

---

---

# DEDICACES

---

Ce travail est dédié à :

- Ma mère NJOUMALEU Elisabeth
- Mes enfants
  - \* DEUMENI Ange Venus
  - \* LEUMENI Freddy Yvan
  - \* LEUDJEU WOUAPPI Beautrel Horssel
- Tous mes amis

---

---

# REMERCIEMENTS

---

Ce travail a été effectué sous la direction du Dr. NDONG NGUEMA Eugène à qui j'exprime ma profonde reconnaissance.

Je remercie également mon encadreur en stage, M.FANDIO TCHABO Ferdinand, Ingénieur Statisticien Economiste, pour toute la compréhension, la disponibilité et la patience dont il a fait montre à mon égard pendant mon stage à l'Institut National de la Statistique (I.N.S.).

Mes sincères remerciements vont également à l'endroit de :

⇒ Tous les enseignants du Master de statistique appliquée de l'Université de Yaoundé I, en particulier :

Pr. Henri GWET pour tout ce qu'il fait pour la réussite du MASTER ;

Pr. Didier DACUNHA-CASTELLE ;

Pr. J.C. THALABARD ;

Pr. Elisabeth GASSIAT.

⇒ Tous mes camarades de promotion.

---

---

# RÉSUMÉ

---

L'objectif de l'indice de la production industrielle (I.P.I) est de retracer la dynamique de la production industrielle Camerounaise selon un rythme trimestriel. Cependant, les variations d'un trimestre sur l'autre des indices bruts sont difficilement interprétables du fait de la présence d'effets saisonniers et d'effets de composition trimestrielle des jours dans le calendrier. Pour faciliter l'interprétation de la production, il est important de désaisonnaliser la série des indices.

Dans notre travail, il est question de désaisonnaliser les indices de la production au CAMEROUN par branche d'activité. La désaisonnalisation d'une série consiste à utiliser certaines techniques mathématiques afin d'expurger la série des variations infra-annuelles périodiques. Pour le faire, nous avons eu besoin des outils mathématiques suivants : les séries temporelles. Ensuite sur deux exemples, nous avons montré comment désaisonnaliser en utilisant deux méthodes : la méthode additive et la méthode multiplicative.

Enfin une modélisation nous a permis de faire une prévision. Les résultats suivants ont été obtenus : - les estimateurs des coefficients saisonniers ; - la série désaisonnalisée ; - les valeurs prévisionnelles sur un an.

Nous sommes arrivés à la conclusion selon laquelle la production industrielle au Cameroun est fortement liée au cycle agricole.

L'intérêt pratique de ce travail est de rendre plus facilement interprétable le rythme de production industrielle au CAMEROUN.

Les méthodes pour parvenir à ces résultats ont, bien sûr, leurs limites, mais elles demeurent des outils utiles et pertinents pour s'adapter et réagir de façon plus éclairée à l'information donnée par les séries chronologiques.

---

---

# ABSTRACT

---

The objective of the index of the industrial production is to recall the CAMEROON dynamic industrial production according to a term rythm. However, the variations of raw term indices are hard to interpret properly because of the presence of seasonal effects and term assessment (evaluation) effects within the days of calendar. With an each term published index, the interpretation of the production is more precise and especially rapid than a simple analysis in terms of annual sliding or yearly average.

Time series have a great importance in all domains of the observation or collection of data evolving with time.

The serial seasonality is an important factor in the explanation of the serial fluctuations and it should be taken into account in all analyses or interpretations. The deseasonalisation of time series consists in using some mathematical techniques in order to expurgate the periodic infra-annual variations. Thus, several methods have been developed to master the seasonal component of the time series in order to explain the present situation and foresee easily how would be the evolution. Two concrete examples of time series data from the cameroonian industrial production indices have allowed us to perform deseasonalisation through an additive model and through a multiplicative model.

From our study, we reach the conclusion that the cameroonian industrial production is strongly correlated with the yearly agricultural cycle.

The methods used for reaching these results have, surely, their limits, but, they remain useful and relevant tools to react and adapt oneself in an enlightened way to the given information by the chronological series.

---

---

# Table des matières

---

<b>DÉDICACES</b>	<b>i</b>
<b>REMERCIEMENTS</b>	<b>ii</b>
<b>RÉSUMÉ</b>	<b>iii</b>
<b>ABSTRACT</b>	<b>iv</b>
<b>INTRODUCTION</b>	<b>2</b>
<b>1 QUELQUES NOTIONS ESSENTIELLES</b>	<b>3</b>
1.1 Rappels sur les séries chronologiques . . . . .	3
1.1.1 Définition d'une série chronologique / temporelle . . . . .	3
1.1.2 Exemples . . . . .	3
1.1.3 Quelques problèmes posés par les séries temporelles . . . . .	3
1.1.4 La modélisation des séries temporelles . . . . .	4
1.1.5 Estimation et élimination des trends et des saisonnalités . . . . .	6
1.2 Eléments de base de statistique . . . . .	7
1.2.1 Espérance-Variance-covariance-Corrélation . . . . .	7
1.2.2 Indépendance-Stationnarité . . . . .	8
1.2.3 Autocovariance-Autocorrélation-Bruit blanc . . . . .	9
1.2.4 Processus stationnaires: Exemples . . . . .	9
1.2.5 Processus stationnaire: Contre-exemple . . . . .	9
1.2.6 Processus moyenne mobile ( <i>moving average</i> ) . . . . .	10
1.2.7 Processus autorégressif . . . . .	10
1.2.8 Etapes pour la modélisation des séries chronologiques . . . . .	11
1.3 Processus linéaires-modèles paramétriques simples . . . . .	11
1.3.1 Processus linéaire . . . . .	11
1.3.2 Estimation de la moyenne et des autocovariances . . . . .	12
1.4 Théorème central limite (T.C.L) cf[13] . . . . .	13
1.5 Exemples . . . . .	13
1.6 Choix de processus . . . . .	14
1.6.1 Processus MA( $q$ ) . . . . .	14
1.6.2 Modèle AR( $p$ ) . . . . .	14
1.6.3 Quelques notes sur les tests . . . . .	15
1.7 Quelques notes sur l'I.P.I . . . . .	16

---

1.7.1	Objet et caractéristique de base . . . . .	16
1.7.2	Utilisation des indices . . . . .	17
<b>2</b>	<b>MODÉLISATION ET DÉSAISONNALISATION</b>	<b>18</b>
2.1	Modèle additif . . . . .	18
2.1.1	Désaisonnalisation par régression linéaire . . . . .	18
2.1.2	Désaisonnalisation par la méthode des moyennes mobiles . . . . .	23
2.2	Modèle multiplicatif . . . . .	25
2.2.1	Lissage . . . . .	26
2.2.2	Calcul des multiplicateurs saisonniers perturbés . . . . .	26
2.2.3	Moyennes des coefficients des trimestres de même nom . . . . .	26
2.2.4	Neutralisation annuelle des coefficients saisonniers . . . . .	26
2.2.5	Calcul de la série CVS . . . . .	27
2.2.6	Vérification de la qualité du modèle par le calcul des valeurs ajustées . . . . .	27
2.2.7	Projections (Prévisions) . . . . .	27
2.2.8	Autre méthode de désaisonnalisation dans le cas du modèle multiplicatif . . . . .	27
2.2.9	Présentation d'un exemple: Série: "Total production industrielle" . . . . .	28
<b>3</b>	<b>APPLICATIONS</b>	<b>32</b>
3.1	Série NO1: TOTAL GENERAL I.P.I . . . . .	32
3.1.1	Analyse exploratoire des données . . . . .	32
3.1.2	Stationnarité et identification du modèle . . . . .	36
3.1.3	Estimation et validation du modèle . . . . .	40
3.1.4	Prévision. . . . .	42
3.2	Série NO2: TOTAL INDUSTRIES MANUFACTURIERES . . . . .	43
3.2.1	Analyse exploratoire des données . . . . .	43
3.2.2	Stationnarité et identification du modèle . . . . .	47
3.2.3	Estimation et validation du modèle . . . . .	51
3.2.4	Prévision . . . . .	55
	<b>CONCLUSION</b>	<b>55</b>
	<b>BIBLIOGRAPHIE</b>	<b>57</b>

---

---

# INTRODUCTION

---

L'étude des séries chronologiques ou séries temporelles correspond à l'analyse statistique d'observations espacées dans le temps. Leur domaine d'application est très vaste et s'étend de l'astronomie à l'économie en passant par la biologie ou la théorie du signal. Elles ont donc suscité un vif intérêt, ce qui a eu pour conséquence le développement de nombreux modèles : AR, ARMA, ARIMA, GARCH, ...

Notre travail porte essentiellement sur les données de l'indice de la production industrielle au Cameroun (IPI). Nous avons eu au total 20 séries chronologiques représentant les indices de production industrielle par branche d'activité. Notre travail porte sur la désaisonnalisation de ces indices. Par la suite nous avons tenté de modéliser au mieux certaines des séries considérées.

Nous avons subdivisé ce travail en trois parties :

- dans un premier temps, nous faisons quelques rappels sur les séries temporelles;
- ensuite nous exposons quelques méthodes de désaisonnalisation;
- en troisième lieu, une application permet d'obtenir nos principaux résultats sur les séries *TOTAL*, *indice de production industrielle* et *TOTAL*, *industries manufacturières*.

# QUELQUES NOTIONS ESSENTIELLES

---

## 1.1 Rappels sur les séries chronologiques

### 1.1.1 Définition d'une série chronologique / temporelle

Une série chronologique (série temporelle, chronique) est une suite d'observations numériques d'une grandeur, qui se distinguent par l'ordre dans lequel elles ont été collectées. C'est donc une suite d'observations répétées d'une variable  $X_t$  (ou de plusieurs variables, c'est-à-dire un vecteur  $(X_t^{(1)}, X_t^{(2)}, \dots)$ ).

**Notation :**

Nous noterons  $\{X_t, t \in I\}$  ou  $\{X_t\}_{t \in I}$  une série temporelle, où  $I$  est un intervalle de temps. Deux cas sont à distinguer :

- (i)  $I$  est un intervalle de temps discret, c'est-à-dire :  $t = 1, 2, 3, \dots, T$ , où  $T$  est le nombre d'observations ( $I \subseteq \mathbb{N}$ );
- (ii)  $I$  est un intervalle de temps continu ( $I \subseteq \mathbb{R}$ ).

### 1.1.2 Exemples

- 1. La série des indices de production industrielle.
- 2. La série des indices mensuels des prix à la consommation.

### 1.1.3 Quelques problèmes posés par les séries temporelles

L'examen d'une ou de plusieurs séries peut donner lieu aux problèmes suivants :

- A. La prévision
- B. La détection de rupture.
- C. L'enlèvement de la tendance.
- D. La correction des variations saisonnières.
- E. La causalité-le décalage temporel-la séparation court et long terme ( Décomposition tendance-cycle).
- F. Diagnostic du modèle (Test d'ajustement).
- G. Estimation des paramètres.

**(1) La prévision**

Prévoir les valeurs futures  $X_{T+h}$ ,  $h \geq 1$ , à partir des valeurs données  $X_1, \dots, X_T$ , par les *valeurs en moyenne* (avec une incertitude à contrôler). Il faut chercher un intervalle de prévision (susceptible de contenir la valeur inconnue de  $X_{T+h}$ ) noté par :  $[\widehat{X}_T^1(h); \widehat{X}_T^2(h)]$ .

Cette méthode est aussi applicable pour des valeurs de  $X_t$ ,  $t \leq T$ , pour traiter des valeurs manquantes et mesurer l'effet d'un phénomène accidentel (grève, ...).

**(2) L'enlèvement de la tendance**

La tendance (trend) traduit l'évolution à moyen terme d'une série. La présence de cette composante dans les séries crée de fortes corrélations (corrélation superficielle) entre les variables  $X_t$  qui n'expriment aucune liaison à caractère explicatif entre elles. C'est pourquoi il faut estimer et enlever cette tendance pour voir si de telles liaisons existent.

**(3) La correction des variations saisonnières**

Comme pour la tendance, il faut corriger les variations saisonnières, car la présence de cette composante oblitère les évolutions significatives.

(2) et (3) sont des exemples pour rendre une série "stationnaire".

**(4) Autres problèmes importants : causalité-décalage temporel**

Dans le cas de plusieurs variables observées simultanément, on voudrait savoir si une influence de l'une sur l'autre existe. Et si oui, avec quel délai et pendant combien de temps la variable explicative influence-t-elle la variable expliquée ?

**Exemple 1.1.** :  $Y_t = a_1 X_{t-1} + a_3 X_{t-3}$ . Ce modèle suppose qu'un choc sur la variable  $X_t$  affecte la variable  $Y_t$  avec des décalages de 1 et 3 périodes.

**1.1.4 La modélisation des séries temporelles**

**(1) Modèle d'ajustement**

**Modèle classique de décomposition :**  $X_t = m_t + S_t + \varepsilon_t$ ,  $t = 1, \dots, T$ .

$m_t$  = fonction (déterministe) de la tendance (soit constante, linéaire, un polynôme, ...);

$S_t$  = saisonnalité (déterministe) périodique du temps, de périodicité  $d$  connue, nulle en moyenne temporelle. On a  $S_t = S_{t+d}$ ,  $\forall t$ , la saisonnalité est définie par  $d$  coefficients saisonniers  $c_1, c_2, \dots, c_d$  avec la contrainte :

$$\sum_{i=1}^d c_i = 0.$$

$\varepsilon_t$  = "bruit aléatoire" ("random noise"), représentant les erreurs non systématiques : (nulles en moyenne), mais pas un "bruit blanc": avec structure de corrélation.

$Y_t$  = partie la plus intéressante : porte le caractère explicatif pour  $X_t$ .

$X_t - m_t$  est la *série détendancialisée*. Elle est égale à *saisonnier + résidu*.

$X_t - S_t$  est la *série désaisonnalisée*. Elle est égale à *tendance + résidu*. Elle est aussi la *série corrigée des variations saisonnières*. On estime le saisonnier sur une *série détendancialisée*.

**Ajustement additif :**

$$f(t, \varepsilon_t) = g(t) + \varepsilon_t \quad (\text{par exemple } X_t = m_t + S_t + \varepsilon_t, t = 1, \dots, T).$$

**Ajustement multiplicatif :**

$$f(t, \varepsilon_t) = g(t) \cdot \varepsilon_t.$$

**Exemple 1.2.**  $m_t = a + bt$ ;  $S_t = 0$ ;  $X_t = a + bt + \varepsilon_t$

Si  $Y_t = BB$  (moyenne nulle, indépendances), on aurait :

en moyenne :  $X_t = a + bt \Rightarrow \varepsilon_t$  n'apporterait aucune information explicative pour  $X_t$  (la prévision, valeur future en moyenne, suivrait une "loi déterministe").

**Principe de la modélisation stochastique**

Observation  $\Rightarrow$  Modélisation  $\Rightarrow$  Estimation  $\Rightarrow$  Diagnostic.

**Exemple 1.3.** *pour diagnostic du modèle*  $X_t = a + \varepsilon_t$

La moyenne théorique  $a$  est-elle différente de zéro c'est-à-dire la moyenne empirique  $\hat{a}$  est-elle significative? (est-elle significativement différente?).

On effectue un test d'hypothèse :

$$H_0 : a = 0 \text{ vs } H_1 : a \neq 0 \text{ basée sur } \hat{a} = \frac{1}{T} \sum_{t=1}^T X_t$$

en supposant que la variance  $\sigma_\varepsilon^2$  des erreurs  $\varepsilon$  est connue.

**Statistique de test :**  $\hat{Z}_t := \sqrt{T} \frac{\hat{a} - 0}{\sigma_\varepsilon} \sim \mathcal{N}(0,1).$

On rejette  $H_0$  au niveau  $\alpha$  (choisi a priori = 0.05) si :

$$|\hat{Z}_t| > z_{1-\frac{\alpha}{2}} = \text{quantile d'une normale centrée réduite.}$$

Cela nous donne un contrôle qui permet de voir si le modèle est bien ajusté ( $a \neq 0$ ). C'est-à-dire si  $\hat{Z}_t$  est "trop" petit, on est tenté de choisir un modèle plus simple ("parcimonieux"); ici, alors  $Y_t = \varepsilon_t$ .

**(2) Modèles autoprojectifs ou autorégressifs**

$$Y_t = h(Y_{t-1}, Y_{t-2}, \dots, \varepsilon_t)$$

où  $\varepsilon_t$  est une variable aléatoire (indépendante de  $Y_t$ ).

Si  $h$  est linéaire, on parle de modèles autorégressifs. Par exemple :

$$Y_t = a_1 Y_{t-1} + a_2 Y_{t-2} + \varepsilon_t; \quad \text{où } \varepsilon_t \text{ est un Bruit Blanc.}$$

**(3) Modèles explicatifs**

$X_t = f(V_t, \varepsilon_t)$  où  $V_t =$  variable exogène (vecteur), nous avons deux types:

- Statique, par exemple :  $X_t = a + bV_t + \varepsilon_t$ .
- Dynamique, par exemple :  $X_t = f(X_{t-1}, \dots, X_{t-p}, V_t, \varepsilon_t)$ .

**Rappel: régression linéaire-méthode des moindres carrés**

$Y_t = \beta_0 + \beta_1 X_t + \varepsilon_t, t = 1, \dots, T, \varepsilon_t =$  Bruit Blanc, c'est-à-dire indépendants et identiquement distribués (i.i.d), de moyenne nulle et de même variance. On cherche les estimateurs de moindres carrés  $\hat{\beta}_0, \hat{\beta}_1$  de  $\beta_0$  et  $\beta_1$  : on cherche la ligne droite de paramètres  $\beta_0, \beta_1$  qui, à partir de l'échantillon, ajuste mieux le nuage de points  $(X_t$  et  $Y_t)$ . Analytiquement, on a:

$$S(\beta_0, \beta_1) := \sum_{t=1}^T [Y_t - (\beta_0 + \beta_1 X_t)]^2 \text{ soit minimal.}$$

Solution (par dérivation de  $S(\beta_0, \beta_1)$ ):

$$\hat{\beta}_1 = \frac{\sum_{t=1}^T (X_t - \bar{X})(Y_t - \bar{Y})}{\sum_{t=1}^T (X_t - \bar{X})^2}, \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}.$$

Les estimateurs des moindres carrés ordinaires (M.C.O.) sont justifiés seulement si  $\varepsilon_t$  est un Bruit Blanc.

**Exemple 1.4.**  $Y_t = a + bt + \varepsilon_t, i.e. X_t = t.$

$$\bar{X} = \frac{T}{2} \text{ et } \sum_{t=1}^T (X_t - \bar{X})^2 = \sum_{t=1}^T \left(t - \frac{T}{2}\right)^2 = \left(\frac{T^2}{2} + 1\right) \frac{T}{6}.$$

Dans le cadre des séries chronologiques,  $\varepsilon_t \neq$  Bruit Blanc.

**1.1.5 Estimation et élimination des trends et des saisonnalités**

Modèle  $X_t = m_t + S_t + \varepsilon_t, \varepsilon_t$  de moyenne nulle. Plusieurs méthodes sont proposées à cet effet:

- A. Méthode paramétrique: Estimation du trend par moindres carrés
- B. Méthode non paramétrique : Estimation du trend | des saisonnalités par moyenne mobile
- C. Méthodes pour enlever le trend | saisonnalités (sans les estimer) par différences (itérées)

Utilisation de l'opérateur de retard  $B$

$$BX_t := X_{t-1} \quad (\implies B^2 X_t := B(BX_{t-1}) = X_{t-2})$$

et l'opérateur de différence  $\nabla$  :

$$\nabla_d X_t = (1 - B^d) X_t = X_t - X_{t-d} \quad (\nabla := \nabla_1); \quad \left(\implies \nabla_d^2 X_t = (1 - B^d)^2 X_t = X_t - 2X_{t-d} + X_{t-2d}\right).$$

**Estimation du trend par moindres carrés**

$X_t = m_t + \varepsilon_t, \varepsilon_t$  de moyenne nulle.

**Exemple 1.5.** :  $m_t = a_0 + a_1 t + a_2 t^2.$

Donnés  $\{X_t\} \longrightarrow$  estimateurs M.C.O de  $a_0, a_1, a_2.$

Résidus :  $\hat{\varepsilon}_t = X_t - \hat{m}_t = X_t - (\hat{a}_0 + \hat{a}_1 t + \hat{a}_2 t^2).$

Prédiction :  $\hat{X}_t$  de  $X_t$  : on utilise  $\hat{X}_t = \hat{m}_t$  (car  $E\varepsilon_t = 0$ ).

**Estimation des trends | saisonnalités par moyenne mobile**

$$X_t = m_t + S_t + \varepsilon_t, E\varepsilon_t = 0 \text{ et } S_{t+d} = S_t, \sum_{j=1}^d S_{t+j} = 0 \text{ (pas de tendance dans } S_t).$$

**Exemple 1.6.** :  $d = 12$ . On utilise une reindiciation :

$X_{jk}, j = 1, \dots, 6$  (ans),  $k = 1, \dots, 12$  (mois) avec

$$X_{jk} := X_{k+12(j-1)}.$$

**Estimation de  $m_t$  par moyenne mobile**

Pour  $j = 1, \dots, 6$  :  $\hat{m}_j = \frac{1}{12} \sum_{k=1}^{12} X_{jk}$  (car trend constante par an; sinon, moyenne plus courte, forme générale:  $\hat{m}_t = (2q + 1)^{-1} \sum_{n=-q}^{n=q} X_{t+n}$ ).

**Estimation de  $S_t$  : cf [5]**

$$\hat{S}_k = \frac{1}{6} \sum_{j=1}^6 (X_{jk} - \hat{m}_j), \quad k = 1, \dots, 12.$$

(Généralisation pour des données avec  $d \neq 12$ ).

**Résidus** :  $\hat{\varepsilon}_{jk} = X_{jk} - \hat{m}_j - \hat{S}_k$

**Méthode pour enlever le trend | la saisonnalité sans estimation**

(1)  $X_t = m_t + \varepsilon_t, E\varepsilon_t = 0$ . Soit  $m_t = \sum_{j=0}^k a_j t^j$ . L'application de  $\nabla^k$  va enlever ce trend:

$$\nabla^k X_t = k! a_k + \nabla^k \varepsilon_t \text{ est de moyenne constante.}$$

(2)  $X_t = m_t + S_t + \varepsilon_t, S_t$  de périodicité  $d$ .

$$\begin{aligned} \nabla_d X_t &= m_t - m_{t-d} + S_t - S_{t-d} + \varepsilon_t - \varepsilon_{t-d}; \quad \nabla_d = 1 - B^d \\ &= m_t - m_{t-d} + \varepsilon_t - \varepsilon_{t-d}. \end{aligned}$$

L'application de (1) sur trend  $m_t - m_{t-d}$  l'enlèvera.

**Exemple 1.7.** :  $d = 12, \nabla_{12} X_t, t = 13, \dots, 72$ . S'il reste un faible trend croissant, on utilise  $\nabla \nabla_{12} X_t$  ( on applique une différence d'ordre un et de périodicité 12).

## 1.2 Eléments de base de statistique

### 1.2.1 Espérance-Variance-covariance-Corrélation

Soient  $X, Y$  des variables aléatoires (réelles).

a)  $p(x)$  est la densité de probabilité de  $X$ , c'est-à-dire

$$P(X \in B) = \int_B p(x) dx \text{ et } P(X \leq b) = \int_{-\infty}^b p(x) dx .$$

**Exemple 1.8.** : La distribution normale  $\mathcal{N}(\mu, \sigma^2)$  :

$$p(x) = (2\pi\sigma^2)^{-\frac{1}{2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

avec

$$EX = \mu \text{ et } Var(X) = \sigma^2.$$

- b)  $EX = \int_{-\infty}^{+\infty} xp(x) dx = \text{espérance de } X$ .
- c)  $Var(X) = E(X - EX)^2 = \int_{-\infty}^{+\infty} (x - EX)^2 p(x) dx = \text{variance de } X$ .
- d) Les moments (centrés) d'ordre  $k$  :  $EX^k$  et  $E(X - EX)^k$ .
- e) La covariance entre  $X$  et  $Y$  :

$$\begin{aligned} r_{XY} &= cov(X, Y) \\ &= cov(Y, X) \\ &= E[(X - EX)(Y - EY)] \\ &= E[XY] - EXEY \\ &= \text{une mesure pour la dépendance linéaire entre } X \text{ et } Y : \end{aligned}$$

**N.B:**  $cov(X, X) = Var(X)$ .

- f) La corrélation entre  $X$  et  $Y$  :

$$corr(X, Y) = \frac{cov(X, Y)}{\sigma_X \sigma_Y} = \rho_{XY}$$

$-1 \leq \rho_{XY} \leq 1$  et  $\rho_{XY} = 0 \iff X$  et  $Y$  sont non corrélés linéairement.

### 1.2.2 Indépendance-Stationnarité

$X$  et  $Y$  sont indépendants ( $X \perp Y$ ) si

$$p_{(X, Y)}(x, y) = p_X(x)p_Y(y)$$

(la distribution conjointe est le produit des distributions marginales)

**Important :**  $X \perp Y \implies cov(X, Y) = 0$ ; mais en général, pas réciproquement, sauf pour une normale bivariée :

si  $(X, Y)$  est normale,  $X \perp Y \iff cov(X, Y) = 0$ .

Dans la suite, si nous parlons du modèle pour une série temporelle  $X_1, \dots, X_T$ , nous utilisons la notation :

$$\{X_t\}_{t \in \mathbb{Z}} = \text{"processus stochastique"}$$

**Définition 1.1.**  $\{X_t\}$  est (faiblement) stationnaire si :

- $EX_t = \mu, \forall t$ ;
- $Var(X_t) = \sigma_X^2 < \infty, \forall t$ ;
- $cov(X_t, X_{t+h}) = r_h^X, \forall t$ , donc les  $r_h^X$  ne dépendent pas de  $t$ .

### 1.2.3 Autocovariance-Autocorrélation-Bruit blanc

**Définition 1.2.** Dans les conditions de la définition 1.1 :

- (1) La suite  $\{r_t^X\}$  est appelée " suite (fonction) des autocovariances" (ACV) de  $\{X_t\}$ .
- (2) La suite  $\{\rho_t^X\}$ ,  $\rho_t^X = \frac{r_t^X}{r_0^X}$ , est appelée " suite (fonction) des autocorrélations" (ACF) de  $\{X_t\}$ .

**N.B :**  $\rho_{-t}^X = \rho_t^X$  et  $|\rho_t^X| \leq \rho_0^X = 1$ .

- (3) Fonction d'autocorrélation partielle (PACF) :

$$\alpha(h) = \frac{\text{cov}[X_t|X_{t-1}, \dots, X_{t-h+1} - E(X_t|X_{t-1}, \dots, X_{t-h+1}), X_{t-h} - E(X_{t-h}|X_{t-1} \dots X_{t-h+1})]}{\text{var}[X_{t-h} - E(X_{t-h}|X_{t-1} \dots X_{t-h+1})]}$$

**Définition 1.3 (Bruit Blanc).**

- (1) Bruit blanc faible :  $\{\varepsilon_t\} \sim BB(0, \sigma^2)$  si :

- $E\varepsilon_t = 0$  ;
- $\text{Var}(\varepsilon_t) = \sigma_\varepsilon^2 < \infty$  ;
- $r_t^\varepsilon = \sigma_\varepsilon^2 \cdot \delta_{t,0}$ , c'est-à-dire  $r_t^\varepsilon = \begin{cases} \sigma_\varepsilon^2 & \text{si } t = 0 \\ 0 & \text{sinon.} \end{cases}$

- (2) Bruit blanc fort : Bruit blanc faible plus  $\{\varepsilon_t\} \sim i.i.d.$

**Notation :**  $\{\varepsilon_t\} \sim IID(0, \sigma^2)$

### 1.2.4 Processus stationnaires : Exemples

- (1) Soit  $X_t = aX_{t-1} + \varepsilon_t$ ,  $\{\varepsilon_t\} \sim BB(0, \sigma^2)$

Processus autorégressif (d'ordre 1) : stationnaire si  $|a| < 1$ . **cf[14]**

- (2) Soit  $X_t = \varepsilon_t + \theta\varepsilon_{t-1}$ ,  $\{\varepsilon_t\} \sim BB(0, \sigma^2)$

Processus moyenne mobile (d'ordre 1).

Son autocovariance :

$$r_h^X = \begin{cases} (1 + \theta^2) \sigma^2 & \text{si } h = 0 \\ \theta \sigma^2 & \text{si } h = \pm 1 \\ 0 & \text{si } |h| > 1 \end{cases}$$

cela implique que son autocorrélation  $\rho_h^X = r_h^X / \sigma_X^2$  est donnée par :

$$\rho_h^X = \begin{cases} 1 & \text{si } h = 0 \\ \frac{\theta}{(1+\theta^2)} & \text{si } h = \pm 1 \\ 0 & \text{si } |h| > 1. \end{cases}$$

### 1.2.5 Processus stationnaire : Contre-exemple

**Définition 1.4.** random walk (marche aléatoire)

$$X_t = \sum_{s=1}^t \varepsilon_s, t \geq 1, \{\varepsilon_t\} \sim BB(0, \sigma^2)$$

ou  $X_t = X_{t-1} + \varepsilon_t, t > 1, X_1 = 0$ .

$$\text{cov}(X_t, X_{t+h}) = \sigma^2 \cdot t, \forall h.$$

$\{X_t\}$  n'est pas stationnaire (car sa variance explose).

Interprétation : marché aléatoire : Dans le domaine financier par exemple, les prix (des actions) d'aujourd'hui donnent de l'information sur les prix de demain mais cette information est *non systématique* (différence aléatoire, avec un certain contrôle de la variabilité ( $\sigma^2 < \infty$ )), cela implique qu'aucune prévision n'est possible.

Il est nécessaire de regarder la structure des corrélations, par exemple la moyenne mobile d'ordre 1 qui permet la prévision de l'horizon  $h = 1$  mais pas de l'horizon  $h > 1$ .

### 1.2.6 Processus moyenne mobile ( *moving average* )

#### (1) Le processus moyenne mobile d'ordre 1 (MA(1))

$$X_t = \mu + \varepsilon_t + \theta\varepsilon_{t-1}, \quad \{\varepsilon_t\} \sim BB(0, \sigma_\varepsilon^2)$$

**Paramètres :** Le coefficient  $\theta$ , la variance  $\sigma_\varepsilon^2$  des "innovations"  $\varepsilon_t$ , la moyenne  $\mu$  (qui peut être supposée =0)

#### (2) Le processus moyenne mobile d'ordre q (MA(q))

$$X_t = \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \dots + \theta_q\varepsilon_{t-q} \quad \{\varepsilon_t\} \sim BB(0, \sigma_\varepsilon^2)$$

**Paramètres :**  $\theta_1, \theta_2, \dots, \theta_q; \sigma_\varepsilon^2$ .

**Interprétation :** Au temps  $t$ , on a un choc purement aléatoire  $\varepsilon_t$  qui est indépendant des chocs aléatoires aux temps différents de  $t$ . La valeur observée  $X_t$  est alors une somme pondérée ( une moyenne mobile) de ce choc et des autres chocs du passé.

**Signification :** Créer des processus non purement aléatoires, c'est-à-dire avec une structure de corrélation intéressante (**Ex : MA(1)**). Nous notons que "la mémoire" de  $X_t$  est finie, c'est-à-dire que la corrélation entre  $X_t$  et  $X_{t+h}$  devient zéro pour  $|h| > q$ .

### 1.2.7 Processus autorégressif

#### (1) Le processus autorégressif d'ordre 1 (AR(1))

$$X_t = aX_{t-1} + \varepsilon_t, \quad \{\varepsilon_t\} \sim BB(0, \sigma_\varepsilon^2)$$

**Paramètres :** Le coefficient  $a$  ( on suppose  $|a| < 1$  pour la stationnarité), la variance des innovations  $\sigma_\varepsilon^2$ .

**N.B : Pour inclure une moyenne non nulle :**  $\{X_t\}$  est AR(1) avec moyenne  $\mu$  si  $\{X_t - \mu\}$  est AR(1).

#### (2) Le processus autorégressif d'ordre p (AR(p))

$$X_t = a_1X_{t-1} + \dots + a_pX_{t-p} + \varepsilon_t \quad \{\varepsilon_t\} \sim BB(0, \sigma_\varepsilon^2).$$

**Paramètres :** les coefficients  $a_1, \dots, a_p; \sigma_\varepsilon^2$ .

**Interprétation : Création de la corrélation intéressante.**

Régression de  $X_t$  sur ses valeurs du passé (fini)  $X_{t-1}, \dots, X_{t-p}$  menant à une mémoire infinie (en pratique: existence d'un temps zéro).

$p = 1$  : la valeur d'aujourd'hui  $X_t$  est un multiple ( $< 1$ ) de la valeur d'hier  $X_{t-1}$  plus un choc  $\varepsilon_t$  purement aléatoire et indépendant du passé de  $X_t$ .

### 1.2.8 Etapes pour la modélisation des séries chronologiques

- Etape 1:** Représentation graphique : Existence des trends, saisonnalités, points de rupture, "valeurs aberrantes" ;
- Etape 2:** Elimination des composantes non stationnaires ;
- Etape 3:** Modélisations des composantes stationnaires ("non bruit blanc") ; modèle AR-MA ;
- Etape 4:** Estimation des paramètres de ce modèle ;
- Etape 5:** Diagnostic (validation du modèle) : ( "tester si les résidus sont compatibles avec l'hypothèse de bruit blanc" ) ;
- Etape 6:** Prédiction :
- (a) de la partie stationnaire
  - (b) de la série originale (mesure de la qualité de prédiction : erreur de prédiction ; donner des intervalles de prédiction).

## 1.3 Processus linéaires-modèles paramétriques simples

On utilise les modèles linéaires pour modéliser la composante aléatoire  $\varepsilon_t$  stationnaire des modèles d'ajustement

$$X_t = g(t) + \varepsilon_t \text{ avec par exemple } g(t) = m_t + S_t.$$

### 1.3.1 Processus linéaire

Le but est de contrôler la mémoire du processus au niveau de ses covariances,  $r_h^X = cov(X_t, X_{t+h}) \rightarrow 0$  assez vite si  $|h| \rightarrow +\infty$ . ("mémoire courte").

Les liaisons linéaires entre  $X_t$  et  $X_{t+h}$  disparaissent assez vite si  $|h| \rightarrow +\infty$

**Définition 1.5.**

- (1) Un processus linéaire est une combinaison linéaire (infinie) de bruits blancs.
- (2) Un processus  $\{X_t\}_t$  de moyenne  $\mu$  est appelé processus linéaire (général) s'il existe une représentation linéaire

$$X_t = \mu + \sum_{k=-\infty}^{+\infty} b_k \varepsilon_{t-k},$$

où  $\{b_k\}_k$  est une suite de coefficients réels, de carré sommable, c-à-d  $\sum_{k=-\infty}^{+\infty} b_k^2 < \infty$  et  $\{\varepsilon_t\}_t$  est un bruit blanc (faible).

L'observation  $X_t$  peut être une combinaison linéaire de ses valeurs du passé et des innovations  $\varepsilon_t$

**Exemple 1.9.**  $AR(p)$ ;  $MA(q)$ .  $ARMA(p,q)$ , où le processus  $ARMA(p,q)$  est défini par:

$$X_t = \sum_{k=1}^p a_k X_{t-k} + \varepsilon_t + \sum_{j=1}^q \theta_j \varepsilon_{t-j}, \quad \{\varepsilon_t\} \sim BB(0, \sigma_\varepsilon^2).$$

### 1.3.2 Estimation de la moyenne et des autocovariances

Soient  $X_1, \dots, X_T$ , des données observées d'un processus  $\{X_t\}_t$  stationnaire,  $EX_t = \mu$ ,  $cov(X_s, X_t) = r_{t-s}$ .

(1) \*\*\* **La moyenne empirique**  $\bar{X}_T$

$$\bar{X}_T = T^{-1} \sum_{t=1}^T X_t.$$

$\bar{X}_T$  est un estimateur consistant pour  $\mu$ , c'est-à-dire  $\bar{X}_T \xrightarrow{p.s.} \mu$  si  $T \rightarrow \infty$ . (loi des grands nombres pour des V.a. corrélées; sous des conditions de "dépendance faible", par exemple  $\sum_h |r_h| < \infty$ ).

Pour un processus linéaire (AR, MA, ARMA), sous des conditions supplémentaires,  $\bar{X}_T$  est asymptotiquement normale, c'est-à-dire

$$(T.C.L.) \quad \sqrt{T} (\bar{X}_T - \mu) \longrightarrow \mathcal{N} \left( 0, \sum_{t=-\infty}^{+\infty} r_t \right), \quad T \longrightarrow \infty.$$

Alors,  $\bar{X}_T$  est un estimateur qui se concentre d'autant plus autour de  $\mu$  que  $T$  est grand.

#### Application

- (a) Tests d'hypothèse :  $H_0 : \mu = 0$  VS  $H_1 : \mu \neq 0$
- (b) Etablir des intervalles de confiance pour  $\mu$ .

#### Propriétés de la moyenne empirique cf[13]

Pour un processus stationnaire  $\{X_t\}$  :

- (a)  $E\bar{X}_T = \mu$ ;
- (b)  $Var(\bar{X}_T) = \frac{1}{T} \sum_{t=-T+1}^{T-1} \left(1 - \frac{|t|}{T}\right) r_t$ ;
- (c) sous des conditions supplémentaires (par exemple :  $\sum_t |t r_t| < \infty$ ) :

$$Var(\bar{X}_T) \sim \frac{1}{T} \longrightarrow 0, \text{ si } T \longrightarrow \infty,$$

alors

$$TVar(\bar{X}_T) \longrightarrow \sum_{t=-\infty}^{+\infty} r_t \text{ si } T \longrightarrow \infty;$$

et, pour un processus linéaire avec  $\sum_k |b_k| < \infty$ ,

$$\sum_{t=-\infty}^{+\infty} r_t = \sigma_\varepsilon^2 \left( \sum_k b_k \right)^2.$$

#### Intervalles de confiance

(a) Exact, si  $\{X_t\}$  est gaussien, en s'appuyant sur :

$$\sqrt{T} (\bar{X}_T - \mu) \sim \mathcal{N} (0, T \cdot \text{Var}(\bar{X}_T)) ;$$

(b) sinon, approximatifs (asymptotiques), en s'appuyant sur :

$$\sqrt{T} (\bar{X}_T - \mu) \xrightarrow{\mathcal{L}} \mathcal{N} \left( 0, \sum_{t=-\infty}^{+\infty} r_t \right), T \rightarrow +\infty.$$

(2) \*\*\* **Les autocovariances empiriques**  $\hat{r}_h$

$\hat{r}_h := T^{-1} \sum_{k=1}^{T-h} (X_k - \bar{X}_T) (X_{k+h} - \bar{X}_T)$ ,  $0 \leq h \leq T - 1$ , et  $\hat{r}_{-h} = \hat{r}_h$  (symétrie).

Comme dans le cas de la moyenne empirique,  $\hat{r}_h$  est un estimateur consistant de  $r_h$ .

De plus,  $E\hat{r}_h \rightarrow r_h$  si  $T \rightarrow \infty$ . Sous les conditions d'un processus linéaire, les a.c.v empiriques sont asymptotiquement normales, donc :

$\sqrt{T} (\hat{r}_h - r_h) \rightarrow \mathcal{N} (0, \Sigma)$ ,  $T \rightarrow \infty$ , où  $\Sigma$  est la variance asymptotique.

## 1.4 Théorème central limite (T.C.L) cf[13]

**Théorème 1.1.** Si  $\sum_t |tr_t| < +\infty$ , alors :  $\sqrt{T} (\bar{X}_T - \mu) \rightarrow \mathcal{N} (0, \sum_{-\infty}^{\infty} r_t)$ ,  $T \rightarrow \infty$

(pour un processus linéaire si  $\sum_k |b_k| < \infty$ ).

**Théorème 1.2.** Si  $\sum_t |t r_t| < +\infty$ , alors :  $\sqrt{T} (\hat{\rho}_h - \rho_h) \rightarrow \mathcal{N} (0, \Sigma)$ ,  $T \rightarrow \infty$

où  $\Sigma_{ij} = \sum_{k \geq 1} (\rho_{k+i} + \rho_{k-i} - 2\rho_i \rho_k) (\rho_{k+j} + \rho_{k-j} - 2\rho_j \rho_k)$

(pour un processus linéaire si  $\sum_k |b_k| < \infty$  et  $E[\varepsilon_t^4] < \infty$ ).  $\rho_h = \frac{r_h}{r_0}$  = autocorrélation empirique.

## 1.5 Exemples

(1) **Bruit blanc** : Soit  $\{X_t\} = \{\varepsilon_t\} \sim IID(0, \sigma^2)$  où  $\sigma^2 = 1$ . Nous savons que  $\rho_h = \delta_{h,0}$ . Le T.L.C donne :  $\sqrt{T} (\hat{\rho}_h - \rho_h) \sim \mathcal{N} (0, 1) \Rightarrow \hat{\rho}_h \sim \mathcal{N} \left( \rho_h, \frac{1}{T} \right) = \mathcal{N} \left( 0, \frac{1}{T} \right)$ ,  $h \neq 0$ . Si on fait un graphique de  $\hat{\rho}_h$  vs  $h$ , on peut "tester" si  $\hat{\rho}_h$ ,  $h > 0$ , est significativement différent de zéro ou non : on sait (T.C.L) qu'au seuil de 5%, un intervalle de confiance des  $\rho_h$  ( $h > 0$ ) est donné par :

$\left[ -\frac{z_{1-\frac{\alpha}{2}}}{\sqrt{T}}, \frac{z_{1-\frac{\alpha}{2}}}{\sqrt{T}} \right] = \left[ -\frac{1.96}{\sqrt{T}}, \frac{1.96}{\sqrt{T}} \right]$ . Si seulement un  $\hat{\rho}_h$  se trouve à l'extérieur de cet intervalle, on ne peut pas rejeter l'hypothèse  $H_0$  que  $\rho_h = 0$ ,  $h > 0$ ; c-à-d  $\{X_t\}$  soit  $IID \Rightarrow$  ceci constitue une méthode ( préliminaire ) pour "tester" un bruit blanc.

(2) **Moyenne mobile d'ordre 1** : Soit  $X_t = \varepsilon_t - 0.8\varepsilon_{t-1}$ , où  $\varepsilon_t \sim IID(0, 1)$ .

$$\text{Rappel : } \rho_h = \begin{cases} 1 & \text{si } h = 0 \\ \frac{\theta}{(1 + \theta^2)}, & \text{si } h = \pm 1 \text{ avec } \theta = -0.8 \implies \rho_1 = \frac{-0.8}{1.64} = -0.4878 ; \\ 0, & \text{si } |h| > 1 \end{cases}$$

$\rho_h = 0$ ,  $h > 1$ . Pour  $T = 200$  par exemple, on fait le graphique de  $(h, \hat{\rho}_h)$ ,  $h = 0, \dots, 40$ . On

trouve :  $\hat{\rho}_1 = -0.5354$ , significativement différent de zéro?

$-0.5354 = -7.57 \times \frac{1}{\sqrt{T}} \notin \left[-\frac{1.96}{\sqrt{T}}, \frac{1.96}{\sqrt{T}}\right] \implies$  il faut rejeter, au niveau 5%,  $H_0 : \{X_t\} \sim IID$ .

Pour le test  $H_0 : \{X_t\}$  suit le modèle  $MA(1)$ , calculer la variance de  $\hat{\rho}_i, i > 1$ , qui vaut  $\frac{1 + 2\rho_1^2}{T}$ , Ce qui donne un nouvel intervalle  $I = \left[-\frac{1.96(1+2\hat{\rho}_1^2)^{\frac{1}{2}}}{\sqrt{T}}, \frac{1.96(1+2\hat{\rho}_1^2)^{\frac{1}{2}}}{\sqrt{T}}\right]$

$\implies \hat{\rho}_h, h = 2, \dots, 40 \in I \implies MA(1)$  accepté.

ceci constitue une méthode préliminaire pour tester l'ordre d'un modèle MA (en pratique, avec bornes  $\pm \frac{1.96}{\sqrt{T}}$ ). Elle ne marche pas pour un modèle AR parce que là  $\rho_h \neq 0 \forall h$  ( $\implies PACF$ ).

## 1.6 Choix de processus

### 1.6.1 Processus MA(q)

Ses autocovariances:  $\forall 0 \leq h \leq q, r_h = \sigma_\varepsilon^2 \sum_{k=0}^{q-h} \theta_k \theta_{k+h}$ , avec  $\theta_0 = 1$ . Pour  $h > q, r_h = 0$ .

Soit  $\{X_t\}$ , un processus stationnaire corrélé d'ordre  $q$ , c'est-à-dire  $r_h^X = 0$  pour chaque  $|h| > q$ . Alors  $\{X_t\}$  possède une représentation comme processus  $MA(q)$ .

Un processus  $MA(q)$  est toujours stationnaire et causal: sa variance  $r_0 = \sigma_\varepsilon^2 \sum_{k=0}^q \theta_k^2$  est toujours finie et sa fonction ACV  $r_h$  est stationnaire.

### 1.6.2 Modèle AR(p)

Tests d'hypothèse: PACF, résidus de *prediction* = *bruit blanc*?

(a) Suite des tests d'hypothèse ( $p = 0, 1, 2, \dots$ ):

$$\begin{aligned} H_0 : X_t &= AR(p), \text{ c.à.d. PACF } \alpha(p+1) = 0 \\ H_1 : \alpha(p+1) &\neq 0 \end{aligned}$$

$\alpha(\cdot)$  est la fonction d'autocorrélation partielle (PACF).

En fait  $|\hat{\alpha}(p+1)| \leq z_{1-\frac{\alpha}{2}}/\sqrt{T} \implies$  on accepte  $H_0$  (vient de  $\sqrt{T}\hat{\alpha}(p+1) \stackrel{H_0}{\sim} \mathcal{N}(0,1)$ ): ceci est une approximation du test avec la statistique de test suivante:

$$Q_1 = \frac{\sqrt{T-p-1} \hat{\alpha}(p+1)}{\sqrt{1 - \hat{\alpha}^2(p+1)}} \sim t_{T-p-1} \text{ sous } H_0;$$

on accepte  $H_0$  au niveau  $\alpha$  si  $|\hat{Q}_1| < t_{T-p-1, 1-\frac{\alpha}{2}}$ .

(b) Tester si les résidus sont des bruits blancs:

(i) Calculer les résidus, pour  $t = p+1, \dots, T$ :

$$e_t = X_t - \sum_{k=1}^p \hat{a}_k X_{t-k},$$

où  $\hat{a}_1, \dots, \hat{a}_p$  sont des estimateurs de Yule-Walker. Si le modèle est correct,  $\{e_t\}$  est un BB.

(ii) Calculer les ACF empiriques de  $\{e_t\}$  :

$$\widehat{\rho}_e(1), \dots, \widehat{\rho}_e(K), \quad K > p \text{ fixé.}$$

Sous  $H_0$ ,  $\rho_\varepsilon(1) = \dots = \rho_\varepsilon(K) = 0$ . Si  $T$  est grand,

$$\sqrt{T} \begin{pmatrix} \widehat{\rho}_\varepsilon(1) - \rho_\varepsilon(1) \\ \widehat{\rho}_\varepsilon(K) - \rho_\varepsilon(K) \end{pmatrix} \sim \mathcal{N}_K(0, I_{K \times K}) \implies T \sum_{j=1}^K \widehat{\rho}_\varepsilon^2(j) \sim \chi_K^2.$$

En fait, on cherche à tester la nullité des  $K$  premières autocorrélations.

(iii) Calculer  $\widehat{Q}_2 = T \sum_{j=1}^K \widehat{\rho}_e^2(j)$ . Ici, les  $\{e_t\}$  sont corrélés, donc correction (Test de portmanteau).

(iv) Si  $\widehat{Q}_2 > \chi_{K-p, 1-\alpha}^2$ , on rejette  $H_0$  au niveau  $\alpha$ .

**N.B :** Choix de  $K$  ni trop petit ( on veut tester à la fois beaucoup de retards  $j$ ) ni trop grand (estimateurs  $\widehat{\rho}_e^2(K)$  mauvaise). On prend habituellement  $K \approx \sqrt{T}$ .

**Critique de ce test:** Il rejette les modèles insuffisants ( $p$  trop petit), mais il est faible en discriminant entre plusieurs modèles plausibles. Pour le test d'un ARMA( $p, q$ ), on remplace  $\chi_{K-p}^2$  par  $\chi_{K-(p+q)}^2$ .

### 1.6.3 Quelques notes sur les tests

a) **Test de racine unité (Test de Dickey-Fuller)**

$X_t$  suit un processus autoregressif d'ordre 1 :

$$X_t = \varphi X_{t-1} + \varepsilon_t, \quad |\varphi| \leq 1, \quad \varepsilon_t \text{ BB}, \quad \text{var}(\varepsilon_t) = \sigma_\varepsilon^2. \quad \text{On veut tester : } (H_0) : \varphi = 1.$$

On considère le rapport

$$\begin{aligned} \widehat{t} &= \frac{(1 - \widehat{\varphi}_T)}{\left( \frac{\widehat{\sigma}_\varepsilon^2}{\sum_{t=1}^T X_{t-1}^2} \right)^{\frac{1}{2}}} \\ &= \frac{T(1 - \widehat{\varphi}_T) \left( \sum_{t=1}^T X_{t-1}^2 / T^2 \right)^{\frac{1}{2}}}{\widehat{\sigma}_\varepsilon} \quad (\text{cf}[5]) \end{aligned}$$

$$\widehat{\sigma}_\varepsilon^2 = \frac{1}{T} \sum_{t=1}^T (X_t - X_{t-1})^2. \quad \text{On a } \frac{1}{T} \sum_{t=1}^T (X_t - \widehat{\varphi}_T X_{t-1})^2 \xrightarrow{(H_0)} \sigma^2$$

On a un test de niveau asymptotique  $\alpha$ , de région critique  $\{\widehat{t} > \alpha\}$  avec respectivement pour  $\alpha = 1\%$ ,  $\alpha = 5\%$ ,  $\alpha = 10\%$ ,  $t_\alpha = 2.58$ ; 1.95; 1.62 .

b) **Test de Portmanteau (Box-Pierce)**

Statistique du test :  $Q = T \sum_{h=1}^K \widehat{\varphi}_h^2[\widehat{\varepsilon}]$ , où  $\widehat{\varphi}_h^2[\widehat{\varepsilon}]$  est la corrélation empirique entre les résidus (distants de  $h$ ). Sous l'hypothèse d'indépendance des erreurs  $\varepsilon_t$ ,  $Q$  suit asymptotiquement une loi du  $\chi^2(K - p - q)$  (on est dans un ARMA ( $p, q$ )).

Si  $Q > \chi^2(K - p - q, 1 - \alpha)$ , on refuse l'hypothèse d'indépendance au niveau  $\alpha$ .

Une modification de cette formule est la statistique de Box-Pierce-Ljung :

$$Q^*(h) = T(T+1) \sum_{k=1}^h \frac{\hat{\varphi}_k^2}{T-k}$$

Si  $\{X_t\}$  est une série stationnaire satisfaisant l'égalité :

$$X_t = c + \sum_{i=1}^q \psi_i \varepsilon_{t-i}$$

où  $\varepsilon_{t-i}$  est BB gaussien, alors  $\hat{\varphi}_l$  est approximativement normalement distribué, de moyenne 0 de variance  $\frac{1+2 \sum_{i=1}^q \varphi_i^2}{T}$ . Ce résultat permet de tester  $H_0 : \varphi_l = 0$  contre  $H_1 : \varphi_l \neq 0$ .

## 1.7 Quelques notes sur l'I.P.I

### 1.7.1 Objet et caractéristique de base

A quoi sert l'indice de la production industrielle (I.P.I)? C'est un instrument statistique qui mesure l'évolution de la production industrielle.

Les agents économiques concernés sont les entreprises, plus principalement les entreprises industrielles exerçant leur activité dans les usines, des chantiers, des carrières et des mines. Si l'on se réfère à la classification traditionnelle des activités économiques en trois secteurs, l'I.P.I. se rapporte au seul secteur secondaire et laisse à d'autres instruments statistiques la description des secteurs primaires (agriculture, pêche, sylviculture) et tertiaire (transports, commerces, services, administration).

Ce que l'I.P.I. se propose de mesurer, c'est ce qui sort en quantités physiques des usines à chaque période (mois, trimestres, années).

L'évolution de la production industrielle en francs courants résulte à la fois de variations des quantités produites et de variations des prix à la production des entreprises: on veut que l'I.P.I ne réagisse qu'aux mouvements sur les quantités produites, et en aucun cas aux mouvements des prix. Aussi les unités de compte recherchées pour suivre la production seront-elles des nombres de pièces ou d'appareils, des mètres de fil, des mètres carrés de tissus, des tonnes de ciment, des litres d'huile, des Kwh d'électricité.

L'indice de la production industrielle actuellement calculé à l'I.N.S. (Institut National de la Statistique) a pour base le trimestre moyen de l'année 1995/1996.

La mesure de la production trimestrielle de branches industrielles ayant des rythmes de fabrication aussi différents les uns des autres constitue l'un des défis majeurs posés par l'indice de la production industrielle.

L'avantage de la présentation sous forme d'indice est de rendre comparable les évolutions de la production de deux familles de produits industriels différents. Un autre avantage du suivi de la production industrielle à l'aide d'indices est de fournir un ensemble hiérarchisé d'évolutions, partant des familles de produits industriels les plus détaillés pour aboutir à l'ensemble de l'industrie, en passant par tous les niveaux intermédiaires souhaités en termes de grandes branches, de branches ou de sous-branches.

### 1.7.2 Utilisation des indices

L'I.P.I autorise à la fois la lecture en termes de cycle économique et en termes d'évolution structurelle : la production de chaque famille élémentaire de produits industriels est retracée sous la forme d'une série chronologique mettant en évidence aussi bien une tendance de fond que des infléchissements cycliques, l'ensemble de ces deux composantes étant appelés "tendance-cycle". L'I.P.I. fournit, par construction, aussi bien le suivi de la production de chaque série-temoin que de celle des branches industrielles et de l'ensemble de l'industrie. L'indice d'ensemble est l'un des quatre grands indicateurs macro-économiques de santé économique, avec les prix, le commerce extérieur et l'emploi.

# MODÉLISATION ET DÉSAISONNALISATION

---

La désaisonnalisation d'une série consiste essentiellement à utiliser certaines techniques mathématiques afin d'expurger la série des variations infra-annuelles périodiques.

L'utilisation du terme saisonnalité provient du fait que les saisons exercent une influence sur l'activité économique et sociale. Pour s'en convaincre, il suffit de vivre au Cameroun pendant une année, où le comportement des agents économiques diffère considérablement dans des secteurs tels que la production agricole, la construction, le transport, les déplacements touristiques, selon la période de l'année.

L'hypothèse sous-jacente à la désaisonnalisation est que la série originale de départ, appelée aussi série brute ou non désaisonnalisée, peut être décomposée en trois composantes : **saisonnaire**, **tendance-cycle** et **irrégulière**. La désaisonnalisation consiste alors à extraire la composante saisonnière de la série brute.

Un modèle de série chronologique est une équation précisant la façon dont ces trois composantes s'articulent les unes par rapport aux autres pour constituer la série chronologique. Il existe de très nombreux modèles, et parmi eux deux modèles classiques simples : le **modèle additif** et le **modèle multiplicatif**.

Il existe plusieurs techniques de désaisonnalisation pour ces deux modèles.

## 2.1 Modèle additif

### 2.1.1 Désaisonnalisation par régression linéaire

#### Présentation du modèle linéaire

La série  $X_t$  est la somme de deux composantes déterministes: une tendance -cycle  $Z_t$ , une saisonnalité  $S_t$  et une composante aléatoire  $\varepsilon_t$ .

$$X_t = Z_t + S_t + \varepsilon_t.$$

Nous supposons que  $Z_t$  et  $S_t$  sont des combinaison linéaires des fonctions connues dans le temps,  $Z_t^i$ ,  $S_t^j$ .

$$X_t = \sum_{i=1}^m Z_t^i \beta_i + \sum_{j=1}^n S_t^j \gamma_j + \varepsilon_t, \quad t = 1, \dots, T.$$

Le but est d'estimer les  $\beta_i$  et les  $\gamma_j$  à partir de  $T$  observations.

**Hypothèses sur les erreurs**

(H) :  $E\varepsilon_t = 0$ ,  $Var(\varepsilon_t) = \sigma^2$ ,  $Cov(\varepsilon_t, \varepsilon_{t-h}) = 0$ ,  $\forall h > 0$ .

**Composante saisonnière du modèle**

On définit :

$$S_t^j = \begin{cases} 1 & \text{si } t = \text{mois } j \\ 0 & \text{si } t \neq \text{mois } j \end{cases}$$

ou plus généralement

$$S_t^j = \begin{cases} 1 & \text{si } t = kj, k \in \mathbb{Z} \\ 0 & \text{sinon} \end{cases}$$

Les  $S_t^j$  sont des fonctions indicatrices.

**Composante tendancielle**

Elle a généralement une forme simple, reflétant la croissance moyenne.

**Exemple 2.1.** *Tendance linéaire* :  $Z_t = \beta_1 + \beta_2 t$ ,  $Z_t^1 = 1$ ,  $Z_t^2 = t$ .

**Estimateurs des moindres carrés ordinaires (MCO)**

$$X_t = \sum_{i=1}^m Z_t^i \beta_i + \sum_{j=1}^n S_t^j \gamma_j + \varepsilon_t, \quad t = 1, \dots, T.$$

La méthode des MCO consiste à choisir les  $\beta_i$  et les  $\gamma_j$  de façon à minimiser la somme des carrés des erreurs :

$$(\hat{\beta}_i, \hat{\gamma}_j) = \arg \min \left\{ \sum_{t=1}^T \varepsilon_t^2 \right\}.$$

Notons  $\beta = (\beta_1, \dots, \beta_m)'$ ,  $\gamma = (\gamma_1, \dots, \gamma_n)'$ , où  $A'$  est la transposée de  $A$ .

$$Z = [Z_t^i]_{\substack{i=1, \dots, m \\ t=1, \dots, T}} \quad S = [S_t^j]_{\substack{j=1, \dots, n \\ t=1, \dots, T}}$$

Le modèle s'écrit sous forme matricielle :

$$X = Z\beta + S\gamma + \varepsilon = \begin{bmatrix} Z & S \end{bmatrix} \begin{bmatrix} \beta \\ \gamma \end{bmatrix} + \varepsilon = Yb + \varepsilon.$$

$\hat{b} = (\hat{\beta}, \hat{\gamma})'$  vérifie :  $Y'Y\hat{b} = Y'X$ , soit :

$$\begin{bmatrix} Z & S \end{bmatrix} \begin{bmatrix} Z' \\ S' \end{bmatrix} \hat{b} = \begin{bmatrix} Z' \\ S' \end{bmatrix} X \text{ et donc : } \begin{bmatrix} \hat{\beta} \\ \hat{\gamma} \end{bmatrix} = \begin{bmatrix} Z'Z & Z'S \\ S'Z & S'S \end{bmatrix}^{-1} \begin{bmatrix} Z'X \\ S'X \end{bmatrix}.$$

D'où les coefficients estimés :

$$\begin{cases} \hat{\beta} = [Z'Z - Z'S(S'S)^{-1}S'Z]^{-1} [Z'X - Z'S(S'S)^{-1}S'X] \\ \hat{\gamma} = [S'S - S'Z(Z'Z)^{-1}Z'S]^{-1} [S'X - S'Z(Z'Z)^{-1}Z'X]. \end{cases}$$

**Exemple : Modèle trimestriel de BUYS-BALLOT**

$$X_t = \beta_1 + \beta_2 t + S_t^1 \gamma_1 + S_t^2 \gamma_2 + S_t^3 \gamma_3 + S_t^4 \gamma_4 + \varepsilon_t.$$

Les  $S_t^j$  sont les composantes trimestrielles :

$$\begin{aligned} S_t^1 &= (1,0,0,0,1,0,0,0,1, \dots) \\ S_t^2 &= (0,1,0,0,0,1,0,0,0,1, \dots) \\ S_t^3 &= (0,0,1,0,0,0,1,0,0,0,1, \dots) \\ S_t^4 &= (0,0,0,1,0,0,0,1,0,0,0,1, \dots). \end{aligned}$$

La contrainte (\*):  $\gamma_1 + \gamma_2 + \gamma_3 + \gamma_4 = 0$  impose l'unicité de la décomposition. Pour expliciter les différents coefficients, on résoud :

$$\left\{ \begin{array}{l} \min_{\beta, \gamma} \sum_{t=1}^T \left[ X_t - \beta_1 - \beta_2 t - \sum_{j=1}^4 S_t^j \gamma_j \right]^2 \\ \text{sous la contrainte (*)} \end{array} \right. ;$$

et, sachant que  $\forall t, \sum_{j=1}^{j=4} S_t^j = 1$ , ce système s'écrit :

$$\left\{ \min_{\beta_2, \delta} \sum_{t=1}^T \left[ X_t - \beta_2 t - \sum_{j=1}^4 S_t^j \delta_j \right]^2 \quad \text{où} \quad \left\{ \begin{array}{l} \beta_1 = [\delta_1 + \delta_2 + \delta_3 + \delta_4] / 4 \\ \gamma_j = \delta_j - \beta_1. \end{array} \right. \right.$$

*Soit*  $N =$  nombre d'années entières,  
 $\tilde{x}_n =$  moyenne des  $X_t$  relatives à l'année  $n$ ,  
 $\bar{x}_j =$  moyenne des  $X_t$  relatives au trimestre  $j$ ,  
 $\bar{x} =$  moyenne de toutes les observations  $X_t$ .

Nous avons les estimations suivantes :

$$(1) \quad \widehat{\beta}_2 = 3 \frac{\sum_{n=1}^N n \tilde{x}_n - \frac{N(N+1)}{2} \bar{x}}{N(N^2 - 1)}$$

$$(2) \quad \widehat{\delta}_j = \bar{x}_j - [j + 2(N - 1)] \widehat{\beta}_2 \quad \text{pour } j = 1, 2, 3, 4.$$

Et finalement :

$$(3) \quad \left\{ \begin{array}{l} \widehat{\beta}_1 = [\widehat{\delta}_1 + \widehat{\delta}_2 + \widehat{\delta}_3 + \widehat{\delta}_4] / 4 \\ \widehat{\gamma}_j = \widehat{\delta}_j - \widehat{\beta}_1. \end{array} \right.$$

**Application: Considérons la série: Total industries manufacturières**

Nous avons la distribution suivante ( tableau TAB 2.1 ):

Graphique des données( cf fig 2.1)

Considérons un modèle de la forme :

$$X_t = \beta_1 + \beta_2 t + S_t^1 \gamma_1 + S_t^2 \gamma_2 + S_t^3 \gamma_3 + S_t^4 \gamma_4 + \varepsilon_t.$$

$n \setminus j$	1	2	3	4	$\tilde{x}_n$
1996	104.54	97.32	105.63	111.81	104.825
1997	127.57	111.89	108.01	117.83	116.325
1998	138.45	125.70	93.41	113.58	117.785
1999	130.31	122.53	99.96	108.29	115.2725
2000	128.14	117.96	97.84	107.76	112.675
2001	130.75	116.60	93.41	113.58	113.585
2002	130.31	129.15	100.91	114.49	118.715
2003	138.56	120.76	112.37	137.18	127.2175
2004	163.54	127.49	120.65	144.99	139.1675
$\tilde{x}_j$	132.463	118.82	103.5766	118.834	118.4241

TAB. 2.1 – Moyennes des indices par trimestre  $j$  et par année  $n$  (industries manufacturières)

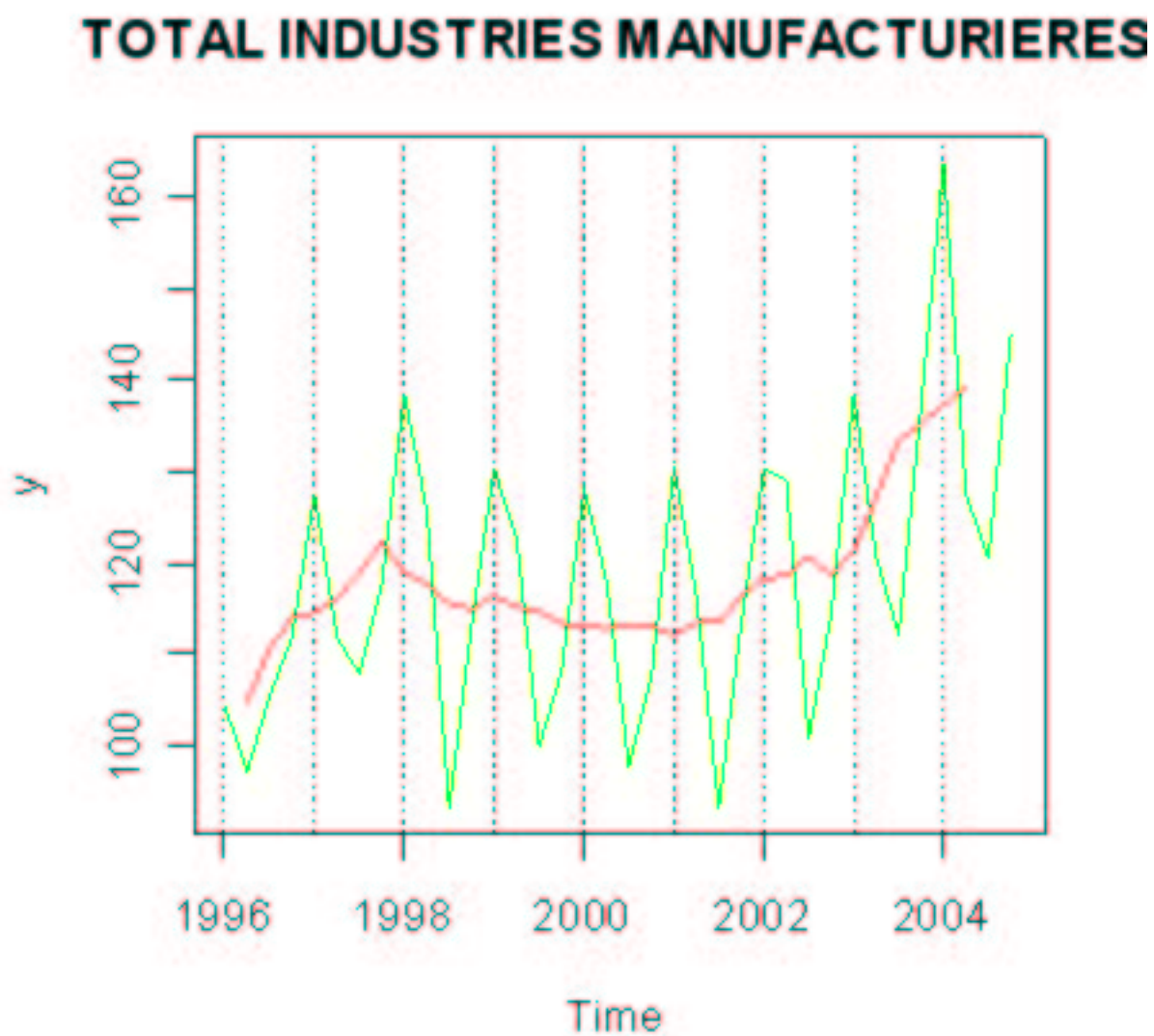


FIG. 2.1 – Evolution chronologique des Industries manufacturières

Nous utilisons (1), (2) et (3) pour calculer  $\hat{\beta}_1, \hat{\beta}_2, \hat{\gamma}_1, \hat{\gamma}_2, \hat{\gamma}_3$  et  $\hat{\gamma}_4$ . On obtient:

$$\begin{aligned}\hat{\beta}_1 &= 105.303, \hat{\beta}_2 = 0.71, \\ \hat{\gamma}_1 &= 15.103, \hat{\gamma}_2 = 0.753, \\ \hat{\gamma}_3 &= -15.202, \hat{\gamma}_4 = -0.654.\end{aligned}$$

On vérifie bien que  $\hat{\gamma}_1 + \hat{\gamma}_2 + \hat{\gamma}_3 + \hat{\gamma}_4 = 0$ .

On écrit alors :

$$\begin{aligned}X_t &= \hat{\beta}_1 + \hat{\beta}_2 t + \hat{\gamma}_1 S_t^1 + \hat{\gamma}_2 S_t^2 + \hat{\gamma}_3 S_t^3 + \hat{\gamma}_4 S_t^4 + \hat{\varepsilon}_t \\ &= 105.303 + 0.71t + 15.103S_t^1 + 0.753S_t^2 - 15.202S_t^3 - 0.654S_t^4 + \hat{\varepsilon}_t\end{aligned}$$

La série ajustée est donnée par :

$$\hat{X}_t = X_t - \hat{\varepsilon}_t = \sum_{i=1}^m Z_t^i \hat{\beta}_i + \sum_{j=1}^4 S_t^j \hat{\gamma}_j.$$

La série corrigée des variations saisonnières (CVS) est donnée par :

$$\hat{Y}_t = X_t - \hat{S}_t = \sum_{i=1}^m Z_t^i \hat{\beta}_i + \hat{\varepsilon}_t$$

### Propriétés statistiques des estimateurs et prévision

Sous l'hypothèse  $E\varepsilon_t = 0$ , les estimations MCO sont sans biais :

$$E\hat{\beta}_i = \beta_i \text{ et } E\hat{\gamma}_j = \gamma_j.$$

La variance des estimateurs peut être estimée par :

$$\hat{V} \begin{pmatrix} \hat{\beta} \\ \hat{\gamma} \end{pmatrix} = \delta^2 \begin{bmatrix} Z'Z & Z'S \\ S'Z & S'S \end{bmatrix}^{-1} \text{ où } \delta^2 = \frac{1}{T - n - m} \sum_{t=1}^T \hat{\varepsilon}_t^2.$$

Ceci permet d'obtenir des intervalles de confiance sur les estimateurs des paramètres.

**Prévision à un horizon  $h$**  : Soit  $h \geq 1$ . Nous supposons que le modèle reste valide en  $T + h$  :

$$X_{T+h} = \sum_{i=1}^m Z_{T+h}^i \beta_i + \sum_{j=1}^n S_{T+h}^j \gamma_j + \varepsilon_{T+h},$$

avec  $E(\varepsilon_{T+h}) = 0$ ,  $Var(\varepsilon_{T+h}) = \sigma^2$ ,  $Cov(\varepsilon_t, \varepsilon_{T+h}) = 0$ ,  $t = 1, \dots, T$ .

$X_{T+h}$  peut être approchée par

$$\hat{X}_T(h) = \sum_{i=1}^m Z_{T+h}^i \hat{\beta}_i + \sum_{j=1}^n S_{T+h}^j \hat{\gamma}_j.$$

Un intervalle de confiance de cette prévision est de la forme :

$$\left[ \hat{X}_T(h) - \phi_{1-\frac{\alpha}{2}} \sqrt{\hat{e}_h}; \hat{X}_T(h) + \phi_{1-\frac{\alpha}{2}} \sqrt{\hat{e}_h} \right]$$

où  $\phi_{1-\frac{\alpha}{2}}$  est le quantile d'ordre  $\alpha$  de la loi de Student à  $T - m - n$  de degré de liberté et

$$\hat{e}_h = \hat{E} \left( \left[ \hat{X}_T(h) - X_{T+h} \right]^2 \right).$$

### 2.1.2 Désaisonnalisation par la méthode des moyennes mobiles

Soit  $X_t = Z_t + S_t + \varepsilon_t, t = 1, \dots, T$ .

Le but est de trouver une transformation du processus  $X_t$  qui annule la composante saisonnière  $S_t$ . On cherche un "filtre"  $\phi$  tel que

$$Y_t = \phi(X_t) = Z_t + \varepsilon_t.$$

#### \*Généralités sur les moyennes mobiles

**Notion d'opérateur retard  $L$**  (rappel)

$$L : X_t \mapsto L(X_t) = LX_t = X_{t-1}, \quad L = \text{Lack ou } B = \text{backward.}$$

**Opérateur avance  $F$** :

$$F : X_t \mapsto F(X_t) = FX_t = X_{t+1} \quad F = \text{forward.}$$

**Définition 2.1.** Une moyenne mobile est un opérateur linéaire, combinaison linéaire finie d'opérateurs retard:  $M = \sum_{i=-m_1}^{m_2} \theta_i L^{-i}, m_1, m_2 \in \mathbb{N}, \theta_i \in \mathbb{R}$ , qui s'écrit aussi:

$$\begin{aligned} M &= L^{m_1} \sum_{i=0}^{m_1+m_2} \theta_{i-m_1} L^{-i} \\ &= L^{m_1} \sum_{i=0}^{m_1+m_2} \theta_{i-m_1} F^i = L^{m_1} \theta(F), \end{aligned}$$

$m_1 + m_2 + 1 =$  ordre de  $M$ , degré de  $\theta(\cdot) = m_1 + m_2$ ,  $\theta(\cdot)$  est un polynome

Si  $m_1 = m_2 = m$ , la moyenne mobile sera dite "centrée". De plus, si  $\theta_i = \theta_{-i}$ , alors la moyenne mobile est dite "symétrique".

**Exemple 2.2.**  $M(X_t) = MX_t = (X_t + X_{t-1})/2$ ;  $M = (L + I)/2 = L(I + F)/2$  est de degré 1, d'ordre 2 et n'est ni centré, ni symétrique.

**Remarque 2.1.1.** : Les moyennes centrées symétriques, sont nécessairement d'ordre impair. Pour  $m$  impair, on considère les moyennes mobiles d'ordre  $m = 2p + 1$  définies par:

$$M_m(X_t) = \frac{1}{2p+1} [X_{t-p} + X_{t-p+1} + \dots + X_t + \dots + X_{t+p+1} + X_{t+p}].$$

$M_m$  est centrée et Symétrique.

**Exemple 2.3.**  $M_3(X_t) = \frac{1}{3} [X_{t-1} + X_t + X_{t+1}]$ ; coefficients:  $\frac{1}{3}, \frac{1}{3}, \frac{1}{3}$ .

Pour  $m$  pair, on peut construire des moyennes mobiles centrées et symétriques comme suit: ( $m = 2p$ )

$$M_{2p}(X_t) = \frac{1}{2p} \left[ X_{t-p+\frac{1}{2}} + \dots + X_{t-\frac{1}{2}} + X_{t+\frac{1}{2}} + \dots + X_{t+p-\frac{1}{2}} \right],$$

où  $X_{t-j-\frac{1}{2}} = \frac{1}{2}(X_{t-j-1} + X_{t-j})$  et  $X_{t-j+\frac{1}{2}} = \frac{1}{2}(X_{t-j+1} + X_{t-j}), j \in [-(p-1); p-1]$ , d'où:

$$\begin{aligned} M_{2p}(X_t) &= \frac{1}{2p} \left[ \frac{1}{2}(X_{t-p} + X_{t-p+1}) + \dots + \frac{1}{2}(X_{t-1} + X_t) + \dots + \frac{1}{2}(X_{t+p-1} + X_{t+p}) \right] \\ &= \frac{1}{2p} \left[ \frac{1}{2}X_{t-p} + X_{t-p+1} + \dots + X_{t-1} + X_t + X_{t+1} + \dots + X_{t+p-1} + \frac{1}{2}X_{t+p} \right] \end{aligned}$$

**Exemple 2.4.** : La moyenne mobile  $2 \times 4$  : elle permet d'estimer des tendances dans le cas de données trimestrielles : elle est d'ordre 5 et de coefficients  $\frac{1}{8}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{8}$ .

$M_{2 \times 4}(X_t) = \frac{1}{8} [X_{t-2} + 2X_{t-1} + 2X_t + 2X_{t+1} + X_{t+2}]$ . Elle élimine les saisonnalités trimestrielles, elle conserve les tendances linéaires, et réduit de 75% la variance d'un bruit blanc. (cf [3])

**Exemple 2.5. Moyenne mobile différence :**  $\Delta_p = (I - L)^p$

Elle transforme un polynôme de degré  $k \leq p$  en une constante. Elle permet d'éliminer une tendance qui serait un polynôme de bas degré.

**Moyenne mobile différence saisonnière :**  $\Delta_{p,s} = (I - L^s)^p$

Elle permet d'absorber les composantes saisonnières de période  $s$ .

**Remarque 2.1.2.** - Une moyenne mobile d'ordre impair peut s'écrire

$$M = \sum_{j=-q}^{+q} \theta_j L^{-j} \text{ où } q \in \mathbb{N} \setminus \{0\} \text{ et } \theta_j = \frac{1}{2q+1}.$$

- Une moyenne mobile d'ordre pair peut s'écrire :

$$M = \sum_{j=-q}^{+q} \theta_j L^{-j} \text{ où } q \in \mathbb{N} \setminus \{0\}, j \neq 0 \text{ et } \theta_j = \begin{cases} \frac{1}{2q} & \text{pour } |j| \neq q \\ \frac{1}{4q} & \text{sinon.} \end{cases}$$

**But des moyennes mobiles :** leur but est :

- (i) d'absorber les composantes saisonnières en laissant invariantes les tendances;
- (ii) de réduire la variance des perturbations (lissage des séries). (cf [3])

**Propriétés :**

- La composée de deux moyennes mobiles est une moyenne mobile ;
- La somme de deux moyennes mobiles est une moyenne mobile.

**\*Applications des moyennes mobiles**

**Cas simples non itératifs** Soit  $X_t$  la série brute,  $X_t$  est trimestrielle.

1°) On calcule la moyenne mobile centrée de  $X_t$  :

$$Y_t = \left( \frac{1}{2} X_{t-2} + X_{t-1} + X_t + X_{t+1} + \frac{1}{2} X_{t+2} \right) / 4;$$

2°) On calcule la différence  $dt = X_t - Y_t$ ;

3°) On calcule les indices saisonniers  $i_k$ : pour la période  $m$  ou  $q$ , c'est la moyenne des  $dt$  en n'utilisant que le trimestre  $q$ . Puis ajuster les indices saisonniers pour que leur somme soit égale à zéro :

$$S_q = i_q - \frac{1}{4} \sum_{k=1}^4 i_k,$$

La série  $X_t$  est supérieure à la série ajustée en période  $j$  de  $S_j$ ;

4°) la série CVS est donnée par  $X_t - S_j$ .

**Exemple d'une moyenne mobile composée : CENSUS X-11** Posons  $X_t = Z_t + S_t + \varepsilon_t$ ,  
 $Z_t$  =tendance,  $S_t$  =saisonnalité et  $\varepsilon_t$  =pertubation.  
 L'algorithme suivant permet de désaisonnaliser  $X_t$  :

étape 1: Estimation de la tendance par moyenne mobile:

$$Z_t^{(1)} = M(X_t).$$

étape 2: Estimation de la composante saisonnière et de la perturbation

$$\sum_t = S_t + \varepsilon_t, \quad \sum_t^{(1)} = X_t - Z_t^{(1)}.$$

étape 3: Estimation de la composante saisonnière par moyenne mobile

$$S_t^{(1)} = M' \left( \sum_t^{(1)} \right) \text{ et } \varepsilon_t^{(1)} = \sum_t^{(1)} - S_t^{(1)}.$$

Il s'agit ici de lisser les valeurs de la composante  $\sum_t$  de chaque mois pour extraire l'évolution du coefficient saisonnier du trimestre concerné. La moyenne mobile  $M'$  utilisée ici devra reproduire la composante saisonnière de chaque trimestre en réduisant au maximum la composante irrégulière. Une contrainte de normalisation des coefficients doit être imposée (somme nulle).

étape 4: Estimation de la série corrigée des variations saisonnières

$$\hat{X}_t = Z_t^{(1)} + \varepsilon_t^{(1)} = X_t - S_t^{(1)}.$$

On choisit  $M$  et  $M'$  de sorte que :

$$M_{2 \times 4} \text{ (ou } M_{2 \times 12}) : M' = M_{3 \times 5} = \text{moyenne mobile sur 7 termes de coefficients :}$$

$$\frac{1}{15}, \frac{2}{15}, \frac{3}{15}, \frac{3}{15}, \frac{3}{15}, \frac{2}{15}, \frac{1}{15}.$$

## 2.2 Modèle multiplicatif

Soit  $X_t = X_{i,j} = Y_t S_j \varepsilon_t$  une série où :

- $t$  = indice du trimestre dans l'ensemble de la série ( $t = 1, 2, \dots, T$ ),
- $i$  = indice de l'année ( $i = 1, 2, \dots, N$ ),
- $j$  = indice du rang du trimestre dans l'année ( $i = 1, 2, 3, 4$ ).

$t = 4(i - 1) + j$ , (pour une serie trimestrielle). Nous avons choisi de commencer la numérotation des années à 1.

- $Y_t$  = tendance ou valeur lissée ou valeur « filtrée » de la série,
- $S_j$  = coefficient spécifique au trimestre  $j$ ,
- $\varepsilon_t$  = coefficient d'aléa (variation résiduelle).

Nous partons de l'hypothèse que le multiplicateur saisonnier est stable, pour les trimestres de même nom d'une année à l'autre.

### 2.2.1 Lissage

Nous déterminons tout d'abord les valeurs de la série «lissée» ( $Y_t$ ) grâce à la méthode des moyennes mobiles d'ordre 4, ce qui fait perdre 4 points (2 en début et 2 en fin de période).

### 2.2.2 Calcul des multiplicateurs saisonniers perturbés

En divisant les données brutes par les données lissées, on obtient pour chaque trimestre une grandeur  $^*S_t$  qui fournit une première estimation du multiplicateur saisonnier, mais perturbée par les aléas car :

$$^*S_t = \frac{X_t}{Y_t} = \frac{Y_t S_j \cdot \varepsilon_t}{Y_t} = S_j \cdot \varepsilon_t$$

$$^*S_t = ^* S_{i,j}$$

### 2.2.3 Moyennes des coefficients des trimestres de même nom

L'étape suivante consiste à synthétiser les coefficients correspondants aux trimestres de même nom dans un coefficient unique.

En tenant compte de l'hypothèse de compensation des aléas sur plusieurs trimestres, on peut les éliminer en déterminant, pour chacun des quatre trimestres, la moyenne des coefficients observés :

$$\boxed{\widehat{S}_j = \frac{\sum_i ^* S_{i,j}}{N-1}} = \text{multiplicateurs moyens non corrigés.}$$

### 2.2.4 Neutralisation annuelle des coefficients saisonniers

On commence par déterminer la moyenne des coefficients calculés à l'étape précédente :

$$\mu = \frac{\sum_{j=1}^4 \widehat{S}_j}{4}.$$

Si cette moyenne  $\mu$  s'avère sensiblement différente de 1 (à  $10^{-3}$  près par exemple), on corrige chacun des quatre multiplicateurs provisoires en les divisant par la moyenne  $\mu$ .

**Multiplicateurs définitifs (corrigés) :**

$$\boxed{S_j = \frac{\widehat{S}_j}{\mu}, \quad j = 1, 2, 3, 4}.$$

On assure ainsi la neutralisation annuelle des quatre coefficients :

$$\frac{\sum_{j=1}^4 S_j}{4} = \frac{\sum_{j=1}^4 \frac{\widehat{S}_j}{\mu}}{4} = \frac{\sum_{j=1}^4 \widehat{S}_j}{4\mu} = \frac{\mu}{\mu} = 1.$$

### 2.2.5 Calcul de la série CVS

Pour déduire de la série brute la série corrigée des variations saisonnières (série CVS), il suffit de diviser chaque valeur observée par le coefficient du trimestre correspondant :

$$\boxed{{}_{(cvs)}X_t = {}_{(cvs)}X_{ij} = \frac{X_{ij}}{S_j} \quad (t = 1, 2, \dots, T)}$$

La série CVS correspond ainsi à la résultante des seules influences de la tendance et des mouvements aléatoires.

### 2.2.6 Vérification de la qualité du modèle par le calcul des valeurs ajustées

On peut vérifier la qualité de la procédure retenue en construisant ensuite la série des valeurs ajustées de la variable par le produit des valeurs tendanciennes par les coefficients trimestriels adéquats :

$$\boxed{X_{ajusté} : \widehat{X}_{ij} = \widehat{X}_t = Y_t \cdot S_j}$$

La série ajustée peut être comparée graphiquement à la série brute : les écarts entre elles ne doivent provenir que des mouvements accidentels et doivent être distribués de manière erratique. Si les écarts augmentent ou diminuent régulièrement, c'est que le modèle d'analyse a été mal choisi.

### 2.2.7 Projections (Prévisions)

- On déduit tout d'abord de la tendance passée ( les  $Y_t$  ) les valeurs tendanciennes futures projetées :  ${}_pY_{T+1}, {}_pY_{T+2}, \dots$ . Très souvent, on fait une simple extrapolation, c'est-à-dire la prolongation d'une droite si la tendance passée est à peu près linéaire.
- On ressaisonalise ensuite en effectuant le produit des valeurs de tendance par le multiplicateur trimestriel adéquat :

$$\boxed{X_{projeté} : {}_p\widehat{X}_{ij} = {}_p\widehat{X}_t = {}_pY_p \cdot S_j}$$

**Remarque 2.2.1.** *La qualité de la prévision de  $X$  dépend crucialemnt de la vraisemblance des hypothèses retenues pour la projection des  $Y_t$ . La projection est peu risquée si l'on a constaté une croissance régulière à taux constant dans le passé. Elle sera arbitraire si on a constaté des ruptures de rythme, des points d'inflexion sur la courbe tendancielle.*

**Remarque 2.2.2.** *Le prévisionniste prendra moins de risque en prolongeant la fin de la courbe, c'est-à-dire en prenant appui sur les données les plus récentes.*

### 2.2.8 Autre méthode de désaisonnalisation dans le cas du modèle multiplicatif

Soit la série  $X_t$ . Les données sont trimestrielles.

1°) On calcule la moyenne mobile centrée de  $X_t$  :

$$Y_t = \left( \frac{1}{2}X_{t-2} + X_{t-1} + X_t + X_{t+1} + \frac{1}{2}X_{t+2} \right) / 4.$$

2°) On calcule le ratio

$$r_t = \frac{X_t}{Y_t}.$$

3°) On calcule les indices saisonniers : pour la période  $q$ , c'est la moyenne des  $r_t$  en n'utilisant que le trimestre  $q$  (**ex** : tous les 1<sup>ers</sup> trimestres), puis on ajuste les indices saisonniers pour que leur produit soit égal à 1 :

$$S_q = \frac{i_q}{4\sqrt{i_1 i_2 i_3 i_4}}.$$

La série  $X_t$  est ainsi  $S_j\%$  supérieure à la série ajustée en période  $j$ .

4°) On divise  $X_t$  par  $S_j$  pour obtenir la série CVS.

## 2.2.9 Présentation d'un exemple: Série : "Total production industrielle"

### Calcul de la tendance $Y_t$

Nous allons nous limiter à la méthode exposée aux paragraphes ci-dessus.

Le cycle étant annuel, nous retiendrons une moyenne mobile sur quatre trimestres. Ce calcul ne fournit des résultats qu'à partir du troisième trimestre.

$$\begin{aligned} Y_3 &= \frac{X_1}{8} + \frac{X_2 + X_3 + X_4}{4} + \frac{X_5}{8} \\ Y_4 &= \frac{X_2}{8} + \frac{X_3 + X_4 + X_5}{4} + \frac{X_6}{8}, \dots \end{aligned}$$

et le dernier trimestre calculable est ici le 34<sup>ème</sup> ( $N - 2 = 34$ ).

$$Y_{34} = \frac{X_{32}}{8} + \frac{X_{33} + X_{34} + X_{35}}{4} + \frac{X_{36}}{8}.$$

Le résultats sont reportés dans le tableau TAB.2.2.

### Calcul des multiplicateurs saisonniers « perturbés » $^*S_t$

$$^*S_t = \frac{X_t}{Y_t}.$$

L'ensemble des résultats se trouve dans le tableau TAB.2.2

Les coefficients supérieurs à 1 correspondent à des trimestres de "haute saison" et ceux inférieurs à 1 caractérisent des trimestres de "basse saison".

### Moyennes par trimestre

On neutralise les variations accidentelles par le calcul des moyennes par trimestre :

$$1. \text{ pour le } 3^e \text{ trimestre : } \hat{S}_3 = \frac{^*S_3 + ^*S_7 + ^*S_{11} + \dots + ^*S_{31}}{8} = 0,871390496,$$

Année	N° du trimestre t	Série brute $X_t$	$MM_4 X_t = Y_t$	$*S_t = \frac{X_t}{Y_t}$	$cv_s X_t$	$X_{ajusté}$
1996	1	107.64			94.84	
1996	2	100.50			98.92	
1996	3	105.30	108.93	0.967	121.03	94.77
1996	4	113.21	112.85	1.003	115.64	110.48
1997	1	125.76	115.13	1.092	110.80	130.67
1997	2	113.78	116.45	0.977	111.99	118.31
1997	3	110.20	118.60	0.929	126.67	103.18
1997	4	118.89	121.25	0.981	121.44	118.70
1998	1	137.31	121.83	1.127	120.98	138.28
1998	2	123.44	122.30	1.009	121.50	124.26
1998	3	105.11	124.58	0.844	120.82	108.38
1998	4	127.81	127.55	1.002	130.55	124.87
1999	1	146.63	130.27	1.126	129.19	147.86
1999	2	137.87	130.45	1.057	135.70	132.54
1999	3	112.48	129.40	0.869	129.29	112.58
1999	4	121.85	128.45	0.949	124.46	125.75
2000	1	144.19	127.51	1.131	127.04	144.72
2000	2	132.73	127.14	1.044	130.64	129.17
2000	3	110.09	127.43	0.864	126.54	110.86
2000	4	121.25	127.61	0.950	123.85	124.93
2001	1	147.12	126.80	1.160	129.62	143.92
2001	2	131.21	127	1.033	129.14	129.03
2001	3	105.11	127.75	0.823	120.82	111.14
2001	4	127.81	129.45	0.987	130.55	126.73
2002	1	146.63	132.27	1.109	129.19	150.13
2002	2	145.32	133.45	1.089	143.03	135.59
2002	3	113.55	134.74	0.843	130.52	117.22
2002	4	128.83	134.73	0.956	131.59	131.90
2003	1	155.92	133.6	1.167	137.37	151.66
2003	2	135.89	134.99	1.007	133.75	137.15
2003	3	114.11	137	0.834	131.16	119.19
2003	4	139.27	136.99	1.019	142.26	134.11
2004	1	161.54	136.84	1.180	142.33	155.31
2004	2	127.83	138.84	0.921	125.82	141.06
2004	3	123.33			141.76	
2004	4	146.07			149.20	

TAB. 2.2 – Série CVS et Série ajustée des données(I.P.I général)

2. pour le 4<sup>e</sup> :  $\widehat{S}_4 = 0,980860281$ ,
3. pour le 1<sup>e</sup> :  $\widehat{S}_1 = 1,136489089$ ,
4. pour le 2<sup>e</sup> :  $\widehat{S}_2 = 1,017096168$ .

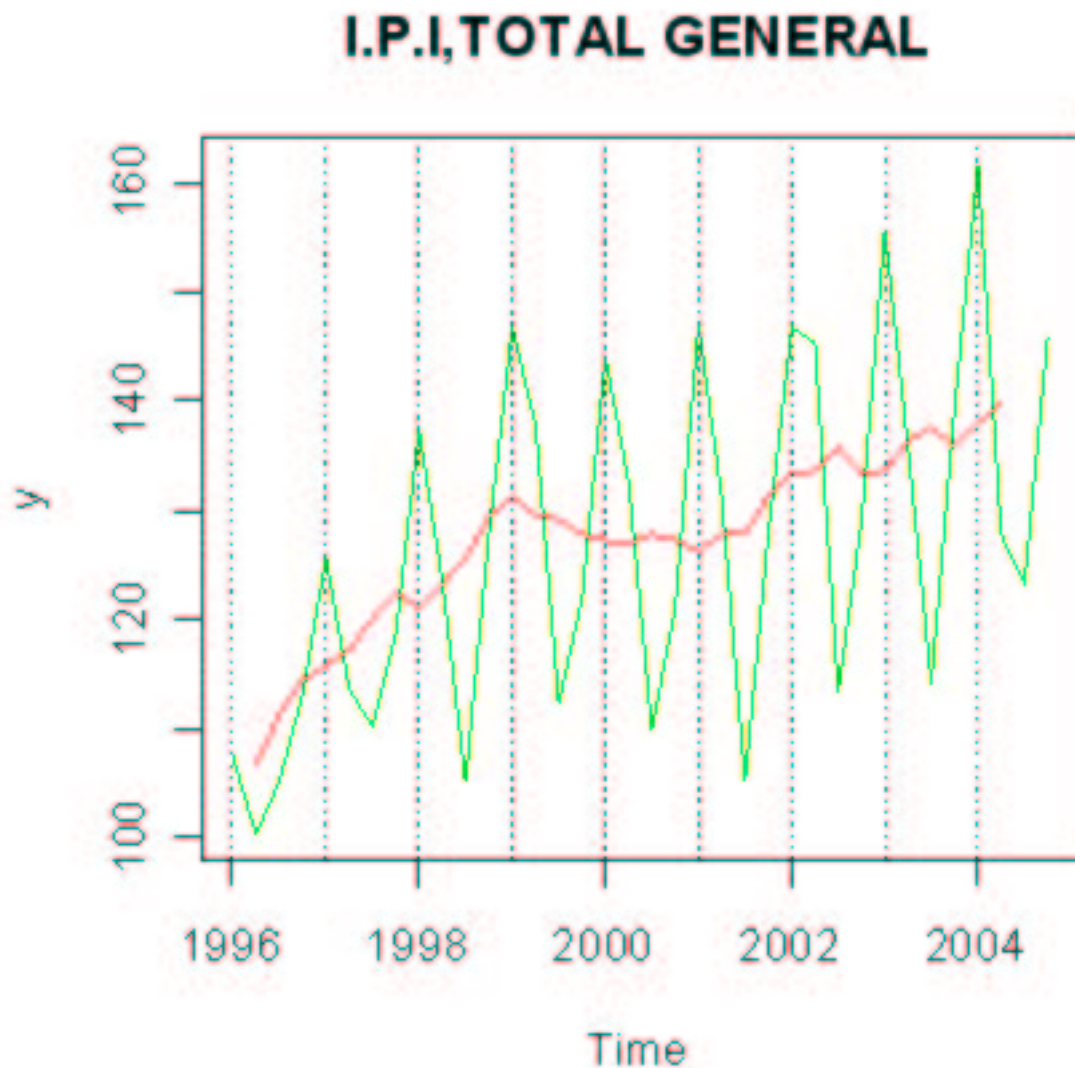


FIG. 2.2 - Evolution annuelle de I.P.I Général

D'où la moyenne :

$$\mu = \frac{\widehat{S}_1 + \widehat{S}_2 + \widehat{S}_3 + \widehat{S}_4}{4} = 1.001459$$

**Neutralisation annuelle des multiplicateurs**

On calcule les rapports :

$$S_j = \frac{\widehat{S}_j}{\mu}$$

On trouve:

$S_1$	$S_2$	$S_3$	$S_4$
1.135	1.016	0.870	0.979

Le premier trimestre correspond à une haute saison. On se trouve en moyenne 13.5% au-dessus de la tendance.

Le deuxième trimestre se situe en moyenne 1.6% au-dessus de la tendance. Le troisième correspond à une basse saison. On se trouve en moyenne 13% en-dessous de la tendance.

Idem pour le quatrième trimestre où on se trouve en moyenne 2% en-dessous de la tendance.

**Calcul des valeurs CVS et des valeurs ajustées (cf TAB.2.2).**

$${}_{CVS}X_t = \frac{X_t}{S_j}, \quad X_{ajusté} = \hat{X}_t = Y_t S_j$$

# APPLICATIONS

---

## 3.1 Série NO1 : TOTAL GENERAL I.P.I

### 3.1.1 Analyse exploratoire des données

y	Qtr1	Qtr2	Qtr3	Qtr4
1996	107.64	100.50	105.30	113.21
1997	125.76	113.78	110.20	118.89
1998	137.31	123.44	105.11	127.81
1999	146.63	137.87	112.48	121.85
2000	144.19	132.73	110.09	121.25
2001	147.12	131.21	105.11	127.81
2002	146.63	145.32	113.55	128.83
2003	155.92	135.89	114.11	139.27
2004	161.54	127.83	123.33	146.07

TAB. 3.1 – *Données brutes de l'I.P.I*

Graphique des données

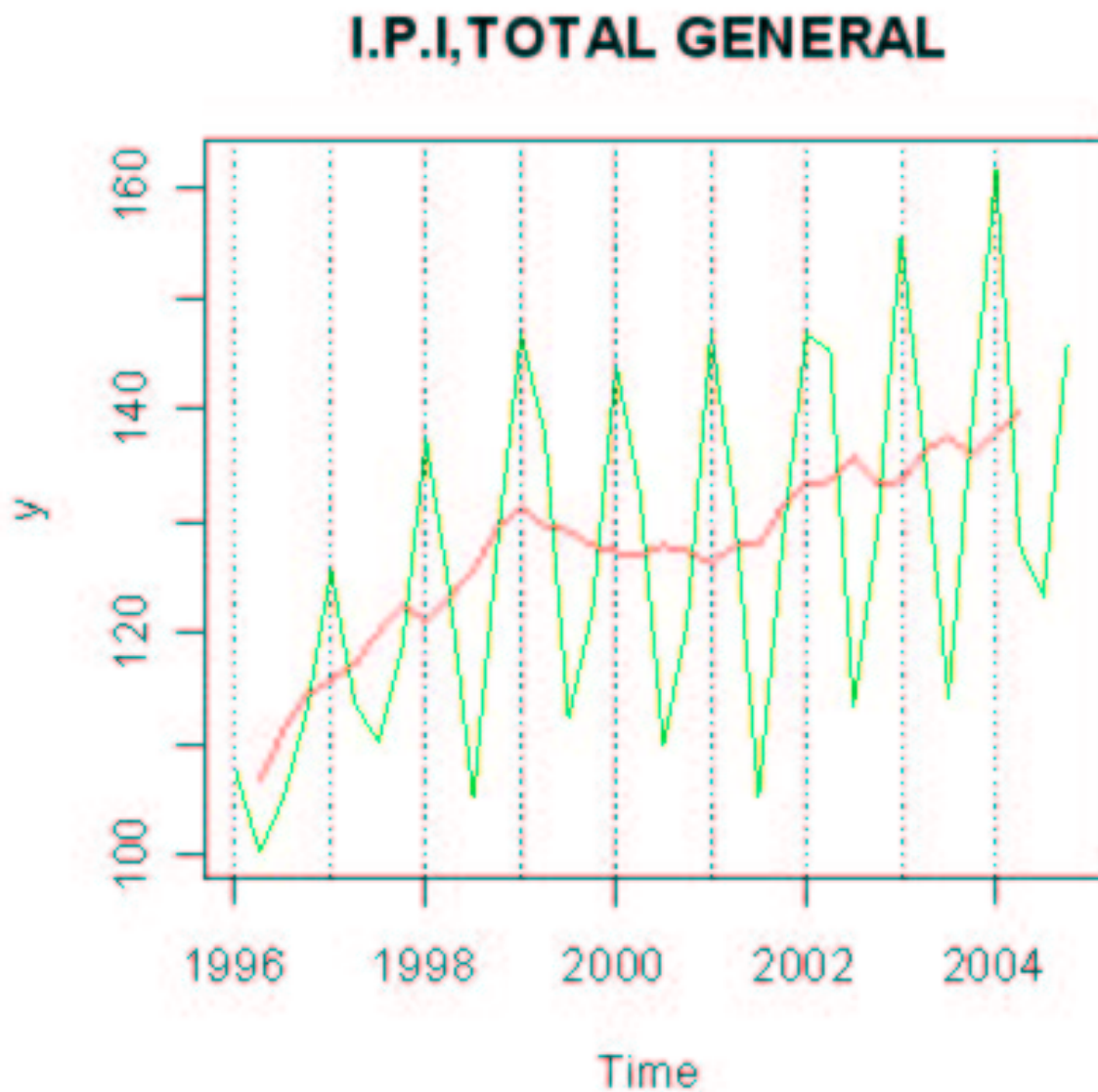
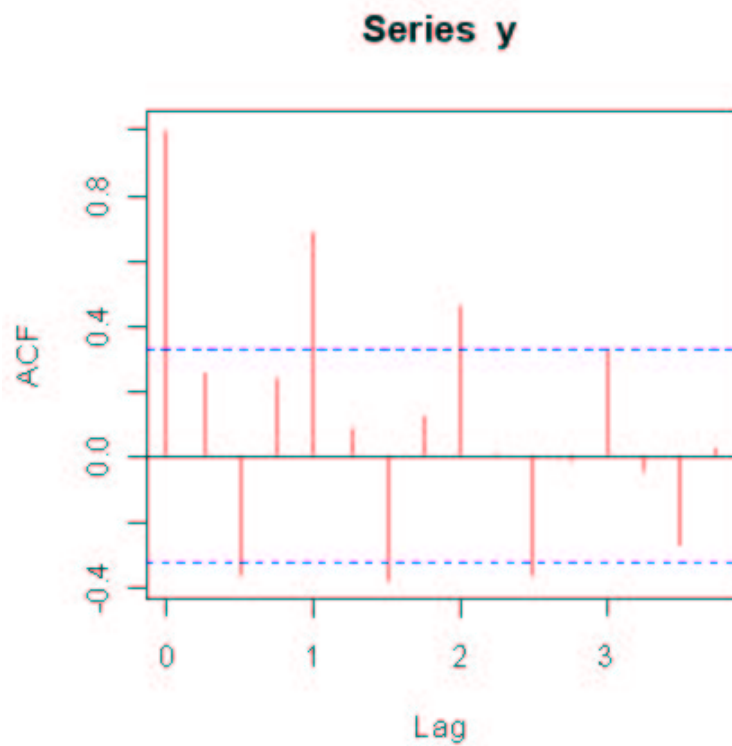


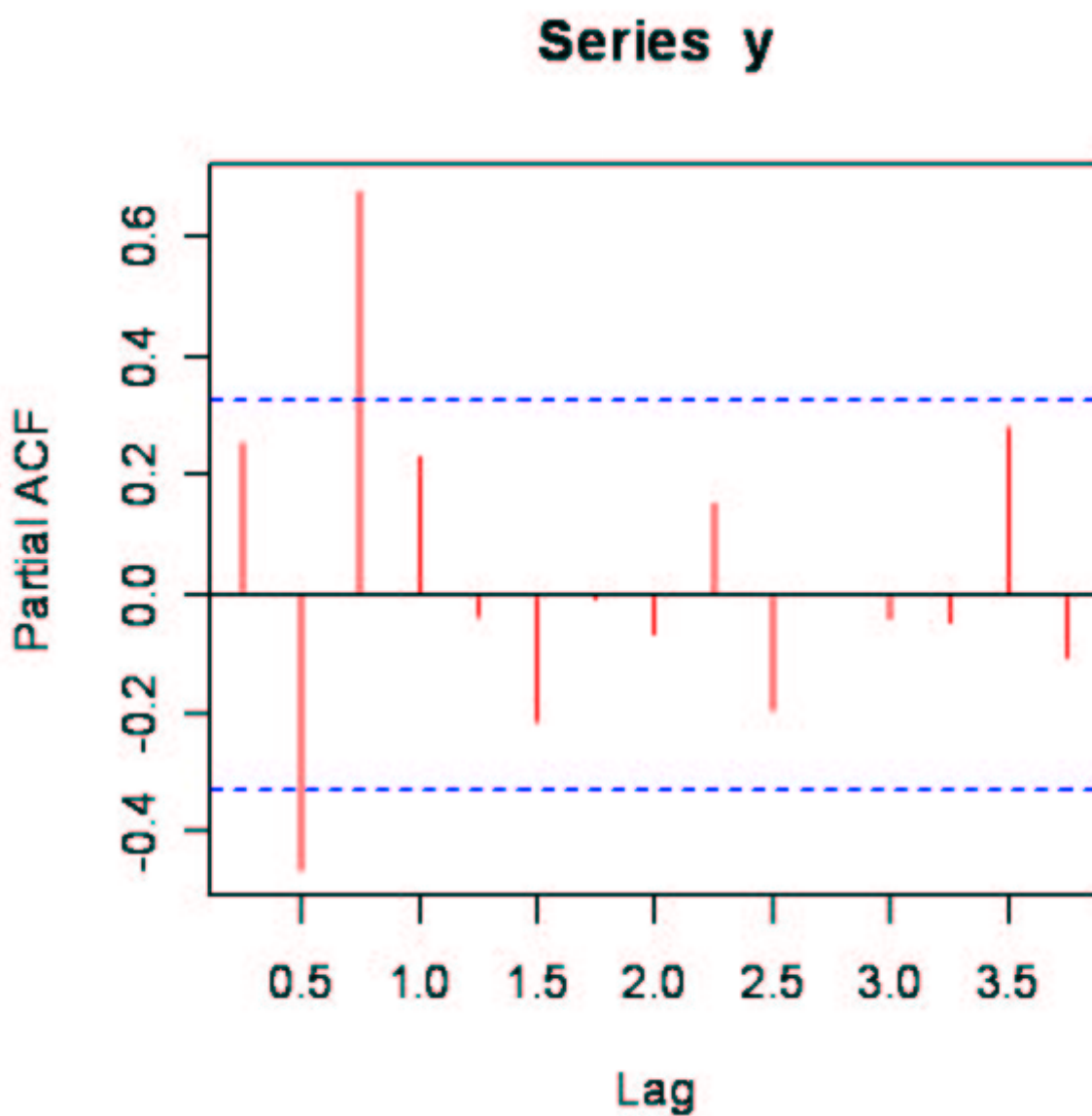
FIG. 3.1 – *Série Total Général I.P.I*

La série dont il est question d'étudier les variations est issue des observations trimestrielles de l'indice général de la production industrielle au Cameroun.

L'examen graphique de la série laisse entrevoir une tendance croissante et presque linéaire. La série présente une saisonnalité de période 4, avec des pics au 1<sup>er</sup> trimestre et des creux au 3<sup>eme</sup>. Par ailleurs, l'amplitude de ses fluctuations semble constante et cela suggère donc une décomposition additive du modèle.

## ACF ET PACF DE LA SERIE BRUTE

FIG. 3.2 – *Corrélogramme simple de y*

FIG. 3.3 – *Corrélogramme partiel de y*

Le corrélogramme de la série brute présente des valeurs significativement non nulles des fonctions d'autocorrélation et d'autocorrélation partielle pour des retards élevés. La série brute n'est donc pas stationnaire. La forte saisonnalité de cette série nous conduit à la désaisonnaliser. Nous notons alors  $Y_{CVS}$  la série désaisonnalisée.

### 3.1.2 Stationnarité et identification du modèle

#### Estimation des coefficients saisonniers

Nous utilisons les moyennes mobiles d'ordre 4

Qtr1	Qtr2	Qtr3	Qtr4
16.614133	1.396347	-16.383302	-1.627208

TAB. 3.2 – *Coefficients saisonniers estimés*

Le coefficient saisonnier au 1<sup>er</sup> trimestre est très élevé, on peut conclure que l'activité a tendance à être très élevée au cours de ce trimestre. Par contre cette activité est à la baisse aux 3<sup>eme</sup> et 4<sup>eme</sup> trimestres. Cela traduit bien évidemment la prédominance de l'agro industrie qui dépend ainsi du cycle agricole.

#### Série désaisonnalisée

ycvs	Qtr1	Qtr2	Qtr3	Qtr4
1996	91.02587	99.10365	121.68330	114.83721
1997	109.14587	112.38365	126.58330	120.51721
1998	120.69587	122.04365	121.49330	129.43721
1999	130.01587	136.47365	128.86330	123.47721
2000	127.57587	131.33365	126.47330	122.87721
2001	130.50587	129.81365	121.49330	129.43721
2002	130.01587	143.92365	129.93330	130.45721
2003	139.30587	134.49365	130.49330	140.89721
2004	144.92587	126.43365	139.71330	147.69721

TAB. 3.3 – Données désaisonnalisées de l'I.P.I Général

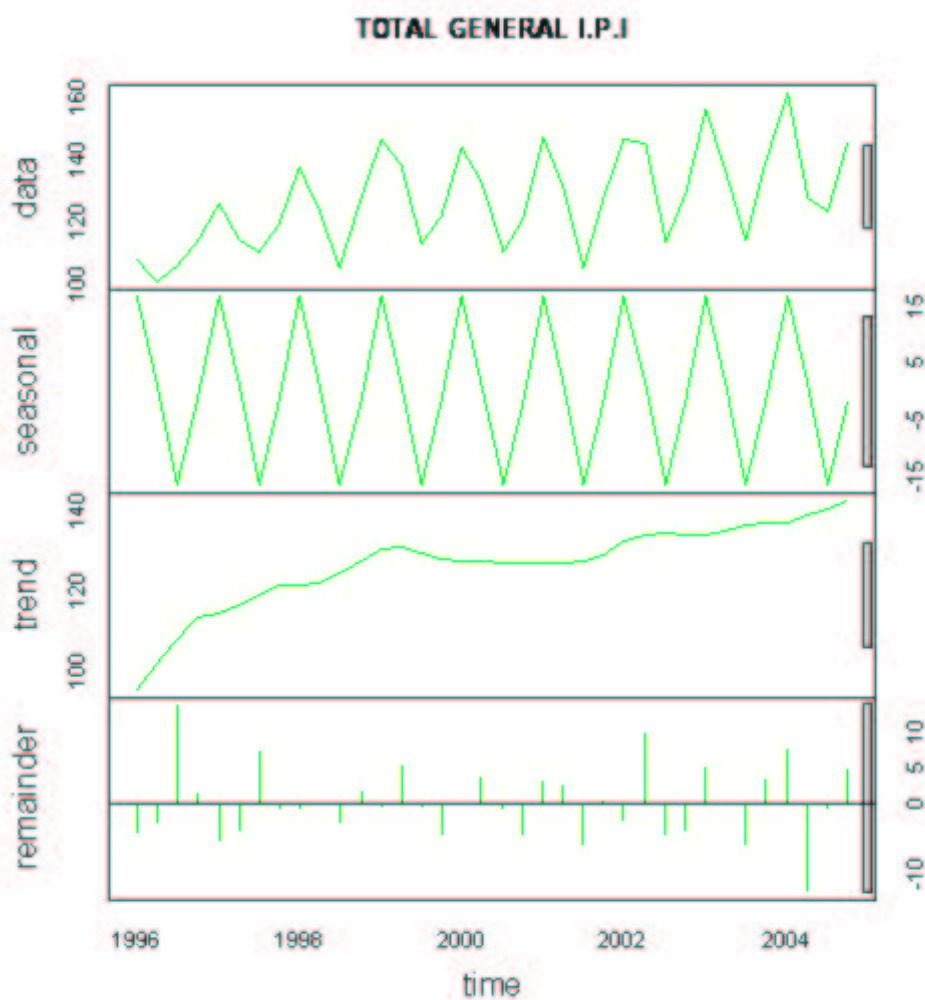


FIG. 3.4 – Série désaisonnalisée de l'I.P.I Général

ACF et PACF de la série désaisonnalisée

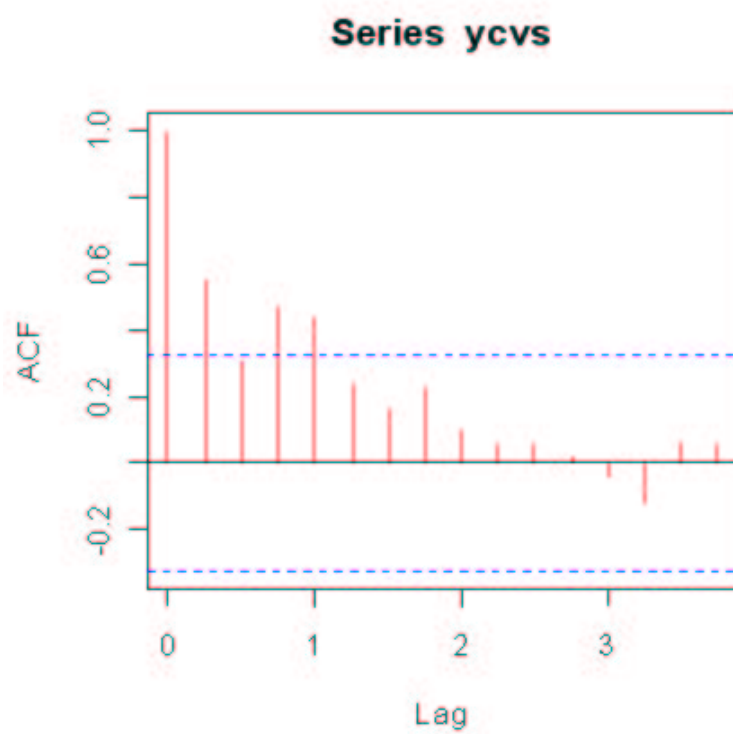
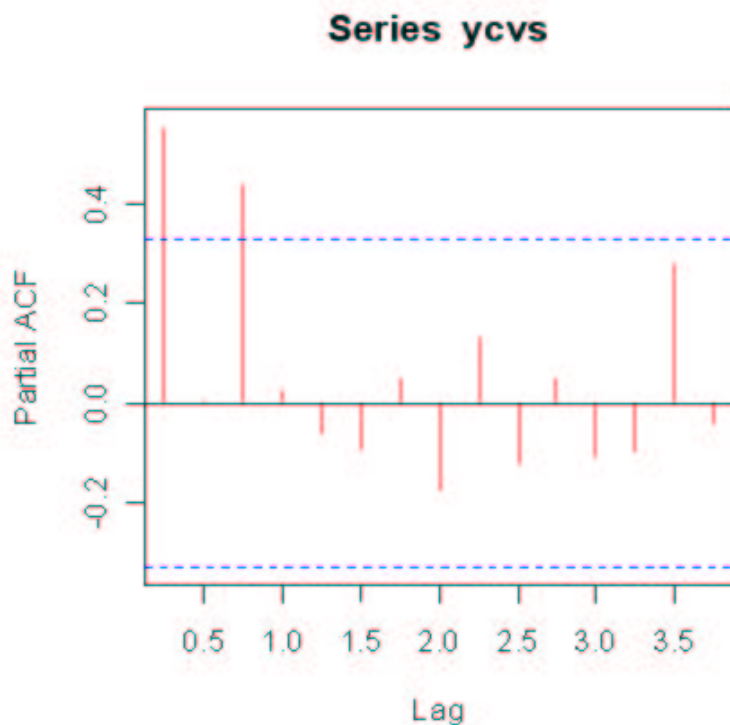


FIG. 3.5 – *Corrélogramme de ycvS*

FIG. 3.6 – *Corrélogramme partiel de ycvS*

Le corrélogramme de la série désaisonnalisée montre que celle-ci présente une forme ARMA(3,0), mais ne nous permet pas de conclure à la stationnarité de la série. Pour le vérifier nous allons effectuer le test de stationnarité de DICKEY-FULLER, ou plutôt sa variante de PHILLIPPE-PERRON.

### Test de stationnarité

*Phillips-Perron Unit Root Test data: ycvS*

*Dickey-Fuller = -5.6834, Truncation lag parameter = 3, p-value = 0.01.*

Le test de DICKEY-FULLER sur la série désaisonnalisée montre que celle-ci n'est pas stationnaire.

La  $p - value < 5\%$ , nous essaierons de stationnariser notre série ( en tendance).

Nous allons estimer notre tendance en utilisant la régression affine.

Coefficients:	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	110.6903	2.4884	44.482	< 2e-16 ***
t	0.8720	0.1173	7.435	1.27e-08 ***

TAB. 3.4 – *Estimation des coefficients de la droite de tendance*

*Residual standard error: 7.31 on 34 degrees of freedom Multiple R-Squared: 0.6192,  
Adjusted R-squared: 0.608  
F-statistic: 55.28 on 1 and 34 DF, p-value: 1.267e-08*

L'équation de la tendance est :  $f(t)=0.8720*t+110.6903$ .

### 3.1.3 Estimation et validation du modèle

$y_{stat}=ycvs-f(t)$

ACF et PACF des résidus(cf FIG.3.7et 3.8)

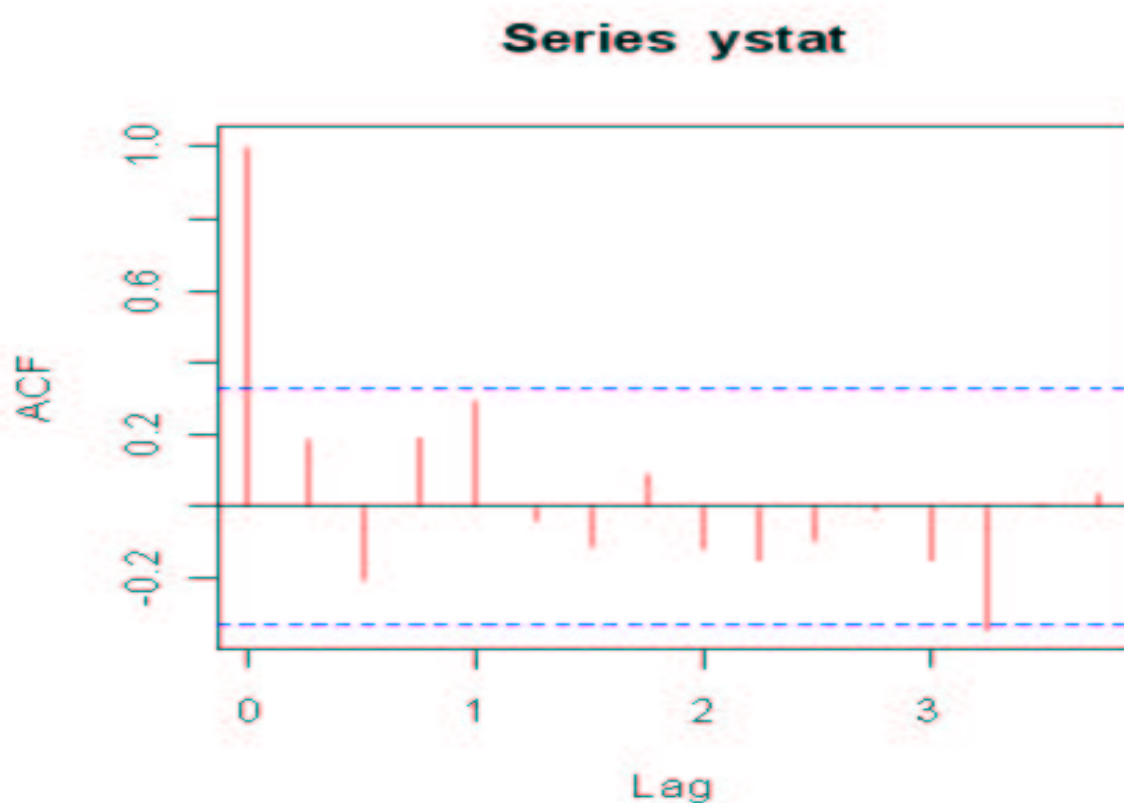
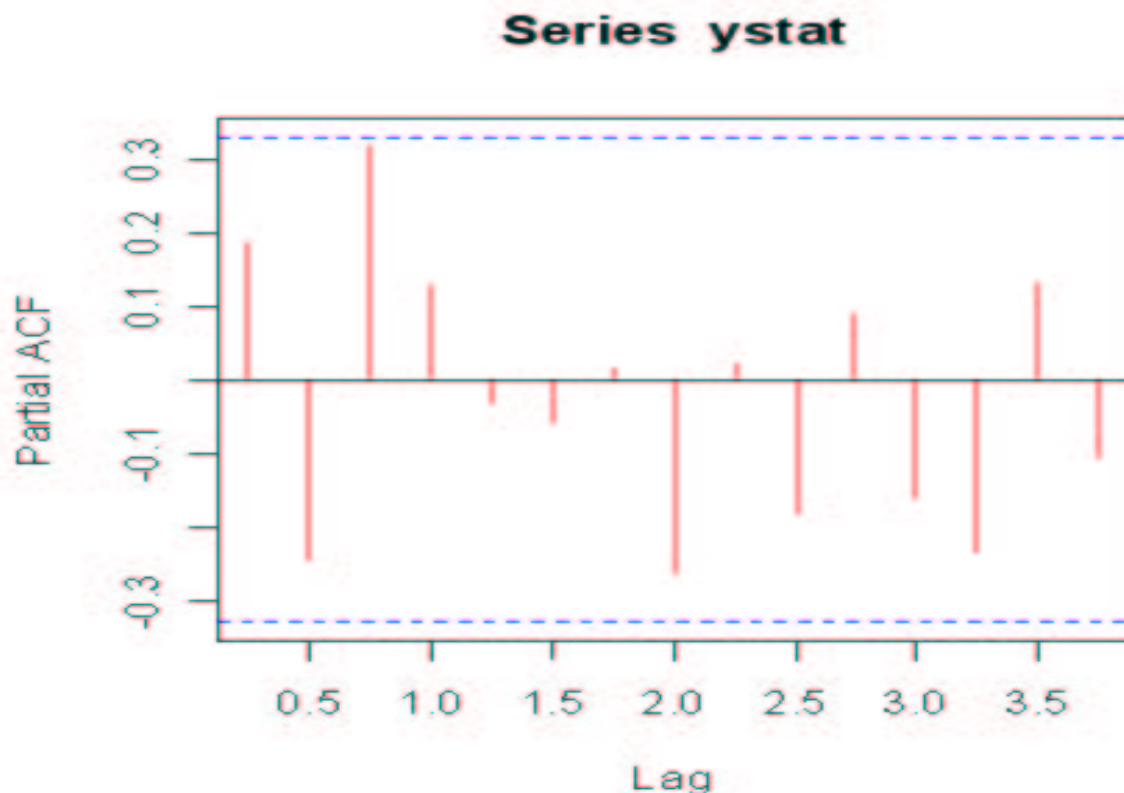


FIG. 3.7 – *Corrélogramme simple de ystat*

FIG. 3.8 – *Corrélogramme partiel de ystat*

L'ACF et le PACF des résidus montrent que ceux-ci suivent un ARIMA(0,0,1)

`t.test(ystat)`

*One Sample t-test*

`data: ystat`

`t = -5e-04, df = 35, p-value = 0.9996 alternative hypothesis: true mean is not equal to 0.95 percent confidence interval: -2.438457; 2.437205`

*sample estimates:*

*mean of x*

`-0.0006260424`

**Conclusion :** les résidus ont une moyenne significativement nulle.

## TEST de PORTMANTEAU SUR LES RESIDUS

Box-Pierce test

`data: ystat X-squared = 1.2722, df = 1, p-value = 0.2594`

Nous avons une  $p$ -value > 5%. Les résidus forment un bruit blanc. En observant l'ACF et le PACF des résidus, nous les modélisons par un processus MA(1)

`arima(ystat, c(0, 0, 1))`

*Call: arima(x = ystat, order = c(0,0,1)) Coefficients:*

	ma1	intercept
	0.4378	-0.1193
s.e.	0.1731	1.5996

$\sigma^2$  estimated as 45.29: log likelihood = -119.82, aic = 245.65

Les résidus s'écrivent :

$y_{stat} = -0.1193 + 0.4378 * \varepsilon(t)$ , où  $\varepsilon(t)$  est un BB

#### -Test de normalité des résidus

`shapiro.test(ystat)`

Shapiro-Wilk normality test

data: ystat  $W = 0.969$ ,  $p\text{-value} = 0.399$  : la  $p$ -value est supérieure à 5%.

Conclusion : les résidus suivent une loi normale centrée.

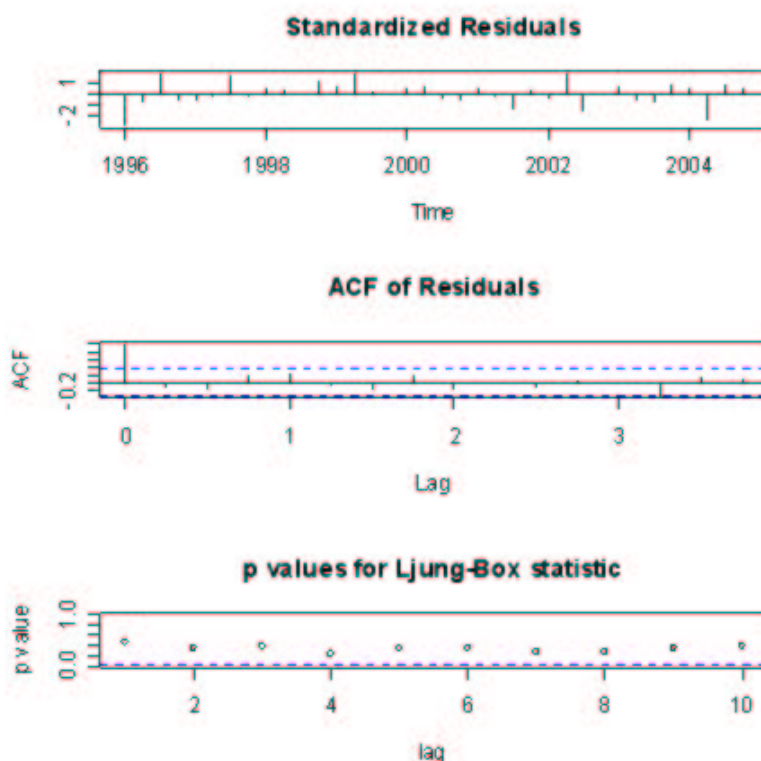


FIG. 3.9 – Test graphique de Ljung-Box

### 3.1.4 Prédiction.

Notre modèle stationnaire étant un ARIMA(0,0,1), nous pouvons prédire des valeurs de ce processus stationnaire sur une période donnée (1 an par exemple). A partir de cet instant, nous remontons aux valeurs de la série brute en suivant les étapes ci-après :

- $y_{stat} = -0.1193 + 0.4378 * \varepsilon(t)$
- $y_{cvs} = 0.8720 * t + 110.6903 + y_{stat}$
- $y = y_{cvs} + sai$

où  $sai$  est l'ensemble des coefficients saisonniers de la série brute. Valeurs du processus stationnaire sur un 1 an

\$ pred	Qtr1	Qtr2	Qtr3	Qtr4
2005	1.3706228	-0.1193319	-0.1193319	-0.1193319
\$ se	Qtr1	Qtr2	Qtr3	Qtr4
2005	6.730006	7.346823	7.346823	7.346823

TAB. 3.5 – Valeurs Prédites sur une année

## 3.2 Serie NO2: TOTAL INDUSTRIES MANUFACTURIERES

### 3.2.1 Analyse exploratoire des données

y	Qtr1	Qtr2	Qtr3	Qtr4
1996	104.54	97.32	105.63	111.81
1997	127.57	11.89	108.01	117.83
1998	138.45	125.70	93.41	113.58
1999	130.31	122.53	99.96	108.29
2000	128.14	117.96	97.84	107.76
2001	130.75	116.60	93.41	113.58
2002	130.31	129.15	100.91	114.49
2003	138.56	120.76	112.37	137.18
2004	163.54	127.49	120.65	144.99

TAB. 3.6 – Données Brutes Industries Manufacturières

Graphique des données brutes (cf FIG. 3.10)

## TOTAL INDUSTRIES MANUFACTURIERES

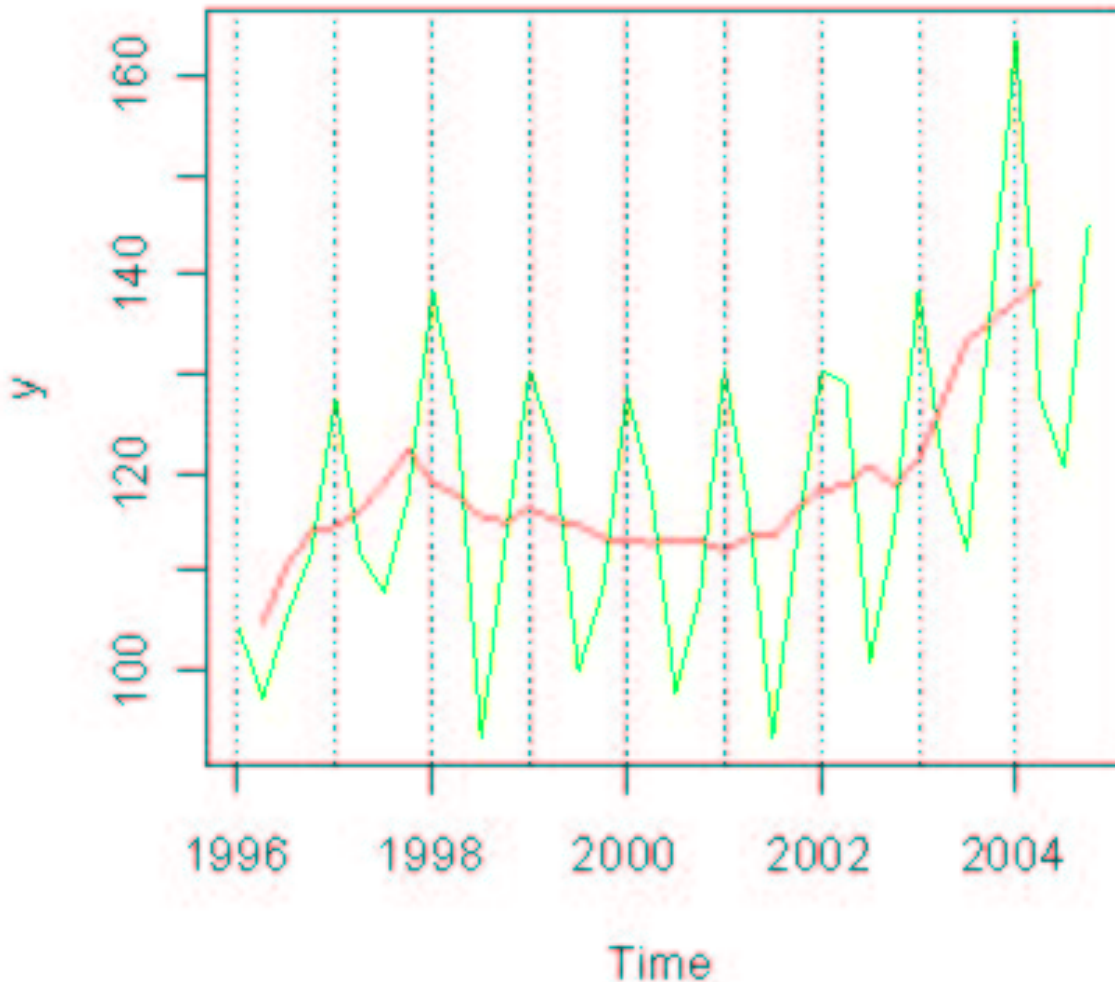


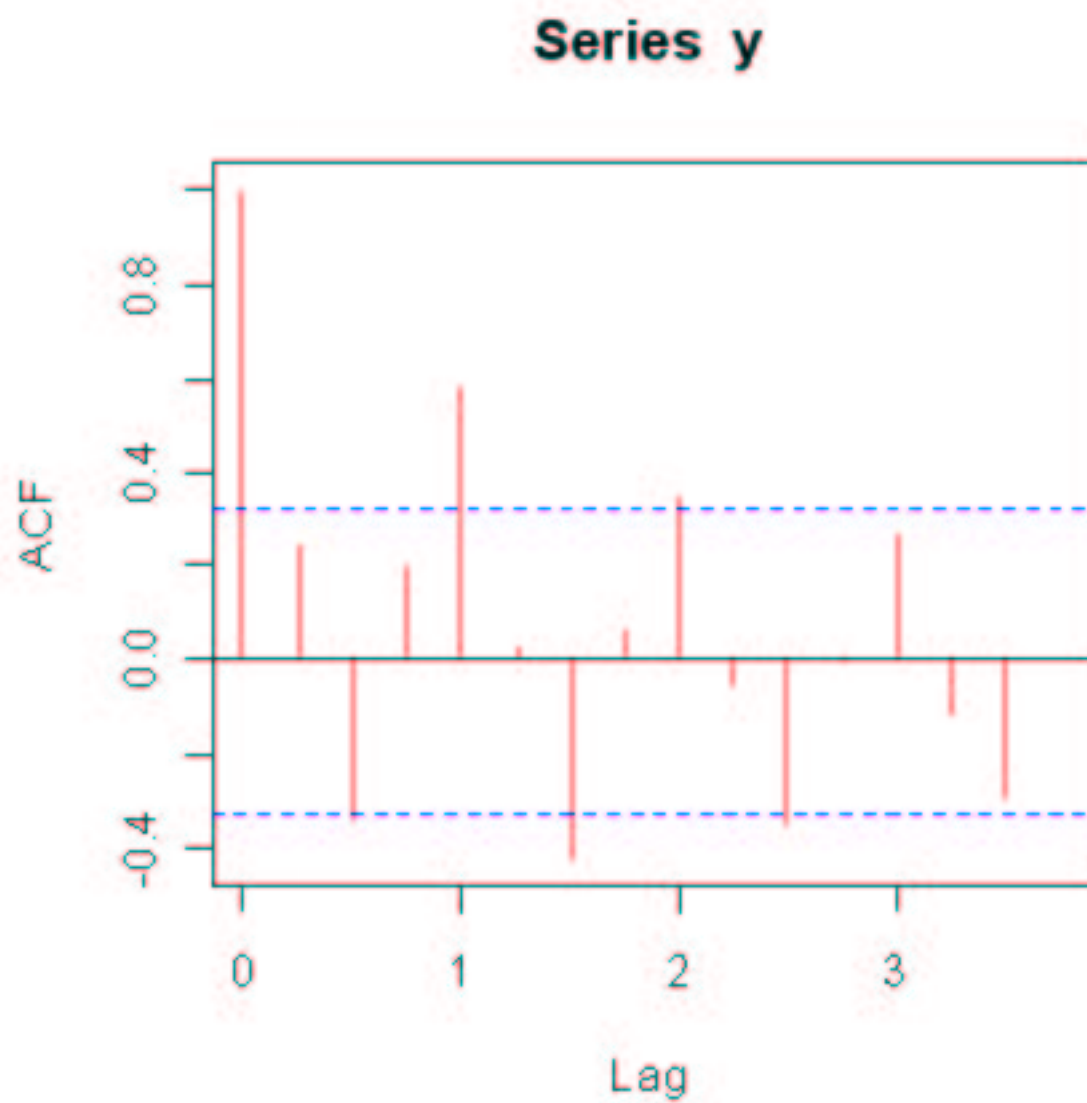
FIG. 3.10 – Série Industries manufacturières

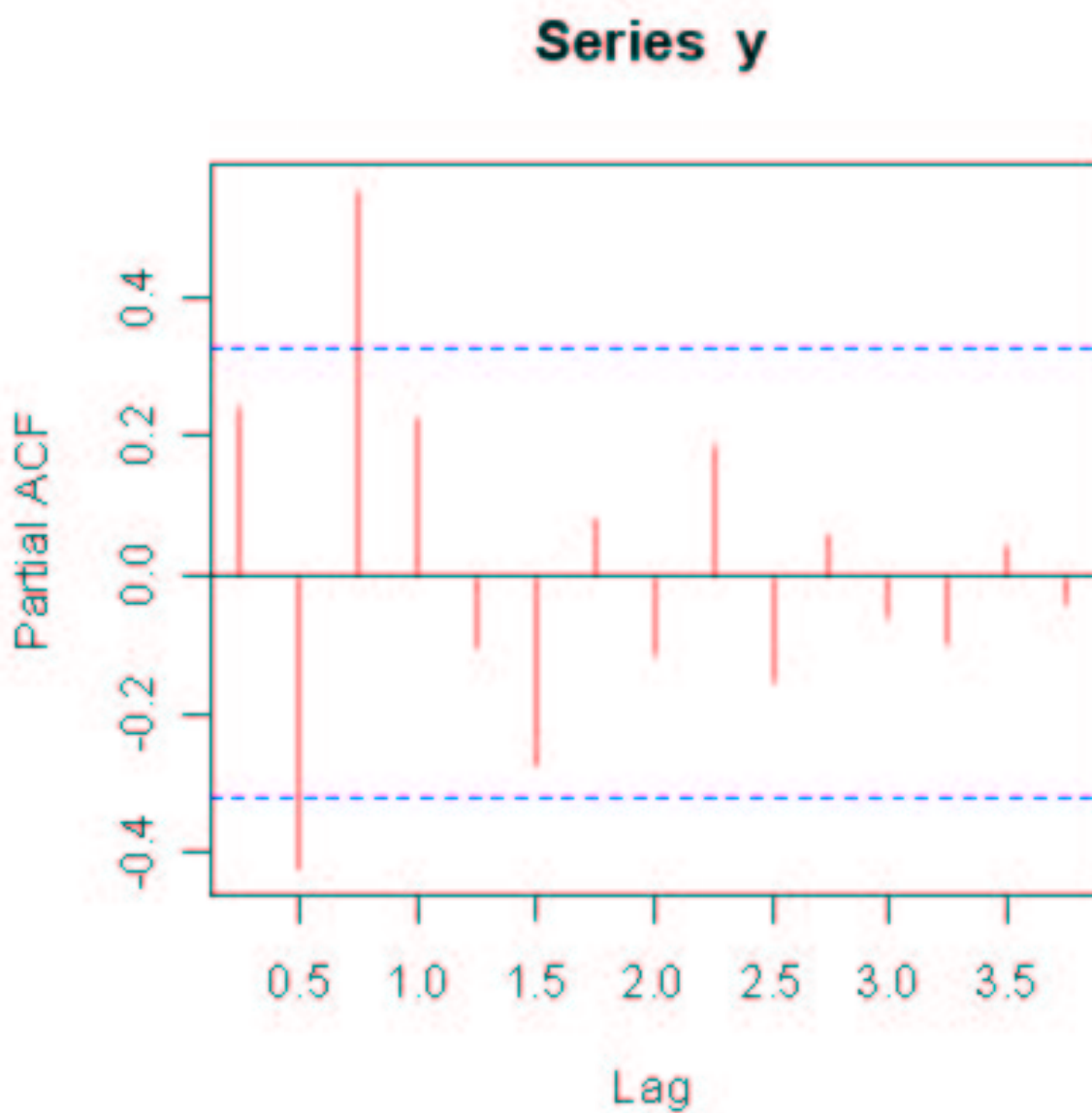
Dans cette série, nous étudions les observations trimestrielles de l'indice de production *total des industries manufacturières* au Cameroun entre 1996 et 2004.

L'examen graphique de la série laisse entrevoir une tendance croissante et presque polynomiale. La série présente une saisonnalité de période 4, avec des pics au 1<sup>er</sup> trimestre et des creux au 3<sup>eme</sup>. Par ailleurs, l'amplitude de ses fluctuations semble croissante et cela suggère donc un modèle de décomposition multiplicative.

**ACF et PACF de la série brute**(cf FIG. 3.11 et 3.12)

Le corrélogramme de la série brute présente des valeurs significativement non nulles

FIG. 3.11 – *Corrélogramme simple de y*

FIG. 3.12 – *Corrélogramme partiel de y*

des fonctions d'autocorrélation et d'autocorrélation partielle. La série brute n'est donc pas stationnaire. La forte saisonnalité de cette série nous conduit à la désaisonnaliser. Notons  $Y_{CVS}$  la série désaisonnalisée.

### 3.2.2 Stationnarité et identification du modèle

#### a)- Estimation des coefficients saisonniers

Qtr1	Qtr2	Qtr3	Qtr4
16.118253	1.026917	-15.471907	-1.673281

TAB. 3.7 – Coefficients saisonniers estimés

Les coefficients saisonniers au 1<sup>er</sup> et au 2<sup>eme</sup> trimestres sont positifs, on peut conclure que l'activité a tendance à être très élevée au cours de ces trimestres. Par contre cette activité est à la baisse aux 3<sup>eme</sup> et 4<sup>eme</sup> trimestres .

#### b)-Série désaisonnalisée

ycvs	Qtr1	Qtr2	Qtr3	Qtr4
1996	88.42175	96.29308	121.10191	113.48328
1997	111.45175	110.86308	123.48191	119.50328
1998	122.33175	124.67308	108.88191	115.25328
1999	114.19175	121.50308	115.43191	109.96328
2000	112.02175	116.93308	113.31191	109.43328
2001	114.63175	115.57308	108.88191	115.25328
2002	114.19175	128.12308	116.38191	116.16328
2003	122.44175	119.73308	127.84191	138.85328
2004	147.42175	126.46308	136.12191	146.66328

TAB. 3.8 – Données désaisonnalisées de l'I.P.I des industries manufacturières( $Y_{CVS}$ )

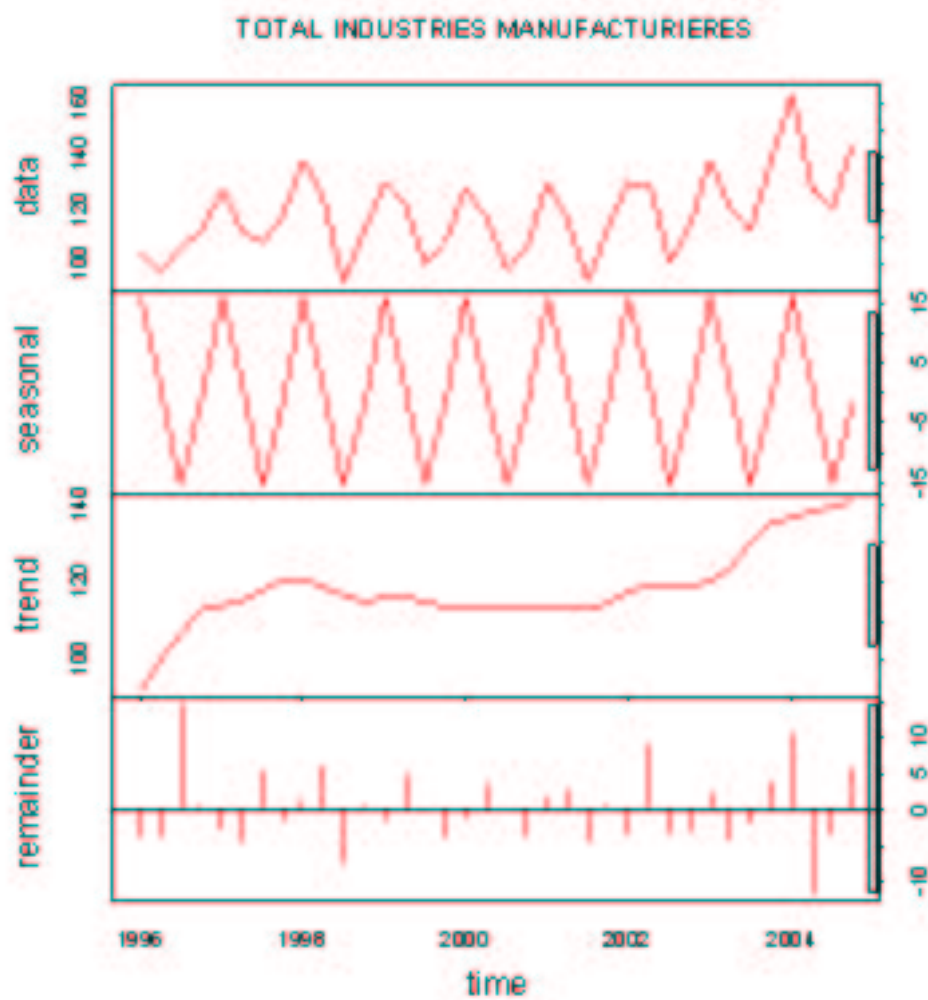


FIG. 3.13 – Série désaisonnalisée de l'I.P.I des industries manufacturières ( $Y_{CVS}$ )

c)- ACF et PACF de la série désaisonnalisée

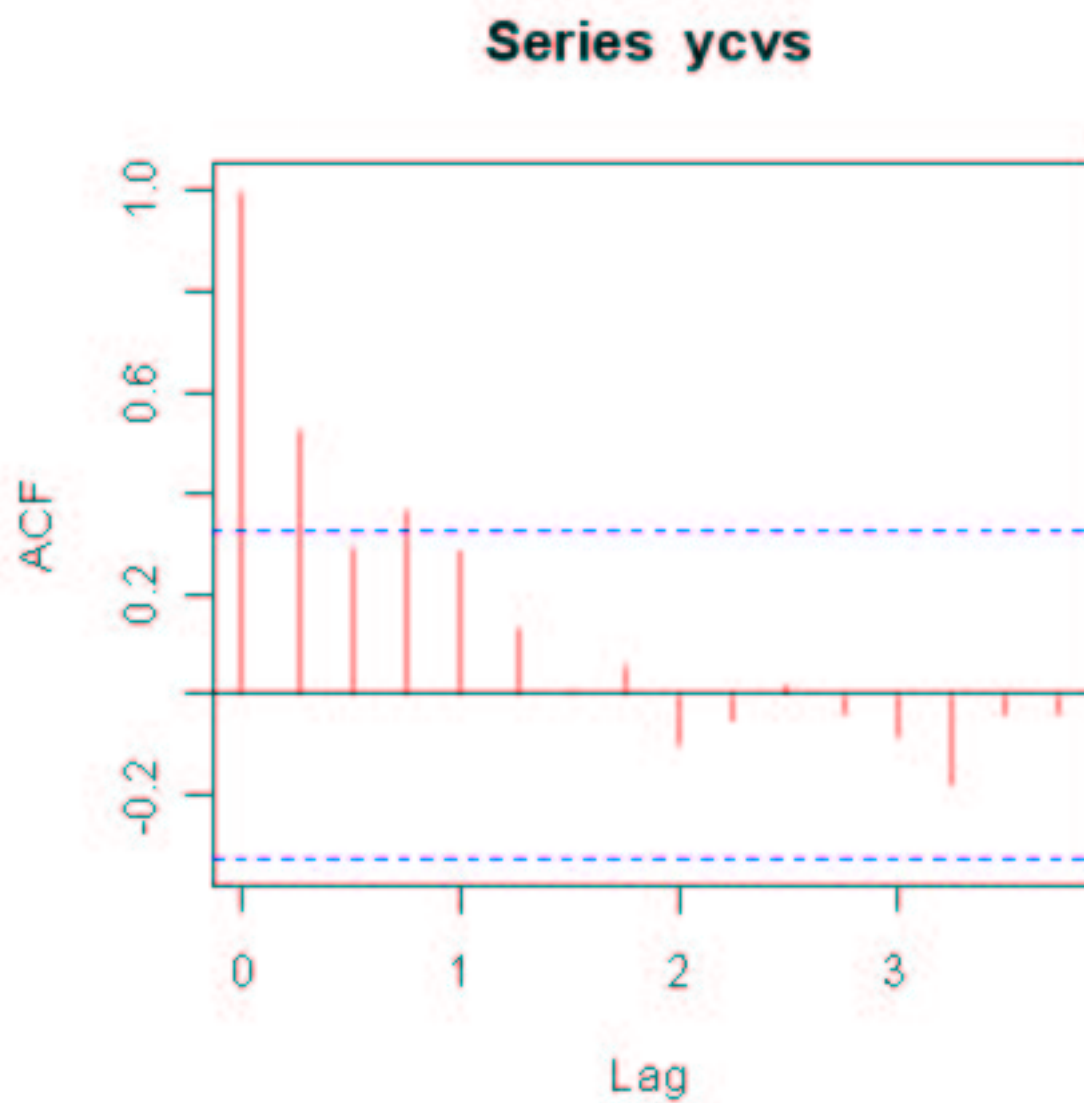
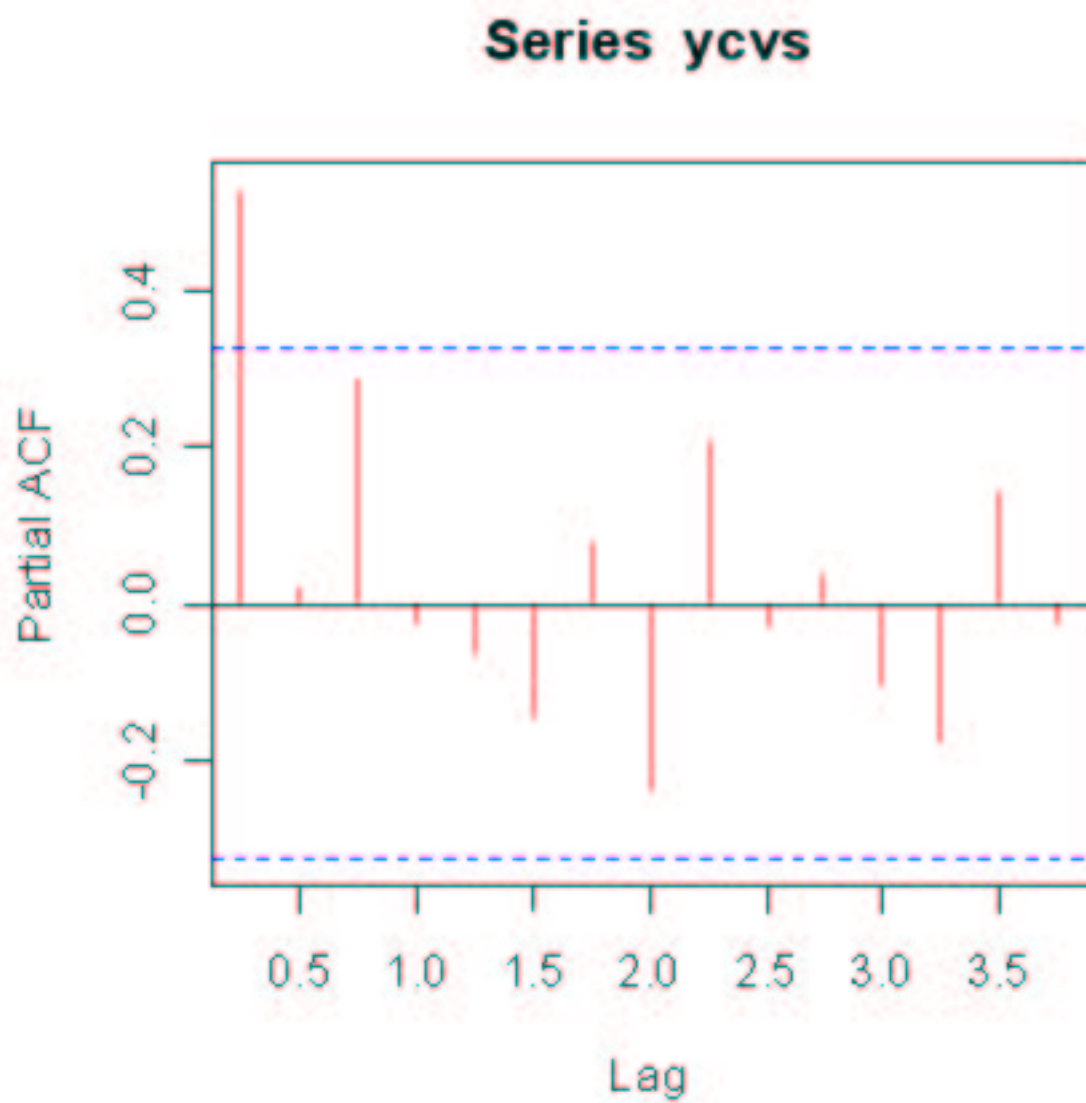


FIG. 3.14 – *Corrélogramme simple de  $Y_{CVS}$*

FIG. 3.15 – *Corrélogramme partiel de  $Y_{CVS}$*

Le corrélogramme de la série désaisonnalisée montre que celle-ci présente une forme ARMA(4,1), mais ne nous permet pas de conclure à la stationnarité de la série. Pour cela nous allons effectuer le test de stationnarité de DICKEY-FULLER

#### d)-Test de stationnarité

*Phillips-Perron Unit Root Test*

*Dickey-Fuller* = -3.6096,

*p-value* = 0.04624

Le test de DICKEY-FULLER sur la série désaisonnalisée montre que celle-ci n'est pas stationnaire.

La *p-value* < 5%, nous essaierons de stationnariser notre série ( tendance). Pour cela nous allons utiliser la méthode de régression pour approcher la tendance par un polynôme de degré 3.

Coefficients	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	95.196753	4.996265	4.996265	< 2e-16 ***
a	4.727698	1.153470	4.099	0.000265 ***
b	-0.309561	0.071912	-4.305	0.000148 ***
c	0.006097	0.001279	4.768	3.89e-05 ***

TAB. 3.9 – Valeurs estimées des coefficients polynomiaux

$a$  = coefficient de  $t$

$b$  = coefficient de  $t^2$

$c$  = coefficient de  $t^3$

L'équation de la tendance est :

$$g(t) = 95.198 + 4.728t - 0.310t^2 + 0.0061t^3$$

### 3.2.3 Estimation et validation du modèle

$$y_{stat} = y_{cvs} - g(t)$$

-ACF et PACF des résidus

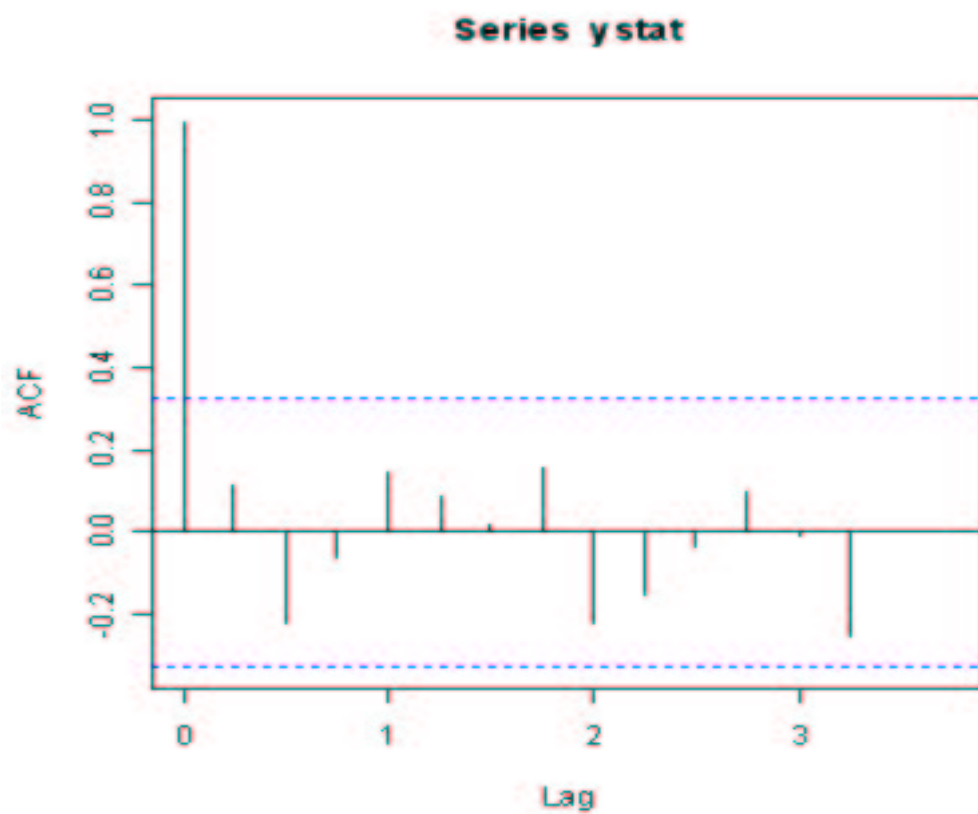


FIG. 3.16 – *Corrélogramme simple de ystat*

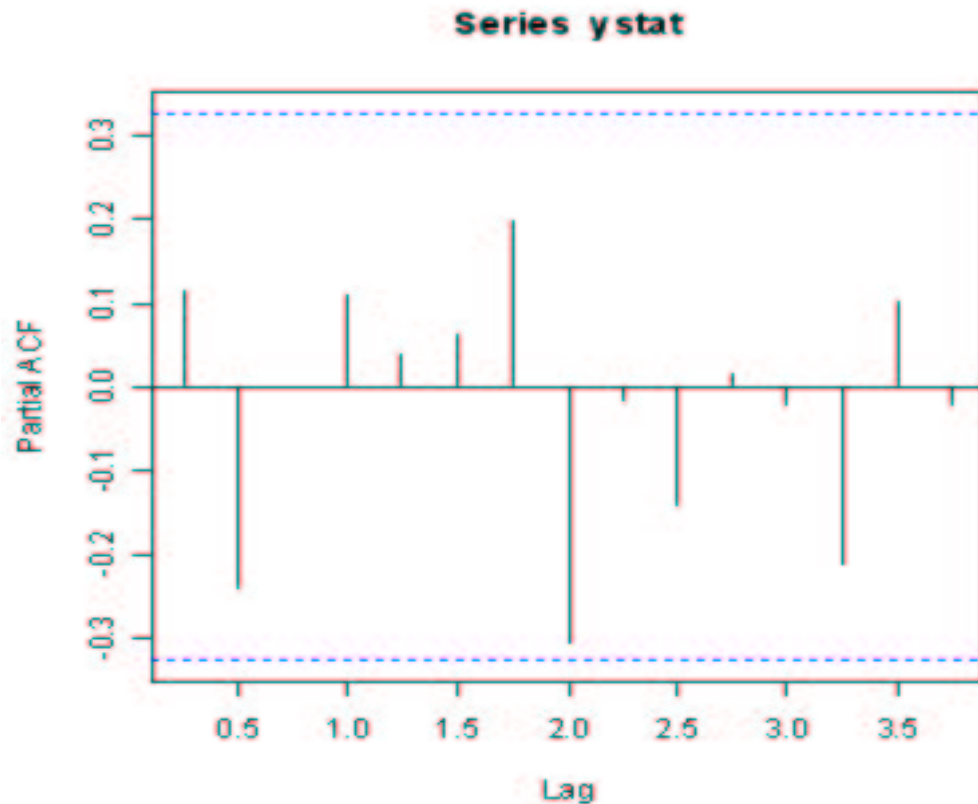


FIG. 3.17 – *Corrélogramme partiel de ystat*

L'ACF et le PACF des résidus montrent que ceux-ci suivent un Bruit Blanc(cf.fig 3.16 et 3.17).

#### -Test sur la moyenne des résidus

*One Sample t-test*

*data: ystat*

$t = -6e - 04$ ,  $df = 35$ ,  $p\text{-value} = 0.9995$

*alternative hypothesis: true mean is not equal to 0 .95 percent*

*confidence interval: -3.003117; 3.001259*

*moyenne des résidus: -0.0009288662*

**-Conclusion : la moyenne des résidus est significativement nulle**

#### **-TEST DE PORTMANTEAU SUR LES RESIDUS**

*Box-Pierce test*

*data: ystat*

$X\text{-squared} = 5.9058,$

$df = 1, p\text{-value} = 0.01509$

$\text{arima}(x = \text{ystat}, \text{order} = c(1, 0, 2))$

*Coefficients :*

	ar1	ma1	ma2	intercept
	0.7517	-0.2503	-0.1813	0.0859
s.e.	0.2368	0.2875	0.1908	2.8077

TAB. 3.10 – *Estimation des coefficients du modèle ARMA(1, 2)*

$\sigma^2$  estimated as 59.47,

log likelihood = -124.78,

aic = 259.55.

$\text{ar1} = \text{arima}(\text{ystat}, c(1, 0, 0))$

$\text{arima}(x = \text{ystat}, \text{order} = c(1, 0, 0))$

*Coefficients :*

	ar1	intercept
	0.4885	-0.04536
s.e.	0.1614	2.4776

TAB. 3.11 – *Coefficients estimés du modèle AR(1)*

$\sigma^2$  estimated as 60.87,

log likelihood = -125.17,

aic = 256.35.

$\text{ma} = \text{arima}(\text{ystat}, c(0, 0, 2))$

*ma*

$\text{arima}(x = \text{ystat}, \text{order} = c(0, 0, 2))$

*Coefficients:*

	ma1	ma2	intercept
	0.5124	0.1216	0.0049
s.e.	0.1832	0.2021	2.1028

TAB. 3.12 – Coefficients estimés du modèle MA(2)

$\sigma^2$  estimated as 61.14,

log likelihood = -125.25,

aic = 258.5.

Nous allons utiliser le Critère d'AKAIKE pour choisir entre les trois modèles suivants: AR(1), MA(2), ARMA(1,2).

Le modèle ayant le plus faible AIC est AR(1)

Les résidus s'écrivent:  $y_{stat}(t) = -0.0453 + 0.4885 * y_{stat}(t - 1)$

#### -Test de normalité des résidus

*shapiro.test(ystat)*

Shapiro-Wilk normality test

data: ystat

$W = 0.9689$ ,

$p - value = 0.3953$

**-Conclusion : les résidus suivent une loi normale centrée.**

### 3.2.4 Prévision

Notre modèle stationnaire étant un ARIMA(1,0,0), nous pouvons prédire des valeurs de ce processus stationnaire sur une période donnée. Prévision sur un an  
`predict(ar,4)`

\$ pred	Qtr1	Qtr2	Qtr3	Qtr4
2005	7.6423986	3.7104702	1.7895558	0.8511073
\$ se	Qtr1	Qtr2	Qtr3	Qtr4
2005	7.801738	8.683001	8.880417	8.926890

TAB. 3.13 – Valeurs prédites sur un an

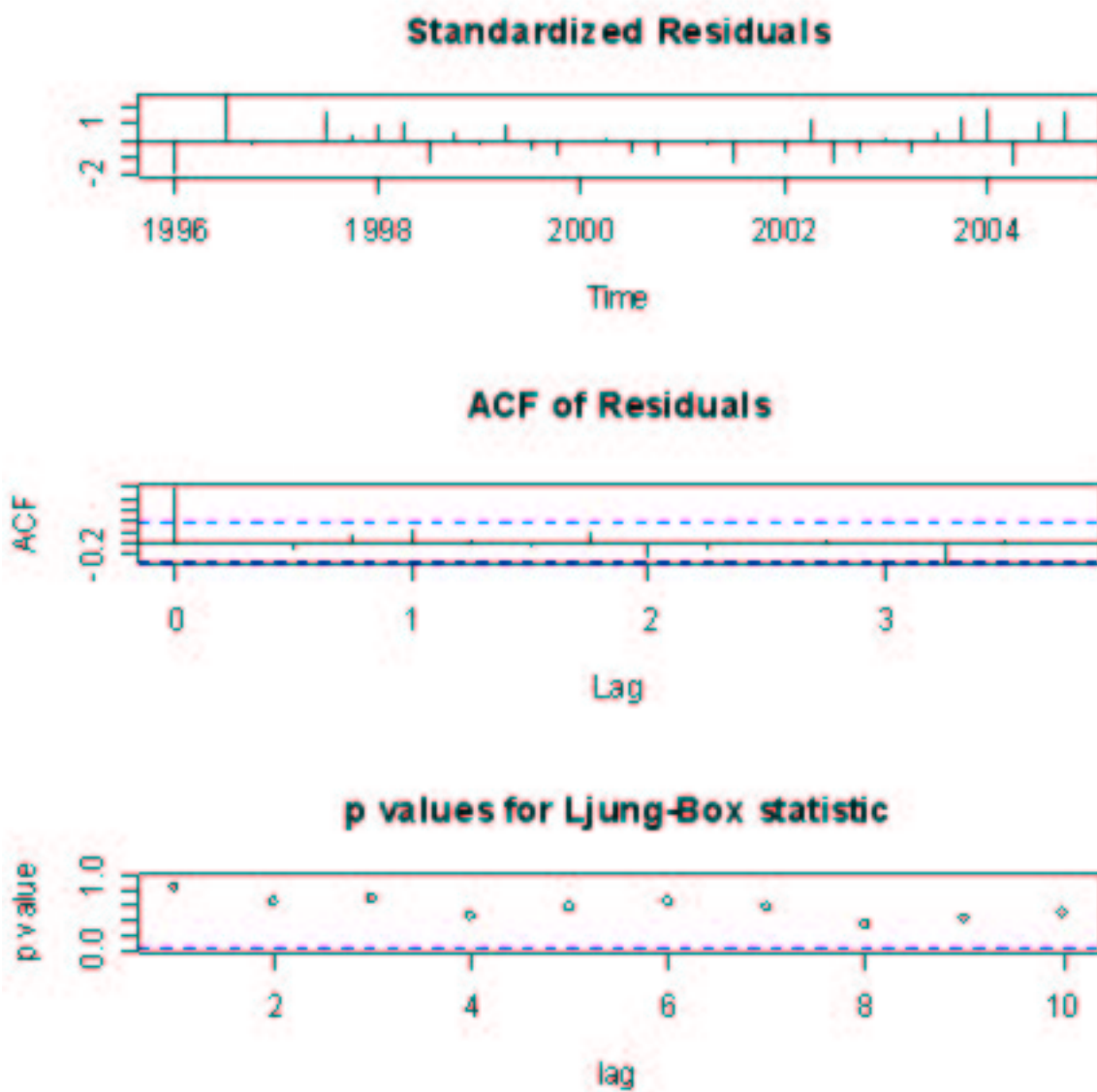


FIG. 3.18 – Test graphique de Ljung-Box de  $y_{stat}$

---

---

# CONCLUSION

---

Ce mémoire a été pour nous l'occasion de réfléchir sur deux aspects du traitement des séries temporelles: la **saisonnalité** et la **modélisation**.

Nous avons tout d'abord utilisé « une méthode classique » de désaisonnalisation, qui consiste à appliquer des moyennes mobiles pour stationnariser la série et à appliquer un modèle ARMA au résidu obtenu. Nous avons choisi d'estimer la composante saisonnière de manière à obtenir une courbe lisse. Cette méthode très couramment utilisée par les logiciels  $X_{11}$  et  $X_{12}$ , a pour but d'évacuer la saisonnalité, comme si elle constituait un phénomène parasite. Le problème de cette méthode est d'une part qu'elle sous-estime l'importance de la saisonnalité ( le résidu obtenu est de variance très élevée), d'autre part qu'elle ne prend pas en compte la persistance à long terme des chocs.

La limite de notre travail a porté sur la correction des effets de jours ouvrables (effets calendaires). Le partage entre effets « *jours ouvrables* » et les effets « *saisonniers* » est très délicat car ceux-ci ne sont pas toujours indépendants, ils se superposent en partie. La correction des effets « *jours ouvrables* » devient parfois compliqué sous l'effet de modifications de l'organisation du travail faisant évoluer l'impact sur l'ensemble de la production de la composition du calendrier en jours ouvrables. La correction CJO est réalisée par des logiciels spécialisés comme TRAMO-SEATS .

La suite logique de notre travail consisterait à faire une correction CJO-CVS en utilisant l'un des logiciels indiqués à cet effet. Les séries calculées par TRAMO-SEATS n'ont pas tendance à être surcorrigées. TRAMO-SEATS utilise des méthodes de désaisonnalisation paramétriques, celles-ci étant fondée sur des modèles ARIMA saisonniers et des méthodes d'analyse spectrale.

---

---

# Bibliographie

---

## A) OUVRAGES

- [1] Alain Saglio, *pour comprendre l'indice de la production industrielle, sources et méthodes de la base 1980*, INSEE
- [2] Dacunha-Castelle D., Duflo M., *Probabilités et statistiques, tome2, problèmes à temps mobile*, Masson 1993.
- [3] Brockwell, P., Davis, R. : *Introduction to time series and Ferocasting*. 1996, Springer.
- [4] Bosq D. : Le Coutre J.P. *Analyse et prévision des séries chronologiques, Méthodes paramétriques et non paramétriques*, Masson 1992.
- [5] Christian Gouriéroux, Alain Monfort : *Séries temporelles et Modèles dynamique*, edition Economica, 1995.
- [6] J. Jacques, DROESKE, Bernard Fichet, Philippe TASSI, *séries chronologiques : Théorie et pratique des modèles ARIMA*, Edition Economica, 1989.
- [7] Pankratz, A. : *Forecasting with univariate, Box-Jenkins Models*. 1983, Wiley.
- [8] Regis Bourbonnais, *Econométrie*, Dunod, 2000.
- [9] Serge Dégerine, *Cours de séries chronologiques*, Septembre 2004 (Université Joseph Fourier).

## B) SITES INTERNET

- [10] <http://www.mathworks.com/access/helpbok/toolbox/stats>.
- [11] <http://pbl.univ.lyon1.fr/splus>.
- [12] <http://www.math.ccsu.edu/dm/>
- [13] <http://www.stat.ucl.ac.be/cours/stat2414>

## C) NOTES DU COURS

- [14] Elisabeth Gassiat, *séries chronologiques*, Master 2, UYI, 2005
- [15] Bernard Rapacchi, *les séries chronologiques*, Grenoble, 1993
- [16] Lafosse, *introduction aux séries chronologiques*, UPS Toulouse, 2003.

---

# ANNEXES

---

## COMMANDES R UTILISEES

```
-AR
-arima
-filter(data)
-lag
-spec.ar : estimate spectral density of a time series from AR fit
-spec.pgram: estimate spectral density of a time series by smoothed periodogram
-start
-stl: seasonal decomposition of time series _ts: time series
-acf: autocorrelation function
-Decompose: classical seasonal decomposition by moving average
-PP.test: Phillips- Perron test for unit roots
-partial autocorrelation function
-plot (y)
-lines(predict(loess(y~time(y))), col='red', lwd=3)
k <- 4
-lines(filter(abs(y), rep(1/k,k)), col='blue', lwd=3, lty=2)
n <- 20
a <- rep(NA, n)
zz <- z
-for(i in 1: n) {
-a[i] <- acf(zz, plot=F)$acf[2]
}
-plot(a,type='h')
-op <- par(mfrow=c(3, 1))
-plot(x)
-abline(h=0, v=1996: 2004, lty=3)
y <- diff(x)
-plot(y)
-abline(h=0, v=1996: 2004, lty=3)
-plot(diff(y, 12,1))
-abline(h=0, v=1996:2004, lty=3)
-par(op)
-plot(x)
-abline(lm(x~time(x)), col='red')
```

```

n <- 2000
x <- arima.sim(model=list(ar=c(.3,.6), ma=c(.8,-.5,.2), order=c(2,1,3)), n)
x <- ts(x)
-plot(x, main="Dériver une fois suffit")
-acf(x)
-acf(diff(x))
-plot(x)
-abline(lm(x~time(x)), col='red')
-abline(h=0,lty=3)

```

### Table de Dickey-Fuller (DF)

Critical values for the Dickey-Fuller t statistics

Sample size	Level of significance		
	0.01	0.05	0.10
<b>Testing <math>\Phi = 0</math> in <math>\Delta x_t = \Phi \cdot x_{t-1} + \eta_t</math> "t-test" statistic</b>			
25	-2.66	-1.95	-1.60
50	-2.62	-1.95	-1.61
100	-2.60	-1.95	-1.61
250	-2.58	-1.95	-1.62
Infinite	-2.58	-1.95	-1.62
<b>Testing <math>\Phi = 0</math> in <math>\Delta x_t = \Phi \cdot x_{t-1} + \beta + \eta_t</math> "t-test" statistic</b>			
25	-3.75	-3.00	-2.62
50	-3.58	-2.93	-2.60
100	-3.51	-2.89	-2.58
250	-3.46	-2.88	-2.57
Infinite	-3.44	-2.87	-2.57
<b>Testing <math>\Phi = 0</math> in <math>\Delta x_t = \Phi \cdot x_{t-1} + \beta + \gamma \cdot t + \eta_t</math> "t-test" statistic</b>			
25	-4.38	-3.60	-3.24
50	-4.15	-3.50	-3.18
100	-4.04	-3.45	-3.15
250	-3.99	-3.43	-3.13
Infinite	-3.98	-3.42	-3.13
<b>Standard t-distribution (Student)</b>			
Infinite	-2.33	-1.65	-1.28
(Source: W. A. Fuller (1976), Table 8.5.2)			