

ANALYSE ET MODÉLISATION DES SÉRIES TEMPORELLES

INTRODUCTION

L'analyse des séries temporelles est essentielle dans un cadre économique car elle permet d'étudier l'évolution des données au fil du temps. Cela aide à faire des prévisions fiables et à prendre des décisions stratégiques éclairées,

reallygreatsite.com



PRÉSENTATION DES DONNÉES

Source des données :

Les données proviennent du site Federal Reserve Economic Data (FRED).

Fichiers analysés :

- Industrial Production electronic S C1.xls
- Job Openings construction S C3.xls

Caractéristiques des séries temporelles :

- Fréquence : Mensuelle
- Période : De janvier 2004 à mars 2024

MÉTHODOLOGIE DE L'ANALYSE

-

Phase 1 -

Analyse Statistique et Graphique : Utilisation des outils statistiques pour identifier les tendances, la saisonnalité et les cycles dans la série temporelle.

Phase 2 -

Ajustement de la Série : Application de techniques de différentiation et modélisation fonctionnelle pour ajuster la série.

Phase 3 -

Modélisation ARIMA : Analyse des résidus de l'ajustement et modélisation avec des modèles ARIMA pour obtenir des prévisions.

PHASE 1

SÉRIE TEMPORELLE 1

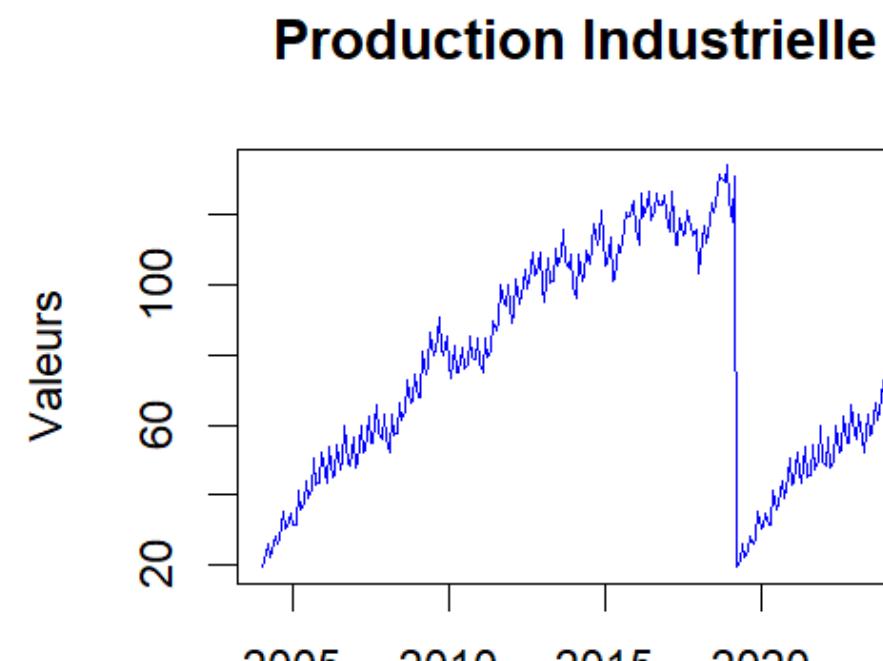
- **Nom du fichier** : Industrial Production electronic S C1.xls
- **Description** : Cette série mesure la production industrielle mensuelle (sous forme d'indice) dans les secteurs des semi-conducteurs et des cartes de circuits imprimés, un élément clé dans l'industrie technologique.

SÉRIE TEMPORELLE 1

```
production_industrielle <- read_excel(file.choose(), skip = 10)
production_industrielle$observation_date <- as.Date(production_industrielle$observation_date)

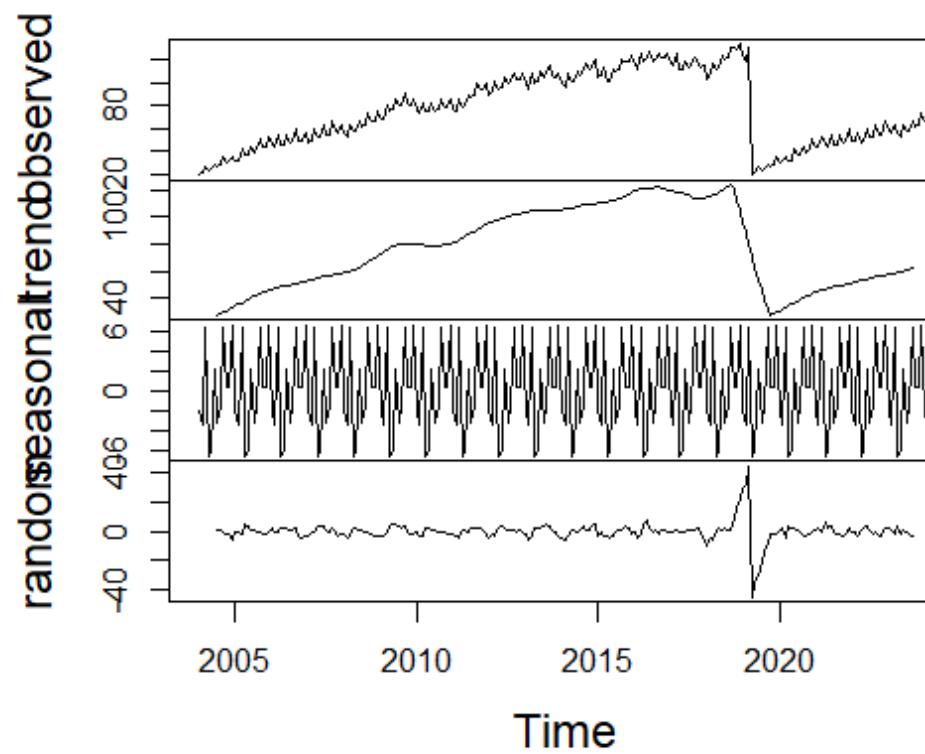
> str(production_industrielle)
tibble [183 × 2] (S3: tbl_df/tbl/data.frame)
$ observation_date: Date[1:183], format: "2009-01-01" "2009-02-01" ...
$ IPB53122N      : num [1:183] 19.5 21.1 26.4 22.4 23.7 ...

# Créer une série temporelle avec fréquence mensuelle
production_industrielle_ts <- ts(production_industrielle$IPB53122N,
                                 start = c(2004, 1), # Début en janvier 2004
                                 end = c(2024, 3),   # Fin en mars 2024
                                 frequency = 12)    # Fréquence mensuelle
```



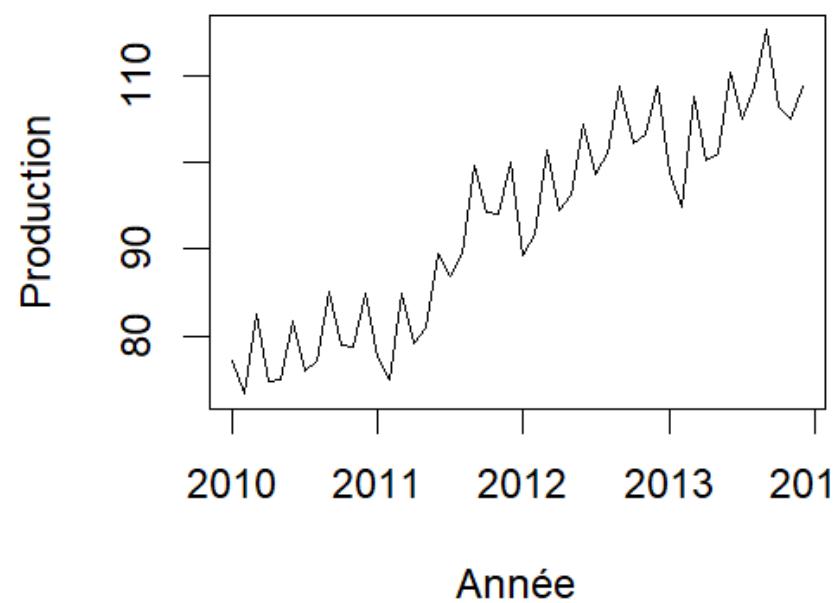
SÉRIE TEMPORELLE 1

Decomposition of additive time series



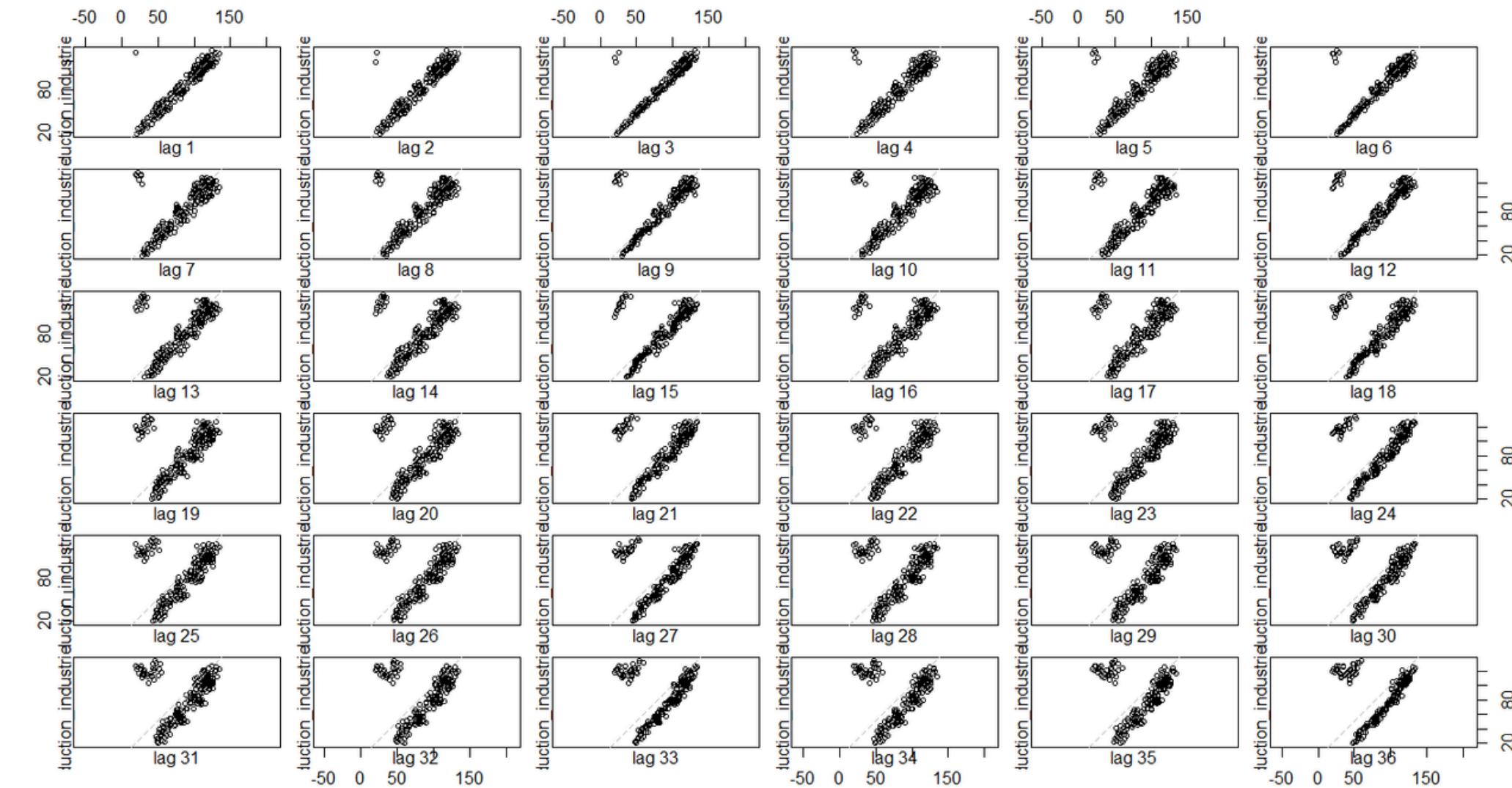
Une tendance à la hausse de la production industrielle, avec un pic en 2018, où la production atteint 133.94.

Les fluctuations saisonnières sont présentes chaque année, suggérant que certains mois, comme décembre, sont plus productifs que d'autres.



Chute brusque à partir de 2019 peut être attribuée à l'impact profond de la pandémie de COVID-19,

SÉRIE TEMPORELLE 1



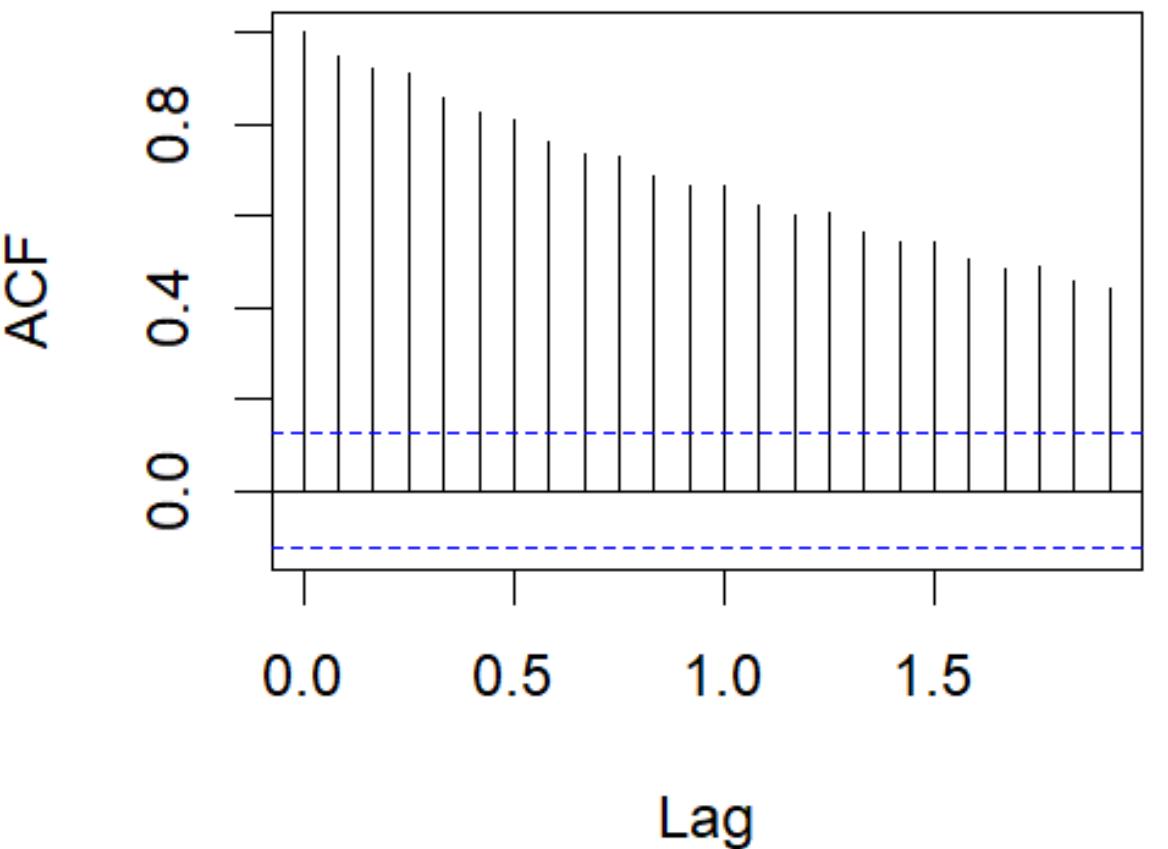
Le graphique de lag montre que les valeurs passées influencent fortement les valeurs futures de la production industrielle, ce qui indique **une tendance persistante et une autocorrélation** dans les données.

SÉRIE TEMPORELLE 1

```
# Calcul de l'autocorrélation (ACF) pour observer la saisonnalité  
acf(production_industrielle_ts)  
acf(production_industrielle_ts , plot=F )
```

```
Autocorrelations of series 'production_industrielle_ts', by lag  
  
0.0000 0.0833 0.1667 0.2500 0.3333 0.4167 0.5000 0.5833 0.6667 0.7500 0.8333 0.9167 1.0000 1.0833 1.1667  
1.000 0.947 0.919 0.909 0.855 0.822 0.810 0.761 0.734 0.731 0.688 0.666 0.666 0.623 0.602  
1.2500 1.3333 1.4167 1.5000 1.5833 1.6667 1.7500 1.8333 1.9167  
0.605 0.565 0.544 0.543 0.503 0.485 0.490 0.455 0.440  
> |
```

Series production_industrielle_ts



L'ACF montre que les valeurs passées de la production industrielle ont une forte influence sur les valeurs futures à court terme, mais cette influence diminue avec le temps. Cela signifie que les effets des valeurs passées sont plus importants à court terme et s'estompent à mesure que le temps passe.

SÉRIE TEMPORELLE 1

```
> adf.test(production_industrielle_ts) # Test de Dickey-
    Augmented Dickey-Fuller Test

data: production_industrielle_ts
Dickey-Fuller = -2.1925, Lag order = 6, p-value = 0.4948
alternative hypothesis: stationary
```

p-value est **0.4948**, qui est supérieure au seuil de significativité habituel de 0.05, On accepte l'hypothèse nulle. Cela signifie que la série temporelle n'est pas stationnaire.

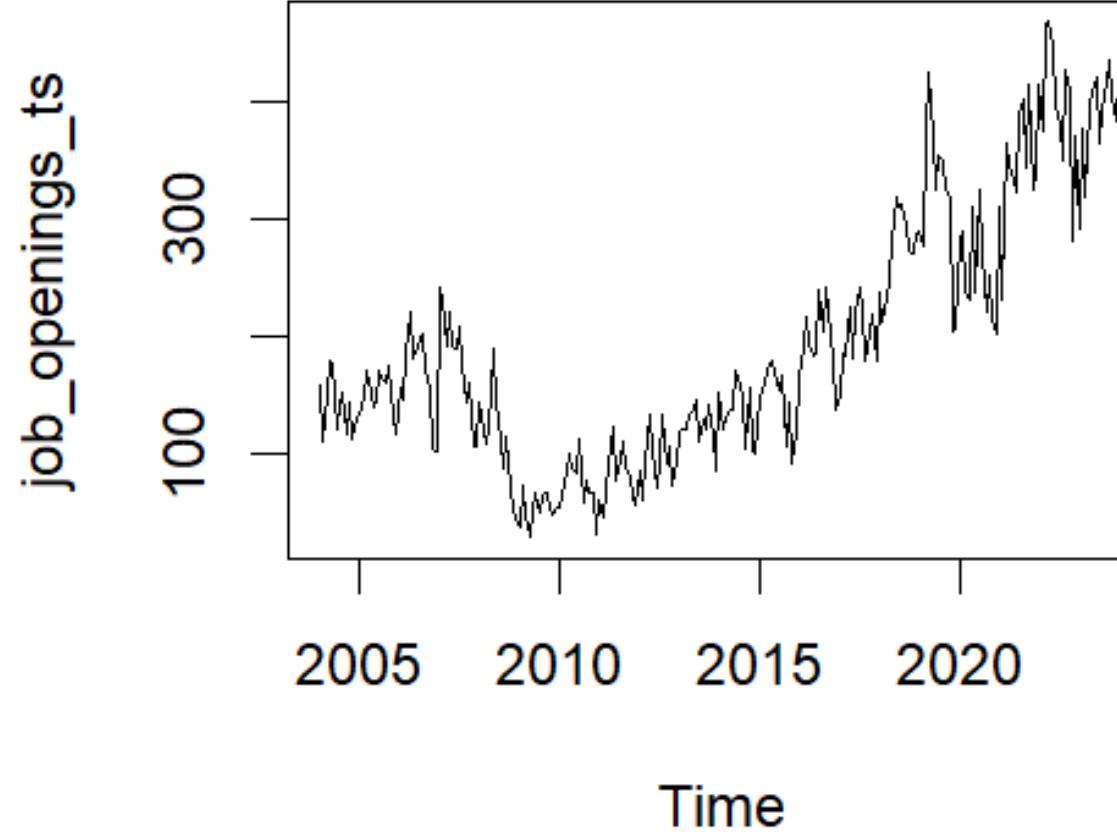
SÉRIE TEMPORELLE 2

- **Nom du fichier** : Job Openings construction S C3.xls
- **Description** : Cette série de données mensuelles suit le nombre d'offres d'emploi disponibles dans le secteur de la construction aux États-Unis, sans ajustement saisonnier. Les données vont de janvier 2004 à mai 2005.

SÉRIE TEMPORELLE 2

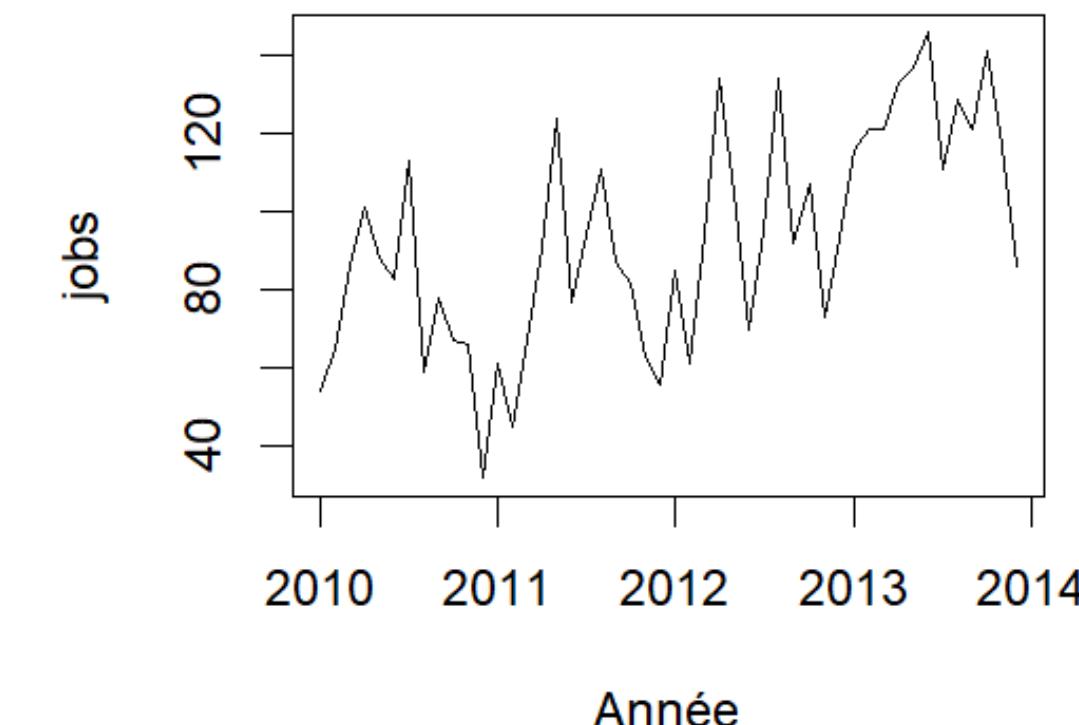
```
data_job_openings <- read_excel(file.choose(), skip = 10)

# Créer une série temporelle avec fréquence mensuelle
job_openings_ts <- ts(data_job_openings$JTU2300JOL,
                        start = c(2004, 1), # Début en janvier 2004
                        end = c(2024, 3),   # Fin en mars 2024
                        frequency = 12)    # Fréquence mensuelle
```



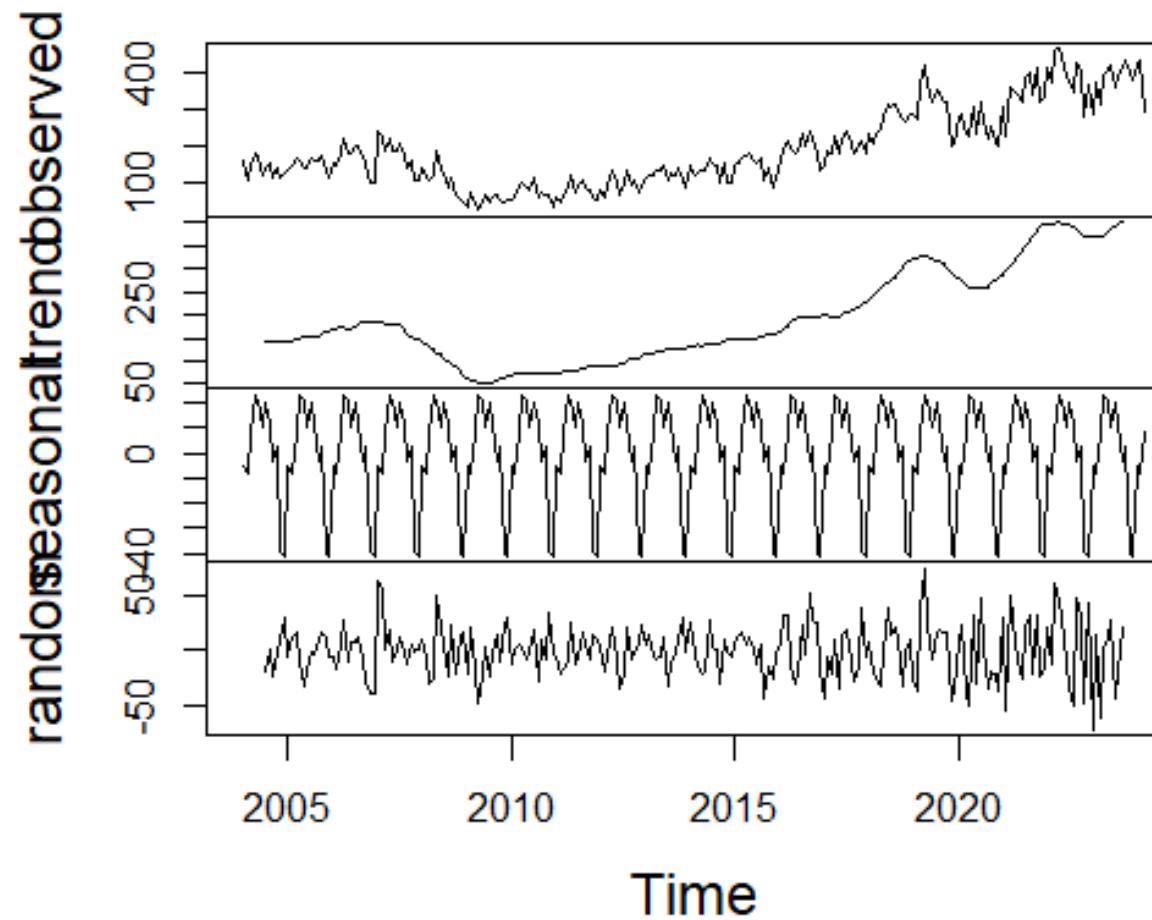
```
# Sélectionner une sous-série de 2010 à 2012
zoom_production_job <- window(job_openings_ts, start=c(2010, 1), end=c(2013, 12))
```

```
# Tracer la sous-série
plot(zoom_production_job, ylab="jobs", xlab="Année")
```



SÉRIE TEMPORELLE 2

Decomposition of additive time series



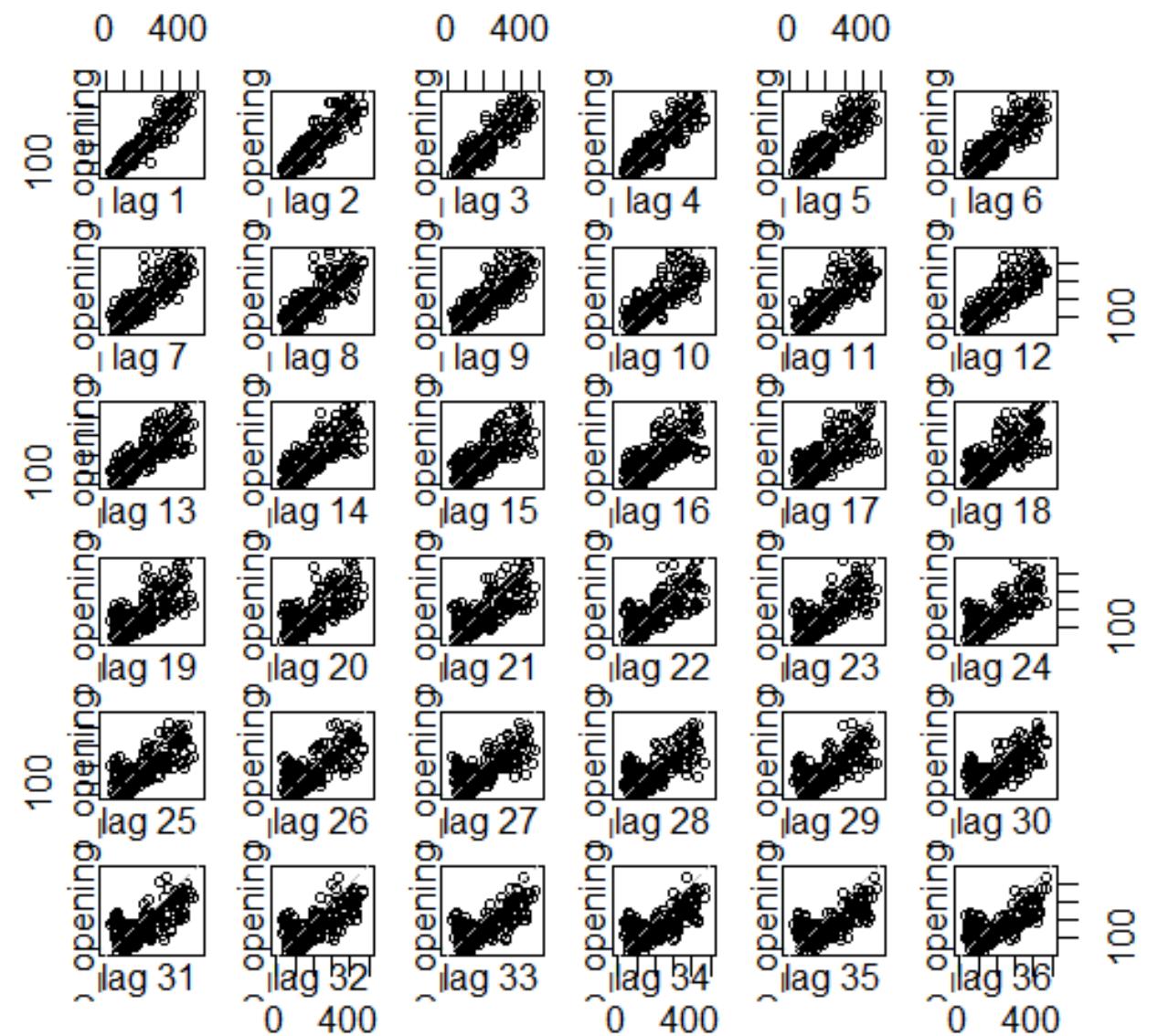
On remarque une tendance générale à la hausse avec des fluctuations saisonnières.

Chute significative entre 2008 et 2010, puis reprise lente mais constante jusqu'en 2023.

Variabilité en fin de série (2023-2024) : on observe un possible changement de structure

Effet saisonnier stable : certains mois comme avril et mai montrent des pics récurrents.

SÉRIE TEMPORELLE 2



Un lag plot est utilisé pour détecter la dépendance temporelle dans une série.

Les points suivent une droite inclinée positive, ce qui suggère une relation linéaire

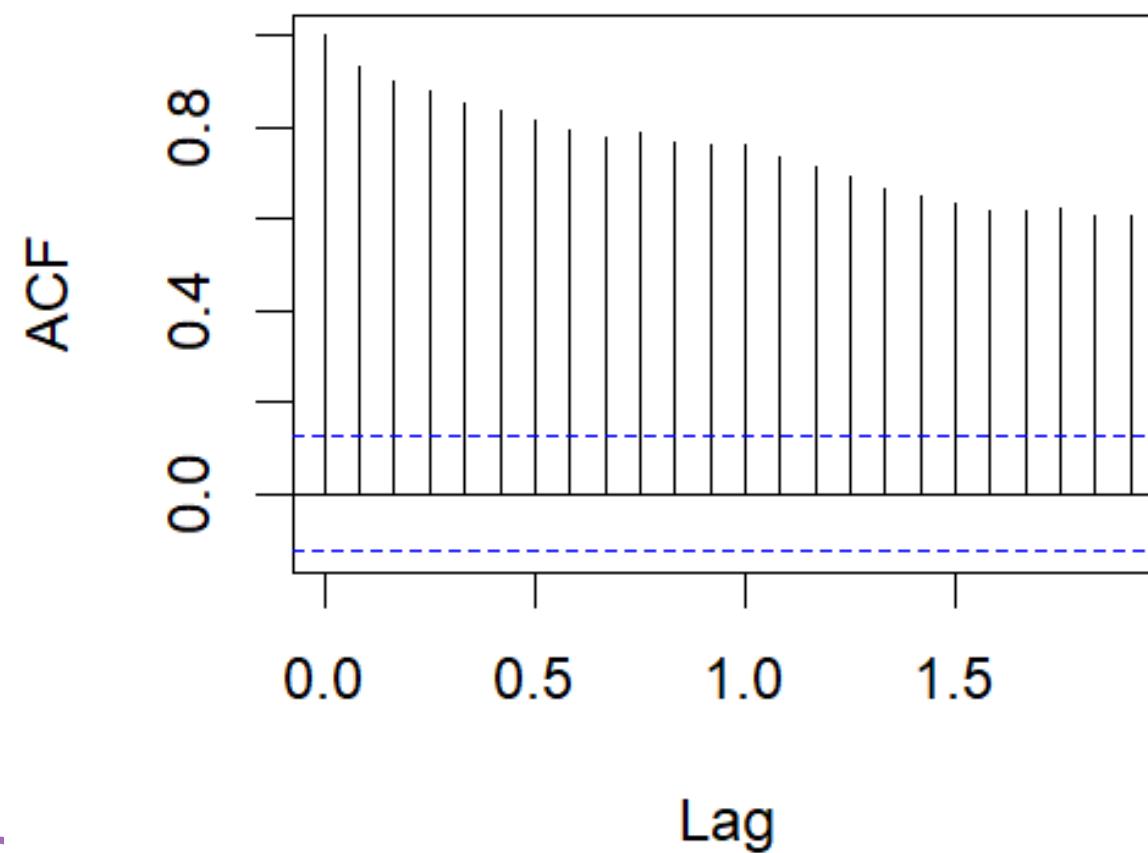
Cette série présente une autocorrélation très forte

==> série hautement prévisible, avec une structure temporelle claire

SÉRIE TEMPORELLE 2

```
# Calcul de l'autocorrélation (ACF) pour observer la saisonnalité  
acf(job_openings_ts)  
acf(job_openings_ts , plot=F )|
```

Series job_openings_ts



La serie est hautement autocorrélée jusqu'à au moins 24 mois.
la dépendance diminue doucement avec le temps,
mais reste notable

SÉRIE TEMPORELLE 2

```
> adf.test(job_openings_ts) # Test de Dickey-
```

```
Augmented Dickey-Fuller Test
```

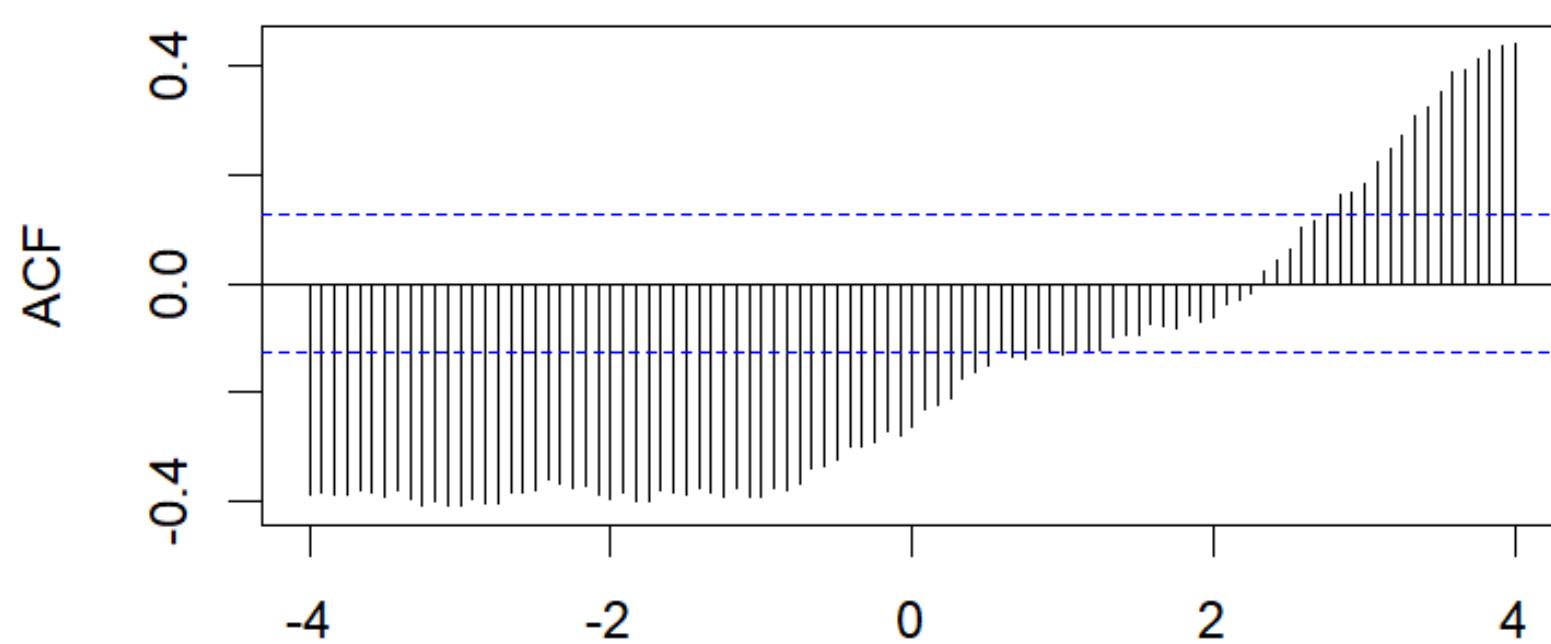
```
data: job_openings_ts
Dickey-Fuller = -2.3316, Lag order = 6, p-value = 0.4363
alternative hypothesis: stationary
```

p-value est 0.4363, qui est supérieure au seuil de significativité habituel de 0.05, On accepte l'hypothèse nulle. Cela signifie que la série temporelle n'est pas stationnaire.

AUTO-CORRELATION CROISE

```
# Calcul de l'autocorrélation croisée  
ccf(job_openings_ts, production_industrielle_ts, lag.max = 48, main = "Autocorrélation Croisée entre Serie1 et Serie2")  
ccf(job_openings_ts, production_industrielle_ts, lag.max = 48, plot=F)
```

Autocorrélation Croisée entre Serie1 et Serie2



les pics sont en dehors des zones de confiance, cela indique que ces corrélations sont significatives.

il semble que les ouvertures d'emplois influencent la production industrielle avec un certain délai, mais la relation directe entre elles reste relativement faible à court terme.

PHASE 2

OBJECTIF DE L'AJUSTEMENT

Pourquoi ajuster la série ?

Les deux séries de production industrielle et Job Openings n'sont pas stationnaires.
Il est donc nécessaire de stabiliser les tendances et capturer la saisonnalité pour faire de bonnes prévisions.

Donc on a utilisé **la différenciation** qui est une technique utilisée pour rendre une série stationnaire

```
# Appliquer une différenciation d'ordre 1  
production_industrielle_diff <- diff(production_industrielle_ts)|
```

```
# Appliquer une différenciation d'ordre 1  
job_openings_ts_diff <- diff(job_openings_ts)
```

APPROCHE 1 – MODÈLE ARIMA

ARIMA – Une modélisation paramétrique

Ce modèle tient compte :

- De l'effet mémoire (MA),
- De la saisonnalité annuelle ,
- De la tendance via la différenciation.

+ Avantages :

- Bonne capacité de modélisation des comportements complexes.
- Fournit des intervalles de confiance pour les prévisions.



APPROCHE 2 – HOLT-WINTERS

Holt-Winters – Une méthode plus intuitive

- Adapté aux séries avec tendance + saisonnalité régulière.
- Pas besoin de rendre la série stationnaire.



Avantages :

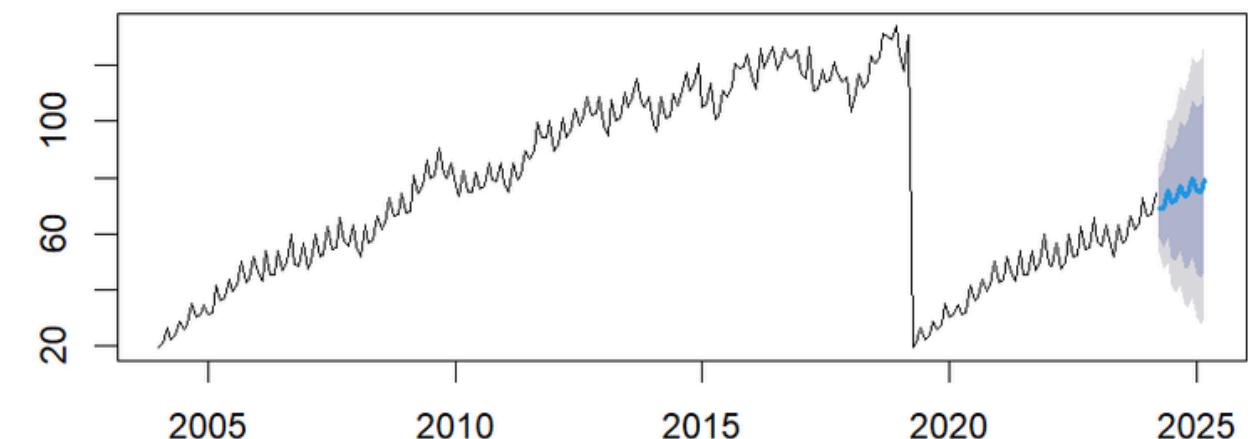
- Simple, rapide et efficace.
- Très bien adapté aux séries mensuelles régulières.

SÉRIE TEMPORELLE 1

```
# Ajustement automatique ARIMA  
model_prod <- auto.arima(production_industrielle_ts, seasonal = TRUE)  
summary(model_prod)
```

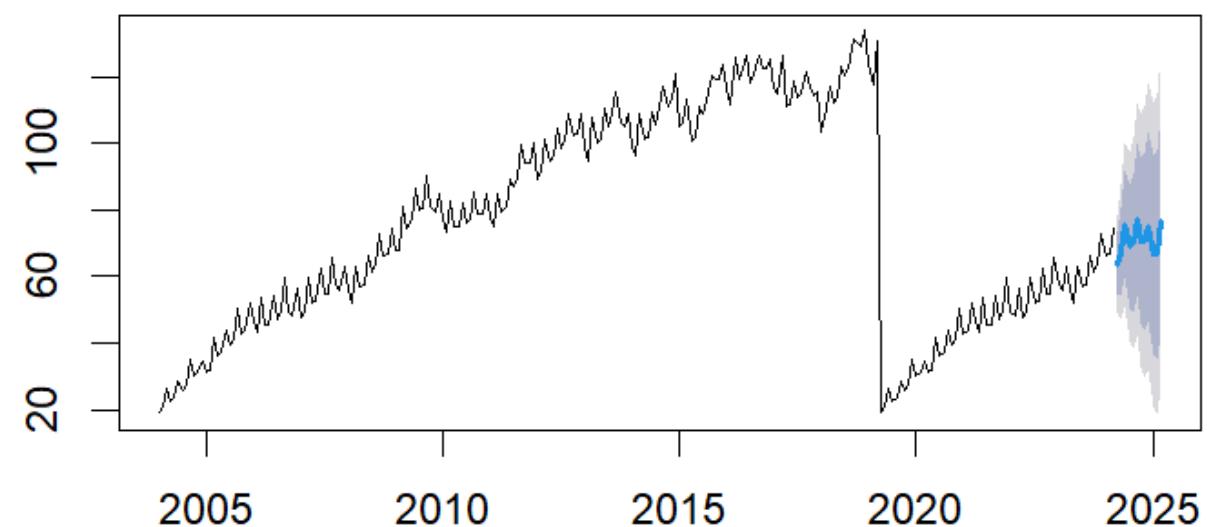
```
# Prévision sur 12 mois  
forecast_prod <- forecast(model_prod, h = 12)  
plot(forecast_prod)  
forecast_prod
```

Forecasts from ARIMA(0,1,1)(2,0,0)[12]



```
# Holt-Winters (saisonalité additive)  
hw_model <- Holtwinters(production_industrielle_ts)  
plot(hw_model)  
hw_model  
forecast_hw <- forecast(hw_model, h = 12)  
plot(forecast_hw)
```

Forecasts from HoltWinters



SÉRIE TEMPORELLE 1

Comparaison des résidus

```
accuracy(model_prod)  
accuracy(forecast_hw)
```

```
> accuracy(model_prod)  
ME      RMSE     MAE      MPE      MAPE     MASE      ACF1  
Training set 0.1242459 8.011999 3.4334085 -1.274731 7.335634 0.2647945 -0.006136804  
> accuracy(forecast_hw)  
ME      RMSE     MAE      MPE      MAPE     MASE      ACF1  
Training set -0.5263412 7.41047 2.605416 -2.259193 5.762796 0.2008977 -0.007483075  
>
```

- En se basant sur les métriques de précision, le modèle Holt-Winters semble légèrement mieux performer en termes de métriques d'erreur (RMSE, MAE et MAPE plus faibles) par rapport au modèle ARIMA. Cependant, les deux modèles sont assez proches, et le meilleur choix pourrait dépendre d'autres facteurs tels que l'interprétabilité ou les préférences en matière de complexité.

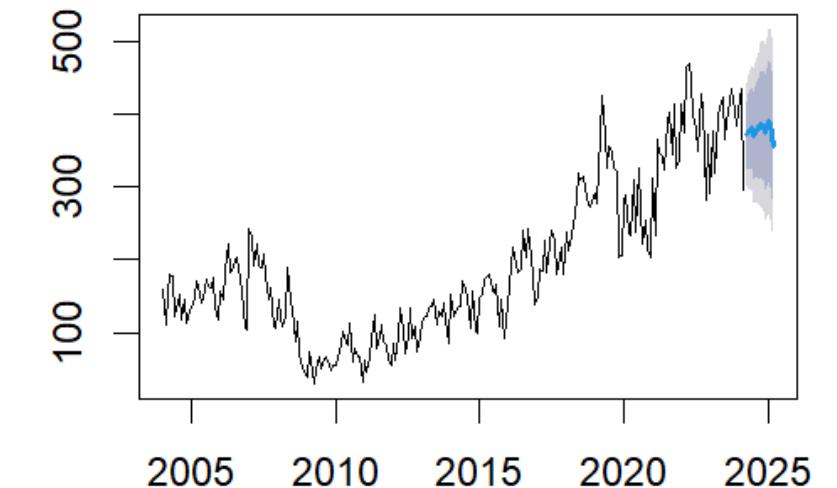
SÉRIE TEMPORELLE 2

```
# Modèle ARIMA automatique  
model_job <- auto.arima(job_openings_ts, seasonal = TRUE)  
summary(model_job)
```

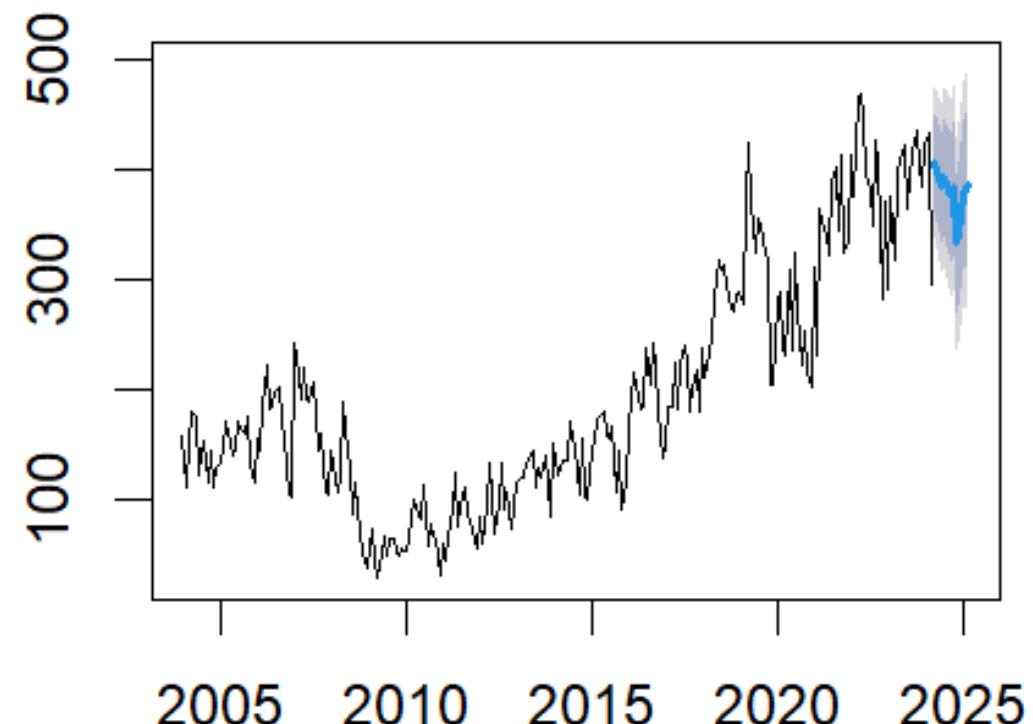
```
# Prévision  
forecast_job <- forecast(model_job, h = 12)  
plot(forecast_job)
```

```
# Holt-Winters  
hw_job <- HoltWinters(job_openings_ts)  
plot(hw_job)  
forecast_hw_job <- forecast(hw_job, h = 12)  
plot(forecast_hw_job)
```

Forecasts from ARIMA(0,1,1)(0,0,1)[12]



Forecasts from HoltWinters



SÉRIE TEMPORELLE 2

Comparaison

```
accuracy(model_job)  
accuracy(forecast_hw_job)
```

```
> accuracy(model_job)  
ME      RMSE     MAE      MPE      MAPE     MASE     ACF1  
Training set 1.425513 35.3769 26.5279 -2.755619 16.57779 0.5937356 0.02959635  
> accuracy(forecast_hw_job)  
ME      RMSE     MAE      MPE      MAPE     MASE     ACF1  
Training set -0.176538 34.40977 24.93633 -4.728665 16.72666 0.5581139 0.03431791
```

- Les deux modèles ont des performances similaires en termes de métriques d'erreur. Cependant, le modèle Holt-Winters présente un léger avantage avec des RMSE et MAE plus faibles, ce qui indique qu'il pourrait être légèrement mieux adapté pour prévoir les offres d'emploi dans ce cas.

PHASE 3



OBJECTIFS

1. Vérifier si les résidus du modèle sont des "bruits blancs" (non corrélés).
2. Choisir le bon modèle ARIMA final via tests et visualisations.

SÉRIE TEMPORELLE 1

```
resid_prod <- residuals(model_prod)

# Test de Ljung-Box pour la stationnarité des résidus (si les résidus sont des bruits blancs)
Box.test(resid_prod, lag = 20, type = "Ljung-Box")

# Histogramme des résidus
hist(resid_prod, main = "Histogramme des Résidus - ARIMA", xlab = "Résidus", breaks = 20)

# Test de normalité des résidus (test de Shapiro-Wilk)
shapiro.test(resid_prod)

> Box.test(resid_prod, lag = 20, type = "Ljung-Box")
    Box-Ljung test

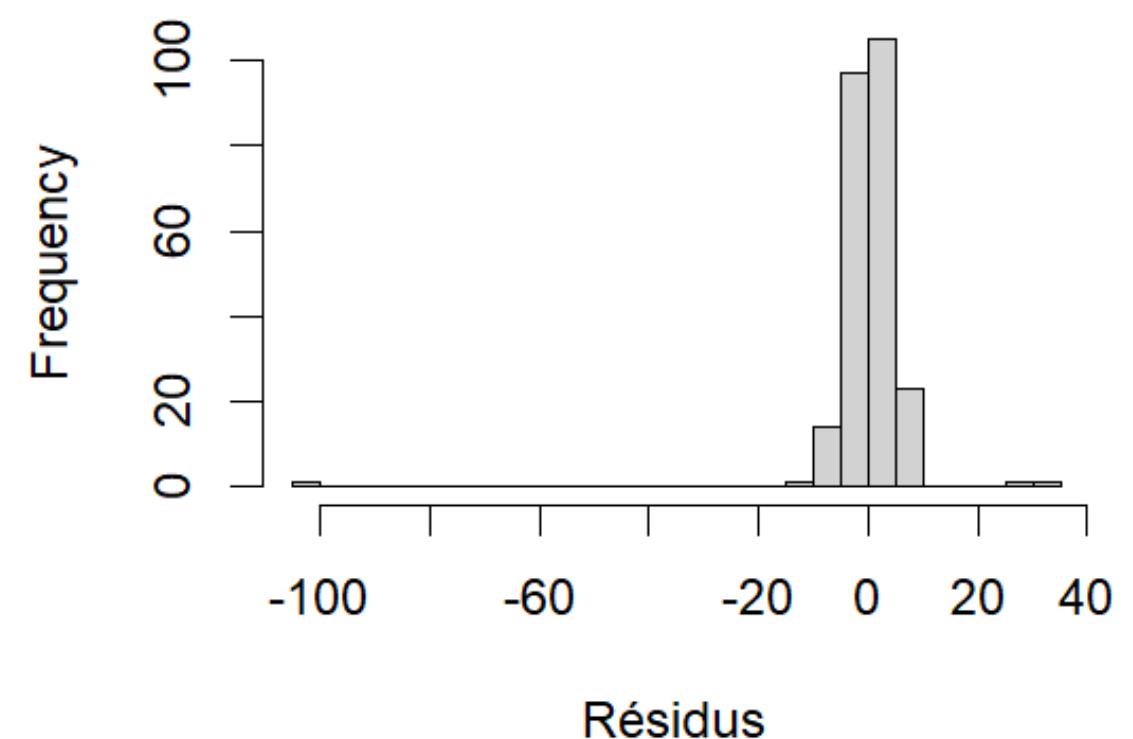
data: resid_prod
X-squared = 30.084, df = 20, p-value = 0.0685

> shapiro.test(resid_prod)

    Shapiro-Wilk normality test

data: resid_prod
W = 0.42405, p-value < 2.2e-16
```

Histogramme des Résidus - ARIMA



SÉRIE TEMPORELLE 1

```
# Comparaison des résidus avec ceux du modèle Holt-Winters
resid_hw_prod <- residuals(hw_model)
|
# Test de Ljung-Box pour les résidus du modèle Holt-Winters
Box.test(resid_hw_prod, lag = 20, type = "Ljung-Box")

# Histogramme des résidus du modèle Holt-Winters
hist(resid_hw_prod, main = "Histogramme des Résidus - Holt-Winters", xlab = "Résidus", breaks = 20)

# Test de normalité des résidus du modèle Holt-Winters
shapiro.test(resid_hw_prod)

> Box.test(resid_hw_prod, lag = 20, type = "Ljung-Box")
    Box-Ljung test

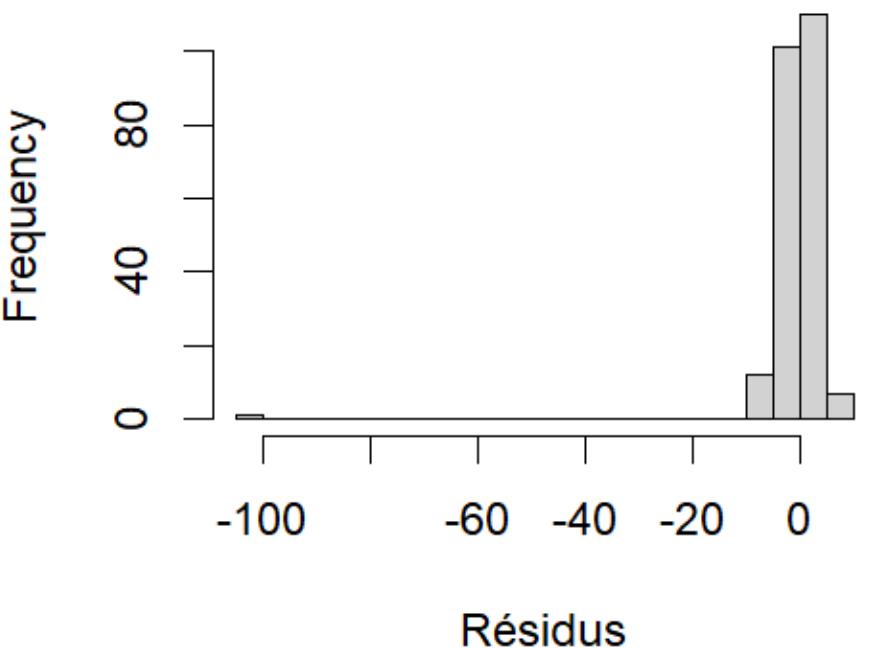
data: resid_hw_prod
X-squared = 9.2591, df = 20, p-value = 0.9797

> shapiro.test(resid_hw_prod)

    Shapiro-Wilk normality test

data: resid_hw_prod
W = 0.31059, p-value < 2.2e-16
```

Histogramme des Résidus - Holt-Wint



SÉRIE TEMPORELLE 1

Modèle ARIMA :

- Le test de Ljung-Box ne montre aucune preuve forte d'autocorrélation dans les résidus, ce qui est positif.
- Le test de Shapiro-Wilk suggère que les résidus ne sont pas distribués normalement, ce qui pourrait poser problème pour certaines hypothèses du modèle.

Modèle Holt-Winters :

- Le test de Ljung-Box indique aucune autocorrélation, ce qui est favorable. Comme pour le modèle ARIMA,
- Le test de Shapiro-Wilk suggère que les résidus ne suivent pas une distribution normale, impliquant que les résidus ne sont pas normalement distribués.

Toutefois, la non-normalité des résidus suggère que le modèle pourrait également laisser certains aspects non modélisés.

SÉRIE TEMPORELLE 2

```
resid_job <- residuals(model_job)

# Test de Ljung-Box pour la stationnarité des résidus (si les résidus sont des bruits blancs)
Box.test(resid_job, lag = 20, type = "Ljung-Box")

# Histogramme des résidus
hist(resid_job, main = "Histogramme des Résidus - ARIMA", xlab = "Résidus", breaks = 20)

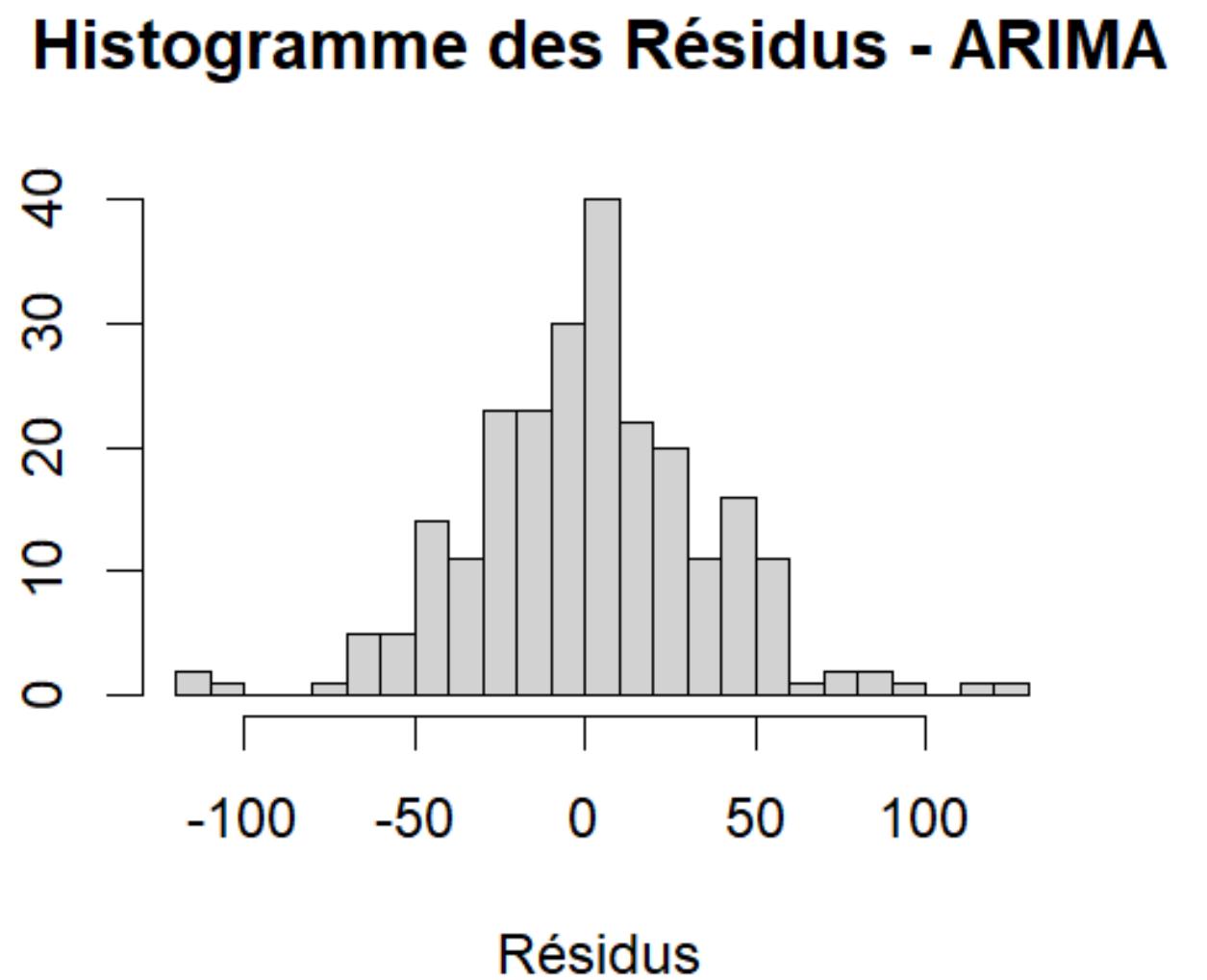
# Test de normalité des résidus (test de Shapiro-Wilk)
shapiro.test(resid_job)

> Box.test(resid_job, lag = 20, type = "Ljung-Box")
    Box-Ljung test

data: resid_job
X-squared = 29.251, df = 20, p-value = 0.08293

> shapiro.test(resid_job)
    Shapiro-Wilk normality test

data: resid_job
W = 0.98284, p-value = 0.00495
```



SÉRIE TEMPORELLE 2

```
resid_hw_job <- residuals(hw_job)

# Test de Ljung-Box pour les résidus du modèle Holt-Winters
Box.test(resid_hw_job, lag = 20, type = "Ljung-Box")

# Histogramme des résidus du modèle Holt-Winters
hist(resid_hw_job, main = "Histogramme des Résidus - Holt-Winters", xlab = "Résidus", breaks = 20)
```

```
# Test de normalité des résidus du modèle Holt-Winters
shapiro.test(resid_hw_job)
```

```
> Box.test(resid_hw_job, lag = 20, type = "Ljung-Box")
```

Box-Ljung test

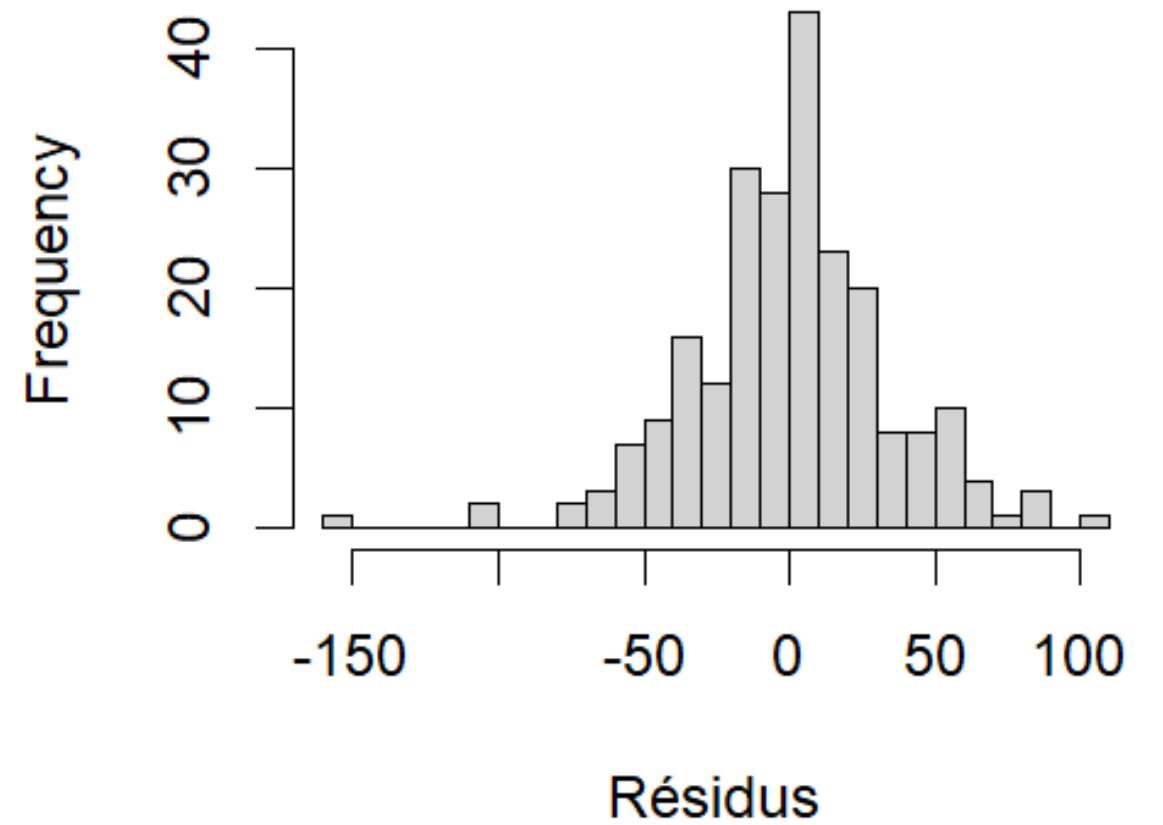
```
data: resid_hw_job
X-squared = 21.469, df = 20, p-value = 0.37
```

```
> shapiro.test(resid_hw_job)
```

Shapiro-Wilk normality test

```
data: resid_hw_job
W = 0.97248, p-value = 0.000182
```

Histogramme des Résidus - Holt-Winters



SÉRIE TEMPORELLE 2

Test de Ljung-Box :

- Le modèle ARIMA montre une légère autocorrélation (p-value de 0.08293), mais cela reste acceptable.
- Le modèle Holt-Winters montre aucune autocorrélation (p-value de 0.37), ce qui indique que les résidus sont bien aléatoires.

Test de normalité :

- ARIMA : Les résidus ne sont pas normalement distribués (p-value de 0.00495).
- Holt-Winters : Les résidus ne suivent pas non plus une distribution normale (p-value de 0.000182).

CONCLUSION

- Autocorrélation : Les résidus du modèle Holt-Winters semblent mieux se comporter en termes d'autocorrélation, ce qui suggère qu'il capture mieux la structure des données sans laisser de patterns non expliqués dans les erreurs.
- Normalité des résidus : Les deux modèles présentent des résidus qui ne sont pas normalement distribués, ce qui pourrait limiter l'interprétation des tests statistiques fondés sur l'hypothèse de normalité, mais cela n'empêche pas nécessairement l'utilisation des modèles dans la prévision.



MERCI !