



FACULTÉ DES SCIENCES ET TECHNIQUES
DE MOHAMMEDIA
UNIVERSITÉ HASSAN II DE CASABLANCA

Rapport

Séries chronologiques

Réalisé par :

Ismail Iaich

Wissal Hattab

Demandé par :

Mme.A.BELMAATI

Filière : GMI2



Remerciement

Avant de commencer la présentation de notre projet

Nous remercions notre chère professeure et encadrante Mme.A.BELMAATI pour leur enseignement passionné, leur soutien précieux tout au long du module « Séries chronologiques » ,leurs remarques pertinentes, leur encouragement, leur disponibilité, leur confiance et les conseils précieuses qui nous avez prodiguer tout le long de la réalisation de notre projet.

Notre sincère remerciement s'adresse à l'ensemble du personnel de FST Mohammedia, en particulier à nos professeurs. Enfin, nous remercions tous ceux et celles qui ont contribué de près ou de loin à la réalisation de ce modeste travail.

Table des matières

Introduction	4
Chapitre 1 : Analyse de la série	5
1.1. Base de données	5
1.2. Chronogramme.....	7
1.3. Décomposition.....	9
1.4. Saisonnalité.....	10
1.5. Stationnarité.....	11
1.6. Elimination de la saisonnalité.....	11
1.7. ACF et PACF	12
Chapitre 2 : Modélisation de la série	13
2.1. Méthode AIC.....	13
2.2. Estimation des paramètres.....	14
Chapitre 3 : Validation du modèle choisi	15
3.1. Test de stabilité de variance (Box).....	15
3.2. Test de normalité des résidus (Shapiro)	15
3.3. Stabilité de la variance et l'indépendance des résidus	16
Chapitre 4 : Prévision	17
Conclusion	19
Bibliographie	20

Introduction

Les séries chronologiques sont un élément fondamental de l'analyse des données dans de nombreux domaines, notamment l'économie, la finance, la météorologie, entre autres. Elles représentent un ensemble de données ordonnées dans le temps, où chaque observation est associée à un instant spécifique. Ces observations peuvent être collectées à des intervalles réguliers ou irréguliers, et sont souvent utilisées pour étudier les tendances, les cycles, les saisons et les modèles de comportement au fil du temps.

L'étude des séries chronologiques implique souvent l'analyse, la modélisation et la prévision des données temporelles afin de comprendre les motifs sous-jacents et de prendre des décisions éclairées.

Au cours de ce rendu, on fera l'étude d'une série chronologique saisonnière et stationnaire porte sur des données de ventes dans les magasins de chaussures disponible sur le site web FRED avec le logiciel R.

Chapitre 1

Analyse de la série

1.1 Base de données

a) Source de données

La base de données "**Retail Sales : Shoe Stores**" disponible sur FRED ; Le site web "Federal Reserve Economic Data" est une ressource précieuse pour les économistes, les chercheurs, les professionnels de la finance et le grand public intéressé par les données économiques et financières. Il est géré par la Réserve fédérale de Saint-Louis et propose une vaste collection de données macroéconomiques, y compris des indicateurs clés sur l'emploi, la production, les prix, le commerce et bien plus encore.

Cette base de données fait partie de cette collection et fournit des données sur les ventes au détail dans les magasins de chaussures. Ces données sont collectées par le US Census Bureau et offrent un aperçu des tendances de consommation dans l'industrie de la vente au détail de chaussures aux États-Unis.

Les données sont disponibles sous forme de séries temporelles et peuvent être téléchargées sous différents formats pour une analyse approfondie.

b) Lire les données « SHOES »

```
data <- read.csv("C:/Users/hp/Documents/Rscripts/Projet/SHOES.csv")
```

Notre data est la suivante :

	DATE	MSALES
1	1992-01-01	1206
2	1992-01-02	1265
3	1992-01-03	1463
4	1992-04-01	1675
5	1992-05-01	1560
6	1992-06-01	1498
7	1992-07-01	1449
8	1992-08-01	1838
9	1992-09-01	1578
10	1992-10-01	1529
11	1992-11-01	1516
12	1992-12-01	2039
13	1993-01-01	1173
14	1993-02-01	1161
15	1993-03-01	1439
16	1993-04-01	1670
17	1993-05-01	1615
18	1993-06-01	1564
19	1993-07-01	1544
20	1993-08-01	1895
21	1993-09-01	1659

Showing 1 to 22 of 383 entries, 2 total columns

➔ On a des observations faites mensuellement comme le montre le tableau ci-dessus.

1.2 Chronogramme

a) Création et présentation de notre série temporelle

On commence par convertir la colonne Date en format Date, puis on sélectionne les lignes avec des dates dans la plage spécifiée, et finalement on crée une série temporelle à partir des données sélectionnées comme suit :

```
data$DATE <- as.Date(data$DATE)

subset_data <- subset(data, DATE >= as.Date("1992-01-01") & DATE <= as.Date("2023-11-01"))

ts_data <- ts(subset_data$MSALES, start = c(1992, 1), end = c(2023, 11), frequency = 12)
```

Pour afficher un graphique interactif de nos données, on va utiliser la bibliothèque « plotly » et puis convertir la série temporelle en un data frame.

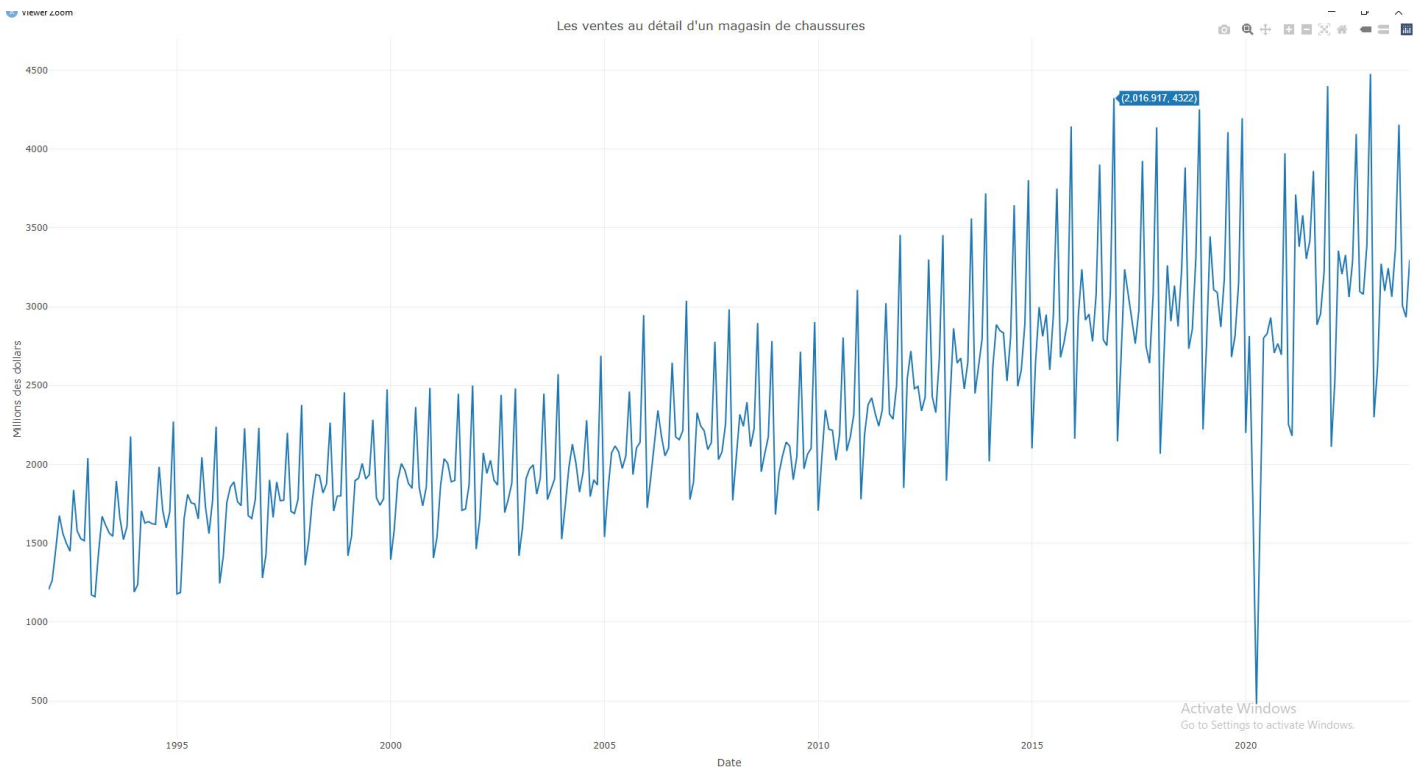
```
library(plotly)

ts_df <- data.frame(Date = time(ts_data), value = ts_data)

# Tracer la série temporelle interactive

p <- plot_ly(data = ts_df, x = ~Date, y = ~value, type = 'scatter', mode = 'lines') %>%
  layout(title = "Les ventes au détail d'un magasin de chaussures", xaxis = list(title = "Date")) %>%
  layout(yaxis = list(title = "Millions des dollars"))
```

Le chronogramme interactif obtenu de notre série :



D'après ce chronogramme, on peut conclure qu'il y a une tendance par ce que la série n'est pas constante par rapport au temps.

On peut observer qu'il y a une saisonnalité, mais on doit décomposer la série pour être sûr.

On remarque aussi qu'on n'a pas une variance constante, donc on va la stabiliser.

b) Stabiliser la variance (appliquant le logarithme) :

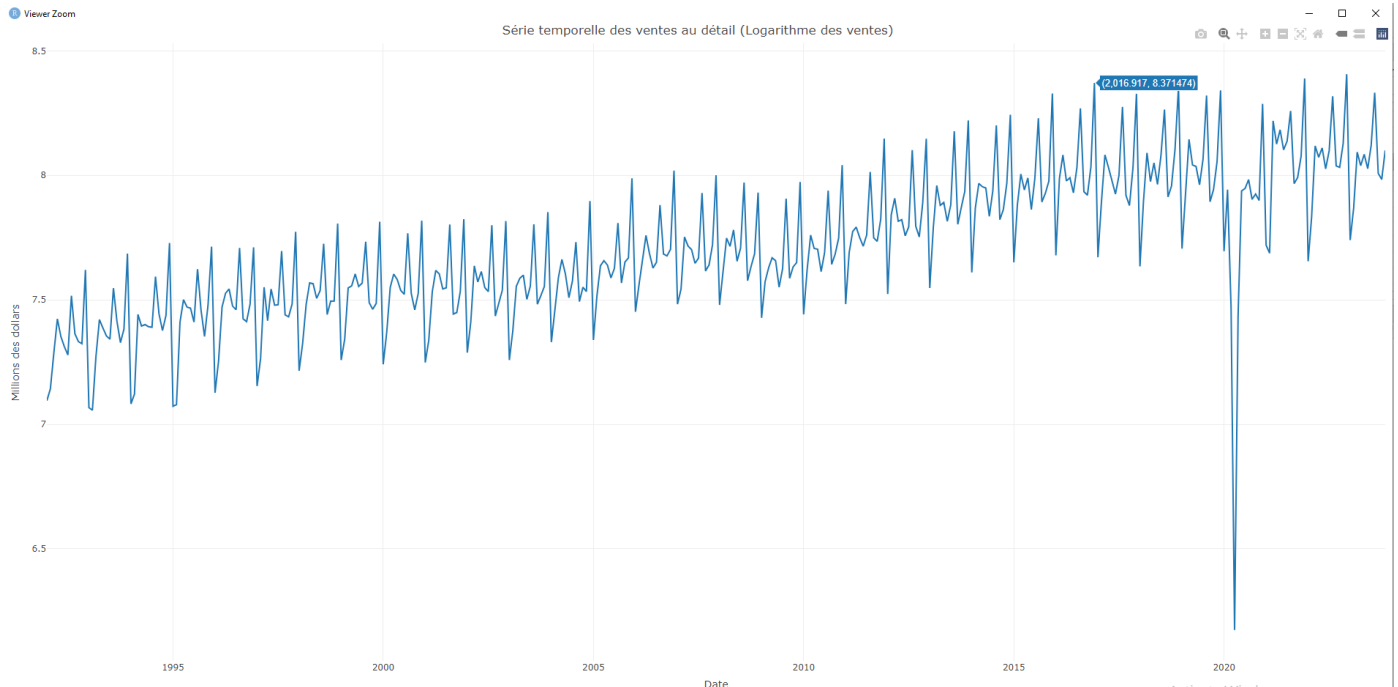
```
lts=log(ts_data)

#Afficher le log des données:

# Convertir la série temporelle en un data frame
ts_df <- data.frame(Date = time(lts), value = lts)

# Tracer la série temporelle interactive
p <- plot_ly(data = ts_df, x = ~Date, y = ~value, type = 'scatter', mode = 'lines') %>%
  layout(title = "série temporelle des ventes au détail (Logarithme des ventes)") %>%
  layout(xaxis = list(title = "Date"), yaxis = list(title = "Millions des dollars"))
```


Le chronogramme interactif obtenu de notre série :



1.3 Décomposition de la série

Appliquer la décomposition à la série temporelle logarithmée pour afficher les composants de notre série (tendance, saisonnalité, résidus).

```
decomposed <- decompose(lts)

# Créer un graphique avec les composantes de la décomposition
par(mfrow=c(4,1)) # Diviser la fenêtre graphique en 4 sous-graphiques

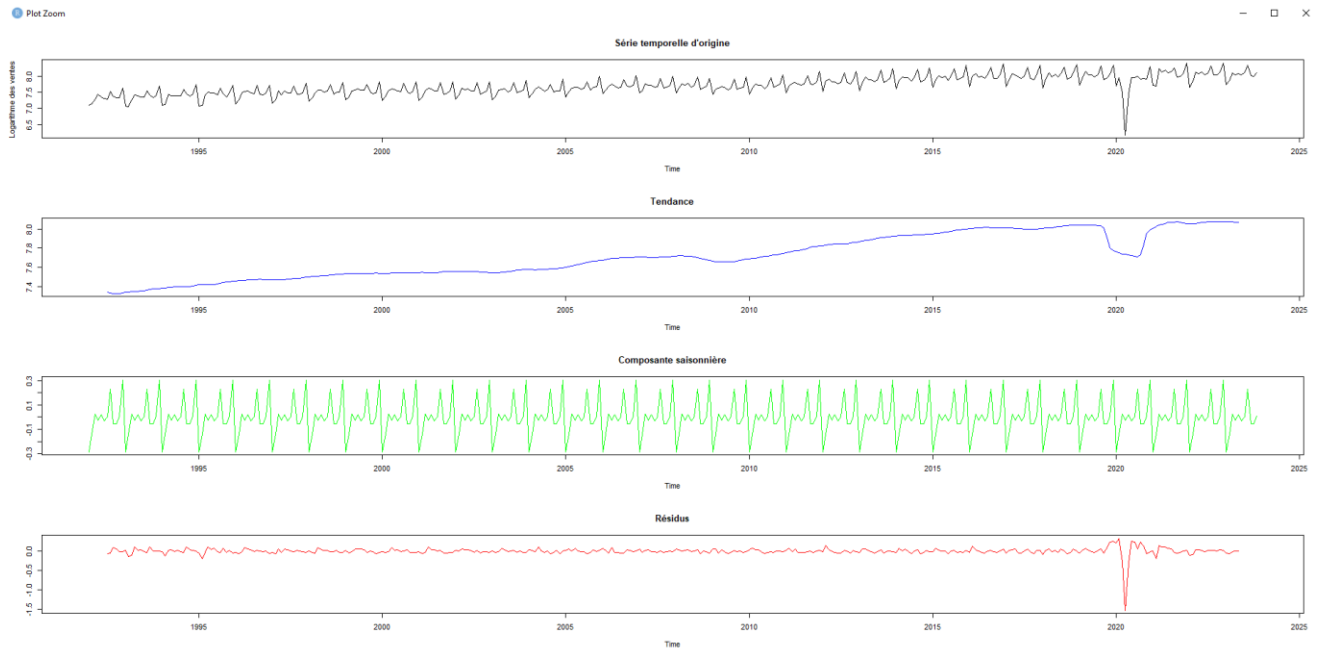
# Tracer la série temporelle d'origine
plot(decomposed$x, type="l", col="black", main="Série temporelle d'origine")%>%
layout(ylab="Logarithme des ventes")

# Tracer la tendance estimée
plot(decomposed$trend, type="l", col="blue", main="Tendance", ylab="")

# Tracer la composante saisonnière
plot(decomposed$seasonal, type="l", col="green", main="Composante saisonnière")%>%
layout(ylab="")

# Tracer les résidus
plot(decomposed$random, type="l", col="red", main="Résidus", ylab="")
```

Ce qui donne :

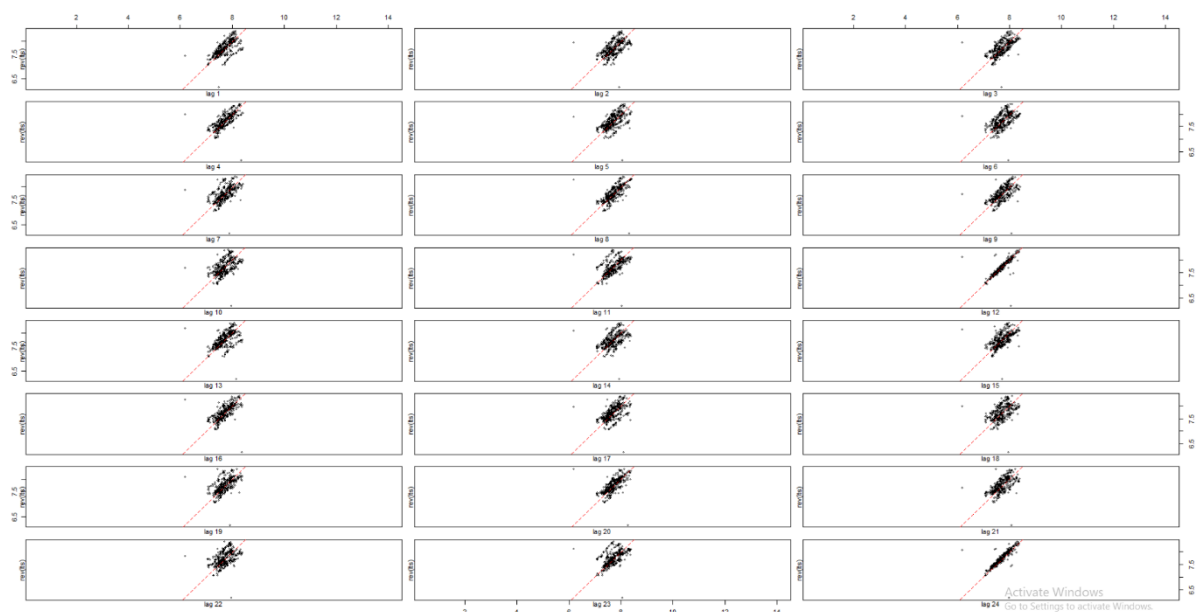


Donc, d'après cette décomposition, on peut conclure que cette série a bien une tendance et une saisonnalité, alors nous devons déterminer la période.

1.4 Saisonnalité de la série

En utilisant :

```
#!e lag plot
lag.plot(rev(lts),24,layout=c(8,3),do.lines=FALSE,diag.col="red",col.main="blue")
```



D'après le lag plot on trouve que notre série est saisonnière d'ordre 12 (période=12).

1.5 Stationnarité de la série

On teste la stationnarité de la série :

```
> adf.test(ts_data)

Augmented Dickey-Fuller Test

data: ts_data
Dickey-Fuller = -5.2397, Lag order = 7, p-value = 0.01
alternative hypothesis: stationary

warning message:
In adf.test(ts_data) : p-value smaller than printed p-value
```

On a $p\text{-value} < 0.05$ alors on rejette l'hypothèse nulle (non stabilité de la série) et donc notre série temporelle est stable et alors **d=0**.

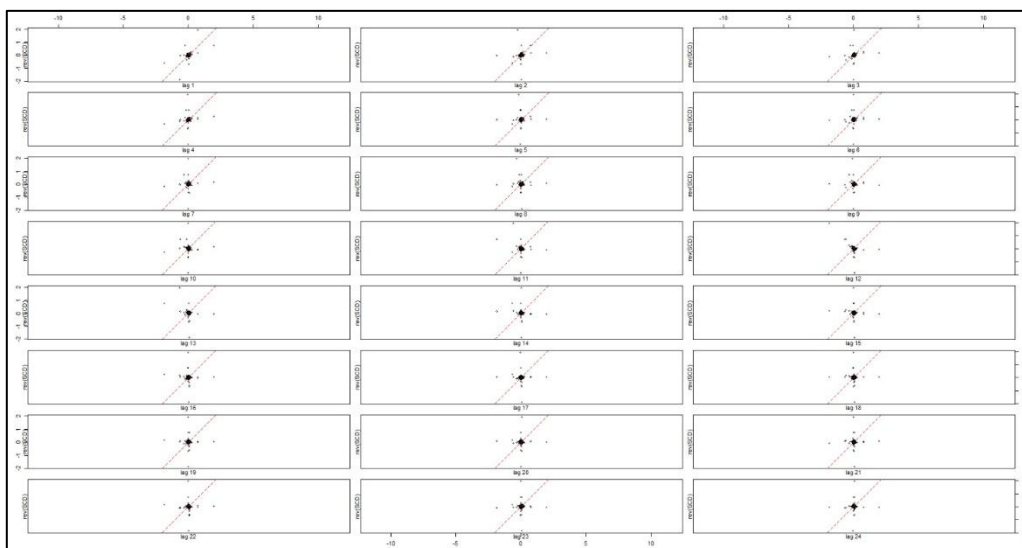
1.6 Elimination de la saisonnalité

a) Méthode de différentiation

Pour éliminer la saisonnalité on applique une différentiation d'ordre 12

```
SCD=diff(1ts,lag=12)
```

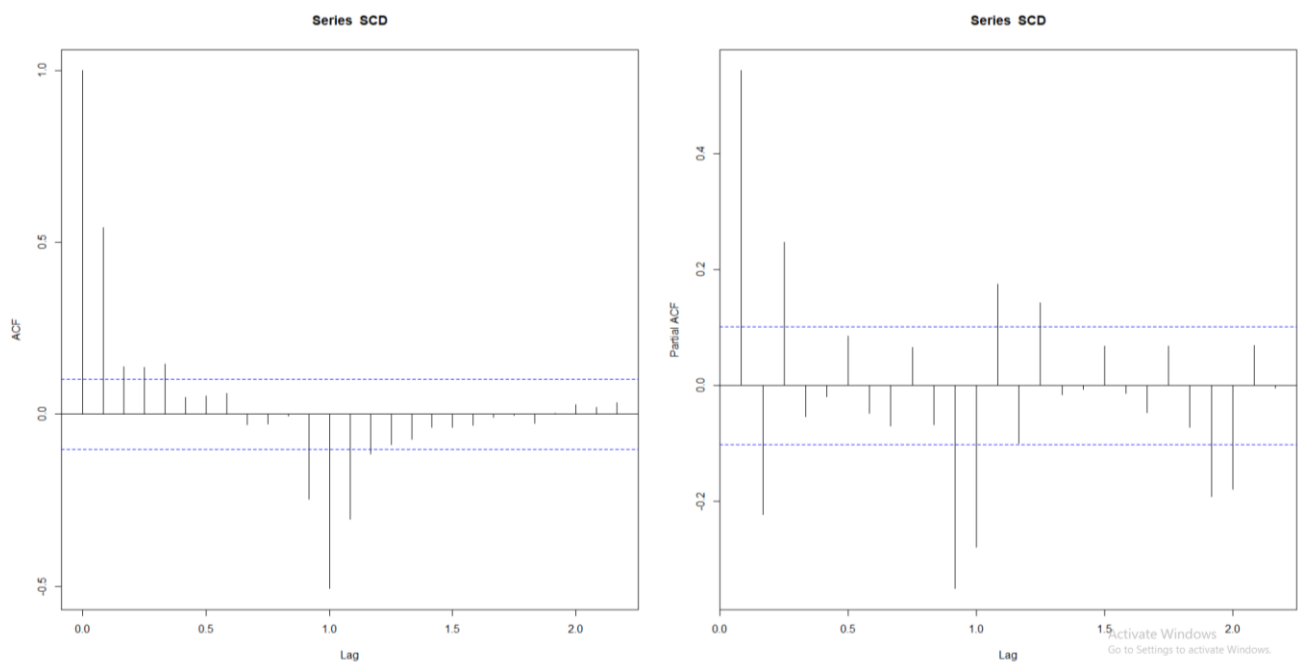
b) Vérifier que la saisonnalité est éliminée



1.7 ACF et PACF

Chercher les paramètres du modèle en utilisant le ACF et le PACF

```
par(mfrow=c(1,2))  
acf(SCD, lag.max = 26)  
pacf(SCD, lag.max = 26)
```



D'après les graphes de ACF et PACF on trouve : **P=3, p=3, Q=1, q=4**.

Chapitre 2

Modélisation de la série

2.1 Méthode AIC :

On va utiliser la boucle suivante qui teste toutes les combinaisons possibles et choisie le modèle avec le plus faible AIC.

```
meilleur_modele <- NULL
meilleur_critere_info <- Inf

for (p in 0:3) {
  for (q in 0:4) {
    for (P in 0:3) {
      for (Q in 0:1) {
        # Exclure les combinaisons non valides
        if (p + q + P + Q <= 11) {
          tryCatch({
            # Ajuster le modèle SARIMA
            modele <- arima(lts, order = c(p, 0, q), seasonal = list(order = c(P, 1, Q), method = "ML", period = 12))

            # Calculer le critère d'information (AIC ici)
            critere_info <- AIC(modele)

            # Mettre à jour le meilleur modèle si nécessaire
            if (critere_info < meilleur_critere_info) {
              meilleur_modele <- modele
              meilleur_critere_info <- critere_info
            }
          }, error = function(e) {
            cat("Erreur lors de l'ajustement du modèle avec les paramètres (p, q, P, Q):", p, q, P, Q, "\n")
          })
        }
      }
    }
  }
}

# Vérifier si le meilleur modèle a été trouvé
if (!is.null(meilleur_modele)) {
  # Afficher l'AIC et les paramètres du meilleur modèle
  cat("AIC du meilleur modèle:", meilleur_critere_info, "\n")
  cat("Paramètres du meilleur modèle SARIMA:\n")
  cat("p =", meilleur_modele$arima[1], "\n")
  cat("d =", 0, "\n") # La différenciation est fixée à 1
  cat("q =", meilleur_modele$arima[2], "\n")
  cat("P =", meilleur_modele$arima[3], "\n")
  cat("D =", 1, "\n") # La différenciation saisonnière est fixée à 1
  cat("Q =", meilleur_modele$arima[4], "\n")
} else {
  cat("Aucun meilleur modèle trouvé.")
}
```

Le meilleur modèle trouvé

```
> # Vérifier si le meilleur modèle a été trouvé
> if (!is.null(meilleur_modele)) {
+   # Afficher l'AIC et les paramètres du meilleur modèle
+   cat("AIC du meilleur modèle:", meilleur_critere_info, "\n")
+   cat("Paramètres du meilleur modèle SARMA:\n")
+   cat("p =", meilleur_modele$arma[1], "\n")
+   cat("d =", 0, "\n") # La différenciation est fixée à 1
+   cat("q =", meilleur_modele$arma[2], "\n")
+   cat("P =", meilleur_modele$arma[3], "\n")
+   cat("D =", 1, "\n") # La différenciation saisonnière est fixée à 1
+   cat("Q =", meilleur_modele$arma[4], "\n")
+ } else {
+   cat("Aucun meilleur modèle trouvé.")
+ }
AIC du meilleur modèle: -612.917
Paramètres du meilleur modèle SARMA:
p = 3
d = 0
q = 4
P = 3
D = 1
Q = 1
```



```
modele <- arima(lts, order = c(3, 0, 4), seasonal = list(order = c(3, 1, 1), period = 12))
```

Le meilleur modèle est SARMA d'ordre (3,0,4) (3,1,1) et période 12.

2.2 Estimation des paramètres :

a) Estimations

```
> #coefficients du modèle:
> modele$coef
      ar1      ar2      ar3      ma1      ma2      ma3      ma4      sar1      sar2      sar3
-0.1818084 0.2283932 0.9065331 0.8716885 0.1088742 -0.9046767 -0.4282735 -0.3217317 -0.2652286 -0.2460792
      sma1
-0.5898650
```

a) Test de significativité des paramètres de modèle

```
> t_stat(modele)
      ar1      ar2      ar3      ma1      ma2      ma3      ma4      sar1      sar2      sar3      sma1
t.stat -6.288052 8.032411 37.39222 15.20668 1.535367 -12.79231 -7.485321 -2.702143 -2.403161 -2.486138 -5.262735
p.val  0.000000 0.000000 0.00000 0.00000 0.124694 0.00000 0.000000 0.006889 0.016254 0.012914 0.000000
```

D'après la sortie précédente, on remarque que les coefficients sont significativement non nuls sauf celui qui est associé à ma2.

Chapitre 3

Validation du modèle choisi

3.1 Test de stabilité de variance (Box) :

```
> Box.test.2(residuals(modele),nlag=10,type = "Box-Pierce")
      Retard  p-value
[1,]      10 0.945013
```

p-value = 0.945013 > 0.05 alors on ne rejette pas l'hypothèse nulle (stabilité de la variance)

Donc nous avons trouvé que le test de box.test.2 est non significatif, donc les résidus forment un bruit blanc.

3.2 Test de normalité des résidus (Shapiro):

```
> #test de normalité des résidus: H0:distribution normale des résidus
> shapiro.test(residuals(modele))

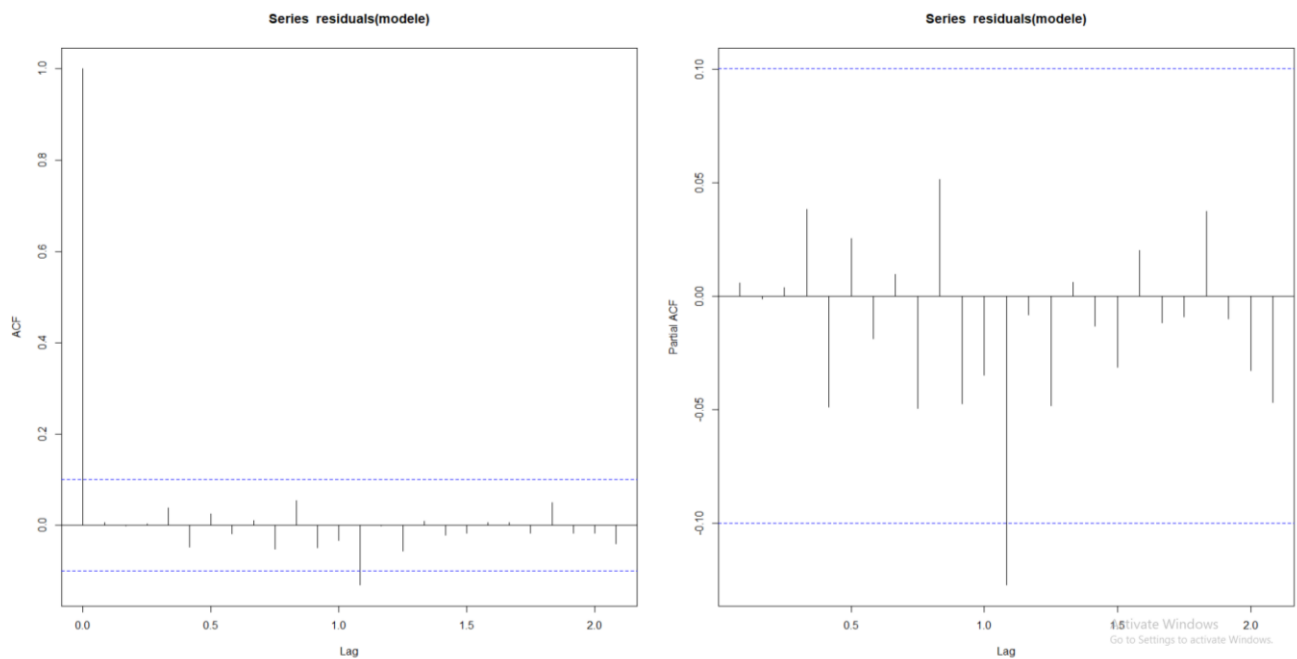
      Shapiro-Wilk normality test

data:  residuals(modele)
W = 0.49142, p-value < 2.2e-16
```

p-value < 0.05 alors on rejette H0 = distribution normale des résidus du modèle.

Test de Shapiro est significatif, donc on n'a pas la normalité, et cela est dû au fait que la variance n'est pas totalement constante.

3.3 Stabilité de la variance et l'indépendance des résidus (bruit blanc):



D'après le ACF et PACF ci-dessus, on remarque que les pics sont tous significativement nuls, on peut conclure donc l'indépendance des résidus. Cela confirme aussi qu'on a un bruit blanc.

Chapitre 4

Prévision

a) Affichage en format numérique :

On va prédire les 24 mois qui suivent en utilisant :

```
par(mfrow=c(1,1))  
P=forecast(modele, level = 95)
```

On présente les valeurs prédites dans le tableau suivant sous R entre 2023 et 2025.

```
> P  
      Point Forecast      Lo 95      Hi 95  
Dec 2023      8.438339 8.241952 8.634726  
Jan 2024      7.774065 7.535797 8.012334  
Feb 2024      7.927451 7.685679 8.169223  
Mar 2024      8.003956 7.760967 8.246945  
Apr 2024      7.558279 7.311420 7.805138  
May 2024      7.913237 7.663936 8.162538  
Jun 2024      8.051644 7.801028 8.302261  
Jul 2024      8.095516 7.841279 8.349753  
Aug 2024      8.242782 7.986823 8.498742  
Sep 2024      8.011527 7.754082 8.268972  
Oct 2024      7.999055 7.738443 8.259668  
Nov 2024      8.068921 7.807033 8.330809  
Dec 2024      8.418246 8.152180 8.684311  
Jan 2025      7.748223 7.477654 8.018792  
Feb 2025      7.894531 7.622616 8.166446  
Mar 2025      8.085976 7.812038 8.359914  
Apr 2025      7.730489 7.454024 8.006954  
May 2025      8.027126 7.749506 8.304747  
Jun 2025      8.074918 7.795218 8.354618  
Jul 2025      8.108838 7.827151 8.390525  
Aug 2025      8.286629 8.003866 8.569392  
Sep 2025      8.006623 7.721822 8.291424  
Oct 2025      8.004849 7.718501 8.291197  
Nov 2025      8.104199 7.816788 8.391610
```

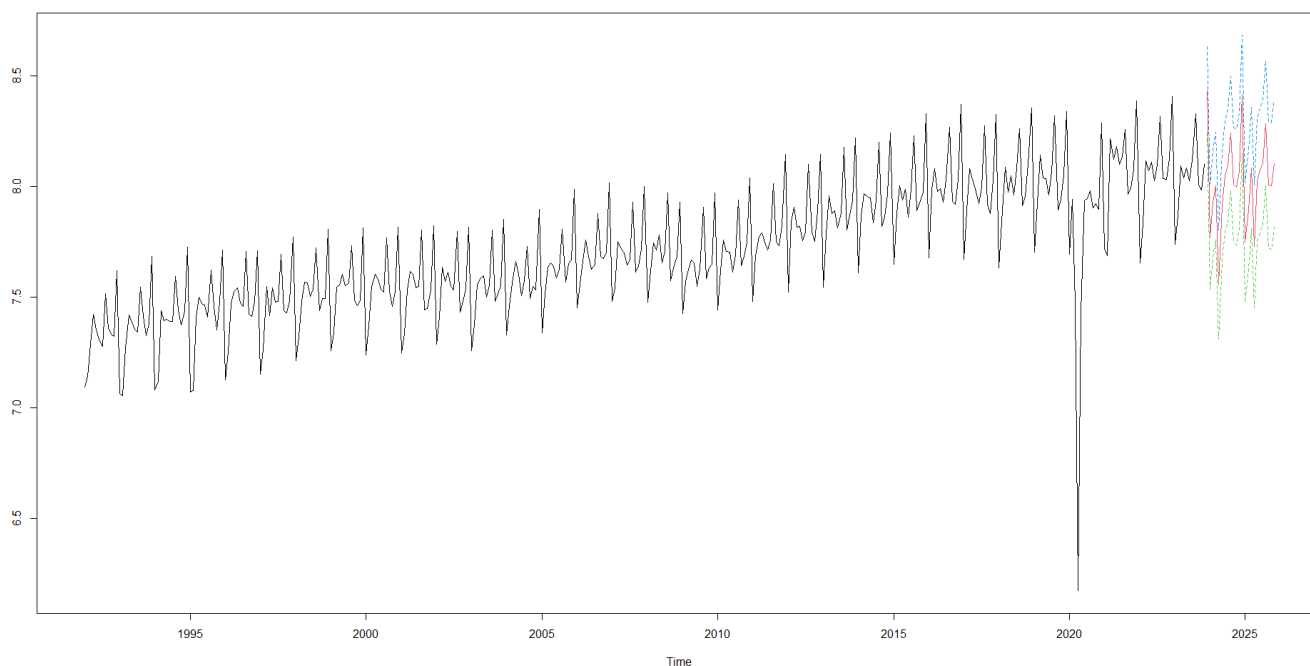
b) Graphe de prévision :

En utilisant :

```
P_l <- ts(P$lower, start = c(2023, 12), frequency = 12)
P_u <- ts(P$upper, start = c(2023, 12), frequency = 12)
P_m <- ts(P$mean, start = c(2023, 12), frequency = 12)

lines(ts_data,col="red")
ts.plot(lts,P_m,P_l,P_u,col=c(1,2,3,4),lty=c(1,1,2,2))
```

Le plot suivant montre la représentation graphique des valeurs observées avec les prévisions :



La partie en bleu représente la borne supérieure de l'intervalle des prédictions et la partie en vert représente sa borne inférieure, et pour la partie rouge représente la moyenne de l'intervalle de prévision.

Conclusion

En conclusion, la modélisation et la prévision des séries chronologiques peuvent représenter un défi de taille en raison de leur complexité intrinsèque. Cependant, l'utilisation du modèle SARMA offre une approche robuste et efficace pour analyser et anticiper les tendances temporelles avec précision.

Dans le cadre de ce projet, l'application du modèle SARMA a démontré sa capacité à capturer les structures autorégressives, intégrées et de moyenne mobile saisonnières présentes dans les données, fournissant ainsi des prévisions fiables et précises. Cette approche a permis d'obtenir des résultats prometteurs avec un niveau de confiance élevé dans les prévisions générées.

Bibliographie

Retail Sales: Shoe Stores (MRTSSM4482USN) | FRED | St. Louis Fed
(stlouisfed.org)