

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

UNIVERSITÉ MOHAMED KHIDER, BISKRA
FACULTÉ DES SCIENCES EXACTES
DÉPARTEMENT DE MATHÉMATIQUES



Mémoire présenté en vue de l'obtention du Diplôme :

MASTER en Mathématiques

Option : **Probabilités et Statistique**

Par

ABDELLAOUI Khedidja

Titre :

Désaisonnalisation par régression linéaire et application

Membres du Comité d'Examen :

Pr. NECIR Abdelhakim	U. Biskra	Président
Pr. YAHIA Djabrane	U. Biskra	Encadrant
Dr. BENAMEUR Sana	U. Biskra	Examinatrice

Biskra le 03 / 06/2025

Dédicace

***Merci à dieu**, qui m'a aidé et m'a inspiré la patience et la force nécessaires pour atteindre ce stade, je lui adresse mes remerciements et ma gratitude pour ses innombrables bénédictions, sans lesquelles je n'aurais pas été en mesure d'achever ce travail.*

*Je dédier ce modeste travail et exprimer ainsi mes vifs remerciements et ma gratitude les plus profonds à **mes parents**, qui ont été mon soutien constant tout au long de ma vie. Je vous remercie pour votre soutien et vos encouragements à toutes les étapes de ma vie et je demande à Dieu de vous protéger et de prolonger votre vie.*

Je remercie aussi mes frères et soeurs pour leur soutien et leurs encouragements constants, qui m'ont incité à aller de l'avant.

Merci infiniment...

Remerciements

*Je tiens tout d'abord á remercier **ALLAH** de m'avoir donné le courage, la patience et la santé pour accomplir ce modeste travail.*

*Mes vifs remerciements vont à mon encadrant, Professeur **YAHIA Djabrane**, pour ses précieux conseils, ses orientations pertinentes jusqu'à la réalisation de ce travail. Aucun mot ne vous rendra justice.*

*Je tiens aussi à remercier les membres de jury : **Prof. NECIR Abdelhakim** en qualité de Président du Jury et **Dr. BENAMEUR Sana** en tant qu'examinatrice, d'avoir accepté d'examiner et d'évaluer ce travail.*

*A tous mes professeurs qui m'ont enseigné à l'école **Bouaziz Mohamed Lakhdar**, au **C.E.M Cheikh Mouloud Zeribi**, au lycée des Frères Martyrs Khadraoui et au Département de Mathématiques de l'université de Biskra, je vous adresse mes sincères remerciements et ma gratitude.*

***A toutes mes amies**, avec les quelles j'ai partagé un parcours commun, vous avez été mes meilleures compagnes.*

Table des matières

Table des figures	v
Liste des tableaux	vi
Introduction	1
1 <i>Modèle d'une série temporelle</i>	3
1.1 Série temporelle	3
1.2 Composantes d'une série temporelle	4
1.2.1 Tendance	4
1.2.2 Saisonnalité	6
1.2.3 Résidus (Bruit Blanc)	7
1.3 Schéma d'une série temporelle	9
1.3.1 Modèle additif	9
1.3.2 Modèle multiplicatif	9
1.3.3 Modèle mixte	10
1.4 Filtration	11
1.4.1 Opérateur de retard et d'avance	11
1.4.2 Opérateur de différentiation	11
1.5 Moyennes mobiles	12
2 <i>Régression et désaisonnalisation</i>	13
2.1 Ecriture d'un modèle linéaire	13
2.2 Estimateur des moindres carrés ordinaires	14
2.3 Généralisation des formules de Buys-Ballot	18
2.3.1 Modèle trimestriel de Buys-Ballot : Cas particulier	18
2.3.2 Modèle général de Buys-Ballot	21
2.4 Le principe de la conservation des aires	26
2.4.1 Le cas d'un schéma additif	26

2.4.2	Le cas d'un schéma multiplicatif	27
2.4.3	La série corrigée des variations saisonnières	27
2.5	Désaisonnalisation de la série temporelle	28
2.5.1	Etapes de la désaisonnalisation	28
2.6	Validation	29
2.6.1	Significativité de la régression	30
2.6.2	Significativité globale	30
2.6.3	Test de Fisher	31
2.6.4	Test de Student	31
2.7	Prévision	32
3	<i>Application aux données réelles</i>	33
3.1	Consommation de gaz naturel aux USA	33
3.1.1	Description de la série	33
3.1.2	Désaisonnalisation de la série	35
3.1.3	Validation du modèle	36
3.1.4	Prévisions	37
3.2	Indice des prix à l'exportation	39
3.2.1	Description de la série	39
3.2.2	Désaisonnalisation par régression linéaire	40
3.2.3	Validation	42
3.2.4	Prévisions	43
	Conclusion	45
	Bibliographie	46
	Annexe : Abréviations et Notations	48

Table des figures

1.1	Représentation graphique de la série USgas : 2000-2019	4
1.2	Tendance linéaire	5
1.3	Tendance polynomiale d'ordre $q=2$	5
1.4	Tendance logarithmique	6
1.5	Tendance exponentielle	6
1.6	Composante de la saisonnalité	7
1.7	Simulation d'un bruit blanc $N(0,1)$ de taille 1000.	8
1.8	Composantes de la série temporelle USVSales (1976 – 2019)	8
1.9	Schéma additif : série co2, 1959-1997.	9
1.10	Schéma multiplicatif : série AirPassengers, 1949-1960	10
1.11	Schéma mixte : série Nile, 1871-1970	10
2.1	Série CVS du Modèle additif (série USgas)	29
3.1	Série de consommation mensuelle du gaz naturel aux USA, 2000-2024	34
3.2	Composantes de la série de consommation du gaz naturel aux USA	35
3.3	Série CVS , de la série de consommation de gaz aux USA.	36
3.4	Désaisonnalisation, CVS et prévision de la série originale de consommation du gaz aux USA.	39
3.5	Série IPE : tarifs des passagers aériens pour l'Europe	39
3.6	Composantes de la série de l'indice des prix à l'exportation	40
3.7	IPE : Série originale et désaisonnalisée	41
3.8	Désaisonnalisation, CVS et prévision de la série originale de l'IPE.	43

Liste des tableaux

2.1	Exemple d'un modèle trimestriel de Buys-Ballot.	20
2.2	Cas général Buys-Ballot pour la série USgas	25
2.3	Estimation des coefficients saisonniers	26
2.4	Coefficients saisonniers définitifs du modèle additif (série USgas)	29
3.1	Estimateurs des coefficients saisonniers de la série de consommation de gaz aux USA.	35
3.2	Coefficients saisonniers définitifs de la série de consommation de gaz aux USA.	36
3.3	Test de Student des paramètres (Modèle additif)	37
3.4	Données initiales et prévisions de l'année 2024	38
3.5	Prévision de la série pour l'année 2025	38
3.6	Coefficients saisonniers : série IPE.	41
3.7	Coefficients saisonniers définitifs de la série IPE.	41
3.8	Test de Student des paramètres (série IPE : Modèle multiplicatif)	42
3.9	Prévisions IPE 2012	43
3.10	Prévisions IPE 2013	43

Introduction



es séries temporelles forment un ensemble de valeurs régulièrement enregistrées à intervalles de temps égaux, tels que le jour, le mois ou l'année, pour décrire et analyser des phénomènes qui évoluent dans le temps. C'est là qu'interviennent les séries chronologiques quotidiennes, mensuelles ou annuelles. Il s'agit en fait d'un outil essentiel qui joue un rôle fondamental dans un grand nombre de domaines, notamment en :

- **Astronomie** : périodicité des taches solaires,
- **Météorologie** : régression temporelle du niveau de la mer sur les conditions météorologiques,
- **Théorie du signal** : bruit dans les récepteurs radio FM,
- **Biologie** : courbes d'autocorrélation des ondes cérébrales des schizophrènes,
- **Économie** : analyse des importations, des exportations et d'autres variables économiques,
- **Assurance** : analyse des sinistres ou incendies,
- **Médecine** : l'évolution des conditions de santé, comme la fréquence cardiaque,
- **Psychologie** : analyse des variations de l'état psychologique pour évaluer l'effet des traitements sur la dépression ou l'anxiété, ... etc.

Le but de l'analyse des séries temporelles est d'extraire des modèles et des tendances invisibles dans les données, ce qui permet de mieux comprendre le phénomène étudié et de fournir des prévisions fiables.

Pour analyser les séries temporelles, les données doivent être converties en un modèle mathématique afin d'estimer les coefficients de saisonnalité et la tendance de la série par : filtrage, moyenne mobile, estimation non paramétrique ou semi paramétrique et régression linéaire. Dans ce mémoire, il sera question de **la désaisonnalisation par régression linéaire et application**. Ainsi, notre objectif dans ce mémoire est de donner une vue générale de la désaisonnalisation des séries temporelles par la régression linéaire.

Le manuscrit est structuré en trois chapitres comme suit :

Chapitre 1 : Introduction aux concepts d'une séries temporelle et de ses propriétés, avec mise en évidence de ses principales composantes.

Chapitre 2 : C'est le coeur du sujet, dont nous estimons les coefficients saisonniers et ceux de la tendance à partir de la méthode des moindres carrés. En outre, nous explorons le modèle de **Buys-Ballot** dans ses cas particuliers et généraux, ainsi que les méthodes de désaisonnalisation utilisant la régression linéaire dans le cadre de modèles additif et multiplicatif. Enfin, nous discutons des techniques de validation et de prévision.

Chapitre 3 : Finalement, nous appliquons la méthode de désaisonnalisation sur des séries temporelles réelles à l'aide de la régression linéaire. En effet, nous utilisons des données mensuelles de consommation de gaz naturel aux États-Unis, pour le modèle additif et les données mensuelles de l'indice des prix à l'exportation (tarifs aériens pour l'Europe) pour le modèle multiplicatif. L'étude de simulation est réalisée par le logiciel statistique **R**.

Chapitre 1

Modèle d'une série temporelle

Dans ce premier chapitre, nous présentons une vue générale des concepts de base liés aux séries temporelles. Nous donnons une définition succincte d'une série temporelle, ces composantes et les types de schémas possibles.

1.1 Série temporelle

Définition 1.1.1. [1, 19]. Une *série temporelle* (ou *série chronologique*), notée X_t , est une suite d'observations d'un phénomène, indicées par un temps t , (jour, mois, semaine, année,...) notée

$$X_t = \{X_t \quad , \quad t = 1, \dots, T\} \quad , \quad T \in \mathbb{R}, \mathbb{Z}, \mathbb{N}, \dots$$

L'importance des séries temporelles est démontrée par ses nombreux **domaines d'application** :

- **Économie** : prévision d'indices économiques (**série USgas**[18, package "TSstudio"]).
- **Finance** : évolution des cours de la bourse (**série EuStockMarkets**[15, package "datasets"]).
- **Météorologie** : analyse de données climatiques (**série airmiles**[5, package "datasets"]).
- **Médecine** : analyse d'électrocardiogrammes (**série esoph**[4, package "datasets"]).

Les **buts** poursuivis dans des études de séries temporelles sont multiples :

- **Description** : détermination de composantes de tendance et saisonnalité et d'erreur.

- **Filtrage** : transformation de la série dans le but d'éliminer certaines caractéristiques ou des valeurs aberrantes .
- **Modélisation** : recherche de causalité ou de stationnarité.
- **Prévision** : consiste à prévoir les valeurs futures de la série.

Exemple 1.1.1. Série USgas : La [Figure 1.1](#) présente l'évolution temporelle de la série de consommation mensuelle de gaz naturel aux États-Unis (2000-2019). Source : [\[18\]](#).

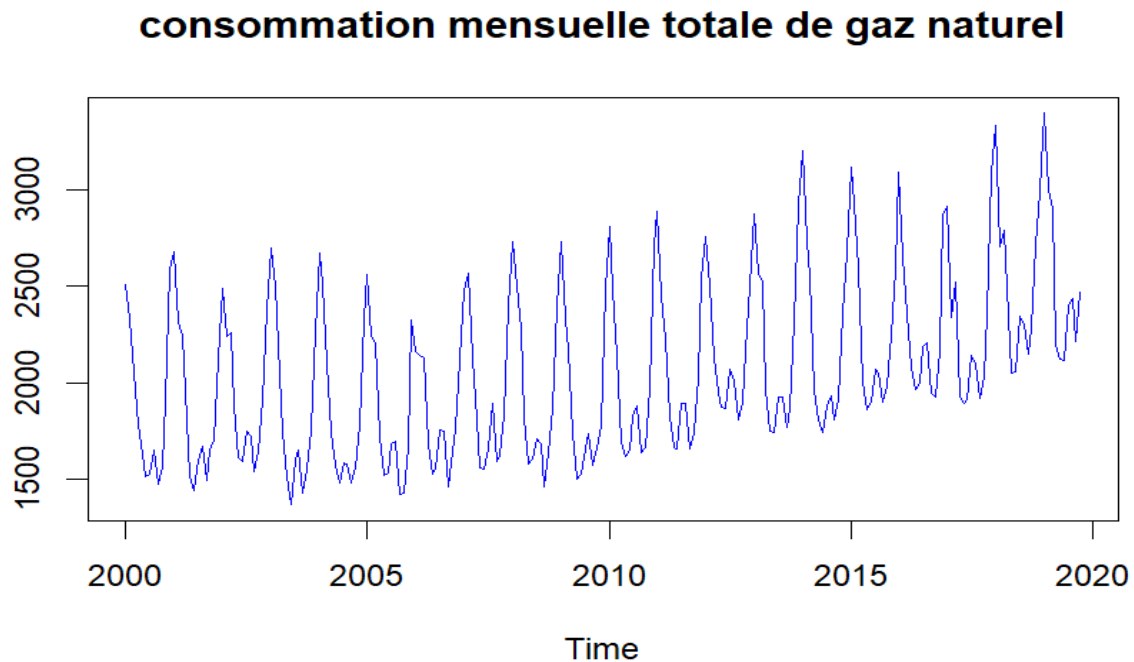


FIGURE 1.1 – Représentation graphique de la série USgas : 2000-2019

1.2 Composantes d'une série temporelle

Selon les auteurs des références [\[6, 11\]](#), une série temporelle X_t se décompose en trois éléments comme suit :

1.2.1 Tendances

La tendance ou (trend en anglais), notée Z_t est une fonction du temps qui capte l'évolution à long terme de la série. On a plusieurs types de tendances :

a)- Tendance linéaire : C'est la tendance la plus simple ([Figure 1.2](#)), dont la forme est :

$$Z_t = a + bt.$$

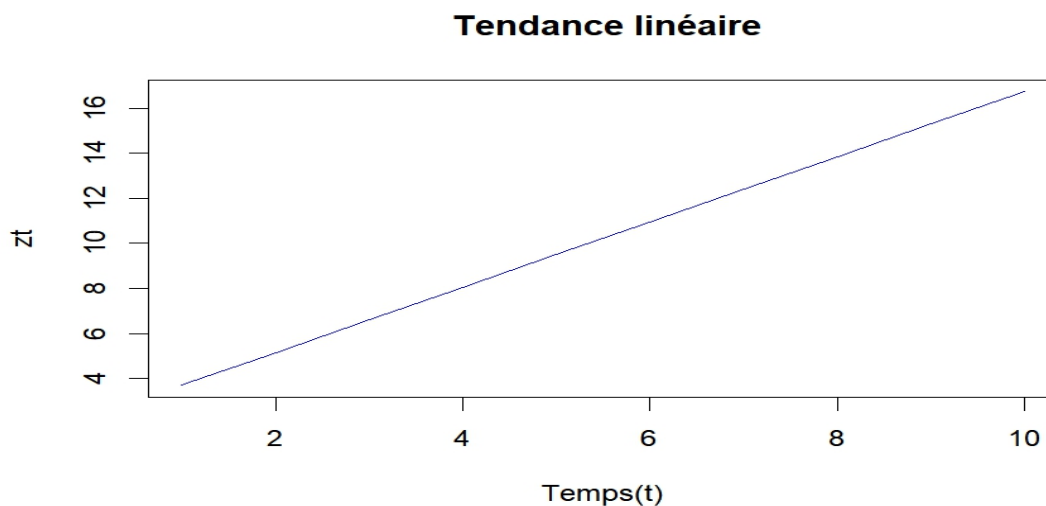


FIGURE 1.2 – Tendance linéaire

b)- **Tendance polynomiale d'ordre ($q > 1$)** : On peut ajuster la série par un polynôme d'ordre q . Les paramètres sont estimés à l'aide de la méthode des moindres carrés, ce qui correspond à une régression avec q variables explicatives :

$$Z_t = a_0 + a_1t + a_2t^2 + \dots + a_qt^q.$$

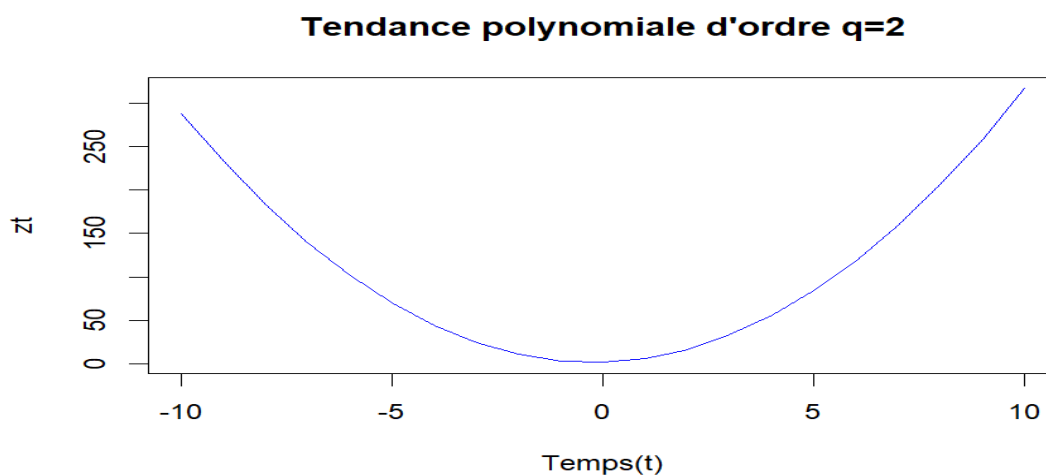


FIGURE 1.3 – Tendance polynomiale d'ordre q=2

c)- **Tendance logarithmique** : $Z_t = \log(t)$.

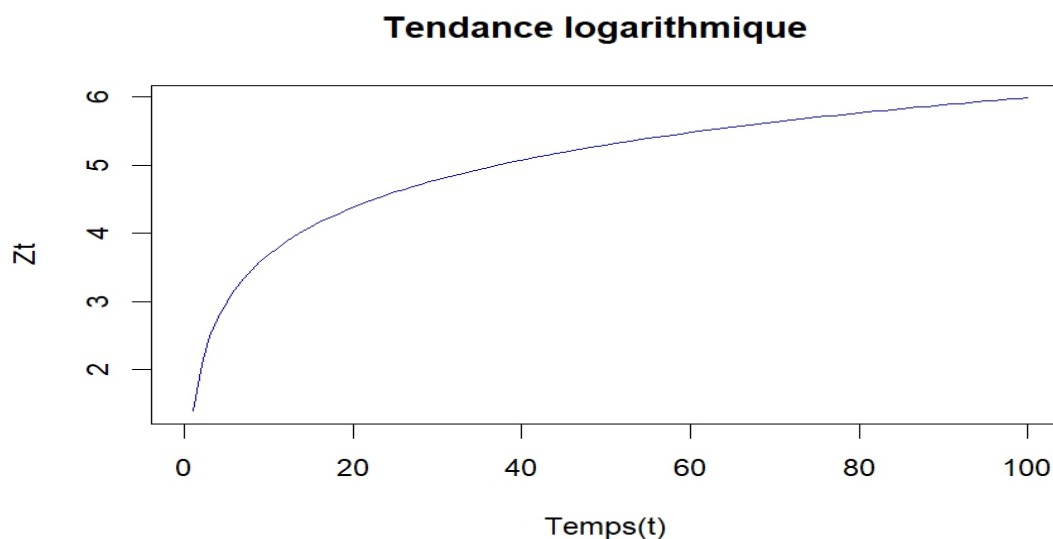


FIGURE 1.4 – Tendance logarithmique

d)- Tendance exponentielle : $Z_t = ae^{bt}$

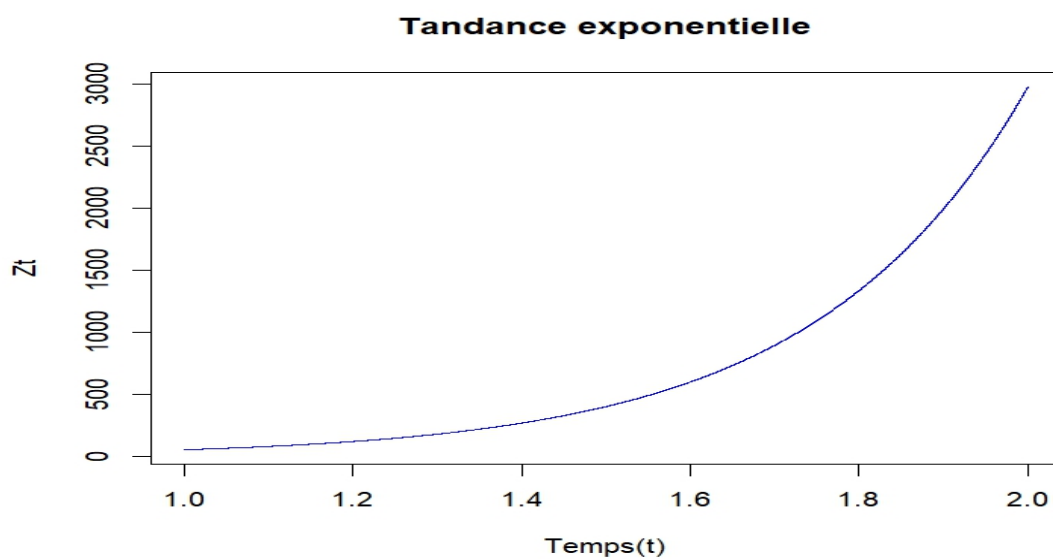


FIGURE 1.5 – Tendance exponentielle

1.2.2 Saisonnalité

La Saisonnalité représente les effets périodiques de la série, d'une période connue p qui se répètent périodiquement. Elle est notée S_t , $t=1, \dots, T$:

$$S_t = \gamma_1 S_t^1 + \gamma_2 S_t^2 + \dots + \gamma_p S_t^p$$

$\gamma_j, \forall j=1, \dots, p$: coefficients saisonniers.

$S_t^j, \forall j=1, \dots, p$: saisons : ($S_{t \pm p} = S_t$)

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

- $\sum_{j=1}^p \gamma_j = 0.$

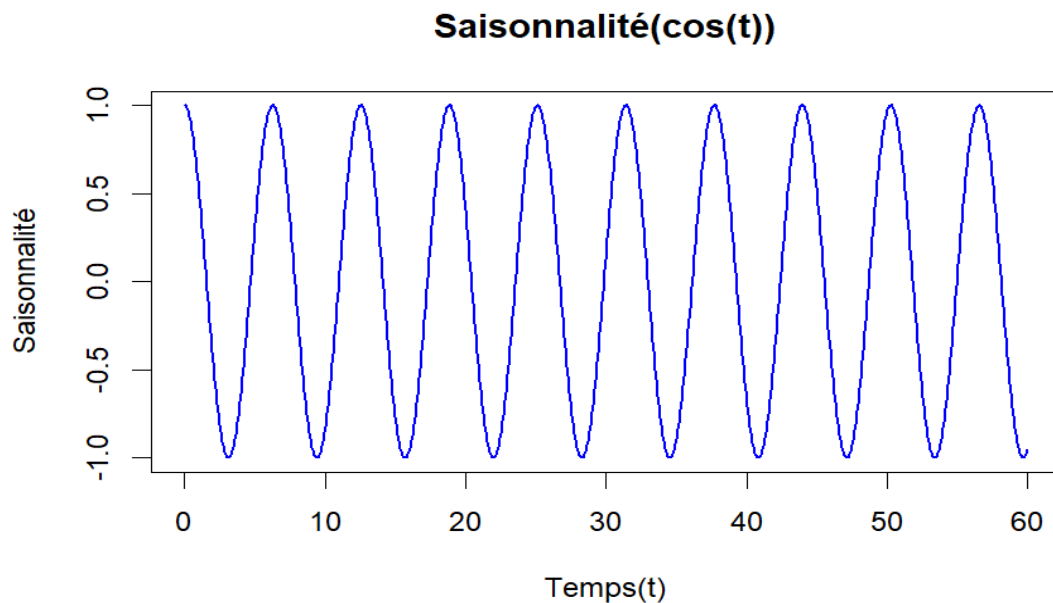


FIGURE 1.6 – Composante de la saisonnalité

1.2.3 Résidus (Bruit Blanc)

C'est une composante aléatoire d'erreur ou bruit blanc (BB) notée ε_t . Idéalement, elle est de faible variabilité par rapport aux autres composantes.

- Les erreurs sont non autocorrélées :

$$E(\varepsilon_t \varepsilon_s) = \begin{cases} \sigma^2 & \text{si } t = s \\ 0 & \text{sinon} \end{cases}$$

- $\varepsilon_t \sim BB(0, \sigma^2)$: il s'agit d'un bruit blanc, centré de variance σ^2 .

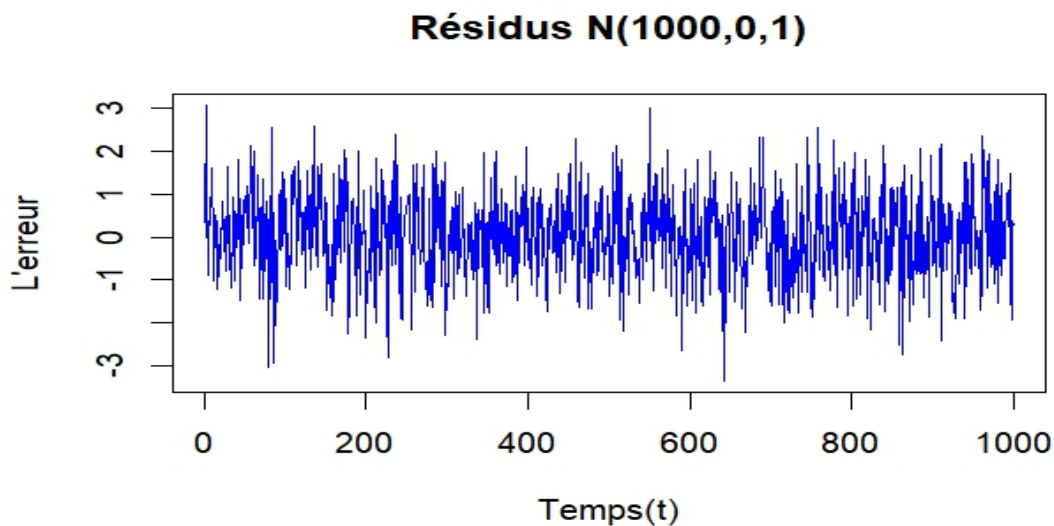


FIGURE 1.7 – Simulation d'un bruit blanc $N(0,1)$ de taille 1000.

Exemple 1.2.1. La *Figure 1.8* suivante, affiche les composantes de la série chronologique **USV-Sales**, des ventes mensuelles totales d'automobiles aux États-Unis (1976 – 2019), source [16].

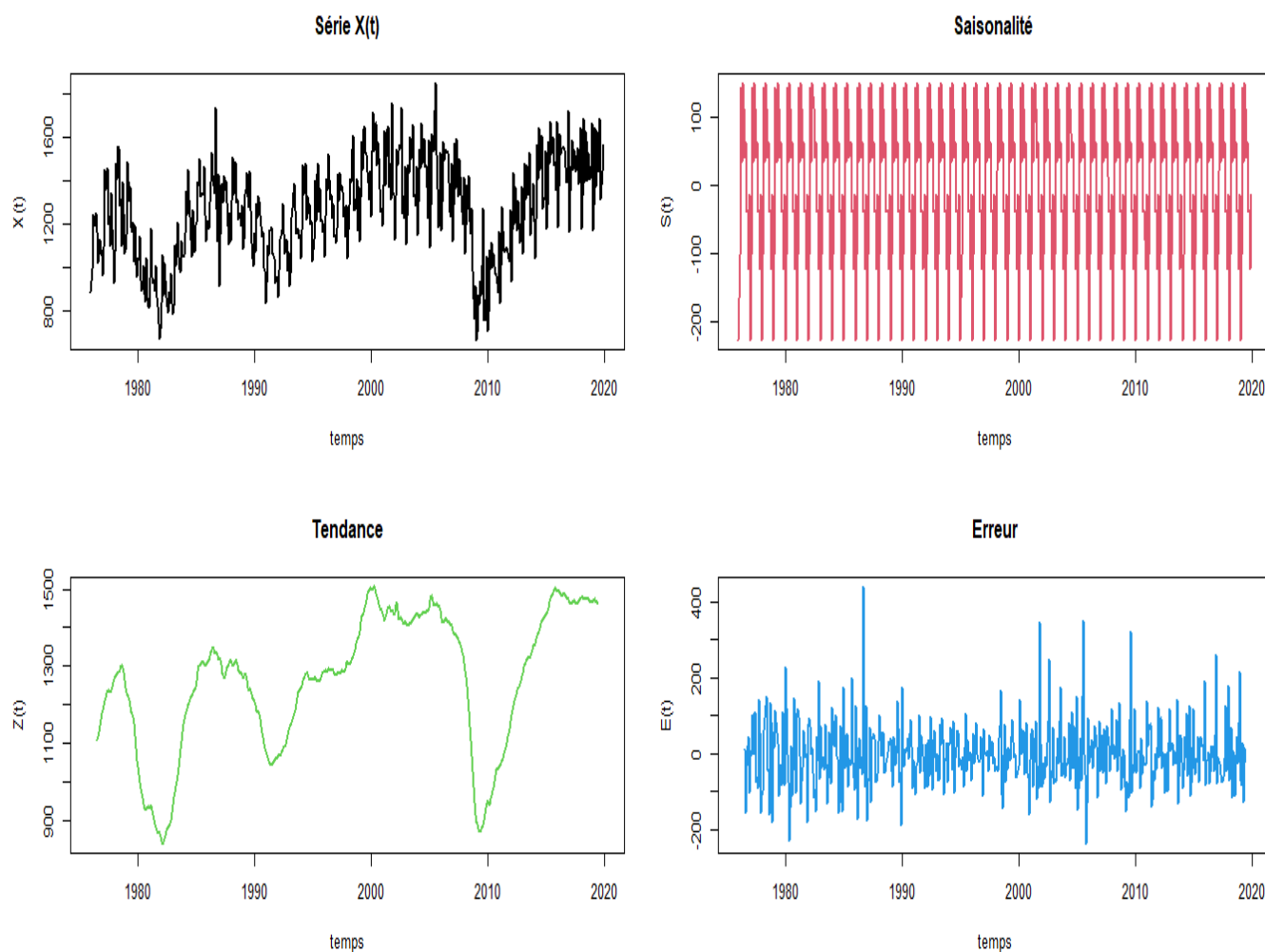


FIGURE 1.8 – Composantes de la série temporelle USV-Sales (1976 – 2019)

1.3 Schéma d'une série temporelle

D'après les références [13, 19] ; une série temporelle peut être conue à l'aide de l'un des modèles suivants :

1.3.1 Modèle additif

C'est le "modèle classique" de décomposition dans le traitement des modèles d'ajustement. La Figure 1.9 de la série des concentrations atmosphériques en CO2 est un exemple de ce modèle. La variable X_t dans le cas d'un Modèle additif s'écrit comme la somme de trois termes :

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

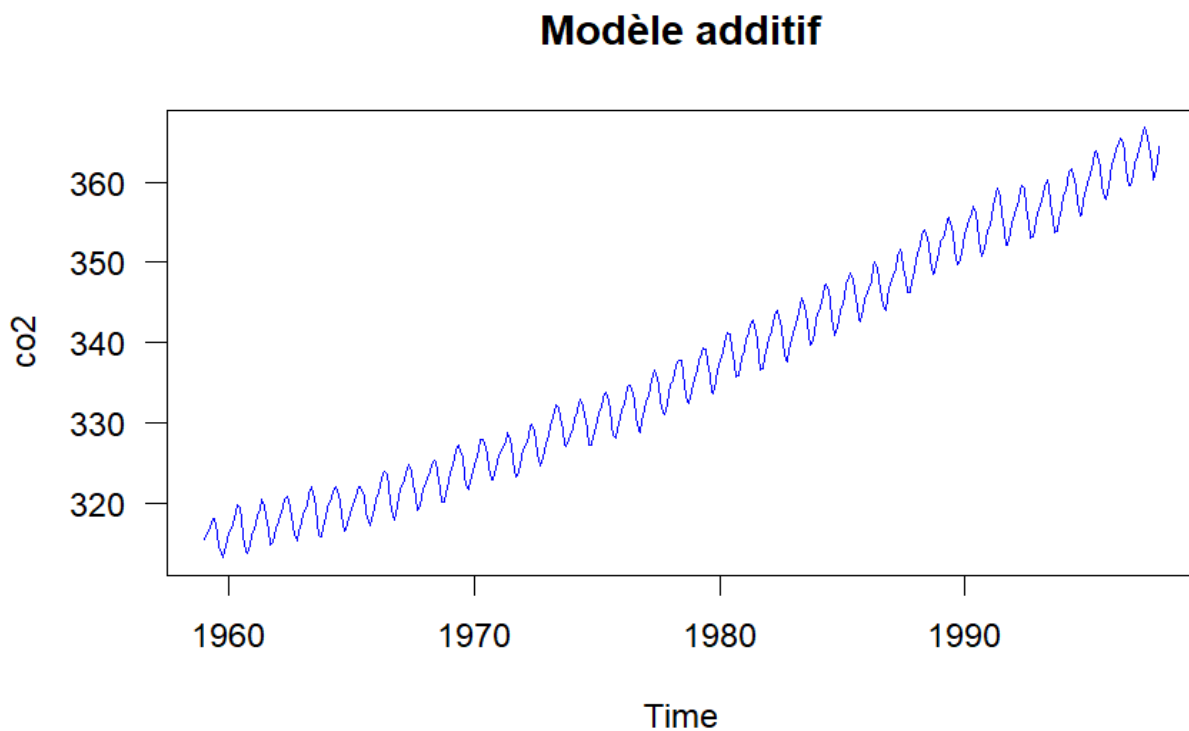


FIGURE 1.9 – Schéma additif : série co2, 1959-1997.

1.3.2 Modèle multiplicatif

La variable X_t s'écrit au terme d'erreur près comme le produit de la tendance et d'une composante de saisonnalité.

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

Comme exemple de modèle multiplicatif, la Figure 1.10 donne une illustration graphique de la série du total mensuel des passagers des compagnies aériennes internationales, de 1949 à 1960.

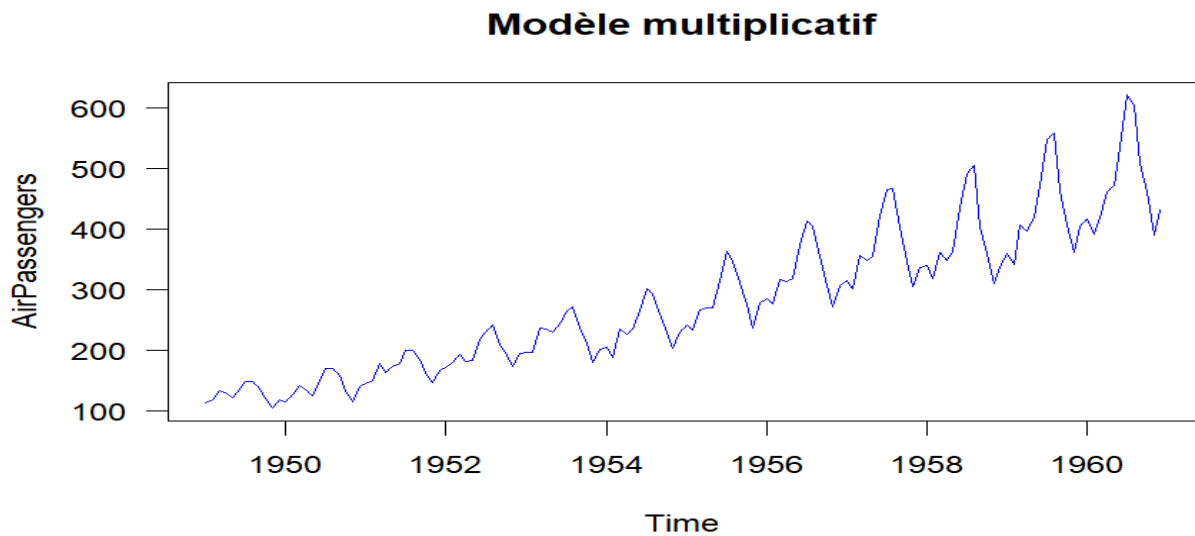


FIGURE 1.10 – Schéma multiplicatif : série AirPassengers, 1949-1960

1.3.3 Modèle mixte

Dans les modèles mixtes, comme dans la Figure 1.11 du débit annuel (en m^3) du Nil de 1871 à 1970, on retrouve les deux opérations multiplication et addition en même temps (à la fois) et cette formule est une équation à l'exemple :

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

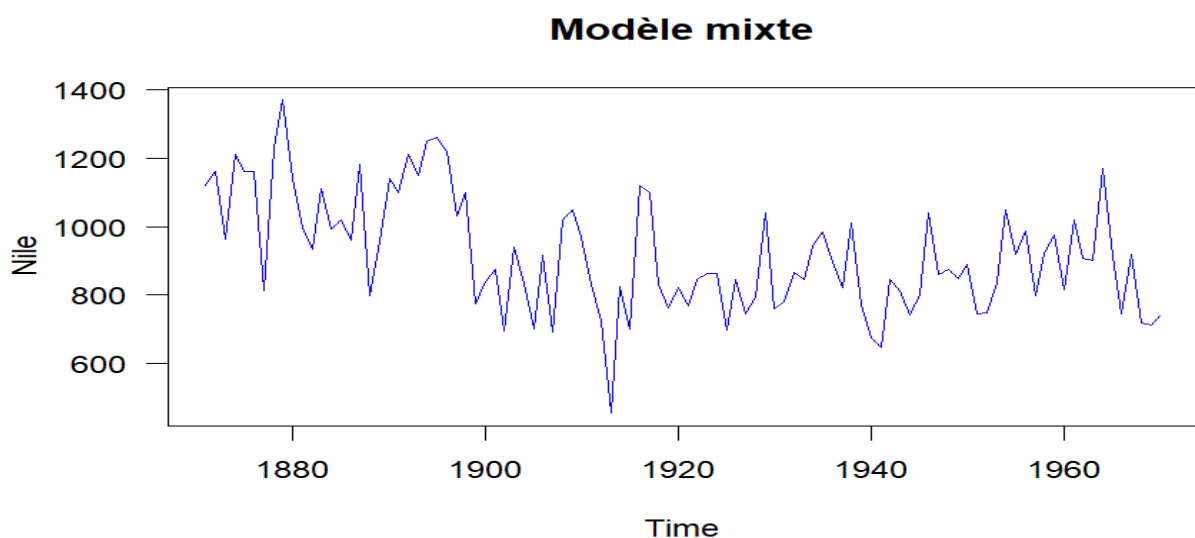


FIGURE 1.11 – Schéma mixte : série Nile, 1871-1970

1.4 Filtration

L'estimation des paramètres de la tendance et de la saisonnalité nécessite la séparation de ces deux composantes pour les évaluer facilement, pour cela nous utilisons la notion de filtration, dont on trouve plus de détails dans les références [1, 6].

1.4.1 Opérateur de retard et d'avance

L'opérateur de retard noté L est un opérateur linéaire tel que :

$$L : X_t \mapsto L(X_t) = LX_t = X_{t-1}.$$

De même, un **opérateur d'avance** noté F est tel que :

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

1.4.2 Opérateur de différentiation

L'opérateur qui fait la différence entre une suite et sa version décalée d'un unité de temps est dit opérateur de différentiation :

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

Proposition 1.4.1. *On a*

- $\Delta^d = (1 - L)^d$ permet d'éliminer les tendances Z_t , polynômiale de degré d .
- $\Delta_p = (1 - L^p)$ permet d'éliminer les saisonnalités S_t de période p .
- $\Delta^d \Delta_p = (1 - L)^d (1 - L^p)$ permet d'éliminer les Z_t et S_t pour un **schéma additif**.
- Δ_p^{d+1} permet d'éliminer les Z_t et S_t pour un **schéma mixte**.

1.5 Moyennes mobiles

Définition 1.5.1. Une **moyenne mobile** est une combinaison linéaire d'opérateurs retard [6] :

$$M = \sum_{i=-m_1}^{m_2} \theta_i L^{-i}, \quad \forall \quad m_1, m_2 \in \mathbb{N},$$

Il s'agit d'un polynôme, appelé **polynôme caractéristique** de M , de **degré** $m_1 + m_2$, et $m_1 + m_2 + 1$ sera appelé **ordre** de M .

Proposition 1.5.1. • Si $m_1 = m_2 = m$, la moyenne mobile sera dite **centrée**.

- Si M est **centrée** et que pour tout i , $\theta_i = \theta_{-i}$, alors la moyenne mobile est dite **symétrique**.
- Les moyennes mobiles d'ordre m sont définies par :

$$M_m(X_t) = \begin{cases} \frac{1}{m}(X_{t-p} + X_{t-p+1} + \cdots + X_t + \cdots + X_{t+p-1} + X_{t+p}) & \text{si } m = 2p + 1, \\ \frac{1}{m}(X_{t-p+\frac{1}{2}} + \cdots + X_{t-\frac{1}{2}} + X_{t+\frac{1}{2}} + \cdots + X_{t+p-\frac{1}{2}}) & \text{si } m = 2p. \end{cases}$$

Remarque 1.5.1. Il existe plusieurs méthodes, dont la filtration et la moyenne mobile, qui permettent d'éliminer soit S_t , soit Z_t , soit les deux en même temps. Dans le **prochain chapitre**, nous apprendrons à connaître et utiliser **la méthode de régression linéaire** en séries temporelles et à estimer les paramètres Z_t et S_t par **la méthode des moindres carrés ordinaires (MCO)**.

Chapitre 2

Régression et désaisonnalisation

Le présent chapitre décrit la méthode de régression linéaire qui permet d'éliminer la saisonnalité des séries temporelles, ainsi que l'estimation des coefficients de tendance et de saisonnalité à l'aide de la méthode des moindres carrés ordinaires (MCO). Puis nous traitons et généralisons la méthode de "**Buys-Ballot**". Enfin, nous discutons de la série corrigée des variations saisonnières, de la validation et de la prévision.

2.1 Ecriture d'un modèle linéaire

Dans un modèle linéaire, la série X_t est la somme de deux composantes déterministes : la tendance Z_t et la saisonnalité S_t et d'une composante aléatoire ε_t , [12] : $X_t = Z_t + S_t + \varepsilon_t$. On écrit Z_t et S_t sous forme de combinaisons linéaires de fonctions connues dans le temps :

$$\begin{cases} Z_t = Z_t^1 \beta_1 + Z_t^2 \beta_2 + \cdots + Z_t^m \beta_m & = \sum_{i=1}^m Z_t^i \beta_i \\ S_t = S_t^1 \gamma_1 + S_t^2 \gamma_2 + \cdots + S_t^p \gamma_p & = \sum_{j=1}^p S_t^j \gamma_j \end{cases}$$

telles que

- Z_t^i est la valeur de la i fonction explicative de la tendance, à l'instant t .

Par exemple, nous voulons modéliser la tendance linéaire où nous choisissons :

$$Z_t = Z_t^1 \beta_1 + Z_t^2 \beta_2 = \beta_1 + \beta_2 t$$

Alors :

- $Z_t^1 = 1$ (constant).
- $Z_t^2 = t$ (le temps).

$\beta_1, \dots, \beta_m$ et $\gamma_1, \dots, \gamma_p$ sont des constantes inconnues qu'il faudra estimer à partir des observations.

Ainsi, le modèle est :

$$X_t = \sum_{i=1}^m Z_t^i \beta_i + \sum_{j=1}^p S_t^j \gamma_j + \varepsilon_t, \forall t = 1, \dots, T. \quad (2.1)$$

- La tendance $\sum_{i=1}^m Z_t^i \beta_i$: est un élément du sous espace vectoriel Z engendré par les vecteurs $Z^1, \dots, Z^m$ de dimension $\mathbf{m}$.
- La saisonnalité $\sum_{j=1}^p S_t^j \gamma_j$: est un élément du sous espace vectoriel S engendré par les vecteurs $S^1, \dots, S^p$ de dimension $\mathbf{p}$.

2.2 Estimateur des moindres carrés ordinaires

La méthode des **MCO**, consiste à choisir comme estimations $\hat{\beta}_i$ et $\hat{\gamma}_j$ des paramètres β_i, γ_j ,

les valeurs qui rendent minimal le carré de la distance entre la série et sa partie déterministe :

$$\begin{aligned} (\hat{\beta}_i, \hat{\gamma}_j) &= \arg \min_{(\beta_i, \gamma_j)} \left\{ \sum_{t=1}^T \varepsilon_t^2 \right\} \\ &= \arg \min_{(\beta_i, \gamma_j)} \left\{ \sum_{t=1}^T \left[X_t - \sum_{i=1}^m Z_t^i \beta_i - \sum_{j=1}^p S_t^j \gamma_j \right]^2 \right\}. \end{aligned}$$

Notations :

$$\beta = [\beta_1, \dots, \beta_m]', \gamma = [\gamma_1, \dots, \gamma_p]', X = [X_1, \dots, X_T]', \varepsilon = [\varepsilon_1, \dots, \varepsilon_T]',$$

$$Z = \begin{bmatrix} | & & | \\ Z^1 & \dots & Z^m \\ | & & | \end{bmatrix} = [Z_t^i]_{t=1, \dots, T}^{i=1, \dots, m}, S = \begin{bmatrix} | & & | \\ S^1 & \dots & S^p \\ | & & | \end{bmatrix} = [S_t^j]_{t=1, \dots, T}^{j=1, \dots, p}$$

- β : vecteur de dimension $\mathbf{m}$.

- γ : vecteur de dimension $\mathbf{p}$.
- X : vecteur aléatoire de dimension $\mathbf{T}$.
- ε : le vecteur des erreurs de dimension $\mathbf{T}$.
- Z : une matrice de dimension $(\mathbf{T}, \mathbf{m})$.
- S : une matrice de dimension $(\mathbf{T}, \mathbf{p})$.

Nous pouvons donc écrire le modèle (2.1) sous forme matricielle comme suit :

$$\begin{aligned} X &= Z\beta + S\gamma + \varepsilon \\ &= \begin{bmatrix} Z & S \end{bmatrix} \begin{bmatrix} \beta \\ \gamma \end{bmatrix} + \varepsilon = Yb + \varepsilon \end{aligned} \quad (2.2)$$

C'est ainsi que :

- Y : matrice de dimension $(\mathbf{T}, \mathbf{m} + \mathbf{p})$.
- b : vecteur de dimension $\mathbf{m} + \mathbf{p}$.

Soit $\hat{b}$ l'estimateur des moindres carrés de b . Le modèle (2.2) est un **modèle de régression linéaire multiple**, et sa solution s'écrit donc comme suit :

$$Y'Y\hat{b} = Y'X \quad , \quad \text{soit} \quad \begin{bmatrix} Z' \\ S' \end{bmatrix} \begin{bmatrix} Z & S \end{bmatrix} \hat{b} = \begin{bmatrix} Z' \\ S' \end{bmatrix} X$$

alors,

$$\hat{b} = (Y'Y)^{-1}Y'X \quad (2.3)$$

$$\begin{bmatrix} \hat{\beta} \\ \hat{\gamma} \end{bmatrix} = \begin{bmatrix} Z'Z & Z'S \\ S'Z & S'S \end{bmatrix}^{-1} \begin{bmatrix} Z'X \\ S'X \end{bmatrix} \quad (2.4)$$

On peut déduire du système des équations normales les expressions de $\widehat{\beta}$ et $\widehat{\gamma}$.

$$Z'Z\widehat{\beta} + Z'S\widehat{\gamma} = Z'X \quad (2.5)$$

$$S'Z\widehat{\beta} + S'S\widehat{\gamma} = S'X \quad (2.6)$$

D'après l'équation (2.6), nous écrivons $\widehat{\gamma}$ en termes de $\widehat{\beta}$ pour trouver :

$$\widehat{\gamma} = (S'S)^{-1}[S'X - S'Z\widehat{\beta}].$$

En remplaçant $\widehat{\gamma}$ dans (2.5), on trouve :

$$\widehat{\beta} = [Z'Z - Z'S(S'S)^{-1}S'Z]^{-1}[Z'X - Z'S(S'S)^{-1}S'X] \quad (2.7)$$

De la même manière, nous pouvons également trouver la valeur de $\widehat{\gamma}$ en remplaçant $\widehat{\beta}$ par $\widehat{\gamma}$ dans l'équation (2.5) :

$$\widehat{\beta} = (Z'Z)^{-1}[Z'X - Z'S\widehat{\gamma}].$$

En remplaçant la valeur de $\widehat{\beta}$ dans l'équation (2.6), on trouve :

$$\widehat{\gamma} = [S'S - S'Z(Z'Z)^{-1}Z'S]^{-1}[S'X - S'Z(Z'Z)^{-1}Z'X]. \quad (2.8)$$

Remarque 2.2.1. *S'il n'y a pas d'effet saisonnier, $X = Z\beta + \varepsilon$, et on retrouve le modèle linéaire usuel, avec pour estimateur **MCO** :*

$$\widehat{\beta} = (Z'Z)^{-1}Z'X.$$

Proposition 2.2.1. *Propriétés des estimateurs [8, p :38-41] :*

1)- Les estimateurs **MCO** sont **sans biais** : $E(\widehat{b}) = E \begin{pmatrix} \widehat{\beta} \\ \widehat{\gamma} \end{pmatrix} = b = \begin{pmatrix} \beta \\ \gamma \end{pmatrix}$, ou $E(\widehat{\beta}_i) = \beta_i$ et $E(\widehat{\gamma}_j) = \gamma_j$.

2)- La variance des estimateurs peut être estimée par :

$$\widehat{Var}(\widehat{b}) = \widehat{Var} \begin{pmatrix} \widehat{\beta} \\ \widehat{\gamma} \end{pmatrix} = \widehat{\sigma}_\varepsilon^2 (Y'Y)^{-1} = \widehat{\sigma}_\varepsilon^2 \begin{bmatrix} Z'Z & Z'S \\ S'Z & S'S \end{bmatrix}^{-1}, \quad \widehat{\sigma}_\varepsilon^2 = \frac{1}{T-p-m} \sum_{t=1}^T \widehat{\varepsilon}_t^2.$$

Preuve :

1) D'après l'équation (2.3) on a : $\widehat{b} = (Y'Y)^{-1}Y'X$ et $X = Yb + \varepsilon$

Alors,

$$\begin{aligned} \widehat{b} &= (Y'Y)^{-1}Y'(Yb + \varepsilon) \\ &= (Y'Y)^{-1}Y'Yb + (Y'Y)^{-1}Y'\varepsilon \\ &= b + (Y'Y)^{-1}Y'\varepsilon \\ E(\widehat{b}) &= b + (Y'Y)^{-1}Y'E(\varepsilon) \end{aligned}$$

et puisque $E(\varepsilon) = 0_T$, il vient

$$E(\widehat{b}) = b$$

2) $\widehat{Var}(\widehat{b}) = \widehat{\sigma}_\varepsilon^2 (Y'Y)^{-1}$:

$$\begin{aligned} \widehat{Var}(\widehat{b}) &= E[(\widehat{b} - E(\widehat{b}))(\widehat{b} - E(\widehat{b}))'] \\ &= E[(\widehat{b} - b)(\widehat{b} - b)'] \quad , \quad \widehat{b} = b + (Y'Y)^{-1}Y'\varepsilon \\ &= E[(Y'Y)^{-1}Y'\varepsilon\varepsilon'Y(Y'Y)^{-1}] \\ &= (Y'Y)^{-1}Y'E(\varepsilon\varepsilon')Y(Y'Y)^{-1} \quad , \quad E(\varepsilon\varepsilon') = \widehat{\sigma}_\varepsilon^2 I_T \\ &= \widehat{\sigma}_\varepsilon^2 (Y'Y)^{-1} = \widehat{\sigma}_\varepsilon^2 \begin{bmatrix} Z'Z & Z'S \\ S'Z & S'S \end{bmatrix}^{-1} \end{aligned}$$

Remarque 2.2.2. La méthode que nous venons de décrire suppose que les vecteurs $Z^1, \dots, Z^m$

et $S^1, \dots, S^p$ constituent un système libre ; elle ne s'applique donc pas directement au modèle de **Buys-Ballot** à tendance linéaire.

2.3 Généralisation des formules de Buys-Ballot

2.3.1 Modèle trimestriel de Buys-Ballot : Cas particulier

Le modèle trimestriel de Buys-Ballot (voir [12]) à tendance linéaire s'écrit comme suit :

$$X_t = \beta_1 + \beta_2 t + \sum_{j=1}^4 \gamma_j S_t^j + \varepsilon_t \quad , \quad \forall t = 1, \dots, T.$$

Il est possible de l'écrire sous **forme de matrices** :

$$\begin{pmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \\ X_5 \\ X_6 \\ X_7 \\ X_8 \\ X_9 \\ \vdots \\ X_t \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 2 & 0 & 1 & 0 & 0 \\ 1 & 3 & 0 & 0 & 1 & 0 \\ 1 & 4 & 0 & 0 & 0 & 1 \\ 1 & 5 & 1 & 0 & 0 & 0 \\ 1 & 6 & 0 & 1 & 0 & 0 \\ 1 & 7 & 0 & 0 & 1 & 0 \\ 1 & 8 & 0 & 0 & 0 & 1 \\ 1 & 9 & 1 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & t & S_t^1 & S_t^2 & S_t^3 & S_t^4 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \gamma_1 \\ \gamma_2 \\ \gamma_3 \\ \gamma_4 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \\ \varepsilon_7 \\ \varepsilon_8 \\ \varepsilon_9 \\ \vdots \\ \varepsilon_t \end{pmatrix} \quad , \quad \text{soit} \quad X = Y\theta + \varepsilon$$

La méthode des **MCO** doit alors être appliquée en tenant compte de cette égalité. Les esti-

mateurs $\widehat{\beta}_1, \widehat{\beta}_2, \widehat{\gamma}_j$ sont obtenus comme solutions du problème de minimisation sous contrainte :

$$\begin{cases} \min_{\beta, \gamma} \sum_{t=1}^T \left[X_t - \beta_1 - \beta_2 t - \sum_{j=1}^4 \gamma_j S_t^j \right]^2 \\ \gamma_1 + \gamma_2 + \gamma_3 + \gamma_4 = 0 \end{cases}$$

Ce problème s'écrit aussi :

$$\begin{cases} \min_{\beta, \gamma} \sum_{t=1}^T \left[X_t - \beta_2 t - \sum_{j=1}^4 \delta_j S_t^j \right]^2 \\ \beta_1 = \frac{1}{4} \sum_{j=1}^4 \delta_j \quad , \quad \gamma_j = \delta_j - \beta_1 \end{cases}$$

Alors, on a les estimateurs suivants :

$$\widehat{\beta}_2 = \frac{3}{N(N^2 - 1)} \left[\sum_{n=1}^N n \bar{x}_n - \frac{N(N+1)}{2} \bar{x} \right] \quad (2.9)$$

$$\widehat{\delta}_j = \bar{x}_j - [j + 2(N-1)] \widehat{\beta}_2 \quad , \quad \forall j = 1, 2, 3, 4 \quad (2.10)$$

D'où finalement,

$$\begin{cases} \widehat{\beta}_1 = [\widehat{\delta}_1 + \widehat{\delta}_2 + \widehat{\delta}_3 + \widehat{\delta}_4] / 4 \\ \widehat{\gamma}_j = \widehat{\delta}_j - \widehat{\beta}_1 \end{cases}$$

En notant :

- T : le nombre d'observations.
- N : le nombre d'années entières $\left(N = \frac{T}{4} \right)$.
- $\bar{x}_n$: moyenne des X_t relatives à l'année n ($n = 1, \dots, N$).
- $\bar{x}_j$: moyenne de X_t relatives au saison j ($j = 1, \dots, 4$).
- $\bar{x}$: moyenne de toutes les observations X_t .
- posons : $t = 4(n-1) + j$

Exemple 2.3.1. Dans le tableau suivant, nous donnons un exemple artificiel du *cas particulier* : du modèle trimestriel de Buys-Ballot :

$\begin{matrix} j \\ n \end{matrix}$	T1	T2	T3	T4	$\bar{x}_n$
2002	4	8	14	19	11.25
2003	3	7	23	25	14.5
2004	11	13	22	24	17.5
$\bar{x}_j$	6	9.33	19.67	22.67	$\bar{x} = 14.42$

TABLE 2.1 – Exemple d'un modèle trimestriel de Buys-Ballot.

Calcul de $\hat{\beta}_2$:

$N=3$,

$$\sum_{n=1}^N n\bar{x}_n = 92.75$$

$$\hat{\beta}_2 = \frac{3}{3(3^2 - 1)} \left[92.75 - \frac{3(3+1)}{2}(14.42) \right] = 0.78$$

Calcul de $\hat{\delta}_j$: $\forall j = 1, 2, 3, 4$

$$\hat{\delta}_1 = \bar{x}_1 - [1 + 2(3-1)]\hat{\beta}_2 = 6 - [1 + 2(2)](0.78) = 2.09$$

$$\hat{\delta}_2 = \bar{x}_2 - [2 + 2(3-1)]\hat{\beta}_2 = 9.33 - [2 + 2(2)](0.78) = 4.64$$

$$\hat{\delta}_3 = \bar{x}_3 - [3 + 2(3-1)]\hat{\beta}_2 = 19.67 - [3 + 2(2)](0.78) = 14.20$$

$$\hat{\delta}_4 = \bar{x}_4 - [4 + 2(3-1)]\hat{\beta}_2 = 22.67 - [4 + 2(2)](0.78) = 16.42$$

Calcul de $\hat{\beta}_1$:

$$\hat{\beta}_1 = [\hat{\delta}_1 + \hat{\delta}_2 + \hat{\delta}_3 + \hat{\delta}_4] / 4 = [2.09 + 4.64 + 14.20 + 16.42] / 4 = 9.34$$

Calcul de $\hat{\gamma}_j$: $\forall j = 1, 2, 3, 4$

$$\hat{\gamma}_1 = \hat{\delta}_1 - \hat{\beta}_1 = 2.09 - 9.34 = -7.25$$

$$\hat{\gamma}_2 = \hat{\delta}_2 - \hat{\beta}_1 = 4.64 - 9.34 = -4.7$$

$$\hat{\gamma}_3 = \hat{\delta}_3 - \hat{\beta}_1 = 14.20 - 9.34 = 4.86$$

$$\hat{\gamma}_4 = \hat{\delta}_4 - \hat{\beta}_1 = 16.42 - 9.34 = 7.08$$

2.3.2 Modèle général de Buys-Ballot

Définition 2.3.1. *La désaisonnalisation par régression linéaire, dans le cas où la tendance est supposée linéaire, équivaut à tester le modèle linéaire [6] :*

$$X_t = \underbrace{\beta_1 + \beta_2 t}_{Z_t} + \underbrace{\gamma_1 S_t^1 + \gamma_2 S_t^2 + \gamma_3 S_t^3 + \dots + \gamma_p S_t^p}_{S_t} + \varepsilon_t.$$

On peut l'écrire sous forme matricielle :

$$\begin{pmatrix} X_1 \\ X_2 \\ X_3 \\ \vdots \\ X_t \end{pmatrix} = \begin{pmatrix} 1 & 1 & S_1^1 & S_1^2 & \dots & S_1^p \\ 1 & 2 & S_2^1 & S_2^2 & \dots & S_2^p \\ 1 & 3 & S_3^1 & S_3^2 & \dots & S_3^p \\ \vdots & \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & t & S_t^1 & S_t^2 & \dots & S_t^p \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \gamma_1 \\ \vdots \\ \gamma_p \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \vdots \\ \varepsilon_t \end{pmatrix}, \quad X = Y\theta + \varepsilon$$

Remarque 2.3.1. *L'estimateur des moindres carrés ordinaires de θ s'écrit $\hat{\theta} = (Y'Y)^{-1}Y'X$.*

Toutefois, cette écriture n'est possible que si $Y'Y$ est inversible, ce qui n'est pas le cas ici car la première colonne (correspondant à la constante) est égale à la somme des $\mathbf{p}$ dernières (les composantes trimestrielles). Deux méthodes sont alors possibles pour faire malgré tout l'identification du modèle.

- ne pas tenir compte de la constante, et identifier le modèle :

$$X_t = \beta_2 t + \gamma_1 S_t^1 + \cdots + \gamma_p S_t^p$$

- rajouter une contrainte, et identifier le modèle :

$$\begin{cases} X_t = \beta_1 + \beta_2 + \sum_{j=1}^p \gamma_j S_t^j \\ \sum_{j=1}^p \gamma_j = 0 \end{cases} \quad (2.11)$$

L'estimateur des paramètres $\theta = [\beta_1, \beta_2, \gamma_j]'$ en utilisant **la méthode des moindres carrés** où :

$$\hat{\beta}_2 = \frac{12}{pN(N^2 - 1)} \left[\sum_{n=1}^N n \bar{x}_n - \frac{N(N+1)}{2} \bar{x} \right] \quad (2.12)$$

$$\hat{\beta}_1 = \bar{x} - \hat{\beta}_2 \frac{Np+1}{2} \quad (2.13)$$

$$\hat{\gamma}_j = \bar{x}_j - \bar{x} - \hat{\beta}_2 \left[j - \frac{p+1}{2} \right] \quad (2.14)$$

En notant :

- N : le nombre d'années entières ($N = \frac{T}{p}$).
- p : le nombre de saisons pour chaque année .
- $\bar{x}_n$: moyenne des X_t relatives à l'année n ($n = 1, \dots, N$) .
- $\bar{x}_j$: moyenne de X_t relatives au saison j ($j = 1, \dots, p$) .
- $\bar{x}$: moyenne de toutes les observations X_t .
- posons : $t = p(n-1) + j$

On a :

$$\begin{cases} \arg \min_{(\beta_1, \beta_2, \gamma_j)} \sum_{t=1}^T \left[X_t - \beta_1 - \beta_2 t - \sum_{j=1}^p \gamma_j S_t^j \right]^2 \\ \sum_{j=1}^p \gamma_j = 0 \end{cases}$$

1- Estimation de $\hat{\beta}_1$:

On pose

$$S(\beta_1, \beta_2, \gamma_j) = \sum_{t=1}^T \left[X_t - \beta_1 - \beta_2 t - \sum_{j=1}^p \gamma_j S_t^j \right]^2 \quad (2.15)$$

La dérivée de l'équation (2.15) par rapport β_1 , donne :

$$\begin{aligned} \frac{\partial S}{\partial \beta_1} = 0 &\Leftrightarrow -2 \sum_{t=1}^T \left[X_t - \hat{\beta}_1 - \hat{\beta}_2 t - \sum_{j=1}^p \hat{\gamma}_j S_t^j \right] = 0 \\ &\Leftrightarrow \sum_{t=1}^T X_t = \hat{\beta}_1 T + \hat{\beta}_2 \sum_{t=1}^T t + \sum_{t=1}^T \sum_{j=1}^p \hat{\gamma}_j S_t^j \end{aligned}$$

On a :

$$\bullet \sum_{t=1}^T t = \frac{T(T+1)}{2}$$

$$\bullet \sum_{j=1}^p \sum_{t=1}^T \hat{\gamma}_j S_t^j = 0 \quad , \text{ car } \sum_{j=1}^p \gamma_j = 0 \text{ et } S_t^j = \begin{cases} 1 & \text{si } t = j \\ 0 & \text{sinon} \end{cases}$$

$$\bullet T = Np \quad , \frac{1}{T} \sum_{t=1}^T X_t = \bar{x}$$

Alors,

$$\sum_{t=1}^T X_t = \hat{\beta}_1 T + \hat{\beta}_2 \sum_{t=1}^T t \quad (2.16)$$

$$\Rightarrow \hat{\beta}_1 = \bar{x} - \hat{\beta}_2 \frac{Np+1}{2}$$

2- Estimation de $\hat{\gamma}_j$: dériver l'équation (2.15) par rapport γ_j :

$$\begin{aligned} \frac{\partial S}{\partial \gamma_j} = 0 &\Leftrightarrow -2 \sum_{t=1}^T S_t^j \left[X_t - \hat{\beta}_1 - \hat{\beta}_2 t - \sum_{j=1}^p \hat{\gamma}_j S_t^j \right] = 0 \\ &\Leftrightarrow \sum_{t=1}^T X_t S_t^j = \hat{\beta}_1 \sum_{t=1}^T S_t^j + \hat{\beta}_2 \sum_{t=1}^T t S_t^j + \sum_{t=1}^T \sum_{j=1}^p \hat{\gamma}_j (S_t^j)^2 \end{aligned}$$

On a :

$$\begin{aligned} \bullet \sum_{t=1}^T S_t^j &= N = \frac{T}{p} \quad \text{et} \quad \sum_{t=1}^T \sum_{j=1}^p \hat{\gamma}_j S_t^j = \sum_{n=1}^N \hat{\gamma}_j = \hat{\gamma}_j N \\ \bullet \sum_{t=1}^T t S_t^j &= \frac{pN(N-1)}{2} + jN = N\bar{t}_j \quad \text{et} \quad \sum_{t=1}^T X_t S_t^j = \sum_{n=1}^N X_{j+(n-1)p} = N\bar{x}_j \end{aligned}$$

Donc :

$$\sum_{t=1}^T X_t S_t^j = \hat{\beta}_1 N + \hat{\beta}_2 N\bar{t}_j + \hat{\gamma}_j N$$

$$\hat{\gamma}_j = \left[\bar{x}_j - \hat{\beta}_1 - \hat{\beta}_2 \bar{t}_j \right]$$

remplaçons $\hat{\beta}_1$ dans l'équation (2.13), on trouve

$$\hat{\gamma}_j = \bar{x}_j - \bar{x} - \hat{\beta}_2 \left[j - \frac{p+1}{2} \right]$$

3- Estimation de $\hat{\beta}_2$: dériver l'équation (2.15) par rapport β_2 :

$$\frac{\partial S}{\partial \beta_2} = 0 \Leftrightarrow -2 \sum_{t=1}^T t \left[X_t - \hat{\beta}_1 - \hat{\beta}_2 t - \sum_{j=1}^p \hat{\gamma}_j S_t^j \right] = 0$$

$$\Leftrightarrow \sum_{t=1}^T t X_t = \hat{\beta}_1 \sum_{t=1}^T t + \hat{\beta}_2 \sum_{t=1}^T t^2 + \sum_{t=1}^T \sum_{j=1}^p t S_t^j \hat{\gamma}_j$$

$$\sum_{t=1}^T t X_t = \hat{\beta}_1 \sum_{t=1}^T t + \hat{\beta}_2 \sum_{t=1}^T t^2 \tag{2.17}$$

On a :

$$\bullet \sum_{t=1}^T t^2 = \frac{T(T+1)(2T+1)}{6}$$

$$\bullet \sum_{t=1}^T tX_t = N \sum_{j=1}^p j\bar{x}_j + p^2 \sum_{n=1}^N n\bar{x}_n - pT\bar{x}$$

En utilisant l'équation (2.16), on trouve finalement :

$$\hat{\beta}_2 = \frac{12}{pN(N^2-1)} \left[\sum_{n=1}^N n\bar{x}_n - \frac{N(N+1)}{2} \bar{x} \right].$$

Exemple 2.3.2. Cas générales Buys-Ballot : Considérons la série *USgas* de la consommation mensuelle de gaz naturel aux *états-Unis* (2000-2019) :

$\begin{smallmatrix} j \\ n \end{smallmatrix}$	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	$\bar{x}_n$	$n\bar{x}_n$
2000	2510.5	2330.7	2050.6	1783.3	1632.9	1513.1	1525.6	1653.1	1475.0	1567.8	1908.5	2587.5	1878.22	1878.22
2001	2677.0	2309.5	2246.6	1807.2	1522.4	1444.4	1598.1	1669.2	1494.1	1649.1	1701.0	2120.2	1853.23	3706.47
2002	2487.6	2242.4	2258.4	1881.0	1611.5	1591.4	1748.4	1725.7	1542.2	1645.9	1913.6	2378.9	1918.92	5756.75
2003	2700.5	2500.3	2197.9	1743.5	1514.7	1368.4	1600.5	1651.6	1428.6	1553.2	1753.6	2263.7	1856.38	7425.50
2004	2675.8	2511.1	2100.9	1745.2	1573.0	1483.7	1584.9	1578.0	1482.2	1557.2	1782.8	2327.7	1866.88	9334.38
2005	2561.9	2243.0	2205.8	1724.9	1522.6	1534.1	1686.6	1695.1	1422.5	1428.2	1663.4	2326.4	1834.54	11007.25
2006	2165.3	2144.4	2126.4	1681.0	1526.3	1550.9	1758.7	1751.7	1462.1	1644.2	1765.4	2122.8	1808.27	12657.87
2007	2475.6	2567.0	2128.8	1810.1	1559.1	1555.2	1659.9	1896.1	1590.5	1627.8	1834.5	2399.2	1925.32	15402.53
2008	2734.0	2503.4	2278.2	1823.9	1576.4	1604.2	1708.6	1682.9	1460.9	1635.8	1868.9	2399.7	1939.74	17457.68
2009	2729.7	2332.5	2170.7	1741.3	1504.0	1527.8	1658.0	1736.5	1575.0	1666.5	1776.2	2491.9	1909.18	19091.75
2010	2809.8	2481.0	2142.9	1691.8	1617.3	1649.5	1825.8	1878.9	1637.5	1664.9	1973.3	2714.1	2007.23	22079.58
2011	2888.6	2452.4	2230.5	1825.0	1667.4	1657.3	1890.5	1891.8	1655.6	1744.5	2031.9	2541.9	2039.78	24477.40
2012	2756.2	2500.7	2127.8	1953.1	1873.8	1868.4	2069.8	2008.8	1807.2	1901.1	2167.8	2503.9	2128.22	27666.82
2013	2878.8	2567.2	2521.1	1967.5	1752.5	1742.9	1926.3	1927.4	1767.0	1866.8	2316.9	2920.8	2179.60	30514.40
2014	3204.1	2741.2	2557.9	1961.7	1810.2	1745.4	1881.0	1933.1	1809.3	1912.8	2357.5	2679.2	2216.12	33241.75
2015	3115.0	2925.2	2591.3	2007.9	1858.1	1899.9	2067.7	2052.7	1901.3	1987.3	2249.1	2588.2	2270.31	36324.93
2016	3091.7	2652.3	2356.3	2083.8	1965.8	2000.7	2186.6	2208.4	1947.8	1925.2	2159.4	2866.3	2287.03	38879.43
2017	2914.2	2340.6	2523.7	1932.5	1892.5	1910.9	2142.1	2094.3	1920.9	2032.0	2357.7	3084.5	2262.16	40718.85
2018	3335.0	2705.9	2792.6	2346.3	2050.9	2058.7	2344.6	2307.7	2151.5	2279.1	2709.9	2993.1	2506.28	47619.23
2019	3399.9	2999.2	2899.9	2201.1	2121.0	2115.2	2407.5	2437.2	2215.6	2472.3	0	0	2526.89	50537.80
$\bar{x}_j$	2805.56	2502.50	2325.42	1885.61	1707.62	1691.11	1863.56	1889.01	1687.34	1788.09	2015.34	2542.63	$\bar{x} = 2056.80$	

TABLE 2.2 – Cas général Buys-Ballot pour la série USgas

N=20 , p=12

$$\sum_{n=1}^N n\bar{x}_n = 1878.22 + \dots + 50537.80 = 455778.5$$

Calcul de $\hat{\beta}_2$:

$$\hat{\beta}_2 = \frac{12}{12 \times 20(20^2 - 1)} \left[455778.5 - \frac{20(20 + 1)}{2} \times 2056.80 \right] = 2.99$$

Calcul de $\hat{\beta}_1$:

$$\hat{\beta}_1 = 2056.80 - (2.99) \times \frac{20 \times 12 + 1}{2} = 1696.63$$

Calcul de $\hat{\gamma}_j$: $\forall j = 1, \dots, 12$

$$\hat{\gamma}_1 = \bar{x}_1 - \bar{x} - \hat{\beta}_2 \times \left(1 - \frac{13}{2} \right) = 2805.56 - 2056.80 - 2.99 \times \left(1 - \frac{13}{2} \right) = 765.20$$

$$\hat{\gamma}_2 = \bar{x}_2 - \bar{x} - \hat{\beta}_2 \times \left(2 - \frac{13}{2} \right) = 2502.50 - 2056.80 - 2.99 \times \left(2 - \frac{13}{2} \right) = 459.15$$

$$\hat{\gamma}_3 = \bar{x}_3 - \bar{x} - \hat{\beta}_2 \times \left(3 - \frac{13}{2} \right) = 2325.42 - 2056.80 - 2.99 \times \left(3 - \frac{13}{2} \right) = 279.08$$

En calculant $\hat{\gamma}_j$, $\forall j = 1, \dots, 12$ de la même manière, nous obtenons :

$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\gamma}_3$	$\hat{\gamma}_4$	$\hat{\gamma}_5$	$\hat{\gamma}_6$	$\hat{\gamma}_7$	$\hat{\gamma}_8$	$\hat{\gamma}_9$	$\hat{\gamma}_{10}$	$\hat{\gamma}_{11}$	$\hat{\gamma}_{12}$
765.20	459.15	279.08	-163.72	-344.69	-364.197	-194.73	-172.27	-376.93	-279.17	-54.91	469.40

TABLE 2.3 – Estimation des coefficients saisonniers

2.4 Le principe de la conservation des aires

L'analyse de la saisonnalité a pour but de trouver une nouvelle répartition du profil *« intra – annuel »* de l'historique, sans modifier le niveau atteint en cumul annuel : les moyennes annuelles de la série brute et de la série **Corrigée des Variations Saisonnières (CVS)** doivent être identiques. Ce principe de base appelé le principe de la conservation des aires permet de calculer des coefficients saisonniers définitifs, voir [2, 14] pour plus de détails.

2.4.1 Le cas d'un schéma additif

Soit γ_j le j -ème coefficient provisoire, on calcule la somme des coefficients saisonniers :

- Si $S = \sum_{j=1}^p \gamma_j = 0$, les γ_j sont les coefficients saisonniers définitifs.
- Si $S \neq 0$, les coefficients saisonniers sont normés afin que leur somme soit nulle.

Les coefficients définitifs sont donnés par :

$$\gamma_j^* = \gamma_j - \bar{S} \quad , \quad \text{tels que} \quad \bar{S} = \frac{1}{p} \sum_{j=1}^p \gamma_j. \quad (2.18)$$

2.4.2 Le cas d'un schéma multiplicatif

Dans le cas d'un schéma multiplicatif la moyenne des coefficients, pour une année donnée, doit être égale à 1. En effet,

- Si $\bar{S} = \frac{1}{p} \sum_{j=1}^p \gamma_j = 1$, les γ_j sont les coefficients saisonniers provisoires.
- Si $\bar{S} \neq 1$, les coefficients définitifs sont donnés par :

$$\gamma_j^* = \gamma_j / \bar{S}. \quad (2.19)$$

2.4.3 La série corrigée des variations saisonnières

La correction des variations saisonnières et des effets de jours ouvrables (**CVS**) est un traitement statistique de la série brute (i.e., initiale) mensuelle ou trimestrielle qui vise à en éliminer les composantes cycliques (saisonnalité, nombre de jours ouvrables, année bissextile) pour permettre de mieux discerner son évolution fondamentale.

Par exemple, la valeur de la consommation d'électricité augmente toujours fortement en été, par rapport aux autres saisons. Une fois la série désaisonnalisée, c'est-à-dire en supprimant la "saison estivale", la consommation d'électricité diminue.

Modèle additif :

$$X_t^{CVS} = X_t - \gamma_j^*. \quad (2.20)$$

Modèle multiplicatif :

$$X_t^{CVS} = X_t / \gamma_j^* \quad (2.21)$$

2.5 Désaisonnalisation de la série temporelle

Lorsque qu'une série temporelle est structurée par une saisonnalité, les comparaisons inter-temporelles du phénomène nécessitent une chronique Corrigée des Variations Saisonnières notée **CVS**. Désaisonnaliser une chronique, c'est éliminer la saisonnalité sans modifier les autres composantes de la chronique. C'est une opération délicate ce qui explique le grand nombre de méthodes de désaisonnalisation. Lorsqu'on procède à la désaisonnalisation de la série, on considère que la composante aléatoire ou résiduelle (ε_t) a un effet nul. Donc $X_t = Z_t + S_t$ (modèle additif) ou $X_t = Z_t \times S_t$ (modèle multiplicatif).

2.5.1 Etapes de la désaisonnalisation

- 1) On calcule les coefficients de tendance de la série en utilisant la méthode des moindres carrés (régression linéaire simple).
- 2) On calcule les coefficients saisonniers définitifs γ_j^* .
- 3) Une fois tous les coefficients saisonniers (définitifs) calculés, on obtient la série corrigée des variations saisonnières (**CVS**) notée X_t^{CVS} .

Exemple 2.5.1. Modèle additif (série USgas) :

1) Dans l'exemple (2.3.2) précédent, nous avons calculé les coefficients de tendance et de sa-

sonnalité, Où : $\sum_{j=1}^{12} \gamma_j = 22.22 \neq 0$

2) Calcul des coefficients saisonniers définitifs γ_j^* :

$$\bar{S} = \frac{1}{p} \sum_{j=1}^p \gamma_j = \frac{1}{12} \sum_{j=1}^{12} \gamma_j = 1.85$$

$$\gamma_1^* = \gamma_1 - \bar{S} = 763.35$$

$$\gamma_2^* = \gamma_2 - \bar{S} = 457.30$$

$$\gamma_3^* = \gamma_3 - \bar{S} = 277.23$$

De la même manière que pour le reste, nous obtenons les résultats suivants :

γ_1^*	γ_2^*	γ_3^*	γ_4^*	γ_5^*	γ_5^*	γ_6^*	γ_7^*	γ_8^*	γ_9^*	γ_{10}^*	γ_{11}^*	γ_{12}^*
763.35	457.30	277.23	-165.57	-346.54	-366.05	-196.58	-174.12	-378.78	-281.02	-281.02	-56.76	467.55

TABLE 2.4 – Coefficients saisonniers définitifs du modèle additif (série USgas)

3) La Figure 2.1 présente la série corrigée (CVS) de la série *USgas*.

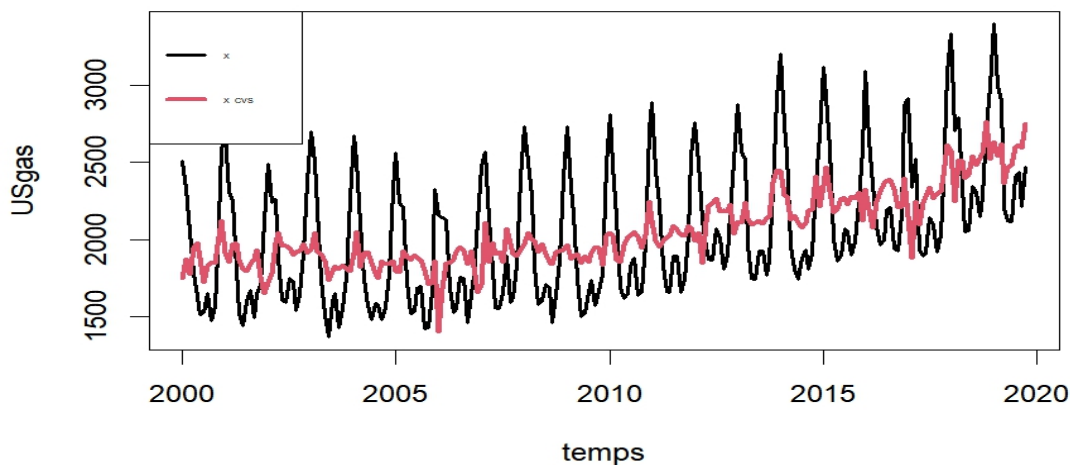


FIGURE 2.1 – Série CVS du Modèle additif (série USgas)

2.6 Validation

La validation croisée des séries chronologiques apparaît comme une méthodologie essentielle, adaptée pour respecter l'ordre temporel inhérent aux données. L'objectif principal de la validation

croisée des séries chronologiques est d'évaluer les performances prédictives d'un modèle d'une manière qui reflète son application future. Ceci est crucial dans divers domaines, tels que la finance, la météorologie et l'épidémiologie, où les décisions reposent sur des prévisions, comme cela été signalé dans [1].

2.6.1 Significativité de la régression

Le coefficient de détermination, R^2 , (**Multiple R-squared**) est défini de façon unique quand il y a une constante dans la régression par

$$R^2 = \frac{\sum_{t=1}^T (\hat{X}_t - \bar{X})^2}{\sum_{t=1}^T (X_t - \bar{X})^2} = 1 - \frac{\sum_{t=1}^T \hat{\varepsilon}_t^2}{\sum_{t=1}^T (X_t - \bar{X})^2} = \frac{SCE}{SCT}.$$

Notons :

- $SCT = \sum_{t=1}^T (X_t - \bar{X})^2$: somme des carrés totale (variation totale des X_t).
- $SCE = \sum_{t=1}^T (\hat{X}_t - \bar{X})^2$: somme des carrés expliqués (variation expliquée).

Ce n'est donc pas un indicateur pertinent de significativité de la régression. Il n'est intéressant que pour comparer des régressions portant sur un même nombre d'observations et de variables. Par contre le **coefficient de détermination ajusté (Adjusted R-squared)** prend en compte le nombre d'observations et de variables et il est donné par :

$$R_{adj}^2 = 1 - (1 - R^2) \frac{T - 1}{T - (m + p)}.$$

2.6.2 Significativité globale

La régression est dite significative, si au moins un des coefficients des variables autres que la constante est non nul. L'hypothèse nulle est : (la régression n'est pas significative), c'est-à-dire :

$$\begin{cases} H_0 : b_i = 0, \forall i=1, \dots, m+p \\ H_1 : b_i \neq 0 \text{ (au moins un de ces coefficients est non nul)} \end{cases}$$

2.6.3 Test de Fisher

La statistique pour tester l'hypothèse $H_0 : (b = 0)$ de significativité globale est :

$$F = \frac{SCE/(m+p) - 1}{SCR/T - (m+p)}.$$

Notons :

- $SCR = \sum_{t=1}^T (X_t - \hat{X}_t)^2$: somme carrés résiduelle (variation des résidus).

Sous H_0 , F suit une **loi de Fisher** à $(m+p-1, T-(m+p))$ **ddl** : $F \sim \mathcal{F}_\alpha^{(m+p-1, T-(m+p))}$.

On rejette l'hypothèse nulle pour de **grandes valeurs de la statistique de test**.

2.6.4 Test de Student

Pour tester la significativité locale d'un coefficient b_i , nous utilisons le test de Student, dont les hypothèses sont

$$\begin{cases} H_0 : b_i = \beta_0 \\ H_1 : b_i \neq \beta_0 (b_i < \beta_0 \text{ ou } b_i > \beta_0) \end{cases}$$

$\hat{b}$ est un sous-vecteur d'un vecteur gaussien est lui-même gaussien, $\hat{b}_i \sim \mathcal{N}(b_i, \sigma_i^2) \forall i = 1, \dots, m+p$

où σ_i^2 est l'élément (i,i) de la matrice $\hat{\sigma}_\varepsilon^2(Y'Y)^{-1}$. Ainsi, la statistique du test est

$$T_i = \frac{(\hat{b}_i - \beta_0)}{\hat{\sigma}(\hat{b}_i)} \sim \mathcal{T}_\alpha^{(T-(m+p))}.$$

- $\hat{\sigma}(\hat{b}_i)$: écart-type de $\hat{b}_i$.
- T_i suit une loi de Student à $(T - (m+p))$ **ddl**.

On rejette l'hypothèse H_0 pour de **grandes valeurs de la statistique de test** T_i .

2.7 Prévision

Les estimateurs $\hat{\beta}_i, \hat{\gamma}_j$ peuvent servir pour prévoir une valeur de X_t non encore observée :

$X_{T+h}, \quad h \geq 1.$

Si on suppose que le modèle est encore vrai à l'instant $T + h$, on a d'après [12] :

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

et la valeur X_{T+h} peut être approchée par :

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

Chapitre 3

Application aux données réelles

Ce dernier chapitre est consacré à l'application des résultats théorique du deuxième chapitre aux données réelles. En effet, nous essayons d'appliquer la méthode de désaisonnalisation par régression à des données sur la consommation de gaz naturel aux États-Unis, dont le modèle est additif, et aux données sur les tarifs aériens en Europe, pour un modèle multiplicatif. Cette étude est réalisée par le logiciel de programmation statistique R.

3.1 Consommation de gaz naturel aux USA

3.1.1 Description de la série

La série temporelle présentée dans la figure (Figure 3.1) est l'évolution au cours du temps des données mensuelles de consommation de gaz naturel aux **États-Unis d'Amérique** non désaisonnalisées du 1-01-2000 au 1-01-2025 (301 observations). Ces données ont été modifiées pour la dernière fois le 12-03-2025 à 9 : 02 am. (Source : U.S. Bureau of Transportation Statistics <https://fred.stlouisfed.org/series/NATURALGAS>).

Nous essayons dans la suite de décrire et comprendre le comportement passé de cette série afin de réaliser des prévisions futures. Pour cela, il faut identifier les tendances et les saisonnalités, préciser le modèle et estimer ces paramètres.

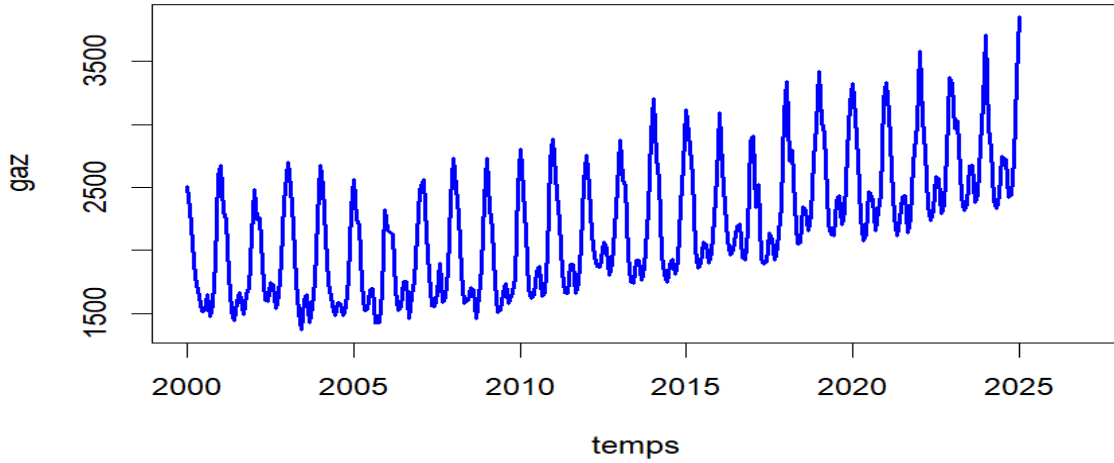


FIGURE 3.1 – Série de consommation mensuelle du gaz naturel aux USA, 2000-2024

La (Figure 3.1) révèle que la série est **un modèle additif** :

$$X_t = \beta_1 t + \sum_{j=1}^{p=12} \gamma_j S_t^j + \varepsilon_t, \quad \forall t = 1, \dots, T = 301.$$

Composantes de la série :

Cette série temporelle se compose en trois éléments selon un modèle additif, comme le montre la Figure 3.2. Ces composantes principales expliquent son comportement de la façon suivante :

- **Tendance** Z_t : Evolution générale à long terme de la série. Ici nous remarquons une augmentation régulière de la consommation du gaz naturel sur plusieurs années (2000-2024), selon un droite linéaire $a + bt$.
- **Saisonnalité** S_t : Il s'agit du Comportement qui se répète à intervalles réguliers (chaque mois) de la série. La consommation de gaz naturel augmente chaque hiver et diminue considérablement en été, de manière récurrente chaque année.
- **Erreur** ε_t : C'est l'irrégularité ou le bruit blanc, il représentent les variations imprévisibles dues au hasard ou à des événements exceptionnels dans la consommation de gaz naturel.

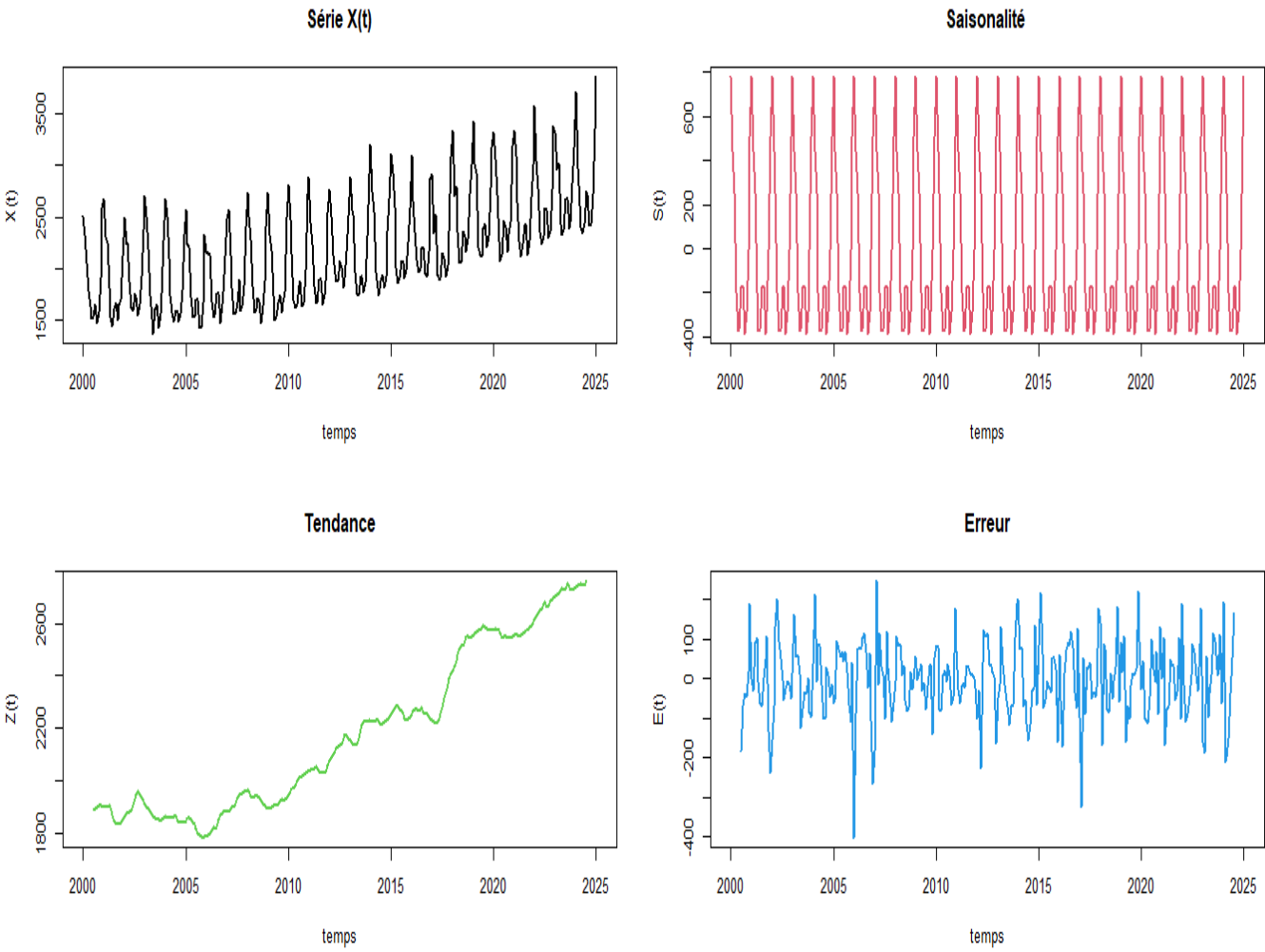


FIGURE 3.2 – Composantes de la série de consommation du gaz naturel aux USA

3.1.2 Désaisonnalisation de la série

L'estimation des coefficients de la tendance et de la saisonnalité par régression linéaire à l'aide de la commande `lm()` nous donne les résultats suivants :

- 1) Coefficient de la tendance : $\hat{\beta}_1 = 3.47$
- 2) Les coefficients saisonniers $\hat{\gamma}_j$:

$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\gamma}_3$	$\hat{\gamma}_4$	$\hat{\gamma}_5$	$\hat{\gamma}_6$	$\hat{\gamma}_7$	$\hat{\gamma}_8$	$\hat{\gamma}_9$	$\hat{\gamma}_{10}$	$\hat{\gamma}_{11}$	$\hat{\gamma}_{12}$
2448.16	2102.36	1910.58	1464.25	1293.28	1292.56	1482.41	1496.11	1274.54	1362.80	1643.54	2162.59

TABLE 3.1 – Estimateurs des coefficients saisonniers de la série de consommation de gaz aux USA.

Nous constatons que $S = \sum_{j=1}^{12} \hat{\gamma}_j = 19933.16 \neq 0$, ce qui nous ramène au calcul des coefficients saisonniers définitifs (2.18) du tableau (Table 3.2). Ce dernier nous aide à créer la série **CVS** dont la représentation graphique est donnée par la Figure 3.3, sous l'utilisation de l'équation (2.20).

3) Coefficients saisonniers définitifs γ_j^* :

γ_1^*	γ_2^*	γ_3^*	γ_4^*	γ_5^*	γ_6^*	γ_7^*	γ_8^*	γ_9^*	γ_{10}^*	γ_{11}^*	γ_{12}^*
787.06	441.27	249.49	-196.85	-367.82	-368.54	-178.69	-164.99	-386.56	-298.30	-17.56	501.49

TABLE 3.2 – Coefficients saisonniers définitifs de la série de consommation de gaz aux USA.

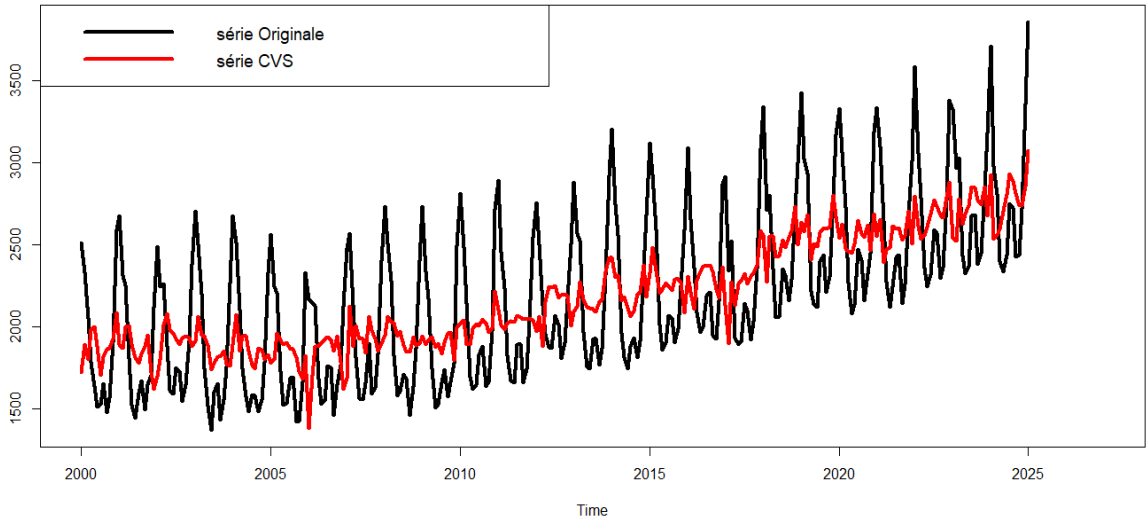


FIGURE 3.3 – Série **CVS**, de la série de consommation de gaz aux USA.

3.1.3 Validation du modèle

Significativité de la régression

Le coefficient de détermination R est un indicateur de la qualité d’ajustement des données de la série. Nous obtenons à l’aide de la **Logiciel R**, les valeurs :

Multiple R-squared : 0.9962, Adjusted R-squared : 0.996.

Alors, le modèle est de grande qualité.

Test de Fisher

Pour évaluer la signification globale du modèle, nous utilisons le test de Fisher exact, qui test vise à vérifier si le modèle utilisé est en mesure d'expliquer la série temporelle de manière significative. En obtenant les résultats suivants de l'analyse de régression.

$$\mathcal{F}_{\alpha}^{(m+p-1, T-(m+p))} = \mathcal{F}_{\alpha=0.05}^{(12, 288)} = 1.785888$$

$$F = 5839$$

$F > \mathcal{F}_{0.05}^{(12, 288)}$, le modèle a donc une signification globale.

Test de Student des paramètres

Dans le but de vérifier que tous les coefficients estimés sont statistiquement différents de zéro, nous effectuons le test de Student :

$$\mathcal{T}_{\alpha}^{(T-(m+p))} = \mathcal{T}_{0.05}^{(288)} = 1.650162$$

Paramètres	$\hat{\beta}_1$	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\gamma}_3$	$\hat{\gamma}_4$	$\hat{\gamma}_5$	$\hat{\gamma}_6$	$\hat{\gamma}_7$	$\hat{\gamma}_8$	$\hat{\gamma}_9$	$\hat{\gamma}_{10}$	$\hat{\gamma}_{11}$	$\hat{\gamma}_{12}$
T_i	37.05	78.77	67.03	60.84	46.56	41.07	41.00	46.95	47.33	40.26	42.99	51.78	68.04

TABLE 3.3 – Test de Student des paramètres (Modèle additif)

Nous pouvons observer que tous les coefficients sont **supérieurs à** $\mathcal{T}_{0.05}^{(288)}$, ce qui **implique** que tous les coefficients sont statistiquement très significatifs. Alors, le modèle estimé est très satisfaisant et peut être utilisé pour réaliser des prévisions futures de la série étudiée.

3.1.4 Prévisions

Après avoir testé la série estimée et sa validité, nous pouvons prédire les valeurs de la série pour l'année 2024 : (1-01-2024, 1-12-2024), puis la comparer aux données anormales pour l'année 2024, où nous obtenons les résultats suivants :

Date	Valeurs de prévision	Valeurs initiales
01-01-2024	3496.115	3710.3
01-02-2024	3153.789	2975.6
01-03-2024	2965.469	2802.9
01-04-2024	2522.609	2396.4
01-05-2024	2355.109	2335.2
01-06-2024	2357.861	2434.0
01-07-2024	2551.185	2748.6
01-08-2024	2568.349	2720.3
01-09-2024	2350.253	2424.4
01-10-2024	2441.985	2438.7
01-11-2024	2726.189	2722.0
01-12-2024	3248.713	3356.3

TABLE 3.4 – Données initiales et prévisions de l’année 2024

Nous observons que les résultats des prévisions et des données sont proches les un des autres.

Les **prévisions pour l’année 2025** de la série de consommation du gaz aux USA sont présentés dans le tableau et la figure suivants :

Date	Valeurs
01-01-2025	3537.756
01-02-2025	3195.430
01-03-2025	3007.110
01-04-2025	2564.250
01-05-2025	2396.750
01-06-2025	2399.502
01-07-2025	2592.826
01-08-2025	2609.990
01-09-2025	2391.894
01-10-2025	2483.626
01-11-2025	2767.830
01-12-2025	3290.354

TABLE 3.5 – Prévision de la série pour l’année 2025

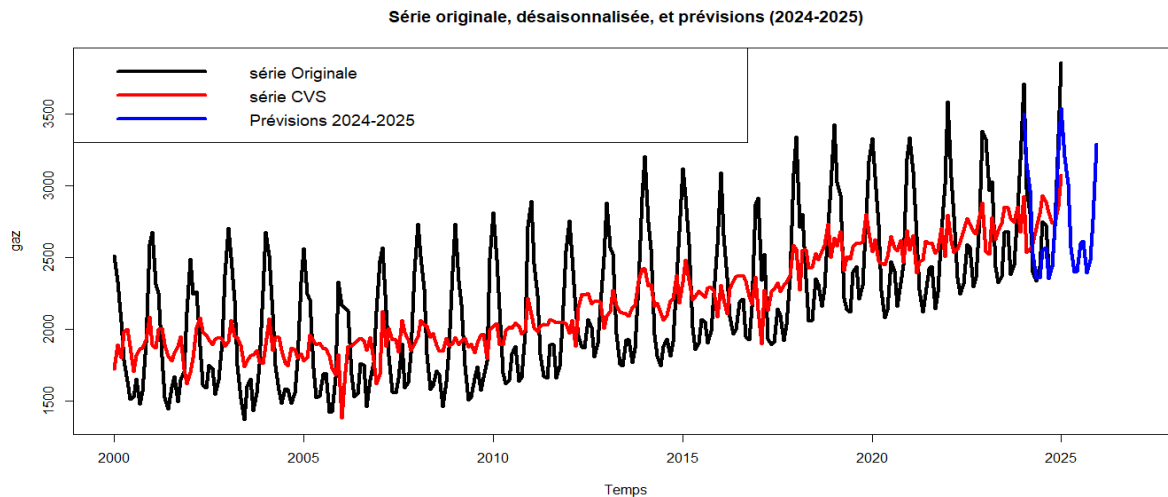


FIGURE 3.4 – Désaisonnalisation, CVS et prévision de la série originale de consommation du gaz aux USA.

3.2 Indice des prix à l'exportation

3.2.1 Description de la série

La série temporelle figurant dans (Figure 3.5), contient des données saisonnières mensuelles de l'indice des prix à l'exportation (IPE) : tarifs aériens vers **l'Europe** non corrigés des variations saisonnières, du 01-01-2002 au 01-12-2011, mises à jour le 15-04-2025, contenant (120 observations), extraite de la source [17].

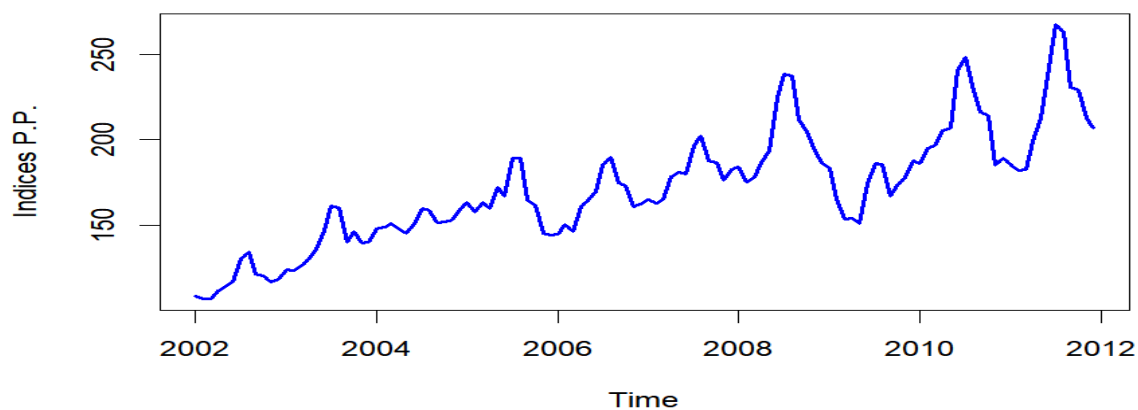


FIGURE 3.5 – Série IPE : tarifs des passagers aériens pour l'Europe

De la Figure 3.5, en déduit que le modèle de la série est **multiplicatif** :

$$X_t = \beta_1 t \times \sum_{j=1}^{p=12} \gamma_j S_t^j \times \varepsilon_t, \quad \forall t = 1, \dots, T = 120.$$

Composantes de la série d'indice des prix à l'exportation

La série de l'indice des prix à l'exportation est composée de trois éléments (voir la Figure 3.6) :

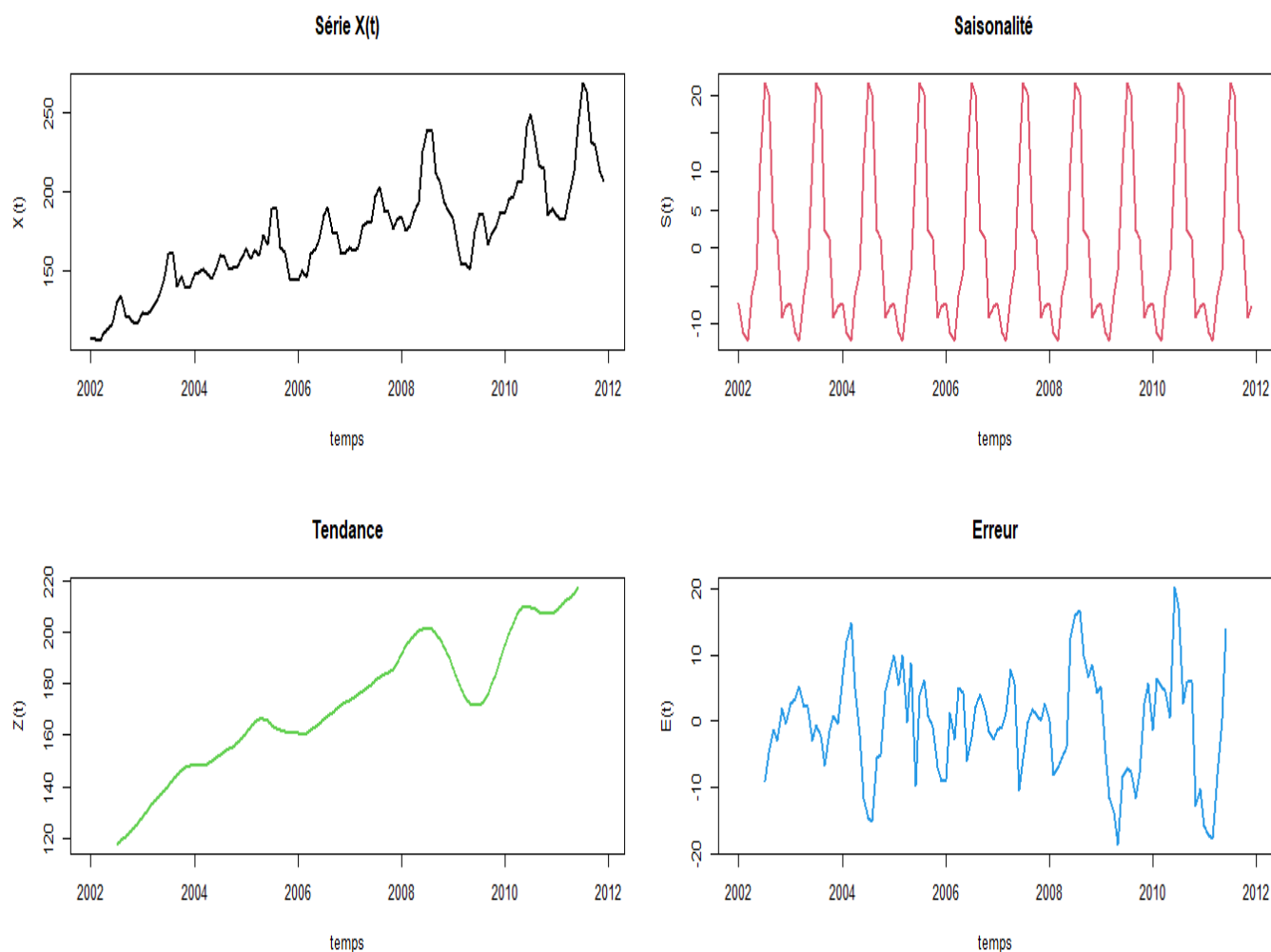


FIGURE 3.6 – Composantes de la série de l'indice des prix à l'exportation

3.2.2 Désaisonnalisation par régression linéaire

L'estimation des coefficients de la tendance et la saisonnalité est donnée par :

- 1) Le coefficient de tendance : $\hat{\beta}_1 = 0.0049$

2) Les coefficients saisonniers $\hat{\gamma}_j$:

$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\gamma}_3$	$\hat{\gamma}_4$	$\hat{\gamma}_5$	$\hat{\gamma}_6$	$\hat{\gamma}_7$	$\hat{\gamma}_8$	$\hat{\gamma}_9$	$\hat{\gamma}_{10}$	$\hat{\gamma}_{11}$	$\hat{\gamma}_{12}$
4.79	4.78	4.76	4.81	4.82	4.88	4.96	4.95	4.85	4.84	4.78	4.79

TABLE 3.6 – Coefficients saisonniers : série IPE.

Remarquons que $\bar{S} = \frac{1}{12} \sum_{j=1}^{12} \hat{\gamma}_j = 4.83 \neq 1$ qui permet de calculer les coefficients saisonniers définitifs (2.19) du tableau (Table 3.7). Ainsi, nous déterminons la série **CVS** (Figure 3.7) en utilisant l'équation (2.21).

3) Les coefficients saisonniers définitifs γ_j^* :

γ_1^*	γ_2^*	γ_3^*	γ_4^*	γ_5^*	γ_6^*	γ_7^*	γ_8^*	γ_9^*	γ_{10}^*	γ_{11}^*	γ_{12}^*
787.06	441.27	249.49	-196.85	-367.82	-368.54	-178.69	-164.99	-386.56	-298.30	-17.56	501.49

TABLE 3.7 – Coefficients saisonniers définitifs de la série IPE.

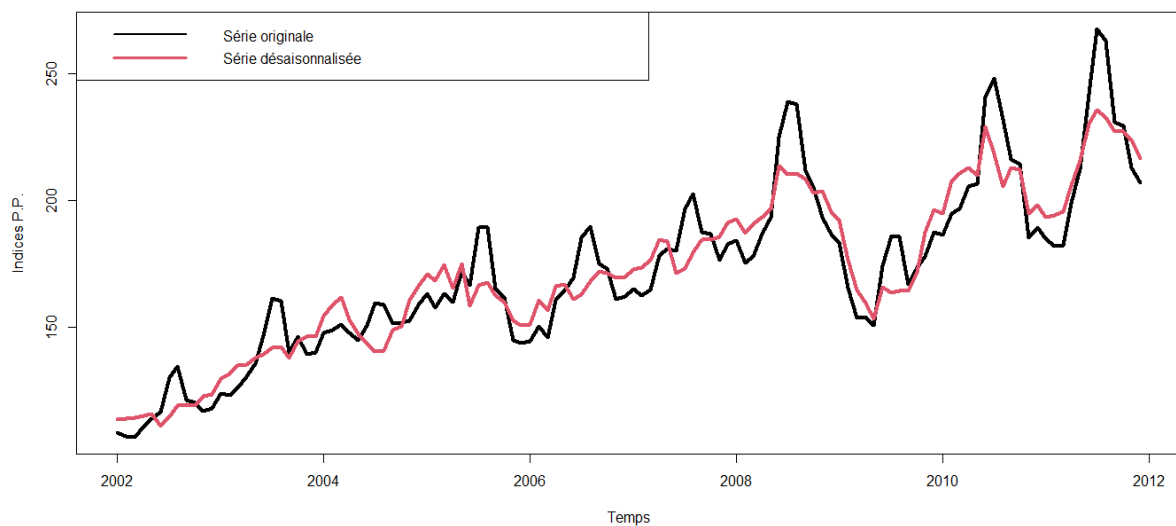


FIGURE 3.7 – IPE : Série originale et désaisonnalisée

3.2.3 Validation

Significativité de la régression

Du logiciel R, nous obtenons **Multiple R-squared** : 0.9998 et **Adjusted R-squared** : 0.9997.

Ces résultats montrent clairement que le modèle est de grande qualité d'ajustement.

Test de Fisher

Le **test de Fisher** est utilisé pour évaluer la validité globale du modèle. Ce test vérifie si tous les coefficients du modèle sont statistiquement significative dans leur globalité. Les Résultats obtenus pour le modèle multiplicatif de la série IPE, sont :

$$\mathcal{F}_{\alpha}^{(m+p-1, T-(m+p))} = \mathcal{F}_{\alpha=0.05}^{(12, 107)} = 1.843745$$

$$F = 3.384 \times 10^4$$

$F > \mathcal{F}_{0.05}^{(12, 107)}$, le modèle a donc une signification globale à 95%.

Test de Student des paramètres

Le **test de Student** est utilisé pour mesurer la signification statistique de chaque coefficient dans un modèle statistique, en particulier pour notre modèle estimé, on a $\mathcal{T}_{\alpha}^{(T-(m+p))} = \mathcal{T}_{0.05}^{(107)} = 1.659219$, et la statistique du test de chaque coefficient est donné par le tableau suivant :

Paramètres	$\hat{\beta}_1$	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\gamma}_3$	$\hat{\gamma}_4$	$\hat{\gamma}_5$	$\hat{\gamma}_6$	$\hat{\gamma}_7$	$\hat{\gamma}_8$	$\hat{\gamma}_9$	$\hat{\gamma}_{10}$	$\hat{\gamma}_{11}$	$\hat{\gamma}_{12}$
T_i	21.74	162.25	161.07	160.40	161.03	161.12	162.81	164.83	164.05	160.02	159.28	156.71	156.33

TABLE 3.8 – Test de Student des paramètres (série IPE : Modèle multiplicatif)

Il est clair que tous les coefficients sont supérieurs à $\mathcal{T}_{0.05}^{107}$, on peut donc dire que **tous les coefficients** sont statistiquement significatifs à 95%.

3.2.4 Prévisions

Finalement, après avoir vérifié la validité du modèle estimé et testé sa significativité, nous avons pu effectuer des prévisions à court terme pour les années **2012 et 2013** et obtenir les résultats sous forme de tableau des valeurs ou par illustration graphique ci-après :

Date	Valeur
01-01-2012	216.4244
01-02-2012	213.2869
01-03-2012	213.3687
01-04-2012	221.9347
01-05-2012	227.2095
01-06-2012	244.0851
01-07-2012	265.0026
01-08-2012	264.5039
01-09-2012	239.3353
01-10-2012	239.1555
01-11-2012	225.9870
01-12-2012	228.2757

TABLE 3.9 – Prévisions IPE **2012**

Date	Valeur
01-01-2013	229.4763
01-02-2013	226.1496
01-03-2013	226.2363
01-04-2013	235.3189
01-05-2013	240.9118
01-06-2013	258.8051
01-07-2013	280.9841
01-08-2013	280.4553
01-09-2013	253.7689
01-10-2013	253.5782
01-11-2013	239.6155
01-12-2013	242.0423

TABLE 3.10 – Prévisions IPE **2013**

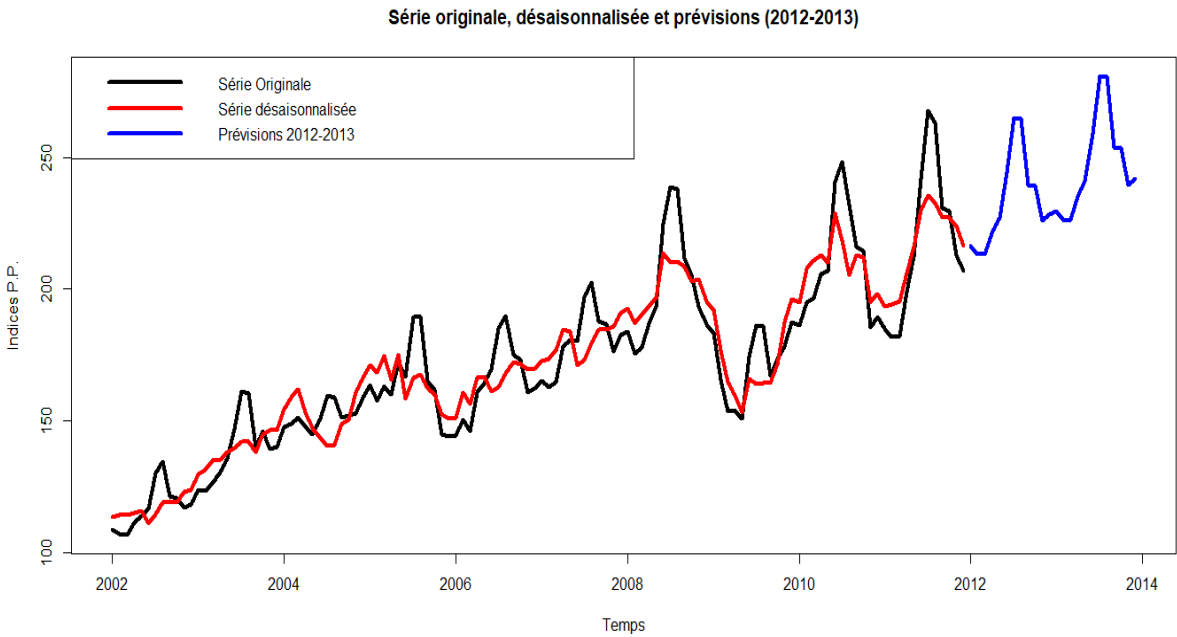


FIGURE 3.8 – Désaisonnalisation, CVS et prévision de la série originale de l’IPE.

Les méthodes de prévision pour les séries temporelles sont nombreuses et dépendent du type de données, de leur structure (saisonnalité, tendance, bruit, etc.), et du contexte d'application. Les modèles choisis ici sont linéaires (additif, multiplicatif), et prend en compte la présence de tendance et de saisonnalité et la longueur de la série temporelle.

En conclusion de ce chapitre, nous constatons que les résultats obtenus sont satisfaisantes pour réaliser des prévisions à courte et moyenne terme des séries temporelles à modèle linéaire, généralisons ainsi les méthodes de lissage exponentiel ou celle de Buyse-Ballot cité dans le deuxième chapitre.

Généralement, les applications aux données réelles soutiennent, enrichissent et valident nos études théoriques. En effet, nous avons formé dans ce dernier chapitre un modèle efficace pour prévoir la demande en énergie (gaz naturel) ou l'évolution d'un marché (indice des prix à l'exportation), ce qui offre un outil puissant pour les décideurs politiques et économiques.

Conclusion

Nous avons abordé dans ce mémoire le thème de la désaisonnalisation à l'aide de la régression linéaire et ces applications aux données réelles des séries temporelles. Le principe de la méthode de désaisonnalisation est basé sur l'estimation des coefficients de la tendance et de la saisonnalité et les corrigés selon le type de modèle, ainsi que de tester leurs signification local et global à l'aide du test de Student et de Fisher respectivement, puis calculé des prévisions à court terme plutôt qu'à long terme du modèle estimé.

Pour l'application aux données réelles, nous nous sommes concentrés sur deux modèles principaux : **le modèle additif** pour analyser les données mensuelles de la consommation de gaz naturel aux Etats-Unis et **le modèle multiplicatif** appliqué aux données saisonnières mensuelles de l'indice des prix à l'exportation : tarifs aériens vers l'Europe.

En conclusion, il faut tenir compte d'un certain nombre de facteurs qui caractérisent cette méthode de désaisonnalisation par régression : le type de données, la nature du modèle, la taille des données et les composantes de la série elle-même. Les résultats de notre étude montrent que cette méthode est efficace à court terme lorsque la structure de la série correspond aux modèles additif ou multiplicatif. Cependant, elle perd de sa précision si la forme de la série varie (non linéaire ou mixte par exemple), ce qui nécessite la recherche d'alternatives plus favorables.

Bibliographie

- [1] Aragon, Y. (2011). *Séries temporelles avec R : Méthodes et cas*. Springer-Verlag France.
- [2] Bourbonnais, R., Terraza, M. (2010). *Analyse des séries temporelles : Applications à l'économie et à la gestion (3rd ed.)*. Dunod.
- [3] Box, G.E.P., Jenkins, G.M., Reinsel, G.C. (1976). *Time Series Analysis, Forecasting and Control*. Third Edition. Holden-Day. Series G.
- [4] Breslow, N.E., Day, N.E. (1980). *Statistical Methods in Cancer Research*. Vol. 1 : The Analysis of Case-Control Studies. IARC Lyon/Oxford University Press.
- [5] Brown, R.G. (1963). *Smoothing, Forecasting and Prediction of Discrete Time Series*. Prentice-Hall.
- [6] Charpentier, A. (2004). *Cours des séries temporelles : Théorie et Applications*. Université de Paris. <https://www.math.u-bordeaux.fr/~hzhang/m2/st/TS1.pdf>.
- [7] Cleveland, W.S. (1993). *Visualizing Data*. New Jersey : Summit Press.
- [8] Cornillon, P.A., Matzner-Løber, É. (2007). *Régression : Théorie et applications*. Springer-Verlag France.
- [9] Durbin, J., Koopman, S.J. (2001). *Time Series Analysis by State Space Methods*. Oxford University Press.
- [10] Florin, A. (2020). *Série temporelles avec R*. https://drive.google.com/file/d/1hakuoucZBUyCuaUcGANDtrLro_zF2ynX/view?usp=drivesdk.

- [11] Goude, Y. (2016). *Introduction aux séries temporelles, tendance et composante saisonnière*. MAP-STA2 : Séries chronologiques. https://www.imo.universite-paris-saclay.fr/~yannig.goude/Materials/time_series/cours2_tendance_composante_saisonniere.pdf.
- [12] Gouriéroux, C., Monfort, A. (1995). *Séries temporelles et modèles dynamiques*. Fenixx.
- [13] Lagnoux, A. (2015). *Renforcement Statistique Séries chronologiques*. Université de Toulouse le Mirail.
- [14] Mazerolle, F. (2005). *Statistique descriptive : Séries statistiques à une et deux variables, Séries chronologiques, Indices*. Gualino.
- [15] R Core Team. (2024). EuStockMarkets dataset. R package "datasets". <https://stat.ethz.ch/R-manual/R-devel/library/datasets/html/EuStockMarkets.html>.
- [16] U.S. Bureau of Economic Analysis. (2018, January 7). Total vehicle sales [TOTALNSA]. Federal Reserve Bank of St. Louis. <https://fred.stlouisfed.org/series/TOTALNSA>.
- [17] U.S. Bureau of Labor Statistics (2025, April 17). Export Price Index (Balance of Payments) : Air Passenger Fares for Europe [IH1421], retrieved from FRED, Federal Reserve Bank of St. Louis. <https://fred.stlouisfed.org/series/IH1421>
- [18] U.S. Bureau of Transportation Statistics. (2018, January 7). Natural gas consumption [NATURALGAS]. Federal Reserve Bank of St. Louis. <https://fred.stlouisfed.org/series/NATURALGAS>.
- [19] Von Sachs, R., Van Bellegem, S. (2005). *Séries chronologiques (4e ed.)*. Université Catholique de Louvain.

Annexe : Abréviations et Notations

Les différentes abréviations et notations utilisées tout au long de ce mémoire sont :

X_t	: Série temporelle.
S_t	: Saisonnalité.
Z_t	: Tendance.
ε_t	: Erreurs.
E	: Espérance.
BB	: Bruit blanc.
σ^2, Var	: Variance.
γ	: Coefficients saisonniers.
L	: Opérateur de retard.
F	: Opérateur d'avance.
Δ	: Opérateur de différentiation.
Δ^d	: Opérateur permet d'éliminer les tendances polynômiale de degré d .
Δ_p	: Opérateur permet d'éliminer les saisonnalités de période p .
$\Delta^d \Delta_p$	: Opérateur permet d'éliminer les polynômiale Z_t et S_t pour un schéma additif .
Δ_p^{d+1}	: Opérateur permet d'élémier les Z_t et S_t pour un schéma mixte .
M	: Moyennes mobiles.
MCO	: Moindres carrés ordinaires.
β_i	: Coefficients de tendance Z_t .
γ_j	: Coefficients saisonniers S_t .
$\hat{\beta}_i$	: Estimateur par la méthode moindres carrés ordinaires de β_i .
$\hat{\gamma}_j$	: Estimateur par la méthode moindres carrés ordinaires de γ_j .

b	: Vecteur de dimension $\mathbf{m}+\mathbf{p}$ contenant les coefficients de saisonnalité et de tendance.
$\hat{b}$	: Estimateur par la méthode MCO de b .
$\hat{b}_i$	: Éléments de vecteur b .
β	: Vecteur de dimension $\mathbf{m}$ contenant les paramètres de la tendance.
γ	: Vecteur de dimension $\mathbf{p}$ contenant les paramètres de la saisonniers.
X	: Vecteur aléatoire de dimension $\mathbf{T}$.
Z	: Matrice engendré par les vecteur $Z^1, \dots, Z^m$ de dimension $(\mathbf{T}, \mathbf{m})$.
S	: Matrice engendré par les vecteur $S^1, \dots, S^p$ de dimension $(\mathbf{T}, \mathbf{p})$.
σ_ε^2	: Variance de ε .
T	: Nombre d'observations.
N	: Nombre d'années entières.
$\bar{x}_n$	: Moyenne des X_t relatives à l'année $\mathbf{n}$.
$\bar{x}_j$	: Moyenne de X_t relatives au saison $\mathbf{j}$.
$\bar{x}, \bar{X}$	: Moyenne de toutes les observations X_t .
CVS	: Corrigée des variations saisonnières.
X_t^{CVS}	: Série corrigée des variations saisonnières.
S	: Somme des coefficient saisonnières.
$\bar{S}$	: Moyenne des coefficients saisonnier.
γ_j^*	: Coefficients définitifs.
R^2	: Coefficient de détermination.
SCT	: Somme des carrés totale.
SCE	: Somme des carrés explique.
SCR	: Somme carrés résiduelle.
ddl	: Degré de liberté.
R_{adj}^2	: Coefficient de détermination ajusté.
H_0	: Hypothèse nulle.
H_1	: Hypothèse alternative.
$\mathcal{F}_\alpha^{(m+p-1, T-(m+p))}$	: Valeur critique de la loi de Fisher.
$\mathcal{T}_\alpha^{(T-(m+p))}$	: Valeur critique de la loi de Student.
α	: Le seuil.

ملخص

نهدف في هذه مذكرة إلى تقديم لمحة عامة حول تعديل المركبة الموسمية للسلاسل الزمنية، وذلك من خلال استخدام طريقة الانحدار الخطي على نموذجي الجمع والجداء، من أجل تقدير معاملات الاتجاه العام والموسمية واختبار مدى أهميتهما. بالإضافة إلى تقييم جودة النموذج والتعرف على خصائص السلسلة المعدلة. وقد بينت نتائج هذه الدراسة أن هذه الطريقة فعالة على المدى القصير عندما تتطابق بنية السلسلة مع أحد النموذجين، غير أنها تفقد دقتها إذا اختلف شكل السلسلة، مما يستدعي البحث عن بدائل أخرى أكثر ملاءمة.

Résumé

Notre objectif dans ce mémoire est de fournir un aperçu général sur la désaisonnalisation des séries temporelles à l'aide de la régression linéaire sur des modèles additifs et multiplicatifs. Nous cherchons à estimer les coefficients de la tendance et de la saisonnalité et à tester leur significativité. Nous visons aussi à évaluer la qualité du modèle et à identifier les caractéristiques des séries corrigées.

Les résultats de cette étude montrent que cette méthode est efficace à court terme lorsque la structure de la série correspond à l'un des deux modèles. Cependant, elle perd de sa précision si la forme de la série varie, ce qui nécessite la recherche d'alternatives plus favorables.

Abstract

This master dissertation aims to provide an introduction to seasonal adjustment of time series by using linear regression on additive and multiplicative models. We seek to estimate the trend and seasonality coefficients and test their significance. We also aim to evaluate the quality of the model, and characterize the properties of the adjusted series.

The results of this study show that this method is effective in the short term when the series structure corresponds to one of the two models. However, it loses its accuracy if the shape of the series varies, which requires the search for more favorable alternatives.