 Processus stochastiques et modèles de séries chronologiques..... | 11 |
| 2.7.1 Définition..... | 11 |
| 2.7.2 Caractéristiques d'un processus stochastique | 11 |
| 2.7.3 Processus stochastique stationnaire..... | 11 |

| | | |
|-------|---|----|
| 2.7.4 | Processus Bruit Blanc | 12 |
| 2.7.5 | Fonction d'auto-covariance | 12 |
| 2.7.6 | Fonction d'auto corrélation..... | 12 |
| 2.7.7 | Les opérateurs | 13 |
| 2.7.8 | Théorème de Wold (décomposition de Wold) | 14 |
| 2.8 | Modèles de série chronologique | 15 |
| 2.8.1 | Les processus linéaires stationnaires..... | 15 |
| 2.8.2 | Les processus aléatoires non-stationnaires | 16 |
| 2.8.3 | Modèle non stationnaire ARIMA et SARIMA..... | 17 |
| 3 | La méthodologie de Holt & Winters | 18 |
| 3.1 | Généralisation sur le lissage exponentiel | 18 |
| 3.2 | Principe des méthodes de lissage exponentiel | 18 |
| 3.3 | Les différents lissages exponentiel..... | 19 |
| 3.3.1 | Lissage exponentiel de Holt & Winters sans saisonnalité | 19 |
| 3.3.2 | Lissage exponentiel de Holt (double)..... | 20 |
| 3.3.3 | Méthodes de Holt & Winters | 20 |
| 3.4 | Application par l'approche de la méthode Holt & Winters | 22 |
| 3.4.1 | Etude de la série Budget..... | 22 |
| 3.4.2 | Etude de la série RETRAITE..... | 24 |
| 3.4.3 | Etude de la série RECETTE | 26 |
| 3.4.4 | Etude de la série DEPENSE | 28 |
| 4 | La méthodologie de Box & Jenkins | 30 |
| 4.1 | Méthodologie de Box –Jenkins | 30 |
| 4.1.1 | Familiarisation avec les données | 31 |
| 4.1.2 | Analyse préliminaire | 31 |
| 4.2 | Stationnarisation de la série..... | 31 |
| 4.2.1 | Test de la tendance (Dickey Fuller)..... | 32 |
| 4.3 | Les étapes de la méthode de Box & Jenkins | 33 |
| 4.3.1 | Identification | 33 |
| 4.3.2 | Estimation du modèle | 34 |
| 4.3.3 | Validation du modèle..... | 34 |
| 4.3.4 | Prévision..... | 37 |
| 4.4 | Application de la méthode Box & Jenkins..... | 38 |
| 4.4.1 | Étude de la série BUDGET | 38 |
| 4.4.2 | Étude de la série RETRAITE..... | 49 |
| 4.4.4 | Etude de la série RECETTE | 58 |

| | | |
|-------|---|-----|
| 4.4.5 | Etude de la série DEPEN | 67 |
| 5 | Approche multi-varie (VAR) | 77 |
| 5.1 | Définition..... | 77 |
| 5.2 | Présentation du modèle (VAR) | 77 |
| 5.3 | La stationnarité | 78 |
| 5.4 | Caractéristiques des processus VAR(p)..... | 78 |
| 5.5 | Représentation canonique..... | 79 |
| 5.6 | Estimation des paramètres | 79 |
| 5.7 | Validation..... | 80 |
| 5.7.1 | Tests de spécification | 80 |
| 5.8 | Prévisions | 82 |
| 5.9 | Application du modèle VAR | 83 |
| 5.9.1 | Etude multivariée des séries (DBUDGETSAt, DRETRAITESAt) | 83 |
| 5.9.2 | Identification de modèle VAR(p) | 83 |
| 5.9.3 | Estimation du modèle VAR(1)..... | 84 |
| 5.9.4 | Validation | 85 |
| 5.9.5 | Test granger de causalité | 88 |
| 5.9.6 | Prévision..... | 88 |
| 6 | Approche de la cointégration bivariée..... | 90 |
| 6.1 | Théorie de de la méthodologie da cointégration bivariée | 90 |
| 6.1.1 | Notations et définitions | 90 |
| 6.1.2 | Notion de régression fallacieuse "illusoire" | 91 |
| 6.1.3 | Cointégration..... | 92 |
| 6.1.4 | Modèles à Correction d'Erreur (MCE)..... | 95 |
| 6.1.5 | Cointégration entre n variables et MCE..... | 98 |
| 6.1.6 | Estimation des Modèles à Correction d'Erreur | 100 |
| 6.2 | Application de la méthodologie de cointégration..... | 101 |
| 6.2.1 | Etude de l'évolution des séries : budget et retraite | 101 |
| 6.2.2 | Etude de l'évolution des séries : recette et dépense..... | 103 |
| 7 | Choix de la méthode de prévision des séries | 104 |
| | Conclusion | 105 |
| | Bibliographie..... | 107 |

Chapitre 1
Présentation de
l'organisme d'accueil

1.1 Historique du système de retraite algérien

L'assurance vieillesse n'a été instituée pour l'ensemble des salariés qu'en 1953, avant cette date, quelques corporations bénéficiaient déjà de régimes de retraite : ce sont les fonctionnaires et les travailleurs jouissant de statuts similaires à ceux des fonctionnaires tels les cheminots, les tramwayistes et les employés de la société d'Etat «Electricité et Gaz d'Algérie».

L'assurance vieillesse pour les travailleurs non-salariés (indépendants) des professions industrielle, commerciale, artisanale, libérale et agricole instituée en 1956 n'a été mise en œuvre qu'en 1958 et constitue au départ l'unique branche de ce régime particulier.

En juillet 1983, il a été mis fin aux régimes de retraite de base ainsi qu'au régime de retraite complémentaire existants. Ainsi, l'avènement de la réforme de juillet 1983 s'est traduit par la fusion des régimes à base professionnelle existants en un régime unifié.

La Caisse Nationale des Retraites (CNR) a été créée par décret n° :85-223 du 20 août 1985 abrogé et remplacé par le décret n°92-07 du 4 janvier 1992 portant statut juridique des Caisses de Sécurité Sociale et organisation administrative et financière de la Sécurité Sociale.

La CNR est le résultat de la fusion de sept (7) caisses (la CAVNOS devenu CASNOS pour les travailleurs non-salariés) en place en 1985, chargée de la gestion des différents régimes de retraite existant avant l'institution en 1983, d'un régime national unique de retraite, offrant les mêmes avantages à tous les travailleurs quel que soit leur secteur d'activité. Il s'agit de :

- La CAAV chargée de la gestion des pensionnés du régime général.
- La CGR chargée de la gestion des pensionnés fonctionnaire.
- La CNMA chargée de la gestion des pensionnés du régime agricole
- La CSSM chargée de la gestion des pensionnés du secteur des mines
- La CAVNOS.... chargée de la gestion des pensionnés non-salariés
- L'EPSGM..... chargée de la gestion des pensionnés de la SONALGAZ
- La Caisse de Retraite des personnels de la SNTF.

1.2 Présentation de la CNR de Boumerdes

La caisse nationale de retraite de la wilaya de BOUMERDES est un établissement public à gestion spécifique régi par les lois applicables en la matière, a été créée en septembre 1986 et elle a été mise en activité en début de 1987, avec un nombre de personnel au 31/01/2018 de 97 employés, nombre de retraités 100636 (en retraite directe 61002, en retraite anticipée 11, en pension de réversion 32736, en allocation directe 9474 et en allocation de réversion 1913). Elle se situe à Oued Tatareg, BP106 Boumerdes.

La caisse de BOUMERDES s'en charge de :

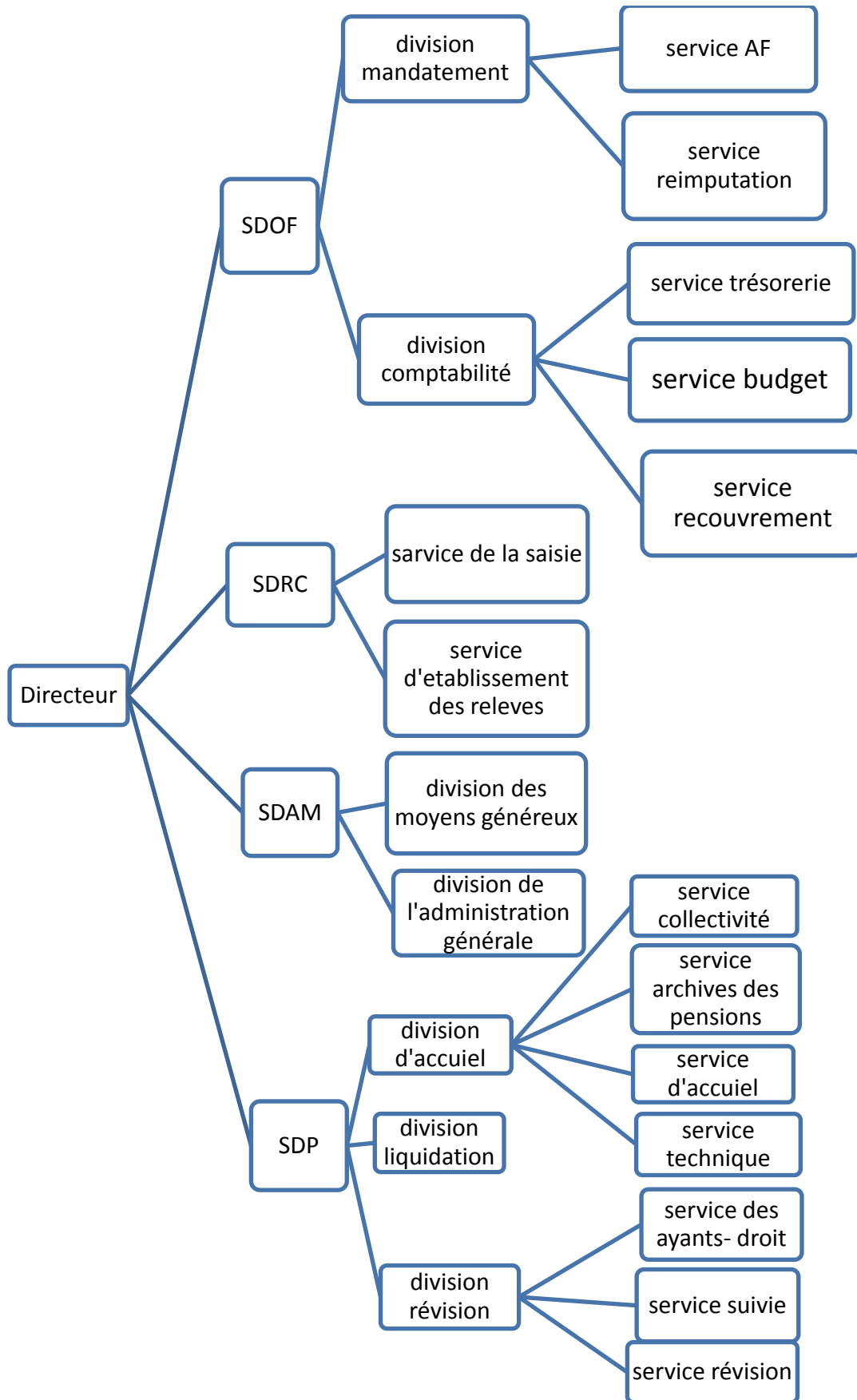
- Gérer la pension et allocation de retraite, ainsi que les pensions et allocations de l'ayant droit.
- Gérer jusqu'à extinction des droits, les pensions et allocations servies au titre de la légalisation antérieure au 1^{er} janvier 1984.
- Assurer le recouvrement, le contrôle et le contentieux du recouvrement des cotisations destinées au financement des pensions de retraites.
- Mettre en application les dispositions relatives prévues par les conventions et accords internationaux de sécurité sociale.
- Assurer l'information des bénéficiaires et des employeurs.
- Gérer le fond d'aide et de secours en application de l'article 52 de la loi N°82-12 du 02 juillet 1983 relative à la retraite.

1.3 Organisation

La caisse de Boumerdes, en plus de la direction qui dirige l'organigramme de la caisse, comprend une sous-direction finances et comptabilité (SDOF), une sous-direction de la reconstitution des carrières (SDRC), sous-direction de l'administration et moyen généraux (SDMA) et la sous-direction des pensions (SDP).

L'agence de Boumerdes compte aussi 3 centres d'accueil et d'orientation (CAO) situés à : Bordj-Ménaiel, Dellys et à Rouïba.

Organigramme est comme tel :



1.4 Financement

Le mode de financement du système découle directement de son caractère professionnel. Les sources de financement sont donc essentiellement des cotisations à la charge des employeurs et des travailleurs et 97% du budget est attribué à la retraite les 3% restent sont partagés sur les achats de l'entreprise et le maintien du matériel. Il y a lieu de signaler que toutes tâches liées à l'affiliation et au recouvrement sont assurés actuellement, pour l'essentiel par CNAS en coordination avec les services de la CNR. Le tableau suivant explique la répartition des cotisations versées à la CNAS.

| Branches | Employeurs | Travailleurs | Œuvre social | Total |
|---|------------|--------------|--------------|--------|
| Assurance sociales | 12% | 1.5% | 1.5% | 15% |
| AT² et MP³ | 1.25% | - | - | 1.25% |
| Retraite | 11.5% | 6.75% | 0.50% | 18.75% |
| Retraite anticipé | 0.25% | 0.25% | - | 0.50% |
| Assurance chômage | 1% | 0.50% | - | 1.50% |
| Promotion logement social | 0.50% | - | - | 0.50% |
| Totale | 26% | 9% | 2% | 37% |

Source : CNAS 2017

{ AT² : Accident de travail
 { MP³ : Maladie Professionnelle

1.5 Notion sur le budget

1.5.1 Notion sur le budget

Le budget est un document récapitulatif des recettes et dépenses prévisionnelles déterminée et chiffrées pour un exercice comptable à venir généralement l'année.

Le budget n'est pas un document normalisé et ne fait pas partie des documents prévus par la méthodologie de la comptabilité ce qui n'empêche pas l'outil <<BUDGET>> d'être largement utilisé par les gestionnaires. La méthode et la procédure budgétaire se révèlent en effet très utiles dans le cadre d'une entreprise ou d'une organisation pour servir de point d'appui aux de prévisions, de pilotage ou de contrôle.

Un budget comporte des données chiffrée qui peuvent être exprimées de façon plus ou moins poussée.

La notion de budget est normalement inséparable d'une période de temps bien définie soit généralement un exercice fiscal de 12 mois (budget dit annuel) de manière à assurer une vision opérationnelle plus concrète pour les opérations impliquées dans la mise en œuvre, le budget est souvent assorti d'un calendrier de réalisation parfois appelé « tableau de marche » ou tableau de progrès contrôlé. Le budget est alors dit « phasé » en fonction des degrés de fitness visés, le chiffrage des recettes- dépenses est ventilé et détaillé par phase de réalisation (mensuelle- trimestrielle ou semestrielle).

1.5.2 Budget prévisionnel

Le budget prévisionnel est un tableau financier regroupant l'ensemble des dépenses (ou charges) et des recettes (ou produits) liées à votre projet pour l'année en cours ou à venir. L'année concernée est appelée « exercice », correspondant généralement à une année civile et fiscale.

L'établissement d'un budget prévisionnel constitue une étape indispensable dans le développement d'une entreprise. C'est la feuille de route que vous allez suivre tout au long de l'année.

1.5.3 Un budget prévisionnel : pour quoi faire ?

Avant toute chose, un budget prévisionnel est fait pour être consulté régulièrement et les objectifs prévus doivent être comparés mois par mois avec ce qui a été réalisé. Il vous sert donc de ligne directrice précise et c'est pour cela qu'il est intéressant de le décomposer en fonction de vos différentes activités ou de vos différents produits. Ainsi, vos objectifs sont précis et segmentés et vous pouvez ajuster les variables qui affectent telle ou telle activité. En identifiant en cours d'année les écarts entre vos résultats et vos prévisions et en y remédiant, vous optimisez la gestion de votre entreprise.

Alors que votre business plan vous a permis, à la création de votre entreprise, de vous fixer des objectifs à moyen et long terme ; le budget prévisionnel vous permet de résumer les principaux indicateurs qui vous feront atteindre ces objectifs et pérenniser votre activité.

1.5.4 Le budget de la CNR

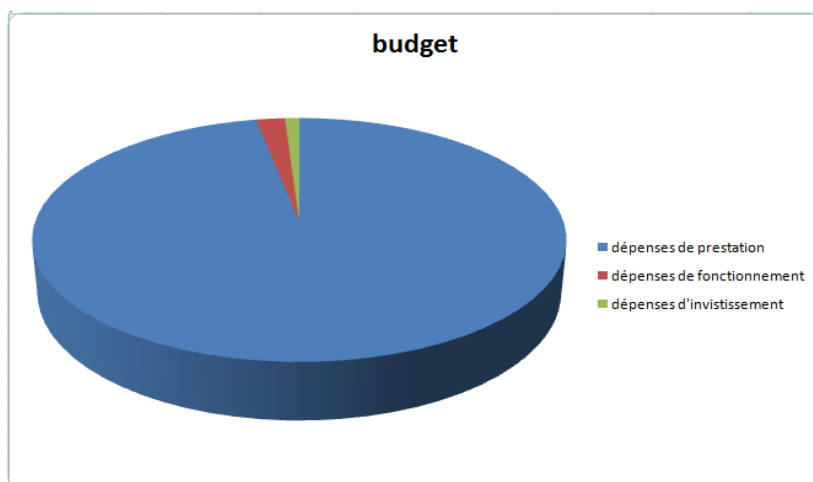


Figure 1-1-1: budget de la CNR

Concernant le budget de la CNR la plus grande partie de ce dernier est réservé pour la retraite et l'autre partie pour le service interne (le fonctionnement et l'investissement).

➤ **Budget prévisionnel de l'exercice 2019**

Le budget prévisionnel de l'exercice 2019 se compose de 3 titres :

| | Désignation | Prévision de budget |
|--------|-----------------------------------|--------------------------|
| Titre1 | Dépenses de prestation | 36 005 500 000 DA |
| Titre2 | Dépenses de fonctionnement | 258 330 000 DA |
| Titre3 | Dépenses d'investissement | 2 580 000 DA |

Tableau 1: Budget prévisionnel de la CNR 2019

On va détailler le budget prévisionnel par suite :

L'élaboration du budget prévisionnel au titre 2019, notamment en matière de dépenses de retraite et de retraite anticipée, a été basée sur deux éléments suivants :

a- Taux moyen d'évolution du coût d'une pension des exercices 2015 ; 2016, 2017 et 1^{er} trimestre 2018.

b- Taux d'évolution du nombre de pensions déterminé en comparant l'évolution du nombre des exercices précités.

➤ **Récapitulation générale des prévisions exercice 2019**

| | Désignation | Prévisions (DA) |
|----------------|-------------------------------------|--------------------------|
| Titre 1 | Dépenses de prestation | 36 005 500 000 DA |
| Chapitre 1 | Pension de retraite | 36 000 000 000 |
| Chapitre 2 | Retraite anticipée | 5 500 000 |
| Titre 2 | Dépenses de fonctionnement | 258 330 000 DA |
| Chapitre 1 | Matière et fourniture | 4 315 000 |
| Chapitre 2 | Services | 49 970 000 |
| Chapitre 3 | Frais de personnel | 202 500 000 |
| Chapitre 4 | Impôts et Taxes | 15 000 |
| Chapitre 5 | Frais financières | 35 000 |
| Chapitre 6 | Frais divers | 1 180 000 |
| Titre 3 | Dépenses d'investissement | 36 266 410 000 DA |
| Chapitre 4 | Agencement et installation | 300 000 |
| Chapitre 6 | Matériels équipements informatiques | 1500 000 |
| Chapitre 7 | Matériels de télécommunication | 30 000 |
| Chapitre 8 | Matériels et mobilier de bureaux | 700 000 |
| Chapitre 10 | Equipement audiovisuels | 50 000 |

Tableau 2 : budget prévisionnel de la CNR 2019

Chapitre 2
Généralité sur les séries
Chronologiques

2.1 Définition d'une série chronologique

Une série chronologique (temporelle), en anglais (time series), est une suite d'observations numériques (X_1, X_2, \dots, X_n) que prendra une variable au cours d'une période définie.

Les observations sont effectuées à intervalle de temps, elles sont ordonnées par un indice $t = 1, \dots, n$; la $t^{\text{ième}}$ observation est notée X_t .

L'indice temps peut être: la minute, l'heure, le jour, l'année, d'où les données sont journalières, hebdomadaires, mensuelles, trimestrielles, annuelles.

La théorie des séries chronologiques est appliquée de nos jours dans des domaines aussi divers que l'économétrie, la médecine ou la démographie, ...etc.

2.2 Représentation graphique d'une série chronologique

Les séries sont généralement représentées dans un repère orthonormé en portant en ordonnée les valeurs des observations et en abscisse les dates correspondantes, les points sont ainsi reliés par les segments de droite et le graphe comme une ligne brisée.

2.3 Composantes d'une série temporelle

2.3.1 La composante tendancielle (T_t)

Elle est censée décrire le mouvement à long terme d'une série chronologique. On admet d'ordinaire que cette projection à long terme est une ligne droite à la hausse ou à la baisse qui élimine toutes les fluctuations (aléatoires) dues aux facteurs saisonniers et cycliques.

2.3.2 La composante cyclique (C_t)

C'est une succession de mouvements persistant des variations de mouvements ascendants (période prospérité) et de mouvements descendants (période de dépression) retraçant le cycle économique, la tendance et le cycle sont regroupés en une seule composante appelée l'extra saisonnier (E_r).

2.3.3 La composante saisonnière (S_t)

Notée $(S_t)_{t \in \mathbb{Z}}$ représente les fluctuations périodiques des événements qui se répètent annuellement, on note par (p) sa période qui correspond à l'année. $p = 4$ série trimestrielle; $p = 12$ mensuelle.

2.3.4 La composante résiduelle ou aléatoire

Notée (ε_t) représente des perturbations irrégulières et imprévisibles due des événements du hasard. Comme : grèves, accidents et autres, elle résume tout ce que les autres composantes n'ont pas pu expliquer du phénomène observée, on les suppose de faible amplitude et présente une allure plus au moins stable autour de la moyenne.

2.4 Modèle de décomposition d'une série chronologique

2.4.1 Modèle additif

Dans un modèle additif, on suppose que les 3 composantes : tendance, variations saisonnières et variations accidentelles sont indépendantes les unes des autres.

On considère que la série X_t s'écrit comme la somme de ces 3 composantes :

$$X_t = T_t + C_t + S_t + \varepsilon_t \quad \text{pour tout } t \in \mathbb{Z}.$$

Avec :

$$\left\{ \begin{array}{l} (T_t): \text{ La tendance.} \\ (S_t): \text{ La composante saisonnière (fonction périodique).} \\ (C_t): \text{ La composante stationnaire.} \end{array} \right.$$

2.4.2 Modèle multiplicatif

On suppose que les variations saisonnières et les variations accidentelles dépendent de la tendance et on a alors :

$$X_t = T_t * C_t * S_t * \varepsilon_t \quad \text{pour tout } t \in \mathbb{Z}.$$

2.5 Identification des composantes de la série chronologique

Dans la prévision à court terme, on se base sur l'identification de la tendance et de la saisonnalité. Pour tester l'existence ou non de la composante saisonnière et tendancielle, pour cela on utilise les tests suivants :

2.5.1 Test graphique

D'après l'examen visuel des graphes ; on peut distinguer l'existence ou l'absence de la tendance et la saisonnalité mais pour plus de certitude ; on fait le test par l'analyse de la variance.

2.5.2 Test par l'analyse de la variance

Soit une analyse de variance à deux facteurs; en effet, toutes observations X_{ij} du tableau sont soumises à des influences de l'année et du mois qui lui correspondent, tel que :

$$i = 1, \dots, m \quad \text{avec :} \quad m : \text{le nombre de mois.}$$

$$j = 1, \dots, n \quad \text{n : le nombres d'année}$$

Le déroulement du test est :

A partir du tableau ANOVA ci-dessous on construit les tests d'hypothèses suivants :

Test de l'influence du facteur de la saisonnalité :

$$\begin{cases} H_0 : \text{pas de saisonnalité.} \\ H_1 : \text{la série est saisonnière.} \end{cases}$$

Calcul de FISHER empirique :

$$f_c = \frac{V_m}{V_r} ; \text{ On le compare à } F_{v_1, v_2}^\alpha \text{ tel que :}$$

$$v_1 = (m - 1) \text{ d. d. l.}$$

$$v_2 = (m - 1)(n - 1) \text{ d. d. l.}$$

Si $f_c > F_{v_1, v_2}^\alpha$ on rejette H_0 .

Test de l'influence du facteur de la tendance :

$$\begin{cases} H_0 : \text{pas de tendance.} \\ H_1 : \text{il existe est tendance.} \end{cases}$$

$$f_c = \frac{V_n}{V_r} ; \text{ On le compare à } F_{v_3, v_2}^\alpha \text{ tel que :}$$

$$v_3 = (n - 1) \text{ d. d. l.}$$

Si $f_c > F_{v_3, v_2}^\alpha$ on rejette H_0 .

Soit le tableau ANOVA pour détecter la tendance et la saisonnalité.

| Somme des carrées | d.d.l | Désignation | variance |
|-------------------|------------------|------------------|--------------------------------|
| S_m | $(m - 1)$ | variance mois | $V_m = \frac{S_m}{(m - 1)}$ |
| S_n | $(n - 1)$ | variance année | $V_n = \frac{S_n}{(n - 1)}$ |
| S_r | $(m - 1)(n - 1)$ | variance résidus | $V_r = \frac{S_r}{(n-1)(m-1)}$ |
| S_t | $nm - 1$ | variance total | / |

Tableau 1: test de FISHER

Calcule de la variance totale (S_t): qui est la somme totale des carrées.

$$S_t = \sum_{i=1}^n \sum_{j=1}^m (X_{ij} - \bar{X})^2 = m \sum_{i=1}^n (X_i - \bar{X})^2 + n \sum_{j=1}^m (X_j - \bar{X})^2 + \sum_{i=1}^n \sum_{j=1}^m (X_{ij} - X_i - X_j + \bar{X})^2$$

Ou:

$$S_n = m \sum_{i=1}^n (X_i - \bar{X})^2 : \text{ Somme des carrées des années.}$$

$$S_m = n \sum_{j=1}^m (X_j - \bar{X})^2 : \text{ Somme des carrées des mois.}$$

$S_r = \sum_{i=1}^n \sum_{j=1}^p (X_{ij} - X_{i.} - X_{.j} + \bar{X})^2$: Somme des carrées des résidus.

Et :

$\bar{X} = \frac{1}{nm} \sum_{i=1}^n \sum_{j=1}^m X_{ij}$; Moyenne arithmétique des (n, m) observations

$X_{i.} = \frac{1}{m} \sum_{j=1}^m X_{ij}$: Moyenne arithmétique pour l'année i.

$X_{.j} = \frac{1}{n} \sum_{i=1}^n X_{ij}$: Moyenne arithmétique pour le mois j.

2.6 Prédiction

La prédiction est une « détermination chiffrée ou non d'un phénomène, d'une grandeur ou d'un ensemble de grandeurs relatifs à « une période future » Les prévisions, qui constituent une « suite d'événements prévisibles et chiffrés suivant des hypothèses d'activité », s'appuient sur des données intérieures à l'entreprise (capacité, rendement...), et extérieures à l'entreprise (marché, coûts des matières...).

La prédiction se base sur des données historiques pour fournir un plan global du budget dans le futur. Il est important de ne négliger aucun facteur dans le calcul du budget futur. Le terme prévoir signifie concevoir, envisager par avance. Cette définition jugée un peu trop simpliste et voir même tautologique ne reflète qu'en partie le sens que nous voulons attribuer à ce terme.

2.7 Processus stochastiques et modèles de séries chronologiques

2.7.1 Définition

On appelle processus stochastique toute famille $(X_t)_{t \in \mathbb{Z}}$, de variables aléatoires définies sur un même espace de probabilité (Ω, \mathcal{A}, P) .

2.7.2 Caractéristiques d'un processus stochastique

Soit un processus stochastique $\{X_t; t \in \mathbb{Z}\}$ on a :

- $E(X_t) = \mu_t$ espérance mathématique ou moyenne de X_t .
- $V(X_t) = E[(X_t - \mu_t)]^2 = \sigma_t^2$ variance de X_t .
- $\text{cov}(X_t, X_{t'}) = E[(X_t - \mu_t)(X_{t'} - \mu_{t'})]$ covariance de X_t .

2.7.3 Processus stochastique stationnaire

2.7.3.1 Stationnarité forte (stricte)

Le processus $\{X_t, t \in \mathbb{Z}\}$ est strictement stationnaire si pour n-uplet du temps $t_1 < t_2 < t_3 < \dots < t_n$ tel que $t_i \in \mathbb{Z}$ et pour tout temps h; $t_{i+h} \in \mathbb{Z}$ et $\forall i = 1 \dots n$. la suite $(X_{t_1+h}, X_{t_2+h}, \dots, X_{t_n+h})$ a la même loi de probabilité que la suite $(X_{t_1}, X_{t_2}, \dots, X_{t_n})$

2.7.3.2 Stationnarité faible (du second ordre)

Un processus $\{X_t; t \in \mathbb{Z}\}$ est dit stationnaire au sens faible (du second ordre) si:

- $\forall t \in \mathbb{Z}: E(X_t^2) < \infty$ (une variance finie)
- $\forall t \in \mathbb{Z}: E(X_t) = m$ indépendant du temps
- $\forall t, h \in \mathbb{Z} : \text{cov}(X_t, X_{t+h}) = E[(X_{t+h} - m)(X_t - m)] = \gamma(h)$ indépendant du temps

2.7.4 Processus Bruit Blanc

Parmi la classe des processus stationnaires, il existe des processus particuliers. Ces processus sont très souvent utilisés en analyse des séries temporelles.

On appelle bruit blanc tout processus $(X_t)_{t \in \mathbb{Z}}$ de variables aléatoires centrées, de même variance σ^2 et non corrélées on le note (ε_t) .

- $E(\varepsilon_t) = 0, \forall t \in \mathbb{Z}$
- $V(\varepsilon_t) = \sigma_{\varepsilon_t}^2, \forall t \in \mathbb{Z}$.
- $\begin{cases} \text{Cov}(\varepsilon_t, \varepsilon_{t+h}) = 0. & \forall h \neq 0, t \neq h \\ \text{Cov}(\varepsilon_t, \varepsilon_{t+h}) = \sigma_{\varepsilon_t}^2. & \text{si } h = 0; t \neq h \end{cases}$

Un bruit Blanc est dit gaussien si la loi de la probabilité du processus est gaussien (ε_t) i.i.d $\sim N(0, \sigma_{\varepsilon_t}^2)$.

2.7.5 Fonction d'auto-covariance

Soit $\{X_t, t \in \mathbb{Z}\}$ un processus aléatoire, la fonction d'auto-covariance qui mesure la covariance pour un couple de valeurs associées à des dates différentes, dont l'intervalle est de longueur h , notée $\gamma(h)$ est définie par :

$$\gamma_h = \text{cov}(X_t; X_{t+h}) = E[(X_t - E(X_t))(X_{t+h} - E(X_{t+h}))] \quad \forall t \in \mathbb{Z}.$$

La fonction d'auto covariance d'un processus stationnaire vérifiés les propriétés suivantes :

- $\gamma_0 = \text{cov}(X_t; X_t) = V(X_t) = E[X_t - (E(X_t))]^2 = \delta_t^2 \geq 0.$
- $|\gamma_h| \leq \gamma_0.$
- $\gamma_{-t} = \gamma_t$; γ_t est une fonction symétrique pour tout $h \in \mathbb{N}$.

2.7.6 Fonction d'auto corrélation

Pour juger la stationnarité d'un processus, on utilise les fonctions d'auto-corrélation simple (FAC) et partielle (FACP).

2.7.6.1 Fonction d'auto corrélation simple

L'auto-corrélation d'un processus stochastique est la corrélation de ce processus avec une version décalée dans le temps de lui-même. Pour un processus stationnaire la fonction d'auto-corrélation (ρ_h) prend la forme suivante :

$$\rho_h = \rho(X_t; X_{t-h}) = \frac{\text{cov}(X_t; X_{t-h})}{\sqrt{V(X_t)V(X_{t-h})}} = \frac{\gamma_h}{\gamma_0}.$$

Le graphe représentant la fonction d'auto-corrélation est appelé corrélogramme.

La fonction d'auto-corrélation simple vérifié les propriétés suivantes :

- $\rho_0 = 1$
- $|\rho_h| \leq \rho_0$
- $\rho_h = \rho_{-h}$, ρ est une fonction symétrique: pour tout $h \in \mathbb{N}$.

2.7.6.2 Fonction d'auto-corrélation partielle

L'auto-corrélation partielle mesure la corrélation entre X_t et X_{t-h} sans toute fois prendre en considération l'influence des variables antérieures à X_{t-h} ainsi, on peut montrer que la fonction d'auto-corrélation partielle d'un processus $\{X_t, t \in \mathbb{Z}\}$ est donnée par

$$\phi_{hh} = \frac{|\rho_h^*|}{|\rho_h|}$$

Avec :

$$\rho_h = \begin{bmatrix} 1 & \cdots & \rho_{h-1} \\ \vdots & \ddots & \vdots \\ \rho_{h-1} & \cdots & 1 \end{bmatrix} ; \quad \rho_h^* = \begin{bmatrix} 1 & \cdots & \rho_1 \\ \vdots & \ddots & \vdots \\ \rho_{h-1} & \cdots & \rho_h \end{bmatrix}$$

Donc la fonction d'auto corrélation partielle :

$$\phi_{hh} = \begin{cases} \frac{\rho_h - \sum_{j=1}^h \phi_{h-1} \phi_{h-j}}{1 - \sum_{j=1}^h \phi_{h-1} \phi_{h-j}} \\ \rho_1 ; h = 1 \end{cases}$$

2.7.7 Les opérateurs

2.7.7.1 L'opérateur de retard (Backward)

Soit le processus $\{X_t; t \in \mathbb{Z}\}$ l'opérateur retard B transforme une observation en fonction de sa valeur passé, tel que :

$$BX_t = X_{t-1}$$

Cet opérateur peut être utilisé récursivement, pour obtenir l'avant-dernière observation avant $t - 1$:

$$B^2(X_t) = B(B(X_t)) = B(X_{t-1}) = X_{t-2} \quad \text{Plus généralement : } B^n X_t = X_{t-n}$$

2.7.7.2 L'opérateur d'avance F (Forward)

L'opérateur d'avance noté F associé à un processus $\{X_t, t \in \mathbb{Z}\}$ transforme une observation en fonction de sa valeur future, tel que :

$$FX_t = X_{t+1}$$

2.7.7.3 L'opérateur de différence Δ

On introduit également l'opérateur de différence Δ dont l'application à X_t fournit la différence entre X_t et la valeur précédente X_{t-1} de la série

$$\Delta X_t = X_t - X_{t-1}$$

En vertu de la définition précédente, on a directement :

$$\Delta X_t = (1 - B)X_t$$

Plus généralement, l'opérateur de différence d'ordre d envoie la valeur X_t de la série observée sur la différence de X_t avec la valeur prise par la série au temps $t - d$:

$$\Delta^d X_t = X_t - X_{t-d}$$

2.7.8 Théorème de Wold (décomposition de wold)

Tout processus stationnaire d'ordre 2, peut être représenté sous la forme suivante:

$$X_t = \sum_{j=0}^{\infty} \theta_j \varepsilon_{t-j} + k_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + k_t$$

Ou :

$$\left\{ \begin{array}{l} \theta_0 = 1; \theta_j \in \sum_{j=0}^{\infty} \theta_j^2 < \infty \\ (\varepsilon_t) \text{ est un BB i. i. d } \sim N(0, \sigma_{\varepsilon_t}^2) \end{array} \right.$$

La composante linéaire aléatoire correspond à la somme de chocs passés, la composante linéaire déterministe est : k_t

$$\text{Cov}(k_t ; \varepsilon_{t-j}) = 0, \forall j.$$

$$E(X_t) = m = k_t$$

D'où : le processus $(X_t ; t \in \mathbb{Z})$ peut s'écrire de la manière suivante

$$X_t = \sum_{j=0}^{\infty} \theta_j \varepsilon_{t-j} + m$$

2.8 Modèles de série chronologique

2.8.1 Les processus linéaires stationnaires

Wold montre que les séries stationnaires peuvent être représentées par les processus ARMA

2.8.1.1 Processus moyenne mobile MA(q)

On appelle un processus $(X_t)_{t \in \mathbb{Z}}$ d'ordre (q) un processus stationnaire s'écrit sous la forme suivante :

$$X_t = \Theta(B) * \varepsilon_t ; \quad t = 1, \dots, T$$

$$\Theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$$

q étant un entier fini et (ε_t) un processus BB

Donc :

$$X_t = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) * \varepsilon_t$$

$$X_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q}$$

$$MA(q) = ARMA(0, q)$$

2.8.1.2 Processus autorégressif AR(p)

On appelle processus autoregressif d'ordre (p) un processus $\{X_t ; t \in \mathbb{Z}\}$

Vérifie une relation de type :

$$\Phi(B) * X_t = \varepsilon_t \quad ; \forall t \in \mathbb{Z}$$

$$\Phi(B) = 1 - \varphi_1 B - \varphi_2 B^2 - \dots - \varphi_p B^p$$

p étant un entier fini et (ε_t) un processus BB.

$$X_t = \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \dots + \varphi_p X_{t-p} + \varepsilon_t$$

$$AR(p) = ARMA(p, 0)$$

2.8.1.3 Processus ARMA(p, q)

On dit qu'un processus X_t est du type **ARMA** s'il est stationnaire et satisfait l'équation :

$$\Phi(B) * X_t = \Theta(B) * \varepsilon_t$$

$$X_t = \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \dots + \varphi_p X_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q}$$

p, q deux entiers finis ε_t un processus BB

2.8.2 Les processus aléatoires non-stationnaires

2.8.2.1 Les processus TS

Un processus TS s'écrit : $X_t = f_t + \varepsilon_t$ où f_t est une fonction polynomiale du temps, linéaire ou non linéaire, et ε_t un processus stationnaire. Le processus TS le plus simple et le plus répandu est représenté par une fonction polynomiale de degré 1. Le processus TS porte alors le nom de linéaire et s'écrit :

$$X_t = a_0 + a_1 t + \varepsilon_t$$

Si (ε_t) est un bruit blanc (gaussien ou non), les caractéristiques de ce processus sont alors :

- $E(X_t) = a_0 + a_1 t + E(\varepsilon_t) = a_0 + a_1 t$.
- $V(X_t) = 0 + V(\varepsilon_t) = \sigma_\varepsilon^2$.
- $\text{Cov}(X_t, X_{t'}) = 0$ pour $t \neq t'$.

Ce processus TS est non stationnaire car $E(X_t)$ dépend du temps. Connaissant \hat{a}_0 et \hat{a}_1 , le processus X_t , peut être stationnarisé en retranchant la valeur de X_t en t la valeur estimée.

Dans ce type de modélisation, l'effet produit par un choc (ou par plusieurs chocs aléatoires) à un instant t est transitoire. Le modèle étant déterministe, la chronique retrouve son mouvement de long terme qui est ici la droite de tendance.

2.8.2.2 Les processus DS

Les processus DS sont des processus que l'on peut rendre stationnaires par l'utilisation d'un filtre aux différences :

$$(1 - B)^d X_t = \delta + \varepsilon_t \Leftrightarrow X_t = X_{t-1} + \delta + \varepsilon_t$$

L'introduction de la constante δ dans le processus DS permet de définir deux processus différents :

- $\delta = 0$: le processus DS est dit sans dérive, Il s'écrit:

$$X_t = X_{t-1} + \varepsilon_t$$

Comme ε_t est un bruit blanc, ce processus DS porte le nom de modèle de marche aléatoire au hasard ou de marche aléatoire (Random Walk Model). Il est très fréquemment utilisé pour analyser l'efficacité des marchés financiers.

- $\delta \neq 0$, le processus porte le nom de processus DS avec dérivé, il s'écrit:

$$X_t = X_{t-1} + \delta + \varepsilon_t$$

2.8.3 Modèle non stationnaire ARIMA et SARIMA

2.8.3.1 Processus ARIMA(p, d, q)

On a vu que si un processus X_t admet une tendance polynomiale de degré d , le processus différencié d fois est stationnaire :

$$Y_t = \Delta^d X_t = (1 - B)^d X_t$$

Le modèle ARIMA revient à appliquer un modèle ARMA sur le processus différencié :

$$Y_t = \text{ARMA}(p, q) \Leftrightarrow X_t = \text{ARIMA}(p, d, q)$$

L'équation d'un modèle ARIMA (p, d, q) est donc donné par :

$$\phi(B)(1 - B)^d X_t = \theta(B)\varepsilon_t$$

Où ϕ et θ sont deux polynômes de degrés respectifs p et q .

- Le 'I' de ARIMA signifie 'integrated' comme réciproque de la différenciation 'd' est un entier positif.
- (ε_t) est un bruit blanc ainsi, ajuster un processus ARIMA (p, d, q) sur la série non stationnaire X_t est équivalent à ajuster un processus ARMA (p, q) sur la série stationnaire $(1 - B)^d X_t$.
- Les processus ARMA apparaissent comme un cas particulier des processus ARIMA dans lesquels $d = 0$.

2.8.3.2 Processus SARIMA

La saisonnalité est un autre facteur de non-stationnarité. On a vu qu'une façon simple d'éliminer une saisonnalité de période s consiste à appliquer l'opérateur Δ_s .

$$Y_t = \Delta_s X_t = X_t - X_{t-s} = (1 - B^s)X_t$$

De façon générale, on peut supposer que l'influence des choses se transmet entre dates distantes d'un nombre entier de périodes selon un processus ARIMA (p, d, q) :

$$\phi_s(B^s)\Delta_s^d X_t = \theta_s(B^s)\varepsilon_t \quad \text{Où } \varepsilon_t \sim \text{BB.}$$

Chapitre 3
Méthodologie de
Holt & Winters

3.1 Généralisation sur le lissage exponentiel

Les méthodes de lissage exponentiel sont très utilisées par les praticiens de la gestion (notamment pour la gestion des stocks) et les économistes. Leur succès est dû à la qualité des résultats. Elles ne reposent sur aucune hypothèse probabiliste et fournissent des prévisions ponctuelles. Compte tenu de leur simplicité,

Introduites par Holt en 1958, Winters en 1960 et popularisées par le livre de Brown en (1963), les méthodes de lissage constituent l'ensemble des techniques empiriques de prévision qui accordent plus ou moins d'importance aux valeurs du passé d'une série temporelle.

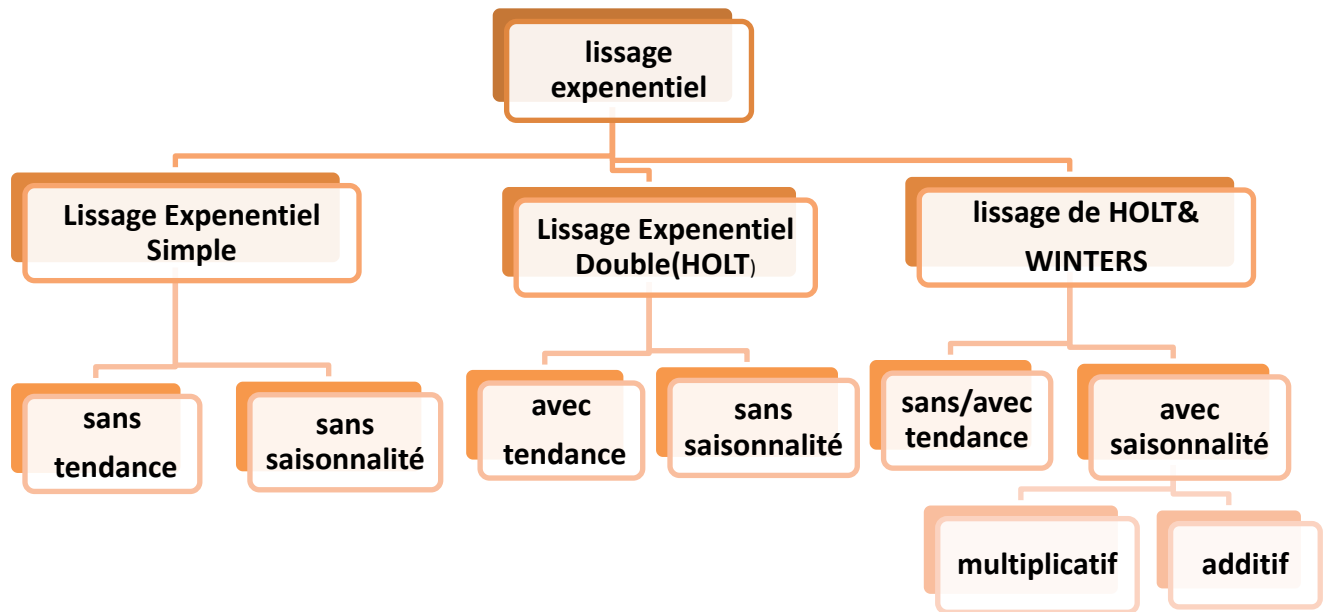
Les méthodes de lissage exponentiel sont :

- Le lissage exponentiel simple dépend d'un **seul** paramètre de lissage.
- Le lissage de Holt dépend de **deux** paramètres: l'un relatif au niveau, l'autre relatif à la tendance.
- Le lissage de Winters dépend de **trois** paramètres: l'un relatif au niveau, un autre relatif à la tendance, et le dernier à la saisonnalité.

3.2 Principe des méthodes de lissage exponentiel

- Les méthodes de lissage exponentiel sont des méthodes de prévision à court terme.
- Elles supposent que le phénomène étudié ne dépend que de ses valeurs passées.
- Ce sont des méthodes d'extrapolation qui donnent un poids prépondérant aux valeurs récentes: les coefficients de pondération décroissent exponentiellement en remontant dans le temps.
- Chacune des méthodes dépend d'un ou plusieurs paramètres (paramètres de lissage) compris entre 0 et 1.
- Le poids de chacune des valeurs passées se calcule à partir de ces paramètres.
- Les méthodes de prévision se sont développées au cours de la seconde.

3.3 Les différents lissages exponentiel



3.3.1 Lissage exponentiel de Holt & Winters sans saisonnalité

Le lissage exponentiel simple (LES) s'applique à des séries chronologiques sans saisonnalité et à tendance localement constante.

Soit $\hat{X}_T(h)$ la prévision à la date (T) pour l'horizon h , c'est-à-dire pour la date (T + h)

Dans le cadre du LES qui s'applique à des séries sans tendance, la prévision faite à la date (T) est une valeur constante indépendante de l'horizon h :

$$\hat{X}_T(h) = \hat{X}_T$$

$$\hat{X}_T = \beta \hat{X}_T + \beta \hat{X}_{T-1} \Rightarrow \hat{X}_T = \beta \sum_{i=0}^{T-1} (1 - \beta)^i \hat{X}_{T-1}$$

Où β est un paramètre compris entre 0 et 1.

Remarque

- β proche de 0, prévision **souple** i.e. fortement influencée par les observations les plus récentes.
- $\beta = 0$, la prévision est alors égale à la dernière valeur observée.
- β est proche de 1, prévision **rigide** i.e. l'influence des observations passées est d'autant plus importante et remonte loin dans le passé.
- $\beta = 1$, alors toutes les prévisions sont identiques.

Formules de mise à jour

$$X_T = \beta X_T + (1 - \beta) \hat{X}_{T-1} = \hat{X}_{T-1} + (1 - \beta)(X_T - \hat{X}_{T-1}).$$

3.3.2 Lissage exponentiel de Holt (double)

La prévision de la série à l'horizon h , $\hat{X}_T(h)$, fournie par la méthode de lissage exponentiel double est donnée par :

$$X_T(h) = \hat{a}_T h + \hat{b}_T$$

Où β est la **constante de lissage** et le couple (\hat{a}_T, \hat{b}_T) est donné par:

$$\begin{cases} \hat{a}_T = \frac{1-\beta}{\beta} (S_1(T) - S_2(T)) \\ \hat{b}_T = 2S_1(T) - S_2(T) \end{cases}$$

$$\text{Tel que : } \begin{cases} S_1(T) = (1 - \beta) \sum_{j=0}^{T-1} \beta^j X_{T-j} \\ S_2(T) = (1 - \beta) \sum_{j=0}^{T-1} \beta^j S(T - j) \end{cases}$$

3.3.3 Méthodes de Holt & Winters

Le lissage de Holt & Winters concerne les séries chronologiques saisonnières. On commence par choisir le modèle de composition, car :

- Il y a une méthode de lissage pour les séries chroniques avec saisonnalité additive.
- Et une autre méthode pour les séries chroniques avec saisonnalité multiplicative.

3.3.3.1 Modèle saisonnier additif

On considère une série chronologique dont on a observé les T premiers instants X_1, \dots, X_T et on suppose que cette série peut être approchée, au voisinage de T par le modèle suivant :

$$a(t - T) + b + S_t$$

Où S_t représente la saisonnalité de période p . La méthode de Holt & Winters propose pour l'estimation de a , b et S_t les formules de mise à jour suivante :

$$\begin{cases} \hat{a}_T = (1 - \beta) \hat{a}_{T-1} + \beta (\mathfrak{b}_T - \mathfrak{b}_{T-1}) \\ \mathfrak{b}_T = \alpha (X_T - \hat{S}_{T-p}) + (1 - \alpha) (\mathfrak{b}_{T-1} + \hat{a}_{T-1}) \\ \hat{S}_T = \gamma (X_T - \mathfrak{b}_T) + (1 - \gamma) \hat{S}_{T-p} \end{cases}$$

Où α , β et γ sont des constantes de lissage appartenant à $]0,1[$.

La prévision à l'horizon h est donc :

$$\hat{X}_T(h) = \begin{cases} \hat{a}_T * h + \mathfrak{b}_T + \hat{S}_{T+h-p} & \text{si } 1 \leq h \leq p \\ \hat{a}_T * h + \mathfrak{b}_T + \hat{S}_{T+h-2p} & \text{si } P + 1 \leq h \leq 2 \end{cases}$$

3.3.3.2 Modèle saisonnier multiplicatif

On considère une série chronologique dont on a observé les T premiers instants X_1, \dots, X_T et on suppose que cette série peut être approchée, au voisinage de T par le modèle suivant :

$$a((t - T) + b) * S_t$$

S_t : La saisonnalité

L'estimation de a, b et S_t sont suivant :

$$\begin{cases} \hat{a}_T = (1 - \beta) \hat{a}_{T-1} + (1 - \beta) (b_T - b_{T-1}) \\ b_T = \alpha (X_T / \hat{S}_{T-p}) + (1 - \alpha) (b_{T-1} + \hat{a}_{T-1}) \\ \hat{S}_T = \gamma (\hat{X}_T / b_T) + (1 - \gamma) \hat{S}_{T-p} \end{cases}$$

La prévision à l'horizon h est donc :

$$\hat{X}_T(h) = \begin{cases} (\hat{a}_T * h + b_T) \hat{S}_{T+h-p} & \text{si } 1 \leq h \leq p \\ (\hat{a}_T * h + b_T) \hat{S}_{T+h-2p} & \text{si } p + 1 \leq h \leq 2p \\ \dots & \dots \end{cases}$$

On prend les valeurs initiales suivantes :

$$\begin{cases} \hat{a}(S) = \frac{1}{S} \left(\sum_{i=1}^S X_T \right) \\ \hat{b}(S) = 0 \\ \hat{S}_s = \frac{X_s}{\hat{a}(S)} \end{cases}$$

3.4 Application par l'approche de la méthode Holt & Winters

3.4.1 Etude de la série Budget

Avant l'étude approfondie de la série BUDGET qui présente le budget de la CNR de janvier 2014 jusqu'à décembre 2018, soit 60 observation. Il convient de vérifier la présence de saisonnalité et de la tendance du modèle par le biais des tests.

Pour tester la saisonnalité d'une série on utilise le test de Fisher.

(Tester est approuver dans l'application Box & Jenkins), et pour la tendance nous avons appliqué le test d'ADF.

Comme notre série est saisonnière et sans tendance, la méthode adéquate du lissage exponentiel est celle de Winters.

La méthode de Winters est caractérisée par deux modèles additif et multiplicatif pour faire la distinction entre ces deux modèles on utilise la méthode analytique de Buys -Ballot.

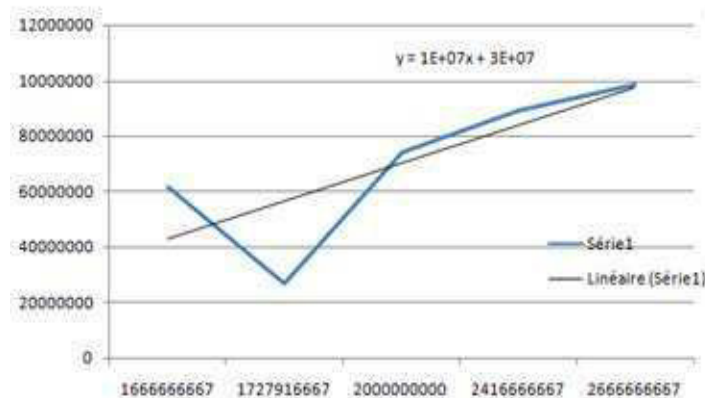


Figure 3.1: écart-type moyenne et droite de régression

- La droite des moindres carrés ordinaires est donnée par : $y = 1E+07x + 3E+07$.
- On a : $a = 1E + 07$ est différent de 0 donc notre modèle est multiplicatif.

L'application de la méthode de Holt & Winters saisonnière sur la série BUDGET a donné à l'aide du logiciel EVIEWS 4.0 les résultats suivants:

3.4.1.1 Coefficient du modèle

```

Date: 05/14/19 Time: 13:32
Sample: 2014M01 2018M12
Included observations: 60
Method: Holt-Winters Multiplicative Seasonal
Original Series: BUDGETT
Forecast Series: BUDGETSM

```

| | | |
|--------------------------|-------|----------|
| Parameters: | Alpha | 1.0000 |
| | Beta | 0.0000 |
| | Gamma | 0.0000 |
| Sum of Squared Residuals | | 6.43E+16 |
| Root Mean Squared Error | | 32733060 |

| | | |
|-----------------------|---------|----------|
| End of Period Levels: | Mean | 2.84E+09 |
| | Trend | 20833333 |
| Seasonals: | 2018M01 | 1.030453 |
| | 2018M02 | 1.019299 |
| | 2018M03 | 1.008392 |
| | 2018M04 | 1.018295 |
| | 2018M05 | 1.007643 |
| | 2018M06 | 0.997218 |
| | 2018M07 | 0.977043 |
| | 2018M08 | 0.967152 |
| | 2018M09 | 0.957466 |
| | 2018M10 | 1.015557 |
| | 2018M11 | 1.005614 |
| | 2018M12 | 0.995868 |

Les paramètres du lissage qui minimisent la somme des carrés résiduels sont :

- Coefficient de lissage de la moyenne $\alpha = 1.00$
- Coefficient de lissage de la tendance $\beta = 0.00$
- Coefficient de lissage de la saisonnalité $\gamma = 0.00$

La somme des carrés résiduels est égale à SCE = 6.43E + 16, d'où la racine moyenne des carrés résiduels est RMSE = 32733060

3.4.1.2 Prévission de la série BUDGET

Les prévisions pour les mois de l'année 2019 de la série BUDGET par la méthode Holt & Winters saisonnière multiplicatif sont les suivantes :

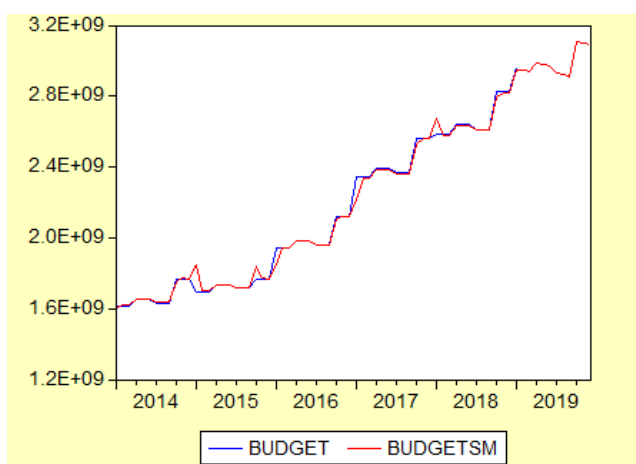


Figure 3.2 : graphe des prévisions de la série BUDGET

| Mois de l'année 2019 | Prévisions (DA) |
|----------------------|-----------------|
| Janvier | 2 946 298 716 |
| Février | 2 935 642 411 |
| Mars | 2 925 237 980 |
| Avril | 2 975 181 118 |
| Mai | 2 965 049 253 |
| Juin | 2 955 148 257 |
| Juillet | 2 915 717 975 |
| Août | 2 906 350 874 |
| Septembre | 2 897 189 822 |
| Octobre | 3 094 125 804 |
| Novembre | 3 084 780 536 |
| Décembre | 3 075 633 780 |

3.4.2 Etude de la série RETRAITE

Considérons la série **RETRAITE** qui regroupe l'évolution du budget de retraite au niveau de CNR de Boumerdes. Les observations de chaque mois durant la période allant De janvier 2014 jusqu'à Décembre 2018 .Les données Forment une série de «60» observations.

La série est saisonnière avec tendance, la méthode adéquate du lissage exponentiel est celle de Winters.

On utilise la méthode analytique de Buys-Ballot, qui nous donne le graphe suivant :

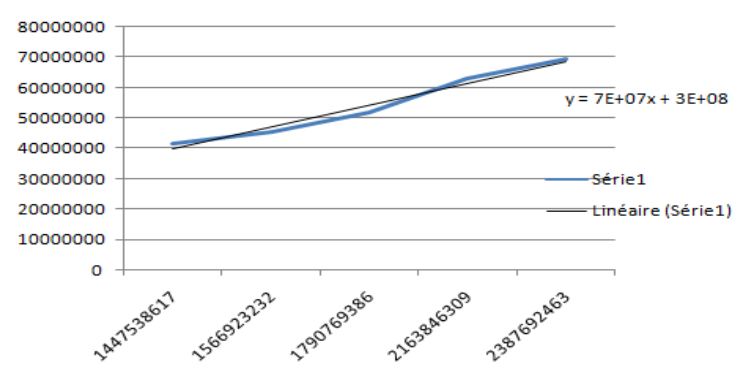


Figure 3. 3: écart-type ,moyenne et droite de régression

- La droite des moindres carrés ordinaires est donnée par : $y = 7E+07x + 3E+08$
- On a : $a = 7E+07$ est différent de 0 Donc notre modèle est multiplicatif.

L'application de la méthode de Holt & Winters sur la série donne les coefficients suivants:

3.4.2.1 Coefficient du modèle

```

Date: 05/14/19 Time: 14:02
Sample: 2014M01 2018M12
Included observations: 60
Method: Holt-Winters Multiplicative Seasonal
Original Series: RETRAITE
Forecast Series: RETRAISM
    
```

| | | |
|-------------|--------------------------|----------|
| Parameters: | Alpha | 1.0000 |
| | Beta | 0.0000 |
| | Gamma | 0.0000 |
| | Sum of Squared Residuals | 3.67E+16 |
| | Root Mean Squared Error | 24733859 |

| | | |
|-----------------------|------------|----------|
| End of Period Levels: | Mean | 2.76E+09 |
| | Trend | 21218750 |
| | Seasonals: | |
| | 2018M01 | 1.037341 |
| | 2018M02 | 1.025446 |
| | 2018M03 | 1.013831 |
| | 2018M04 | 1.010687 |
| | 2018M05 | 0.999517 |
| | 2018M06 | 0.988599 |
| | 2018M07 | 1.009738 |
| | 2018M08 | 0.998982 |
| | 2018M09 | 0.988459 |
| | 2018M10 | 0.985882 |
| | 2018M11 | 0.975729 |
| | 2018M12 | 0.965790 |

Les paramètres du lissage qui minimisent la somme des carrés résiduels sont :

- Coefficient de lissage de la moyenne $\alpha = 1.00$.
- Coefficient de lissage de la tendance $\beta = 0.00$.

- Coefficient de lissage de la saisonnalité $\gamma = 0.00$.

La somme des carrés résiduels est égale à $SCE = 3.67E + 16$, d'où la racine moyenne des carrés résiduels est $RMSE = 24733859$.

3.4.2.2 Prédiction de la série retraite

Les prévisions pour les mois de l'année 2019 de la série RETRAITE par la méthode Holt & Winters saisonnière multiplicatif sont les suivantes :

| Mois de l'année 2019 | Prévisions (da) |
|----------------------|-----------------|
| janvier | 2883659871 |
| février | 2872352589 |
| mars | 2861329686 |
| avril | 2873904336 |
| mai | 2863349360 |
| juin | 2853049523 |
| juillet | 2935482017 |
| août | 2925407025 |
| septembre | 2915566174 |
| octobre | 2928883555 |
| novembre | 2919426307 |
| décembre | 2910180968 |

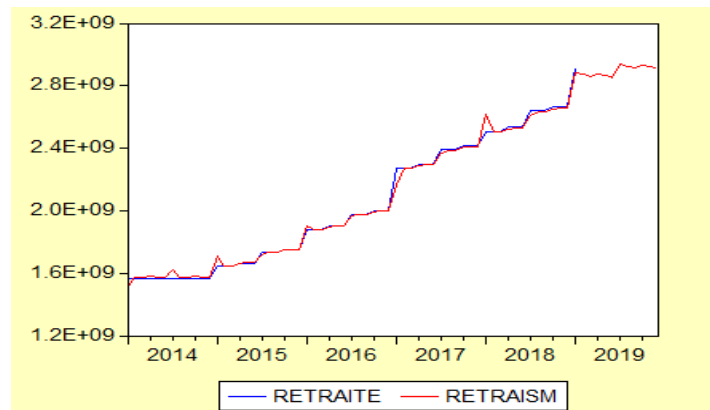


Figure 3.4: prédiction des données de la série RETRAITE

3.4.3 Etude de la série RECETTE

Nous disposons d'une série mensuelle évoluant de janvier 2009 jusqu' à décembre 2018, ou toutes la variables sont présentées en une série de 120 observations .Elle représente l'évolution de quote -part (11.75% parmi 26% déclaré par l'employeur à la CNAS) au sein de CNR de Boumerdes.

On utilise la méthode analytique de Buys -Ballot, qui nous donne le graphe suivant :

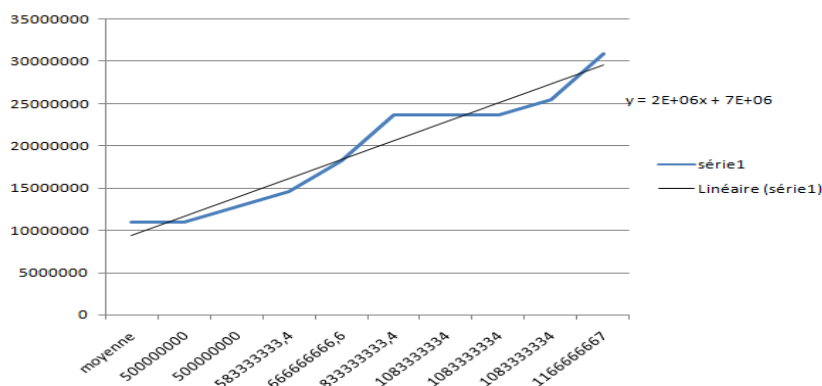


Figure 3. 5: écart-type,moyenne et la droite de régression

- La droite des moindres carrés ordinaires est donnée par : $y = 2E + 06x + 7E + 06$.
- On a : $a=2^E+06$ est différent de 0 donc notre modèle est multiplicatif.

L'application de la méthode de Winters sur la série donne les coefficients suivants:

3.4.3.1 Coefficient du modèle

| Included observations: 120 | | |
|--|---------|----------|
| Method: Holt-Winters Multiplicative Seasonal | | |
| Original Series: RECETTE | | |
| Forecast Series: RECETTSM | | |
| Parameters: | Alpha | 1.0000 |
| | Beta | 0.0000 |
| | Gamma | 0.0000 |
| Sum of Squared Residuals | | 8.28E+16 |
| Root Mean Squared Error | | 26264138 |
| End of Period Levels: | Mean | 1.50E+09 |
| | Trend | 8487654. |
| Seasonals: | 2018:01 | 1.041582 |
| | 2018:02 | 1.029543 |
| | 2018:03 | 1.017819 |
| | 2018:04 | 1.021800 |
| | 2018:05 | 1.010495 |
| | 2018:06 | 0.999471 |
| | 2018:07 | 0.983747 |
| | 2018:08 | 0.973304 |
| | 2018:09 | 0.963107 |
| | 2018:10 | 0.996473 |
| | 2018:11 | 0.986300 |
| | 2018:12 | 0.976358 |

Les paramètres du lissage qui minimisent la somme des carrées résiduels sont :

- Coefficient de lissage de la moyenne $\alpha = 1.00$.
- Coefficient de lissage de la tendance $\beta = 0.00$.
- Coefficient de lissage de la saisonnalité $\gamma = 0.00$.

La somme des carrés résiduelles est égale a $SCE = 8.28E + 16$ D'où la racine moyenne des carrés résiduels est : $RMSE = 26264138$.

3.4.3.2 Prédiction de la série RECETTE

Les prévisions pour les mois de l'année 2019 de la série recette par la méthode Holt & Winters saisonnière multiplicatif sont les suivantes :

| Mois de l'année 2019 | Prévisions (DA) |
|----------------------|-----------------|
| janvier | 1573 041 562 |
| février | 1 563 597 567 |
| mars | 1 554 431 692 |
| avril | 1 569 184 248 |
| mai | 1 560 399 792 |
| juin | 1 551 859 286 |
| juillet | 1 535 795 519 |
| août | 1 527 752 172 |
| septembre | 1 519 921 566 |
| octobre | 1 581 035 527 |
| novembre | 1 573 266 581 |
| décembre | 1 565 693 838 |

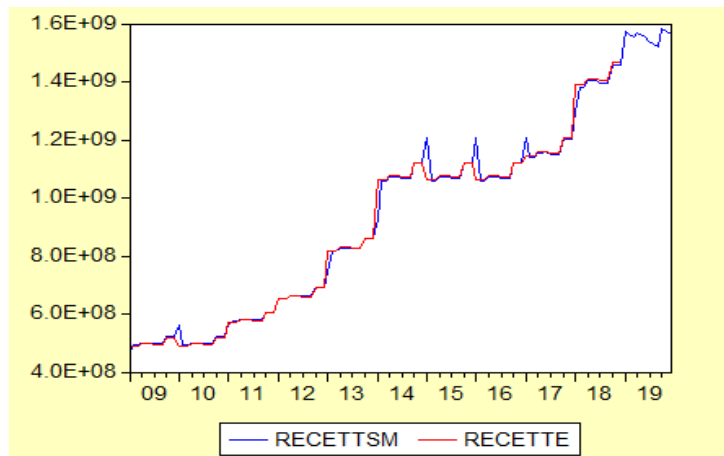


Figure 3.6 : les prévisions de la série RECETTE

3.4.4 Etude de la série DEPENSE

Considérons la série **DEPEN** qui regroupe l'évolution des dépenses de l'agence (CNR de Boumerdes). Les observations de chaque mois durant la période allant de janvier 2009 jusqu'à décembre 2018. Les données forment une série de « 120 » observations.

On utilise la méthode analytique de Buys & Ballot, qui nous donne le graphe suivant :

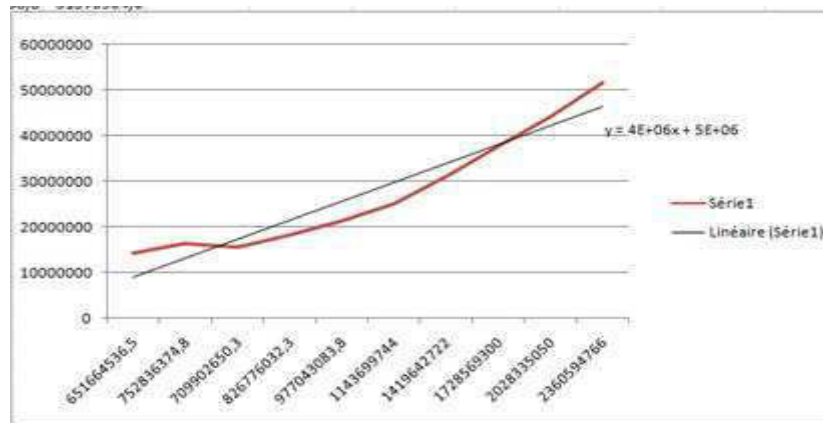


Figure 3.7 : écart-type, moyenne et la droite de régression

- La droite des moindres carrés ordinaires est donnée par : $y = 4E+06x + 5E+06$
- On a : $a=4^E+06$ est différent de 0 donc notre modèle est Multiplicatif.

L'Application de la méthode Holt & Winters saisonnière multiplicatif sur la série **DEPEN** a l'aide du logiciel EVIEWS4.0 les résultats suivant :

3.4.4.1 Coefficient du modèle

| Method: Holt-Winters Multiplicative Seasonal | | |
|--|--------------------------|----------|
| Original Series: DEPEN | | |
| Forecast Series: DEPENSM | | |
| Parameters: | Alpha | 0.8700 |
| | Beta | 0.0600 |
| | Gamma | 0.7602 |
| | Sum of Squared Residuals | 9.74E+16 |
| | Root Mean Squared Error | 28369542 |
| End of Period Levels: | Mean | 2.41E+09 |
| | Trend | 18609578 |
| | Seasonals: | |
| | 2018:02 | 1.041933 |
| | 2018:03 | 1.032088 |
| | 2018:04 | 1.034715 |
| | 2018:05 | 1.019610 |
| | 2018:06 | 1.004029 |
| | 2018:07 | 0.983705 |
| | 2018:08 | 0.968880 |
| | 2018:09 | 0.954553 |
| | 2018:10 | 0.983476 |
| | 2018:11 | 0.969497 |
| | 2018:12 | 0.955979 |
| | 2019:01 | 1.051536 |

Les paramètres du lissage qui minimisent la somme des carrés résiduels sont :

- Coefficient de lissage de la moyenne $\alpha = 0.87$
- Coefficient de lissage de la tendance $\beta = 0.06$
- Coefficient de lissage de la saisonnalité $\gamma = 0.76$

La somme des carrés résiduelles est égale à $SCE = 9.74E + 16$. D'où la racine moyenne des carrés résiduels est : $RMSE = 28369542$.

3.4.4.2 Prédiction de la série DEPENSE

Les prévisions pour les mois de l'année 2019 de la série dépense par la méthode Holt & Winters saisonnière multiplicatif sont les suivantes :

| Mois de l'année 2019 | Prévisions (DA) |
|----------------------|-----------------|
| Janvier | 2 716 906 097 |
| Février | 2 535 518 550 |
| Mars | 2 530 766 265 |
| Avril | 2 556 463 815 |
| Mai | 2 538 117 539 |
| Juin | 2 518 016 897 |
| Juillet | 2 485 353 395 |
| Août | 2 465 927 136 |
| Septembre | 2 447 228 233 |
| Octobre | 2 539 680 370 |
| Novembre | 2 521 623 457 |
| Décembre | 2 504 253 845 |

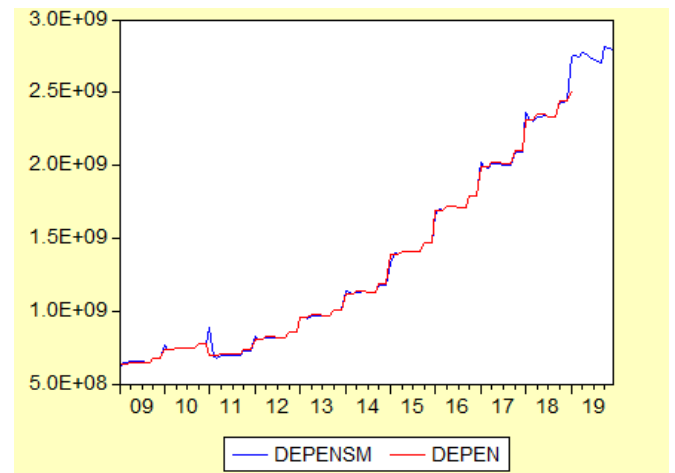


Figure 3.8 : les prévisions des données Dépense

Chapitre 4
Méthodologie de
Box & Jenkins

Introduction

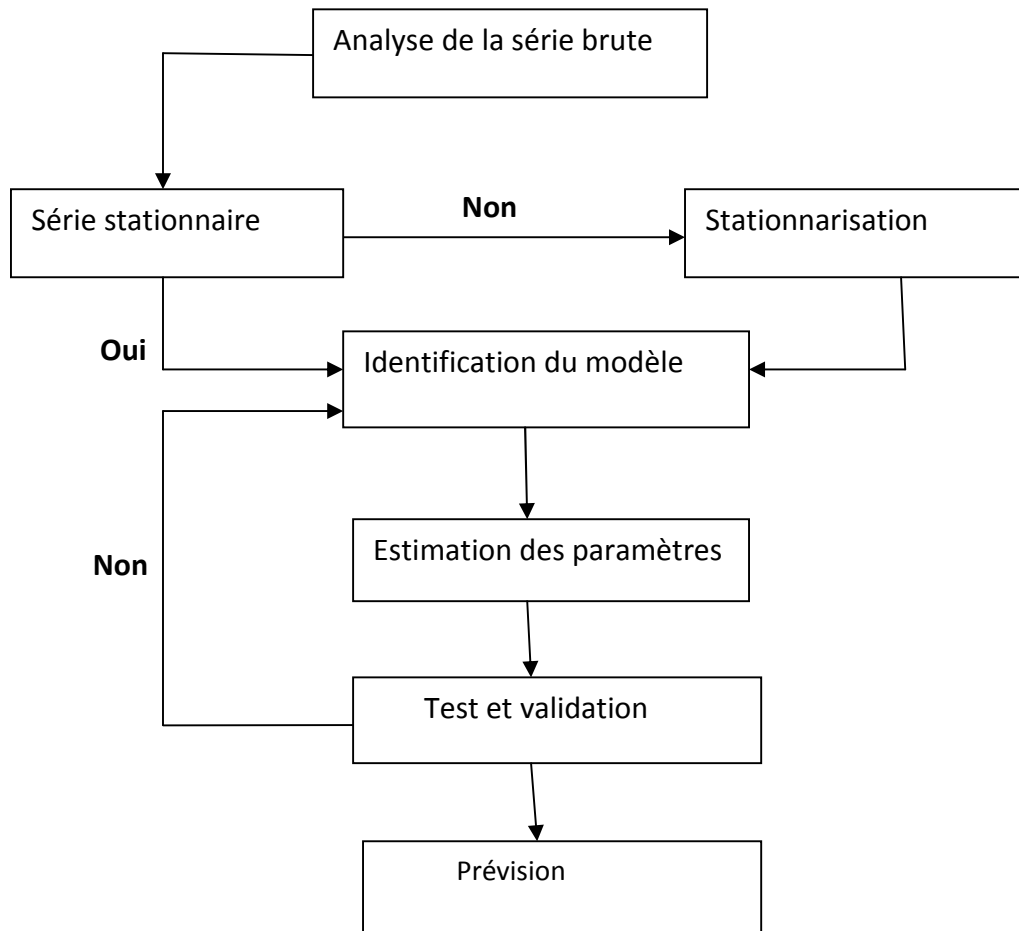
Dans ce chapitre nous allons finalement résoudre le problème de la construction en toute généralité des prédicteurs pour des processus soit stationnaires, soit non stationnaires à cause de la présence éventuelle d'une tendance ou d'une saisonnalité. Nous allons voir surtout comment l'approche paramétrique de Box-Jenkins, c'est-à-dire la modélisation du type ARMA à un nombre fini des paramètres, est capable de fournir de bons prédicteurs (linéaires) si le nombre d'observations d'une période assez homogène de la fin de notre série est suffisamment grand.

4.1 Méthodologie de Box –Jenkins

La partie autorégressive d'un processus, notée AR, est constituée par une combinaison linéaire finie des valeurs passées du processus. La partie moyenne mobile, notée MA, constituée par une combinaison linéaire finie en t des valeurs passées d'un Bruit Blanc. Wold (1954) montre que les modèle ARMA de représenter la plupart du processus stationnaire.

L'approche de box Jenkins (1970) consiste en une méthodologie d'étude systématique des séries chronologiques à partir de leurs caractéristiques afin de déterminer, dans la famille des modèles **ARMA**, la plus adapté à représenter le phénomène étudié, en fait elle représente une similitude très prononcé avec la méthode scientifique qui rappelons nous, consiste à formuler la supposition sous forme d'un modèle et à le mettre à l'épreuve et à réviser le modèle en conséquence, ces étapes étant répétées autant de fois que nécessaire. Une fois le modèle **ARMA** connu, on détermine mécaniquement les prévisions à court terme des valeurs futures du processus étudié, comme il faut encore pouvoir représenter la tendance et la saisonnalité, on étend dans ce cas la classe du modèle aux modèles **ARIMA** et **SARIMA**.

La démarche suivie par la méthode de Box-Jenkins est celle en quatre étapes de toutes les méthodes d'analyse prévisionnelle. A ce titre elle peut être schématisée par le diagramme suivant :



4.1.1 Familiarisation avec les données

Cette étape permet à l'utilisateur de s'informer sur le domaine dont il relève les données à savoir : leur nature (journalières, hebdomadaires, mensuelle, annuelle,...) les éléments influents et les objectifs poursuivis (ponctuels ou par intervalle).

La familiarisation avec les données se fait par l'examen graphique de la série qui peut révéler des changements structurelles, tel que :

- L'existence d'une saisonnalité ou d'une tendance.
- L'existence d'un point de rupture.

4.1.2 Analyse préliminaire

Cette étape fournit à l'utilisateur des outils lui permettant de corriger les valeurs aberrantes, compléter celle manquantes et enfin de transformer les données et cela pour se ramener à un modèle stationnaire, car la méthodologie de Box Jenkins est conditionnée par la stationnarité de la série.

4.2 Stationnarisation de la série

L'examen du graphique ne suffit pas très souvent pour mettre en évidence une saisonnalité, donc pour confirmer la non stationnarité de la série on utilise les tests suivants :

4.2.1 Test de la tendance (Dickey Fuller)

Le test de Dickey-Fuller (DF) permet de mettre en évidence le caractère stationnaire ou non d'une série par la détermination d'une tendance de type déterministe (TS) ou aléatoire (DS).

Pour la construction de ce test, Dickey et Fuller ont proposé trois modèles de base :

- $(1 - \phi B) X_t = \varepsilon_t$: Modèle autorégressif d'ordre 1.....**Modèle [1]**
- $(1 - \phi B) (X_t - c) = \varepsilon_t$: Modèle autorégressif d'ordre 1 avec constante...**Modèle [2]**
- $(1 - \phi B) (X_t - c - bt) = \varepsilon_t$: Modèle autorégressif d'ordre 1 avec tendance....**Modèle [3]**

Les hypothèses à tester sont :

$$\begin{cases} H_0: \phi_1 = 1 \\ H_1: \phi_1 < 1 \end{cases}$$

C'est un test itératif qui se déroule comme suit :

- Si Dans l'un des modèles la statistique relative à ϕ_1 est supérieure aux valeurs tabulées

$t_{0.05}$, On déduit qu'il existe une racine unitaire. On estime en premier le modèle [3] ; on commence par tester la significativité de la tendance :

- Si la tendance n'est pas significative alors on teste la significativité de la constante.
- Si elle est significative le processus est de type **TS** avec dérive, pour stationnariser la chronique on applique la méthode moindres carrés ordinaires pour l'estimation de la tendance.
- Si elle n'est pas significative, le processus est donc une marche aléatoire dit **DS** sans dérive, on différencie la série une fois et on recommence jusqu'à ce que la série soit stationnaire.

➤ Désaisonnalisation

Dans le cas où le test confirme que la série est affectée d'un mouvement saisonnier, il convient de la retirer préalablement avant tout traitement statistique :

- Si la série est saisonnière, dans ce cas là on applique la différenciation saisonnière.
- En revanche si elle est affectée d'une tendance on applique la différenciation ordinaire ou un ajustement par fonction (exponentiel, logarithmique, linéaire ou autre ...), dans le cas où elle est déterministe.

4.3 Les étapes de la méthode de Box & Jenkins

4.3.1 Identification

Après s'être assuré que la série ne contient ni racine unitaire, ni composante saisonnière, le choix de p et q repose sur l'observation des corrélogramme, c'est-à-dire de l'ensemble des fonctions d'auto-corrélation simple et des fonctions partielles pour déterminer dans la famille des modèles ARMA le modèle le plus susceptible de représenter la série.

La sélection d'un modèle parmi l'ensemble des modèles proposés se fait à l'aide de quelques critères d'information qui permettent de ce fait de réduire considérablement le champ de recherche du modèle.

Les critères de choix de modèle

➤ Critères de pouvoir prédictif

- L'estimateur $\hat{\sigma}^2$ de la variance du résidu ... (1).
- Le coefficient de détermination $R^2 = 1 - \frac{\hat{\sigma}^2}{\gamma}$, γ étant la variance empirique de la série... (2).
- Le coefficient de détermination modifié $\bar{R}^2 = 1 - \frac{\hat{\sigma}^2/(N-p-q)}{\gamma/(N-1)}$ qui prend en compte les degrés des polynômes auto régressifs et moyenne mobile..... (3).
- La statistique de Fisher $F = \frac{\gamma - \hat{\sigma}^2/(p+q)}{\hat{\sigma}^2/(N-p-q)}$ (4).

Le but est de minimiser (1) et maximiser (2), (3) et (4).

➤ Critères d'information

Ils existent des critères d'informations qui sont utilisés, comme guide, dans le choix du modèle. Parmi ces critères, ils existent les critères d'information qui mesurent l'écart entre la vraie loi inconnue et celle du modèle proposé. Les estimateurs de la quantité d'information qui ont été proposés sont :

$$\text{AIC} = \log \hat{\sigma}_\varepsilon^2 + \frac{2(p+q)}{n} \text{ (Akaike Information Criterion).}$$

$$\text{SCH} = \log \hat{\sigma}_\varepsilon^2 + (p+q) \frac{\log n}{n} \text{ (Critère de Schwartz 1978).}$$

Les critères de choix consistent à minimiser AIC et SCH

4.3.2 Estimation du modèle

Il s'agit, dans cette étape, d'estimer les paramètres du ou des modèles ARMA sélectionnée à l'étape d'identification. Les méthodes d'estimation diffèrent selon le type de processus diagnostiqué.

Dans le cas d'un modèle **AR**, nous pouvons appliquer une méthode des moindres carrés ou bien nous pouvons utiliser les relations existantes entre l'auto corrélation et les coefficients des modèles (équations de Yule-Walker).

Par contre l'estimation des processus **MA** ou **ARMA** repose généralement sur des techniques d'estimation fondées sur la maximisation d'une fonction de vraisemblance.

4.3.3 Validation du modèle

La validation du modèle consiste à vérifier si le modèle retenu et estimé est validé ou non.

4.3.3.1 Test sur les paramètres

Les coefficients du modèle doivent être significativement non nulle (le test de student s'applique de manière classique). Si un modèle n'est pas significativement différent de zéro, il convient alors de procéder à une nouvelle identification en éliminant l'ordre du modèle **AR** ou **MA** non valide.

4.3.3.2 Test sur résidus

Le principe de ce test est de vérifier l'hypothèse du non corrélation des perturbations aléatoires (les résidus entre les valeurs observées et les valeurs estimées doivent se comporter comme un bruit blanc).

➤ Test de Box-Pierce

Le test de Box-Pierce « portemanteau » permet d'identifier les processus sans mémoire (suite de variables aléatoire indépendantes entre elles). Nous devons donc identifier :

$$\text{Cov}(X_t; X_{t-h}) = 0 \text{ ou encore } \rho_h = 0 ; \forall h$$

Soient les hypothèses :

$$\begin{cases} H_0: \rho_h = 0 \quad \forall h \\ H_1: \text{il existe au moins } \rho_i \text{ significativement différent de zéro} \end{cases}$$

Pour effectuer ce test, on a recours la statique Q qui est donnée par :

$$Q = n \sum_{h=1}^k \hat{\rho}_h^2$$

K : Nombre de retards.

$\hat{\rho}_h^2$: Le coefficient d'auto corrélation d'ordre h.

n : Nombre d'observations.

Q : suit asymptotiquement une loi de (χ^2) à (K) degrés de liberté.

Si la statistique Q est supérieure au χ^2 lu dans la table au seuil $(1 - \alpha)$ et k degrés de liberté donc les résidus ne forment pas un bruit blanc.

➤ Test de Box-Ljung

Ce test est basé sur l'hypothèse nulle : les auto-corrélations au pas K ne sont pas nulles, tel que $k = n/4$.

On teste :

$$\begin{array}{l} H_0: \rho_1 = \rho_2 = \rho_3 = \dots = \rho_k = 0 \\ \text{contre} \\ H_1: \rho_j ; j = 1, \dots, k \text{ tq: } \rho_j \neq 0 \end{array}$$

Box et Perce propose la statistique $Q = n(n + 2) \sum_{h=1}^k \frac{\hat{\rho}_h^2}{(n-h)}$

Si $Q < \chi^2_{(1-\alpha), (k-p-q)}$ on accepte au seuil de $\alpha = 5\%$, l'hypothèse H_0 : les résidus forment un bruit blanc.

➤ Test de Durbin-Watson

Ce test permet de tester l'auto-corrélation d'ordre 1.

$$\left\{ \begin{array}{l} H_0: \rho = 0 \text{ «Absence d'auto-corrélation à l'ordre 1 des résidus»} \\ H_1: \rho \neq 0 \text{ «Présence d'auto-corrélation à l'ordre 1 des résidus»} \end{array} \right.$$

La statistique de Durbin Watson est obtenue après estimation du modèle par les méthodes des moindres carrés ordinaires, notée DW, est donnée par :

$$DW = \frac{\sum_{t=2}^n (\varepsilon_t - \varepsilon_{t-1})^2}{\sum_{t=1}^n \varepsilon_t^2}$$

Avec (ε_t) : sont les résidus de l'estimation du modèle. Cette statistique varie entre 0 et 4 et nous avons $DW=2$ lorsque $\hat{\rho}=0$ ($\hat{\rho}$ Est l'estimateur de ρ) d'où :

- $DW = 0$, il existe une auto-corrélation positive.
- $DW = 4$, il existe une auto-corrélation négative.
- $DW \approx 2$, indique l'absence d'auto-corrélation.

➤ Test de normalité

Pour vérifier si le processus des résidus $\{\varepsilon_t, t \in \mathbb{Z}\}$ forme un bruit blanc gaussien, plusieurs tests peuvent être utilisés, mais le test le plus fréquent qui permet de vérifier la normalité d'une distribution statistique est celui de Jaque et Bera (1984).

i. Les tests de Skewness et de Kurtosis

Sous l'hypothèse d'une loi normale et si le nombre d'observation assez grand (pour $n > 50$), on peut écrire :

$$\beta_1 \rightarrow N\left(0, \sqrt{\frac{6}{n}}\right), \quad \beta_2 \rightarrow N\left(3, \sqrt{\frac{24}{n}}\right)$$

Soit les statistiques : $\vartheta_1 = \left| \frac{\beta_1}{\sqrt{\frac{6}{n}}} \right|$ et $\vartheta_2 = \left| \frac{\beta_2 - 3}{\sqrt{\frac{24}{n}}} \right|$ qui sont comparées à 1.96.

Soient les hypothèses :

$$\begin{cases} H_0: \vartheta_1 = 0 \text{ (symétrique) et } \vartheta_2 = 0 \text{ (aplatissement normale).} \\ H_0: \vartheta_1 \neq 0 \text{ et } \vartheta_2 \neq 0. \end{cases}$$

On acceptera pour un risque $\alpha = 5\%$ l'hypothèse H_0 si $|\vartheta_1| < 1.96$ et $|\vartheta_2| < 1.96$.

ii. Test de Jarque-Bera

Ce test regroupe en lui seul les tests de Skewness et Kurtosis. On définit la statistique par :

$$JB = \frac{n}{6} Sk + \frac{n}{24} (ku - 3)^2$$

Pour un seuil $\alpha = 5\%$, on teste les hypothèses:

$$\begin{cases} H_0 : \text{Les résidus sont gaussiens.} \\ H_1 : \text{Les résidus ne sont pas gaussiens.} \end{cases}$$

On rejette H_0 si $JB > \chi_2^2$.

➤ Test d'hétéroscédasticité (l'existence d'effet ARCH)

Pour ce faire, on utilise les corrélogrammes des carrés des résidus, si un ou plusieurs termes

sont significativement différents de zéro, on déduit qu'il y a effet **ARCH** qui est détecté

également par la statistique de Box et Ljung, et qui est confirmé par la statistique du Multiplicateur de Lagrange $LM = nR^2$ avec n le nombre d'observations servant au calcul

de la régression et R^2 est le coefficient de détermination.

Soit une spécification de type ARCH pour les erreurs (ε_t) tel que :

$$\begin{cases} \varepsilon_t = u_t h_t \\ h_t^2 = \alpha_0 + \sum_{i=1}^p \alpha_i \varepsilon_{t-i}^2 \end{cases} \quad \text{Avec : } u_t \sim N(0,1)$$

Soit l'hypothèse

$$\begin{cases} H_0: \alpha_1 = \alpha_2 = \dots = \alpha_p = 0 \\ H_1: (\text{les } \alpha_i \text{ tous non nulles}). \end{cases}$$

- Si $LM < \chi^2(p)$ on accepte H_0 . La variance de l'erreur est constante $h_t^2 = \alpha_0$
- Dans le cas contraire $LM > \chi^2(p)$, p degrés de liberté compris entre 1 et 3, on

rejette H_0 et le processus est justifiable d'un modèle ARCH(p).

- Si $p > 3$ le modèle sera justifié d'un modèle de type GARCH.

➤ Test de comparaison du modèle

Une fois le modèle identifié, on se demande parfois si un autre modèle n'aurait pas été plus préférable, il s'agit donc de tester le modèle retenu ARMA(p, q) contre un autre modèle en augmentant l'ordre de (p ou q), en ajoutant ou en éliminant une constante, puis choisir le modèle le plus qualifié à l'aide des critères : critère d'information d'Akaike(AIC), critère de Schwartz (SC), le coefficient de détermination R^2 .

4.3.4 Préviation

A un instant (t), le processus possède une valeur X_t que l'on désire prévoir à l'instant $t + 1$.

Sachant que le processus est un **ARMA (p, q)**, on note cette prévision $\hat{X}_t(l)$, l est dit horizon de prévision.

$$X_t(h) = E(X_{t+h}/y_t) \quad h = 1; 2; \dots; 12$$

Où: $y_t = \{X_1, \dots, X_t\}$ les valeurs passés de la réalisation.

4.4 Application de la méthode Box & Jenkins

4.4.1 Étude de la série BUDGET

4.4.1.1 Représentation graphique de la série

La première étape de l'étude d'une série chronologique est la représentation graphique. Cette visualisation donne des indications très précieuses pour choisir un modèle. Pour illustrer cette première phase de modélisation, nous examinons le graphique de la série nommée BUDGET.

➤ Représentation graphique

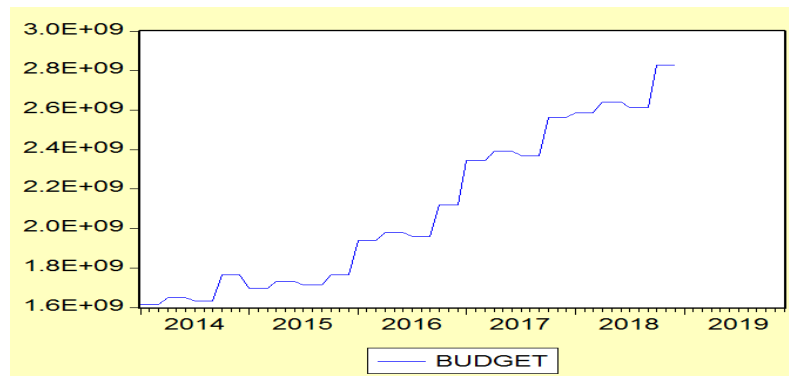


Figure Erreur ! Il n'y a pas de texte répondant à ce style dans ce document.-1: graphe des données de la série BUDGET

La représentation graphique de la série caractérisée par une tendance haussière, et un phénomène périodique caractérisé par des mouvements ascendants et descendants, ces mouvements périodiques sont dus au rythme des saisons. Dans ce cas nous dirons que la série BUDGET n'est pas stationnaire,

L'examen graphique ne permet pas toujours de déterminer avec certitude l'existence d'une tendance et la saisonnalité. Afin d'éloigner l'incertitude, nous utilisons le test de Fisher et le test de DICKEY-FULLER.

➤ Vérification du facteur saisonnier et tendanciel

- Test d'influence du facteur lignes (mois)

$$\begin{cases} H_0 : \text{l'effet saisonnier non significatif} \\ H_1 : \text{l'effet saisonnier significatif} \end{cases}$$
- Test d'influence du facteur colonne (années)

$$\begin{cases} H_0 : \text{l'effet de la tendance non significatif} \\ H_1 : \text{l'effet de la tendance significatif} \end{cases}$$

Ce test, nous permettons de montrer si la série est influencée par un effet saisonnier et un effet tendanciel ou non, pour un seuil de 5%

Les résultats du test de Fisher sont regroupés dans le tableau suivant :

| ANALYSE DE VARIANCE | | | | | | |
|-----------------------|------------------|------------------|--------------------|-------------|--------------|------------------------|
| Source des variations | Somme des carrés | Degré de liberté | Moyenne des carrés | F | Probabilité | Valeur critique pour F |
| Lignes | 2,67481E+17 | 11 | 2,43165E+16 | 29,47451801 | 1,033107E-16 | 2,014046013 |
| Colonnes | 9,09018E+18 | 4 | 2,27255E+18 | 2754,600507 | 3,59913E-52 | 2,583667427 |

Tableau 1: Fisher de la série saisonnière BUDGET

Tableau d’ANOVA série (BUDGET)

Au seuil 5% les résultats de l’ANOVA obtenus par Excel sont les suivants :

- ✓ Résultat du test d’influence de facteur mois (Saisonnalité) :

$$F_{cal} = 29.4745 > F^{0.05}(11; 59) = 2.0140$$

On rejette l’hypothèse H_0 , donc la série affectée d’une saisonnalité.

- ✓ Résultat du test d’influence de facteur année (Tendance) :

$$F_{cal} = 2754.600507 > F^{0.05}(9; 59) = 2.5836$$

Donc, nous rejetons l’hypothèse nulle. La série est donc affectée d’une tendance.

➤ **Dessaisonnalisation de la série (BUDGET)**

Une dessaisonnalisation de la série (BUDGET) s’impose pour la suite de notre étude. Notons (BUDGETTSA) la série désaisonnalisée.

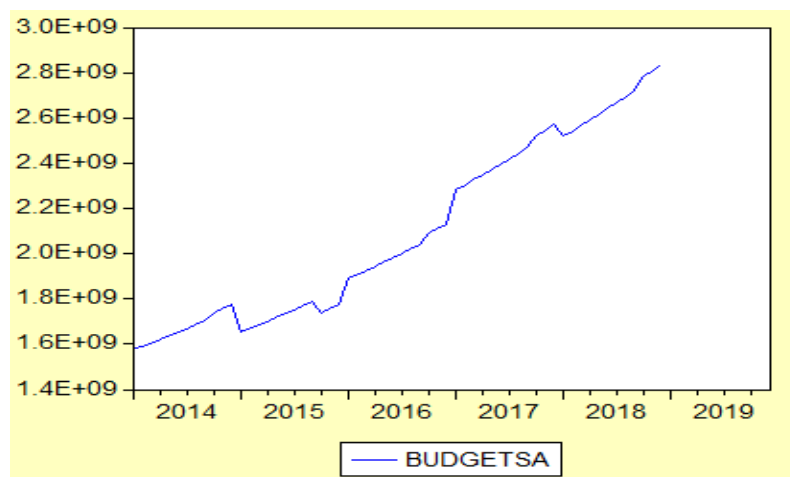


Figure Erreur ! Il n'y a pas de texte répondant à ce style dans ce document.-2:graphe série BUDGET désaisonnalisée

La représentation graphique de la série (BUDGETTSA) monte la neutralité de l’effet saisonnier.

➤ **Vérification de l'élimination du facteur saisonnier**

On applique le test ANOVA sur la série (BUDGET) pour voir si l'effet saisonnier a été éliminé

| ANALYSE DE VARIANCE | | | | | | |
|-----------------------|------------------|------------------|--------------------|-------------|-------------|------------------------|
| Source des variations | Somme des carrés | Degré de liberté | Moyenne des carrés | F | Probabilité | Valeur critique pour F |
| Lignes | 7,39004E+17 | 11 | 6,71822E+16 | 1,761150792 | 0,091126236 | 2,014046013 |
| Colonnes | 6,80482E+18 | 4 | 1,7012E+18 | 44,59630693 | 6,16938E-15 | 2,583667427 |

Tableau 2 : Fisher de la série désaisonnalisée BUDGET

Au seuil 5% les résultats de l'ANOVA obtenus par Excel sont les suivants :

✓ Résultat du test d'influence de facteur mois (Saisonnalité) :

On a $F_c = 1.76115 < F^{0.05}(11,59) = 2.0140$

On accepte l'hypothèse H_0 : la série n'est pas saisonnière. A travers le tableau et le graphe de la série désaisonné qui sont affichés, nous remarquons que l'effet saisonnier a disparu, mais on ne peut dire que le processus qui génère cette série est stationnaire qu'après avoir fait un recours au test de racine unitaire (Dickey-Fuller) qui permet d'affirmer ou d'infirmer notre hypothèse.

4.4.1.2 Etude de la stationnarité de la série dessaisonnée (BUDGETSA)

Le principe général du test de Dickey Fuller consiste à tester l'hypothèse nulle de la présence d'une racine unitaire. Ce qui met en évidence le caractère stationnaire ou non d'une série par la détermination d'une tendance de type déterministe (TS) ou aléatoire (DS).

Afin d'étudier la stationnarité de cette série, nous allons y procéder par les différents tests suivants :

➤ **Test de la racine unitaire (Dickey-Fuller) sur la série (BUDGETSA)**

Pour vérifier si oui ou non la série possède une tendance et qu'elle nature est cette tendance, on applique la méthode de test de racine unitaire (test de Dickey-Fuller). On commence par le modèle [3] : test sur la tendance.

Estimation du modèle [3]

Dans ce modèle nous testons les deux hypothèses :

$$\begin{cases} H_0 : \text{Absence de la tendance.} \\ H_1 : \text{L'existence d'une tendance.} \end{cases}$$

| ADF Test Statistic | -1.699482 | 1% Critical Value* | -4.1190 | |
|--------------------|-------------|-----------------------|-------------|--------|
| | | 5% Critical Value | -3.4862 | |
| | | 10% Critical Value | -3.1711 | |
| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
| BUDGETSA(-1) | -0.080328 | 0.047266 | -1.699482 | 0.0948 |
| C | 1.24E+08 | 67470727 | 1.831855 | 0.0723 |
| @TREND(2014:01) | 2172878. | 1077739. | 2.016145 | 0.0486 |
| R-squared | 0.089012 | Mean dependent var | 21439496 | |
| Adjusted R-squared | 0.056477 | S.D. dependent var | 32486600 | |
| S.E. of regression | 31555900 | Akaike info criterion | 37.42193 | |
| Sum squared resid | 5.58E+16 | Schwarz criterion | 37.52757 | |
| Log likelihood | -1100.947 | F-statistic | 2.735860 | |
| Durbin-Watson stat | 1.963694 | Prob(F-statistic) | 0.073511 | |

D'après le modèle [3] on a les résultats suivants :

- La statistique $t = -1.6994$ est supérieure aux valeurs critiques relatives aux seuils de 1% , 5% et 10% donc la série admet une racine unitaire
- La tendance n'est pas significativement différent de zéro car la valeur estimée de la tendance $=2.016$ est inférieure à la valeur critique $=2.79$ (donnée par la table de Dickey-Fuller) au seuil de 5%.

On passe alors au modèle [2]:test sur la constante.

Estimation du modèle [2]

Dans ce modèle nous testons les hypothèses suivantes :

- H_0 : Absence de la constante.
- H_1 : L'existence de la constante.

| ADF Test Statistic | 1.155462 | 1% Critical Value* | -3.5437 | |
|--------------------|-------------|-----------------------|-------------|--------|
| | | 5% Critical Value | -2.9109 | |
| | | 10% Critical Value | -2.5928 | |
| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
| BUDGETSA(-1) | 0.012549 | 0.010861 | 1.155462 | 0.2527 |
| C | -4703926. | 23015610. | -0.204380 | 0.8388 |
| R-squared | 0.022887 | Mean dependent var | 21439496 | |
| Adjusted R-squared | 0.005744 | S.D. dependent var | 32486600 | |
| S.E. of regression | 32393159 | Akaike info criterion | 37.45810 | |
| Sum squared resid | 5.98E+16 | Schwarz criterion | 37.52853 | |
| Log likelihood | -1103.014 | F-statistic | 1.335093 | |
| Durbin-Watson stat | 2.008531 | Prob(F-statistic) | 0.252722 | |

- La statistique (1.155) est supérieure aux valeurs critiques relatives (-3.5437 ; -2.9109 ; -2.5928) au seuils d 1% 5% 10% (donnée par la table de Dickey-Fuller Augmenté) donc notre série admet une racine unitaire.
- la statistique relative à la constante $t = -0.2043$ est inférieure à valeur critique (2.54) donc la constante n'est pas significativement différent de zéro on passe au modèle [1] sans constante et sans tendance.

Estimation du modèle [1]

Dans ce modèle nous testons les deux hypothèses suivantes :

- H_0 : Le processus est non stationnaire.
- H_1 : Le processus est stationnaire.

| ADF Test Statistic | 5.253008 | 1% Critical Value* | -2.6019 | |
|--------------------|-------------|-----------------------|-------------|--------|
| | | 5% Critical Value | -1.9460 | |
| | | 10% Critical Value | -1.6187 | |
| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
| BUDGETSA(-1) | 0.010367 | 0.001974 | 5.253008 | 0.0000 |
| R-squared | 0.022171 | Mean dependent var | 21439496 | |
| Adjusted R-squared | 0.022171 | S.D. dependent var | 32486600 | |
| S.E. of regression | 32124458 | Akaike info criterion | 37.42494 | |
| Sum squared resid | 5.99E+16 | Schwarz criterion | 37.46015 | |
| Log likelihood | -1103.036 | Durbin-Watson stat | 2.002681 | |

Comme la valeur estimée de la statistique qui est égale à (5.2530) est supérieure aux différentes valeurs critiques (-2.6019); (-1.9460) ; (-1.6187) relatives aux seuils 1%, 5% et 10% respectivement, donc la série possède de racine unitaire. Par conséquent, le processus qui génère la série (BUDGETTESA) n'est pas stationnaire de type DS.

Pour stationnariser notre série en utilisant le filtre aux différences première on la note (DBUDGETSA).

➤ **Test de la racine unitaire (Dickey-Fuller Augmenté) sur la série désaisonnalisée différenciée (DBUDGETSA)**

On commence par le modèle [3] : test sur la tendance

| ADF Test Statistic | -7.505840 | 1% Critical Value* | -4.1219 | |
|--------------------|-------------|-----------------------|-------------|--------|
| | | 5% Critical Value | -3.4875 | |
| | | 10% Critical Value | -3.1718 | |
| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
| D(BUDGETSA(-1)) | -1.012034 | 0.134833 | -7.505840 | 0.0000 |
| C | 9555241. | 8994309. | 1.062365 | 0.2927 |
| @TREND(2014:01) | 401599.2 | 261574.6 | 1.535314 | 0.1304 |
| R-squared | 0.506007 | Mean dependent var | 211896.8 | |
| Adjusted R-squared | 0.488044 | S.D. dependent var | 45621202 | |
| S.E. of regression | 32642480 | Akaike info criterion | 37.49047 | |
| Sum squared resid | 5.86E+16 | Schwarz criterion | 37.59704 | |
| Log likelihood | -1084.224 | F-statistic | 28.16882 | |
| Durbin-Watson stat | 2.000287 | Prob(F-statistic) | 0.000000 | |

- La statistique $t = -7.5058$ est inférieure aux valeurs critiques relatives à 1% 5% 10% donc la série ne possède pas une racine unitaire
- la t-statistique associée à la tendance qui égale à 1.53 est inférieure à la valeur critique par Dickey – Fuller au seuil de 5% avec une probabilité supérieure à 5% donc le coefficient de la tendance n'est pas significativement différent de zéro. On passe alors au modèle [2]:test sur la constante.

| ADF Test Statistic | -7.259694 | 1% Critical Value* | -3.5457 | |
|--------------------|-------------|-----------------------|-------------|--------|
| | | 5% Critical Value | -2.9118 | |
| | | 10% Critical Value | -2.5932 | |
| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
| D(BUDGETSA(-1)) | -0.969650 | 0.133566 | -7.259694 | 0.0000 |
| C | 20899730 | 5190103. | 4.026843 | 0.0002 |
| R-squared | 0.484836 | Mean dependent var | 211896.8 | |
| Adjusted R-squared | 0.475636 | S.D. dependent var | 45621202 | |
| S.E. of regression | 33035666 | Akaike info criterion | 37.49795 | |
| Sum squared resid | 6.11E+16 | Schwarz criterion | 37.56900 | |
| Log likelihood | -1085.440 | F-statistic | 52.70316 | |
| Durbin-Watson stat | 2.001863 | Prob(F-statistic) | 0.000000 | |

- Dans le modèle [2], la t-statistique associée à la constante qui égale 4.026 est supérieure à la valeur critique au seuil de 5% donc la constante est significativement différente de zéro

On conclure que notre série (DBUDGETSA) est stationnaire à la première différence.

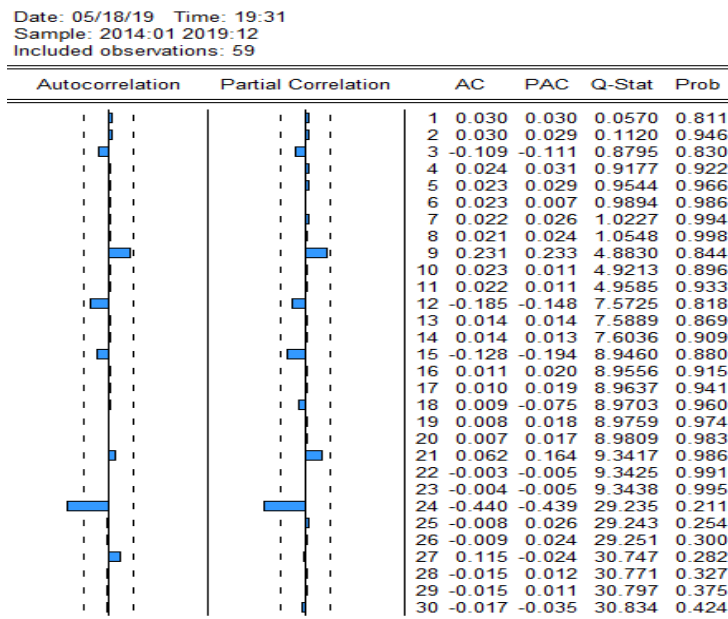
Nous passons maintenant à l'étape du choix du modèle adéquat

4.4.1.3 Identification du modèle

➤ Corrélogramme de la série (DBUDGETSA)

Cette phase consiste à déterminer les modèles adéquats dans la famille ARIMA. Elle est fondée sur l'étude des corrélogrammes simples et partiels de la série étudiée (éventuellement stationnarisée).

Pour connaître les ordres du modèle, on regarde le corrélogramme de la série stationnaire DBUDGETSA. Le corrélogramme simple permet d'identifier un modèle MA(q), tandis que le corrélogramme partiel permet d'identifier un modèle AR(p).



L'analyse du correlogramme simple et partiel de la série stationnaire (DBUDGETTSA) montre qu'au retard ($k=24$) le terme est significativement différent de 0. Par conséquent nous avons trois modèles candidats ARMA(24,0) ARMA(0,24) ; ARMA(24,24). Et pour pouvoir choisir un bon modèle parmi ceux présente, on construit un tableau de comparaison de leurs critères d'information et nous retenons le modèle qui minimise ces critères.

➤ **Choix du modèle**

| Critères | ARMA(24,0) | ARMA(0,24) |
|----------------------|-----------------------|-----------------------|
| R² | 0.76 | 0.59 |
| SC | 36.66 | 36.58 |
| AIC | 36.76 | 36.65 |
| SCR | 1.30 ^E +16 | 2.50 ^E +16 |

Tableau 3: choix de modèle optimal

D'après le tableau ci-dessus, le modèle ARMA(0,24) est le meilleur car il possède les meilleurs critères d'informations (les critères d'Akaike et de Schwarz sont minimum, et le R² soit maximum).

4.4.1.4 Estimation des paramètres du modèle optimal

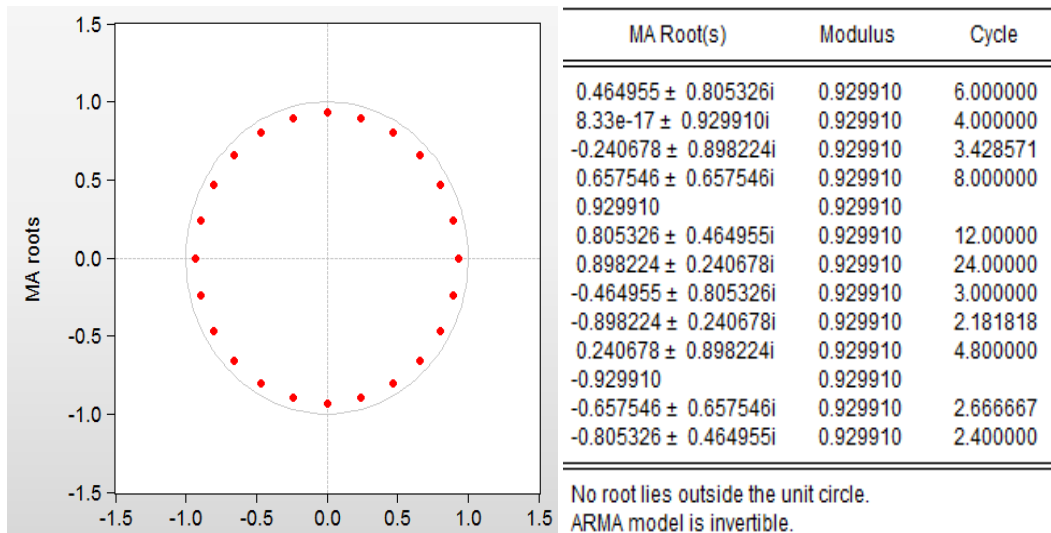
| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|--------------------|-------------|-----------------------|-------------|----------|
| C | 21439595 | 3683681. | 5.820155 | 0.0000 |
| MA(24) | -0.863240 | 0.057058 | -15.12912 | 0.0000 |
| R-squared | 0.590978 | Mean dependent var | | 21439496 |
| Adjusted R-squared | 0.583802 | S.D. dependent var | | 32486600 |
| S.E. of regression | 20958211 | Akaike info criterion | | 36.58727 |
| Sum squared resid | 2.50E+16 | Schwarz criterion | | 36.65769 |
| Log likelihood | -1077.324 | F-statistic | | 82.35678 |
| Durbin-Watson stat | 1.923711 | Prob(F-statistic) | | 0.000000 |

4.4.1.5 validation du modèle

➤ **Test sur les paramètres du modèle**

D'après le tableau suivant et le tableau d'estimation nous constatons que : Les coefficients des paramètres du modèle sont significativement différent de zéro car les t-statistiques ($|-15.129|, |5.82|$) du modèle est supérieure à la valeur critique (de Student =1.96) au seuil de 5% ce qui est confirmé par la probabilité de coefficient qui est inférieure à 0,05.

Les inverses des racines sont distincts et inférieurs en module à 1 .Par conséquent, les conditions d'invisibilité et de stationnarité sont vérifiées.



Test sur les résidus

➤ **Corrélogramme des résidus**

D’après l’analyse du corrélogramme, on remarque que tous les termes sont à l’extérieur de l’intervalle de confiance et les probabilités correspondantes sont supérieure à 0.05, donc on estime que les résidus forment un bruit blanc.

| Autocorrelation | Partial Correlation | AC | PAC | Q-Stat | Prob |
|-----------------|---------------------|-----------|--------|--------|-------|
| | | 1 -0.053 | -0.053 | 0.1751 | |
| | | 2 -0.060 | -0.063 | 0.4000 | 0.527 |
| | | 3 0.088 | 0.082 | 0.8999 | 0.638 |
| | | 4 -0.052 | -0.047 | 1.0767 | 0.783 |
| | | 5 -0.057 | -0.053 | 1.2942 | 0.862 |
| | | 6 0.110 | 0.094 | 2.1173 | 0.833 |
| | | 7 -0.047 | -0.036 | 2.2693 | 0.893 |
| | | 8 -0.052 | -0.039 | 2.4582 | 0.930 |
| | | 9 0.071 | 0.043 | 2.8178 | 0.945 |
| | | 10 -0.060 | -0.048 | 3.0781 | 0.961 |
| | | 11 -0.064 | -0.051 | 3.3830 | 0.971 |
| | | 12 0.028 | -0.008 | 3.4412 | 0.984 |
| | | 13 -0.013 | -0.002 | 3.4539 | 0.991 |
| | | 14 -0.017 | -0.001 | 3.4768 | 0.996 |
| | | 15 0.013 | -0.015 | 3.4897 | 0.998 |
| | | 16 -0.004 | 0.003 | 3.4914 | 0.999 |
| | | 17 -0.007 | 0.005 | 3.4960 | 1.000 |
| | | 18 0.014 | -0.001 | 3.5131 | 1.000 |
| | | 19 -0.017 | -0.016 | 3.5402 | 1.000 |
| | | 20 -0.021 | -0.017 | 3.5796 | 1.000 |
| | | 21 0.081 | 0.072 | 4.2057 | 1.000 |
| | | 22 -0.011 | -0.007 | 4.2186 | 1.000 |
| | | 23 -0.014 | -0.006 | 4.2382 | 1.000 |
| | | 24 -0.168 | -0.192 | 7.1412 | 0.999 |

➤ **Test de Ljung-Box (test d’autocorrélation)**

Le corrélogramme des résidus montre bien que les résidus forment un bruit blanc, ce qui est confirmé par la statistique de Ljung-Box .Ce test est basé sur la comparaison entre la dernière valeur de Q-stat calculée (24= sur le corrélogramme) et la valeur tabulée de Khi-deux de $(K - p - q)$: Degré de liberté

- K : le nombre de retard.
- p : l'ordre d'autorégressive.
- q : l'ordre de Moyen mobile.

Si la Q-statistique : $Q - \text{stat} < \chi^2_{(K-p-q)}$. On accepte l'existence de la nullité d'auto-corrélation des résidus, alors les résidus constituent un bruit blanc ce qui nous donne un modèle valide

$Q - \text{stat}(24) = 7.1412 < \chi^2_{(23)} = 35.17$ au seuil de 5%, D'où la validité du modèle

➤ **Test du Durbin Watson**

On teste les hypothèses:

- $H_0 : \rho = 0$ «les résidus non corrélés »
- $H_1 : \rho \neq 0$ «les résidus corrélés»

Si DW est proche de 2 alors on accepte H_0 . La statistique $DW=1.92$, on accepte l'hypothèse de non corrélation des résidus.

➤ **Test de normalité des résidus**

Les tests sont effectués à partir des valeurs empiriques des coefficients de Skewness, Kurtosis et la statistique de Jarque-Berra données par histogramme des résidus de la série(DBUDGETSA).

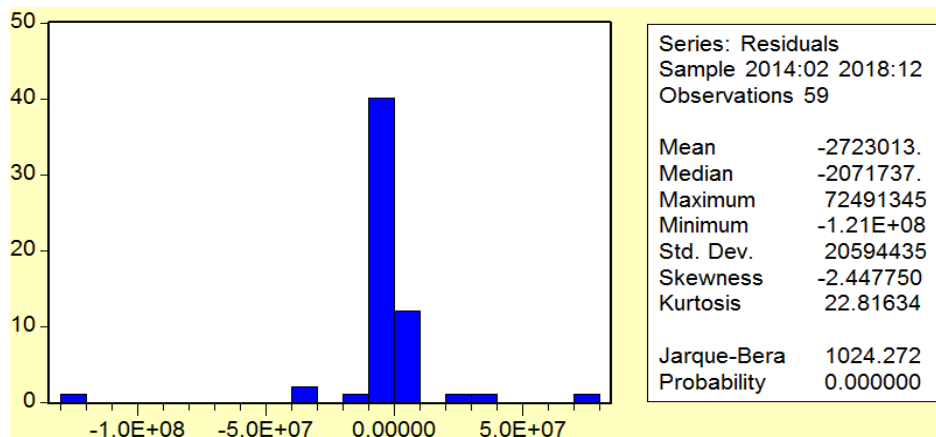


Figure Erreur ! Il n'y a pas de texte répondant à ce style dans ce document.-3: la normalité des résidus

- **Test Skewness:** $Sk^{1/2} = -2.44 \Rightarrow \vartheta_1 = \frac{|-2.44-0|}{\left(\frac{6}{59}\right)^{1/2}} = 7.65 > 1.96$ (la distribution est dissymétrie)
- **Test Kurtosis:** $ku = 22.81 \Rightarrow \vartheta_2 = \frac{|22.81-3|}{\left(\frac{24}{59}\right)^{1/2}} = 31.06 > 1.96$ (la distribution est leptokurtique)
- **Test Jarque-Bera:** $JB = 1024.272 > \chi^2_2 = 5.99$

Donc on rejette l'hypothèse H_0 , donc les résidus non gaussiens.

➤ **Test d'heteroscédasticité (l'existence d'effet ARCH)**

| F-statistic | 0.105244 | Probability | 0.746834 | |
|---|-------------|-----------------------|-------------|--------|
| Obs*R-squared | 0.108798 | Probability | 0.741516 | |
| Test Equation: | | | | |
| Dependent Variable: RESID^2 | | | | |
| Method: Least Squares | | | | |
| Date: 06/03/19 Time: 00:49 | | | | |
| Sample(adjusted): 2014:03 2018:12 | | | | |
| Included observations: 58 after adjusting endpoints | | | | |
| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
| C | 4.50E+14 | 2.76E+14 | 1.629647 | 0.1088 |
| RESID^2(-1) | -0.043311 | 0.133504 | -0.324414 | 0.7468 |
| R-squared | 0.001876 | Mean dependent var | 4.32E+14 | |
| Adjusted R-squared | -0.015948 | S.D. dependent var | 2.04E+15 | |
| S.E. of regression | 2.06E+15 | Akaike info criterion | 73.39309 | |
| Sum squared resid | 2.37E+32 | Schwarz criterion | 73.46414 | |
| Log likelihood | -2126.400 | F-statistic | 0.105244 | |
| Durbin-Watson stat | 2.003935 | Prob(F-statistic) | 0.746834 | |

A partir de la figure on remarque que la probabilité de RESID2 est supérieur a 0.05 donc on accepte l'hypothèse nulle d'hémoscédasticité il n'existe pas un effet ARCH.

Conclusion: D'après les résultats on peut valider le modèle de la série (BUDGETSA) s'écrit sous la forme :

$$(1 - B)Budgetsa_t = 21439594.73 + (1 + 0.863B^{24})\varepsilon_t$$

$$Budgetsa_t = 21439594.73 + budgetsa_t + 0.863 * \varepsilon_{t-24} + \varepsilon_t$$

Ou (ε_t) est un processus bruit blanc.

4.4.1.6 Prévision

Rappelons que la série sous étude "BUDGET" comprend 60 observation de Janvier 2014 à décembre 2018, donc l'instant t est décembre 2018.

On souhaitera prédite les 12 mois de l'année 2019, soit un horizon $h = 12$:

On remplaçant t par $t + h$ dans l'équation précédente on obtient :

$$Budgetsa_{t+h} = 21439594.73 + budgetsa_{t+h-1} + 0.863 * \varepsilon_{t+h-24} + \varepsilon_{t+h}$$

Avec $h = 1, \dots, 12$. On obtient les prévisions suivantes :

| Mois de l'année 2019 | Prévision (DA) |
|----------------------|----------------|
| Janvier | 2 960 160 984 |
| Février | 2 954 043 464 |
| Mars | 2 948 141 502 |
| Avril | 3 002 091 555 |
| Mai | 2 994 584 051 |
| Juin | 2 987 459 763 |
| Juillet | 2 949 533 497 |
| Aout | 2 941 040 262 |
| Septembre | 2 932 833 539 |
| Octobre | 3 125 532 070 |
| Novembre | 3 116 779 625 |
| Décembre | 3 108 316 996 |

Le graphe des prévisions est donné comme suit :

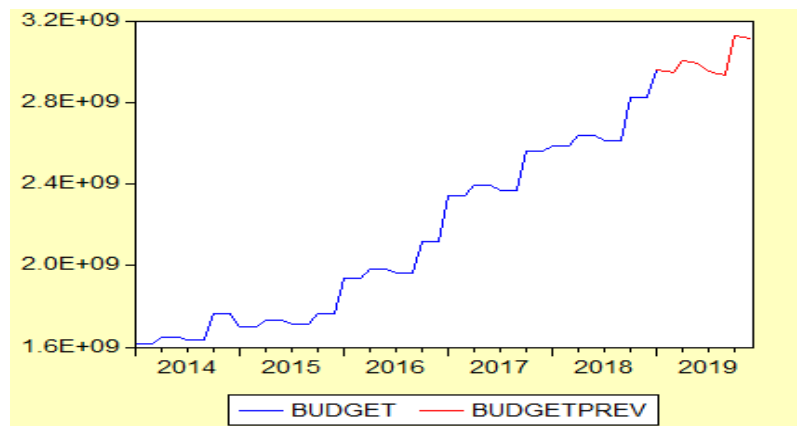


Figure Erreur ! Il n'y a pas de texte répondant à ce style dans ce document.-4:graphe des prévisions de la série BUDGET

4.4.2 Étude de la série RETRAITE

4.4.2.1 Représentation graphique de la série

Pour illustrer cette première phase de modélisation, nous examinons le graphique de la série nommée RETRAITE .

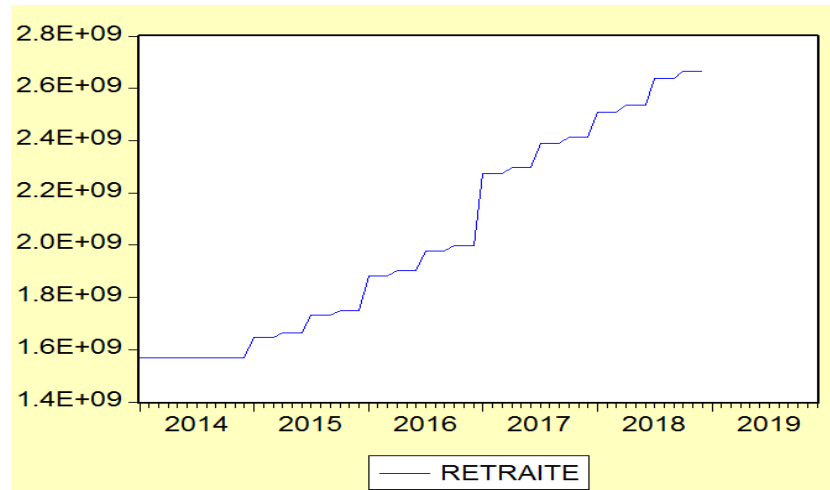


Figure Erreur ! Il n'y a pas de texte répondant à ce style dans ce document.-5: graphe des données de la série RETRAITE

De ce graphique, nous observons que la série évolue de manière croissante dans le temps. Ce qui nous permet de tirer quelques conclusions intéressantes au vu de l'allure de la courbe ci-dessus :

- Cette évolution semble être caractérisée par une tendance haussière, dans ce cas nous dirons que la série n'est pas stationnaire
- Un phénomène périodique caractérisé par des mouvements ascendants et descendants, ces mouvements périodiques sont dus au rythme des saisons.

Dans ce cas nous dirons que la série RETRAITE n'est pas stationnaire,

➤ Vérification du facteur saisonnier et tendanciel

- Test d'influence du facteur lignes (mois)
 - $\left\{ \begin{array}{l} H_0 : \text{l'effet saisonnier non significatif} \\ H_1 : \text{l'effet saisonnier significatif} \end{array} \right.$
- Test d'influence du facteur colonne (années)
 - $\left\{ \begin{array}{l} H_0 : \text{l'effet de la tendance non significatif} \\ H_1 : \text{l'effet de la tendance significatif} \end{array} \right.$

Ce test, nous permettons de montrer si la série est influencée par un effet saisonnier et un effet tendanciel ou non, pour un seuil de 5%

Les résultats du test de Fisher sont regroupés dans le tableau suivant :

| ANALYSE DE VARIANCE | | | | | | |
|-----------------------|------------------|------------------|--------------------|-------------|-------------|------------------------|
| Source des variations | Somme des carrés | Degré de liberté | Moyenne des carrés | F | Probabilité | Valeur critique pour F |
| Lignes | 1,14529E+17 | 11 | 1,04118E+16 | 14,1600504 | 4,29045E-11 | 2,014046013 |
| Colonnes | 8,88586E+18 | 4 | 2,22146E+18 | 3021,201099 | 4,7009E-53 | 2,583667427 |

Tableau 4: tableau d'ANOVA série(RETRAITE)

- ✓ Résultat du test d'influence de facteur mois (saisonnalité):

$$F_{\text{cal}} = 14.160 > F^{0.05}(11; 59) = 2.0140;$$

On rejette l'hypothèse H_0 , donc la série affectée d'une saisonnalité.

- ✓ Résultat du test d'influence de facteur année (Tendance):

$$F_{\text{cal}} = 3021.201 > F^{0.05}(4; 59) = 2.5836$$

Donc, nous rejetons l'hypothèse nulle. La série est donc affectée d'une tendance.

- **Dessaisonnalisation de la série (RETRAITE)**

Une dessaisonnalisation de la série (RETRAITE) s'impose pour la suite de notre étude. Notons (RETRAITESA) la série désaisonnalisée

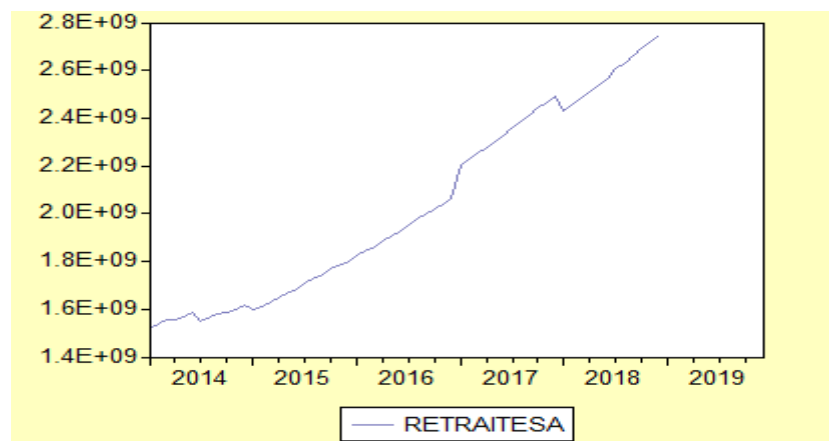


Figure Erreur ! Il n'y a pas de texte répondant à ce style dans ce document.-6: graphe des données de la série(RETRAITESA)

La représentation graphique de la série nous montre l'absence de l'effet saisonnier.

➤ **Vérification de l'élimination du facteur saisonnier**

On applique le test ANOVA sur la série (RETRAITSA) pour voir si l'effet saisonnier a été éliminé.

| ANALYSE DE VARIANCE | | | | | | |
|-----------------------|------------------|------------------|--------------------|-------------|-------------|------------------------|
| Source des variations | Somme des carrés | Degré de liberté | Moyenne des carrés | F | Probabilité | Valeur critique pour F |
| Lignes | 4,70361E+17 | 11 | 4,27601E+16 | 0,8506877 | 0,592500718 | 2,014046013 |
| Colonnes | 1,0252E+19 | 4 | 2,563E+18 | 50,98946665 | 5,76222E-16 | 2,583667427 |

Figure Erreur ! Il n'y a pas de texte répondant à ce style dans ce document.-7:Fisher de série désaisonnalisée (RETRAITESA)

✓ Résultat du test d'influence de facteur mois (saisonnalité) au seuil de 5:

On a $F_c = 0.8506 < F^{0.05}(11,59) = 2.0140$

On accepte l'hypothèse H_0 : la série n'est pas saisonnière. A travers le tableau et le graphe de la série désaisonnalisé qui sont affichés, nous remarquons que l'effet saisonnier a disparu, mais on ne peut dire que le processus qui génère cette série est stationnaire qu'après avoir fait un recours au test de racine unitaire (Dickey-Fuller) qui permet d'affirmer ou d'infirmer notre hypothèse.

4.4.3 Etude de la stationnarité de la série dessaisonalisé (RETRAITESA)

A fin d'étudier la stationnarité de cette série, nous allons y procéder par les différents tests suivants :

➤ **Test de racine unitaire (Dickey-Fuller) sur la série (RETRAITESA)**

Estimation du Modèle [3]

| | | t-Statistic | Prob.* |
|--|-----------|-------------|--------|
| Augmented Dickey-Fuller test statistic | | -2.077310 | 0.5473 |
| Test critical values: | 1% level | -4.121303 | |
| | 5% level | -3.487845 | |
| | 10% level | -3.172314 | |

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|-------------------|-------------|------------|-------------|--------|
| RETRAITSA(-1) | -0.081460 | 0.039214 | -2.077310 | 0.0424 |
| C | 1.20E+08 | 53215020 | 2.248830 | 0.0285 |
| @TREND("2014M01") | 2178266. | 885468.0 | 2.460016 | 0.0170 |

| | | | |
|--------------------|-----------|-----------------------|----------|
| R-squared | 0.146621 | Mean dependent var | 20804013 |
| Adjusted R-squared | 0.116144 | S.D. dependent var | 22303891 |
| S.E. of regression | 20968700 | Akaike info criterion | 36.60447 |
| Sum squared resid | 2.46E+16 | Schwarz criterion | 36.71011 |
| Log likelihood | -1076.832 | Hannan-Quinn criter. | 36.64571 |
| F-statistic | 4.810758 | Durbin-Watson stat | 2.028760 |
| Prob(F-statistic) | 0.011802 | | |

D'après le modèle [3] on a les résultats suivants :

- La statistique $t = -2.077310$ est supérieure aux valeurs critiques relatives à 1% 5% 10% donc la série admet une racine unitaire

- La tendance n'est pas significativement différent de zéro car la valeur estimée de la tendance = 2.46 < 2.79 la valeur critique (donnée par la table de Dickey-Fuller) au seuil de 5%.

On passe alors au modèle [2]: test sur la constante

Estimation du Modèle [2] :

| | | t-Statistic | Prob.* |
|---|-----------|-----------------|---------------|
| Augmented Dickey-Fuller test statistic | | 1.810860 | 0.9997 |
| Test critical values: | | | |
| | 1% level | -3.546099 | |
| | 5% level | -2.911730 | |
| | 10% level | -2.593551 | |

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|---------------|-------------|------------|-------------|--------|
| RETRAITSA(-1) | 0.013413 | 0.007407 | 1.810860 | 0.0754 |
| C | -6236236. | 15201496 | -0.410238 | 0.6832 |

| | | | |
|--------------------|-----------|-----------------------|----------|
| R-squared | 0.054400 | Mean dependent var | 20804013 |
| Adjusted R-squared | 0.037811 | S.D. dependent var | 22303891 |
| S.E. of regression | 21878163 | Akaike info criterion | 36.67319 |
| Sum squared resid | 2.73E+16 | Schwarz criterion | 36.74361 |
| Log likelihood | -1079.859 | Hannan-Quinn criter. | 36.70068 |
| F-statistic | 3.279213 | Durbin-Watson stat | 2.012171 |
| Prob(F-statistic) | 0.075434 | | |

- La statistique (1.8101860) est supérieure aux valeurs critiques relatives (-3.546099 ; -2.911730 ; -2.593551) au seuils d 1% 5% 10% (donnée par la table de Dickey-Fuller) donc notre série admet une racine unitaire.
- la statistique relative à la constante t=- 0.4102 est inférieure à valeur critique (2.54) au seuil de 5%, donc la constante n'est pas significativement différent de zéro on passe au modèle [1] sans constante et sans tendance.

Estimation du Modèle [1]

| | | t-Statistic | Prob.* |
|---|-----------|-----------------|---------------|
| Augmented Dickey-Fuller test statistic | | 7.568428 | 1.0000 |
| Test critical values: | | | |
| | 1% level | -2.604746 | |
| | 5% level | -1.946447 | |
| | 10% level | -1.613238 | |

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|---------------|-------------|------------|-------------|--------|
| RETRAITSA(-1) | 0.010429 | 0.001378 | 7.568428 | 0.0000 |

| | | | |
|--------------------|-----------|-----------------------|----------|
| R-squared | 0.051608 | Mean dependent var | 20804013 |
| Adjusted R-squared | 0.051608 | S.D. dependent var | 22303891 |
| S.E. of regression | 21720733 | Akaike info criterion | 36.64224 |
| Sum squared resid | 2.74E+16 | Schwarz criterion | 36.67745 |
| Log likelihood | -1079.946 | Hannan-Quinn criter. | 36.65598 |
| Durbin-Watson stat | 2.000248 | | |

Comme la valeur estimée de la statistique qui est égale à (7.568428) est supérieure aux différentes valeurs critiques relatives aux seuils 1%, 5% et 10%. Donc la série possède de racine unitaire. Par conséquent, le processus qui génère la série (RETRAITESA) n'est pas stationnaire de type DS.

Pour stationnariser notre série en utilisant le filtre aux différences première on la note (DRETRAITESA).

➤ **Test de racine unitaire (Dickey-Fuller) sur la série différenciée (DRETRAITESA)**

On commence par le modèle [3] : test sur la tendance

| | | t-Statistic | Prob.* | |
|---|-------------|-----------------------|-------------|----------|
| Augmented Dickey-Fuller test statistic | | | | |
| Test critical values: | 1% level | -4.124265 | | |
| | 5% level | -3.489228 | | |
| | 10% level | -3.173114 | | |
| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
| D(RETRAITSA(-1)) | -1.022562 | 0.134738 | -7.589237 | 0.0000 |
| C | 9448837. | 6112115. | 1.545919 | 0.1279 |
| @TREND("2014M01") | 389892.5 | 179390.8 | 2.173425 | 0.0341 |
| R-squared | 0.511533 | Mean dependent var | | 177359.0 |
| Adjusted R-squared | 0.493770 | S.D. dependent var | | 30826793 |
| S.E. of regression | 21933211 | Akaike info criterion | | 36.69524 |
| Sum squared resid | 2.65E+16 | Schwarz criterion | | 36.80182 |
| Log likelihood | -1061.162 | Hannan-Quinn criter. | | 36.73675 |
| F-statistic | 28.79854 | Durbin-Watson stat | | 2.002187 |
| Prob(F-statistic) | 0.000000 | | | |

- La statistique $t=-7.589237$ est inférieure aux valeurs critiques relatives à 1% 5% 10% donc la série ne possède pas une racine unitaire
- la t-statistique associée à la tendance qui égale à 2.17 est inférieure à la valeur critique par Dickey – Fuller aux seuils de 5% avec une probabilité supérieure à 5% donc le coefficient de la tendance n’est pas significativement différent de zéro.

On passe alors au modèle [2]:test sur la constante.

| | | t-Statistic | Prob.* | |
|---|-------------|-----------------------|-------------|----------|
| Augmented Dickey-Fuller test statistic | | | | |
| Test critical values: | 1% level | -3.548208 | | |
| | 5% level | -2.912631 | | |
| | 10% level | -2.594027 | | |
| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
| D(RETRAITSA(-1)) | -0.939561 | 0.133440 | -7.041068 | 0.0000 |
| C | 19622748 | 4058671. | 4.834771 | 0.0000 |
| R-squared | 0.469580 | Mean dependent var | | 177359.0 |
| Adjusted R-squared | 0.460108 | S.D. dependent var | | 30826793 |
| S.E. of regression | 22650711 | Akaike info criterion | | 36.74315 |
| Sum squared resid | 2.87E+16 | Schwarz criterion | | 36.81420 |
| Log likelihood | -1063.551 | Hannan-Quinn criter. | | 36.77083 |
| F-statistic | 49.57664 | Durbin-Watson stat | | 2.006118 |
| Prob(F-statistic) | 0.000000 | | | |

- Dans le modèle [2], la t-statistique associée à la constante qui égale 4.834 est supérieure à la valeur critique au seuils de 5% donc la constante est significativement différente de zéro.

On conclure que notre série (DRETRAITESA) est stationnaire à la première différence. Nous passons maintenant à l’étape du choix du modèle adéquat.

4.4.3.1 Identification de modèle

Corrélogramme de série (RETRAITESA)

| Autocorrelation | Partial Correlation | AC | PAC | Q-Stat | Prob | |
|-----------------|---------------------|----|--------|--------|--------|-------|
| | | 1 | 0.060 | 0.060 | 0.2259 | 0.635 |
| | | 2 | 0.059 | 0.055 | 0.4436 | 0.801 |
| | | 3 | 0.134 | 0.128 | 1.5946 | 0.661 |
| | | 4 | 0.054 | 0.037 | 1.7825 | 0.776 |
| | | 5 | 0.052 | 0.035 | 1.9624 | 0.854 |
| | | 6 | 0.124 | 0.101 | 3.0062 | 0.808 |
| | | 7 | 0.033 | 0.008 | 3.0794 | 0.878 |
| | | 8 | 0.031 | 0.008 | 3.1479 | 0.925 |
| | | 9 | 0.050 | 0.017 | 3.3289 | 0.950 |
| | | 10 | 0.022 | 0.003 | 3.3660 | 0.971 |
| | | 11 | 0.021 | 0.003 | 3.3992 | 0.984 |
| | | 12 | -0.311 | -0.348 | 10.820 | 0.544 |
| | | 13 | 0.040 | 0.065 | 10.948 | 0.615 |
| | | 14 | 0.039 | 0.058 | 11.067 | 0.681 |
| | | 15 | 0.033 | 0.105 | 11.155 | 0.742 |
| | | 16 | 0.036 | 0.038 | 11.262 | 0.793 |
| | | 17 | 0.034 | 0.034 | 11.361 | 0.837 |
| | | 18 | 0.069 | 0.142 | 11.784 | 0.858 |
| | | 19 | 0.020 | -0.013 | 11.822 | 0.893 |
| | | 20 | 0.019 | -0.011 | 11.854 | 0.921 |
| | | 21 | 0.037 | 0.016 | 11.983 | 0.940 |
| | | 22 | 0.014 | -0.014 | 12.003 | 0.957 |
| | | 23 | 0.012 | -0.012 | 12.018 | 0.970 |

L'analyse du corrélogramme simple et partiel de la série stationnaire (DRETRAITESA) montre qu'au retard (k=12) le terme est significativement différent de 0. Par conséquent nous avons trois modèles candidats.

ARMA(12,0) ARMA(0,12) ; ARMA(12,12) Et pour pouvoir choisir un bon modèle parmi ceux présente, on construit un tableau de comparaison de leurs critères d'information et nous retenons le modèle qui minimise ces critères

➤ **Choix de modèle optimal**

| Modèle | R ² | SSR | SC | AIC |
|------------------|----------------|-----------------------|-------|-------|
| ARIMA (12,1, 12) | 0.58 | 9.20 ^E +16 | 35.99 | 35.87 |
| ARIMA (12, 1, 0) | 0.16 | 1.84 ^E +16 | 36.60 | 36.52 |
| ARIMA (0, 1, 12) | 0.51 | 1.40 ^E +16 | 36.07 | 36.00 |

Tableau 5:choix du modèle optimal

On choisit le modèle ARIMA (0,1,12) parmi les modèles proposés et on rejette ARIMA(12,1,12) car les paramètres sont significativement nul.

4.4.3.2 Estimation des paramètres du modèle optimal

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|--------------------|-------------|-----------------------|-------------|--------|
| C | 23552177 | 1343820. | 17.52629 | 0.0000 |
| MA(12) | -0.969408 | 0.001131 | -857.4218 | 0.0000 |
| R-squared | 0.515041 | Mean dependent var | 20804013 | |
| Adjusted R-squared | 0.506533 | S.D. dependent var | 22303891 | |
| S.E. of regression | 15667856 | Akaike info criterion | 36.00543 | |
| Sum squared resid | 1.40E+16 | Schwarz criterion | 36.07586 | |
| Log likelihood | -1060.160 | F-statistic | 60.53578 | |
| Durbin-Watson stat | 1.993612 | Prob(F-statistic) | 0.000000 | |

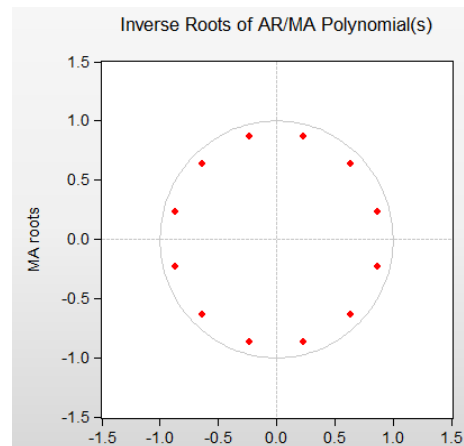
4.4.3.3 Validation du modèle

➤ Test sur les paramètres

Les coefficients des paramètres du modèle ARMA(0,12) sont significativement différent de 0, d’ailleurs le test de Student le confirme puisque les t-statistique associées aux paramètres du modèle (17.52; -857.42) en valeurs absolues sont supérieures aux valeurs théoriques (1.96) au seuil de 5%. Leurs inverses sont tous inférieures à 1 par conséquent, les conditions de stationnarité et d’invisibilité sont vérifiées.

| MA Root(s) | Modulus | Cycle |
|-----------------------|----------|----------|
| -0.637190 ± 0.637190i | 0.901123 | 2.666667 |
| -0.233228 ± 0.870418i | 0.901123 | 3.428571 |
| 0.637190 ± 0.637190i | 0.901123 | 8.000000 |
| 0.870418 ± 0.233228i | 0.901123 | 24.00000 |
| 0.233228 ± 0.870418i | 0.901123 | 4.800000 |
| -0.870418 ± 0.233228i | 0.901123 | 2.181818 |

No root lies outside the unit circle.
ARMA model is invertible.



➤ Test sur les résidus

➤ Correlogramme des résidus

| Autocorrelation | Partial Correlation | AC | PAC | Q-Stat | Prob |
|-----------------|---------------------|-----------|--------|--------|-------|
| | | 1 -0.053 | -0.053 | 0.1751 | |
| | | 2 -0.060 | -0.063 | 0.4000 | 0.527 |
| | | 3 0.088 | 0.082 | 0.8999 | 0.638 |
| | | 4 -0.052 | -0.047 | 1.0767 | 0.783 |
| | | 5 -0.057 | -0.053 | 1.2942 | 0.862 |
| | | 6 0.110 | 0.094 | 2.1173 | 0.833 |
| | | 7 -0.047 | -0.036 | 2.2693 | 0.893 |
| | | 8 -0.052 | -0.039 | 2.4582 | 0.930 |
| | | 9 0.071 | 0.043 | 2.8178 | 0.945 |
| | | 10 -0.060 | -0.048 | 3.0781 | 0.961 |
| | | 11 -0.064 | -0.051 | 3.3830 | 0.971 |
| | | 12 0.028 | -0.008 | 3.4412 | 0.984 |
| | | 13 -0.013 | -0.002 | 3.4539 | 0.991 |
| | | 14 -0.017 | -0.001 | 3.4768 | 0.996 |
| | | 15 0.013 | -0.015 | 3.4897 | 0.998 |
| | | 16 -0.004 | 0.003 | 3.4914 | 0.999 |
| | | 17 -0.007 | 0.005 | 3.4960 | 1.000 |
| | | 18 0.014 | -0.001 | 3.5131 | 1.000 |
| | | 19 -0.017 | -0.016 | 3.5402 | 1.000 |
| | | 20 -0.021 | -0.017 | 3.5796 | 1.000 |
| | | 21 0.081 | 0.072 | 4.2057 | 1.000 |
| | | 22 -0.011 | -0.007 | 4.2186 | 1.000 |
| | | 23 -0.014 | -0.006 | 4.2382 | 1.000 |
| | | 24 -0.168 | -0.192 | 7.1412 | 0.999 |

Le corrélogramme des résidus du modèle retenu ARIMA (0,1,12) ne fait apparaître aucun terme en dehors de l'intervalle de confiance au seuil de 5% et la statistique de **Box-Ljung Q – stat = 7.141(k = 24) < $\chi^2_{23} = 35.17$** au seuil de 5%. Donc les résidus forment un bruit blanc.

➤ **Test de normalité**

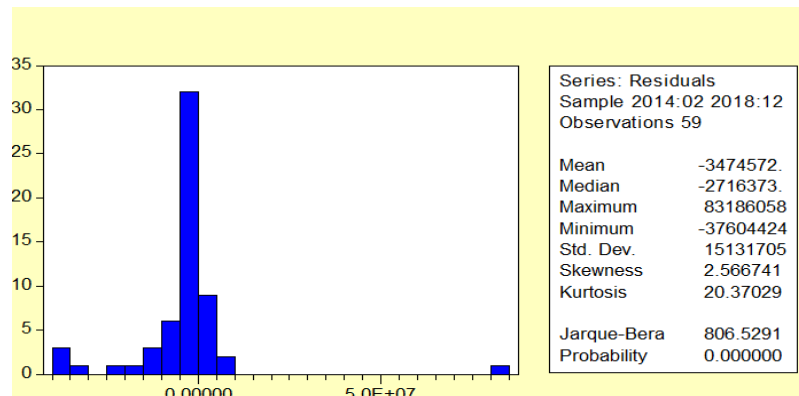


Figure Erreur ! Il n'y a pas de texte répondant à ce style dans ce document.-8:Test de normalité

➤ **Test skewness:** $Sk^{1/2} = 2.56 \rightarrow \vartheta_1 = \frac{|2.56|}{\sqrt{\frac{6}{59}}} = 8.02 > 1.96$

➤ **Test Kurtosis:** $Kur = 25.64 \rightarrow \vartheta_2 = \frac{|20.37-3|}{\sqrt{\frac{24}{59}}} = 4.61 > 1.96$

D'après les résultats on rejette H_0 concerne la normalité et l'aplatissement ce qui est confirmé par la statistique de Jarque-Berra.

➤ **Statistique de jarque-Berra**

$$JB = \left(\frac{59}{6}\right) SK + \left(\frac{59}{24}\right) (Ku - 3)^2 = 1386.95 > 5.99.$$

D'où les résidus sont bruit blanc non gaussien

➤ **Test d'hétéroscédasticité (l'existence d'effet ARCH)**

| Heteroskedasticity Test: ARCH | | | | |
|---|-------------|-----------------------|-------------|--------|
| F-statistic | 0.099815 | Prob. F(1,56) | 0.7532 | |
| Obs*R-squared | 0.103196 | Prob. Chi-Square(1) | 0.7480 | |
| Test Equation: | | | | |
| Dependent Variable: RESID^2 | | | | |
| Method: Least Squares | | | | |
| Date: 05/28/19 Time: 00:46 | | | | |
| Sample (adjusted): 2014M03 2018M12 | | | | |
| Included observations: 58 after adjustments | | | | |
| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
| C | 4.30E+14 | 2.64E+14 | 1.625189 | 0.1097 |
| RESID^2(-1) | -0.042179 | 0.133507 | -0.315935 | 0.7532 |
| R-squared | 0.001779 | Mean dependent var | 4.12E+14 | |
| Adjusted R-squared | -0.016046 | S.D. dependent var | 1.95E+15 | |
| S.E. of regression | 1.97E+15 | Akaike info criterion | 73.30495 | |
| Sum squared resid | 2.17E+32 | Schwarz criterion | 73.37600 | |
| Log likelihood | -2123.844 | Hannan-Quinn criter. | 73.33263 | |
| F-statistic | 0.099815 | Durbin-Watson stat | 2.003798 | |
| Prob(F-statistic) | 0.753226 | | | |

A partir de la figure on remarque que la probabilité de RESID2 est supérieur a 0.05 donc on accepte l'hypothèse nulle d'hémoscédasticité il n'existe pas un effet ARCH.

Conclusion

$$REETRAITESA_t = c + (1 - \theta B^{12})\varepsilon_t$$

$$REETRAITESA_t = 23552176.71 - 0.9694084866\varepsilon_{t-12} + \varepsilon_t$$

4.4.3.4 Prévision

Rappelons que la série sous étude "retraite" comprend 60 observations de Janvier 2014 à décembre 2018, donc l'instant t est décembre 2018. On souhaitera prédire les 12 mois de l'année 2019, soit un horizon $h = 12$:

On remplaçant t par $t + h$ dans l'équation précédente on obtient :

$$\text{REETRAITESA}_{t+h} = 23552176.71 - 0.9694084866\varepsilon_{t+h-12} + \varepsilon_{t+h}$$

Avec $h = 1, \dots, 12$. On obtient les prévisions suivantes :

| Mois de l'année 2019 | Prévision |
|----------------------|------------|
| Janvier | 2914543471 |
| Février | 2904929665 |
| Mars | 2895577450 |
| Avril | 2915799168 |
| Mai | 2905995792 |
| Juin | 2896442673 |
| Juillet | 2990771914 |
| Aout | 2982464695 |
| Septembre | 2974440123 |
| Octobre | 2987993765 |
| Novembre | 2979481822 |
| décembre | 2971226825 |

Le graphe des prévisions est donné comme suit :

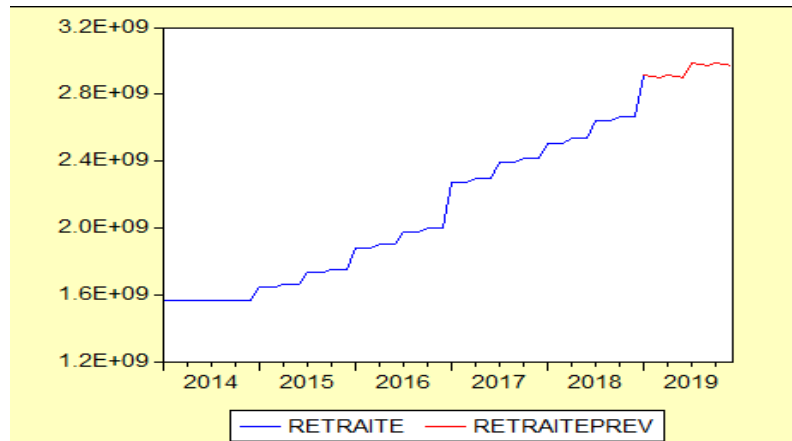


Figure Erreur ! Il n'y a pas de texte répondant à ce style dans ce document.-9:prévisions de la série (RETRAITE)

4.4.4 Etude de la série RECETTE

4.4.4.1 Étude préliminaire de la série

➤ **Représentation graphique de la série:**

Pour illustrer cette première phase de modélisation, nous examinons le graphique de la série nommée RECETTE

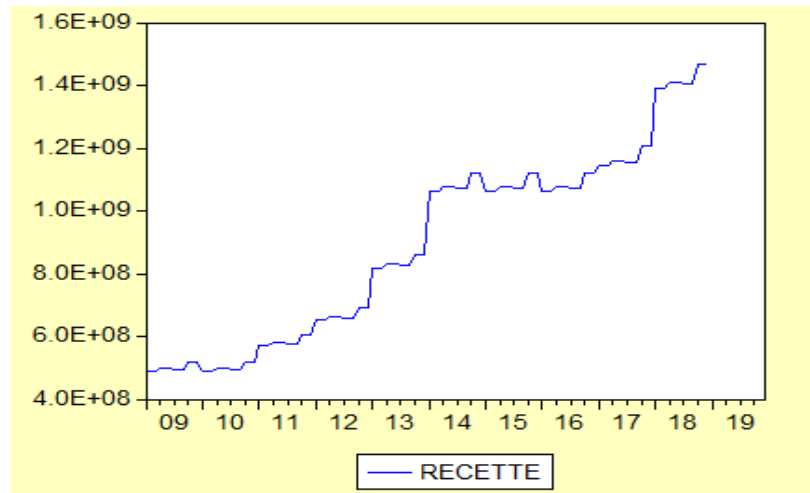


Figure Erreur ! Il n'y a pas de texte répondant à ce style dans ce document.-10:graphe de la série recette

La représentation graphique de la série Caractérisée par une tendance haussière, et Un phénomène périodique caractérisé par des mouvements ascendants et descendants, ces mouvements périodiques sont dus au rythme des saisons.

Dans ce cas nous dirons que la série RECETTE n'est pas stationnaire, Afin d'éloigner l'incertitude, nous utilisons le test de Fisher et le test de DICKEY-FULLER.

➤ **Vérification du facteur saisonnier et tendanciel :**

- Test d'influence du facteur lignes (mois)
 - { H_0 : l'effet saisonnier non significatif
 - { H_1 : l'effet saisonnier significatif

- Test d'influence du facteur colonne (années)
 - { H_0 : l'effet de la tendance non significatif
 - { H_1 : l'effet de la tendance significatif

Les résultats du test de Fisher sont regroupés dans le tableau suivant :

| ANALYSE DE VARIANCE | | | | | | |
|-----------------------|------------------|------------------|--------------------|------------|-------------|------------------------|
| Source des variations | Somme des carrés | Degré de liberté | Moyenne des carrés | F | Probabilité | Valeur critique pour F |
| Lignes | 1,70812E+17 | 11 | 1,55283E+16 | 8,61893037 | 1,6406E-10 | 1,886683603 |
| Colonnes | 1,14042E+19 | 9 | 1,26713E+18 | 703,315045 | 1,5526E-85 | 1,975806073 |

Tableau 6:Fisher de la série recette

✓ Résultat du test d'influence de facteur mois (saisonnalité):

$$F_{cal} = 8.6189 > F^{0.05}(11; 119) = 1.8866$$

On rejette l'hypothèse H_0 , donc la série affectée d'une saisonnalité.

✓ Résultat du test d'influence de facteur année (tendance):

$$F_{cal} = 703.315 > F^{0.05}(9; 119) = 1.9758$$

Donc, nous rejetons l’hypothèse nulle. La série est donc affectée d’une tendance.

➤ **Dessaisonalisation de la série (RECETTE)**

Une dessaisonalisation de la série (RECETTE) s’impose pour la suite de notre étude. Notons (RECETTESA) la série désaisonnalisée

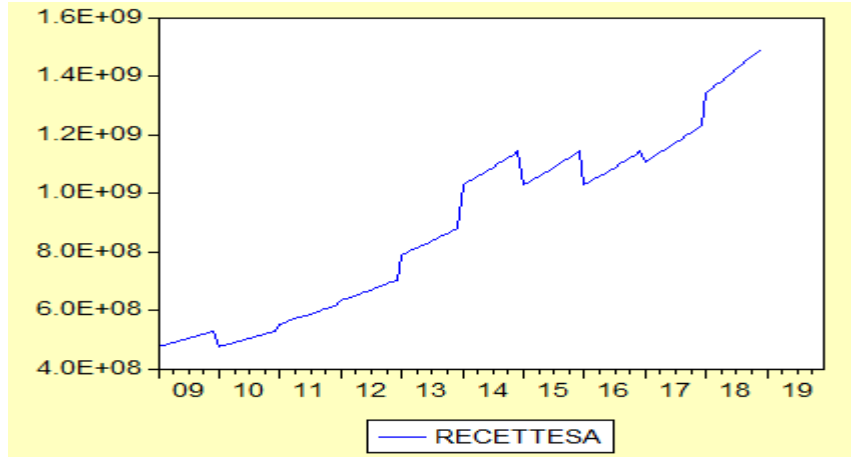


Figure Erreur ! Il n'y a pas de texte répondant à ce style dans ce document.-11:graphe de la série recettesa

La représentation graphique de la série (RECETTESA) fait monter la neutralité de l’effet saisonnier.

➤ **Vérification de l’élimination du facteur saisonnier**

On applique le test ANOVA sur la série (RECETTESA) pour voir si l’effet saisonnier a été éliminé.

| ANALYSE DE VARIANCE | | | | | | |
|-----------------------|------------------|------------------|--------------------|------------|-------------|------------------------|
| Source des variations | Somme des carrés | Degré de liberté | Moyenne des carrés | F | Probabilité | Valeur critique pour F |
| Lignes | 1,96334E+17 | 11 | 1,78485E+16 | 1,10186103 | 0,36792044 | 1,886683603 |
| Colonnes | 1,01171E+19 | 9 | 1,12412E+18 | 69,396422 | 8,9873E-39 | 1,975806073 |

Tableau 7: Fisher de la série recettesa

✓ Résultat du test d’influence de facteur mois (saisonnalité):

On a $F_c = 1.10186 < F^{0.05}(11,119) = 1.88668$

On accepte l’hypothèse H_0 : la série n’est pas saisonnière.

A travers le tableau et le graphe de la série désaisonnalisée qui sont affichés, nous remarquons que l’effet saisonnier a disparu, mais on ne peut dire que le processus qui génère cette série est stationnaire qu’après avoir fait un recours au test de racine unitaire (Dickey-Fuller) qui permet d’affirmer ou d’infirmier notre hypothèse.

4.4.4.2 Etude de la stationnarité de la série dessaisonnalisée(RECETTESA)

Afin d’étudier la stationnarité de cette série, nous allons y procéder par les différents tests suivants :

➤ **Test de la racine unitaire (Dickey -Fuller) sur la série (RECETTESA)**

On commence par le modèle 3 : Test sur la tendance

Estimation du modèle 3

| | | t-Statistic | Prob.* |
|---|-----------|------------------|---------------|
| Augmented Dickey-Fuller test statistic | | -1.893432 | 0.6517 |
| Test critical values: | 1% level | -4.036983 | |
| | 5% level | -3.448021 | |
| | 10% level | -3.149135 | |

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|-------------------|-------------|------------|-------------|--------|
| RECETTESA(-1) | -0.060329 | 0.031862 | -1.893432 | 0.0608 |
| C | 26563546 | 12837805 | 2.069166 | 0.0408 |
| @TREND("2009M01") | 591836.8 | 279031.3 | 2.121041 | 0.0360 |

| | | | |
|--------------------|-----------|-----------------------|----------|
| R-squared | 0.041284 | Mean dependent var | 8567317. |
| Adjusted R-squared | 0.024754 | S.D. dependent var | 25111642 |
| S.E. of regression | 24798884 | Akaike info criterion | 36.91538 |
| Sum squared resid | 7.13E+16 | Schwarz criterion | 36.98544 |
| Log likelihood | -2193.465 | Hannan-Quinn criter. | 36.94383 |
| F-statistic | 2.497573 | Durbin-Watson stat | 1.942145 |
| Prob(F-statistic) | 0.086701 | | |

D'après le modèle [3] on a les résultats suivants :

- La statistique $t=-1.8934$ est supérieure aux valeurs critiques relatives à 1% 5% 10% donc la série admet une racine unitaire
- La tendance n'est pas significativement différent de zéro car la valeur estimé de la tendance $=2.1210 < 2.79$ la valeur critique (donnée par la table de Dickey-Fuller) au seuil de 5%. On passe alors au modèle [2]:test sur la constante.

Estimation du modèle 2

| | | t-Statistic | Prob.* |
|---|-----------|-----------------|---------------|
| Augmented Dickey-Fuller test statistic | | 0.694205 | 0.9916 |
| Test critical values: | 1% level | -3.486064 | |
| | 5% level | -2.885863 | |
| | 10% level | -2.579818 | |

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|---------------|-------------|------------|-------------|--------|
| RECETTESA(-1) | 0.005324 | 0.007669 | 0.694205 | 0.4889 |
| C | 3845520. | 7182342. | 0.535413 | 0.5934 |

| | | | |
|--------------------|-----------|-----------------------|----------|
| R-squared | 0.004102 | Mean dependent var | 8567317. |
| Adjusted R-squared | -0.004410 | S.D. dependent var | 25111642 |
| S.E. of regression | 25166950 | Akaike info criterion | 36.93663 |
| Sum squared resid | 7.41E+16 | Schwarz criterion | 36.98333 |
| Log likelihood | -2195.729 | Hannan-Quinn criter. | 36.95559 |
| F-statistic | 0.481921 | Durbin-Watson stat | 1.996372 |
| Prob(F-statistic) | 0.488930 | | |

- La statistique (0.694205) est supérieure aux valeurs critiques relatives

(-3.4860 ; -2.8858 ; -2.5798) au seuils d 1% 5% 10% (donnée par la table de Dickey-Fuller Augmenté) donc notre série admet une racine unitaire

- la statistique relative à la constante $t=0.535413$ est inférieure à la valeur critique (2.54) donc la constante n'est pas significativement différent de zéro on passe au modèle [1] sans constante et sans tendance.

Estimation du modèle 1

| | t-Statistic | Prob.* |
|--|-------------|--------|
| Augmented Dickey-Fuller test statistic | 3.751089 | 0.9999 |
| Test critical values: | | |
| 1% level | -2.584539 | |
| 5% level | -1.943540 | |
| 10% level | -1.614941 | |

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|--------------------|-------------|-----------------------|-------------|----------|
| RECETTESA(-1) | 0.009212 | 0.002456 | 3.751089 | 0.0003 |
| R-squared | 0.001662 | Mean dependent var | | 8567317. |
| Adjusted R-squared | 0.001662 | S.D. dependent var | | 25111642 |
| S.E. of regression | 25090765 | Akaike info criterion | | 36.92227 |
| Sum squared resid | 7.43E+16 | Schwarz criterion | | 36.94562 |
| Log likelihood | -2195.875 | Hannan-Quinn criter. | | 36.93175 |
| Durbin-Watson stat | 1.999249 | | | |

Comme la valeur estimée de la statistique qui est égale à (3.751089) est supérieure aux différentes valeurs critiques (-2.5845); (-1.9435); (-1.6149) relatives aux seuils 1%, 5% et 10% respectivement, donc la série possède de racine unitaire. Par conséquent, le processus qui génère la série (RECETTESA) n'est pas stationnaire de type DS.

Pour stationnariser notre série en utilisant le filtre aux différences première on la note (DRECETTESA).

➤ Test racine unitaire (Dickey-Fuller) de la série différenciée

On commence par le modèle 3 : test sur la tendance

| | t-Statistic | Prob.* |
|--|-------------|--------|
| Augmented Dickey-Fuller test statistic | -10.72893 | 0.0000 |
| Test critical values: | | |
| 1% level | -4.037668 | |
| 5% level | -3.448348 | |
| 10% level | -3.149326 | |

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|--------------------|-------------|-----------------------|-------------|----------|
| D(RECETTESA(-1)) | -1.000479 | 0.093251 | -10.72893 | 0.0000 |
| C | 3830546. | 4758269. | 0.805029 | 0.4225 |
| @TREND("2009M01") | 78915.86 | 68732.96 | 1.148152 | 0.2533 |
| R-squared | 0.500239 | Mean dependent var | | 79271.06 |
| Adjusted R-squared | 0.491548 | S.D. dependent var | | 35464602 |
| S.E. of regression | 25288337 | Akaike info criterion | | 36.95468 |
| Sum squared resid | 7.35E+16 | Schwarz criterion | | 37.02512 |
| Log likelihood | -2177.326 | Hannan-Quinn criter. | | 36.98328 |
| F-statistic | 57.55500 | Durbin-Watson stat | | 2.000000 |
| Prob(F-statistic) | 0.000000 | | | |

- La statistique $t=-10.7289$ est inférieure aux valeurs critiques relatives à 1% 5% 10% donc la série ne possède pas une racine unitaire.
- la t-statistique associée à la tendance qui égale à 1.14 est inférieure à la valeur critique donnée par Dickey – Fuller au seuil de 5% avec une probabilité supérieure à 5% donc

le coefficient de la tendance n'est pas significativement différent de zéro .On passe alors au modèle [2]:test sur la constante

| | | | t-Statistic | Prob.* |
|--|-------------|-----------------------|-------------|----------|
| Augmented Dickey-Fuller test statistic | | | -10.65272 | 0.0000 |
| Test critical values: | | | | |
| | 1% level | | -3.486551 | |
| | 5% level | | -2.886074 | |
| | 10% level | | -2.579931 | |
| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
| D(RECETTESA(-1)) | -0.989111 | 0.092851 | -10.65272 | 0.0000 |
| C | 8508089. | 2461791. | 3.456057 | 0.0008 |
| R-squared | 0.494510 | Mean dependent var | | 79271.06 |
| Adjusted R-squared | 0.490153 | S.D. dependent var | | 35464602 |
| S.E. of regression | 25323003 | Akaike info criterion | | 36.94913 |
| Sum squared resid | 7.44E+16 | Schwarz criterion | | 36.99609 |
| Log likelihood | -2177.999 | Hannan-Quinn criter. | | 36.96820 |
| F-statistic | 113.4804 | Durbin-Watson stat | | 2.000060 |
| Prob(F-statistic) | 0.000000 | | | |

- Dans le modèle [2], la t-statistique associée à la constante qui égale 3.45 est supérieure à valeur critique au seuil de 5% donc la constante est significativement différente de zéro.
- On conclure que notre série (DRECETTSA) est stationnaire au première différence.

Nous passons maintenant à l'étape du choix du modèle adéquat

➤ **Corrélogramme de la série (DRECETTSA)**

| Autocorrelation | Partial Correlation | AC | PAC | Q-Stat | Prob |
|-----------------|---------------------|-----------|--------|--------|-------|
| | | 1 0.011 | 0.011 | 0.0144 | 0.904 |
| | | 2 0.010 | 0.010 | 0.0276 | 0.986 |
| | | 3 0.010 | 0.010 | 0.0397 | 0.998 |
| | | 4 0.009 | 0.009 | 0.0510 | 1.000 |
| | | 5 0.009 | 0.009 | 0.0610 | 1.000 |
| | | 6 0.009 | 0.009 | 0.0715 | 1.000 |
| | | 7 0.009 | 0.009 | 0.0825 | 1.000 |
| | | 8 0.009 | 0.008 | 0.0926 | 1.000 |
| | | 9 0.008 | 0.008 | 0.1020 | 1.000 |
| | | 10 0.008 | 0.007 | 0.1109 | 1.000 |
| | | 11 0.008 | 0.007 | 0.1188 | 1.000 |
| | | 12 0.137 | 0.136 | 2.6499 | 0.998 |
| | | 13 -0.004 | -0.007 | 2.6519 | 0.999 |
| | | 14 -0.004 | -0.007 | 2.6543 | 1.000 |
| | | 15 -0.005 | -0.008 | 2.6573 | 1.000 |
| | | 16 -0.005 | -0.008 | 2.6611 | 1.000 |
| | | 17 -0.005 | -0.008 | 2.6652 | 1.000 |
| | | 18 -0.006 | -0.009 | 2.6712 | 1.000 |
| | | 19 -0.007 | -0.009 | 2.6791 | 1.000 |
| | | 20 -0.008 | -0.009 | 2.6874 | 1.000 |
| | | 21 -0.008 | -0.009 | 2.6960 | 1.000 |
| | | 22 -0.008 | -0.009 | 2.7050 | 1.000 |
| | | 23 -0.008 | -0.009 | 2.7142 | 1.000 |
| | | 24 -0.431 | -0.457 | 30.832 | 0.159 |
| | | 25 -0.004 | 0.006 | 30.835 | 0.195 |
| | | 26 -0.004 | 0.005 | 30.838 | 0.234 |
| | | 27 -0.005 | 0.005 | 30.841 | 0.278 |
| | | 28 -0.005 | 0.004 | 30.846 | 0.324 |
| | | 29 -0.005 | 0.003 | 30.851 | 0.372 |
| | | 30 -0.006 | 0.002 | 30.857 | 0.422 |

Nous remarquons que la fonction d'auto-corrélation simple (AC) possède des valeurs importantes aux retards $q = 24$ et la fonction d'auto-corrélation partielle (PAC) possède des valeurs importantes aux retards $p = 24$.

Par conséquent nous avons sélectionné les modèles suivant:

ARIMA(0 ;1 ;24) ; ARIMA(24 ;1 ;0) Et pour pouvoir choisir un bon modèle parmi ceux présente, on construit un tableau de comparaison de leurs critères d'information et nous retenons le modèle qui minimise ces critères.

4.4.4.3 Estimation des paramètres du modèle optimal

Après avoir identifié le modèle, il convient de passer à la phase d'estimation des paramètres de ce modèle, nous avons :

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|--------------------|-------------|-----------------------|-------------|----------|
| C | 8624285. | 2282634. | 3.778217 | 0.0003 |
| AR(24) | -0.521049 | 0.059766 | -8.718116 | 0.0000 |
| AR(12) | 0.298450 | 0.047457 | 6.288899 | 0.0000 |
| SIGMASQ | 4.26E+14 | 1.66E+13 | 25.61570 | 0.0000 |
| R-squared | 0.318390 | Mean dependent var | | 8567317. |
| Adjusted R-squared | 0.300609 | S.D. dependent var | | 25111642 |
| S.E. of regression | 21000770 | Akaike info criterion | | 36.65889 |
| Sum squared resid | 5.07E+16 | Schwarz criterion | | 36.75231 |
| Log likelihood | -2177.204 | Hannan-Quinn criter. | | 36.69683 |
| F-statistic | 17.90604 | Durbin-Watson stat | | 1.978375 |
| Prob(F-statistic) | 0.000000 | | | |

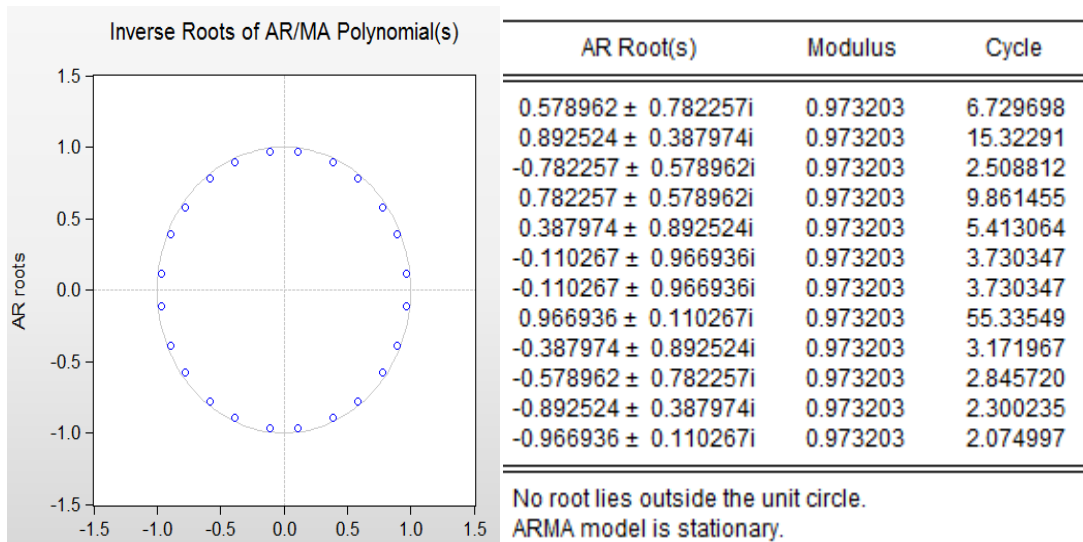
4.4.4.4 Validation du modèle

➤ Test sur les paramètres du modèle

D'après le tableau suivant et le tableau d'estimation nous constatons que :

Les coefficients des paramètres du modèle sont significativement différent de zéro car les t-statistiques (8.71) et (6.29) du modèle sont supérieure à la valeur critique (de Student =1.96) au seuil de 5% ce qui est confirmé par les probabilités des coefficients qui sont tous inférieurs a 0,05.

Les inverses des racines sont distincts et inférieurs en module à 1 .par conséquent, les conditions d'invisibilité et de stationnarité sont vérifiées.



➤ Test sur les résidus

Corrélogramme des résidus

| Autocorrelation | Partial Correlation | AC | PAC | Q-Stat | Prob |
|-----------------|---------------------|----|--------|--------|--------|
| | | 1 | 0.004 | 0.004 | 0.0019 |
| | | 2 | 0.004 | 0.003 | 0.0031 |
| | | 3 | 0.003 | 0.003 | 0.0039 |
| | | 4 | 0.002 | 0.002 | 0.0043 |
| | | 5 | 0.001 | 0.001 | 0.0044 |
| | | 6 | 0.001 | 0.001 | 0.0045 |
| | | 7 | 0.001 | 0.001 | 0.0047 |
| | | 8 | 0.000 | 0.000 | 0.0047 |
| | | 9 | 0.000 | 0.000 | 0.0047 |
| | | 10 | -0.001 | -0.001 | 0.0047 |
| | | 11 | -0.001 | -0.001 | 0.0049 |
| | | 12 | -0.235 | -0.235 | 6.1363 |
| | | 13 | -0.011 | -0.009 | 6.1492 |
| | | 14 | -0.011 | -0.010 | 6.1634 |
| | | 15 | -0.012 | -0.011 | 6.1800 |
| | | 16 | -0.013 | -0.012 | 6.1991 |
| | | 17 | -0.013 | -0.013 | 6.2198 |
| | | 18 | -0.015 | -0.015 | 6.2457 |
| | | 19 | -0.016 | -0.016 | 6.2770 |
| | | 20 | -0.016 | -0.017 | 6.3097 |
| | | 21 | -0.017 | -0.017 | 6.3443 |
| | | 22 | -0.017 | -0.018 | 6.3809 |
| | | 23 | -0.017 | -0.019 | 6.4191 |
| | | 24 | -0.096 | -0.161 | 7.6247 |
| | | 25 | 0.001 | -0.005 | 7.6248 |
| | | 26 | 0.000 | -0.007 | 7.6248 |
| | | 27 | -0.001 | -0.008 | 7.6249 |
| | | 28 | -0.001 | -0.010 | 7.6250 |
| | | 29 | -0.002 | -0.011 | 7.6256 |
| | | 30 | -0.001 | -0.012 | 7.6258 |

D’après l’analyse du corrélogramme, on remarque que tous les termes sont à l’extérieur de l’intervalle de confiance et les probabilités correspondantes sont supérieure a 0.05, donc on estime que les résidus forment un bruit blanc. Ce qui est confirmé par la statistique de Ljung-Box : $Q - stat(30) = 7.6258 < \chi^2_{(29)} = 42.56$ au seuil 5%. D’où la validité du modèle de plus la statistique $DW = 1.978 \approx 2$, on accepte l’hypothèse de non corrélation des résidus.

➤ Test de normalité des résidus

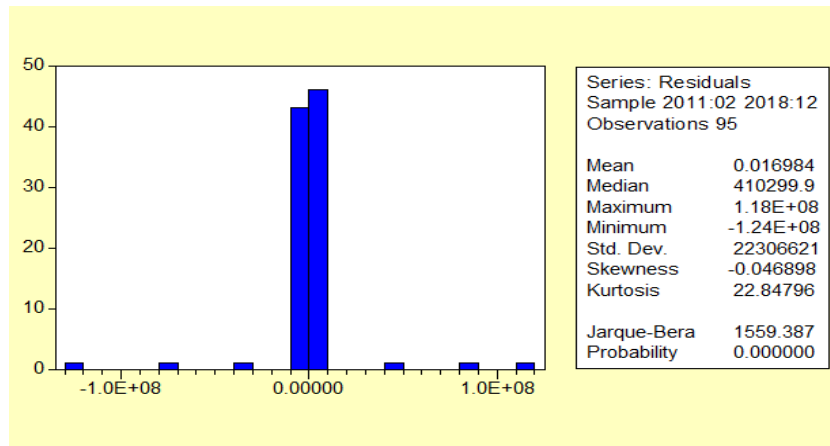


Figure Erreur ! Il n'y a pas de texte répondant à ce style dans ce document.-12 : normalité des résidus

- **Test Skewness:** $Sk^{\frac{1}{2}} = 0.14 \rightarrow v_1 = \frac{Sk^{\frac{1}{2}}}{\sqrt{\frac{6}{n}}} = \frac{|-0.04|}{(\frac{6}{119})^{\frac{1}{2}}} = 0.014 < 1.96$ (distribution symétrique)
- **Test Kurtosis:** $ku = 25.42 \rightarrow v_2 = \frac{Ku-3}{\sqrt{\frac{24}{n}}} = \frac{|22.84-3|}{(\frac{24}{119})^{\frac{1}{2}}} = 44.178 > 1.96$ (distribution leptokurtique)

D'après les résultats du test, nous rejetons l'hypothèse de normalité.

- **Test de Jarque-Berra :**

$$JB = \left(\frac{119}{6}\right) Sk + \left(\frac{119}{24}\right) (Ku - 3)^2 = 1559.387 > \chi_2^2 = 5.99$$

On rejette H_0 Donc les résidus forment un bruit blanc non gaussien.

- **Test d'hétéroscédasticité (l'existence d'effet ARCH)**

| Heteroskedasticity Test: ARCH | | | | |
|--|-------------|-----------------------|-------------|--------|
| F-statistic | 0.190858 | Prob. F(1,116) | 0.6630 | |
| Obs*R-squared | 0.193830 | Prob. Chi-Square(1) | 0.6597 | |
| Test Equation: | | | | |
| Dependent Variable: RESID^2 | | | | |
| Method: Least Squares | | | | |
| Date: 06/07/19 Time: 12:52 | | | | |
| Sample (adjusted): 2009M03 2018M12 | | | | |
| Included observations: 118 after adjustments | | | | |
| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
| C | 4.47E+14 | 2.00E+14 | 2.233673 | 0.0274 |
| RESID^2(-1) | -0.040529 | 0.092770 | -0.436873 | 0.6630 |
| R-squared | 0.001643 | Mean dependent var | 4.30E+14 | |
| Adjusted R-squared | -0.006964 | S.D. dependent var | 2.12E+15 | |
| S.E. of regression | 2.13E+15 | Akaike info criterion | 73.44539 | |
| Sum squared resid | 5.27E+32 | Schwarz criterion | 73.49235 | |
| Log likelihood | -4331.278 | Hannan-Quinn criter. | 73.46446 | |
| F-statistic | 0.190858 | Durbin-Watson stat | 2.003456 | |
| Prob(F-statistic) | 0.663016 | | | |

A partir de la figure on remarque que la probabilité de RESID2 est supérieur a 0.05 donc on accepte l'hypothèse nulle d'hémoscédasticité il n'existe pas d'un effet ARCH.

Conclusion

D'après les résultats on peut valider le modèle de la série (DRECETTESA) s'écrit sous la forme :

$$(1 - B - \theta_1 B^{12} - \theta_2 B^{24})(RECETTESA)_t = C + \varepsilon_t$$

$$RECETTESA_t = 862485 + RECETTESA_{t-1} + 0.29RECETTESA_{t-12} - 0.52 RECETTESA_{t-24} + \varepsilon_t$$

4.4.4.5 Prévision

Rappelons que la série sous étude "recette" comprend 120 observation de Janvier 2009 à décembre 2018, donc l'instant t est décembre 2018.

On souhaitera prédire les 12 mois de l'année 2019, soit un horizon h = 12:

On remplaçant t par t+h dans l'équation précédant on obtient :

$$RECETTESA_{t+h} = 862485 + 0.29RECETTESA_{t+h-12} - 0.52 RECETTESA_{t+h-24} + \varepsilon_{t+h}$$

Avec h=1,.....,12. On obtient les prévisions suivantes :

| Mois de l'année 2019 | Prévision (DA) |
|----------------------|----------------|
| Janvier | 1 477 197 263 |
| Février | 1 471 565 780 |
| Mars | 1 466 209 572 |
| Avril | 1 483 174 522 |
| Mai | 1 477 669 671 |
| Juin | 1 472 431 373 |
| Juillet | 1 459 105 430 |
| Aout | 1 452 538 330 |
| Septembre | 1 446 309 152 |
| Octobre | 1 505 761 888 |
| Novembre | 1 499 687 183 |
| décembre | 1 493920 731 |

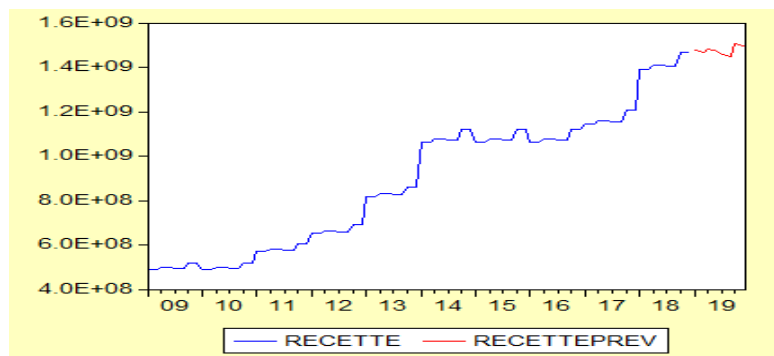


Figure Erreur ! Il n'y a pas de texte répondant à ce style dans ce document.-13:prévisions de recette

4.4.5 Etude de la série DEPEN

4.4.5.1 Étude préliminaire de la série

➤ **Représentation graphique de la série**

Pour illustrer cette première phase de modélisation, nous examinons le graphique de la série nomméeDEPEN

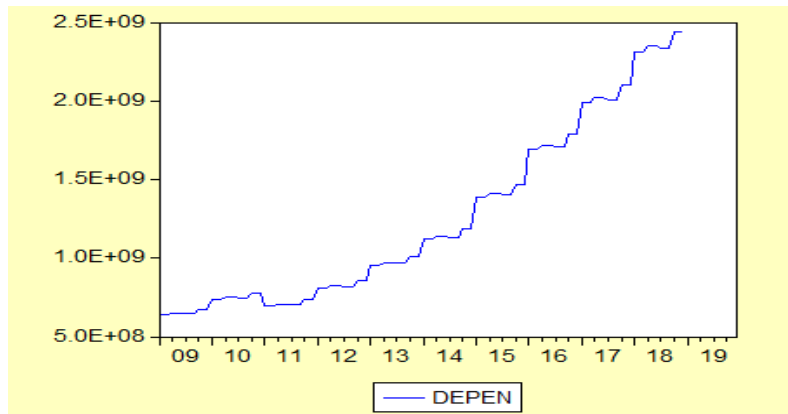


Figure Erreur ! Il n'y a pas de texte répondant à ce style dans ce document.-14: **graphe des données de la série dépense**

L'analyse visuelle du graphe montre à première vue la non stationnarité de la série du fait de la présence d'un mouvement de tendance générale et Un phénomène périodique caractérisé par des mouvements ascendants et descendants, ces mouvements périodiques sont dus au rythme des saisons.

Pour confirmer ou infirmer notre hypothèse nous appliquerons le test de Fisher et le test de DICKEY-FULLER.

➤ **Vérification du facteur saisonnier et tendanciel**

- Test d'influence du facteur lignes (mois)
 - { H_0 : l'effet saisonnier non significatif
 - { H_1 : l'effet saisonnier significatif

- Test d'influence du facteur colonne (années)
 - { H_0 : l'effet de la tendance non significatif
 - { H_1 : l'effet de la tendance significatif

Les résultats du test de Fisher à l'aide de logiciel Excel sont regroupés dans le tableau suivant :

| ANALYSE DE VARIANCE | | | | | | |
|-----------------------|------------------|------------------|--------------------|------------|-------------|------------------------|
| Source des variations | Somme des carrés | Degré de liberté | Moyenne des carrés | F | Probabilité | Valeur critique pour F |
| Lignes | 8,33366E+16 | 11 | 7,57606E+15 | 43,8517516 | 3,969E-33 | 1,886683603 |
| Colonnes | 3,90943E+19 | 9 | 4,34381E+18 | 25142,8569 | 4,479E-162 | 1,975806073 |

Tableau 8: Fisher de la série dépense

✓ Résultat du test d'influence de facteur mois (Saisonnalité) :

$$F_{cal} = 43.8517 > F^{0.05}(11; 119) = 1.8866;$$

On rejette l'hypothèse H_0 , donc la série affectée d'une saisonnalité.

✓ Résultat du test d'influence de facteur année (Tendance) :

$$F_{cal} = 25142.85 > F^{0.05}(9; 119) = 1.9758$$

Donc, nous rejetons l’hypothèse nulle. La série est donc affectée d’une Tendence.

➤ **Dessaisonalisation de la série(DEPEN)**

Une dessaisonalisation de la série (DEPEN) s’impose pour la suite de notre étude. Notons (DEPENSA) la série désaisonnalisée.

➤ **Graphe de série dépense désaisonnalisée (dépensa)**

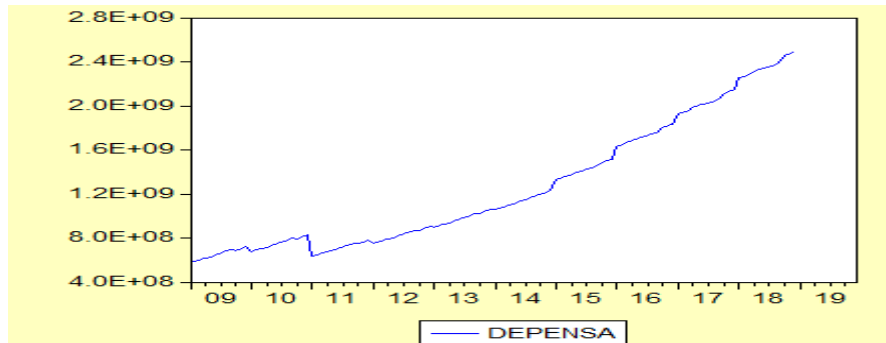


Figure Erreur ! Il n'y a pas de texte répondant à ce style dans ce document.-15: **graphe de la série depense désaisonnalisée**

La représentation graphique de la série monter la neutralité de l’effet saisonnier.

➤ **Vérification de l’élimination du facteur saisonnier**

On applique le test ANOVA sur la série (DEPEN) pour voir si l’effet saisonnier a été éliminé

| ANALYSE DE VARIANCE | | | | | | |
|-----------------------|------------------|------------------|--------------------|-------------|-------------|------------------------|
| Source des variations | Somme des carrés | Degré de liberté | Moyenne des carrés | F | Probabilité | Valeur critique pour F |
| Lignes | 1,7451E+19 | 11 | 1,58645E+18 | 0,900432035 | 0,542912388 | 1,886683603 |
| Colonnes | 6,87434E+19 | 9 | 7,63815E+18 | 4,335233464 | 8,8824E-05 | 1,975806073 |

Tableau 9: test de Fisher

D’après le tableau ci-dessus on a les résultats suivants :

✓ Résultat du test d’influence de facteur mois (Saisonnalité) :

On a $F_c = 0.90043 < F^{0.05}(11,119) = 1.8866$ au seuil de 5%

Dans ce cas on accepte l'hypothèse H_0 , donc l’effet saisonnier est disparu

4.4.5.2 Etude de la stationnarité de la série dessaisonalisé(DEPENSA)

➤ **Test de la racine unitaire (Dickey-Fuller) sur la série (DEPEN)**

Estimation du modèle [3]

| ADF Test Statistic | -0.678597 | 1% Critical Value* | -4.0373 | |
|--------------------|-------------|-----------------------|-------------|--------|
| | | 5% Critical Value | -3.4478 | |
| | | 10% Critical Value | -3.1488 | |
| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
| DEPENSA(-1) | -0.009841 | 0.014502 | -0.678597 | 0.4987 |
| C | 5095413. | 6586451. | 0.773620 | 0.4407 |
| @TREND(2009:01) | 387872.5 | 238436.7 | 1.626731 | 0.1065 |
| R-squared | 0.090124 | Mean dependent var | 16070543 | |
| Adjusted R-squared | 0.074437 | S.D. dependent var | 27356104 | |
| S.E. of regression | 26318270 | Akaike info criterion | 37.03431 | |
| Sum squared resid | 8.03E+16 | Schwarz criterion | 37.10437 | |
| Log likelihood | -2200.542 | F-statistic | 5.744958 | |
| Durbin-Watson stat | 2.177399 | Prob(F-statistic) | 0.004178 | |

D'après le modèle [3] on remarque que :

- la t- statistique supérieurs aux valeurs critiques aux seuils de 1% ,5% et 10% , d'où il existe une racine unitaire donc la série n'est pas stationnaire
- la t-statistique associée à la tendance (trend=1.62) est inférieure à valeur tabulée (2.79) donc la tendance est significativement nulle.

Nous passons alors à l'étude du modèle(2) avec constante seulement.

Estimation du modèle [2]

| ADF Test Statistic | 2.953126 | 1% Critical Value* | -3.4861 | |
|--------------------|-------------|-----------------------|-------------|--------|
| | | 5% Critical Value | -2.8857 | |
| | | 10% Critical Value | -2.5795 | |
| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
| DEPENSA(-1) | 0.012703 | 0.004302 | 2.953126 | 0.0038 |
| C | 196741.7 | 5898801. | 0.033353 | 0.9735 |
| R-squared | 0.069368 | Mean dependent var | 16070543 | |
| Adjusted R-squared | 0.061413 | S.D. dependent var | 27356104 | |
| S.E. of regression | 26502779 | Akaike info criterion | 37.04006 | |
| Sum squared resid | 8.22E+16 | Schwarz criterion | 37.08677 | |
| Log likelihood | -2201.884 | F-statistic | 8.720952 | |
| Durbin-Watson stat | 2.177355 | Prob(F-statistic) | 0.003803 | |

D'après le modèle [2] ,on remarque que la t- statistique supérieure aux valeurs critiques aux seuils de 1% ,5% et 10%,d'où il existe une racine unitaire.

la constante est significativement nulle car la t-statistique (0.033) est inférieure à valeur tabulée (2.54). Nous passons alors au modèle (1) sans constante et sans tendance.

Estimation du modèle [1]

| ADF Test Statistic | 7.274791 | 1% Critical Value* | -2.5830 | |
|--------------------|-------------|-----------------------|-------------|--------|
| | | 5% Critical Value | -1.9426 | |
| | | 10% Critical Value | -1.6171 | |
| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
| DEPENSA(-1) | 0.012834 | 0.001764 | 7.274791 | 0.0000 |
| R-squared | 0.069359 | Mean dependent var | 16070543 | |
| Adjusted R-squared | 0.069359 | S.D. dependent var | 27356104 | |
| S.E. of regression | 26390366 | Akaike info criterion | 37.02326 | |
| Sum squared resid | 8.22E+16 | Schwarz criterion | 37.04662 | |
| Log likelihood | -2201.884 | Durbin-Watson stat | 2.177620 | |

- Nous avons 7.27 est supérieure a toutes les valeurs critiques aux seuils de 1%, 5% et 10% on accepte l'hypothèse H_0 , la série est donc non stationnaire. La série est engendrée par un processus DS sans dérive, elle comporte donc une racine unitaire.
- Pour déterminer l'ordre d'intégration de la série, nous allons appliquer le test ADF à la série en différenciation première.

➤ **Test de la racine unitaire(Dickey-Fuller) sur la série différenciée**

On commence par le modèle [3] : test sur la tendance

| ADF Test Statistic | -11.82711 | 1% Critical Value* | -4.0380 | |
|--------------------|-------------|-----------------------|-------------|--------|
| | | 5% Critical Value | -3.4481 | |
| | | 10% Critical Value | -3.1489 | |
| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
| D(DEPENSA(-1)) | -1.097745 | 0.092816 | -11.82711 | 0.0000 |
| C | 1766802. | 4942420. | 0.357477 | 0.7214 |
| @TREND(2009:01) | 262450.6 | 74542.03 | 3.520841 | 0.0006 |
| R-squared | 0.548808 | Mean dependent var | 670.4842 | |
| Adjusted R-squared | 0.540961 | S.D. dependent var | 38859008 | |
| S.E. of regression | 26327909 | Akaike info criterion | 37.03525 | |
| Sum squared resid | 7.97E+16 | Schwarz criterion | 37.10569 | |
| Log likelihood | -2182.080 | F-statistic | 69.94022 | |
| Durbin-Watson stat | 2.021006 | Prob(F-statistic) | 0.000000 | |

- D'après le modèle [3] on n'observe que la t-satistic (3.52) associée à la tendance qui est supérieure à valeur critique (2.79), donc la tendance est significativement différente de 0. D'où la série possède une tendance déterministe après la première différenciation.

Pour éliminer l'effet tendanciel on fait la régression sur la tendance déterministe.

4.4.5.3 Estimation du modèle (estimation de la tendance)

On passe à la modélisation par la méthode de Box-Jenkins.

Le modèle s'écrit : Soit avec une tendance linéaire : $R_t = \alpha t + C$

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|--------------------|-------------|-----------------------|-------------|--------|
| C | 2075621. | 4844547. | 0.428445 | 0.6691 |
| TEMPS | 233248.7 | 70071.15 | 3.328741 | 0.0012 |
| R-squared | 0.086512 | Mean dependent var | 16070543 | |
| Adjusted R-squared | 0.078705 | S.D. dependent var | 27356104 | |
| S.E. of regression | 26257521 | Akaike info criterion | 37.02147 | |
| Sum squared resid | 8.07E+16 | Schwarz criterion | 37.06817 | |
| Log likelihood | -2200.777 | F-statistic | 11.08052 | |
| Durbin-Watson stat | 2.190240 | Prob(F-statistic) | 0.001167 | |

Figure Erreur ! Il n'y a pas de texte répondant à ce style dans ce document.-16:estimation de la Tendance linéaire

$$R_t = 233248.7 * t \Rightarrow Z_t = \text{depensa}_t - \text{depensa}_{t-1} - 233248.7 * t$$

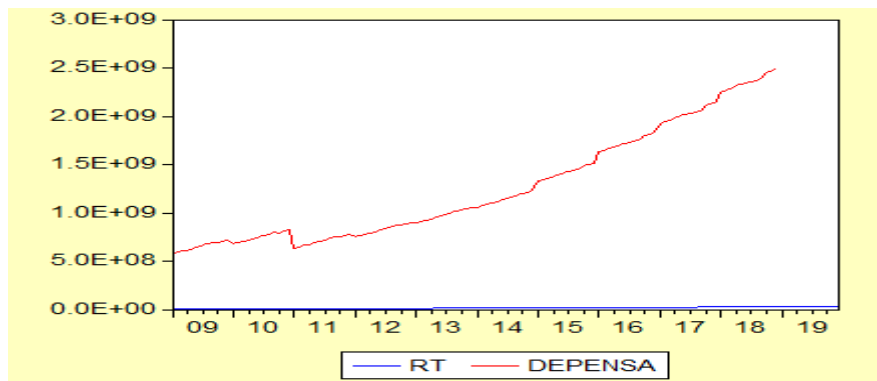


Figure Erreur ! Il n'y a pas de texte répondant à ce style dans ce document.-17:graphe de la tendance linéaire

Soit avec une Tendence quadratique : $Q_t = c + \alpha t + \beta^2 t$

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|--------------------|-------------|-----------------------|-------------|----------|
| C | 3316011. | 7374150. | 0.449680 | 0.6538 |
| TEMPS | 171741.8 | 283684.7 | 0.605397 | 0.5461 |
| TEMPS*2 | 512.5576 | 2290.179 | 0.223807 | 0.8233 |
| R-squared | 0.086906 | Mean dependent var | | 16070543 |
| Adjusted R-squared | 0.071163 | S.D. dependent var | | 27356104 |
| S.E. of regression | 26364765 | Akaike info criterion | | 37.03784 |
| Sum squared resid | 8.06E+16 | Schwarz criterion | | 37.10790 |
| Log likelihood | -2200.752 | F-statistic | | 5.520322 |
| Durbin-Watson stat | 2.191188 | Prob(F-statistic) | | 0.005127 |

Figure Erreur ! Il n'y a pas de texte répondant à ce style dans ce document.-18:estimation de la Tendence quadratique

$$Q_t = 3316011 + 171741.8t + 512.55t^2$$

$$\Rightarrow Z_t = \text{depensa}_t - \text{depensa}_{t-1} - (3316011 + 171741.8t + 512.55t^2)$$

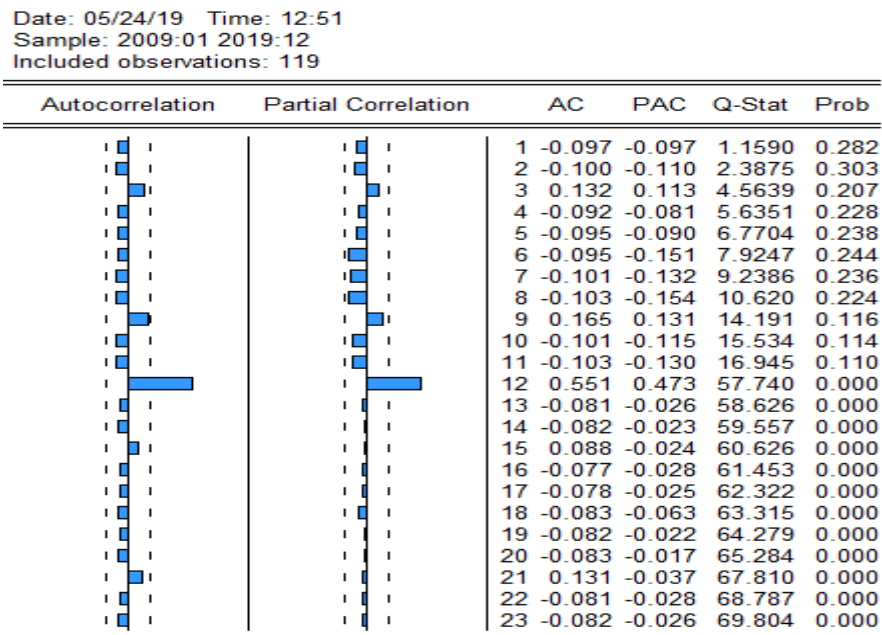
➤ **Choix de la tendance optimal**

| | R ² | SSR | SC | AIC |
|--------------------------|----------------|-----------------------|-------|-------|
| Forme linéaire | 0.086 | 8.07 ^E +16 | 37.06 | 37.02 |
| Forme quadratique | 0.086 | 8.06 ^E +16 | 37.10 | 37.03 |

Tableau 10:choix de la tendance

On rejette l'estimation de la tendance cubique et quadratique et on retient le modèle avec une tendance linéaire

➤ **Corrélogramme de (Z)**



Les modèles proposés sont ARMA(12,12), ARMA(12,0), ARMA(0,12) .

➤ **Choix de modèle optimal**

| Modèle | R ² | SSR | SC | AIC |
|-----------------------|----------------|-----------------------|-------|-------|
| ARIMA(12,1,12) | 0.86 | 1.05 ^E +16 | 35.14 | 35.09 |
| ARIMA (12,1,0) | 0.36 | 4.89 ^E +16 | 36.63 | 36.61 |
| ARIMA (0,1,12) | 0.51 | 3.93 ^E +16 | 36.28 | 36.30 |

Tableau 11:choix du modèle optimal

Parmi les modèles proposés, on choisit ceux qui minimise les critères **AIC**, **SC**, **SSR** et maximiser le critère **R²**. On opte pour le modèle **ARIMA(12,1,12)**

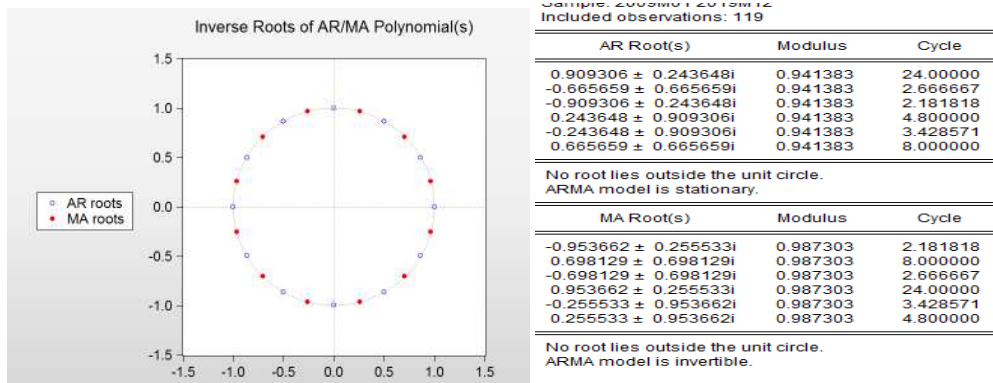
4.4.5.4 Estimation du modèle optimal

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|--------------------|-------------|-----------------------|-------------|--------|
| AR(12) | 0.429250 | 0.053915 | 7.961635 | 0.0000 |
| MA(12) | 0.952644 | 0.008270 | 115.1887 | 0.0000 |
| R-squared | 0.862463 | Mean dependent var | 1551813. | |
| Adjusted R-squared | 0.861153 | S.D. dependent var | 26892457 | |
| S.E. of regression | 10020712 | Akaike info criterion | 35.09672 | |
| Sum squared resid | 1.05E+16 | Schwarz criterion | 35.14668 | |
| Log likelihood | -1875.675 | Durbin-Watson stat | 2.108489 | |

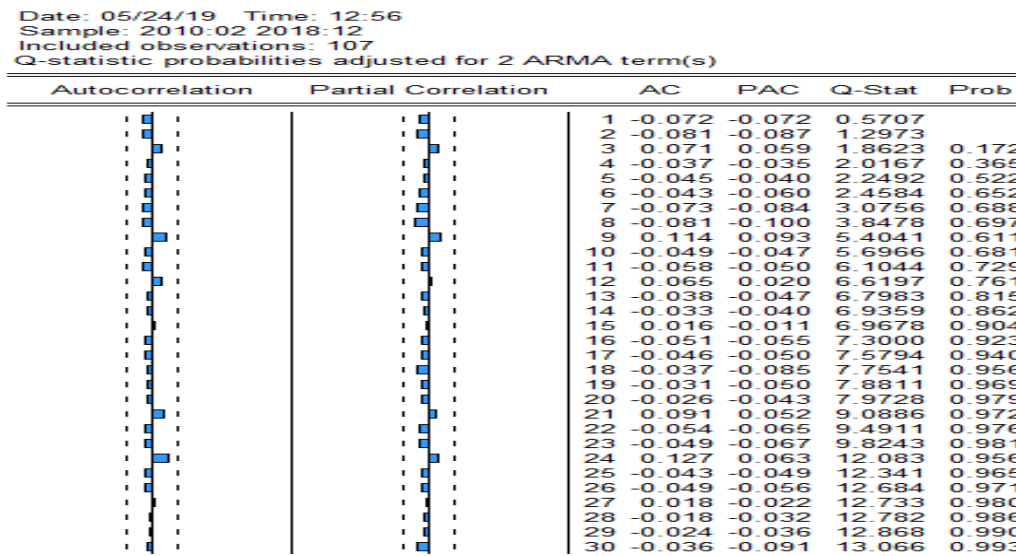
4.4.5.5 validation du modèle

➤ Test sur les paramètres

les coefficients du modèle sont tous significativement différents de zéro au seuil de 5%, et les inverses des racines sont distincts et inférieurs en module à 1 par conséquent, les conditions de stationnarité et d'invisibilité sont variées.



➤ Test sur les résidus



Le Corrélogramme des résidus du modèle retenu ARIMA (12,1,12) ne fait apparaitre aucun terme en dehors de l'intervalle de confiance au seuil de 5% et la statistique de **Box- Ljung Q=13.066** est inférieure à la valeur tabulée de khi-deux qu'est de **42.56** au seuil de 5%. Donc les résidus forment un bruit blanc.

- La statistique DW=2.10, on accepte l'hypothèse de non corrélation des Résidus.

➤ **Test de normalité sur les résidus**

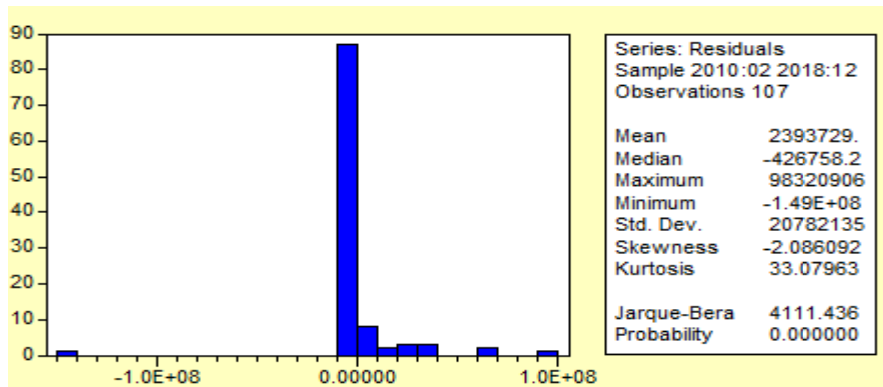


Figure Erreur ! Il n'y a pas de texte répondant à ce style dans ce document.-19 :normalité des résidus

- **Test skeweness** : $\vartheta_1 = \frac{|-2.08|}{\sqrt{\frac{6}{119}}} = 9.26 > 1.96$
- **Test Kurtosis**: $\vartheta_2 = \frac{|33.07-3|}{\sqrt{\frac{24}{119}}} = 66.95 > 1.96$
- **Statistique de Jarque-Berra** : $JB = 4111.43 > \chi_2^2 = 5.99$

D’où les résidus sont bruit blanc non gaussien

- **Test d’heteroscédasticité(l’existence d’effet ARCH)**

| Heteroskedasticity Test: ARCH | | | | |
|--|-------------|-----------------------|-------------|--------|
| F-statistic | 0.055382 | Prob. F(1,116) | 0.8144 | |
| Obs*R-squared | 0.056310 | Prob. Chi-Square(1) | 0.8124 | |
| Test Equation: | | | | |
| Dependent Variable: RESID^2 | | | | |
| Method: Least Squares | | | | |
| Date: 05/28/19 Time: 00:50 | | | | |
| Sample (adjusted): 2009M03 2018M12 | | | | |
| Included observations: 118 after adjustments | | | | |
| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
| C | 2.90E+14 | 1.78E+14 | 1.628070 | 0.1062 |
| RESID^2(-1) | -0.021845 | 0.092827 | -0.235334 | 0.8144 |
| R-squared | 0.000477 | Mean dependent var | 2.83E+14 | |
| Adjusted R-squared | -0.008139 | S.D. dependent var | 1.90E+15 | |
| S.E. of regression | 1.91E+15 | Akaike info criterion | 73.22792 | |
| Sum squared resid | 4.24E+32 | Schwarz criterion | 73.27488 | |
| Log likelihood | -4318.447 | Hannan-Quinn criter. | 73.24699 | |
| F-statistic | 0.055382 | Durbin-Watson stat | 2.000951 | |
| Prob(F-statistic) | 0.814365 | | | |

A partir de la figure on remarque que la probabilité de RESID2 est supérieur a 0.05 donc on accepte l’hypothèse nulle d’hémoscédasticité il n’existe pas un effet ARCH.

Conclusion: D’après les résultats on peut valider le modèle de la série (DEPENSA) s’écrit sous la forme :

$$(1 - B) \text{DEPENSA}_t = 233248.7 * \text{temps} + Z_t$$

$$Z_t = 0.429Z_{t-12} + 0.952\varepsilon_{t-12} + e_t \quad \text{tq } e_t \sim \text{BB}$$

4.4.5.6 Prévision

Rappelons que la série sous étude "dépense" comprend 120 observation de Janvier 2009 à décembre 2018, donc l'instant t est décembre 2018.

On souhaitera prédire les 12 mois de l'année 2019, soit un horizon h = 12:

On remplaçant t par t+h dans l'équation précédente on obtient :

$$Z_{t+h} = 0.429Z_{t+h-12} + 0.952\varepsilon_{t+h-12} + \varepsilon_{t+h}$$

Avec h=1,.....,12. On obtient les prévisions suivantes :

| Mois de l'année 2019 | Prévision |
|----------------------|------------|
| janvier | 2512790233 |
| Février | 2506973355 |
| Mars | 2726441556 |
| Avril | 2718215972 |
| Mai | 2699284051 |
| Juin | 2745389704 |
| Juillet | 2744756394 |
| Aout | 2736465036 |
| Septembre | 2721834538 |
| Octobre | 2713728050 |
| Novembre | 2709590606 |
| décembre | 2814673867 |

Le graphe des prévisions est donné comme suit :

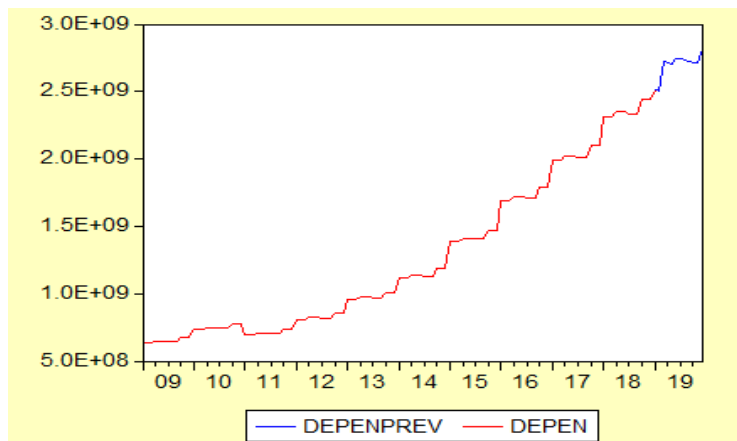


Figure Erreur ! Il n'y a pas de texte répondant à ce style dans ce document.-20:graphe des prévisions de série dépense

La représentation graphique des 12 valeurs prédites sont en harmonie avec l'allure générale de la série étudiée puisque le phénomène de périodicité est reproduit

Chapitre 5
Modèle VAR

Introduction

Les processus autorégressifs vectoriels VAR (p) (vecteur autorégressive) constituent une généralisation des processus autorégressifs uni-variés présentés auparavant.

Nous parlons d'approche multi-variée lorsque la description et l'analyse portent sur plusieurs variables considérées simultanément. En général, il est possible de faire cette étude, lorsque les variables en jeu sont en relation de dépendance, et sont toutes stationnaires avec un nombre de décalage identique.

Depuis les travaux initiaux de Sims (1980), les techniques économétriques basées sur les modèles VAR ont connu de nombreux développements, leur popularité est due à leur caractère flexible et leur facilité d'utilisation pour produire les modèles ayant des caractéristiques descriptives utilisées pour tester des hypothèses économiques et modéliser les interactions existantes entrées variables.

Dans le chapitre présent nous développons les modèles de séries chronologiques multi-variées, en particulier le modèle VAR.

La modélisation de VAR(p)

5.1 Définition

Un VAR est une forme de modèle dynamique linéaire. Il comporte plusieurs équations. Chaque équation dénote une relation linéaire où une variable est exprimée comme une combinaison de ses propres valeurs passées et des valeurs passées des autres variables. L'ensemble de ces variables du modèle sont endogènes. Chaque équation est complétée par un terme d'erreur qui est soit endogène, soit exogène.

5.2 Présentation du modèle (VAR)

Un modèle VAR d'ordre(p) se définit comme la représentation autorégressive vectorielle d'une variable aléatoire de dimension (n, 1):

$$Y_t = \mu + \varphi Y_{t-1} + \varphi Y_{t-2} + \dots + \varphi_p Y_{t-p} + V_t$$

$$Y_t = \begin{bmatrix} Y_{1,t} \\ Y_{2,t} \\ \vdots \\ Y_{n,t} \end{bmatrix} \quad \mu = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{bmatrix} \quad \varphi = \begin{pmatrix} \varphi_{11} & \varphi_{12} & \varphi_{1n} \\ \varphi_{21} & \varphi_{22} & \varphi_{2n} \\ \vdots & \vdots & \vdots \\ \varphi_{n1} & \varphi_{n2} & \varphi_{nn} \end{pmatrix} \quad V = \begin{bmatrix} V_{1,t} \\ V_{2,t} \\ \vdots \\ V_{n,t} \end{bmatrix}$$

Les variables $Y_{1,t}, Y_{2,t}, \dots, Y_{n,t}$ sont stationnaires. Les perturbations $V_{1,t}, V_{2,t}, \dots, V_{n,t}$ sont des bruits blancs de variances constantes et non auto-corrélées.

5.3 La stationnarité

Un processus vectoriel $\{X_t \ t \in \mathbb{Z}\}$ de dimension $(n,1)$ satisfait une représentation VAR(p) tel que $t \in \mathbb{Z}$;

$$\varphi(L) X_t = \varphi_0 X_t - \varphi_1 X_{t-1} - \varphi_2 X_{t-2} \dots - \varphi_p X_{t-p} = \mu + V_t$$

est stationnaire si et seulement si les racines du déterminant noté :

$$\det (I - \varphi_1 \lambda_i - \varphi_2 \lambda_i^2 - \dots - \varphi_p \lambda_i^p = 0) \text{ du polynôme matriciel } \varphi(L) \text{ noté } \lambda_i \text{ avec}$$

$i \in [1, n]$ sont toutes supérieures à l'unité en module.

$$\det [\varphi_1(\lambda_i)] = \varphi_1(\lambda_i) \text{ avec } |\lambda_i| > 1 \quad \forall i \in [1, n]$$

L'ensemble de ces considérations implique qu'un VAR stationnaire ne doit pas avoir de tendance, de saisonnalité, ni de racine unitaire.

5.4 Caractéristiques des processus VAR(p)

Étudions les principales caractéristiques des processus

VAR. Considérons un processus VAR (1)

$$X_t = \varphi_0 X_t + \varphi_1 X_{t-1} + V_t \quad \text{Où : } V_t \sim \text{BB}(0, \Sigma)$$

➤ L'espérance

$$\text{On a : } E[X_t] = E[\varphi_0 + \varphi_1 X_{t-1} + \varepsilon_t]$$

Le processus étant stationnaire, on a : $E[X_t] - E[X_{t-1}] = 0$

On peut donc écrire (sachant que $E[\varepsilon_t] = 0$) : $E[X_t] = \varphi_0 + \varphi_1 E[X_t]$

$$\text{D'où } E[X_t] = (1 - \varphi_1)^{-1} \varphi_0$$

➤ Fonction d'autocovariance

Considérons le processus centré : $Y_t = X_t - E[X_t]$ soit $Y_t = \varphi Y_{t-1} + \varepsilon_t$

covariance Γ est donnée par :

$$\Gamma(0) = E[Y_t Y_t'] = E[\varphi Y_{t-1} Y_t' + \varepsilon_t Y_t'] \quad (*)$$

$$\text{Or } E[\varepsilon_t Y_t'] = E[\varepsilon_t (\varphi_1 Y_{t-1}' + \varepsilon_t')] = \varphi_1 E[\varepsilon_t Y_{t-1}'] + E[\varepsilon_t \varepsilon_t']$$

Comme ε_t est BB alors on a donc : $E[\varepsilon_t Y_t'] = E[\varepsilon_t \varepsilon_t'] = \Sigma_\varepsilon$

On remplaçant dans (*) on aura $\Gamma(0) = \varphi_1 E[Y_{t-1} Y_t'] + \Sigma$

On remarque que $\Gamma(0) = E[Y_{t-1} Y_t'] = \Gamma(1)$, on déduit $\Gamma(0) = \varphi_1 \Gamma(1)'$

On calcule la matrice d'auto covariance d'ordre 1 :

$$\Gamma(1) = E[Y_{t-1}' Y_t] = E[(\varphi_1 Y_{t-1} + \varepsilon_t) + Y_{t-1}'] = \varphi_1 E[Y_{t-1}' Y_{t-1}] = \varphi_1 \Gamma(0)$$

On en déduit la formule de récurrence suivante pour la matrice d'auto covariance d'ordre h

D'un processus VAR(1) : $\Gamma(h) = \varphi_1 \Gamma(h-1) \quad \forall h \geq 1$

5.5 Représentation canonique

Considérons un processus VAR centré, c'est à dire avec $\varphi_0 = 0$

$$\Phi(L) X_t = \varepsilon_t$$

On peut écrire : $X_t = \Phi^{-1}(L) \varepsilon_t = \frac{\tilde{\phi}(L)}{\det \Phi(L)} \varepsilon_t$

Définition

Si toutes les racines du déterminant de $\phi(L)$ sont du module supérieur à 1, Alors l'équation $\Phi(L) X_t = \varepsilon_t$, définit un unique processus VAR(p) stationnaire. On dit que X_t est en représentation canonique et ε_t est appelé résidu du processus.

Remarque 1

Si au moins une des racines de déterminant est égale à 1, le processus n'est plus stationnaire et on ne peut pas se ramener à une représentation canonique.

Remarque 2

En représentation canonique, la prévision s'écrit : $E[X_{t+1}/X_t] = \sum_{i=1}^p \phi_i X_{t+1-i}$

5.6 Estimation des paramètres

La méthode du maximum de vraisemblance

Considérons un processus VAR (p) $X_t = \varphi_1 X_{t-1} + \dots + \varphi_p X_{t-p} + \varepsilon_t$

Où ε_t est un bruit blanc de matrice de variance covariance Σ .

On écrit la vraisemblance conditionnellement à toutes valeurs passées du processus :

$$L(X_1 \dots X_t) = \prod_{t=1}^T L(X_t | X_{t-1})$$

Où X_{t-1} désigne tout le passé de X_t jusqu'à la date $(t - 1)$ incluse la vraisemblance s'écrit alors :

$$L(X_1 \dots X_T) = \prod_{t=1}^T \frac{1}{(\sqrt{2\pi})^n \sqrt{\det \Sigma}} \times \exp\left[-\frac{1}{2} \sum_{t=1}^T (X_t - \varphi_1 X_{t-1} \dots - \varphi_p X_{t-p})' \Sigma (X_t - \varphi_1 X_{t-1} \dots - \varphi_p X_{t-p})\right]$$

On en déduit l'expression de la log-vraisemblance :

$$\text{Log } L((X_1 \dots X_T)) = -\frac{nT}{2} \text{Log } 2\pi - \text{Log det } \Sigma - \frac{1}{2} \sum_{t=1}^T \varepsilon_t' \Sigma^{-1} \varepsilon_t$$

On maximise ensuite cette expression afin d'obtenir les estimations $\varphi_1 \dots \varphi_p$ et de Σ .

Remarque

On peut également estimer ses paramètres par la méthode des **MCO**.

5.7 Validation

5.7.1 Tests de spécification

➤ Test du rapport de maximum de vraisemblance :

On peut effectuer des tests sur l'ordre p du VAR. Considérons le test suivant :

$$\begin{cases} H_0 : \Phi_{p+1} = 0 : \text{Processus VAR}(p) \\ H_1 : \Phi_{p+1} \neq 0 : \text{Processus VAR}(p+1) \end{cases}$$

La matrice d'information de Fisher est difficile à calculer, ce qui explique que l'on utilise un test du rapport du maximum de vraisemblance. La technique consiste à estimer un modèle contraint VAR (p) et un modèle non contraint VAR ($p+1$) et à effectuer le rapport des log-vraisemblances. Rappelons que la log-vraisemblance d'un processus VAR s'écrit :

$$\text{Log } L((X_1 \dots X_T)) = -\frac{nT}{2} \text{Log } 2\pi - \text{Log det } \Sigma - \frac{1}{2} \sum_{t=1}^T \varepsilon_t' \Sigma^{-1} \varepsilon_t$$

Avec $\sum_{t=1}^T \varepsilon_t' \Sigma^{-1} \varepsilon_t$ Est un scalaire, on a donc, on notant Tr la trace :

$$\begin{aligned} \sum_{t=1}^T \varepsilon_t' \Sigma^{-1} \varepsilon_t &= \text{Tr} \left(\sum_{t=1}^T \varepsilon_t' \Sigma^{-1} \varepsilon_t \right) \\ &= \text{Tr} \left(\sum_{t=1}^T \Sigma^{-1} \varepsilon_t \varepsilon_t' \right) \\ &= \text{Tr} \left(\Sigma^{-1} \frac{1}{T} \sum_{t=1}^T \varepsilon_t \varepsilon_t' \right) \\ &= \text{Tr} \left(\text{Tr} \Sigma^{-1} \Sigma \right) = \text{Tr} (T I_n) = nT \end{aligned}$$

Soient $\text{Log } L^c$ le log vraisemblance estimée du modèle contraint :

$$\text{Log } L^c = -\frac{nT}{2} \text{Log } 2\pi - \frac{T}{2} \text{Log det } \hat{\Sigma}^c - \frac{1}{2} nT$$

Soient $\log L^{nc}$ le log vraisemblance estimée du modèle non contraint :

$$\text{Log } L^{nc} = -\frac{NT}{2} \text{Log } 2\pi - \frac{T}{2} \text{Log } \det \hat{\Sigma}^{nc} - \frac{1}{2} nT$$

Où $\hat{\Sigma}^c$ (respectivement $\hat{\Sigma}^{nc}$) désigne l'estimateur de la matrice de variance covariance des résidus du modèle contraint (respectivement non contraint).

On calcule la statistique de test $\xi = T \text{RMV}$ où RMV désigne le rapport du maximum de vraisemblance.

$$\xi = T \log \left(\frac{\det \hat{\Sigma}^c}{\det \hat{\Sigma}^{nc}} \right)$$

Sous l'hypothèse nulle, cette statistique suit une loi de khi-deux à r degrés de liberté où r désigne le nombre de contraintes. Si l'on accepte l'hypothèse nulle, on peut effectuer un deuxième test :

$$\begin{cases} H_0 : \Phi_p = \text{processus VAR}(P) \\ H_1 : \Phi_p \neq \text{Processus VAR}(P+1) \end{cases}$$

Ce test s'effectue de la même façon que précédemment. On a ainsi une séquence de tests emboîtés dont le but est de déterminer l'ordre p du processus VAR.

Remarque :

Dans le cas d'un processus AR, en plus des tests sur les paramètres, on effectue des tests sur les résidus afin de valider le processus. Dans le cas des processus VAR, ces tests ne sont pas très puissants et l'on préfère réaliser un graphe des résidus. Notons cependant qu'il convient d'examiner attentivement les résidus surtout lors d'utilisation des modèles VAR pour l'analyse de réponses pulsionnelles où l'absence de corrélation des résidus est cruciale pour l'interprétation.

Afin de déterminer l'ordre p du VAR, on peut également utiliser des Critères d'information. Ainsi, on estime un certain nombre de modèles VAR pour un ordre p allant de 0 à h , où h est le retard maximum. On retient le retard p qui minimise les Critères **AIC**, **SIC**. Définis comme :

$$\begin{cases} \text{AIC} = \log \det \hat{\Sigma} + \frac{2pn^2}{T} \\ \text{SIC} = \log \det \hat{\Sigma} + n^1 p \frac{\log T}{T} \end{cases}$$

Où n est le nombre de variables du système, T est le nombre d'observations et $\hat{\Sigma}$ est un estimateur de la matrice de variance covariance des résidus.

Remarque

Le Critère **SIC** conduit à des estimateurs convergents de **p**, le critère **AIC** donnant des estimateurs efficaces de **p**.

5.8 Prévisions

Considérons un processus VAR (p) : $X_t = \widehat{\varphi}_1 X_{t-1} - \widehat{\varphi}_2 X_{t-2} \dots - \widehat{\varphi}_p X_{t-p} + \varepsilon_t$

On suppose que **p** a été choisi, que les φ_i ont été estimés et que la matrice de variance-covariance associée à ε_t a été estimée.

Afin de réaliser des prévisions, il est nécessaire de vérifier que le modèle est bien en représentation canonique. Pour cela, on calcule le déterminant du polynôme et on regarde si les racines sont bien à l'extérieur du disque unité. Si tel est le cas, alors la prévision en (T + 1) du processus est :

$$E[X_{T+1} / X_t] = \widehat{\varphi}_1 X_{t-1} - \widehat{\varphi}_2 X_{t-2} \dots - \widehat{\varphi}_p X_{T-p+1}$$

Où X_t désigne le passé jusqu'à la date T incluse.

5.9 Application du modèle VAR

Suite à l'application de la méthodologie de Box & Jenkins sur les deux séries $DBUDGETSA_t$, $DRETRAITESA_t$, nous pouvons donc appliquer la modélisation multivariée VAR.

5.9.1 Étude multivariée des séries ($DBUDGETSA_t$, $DRETRAITESA_t$)

Dans la suite de notre étude nous considérons le vecteur $Z_t = (DBUDGETSA_t, DRETRAITESA_t)'$ et $\varepsilon_t = (e_{1t}, e_{2t})'$ vecteur résiduel associée à Z_t de matrice de variance-covariance.

➤ Représentation graphique

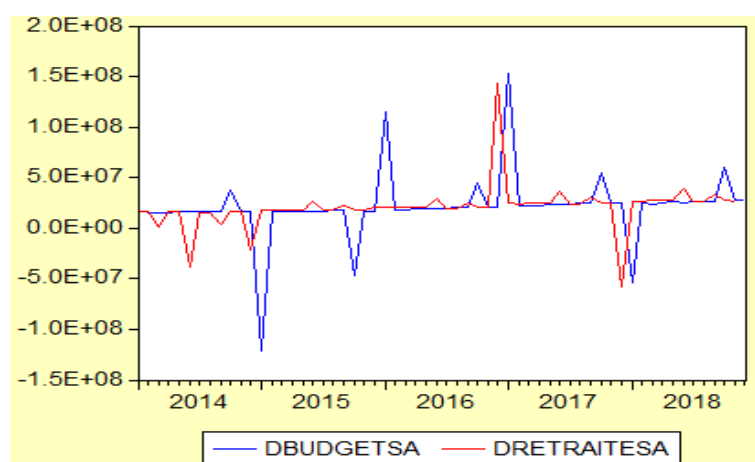


Figure 5-1: graphes des séries ($DBUDGETSA_t$, $DRETRAITESA_t$)

5.9.2 Identification de modèle VAR(p)

A l'instar de la méthodologie de Box & Jenkins l'identification est une étape cruciale, pour mener à bien cette démarche on a recours aux critères d'information LR, FPE, AIC, SC et HQ pour déterminer le nombre de décalages p .

A cette fin, nous avons estimé divers processus VAR pour des ordres de retards p allant de 1 à 4. Pour chaque modèle, nous avons calculé les critères d'informations précédemment cités.

Le tableau ci-dessous reporte les résultats obtenus. Nous optons à retenir un modèle VAR(1).

| Lag | LogL | LR | FPE | AIC | SC | HQ |
|-----|-----------|----------|-----------|-----------|-----------|-----------|
| 1 | -1992.964 | NA | 4.53E+29* | 73.96164* | 74.10897* | 74.01846* |
| 2 | -1990.803 | 4.001899 | 4.85E+29 | 74.02975 | 74.32442 | 74.14339 |
| 3 | -1987.276 | 6.270219 | 4.95E+29 | 74.04727 | 74.48927 | 74.21773 |
| 4 | -1985.961 | 2.241725 | 5.48E+29 | 74.14669 | 74.73601 | 74.37397 |

* indicates lag order selected by the criterion

LR: sequential modified LR test statistic (each test at 5% level)

FPE: Final prediction error

AIC: Akaike information criterion

SC: Schwarz information criterion

HQ: Hannan-Quinn information criterion

Les comparaisons des modèles suivants les critères AIC, SC et HQ nous pousse à choisir le modèle VAR(1).

5.9.3 Estimation du modèle VAR(1)

5.9.3.1 Estimation du modèle VAR(1) avec constante

Le modèle VAR(1) avec constante s'écrit sous la forme suivante :

$$Z_t = \phi_0 + \phi_1 Z_{t-1} + \varepsilon_t$$

Où $\phi_0 = (a_1^0, a_2^0)$ représente l'estimation de la constante et $\phi_{p(p=1)}$ sont des matrices carrées d'ordre 2. On effectue l'estimation nous obtenons le tableau qui sera présenté comme suit :

-Le tableau contient 3 colonnes au nombre des variables du modèle VAR.

-Ce tableau peut être décomposé en 3 blocs, chaque bloc est associé à une série.

-Chaque bloc contient p=1 lignes.

-La ligne (i=1) d'un bloc précis correspond à la série associée à l'instant t-i.

-Chaque ligne contient les coefficients de retard i (donnés en haut), ainsi que les t-statistiques associées (données en bas entre crochets).

| DBUDGETSA DRETRAITE... | | |
|------------------------|--------------------------------------|------------------------------------|
| DBUDGETSA(-1) | -0.000755 (0.09605) [-0.00787] | 0.027640 (0.09328) [0.29633] |
| DRETRAITESA(-1) | 1.033127 (0.13995) [7.38197] | 0.058165 (0.13590) [0.42798] |
| C | 6191.025 (4687193) [0.00132] | 19151645 (4551622) [4.20765] |

A la lecture du tableau on constate est non significative aux seuils 1%, 5% et 10% puisque les t-statistiques (données par valeurs entre crochet) sont inférieures aux différentes valeurs critiques, donc on réestime le modèle sans constante.

5.9.3.2 Estimation du modèle VAR(1) sans constante

| DBUDGETSA DRETRAITE... | | |
|------------------------|--------------------------------------|-------------------------------------|
| DBUDGETSA(-1) | -0.000704 (0.08688) [-0.00810] | 0.187884 (0.09722) [1.93256] |
| DRETRAITESA(-1) | 1.033238 (0.11080) [9.32491] | 0.402014 (0.12399) [3.24232] |

A partir de la table d'estimation on obtient l'équation suivante :

$$Z_t = \begin{pmatrix} -0.0007 & 1.03 \\ 0.18 & 0.40 \end{pmatrix} Z_{t-1} + \varepsilon_t$$

Bien entendu cette écriture du modèle subira un remaniement après l'épreuve des tests suite à l'étape de validation.

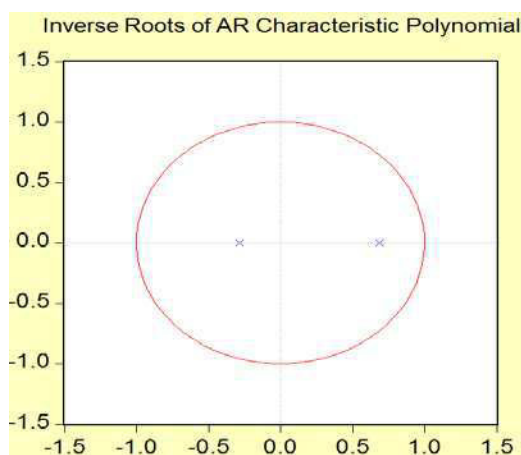
Nous obtenons également la matrice suivante de variance-covariance :

$$\hat{\Sigma} = \begin{pmatrix} 1.05E + 15 & 3.18E + 13 \\ 3.18E + 13 & 4.97E + 14 \end{pmatrix}$$

5.9.4 Validation

➤ **Test sur les racines**

Les racines des polynômes autorégressifs des séries (DRETRAITESA_t ; DBUDGETSA_t) sont supérieures en module à 1, et leurs inverses calculés par Eviews sont tous inférieures à 1, ainsi les conditions de stationnarité et d'inversibilité sont vérifiées. Voir la figure ci-dessous.



| Root | Modulus |
|-----------|----------|
| 0.685087 | 0.685087 |
| -0.283777 | 0.283777 |

No root lies outside the unit circle.
VAR satisfies the stability condition.

➤ **Test sur les résidus**

De la même façon que la méthodologie de Box & Jenkins, il convient de vérifier si les résidus forment un bruit blanc, une observation des corrélogrammes des résidus des deux séries s'impose.

Corrélogrammes des résidus des séries

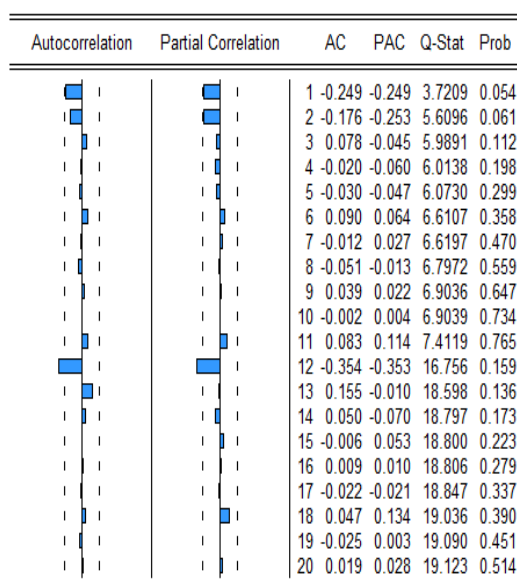


Figure 5-2: corrélogramme des résidus (DRETRAITESA)

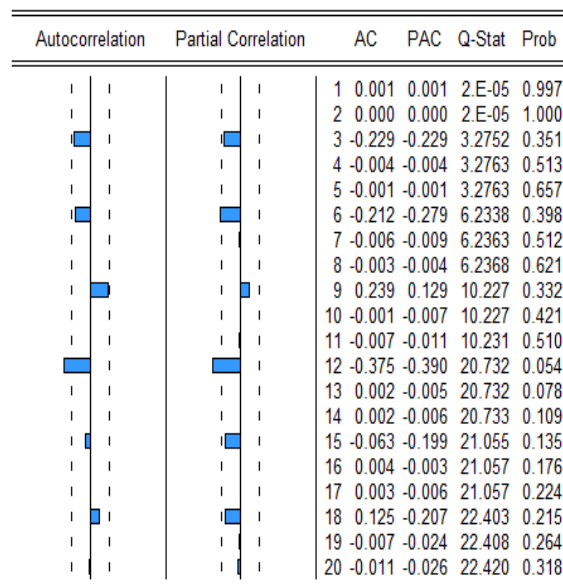


Figure 5-3: corrélogramme des résidus (DBUDGETSA)

L'observation du corrélogramme des résidus des deux séries : (DBUDGETSA_t ; DRETRAITESA_t) nous montre que les résidus forment un bruit blanc car toutes les probabilités sont supérieures à 0.05 .

Globalement, l'analyse des deux corrélogrammes nous montre que tous les termes sont à l'intérieur de l'intervalle de confiance de là on déduit une absence de corrélation, donc les résidus des séries une à une forment un processus bruit blanc.

Etant donné que nous sommes dans l'obligation d'avoir un vecteur résiduel de type bruit blanc, l'observation des corrélogrammes simples ne suffit pas, c'est ainsi qu'on examinera les corrélogrammes croisés des résidus.

➤ Corrélogrammes des résidus croisés entre (DBUDGETSA) et (DRETRAITESA)

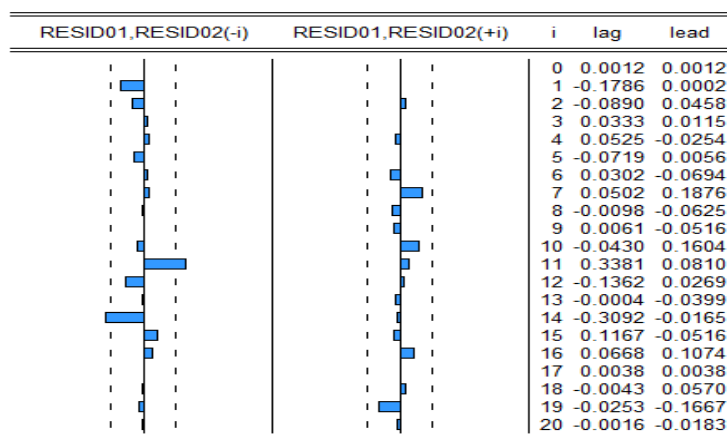


Figure 5-4: corrélogramme des résidus croisés (DBUDGETSA_t, DRETRAITESA_t)

En observant la Figure (5-4) on remarque que tous les pics sont à l'intérieur de l'intervalle de confiance sauf les pics (11,14). On fait le test de **Portemanteau** qui nous donne les Q-stat et on les compare avec la valeur de khi-deux.

VAR Residual Portmanteau Tests for Autocorrelations
H0: no residual autocorrelations up to lag h
Date: 06/10/19 Time: 12:43
Sample: 2014:01 2018:12
Included observations: 57

| Lags | Q-Stat | Prob. | Adj Q-Stat | Prob. | df |
|------|----------|--------|------------|--------|-----|
| 1 | 5.342804 | NA* | 5.438211 | NA* | NA* |
| 2 | 7.673655 | 0.1043 | 7.853821 | 0.0971 | 4 |
| 3 | 11.09033 | 0.1966 | 11.46031 | 0.1770 | 8 |
| 4 | 11.30754 | 0.5028 | 11.69391 | 0.4706 | 12 |
| 5 | 11.65579 | 0.7673 | 12.07565 | 0.7388 | 16 |
| 6 | 15.00305 | 0.7762 | 15.81670 | 0.7279 | 20 |
| 7 | 17.16316 | 0.8416 | 18.27923 | 0.7893 | 24 |
| 8 | 17.53860 | 0.9372 | 18.71596 | 0.9067 | 28 |
| 9 | 21.02702 | 0.9311 | 22.85847 | 0.8827 | 32 |
| 10 | 22.59918 | 0.9601 | 24.76513 | 0.9212 | 36 |
| 11 | 29.88141 | 0.8785 | 33.78876 | 0.7449 | 40 |
| 12 | 46.10528 | 0.3852 | 54.33899 | 0.1366 | 44 |
| 13 | 47.57068 | 0.4903 | 56.23736 | 0.1938 | 48 |
| 14 | 53.18381 | 0.4284 | 63.67801 | 0.1285 | 52 |

*The test is valid only for lags larger than the VAR lag order.
df is degrees of freedom for (approximate) chi-square distribution

On a Q-stat = 29.88 (au retard 11) est inférieur à $\chi^2_{0.05}(40) = 66.80$, et la Q-stat=53.18 (au retard 14) est inférieur à $\chi^2_{0.05}(52) = 79.52$ et les probabilités sont toutes supérieures à 0.05. Donc les résidus croisés se comportent comme un bruit blanc.

A la base de l'étude des corrélogrammes simples et croisés des résidus, on déduit que les résidus associés au modèle VAR(1) forment un vecteur bruit blanc.

➤ **Graphe des résidus**



Figure : graphes multiples des résidus

➤ **Multiple graphe**

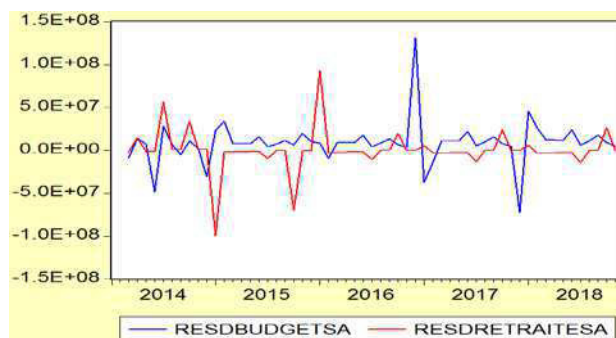


Figure 5-5:graphes des résidus

➤ Test sur les estimations

En tenant compte du tableau des estimations et le fait qu'un coefficient est significativement différent de zéro au seuil de 5%, si la t-stat associée est supérieure à 1.96 et en vertu des résultats des testes précédentes, le modèle VAR(1) est validé et il s'écrit de la façon suivante :

$$\begin{cases} \text{DRETRAITESA}_t = 0.4020137585 * \text{DRETRAITESA}_{t-1} \\ \text{DBUDGETSA}_t = 1.033238075 * \text{DRETRAITESA}_{t-1} \end{cases}$$

5.9.5 Test granger de causalité

Dans notre étude nous retenons le concept de causalité définie par Granger (1969), car il a l'avantage d'être très opérationnel dans les travaux appliqués et s'inscrit parfaitement dans le cadre des modèles vectoriels autorégressifs : il est en effet facilement testable, puisqu'il s'agit ici seulement de tester la nullité jointe de certains coefficients. Nous allons donc tester la causalité entre les deux séries, une à l'autre.

| Null Hypothesis: | Obs | F-Statistic | Probability |
|---|-----|-------------|-------------|
| DRETRAITESA does not Granger Cause ... | 56 | 25.6760 | 1.9E-08 |
| DBUDGETSA does not Granger Cause DRETRAI... | | 0.01254 | 0.98754 |

D'après les résultats du test on a:

- DRETRAITESA_t cause DBUDGETSA_t car la probabilité critique du test $p=1.9^E-08$ est inférieure à 0.05.
- DBUDGETSA_t ne cause pas DRETRAITESA_t car la probabilité critique du test $p=0.98$ est supérieure à 0.05.

Analyse : on remarque que la série (DBUDGETSA_t) est influencée par la série (DRETRAITESA_t) de façon directe. Cette analyse montre effectivement l'existence de réseau d'influence entre le Budget et le budget de retraite de CNR.

5.9.6 Prévision

Le modèle étant validé, l'expression de la prévision pour horizon $h = 1$ à partir de l'instant présent t est donnée par :

$$\begin{cases} \text{DRETRAITESA}_t = 0.4020137585 * \text{DRETRAITESA}_{t-1} \\ \text{DBUDGETSA}_t = 1.033238075 * \text{DRETRAITESA}_{t-1} \end{cases}$$

Pour un horizon $h = 12$ on obtient :

$$\begin{cases} \text{DRETRAITESA}_{t+h} = 0.4020137585 * \text{DRETRAITESA}_{t+h-1} \\ \text{DBUDGETSA}_{t+h} = 1.033238075 * \text{DRETRAITESA}_{t+h-1} \end{cases}$$

Les graphes des prévisions des deux séries et les tableaux associés sont donnés ci-dessous.

| Mois de l'année 2019 | Prévision |
|----------------------|---------------|
| Janvier | 2 806 903 831 |
| Février | 2 803 827 615 |
| Mars | 2 785 119 525 |
| Avril | 2 788 206 960 |
| Mai | 2 760 073 319 |
| Juin | 2 731 493 323 |
| Juillet | 2 799 997 928 |
| Aout | 2 772 022 536 |
| Septembre | 2 744 613 434 |
| Octobre | 2 737 267 845 |
| Novembre | 2 709 957 750 |
| décembre | 2 683 267 395 |

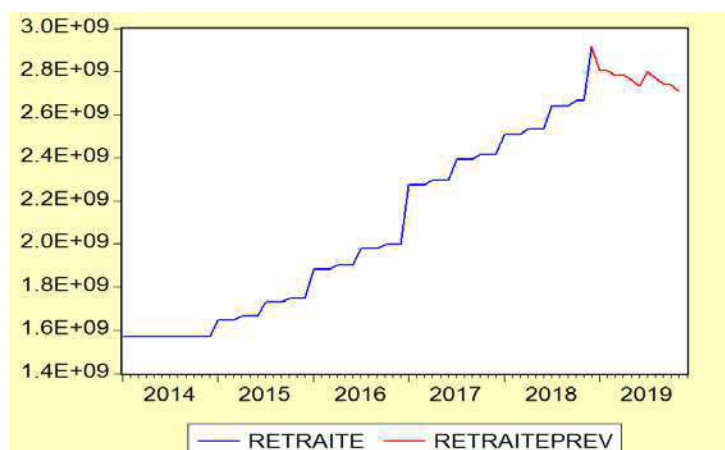


Figure 5-6 : prévision des données retraite

| Mois de l'année 2019 | Prévision |
|----------------------|------------|
| Janvier | 2929494977 |
| Février | 2909056630 |
| Mars | 2882550750 |
| Avril | 2882855850 |
| Mai | 2852601453 |
| Juin | 2822441060 |
| Juillet | 2893115268 |
| Aout | 2864002464 |
| Septembre | 2835680595 |
| Octobre | 2828091277 |
| Novembre | 2799875024 |
| décembre | 2772299074 |

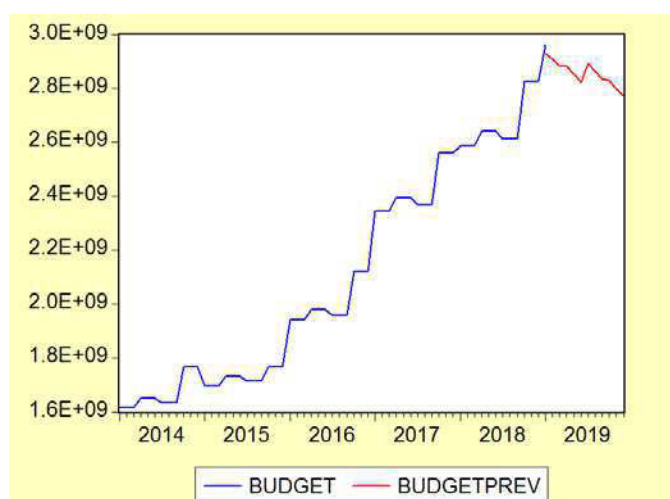


Figure 5-7:prévision des données budget

Chapitre 6
Approche de
cointégration bivariée

6.1 Méthodologie de cointégration bivariée

Introduction

Depuis près d'un demi-siècle, il a été démontré que la non stationnarité semble être le problème des séries macroéconomiques et financières. En effet, de nombreuses séries économiques sont, ou du moins apparaissent être intégrées d'ordre 1. La tendance habituelle est d'utiliser les différences premières et de ce fait régresser les variables stationnaires. Cependant, cette démarche implique malheureusement une perte importante d'informations de long terme sur les variables aux travers des relations existantes entre leurs niveaux.

Pour se rendre compte de ce problème, l'économétrie des séries non stationnaires a connu de nombreux développements à travers notamment de la théorie de la cointégration.

L'analyse de la cointégration permet d'identifier la relation entre plusieurs variables. Cette notion a été introduite donc par Engle et Newbold sous le nom de "spurious régressions", ou régressions fallacieuses puis formalisée par Engle et Granger en 1987, et enfin par Johansen en 1991 et 1995.

6.1.1 Notations et définitions

Séries intégrées

Soient X_t et Y_t deux séries aléatoires notées:

$$X_t = X_{t-1} + \mu_t \dots \dots (1)$$

$$Y_t = Y_{t-1} + \vartheta_t \dots \dots (2)$$

Avec μ_t et ϑ_t considérés comme deux bruits blancs indépendants. Il est à noter que X_t et Y_t sont deux processus autorégressifs d'ordre 1, c'est-à-dire ils suivent un processus AR(1).

Considérons les opérations suivantes :

$$X_t - X_{t-1} = \mu_t \dots \dots (3)$$

$$Y_t - Y_{t-1} = \vartheta_t \dots \dots (4)$$

Ces types d'opérations qui viennent d'être posées sont appelées différenciations.

Les deux processus(1) et (2) ont été différenciés une fois et ont donnés les résultats (3)et (4) on dit que X_t et Y_t sont intégrés d'ordre 1 et on écrit I(1).

Une intégration (I) est une opération de différenciation notée d à laquelle est soumis un processus stochastique pour déterminer sa stationnarité, d'où le symbole I(d).

Dans les processus étudiés aux chapitres précédents, il a été montré que le processus ARMA est un cas particulier du processus non stationnaire ARIMA avec un ordre d'intégration $d = 0$, ce qui lui donne le qualificatif ARMA~I(0).

Finalement, un processus I(0) est un processus stationnaire. En d'autres termes, il n'est pas nécessaire de différencier le processus pour le rendre stationnaire.

➤ **Propriétés des séries intégrées**

Les propriétés suivantes peuvent être dégagées des séries intégrées :

- Si X_t est intégré d'ordre d noté $X_t \sim I(d)$, il s'ensuit que $\alpha + \beta t \sim I(d)$ où α et β sont des constantes, avec comme condition $\beta \neq 0$.
- Si X_t est intégré d'ordre 0 noté $X_t \sim I(0)$ et Y_t est intégré d'ordre 0 noté $Y_t \sim I(0)$, leur combinaison linéaire $\alpha X_t + \beta Y_t$ aura aussi le même ordre d'intégration noté $I(0)$ avec toujours α et β restent constants.
- Si les ordres d'intégration sont différent, par exemple $X_t \sim I(0)$ et $Y_t \sim I(1)$ la combinaison linéaire $\alpha X_t + \beta Y_t$ aura l'ordre d'intégration le plus élevé.
- Si $X_t \sim I(0)$ elle possède une tendance qui fluctue autour de son moment d'ordre 1, sans s'en éloigner grandement.
- Si $X_t \sim I(1)$ c'est-à-dire $X_t - X_{t-1} = \mu_t$, cela veut dire que le moment centré d'ordre 2 dépend du temps et tend vers l'infini.

$$\begin{cases} E(X_t) = t\mu_t \\ V(X_t) = t\sigma_t^2 \end{cases} \text{ Si } |\rho| = 1$$

- Pour $\beta \neq 0$ et μ_t égale à l'espérance mathématique de X_t et σ_t^2 sa variance étant les caractéristiques d'un AR(1).

Enfin, considérons la combinaison linéaire de Y_t et X_t donnant :

$$Y_t = \alpha_0 + \alpha_1 X_t + \varepsilon_t$$

En se référant aux hypothèses classiques des MCO on sait que le paramètre estimé de X_t est $\hat{\alpha}_1$.

$$\hat{\alpha}_1 = \frac{\sum x_t y_t}{\sum x_t^2} \dots \dots \dots (5)$$

Avec x_t et y_t étant les différences à la moyenne. Formulant l'hypothèse que $Y_t \sim I(0)$ et $X_t \sim I(0)$ Dans la mesure où X_t n'est pas stationnaire, sa variance va augmenter infiniment, dominant de la sorte le terme du numérateur de (5) avec pour résultat que $\hat{\alpha}_1$ convergera asymptotiquement vers 0.

6.1.2 Notion de régression fallacieuse "illusoire"

C'est un problème que l'on rencontre dans les séries temporelles, notamment, dans la régression d'une variable sur une ou plusieurs autres.

La situation qui se présente est qu'il arrive dans ces opérations que l'on ait un R^2 très élevé malgré l'absence de relation évidente entre les deux variables.

Régressons les équations (1) et (2).

$$X_t = X_{t-1} + \mu_t \dots \dots (1)$$

$$Y_t = Y_{t-1} + \vartheta_t \dots \dots (2)$$

Avec toujours μ_t et ϑ_t considérés comme deux bruits blancs (BB) indépendants. Par ailleurs, on sait que les deux distributions ne sont pas stationnaires, puisqu'elles sont $I(1)$, le R^2 de la régression devrait en principe être assez faible et tendre vers 0.

Toutefois, il se peut qu'il n'en soit pas ainsi. Ce constat a été fait pour la première fois par Yule (1926) pour qu'il existe une relation statistiquement significative entre Y et X , alors qu'en réalité il n'en est rien : d'où l'appellation "régression fallacieuse". Il a particulièrement fait le constat que ce genre de corrélation persiste dans les séries chronologiques non stationnaires même si l'échantillon est de très grande taille.

Généralement, l'anormalité dans une régression peut être donnée par un d de Durbin Watson très faible; ce qui laisse croire à l'existence d'une auto-corrélation très forte du premier ordre.

Selon Granger et Newbold (1974), un $R^2 > DW$ est une bonne règle empirique pour douter que la régression estimée est fautive.

Dans la mesure où les variables sont $I(1)$, donc non stationnaires en niveau, on peut poser l'analyse de la régression fallacieuse plus loin en stationnarisant les variables en différences premières ΔX_t et ΔY_t . Il apparaît que R^2 devient plus faible et DW plus grand.

6.1.3 Cointégration

Considérons les deux processus X_t et Y_t intégrés d'ordre d . Conformément aux propriétés précédemment énoncées sur les séries intégrées, si X_t et Y_t sont intégrées $I(d)$, leur combinaison linéaire $Z_t = \Delta X_t + \Delta Y_t$ sera aussi $I(d)$. Toutefois, il peut se produire des situations où Z_t n'a pas le même ordre d'intégration.

6.1.3.1 Définition

Soient les processus aléatoires $\{X_1; X_2; \dots; X_n\}$ ayant le même ordre d . On dit qu'ils sont cointégrés s'il peut se former une combinaison linéaire de ces processus qui présente un ordre d'intégration inférieur à d . Il est nécessaire qu'il ait une valeur $b > 0$ et des valeurs $\{\beta_1; \beta_2; \dots; \beta_n\}$ vérifiant:

- $\beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n$ est $I(d - b)$. On peut écrire aussi que cette combinaison linéaire est Z_t et de ce fait poser:

$$Z_t = \sum_{i=1}^n \beta_i X_i$$

Chaque variable $X_1; X_2; \dots; X_n$ est $I(d)$.

6.1.3.2 Vecteurs cointégrés ou relation de long terme

A partir de cette relation de long terme, on peut dégager un vecteur noté :

$\beta = (\beta_1; \beta_2; \dots; \beta_n)'$ que l'on appelle vecteur cointégrant. Ce qui permet de poser au plan vectoriel la notion cointégrée suivante:

$$\beta'X_1 = \beta_1X_1 + \beta_2X_2 + \dots + \beta_nX_n \dots (6) \text{ Où } \beta' \text{ est le vecteur transposé de } \beta.$$

Plusieurs vecteurs cointégrés linéairement indépendants peuvent exister pour les mêmes n variables. On note le nombre de ces vecteurs r . Toutefois, il existe une condition à savoir que le nombre de vecteurs (r) doit être inférieur au nombre de variables (n), $r < n$. De façon plus explicite, on ne peut avoir plus de $(n - 1)$ vecteurs de cointégration linéairement indépendants. Ce qui veut dire que dans une relation linéaire ($n = 2$), le rang de cointégration ne peut dépasser 1.

➤ Propriétés

- Toute combinaison linéaire de vecteurs de cointégration pour les variables est aussi un vecteur de cointégration pour ces variables.
- Une même direction vectorielle est représentée par une infinité de vecteurs, tous linéairement indépendants, mais aussi, combinaisons linéaires les uns des autres.
- Si $r > 1$, même lorsque chacune des r directions vectorielles a été normalisée et de ce fait, caractérisée par un vecteur unique normalisé, toute combinaison linéaire de ces r vecteurs est également un vecteur de cointégration.

6.1.3.3 Test de cointégration

Tester la cointégration entre variables se ramène à estimer par MCO une équation linéaire où l'une des variables est exprimée en fonction des autres et à tester si le résidu Z_t a une racine unitaire. Si on ne rejette pas l'hypothèse d'existence d'une racine unitaire pour le résidu estimé, on conclut que les séries ne sont pas cointégrées (existence d'une racine unitaire); dans le cas contraire on conclut à la cointégration (il n'y a pas de racine unitaire).

Estimation

Généralement, on présente la méthode d'Engle et Granger en deux étapes. Toutefois, cette méthode n'est valable que dans le cas de l'existence d'un seul vecteur cointégrant entre les variables utilisées, et que $b = d$, c'est-à-dire même ordre d'intégration. Les deux étapes sont d'une part, montrer l'existence d'une relation de long terme; d'autre part exprimer les variables cointégrées sous la forme d'un modèle à correction d'erreur.

1^{ière} étape: estimation de la relation de long terme

On pense que si les variables $X_1; X_2; \dots; X_n$ sont cointégrées, il existe des paramètres

$\beta_1; \beta_2; \dots; \beta_n$ telle que la combinaison linéaire :

$$\beta_1X_1 + \beta_2X_2 + \dots + \beta_nX_n \sim I(0)$$

permette de normaliser l'équation par l'écriture suivante :

$$X_1 + \frac{\beta_2}{\beta_1} X_2 + \dots + \frac{\beta_n}{\beta_1} X_n \sim I(0) \dots \dots (7)$$

En posant $\hat{\beta}_i = \frac{\beta_i}{\beta_1}$, la relation (7) s'écrit :

$$X_1 + \hat{\beta}_2 X_2 + \dots + \hat{\beta}_n X_n \sim I(0)$$

Cette méthode est appelée normalisation des coefficients de façon à résoudre le problème de sur identification qui peut se poser dans les modèles d'équations simultanées. Ces coefficients normalisés sont :

$$\hat{\beta}_1 = \frac{\beta_1}{\beta_1} = 1 ; \hat{\beta}_2 = \frac{\beta_2}{\beta_1} ; \dots ; \hat{\beta}_n = \frac{\beta_n}{\beta_1}$$

Tout processus $I(0)$ est le résultat d'une constante caractérisée par son espérance mathématique à laquelle on ajoute un processus aléatoire $I(0)$ de moyenne nulle. Dans le cas présent, la stationnarité du processus entraîne que :

$$X_{1t} + \hat{\beta}_2 X_{2t} + \dots + \hat{\beta}_n X_{nt} = \mu + Z_t \dots \dots (8)$$

Avec μ une constante et Z_t un processus aléatoire stationnaire $I(0)$ de moyenne nulle.

On rappelle que Z_t ne suit pas obligatoirement un bruit blanc (BB). Il peut être autocorrélé notamment dans le cas des petits échantillons.

Finalement, la cointégration signifie que :

$$X_{1t} = \mu + (-\hat{\beta}_2)X_{2t} + \dots + (-\hat{\beta}_n)X_{nt} + Z_t \dots \dots (9)$$

Dans laquelle les coefficients cointégrants, normalisés $\hat{\beta}_i$ sont seuls pour lesquels Z_t a les propriétés équivalentes ci-après selon **Eric DOR** (2004).

- Cette étape estime en fait la relation de long terme qui est fondée aussi sur un échantillon de grande taille.

Test de cointégration

La plupart des tests de cointégration proposés par **Engle** et **Granger** (1987) s'appuient principalement sur les résidus estimés \hat{Z}_t de la relation statistique de façon à vérifier s'ils sont réellement $I(0)$ tel que le stipule le théorème de représentation de **Granger** lorsque les variables sont cointégrées.

Les tests proposés par les deux auteurs posent l'hypothèse nulle en cas d'absence d'une relation de cointégration, c-à-d d'une non stationnarité des résidus. Ces tests ne sont valables que dans le cas d'un modèle bivarié. Parmi les nombreux tests proposés, nous nous

intéressons particulièrement aux tests de Dickey-Fuller simple (DF), Dickey-Fuller Augmenté (ADF) et celui de CRDW, mais **Engle** et **Granger** recommandent l'usage des tests ADF, car ils se révèlent être les plus puissants.

Tests DF et ADF

L'objectif ici est de tester l'existence d'une racine unitaire dans les résidus estimés Z_t de la Relation :

$$y_t = \alpha_0 + \alpha_1 x_t + Z_t$$

Soit le test DF : $\Delta \hat{z}_t = \phi \hat{z}_{t-1} + \hat{u}_t$

Avec : $\phi = \rho - 1$ (processus AR(1))

$$\text{Test ADF: } \Delta \hat{z}_t = \phi \hat{z}_{t-1} + \sum_{i=2}^p \theta_i \Delta \hat{z}_{t-i} + \hat{u}_t$$

L'interprétation se fait comme suit : si dans chacun des cas, le t de Student calculée (t_c) du coefficient ϕ est supérieur à la valeur théorique, on accepte $H_0 : \phi = 0$ (conduisant $\rho = 1$), et donc les séries X_t et Y_t ne sont pas cointégrées du fait de leur non stationnarité.

6.1.4 Modèles à Correction d'Erreur (MCE)

Considérons deux séries X_t et Y_t cointégrées, traduisant qu'il existe une relation de long terme ou d'équilibre entre les deux variables. Toutefois cet équilibre de long terme ne signifie pas qu'il ne peut pas apparaître à court terme un déséquilibre. Le terme d'erreur Z_t exprime cette "erreur de déséquilibre".

La question qui se pose, était comment expliquer ou établir le lien entre ce déséquilibre de court terme et l'équilibre de long terme ? C'est à cette question le MCE tente de répondre.

6.1.4.1 Conditions de recours au MCE

Les développements précédents ont montré qu'il existe un certain nombre de conditions pour recourir au MCE. Parmi ces conditions, la principale est incarnée par le théorème de Représentation de Granger qui s'énonce comme suit:

Théorème

Si deux séries X_t et Y_t sont cointégrées, la relation entre les deux peut s'exprimer comme un MCE pour capturer la structure dynamique de l'évolution de court terme de ces séries.

6.1.4.2 Présentation des Modèles à Correction d'erreur

Soit le modèle : $Y_t = \beta_0 + \beta_1 X_t + Z_t \dots \dots \dots (10)$

et considérons la relation de long terme :

$$Y_t = \beta_0 + \beta_1 X_t \dots \dots (11)$$

Admettons que les deux variables sont I(1), avec la combinaison linéaire cointégrante .

$$Y_t - \beta_0 + \beta_1 X_t = Z_t \neq 0$$

En principe, Z_t est I(0) donc stationnaire. Le vecteur cointégrant est $(1, -\beta_0, -\beta_1)$, Z_t représente l'erreur de déséquilibre, le coefficient β_1 est le paramètre de long terme. Si X_t et Y_t sont en équilibre, $Z_t = 0$, dans le cas contraire est soit positif, soit négatif.

Supposons que X_t et $Y_t \sim CI(1; 1)$ (c-à-dire sont cointégrées d'ordre (1,1)) et considérons la relation ci-après :

$$y_t = k + aX_t + bX_{t-1} + \gamma Y_{t-1} + \varepsilon_t \dots (12)$$

Avec $k =$ constante et $0 < \gamma < 1$

La représentation de (12~) permet d'obtenir une représentation dynamique appelée MCE.

$$y_t - Y_{t-1} = k + aX_t + bX_{t-1} + (\gamma - 1)Y_{t-1} + \varepsilon_t$$

On retranche au membre de droite et on lui ajoute en même temps aX_{t-1} , on obtient :

$$\begin{aligned} \Delta Y_t &= k + a\Delta X_t + (a + b) X_{t-1} + (\gamma - 1)Y_{t-1} + \varepsilon_t \\ &= a\Delta X_t + (1 - \gamma) \left[Y_{t-1} - \frac{a + b}{1 - \gamma} X_{t-1} - \frac{k}{1 - \gamma} \right] + \varepsilon_t \end{aligned}$$

$$\text{Posons : } \frac{a+b}{1-\gamma} = \hat{\beta}_1 \quad \text{et} \quad \frac{k}{1-\gamma} = \hat{\beta}_0$$

Finalement les résultats :

$$\Delta Y_t = a\Delta X_t + (1 - \gamma) \left[Y_{t-1} - \hat{\beta}_1 X_{t-1} - \hat{\beta}_0 \right] + \varepsilon_t$$

$$\Delta Y_t = k + a\Delta X_t + (\lambda) \left[Y_{t-1} - \hat{\beta}_1 X_{t-1} \right] + \varepsilon_t \quad (13)$$

Avec : $\lambda = - (1 - \gamma)$ et $k = \hat{\beta}_0 - \hat{\beta}_0 \gamma$

Ce modèle décrit la variation de Y_t autour de son trend $\hat{\beta}_1$ est l'estimateur des MCO de la relation de longue période et le modèle présente une certaine cohérence à condition que les variables soient intégrées, traduisant dans un tel cas que les différences premières ΔX_t et ΔY_t et $(Y_{t-1} - \hat{\beta}_1 X_{t-1})$ sont I(0), c.-à-d. stationnaires. S'il y a équilibre, le terme $Y_{t-1} - \hat{\beta}_1 X_{t-1} = 0$. Au cours des périodes de déséquilibre, il mesure la distance à laquelle le système s'est écarté de son équilibre à la date t:

Afin d'avoir une indication sur la vitesse d'ajustement de y_t vers son niveau d'équilibre, on recourt aux MCO du paramètre γ ($\gamma > 0$). Toutefois, le coefficient $= (-1+\gamma) < 0$.

On peut réécrire la relation (13) : comme suit

$$\Delta Y_t = k + a \Delta X_t + \lambda Z_{t-1} + \varepsilon_t \dots (14).$$

Avec : $Z_{t-1} = Y_{t-1} - \widehat{\beta}_1 X_{t-1} - \widehat{\beta}_0$ et $\lambda = -(1-\gamma)$

Le modèle (13) signifie que Δy_t est en fonction de Δx_t et également du terme d'équilibre Z_{t-1} . Si Z_{t-1} n'est pas nul, le modèle se place hors de l'équilibre. Admettons que $\Delta X_t = 0$ et que $Z_{t-1} > 0$. Cela implique que Y_{t-1} est très élevé pour être à l'équilibre, en d'autres termes, Y_{t-1} est au dessus de sa valeur d'équilibre ($k - b X_{t-1}$). Du fait qu'on s'attend à un $\lambda < 0$ pour ramener l'équilibre initiale. Cela veut dire que si Y_t est au-dessus de sa valeur d'équilibre, il commence à diminuer dans la période qui suit pour corriger l'équilibre d'où l'appellation de MCE. A l'inverse, si $Z_{t-1} < 0$ puisque $\lambda < 0$, (conduisant Y_t inférieur à sa valeur d'équilibre) $\lambda Z_{t-1} > 0$ entraînant un accroissement de Y_t à la période suivante.

Finalement, c'est la valeur absolue de λ qui oriente la rapidité avec laquelle l'équilibre est établi. Notons que α est le paramètre de court terme. En définitive, suivant les deux étapes Définies par Engle et Granger, il ressort les deux paramètres qui sont : $\widehat{\beta}_1$ de l'équation (11) qui exprime la relation de long terme et α de l'équation (14) qui caractérise la relation de court terme.

Finalement, les étapes d'Engle et de Granger se résument de la façon suivante :

1 ère étape : Tester l'ordre d'intégration des séries.

Une condition nécessaire de cointégration est que les séries doivent avoir le même ordre d'intégration. Dans le cas contraire, elles ne peuvent pas être cointégrées. Pour se faire, il faut recourir aux tests DF et ADF de manière à déterminer le type de tendance (déterministe ou stochastique) des variables.

Si on a deux variables X_t et Y_t pour parler de cointégration, il faut que $X_t \sim I(d)$ et $Y_t \sim I(d)$.

2 ème étape : Estimer la relation de long terme.

Soit à considérer le modèle : $Y_t = \alpha + \beta X_t + \varepsilon_t$: La seconde condition d'une existence de relation de cointégration entre X_t et Y_t est que le résidu Z_t issu de la régression des MCO soit stationnaire, c.-à-d. : $Y_t - \widehat{\alpha} - \widehat{\beta} X_t = Z_t \sim I(0)$

La stationnarisation de Z_t ne se fait plus à l'aide des tests DF et ADF. Du fait que le

test porte sur les résidus estimés et non pas sur les vraies valeurs des résidus, on recourt aux tables de Mackinnon (1991). Si les résidus sont stationnaires, on passe à l'estimation des MCE.

3ème étape : Estimer le MCE

Cette procédure se fait en deux phases.

Phase 1 : Estimation par les MCO de la relation de long terme :

$$Y_t = \hat{\alpha} + \hat{\beta}X_t + Z_t$$

Phase 2 : Estimation par les MCO de la relation du modèle dynamique (court terme) :

$$\Delta Y_t = \alpha_1 \Delta X_t + \alpha_2 Z_{t-1} + u_t \quad (15)$$

Le coefficient α_2 est la force de rappel et doit être significativement négatif. Si tel n'est pas le cas, on rejette la spécification de nature MCE. Concrètement, si $\alpha_2 > 0$ cela voudrait dire que Y_{t-1} est au dessus de sa valeur d'équilibre (la tendance) et s'éloignerait d'elle. Le modèle (15) peut être généralisé pour saisir un processus dynamique plus complexe et ceci en introduisant des opérateurs de retard $A(L)$ et $B(L)$:

$$A(L)\Delta Y_t = B(L)\Delta X_t + \lambda(Y_{t-1} - \hat{\beta}X_{t-1}) + \varepsilon_t \dots (16)$$

avec

$$A(L) = 1 - \rho_1 L - \dots - \rho_p L^p$$

$$B(L) = \alpha_0 + \alpha_1 L + \dots + \alpha_p L^p$$

$$\lambda = 1 - \rho_1 - \rho_2 - \dots - \rho_p$$

6.1.5 Cointégration entre n variables et MCE

Si on se réfère aux théories macroéconomiques récentes et notamment celles portant sur l'équilibre, il ressort qu'une variable stationnaire est susceptible d'être expliquée par des Variables non stationnaires. De tout temps d'ailleurs on sait qu'un phénomène économique peut être la combinaison de plusieurs variables. C'est la raison pour laquelle, il est nécessaire de porter l'analyse de la cointégration à n variables et de voir l'implication pour le MCE.

6.1.5.1 Cointégration entre n variables

$$\text{Soit le modèle : } Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \dots + \beta_n X_{nt} + Z_t$$

Et considérons que toutes ces variables sont $I(0)$ avec une certaine possibilité de cointégration. En se référant à la situation de deux variables dont la combinaison linéaire est stationnaire, si on généralise, on dira que la combinaison linéaire pour n variables est aussi stationnaire, ce qui entraîne leur cointégration. Le recours aux MCO du modèle permet d'estimer le résidu Z_t

$$Z_t = Y_t - \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \dots + \beta_n X_{nt}$$

Dans le cas d'une stationnarité de Z_t , on admet l'hypothèse d'une cointégration entre les variables. Les tests DF et ADF de stationnarité sur le résidu Z_t s'effectuent comme dans le cas de deux variables sur les valeurs critiques tabulées par Mackinnon (1991) en fonction du nombre de variables total du modèle. Il ressort le vecteur suivant de cointégration $\{1, -\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_n\}$

Il demeure toutefois des possibilités de cas complexes. On suppose, par exemple, que les variables y_t, x_{1t}, x_{2t} soient cointégrées deux à deux tel que par exemple: $y_t, x_{1t} \sim CI(1; 1)$

et $x_{2t}, x_{3t} \sim CI(1; 1)$ donnant leur combinaison linéaire $I(0)$.

Nous posons :

$$Z_t^{(1)} = y_t - \hat{\alpha}_0 - \hat{\alpha}_1 x_{1t} \quad \text{et} \quad Z_t^{(2)} = x_{2t} - \hat{\gamma}_0 - \hat{\gamma}_1 x_{3t}$$

D'où les vecteurs cointégrants suivants : $r_1 = \{1, -\hat{\alpha}_0, -\hat{\alpha}_1\}$ et $r_2 = \{1, -\hat{\gamma}_0, -\hat{\gamma}_1\}$

Ce qui engendre deux vecteurs de cointégration r_1 et r_2 . Pour la formulation de l'estimation du résidu Z_t , on aura :

$$Z_t - Z_t^{(1)} + Z_t^{(2)} = y_t - \hat{\alpha}_0 - \hat{\alpha}_1 x_{1t} + x_{2t} - \hat{\gamma}_0 - \hat{\gamma}_1 x_{3t} \sim I(0).$$

Ce qui génère un autre vecteur de cointégration qui est : $r_3 = \{1, -\hat{\alpha}_0, -\hat{\alpha}_1, 1, -\hat{\gamma}_0, -\hat{\gamma}_1\}$

La conclusion qui s'impose est que dans un modèle à une variable endogène et n variables exogènes (donnant au total $n + 1$ variables), il peut se former n vecteurs de cointégration linéairement indépendants. Le nombre de vecteurs linéairement indépendants appelé rang de la cointégration est r : Dans le cas présent on a 3 variables exogènes et une variable endogène. Il ressorti un $n = 3$.

De façon générale, l'existence d'un vecteur unique de cointégration est possible lorsque les séries ont le même ordre d'intégration tel que $I(1)$: Dans le cas contraire, si les ordres sont différents, il y a toute certitude qu'il n'y ait pas un seul vecteur de cointégration. La démarche est la suivante pour tester une probable cointégration entre plusieurs variables :

- On teste la cointégration sur les $n + 1$ variables.
- Si l'hypothèse de cointégration est vérifiée, on teste les variables en les combinant.

6.1.6 Estimation des Modèles à Correction d'Erreur

La méthode d'estimation en deux étapes par Engle et Granger (1987). Son principal avantage réside dans sa simplicité de mise en œuvre. Il est important de noter que cette technique n'est valable que pour les séries cointégrées $CI(1; 1)$

Si les tests précédents confirment l'existence d'une cointégration, il apparaît deux possibilités :

- Il existe un vecteur unique de cointégration;
- Il existe plusieurs vecteurs de cointégration.

Si le premier cas de figure se confirme, on peut recourir aux méthodes d'estimation en deux étapes d'Engle et de Granger.

1) Estimation par MCO de la relation de long terme et détermination du résidu Z_t

$$\hat{Z}_t = y_t - \hat{\beta}_1 + \hat{\beta}_0 x_{1t} + \hat{\beta}_2 x_{2t} + \dots + \hat{\beta}_r x_{rt}$$

2) Estimation par MCO par la relation de court terme

$$\Delta y_t = \alpha_1 \Delta x_{1t} - \alpha_2 \Delta x_{2t} - \dots - \alpha_r \Delta x_{rt} - \lambda \hat{Z}_{t-1} + v_t$$

Le coefficient λ est la force de rappel et comme dans le cas à une variable exogène doit être négatif.

6.2 Application de la méthodologie de cointégration

Dans le chapitre 3, nous avons étudié les séries (retraite ; budget). Cette étude, Menée dans un cadre uni varié, n'explique pas l'influence de la série retraite sur la série budget. Afin de trouver la relation qui existe entre ces deux séries, nous avons proposé de faire une étude dans le cadre bi varié par le biais de la théorie de la cointégration qui exige que les séries soient intégrées d'ordre 1, ce qui est vérifié (voir le chapitre 1).

6.2.1 Etude de l'évolution des séries : budget et retraite

Considérons les séries mensuelles $RETRAITESA_t$ et $BUDGETSA_t$ sur la période allant de janvier 2014 à Décembre 2018. Ces deux séries, comme on l'a déjà vu dans le chapitre 4, exhibent une tendance commune à la hausse .

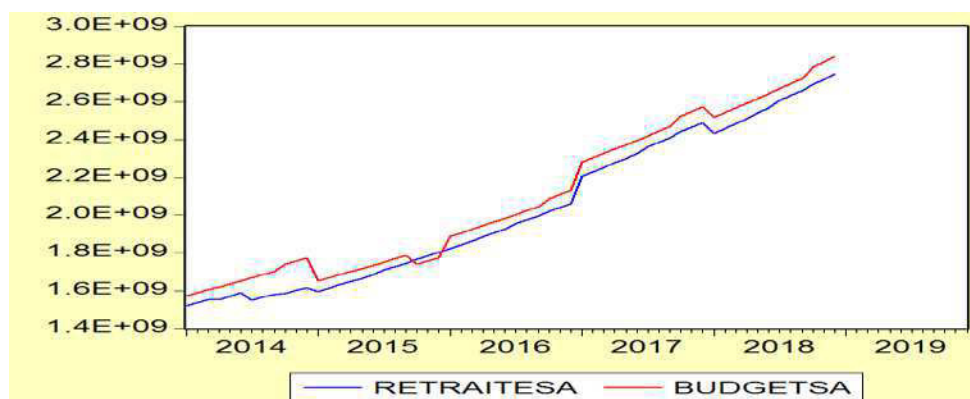


Figure 6-1: graphe des séries RETRAITESA; BUDGETSA

Afin de trouver la relation de cointégration entre ces deux séries, nous avons appliqué la procédure en deux étapes d'**Engle et Granger**.

➤ Détermination de l'ordre d'intégration des séries

D'après l'application de Box-Jenkins sur les deux séries $RETRAITESA_t$ et $BUDGETSA_t$, on a Trouvé que les deux séries avaient le même ordre d'intégration qui est $d = 1$.

➤ Estimation de la relation statique entre $RETRAITESA_t$ et $BUDGETSA_t$.

Cette étape consiste à estimer la relation de long terme suivante :

$$BUDGETSA_t = \beta + \alpha * RETRAITESA_t + Z_t$$

L'estimation de cette relation est donnée par la table suivante :

Dependent Variable: BUDGETSA
 Method: Least Squares
 Date: 06/13/19 Time: 15:11
 Sample (adjusted): 2014M01 2018M12
 Included observations: 60 after adjustments

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|--------------------|-------------|-----------------------|-------------|----------|
| RETRAITSA | 1.007490 | 0.011294 | 89.20270 | 0.0000 |
| C | 52570219 | 23331302 | 2.253206 | 0.0280 |
| R-squared | 0.992764 | Mean dependent var | | 2.10E+09 |
| Adjusted R-squared | 0.992639 | S.D. dependent var | | 4.00E+08 |
| S.E. of regression | 34349789 | Akaike info criterion | | 37.57486 |
| Sum squared resid | 6.84E+16 | Schwarz criterion | | 37.64467 |
| Log likelihood | -1125.246 | Hannan-Quinn criter. | | 37.60216 |
| F-statistic | 7957.122 | Durbin-Watson stat | | 0.444799 |
| Prob(F-statistic) | 0.000000 | | | |

Nous remarquons que les coefficients α et β sont significativement différents de zéro au seuil 5% car leurs statistiques de Student sont en valeurs absolues supérieures à 1.96. Donc la relation de long terme est écrite sous la forme suivante :

$$\text{BUDGETSA}_t = 525770219 + 1.007 \text{RETRAITSESA} + Z_t$$

Avec Z_t les résidus de la relation estimée.

Notons que la condition nécessaire pour que la relation de long terme soit une relation de cointégration et que la série de ces résidus soit stationnaire (dans le cas contraire on dit qu'il s'agit d'une régression fallacieuse).

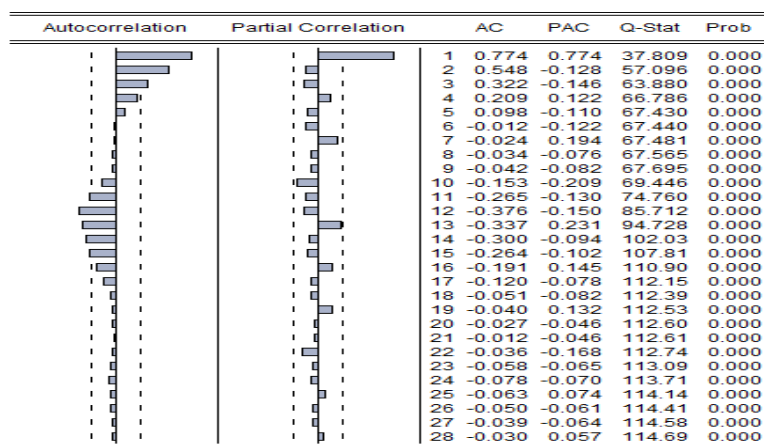


Figure 6-2: corrélogramme des résidus \hat{Z}_t

En analysant le corrélogramme des résidus estimés \hat{Z}_t , nous constatons que la série \hat{Z}_t n'est pas stationnaire. Pour confirmer la non stationnarité des résidus, nous avons appliqué le test ADF.

Choix du nombre de retards

| Lag | Akaike Criterion |
|-----|------------------|
| 0 | 36.63 |
| 1 | 36.66 |
| 2 | 36.69 |

Le nombre de retard est définitivement $p = 0$.

Pour confirmer la stationnarité des résidus, nous avons appliqué le test ADF dont les résultats sont reportés dans la table suivante :

| Augmented Dickey-Fuller test statistic | | -2.676197 | 0.0082 | |
|---|-------------|-----------------------|-------------|--------|
| Test critical values: | | | | |
| 1% level | | -2.604746 | | |
| 5% level | | -1.946447 | | |
| 10% level | | -1.613238 | | |
| *Mackinnon (1996) one-sided p-values. | | | | |
| Augmented Dickey-Fuller Test Equation | | | | |
| Dependent Variable: D(RESID03) | | | | |
| Method: Least Squares | | | | |
| Date: 06/13/19 Time: 16:18 | | | | |
| Sample (adjusted): 2014M02 2018M12 | | | | |
| Included observations: 59 after adjustments | | | | |
| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
| RESID03(-1) | -0.221670 | 0.082830 | -2.676197 | 0.0097 |
| R-squared | 0.109514 | Mean dependent var | 479664.9 | |
| Adjusted R-squared | 0.109514 | S.D. dependent var | 22903877 | |
| S.E. of regression | 21613373 | Akaike info criterion | 36.63233 | |
| Sum squared resid | 2.71E+16 | Schwarz criterion | 36.66754 | |
| Log likelihood | -1079.654 | Hannan-Quinn criter. | 36.64607 | |
| Durbin-Watson stat | 1.803146 | | | |

Afin d'interpréter ces résultats, nous utilisons la table des valeurs critiques de Mackinnon. Nous constatons que la statistique ADF estimée (2.676) est supérieur à la valeur Tabulée de Mackinnon (3.34) au seuil 5% . Nous en déduisons donc que les résidus de la relation entre $RETRAITESA_t$ et $BUDGETSA_t$ ne sont pas stationnaires. Par conséquent, les deux séries ne sont pas cointégrées.

Après avoir estimé la relation de long terme entre les deux séries, nous ne pouvons pas passer au MCE puisque les résidus possèdent une racine unitaire donc ne sont pas stationnaires ce qui inclut une absence de relation de cointégration entre les deux séries $RETRAITESA_t$ et $BUDGETSA_t$

6.2.2 Etude de l'évolution des séries : recette et dépense

Considérons les séries mensuelles $RECETTESA_t$ et $DEPENSA_t$ sur la période allant de janvier 2009 à Décembre 2018. Ces deux séries, comme on l'a déjà vu dans le chapitre 4, exhibent une tendance commune à la hausse .

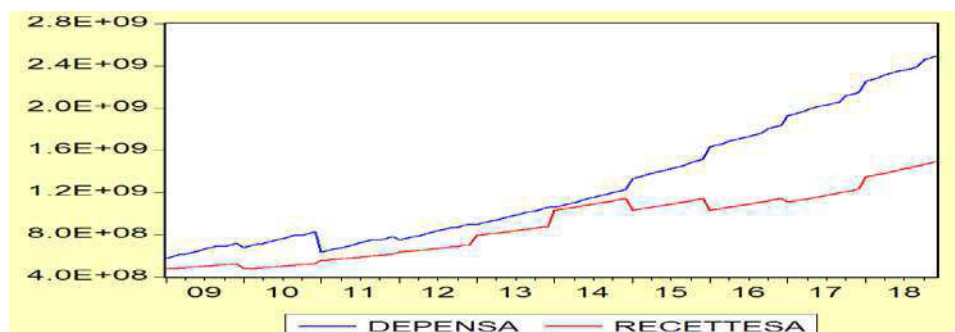


Figure 6-3: évolution des séries $RECETTESA_t$ et $DEPENSA_t$

Afin de trouver la relation de cointégration entre ces deux séries, nous avons appliqué la Procédure en deux étapes **d'Engle et Granger** .

➤ **Détermination de l'ordre d'intégration des séries**

D'après l'application de Box-Jenkins sur les deux séries $RECETTESA_t$ et $DEPENSA_t$, on a Trouvé que les deux séries avaient le même ordre d'intégration qui est $d = 1$.

Estimation de la relation statique entre $RECETTESA_t$ et $DEPENSA_t$.

Cette étape consiste à estimer la relation de long terme suivante :

$$RECETTESA_t = \beta + \alpha * DEPENSA_t + Z_t$$

L'estimation de cette relation est donnée par la table suivante :

| Dependent Variable: RECETTESA | | | | |
|--|-------------|-----------------------|-------------|--------|
| Method: Least Squares | | | | |
| Date: 06/15/19 Time: 10:09 | | | | |
| Sample (adjusted): 2009M01 2018M12 | | | | |
| Included observations: 120 after adjustments | | | | |
| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
| DEPENSA | 0.493830 | 0.017889 | 27.60572 | 0.0000 |
| C | 2.69E+08 | 24779814 | 10.87158 | 0.0000 |
| R-squared | 0.865921 | Mean dependent var | 8.92E+08 | |
| Adjusted R-squared | 0.864784 | S.D. dependent var | 3.06E+08 | |
| S.E. of regression | 1.12E+08 | Akaike info criterion | 39.93096 | |
| Sum squared resid | 1.49E+18 | Schwarz criterion | 39.97742 | |
| Log likelihood | -2393.858 | Hannan-Quinn criter. | 39.94983 | |
| F-statistic | 762.0758 | Durbin-Watson stat | 0.069695 | |
| Prob(F-statistic) | 0.000000 | | | |

Nous remarquons que les coefficients α et β sont significativement différent de zéro au seuil 5% car leurs statistiques de Student sont en valeurs absolues supérieures à 1.96. Donc la relation de long terme est écrit sous la forme suivante :

$$RECETTESA_t = 2.69E + 08 + 0.49 * DEPENSA_t + Z_t$$

Avec Z_t les résidus de la relation estimée.

Notons que la condition nécessaire pour que la relation de long terme soit une relation de cointégration et que la série de ces résidus soit stationnaire (dans le cas contraire on dit qu'il s'agit d'une régression fallacieuse).

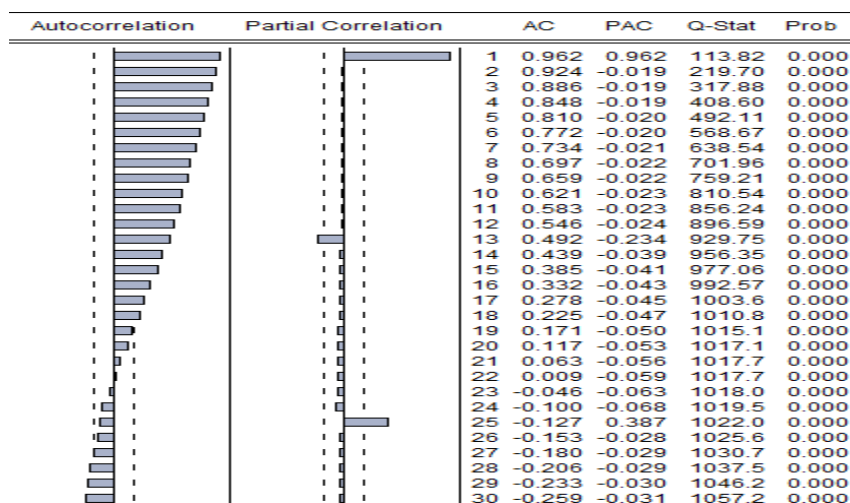


Figure 6-4: corrélogramme des résidus estimés

En analysant le corrélogramme des résidus estimés \widehat{Z}_t , nous constatons que la série \widehat{Z}_t n'est pas stationnaire. Pour confirmer la non stationnarité des résidus, nous avons appliqué le test ADF.

Choix du nombre de retards

| Lag | Akaike Criterion |
|-----|------------------|
| 0 | 38.12 |
| 1 | 38.52 |
| 2 | 38.44 |

Le nombre de retards est définitivement $p = 0$.

Pour confirmer la stationnarité des résidus, nous avons appliqué le test ADF dont les Résultats sont reportés dans la table suivante :

| Augmented Dickey-Fuller test statistic | -1.573251 | 0.1085 | | |
|--|-------------|-----------------------|-------------|----------|
| Test critical values: | 1% level | -2.584539 | | |
| | 5% level | -1.943540 | | |
| | 10% level | -1.614941 | | |
| *Mackinnon (1996) one-sided p-values. | | | | |
| Augmented Dickey-Fuller Test Equation | | | | |
| Dependent Variable: D(RESID01) | | | | |
| Method: Least Squares | | | | |
| Date: 06/15/19 Time: 10:28 | | | | |
| Sample (adjusted): 2009M02 2018M12 | | | | |
| Included observations: 119 after adjustments | | | | |
| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
| RESID01(-1) | -0.037845 | 0.024055 | -1.573251 | 0.1183 |
| R-squared | 0.020089 | Mean dependent var | | 637928.1 |
| Adjusted R-squared | 0.020089 | S.D. dependent var | | 29687304 |
| S.E. of regression | 29387603 | Akaike info criterion | | 37.23841 |
| Sum squared resid | 1.02E+17 | Schwarz criterion | | 37.26177 |
| Log likelihood | -2214.686 | Hannan-Quinn criter. | | 37.24790 |
| Durbin-Watson stat | 1.965376 | | | |

Afin d'interpréter ces résultats, nous utilisons la table des valeurs critiques de Mackinnon. Nous constatons que la statistique ADF estimée 1.57 est supérieur à la valeur Tabulée de Mackinnon au seuil 5% (3.34). Nous en déduisons donc que les résidus de la Relation entre $RECETTES_t$ et $DEPENSA_t$ ne sont pas stationnaires. Par conséquent, les deux séries ne sont pas cointégrées.

Après avoir estimé la relation de long terme entre les deux séries, nous ne pouvons pas passer au MCE puisque les résidus possèdent une racine unitaire donc ne sont pas stationnaire ce qui inclut une absence de relation de cointégration entre les deux séries $RECETTES_t$ et $DEPENSA_t$.

Chapitre 7
étude comparative

7 Choix de la méthode de prévision des séries

Pour comparer entre les trois méthodes de prévisions, on utilise le critère RMSE: racine de l'erreur quadratique moyenne (Root Mean Squar Error) :

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n \hat{\varepsilon}_t^2}$$

| Série | Box & Jenkins | Holt & Winters | Var(1) |
|-----------------|---------------|----------------|-------------|
| RMSE (dépense) | 9354143.46 | 28369542 | - |
| RMSE (recette) | 20534523.77 | 26264138 | - |
| RMSE (retraite) | 15275252.31 | 24733859 | 2113489.97 |
| RMSE (budget) | 20412414.52 | 32733060 | 29748949.56 |

Tableau 17: comparaison RMSE

| Série | Box & Jenkins | Méthode financière(CNR) | Réelle | Erreur Relatif | |
|----------------------|----------------|-------------------------|----------------|----------------|--------------------|
| | | | | Box & Jenkins | Méthode financière |
| Retraite (DA) | 35 319 667 363 | 36 005 500 000 | 35 890 000 000 | 570332637 | -115 500 000 |
| Budget (DA) | 36 020 517 308 | 36 293 410 000 | 37 000 000 000 | 979582692 | 706590000 |

Tableau 28: comparaison entre les méthodes statistiques et la méthode financière

D'après les tableaux (17,18) ci-dessus on a les résultats suivants:

- Pour la série dépense, RMSE de Box & Jenkins < RMSE de Holt & Winters.
- Pour la série recette, RMSE de Box & Jenkins < RMSE de Holt & Winters.
- Pour la série retraite, RMSE de Box & Jenkins < RMSE de VAR(1) < RMSE de Holt & Winters.
- Pour la série budget, RMSE de Box & Jenkins < RMSE de VAR(1) < RMSE de Holt & Winters.

Pour les quatre séries nous concluons que les prévisions obtenues par la méthode de Box & Jenkins sont meilleurs par rapport les deux autres méthodes.

Les prévisions par les deux méthodes statistiques et financière sont plus proches de la réalité pour les différentes séries .

Conclusion

La réalisation de ce projet constitue notre premier contact avec la vie professionnelle (active). Nous avons non seulement pu voir la complexité rencontrée dans la pratique, mais aussi mettre en pratique les connaissances acquises pendant notre cursus en l'occurrence la modélisation mathématique et les différentes méthodes de résolution des problèmes des prévisions.

Tout au long de ce travail nous avons essayé d'atteindre l'objectif fixé est de fournir un outil d'aide à la décision par des prévisions budgétaires à court terme pour, à savoir, trouver la méthode et le modèle qui convient au mieux à nos séries.

Pour cela nous avons utilisé quatre approches économétriques traitant les séries chronologiques : « l'approche univariée de Box & Jenkins », « l'approche de Holt & Winters », « l'approche du modèle VAR » et « l'approche bivariée de cointégration ».

Dans un premier temps, il nous a fallu d'étudier individuellement les séries en utilisant l'approche univariée qui comporte deux méthodes Holt & Winters et Box & Jenkins. Pour l'évolution de budget, pension de retraite, dépense et recette de la CNR. L'intérêt de la modélisation des séries individuelles est de permettre au passé du processus de données une description de la structure ayant généré la série elle-même. En suivant ces démarches de modélisation on a abordé les résultats suivants :

- Les séries : « budget », « retraite », « recette » sont de type DS et la série « dépense » est de type TS
- On a l'effet saisonnier dans les 4 séries.

Cependant les deux méthodes précédentes ne prennent pas en compte les relations existantes entre plusieurs séries, pour cette raison nous avons proposé, par la suite, l'approche multivariée (VAR) et l'approche de cointégration bivariée.

Nous achevons notre étude par une comparaison entre les trois méthodes à la base du critère RMSE (Root Mean Square Error) qui a permis de conclure que : l'approche univariée de Box et Jenkins a donné de meilleurs résultats.

Bibliographie

- 1) **ADRAR. S et BEN BENNAI .N** , « Etude prévisionnelle des ventes d'huile par la méthode de BOX et JENKINS ,Cas pratique SPA CEVITAL ». UNIVERSITE ABDERRAHMANE MIRA DE BEJAIA ,Mémoire-(2012-2013) .
- 2) **AILIA, KESANI .F** , «Etude previsionnelle des ventes des produits laitiers au sein du LFB(Laiterie fromagerie de Boudouaou)» .Université M'hammed BOUGARA Boumerdes ,faculté des sciences , Boumerdes,Mémoire-(2006-2007).
- 3) **Bourbonnais R. Terraza . M** , « Analyse des séries Temporelles ». Dunod, Paris , Livre-(2004).
- 4) **Charpentier, A**, « Cours des séries temporelles. Théorie et Applications ». Université de Paris , Article-(2004) .
- 5) **Charpentier, A**, « Modèles de prévision Séries temporelles » .Article-(2012)
- 6) **Damodar N. Gujarati** , « Econométrie » .de boeck. Belgique , (2004) .
- 7) **Goujet Brigitte Doriath et Christian** ,«Gestion prévisionnelle et mesure de la performance » Dunod , Paris , Livre-(2007).
- 8) **HAMRITA.M H** , «Series temporelles » , Institut Supérieur de Mathematiques Appliqués Informatique, Kairaoun , Article-(Décembre 2013).
- 9) **Hamisultane Hélène**, «Modèle à Correction d'Erreur et Applications » Article.
- 10) **Jonas Kibala Kuma**, « Cointégration et Modèle à Correction d'Erreur : Pratiques sur Eviews ». Licence. Congo-Kinshasa. (2018) .
- 11) **LAGNOUX Agnes** , «Series chronologiques - Prevision par lissage exponentiel »Article.
- 12) **MAZIGH. I HADJOU. S** , «Prévision à court terme par l'approche de la cointégration bivariée et multivariée ,Boumerdes » ,UNIVERSITÉ DE M'HAMED BOUGERRA, Mémoire- (2016-2017).
- 13) **PARDOUX .C et GOLDFARB.B** , « Prévision a court terme : méthode de lissage exponentie»l, Université Paris -Dauphine .