Université Abderrahmane Mira-Bejaia

Faculté des Sciences Économiques, Commerciales et des Sciences de Gestion

Département des Sciences Commerciales

Polycopié pédagogique

Préparé par : Dr TOUATI Karima

Titre

Econométrie et statistiques appliquées

Cours destiné aux étudiants de

M1 Finance et Commerce International du département des sciences Commerciales

Année : 2022/2023

Sommaire

Introduction générale

L'économétrie est le principal outil d'analyse quantitative utilisé par les économistes et gestionnaires dans divers domaines d'application, comme la macroéconomie, la finance, le commerce ou le marketing. Les méthodes de l'économétrie permettent de vérifier l'existence de certaines relations entre des phénomènes économiques, et de mesurer concrètement ces relations, sur la base d'observations de faits réels.

Ce cours intitulé «Econométrie et statistiques appliquées » constitue une introduction aux méthodes statistiques qui permettent de tester la validité des théories économiques. Il vise à rendre les utilisateurs de ces méthodes aptes à choisir les techniques les plus adéquates pour résoudre un problème donné, à interpréter les résultats obtenus lors de leur application ainsi qu'à évaluer la validité des hypothèses sur lesquelles leurs propriétés optimales reposent.

Il permet de se familiariser avec les séries temporelles et les modèles nécessaires pour une meilleure représentation des phénomènes financiers et commerciaux et une bonne estimation des modèles, afin d'avoir des prévisions viables permettant la prise de décision.

Il a pour objet l'explication de différentes techniques statistiques et économétriques permettant d'estimer et de prévoir les séries macroéconomiques et financières.

Pour pouvoir tirer le maximum de ce cours, les étudiants doivent :

- Avoir des connaissances mathématiques (en particulier les calculs matriciels) ainsi que en statistiques descriptives, notamment les paramétrés de dispersion
- Avoir des connaissances en lois de probabilité

Le cours s'articule autour de cinq grands chapitres qui sont enrichis par des illustrations et exercices d'applications, selon la composition suivante :

Chapitre 1 portant intitulé « le modèle de régression simple », sera consacré à la présentation des formules de base permettant d'estimer les paramètres du modèle, les hypothèses stochastiques ainsi que l'examen de l'estimation d'un modèle à l'aide des premiers tests statistiques (Student, Fisher). Les exercices proposés se concentrent sur

l'estimation des relations linéaires par les moindres carrés ordinaires ainsi que l'application du test de signification individuel.

Chapitre 2 intitulé « *le modèle de régression simple* » constitue un prolongement du chapitre 1. Il sera présenté le modèle linéaire général, la procédure d'estimation des paramètres en étudiant les propriétés statistiques des estimateurs, les différents tests d'hypothèses concernant les coefficients du modèle et l'analyse de la variance ainsi qu'aux tests s'y rattachant.

Chapitre 3 : portant sur l'*Analyse des séries temporelles*, a pour objet l'analyse des séries chronologiques à travers, la définition de leurs caractéristiques, la détermination de leurs composantes, leurs représentations sous différents modèles et le calcul de leurs prévisions. Les exercices proposés se concentrent sur l'identification de la typologie des modèles et les méthodes adéquates pour corriger la tendance et les mouvements saisonniers.

Chapitre 4 : intitulé « Le modèle ARIMA et méthodologie de Box & Jenkins » traite des caractéristiques statistiques (en termes de stationnarité) des séries temporelles en présentant les différents tests (Dickey-Fuller**).** Les différentes classes de modèles (AR, MA, ARMA) en étudiant leurs propriétés. Enfin, la méthode Box et Jenkins qui systématise une démarche d'analyse des séries temporelles

Chapitre 5 portant sur « **Modélisation VAR et cointegration. »** traitera des modèles autorégressifs (VAR), VECM et ARDL.On se focalisera sur leurs représentations générales, estimation des paramètres, étude de la causalité et la dynamique du VAR.

Enfin, le module est dispensé sous forme de cours magistral et de travaux dirigés. Il est programmé pour un seul semestre. Le présent cours est destiné aux étudiants de master 1, option, Finance et Commerce International du département des sciences Commerciales. Notons que ces étudiants n'ont pas eu l'occasion d'étudier les notions de base relatives à l'économétrie.

L'évaluation de ce cours est sommative par un examen final pour établir un bilan de ce que les apprenants ont appris, puis ils sont classés par rapport un seuil d'acceptabilité (la moyenne). Toutefois, durant le cours, il est primordial de tester les prérequis de l'étudiant afin de s'assurer de sa prédisposition à assimiler le cours (évaluation pronostic). Concernant les travaux dirigés, l'évaluation est faite par une interrogation qui permet de situer le degré d'assimilation. Nous procédons, également, à une évaluation continue en prenant en considération l'assiduité et la participation de l'étudiant durant le semestre.

Chapitre 1 : Le modèle de régression simple

Dans le cadre de l'économétrie, un modèle consiste en une présentation formalisée d'un phénomène sous forme d'équations dont les variables sont des grandeurs économiques. L'objectif du modèle est de représenter les traits les plus marquants d'une réalité qu'il cherche à styliser. Le modèle est donc l'outil que le modélisateur utilise lorsqu'il cherche à comprendre et à expliquer des phénomènes. Pour ce faire, il émet des hypothèses et explicite des relations. [1]

Le modèle est donc une présentation schématique et partielle d'une réalité naturellement plus complexe. Toute la difficulté de la modélisation consiste à ne retenir que la ou les représentations intéressantes pour le problème que le modélisateur cherche à expliciter. Ce choix dépend de la nature du problème, du type de décision ou de l'étude à effectuer. La même réalité peut ainsi être formalisée de diverses manières en fonction des objectifs.[2]

Dans le modèle de régression simple, une variable endogène est expliquée par une variable exogène[3].

Il sera présenté dans ce chapitre les formules de base permettant d'estimer les paramètres du modèle, les hypothèses stochastiques et leurs conséquences ainsi que la qualité de l'estimation d'un modèle qui sera examinée à l'aide des premiers tests statistiques.

I. Présentation du modèle

A. Exemple introductif

Selon la théorie keynésienne, la consommation des manages est représentée par la relation : $C = f(Y)$ dont la forme est supposée linéaire $C = a_0 + a_1 Y$ dont la série d'observation $(C_t , Y_t)/t \in (1, T)$ est disponible. Le modèle qui est déterministe[4] n'est pas un modèle économétrique. Pour construire la relation qui existe entre le revenu Y et la consommation C , il est supposé que les (C_t , Y_t) correspondent à un échantillon de

[2] Bourbonnais Régis (2018) Chapitre 1. Qu'est-ce que l'économétrie ? Pages 1 à 12 https://www.cairn.info/econometrie--9782100773459-page-1.htm

[3] Bourbonnais Régis « économétrie : cours et exercices corrigés», 9eme édition Dunod, Paris, 2015.

[4] Les modèles déterministes sont basés sur une loi connue ou hypothétique de la physique, des mathématiques ou d'une quelconque autre discipline, de sorte que des valeurs d'input données produisent toujours le même résultat. Par contre, le modèle stochastique accepte une certaine distribution de probabilité associée à des inputs donnés, dans les processus au sein du modèle et donc dans l'output, de sorte que le même input peut amener à différentes valeur d'output.

ménages caractéristique d'une population de ménages beaucoup plus vaste ; c.à.d que (y_t , x_t) est un échantillon dans une distribution à deux dimension, c.à.d que pour une valeur du revenu x déterminée, on observera des valeurs différents de la variable y , correspondante chacune à une observation particulière.

Soit la fonction de consommation keynésienne [5] : $C = a_0 + a_1 Y$

où : C = consommation, Y = revenu, a_1= propension marginale à consommer,

a_0 = consommation autonome ou incompressible.

La variable consommation est appelée « variable à expliquer » ou « variable endogène ». La variable revenu est appelée « variable explicative » ou « variable exogène » (c'est le revenu qui explique la consommation). a_1 et a_0 sont les paramètres du modèle ou encore les coefficients de régression.

Deux types de spécifications peuvent être distingués :

- Les modèles en série temporelle, les variables représentent des phénomènes observés à intervalles de temps réguliers, par exemple la consommation et le revenu annuel sur 20 ans pour un pays donné. Le modèle s'écrit alors :$C_t = a_0 + a_1 Y_t$ t=1,...,20 où :C_t=consommation au temps t, Y_t = revenu au temps t.
- Les modèles en coupe instantanée, les variables représentent des phénomènes observés au même instant mais concernant plusieurs individus, par exemple la consommation et le revenu observés sur un échantillon de 20 pays. Le modèle s'écrit alors :$C_i = a_0 + a_1 Y_i$ i =1,...,20 où : C_i=consommation du pays *i* pour une année donnée, Y_i= revenu du pays *i* pour une année donnée.

B. Rôle du terme aléatoire

Le modèle tel qu'il vient d'être spécifié n'est qu'une caricature de la réalité. En effet ne retenir que le revenu pour expliquer la consommation est insuffisant ; il existe une multitude d'autres facteurs susceptibles d'expliquer la consommation. C'est pourquoi il est ajouté un terme (ε_t) qui synthétise l'ensemble de ces informations non explicitées dans le modèle : $C_t = a_0 + a_1 Y_t + \varepsilon_t$ si le modèle est spécifié en série temporelle ($C_i = a_0 + a_1 Y_i + \varepsilon_i$ si le modèle est spécifié en coupe instantanée), où ε_t représente l'erreur de spécification du modèle, c'est-à-dire l'ensemble des phénomènes explicatifs de la consommation non liés au revenu. Le terme ε_t mesure la

[5] Idem

différence entre les valeurs réellement observées de C_t et les valeurs qui auraient été observées si la relation spécifiée avait été rigoureusement exacte. Le terme ε_t regroupe donc trois erreurs :

- une erreur de spécification, c'est-à-dire le fait que la seule variable explicative n'est pas suffisante pour rendre compte de la totalité du phénomène expliqué ;
- une erreur de mesure, les données ne représentent pas exactement le phénomène
- une erreur de fluctuation d'échantillonnage, d'un échantillon à l'autre les observations, et donc les estimations, sont légèrement différentes.

Les valeurs vraies a_0 et a_1 ne sont pas connues mais seulement les deux séries d'observations C_t et R_t. Les estimateurs de a_0 et a_1 , notés respectivement $\hat{a}_0$ et , $\hat{a}_1$ sont des variables aléatoires, qui suivent les mêmes lois de probabilité, celle de ε_t, puisqu'ils sont fonctions de la variable aléatoire ε_t. Les caractéristiques de moyenne et d'écart type de ces coefficients permettent de construire des tests de validité du modèle estimé.

II. Estimation des paramètres

A. Modèle et hypothèses

Soit le modèle suivant : $y_t = a_0 + a_1 x_t + \varepsilon_t$ pour $t = 1, \ldots\ldots, n$ avec :

y_t = variable à expliquer au temps t;

x_t = variable explicative au temps t;

a_0, a_1 = paramètres du modèle ;

ε_t = erreur de spécification (différence entre le modèle vrai et le modèle spécifié), cette erreur est inconnue et restera inconnue ;

n= nombre d'observations.

Hypothèses

• H1 : le modèle est linéaire en x_t (ou en n'importe quelle transformation de x_t).

• H2 : les valeurs x_t sont observées sans erreur (x_t non aléatoire).

• H3 : $E(\varepsilon_t) = 0$, l'espérance mathématique de l'erreur est nulle : en moyenne le modèle est bien spécifié et donc l'erreur moyenne est nulle. Cette hypothèse signifie que les facteurs secondaires n'ont pas un effet systématique jouant à la hausse ou à la baisse sur la variable y .

• H4 : $E(\varepsilon t^2) = \sigma_\varepsilon^2$, la variance de l'erreur est constante[6]: le risque de l'amplitude de l'erreur est le même quelle que soit la période. Cette hypothèse, souvent appelée la propriété de l'homoscédasticité, traduit l'idée que l'amplitude de la variabilité de l'aléa provenant des facteurs secondaires est invariante à travers les individus ou à travers le temps.[7]

• H5 : $E(\varepsilon_t \varepsilon_t') = 0$ si $t \neq t'$, les erreurs sont non corrélées (ou encore indépendantes) : une erreur à l'instant t n'a pas d'influence sur les erreurs suivantes.

• H6 : $Cov(x_t, \varepsilon_t) = 0$, l'erreur est indépendante de la variable explicative.

B. Formulation des estimateurs

En traçant un graphique (1) des couples de données liant le revenu et la consommation observée, nous obtenons un nuage de points que nous pouvons ajuster à l'aide d'une droite. Le principe de base de la méthode des MCO est de choisir parmi toutes les droites possibles celle qui minimise l'écart entre les réalisations de la variable expliquée et les valeurs prévus par le modèle estimé. Mais pour éviter la compensation entre les écarts négatifs et positifs, la minimisation porte sur les erreurs quadratiques comptées parallèlement à l'axe de la variable expliqué.[8] On cherche une droite, d'équation : $y_t = \hat{a}_0 + \hat{a}_1 x_t$ qui approche « au mieux » les données. On l'appelle droite des moindres carrées de y en x ou droite de régression de y en x.

L'estimateur des coefficients a_0 et a_1 est obtenu donc en minimisant la distance au carré entre chaque observation et la droite, d'où le nom d'estimateur des moindres carrés ordinaires (MCO).

La résolution analytique est la suivante :

$$Min \sum_{t=1}^{n} \varepsilon_t{}^2 = Min \sum_{t=1}^{n} (y_t - a_0 - a_1 x_t)^2 = Min\ S$$

En opérant par dérivation par rapport à a0et a1afin de trouver le minimum[9] de cette fonction, on obtient les résultats suivants :

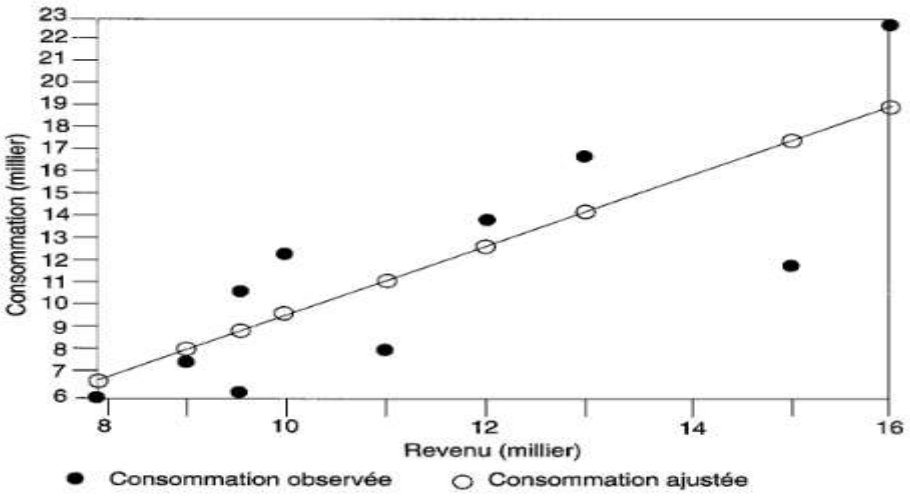
$$\frac{\partial S}{\partial a_0} = -2 \sum_t (y_t - \hat{a}_0 - \hat{a}_1 x_t) = 0$$

[6] Dans le cas où cette hypothèse n'est pas vérifiée, on parle alors de modèle hétéroscédastique.
[7] Sami Mestiri (2021) Le modèle de régression linéaire simple, Faculté des sciences économiques et de gestion de Mahdia, p.8.
[8] Sami Mestiri (2021), *Op.cit.*, p.9.
[9] Les conditions du deuxième ordre sont considérées comme vérifiées car la fonction est convexe.

Et $\frac{\partial S}{\partial a_1} = -2 \sum_t x_t \, (y_t - \hat{a}_0 - \hat{a}_1 \, x_t \,) = 0$

Graphiquement, il s'agira de tracer une droite linéaire qui minimisera l'écart (mis au carré) entre chaque couple (*X*,*Y*) et son point correspondant sur la droite de régression.

Graphique 1 – Ajustement d'un nuage de points par une droite



Source : Bourbonnais Régis, « économétrie : cours et exercices corrigés», 9eme édition Dunod, Paris, 2015

Sommant par rapport à t, il vient : $\sum_t x_t \; y_t - \; \hat{a}_0 \; \sum_t x_t \; - \hat{a}_1 \; \sum_t {x_t}^2 = 0$

$\sum_t y_t \; - n \; \hat{a}_0 - \hat{a}_1 \; \sum_t x_t = 0$

Qui est sont appelées les équations normales et qui impliquent que :

$$\hat{a}_1 = \frac{\sum_{t=1}^{n} (x_t - \bar{x})(y_t - \bar{y})}{\sum_{t=1}^{n} (x_t - \bar{x})^2}$$

$$\hat{a}_0 = \bar{y} - \hat{a}_1 \, \bar{x}$$

La théorie économique postule parfois des relations dans lesquelles $a_0 = 0$: c'est le cas par exemple pour une fonction de production de produit industriel où le facteur de production (unique) nul entraîne une production nulle. L'estimation de a_1 est alors donnée par la formule suivante :

$$\hat{a}_1 = \frac{\sum_{t=1}^{n} x_t y_t}{\sum_{t=1}^{n} {x_t}^2}$$

C. Les différentes écritures du modèle : erreur et résidu

Le modèle de régression simple peut s'écrire sous deux formes selon qu'il s'agit du modèle théorique spécifié par l'économiste ou du modèle estimé à partir d'un échantillon.

• Modèle théorique spécifié par l'économiste avec ε_t l'erreur inconnue :

$$y_t = a_0 + a_1 x_t + \varepsilon_t$$

• Modèle estimé à partir d'un échantillon d'observations :

$$y_t = \hat{a}_0 + \hat{a}_1 x_t + e_t = \widehat{y_t} + e_t \qquad e_t = \text{résidu}$$

Le résidu observé et est donc la différence entre les valeurs observées de la variable à expliquer et les valeurs ajustées à l'aide des estimations des coefficients du modèle ; ou encore : $\widehat{y}_t = \hat{a}_0 + \hat{a}_1 x_t$

III. Construction des tests

A. Hypothèse de normalité des erreurs

Cette hypothèse n'est pas indispensable afin d'obtenir des estimateurs convergents mais elle permet de construire des tests statistiques concernant la validité du modèle estimé.

B. Conséquences de l'hypothèse de normalité des erreurs

L'estimateur de la variance de l'erreur (σ_ε^2) noté $\hat{\sigma}_\varepsilon^2$ est donc égal à :

$$\widehat{\sigma}^2{}_\varepsilon = \frac{1}{(n-2)} \sum_t e_t^2$$

Les estimateurs empiriques de la variance de chacun des coefficients

$$\widehat{\sigma}^2{}_{\hat{a}_1} = \frac{\widehat{\sigma}^2{}_\varepsilon}{\sum_{t=1}^{n} (x_t - \bar{x})^2}$$

$$\widehat{\sigma}^2{}_{\hat{a}_0} = \widehat{\sigma}^2{}_\varepsilon \left(\frac{1}{n} + \frac{\bar{x}^2}{\sum_{t=1}^{n} (x_t - \bar{x})^2} \right)$$

Ces estimateurs sont convergents et sans biais.

L'hypothèse de normalité des erreurs implique que : $\frac{\hat{a}_1 - a_1}{\sigma_{\hat{a}_1}}$ et $\frac{\hat{a}_0 - a_0}{\sigma_{\hat{a}_0}}$ suivent une loi normale centrée réduite N (0, 1).

Les estimateurs sont convergents et sans biais.

Valeurs ajustées, résidus et somme des carrés des résidus : Une fois les coefficients de la droite estimés, on calcule pour chaque individu : [10]

- $\hat{y}_t = \hat{a}_0 + \hat{a}_1 x_t$ s'appelle la valeur ajustée ou prédite de *y* par le modèle.
- $e_t = y_t - \hat{y}_t$ s'appelle le résidu de l'observation. C'est l'écart entre la valeur de Y observée sur l'individu et la valeur prédite. Le résidu e_t est une approximation du terme d'erreur εi .
- la somme des carrés des résidus est SCR =$\sum_t e_t^2$. Elle mesure la distance de la droite de régression aux points du nuage de points qui est minimale au sens des moindres carrés.
- La statistique $\hat{\sigma}^2_{\varepsilon} = \frac{\sum_t e_t^2}{(n-2)}$ est un estimateur sans biais de σ 2

C. Test bilatéral, test unilatéral et probabilité critique d'un test

1) Test bilatéral

Soit à tester, à un seuil de 5 %, l'hypothèse H0 :a_1= 0 contre l'hypothèseH1 :$a_1 \neq 0$.

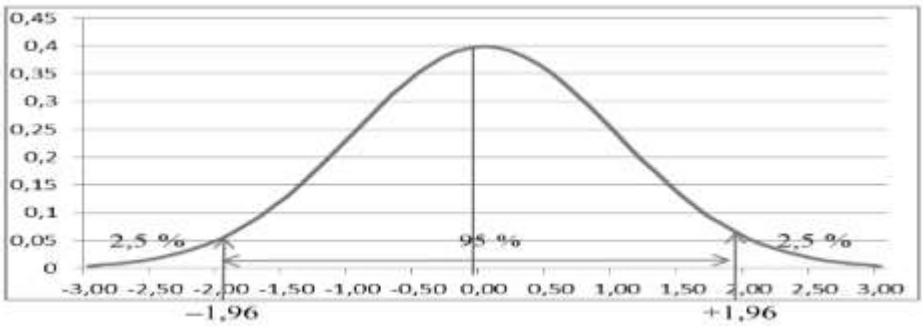
Nous savons que $\frac{\hat{a}_1 - a_1}{\sigma_{\hat{a}_1}}$ suit une loi de student à (n - 2) degrés de liberté**.**

Sous H0 (a_1= 0) le ratio appelé ratio de Student $\frac{\hat{a}_1 - 0}{\sigma_{\hat{a}_1}}$ suit donc une loi de Student à (n– 2) degrés de liberté[11]. Le test d'hypothèses bilatéral consiste donc à comparer le ratio de Student empirique $t^* = \frac{|\hat{a}_1|}{\hat{\sigma}_{\hat{a}_1}}$ à la valeur du t de Student lue dans la table à n– 2 degrés de liberté1et pour un seuil de probabilité égal à 5 %, soit si n–2 > 30, $t^{0,05}_{\infty}$ =1,96. Si t*> $t^{0,05}_{\infty}$ =1,96, nous rejetons l'hypothèse H0 (graphique 2), le coefficient théorique et inconnu $a_1$est significativement différent de 0.

[10] Cours de Méthodes statistiques pour l'analyse des données en psychologie, Master 1, Université Paris Ouest Nanterre La Défense UFR SPSE- PMP STA 21 https://fermin.perso.math.cnrs.fr/Files/Chap3.pdf, p 6.

[11] La notion de degré de liberté correspond au nombre de valeurs restant réellement à disposition après une procédure d'estimation statistique. Si un échantillon comprend 10 observations etqu'on dispose en plus de la moyenne de cet échantillon, on ne peut choisir librement les valeurs que pour 9 de ces observations, la dixième se déduisant de la valeur de la moyenne. Dans le cas présent, le modèle de régression simple, le nombre de degrés de liberté est donc de n–2 car nous avons estimé deux paramètres $a_0$et a_1.

Graphique 2 – Test bilatéral à 5 %

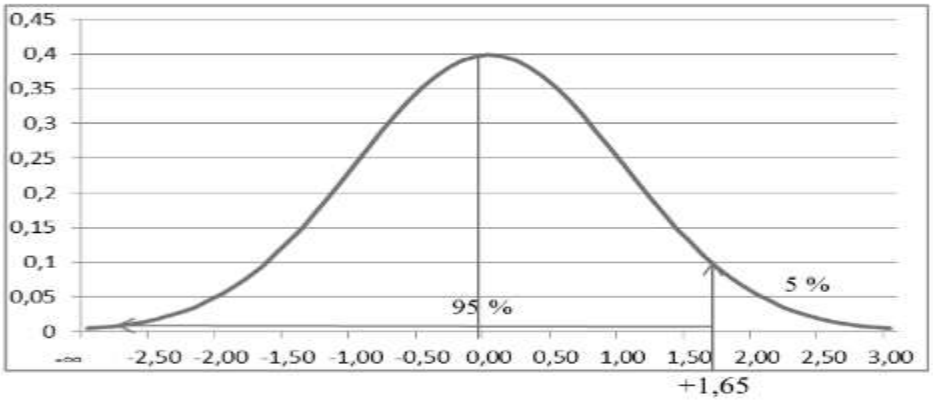


2) Test unilatéral

Soit à tester, à un seuil de 5 %, l'hypothèse H0 :a_1= 0 contre l'hypothèseH1 :a_1> 0 ou a_1< 0 selon que le coefficient estimé soit positif ou négatif.

Le test d'hypothèses unilatéral consiste donc à comparer le ratio de Student empirique $t^* = \left|\frac{\widehat{a_1}}{\sigma_{\widehat{a_1}}}\right|$ à la valeur du t de Student lue dans la table à n– 2 degrés de liberté et pour un seuil de probabilité égal à 5 %, soit si n– 2 > 30, $t^{0,05}_{\infty}$ =1,65. Si t*> $t^{0,05}_{\infty}$ =1,65 nous rejetons l'hypothèse H0 (voir graphique 3), le coefficient théorique et inconnu $a_1$est significativement différent de 0.[12]

Graphique 3 – Test unilatéral à 5 % (H1 : a1> 0)



IV. Équation et tableau d'analyse de la variance

A. Équation d'analyse de la variance

L'équation fondamentale d'analyse de la variance :

$$\sum_t (y_t - \bar{y})^2 = \sum_t (\hat{y}_t - \bar{y})^2 + \sum_t e_t^2$$

[12] La table de Student (donnée en annexe 1) est tabulée pour les tests bilatéraux, il faut donc lire à 10 % = 2 ×0,05.

$$\mathbf{SCT} \quad = \quad \mathbf{SCE} \quad + \mathbf{SCR}$$

La variabilité totale (SCT) est égale à la variabilité expliquée (SCE) + la variabilité des résidus (SCR).

Cette équation permet de juger de la qualité de l'ajustement d'un modèle. En effet, plus la variance expliquée est proche de la variance totale, meilleur est l'ajustement du nuage de points par la droite des moindres carrés. Il est d'usage de calculer le rapport :

$$R^2 = \frac{\sum_t (\widehat{y_t} - \bar{y})^2}{\sum_t (y_t - \bar{y})^2} = 1 - \frac{\sum_t e_t^2}{\sum_t (y_t - \bar{y})^2}$$

R^2 est appelé le coefficient de détermination, et R le coefficient de corrélation multiple (dans le cas particulier du modèle de régression à une seule variable explicative, il est égal au coefficient de corrélation linéaire simple entre *x* et *y*).

B. Tableau d'analyse de la variance

Le tableau présente l'analyse de la variance pour un modèle de régression simple.

Tableau : Analyse de la variance par une régression simple.

Source de variation	**Sommes des carrées**	**Degrés de liberté**	**Carées moyen**
x	$SCE = \sum_t (\hat{y}_t - \bar{y})^2$	1	SCE/1
Résidu	$SCR = \sum_t e_t^2$	n -2	SCR/n-2
Total	$SCT = \sum_t (y_t - \bar{y})^2$	n-1	

Les degrés de liberté correspondent au nombre de valeurs que nous pouvons choisir arbitrairement (par exemple, pour la variabilité totale, connaissant n−1 valeurs, nous pourrons en déduire la n-ième, puisque nous connaissons la moyenne $\bar{y}$.

Le test H0: a_1=0 est équivalent au test d'hypothèse[13]H0: SCE= 0 (la variable explicative x_t ne contribue pas à l'explication du modèle).

Soit le test d'hypothèses H0: SCE = 0 contre l'hypothèse H1:SCE≠ 0.

La statistique[14]de ce test est donnée par

[13] Cela n'est vrai que dans le cas du modèle de régression simple

$$F^* = \frac{\frac{SCE}{ddl_{SCE}}}{\frac{SCR}{ddl_{SCR}}} = \frac{\frac{\sum_t(\widehat{y_t}-\bar{y})^2}{1}}{\frac{\sum_t e_t^2}{n-2}}$$

Ou encore : $F^* \frac{\frac{SCE}{ddl_{SCE}}}{\frac{SCR}{ddl_{SCR}}} = \frac{\frac{\sum_t(\widehat{y_t}-\bar{y})^2}{1}}{\frac{\sum_t e_t^2}{n-2}} = \frac{\frac{R^2}{1}}{\frac{1-R^2}{n-2}}$

La statistique F^* est le rapport de la somme des carrés expliqués par x_t sur la somme des carrés des résidus, chacune de ces sommes étant divisée par son degré de liberté respectif. Ainsi, si la variance expliquée est significativement supérieure à la variance résiduelle, la variable x_t est considérée comme étant une variable réellement explicative.

F^* suit une statistique de Fisher à 1 et n−2 degrés de liberté. Si $F^* > F^{\alpha}_{1\,;\,n-2}$ nous rejetons au seuil α l'hypothèse H_0 d'égalité des variances, la variable x_t est significative ; dans le cas contraire, nous acceptons l'hypothèse d'égalité des variances, la variable x_t n'est pas explicative de la variable y_t.

C. **La prévision dans le modèle de régression simple**

Lorsque les coefficients du modèle ont été estimés, il est possible de calculer une prévision à un horizon *h* . Soit le modèle estimé sur la période t = 1,. . . , n : $\hat{y}_t = \hat{a}_0 + \hat{a}_1 x_t + e_t$, si la valeur de la variable explicative x_t est connue en n+1 (x_{n+1}), la prévision est donnée par $\hat{y}_{n+1} = \hat{a}_0 + \hat{a}_1 x_{n+1}$. Il est montré que cette prévision est sans biais[15]

Série de TD N°1

[14] Nous comparons la somme des carrés expliqués SCE à la somme des carrés des résidus SCR qui est représentative de la somme des carrés théoriquement la plus faible.

[15] Source : Bourbonnais Régis, « économétrie : cours et exercices corrigés», 9eme édition Dunod, Paris, 2015, P 39.

Exercice 1 : Un agronome s'intéresse à la liaison pouvant exister entre le rendement de maïs y (en quintal) d'une parcelle de terre et la quantité d'engrais *x* (en kilo). Il relève 10 couples de données consignés dans le tableau suivant :

	X	**Y**
1	16	20
2	18	24
3	23	28
4	24	22
5	28	32
6	29	28
7	26	32
8	31	36
9	32	41
10	34	41

Travail à faire :

a) Tracer le nuage de points et le commenter.

b) Estimer les paramètres du modèle de régression linéaire par la méthode des MCO.

c) Tester au seuil de 5% la significativité du coefficient associé à la variable X

d) Construire le tableau d'analyse de la variance correspondant à cette régression.

Exercice 2 :

Soit les résultats d'une estimation économétrique :

y_t = 1,251- 32,95 +e_t ; n = 20 ; R^2 =0,23 . $\widehat{\sigma}_{\varepsilon}$ = 10,66

a) A partir des informations connues, on demande de retrouver les statistiques suivantes : la somme des carrés des résidus (SCR), la somme des carrés totaux (SCT), la somme des carrés expliqués (SCE), la valeur de la statistique du Fisher empirique (F^*) et l'écart type du coefficient $\widehat{\sigma}_{\hat{a}_1}$

b) Le coefficient de la variable x est-il significativement différent de 0 ?

Exercice 3

Le tableau 2 présente le revenu moyen par habitant sur 10 ans exprimé en dollars pour un pays (x_t) et la consommation (y_t)

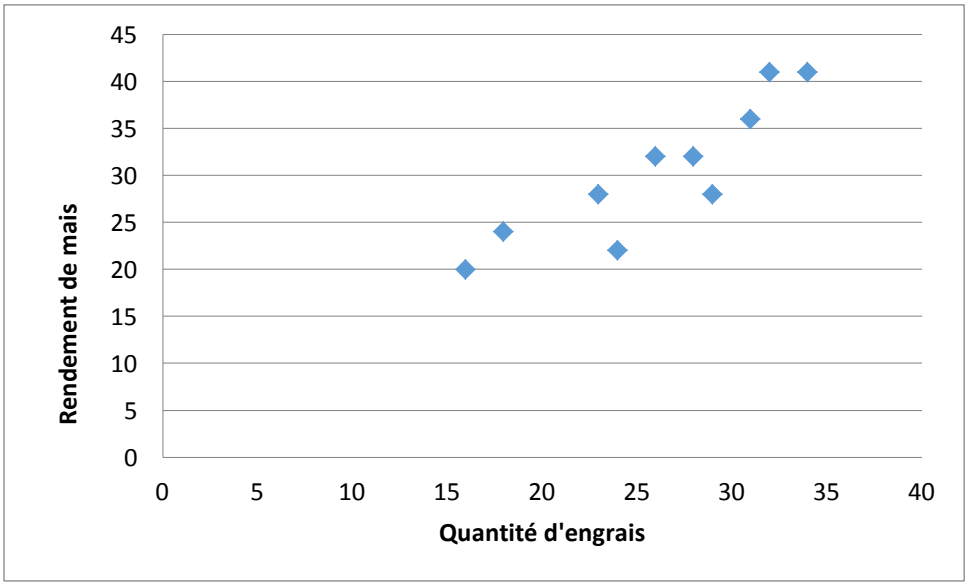
(1) t	(2) y_t	(3) x_t	(4) $y_t - \overline{y}$	(5) $x_t - \overline{x}$	(6) (5)* (5)	(7) (4)* (5)
1	7 389,99	8 000	– 2 595,59	– 3 280	10 758 400	8 513 518
2	8 169,65	9 000	– 1 815,93	– 2 280	5 198 400	4 140 300
3	8 831,71	9 500	– 1 153,87	– 1 780	3 168 400	2 053 879
4	8 652,84	9 500	– 1 332,74	– 1 780	3 168 400	2 372 268
5	8 788,08	9 800	– 1 197,50	– 1 480	2 190 400	1 772 292
6	9 616,21	11 000	– 369,37	– 280	78 400	103 422
7	10 593,45	12 000	607,88	720	518 400	437 670
8	11 186,11	13 000	1 200,54	1 720	2 958 400	2 064 920
9	12 758,09	15 000	2 772,52	3 720	13 838 400	10 313 755
10	13 869,62	16 000	3 884,05	4 720	22 278 400	18 332 692
Somme	99 855,75	112 800	0	0	64 156 000	50 104 729
Moyenne	9 985,57	11 280	0	0	6 415 600	5 010 472

- À partir des données du tableau 2 de l'exercice 1, on demande de calculer les estimations de $\hat{a}_0$ et $\hat{a}_1$.
- La propension marginale à consommer est-elle significativement différente de 0 ?
- Quel est l'intervalle de confiance au seuil (ou niveau) de 95 % pour la propension marginale à consommer ?

Corrigé de la Série de TD N°1

Exercice 1

a) Graphique nuage de points "Rendements Agricoles"



b) Estimation les paramètres du modèle de régression linéaire par la méthode des MCO

	x_t	y_t	$x_t - \bar{x}$	$y_t - \bar{y}$	$(x_t - \bar{x})(y_t - \bar{y})$	$(x_t - \bar{x})^2$	$(y_t - \bar{y})^2$
1	16,00	20,00	-10,10	-10,40	105,04	102,01	108,16
2	18,00	24,00	-8,10	-6,40	51,84	65,61	40,96
3	23,00	28,00	-3,10	-2,40	7,44	9,61	5,76
4	24,00	22,00	-2,10	-8,40	17,64	4,41	70,56
5	28,00	32,00	1,90	1,60	3,04	3,61	2,56
6	29,00	28,00	2,90	-2,40	-6,96	8,41	5,76
7	26,00	32,00	-0,10	1,60	-0,16	0,01	2,56
8	31,00	36,00	4,90	5,60	27,44	24,01	31,36
9	32,00	41,00	5,90	10,60	62,54	34,81	112,36
10	34,00	41,00	7,90	10,60	83,74	62,41	112,36
Somme	261,00	304,00	0,00	0,00	**351,6**	**314,9**	**492,4**
Moyenne	$\bar{x}$ = 26,10	$\bar{y}$ = 30,40					

$\hat{a}_1 = \frac{\sum_{t=1}^{n}(x_t-\bar{x})(y_t-\bar{y})}{\sum_{t=1}^{n}(x_t-\bar{x})^2} = \frac{351,6}{314,9}$ = **1,1165**

$\hat{a}_0 = \bar{y} - \hat{a}_1 \bar{x}$ = 30,40- 1,1165 (26,10)= 30,40- 29,1418 = **1,26**

$\hat{y}_t = \hat{a}_0 + \hat{a}_1 x_t$ = 1,26 +1,1165 x_t ; $e_t = y_t - \hat{y}_t$

Exemple : $\hat{y}_1$ = 1,26 +1,1165(16)= **19,124**

$e_1 = y_1 - \hat{y}_{t1}$ = 20 - 19,124 = **0,88**.

Tableau – Calcul du résidu d'estimation

t	$\hat{y}_t = \hat{a}_0 + \hat{a}_1 x_t$	$e_t = y_t - \hat{y}_t$	e_t^2
1	19,124	0,88	0,767376
2	21,357	2,64	6,985449
3	26,9395	1,06	1,12466025
4	28,056	-6,06	36,675136
5	32,522	-0,52	0,272484
6	33,6385	-5,64	31,7926823
7	30,289	1,71	2,927521
8	35,8715	0,13	0,01651225
9	36,988	4,01	16,096144
10	39,221	1,78	3,164841
Somme	292,6665	0,00	**99,82280**

c) **Tester de au seuil de 5% la significativité du coefficient associé à x_t**

H0 = $\hat{a}_1 = 0$; H1= $\hat{a}_1 \neq 0$

Sous l'hypothèse H0, cette relation devient : $\frac{\hat{a}_1 - 0}{\hat{\sigma}_{\hat{a}_1}} = \frac{\hat{a}_1}{\hat{\sigma}_{\hat{a}_1}} = t^*_{\hat{a}_1}$ → loi de Student à (n– 2) degrés de liberté.

$\hat{\sigma}^2{}_{\hat{a}_1} = \frac{\hat{\sigma}^2{}_\varepsilon}{\sum_{t=1}^{n}(x_t - \bar{x})^2}$; Nous savons que $\hat{\sigma}^2{}_\varepsilon = \frac{\sum_t e_t^2}{(n-2)}$

$\hat{\sigma}^2{}_\varepsilon = \frac{\sum_t e_t^2}{(n-2)} = \frac{99{,}82280}{10-2} = 12{,}4778$

$\hat{\sigma}^2{}_{\hat{a}_1} = \frac{12{,}4778}{314{,}9} = 0{,}0396$; $\sigma_{\hat{a}_1} = \sqrt{\hat{\sigma}^2{}_{\hat{a}_1}} = \sqrt{0{,}0396} = 1{,}198$

Donc : $t^* = \frac{|\hat{a}_1|}{\hat{\sigma}_{\hat{a}_1}} = \frac{1{,}11}{0{,}198} = 5{,}60$

Donc $t^* = 5{,}60 > T_{2.8}{}^{0.05} = 2{,}306$

Le coefficient $\hat{a}_1$ est donc significativement différente de 0, la variable explicative X est est bien explicative de la variable Y

Le tableau d'analyse de la variance correspondant à cette régression

Source de variation	**Sommes des carrées**	**Degres de liberté**	**Carées moyen**
X	**492,4**	**1**	**492,4**
Résidu	**99,82280**	**8**	**12,47785**
Total	**592,2228**	**9**	

Exercice 2 :

Nous avons $\delta(\varepsilon) = \sqrt{SCR/(n-2)} = 10{,}66 \Rightarrow$ SCR = (10,66)2 × 18 = 2 045,44

Nous pouvons calculer SCE et SCT à l'aide du coefficient de détermination.
R2 = 0,23 = 1 − SC R/SCT

SC T = SC R/(1 − R2) = 2045,44/(1 − 0,23) = 2 656,42

Or SCT = SCE + SCR ⇒ SCE = 610,98

Nous pouvons calculer maintenant $F^* = \frac{R^2}{(1-R^2)(n-2)} = \frac{SCE}{SCR/(n-2)} = 5{,}40$

Dans le cas d'un modèle de régression simple , $t^{*2} = F^*$; $t^* = \sqrt{F^*} = 2{,}32$

Nous avons $t^*_{a_1} = \frac{\hat{a}_1}{\sigma_{(\hat{a}_1)}}$, L'écart type du coefficient : $\sigma_{(\hat{a}_1)} = \frac{\hat{a}_1}{t^*_{a_1}} = \frac{1,251}{2,32} = 0,54$

On pose le test d'hypothèses : H0 : $a_1 = 0$ contre l'hypothèse $H_1 : a_1 = \neq 0$

Sous H0, nous pouvons écrire $t^*_{a_1} = \frac{\hat{a}_1}{\sigma_{(\hat{a}_1)}} = 0,46 < t^{5\%}_8$

Nous sommes donc dans la zone de l'acceptation de H_0, le coefficient a_1 n'est pas significativement différent de 0.

Exercice 3

1) **La propension marginale à consommer est-elle significativement différente de 0 ?**

$\hat{a}_1 = \frac{\sum_{t=1}^{n}(x_t - \bar{x})(y_t - \bar{y})}{\sum_{t=1}^{n}(x_t - \bar{x})^2} = \frac{50104729}{61156000} = 0,78$

$\hat{a}_0 = \bar{y} - \hat{a}_1 \bar{x}$ = 9985,57- 0,78 (11280) = 1176, 08

La significativité est très importante en économétrie. En effet, dans le cas d'une réponse négative (le coefficient n'est pas significativement différent de 0) la variable explicative Revenu ne sera pas considérée comme étant explicative de la consommation puisque son coefficient de pondération est nul.

Si nous rejetons l'hypothèse H0, à un seuil α fixé, alors la propension marginale à consommer est considérée comme étant significativement différente de 0. Le seuil le plus communément employé est α =0,05, soit un risque de rejeter à tort H0 de 5 %.

Nous savons que : $\frac{\hat{a}_1 - a_1}{\hat{\sigma}_{\hat{a}_1}}$ suit donc une loi de Student à (n– 2) degrés de liberté

Sous l'hypothèse H0, cette relation devient : $\frac{\hat{a}_1 - 0}{\hat{\sigma}_{\hat{a}_1}} = \frac{\hat{a}_1}{\hat{\sigma}_{\hat{a}_1}} = t^*_{\hat{a}_1}$ → loi de Student à (n– 2) degrés de liberté.

Nous savons que $\hat{\sigma}^2_{\hat{a}_1} = \frac{\hat{\sigma}^2_{\varepsilon}}{\sum_{t=1}^{n}(x_t - \bar{x})^2}$, nous connaissons $\sum_{t=1}^{n}(x_t - \bar{x})^2 = 64156000$

On doit déterminer la valeur de $\hat{\sigma}^2_{\varepsilon}$. L'estimateur de la variance de l'erreur nous est donné comme suit : $\hat{\sigma}^2_{\varepsilon} = \frac{1}{(n-2)}\sum_t e_t^2$

Nous devons caliculer la valeur de e_t^2. Pour cela il faut déterminer la série ajustée $\hat{y}_t$

Calcul de $\hat{y}_t$ e_t

La série ajustée $\hat{y}_t$ est calculée par application des estimations $\hat{a}_0$ et $\hat{a}_1$

$\hat{y}_t = \hat{a}_0 + \hat{a}_1 x_t$, exemple : $\hat{y}_1 = 1\,176{,}08 + 0{,}78 \times 8\,000 = 7\,423{,}95$;

$e_t = y_t - \hat{y}_t$, exemple : $e_1 = y_1 - \hat{y}_{t1} = 7\,389{,}99 - 7\,423{,}95 = -33{,}96$

Les résultats sont illustrés dans le tableau suivant :

Tableau – Calcul du résidu d'estimation

$\widehat{y_t}$	e_t	e_t^2
7 423,95	– 33,96	1 153,38
8 204,93	– 35,28	1 244,98
8 595,43	236,28	55 830,26
8 595,43	57,41	3 296,40
8 829,72	– 41,64	1 733,93
9 766,90	– 150,69	22 707,42
10 547,88	45,57	2 076,39
11 328,87	– 142,76	20 379,08
12 890,83	– 132,74	17 620,12
13 671,81	197,81	39 127,38
Somme	0,00	165 169,3
Moyenne	0,00	16 516,93

b) Calcul de l'estimation de la variance de l'erreur et de l'écart type du coefficient de régression. L'estimation de la variance de l'erreur est donc égale à

$$\hat{\sigma}^2{}_{\varepsilon} = \frac{\sum_t e_t^2}{(n-2)} = \frac{165\,169{,}3}{8} = 20\,646{,}16$$

Ce qui nous permet de calculer la variance estimée de $\hat{a}_1$

$$\hat{\sigma}^2{}_{\hat{a}_1} = \frac{\hat{\sigma}^2{}_{\varepsilon}}{\sum_{t=1}^{n}(x_t - \bar{x})^2} = \frac{20\,646{,}16}{64\,156\,000} = 0{,}000\,321\,8 \text{ Donc } \sigma_{\hat{a}_1} = 0{,}017\,9$$

c) Calcul du ratio de *Student* et règle de décision.

Ce test permet de tester la pertinence d'une variable explicative qui figure dans un modèle et sa contribution à l'explication du phénomène que l'on cherche à modéliser. Dans notre exemple, nous calculons le ratio de Student :

Nous avons $t^*{}_{a_1} = \frac{\hat{a}_1}{\delta(\hat{a}_1)} = \frac{0{,}78}{0{,}0179} = 43{,}7 : t^*{}_{a_1} > t^{5/2\%}{}_8 = 2{,}306$

La propension marginale à consommer est donc significativement différente de 0, la variable Revenu est bien explicative de la variable.

Chapitre 2. Le modèle de régression multiple

Le modèle de régression multiple est une extension du modèle de régression simple. Dans ce présent chapitre, il sera présenté le modèle linéaire général, la procédure d'estimation des paramètres en étudiant les propriétés statistiques des estimateurs, les différents tests d'hypothèses concernant les coefficients du modèle et l'analyse de la variance ainsi qu'aux tests s'y rattachant.

I. Le modèle linéaire général

1.1 . Présentation

Lors du chapitre précédent, nous avons considéré qu'une variable endogène est expliquée à l'aide d'une seule variable exogène. Cependant, il est extrêmement rare qu'un phénomène économique ou social puisse être appréhendé par une seule variable. Le modèle linéaire général est une généralisation du modèle de régression simple dans lequel figurent plusieurs variables explicatives :

$$y_t = a_0 + a_1 x_{1t} + a_2 x_{2t} + \cdots . a_k x_{kt} + \varepsilon_t \quad \text{pour t =1,…....,n}$$

avec :

y_t = variable à expliquer à la date t;

x_{1t} = variable explicative 1 à la date t;

x_{2t} = variable explicative 2 à la date t;

...x_{kt} = variable explicative k à la date t;

$a_0, a_1, ..., a_k$ = paramètres du modèle ;

ε_t = erreur de spécification (différence entre le modèle vrai et le modèle spécifié), cette erreur est inconnue et restera inconnue;

n = nombre d'observations.

1.2 Forme matricielle

L'écriture précédente du modèle est d'un maniement peu pratique. Afin d'en alléger l'écriture et de faciliter l'expression de certains résultats, on a habituellement recours aux notations matricielles.

En écrivant le modèle, observation par observation, nous obtenons :

$$y_1 = a_0 + a_1 x_{11} + a_2 x_{21} + \cdots . a_k x_{k1} + \varepsilon_t$$

$$y_2 = a_0 + a_1 x_{12} + a_2 x_{22} + \cdots . a_k x_{k2} + \varepsilon_t$$

…….

$$y_t = a_0 + a_1 x_{1t} + a_2 x_{2t} + \cdots . a_k x_{kt} + \varepsilon_t$$

…….

$$y_n = a_0 + a_1 x_{1n} + a_2 x_{2n} + \cdots . a_k x_{kn} + \varepsilon_n$$

Soit, sous forme matricielle : $Y \quad = \quad X \quad a \quad + \quad \varepsilon$ (1)

$(n,1) \quad (n, k+1) \quad (k+1, 1) \quad (n,1)$

Avec :

$$Y = \begin{pmatrix} y_1 \\ . \\ : \\ y_t \\ . \\ . \\ y_n \end{pmatrix}; \quad X = \begin{pmatrix} 1 & x_{11} & x_{21} \cdots & x_{k1} \\ & & & \\ & & & \\ & & & \\ \vdots & & \ddots & \vdots \\ 1 & x_{1n} & x_{2n} \cdots & x_{kn} \end{pmatrix}; \quad a = \begin{pmatrix} a_0 \\ a_1 \\ : \\ a_t \\ . \\ . \\ a_k \end{pmatrix}; \quad \varepsilon \begin{pmatrix} \varepsilon_1 \\ . \\ : \\ \varepsilon_t \\ . \\ . \\ \varepsilon_n \end{pmatrix}$$

La première colonne de la matrice X, composée de 1, qui correspond au coefficient a_0 (coefficient du terme constant). La dimension de la matrice X est donc de *n* lignes et *k+1* colonnes (*k* étant le nombre de variables explicatives réelles, c'est-à-dire constante exclue).

II. Estimation et propriétés des estimateurs

2.1 . Estimation des coefficients de régression

Soit le modèle sous forme matricielle à *k* variables explicatives et *n* observations : $Y - Xa + \varepsilon$ (1)

Afin d'estimer le vecteur *a* composé des coefficients a_0, $a_1 ... a_k$, nous appliquons la méthode des Moindres Carrés Ordinaires (MCO) qui consiste à minimiser la somme des carrés des erreurs, soit :

$$Min \sum_{t=1}^{n} {\varepsilon_t}^2 = Min \quad \varepsilon'\varepsilon = Min \, (Y - Xa)'(Y - Xa) = Min \, S \qquad (2)$$

avec ε′ transposé du vecteur ε.

Pour minimiser cette fonction par rapport à *a*, *S* sera différenciée par rapport à a

$$\frac{\partial S}{\partial a} = -2\,X'Y + 2\,X'X\hat{a} = 0 \;\rightarrow \hat{a} = \;(X'X)^{-1}\;X'Y \qquad (3)$$

Cette solution est réalisable si la matrice carrée *X'X* de dimension (*k+1*, *k+1*) est inversible. La matrice *X'X* est la matrice des produits croisés des variables explicatives ; en cas de colinéarité parfaite entre deux variables explicatives, la matrice *X'X* est singulière et la méthode des MCO défaillante.

On appelle équations normales les équations issues de la relation :

$$X'\,X\hat{a} = X'Y$$

Le modèle estimé s'écrit :

$$\hat{y}_t = \hat{a}_0 + \hat{a}_1\,x_{1t} + +\hat{a}_2\,x_{2t} + \cdots.\,\hat{a}_k x_{kt} + e_t$$

Effet de la variation d'une seule des variables explicatives

Si la variable x_2 passe de la valeur x_{2t} à $(x_{2t}+\Delta x_{2t})$, toutes choses étant égales par ailleurs (les *k−1* autres variables restant constantes), alors la variable à expliquer varie de $\hat{a}_2\,\Delta x_2$.

Les coefficients s'interprètent donc directement en termes de propension marginale.

2.2 Hypothèses et propriétés des estimateurs

Par construction, le modèle est linéaire en X (ou sur ces coefficients) et nous distinguons les hypothèses stochastiques (liées à l'erreur ε) des hypothèses structurelles.

1) Hypothèses stochastiques

– H1 : les valeurs $x_{i,t}$sont observées sans erreur.

–H2 : $E(\varepsilon_t)=0$, l'espérance mathématique de l'erreur est nulle.

–H3 : $E(\varepsilon_{2t})=\sigma_{2\varepsilon}$, la variance de l'erreur est constante (∀t)(homoscédasticité).

–H4 : $E(\varepsilon_t\varepsilon_{t'})=0$si t =t', les erreurs sont non corrélées (ou encore indépendantes).

– H5 : $Cov(x_{it},\varepsilon_t)=0$, l'erreur est indépendante des variables explicatives (problème de l'endogénéïté.)

2) Hypothèses structurelles

– H6 : absence de colinéarité entre les variables explicatives, cela implique que la matrice (X′X) est régulière et que la matrice inverse $(X'X)^{-1}$ existe.

–H7:(X′X)/n tend vers une matrice finie non singulière.

–H8:n > k+1, le nombre d'observations est supérieur1au nombre des séries explicatives.

3) Propriétés des estimateurs

L'estimateur est sans biais : $E(\hat{a}\ ) = a$

La matrice des variances et covariances de l'erreur ε est donnée comme suit :

$$\Omega_{\hat{a}} = \hat{\sigma}^2{}_{\varepsilon}\,(X'X)^{-1}$$

L'estimateur est convergent

2.3. Équation d'analyse de la variance et qualité d'un ajustement

Comme pour le modèle de régression simple, l'équation fondamentale d'analyse de la variance est donnée comme suit :

$$\sum_t (y_t - \bar{y})^2 = \sum_t (\hat{y}_t - \bar{y})^2 + \sum_t e_t^2$$

$$\text{SCT} \quad = \quad \text{SCE} \quad + \ \text{SCR}$$

La variabilité totale (SCT) est égale à la variabilité expliquée (SCE) + la variabilité des résidus (SCR).

Cette équation permet de juger de la qualité de l'ajustement d'un modèle ; en effet, plus la variance expliquée est « proche » de la variance totale, meilleur est l'ajustement global du modèle. C'est pourquoi nous calculons le rapport SCE sur SCT:

$$R^2 = \frac{\sum_t (\hat{y}_t - \bar{y})^2}{\sum_t (y_t - \bar{y})^2} = 1 - \frac{\sum_t e_t^2}{\sum_t (y_t - \bar{y})^2}$$

R^2 est appelé le coefficient de détermination, et R le coefficient de corrélation multiple. R^2 mesure la proportion de la variance de Y expliquée par la régression de Y sur X. Cette qualité de l'ajustement et l'appréciation que l'on a du R^2 doivent être tempérées par le degré de liberté de l'estimation. En effet, lorsque le degré de liberté est faible1, il convient de corriger le R^2 afin de tenir compte du relativement faible nombre d'observations comparé au nombre de facteurs explicatifs par le calcul d'un R^2« corrigé »

$$\bar{R}^2 = 1 - \frac{n-1}{n-k-1}(1 - R^2$$

III. Les tests statistiques

3.1 Comparaison d'un paramètre a_i à une valeur fixée a

Le test d'hypothèses est le suivant

$$\text{H0} : a_i = \bar{a}$$

$$H0 : a_i \neq \bar{a}$$

Nous savons que : $\frac{\hat{a}_i - a_i}{\hat{\sigma}_{\hat{a}_i}}$ suit donc une loi de Student à $(n-k-1)$ degrés de liberté

Sous l'hypothèse H0, cette relation devient : $\frac{|\hat{a}_i - \bar{a}|}{\hat{\sigma}_{\hat{a}_i}} = t^*_{\hat{a}_i} \rightarrow$ loi de Student à (n– k-1) degrés de liberté.

Si $t^*_{\hat{a}_i} > t^{\alpha/2}_{(n-k-1)}$ alors nous rejetons l'hypothèse H0, a_i est significativement différent de a (au seuil de α).
Si $t^*_{\hat{a}_i} < t^{\alpha/2}_{(n-k-1)}$ alors nous acceptons l'hypothèse H0, ai n'est pas significativement différent de a (au seuil de α).

Cas particulier : test par rapport à une valeur particulière a=0.

Si nous désirons savoir si une variable explicative figurant dans un modèle est réellement – significativement – contributive pour expliquer la variable endogène, il convient de tester si son coefficient de régression est significativement différent de 0 pour un seuil choisi, en général α=5%.
Sous H0 (a_i=0), devient :

$\left|\frac{\hat{a}_i}{\hat{\sigma}_{\hat{a}_i}}\right| = t^*_{\hat{a}_i} \rightarrow$ loi de Student à (n– k-2) degrés de liberté.

$t^*_{\hat{a}_i}$ est appelé le ratio de Student, les règles de décision citées plus haut s'appliquent alors.

Ce test est très important ; en effet, si dans un modèle estimé, un des coefficients (hormis le terme constant) n'est pas significativement différent de 0, il convient d'éliminer cette variable1et de ré-estimer les coefficients du modèle. La cause de cette non-significativité, est due :

- soit à une absence de corrélation avec la variable à expliquer,
- soit à une colinéarité trop élevée avec une des variables explicatives.

IV. L'analyse de la variance

A. Construction du tableau d'analyse de la variance et test de signification globale d'une régression

Dans cette section, nous allons nous interroger sur la signification globale du modèle de régression, c'est-à-dire si l'ensemble des variables explicatives a une influence

sur la variable à expliquer. Ce test peut être formulé de la manière suivante : existe-t-il au moins une variable explicative significative ? Soit le test d'hypothèses :

H0 :$a_1 = a_2 = ... = a_k = 0$ (tous les coefficients sont nuls1)

H1 : il existe au moins un des coefficients non nul

Nous ne testons pas le cas où le terme constant a0est nul, car seules nous intéressent les variables explicatives. Un modèle dans lequel seul le terme constant est significatif n'a aucun sens économique.

Le cas où l'hypothèse H0 est acceptée signifie qu'il n'existe aucune relation linéaire significative entre la variable à expliquer et les variables explicatives (ou encore que la Somme des Carrés Expliqués n'est pas significativement différente de 0).

D'après l'équation fondamentale d'analyse de la variance, la régression est jugée significative si la variabilité expliquée est significativement différente de 0. Le tableau 1 présente le tableau d'analyse de la variance permettant d'effectuer le test de Fisher.

Ou encore : $F^* = \frac{\sum_t (\widehat{y_t} - \bar{y})^2 / k}{\sum_t e_t^2 / (n-k-1)} = 1 - \frac{R^2/k}{(1-R^2)/(n-k-1)}$

Tableau 1 – Analyse de la variance pour une régression multiple

Source de variation	**Somme des carrées**	**Degré de liberté**	**Carées moyen**
$x_1, x_2, \ldots . x_{1k}$	$SCE = \sum_t (\hat{y}_t - \bar{y})^2$	k	SCE/k
Résidu	$SCR = \sum_t e_t^2$	n-k-1	SCR/n-k-1
Total	$SCT = \sum_t (y_t - \bar{y})^2$	n-1	

L'hypothèse de normalité des erreurs implique que sous l'hypothèse H0, F^* suit une loi de Fisher (rapport de deux chi-deux). Nous comparons donc ce F^*calculé au F théorique à *k* et *(n−k−1)* degrés de liberté : si $F^* > F$ nous rejetons l'hypothèse H0, le modèle est globalement explicatif.

Dans la pratique, ce test est effectué immédiatement grâce à la connaissance du coefficient de détermination R^2(seulement si le modèle comporte un terme constant) qui permet de calculer le Fisher empirique (calculé).

Pour améliorer la qualité d'ajustement, il est nécessaire d'intégrer une variable indicatrice[16]. Une variable indicatrice est une variable explicative particulière qui n'est composée que de 0 ou de 1. Cette variable est utilisée lorsque, dans un modèle, nous désirons intégrer un facteur explicatif binaire : « le phénomène a lieu ou n'a pas lieu » pour corriger, par exemple, d'une valeur anormale ; ou bien lorsque le facteur explicatif est qualitatif : « le genre d'un individu, homme ou femme ». Il s'agit donc d'incorporer une ou des variables explicatives supplémentaires au modèle spécifié initialement et d'appliquer les méthodes classiques d'estimation. Le modèle de régression diffère selon l'apparition du phénomène par les valeurs d'un ou plusieurs coefficients alors que les autres paramètres sont identiques. En cas de modification structurelle d'un coefficient de régression, la variable muette affecte alors le coefficient de la ou des variables explicatives considérées.
Le domaine d'utilisation des variables indicatrices est très vaste, nous pouvons citer : la correction des valeurs anormales, la modification structurelle (0 pour la période avant le changement structurel, 1 après le changement structurel), l'intégration de la saisonnalité, la caractérisation d'un individu (genre, situation matrimoniale...), l'intégration de facteurs qualitatifs (appartenance d'un pays à la zone euro, promotion non quantifiable...), etc.

V : Problèmes associés à l'analyse de la régression

1. La relation qui unit la variable dépendante Y aux variables X est linéaire (les variables X et Y n'ont pas besoin d'être linéaires)
2. Le terme d'erreur est une variable aléatoire distribuée normalement avec une moyenne de 0 et une variance σ^2 constante.
    - Variance pas constante → Hétéroscédasticité
3. $\text{Cov}(X, \varepsilon) = 0$

   Termes d'erreurs corrélés avec les variables explicatives → Endogénéïté
4. $\text{Cov}(\varepsilon_i, \varepsilon_j) = 0$

   Termes d'erreurs corrélés entre eux → Autocorrélation
5. $\text{Cov}(X_j, X_k) = 0$

   Variables indépendantes corrélées → Multicollinéarité

1- l' hétéroscédasticité

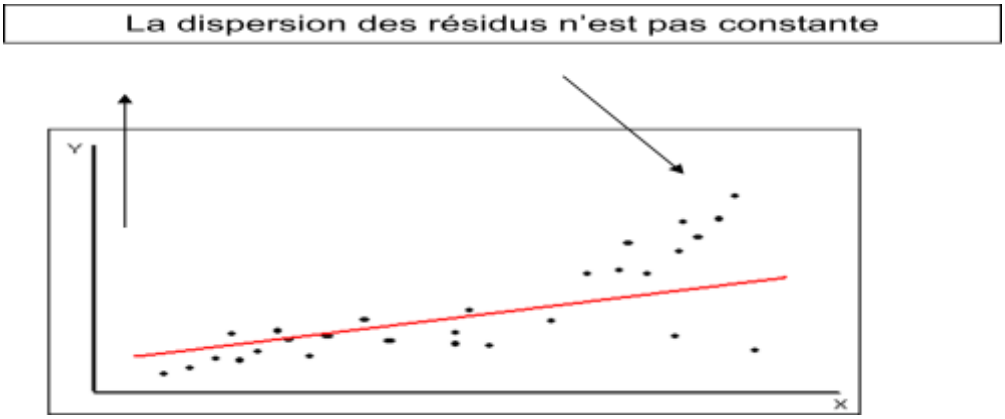
[16] Les termes de variables indicatrices, de variables auxiliaires ou de variables muettes sont indifféremment employés en français. Le terme anglo-saxon *dummy* est le plus couramment utilisé.

Le problème d'hétéroscédasticité est fréquent dans le cas d'une série de données en coupe transversale, notamment dans les données microéconomiques avec un effet de taille (l'existence d'une différence en termes de grandeur entre les observations).

La variance des termes d'erreur diffère entre les observations

Les paramètres estimés demeurent valides

Les tests de signification du modèle (*F*-test et *t*-test) ne sont pas fiables



Dans l'estimation MCO les erreurs sont supposées être homoscédastique, c'est-à-dire que la variance de l'erreur est constante pour l'ensemble des observations. Lorsque cette hypothèse n'est pas satisfaite l'estimation par MCO n'est pas optimale. Tandis que dans un modèle avec hétéroscédasticité la variance de l'erreur n'est pas constante d'une observation à l'autre, la matrice variance-covariance s'écrit comme suit :

$$\Omega_{\varepsilon} = \begin{pmatrix} E(\varepsilon_1 \varepsilon_1) & E(\varepsilon_1 \varepsilon_2) & E(\varepsilon_1 \varepsilon_n) \\ E(\varepsilon_2 \varepsilon_2) & E(\varepsilon_2 \varepsilon_2) & E(\varepsilon_2 \varepsilon_n) \\ E(\varepsilon_n \varepsilon_2) & E(\varepsilon_n \varepsilon_2) & E(\varepsilon_n \varepsilon_n) \end{pmatrix} = \begin{pmatrix} {\delta^2}_{\varepsilon_1} & 0 & 0 \\ 0 & {\delta^2}_{\varepsilon_2} & 0 \\ 0 & 0 & {\delta^2}_{\varepsilon_n} \end{pmatrix}$$

En présence d'hétéroscédasticité, la valeur du statistique *t* et du statistique *F* seront surestimés. Nous aurons donc tendance à rejeter plus souvent qu'il ne le faudrait l'hypothèse nulle

Test pour détecter l'hétéroscédasticité conditionnelle :

Test Breusch-Pagan : Cette méthode examine si la variance estimée des résidus d'une régression dépend de la valeur des variables explicatives.

Supposons un modèle de régression linéaire pour lequel nous avons les résidus

Le test de Breusch-Pagan consiste à régresser les résidus du modèle de régression initial par les variables explicatives de ce modèle

Modèle de régression initial : $Y_i = b_0 + b_1 X_{1i} + \ldots + \varepsilon_i$

Régression selon Breusch-Pagan : $\hat{\varepsilon}_i^2 = \gamma_0 + \gamma_1 X_{1i} + \ldots + \nu_i$

H_0 : Homoscédasticité contre H_1 : Présence d'hétéroscédasticité

Nous calculerons « $n*R^2$ » où n = nombre d'observations et R^2 = Coefficient de détermination de la régression : $\hat{\varepsilon}_i^2 = b_0 + b_1 X_{1i} + \ldots + \nu_i$

« $n*R^2$ » suivra une distribution du Khi carré (test unilatéral) avec un nombre de degrés de liberté égal au nombre de variables indépendantes de la régression

Nous rejetterons l'hypothèse nulle lorsque la valeur de « $n*R^2$ » sera supérieure à la valeur critique

Les méthodes de correction pour l'hétéroscédasticité les plus populaires sont :

- ✓ Régression robustes (méthode de White)
- ✓ Les moindres-carrés pondérés
- ✓ Les moindres-carrés quasi-généralisés

Le test de White

Ce test est fondé sur l'existence d'une relation significative entre le carré des résidus et les variables explicatives en niveau et au carré dans une seule équation de régression.

Une fois cette équation est estimé, l'étape suivante consiste à appliquer soit le test de significativité globale de Ficher, soit le test LM.

Plusieurs techniques de corrections peuvent être appliquées. Le principe de base de ces techniques consiste à effectuer une transformation sur les données du modèle estimé afin de rendre les écarts des résidus homoscédastique. En effet, une transformation peut s'effectuer en divisant les termes de l'équation de régression sur l'écart type du résidu.

Test d'hétéroscédasticité de White [WHI 1980]. Il est basé sur la régression du résidu au carré sur une constante et les produits croisés entre toutes les paires de variables explicatives différentes (constante comprise). On teste l'hypothèse nulle que tous les

coefficients de cette régression sont nuls, sauf le terme constant. Sous cette hypothèse, le test est distribué selon une loi Chi-2 à (k + 1)k/2 − 1 degrés de liberté.

2- Multicolinéarité : conséquences et détection

Le terme de multicolinéarité est employé dans le cas d'un modèle incorporant des séries explicatives qui sont liées entre elles. À l'opposé, pour des séries explicatives de covariance nulle ($Cov(x_1,x_2) = 0$), nous dirons qu'elles sont orthogonales. Si, pour des études théoriques, nous pouvons supposer que deux séries statistiques sont orthogonales, dans la pratique, lorsque l'économiste modélise des phénomènes économiques, les séries explicatives sont toujours plus ou moins liées entre elles.

Les conséquences de la multicolinéarité :

a) augmentation de la variance estimée de certains coefficients lorsque la colinéarité entre les variables explicatives augmente (le t de Student diminue) ;

b) b) instabilité des estimations des coefficients des moindres carrés, des faibles fluctuations concernant les données entraînent des fortes variations des valeurs estimées des coefficients ;

c) c) en cas de multicolinéarité parfaite, la matrice X′X est singulière (le déterminant est nul), l'estimation des coefficients est alors impossible et leur variance est infinie

Tests de détection d'une multicolinéarité

Test de Klein **:** Le test de Klein est fondé sur la comparaison du coefficient de détermination R^2_y calculé sur le modèle à *k* variables :

$$y = \hat{a}_0 + \hat{a}_1 x_1 + \hat{a}_2 x_2 + \cdots \hat{a}_k x_k + e$$

et les coefficients de corrélation simple $r^2_{xi,xj}$ entre les variables explicatives pour $i \neq j$.

Si $R^2_y < r^2_{xi\,;\,xj}$, il y a présomption de multicolinéarité. Il ne s'agit pas d'un test statistique au sens test d'hypothèses mais simplement d'un critère de présomption de multicolinéarité.

3- Problème de l'autocorrélation

Les résidus sont corrélés entre eux. C'est un problème fréquent des séries chronologiques

- Les paramètres estimés demeurent valides (sauf si une variable *X* est un« lag » de la variable *Y*)
- Les tests de signification du modèle (*F*-test et *t*-test) ne sont pas fiables

Autocorrélation positive : un résidu positif pour une observation accroît les probabilités d'obtenir un résidu positif pour l'observation suivante. C'est le type d'autocorrélation le plus fréquent.

Autocorrélation négative : un résidu positif pour une observation accroît les probabilités d'obtenir un résidu négatif pour l'observation suivante

En présence d'autocorrélation positive, la valeur du statistique *t* et du statistique *F* seront surestimés

Nous aurons donc tendance à rejeter plus souvent qu'il ne le faudrait l'hypothèse nulle

b) Test de Durbin et Watson

Le test de Durbin et Watson (DW) permet de détecter une autocorrélation des erreurs d'ordre 1 selon la forme :

$$\boldsymbol{e_t = e_{t-1} + v_t} \text{ avec } v_t \rightarrow N(0, \delta_v^2)$$

Le test d'hypothèses est le suivant : H0 : $\rho = 0$ contre H0 : $\rho \neq 0$

Pour tester l'hypothèse nulle H0, la statistique de Durbin etWatson est calculée :

$$\frac{\sum_{t=2}^{n}(\boldsymbol{e_t - e_{t-1}})^2}{\sum_{t=1}^{n} \boldsymbol{e_t}^2}$$

où e_t sont les résidus de l'estimation du modèle.

De par sa construction, cette statistique varie entre 0et 4et nous avons DW= 2 lorsque $\hat{\rho}= 0$ ($\hat{\rho}$ est le ρ observé).

Afin de tester l'hypothèse H0,Durbin et Watson ont tabulé les valeurs critiques de DW au seuil de 5% en fonction de la taille de l'échantillon net du nombre de variables explicatives (k). La lecture de la table permet de déterminer deux valeurs $d_1$et d_2 comprises entre 0 et 2 qui délimitent l'espace entre 0 et 4 selon le schéma 1 :

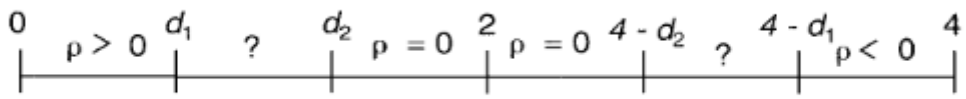


Schéma 1 – Interprétation du test de Durbin et Watson

Selon la position du DW empirique dans cet espace, nous pouvons conclure :

- $d_2 < DW < 4 - d_2$, on accepte l'hypothèse H0 $\rightarrow \rho = 0$;
- $0 < DW < d_1$, on rejette l'hypothèse H0 $\rightarrow \rho > 0$;
- $4 - d_1 < DW < 4$, on rejette l'hypothèse H0 $\rightarrow \rho < 0$

- $d_1 < DW < d_2$ ou $4 - d_2 < DW < 4 - d_1$; Nous sommes dans une zone d'indétermination, ou zone de doute, c'est-à-dire que nous ne pouvons pas conclure.

Conditions d'utilisation :

- le modèle doit comporter impérativement un terme constant;
- la variable à expliquer ne doit pas figurer parmi les variables explicatives (entant que variable retardée);
- pour les modèles en coupe instantanée, les observations doivent être ordonnées en fonction des valeurs croissantes ou décroissantes de la variable à expliquer ou d'une variable explicative soupçonnée être la cause de l'auto-corrélation ;
- le nombre d'observations doit être supérieur ou égal à 15.

Le test de Durbin et Watson est un test présomptif d'indépendance des erreurs du fait qu'il utilise les résidus ; de plus, il ne teste qu'une autocorrélation d'ordre 1.

Annexe 1 relative au chapitre 2 : La lecture des résultats d'estimation d'un modèle de régression multiple sous eviews

L'estimation de l'équation des déterminants des exportations (EXPO), sous eviews, par la méthode des MCO, donne les résultats figurant dans tableau suivant. Les variables explicatives sont le Produit Intérieur Brut (PIB), le taux de change (TCH) et l'inflation (INF).

Dependent Variable: EXPO
Method: Least Squares
Date: 02/25/19 Time: 16:21
Sample: 1970 2017
Included observations: 48

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3.18E+08	2.99E+09	0.106440	0.9157
PIB	0.346619	0.031454	11.01981	0.0000
TCH	2034941.	53991394	0.037690	0.9701
INF	-2.44E+08	1.77E+08	-1.381044	0.1742

R-squared	0.868952	Mean dependent var	2.45E+10
Adjusted R-squared	0.860017	S.D. dependent var	2.33E+10
S.E. of regression	8.70E+09	Akaike info criterion	48.69044
Sum squared resid	3.33E+21	Schwarz criterion	48.84638
Log likelihood	-1164.571	F-statistic	97.25199
Durbin-Watson stat	0.332981	Prob(F-statistic)	0.000000

Description des résultats de la sortie de la régression

La première zone de la fenêtre « **Equation** » contient des informations générales sur la modèle de régression étudié :

- ✓ **Ligne 01** : Nom de la variable dépendante (ou de la variable à expliquer).
- ✓ **Ligne 02** : Méthode de la régression utilisée.
- ✓ **Ligne 03** : Date et l'heure de l'exécution de la régression.
- ✓ **Ligne 04** : La taille de l'échantillon sur laquelle la régression est exécutée.
- ✓ **Ligne 05** : Nombre d'observation de l'échantillon sur laquelle la régression est exécutée.

La deuxième zone de la fenêtre « **Equation** » contient les différentes valeurs estimées pour les coefficients de modèle de régression étudié.

- ✓ **Colonne 01** : Noms des variables explicatives (indépendantes ou exogènes) de modèle de régression étudié.
- ✓ **Colonne 02** (« **Coefficients** ») : Valeurs estimées pour les coefficients de modèle de régression étudié : $\hat{a}_1, \hat{a}_2, \dots, \hat{a}_k$.
- ✓ **Colonne 03** (« **Std. Error** » ou « **Standard Error** ») : Valeurs estimées pour les écart-types des coefficients estimés : $\delta(\hat{a}_1),\ \delta(\hat{a}_2), \dots,\ \delta(\hat{a}_k)$.
- ✓ **Colonne 04** (« **t-Statistic** » ou « **Statistique de Student** ») : Valeur du **t** de Student calculé : $t_{a_k} = \frac{\hat{a}_k}{\delta(\hat{a}_k)}$
- ✓ **Colonne 05** (« **Prob** ») : Probabilité critique (*p-value*) du test de nullité des coefficients de modèle de régression étudié. Pour un risque de 5%, si Prob < 0,05 → on rejette l'hypothèse $\{H_0 : a_i = 0\}$ et on accepte l'hypothèse $\{H_1 : a_i \neq 0\}$.

La troisième zone de la fenêtre de l'équation contient les valeurs calculées pour les différents calculs statistiques.

- ✓ **R-squared** (« **R au carré** »): Valeur calculée pour le coefficient de détermination (R^2).
- ✓ **Ajusted R-squared** (« **R au carré ajusté** »): Valeur calculée pour le coefficient de détermination ajusté ($\bar{R}^2$).
- ✓ **S. E. of regression** (« **Standard Error of regression** ») : Valeur estimée pour l'écart-type des résidus de la régression ($\hat{\delta}$).

- ✓ **Sum squared resid** (« **Somme des carrés des résidus** ») : Valeur calculée pour la Somme des Carrés des Résidus (SCR= $\sum e_t^2$).
- ✓ **Log likelihood** : Valeur calculée pour le Log-vraisemblance pour les paramètres estimés.
- ✓ **Mean dependent var** : Moyenne de la variable dépendante.
- ✓ **S. D. dependent var** (« **Standard Deviation of the dependent var** ») : Écart-type de la variable dépendante.
- ✓ **Akaike info criterion** : Critère d'Akaike (*AIC*).
- ✓ **Schwarz criterion** : Critère de Schwarz (*BIC* ou *SC*).
- ✓ **Durbin-Watson stat** : Statistique du Durbin-Watson.

Série de TD N°2

Exercice 1 :
Pendant dix ans, une ferme a expérimenté le rendement du maïs Y associé à l'emploi de quantités croissantes d'un fertilisantX_1 et d'un insecticide X_2 .

X_{1t}	6	10	12	14	16	18	22	24	26	32
X_{2t}	4	4	5	7	9	12	14	20	21	24
Y_t	40	44	46	48	52	58	60	68	74	80

- Estimer les paramètres du modèle $Y_t = a_0 + a_1 X_{1t} + a_2 X_{2t}$
- Tester la signification des paramètres estimés au seuil de signification de 5%.
- Tester la signification globale du modèle

NB : On vous donne les informations suivantes :

La matrice $(X'X)=\begin{pmatrix} 10 & 180 & 120 \\ 180 & 3816 & 2684 \\ 120 & 2684 & 1944 \end{pmatrix}$; La matrice $(X'Y)=\begin{pmatrix} 570 \\ 11216 \\ 7740 \end{pmatrix}$

$$(X'X)^{-1} = \begin{pmatrix} 1{,}36347915 & -0{,}17700916 & 0{,}1602238 \\ -0{,}17700916 & 0{,}03204476 & -0{,}03331638 \\ 0{,}1602238 & -0{,}03331638 & 0{,}0362258 \end{pmatrix}$$

Exercice 2 : L'estimation de l'équation des déterminants des exportations (EXPO), sous eviews, par la méthode des MCO, nous donne les résultats figurant dans tableau N°1. Les variables explicatives sont le Produit Intérieur Brut (PIB) et le taux de change (TCH).

Tableau N°1 :

Dependent Variable: EXPO
Method: Least Squares
Date: 11/03/19 Time: 18:58
Sample: 1970 2017
Included observations: 48

Variable	Coefficient	Std. Error	t-Statistic	Prob.
PIB	0.349562	0.031697	**11.02833**	0.0000
TCH	16489431	53498509	**0.308222**	0.7593
C	-2.68E+09	2.08E+09	**-1.288386**	0.2042

R-squared	0.863272	Mean dependent var	2.45E+10
Adjusted R-squared	**0.85**7195	S.D. dependent var	2.33E+10
S.E. of regression	8.79E+09	Akaike info criterion	48.69121
Sum squared resid	3.47E+21	Schwarz criterion	48.80816
Log likelihood	-1165.589	Hannan-Quinn criter.	48.73541
F-statistic	**142**.0602	**Durbin-Watson stat**	0.304526
Prob(F-statistic)	0.000000		

a) Commenter tout d'abord, d'un point de vue économique, les résultats obtenus
b) Ecrire les résultats de la régression sous forme d'une équation et interpréter les coefficients
c) Tester au seuil de 5% la significativité de chacun des coefficients, pris un par un.
d) Tester au seuil de 5% l'hypothèse selon laquelle tous les coefficients seraient nuls. Expliquer au préalable la construction de la statistique utilisée pour le test et la loi quelle suit, sous l'hypothèse nulle
e) Quelle statistique de test figurant dans le tableau précédent permet d'appréhender la question de l'autocorrélation des résidus ? Rappelez de façon détaillée le principe de test. Quelle est votre conclusion au seuil de 5% ?

Nb : $t_{2.45}^{0.05/2} = 1{,}96$ $f_{2.45}^{0.05/2} = 3{,}23$; pour (n=48 et k=2) $d_1 = 1{,}43$; d2= 1,62

Corrigé de la Série de TD N°2

$$\hat{a} = (X'X)^{-1}X'Y = \begin{pmatrix} 1{,}363 & -0{,}177 & 0{,}160 \\ -0{,}177 & 0{,}032 & -0{,}033 \\ 0{,}160 & -0{,}033 & 0{,}036 \end{pmatrix}.\begin{pmatrix} 570 \\ 11216 \\ 7740 \end{pmatrix}$$

$$\hat{a} = \begin{pmatrix} 31{,}08 \\ 0{,}65 \\ 1{,}11 \end{pmatrix}$$

$\hat{\sigma}^2{}_{\varepsilon} = \frac{\sum_t e_t^2}{(n-k-1)} = \frac{e\, e'}{(n-k-1)}$; Nous devons calculer la valeur de e

$e = Y - Y' = Y - X\,\hat{a}$; $e_t = y_t - \hat{y_t} = y_t - (31{,}98 + 0{,}65\, x_{1t} + 1{,}11\, x_{2t})$

Les valeurs de la serie ajustée $\widehat{y_t}$ et celles des résidus e_t sont illustrés dans ce tableau 2.2 :

Tableau 2 .2

y_t	x_{1t}	x_{2t}	$\widehat{y_t} = 31{,}98 + 0{,}65\, x_{1t} + 1{,}11\, x_{2t}$	e_t	${e_t}^2$
40	6	4	40,32	-0,32	0,1024
44	10	4	42,92	1,08	1,1664
46	12	5	45,33	0,67	0,4489
48	14	7	48,85	-0,85	0,7225
52	16	9	52,37	-0,37	0,1369
58	18	12	57	1	1
60	22	14	61,82	-1,82	3,3124
68	24	20	69,78	-1,78	3,1684
74	26	21	72,19	1,81	3,2761
80	32	24	79,42	0,58	0,3364
Total					**13,6704**

$\sum_t e_t^2 = 13{,}6704$

$$\hat{\sigma}^2{}_\varepsilon = \frac{\sum_t e_t^2}{(n-k-1)} = \frac{\mathbf{13,6704}}{7} = 1{,}95$$

La matrice des variances et covariances de l'erreur ε est donnée comme suit :

$$\Omega_{\hat{a}} = \hat{\sigma}^2{}_\varepsilon (X'X)^{-1} = 1{,}95 \begin{pmatrix} 1{,}363 & -0{,}177 & 0{,}160 \\ -0{,}177 & 0{,}032 & -0{,}033 \\ 0{,}160 & -0{,}033 & 0{,}036 \end{pmatrix}$$

Les variances des coefficients de régression se trouvent sur la première diagonale

$\hat{\sigma}^2{}_{\hat{a}_0} = 1{,}95(1{,}363) = 2{,}66 \rightarrow \hat{\sigma}_{\hat{a}_0} = 1{,}63$

$\hat{\sigma}^2{}_{\hat{a}_1} = 1{,}95(0{,}032) = 0{,}06 \rightarrow \hat{\sigma}_{\hat{a}_1} = 0{,}25$

$\hat{\sigma}^2{}_{\widehat{a_2}} = 1{,}95(0{,}036) = 0{,}07 \rightarrow \hat{\sigma}_{\hat{a}_2} = 0{,}26$

Test de signification pour les paramètres estimés :

Le test peut être formulé à partir des deux hypothèses suivantes :0:H

Nous savons que : $\frac{\hat{a}_i - a_i}{\hat{\sigma}_{\hat{a}_i}}$ suit donc une loi de Student à (*n−k−1*) degrés de liberté

Sous l'hypothèse H0, cette relation devient : $\frac{|\widehat{a_i} - 0|}{\hat{\sigma}_{\hat{a}_i}} = t^*_{\hat{a}_i}$ → loi de Student à (n– k-1) degrés de liberté.

$t^*_{\hat{a}_0} = \frac{31{,}98}{1{,}63} = 203{,}55$; $t^*_{\hat{a}_1} = \frac{0{,}65}{0{,}25} = 2{,}6$; $t^*_{\hat{a}_2} = \frac{1{,}11}{0{,}26} = 4{,}62$

Comme les valeurs de t Student dépassent tous trois $t_7^{5\%} = 2,\ 365,$ les coefficients $a_0\ , a_1\ et\ a_2$ sont statistiquement significatifs au seuil de 5%.

$$R^2 = \frac{\sum_t(\widehat{y_t}-\bar{y})^2}{\sum_t(y_t-\bar{y})^2} = 1 - \frac{\sum_t e_t{}^2}{\sum_t(y_t-\bar{y})^2} = 1 - \frac{13,6704}{1634} = 1 - 0,0084 = 0,9916$$

$$R^2 = 99,16\%$$

$$F^* = \frac{\sum_t(\widehat{y_t}-\bar{y})^2/k}{\sum_t e_t{}^2/(n-k-1)} = 1 - \frac{R^2/k}{(1-R^2)/(n-k-1)} = \frac{0,9916/2}{(1-0,9916)/7} = 413,17$$

La valeur calculée de F dépasse largement la valeur tabulée $f_{2.7}^{5\%} = 4,7\ 4$ au seuil de %5 donc on accepte H1 et le modèle est globalement significatif.

Exercice 2 :

a) C'est une fonction des exportations expliquée par le PIB et le TCH. Les résultats obtenus nous montrent l'effet du PIB et du taux de change sur la variation des exportations. Lle coefficient de détermination (R^2=0.86) obtenue dans la régression montre que la variation des EXPO **est bien expliquée** par la combinaison linéaire des variables explicatives (PIB et TCH). En d'autre terme, le PIB et le TCH expliquent 85% de la variabilité des exportations.

b) **EXPO** = -2.68 $(10)^9$ + 0.34***PIB** + 1648941***TCH.**

Ces résultats indiquent que :

- Une augmentation de 1 DA du PIB entraine une augmentation de 0,34 DA des exportations
- Une augmentation de 1 DA du TCH engendre une augmentation de 1648941 DA des exportations (Cependant cette variable n'est pas significative)

c) Test de Student

- H0 : $\hat{a}_1 = 0$ contre H1 : $\hat{a}_1 \neq 0$. Nous avons $t^* = \hat{a}_1 / \sigma\hat{a}_1 = 11,02$

$T^* = 11,02 > t_{2.45}^{0.05/2} = 1,96$; Alors **on rejette H0** et on accepte H1, donc a_1 est significativement différent de 0, le PIB contribue **significativement** dans l'explication de l'exportation.

- H0 : $\hat{a}_2 = 0$; H1 : $\hat{a}_2 \neq 0$; $t^* = \hat{a}_2 / \sigma\hat{a}_2 = 0.3082$

  $t^* = 0.3082 < t_{2.45}^{0.05/2} = 196$; Donc **on accepte H0**, a_2 est significativement nul, donc le TCH **n'est pas significativement** contributif à l'explication des exportations.

– H0 : $a_0 = 0$; contre H1 : $a_0 \neq 0$. Nous avons $t^*a_0 = \left|\frac{\hat{a}0}{\sigma\hat{a}0}\right| = |-1.28| = 1.28$

$t^* = 1.28 < t_{2.45}^{0.05/2} = 1.96$ Donc on accepte H_0, **la constante a_0 est significativement nulle**.

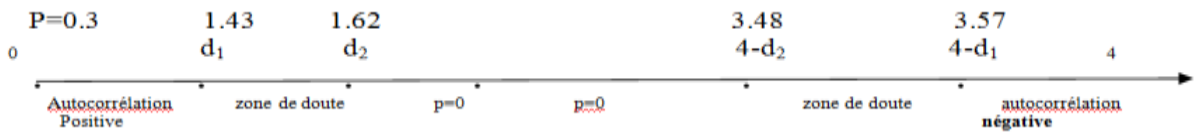
d) Le test de Fisher

H0 : = $a_0 = \hat{a}_1 = \hat{a}_2 = 0$ contre H1 : il existe au moins un coefficient non nul

$F^* = (R^2/k) / (1-R^2) / (n-k-1)$ suit la loi de Fisher à k et (n-k-1) ddl (degré de liberté)

$F^*=142,06 > f_{2.45}^{0.05/2} = 3,23$. Alors **on rejette H0** et on accepte H1, donc il existe au moins un coefficient non nul, le modèle est significatif.

e) La statistique de Durbin Watson sert à vérifier l'absence d'autocorrélation des erreurs c'est-à-dire l'indépendance de chaque erreur par rapport à la précédente. Dans notre cas cette statistique égale à 0,3, que l'on compare à celles lue dans la table de Durbin Watson à n=48 et k=2 (n : nombre d'observation ; k nombres de variables) explicatives), soit (d1=1 ,43 et d2=1,62). La valeur de DW se situe dans la zone d'autocréation positive. .Nous pouvons donc conclure une dépendance des erreurs.



Chapitre 3: Analyse des séries temporelles

L'analyse des séries temporelles a considérablement évolué ces dernières décennies: les progrès méthodologiques alliés à l'utilisation banalisée des logiciels font que la présentation et l'enseignement de cette discipline a connu de profonds bouleversements. Ce chapitre a pour but de familiariser les étudiants avec le traitement d'une série chronologique de type additive et multiplicative.

I-Définition et composantes d'une série temporelle

1-1- Définition d'une série temporelle (chronologique)

Une série temporelle (ou chronologique) est une suite (Y_t) d'observations chiffrées d'un même phénomène, ordonnées dans le temps [17]. Une série temporelle ou encore chronique est une succession d'observations au cours du temps représentant un phénomène économique (prix, ventes...) ; par hypothèse, le pas du temps est considéré constant : l'heure, le jour, le mois, le trimestre, l'année[18].

Les dates d'observations sont généralement ordonnées de manière régulière dans le temps La périodicité des observations est variable : mensuelle (p =12) comme les ventes d'une société, trimestrielle (p = 4) comme la consommation trimestrielle d'électricité, semestrielle (p =2)….

1-2- Les composantes d'une série temporelle

L'objectif de la décomposition d'une série chronologique est de distinguer dans l'évolution de la série, une tendance « générale », des variations saisonnières, et des variations accidentelles imprévisibles. [19]Cela permet de mieux comprendre, de décrire l'évolution de la série et de prévoir son évolution (à partir de la tendance et des variations saisonnières).[20]

1-2-1- La tendance (trend) Tt

La tendance représente l'évolution à long terme de la série étudiée, l'évolution fondamentale de la série. Elle traduit le comportement moyen de la série (tendance à la hausse ou à la baisse).

---

[17] Florence NICOLEAU, (2006), « séries chronologiques », Polycopié de cours, IUT de Nice Côte d'Azur, Département STID, 2005/2006, p 1.

[19] On peut distinguer aussi la composante cyclique sui est une variation se trouvant généralement dans les séries de longue durée et traduit des phases successives de croissance et de récession qui constitue le cycle économique

[20] Florence NICOLEAU(2006), *Op.cit.*, p 4.

1-2-2- La saisonnalité St

Les variations saisonnières sont des fluctuations périodiques à l'intérieur d'une année, et qui se reproduisent de façon plus ou moins permanente d'une année sur l'autre. Elle correspond au phénomène qui se répète à un intervalle de temps régulier (périodique).

Exemple : quasi stagnation entre le 1° et le 3° trimestre forte augmentation au 4° trimestre.

1-2-3- La composante résiduelle (résidus, erreur) et

Ce sont des fluctuations accidentelles irrégulières dues par exemple aux : guerre, grèves….elle sont de nature aléatoire. Elles sont supposées en général de faible amplitude..

1-3 Les tests de détection de la saisonnalité

1.3.1 La représentation graphique et le tableau de Buys-Ballot

L'analyse graphique d'une chronique suffit, parfois, pour mettre en évidence une saisonnalité. Néanmoins, si cet examen n'est pas révélateur ou en cas de doute, le tableau de Buys-Ballot permet d'analyser plus finement l'historique.

Le tableau de Buys-Ballot est un tableau à deux entrées dans lequel sont consignées les valeurs de x. Il est constitué en ligne par les années et en colonne par le facteur à analyser (mois, trimestre...). Les moyennes et les écarts types des années et des trimestres (ou des mois selon le cas) sont calculés ainsi que pour l'ensemble des observations de la chronique.

1.3 .2. Analyse de la variance et test de Fisher

L'examen visuel du graphique ou du tableau ne permet pas toujours de déterminer avec certitude l'existence d'une saisonnalité, de surcroît il interdit l'automatisme de traitement qui peut s'avérer nécessaire dans le cas d'un nombre important de séries à examiner. Le test de Fisher à partir de l'analyse de la variance permet de pallier ces deux inconvénients.

Ce test suppose la chronique sans tendance ou encore sans extra saisonnalité. Dans le cas contraire cette composante sera éliminée par une régression sur le temps (extra-saisonnalité déterministe), ou par une procédure de filtrage (extra-saisonnalité aléatoire).[21]

Soit : N le nombre d'années, p le nombre d'observations (la périodicité) dans l'année (trimestre p = 4, mois p = 12, etc.).

x_{ij} la valeur de la chronique pour la i-ème année (i = 1, ..., N) et la j-ème période (j = 1, ..., p) supposée telle que $x_{ij} = m_{ij} + e_{ij}$; les e_{ij} sont les résidus considérés comme aléatoires formés d'éléments indépendants : $e \rightarrow N(0 ; \sigma^2)$. Les m_{ij} sont les éléments d'une

[21] Regis Bourbounis et Michal Thereza (1998), *Analyse des séries temporelles en économie*, Press Universitaire de France , p 17.

composante de la chronique qui s'écrivent : m_{ij} = a_i + b_j avec b_j qui mesure l'effet période en colonne du tableau et a_i qui mesure l'effet année en ligne du tableau.

Deux effets absents sont testés contre deux effets significativement présents :

- si l'effet période est significatif, la série est saisonnière ;
- si l'effet année est significatif, ceci suggère deux interprétations.

1. La chronique de départ n'a pas été transformée, elle possède alors des paliers horizontaux.
2. La chronique a été transformée, des changements de tendance existent dans la chronique.

Le déroulement du test est le suivant : [22]

a) Calcul de la variance totale du tableau

Soit S_T la somme totale des carrés : $S_T = \sum_{i=1}^{n}\sum_{j=1}^{p}(x_{ij} - x_{..})^2$

Avec $x_{..} = \frac{1}{N.p}\sum_{i=1}^{n}\sum_{j=1}^{p}(x_{ij})$ la moyenne générale de la chronique sur les N. p observations.

$x_{i.} = \frac{1}{p}\sum_{j=1}^{p} x_{ij}$ la moyenne de l'anneé i

$x_{.j} = \frac{1}{N}\sum_{i=1}^{N} x_{ij}$ la moyenne de la periode j

Analyse de la variance pour détecter une saisonnalité et/ou une tendance.

Somme des carrés	Degré de liberté	Désignation	Variance
$SP = N_j \sum_{j=1}^{P}(X_{.j} - X_{..})^2$	P-1	Variance période	$V_P = \frac{SP}{P-1}$
$S_A = P\sum_{i=1}^{N}(X_{i.} - X_{..})^2$	N-1	Variance année	$V_A = \frac{SA}{N-1}$
$S_A = ST - SA - SP$	(P-1)(N-1)	Variance résiduelle	$V_R = \frac{SR}{(P-1)(N-1)}$
		Variance totale	$V_T = V_P + V_A + V_R$

Source : Abderrahmani Fares (2018), p.11 et Regis Bourbounis et Michal Thereza (1998)

b) Test de l'influence du facteur colonne (période, mois ou trimestre : Ho = pas d'influence)

Calcul du Fisher empirique $F_c = \frac{V_P}{V_R}$

$F^{\alpha}{}_{v1;v\,2}$ à v_1 = p -1 et v_2 = (N -1) (p -1) degrés de liberté

Si le Fisher empirique est supérieur au Fisher lu dans la table, on rejette l'hypothèse H0, la série est donc saisonnière.

c) Test de l'influence du facteur ligne (année : Ho = pas d'influence)

[22] Regis Bourbounis et Michal Thereza (1998), *Op.cit*., p.18

Calcul du Fisher empirique $F_c = \frac{V_A}{V_R}$ que l'on compare au Fisher lu dans la table

$F^{\alpha}{}_{v3;v\,2}$ à v_3 = N -1 et v_2 = (N -1) (p -1) degrés de liberté

Si le Fisher empirique est supérieur au Fisher lu, on rejette l'hypo- thèse Ho, la série est donc affectée d'une tendance.

Ces étapes sont synthétisées dans le tableau ci-après

	Test de la tendance	Test de saisonnalité
Les hypothèses	H_0 : la série n'a pas de tendance H_1 : la série possède une tendance	H_0 : la série n'a pas de saisonnalité H_1 : la série possède une saisonnalité
La statistique du test	:$F_c = \frac{V_A}{V_R}$	$F_c = \frac{V_P}{V_R}$
La règle de décision	Si $F_c > F_{(V1,V2)}$ on accepte H_1 Si $F_c \leq F_{(V1,V2)}$ on accepte H	Si $F_c > F_{(V3,V4)}$ on accepte H_1 Si $F_c \leq F_{(V3,V4)}$ on accepte H_0
Le degré de liberté	V1=N-1 et V2=(P-1)(N-1)	V3=P-1 et V4=(N-1)(P-1)

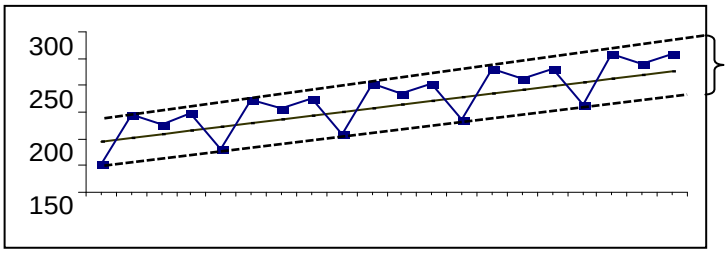
Source : ABDERRAHMANI F, (2018) « Guide pratique des séries temporelles macro-économiques et financières avec eviews 9.5 », *polycopié de cours à caractère pédagogique*, université de Bejaia, p.11.

II- Modèles de décomposition d'une série chronologique

2-1 Modèle de décomposition d'une série chronologique

La technique de décomposition-recomposition repos sur un modèle qui l'autorise. Ce modèle porte le nom de schéma de décomposition. Il en existe essentiellement trois grands types :[23]

- Le schéma additif qui suppose l'orthogonalité (indépendance) des différentes composantes. Il s'écrit : $x = T_t + S_t + C_t + e_t$. Dans ce schéma la saisonnalité est rigide en amplitude et en période. Le modèle additif est engendré par deux lignes parallèles. L'amplitude de variation dans le modèle additif est constante.

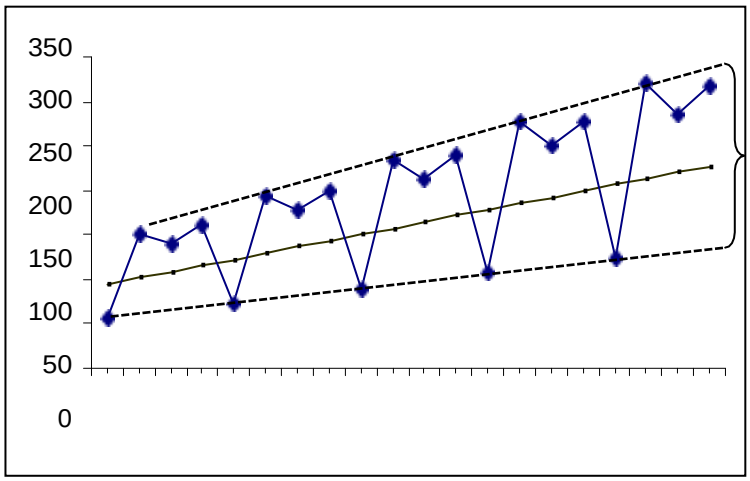


- Le schéma multiplicatif : $Xt = Tt * St + et$, dans lequel la composante saisonnière est liée à l'extra-saisonnier (saisonnalité souple avec variation de l'amplitude au cours du temps). Graphiquement, l'amplitude des variations (saisonnières) varie

[23]Regis Bourbounis et Michal Thereza (1998), *Op.cit*,p .16

- Le schéma multiplicatif complet : $X_t = T_t * S_t * e_t$ (interaction générale des trois composantes). Il est actuellement le plus utilisé en économie. Il est commode puisque le logarithme de la chronique conduit au schéma additif. Dans le cas d'une série (Yt) à valeurs positives, ce 2e modèle multiplicatif se ramène à unmodèle additif en considérant la série (ln(Yt)) : $\ln(Y_t) = \ln(C_t) + \ln(S_t) + \ln(\varepsilon_t)$.

- La seule différence entre les 2 modèles multiplicatifs est dans l'estimation des ε_t , qui n'apas une grande importance.



2-2. Les tests de décomposition

2.1 Le test de la bande

Le « test de la bande » consiste à partir de l'examen visuel du graphique de l'évolution de la série brute à relier, par une ligne brisée, toutes les valeurs « hautes » et toutes les valeurs « basses » de la chronique. Si les deux lignes sont parallèles, la décomposition de la chronique peut se faire selon un schéma additif ; dans le cas contraire, le schéma multiplicatif semble plus adapté.[24]

2.2. Le test de Buys-Ballot

Le test de Buys-Ballot est fondé sur les résultats du tableau 1 (calcul des moyennes et des écarts-types par année). Le schéma est, par définition, additif si l'écart-type et la moyenne sont indépendants ; il est multiplicatif dans le cas contraire. Lorsque le nombre d'années est suffisant, nous pouvons estimer par la méthode des MCO les paramètres de l'équation.

[24] Regis Bourbounis et Michal Thereza (1998), *Op.cit*.

Dans le cas, où le coefficient a1 n'est pas significativement différent de 0 (test de Student) alors on accepte l'hypothèse d'un schéma additif ; dans le cas contraire, nous retenons un schéma multiplicatif.

III. Dessaisonalisation des séries chronologiques

3.1 Construction des séries corrigées des variations saisonnières ou séries CVS.

La correction des variations saisonnières suppose que certaines hypothèses soient vérifiées et cela dans le but de simplifier certaines écritures.[25]

Les hypothèses :

1.1.1.1. ***Hypothèse 1: La décomposition de la série***

Nous admettrons que la série chronologique n'est constituée que de trois composantes, qui sont : la tendance notée $\mathbf{T_t}$, la saisonnalité notée $\mathbf{S_t}$ et le résidu noté $\mathbf{R_t}$

1.1.1.2. ***Hypothèse 2: Les modèles***

Nous nous intéresserons à trois types de modèles :

- le modèle additif,

$$y_t = T_t + S_t + R_t$$

- le modèle multiplicatif général,

$$y_t = T_t\,(1+S_t) + R_t$$

- le modèle multiplicatif particulier de la forme :

$$y_t = T_t * S_t * R_t$$

ce dernier modèle devient additif lorsque nous appliquons un logarithme :

$$\ln(y_t) = \ln(T_t) + \ln(S_t) + \ln(R_t)$$

Si nous posons le changement de variable :

$$y'_t = \ln(y_t) \qquad T'_t = \ln(T_t) \qquad S'_t = \ln(S_t) \qquad R'_t = \ln(R_t)$$

alors le modèle peut être traité de façon additive :

$$y'_t = T'_t + S'_t + R'_t$$

[25] Jean-Louis MONINO - Jean-Michel KOSIANSKI - François LE CORNU , Travaux diriges – statistique descriptive – Polycopier de cours

1.1.1.3. ***Hypothèse 3: La composante tendancielle***

Nous admettons que l'opérateur moyenne mobile laisse passer la composante tendancielle sans la modifier si la tendance est un polynôme de degré au plus égal à un de la forme :

$$T_t = a\,t + b$$

1.1.1.4. ***Hypothèse 4: La composante saisonnière***

La saison est rigoureusement identique de période en période, hypothèse que l'on écrit de la façon suivante :

$$S_t = S_{ij} = S_j \quad \text{avec} \quad S_{t+p} = S_{i+1,j} = S_j \quad \text{avec} \quad j = 1,....,p$$

Il y a compensation des p composantes saisonnières S_j . Nous pouvons écrire cette hypothèse sous la forme :

$$\sum_{j=1}^{j=p} S_j = 0$$

Cette hypothèse est également connue sous le nom de compensation des aires.

1.1.1.5. Hypothèse 5 : La composante résiduelle

Les variations résiduelles possèdent les propriétés suivantes :

$$\overline{R} = 0 \qquad S_R^2 \approx 0$$

La moyenne mobile appliquée aux variations résiduelles a des fluctuations très faibles autour de zéro.

$$mm_t^p(R) \approx 0$$

3.2 Le calcul des coefficients saisonniers et la correction des variations saisonnières des séries chronologiques

Plusieurs étapes sont nécessaires pour trouver les coefficients saisonniers et corriger les séries chronologiques des variations saisonnières :

1.2.1. ETAPE 1: *La représentation graphique*

La série chronologique est représentée pour observer les trois composantes de la série et éventuellement pour repérer les points aberrants.

1.2.2. ETAPE 2 : *Les corrections des points aberrants*

Cette étape consiste à éliminer par un calcul simple les points aberrants pour qu'ils ne soient pas pris en compte dans les calculs. Estimation graphique, ou estimation par une moyenne ou par toute autre méthode. Nous retiendrons comme méthode la demi-somme des deux points qui encadrent le point aberrant.

1.2.3. **ETAPE 3 :** *Le choix du modèle*

Nous déterminerons le type de modèle à utiliser pour la correction des variations saisonnières. Deux grands modèles sont à notre disposition ; le modèle additif ou multiplicatif. Plusieurs méthodes: La méthode du profil, La méthode de la bande, La méthode du tableau de Buys-Ballot.

1.2.4. **ETAPE 4 :** *Le filtrage de la série*

Dans cette étape, la composante saisonnière est supprimée en appliquant un filtre. Ainsi, nous devons déterminer la longueur p de la moyenne mobile que nous devons appliquer afin d'éliminer les variations saisonnières. Nous conviendrons qu'elle doit être au moins égale à la saison de la série. Nous envisagerons deux cas :

- ***Cas d'un modèle additif***

$$y_t = T_t + S_t + R_t$$

Soit X la série des moyennes mobiles définie :

- si p est paire par:

$$mm_t^p(y_t) = x_t = \frac{1}{p}\left[\sum_{i=-k+1}^{i=k-1} y_{t+i} + \frac{1}{2}y_{t-k} + \frac{1}{2}y_{t+k}\right]$$

- si p est impaire par :

$$mm_t^p(y_t) = x_t = \frac{1}{p}\sum_{i=-k}^{i=k} y_{t+i}$$

nous calculons la série des différences saisonnières définies par :

$$d_{ij} = y_{ij} - mm_t^p(y_t) = y_t - x_t$$

Pour le mois ou trimestre j, on obtient une première approximation du coefficient saisonnier S'_j en calculant la moyenne ou la médiane des différences saisonnières. Mais l'hypothèse dite de « conservation des aires » ne peut être vérifiée puisque les coefficients saisonniers sont obtenus de façon indépendante. Ainsi, nous corrigeons les coefficients S'_j en procédant de la manière suivante :

- recherche de la moyenne des coefficients saisonniers

$$\bar{S}' = \frac{1}{p}\sum_{j=1}^{p} S'_j$$

- correction des coefficients saisonniers

$$S_j^* = S'_j - \bar{S}'$$

- ***Cas d'un modèle multiplicatif***

Le seul modèle multiplicatif que nous verrons sera de la forme :

$$y_t = T_t\,(1+S_t) + R_t$$

que l'on peut écrire de la façon suivante, si nous remplaçons l'indice t par sa décomposition en année et saison (cf. TD 8) :

$$y_{ij} = T_{ij}\,(1+S_j) + R_{ij}$$

Posons un changement de variable simple:

$$s_j = 1 + S_j$$

où s_j est appelé coefficient saisonnier.

Le modèle s'écrit alors de la façon suivante :

$$y_{ij} = T_{ij}\,s_j + R_{ij}$$

On calcule les rapports saisonniers

$$r_{ij} = \frac{y_{ij}}{mm_t^p(Y)}$$

Pour le mois ou le trimestre j on obtient une première approximation du coefficient saisonnier S'j en calculant la moyenne ou la médiane des différences saisonnières.

Mais l'hypothèse dite de « conservation des aires » ne peut être vérifiée puisque les coefficients saisonniers sont obtenus de façon indépendante. Avant de corriger les coefficients saisonniers nous rappelons les hypothèses suivantes :

- Hypothèse sur les coefficients saisonniers

$$\sum_{j=1}^{p} S_j = 0$$

- Hypothèse sur la moyenne des coefficients saisonniers

$$s_j = 1 + S_j$$

$$\sum_{j=1}^{p} s_j = \sum_{j=1}^{p}\left(1+S_j\right) = p$$

Ainsi, nous pouvons maintenant corriger les coefficients S'$_j$ en procédant de la manière suivante :

- recherche de la moyenne des coefficients saisonniers

$$\bar{S}' = \frac{1}{p}\sum_{j=1}^{p} S'_j$$

- correction des coefficients saisonniers

$$S_j^* = \frac{S'_j}{\bar{S}'}$$

1.2.5. ETAPE 5 *:La série corrigée des variations saisonnières*

C'est la dernière étape de la construction des séries chronologiques corrigées des variations saisonnières. Deux cas se présentent :

1.2.5.1.Cas d'un modèle additif

La série CVS s'écrit:

$$y_{i,j}^{CVS} = y_{i,j} - S_j^*$$

1.2.5.2.Cas d'un modèle multiplicatif

La série CVS s'écrit:

$$y_{i,j}^{CVS} = \frac{y_{i,j}}{S_j^*}$$

L'objectif de ce chapitre était d'aborder essentiellement la correction des variations saisonnières des séries. Quand la série chronologique est de type déterministe, il est possible de voir sur un graphique les trois grandes composantes d'une série. Il nous faudra également connaître le type de modèle additif ou multiplicatif. Alors, nous pourrons procéder à l'élimination la composante tendancielle à l'aide du filtre des moyennes mobiles, puis estimer la composante saisonnière dans les meilleures conditions statistiques, pour enfin restituer le plus fidèlement possible la composante tendancielle et la composante saisonnière et en déduire la composante résiduelle.

Série de TD N°3

Exercice 1

Le tableau ci-dessus représente les données trimestrielles d'une série (notée X) chronologique

Tableau 1

	T1	T2	T3	T4	La moyenne	L'ecart-type
2019	6	5	2	20	8.25	8.015610
2020	2	4	5	19	7.5	7.767453
2021	5	8	6	22	10.25	7.932003

a : Tracer la graphe de la série X. La série est-elle affectée par un mouvement saisonnier ?

La régression de l'écart-type sur la moyenne nous donne les résultats consignés dans le tableau suivant :
Tableau 2 :

Dependent Variable: ECARTTYPE
Method: Least Squares
Date: 12/19/22 Time: 14:43
Sample: 2001 2003
Included observations: 3

Variable	Coefficient	Std. Error	t-Statistic	Prob.
MOYENNE	0.038880	0.079856	0.486875	0.7116
C	7.568065	0.698261	10.83844	0.0586

R-squared	0.191623	Mean dependent var	7.905022
Adjusted R-squared	-0.616754	S.D. dependent var	0.126259
S.E. of regression	0.160541	Akaike info criterion	-0.585814
Sum squared resid	0.025773	Schwarz criterion	-1.186739
Log likelihood	2.878721	Hannan-Quinn criter.	-1.793750
F-statistic	0.237047	Durbin-Watson stat	1.989691
Prob(F-statistic)	0.711552		

b : Quel est le modèle de décomposition de la série X?

c : Dessaisonnaliser la série

d : Calculer la prévision pour l'année 2022

Exercice 2

Le tableau suivant représente les données trimestrielles d'une série chronologique

	T1	T2	T3	T4	La moyenne	Ecart-type
2018	52	36	69	89	61.5	19.70
2019	65	45	86	111	76.75	24.51
2020	81	56	108	139	96	30.89
2021	102	70	135	174	120.25	38.61

a - La série est-elle affectée par un mouvement saisonnier ?

La régression de l'écart-type en fonction de la moyenne nous donne les résultats suivants

Dependent Variable: ECAR
Method: Least Squares
Date: 12/03/22 Time: 12:26
Sample: 2018 2021
Included observations: 4

Variable	Coefficient	Std. Error	t-Statistic	Prob.
MOY	0.322804	0.001989	**162.2713**	**0.0000**
C	-0.181043	0.181639	-0.996723	0.4239

R-squared	0.999924	Mean dependent var	28.42750
Adjusted R-squared	0.999886	S.D. dependent var	8.190712

b : Quel est le modèle de décomposition de cette série ?

c : Dessaisonnaliser la série (sachant que y = 39 .15 +5 ,82 t)

Exercice 3 : Un modèle additif sur une serie trismetrielle[26]

Le tableau ci-dessous donne le chiffre d'affaires d'une entreprise sur la période 1994 à 1997.

Tableau 1 - Chiffre d'affaires trimestriel

	Trimestre 1	Trimestre 2	Trimestre 3	Trimestre 4
1994	120	181	71	119
1995	128	190	73	124
1996	140	196	84	133
1997	145	206	96	142

1 - Calculer les coefficients saisonniers et donner la série Z corrigée des variations saisonnières. Vérifier que ces coefficients vérifient bien les hypothèses de départ.

2 - Tracer sur le même graphique la série X et la série Z corrigée des variations saisonnières.

Exercice 4 : un modèle multiplicatif sur une série mensuelle

La compagnie aérienne régionale Air-Hub désire connaître la structure du trafic aérien d'une de ses lignes. Pour cela la compagnie fournit la série mensuelle du nombre de passagers entre 1990 et 1994.

Tableau 2 - Trafic mensuel d'une ligne aérienne - en nombre de passagers

[26] Tiré de Jean-Louis MONINO - Jean-Michel KOSIANSKI - François LE CORNU , Travaux diriges – statistique descriptive – Polycopier de cours

	janvier	février	mars	avril	mai	juin	juillet	août	septembre	octobre	novembre	décembre
1990	713	756	1 042	903	905	1 240	812	160	997	1 180	1 160	1 022
1991	1 026	989	1 161	1 074	980	1 480	1 010	570	1 110	1 248	1 220	1 120
1992	1 006	1 037	1 220	1 227	1 040	1 730	1 034	540	1 203	1 310	1 340	1 140
1993	1 092	1 081	1 284	1 236	1 068	1 910	1 203	490	1 282	1 360	1 370	1 160
1994	1 080	1 067	1 279	1 228	1 059	2 160	1 190	430	1 278	1 282	1 163	1 365

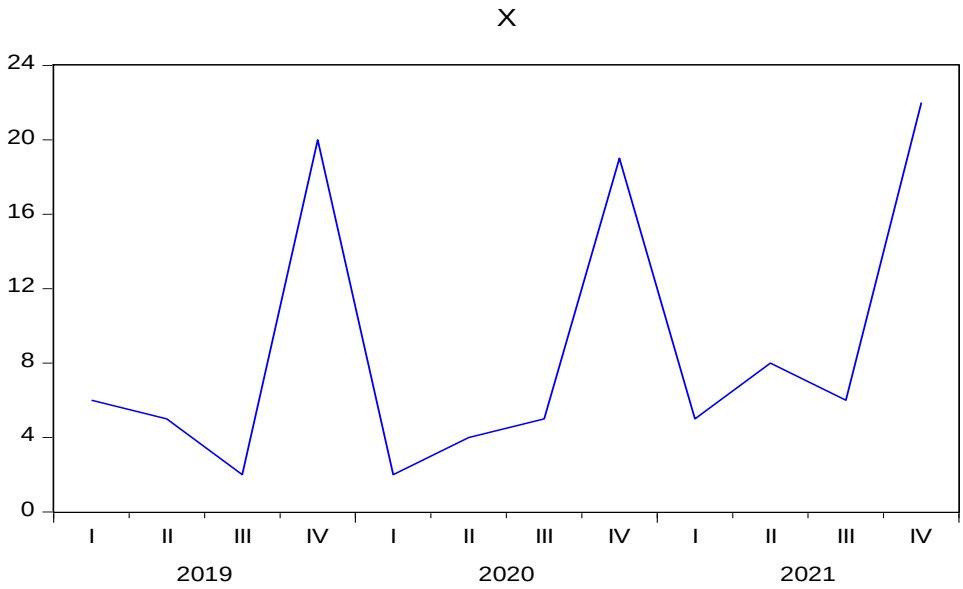
1 - Calculer les coefficients saisonniers et donner la série Z corrigée des variations saisonnières. Vérifier que ces coefficients vérifient bien les hypothèses de départ.

2 - Tracer sur le même graphique la série X et la série Z corrigée des variations saisonnières.

Corrigé-type de la série 3

Corrigé-Type de l'exercice 1

a)



Le graphe montre un mouvement saisonnier ou les ventes augmentent dans le quatrième trimestre et baissent dans le deuxième et le troisième trimestre. La lecture du tableau indique la persistance du trimestre T4 à se classer en première position quelle que soit l'année et la position de « creux » occupée par le trimestre T1, ce qui nous conduit à retenir l'existence d'une saisonnalité.

b) le modèle de décomposition

D'après le tableau donnant les résultats de la régression de l'écart-type sur la moyenne, la probabilité associée au coefficient de la moyenne (0,7116) est supérieur à 5%. La moyenne est indépendante de l'écart- type. **Le modèle est donc additif.**

c : La dessaisonnalisation de la série X

Nous devons commencer par le calcul des donner sans tendance.

L'estimation de la tendance par la methode des MCO nous donne les résultats consignés dans ce tableau (voir la conne n°5).

$$\widehat{a_1} = \frac{\sum_{i=1}^{12}(x_i t_i) - N\bar{x}\,\bar{t}}{\sum_{t=1}^{15} t_i^2 - N\,\bar{t}^2} = \mathbf{0,71}$$

$$\widehat{a_0} = \bar{x} - \widehat{a_1}\,\bar{t} = 4{,}03$$

$\hat{x}_i$ = 4,03+ 0,71 t_i

x_i	t_i	$x_i \,.t_i$	$t_i\, t_i$	$\hat{x}_i$ = 4,03+ 0,71 t_i	$S_i = x_i - \hat{x}_i$	$\bar{S}$	$S^* = S_i - \bar{S}$	$X_{CVS} = x_i - S^*$
6	1	6	1	4,74	1,26	2,445	-1,185	7,185
5	2	10	4	5,45	-0,45	2,445	-2,895	7,895
2	3	6	9	6,16	-4,16	2,445	-6,605	8,605
20	4	80	16	6,87	13,13	2,445	10,685	9,315
2	5	10	25	7,58	-5,58	-1,145	-4,435	6,435
4	6	24	36	8,29	-4,29	-1,145	-3,145	7,145
5	7	35	49	9	-4	-1,145	-2,855	7,855
19	8	152	64	9,71	9,29	-1,145	10,435	8,565
5	9	45	81	10,42	-5,42	-1,235	-4,185	9,185
8	10	80	100	11,13	-3,13	-1,235	-1,895	9,895
6	11	66	121	11,84	-5,84	-1,235	-4,605	10,605
22	12	264	144	12,55	9,45	-1,235	10,685	11,315

X_{CVS} ; *est est la série corrigée des variations saisonnières* : $X_{CVS} = x_i - S^*$

d : Calculer de la prévision pour l'année 2022

Pour calculer la prévision en tenant compte de l'effet saisonnier, nous devons calculer les coefficients saisonniers.

Le premier coefficient est égale à la moyenne des premiers trimestres S* de chaque année.

Le coefficient saisonnier est donné par : $C_{Si} = \sum_i^N \frac{S_i}{N}$

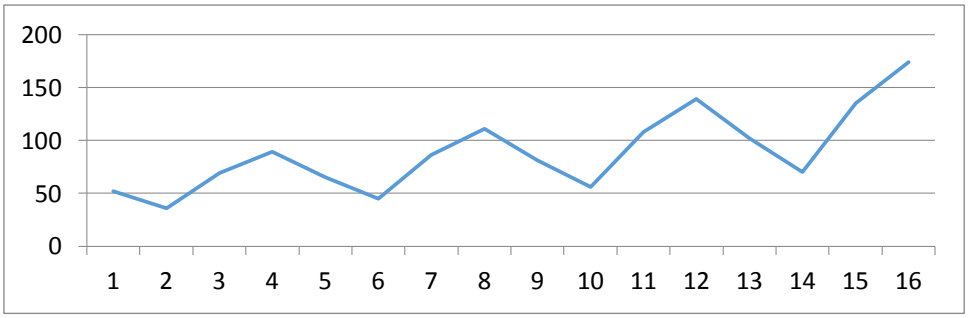
$$C_{S1} = \frac{-1{,}185 - 4{,}435 - 4{,}185}{3} = -3.268$$

La prévision est donnée par : $X_{t+h} = \hat{x}_{t+h} + C_{Si}$

Donc $X_{13} = \hat{x}_{13} + C_{S1}$. $\hat{x}_{13}$ = 4,03+ 0,71 (13) = 13,26.

Donc X_{13} = 13,26 + $(-3.268) = 9{,}992$

Corrige de l'exercice 2 :



Le graphe montre un mouvement saisonnier ou les ventes augmentent dans le deuxième et le quatrième trimestre et baissent dans le deuxième et le troisième trimestre.

b) le modèle de décomposition

D'après le tableau donnant les résultats de la régression de l'écart-type sur la moyenne, la probabilité associée au coefficient de la moyenne (0,000) est inferieur à 5%. Donc la moyenne est dépendante de l'écart- type. **Le modèle est donc multiplicatif.**

Corrige de l'exercice 2 :

Nous devons commencer par le calcul des donner sans tendance.

L'estimation de la tendance par la méthode des MCO nous donne les résultats consignés dans ce tableau

Dependent Variable: Y
Method: Least Squares
Date: 12/13/22 Time: 10:31
Sample: 2001 2016
Included observations: 16

Variable	Coefficient	Std. Error	t-Statistic	Prob.
T	5.820588	1.445567	4.026508	0.0012
C	39.15000	13.97797	2.800836	0.0142

R-squared	0.536620	Mean dependent var	88.62500
Adjusted R-squared	0.503521	S.D. dependent var	37.82922
S.E. of regression	26.65495	Akaike info criterion	9.520295
Sum squared resid	9946.806	Schwarz criterion	9.616869
Log likelihood	-74.16236	Hannan-Quinn criter.	9.525240
F-statistic	16.21276	Durbin-Watson stat	1.753725
Prob(F-statistic)	0.001249		

Donc : $\hat{y}_i$ = 39,15+ 5,82t_i

Pour calculer les donner sans tendance, nous devons diviser les donner $S_i = y_i/\hat{y}_i$

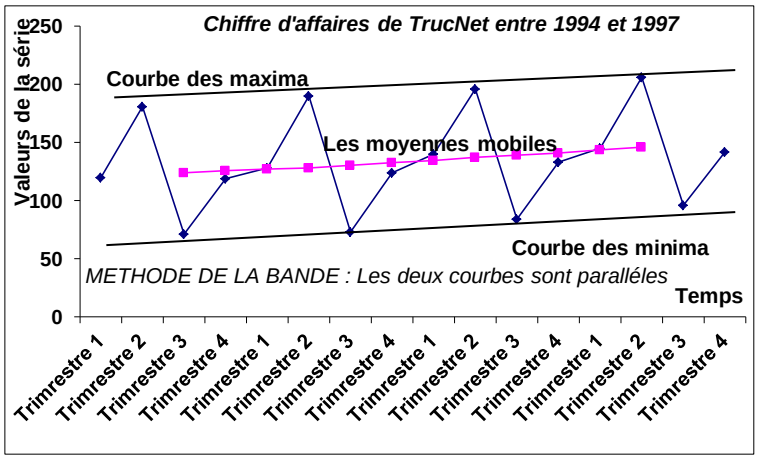
y	t	$\hat{y}_i$ =39.15+5.82t	$S_i = y_i/\hat{y}_i$	$\bar{S}$	$S^* = S_i/\bar{S}$	$Y_{CVS} = y_i/S^*$
52	1	44,97	1,1563264	1,127	1,0260217	50,68119

36	2	50,79	0,7088009	1,127	0,6289272	57,24033
69	3	56,61	1,2188659	1,127	1,0815137	63,79947
89	4	62,43	1,4255967	1,127	1,2649482	70,35861
65	5	68,25	0,952381	0,982	0,969838	67,0215
45	6	74,07	0,6075334	0,982	0,6186695	72,73674
86	7	79,89	1,0764802	0,982	1,096212	78,45198
111	8	85,71	1,2950648	0,982	1,3188032	84,16722
81	9	91,53	0,8849558	0,945	0,9364611	86,49585
56	10	97,35	0,575244	0,945	0,6087238	91,99575
108	11	103,17	1,0468159	0,945	1,1077417	97,49565
139	12	108,99	1,2753464	0,945	1,3495729	102,99555
102	13	114,81	0,8884244	0,962	0,923518	110,44722
70	14	120,63	0,5802868	0,962	0,6032088	116,04606
135	15	126,45	1,0676157	0,962	1,1097876	121,6449
174	16	132,27	1,315491	0,962	1,3674543	127,24374

Exercice 3 : Un modèle additif sur une série trimestrielle (

La représentation graphique des séries.

Graphique 1 - Chiffre d'affaires et moyennes mobiles de l'entreprise



Eléments de réponse à la question 1 :

- Pour calculer les coefficients saisonniers, il faut enlever la tendance de la série. Cette tendance est approchée par les moyennes mobiles centrées. Il faut calculer les

différences saisonnières, car le modèle est additif. Appelons Y les différences saisonnières, nous obtenons :

$$y_t = x_t - mm^4_t\,(X)$$

où $\mathbf{mm^4_t\,(X)}$ est la moyenne mobile de longueur 4, au temps t de la série X des chiffres d'affaires de TrucNet.

Les calculs du tableau 3 sont obtenus de la manière suivante :

- Les deux premières lignes et les deux dernières lignes du tableau pour les deux dernières colonnes n'existent pas. En effet, le calcul des moyennes mobiles centrées entraîne une perte de valeurs aux deux extrémités de la série. Cette perte de valeurs est égale à la longueur de la moyenne mobile, comme la moyenne mobile employée est centrée, la perte de points est par moitié repartie aux deux extrémités.
- La troisième ligne et les suivantes, pour les deux colonnes

Recherche de la première valeur de la seconde colonne du tableau 3 (moyenne mobile de longueur 4)

$$(0.5*120 + (181 + 71 + 119) + 0.5*128) / 4 = 123{,}75$$

Recherche de la première valeur de la troisième colonne du tableau 3 (différences saisonnières)

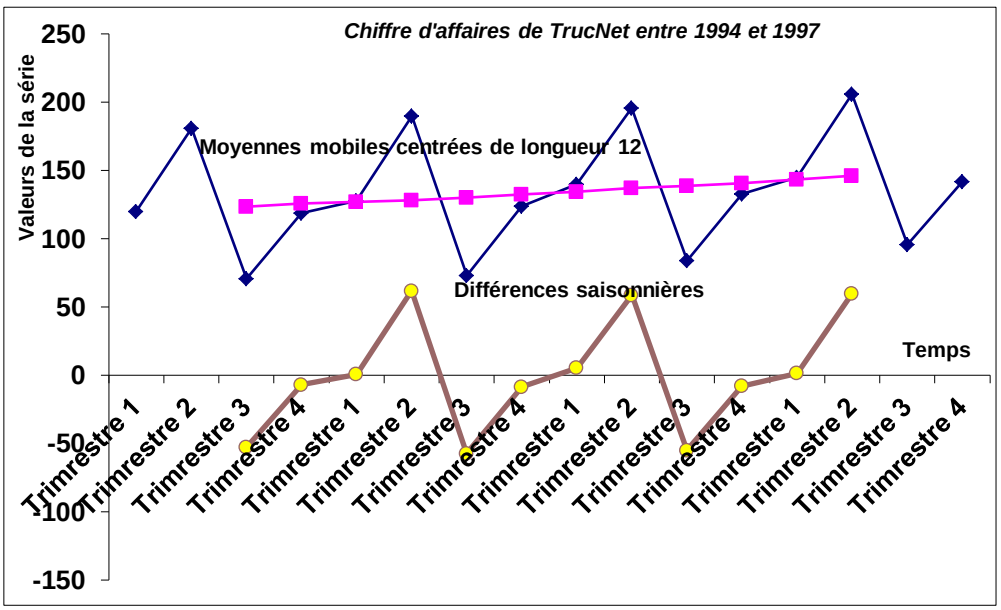
Tableau 3 - Exercice 1 - Différences saisonnières

	Série X	$mm_4(X)$	$X - mm_4(X)$
Trimrestre 1	120	X	X
Trimrestre 2	181	X	X
Trimrestre 3	71	**123.75**	**-52.75**
Trimrestre 4	119	**125.88**	**-6.88**
Trimrestre 1	128	**127.25**	**0.75**
Trimrestre 2	190	**128.13**	**61.88**
Trimrestre 3	73	**130.25**	**-57.25**
Trimrestre 4	124	**132.50**	**-8.50**
Trimrestre 1	140	**134.63**	**5.38**
Trimrestre 2	196	**137.13**	**58.88**
Trimrestre 3	84	**138.88**	**-54.88**
Trimrestre 4	133	**140.75**	**-7.75**
Trimrestre 1	145	**143.50**	**1.50**
Trimrestre 2	206	**146.13**	**59.88**
Trimrestre 3	96	X	X
Trimrestre 4	142	X	X

Commentaires :

Observons que cette opération nous permet d'éliminer la tendance, pour cela représentons la séries des différences saisonnières sur le même graphique que la série X.

Graphique 2 -Représentation des différences saisonnières



- Calculons les coefficients saisonniers. Notons que les calculs sont effectués pour chaque trimestre et de façon indépendante. Pour le premier trimestre nous devons calculer la moyenne des différences saisonnières de la façon suivante :

$$(0{,}75 + 5{,}38 + 1{,}50) / 3 = 2{,}54$$

Tableau 4 - Tableau des coefficients saisonniers

Les différences saisonnières permettent de trouver de façon indépendante les quatre coefficients saisonniers. Ainsi, nous sommes obligés de vérifier que leur somme est égale à zéro (voir tableau des coefficients saisonniers première colonne). Nous devons retrancher la moyenne des coefficients saisonniers pour les corriger. Nous obtenons la deuxième colonne du tableau des coefficients saisonniers.

Eléments de réponse à la question 2 :

Pour donner la série corrigée des variations saisonnières il nous faut retirer de la série les coefficients saisonniers correspondants.

Nous obtenons les coefficients définitifs S^*_j.

$$S_j^* = S_j' - \bar{S}'$$

La correction du premier coefficient saisonnier est la suivante :

$$2{,}54 - 0{,}02 = 2{,}52$$

Nous pouvons maintenant calculer la série corrigée des variations saisonnières. Ainsi, pour le premier trimestre de la première année, nous obtenons :

120 - 2,52 = 117,5

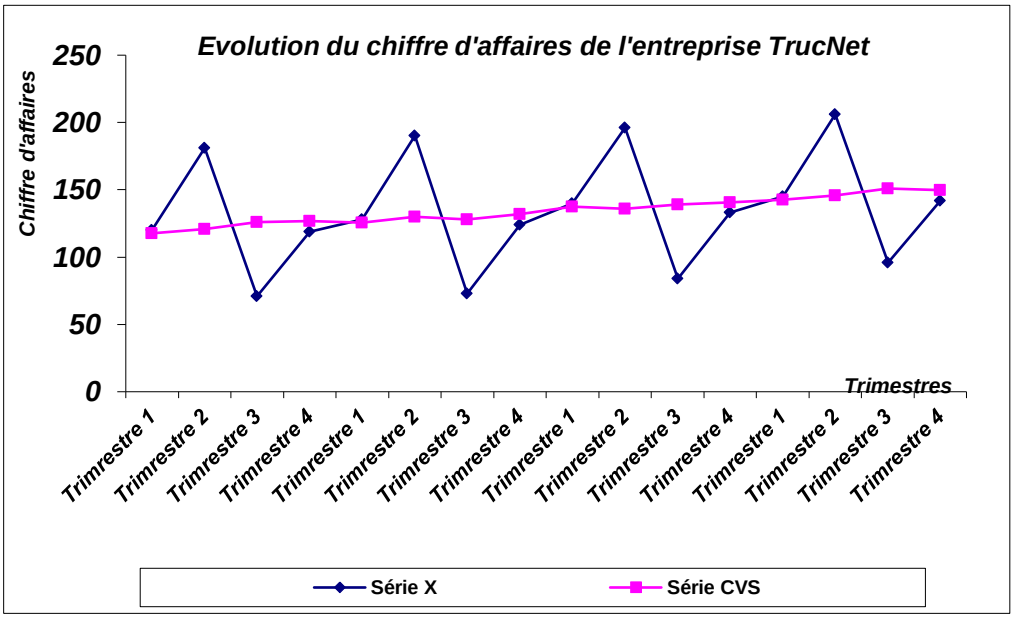
Nous retrancherons à tous les premiers trimestres des années suivantes le même coefficient saisonnier. Nous procéderons de la même manière pour les coefficients saisonniers suivants. Nous obtenons le tableau des calculs ci-dessous.

Tableau 5 - Tableau de la série corrigée des variations saisonnières

	Série X	Série CVS
Trimrestre 1	120	117.5
Trimrestre 2	181	120.8
Trimrestre 3	71	126.0
Trimrestre 4	119	126.7
Trimrestre 1	128	125.5
Trimrestre 2	190	129.8
Trimrestre 3	73	128.0
Trimrestre 4	124	131.7
Trimrestre 1	140	137.5
Trimrestre 2	196	135.8
Trimrestre 3	84	139.0
Trimrestre 4	133	140.7
Trimrestre 1	145	142.5
Trimrestre 2	206	145.8
Trimrestre 3	96	151.0
Trimrestre 4	142	149.7

Traçons sur le même graphique la série X et la série Z corrigée des variations saisonnières.

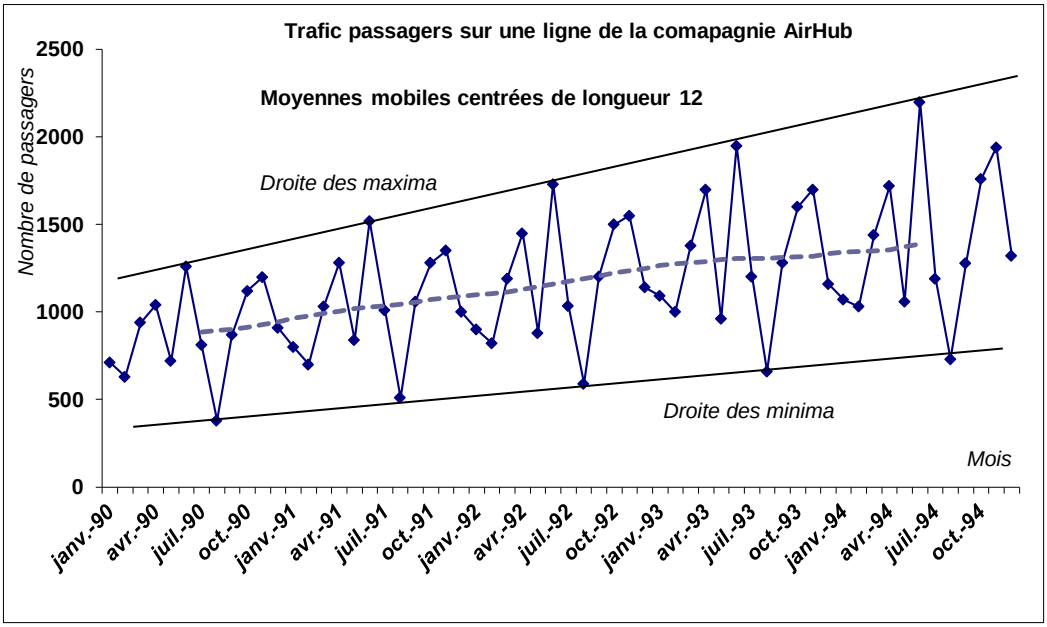
Graphique 3 - Série CVS des chiffres d'affaires de l'entreprise TrucNet



Corrigé de l'**exercice 4 : un modele multiplicatif sur une serie annuelle**

La représentons graphiquement la série.

Graphique 4 - Trafic de passagers d'une ligne de la compagnie AirHub



Eléments de réponse à la question 1 :

Pour calculer les coefficients saisonniers , il faut enlever la tendance (qui approchée par les moyennes mobiles centrées) de la série. Pour cela il nous faut calculer les rapports saisonnières (le modèle est multiplicatif). Appelons Y les rapports saisonniers que nous calculons en effectuant :

$$y_t = x_t / [mm_{12}(X)]_t$$

où $[mm_{12}(X)]_t$ est la moyenne mobile au temps t de la série X du trafic de passagers d'un ligne de la compagnie AirHub.

Nous obtenons pour le mois de juillet 1990 la valeur suivante :

$$812 / 886.5 = 0.9159$$

Nous procédons de la même manière pour tous les autres rapports saisonniers. Notons que comme les moyennes mobiles, les rapports saisonniers sont tronqués des 12 valeurs (6 valeurs en début de série et 6 valeurs en fin de série)

Tableau 6 - Tableau des rapports saisonniers de la compagnie AirHub

	janvier	février	mars	avril	mai	juin	juillet	août	septembre	octobre	novembre	décembre
1990	-	-	-	-	-	-	0.9159	0.4255	0.9669	1.2261	1.2924	0.9636
1991	0.8304	0.7164	1.0400	1.2736	0.8252	1.4787	0.9751	0.4880	1.0032	1.1958	1.2510	0.9178
1992	0.8187	0.7430	1.0692	1.2854	0.7682	1.4917	0.8811	0.4962	0.9988	1.2267	1.2535	0.9127
1993	0.8630	0.7842	1.0769	1.3189	0.7388	1.4926	0.9209	0.5051	0.9783	1.2179	1.2891	0.8700
1994	0.7966	0.7655	1.0680	1.2695	0.7729	1.5848	-	-	-	-	-	-

Le coefficient saisonnier du mois de janvier est obtenu en calculant la moyenne des rapports saisonniers des mois de janvier (voir tableau ci-dessous colonne mois de janvier). Notons que la moyenne s'effectue sur quatre années puisque nous avons perdu des valeurs

aux deux extrémités de la série des rapports saisonniers.

(0.8304+0.8187+0.8630+0.7966)/4=0.827

Nous procédons de la même façon pour tous les autres mois. Ainsi, les premiers coefficients sont calculés de manière indépendante. La somme des coefficients saisonniers ne peut pas être égale à la longueur de la moyenne mobile. Nous procédons à une correction de ces coefficients en calculant la moyenne et en divisant chacun des coefficients par la moyenne.

La somme des coefficients saisonniers :

0,827+0,752+1,064+1,287+0,776+1,512+0,923+0,479+0,987+1,217+1,271+0,916 = 12,011

Notons que la moyenne est proche de 12 :

(0,827+0,752+1,064+1,287+0,776+1,512+0,923+0,479+0,987+1,217+1,271+0,916)/12=1,001

Le coefficient saisonnier corrigé pour le mois de janvier est obtenu en effectuant le calcul suivant :

0,827 / 1,001 = 0,826

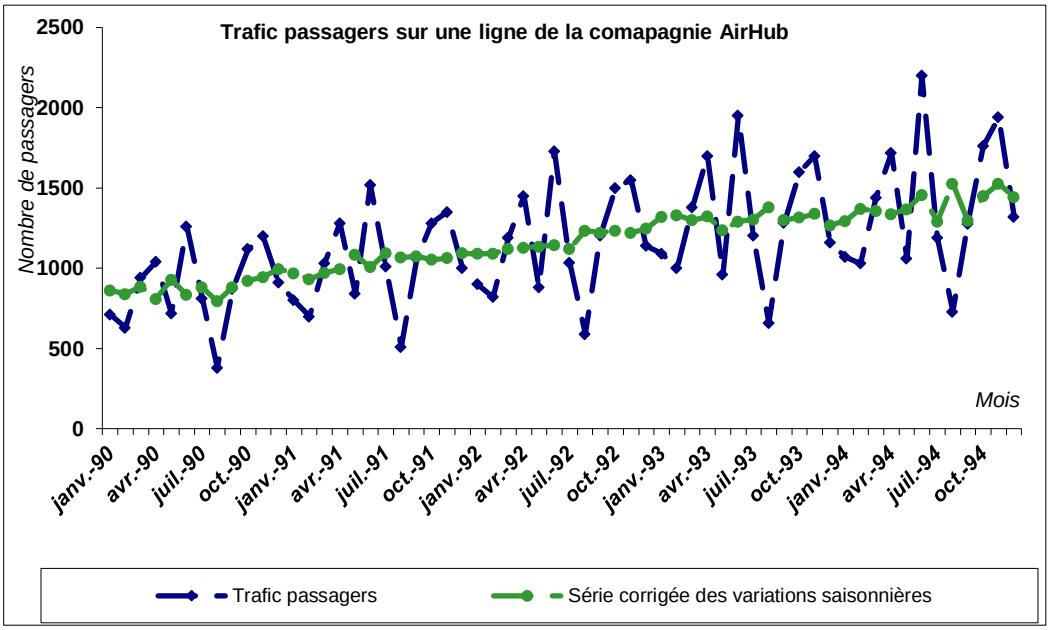
Nous procédons de la même façon pour tous les autres coefficients saisonniers.

Tableau 7 - Coeffcients saisonniers d'une série avec modèle multiplicatif

	Les coefficients saisonniers	
	non corrigés	*corrigés*
janvier	0,827	0,826
février	0,752	0,752
mars	1,064	1,063
avril	1,287	1,286
mai	0,776	0,776
juin	1,512	1,511
juillet	0,923	0,922
août	0,479	0,478
septembre	0,987	0,986
octobre	1,217	1,215
novembre	1,271	1,270
décembre	0,916	0,915
Total	12,011	12,000
moyenne	1,001	1,000

Notons que dans notre exemple la somme des coefficients saisonniers non corrigés est très voisine de la somme théorique (voir la première colonne du tableau 7).

Graphique 5 - Série corrigée des variations saisonnières



La série corrigée de variations saisonnières est obtenue en divisant chaque valeur de la série initiale par son coefficient saisonnier respectif. Ainsi, la première valeur de la série CVS (Corrigée des Variations Saisonnières) s'obtient :

$$713 / 0.826 = 863$$

(la valeur 0.826 correspond au coefficient saisonnier corrigé du mois de janvier)

Tableau 8 – Série corrigée des variations saisonnières – Série C.V.S

	Janvier	Février	Mars	Avril	Mai	Juin	Juillet	Août	Septembre	Octobre	Novembre	Décembre
1 990	863	838	885	809	928	834	880	795	882	921	945	994
1 991	968	931	969	996	1 083	1 006	1 095	1 066	1 075	1 053	1 063	1 093
1 992	1 089	1 091	1 120	1 128	1 135	1 145	1 121	1 234	1 220	1 234	1 220	1 246
1 993	1 321	1 331	1 299	1 322	1 238	1 291	1 304	1 380	1 300	1 316	1 338	1 267
1 994	1 295	1 370	1 355	1 338	1 367	1 456	1 290	1 526	1 296	1 448	1 527	1 442

Chapitre 4 : **Les modèles ARIMA et la méthodologie de Box & Jenkins**

Avant le traitement d'une série chronologique, il convient d'en étudier les caractéristiques stochastiques. Si ces caractéristiques – c'est-à-dire son espérance et sa variance – se trouvent modifiées dans le temps, la série chronologique est considérée comme non stationnaire ; dans le cas d'un processus stochastique invariant, la série temporelle est alors stationnaire. De manière formalisée, le processus stochastique y_t est stationnaire si : la moyenne est constante et indépendante du temps, la variance est finie et indépendante du temps ; et la covariance est indépendante du temps.

Ce chapitre définit les processus stochastiques stationnaires et non stationnaires, et insiste sur la distinction entre processus à tendance déterministe et à tendance stochastique. Il présente également les principaux tests de racine unitaire, décrit les modèles ARMA et ARIMA et explique comment les utiliser pour modéliser l'évolution d'une série temporelle.

I. Etude de la stationnarité

Avant le traitement d'une série chronologique, il convient de s'assurer de sa stationnarité car la stationnarité constitue une condition nécessaire pour éviter les régressions fallacieuses, de telles régissions se réalisent lorsque les variables ne sont pas stationnaires, l'estimation des coefficients par la méthode des moindre carrés ordinaires (MCO) ne converge pas vers les vrais coefficients et les tests usuels des **t** de Student et **f** Fisher ne sont plus valides.

Quelques concepts méritent d'être définis.

1.1 La fonction d'autocorrélation (FAC) est la fonction notée ρ_k qui mesure la corrélation de la série avec elle-même décalée de *k* périodes. Sa formulation est la suivante :

$$\rho_k = \frac{cov(y_t\ , y_{t-1})}{\sigma_{yt}\sigma_{yt-k}} = \frac{\sum_{t=k+1}^{n}(y_t - \bar{y})(y_{t-k} - \bar{y})}{\sqrt{\sum_{t=k+1}^{n}(y_t - \bar{y})^2}\ \sqrt{\sum_{t=k+1}^{n}(y_{t-k} - \bar{y})^2}}$$

Cette formule est difficile à manier puisqu'elle exige de recalculer pour chaque terme ρ_k les moyennes et les variances. il lui est préféré d'utiliser la fonction d'autocorrélation d'échantillonnage :

$$\hat{\rho}_k = \frac{\sum_{t=k+1}^{n}(y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t=1}^{n}(y_t - \bar{y})^2}$$

La fonction d'autocorrélation partielle (FAP) s'apparente à la notion de corrélation partielle. Le coefficient de corrélation partielle est définie comme étant le calcul de l'influence de x_1 sur x_2 en éliminant les influences des autres variables x_3 , x_4 , . . . , x_k .

Par analogie, nous pouvons définir l'autocorrélation partielle de retard k comme le coefficient de corrélation partielle entre y_t et y_{t-k} , c'est-à-dire comme étant la corrélation entre y_t et y_{t-k} l'influence des autres variables décalées de k périodes (y_{t-1} , y_{t-2} , . . . , y_{t-k+1}) ayant été retirée.

Afin d'éviter par la suite toutes ambiguïtés entre les deux fonctions d'autocorrélation, nous appelons fonction d'autocorrélation simple, la fonction d'autocorrélation.

1.2 Tests de « bruit blanc » et de stationnarité

L'étude de stationnarité s'effectue essentiellement à partir de l'étude des fonctions d'autocorrélation (ou de leur représentation graphique appelée « corrélogramme »). Une série chronologique est stationnaire si elle ne comporte ni tendance ni saisonnalité. Nous allons donc, à partir de l'étude du corrélogramme d'une série, essayer de montrer de quelle manière nous pouvons mettre en évidence ces deux composantes. Différents types de séries stationnaires peuvent être distingués :

- à mémoire, c'est-à-dire dont on peut modéliser, par une loi de reproduction, le processus ;
- identiquement et indépendamment distribuée notée i.i.d. ou appelée Bruit Blanc (« White Noise »);
- normalement (selon une loi normale) et indépendamment distribuée notée n.i.d. ou appelée Bruit Blanc gaussien.

Tests de « bruit blanc » sous eviws

Eviews fournit les résultats des fonctions d'autocorrélation simple (colonne AC) et partielles (colonne PAC), avec les correlogrammes respectifs. Les bornes de l'intervalle de confiance sont stylisées par des pointillés horizontaux ; chaque terme qui sort de cet intervalle est donc significativement différent de zéro au seuil de 5%.

1.3 Série stationnaire

Une série est stationnaire si ses caractéristiques (espérance et variance) se trouvent invariantes dans le temps. Une série pour t=1, ..., t est dite stationnaire si :

- La moyenne est constante et indépendante du temps ; $E(X_t) = E(X_{t+k}) = \mu$
- La variance est définie et indépendante du temps ; $V(X_t) < \infty$
- La covariance est indépendante du temps ;

$$Cov(X_t, X_{t+k}) = E[(Y_t - \mu)(Y_{t+k} - \mu)] = \gamma_k$$

Il existe deux types de séries temporelles :

Le bruit blanc est un cas particulier de séries temporelles stochastiques pour lequel la valeur prise par *X* à la date *t* s'écrit :

$$X_t = \varepsilon_t$$

Un processus stochastique X ou (X_t) est un bruit blanc si[27] :

- **$E(X_t) = 0$** ; quelque soit t ;
- **$V(X_t) = \sigma_x^2$;** quelque soit t ;
- **$Cov(X_t, X_\Theta) = 0$;** quelque soit $t \neq \Theta$.

Les principales propriétés d'une série de bruit blanc sont :

- Il n'ya pas de corrélation entre les termes de la série ;
- Les valeurs passées de la série ne permettent pas de prévoir les valeurs futures de la série.

✓ **Série marche au hasard (aléatoire)**

C'est un autre cas particulier de processus stochastique pour lequel la valeur prise par la variable Y à la date « t » est régie par l'équation ;

$$Y_t = Y_{t-1} + \varepsilon_t$$

Où : ε_t est une variable aléatoire qui présente les mêmes propriétés.

1.3 Série non stationnaire

Il existe deux types de processus non stationnaires :

✓ **Processus TS (Trend Stationary)**

Il représente une non-stationnarité de nature déterministe. Le processus TS s'écrit :

$$X_t = f(t) + \varepsilon_t$$

f : est une fonction polimoniale du temps ;

[27] Eric DOR, op.cit, p 163.

ε_t : est un processus stationnaire.

- ✓ **Processus DS (Différence Stationary)**

Le processus DS est un processus qu'ont peut rendre stationnaire par l'utilisation de la différenciation :

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

On peut définir deux types de processus DS :

- Le processus DS avec dérive ($\beta \neq 0$) s'exprime comme suit :

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

- Le processus DS sans dérive ($\beta = 0$) s écrit :

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

1.4 Test de racine unitaire

La stationnarité est une condition nécessaire pour l'étude de toute série chronologique dans l'approche classique, car les analyses économétriques ne s'appliquent qu'à des séries stationnaires. Les tests de racine unitaire « *Unit Root Test* » permettent non seulement de détecter l'existence d'une non-stationnarité mais aussi de déterminer de quelle non-stationnarité il s'agit (processus TS ou DS), et donc la bonne méthode pour stationnariser la série.[28]

- ➢ **Tests de Dickey-Fuller simples (1979)**

Les tests de Dickey-Fuller (DF) permettent de mettre en évidence le caractère stationnaire ou non d'une chronique par la détermination d'une tendance déterministe ou stochastique. Les modèles servant de base à la construction de ces tests sont au nombre de trois :

Modèle [1] : $\boldsymbol{\Phi_1 X_{t-1}} + \varepsilon_t$ Modèle autorégressif d'ordre 1 ;

Modèle [2] : $\boldsymbol{\Phi_1 X_{t-1}} + \boldsymbol{c} + \varepsilon_t$ Modèle autorégressif avec constante ;

Modèle [3] : $\boldsymbol{\Phi_1 X_{t-1}} + \boldsymbol{c} + \boldsymbol{\beta_t} + \varepsilon_t$ Modèle autorégressif avec tendance.

Le principe des tests est simple : si l'hypothèse H_0 : $\varphi = 1$ est retenue dans l'un de ces trois modèles, le processus est alors non stationnaire[29]. En effet, si l'hypothèse H_0 est vérifiée, la chronique X_t n'est pas stationnaire quel que soit le modèle retenu.

[28] BOURBONNAIS Régis (2015), Econometrie, 9e édition , *2015*, p 248.

[29] Régis BOURBONNAIS, « économétrie », Dunod, 7ème édition, Paris, 2009, P233.

- **Tests de Dickey-Fuller Augmentés**

Dans les modèles précédents, utilisés pour les tests de Dickey-Fuller simple, le processus εt est par hypothèse, un bruit blanc. Or il n'y a aucune raison pour que, à priori, l'erreur soit corrélée ; on appelle tests de Dickey et Fuller Augmentés (ADF, 1981) la prise en compte de cette hypothèse.

Les tests ADF sont fondés, sous l'hypothèse alternative $|\boldsymbol{\Phi}_1| < 1$, sur l'estimation par les MCO des trois modèles :

Modèle [4] : $\Delta \mathbf{X} = \boldsymbol{\rho X_{t-1}} - \sum_{j=2}^{p} \boldsymbol{\Phi_j} \Delta \boldsymbol{X_{t-j+1}} + \varepsilon_t$;

Modèle [5] : $\Delta \mathbf{X} = \mathbf{\rho X_{t-1}} - \sum_{j=2}^{p} \mathbf{\Phi_j} \Delta \mathbf{X_{t-j+1}} + \mathbf{c} + \varepsilon_t$;

Modèle [6] : $\Delta \mathbf{X} = \mathbf{\rho X_{t-1}} - \sum_{j=2}^{p} \mathbf{\Phi_j} \Delta \mathbf{X_{t-j+1}} + \mathbf{c} + \mathbf{b_t} + \varepsilon_t$.

Le test se déroule de manière similaire aux tests DF simples, seules les tables statistiques diffèrent.

Stratégie simplifiée des tests de racine unitaire

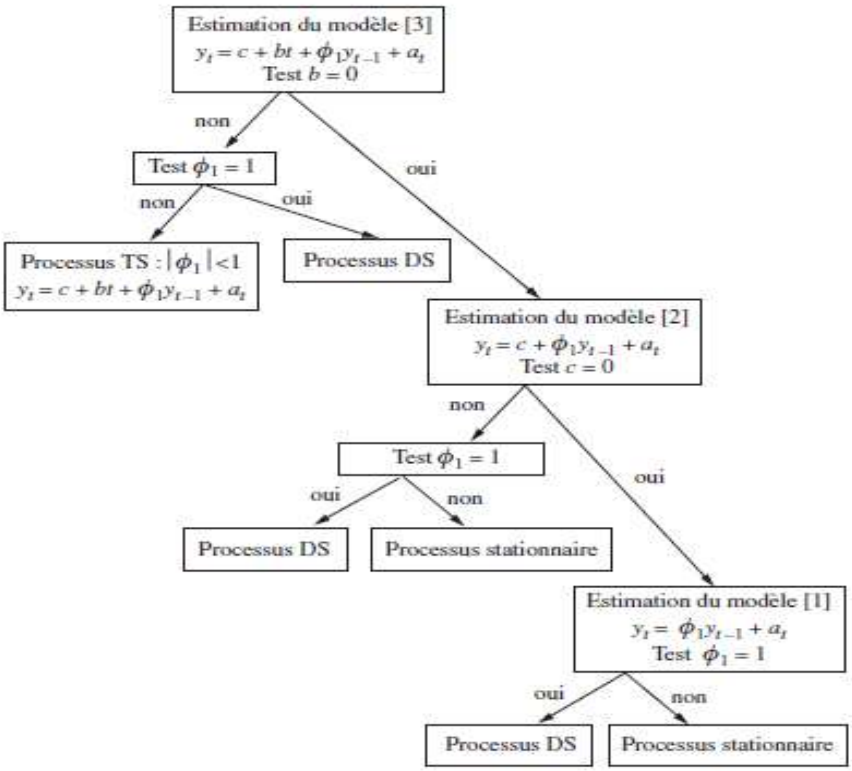


Schéma 1 – Stratégie simplifiée des tests de racine unitaire

Source : BOURBONNAIS Régis (2015), Econometrie, 9e édition , *2015, P251*

Exemple 1 : Le Test ADF

Suivant la stratégie simplifiée des tests de racine unitaire, nous commençons par tester l'hypothèse de l'existence de la tendance, On estimera alors le modèle 3 de ADF.

Application du modèle 3:

Les résultats du test ADF sur la série EXPO nous donne les résultats suivants :

Null Hypothesis: EXPO has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 0 (Automatic - based on SIC, maxlag=9)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-1.961616	0.6065
Test critical values:	1% level	-4.165756	
	5% level	-3.508508	
	10% level	-3.184230	

*MacKinnon (1996) one-sided p-values.
Augmented Dickey-Fuller Test Equation
Dependent Variable: D(EXPO)
Method: Least Squares
Date: 02/25/19 Time: 16:27
Sample (adjusted): 1971 2017
Included observations: 47 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
EXPO(-1)	-0.175752	0.089595	-1.961616	0.0562
C	-1.31E+08	2.60E+09	-0.050274	0.9601
@TREND(1970)	2.16E+08	1.53E+08	**1.408512**	0.1660

R-squared	0.082266	Mean dependent var	7.84E+08
Adjusted R-squared	0.040551	S.D. dependent var	8.51E+09
S.E. of regression	8.33E+09	Akaike info criterion	48.58631
Sum squared resid	3.05E+21	Schwarz criterion	48.70440
Log likelihood	-1138.778	Hannan-Quinn criter.	48.63075
F-statistic	1.972083	Durbin-Watson stat	1.902371
Prob(F-statistic)	0.151277		

Test du trend:

H_0 : B=0

H_1 : B≠0

Tb = |1.40| < T^{ADF}= **3.18**. On accepte H_0 : B=0, la tendance n'est pas significative. On passe à l'estimation du modèle 02

Modèle 2 : $X_t = c + \phi_1 X_{t-1} + a_t$

Null Hypothesis: EXPO has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic - based on SIC, maxlag=9)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-1.385033	0.5817
Test critical values:	1% level	-3.577723	

5% level		-2.925169
10% level		-2.600658

*MacKinnon (1996) one-sided p-values.
Augmented Dickey-Fuller Test Equation
Dependent Variable: D(EXPO)
Method: Least Squares
Date: 02/25/19 Time: 16:28
Sample (adjusted): 1971 2017
Included observations: 47 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
EXPO(-1)	-0.073453	0.053033	-1.385033	0.1729
C	2.56E+09	1.78E+09	1.442181	0.1562

R-squared	0.040886	Mean dependent var	7.84E+08
Adjusted R-squared	0.019573	S.D. dependent var	8.51E+09
S.E. of regression	8.42E+09	Akaike info criterion	48.58786
Sum squared resid	3.19E+21	Schwarz criterion	48.66659
Log likelihood	-1139.815	Hannan-Quinn criter.	48.61748
F-statistic	1.918317	Durbin-Watson stat	2.015584
Prob(F-statistic)	0.172871		

Test de la constante :

$H_0 : C=0$
$H_1 : C\neq 0$

$Tc = |1.44| < T^{ADF} = \mathbf{2.89}$, on accepte $H_0 : C= 0$, la constante n'est pas significative. On passe à l'estimation du modèle 01.

Modèle 1 : $X_t = \phi_1 X_{t-1} + a_t$

Null Hypothesis: EXPO has a unit root
Exogenous: None
Lag Length: 0 (Automatic - based on SIC, maxlag=9)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-0.490171	0.4984
Test critical values:	1% level	-2.615093	
	5% level	**-1.94**7975	
	10% level	-1.612408	

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

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(EXPO)
Method: Least Squares
Date: 02/25/19 Time: 16:28
Sample (adjusted): 1971 2017
Included observations: 47 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
EXPO(-1)	-0.018176	0.037081	-0.490171	0.6263

R-squared	-0.003444	Mean dependent var	7.84E+08
Adjusted R-squared	-0.003444	S.D. dependent var	8.51E+09

S.E. of regression	8.52E+09	Akaike info criterion	48.59049
Sum squared resid	3.34E+21	Schwarz criterion	48.62985
Log likelihood	-1140.876	Hannan-Quinn criter.	48.60530
Durbin-Watson stat	2.035727		

Test de ϕ:

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

$T\phi = -0.49 > T^{ADF}(5\%) = -1.94$. On accepte H_0 $\phi = 1$, le processus est **non stationnaire**
Le processus de cette série est un processus **« DS sans dérive »**

La stationnarisation de la série et récupération de l'ordre d'intégration :

Null Hypothesis: D(EXPO) has a unit root
Exogenous: None
Lag Length: 0 (Automatic - based on SIC, maxlag=9)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-6.921577	0.0000
Test critical values:	1% level	-2.616203	
	5% level	**-1.948140**	
	10% level	-1.612320	

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

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(EXPO,2)
Method: Least Squares
Date: 02/25/19 Time: 16:29
Sample (adjusted): 1972 2017
Included observations: 46 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(EXPO(-1))	-1.034365	0.149441	**-6.921577**	0.0000

R-squared	0.515618	Mean dependent var	1.01E+08
Adjusted R-squared	0.515618	S.D. dependent var	1.24E+10
S.E. of regression	8.63E+09	Akaike info criterion	48.61695
Sum squared resid	3.35E+21	Schwarz criterion	48.65670
Log likelihood	-1117.190	Hannan-Quinn criter.	48.63184
Durbin-Watson stat	1.997989		

Test du ϕ:

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

$T\phi = -6.92 < T^{ADF}(5\%) = -1.94$. On accepte H_1 $\phi < 1$, le processus est **stationnaire**
Le processus EXPO est devenu stationnaire avec une seule différenciation. Donc **EXPO→I(1)**

<u>Exemple 2 : Une serie comportant une tendance</u>

L'Application du test ADF, modèle 3, sur la série Y donne les résultats de l'estimation sous eviews, sont consignés dans le tableau suivant :

Null Hypothesis: Y has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 6 (Automatic - based on SIC, maxlag=9)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-5.041157	0.0016
Test critical values:	1% level	-4.284580	
	5% level	-3.562882	
	10% level	-3.215267	

*MacKinnon (1996) one-sided p-values.
Augmented Dickey-Fuller Test Equation
Dependent Variable: D(Y)
Method: Least Squares
Date: 04/27/19 Time: 19:13
Sample (adjusted): 1987 2017
Included observations: 31 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
Y(-1)	-0.534336	0.105995	-5.041157	0.0000
D(Y(-1))	0.430805	0.145076	2.969523	0.0071
D(Y(-2))	-0.031805	0.169304	-0.187860	0.8527
D(Y(-3))	0.557266	0.206000	2.705175	0.0129
D(Y(-4))	0.331010	0.244879	1.351729	0.1902
D(Y(-5))	0.402116	0.215548	1.865552	0.0755
D(Y(-6))	0.674157	0.232865	2.895058	0.0084
C	-9.686533	2.969369	-3.262152	0.0036
@TREND("1980")	1.692245	0.340462	**4.970432**	0.0001

R-squared	0.650021	Mean dependent var	3.428087
Adjusted R-squared	0.522755	S.D. dependent var	5.468270
S.E. of regression	3.777639	Akaike info criterion	5.733776
Sum squared resid	313.9523	Schwarz criterion	6.150095
Log likelihood	-79.87353	Hannan-Quinn criter.	5.869486
F-statistic	5.107604	Durbin-Watson stat	2.155190
Prob(F-statistic)	0.001106		

Test du trend:

$\begin{cases} H_0 : B=0 \\ H_1 : B\neq 0 \end{cases}$

Tb $= |4.97| < T^{ADF} = 3.18$, on rejette $H_0 : B=0$, la tendance est significative.

On passe au test de **ϕ**

Test du ϕ:

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

$T\phi = -5.04 < T^{ADF}(5\%) = -3.65$; on rejette H_0 $\phi = 1$, le processus est un TS**.** Il convient de le stationnariser en retranchant la tendance de la serie Y par la methode des MCO :

L'estimation de l'équation de la tendance (par les (MCO) ,

Les résultats de l'estimation sont donnés dans le tableau ci-apres :

Dependent Variable: T
Method: Least Squares
Date: 04/27/19 Time: 19:45
Sample: 1980 2017
Included observations: 38

Variable	Coefficient	Std. Error	t-Statistic	Prob.
@TREND(1980)	0.241394	0.036592	6.596920	0.0000

R-squared	-1.043790	Mean dependent var	6.217105
Adjusted R-squared	-1.043790	S.D. dependent var	3.393242
S.E. of regression	4.851019	Akaike info criterion	6.022218
Sum squared resid	870.6983	Schwarz criterion	6.065312
Log likelihood	-113.4221	Hannan-Quinn criter.	6.037551
Durbin-Watson stat	0.059825		

- **Tester la stationnarité des résidus en menant le test ADF avec le premier modèle.**

Null Hypothesis: RESID01Y has a unit root
Exogenous: None
Lag Length: 3 (Automatic - based on SIC, maxlag=9)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-2.762418	0.0072
Test critical values:	1% level	-2.634731	
	5% level	-1.951000	
	10% level	-1.610907	

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

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(RESID01Y)
Method: Least Squares
Date: 04/27/19 Time: 20:53
Sample (adjusted): 1984 2017
Included observations: 34 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
RESID01Y(-1)	-0.201467	0.072931	-2.762418	0.0097
D(RESID01Y(-1))	0.466225	0.155104	3.005886	0.0053
D(RESID01Y(-2))	-0.124718	0.168120	-0.741838	0.4640
D(RESID01Y(-3))	0.584671	0.196173	2.980377	0.0057

R-squared	0.381199	Mean dependent var	0.375246
Adjusted R-squared	0.319319	S.D. dependent var	5.308341
S.E. of regression	4.379561	Akaike info criterion	5.901905
Sum squared resid	575.4166	Schwarz criterion	6.081477
Log likelihood	-96.33238	Hannan-Quinn criter.	5.963144
Durbin-Watson stat	2.060358		

Donc, le processus **resid01y**→I(0), par contre la série T est intégré d'ordre 1.

Exemple 3: **Une serie comportant une contante**

Apres avoir testé la tendance, nous l'avons trouvé non significative, nous avons ainsi passe à l'estimation du modèle 2. Le résultat de l'estimation de ce modèle est donné dont le tableau suivant :

Null Hypothesis: M has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic - based on SIC, maxlag=9)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-4.899478	0.0003
Test critical values:	1% level	-3.621023	
	5% level	-2.943427	
	10% level	-2.610263	

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

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(M)
Method: Least Squares
Date: 04/27/19 Time: 20:13
Sample (adjusted): 1981 2017
Included observations: 37 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
M(-1)	-0.820393	0.167445	-4.899478	0.0000
C	12.30044	2.990904	4.112615	0.0002

R-squared	0.406829	Mean dependent var	-0.246343
Adjusted R-squared	0.389881	S.D. dependent var	12.03302
S.E. of regression	9.399008	Akaike info criterion	7.371624
Sum squared resid	3091.947	Schwarz criterion	7.458700
Log likelihood	-134.3750	Hannan-Quinn criter.	7.402322
F-statistic	24.00489	Durbin-Watson stat	1.977555
Prob(F-statistic)	0.000022		

Test de la constante :

$$\begin{cases} H_0 : C=0 \\ H_1 : C\neq 0 \end{cases}$$

$Tc = |4.11| > T^{ADF}$, donc on accepte $H_1 : C\neq 0$, la constante est significative. On passe au test de ϕ

Test de ϕ:

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

$T\phi = -4.899478 < T^{ADF}(5\%) = -2.96$ on rejette H_0 $\phi = 1$, le processus est **stationnaire**
Le processus M est intégré d'ordre 0 ; Donc, M→I(0)

II- Le processus ARMA

2.1 Analyse des fonctions d'autocorrélation

L'analyse de la fonction d'autocorrélation d'une série chronologique permet de savoir quels sont les termes ρ_k qui sont significativement différents de 0. En effet, par exemple, si aucun terme n'est significativement différent de 0, on peut en conclure que le processus étudié est sans mémoire et donc qu'à ce titre il n'est affecté ni de tendance ni de saisonnalité. Ou encore si une série mensuelle présente une valeur élevée pour ρ_{12} (corrélation entre y_t et y_{t-12}), la série étudiée est certainement affectée d'un mouvement saisonnier.

Le test d'hypothèses pour un terme ρ_k est le suivant :

H0 : $\rho_k = 0$

H1 : $\rho_k =/ 0$

Quenouille a démontré que pour un échantillon de taille importante (n > 30) , le coefficient ρ_k tend de manière asymptotique vers une loi normale de moyenne 0 et d'écart type $1/\sqrt{n}$.

L'intervalle de confiance du coefficient ρ_k est alors donné par :$\rho_k = 0 \pm t^{\frac{\alpha}{2}} \frac{1}{\sqrt{n}}$

Si le coefficient calculé $\hat{\rho}_k$ est à l'extérieur de cet intervalle de confiance, il est significativement différent de 0 au seuil α (en général α = 0,05 et $t^{\alpha/2}$ = 1,96). La plupart des logiciels fournissent, avec le corrélogramme, l'intervalle de confiance, ce qui autorise une interprétation instantanée.

Dans le cas où le corrélogramme ne laisse apparaître aucune décroissance de ses termes (absence de « cut off »), nous pouvons en conclure que la série n'est pas stationnaire en tendance.

2.2 Statistiques de Box-Pierce et Ljung-Box

Le test de Box-Pierce permet d'identifier les processus sans mémoire (suite de variables aléatoires indépendantes entre elles). La $cov(y_t , y_{t-1}) = 0$ doit être identifiée (ou encore $\rho_k = 0$ ∀k.

Un processus de bruit blanc implique que $\rho_1 = \rho_2 = \ldots = \rho_h = 0$, soit les hypothèses :

H0 : $\rho_1 = \rho_2 = \ldots = \rho_h = 0$

H1 : il existe au moins un ρi significativement différent de 0.

Pour effectuer ce test, on recourt à la statistique Q (due à Box-Pierce1) qui est donnée par :

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

h = nombre de retards, $\hat{\rho}_k^2$ = autocorrélation empirique d'ordre k, n = nombre d'observations. La statistique Q est distribuée de manière asymptotique comme un $\chi 2$ (chi deux) à h degrés de liberté. L'hypothèse de bruit blanc est rejetée, au seuil α, si la statistique Q est supérieure au $\chi 2$ lu dans la table au seuil $(1 - \alpha)$ et h degrés de liberté.

Une autre statistique peut aussi etre utilisée, dont les propriétés asymptotiques sont meilleures, dérivée de la première qui est le Q′ de Ljung et Box 2 :

$$Q' = n\,(n+2) \sum_{k=1}^{h} \frac{\hat{\rho}_k^2}{n-k}$$

qui est distribuée selon un $\chi 2$ à h degrés de liberté et dont les règles de décisions sont identiques au précédent. Ces tests sont appelés par les anglosaxons : « portmanteau test ».

2.3 Typologie des modèles AR, MA et ARMA

Il sera présenté les processus autorégressifs et les processus de moyenne mobile.

1) Modèle AR (Auto Régressif)

Processus stochastiques autorégressifs stationnaires

Les processus autorégressifs peuvent s'écrire comme la somme d'une constante, de la valeur courante d'un bruit blanc et d'une combinaison linéaire finie de leurs valeurs passées. Un processus stochastique {Xt} est dit autorégressif d'ordre p, et noté AR(p) ou ARMA (p, 0), si : $\forall t : X_t = \mu + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + a_t$ où le processus stochastique $\{a_t\}$ est un bruit blanc

Exemple : Processus AR(1) ou ARMA(1, 0) : $\forall t : X_t = \mu + \phi_1 X_{t-1} + a_t$ où $\{a_t\}$ est un bruit blanc.

Exemple Processus AR(2): $\forall t : X_t = \mu + \phi_1 X_t - 1 + \phi_2 X_{t-2} + a_t$ où $\{a_t\}$ est un bruit blanc.

Dans le processus autorégressif d'ordre p, l'observation présente y_t est générée par une moyenne pondérée des observations passées jusqu'à la p-ième période sous la forme suivante :

AR(1) : $y_t = \theta_1 y_{t-1} + \varepsilon_t$

AR(2) : $y_t = \theta_1 y_{t-1} + \theta 2\, y_{t-2} + \varepsilon_t$

. . .

AR(p) : $y_t = \theta_1 y_{t-1} + \theta 2\, y_{t-2} + \ldots + \theta p\, y_{t-p} + \varepsilon t$ ….[4]

où $\theta_1, \theta_2, \ldots, \theta_p$ sont des paramètres à estimer positifs ou négatifs, ε_t est un aléa gaussien. Nous pouvons ajouter à ce processus une constante qui ne modifie en rien les propriétés stochastiques. L'équation [4] peut aussi s'écrire à l'aide de l'opérateur décalage D :

$(1 - \theta_1 D - \theta_2 D_2 - \ldots - \theta_p D\, p)y_t = \varepsilon_t$

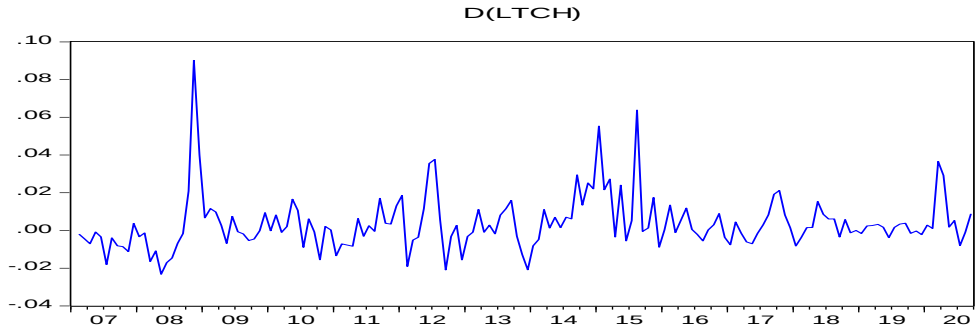
Il est démontré que le corrélogramme simple d'un processus AR(p) est caractérisé par une décroissance géométrique de ses termes de type : $\rho_k = \rho^k$

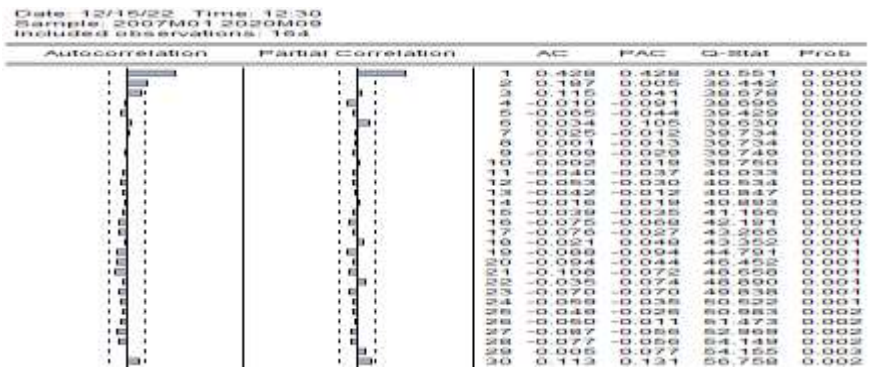
Le corrélogramme partiel a ses seuls *p* premiers termes différents de 0.

Exemple :

Selon les informations données dans la figure (2) et le tableau (1) de cet exemple, la série stationnaire (en première difference) D(LTCH) (voir figure 1) est-elle générée par AR(1)

Figure 1 : Evolution de la série D(LTCH)





L'identification des termes AR est basée sur l'examen des fonctions d'autocorrélation et d'autocorrélation partielle. Si le correlogramme partiel n'a que ses q premiers retards différents de zéro et que les termes du correlogrammes simple diminuent lentement, cela caractérise un AR(P)

Dependent Variable: D(LTCH)
Method: ARMA Maximum Likelihood (OPG - BHHH)
Date: 12/15/22 Time: 12:31
Sample: 2007M02 2020M09
Included observations: 164
Convergence achieved after 9 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.003604	0.002426	1.485288	0.1394
AR(1)	0.425818	0.070187	6.066887	0.0000
SIGMASQ	0.000172	9.90E-06	17.32579	0.0000

R-squared	0.183223	Mean dependent var	0.003605
Adjusted R-squared	0.173077	S.D. dependent var	0.014539
S.E. of regression	0.013221	Akaike info criterion	-5.794636
Sum squared resid	0.028143	Schwarz criterion	-5.737931
Log likelihood	478.1602	Hannan-Quinn criter.	-5.771616
F-statistic	18.05811	Durbin-Watson stat	1.997100
Prob(F-statistic)	0.000000		

Inverted AR Roots	.43

Le coefficient AR(1) est significativement différent de zéro. Les autres statistiques DW (1, 99) , F (18,05) empirique laissent présager d'un bon ajustement. Il convient maintenant d'analyser les résidus à partir de sa fonction d'autocorrélation.

Date: 12/19/22 Time: 18:48
Sample: 2007M01 2020M09
Included observations: 164
Q-statistic probabilities adjusted for 1 ARMA term

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.000	0.000	1.E-06	
		2	-0.012	-0.012	0.0253	0.874
		3	0.073	0.073	0.9279	0.629
		4	-0.041	-0.041	1.2071	0.751
		5	-0.109	-0.107	3.2229	0.521
		6	0.072	0.067	4.1069	0.534
		7	0.020	0.024	4.1736	0.653
		8	-0.007	0.008	4.1814	0.759
		9	-0.015	-0.034	4.2204	0.837
		10	0.028	0.020	4.3569	0.886
		11	-0.031	-0.015	4.5315	0.920
		12	-0.034	-0.032	4.7428	0.943
		13	-0.025	-0.034	4.8544	0.963
		14	0.019	0.019	4.9189	0.977
		15	-0.009	0.002	4.9337	0.987
		16	-0.049	-0.057	5.3733	0.988
		17	-0.060	-0.072	6.0500	0.988
		18	0.054	0.058	6.5979	0.988
		19	-0.067	-0.051	7.4335	0.986

Tous les termes des fonctions d'autocorrélation simple et partielle sont tous situés dans l'intervalle de confiance matérialisé par les traits verticaux. Le résidu peut être assimilé à

un processus de bruit blanc. L'estimation du modèle AR(1) est donc validée, la série peut être valablement représentée par un processus de type AR(1) sur la série différenciée.

2) Modèle MA (Moving Average : Moyenne Mobile)

Dans le processus de moyenne mobile d'ordre *q*, chaque observation y_t est générée par une moyenne pondérée d'aléas jusqu'à la *q*-ième période.

MA(1) : $y_t = \varepsilon_t - \alpha_1\varepsilon_{t-1}$

MA(2) : $y_t = \varepsilon_t - \alpha_1\varepsilon_{t-1} - \alpha_2\varepsilon_{t-2}$

. . .

MA(q) : yt = εt − α1εt−1 − α2εt−2 − . . . − αq εt−q

où $\alpha_1,\alpha_2,\ldots,\alpha_q$ sont des paramètres pouvant être positifs ou négatifs et εt est un aléa gaussien.

L'équation [5] peut aussi s'écrire :

$(1 - \alpha_1 D - \alpha_2 D_2 - \ldots - \alpha q\, D\, q)\varepsilon t = yt$.

Dans ce processus, tout comme dans le modèle autorégressif AR, les aléas sont supposés être engendrés par un processus de type bruit blanc. Le modèle MA est interprété comme étant représentatif d'une série chronologique fluctuant autour de sa moyenne de manière aléatoire, d'où le terme de moyenne mobile car celle-ci, en lissant la série, gomme le bruit créé par l'aléa.

Il est à noter qu'il y a équivalence entre un processus MA(1) et un processus AR d'ordre p infini : MA(1) = AR(∞).

Le corrélogramme simple d'un processus MA(q) est de la forme générale :

$\rho_k = \frac{\sum_{i=0}^{i=q-k} \alpha_i \alpha_{i+k}}{\sum_{i=0}^{i=q} \alpha_i^2}$ pour k = 0,1,. . . ,q et $\rho_k = 0$ pour k > q.

C'est-à-dire que seuls les q premiers termes du corrélogramme simple sont significativement différents de 0.

Le corrélogramme partiel est caractérisé par une décroissance géométrique des retards.

Exemple 1 (suite) : le processus est –il un MA(1), MA(2)

Dependent Variable: D(LTCH)
Method: ARMA Maximum Likelihood (OPG - BHHH)
Date: 12/15/22 Time: 12:32
Sample: 2007M02 2020M09
Included observations: 164
Convergence achieved after 10 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.003608	0.001879	1.920209	0.0566
MA(1)	0.389172	0.069681	5.585096	0.0000
SIGMASQ	0.000176	1.14E-05	15.42916	0.0000

R-squared	0.162133	Mean dependent var	0.003605
Adjusted R-squared	0.151725	S.D. dependent var	0.014539
S.E. of regression	0.013391	Akaike info criterion	-5.769361
Sum squared resid	0.028870	Schwarz criterion	-5.712657
Log likelihood	476.0876	Hannan-Quinn criter.	-5.746341
F-statistic	15.57729	Durbin-Watson stat	1.890054
Prob(F-statistic)	0.000001		

Inverted MA Roots	-.39

Le coefficient moyenne mobile d'ordre 1 est significativement différent de zéro dans la mesure où le t de Student **5.58** est supérieur à la valeur critique 1.96 .

Dependent Variable: D(LTCH)
Method: ARMA Maximum Likelihood (OPG - BHHH)
Date: 12/15/22 Time: 12:34
Sample: 2007M02 2020M09
Included observations: 164
Convergence achieved after 8 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.003595	0.001861	1.931510	0.0552
MA(2)	0.201727	0.085588	2.356969	0.0196
SIGMASQ	0.000202	1.24E-05	16.30369	0.0000

R-squared	0.037949	Mean dependent var	0.003605
Adjusted R-squared	0.025998	S.D. dependent var	0.014539
S.E. of regression	0.014349	Akaike info criterion	-5.631648
Sum squared resid	0.033149	Schwarz criterion	-5.574943
Log likelihood	464.7952	Hannan-Quinn criter.	-5.608628
F-statistic	3.175405	Durbin-Watson stat	1.303176
Prob(F-statistic)	0.044406		

Inverted MA Roots	-.00+.45i	-.00-.45i

3) Modèle ARMA (mélange de processus AR et MA)

Les modèles ARMA sont représentatifs d'un processus généré par une combinaison des valeurs passées et des erreurs passées. Ils sont définis par l'équation :

$$ARMA_{(p,q)} = \left(1 - \theta_1 D - \theta_2 D^2 - \cdots \theta_p D^p\right) y_t$$
$$= \left(1 - \alpha_1 D - \alpha_2 D^2 - \ldots \alpha_p D^p\right) \varepsilon_t$$

Nous avons :

ARMA(1,0) = AR(1); ARMA(0,1) = MA(1).

Dans le cas d'un processus ARMA (p, q) avec constante :

$$y_t = \mu + \theta_1 x_{t-1} + \theta_2 x_{t-2} + \cdots - \cdots \theta_p x_{t-P} + \varepsilon_t - \alpha_1 \varepsilon_{t-1} - \alpha_2 \varepsilon_{t-2} - \cdots . - \alpha_q \varepsilon_{t-q}$$

L'espérance du processus est donnée par :

$$E(x_t) = \frac{\mu}{1 - \theta_1 - \theta_2 - \cdots \theta_p}$$

Donc connaissant l'espérance du processus, la constante du processus ARMA est déterminée par :$\mu = E(x_t)\ (1 - \theta_1 - \theta_2 - \cdots \theta_p)$

Les corrélogrammes simples et partiels sont, par voie de conséquence, un mélange des deux corrélogrammes des processus AR et MA purs. Il s'avère ainsi plus délicat d'identifier ces processus à partir de l'étude des fonctions d'autocorrélation empiriques.

Exemple 1 (suite) : Selon les resulats consignés dans le tableau suivant, le processus est –il un ARMA(1,1)

Dependent Variable: D(LTCH)
Method: ARMA Maximum Likelihood (OPG - BHHH)
Date: 12/15/22 Time: 12:33
Sample: 2007M02 2020M09
Included observations: 164
Convergence achieved after 14 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.003603	0.002461	1.463875	0.1452
AR(1)	0.437886	0.173307	2.526652	0.0125
MA(1)	-0.014784	0.184349	-0.080195	0.9362
SIGMASQ	0.000172	1.05E-05	16.30125	0.0000

R-squared	0.183250	Mean dependent var	0.003605
Adjusted R-squared	0.167936	S.D. dependent var	0.014539

S.E. of regression	0.013262	Akaike info criterion	-5.782474
Sum squared resid	0.028142	Schwarz criterion	-5.706868
Log likelihood	478.1629	Hannan-Quinn criter.	-5.751781
F-statistic	11.96613	Durbin-Watson stat	1.992183
Prob(F-statistic)	0.000000		
Inverted AR Roots	.44		
Inverted MA Roots	.01		

Le coefficient moyen mobile d'ordre 1 est significativement nul. Donc le processus D(LTCH) n'est pas générée par un ARMA(1,1).

III- La méthode de Box et 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, est constituée d'une combinaison linéaire finie en *t* des valeurs passées d'un bruit blanc. Wold (1954) montre que les modèles ARMA permettent de représenter la plupart des processus stationnaires. L'approche de Box et Jenkins (1976) consiste en une méthodologie d'étude systématique des séries chronologiques à partir de leurs caractéristiques afin de déterminer, dans la famille des modèles ARIMA, le plus adapté à représenter le phénomène étudié. Trois étapes principales sont définies.

A. Recherche de la représentation adéquate : **l'identification.**

La phase d'identification consiste à déterminer le modèle adéquat dans la famille des modèles ARIMA. Elle est fondée sur l'étude des corrélogrammes simple et partiel.

1) Désaisonnalisation.

Dans le cas d'une série affectée d'un mouvement saisonnier, il convient de la retirer préalablement à tout traitement statistique. Cette saisonnalité est ajoutée à la série prévue à la fin du traitement afin d'obtenir une prévision en terme brut.

2) Recherche de la stationnarité en terme de tendance.

Si l'étude du corrélogramme simple et les tests statistiques s'y rapportant (statistique Q) présagent d'une série affectée d'une tendance, il convient d'en étudier les caractéristiques selon les tests de Dickey-Fuller. La méthode d'élimination de la tendance est fonction du processus DS ou TS sous-jacent à la chronique étudiée.

Après stationnarisation, nous pouvons identifier les valeurs des paramètres p, q du modèle ARMA.

- Si le corrélogramme simple n'a que ses q premiers termes (q = 3 maximum) différents de 0 et que les termes du corrélogramme partiel diminuent lentement, nous pouvons pronostiquer un MA(q) .
- Si le corrélogramme partiel n'a que ses p premiers termes (p = 3 maximum) différents de 0 et que les termes du corrélogramme simple diminuent lentement, cela caractérise un AR(p) .
- Si les fonctions d'autocorrélation simple et partiel ne paraissent pas tronquées, il s'agit alors d'un processus de type ARMA, dont les paramètres dépendent de la forme particulière des corrélogrammes

B. Estimation des paramètres

Les méthodes d'estimation diffèrent selon le type de processus diagnostiqué. Dans le cas d'un modèle AR(p), une méthode des moindres carrés est appliquée ou bien les relations existantes entre les autocorrélations et les coefficients du modèle (équations de Yule Walker) seront utilisées.

L'estimation des paramètres d'un modèle MA(q) s'avère plus complexe. Box et Jenkins suggèrent d'utiliser une procédure itérative de type balayage.

C. Tests d'adéquation du modèle et prévision

Les paramètres du modèle étant estimés, les résultats d'estimation suivant seront examinés.

• Les coefficients du modèle doivent être significativement différents de 0 (le test du t de Student s'applique de manière classique). Si un coefficient n'est pas significativement différent de 0, il convient d'envisager une nouvelle spécification éliminant l'ordre du modèle AR ou MA non valide.

• L'analyse des résidus : si les résidus obéissent à un bruit blanc, il ne doit pas exister d'autocorrélation dans la série et les résidus doivent être homoscédastiques.

Les tests suivants peuvent être utilisés.

- le test de Durbin Watson, bien qu'il ne permette de détecter que des autocorrélations d'ordre 1 ;
- les tests de Box et Pierce et de Ljung et Box permettent de tester l'ensemble des termes de la fonction d'autocorrélation. Si le résidu est à mémoire, cela signifie que la spécification du modèle est incomplète et qu'il convient d'ajouter au moins un ordre au processus ;

- le test ARCH d'hétéroscédasticité effectué à partir de la fonction d'autocorrélation du résidu au carré.

La phase de validation du modèle est très importante et nécessite le plus souvent un retour à la phase d'identification.

Lorsque le modèle est validé, la prévision peut alors être calculée à un horizon de quelques périodes, limitées car la variance de l'erreur de prévision croît très vite avec l'horizon.

La méthodologie de Box & Jenkins vise à formuler un modèle permettant de représenter une chronique avec comme finalité de prévoir des valeurs futures. De ce fait, l'objet de cette méthodologie est de modéliser une série temporelle en fonction de ses valeurs passées et présentes afin de déterminer le processus ARIMA adéquat par principe de parcimonie. Cette méthodologie suggère une procédure à trois étapes : Identification du modèle ; Estimation du modèle ; et Validation du modèle (Test de diagnostique).

L'application de cette méthode de Box & Jenkins sur des series chronologique est illustrée dans le Abderrahmani (2018)[30]

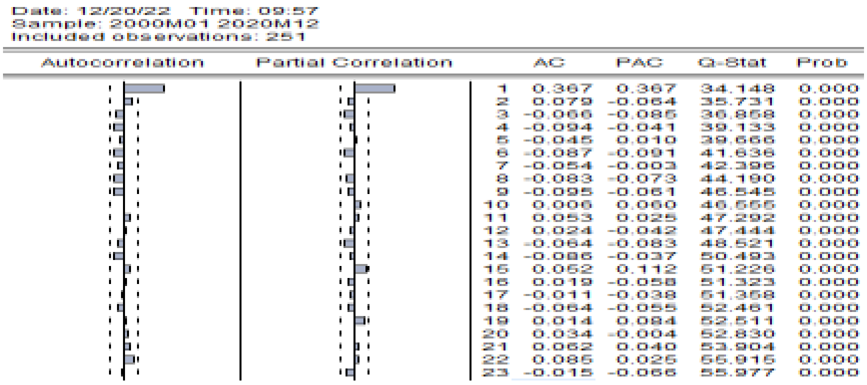
[30] ABDERRAHMANI F, (2018) « Guide pratique des séries temporelles macro-économiques et financières avec eviews 9.5 », *polycopié de cours à caractère pédagogique*, université de Bejaia.

Série de TD N°4

Exercice 01 :

On vous donne les informations consignées dans les figures et tableaux ci après :

Figure 1 : Le correlograme de la série OILP stationarisé (en première différence)



Date: 12/20/22 Time: 09:57
Sample: 2000M01 2020M12
Included observations: 251

	AC	PAC	Q-Stat	Prob
1	0.367	0.367	34.148	0.000
2	0.079	-0.064	35.731	0.000
3	-0.066	-0.085	36.858	0.000
4	-0.094	-0.041	39.133	0.000
5	-0.045	0.010	39.666	0.000
6	-0.087	-0.091	41.636	0.000
7	-0.054	-0.003	42.396	0.000
8	-0.083	-0.073	44.190	0.000
9	-0.095	-0.061	46.545	0.000
10	0.006	0.060	46.555	0.000
11	0.053	0.025	47.292	0.000
12	0.024	-0.042	47.444	0.000
13	-0.064	-0.083	48.521	0.000
14	-0.086	-0.037	50.493	0.000
15	0.052	0.112	51.226	0.000
16	0.019	-0.058	51.323	0.000
17	-0.011	-0.038	51.358	0.000
18	-0.064	-0.055	52.461	0.000
19	0.014	0.084	52.511	0.000
20	0.034	-0.004	52.830	0.000
21	0.062	0.040	53.904	0.000
22	0.085	0.025	55.915	0.000
23	-0.015	-0.066	55.977	0.000

Tableau 01 : Les résultats de la régression de D(OILP) sur AR(1)

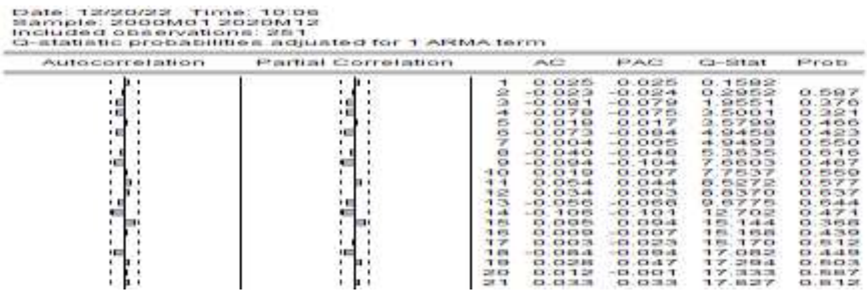
Dependent Variable: D(OILP)
Method: ARMA Maximum Likelihood (OPG - BHHH)
Date: 12/15/22 Time: 12:23
Sample: 2000M02 2020M12
Included observations: 251
Convergence achieved after 18 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.119327	0.612170	0.194925	0.8456
AR(1)	0.367580	0.042051	8.741362	0.0000
SIGMASQ	30.01016	2.053149	14.61665	0.0000

R-squared	0.135312	Mean dependent var	0.097649
Adjusted R-squared	0.128339	S.D. dependent var	5.902981
S.E. of regression	5.511187	Akaike info criterion	6.263896
Sum squared resid	7532.550	Schwarz criterion	6.306033
Log likelihood	-783.1189	Hannan-Quinn criter.	6.280853
F-statistic	19.40437	Durbin-Watson stat	1.944302
Prob(F-statistic)	0.000000		

Inverted AR Roots	.37

La figure 2 : Le correlograme des residus de l’estimation du processus AR(1)



Date: 12/20/22 Time: 10:06
Sample: 2000M01 2020M12
Included observations: 251
Q-statistic probabilities adjusted for 1 ARMA term

	AC	PAC	Q-Stat	Prob
1	0.025	0.025	0.1582	
2	-0.023	-0.024	0.2952	0.587
3	-0.081	-0.079	1.9551	0.376
4	-0.078	-0.075	3.5001	0.321
5	0.019	0.017	3.5799	0.466
6	-0.073	-0.084	4.9458	0.423
7	0.004	-0.005	4.9493	0.550
8	-0.040	-0.048	5.3635	0.616
9	-0.094	-0.104	7.6603	0.467
10	0.019	0.007	7.7537	0.559
11	0.054	0.044	8.5272	0.577
12	0.034	0.003	8.8370	0.637
13	-0.056	-0.066	9.6775	0.644
14	-0.106	-0.101	12.702	0.471
15	0.085	0.094	15.144	0.368
16	0.009	-0.007	15.168	0.439
17	0.003	-0.023	15.170	0.512
18	-0.084	-0.084	17.082	0.449
19	0.028	0.047	17.284	0.503
20	0.012	-0.001	17.333	0.567
21	0.033	0.033	17.827	0.612

Tableau 02 : Les résultats de la régression de D(OILP) sur MA(1)

Dependent Variable: D(OILP)
Method: ARMA Maximum Likelihood (OPG - BHHH)
Date: 12/15/22 Time: 12:24
Sample: 2000M02 2020M12
Included observations: 251
Convergence achieved after 7 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.108747	0.528278	0.205852	0.8371
MA(1)	0.347391	0.054531	6.370497	0.0000
SIGMASQ	30.26075	2.071012	14.61158	0.0000

R-squared	0.128092	Mean dependent var	0.097649
Adjusted R-squared	0.121060	S.D. dependent var	5.902981
S.E. of regression	5.534149	Akaike info criterion	6.272145
Sum squared resid	7595.449	Schwarz criterion	6.314282
Log likelihood	-784.1542	Hannan-Quinn criter.	6.289102
F-statistic	18.21683	Durbin-Watson stat	1.916854
Prob(F-statistic)	0.000000		

Inverted MA Roots	-.35

Tableau 3 : Les résultats de la régression de D(OILP) sur AR(1) et MA (1)

Dependent Variable: D(OILP)
Method: ARMA Maximum Likelihood (OPG - BHHH)
Date: 12/20/22 Time: 10:03
Sample: 2000M02 2020M12
Included observations: 251
Convergence achieved after 25 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.116527	0.590111	0.197466	0.8436
AR(1)	0.264636	0.106653	2.481275	0.0138
MA(1)	0.120587	0.132302	0.911452	0.3629
SIGMASQ	29.92371	2.062107	14.51123	0.0000

R-squared	0.137803	Mean dependent var	0.097649
Adjusted R-squared	0.127331	S.D. dependent var	5.902981
S.E. of regression	5.514372	Akaike info criterion	6.268998
Sum squared resid	7510.850	Schwarz criterion	6.325181
Log likelihood	-782.7593	Hannan-Quinn criter.	6.291608
F-statistic	13.15919	Durbin-Watson stat	1.986549
Prob(F-statistic)	0.000000		

Inverted AR Roots	.26
Inverted MA Roots	-.12

Travail à faire :

1) Le processus est-il un AR(1) ?

2) Le processus est-il un MA(1) ?

3) Le processus est-il un ARMA(1) ?

Exercice 02 :

L'application du test ADF, sous eviews, sur le PIB donne les principaux résultats suivants:

Application du Modèle 3 sur la série PIB

ADF Test Statistic	-1.700090	1% Critical Value*	-4.1630
		5% Critical Value	-3.5066
		10% Critical Value	-3.1828

*MacKinnon critical values for rejection of hypothesis of a unit root.
Augmented Dickey-Fuller Test Equation
DependentVariable: D(PIB)
Method: Least Squares
Date: 05/29/22 Time: 12:25
Sample(adjusted): 1972 2018
Included observations: 47 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
PIB(-1)	-0.115598	0.067995	-1.700090	0.0963
D(PIB(-1))	0.116027	0.150930	0.768749	0.4462
C	-86605630	4.35E+09	-0.019896	0.9842
@TREND(1970)	4.89E+08	2.96E+08	1.654019	0.1054

Application du Modèle 2 sur la série PIB

ADF Test Statistic	-0.520800	1% Critical Value*	-3.5745
		5% Critical Value	-2.9241
		10% Critical Value	-2.5997

*MacKinnon critical values for rejection of hypothesis of a unit root.
Augmented Dickey-Fuller Test Equation
DependentVariable: D(PIB)
Method: Least Squares
Date: 05/29/22 Time: 12:29
Sample(adjusted): 1972 2018
Included observations: 47 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
PIB(-1)	-0.017863	0.034299	-0.520800	0.6051
D(PIB(-1))	0.075872	0.151874	0.499575	0.6199
C	4.71E+09	3.31E+09	1.422621	0.1619
R-squared	0.010119	Meandependent var		3.59E+09

Travail à faire :
Apres avoir expliqué (brièvement) le principe du test ADF,commenter les résultats de l'application du modèle 3 et 2 du test ADF sur la série(PIB)

Eléments de réponses

Corrigé de l'exercice 1 :

Le test de significativité pour la variable AR(1) permet de déterminer si la variable explicative est pertinente, c'est à dire qu'il s'agit bien d'un processus AR(1).

On pose les hypothèses suivantes : H0 : a = 0 H1 : a ≠ 0.

Il faut comparer le t* à avec le t théorique lue dans la table. Si t* > t alors on rejette H0, dans ce cas la variable explicative est significativement différente de 0, donc la variable explicative contribue significativement à expliquer la variable D(OILP). On dit donc qu'elle est significative. Le logiciel EVIEWS nous donne la probabilité critique du t-Statistic (t*). Elle peut être trouvée dans la table en cherchant la probabilité correspondant

à la valeur du t* pour un degré de liberté de (T - K). Les résultats consignés dans le tableau N° 1 montrent que la probabilité critique est inferieur au seuil $\alpha = 5\%$, on rejette H0 et le coefficient associé à la variable AR(1) est significativement différent de 0.

D'apres la figure 2, tous les termes des fonctions d'autocorrélation simple et partielle sont tous situés dans l'intervalle de confiance matérialisé par les traits verticaux. L'estimation du modèle AR(1) est donc validée, la série peut être valablement représentée par un processus de type AR(1) sur la série différenciée.

Le test de significativité pour la variable MA(1) permet de déterminer si la variable explicative est pertinente, c'est à dire qu'il s'agit bien d'un processus AM(1). On pose les hypothèses suivantes : H0 : a = 0 contre H1 : $a \neq 0$. Nous obtenons les résultats du tableau 3. D'après le tableau 3, on constate que la probabilité critique du t-Statistic (t*) associé à la variable MA(1) est inferieur au seuil $\alpha = 5\%$, on rejette donc H0. La variable MA(1) contribue à expliquer de manière significative la variable D(OILP), on peut en déduire qu'elle suit un processus MA(1). Cependant, On note que la constante n'est pas significative car sa probabilité critique associée est elle aussi supérieure au seuil de 5 %.

Le test de significativité pour la variable AR(1) et AM(1) en même temps permet de déterminer si les variables explicatives sont pertinentes, c'est à dire qu'il s'agit bien d'un processus ARMA(1,1). On constate que les probabilités critiques des t-Statistic (t*) associés aux variables MA(1) et à la constante sont toutes supérieures au seuil $\alpha = 5\%$. Par conséquent , le processus D(OILP) n'est pas un processus ARMA(1,1).

Corrigé de l'exercice 2 :

Les résultats du Modèle 3 appliqué sur la série PIB

Test du trend:

$$\begin{cases} H_0 : B=0 \\ H_1 : B\neq 0 \end{cases}$$

$Tb = |1.65| < T^{ADF} = 2.78$ donc on accepte $H_0 : B=0$, la tendance est non significative.

On passe à l'estimation du modèle 02

Les résultats du Modèle 2 appliqué sur la série PIB

Test de la constante

$$\begin{cases} H_0 : C=0 \\ H_1 : C\neq 0 \end{cases}$$

$Tc = |0.17| < T^{ADF} = 2.52$ donc on accepte $H_0 : C= 0$, la constante est non significative. On passe à l'estimation du premier modè

Chapitre 5 : Modélisation VAR et conitegtarion

Les modèles Vector Auto Regressive (VAR) ont été introduits par Sims (1980) « *comme réponse à la critique des méthodes d'identification généralement utilisées dans les modèles économétriques traditionnels* » (Hairault (1995), p.89). La modélisation consistera dans le cadre de cette analyse à modéliser les interactions existantes entre les variables stationnaires à partir de leur passé et de celui des autres variables. Autrement dit, elle « permet de résumer les corrélations entre les variables sans préjuger de la forme des liens entre celles-ci » (Garcia et Verdelhan (2001)[31].

Ces modèles sont utilisés pour analyser l'efficacité et la dynamique générale des variables. Ils fournissent une méthode commode d'analyse de l'impact d'une variable donnée sur elle-même et sur les autres variables en utilisant des instruments d'analyse portant sur les tests de causalité, la décomposition de la variance de l'erreur de prévision et les réponses impulsionnelles. Ils permettent également d'analyser les interactions des variables entre elles en faisant abstraction aux contraintes liées à la structure théorique.

I- La modélisation VAR

La modélisation VAR est nécessaire dans une analyse économétrique, car elle exploite sans contrainte tous les liens de causalité entre les déterminants d'un phénomène[32].

1.1 **Présentation du modèle VAR**

Un groupe de variables aléatoires temporelles est généré par un modèle VAR si chacune de ses variables est une fonction linéaire de ses propres valeurs passées et des valeurs passées des autres variables du groupe, à laquelle s'ajoute un choc aléatoire de type bruit blanc[33]. Ce modèle comporte trois avantages :

- Il permet d'expliquer une variable par rapport à ses retards et en fonction de l'information contenue dans d'autres variables pertinentes.
- On dispose d'un espace d'information très large.
- Cette méthode est assez simple à mettre en œuvre, et comprend des procédures d'estimation et des tests.

[31] William Green. Econométrie. Paris: Pearson Education ,5e Edition,2006.

[32] Véronique MEURIOT, « Réflexions méthodologiques sur la modélisation non structurelle », Montpelier, 2008, P51.

[33] Eric DOR, « économétrie », Op.cit, P220.

La construction du modèle VAR se fait d'abord par la sélection des variables d'intérêts en se référant évidemment à la théorie économique, ensuite par le choix de l'ordre des retards des variables et en enfin par l'estimation des paramètres.

La représentation du modèle VAR à (k) variables et (p) décalages noté VAR(p) s'écrit :

$$Y_t = \Phi_o + \Phi_1 Y_{t-1} + \Phi_2 Y_{t-2+\ldots} \Phi_p Y_{t-p} + \mathcal{E}_t$$

Avec : Φ_o vecteur du terme constant ;

$\Phi_1, \Phi_2 \ldots, \Phi_p$: sont des matrices.

1.2 Estimation et détermination du nombre de retards (p)

Les paramètres du modèle VAR ne peuvent être estimés que sur des séries temporelles stationnaires. Deux techniques d'estimation sont possibles :

- Estimation de chaque équation du modèle VAR par la méthode des Moindres Carrés Ordinaires (MCO).
- Estimation par la méthode du maximum de vraisemblance.

L'estimation d'un modèle VAR nécessite le choix du nombre de retard (p), la sélection de l'ordre des retards détermine la période maximum d'influence des variables explicatives sur la série à expliquée. Lorsque la valeur de p du nombre de retards du modèle VAR(p) est inconnue, il existe des critères statistiques pour la définir, il s'agit du critère d'AKAIKE et du critère de SCHWARZ. La procédure consiste à définir un ordre jugé suffisamment bas P_{min} (généralement est égal à 1) et ensuite à tester successivement si on peut admettre l'ordre immédiatement supérieur. On s'arrête au retard P* pour lequel la valeur de l'une des statistiques AKAIKE ou de SCHWARZ est minimisé.

$$AIC(p) = Ln\ [det|\mathcal{E}_\varepsilon|] + \frac{2k^2p}{n}$$

$$SC(p) = Ln\ [det|\mathcal{E}_\varepsilon|] + \frac{k^2p\ Ln(n)}{n}$$

Où :

det : déterminant de la matrice variance-covariance ;

k : est le nombre de variables du système;

n : le nombre d'observations ;

$\mathcal{E}_\varepsilon$: matrice variance covariance des erreurs ;

Ln : logarithme népérien.

1.3 Les applications du modèle VAR

Au niveau théorique, la mise en évidence de relation causale entre les variables économiques fournit des éléments de réflexion propices à une meilleure compréhension des phénomènes économiques.

- **La causalité au sens de GRANGER**

La causalité consiste à étudier l'évolution de l'ensemble des variables, et d'examiner si le passé des unes apporte une information supplémentaire sur la valeur présente et future des autres. Cette approche est formalisée comme suit :

Soit un processus VAR(1) pour 2 variables Y_{1t}, Y_{2t} :

$$\begin{cases} \boldsymbol{Y_{1t} = \beta_0 + \beta_1 Y_{1t-1} + \beta_2 Y_{2t-1} + \mathcal{E}_{1t}} \\ \boldsymbol{Y_{2t} = \alpha_0 + \alpha_1 Y_{1t-1} + \alpha_2 Y_{2t-1} + \varepsilon_{2t}} \end{cases}$$

Le test consiste à poser ces deux hypothèses :

- Y_{2t} ne cause pas Y_{1t}, si l'hypothèse H_0 suivante est acceptée :

  $\mathbf{H_0 : \beta_1 = \beta_2 = 0}$

- Y_{1t} ne cause pas Y_{2t}, si l'hypothèse H_0 suivante est acceptée :

  $\mathbf{H_0 : \alpha_1 = \alpha_2 = 0}$

On test ces deux hypothèses à l'aide d'un test de Fisher classique de nullité des coefficients. La statistique du test est notée : $\mathbf{F^* = \frac{SCR_c - SCR_{nc}\,/c}{\frac{SCR_{nc}}{n-k-1}}}$

Avec :

c : le nombre de coefficient dont on teste la nullité ;

SCR_c : somme des carrés des résidus du modèle contraint ;

SCR_{nc} : somme des carrés des résidus du modèle non-contraint.

- **La règle de décision :**

Si F^* > à la valeur de la table => on rejette H_0.

- **L'analyse des chocs**

Elle mesure l'impact de la variation d'une innovation sur les valeurs actuelles et futures des variables endogènes. Un choc sur la $i^{ème}$ variable peut avoir une conséquence immédiate sur cette même variable, et également sur les autres variables exogènes à travers la structure dynamique du modèle VAR. **Une fonction de réponse impulsionnelle** trace l'effet d'un choc ponctuel sur l'une des innovations sur les valeurs actuelles et futures des variables endogènes.

Les fonctions de réponse impulsionnelle reflètent la réaction dans le temps des variables aux chocs contemporains identifiés. Leur examen fournit des informations sur les conséquences dans le temps des chocs.

La simulation des chocs structurels est une méthode puissante pour l'analyse de la dynamique entre un groupe de variables. En identifiant un modèle VAR (1), l'analyse impulsionnelle permet d'expliquer les influences des chocs structurels d'une des variables sur les autres variables du système. Les réponses aux impulsions demeure l'un des instruments le mieux indiqué pour expliquer les sources d'impulsion. Elles reflètent la réaction dans le temps des variables aux chocs contemporains identifiés. Leur examen fournit des informations sur les conséquences dans le temps des chocs.

Fonctions de réponses impulsionnelles

Les figures qui suivent retracent les réponses à des chocs sur les résidus des variables étudiées. Les courbes en pointillés représentent l'intervalle de confiance. L'amplitude du choc est égale à l'écart type des erreurs de la variable et l'on s'intéresse aux effets du choc sur dix périodes. L'horizon temporel des réponses est fixé sur ces dix périodes et il représente le délai nécessaire pour que les variables retrouvent leurs niveaux de long terme.

Décomposition de la variance de l'erreur de prévision : Alors que les fonctions de réponse impulsionnelle retracent les effets d'un choc sur une variable endogène sur les autres variables du VAR, la décomposition de la variance sépare la variation d'une variable endogène dans les chocs constitutifs du VAR. Ainsi, la décomposition de la variance fournit des informations sur l'importance relative de chaque innovation aléatoire dans l'affectation des variables du VAR. L'objectif est de calculer la contribution de chacune des innovations à la variance de l'erreur en pourcentage. Quand une innovation explique une part importante de la variance de l'erreur de prévision, nous en déduisons que l'économie étudiée est très sensible aux chocs affectant cette série.

1.4 Validation du modèle VAR

- **Test d'autocorrélation des erreurs**

Il existe un grand nombre de tests d'autocorrélation, les plus connus sont ceux de Box et Pierce (1970) et Ljung et Box (1978). Nous n'étudierons ici que le test de Box et Pierce. Le test de Ljung et Box est à appliquer lorsque l'échantillon est de petite taille[34].

Soit une autocorrélation des erreurs d'ordre K (K>1) :

$\varepsilon_t = \rho_1 \varepsilon_{t-1} + \rho_2 \varepsilon_{t-2} + \cdots + \rho_k \varepsilon_{t-k} + \nu_t$ Où $\nu_t \sim N(0, \sigma_\nu^2)$

Les hypothèses du test de Box-Pierce sont les suivantes :

$\begin{cases} H_0: \rho_1 = \rho_2 = \cdots = \rho_k = 0. \\ H_1: \text{Il existe au moins un } \rho_i \text{ significativement différent de 0.} \end{cases}$

Pour effectuer ce test, on fait recours à la statistique *Q* qui est donnée par :

$$\varrho = n \sum_{k=1}^{k} \hat{\rho}_k^2$$

Où n est le nombre d'observations et $\hat{\rho}_k^2$ est le coefficient d'autocorrélations d'ordre k des résidus estimés e_t.
Sous l'hypothèse H_0 vraie, *Q* suit la loi du Khi-deux avec K degrés de liberté :

$$\varrho = n \sum_{k=1}^{k} \hat{\rho}_k^2 \sim > x^2(k)$$

La règle de décision est la suivante :

Si *Q* > k* où k* est la valeur donnée par la table du Khi-deux pour un risque fixé et un nombre K de degrés de liberté bien précis, on rejette H_0 et on accepte H_1 (autocorrélation des erreurs).
Il existe un grand nombre de test d'absence de corrélation, nous allons utiliser « l'autocorrélation LM test » qui fait l'objet de tester le caractère non autocorrélation des résidus. L'hypothèse nulle est qu'il y a absence d'autocorrélation contre l'hypothèse alternative d'existence d'autocorrélation.

- **Test d'héthéroscédasticité**

Il existe plusieurs tests possibles: test de Goldfeld et Quandt, test de White, test de Gleisjer et test ARCH. Nous n'étudierons ici que le test de White. Ce test est fondé sur une

[34] Bourbonnais Régis et Terraza Michel. Analyse des séries temporelles : Application à l'économie et à la gestion. Paris : Dunod, 2002.

relation significative entre le carré des résidus et une ou plusieurs variables explicatives en niveau et au carré au sein de la même équation de régression[35].

$$e_t{}^2 = \alpha_0 + \alpha_1 x_1 t + \alpha_2 x_1 t^2 + \alpha_3 x_2 t + \alpha_4 x_2 t^2 \ldots + \alpha_k x_k t + \alpha_k x_{k+1} t^2 + v_t$$

On considère le "R^2" le coefficient de détermination obtenu de cette régression et "n" le nombre d'observation. Si l'un de ses coefficients est significativement différent de zéro donc il y a un problème d'hétéroscédasticité. Nous pouvons également procéder à ce test à l'aide du test de Fisher de nullité des coefficients.

$H_0: \alpha_1 = \alpha_2 = \alpha_3 = \alpha_k = 0.$

Soit la statistique LM qui est distribuée comme un χ^2 à $p = 2 * k$ degré de liberté.

Si $n * R^2 > \chi^2{}_p$ lue dans la table de χ^2 donc on accepte H_1.

Si $n * R^2 < \chi^2{}_p$ lue dans la table de χ^2 donc on accepte H_0.

Un exemple d'application du modèle VAR portant sur le lien entre la diversification des exportations et la croissance économique est illustré dans Touati et Keddari(2021)[36].

II- La cointégration et modèles à correction d'erreurs

La cointégration désigne l'existence d'une réelle relation à long terme entre des variables intégrées. En effet, le risque d'estimer des relations fallacieuses[37] et d'interpréter les résultats de manière erronée est très élevé.

L'étude de la relation de long terme en utilisant les techniques de cointegration prend, depuis la fin des années 80's, une place particulière dans l'économétrie .Nous distinguons essentiellement deux grandes approches : la première approche est celle de Engel et Granger (1987) et Phillips et Ouliaris (1990), basée sur les résidus en deux étapes afin de tester l'hypothèse nulle de non cointegration , la seconde approche est celle de Johansen (1991-1995) qui décrit une régression de système basée sur un rang réduit ;cependant , le test de Johansen (1988)et de Johansen et Juselius(1990) s'avère le plus efficient car il a l'avantage d'identifier le nombre de vecteurs cointégrés entre les

[35] Lardic sandrine et Mignon Valérie. Econométrie des séries temporelles macroéconomiques et financières. Paris : Economica, 2002.

[36] Touati Karima, Keddari Nassim (2021), Impact de la diversification des exportations sur la croissance économique en Algérie : Modélisation VAR, in Ouvrage collectif national intitulé *La problématique de la diversification économique dans le cadre du développement durable en Algérie, entre opportunités et défis*, Université de Constantine, Algérie.

[37] Une estimation par MCO peut donner des résultats qui font croire faussement qu'une relation de long terme existe (R^2 > Durbin Watson).

variables non stationnaires en niveau dans le cadre d'un VECM (Vectoriel Error Correction Model)[38].

Les conditions de cointégrations sont :

- Il faut que les séries soient intégrées de même ordre ;
- la combinaison linéaire de ces deux séries permet de se ramener à une série d'ordre d'intégration inférieur.

2.1 L'approche d'Engle et Granger (1987)

Le test d'Engle et Granger est une méthode de vérification de l'existence d'une relation de cointégration entre deux variables intégrées et d'estimation de cette relation. Cette méthode est valable sous l'hypothèse arbitraire qu'il existe un seul vecteur de cointégration entre les variables utilisées.[39] La généralisation de deux à k variables s'avère assez complexe du fait du nombre de vecteurs de cointégrations possibles.

La méthode d'Engle et Granger nous permet d'estimer d'estimer facilement un MCE en deux étapes, elle fournit également un certain nombre de tests de cointégration faciles à mettre en œuvre. L'inconvenants de cette approche c'est qu'elle ne permet pas de distinguer plusieurs vecteurs de cointégration.

2.2 Approche multivariée de cointégration de Johansen (2001)

Les tests de Johansen permettent de vérifier des hypothèses sur le nombre de vecteurs de cointégration dans un système VAR(p) reliant des variables qui sont toutes intégrées du même ordre[40]. Ainsi, si on analyse un comportement de N variables, on peut avoir jusqu'à N-1 relations de cointégrations.

2. 3 Estimation d'un modèle VECM

Le point de départ d'un modèle VECM est un modèle VAR(P). On peut réécrire le modèle VAR(2) sous la forme d'un VECM comme suit :

$$X_t = A_1 X_{t-1} + A_2 X_{t-2} + \varepsilon_t \Longrightarrow \text{VAR}(2)$$

$$\Delta X_t = \beta \Delta X_{t-1} + \pi X_{t-1} + \varepsilon_t \Longrightarrow \text{VECM}$$

Avec:

$\pi = A_1 + A_2 - I$;
$\beta = -A_2$;
I : l'identité de X_{t-1}

[38] Abdallah Ali.Taux de change et performances économiques dans las pays en développement : l'exemple du Maghreb. Thèse de Doctorat. Université Val de Marne, Paris XII. 2006.
[39] Eric DOR, op.cit, p 215.
[40] Eric DOR, op.cit, p 226.

Le test de cointégration est fondé sur le rang de la matrice. Le rang de la matrice détermine le nombre de relations de cointégration (relations de long terme). Johansen propose un test fondé sur les vecteurs propres correspondant aux valeurs propres maximales de la matrice.

A partir des valeurs propres de la matrice on calcule une statistique notée :

$$\lambda_{trace} = -n \sum_{i=r+1}^{\infty k} Ln\,(1 - \lambda i)$$

Avec : λ_i : la $i^{ème}$ valeur propre de la matrice (π) ;

k : le nombre de variables ;

r : le rang de la matrice (π) ;

n : nombre d'observations.

Cette statistique suit une loi de Khi-deux tabulée par Johansen. Le test fonctionne de la manière suivante :

- Le rang de la matrice $\pi = 0$: r = 0

On teste les deux hypothèses suivantes : $\begin{cases} H_0 : r = 0 \\ H_1 : r > 0 \end{cases}$

Si H_0 est refusée, on passe au test suivant (r = 1).

Règle de décision :

Si $\lambda_{trace} >$ à la valeur critique de la table de Johansen ==> on rejette H_0.

Si H_0 est acceptée, on ne peut estimer un modèle VECM.

Le rang de la matrice $\pi = 1$: r = 1 : $\begin{cases} H_0 : r = 1 \\ H_1 : r > 1 \end{cases}$

Si H_0 est refusée, on passe au test suivant (r = 2), et ainsi de suite.

La procédure s'arrête à r = k-1.

Un exemple d'application du modèle VECM portant sur les déterminants des taux d'intérêts débiteurs en Algerie est illustré dans Moussi et Touati (2021)[41].

III- Le modèle ARDL

3.1 Définitions

Le modèle ARDL[42] permet d'identifier et d'analyser la relation de long-terme et de court-terme entre les variables explicatives et la variable à expliquer sur des séries qui ne

[41] Moussi Froudja, Touati Karima, (2021), Etude économétrique des déterminants du taux d'intérêt débiteur en Algérie", Revue algérienne d'économie et gestion, Volume 15, Numéro 2, Pages 865-883, Disponible sur https://www.asjp.cerist.dz/en/downArticle/154/15/2/176718

[42] Bouznit, Mohammed. « Rendement du capital humain et dynamique de la croissance au sein des pays sous développées» thèse de doctorat, ENSSEA, 2016, p 73-75.

sont pas intégrées de même ordre et, d'autre part d'obtenir des meilleures estimations sur des échantillons de petite taille.

L'avantage du modèle ARDL se manifeste dans sa flexibilité, car ce dernier peut être appliqué même sur les variables qui ne sont pas intégrées de même ordre, mais il suffit de s'assurer qu'aucune des variables n'est intégrée d'ordre deux et plus. En outre, les estimateurs obtenus du modèle ARDL sont robustes et sans biais même pour le cas d'un échantillon de taille faible (Harris et Sollis, 2003). De ce fait, le modèle ARDL[43] mettant en relation la variable à expliquer ($y_{(t)}$) , et les variables explicatives ($x_{(t)}$) peut s'écrire de la façon suivante : $y_{(t)} = \alpha + \beta x_{(t)} + u_{(t)}$

La procédure ARDL à long terme implique deux étapes. A la première étape, on teste l'existence d'une relation de long terme. La présence de la relation à long terme entre les variables est testée en calculant les F-statistiques pour tester la signification des niveaux décales des variables sous la forme de correction d'erreur du modèle ARDL sous-jacent. Le modèle à correction d'erreur est le suivant :

$$Dy(t) = \alpha_O + \sum_{i=1}^{p} \delta_i Dy_{t-i} + \sum_{i=1}^{p} y_i Dx_{t-i} + \beta_1 y_{t-1} + \beta_2 x_{t-1} + s_t$$

Où ð et y représentent la dynamique à court terme du modèle tandis que β_1 et β_2 représentent la relation de long terme et ε est le terme d'erreur du bruit blanc. Les valeurs actuelles de Dx, de l'équation sont exclues en suivant le modèle de Pesaran et Shin (1998). L'hypothèse nulle du test F est la non-existence de la relation de cointégration:

$$\begin{cases} H_O: \beta_1 = \beta_2 = 0 \\ H_1: \beta_1 \neq \beta_2 \neq 0 \end{cases}$$

Les statistiques pertinentes sont les statistiques F pour la signification conjointe de β_1 et β_2, et la distribution asymptotique de F est non-standard, et calculé indépendamment de l'ordre d'intégration des variables explicatives. Peseran et al (1996) ont calculé les valeurs critiques appropriées; en conséquence, il existe deux ensembles de valeurs critiques. Un ensemble en supposant que toutes les variables sont I (0) et une en supposant que toutes les variables sont I(1).

1. Si la valeur de la F-stat dépasse la borne supérieure, alors on rejette H0 et on

[43] Jaouad OBAD1 and Youssef JAMAL, « L'impact des dépenses publiques sur la croissance économique au Maroc : Application de l'approche ARDL », ISSN 2028-9324 Vol. 16 No. 2 Jun. 2016, pp. 444-455, pdf, page 3 à 6. http://www.ijias.issr-journals.org/

conclut à l'existence d'une relation de long terme entre les variables considérées.

2. Si la valeur de la F-stat est inférieure à la borne inférieure, alors on ne rejette pas H0 et on conclut à l'absence de relation de long terme entre les variables considérées.

3. Si la valeur de la F-stat est comprise entre les deux bornes, alors on ne peut pas conclure. Le résultat dépend du fait que les variables sont I (0) ou I (1). Une fois que les résultats des tests rejettent l'hypothèse nulle de la «non-existence de la relation de long terme », alors il est possible de procéder à la prochaine étape de la procédure ARDL d'estimation, qui est l'estimation des coefficients de long terme.

Dans la deuxième étape, on détermine les ordres des retards dans le modèle ARDL en utilisant le critère d'information Schwartz (SIC) et ensuite, le modèle choisi est estimé par la méthode des moindres carrés ordinaires pour obtenir une estimation de long terme. Cette estimation de long terme, de la spécification ARDL choisie, donne une estimation des coefficients de la relation de cointégration. Il est important de noter, cependant, que cette étape n'est viable que si les résultats des tests de F rejettent l'inexistence d'une relation de long terme entre les variables, donc la variable x peut être considérée comme la variable qui explique y à long terme.

3.2 La méthodologie du modèle ARDL

Les étapes à suivre pour l'analyse de la cointégration dans le modèle ARDL sont :

a) Sélectionner le nombre de retard optimal

Afin de choisir un retard optimal pour chaque variable, la méthode ARDL estime régressions, où (p) est le nombre maximal de retard et k est le nombre de variables dans l'équation. Le modèle peut être choisi sur la base du Schawrtz-Bayesian criteria (SBC) et du critère d'information d'Akaike (AIC). Le SBC permet de sélectionner un nombre plus réduit de retards alors que l'AIC permet de sélectionner le nombre maximum de retards. Après la sélection du modèle ARDL par l'AIC ou la SBC.

b) Tester de la stationnarité des séries temporelles

Afin de déterminer l'ordre d'intégration des séries temporelles et la stationnarité des séries étudiées, le test de stationnarité de Dickey Fuller Augmenté (ADF) est largement utilisé. En effet, afin d'utiliser l'approche du Bound-Test développé par Pesaran et al (2001), il faut s'assurer préalablement qu'aucune des séries n'est intégrée d'ordre 2 ou plus car les

valeurs critiques fournies par Pesaran et al. (2001) concernent uniquement les niveaux d'intégration 0 et 1.

c) Tester la cointégration par (Bounds-test) :

Le test de cointégration selon l'approche de Pesaran et al (2001) dans les modèles ARDL consiste à tester la nullité conjointe des coefficients des variables en niveau et retardées du modèle. En fait, l'hypothèse nulle du test de cointégration (Wald-test) s'écrit :

H0 : $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = \alpha_5 = 0$; (Pas de relation de cointégration).

H1 : au moins un des coefficients est significativement différent de zéro (présence cde relation de cointégration).

Si l'hypothèse nulle est rejetée, alors il y'a une relation de long terme entre les variables, sinon il n'y a aucune relation de long terme entre les variables. La statistique du test F-stat ou statistique de Wald suit une distribution non standard qui dépend du caractère non stationnaire des variables régressées, du nombre de variables dans le modèle ARDL, de la présence ou non d'une constante et d'une tendance ainsi que de la taille de l'échantillon. Deux valeurs critiques sont générées avec plusieurs cas et différents seuils : la première correspondant au cas où toutes les variables du modèle sont I (1) : CV-I (1) qui représente la borne supérieure ; la seconde correspond au cas où toutes les variables du modèles sont I (0) : CVI (0) qui est la borne inférieure. (D'où le nom de « Bound testing approach cointegration » ou « approche de test de cointégration par les bornes »). Alors la règle de décision pour le test de cointégration est la suivante :

- Si **F-stat > CV-I (1)**, alors l'hypothèse nulle est rejetée et donc il y'a Cointégration.
- Si par contre **F-stat < CV-I (0)**, alors l'hypothèse nulle de non cointégration est acceptée.
- Si la F-stat est incomprise entre les deux (2) valeurs critiques, rien ne peut être conclu.

Un exemple d'application du modèle ARDL portant sur le lien taux de change et prix de pétrole est illustré dans Touati (2021)[44].

[44] Touati Karima (2021), Les effets à court et long termes du prix de pétrole sur le taux de change en Algérie : Modèle ARDL sur données mensuelles (2012-2019), *Revue des Sciences Economiques, de Gestion et Sciences Commerciales,*

Série de TD N°5 et quelques éléments de reponses

Exercice 01

L'estimation du modèle VAR(1) sur les variables TCH et PIB (stationarisées en première différence) donne les résultats figurant dans le tableau 1.

Tableau 01 : Résultats de l'estimation VAR(1)

VectorAutoregressionEstimates
Date: 05/29/22 Time: 12:27
Sample(adjusted): 1972 2018
Includedobservations: 47 afteradjusting endpoints
Standard errors in ( ) & t-statistics in [ ]

	D(TCH)	D(PIB)
D(TCH(-1))	0.521827	-80805877
	(0.19529)	(6.1E+08)
	[2.67206]	[-0.13155]
D(PIB(-1))	6.60E-11	0.043549
	(6.8E-11)	(0.21290)
	[0.97497]	[0.20455]
C	0.970697	3.62E+09
	(0.89959)	(2.8E+09)
	[1.07904]	[1.27963]
R-squared	0.167019	0.004409

Apres avoir expliqué (brièvement) le principe de modélisation VAR, interpréter les résultats illustrés dans le tableau

La modélisation VAR consiste à modéliser les interactions existantes entre les variables stationnaires, à partir de leur passé et de celui des autres variables. Avant le traitement d'une série chronologique, il convient de s'assurer de la stationnarité des variables retenues.

L'interprétation : Chaque colonne du tableau correspond à une équation du VAR

Les résultats de l'estimation montrent qu'un grand nombre de coefficient associé à chaque variable sont non significatif d'un point de vue statistique, à l'exception du coefficient de D(IDE(-1)) dans l'équation du D(IDE) qui est significatf (t-sta |3.37| est superieur à 1,96 ; et le coefficient de D(TCH(-1)) dans l'équation du D(TCH). Et la constante. Donc l'IDE retardé d'une année affecte positivement D(IDE). De même, D(TCH) retardé d'une année a un effet positif sur D(TCH).

Exercice 2 :

L'application du test de Dickey-Fuller augmenté a été faite sur chacune des trois séries. Le tableau suivant est issu de la mise en œuvre de ce test sur le logiciel EVIEWS. Pour chacune des séries précisez le modèle retenu et donner votre conclusion au seuil de 5%.

Table N°1. Résultats du test de stationnarité (ADF)

Variables	Niveau						ADF Difference Test	
	t-statistics and tabulated value	Modèle 3 Constant and Trend		Modèle 2 Constant		Modèle 1 None	Modèle 1 None	Orderr d' intégration
		T de ADF	Ttrend	T de ADF	Tconst	T de ADF	T de ADF	
LPPT	t-statistics	-1.82	0.041	-2.01	1.99	-0.40	-7.39	**I (1)**
	tabulated value	-3.45	3.18	-2.93	2.89	-1.94	-1.94	
LTCH	t-statistics	-1.26	0.92	-1.22	1.32	2.37	-4.56	**I (1)**
	tabulated value	-3.45	3.18	-2.93	2.89	-1.94	-1.94	
LMM	t-statistics	-1.72	1.30	-2.04	2.11	4.28	-5.41	**I (1)**
	tabulated value	-3.45	3.18	-2.93	2.89	-1.94	-1.94	
LINT	t-statistics	-3.45	2.82	-2.50	-0.74	-2.40	/	**I(0)**
	tabulated value	-4.15	3.18	-2.93	2.89	-1.94	-1.94	

Source : TOUATI (2021), Les effets à court et long termes du prix de pétrole sur le taux de change en Algérie : Modèle ARDL sur données mensuelles (2012-2019), *Revue des Sciences Economiques, de Gestion et Sciences Commerciales,*

Travail à faire :

- Quelles (s) conclusions tirez-vous du tableau 1?
- Quelles (s) conclusions tirez-vous du tableau 2?
- Peut-on appliquer le test de cointégration de Johansen dans ces conditions ? justifier votre réponse
- Quel modèle peut-on alors appliqué

D'après les résultats de la stationnarité, les séries LPPT, LTCH, LMM sont intégrées d'ordre 1 (stationnaire après la première différence), alors que le taux d'intérêt LTIN reste stationnaire en niveau (sans différenciation).

L'observation des résultats d'estimation VAR (1) montre que tous les coefficients sont non significatifs, mais ce qui nous intéresse en fait dans cette estimation du modèle VAR (1) c'est d'exprimer le taux de change en fonction des autres variables du modèle. Les résultats indiquent que le taux de change dépend positivement de son taux passé, ce qui est expliqué par la tendance à la hausse du taux de change réel (dépréciation continue de la monnaie nationale). Cependant, le coefficient associé au prix de pétrole n'est pas significatif. Donc le lien entre ces deux variables n'est pas vérifié.

Tableau 2 : résultats de l'estimation du VAR(1)

Vector Autoregression Estimates
Date: 12/20/22 Time: 10:50
Sample (adjusted): 2012M03 2019M12
Included observations: 94 after adjustments
Standard errors in ( ) & t-statistics in [ ]

	D(LTCH)	D(LPPT)	D(LMM)	LINT
D(LTCH(-1))	0.316542	0.760769	-0.191075	-5.176477
	(0.09827)	(0.60636)	(0.08959)	(3.34014)
	[3.22119]	[1.25465]	[-2.13285]	[-1.54978]
D(LPPT(-1))	-0.026130	0.279620	0.005215	-0.411027
	(0.01705)	(0.10523)	(0.01555)	(0.57964)
	[-1.53228]	[2.65734]	[0.33545]	[-0.70911]
D(LMM(-1))	0.029467	0.171822	-0.072626	-0.481443
	(0.11354)	(0.70062)	(0.10351)	(3.85938)
	[0.25952]	[0.24524]	[-0.70161]	[-0.12475]
LINT(-1)	0.000550	0.004257	0.000588	0.868131
	(0.00160)	(0.00989)	(0.00146)	(0.05450)
	[0.34300]	[0.43030]	[0.40204]	[15.9297]
C	0.003407	-0.005306	0.006517	-0.008129
	(0.00159)	(0.00979)	(0.00145)	(0.05391)
	[2.14823]	[-0.54219]	[4.50731]	[-0.15080]
R-squared	0.147205	0.083682	0.061519	0.742940
Adj. R-squared	0.108877	0.042499	0.019340	0.731387
Sum sq. resids	0.014110	0.537221	0.011727	16.30139
S.E. equation	0.012591	0.077693	0.011479	0.427974
F-statistic	3.840674	2.031952	1.458518	64.30569
Log likelihood	280.4159	109.3579	289.1106	-51.03415
Akaike AIC	-5.859913	-2.220381	-6.044907	1.192216
Schwarz SC	-5.724631	-2.085099	-5.909625	1.327498
Mean dependent	0.004997	-0.003037	0.005014	-0.310792
S.D. dependent	0.013338	0.079398	0.011591	0.825760

Determinant resid covariance (dof adj.)	2.01E-11
Determinant resid covariance	1.61E-11
Log likelihood	634.4356
Akaike information criterion	-13.07310
Schwarz criterion	-12.53197

Le test de cointegration de Johenson ne peut pas s'appliquer car les séries ne sont pas intégrées du même ordre. Comme les séries sont intégrées d'ordre I(1) et I(0), nous pouvons appliquer le modèle ARDL.

Exercice 3 :

Les résultats du test de cointegration (le test de la trace) sur les séries chronologiques stationnaire (en première différence) sont consignés dans la tableau ci après :

Date: 12/20/22 Time: 16:51
Sample (adjusted): 1982 2017
Included observations: 36 after adjustments
Trend assumption: Linear deterministic trend
Series: INF CMM TRESC
Lags interval (in first differences): 1 to 1

Unrestricted Cointegration Rank Test (Trace)

Hypothesized No. of CE(s)	Eigenvalue	Trace Statistic	0.05 Critical Value	Prob.**
None *	0.467036	41.77981	29.79707	0.0013
At most 1 *	0.336745	19.12498	15.49471	0.0135
At most 2 *	0.113660	4.343557	3.841466	0.0371

Trace test indicates 3 cointegrating eqn(s) at the 0.05 level
* denotes rejection of the hypothesis at the 0.05 level
**MacKinnon-Haug-Michelis (1999) p-values

Les résultats du test de la trace figurant dans le tableau ci-dessus montrent que les variables INF, CMM et TRESC sont cointégrées au seuil de 5%. L'hypothèse nulle d'absence de cointégration est rejetée du fait que le test de la trace indique l'existence de trois relations de cointégration.

Nous nous intéressons aux déterminants de l'inflation ; L'estimation du VECM est illustrée dans le tableau suivant :

Vector Error Correction Estimates
Date: 12/20/22 Time: 16:52
Sample (adjusted): 1982 2017
Included observations: 36 after adjustments
Standard errors in ( ) & t-statistics in [ ]

Cointegrating Eq:	CointEq1		
INF(-1)	1.000000		
CMM(-1)	-0.404989 (0.17846) [-2.26941]		
TRESC(-1)	-0.681362 (0.38062) [-1.79014]		
C	1.247807		
Error Correction:	**D(INF)**	**D(CMM)**	**D(TRESC)**
CointEq1	-0.358615 (0.12670) [-2.83046]	0.126604 (0.31328) [0.40413]	0.041029 (0.02921) [1.40459]

Les coefficients associés à la varaible CMM est significativement défirent de zéro d'un point de vue statistique, telle que l'indique la statistique de student calculée supérieure à la valeur critique au seuil de 5%. CointEq1 indique les résidus retardés d'une période de la relation de cointégration qui figure dans le tableau ci-dessus. Les statistiques de Student sont ceux mises entre crochet. Ainsi, les résultats obtenus montrent que le terme à correction d'erreur est négatif et significativement différent de zéro, ce qui signifie que les variables sont caractérisées par un retour vers la cible de long terme (vers l'équilibre).

Conclusion générale

Au terme de ce cours, nous avons mis l'accent sur les séries temporelles qui sont des réalisations de processus stochastiques. Ce processus peut être stationnaire, sous certaines conditions, ou non stationnaire avec une tendance déterministe uniquement ou avec une tendance stochastique. Les tests de racine unitaire vérifient l'hypothèse de racine unitaire, contre l'hypothèse de stationnarité ou de non-stationnarité à tendance déterministe uniquement. La classification des processus est importante en pratique : le type de méthode d'inférence statistique à utiliser dépend des propriétés de ces processus. Les tests de racine unitaire permettent également de déterminer une transformation stationnaire d'une série non stationnaire. Cette transformation stationnaire est nécessaire pour la spécification l'estimation de processus type ARMA.

Dans ce support de cours nous avons présenté une initiation aux fondements de l'économétrie. Ces fondements sont indispensables aux étudiants pour pouvoir élargir leurs connaissances. Ce cours étant semestriel, ne nous permet pas d'aborder tous les aspects de cette branche. En effet, les modèles non linaires n'ont pas été traité. Ainsi, les étudiants sont invités à approfondir certains aspects de ce cours par des lectures complémentaires.

Bibliographie

1. ABDALLAH A.Taux de change et performances économiques dans las pays en développement : l'exemple du Maghreb. Thèse de Doctorat. Université Val de Marne, Paris XII. 2006.
2. ABDERRAHMANI F, (2018) « Guide pratique des séries temporelles macro-économiques et financières avec eviews 9.5 », *polycopié de cours à caractère pédagogique*, université de Bejaia.
3. BOURBONNAIS R, (2009), « *Econométrie, Manuel et exercices corrigés* », 7ème Edition, DUNOD, Paris

4. BOURBOUNIS R et MICHAL T (1998), *Analyse des séries temporelles en économie*, Press Universitaire de France
5. BOURBONNAIS R, *« économétrie : cours et exercices corrigés*», 9eme édition dunod, Paris, 2015.
6. BOUZNIT, M. « Rendement du capital humain et dynamique de la croissance au sein des pays sous développées» thèse de doctorat, ENSSEA, 2016, p 73-75.
7. DOR E, « *Econométrie* », Pearson Education France, 2009.
8. DEGERINE Serge, (2007) cour de séries chronologique. Université Joseph Fourier
9. HAMISULTANE Hélène, Econométrie des séries temporelles, Consulté le 10/05/2020 sur https://shs.hal.science/cel-01261174/document
10. LARDIC, S et MIGNON V. *Econométrie des séries temporelles macroéconomiques et financières*. Paris : Economica, 2002.
11. MONBET. V, (2017), « Modélisation des séries temporelles Master Statistique et Économétrie Notes de cours », Consulté le 15/05/2020 sur https://perso.univ-rennes1.fr/valerie.monbet/ST_M1/CoursST2017_1.pdf
12. MONINO JL, KOSIANSKI JM, LE CORNU F , Travaux diriges – statistique descriptive – Polycopier de cours
13. MOUSSI F, TOUATI K, (2021), Etude économétrique des déterminants du taux d'intérêt débiteur en Algérie", Revue algérienne d'économie et gestion, Volume 15, Numéro 2, Pages 865-883, Disponible sur https://www.asjp.cerist.dz/en/downArticle/154/15/2/176718
14. NICOLEAU Florence, « séries chronologiques », Polycopié de cours, IUT de NICE CÔTE D'AZUR, Département STID, 2005/2006
15. OBAD Jaouad and Youssef JAMAL, « L'impact des dépenses publiques sur la croissance économique au Maroc : Application de l'approche ARDL », ISSN 2028-9324 Vol. 16 No. 2 Jun. 2016, pp. 444-455, pdf, page 3 à 6. http://www.ijias.issr-journals.org/
16. PERRAUDIN Corinne, « *SERIES CHRONOLOGIQUES* », Université Paris I, Cours de Magistère d'Economie – Deuxième année, 2004-2005
17. TOUATI (2021), Les effets à court et long termes du prix de pétrole sur le taux de change en Algérie : Modèle ARDL sur données mensuelles (2012-2019), *Revue des Sciences Economiques, de Gestion et Sciences Commerciales,*
18. TOUATI K, KEDDARI N (2021), Impact de la diversification des exportations sur la croissance économique en Algérie : Modélisation VAR, in Ouvrage collectif national intitulé *La problématique de la diversification économique dans le cadre du développement durable en Algérie, entre opportunités et défis,* Université de Constantine, Algérie
19. TOUATI , (2017)."The impact of oil price shock of 2014 on the exchange rate in Algeria: Vector Autoregressive Model" Revue Finance & marchés, Volume 4, Numéro 1, Pages 200-235
20. Cours de Méthodes statistiques pour l'analyse des données en psychologie, Master 1, Université Paris Ouest Nanterre La Défense UFR SPSE- PMP STA 21 https://fermin.perso.math.cnrs.fr/Files/Chap3.pdf
21. http://bibliotheque.pssfp.net/livres/SCIENCES_DE_GESTION_SYNTHESE_DE_COURS_EXERCICES_CORRIGES.pdf