

1961-2024

Analyse des Séries Chronologiques du Chômage et du PIB Nominal du Canada



Noms des membres :

*Gloria Gbossou

*Ralph Kplé

* Hanniel Kouamé

*Alaine Kamanzi

Table des matières

Description des Données	2
Analyse du PIB Nominal	2
Analyse du Taux de Chômage	4
Filtrage	7
Filtrage linéaire	7
Filtrage exponentiel	9
Différenciation	19
Application de la différenciation	19
Analyse de la stationnarité des séries différenciées	25
Stationnarité	26
Obtention des séries de moyennes constantes	26
Analyse de l'autocorrélation	27
Modélisation et Prévision avec les Processus ARMA	30
Sélection des Ordres (p,q) des modèles ARMA	30
Analyse de l'ACF et du PACF	31
Estimation et comparaison des modèles ARMA	33
Estimation par Maximum de Vraisemblance	33
Analyse des Paramètres et Modèles Imbriqués	38
Validation Diagnostique et Performance des Modèles	40
Critères d'Information (AIC/BIC)	40
Test du Ratio de Vraisemblance	41
Erreurs Quadratiques Moyennes (EQM)	42
Modélisation Bivariée et Corrélations Croisées	43
Analyse des innovations	44
Corrélation croisée (CCF) entre les innovations	47
Conclusion	48
Références	49
Annexe	49
Codes	49

Description des Données

Intérêt pour les séries

Avant d'entrer dans l'analyse statistique, il est important de comprendre pourquoi nous avons choisi d'étudier ces deux séries : le PIB nominal et le taux de chômage. Le PIB nominal est un indicateur clé de l'activité économique, représentant la valeur totale des biens et services produits au Canada. Quant au taux de chômage, il reflète l'état du marché du travail et a un impact direct sur la croissance économique et les politiques publiques.

Nous avons choisi ces séries parce qu'elles sont étroitement liées selon la loi d'Okun, une relation économique qui suggère que lorsque l'économie croît rapidement, le chômage tend à diminuer, et inversement. Cette dynamique entre le PIB et le chômage est souvent analysée pour mieux comprendre les cycles économiques et guider les décisions politiques visant à stimuler la croissance et réduire les inégalités. Il est donc crucial d'étudier ces deux variables dans un contexte plus large, en prenant en compte les facteurs externes qui peuvent influencer cette relation, tels que les crises économiques ou les changements technologiques.

On utilise d'ailleurs ces deux séries en raison du lien entre le PIB et le taux de chômage dans une économie, établi par l'économiste américain Arthur Melvin Okun. Le grand économiste et chercheur Edward S. Knotek II (quand il était à la Federal Reserve Bank of Kansas City) en a d'ailleurs fait une étude plus poussée afin d'en démontrer l'utilité (Knotek, 2007).

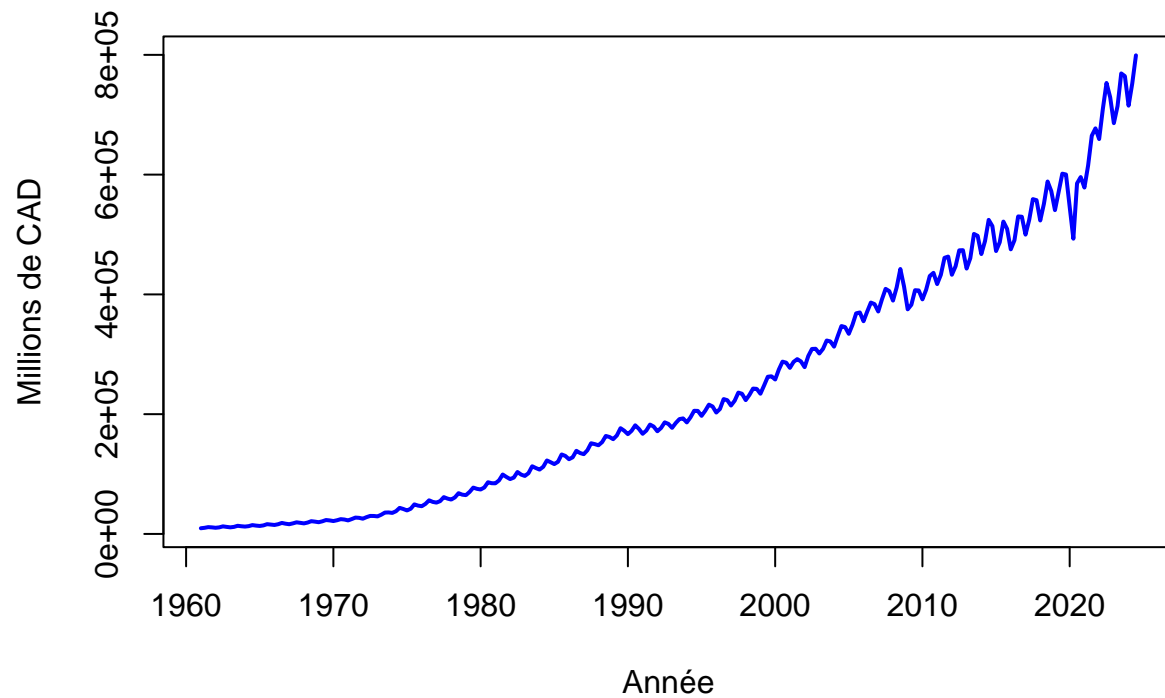
Pour notre analyse, nous avons choisi deux séries chronologiques économiques du Canada, issues de la base de données **FRED (Federal Reserve Economic Data)**:

Analyse du PIB Nominal

Source et Fréquence

La première série concerne le **Produit Intérieur Brut (PIB) Nominal**, qui représente la valeur totale de tous les biens et services produits au Canada sans ajustement pour l'inflation. Les données proviennent du **Fonds Monétaire International (FMI)**, extraite de la base de données **International Financial Statistics (IFS)** (Accès à la série sur FRED) et ont été consultées le 4 *Février*. Elles sont enregistrées à une **fréquence trimestrielle**, couvrant la période de **1961 à 2024**, et sont exprimées en **millions de dollars canadiens (CAD)**, sans ajustement aux variations saisonnières.

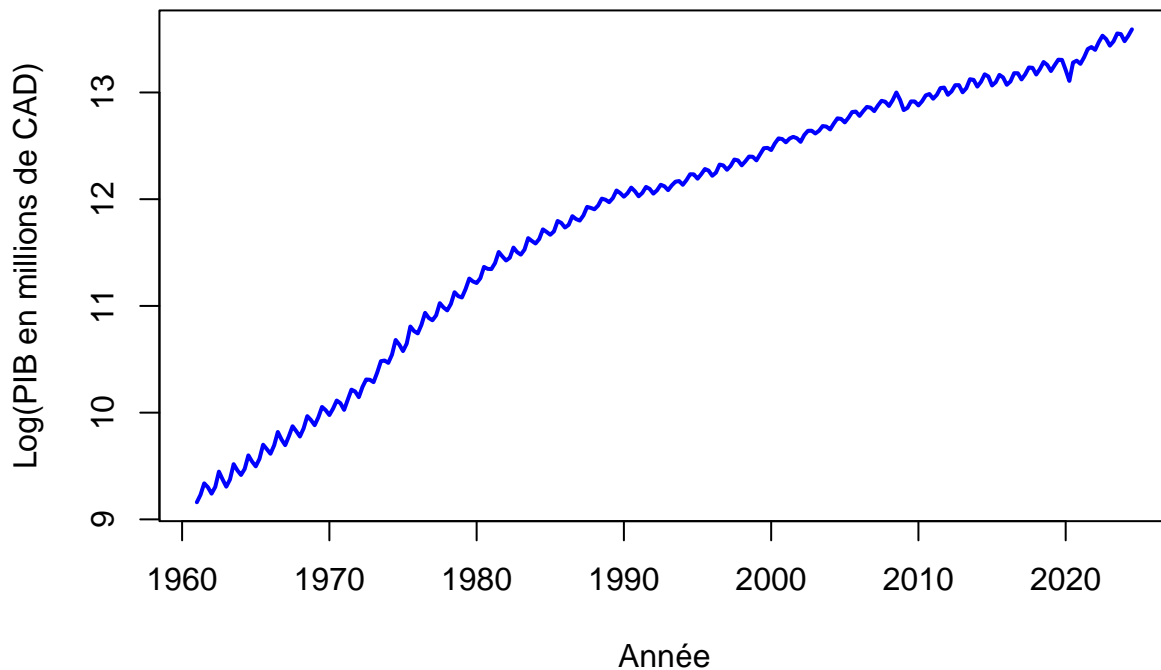
PIB Nominal du Canada (1961–2024)



Transformation de la Série

Le Produit Intérieur Brut (PIB) du Canada suit une croissance exponentielle au fil du temps. En prenant le logarithme naturel du PIB, cette croissance exponentielle est transformée en une tendance linéaire, ce qui facilite l'interprétation des taux de croissance. Cette transformation permet également de stabiliser la variance des données, rendant ainsi notre modèle économique plus précis et fiables.

Logarithme du PIB Nominal du Canada (1961–2024)



Statistiques Descriptives et Traitement des Données

```
##      Statistique Valeur
## 1      Minimum    9.16
## 2  1er Quartile  10.90
## 3      Médiane   12.12
## 4      Moyenne   11.83
## 5  3ème Quartile  12.91
## 6      Maximum   13.59
## 7      Écart-type  1.26
```

```
## Valeurs manquantes : 0
```

La moyenne du logarithme du PIB nominal est de 11,83 millions de dollars canadiens, ce qui est légèrement inférieur à la médiane. Cela suggère une légère baisse de la croissance économique au fil du temps. Son écart-type, quant à lui, est de 1,26 millions de dollars, indiquant une faible variabilité et donc une stabilité dans la croissance économique. De plus, il n'y a aucune valeur manquante dans cette série, ce qui signifie que toutes les observations sont complètes et fiables pour l'analyse.

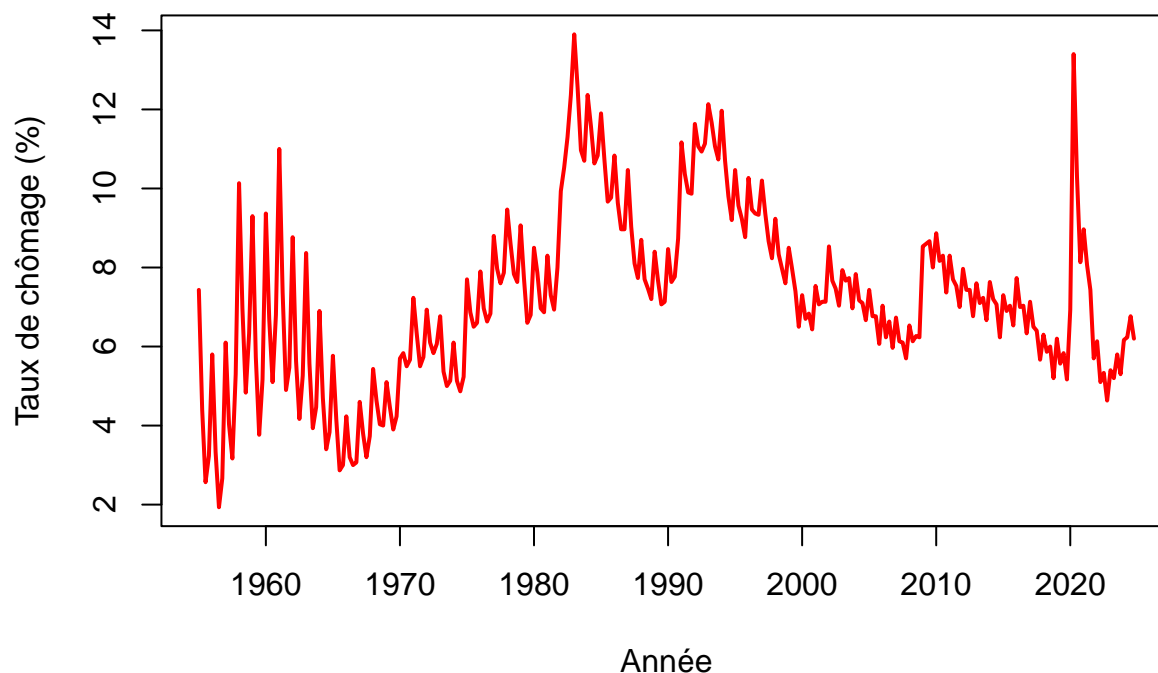
Analyse du Taux de Chômage

Source et Fréquence

La seconde série étudiée porte sur le **Taux de Chômage Total des personnes âgées de 15 ans et plus**, qui mesure la proportion de la population active actuellement sans emploi mais en recherche active. Ces

données sont également issues de l'**Organisation de coopération et de développement économiques (OCDE)**, extraites de l'indicateur "**Main Economic Indicators**" (Accès à la série sur FRED) et ont été consultées le 4 *Fevrier*. Comme la première série, elles sont publiées à une **fréquence trimestrielle** et couvrent la période de **1961 à 2024**. L'unité de mesure utilisée est le **pourcentage (%) de la population active, sans ajustement aux variations saisonnières**.

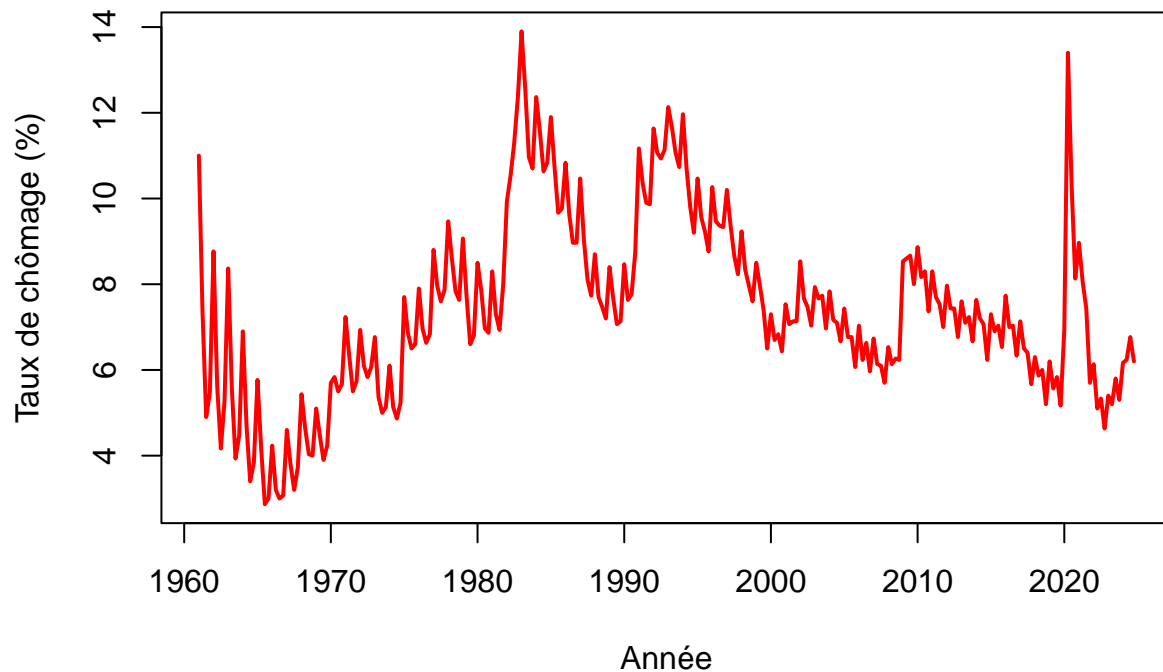
Taux de Chômage au Canada (1955–2024)



Ajustements sur la Série

Étant donné que la série du chômage commence en 1955, tandis que la série du PIB commence en 1961, il y a une différence de période d'observation entre les deux. Afin d'assurer une comparabilité optimale entre ces deux séries et d'éviter toute distorsion dans l'analyse, nous avons décidé d'éliminer les observations du taux de chômage antérieures à 1961.

Taux de Chômage au Canada (1961–2024)



Statistiques Descriptives et Traitement des Données

##	Statistique	Valeur
## 1	Minimum	2.87
## 2	1er Quartile	6.10
## 3	Médiane	7.15
## 4	Moyenne	7.41
## 5	3ème Quartile	8.55
## 6	Maximum	13.90
## 7	Écart-type	2.11

Valeurs manquantes : 0

La moyenne du taux de chômage est de 7,41 %, ce qui est légèrement supérieur à la médiane. Cela suggère qu'il y a eu des périodes de chômage plus élevé au cours de la période étudiée. L'écart-type est de 2,11 %, ce qui est relativement élevé et indique de grandes variations du taux de chômage. Ces variations peuvent être attribuées à des récessions et à des reprises économiques. De plus, il n'y a aucune valeur manquante dans cette série, ce qui garantit que toutes les observations sont disponibles et complètes pour l'analyse.

Hypothèses et Interprétation

Tendance et Saisonnalité

Hypothèses sur la tendance

PIB Nominal : À vue d'œil, la série du PIB nominal semble présenter une tendance croissante, probablement exponentielle, ce qui justifie l'application du logarithme. Cette croissance peut être attribuée à l'expansion économique, à l'inflation et aux variations des investissements au fil du temps.

Taux de chômage : La série du taux de chômage ne montre pas de tendance claire sur l'ensemble de la période. Cependant, on observe des cycles économiques marqués par des périodes de hausse et de baisse, ce qui pourrait indiquer la présence de cycles plutôt qu'une véritable tendance à long terme.

Hypothèses sur la saisonnalité

PIB Nominal : Le PIB nominal étant mesuré trimestriellement, il est possible qu'il y ait des fluctuations saisonnières (par exemple, une hausse en fin d'année en raison de la consommation des fêtes).

Taux de chômage : Le chômage peut théoriquement être sujet à des variations saisonnières, notamment en raison du marché du travail saisonnier (ex. : embauches en été et licenciements après la période des fêtes). Toutefois, à première vue, la série ne semble pas montrer de saisonnalité évidente.

Conclusion

En résumé, on formule l'hypothèse d'une tendance croissante pour le PIB, d'une absence de tendance marquée pour le chômage, et d'une saisonnalité potentielle pour les deux séries.

Filtrage

Passons maintenant au filtrage, le filtrage est une étape essentielle dans l'analyse de nos séries temporelles, car il permet de séparer les composantes sous-jacentes d'une série, notamment la tendance et les fluctuations aléatoires.

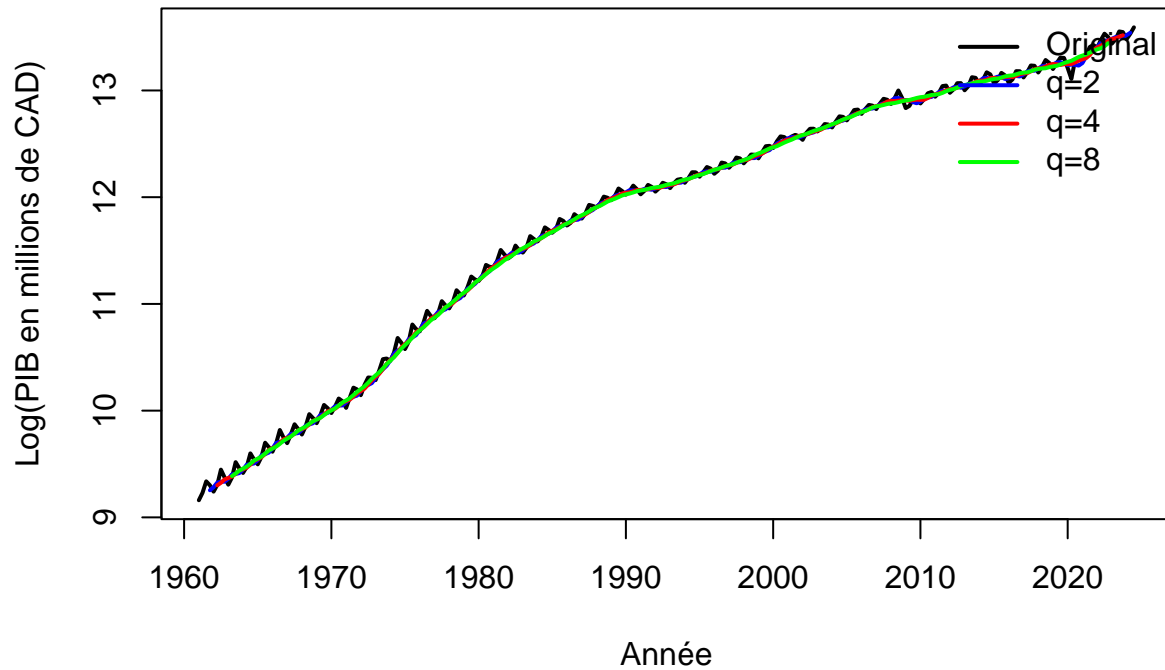
En appliquant des méthodes de lissage comme le lissage linéaire et exponentielle, nous cherchons à mieux comprendre la structure des données et à faciliter l'interprétation des variations.

Filtrage linéaire

On propose arbitrairement $q=2$, $q=4$ et $q=8$ comme paramètres de lissage linéaire pour chacune de nos séries.

Série du PIB

Lissage linéaire du PIB

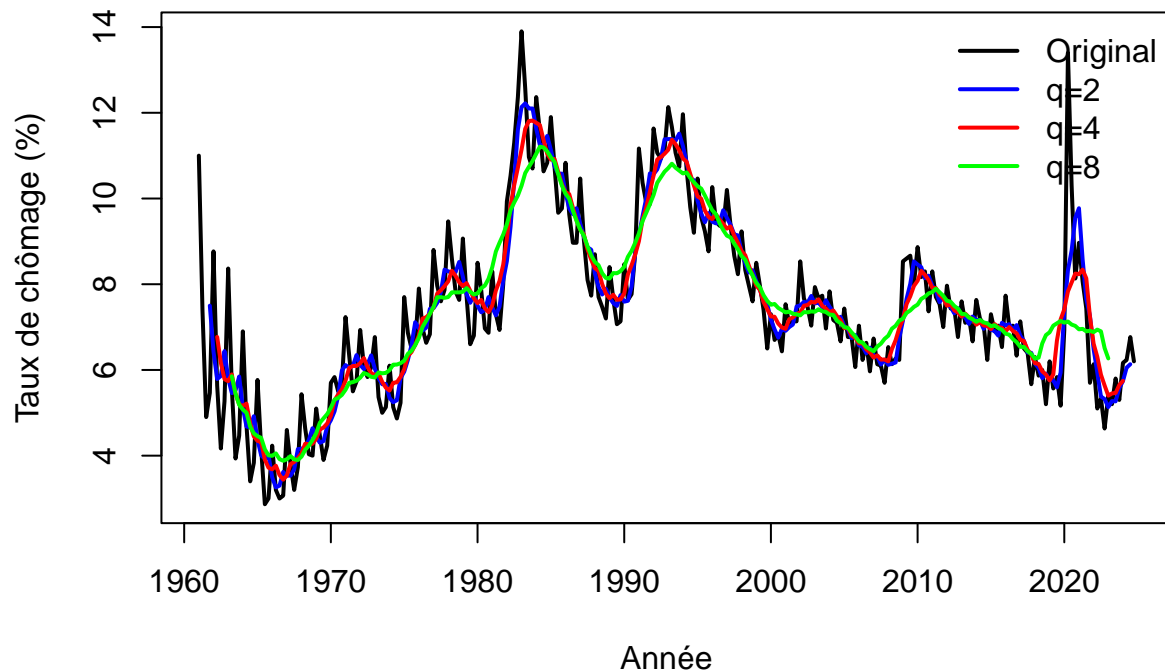


En observant les résultats de nos différents lissages linéaires sur la série du PIB les courbes se chevauchent presque complètement, il est assez difficile de voir la différence entre les niveaux de lissage.

Mais en observant plus en profondeur, nous avons choisi de garder $q=4$ comme paramètre de lissage linéaire. Dans l'ensemble, ce graphique montre que le PIB suit une tendance croissante stable, et que le choix du paramètre q n'a qu'un impact mineur sur l'allure générale de la série.

Série du Chômage

Lissage linéaire du taux de chômage



Contrairement au graphique du PIB, ici, on observe plus de distinction entre les différentes valeurs de q .

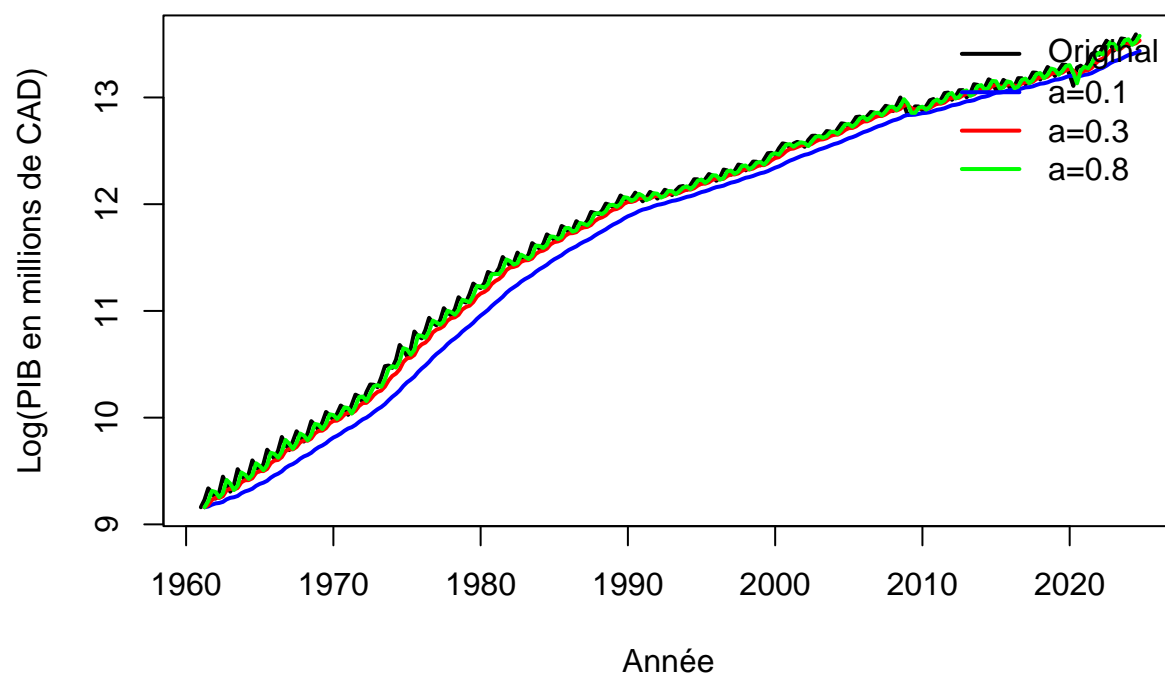
Lorsqu'on applique pour ces différents paramètres, le lissage linéaire à la série du taux de chômage, l'on observe que le lissage n'est pas si efficace quand $q=2$, et que des variations importantes dans notre série sont "minimisées" ou "éliminées" quand $q=8$. Par contre lorsque $q=4$, le lissage effectué est plus ou moins bon et les variations importantes dans notre série sont conservées. Ainsi, prendre $q=4$ comme paramètre de lissage semble donner le meilleur compromis entre la qualité du lissage et la conservation des informations importantes.

Filtrage exponentiel

Pour le lissage exponentiel, l'on propose arbitrairement différentes valeurs pour le paramètre a notamment 0.1, 0.3 et 0.8. Et l'on comparera les résultats selon chaque scénario afin de retenir le paramètre de lissage au résultat le plus approprié.

Série du PIB

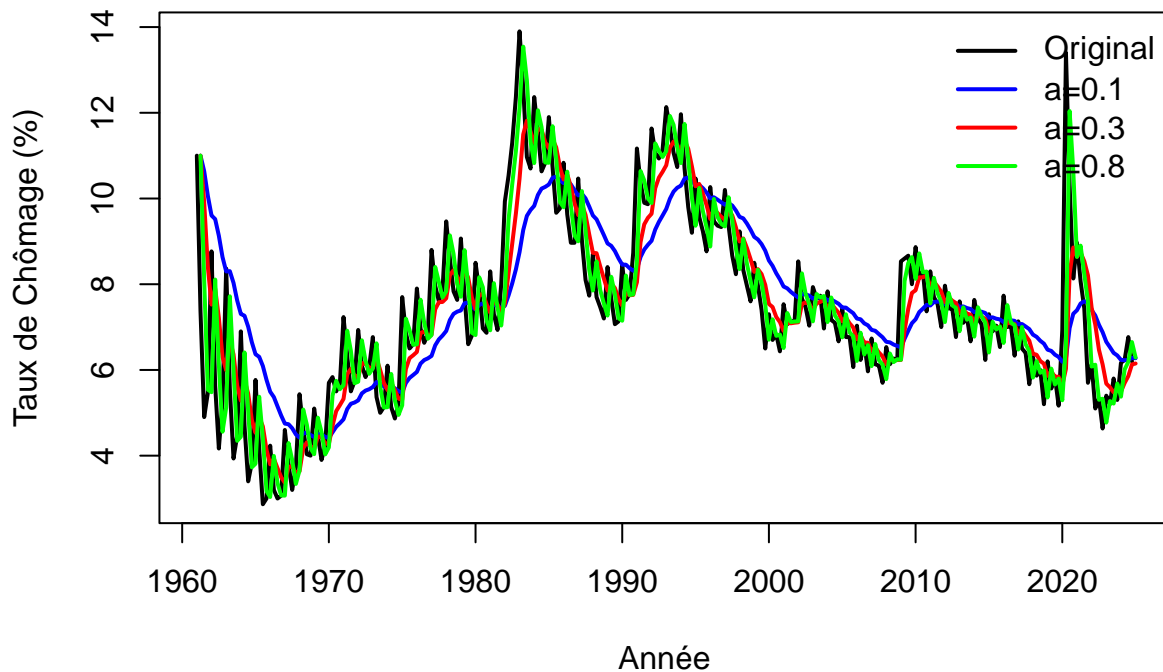
Lissage exponentiel du PIB



Pour la série du PIB, le lissage exponentiel à paramètre $\alpha=0.3$ est celui qui nous permet de toujours identifier les variations présentes dans notre série tout en fournissant un bon lissage. Il faudrait donc conserver $\alpha=0.3$ pour notre lissage exponentiel par rapport à tous les paramètres proposés.

Série du Chômage

Lissage exponentiel du taux de chômage



Pour la série du taux de chômage, le lissage exponentiel avec $a=0.3$ semble donner un meilleur équilibre entre réduction de bruit et préservation de données tout en préservant les variations importantes de la tendance. Le lissage pour $a=0.8$ présente trop de bruit et celui pour $a=0.1$ perd des informations significatives sur nos données (comme des variations importantes).

On a finalement conservé les séries suivantes selon chaque type de lissage :

Pour chacune de nos séries, le choix entre les séries lissées selon le type de lissage réalisé n'est pas aussi évident.

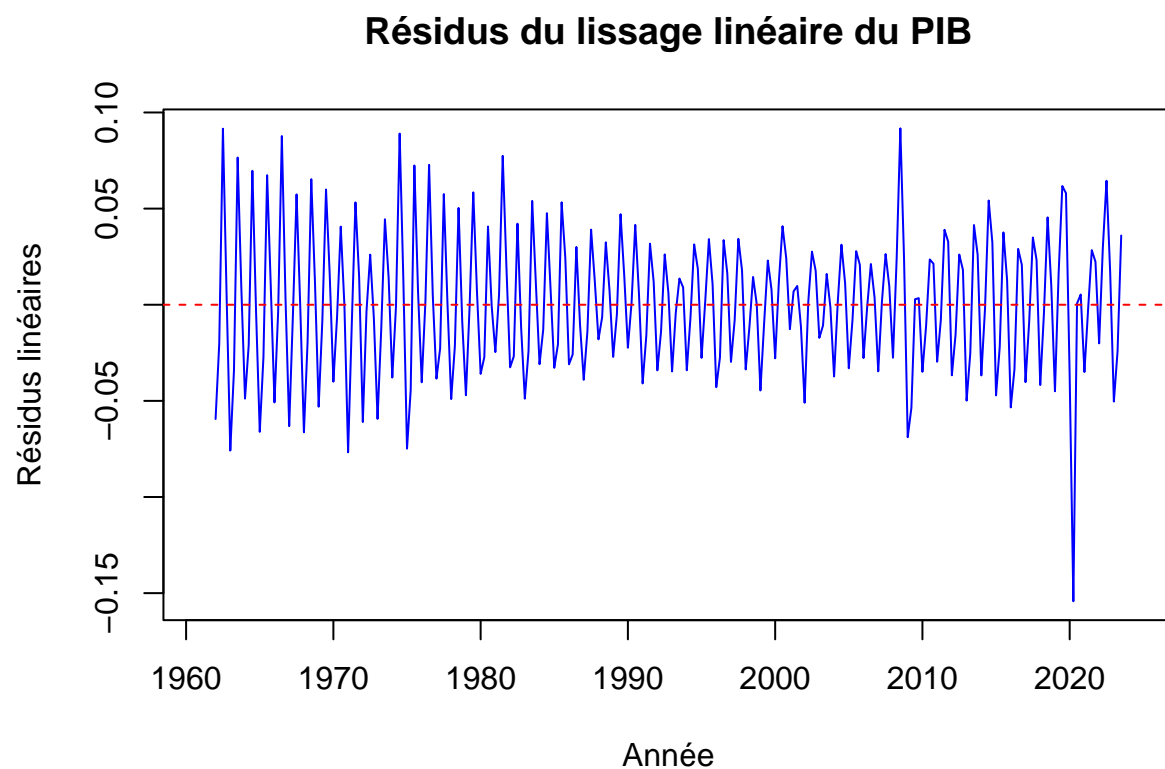
En effet, nous devons choisir une technique de lissage, mais nous sommes assez confus en raison des différences des tendances et des variations. Certaines méthodes préservent mieux les fluctuations locales, tandis que d'autres mettent davantage en avant la tendance générale.

Tout ceci complique la décision.

Pour cela, nous allons émettre une petite analyse sur les résidus correspondant à chaque type de lissage.

Analyse des Résidus

Résidus Lissage linéaire ~ Série du PIB

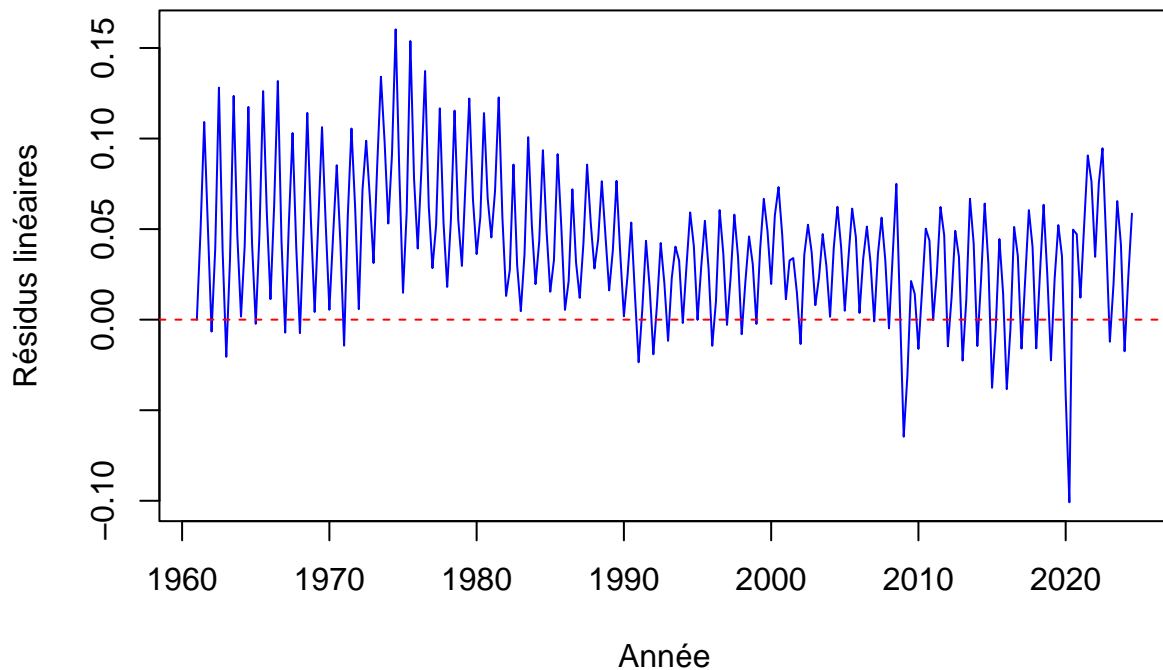


Moyenne des résidus du lissage linéaire : 2.170445e-05

La moyenne des résidus linéaires est proche de 0, ce qui suggère un bon ajustement du lissage linéaire.

Résidus Lissage exponentiel Série du PIB

Résidus du lissage exponentiel du PIB



Moyenne des résidus du lissage exponentiel : 0.0400076

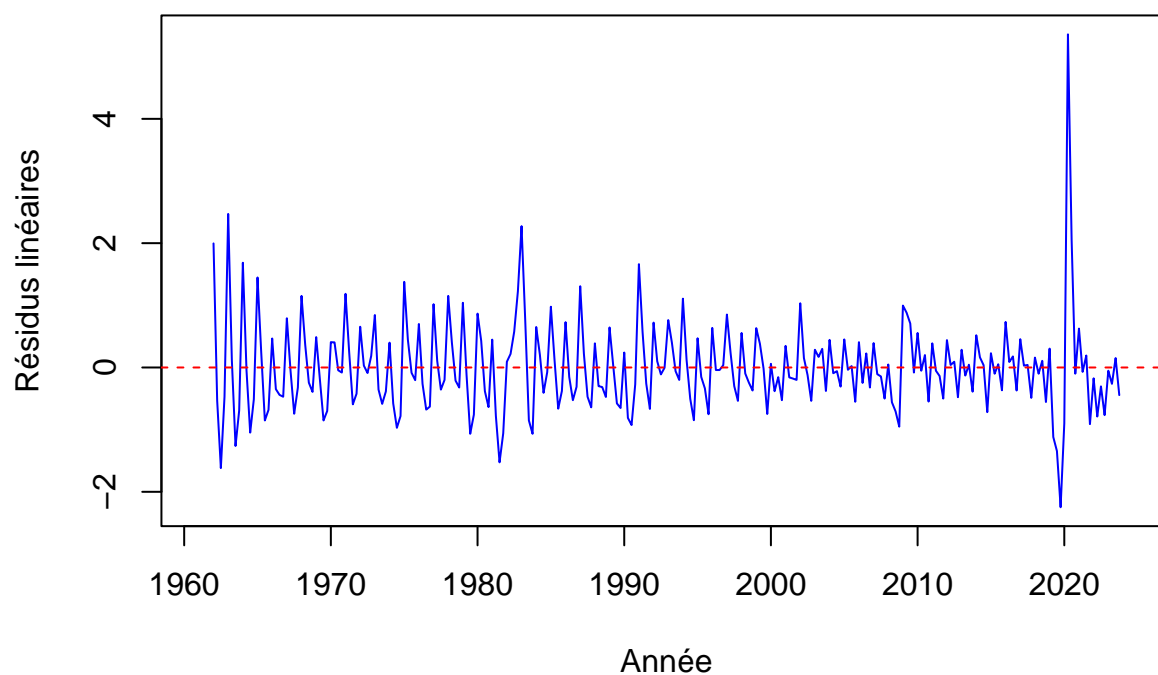
La moyenne des résidus exponentiels est plus élevée (0.0400076), indiquant un possible biais dans l'estimation.

Conclusion

Pour notre série reliée à l'évolution du PIB, la série résiduelle associée au lissage linéaire semble prendre des valeurs plus proches de 0 par rapport celle associée au lissage exponentiel. Cela présagerait une moyenne plus proche de 0 pour cette série. On observe également que pour cette série, la moyenne des observations est plus proche de 0 que celle de la deuxième. Ainsi, le lissage linéaire serait la méthode de lissage la plus adaptée pour notre série.

Résidus Lissage linéaire ~ Série du taux Chômage

Résidus du lissage linéaire du taux de chômage

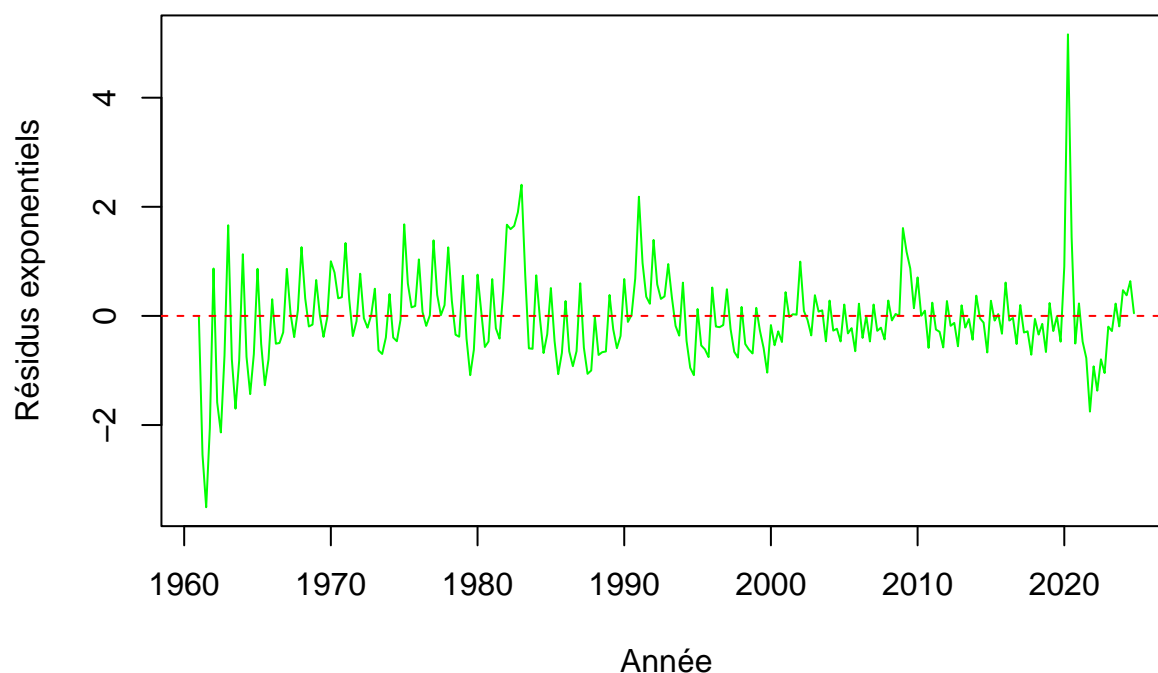


Moyenne des résidus du lissage linéaire : -0.002434292

Cette valeur est très proche de zéro, ce qui suggère que le lissage linéaire ne présente pas de biais systématique significatif.

Résidus Lissage exponentiel ~ Série du taux Chômage

Résidus du lissage exponentiel du taux de chômage



Moyenne des résidus du lissage exponentiel du taux de chômage : -0.04420147

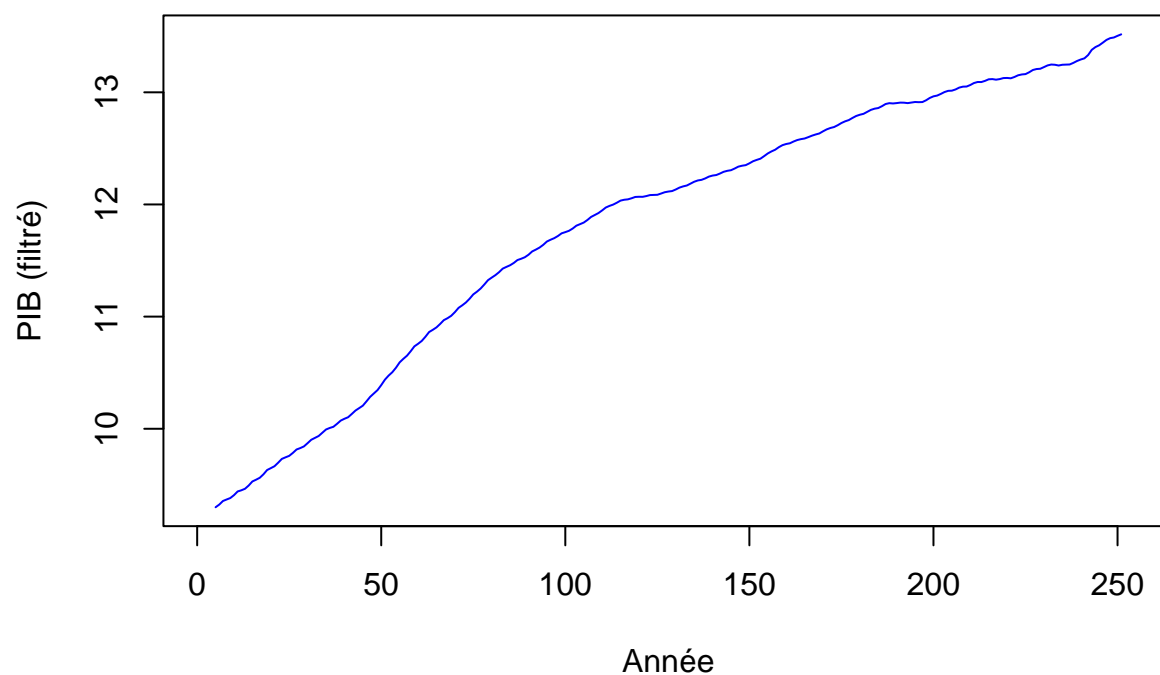
La moyenne des résidus est légèrement plus éloignée de zéro, ce qui pourrait indiquer un léger biais dans le modèle exponentiel. Cela signifie que cette méthode pourrait avoir tendance à sous-estimer ou surestimer certaines valeurs.

Conclusion

Pour notre série reliée au taux de chômage, la série résiduelle associée au lissage linéaire semble prendre des valeurs plus proches de 0 par rapport celle associée au lissage exponentiel. Cela présagerait une moyenne plus proche de 0 pour cette série. On observe également que pour cette série la moyenne des observations est plus proche de 0 que celle de la deuxième. Ainsi, le lissage linéaire serait la méthode de lissage la plus adaptée pour notre série.

Série filtrée du PIB

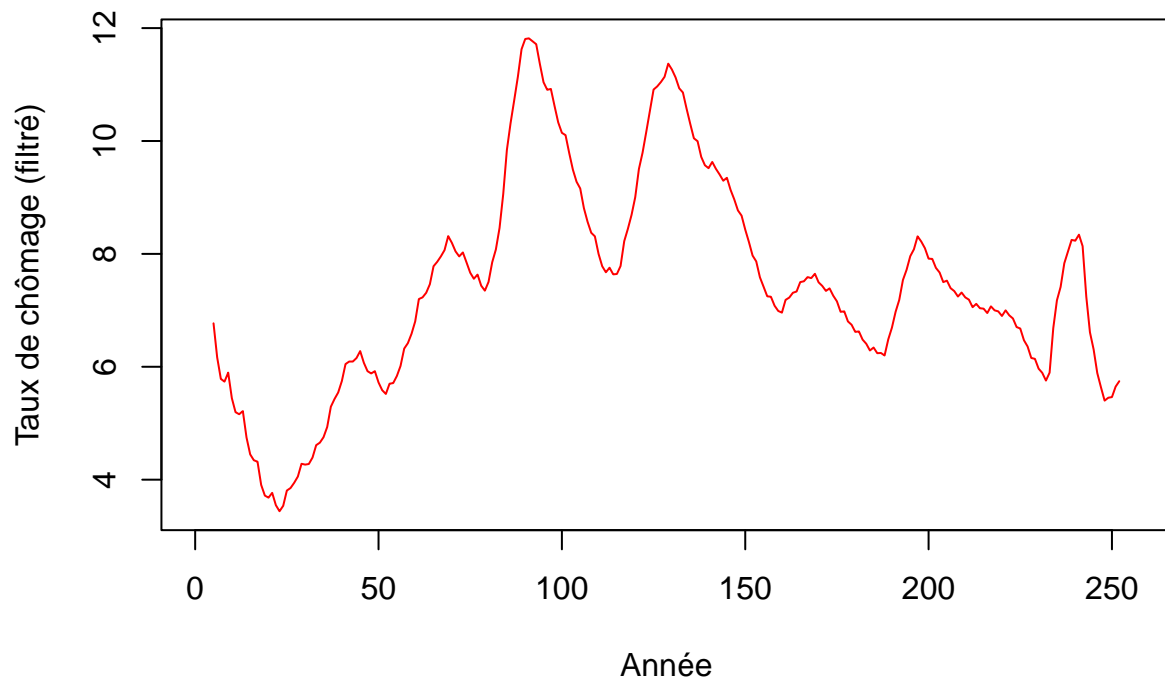
Série du PIB après lissage linéaire



La moyenne des résidus est : 2.170445e-05 et la variance des résidus est : 0.001325118

Série filtrée du taux de chômage

Série du taux de chômage après lissage linéaire



La moyenne des résidus est : -0.002434292 et la variance des résidus est : 0.5508259

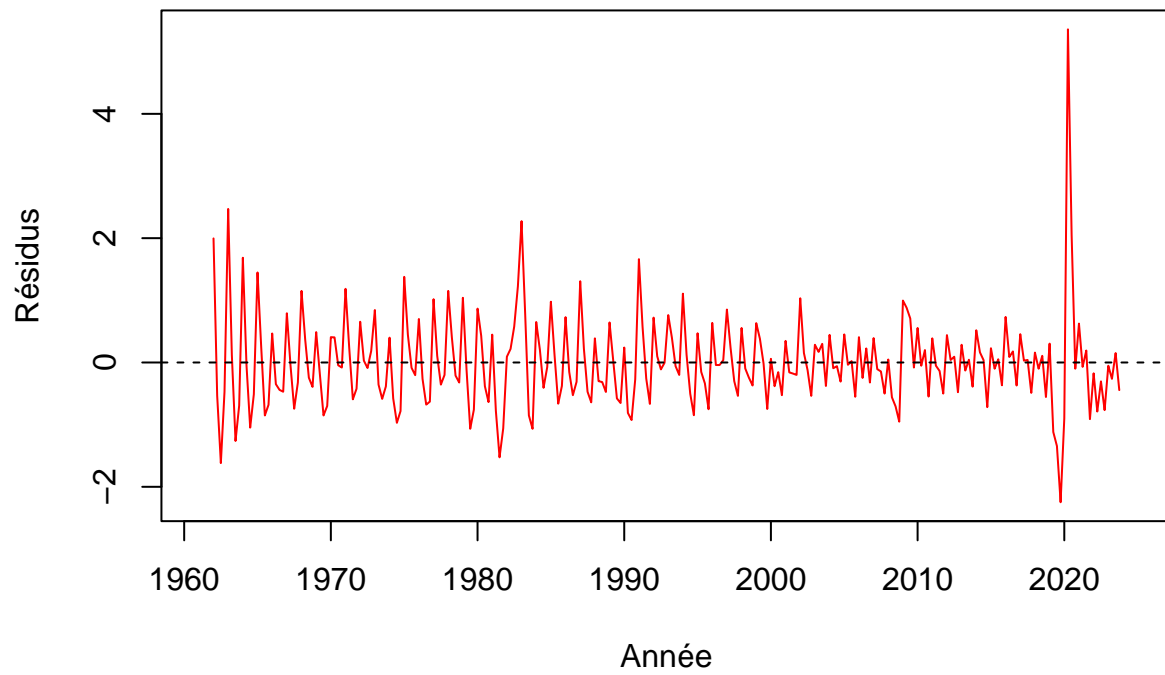
Stationnarité des résidus

En observant la série des résidus, on remarque la présence de composantes saisonnières, ce qui indique une variabilité périodique. Cette répétition semble montrer que la moyenne des résidus n'est pas constante au cours du temps. Aussi l'amplitude des variations change cela signifie que la variance n'est pas constante non plus.

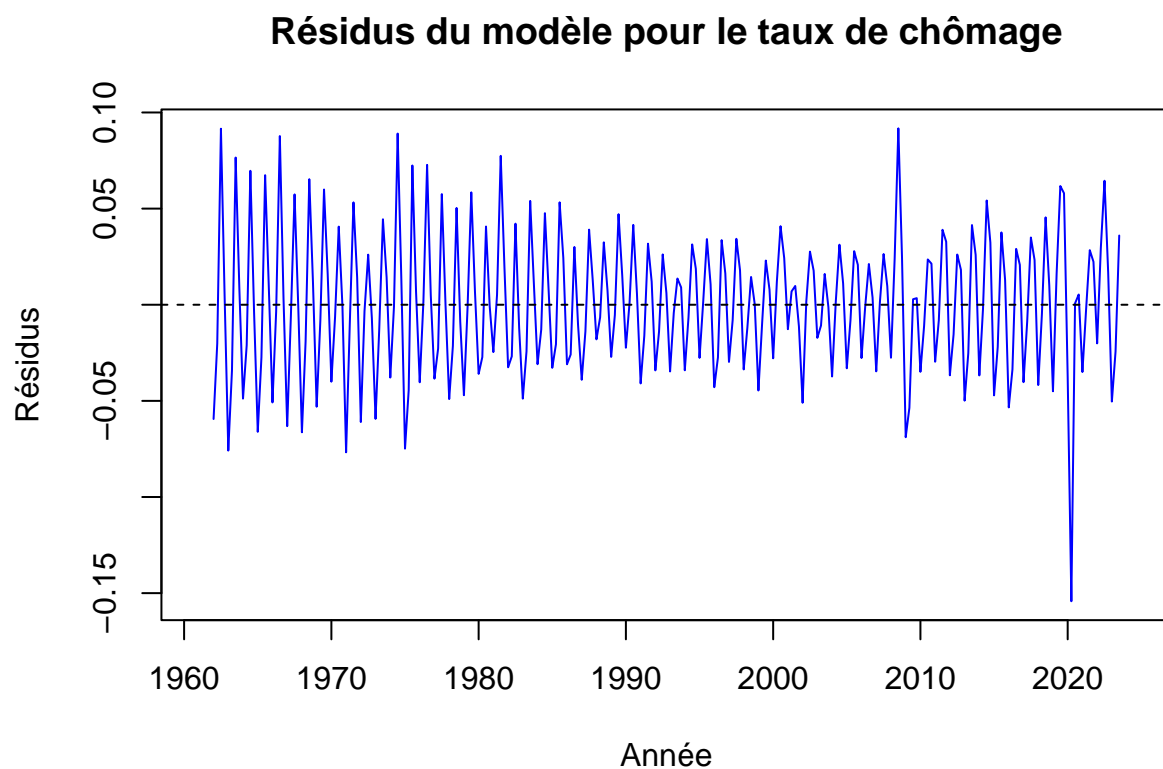
Or, pour qu'une série soit stationnaire au sens faible, il faut que la moyenne et la variance restent constantes dans le temps et que l'autocorrélation ne dépende que du décalage temporel et non du temps lui-même. Vu que l'observation des résidus montre des fluctuations régulières de la moyenne et/ou de la variance, on peut conclure que les résidus ne sont pas stationnaires au sens faible. On pourra effectuer par la suite d'autres filtrages pour rendre nos séries stationnaires.

Série du PIB

Résidus du modèle pour le PIB



Série du taux de chômage



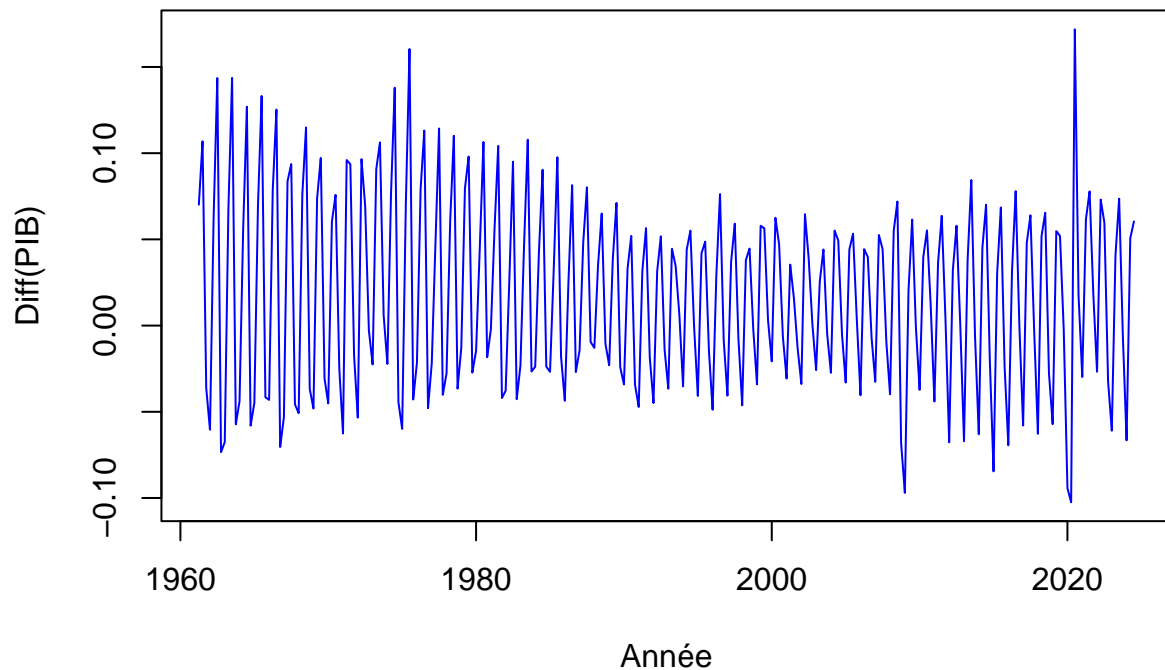
Différenciation

Application de la différenciation

Série du PIB

Dans cette section, nous allons appliquer la première différenciation $Y'_t = Y_t - Y_{t-1}$ sur notre série dans le but d'éliminer sa tendance croissante.

Différenciation Ordre 1



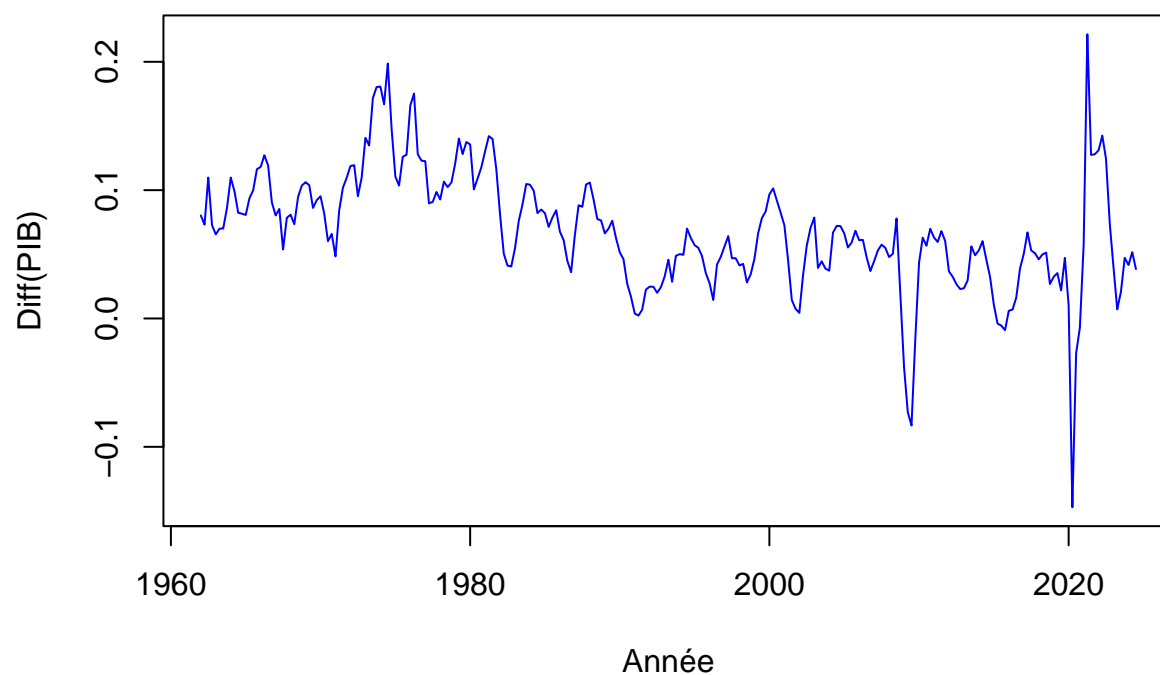
On remarque une forte composante saisonnière tout au long de la période, probablement dû à des variations trimestrielles du PIB. Même après une différenciation d'ordre 1, la saisonnalité reste marquée, ce qui suggère d'appliquer une différenciation saisonnière.

Il existe des périodes où la moyenne semble être légèrement positive, suivies de périodes où elle tend vers zéro.

En résumé, cette différenciation de premier ordre réduit la tendance mais ne stabilise pas totalement la série. Il reste une forte saisonnalité, une moyenne non constante.

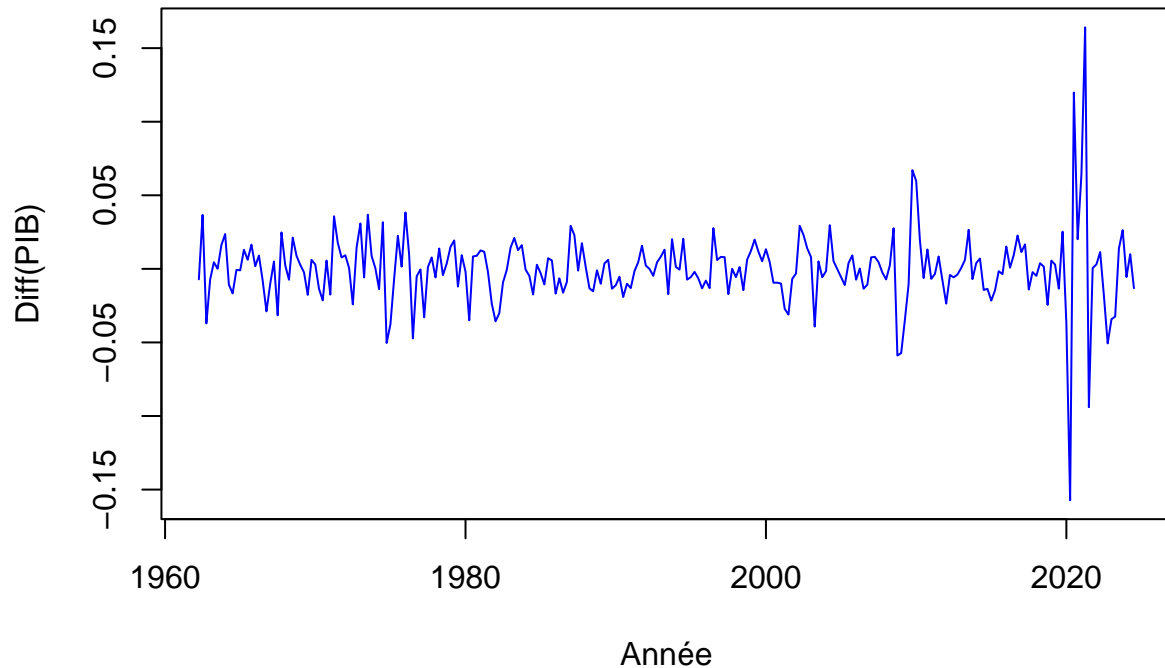
Appliquons la différenciation d'ordre 4, pour éliminer l'effet saisonnier de la série. Nous avons choisi l'ordre 4 car nous avons des données trimestrielles, les fluctuations récurrentes, liées à la saisonnalité, se manifestent généralement tous les 4 trimestres.

Différenciation Ordre 4



Maintenant que nous avons éliminé la saisonnalité grâce à la différenciation d'ordre 4, il reste à nous attaquer à la tendance présente dans la série. En effet, même après avoir retiré l'effet saisonnier, notre série présente une tendance linéaire qui empêche l'obtention d'une série assez stationnaire.

Différenciation Ordre 4+1



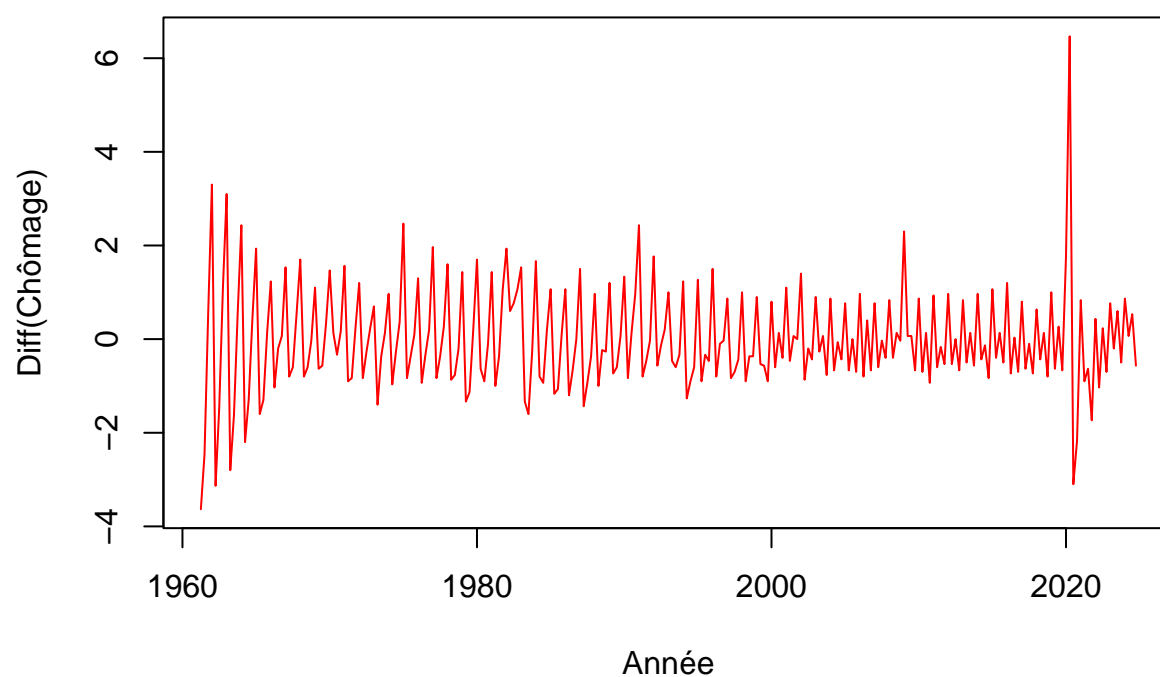
La transformation par différenciation d'ordre 4, puis d'ordre 1 a réussi à atténuer les composantes non stationnaires. Le graphique obtenu après avoir appliqué la différenciation d'ordre 4 et d'ordre 1 semble montrer une moyenne relativement constante au cours du temps. Même lorsque la série présente des fluctuations significatives, nous pouvons observer qu'elles suivent un schéma régulier, tant à la hausse qu'à la baisse. Cela signifie que les pics et les creux ont une certaine réciprocity dans leur apparition, ce qui tend à entraîner une annulation relative des variations extrêmes.

Bien que le graphique suggère une moyenne constante, il est toujours prudent de ne pas se fier uniquement à l'aspect visuel. Des fluctuations résiduelles ou la présence de valeurs aberrantes pourraient masquer des irrégularités qui ne sont pas immédiatement apparentes.

Série du Chômage

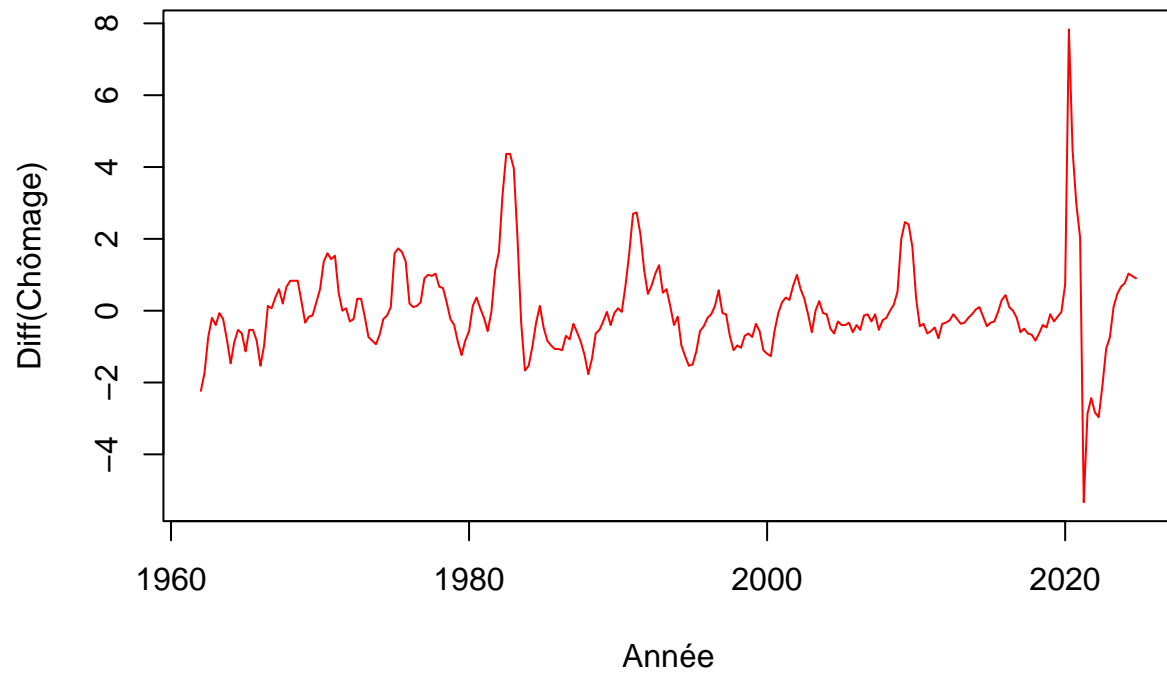
Dans cette section, nous allons appliquer la première différenciation $Y'_t = Y_t - Y_{t-1}$ sur notre série dans le but d'éliminer sa tendance linéaire.

Différenciation Ordre 1



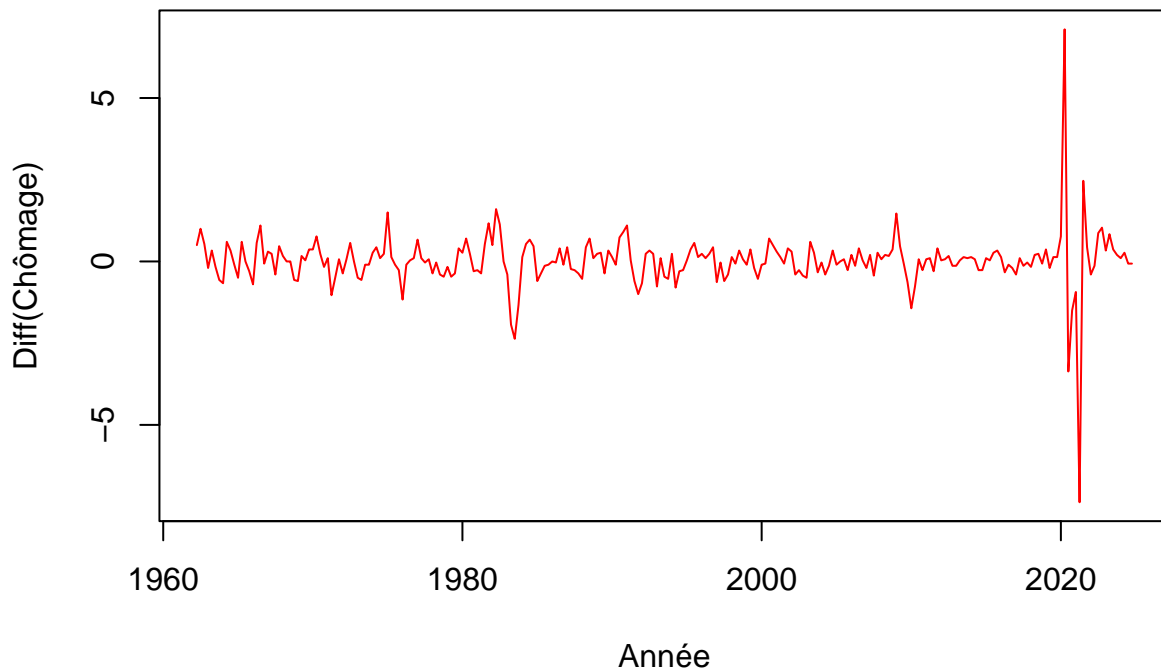
Il reste encore une composante saisonnière significative dans la série, ce qui suggère que la différenciation d'ordre 1 n'a pas suffi à éliminer complètement la saisonnalité. Comme nous l'avons fait pour la série du PIB, nous allons appliquer une différenciation d'ordre 4, car nos données sont trimestrielles. Cette approche permet de capturer et d'éliminer les variations cycliques qui se répètent.

Différenciation Ordre 4



La moyenne n'est pas encore stable, une autre différenciation sera donc nécessaire pour supprimer cette tendance et rendre notre série plus stationnaire.

Différenciation d'ordre 4 + 1



Après avoir appliqué les différenciations, nous constatons que la série oscille autour d'une moyenne relativement constante, sans vraiment de tendance apparente.

Bien que des fluctuations subsistent, elles semblent s'annuler sur des périodes plus longues, suggérant un équilibre global. Cela signifie que les variations observées suivent un mouvement autour d'un niveau moyen plutôt stable.

Analyse de la stationnarité des séries différenciées

Avant d'évaluer la stationnarité des séries différenciées, nous commençons par calculer leur moyenne et leur variance. Ces deux mesures nous donnent un premier aperçu de la stabilité de la série.

```
## Moyenne de la différenciation d'Ordre 5 du PIB : -0.0001670447
```

```
## Variance de la différenciation d'Ordre 5 du PIB : 0.0006343168
```

```
## Moyenne de la différenciation d'Ordre 5 du Chômage : 0.0124834
```

```
## Variance de la différenciation d'Ordre 5 du Chômage : 0.7419771
```

La moyenne de la différenciation du PIB est proche de zéro (-0.000167), ce qui suggère que la série ne présente pas de tendance marquée après différenciation.

La variance de la différenciation du PIB est relativement faible (0.000634), indiquant des fluctuations modérées autour de la moyenne.

La moyenne de la différenciation du taux de chômage (0.01248) est légèrement positive, ce qui pourrait indiquer un léger biais dans la tendance des variations du taux de chômage.

La variance de la différenciation du taux de chômage (0.7419) est nettement plus élevée que celle de la différenciation du PIB, ce qui signifie que les variations du taux de chômage sont plus importantes et plus irrégulières.

Pour tester la stationnarité de nos séries, nous avons déjà effectué une première étape importante : montrer que la moyenne de nos séries semble constante au fil du temps.

En complément de cela, nous allons maintenant passer au test de Box-Ljung. En effectuant ce test, nous pourrions ainsi évaluer l'hypothèse nulle selon laquelle il n'y a pas d'autocorrélation significative, ce qui confirme le fait de ne pas rejeter l'hypothèse de la stationnarité de nos séries.

Série du PIB

```
## Statistique Q : 0.1008431 | p-value : 0.939771
```

Le test de Box-Ljung donne un résultat de 0.10084 pour la statistique de ce test avec un degré de liberté de 1 et une valeur p de 0.7508. La valeur p étant bien supérieure au seuil de signification 0.05, nous ne rejetons pas l'hypothèse nulle, ce qui suggère qu'il n'y a pas d'autocorrélation significative dans les résidus.

Il est donc possible que la série du PIB soit stationnaire, car l'absence d'autocorrélation est un indicateur de stationnarité.

Série du Chômage

```
## Statistique Q : 0.1180208 | p-value : 0.9354001
```

Le test de Box-Ljung donne un résultat de 0.11802 pour la statistique de ce test avec un degré de liberté de 1 et une valeur p de 0.7312. Comme la valeur p est bien supérieure à 0.05, nous ne rejetons pas l'hypothèse nulle. Cela suggère qu'il n'y a pas d'autocorrélation significative dans les résidus.

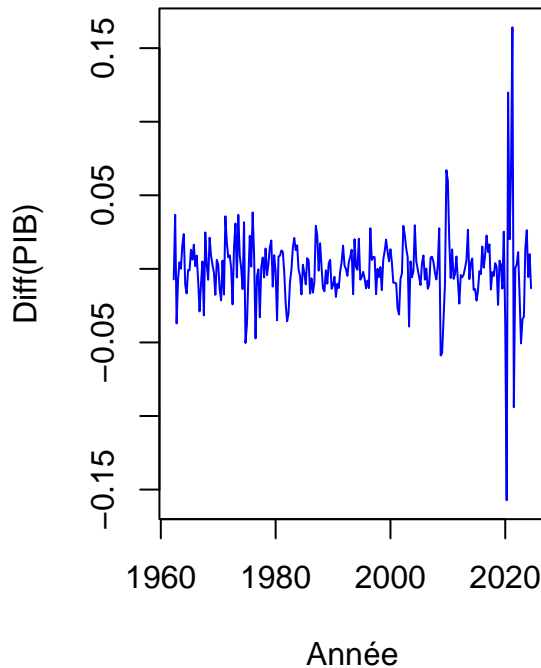
L'absence d'autocorrélation indique une possible stationnarité de la série du taux de chômage.

Stationnarité

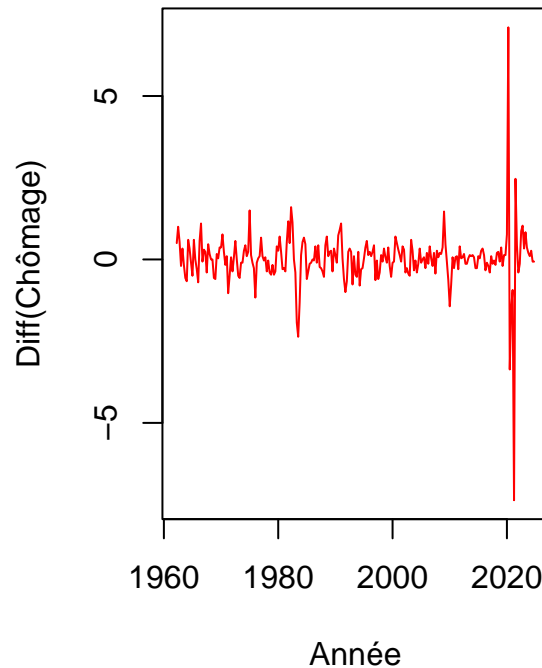
Obtention des séries de moyennes constantes

Pour les différentes méthodes appliquées à nos deux séries, l'on constate que la différenciation est jusque-là la meilleure méthode pour rendre la moyenne constante. En effet, l'on remarque qu'avec le filtrage, nos séries ne semblent pas tourner autour d'une constante, ce qui ne présagerait pas une moyenne constante. On utilise ainsi les mêmes différenciations que ci-haut.

Différenciation Ordre 4 + 1



Différenciation Ordre 4 + 1



Analyse de l'autocorrélation

Nous allons calculer les autocorrélations pour les séries différenciées, puis de tester la significativité de ces autocorrélations à travers le test de Portemanteau : le test de Box-Ljung. Nous allons commencer par calculer les autocorrélations pour différentes valeurs de décalage (lag), allant de $h = 0$ à $h = 20$, sur les séries différenciées.

Les autocorrélations à différents retards indiquent dans quelle mesure les valeurs passées influencent les valeurs actuelles de la série.

Le test de Box-Ljung est utilisé pour tester l'hypothèse nulle selon laquelle il n'y a pas d'autocorrélation significative dans une série temporelle à un ensemble de lags.

La statistique du test de Box-Ljung se calcule comme suit :

$$Q = n(n+2) \sum_{j=1}^n \frac{\hat{\rho}_j^2}{n-j}$$

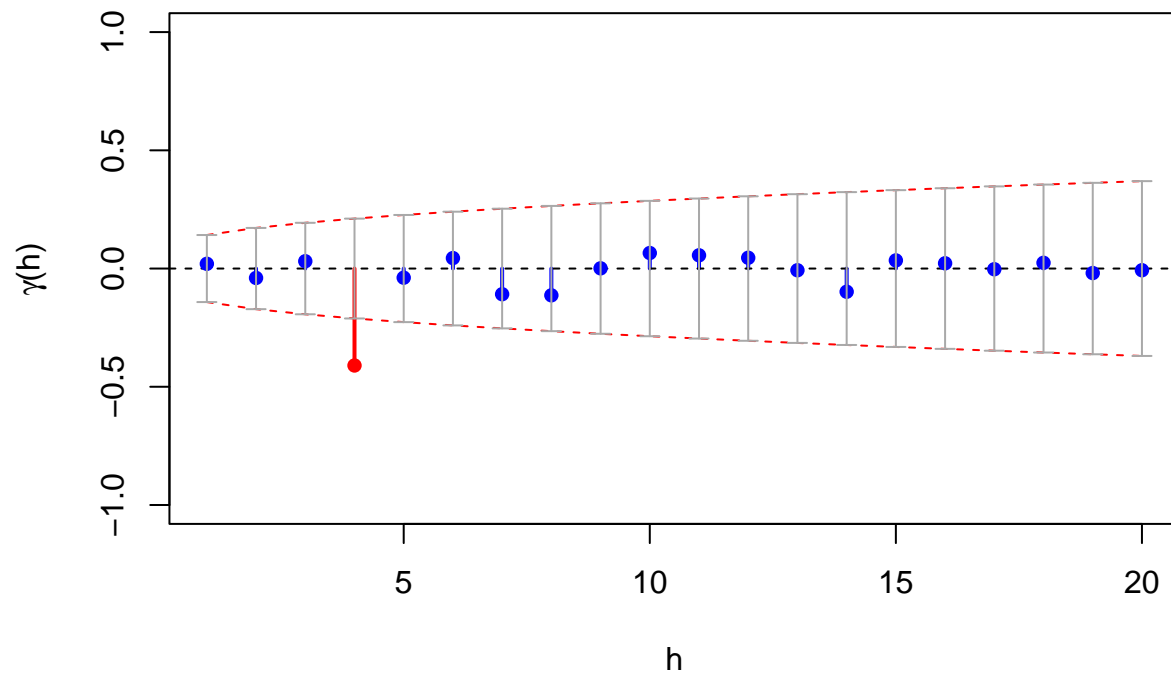
Cette statistique suit une loi du khi-deux avec h degrés de liberté

Après avoir calculé la statistique Q du test de Box-Ljung, nous allons comparer sa valeur à une loi du khi-deux avec h degrés de liberté pour évaluer la significativité de l'autocorrélation, nous allons également calculer des intervalles de confiance à 95% pour les autocorrélations.

Ces intervalles vont nous indiquer si les autocorrélations calculées sont statistiquement significatives.

Une fois que nous avons calculé les autocorrélations et les intervalles de confiance, nous éliminerons les autocorrélations qui ne se trouvent pas à l'intérieur de leurs intervalles de confiance.

Autocorrélation avec Intervalles de Confiance

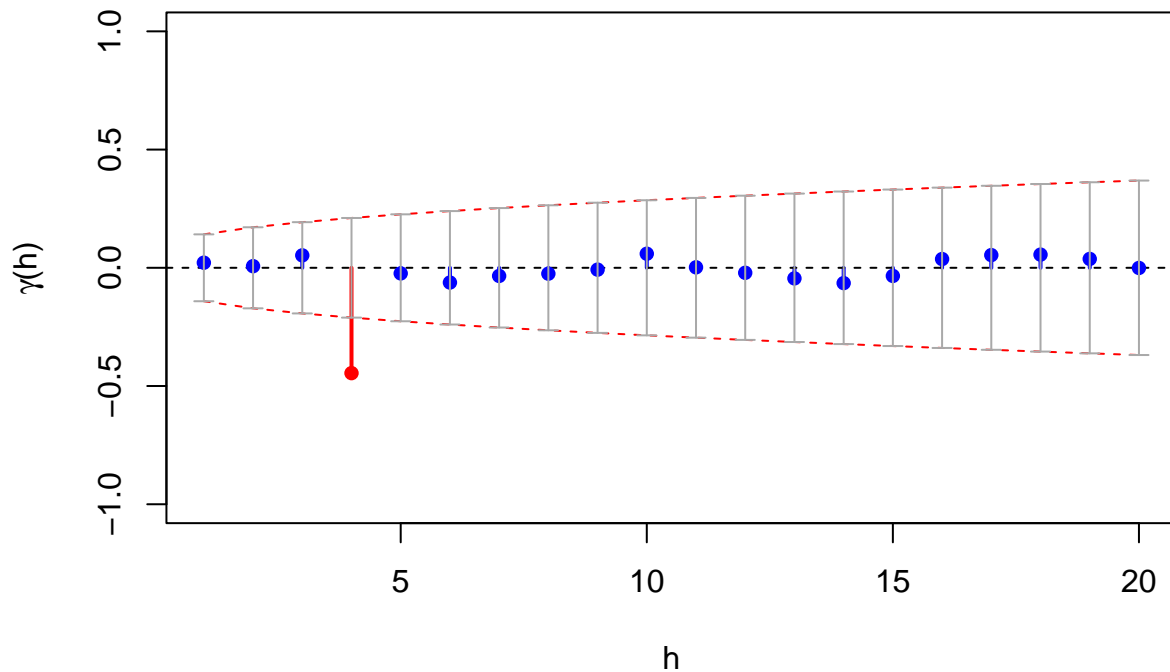


Nous avons une autocorrélation non significative à un lag de $h=4$ (environ), ce pourrait indiquer que la valeur de la série du PIB à ce lag ne présente pas de dépendance significative.

Il n'y a pas de preuve statistique suffisante pour affirmer qu'il existe une relation temporelle notable entre les valeurs du PIB à $t+4$.

Un retard de 4 périodes dans le PIB pourrait ne pas avoir d'impact significatif sur la valeur future, ce qui pourrait suggérer que l'effet de la variation du PIB se dissipe avant d'atteindre ce lag.

Autocorrélation avec Intervalles de Confiance



Le point rouge au lag $h=4$ pour le chômage suggère qu'il y a une autocorrélation significative à ce lag. Cela signifie que les valeurs passées du chômage à t influencent de manière significative les valeurs du chômage à $t+4$.

Cela pourrait indiquer que les variations du chômage persistent au-delà de 4 périodes, suggérant une dépendance temporelle sur cette période.

Conclusion

Les points rouges au lag $h=4$ pour les deux séries du PIB et du chômage indiquent que l'impact des variations passées (que ce soit dans la croissance économique ou le taux de chômage) est significatif à ce lag. Cela signifie que les variations passées influencent encore les valeurs futures après 4 périodes, et suggère une certaine forme de dépendance à ce lag.

Cette significativité au lag 4 souligne que nos séries temporelles présentent des effets persistants qui ne se dissipent pas immédiatement, ce qui peut être lié à des dynamiques économiques à moyen terme.

Il peut être pertinent de tester ces relations sur d'autres horizons temporels pour identifier des patterns spécifiques ou pour mieux comprendre la dynamique de chaque série.

Cela ouvre la porte à des approches de modélisation plus spécifiques, comme les modèles ARIMA saisonniers ou des méthodes de décomposition des séries temporelles, qui peuvent mieux capter les influences saisonnières et cycliques.

À ce propos, une question intéressante qui pourrait enrichir la réflexion serait : **Comment les événements macroéconomiques peuvent-ils perturber ou renforcer les cycles économiques saisonniers dans les séries temporelles ?**

Modélisation et Prédiction avec les Processus ARMA

Cette seconde partie approfondit l'analyse des séries du **produit intérieur brut (PIB)** et du **taux de chômage** canadien, préalablement rendues stationnaires dans la Partie 1. Notre objectif est d'explorer leur dynamique à l'aide de modèles **ARMA (Autorégressifs à Moyenne Mobile)**, d'identifier les ordres optimaux (p, q) via les fonctions ACF/PACF, et de comparer les performances des modèles avec des critères comme l'AIC et le BIC. Nous examinerons également les corrélations croisées entre ces séries pour évaluer leur interdépendance, conformément à la loi d'Okun, qui postule un lien inverse entre croissance du PIB et évolution du chômage.

Les chocs structurels, comme ceux induits par la **pandémie de COVID-19**, ont perturbé les équilibres économiques traditionnels, introduisant des ruptures de tendance et une volatilité accrue. Bien que ces événements puissent complexifier la modélisation, nous conservons l'intégralité des données pour préserver la représentativité des dynamiques à long terme. Cette approche nous permettra d'évaluer la résilience des modèles ARMA face à des perturbations exogènes, tout en respectant les exigences de stationnarité validées précédemment.

Sélection des Ordres (p, q) des modèles ARMA

Les modèles **ARMA** sont des outils fondamentaux en analyse des séries temporelles stationnaires. Ils combinent deux composantes essentielles :

La partie autoRegressive (AR) capture la relation entre une observation et ses valeurs passées. Autrement dit, elle modélise l'évolution de la série en fonction de ses propres antécédents.

$$X_t = \sum_{i=1}^p \phi_i X_{t-i} + \epsilon_t \quad \text{où} \quad \epsilon_t \sim \text{WN}(0, \sigma^2)$$

X_t est la valeur de la série à l'instant t ; ϕ_i sont les coefficients autorégressifs; ϵ_t est un bruit blanc.

La partie moyenne mobile (MA) intègre l'impact des chocs aléatoires passés, en lissant les fluctuations imprévues pour mieux comprendre la dynamique sous-jacente des données.

$$X_t = \epsilon_t + \sum_{j=1}^q \theta_j \epsilon_{t-j} \quad \text{avec} \quad \epsilon_t \sim \text{WN}(0, \sigma^2)$$

θ_j sont les coefficients de moyenne mobile.

Le modèle ARMA(p, q) combine les deux composantes :

$$X_t = \sum_{i=1}^p \phi_i X_{t-i} + \epsilon_t + \sum_{j=1}^q \theta_j \epsilon_{t-j}$$

L'un des défis majeurs réside dans le choix des ordres p et q . Une sélection inadéquate conduit à un modèle **sous-ajusté**, qui ne capture pas correctement les tendances et les dépendances temporelles ; ou **sur-ajusté**, trop sensible aux variations aléatoires, ce qui le rend peu généralisable.

Analyse de l'ACF et du PACF

Pour affiner cette sélection, nous allons analyser les fonctions d'autocorrélation (ACF) et d'autocorrélation partielle (PACF) sur différents intervalles de confiance.

La fonction d'autocorrélation (ACF) nous permet d'évaluer la corrélation entre une observation X_t et ses valeurs passées X_{t-h} , à différents retards h , sans contrôler pour les retards intermédiaires, afin de mieux comprendre la structure de la série.

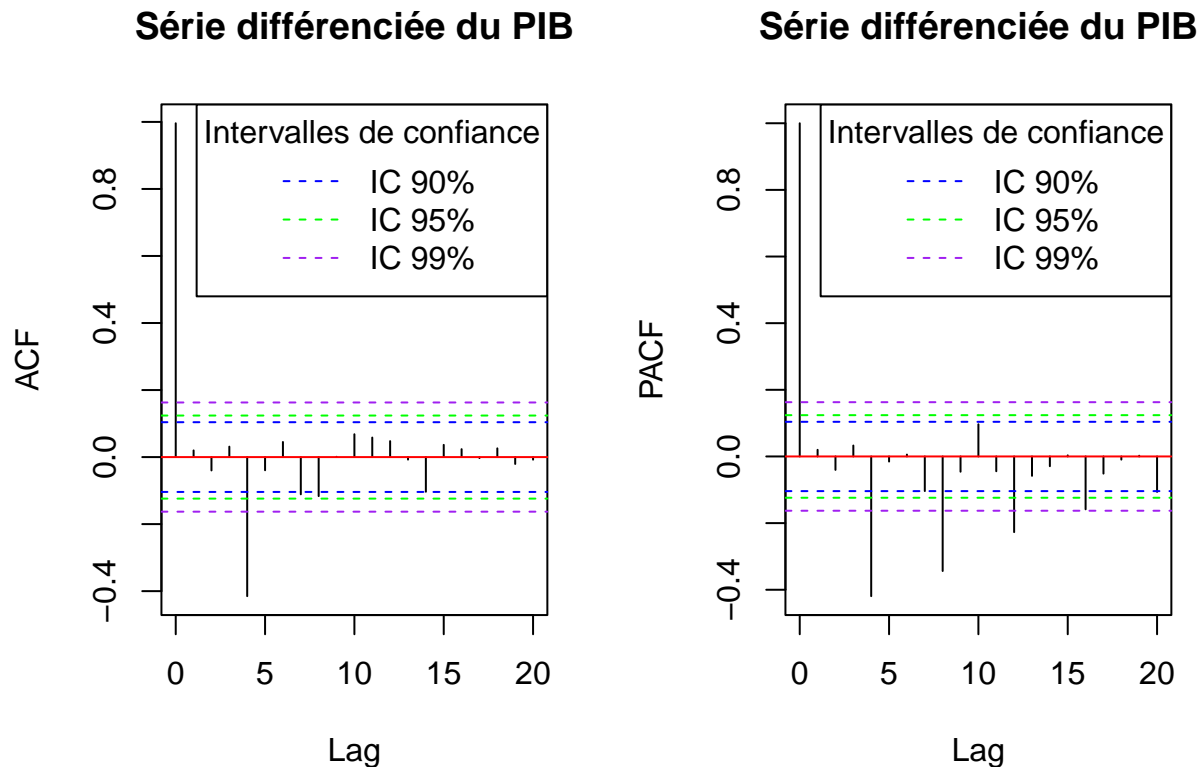
$$\rho(h) = \frac{\text{Cov}(X_t, X_{t-h})}{\text{Var}(X_t)}$$

La fonction d'autocorrélation partielle (PACF) mesure la corrélation entre X_t et X_{t-h} en contrôlant pour l'influence des retards.

$$\pi(h) = \text{Corr}(X_t, X_{t-h} | X_{t-1}, \dots, X_{t-h-1})$$

En examinant ces fonctions sur plusieurs intervalles de confiance, nous pourrions mieux distinguer les corrélations significatives des fluctuations aléatoires. Cela nous aidera à éviter un sur-ajustement du modèle en nous concentrant uniquement sur les dynamiques réellement présentes dans les données.

Série du PIB



En examinant les fonctions ACF et PACF sous trois niveaux de confiance (90%, 95%, 99%), nous pouvons identifier les ordres potentiels du modèle ARMA.

ACF

- À 90%, plusieurs lags (4, 7, 8) apparaissent comme significatifs.
- À 95%, seul le lag 4 conserve sa significativité.
- À 99%, seul le lag 4 est significatif.

L'importance du lag 4, même à un seuil strict de 99%, indique que la composante MA(4) est un choix solide.

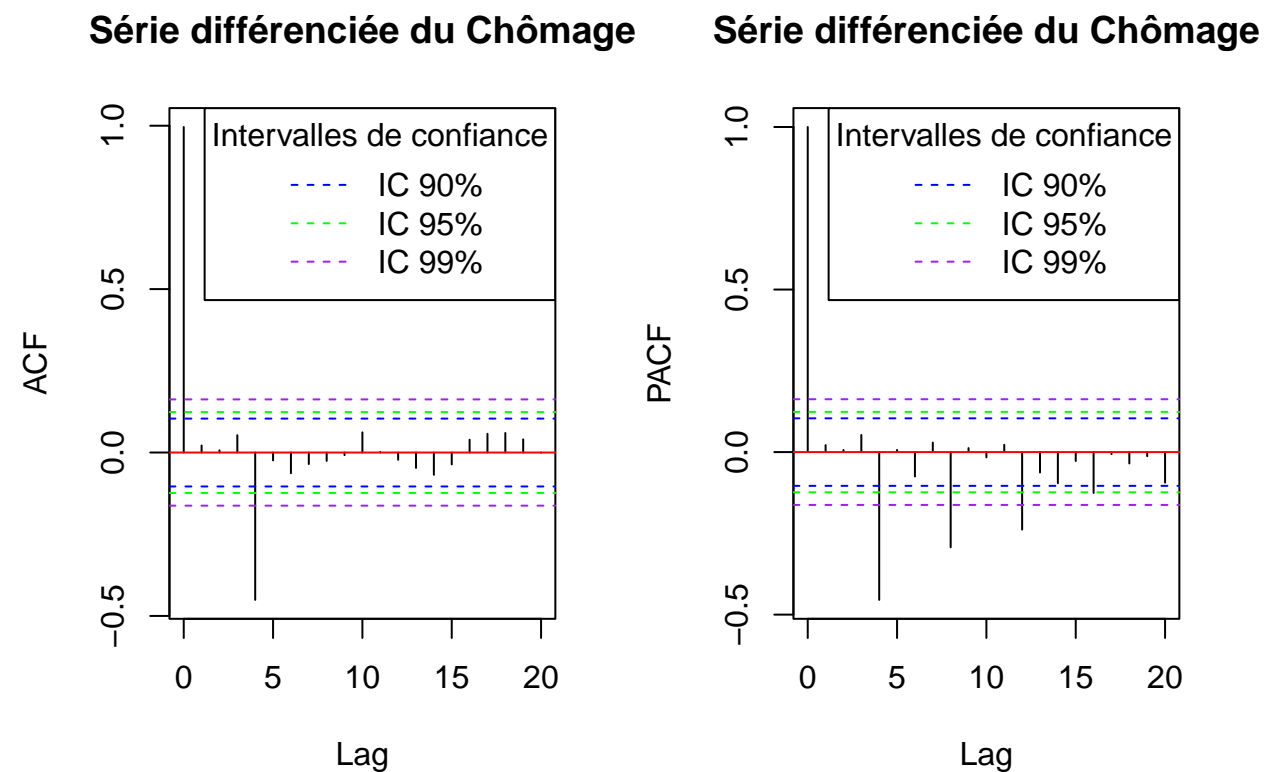
PACF

- À 90%, les lags 4, 8, 12, 16 sont significatifs.
- À 95%, la même tendance se maintient, mais la confiance diminue pour les lags les plus élevés.
- À 99%, seuls les lags 4, 8 et 12 restent significatifs.

La répétition de la significativité à des multiples de 4 renforce l'idée d'un processus AR(4).

L'analyse des corrélations temporelles indique qu'un modèle **ARMA(4,4)** est le plus approprié pour notre série du PIB.

Série du Chômage



De la même manière, nous analysons les ACF et PACF sous différents niveaux de confiance pour estimer les ordres optimaux du modèle ARMA.

ACF

- À 90%, seul le lag 4 est significatif.

- À 95%, la tendance reste la même : lag 4 significatif.
- À 99%, le même lag 4 est significatif.

La stabilité de ce lag à tous les niveaux de confiance confirme qu'un modèle MA(4) est un bon candidat.

PACF

- À 90%, les lags 4, 8, 12, 16 sont significatifs.
- À 95%, seuls les lags 4, 8 et 12 restent significatifs, le lag 16 n'étant plus pertinent.
- À 99%, cette tendance persiste, mais avec une perte de confiance progressive sur les lags les plus élevés.

La significativité répétée aux multiples de 4 renforce l'hypothèse d'un processus AR(4).

Comme pour le PIB, la structure des autocorrélations suggère un modèle ARMA(4,4), intégrant un processus AR(4) et un MA(4).

Estimation et comparaison des modèles ARMA

Dans notre démarche d'analyse des séries temporelles, nous passons maintenant à l'estimation du modèle ARMA afin de mieux comprendre la dynamique de notre série et d'obtenir des prévisions fiables.

Après avoir analysé les fonctions d'autocorrélation (ACF) et d'autocorrélation partielle (PACF), nous avons pu identifier des valeurs pertinentes pour les paramètres du modèle p et q.

Nous commençons par diviser nos séries stationnaires en deux ensembles : un ensemble d'entraînement et un ensemble de test. En général, nous réservons environ 80 % des observations pour l'entraînement, tandis que les 20 % restants servent à évaluer la performance du modèle. Cette séparation nous permet de construire un modèle robuste sur l'ensemble d'entraînement et de tester sa capacité prédictive sur l'ensemble de test.

Estimation par Maximum de Vraisemblance

L'estimation du modèle ARMA repose sur la méthode du maximum de vraisemblance, qui optimise les paramètres afin de maximiser la probabilité d'observer nos données. Une fois le modèle ajusté, nous pourrions vérifier sa pertinence et affiner notre approche si nécessaire, en fonction des performances observées sur l'ensemble de test.

Grâce à cette approche rigoureuse, nous nous assurons que notre modèle capte au mieux les tendances de la série temporelle et qu'il est capable de fournir des prévisions fiables.

Série du PIB

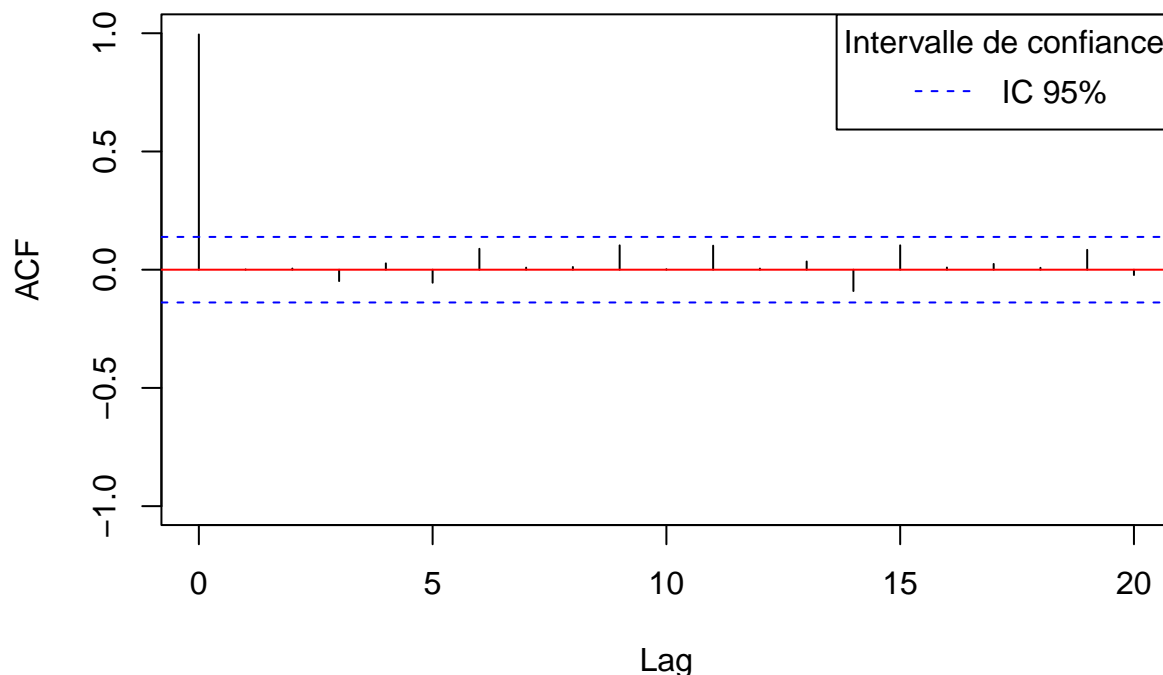
Tous les coefficients du modèle ARMA(4,4) sont significatifs, sauf AR(4), dont le coefficient $\phi_4 = -0.1686$ est de faible amplitude, ce qui suggère qu'il pourrait être retiré pour éviter le surajustement. La variance résiduelle $\sigma^2 = 0.0002358$ est très faible, indiquant que le modèle capture bien la variabilité des données.

Les critères AIC (-1078.2) et BIC (-1045.219) montrent que le modèle est performant, mais le BIC plus élevé que l'AIC suggère une possible surcomplexité par rapport à la taille de l'échantillon.

En conclusion, bien que le modèle soit efficace, la non-significativité d'AR(4) justifie l'exploration de modèles plus simples.

L'analyse des résidus est essentielle pour valider la qualité de notre modèle ARMA. Elle permet de vérifier si nous avons correctement capturé les dynamiques temporelles, laissant des résidus ressemblant à un bruit blanc, c'est-à-dire sans autocorrélation ni structure discernable.

ACF des Résidus du Modèle ARMA(4,4) – PIB



L'ACF montre qu'aucun pic ne dépasse les intervalles de confiance. Cela indique que notre modèle ARMA(4,4) a bien absorbé les dépendances temporelles, et aucune structure non modélisée n'est présente.

Test de Ljung-Box

Le test de Ljung-Box donne une statistique $Q = 4.87$ avec une p-value de 0.780, bien au-dessus du seuil de 5%. Cela confirme que les résidus sont indépendants, validant ainsi la performance de notre modèle.

Test des Points Tournants

Le test des points de retournement montre 135 points avec une p-value de 0.613, bien supérieure à 0.05. Cela indique que nos résidus oscillent de manière aléatoire, sans tendance ni fluctuations anormales.

Les diagnostics suggèrent que notre modèle ARMA(4,4) est valide. Les résidus se comportent comme un bruit blanc, sans autocorrélation ni motifs résiduels. Les termes AR(1-3) et MA(1-4) ont efficacement capturé les dépendances. Toutefois, il serait utile de comparer ce modèle avec des modèles plus simples afin d'éviter tout surajustement.

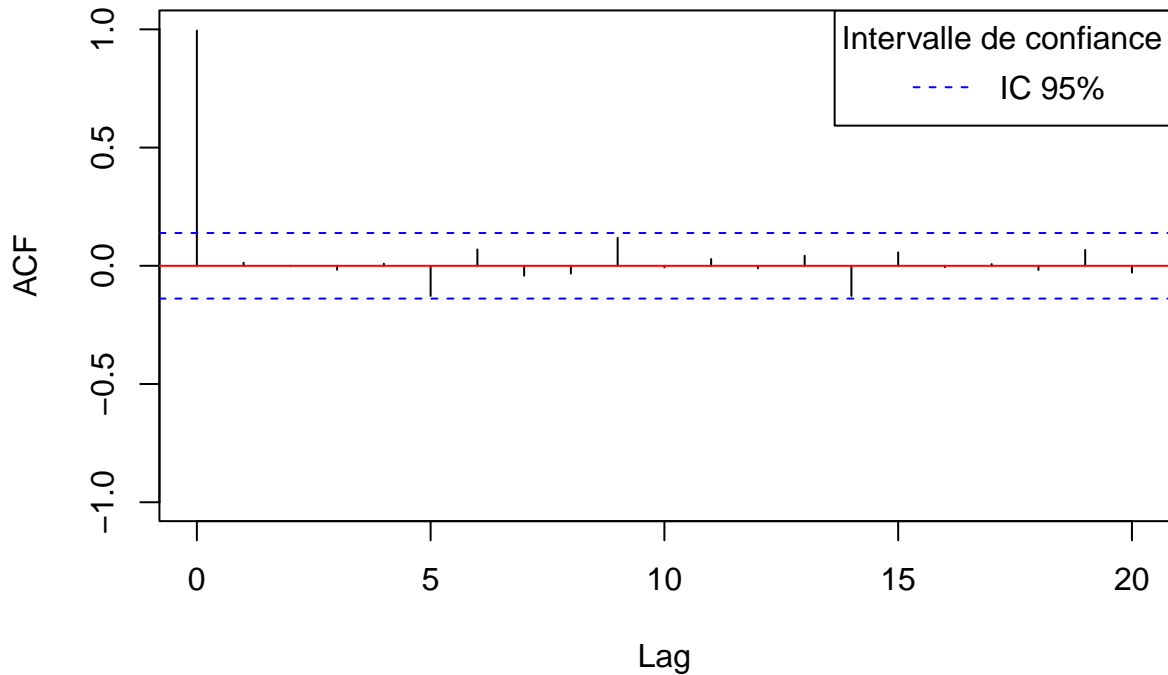
Pour alléger le modèle tout en conservant ses performances, nous estimons une version simplifiée **ARMA(3,4)**

Le modèle simplifié présente une légère amélioration en termes de critère d'information :

AIC = -1078.36 contre -1078.2 pour ARMA(4,4), indiquant une qualité d'ajustement quasi-identique.

BIC = -1048.676 contre -1045.219, ce qui montre une amélioration significative en faveur du modèle ARMA(3,4), en raison de la réduction du nombre de paramètres.

ACF des Résidus du Modèle ARMA(3,4) – PIB



Aucun pic significatif, ce qui indique que les dépendances temporelles ont bien été prises en compte par les termes AR(1-3) et MA(1-4).

Test de Ljung-Box

La statistique Q obtenue est de 36.91, avec une p-value de 0.939, indiquant que l'hypothèse d'indépendance des résidus jusqu'au lag 20 ne peut être rejetée. Cela valide ainsi l'hypothèse de bruit blanc.

Test des Points Tournants

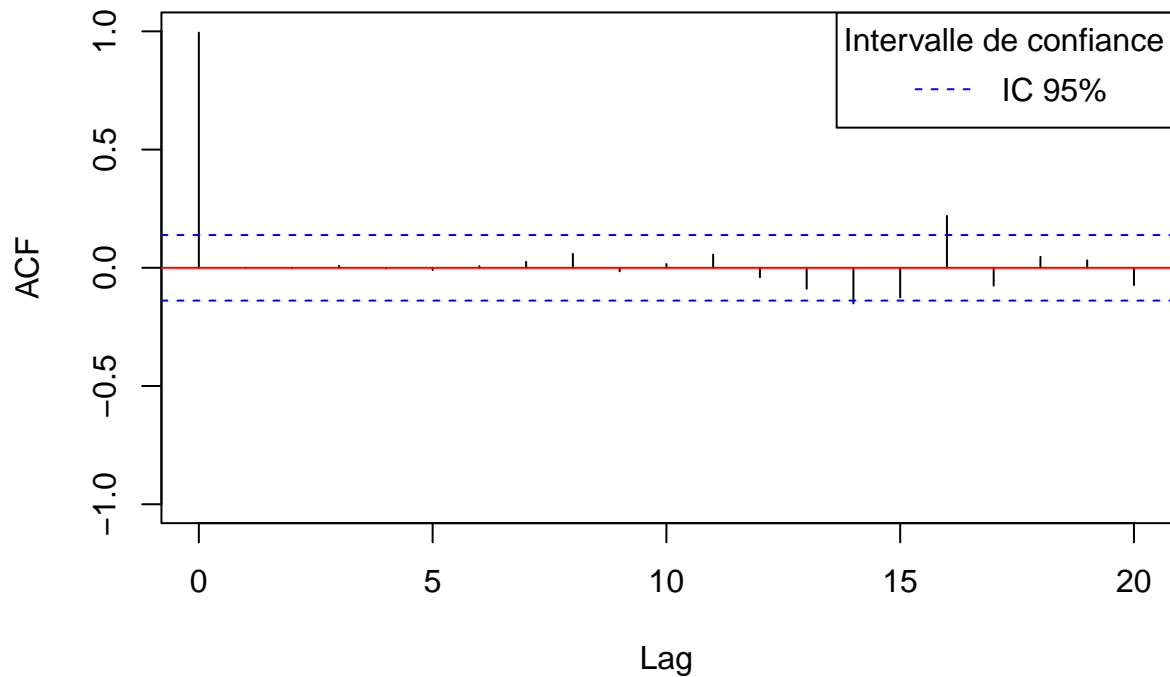
Le test des points de retournement a donné 127 occurrences, avec un z-score de -0.84 et une p-value de 0.40. Ces résultats indiquent que les variations des résidus sont aléatoires, sans tendance ni cyclicité significative.

En résumé, l'ARMA(3,4) semble un meilleur choix car il offre une modélisation plus efficace et plus robuste en conservant la qualité d'ajustement tout en réduisant le risque de surajustement grâce à un modèle plus parcimonieux.

Série du Chômage

Avec une AIC de 235.62 et une BIC de 268.60, ces valeurs relativement élevées suggèrent que le modèle pourrait être trop complexe pour les données. De plus, la variance des résidus $\sigma^2 = 0.1682$ est relativement élevée par rapport aux modèles utilisés pour le PIB, ce qui indique un ajustement moins précis et suggère une capacité d'explication limitée du modèle. En ce qui concerne la significativité des coefficients, la composante MA présente un coefficient relativement élevé en valeur absolue avec $\theta_4 = -0.372$, suggérant une contribution importante des chocs passés à l'évolution de la série. De même, $\theta_2 = -0.290$ semble également significatif. Pour la composante AR, les coefficients $\phi_1 = 0.468$ et $\phi_4 = -0.1686$ indiquant une influence notable des valeurs passées sur la dynamique actuelle de la série. Compte tenu de la complexité du modèle et des performances comparativement plus faibles, le modèle ARMA(4,4) ne semble pas être le plus adapté pour représenter la série. Il pourrait être avantageux d'envisager un modèle plus parcimonieux offrant un meilleur compromis entre ajustement et simplicité.

ACF des Résidus du Modèle ARMA(4,4) – Chômage



L'ACF des résidus présente un pic significatif au lag 16, cela suggère qu'il pourrait s'agir d'un bruit aléatoire.

Test de Ljung-Box

Le test de Ljung-Box, avec une statistique Q de 22.24 et une p-value de 0.410, montre globalement que les résidus sont indépendants, bien qu'il masque le problème spécifique au lag 16.

Test des Points Tournants

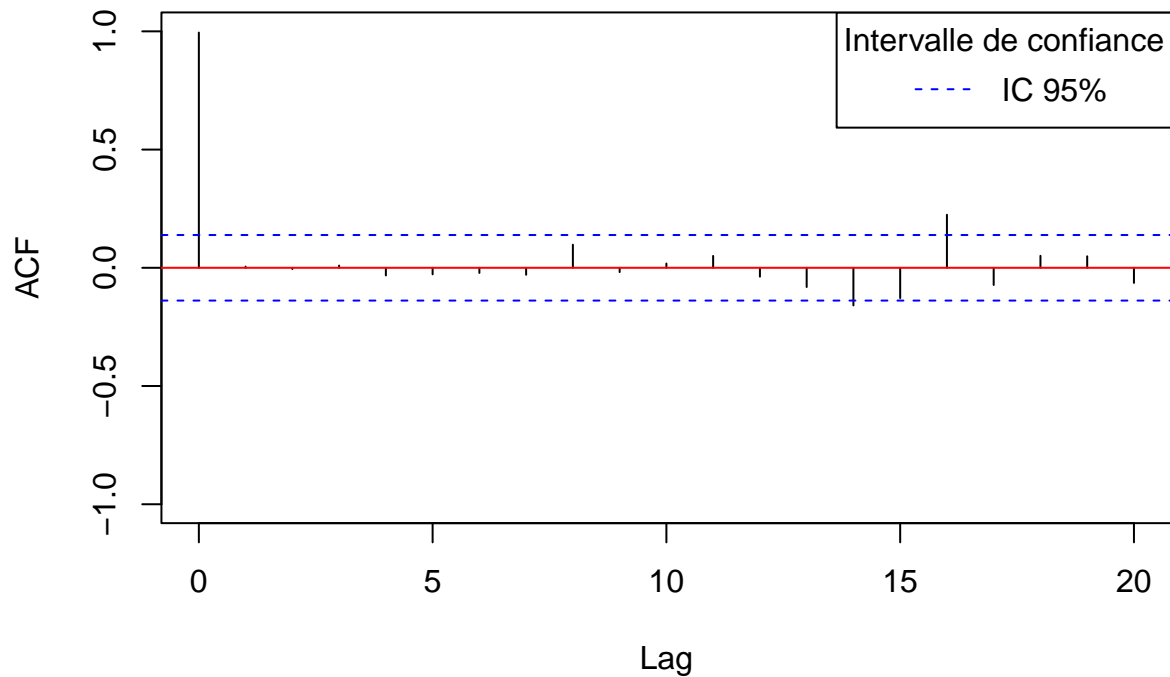
Le test des points de retournement, avec 133 points de retournement et une p-value de 0.866, indique un comportement globalement aléatoire des résidus.

Les problèmes identifiés dans ce modèle incluent un surajustement, dû au nombre excessif de paramètres non significatifs, et une saisonnalité résiduelle non modélisée, visible dans le pic au lag 16. De plus, la variance résiduelle élevée suggère que le modèle a un faible pouvoir explicatif pour les données étudiées.

Afin d'améliorer l'ajustement, nous allons maintenant tester un modèle ARMA(1,4), qui pourrait mieux réduire la complexité du modèle.

Le modèle ARMA(1,4) démontre une amélioration significative par rapport aux versions précédentes. En termes de critères de sélection de modèle, l'AIC atteint 230.78, une réduction notable par rapport à 235.62 pour le modèle ARMA(4,4). Le BIC est également plus faible à 253.87 contre 268.60 pour ARMA(4,4), indiquant un meilleur compromis entre ajustement et complexité.

ACF des Résidus du Modèle ARMA(1,4) – Chômage



Le diagnostic des résidus met en évidence deux points d'amélioration. D'une part, bien que les pics significatifs aient disparu pour la majorité des lags, un pic subsiste encore au lag 16, qui semble être du bruit et peut donc être négligé.

Test de Ljung-Box

D'autre part, le test de Ljung-Box donne une statistique Q de 24.69, avec une p-value de 0.507, ce qui indique une absence d'autocorrélation significative dans les résidus à un niveau global.

Test des Points Tournants

En outre, le test des points de retournement révèle que le comportement est globalement aléatoire, avec une p-value de 0.613, ce qui confirme l'absence de tendances systématiques dans les données.

Il est possible que le modèle ARMA(1,4) soit une option adéquate pour modéliser la série, étant donné sa simplicité et son ajustement relativement satisfaisant, il pourrait constituer un choix raisonnable.

Les résultats des modèles ARMA pour les séries de PIB et de chômage révèlent des caractéristiques distinctes mais intéressantes.

Analyse des Paramètres et Modèles Imbriqués

Série du PIB

```
## Parameter Estimate_PIB_ARMA(4,4) StdError_PIB_ARMA(4,4) Estimate_PIB_ARMA(3,4)
##      ar1              0.6334              0.1438              0.5168
##      ar2             -0.4836              0.1590             -0.3731
##      ar3              0.5126              0.1558              0.2475
##      ar4             -0.1686              0.1127              NA
## intercept            -0.0004              0.0001             -0.0003
##      ma1             -0.5169              0.1371             -0.4088
##      ma2              0.3889              0.1464              0.3115
##      ma3             -0.5231              0.1436             -0.2892
##      ma4             -0.3489              0.1211             -0.4621
## StdError_PIB_ARMA(3,4)
##              0.1559
##              0.1429
##              0.1270
##              NA
##              0.0003
##              0.1408
##              0.1226
##              0.1134
##              0.0712
```

Nous avons estimé deux modèles ARMA pour la série du PIB : un modèle ARMA(4,4) et un modèle ARMA(3,4). Le modèle ARMA(4,4) comporte un paramètre autorégressif supplémentaire par rapport au modèle ARMA(3,4), ce qui lui permet de capturer des dynamiques temporelles plus complexes. Cependant, cette augmentation du nombre de paramètres peut aussi introduire un risque de surajustement, rendant le modèle moins généralisable aux nouvelles données.

L'examen des paramètres estimés montre que, dans le modèle ARMA(4,4), le coefficient AR(1) de 0.6334, avec un écart-type de 0.1438, indique une persistance significative des chocs, tandis que le AR(2) de -0.4836 (écart-type: 0.1590) suggère un effet correctif. À l'inverse, le modèle ARMA(3,4) présente un AR(1) plus modéré à 0.5168 (écart-type: 0.1559), avec un AR(3) moins marqué à 0.2475 (écart-type: 0.1270).

La composante moyenne mobile montre également des écarts importants. Le terme MA(1) du ARMA(4,4) atteint -0.5169 (écart-type: 0.1371), alors que dans le modèle ARMA(3,4), le MA(4) se distingue par une estimation plus précise de -0.4621 avec un écart-type relativement faible de 0.0712. L'absence du terme AR(4) dans le modèle simplifié ne semble pas compromettre sa performance, comme en témoignent les écarts-types généralement comparables ou inférieurs à ceux du modèle complet.

En résumé, bien que le modèle ARMA(4,4) permette une meilleure prise en compte des dynamiques temporelles, le modèle ARMA(3,4) semble offrir un meilleur compromis entre ajustement et simplicité.

Série du Chômage

```
## Parameter Estimate_Unemp_ARMA(4,4) StdError_Unemp_ARMA(4,4)
##      ar1              0.4680              0.2385
##      ar2              0.0837              0.2128
##      ar3              0.1205              0.2174
##      ar4             -0.1713              0.1662
## intercept            0.0006              0.0028
##      ma1             -0.0841              0.2358
##      ma2             -0.2896              0.1580
##      ma3             -0.2544              0.2385
##      ma4             -0.3720              0.1150
## Estimate_Unemp_ARMA(1,4) StdError_Unemp_ARMA(1,4)
##              0.5801              0.0821
##              NA              NA
##              NA              NA
##              NA              NA
##              0.0008              0.0032
##             -0.2026              0.0878
##             -0.2418              0.0641
##             -0.0978              0.0744
##             -0.4578              0.0680
```

Pour la série du chômage, nous avons estimé un modèle ARMA(4,4) et un modèle ARMA(1,4).

Nous observons que, le coefficient AR(1) apparaît significatif dans les deux spécifications, bien qu'avec une intensité variable : 0,5801 (écart-type : 0,0821) pour le modèle ARMA(1,4) contre 0,4680 (écart-type : 0,2385) pour le ARMA(4,4). Cette différence suggère que la persistance des chocs sur le chômage est mieux capturée par le modèle plus parcimonieux.

Les termes autorégressifs supplémentaires du modèle ARMA(4,4) présentent des résultats mitigés. Les coefficients AR(2) à AR(4), compris entre -0,1713 et 0,1205, montrent des écarts-types élevés (de 0,1662 à 0,2385) par rapport à leurs valeurs estimées, indiquant une faible significativité statistique. Cette observation remet en question l'utilité des retards supplémentaires dans la modélisation du chômage.

La composante moyenne mobile offre des enseignements plus contrastés. Le coefficient MA(4) se révèle particulièrement robuste dans les deux spécifications : -0,4578 (écart-type : 0,0680) pour ARMA(1,4) contre -0,3720 (écart-type : 0,1150) pour ARMA(4,4). Cependant, les termes MA(1) à MA(3) du modèle complet présentent des incertitudes importantes, avec des écarts-types représentant 50% à 100% de la valeur estimée.

Ainsi, pour la série du chômage, nous privilégierons le modèle ARMA(1,4), qui présente des performances comparables tout en étant plus simple et plus interprétable.

Conclusion

Nos résultats mettent en évidence l'importance d'un compromis entre précision et parcimonie dans le choix d'un modèle ARMA. Pour la série du PIB, bien que le modèle ARMA(4,4) puisse sembler plus précis, son coût en termes de complexité n'est pas forcément justifié. Le modèle ARMA(3,4) apparaît ainsi comme une meilleure alternative.

En revanche, pour la série du chômage, le modèle ARMA(1,4) se révèle aussi performant que le modèle ARMA(4,4) tout en utilisant moins de paramètres. Il constitue donc un choix plus robuste et plus fiable pour la prévision.

Finalement, notre analyse confirme que l'ajout de complexité dans un modèle ne se traduit pas toujours par une meilleure performance. Il est essentiel d'évaluer la stabilité des paramètres estimés ainsi que les critères d'information pour sélectionner le modèle le plus pertinent.

Validation Diagnostique et Performance des Modèles

Afin de comparer la performance des modèles ARMA(3,4) et ARMA(1,4) appliqués aux séries du PIB et du chômage, nous calculons les critères d'information d'Akaike (AIC) et de Schwarz (BIC) sur deux ensembles de données distincts : l'échantillon d'entraînement et l'échantillon de test.

L'AIC et le BIC sont des critères couramment utilisés pour évaluer l'ajustement d'un modèle tout en pénalisant sa complexité. Un AIC plus faible indique un meilleur compromis entre la qualité de l'ajustement et le nombre de paramètres, tandis que le BIC impose une pénalisation plus forte pour éviter le surajustement.

En comparant ces valeurs sur les deux échantillons, nous pourrions déterminer quel modèle offre la meilleure capacité de généralisation. Un modèle performant devrait présenter des valeurs d'AIC et de BIC faibles sur l'ensemble d'entraînement, mais aussi éviter une forte dégradation sur l'ensemble de test.

Nous analysons donc ces résultats afin d'identifier le modèle le plus adapté aux données et d'assurer un bon équilibre entre précision et complexité.

Critères d'Information (AIC/BIC)

Série du PIB

Les résultats montrent que les deux modèles ARMA(4,4) et ARMA(3,4) obtiennent des valeurs d'AIC et de BIC similaires sur l'échantillon d'entraînement, avec un léger avantage pour l'ARMA(3,4) qui affiche un AIC de -1078.36 contre -1078.20 pour l'ARMA(4,4). De plus, son BIC est plus bas (-1048.68 contre -1045.22), indiquant qu'il parvient à obtenir un bon ajustement avec moins de paramètres.

Lorsqu'on examine la performance sur l'échantillon de test, nous observons que l'ARMA(3,4) continue à mieux se comporter, avec un AIC de -156.89 et un BIC de -141.59, contre -154.83 et -137.62 pour l'ARMA(4,4). Ces résultats confirment que l'ARMA(3,4) est légèrement plus performant, offrant un meilleur compromis entre ajustement et complexité.

Série du Chômage

Concernant la série du chômage, nous constatons que le modèle ARMA(1,4) surpasse l'ARMA(4,4) sur les deux échantillons. En entraînement, il présente un AIC de 230.78 et un BIC de 253.87, inférieurs à ceux de l'ARMA(4,4) (AIC = 235.62, BIC = 268.60), ce qui indique un meilleur ajustement avec une complexité réduite.

Cette tendance se confirme sur l'échantillon de test, où l'ARMA(1,4) affiche un AIC de 204.54 contre 210.55 pour l'ARMA(4,4), et un BIC de 216.13 contre 227.93. Ces résultats suggèrent que l'ARMA(1,4) offre une meilleure capacité de généralisation, tout en évitant le risque de surajustement.

Conclusion

Les résultats des critères d'information nous permettent de tirer des conclusions claires sur le choix des modèles. Pour la série du PIB, bien que les deux modèles obtiennent des performances proches, l'ARMA(3,4) se distingue légèrement par des valeurs d'AIC et de BIC plus faibles, ce qui en fait une meilleure option.

Pour la série du chômage, l'ARMA(1,4) est clairement plus performant, avec de meilleures valeurs d'AIC et de BIC sur les deux échantillons. Son ajustement est presque aussi bon que celui de l'ARMA(4,4), tout en étant plus simple et plus robuste face aux données hors échantillon.

En conclusion, nous retenons l'ARMA(3,4) pour le PIB et l'ARMA(1,4) pour le chômage, car ces modèles offrent le meilleur équilibre entre précision et complexité, garantissant une meilleure interprétation et une meilleure capacité prédictive.

Test du Ratio de Vraisemblance

Pour comparer les modèles estimés, nous utilisons le test du ratio de vraisemblance, qui permet d'évaluer si l'ajout de paramètres supplémentaires améliore significativement l'ajustement du modèle. Ce test repose sur la comparaison des log-vraisemblances des modèles imbriqués et mesure si la complexité accrue d'un modèle est justifiée par une amélioration statistiquement significative de la qualité de l'ajustement.

Nous calculons ainsi la statistique du ratio de vraisemblance

$$LR = -2 \left(\log L_{H_0} - \log L_{H_1} \right)$$

$$LR \sim \chi_{df}^2, \quad \text{avec } df = k_{H_1} - k_{H_0}$$

Série du PIB

La statistique du ratio de vraisemblance entre l'ARMA(4,4) et l'ARMA(3,4) est de 1.84, avec une valeur-p de 0.175. Cette valeur-p étant relativement élevée, elle ne permet pas de rejeter l'hypothèse nulle selon laquelle le modèle ARMA(3,4) explique aussi bien les données que l'ARMA(4,4). Autrement dit, l'ajout d'un paramètre supplémentaire dans l'ARMA(4,4) ne conduit pas à une amélioration significative de l'ajustement.

Puisque l'ARMA(3,4) présente déjà de meilleures valeurs d'AIC et de BIC et que le test du ratio de vraisemblance ne justifie pas la complexité supplémentaire du modèle ARMA(4,4), nous concluons que l'ARMA(3,4) est le modèle le plus approprié pour modéliser la série du PIB.

Série du Chômage

Pour la série du chômage, la statistique du ratio de vraisemblance entre l'ARMA(4,4) et l'ARMA(1,4) est de 1.16, avec une valeur-p de 0.762. Cette valeur-p très élevée indique clairement que le modèle ARMA(4,4) n'apporte aucune amélioration significative par rapport à l'ARMA(1,4). Cela signifie que l'ARMA(1,4) est suffisant pour capturer la dynamique de la série du chômage et que l'ajout de paramètres supplémentaires dans l'ARMA(4,4) n'est pas justifié.

Cette conclusion est cohérente avec nos résultats précédents basés sur l'AIC et le BIC, qui montraient que l'ARMA(1,4) était le modèle le plus performant pour cette série.

Conclusion

La cohérence de ces résultats avec les analyses précédentes par critères d'information est remarquable. Pour les deux séries, les tests du rapport de vraisemblance confirment la suffisance des modèles réduits, validant ainsi le principe de parcimonie.

Pour le PIB, l'ARMA(3,4) est préféré à l'ARMA(4,4) car la différence d'ajustement n'est pas statistiquement significative. De même, pour le chômage, l'ARMA(1,4) est clairement le modèle le plus adapté, car l'ARMA(4,4) n'apporte aucune amélioration substantielle.

Ainsi, en tenant compte à la fois des critères d'information et du test du ratio de vraisemblance, nous retenons l'ARMA(3,4) pour le PIB et l'ARMA(1,4) pour le chômage comme les modèles les plus appropriés.

Erreurs Quadratiques Moyennes (EQM)

Dans cette section, nous évaluons la performance des modèles ARMA en calculant les valeurs ajustées sur l'échantillon d'entraînement et en faisant des prédictions sur l'échantillon de test. Pour cela, nous utilisons la fonction de prévision pour estimer les valeurs hors échantillon et comparons ces prédictions aux valeurs réelles sur l'échantillon de test.

L'évaluation se fait à l'aide de l'erreur quadratique moyenne (RMSE), qui mesure la différence entre les valeurs observées et les valeurs prédites. Cette métrique permet d'apprécier la précision des prévisions de chaque modèle, plus l'erreur quadratique moyenne est faible, plus le modèle est performant.

En comparant les RMSE obtenus pour chaque modèle, nous pourrions déterminer quel modèle fournit les meilleures prévisions hors échantillon et, par conséquent, quel modèle est le plus approprié pour chacune des séries (PIB et chômage). Cette analyse nous permettra de mieux comprendre la capacité de généralisation de chaque modèle et de faire un choix éclairé entre les différentes spécifications testées.

Série du PIB

Pour la série du PIB, nous observons les résultats suivants :

Modèle ARMA(4,4) : L'erreur quadratique moyenne (MSE) sur l'échantillon d'entraînement est de 0.0002358, et sur l'échantillon de test, elle est de 0.0018460.

Modèle ARMA(3,4) : L'erreur quadratique moyenne sur l'échantillon d'entraînement est de 0.0002415, et sur l'échantillon de test, elle est de 0.0018441.

Nous remarquons que les MSE pour les modèles ARMA(4,4) et ARMA(3,4) sont très proches l'un de l'autre, aussi bien sur l'échantillon d'entraînement que sur l'échantillon de test. Cela signifie que les deux modèles offrent un ajustement similaire aux données d'entraînement et une capacité de prévision comparable sur l'échantillon de test. Toutefois, le modèle ARMA(4,4) présente légèrement une meilleure performance avec un MSE d'entraînement légèrement plus bas. Cependant, les différences entre les deux modèles sont très petites, ce qui indique qu'aucun des deux modèles n'est de manière décisive meilleur que l'autre pour cette série.

Série du Chômage

Pour la série du chômage, les résultats sont les suivants :

Modèle ARMA(4,4) : L'erreur quadratique moyenne sur l'échantillon d'entraînement est de 0.1682249, et sur l'échantillon de test, elle est de 2.553193.

Modèle ARMA(1,4) : L'erreur quadratique moyenne sur l'échantillon d'entraînement est de 0.1694584, et sur l'échantillon de test, elle est de 2.552905.

Les MSE pour la série du chômage sont bien plus élevés que ceux du PIB, ce qui est attendu en raison de la nature plus volatile et plus complexe de la série du chômage. Il est intéressant de noter que les différences de MSE entre l'échantillon d'entraînement et celui de test sont beaucoup plus marquées pour le chômage, indiquant que les modèles peuvent avoir plus de difficulté à prédire avec précision les données hors échantillon. Cependant, les deux modèles ARMA(4,4) et ARMA(1,4) offrent des performances presque identiques, avec des MSE de test très proches.

Conclusion

En résumé, pour la série du PIB, les deux modèles ARMA(4,4) et ARMA(3,4) présentent des performances similaires, tant sur l'échantillon d'entraînement que sur l'échantillon de test. Il n'y a pas de différence substantielle en termes de MSE, ce qui nous amène à privilégier le modèle le plus simple, l'ARMA(3,4), sur la base de la parcimonie et des critères d'information (AIC, BIC).

Pour la série du chômage, bien que les deux modèles aient des MSE presque identiques, la capacité de prévision de ces modèles est relativement faible par rapport à celle du PIB, ce qui reflète la complexité de la

série. Les résultats suggèrent que l'ARMA(1,4) est suffisant pour capturer la dynamique de cette série, tout en évitant le surajustement.

Dans l'ensemble, ces résultats confirment que l'ARMA(3,4) est le modèle le plus approprié pour le PIB et l'ARMA(1,4) pour le chômage.

Recommandation finale des modèles ARMA

L'analyse approfondie réalisée à travers divers critères d'évaluation permet de formuler des recommandations claires sur les spécifications ARMA les plus appropriées pour chaque série économique.

Série du PIB

Pour la série du PIB, l'ensemble des résultats converge vers le modèle ARMA(3,4) comme étant le choix optimal. Ce modèle se distingue par une performance prédictive légèrement supérieure en échantillon de test, avec une erreur quadratique moyenne de 0.001844 contre 0.001846 pour le modèle ARMA(4,4). De plus, les critères d'information (AIC et BIC) sont plus favorables pour l'ARMA(3,4), et ce modèle présente l'avantage d'une parcimonie accrue, avec un paramètre AR en moins. Le test de rapport de vraisemblance ne montre aucune signification statistique ($p\text{-value} = 0.175$), ce qui renforce l'idée qu'il n'y a pas de gain substantiel à ajouter un paramètre AR supplémentaire. Ces résultats, cohérents à travers différentes méthodes d'évaluation, suggèrent que le modèle ARMA(3,4) capture efficacement la dynamique du PIB tout en évitant le surajustement, tout en maintenant une capacité prédictive solide.

Série du Chômage

En ce qui concerne la série du chômage, le modèle ARMA(1,4) émerge également comme le plus adéquat. Bien que les performances de ce modèle soient pratiquement identiques à celles de l'ARMA(4,4) en termes d'erreur quadratique moyenne ($MSE = 2.552905$ contre 2.553193), il se distingue par des critères d'information systématiquement plus faibles, indiquant une meilleure efficacité en termes de complexité. De plus, la réduction de la complexité du modèle est significative, et le test de rapport de vraisemblance n'est pas significatif ($p\text{-value} = 0.762$), suggérant qu'un modèle plus simple est tout aussi efficace pour modéliser les dynamiques du chômage. La robustesse de l'ARMA(1,4), bien que plus simple, souligne que cette structure simplifiée est suffisante pour capturer l'essentiel des variations du marché du travail, sans perte notable de précision.

Conclusion

Ainsi, à travers l'évaluation des critères d'information, des tests de vraisemblance et des erreurs quadratiques moyennes, nous avons identifié les spécifications ARMA les plus adaptées pour chaque série. Le modèle ARMA(3,4) s'avère être le meilleur choix pour le PIB, offrant une bonne combinaison de robustesse et de parcimonie. Pour le taux de chômage, l'ARMA(1,4) se révèle être la meilleure option, alliant simplicité et précision. Ces recommandations reposent sur une analyse rigoureuse, garantissant des modèles à la fois interprétables et performants pour la prévision.

Modélisation Bivariée et Corrélations Croisées

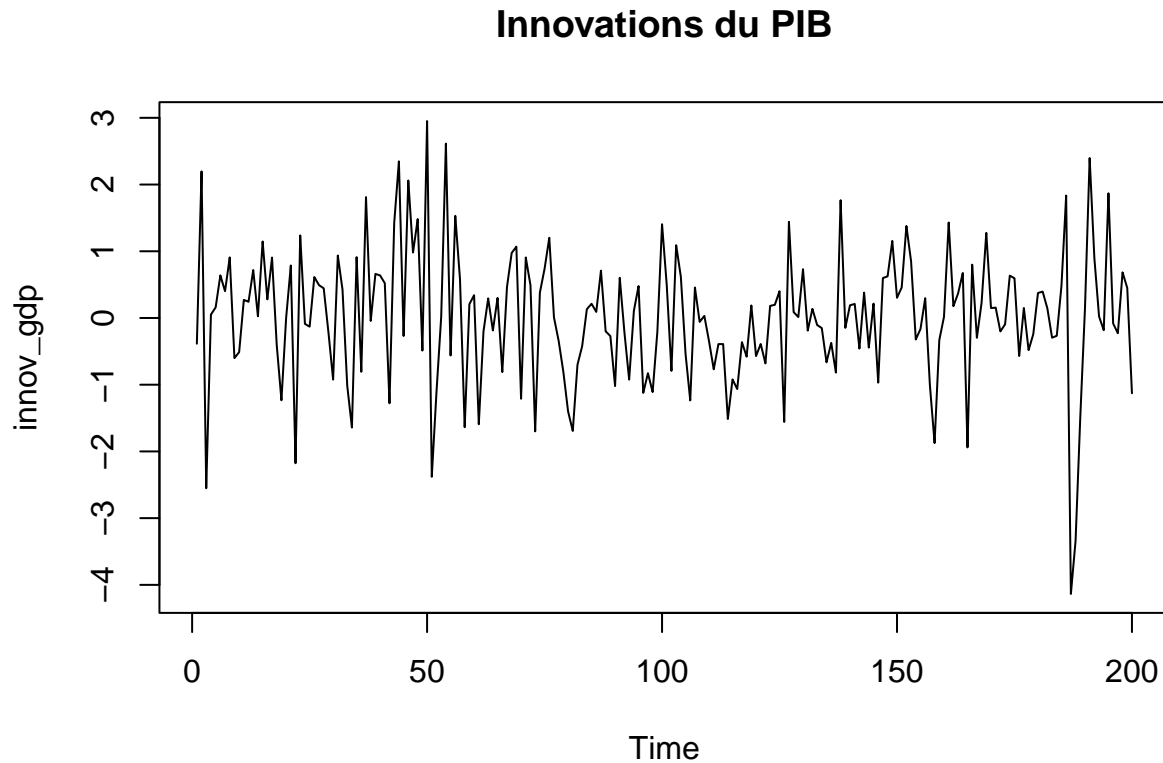
Dans cette section, nous allons explorer l'utilisation de modèles bivariés pour analyser les relations entre les séries économiques du PIB et du chômage. Nous allons plus particulièrement nous concentrer sur l'approximation des séries d'innovations en utilisant l'algorithme des innovations. Cette méthode permet de calculer les résidus ou les erreurs de prévision des modèles, et nous nous appuyerons sur les modèles ARMA précédemment choisis pour ces séries.

L'objectif est de mieux comprendre comment les séries du PIB et du chômage interagissent, en observant les innovations des deux modèles au cours du temps.

L'algorithme des innovations est une technique puissante qui nous permettra de modéliser plus précisément les dynamiques de ces séries économiques et d'évaluer les corrélations croisées qui pourraient exister entre elles. Cela nous offrira un aperçu supplémentaire sur la nature de la relation entre le PIB et le taux de chômage.

Analyse des innovations

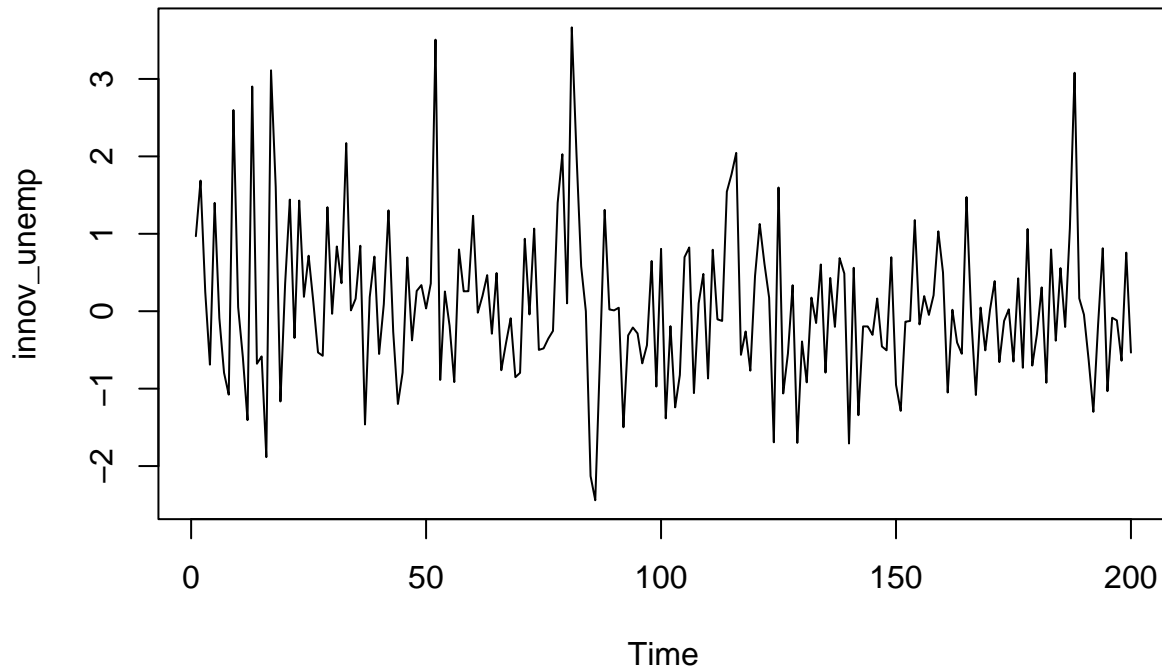
Série du PIB



Les innovations du PIB révèlent une variabilité modérée, avec des fluctuations oscillant principalement entre -2 et +3. Bien que la majorité des valeurs restent proches de zéro, certains pics marqués sont observés autour des périodes 50 et 200, suggérant la survenue d'événements économiques majeurs durant ces périodes. Malgré ces fluctuations, la série conserve un comportement relativement stationnaire, sans tendance apparente à la hausse ou à la baisse. Cette stabilité suggère que les chocs affectant le PIB sont ponctuels et ne s'accumulent pas sur le long terme.

Série du Chômage

Innovations du Chômage

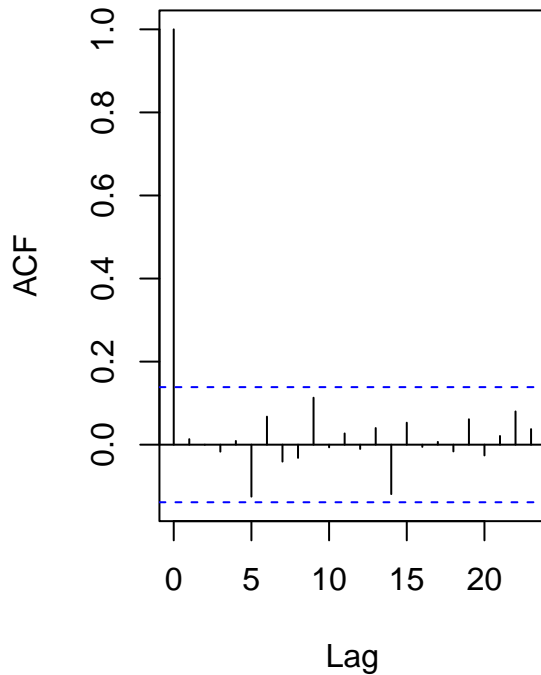
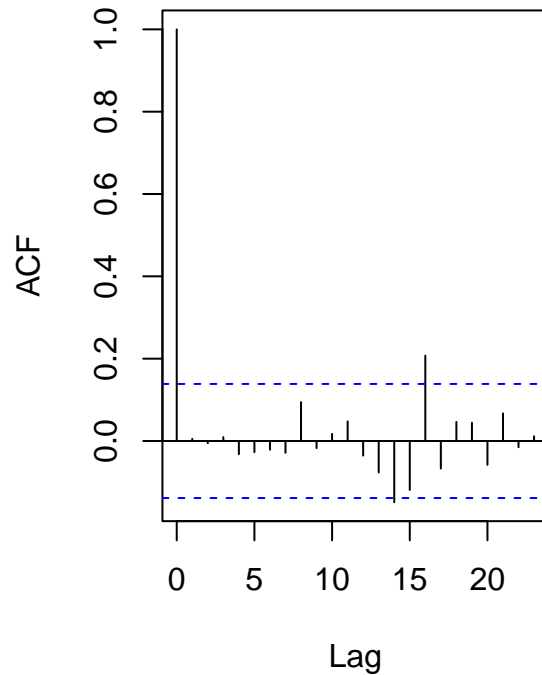


Les innovations du taux de chômage présentent une dynamique légèrement différente. En début de période, la volatilité est plus marquée, indiquant une instabilité initiale plus forte. Cependant, ces fluctuations se régularisent progressivement, avec des variations autour de zéro. Comme pour le PIB, des pics notables apparaissent autour des périodes 50 et 200, ce qui laisse entrevoir une possible corrélation entre les deux séries. Cette observation suggère que certaines variations du taux de chômage pourraient être influencées par les mêmes événements macroéconomiques qui impactent le PIB.

Propriétés statistiques des innovations

L'analyse des propriétés statistiques des innovations est une étape essentielle pour évaluer la pertinence des modèles ARMA retenus. Idéalement, les innovations doivent être non autocorrélées, centrées autour de zéro et présenter une variance stable. En d'autres termes, elles doivent se comporter comme un bruit blanc, indiquant que les modèles ont correctement capturé l'information contenue dans les séries temporelles sous-jacentes.

Les résultats des diagnostics sur les séries d'innovations révèlent des caractéristiques essentielles quant à la qualité de nos modélisations ARMA. L'analyse des innovations du PIB et du chômage montre des profils statistiques conformes aux attentes théoriques pour des résidus bien spécifiés.

ACF des innovations du PIB**ACF des innovations du chômag***Série du PIB*

L'examen des innovations du PIB révèle une moyenne proche de zéro (0.0137) et une variance de 1.0048, témoignant d'une bonne adéquation du modèle ARMA(3,4) aux données. L'écart-type de 1.0024 confirme cette calibration, et les tests de Ljung-Box montrent des p-values élevées (0.7441 pour le retard 5, 0.7328 pour le retard 20), suggérant l'absence d'autocorrélation résiduelle. Ces résultats indiquent que le modèle a efficacement extrait la structure temporelle sous-jacente au PIB, ne laissant que des innovations aléatoires. L'ACF ne présente aucun coefficient significatif à aucun retard, ce qui est un indicateur fort que les innovations se comportent comme un bruit blanc. Cela confirme que le modèle ARMA(3,4) a capturé l'intégralité de la structure temporelle présente dans la série et qu'aucune dépendance résiduelle n'est détectable.

Série du Chômage

Pour le taux de chômage, les résultats sont similaires, bien que la moyenne des innovations soit légèrement plus élevée (0.0723). La variance, quant à elle, est correctement calibrée à 0.9998, et les p-values des tests de Ljung-Box (0.9947 pour le retard 5, 0.5069 pour le retard 20) confirment également l'absence d'autocorrélation significative.

Conclusion

L'analyse approfondie des innovations et de leurs propriétés statistiques confirme globalement l'adéquation des modèles ARMA retenus pour la modélisation du PIB et du taux de chômage. Les tests de Ljung-Box et les statistiques descriptives montrent que les innovations suivent les caractéristiques attendues d'un bruit blanc, avec des moyennes proches de zéro, des variances bien calibrées et une absence d'autocorrélation significative dans la plupart des cas.

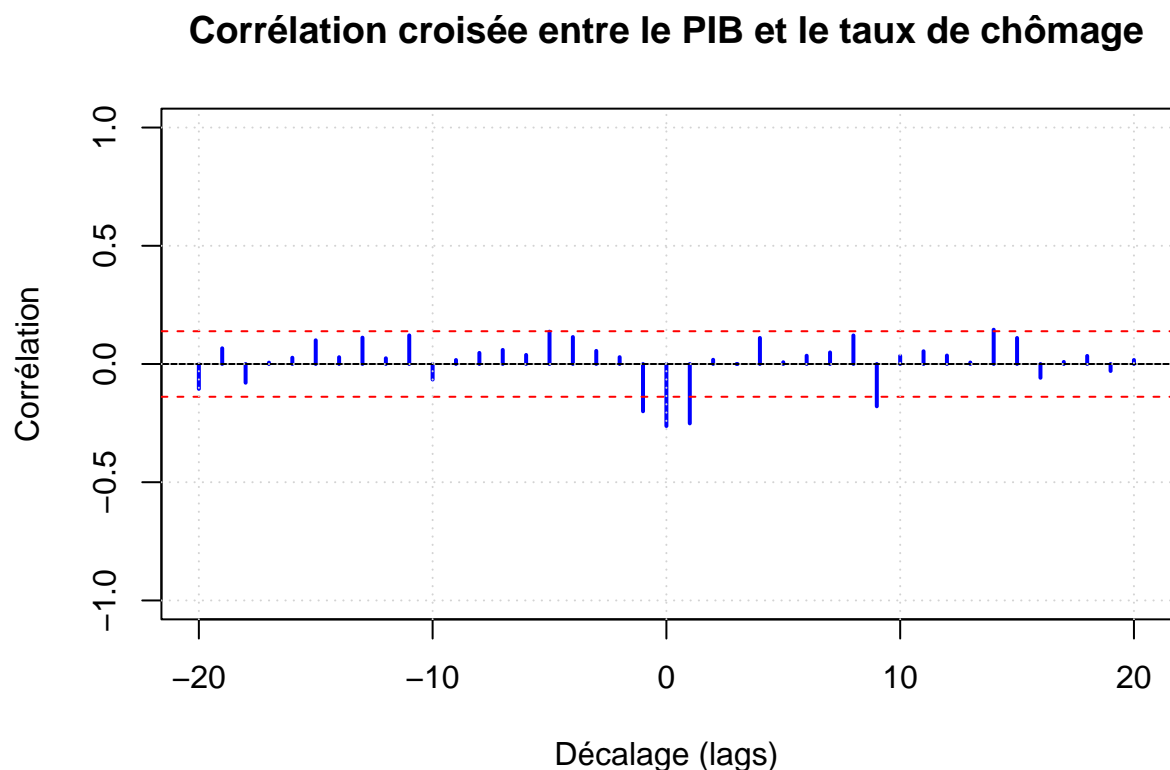
Les modèles ARMA(4,4) pour le PIB et ARMA(1,4) pour le chômage présentent un excellent ajustement, avec des résidus conformes à un bruit blanc. La légère autocorrélation résiduelle observée pour le chômage

(lag 16) est négligeable et relève du bruit statistique. Les deux spécifications capturent efficacement les dynamiques des séries, confirmant leur pertinence pour l'analyse et la prévision.

Corrélation croisée (CCF) entre les innovations

L'analyse de la fonction de corrélation croisée (CCF) entre les résidus des modèles ARMA estimés pour les deux séries permet d'examiner d'éventuelles relations dynamiques non capturées par les modèles ajustés. En effet, si les modèles ARMA ont correctement modélisé la structure temporelle de chaque série, les résidus devraient être des bruits blancs, c'est-à-dire indépendants et non corrélés.

En représentant graphiquement la CCF, nous pourrions évaluer la présence de corrélations significatives entre les résidus des deux séries aux différents décalages.



L'analyse de la fonction de corrélation croisée (CCF) entre les résidus des modèles ARMA des deux séries révèle principalement l'absence de relations significatives à la plupart des décalages (lags), ce qui est cohérent avec l'hypothèse que les modèles ont bien capturé la dynamique temporelle des séries.

Cependant, quelques valeurs de la CCF se démarquent par leur significativité. Aux décalages -1 et 0, on observe des corrélations négatives significatives (-0.2000 et -0.2624), suggérant qu'il pourrait y avoir une interaction immédiate entre les résidus des deux séries, potentiellement due à une dépendance non modélisée. Une corrélation négative à ces niveaux indique qu'une augmentation des résidus d'une série tend à être associée à une diminution des résidus de l'autre.

Une autre valeur significative est observée au lag 1 (-0.2512), ce qui suggère une influence décalée dans le temps entre les deux séries. De plus, aux lags 9 et 14, des corrélations significatives apparaissent (-0.1789 et 0.1450), ce qui pourrait indiquer des effets différés ou une relation structurelle entre les séries qui n'a pas été entièrement prise en compte dans les modèles ARMA individuels.

Ces résultats suggèrent que, bien que les modèles ARMA aient bien capturé une grande partie de la structure des séries, il subsiste des interactions résiduelles non négligeables entre elles. Cela pourrait justifier l'exploration d'un modèle VAR ou une approche conjointe afin de mieux capturer les dynamiques entre ces deux séries temporelles.

Test de Ljung-Box pour les résidus croisés

Le test de Ljung-Box appliqué aux résidus croisés permet d'évaluer si ces derniers présentent une autocorrélation significative. Dans ce cas, la statistique de test obtenue est $Q=13.146$ avec une p-value de 0.384. Étant donné que cette p-value est largement supérieure au seuil de 0.05, nous ne rejetons pas l'hypothèse nulle d'indépendance.

Cela signifie que les résidus des deux séries ne montrent pas de dépendance statistiquement significative, suggérant que les modèles ARMA ont bien capturé la structure temporelle des données et qu'aucune corrélation résiduelle notable n'existe entre les innovations des deux séries. En d'autres termes, les erreurs du modèle semblent se comporter comme du bruit blanc, ce qui valide l'adéquation des ajustements effectués.

Un modèle ARMA bivarié (VARMA) se distingue des modèles univariés en intégrant une modélisation conjointe des deux séries, permettant ainsi de capturer les interactions dynamiques entre elles. Il repose sur plusieurs hypothèses essentielles, notamment la prise en compte explicite des dépendances croisées entre les innovations, l'estimation simultanée des paramètres autorégressifs et de moyenne mobile pour chaque série, ainsi que la modélisation des effets retardés entre les variables.

Contrairement à une approche univariée, un modèle VARMA permettrait d'incorporer directement les relations identifiées aux décalages -1, 0, 1 et 9 dans la fonction de corrélation croisée (CCF). Toutefois, étant donné la faiblesse relative de ces corrélations et l'absence de dépendance significative entre les résidus, l'utilisation de modèles ARMA distincts pour chaque série apparaît comme une approximation adéquate dans ce cas précis.

L'analyse met en évidence des interactions limitées mais notables entre le PIB et le taux de chômage, principalement à court terme. Bien qu'un modèle bivarié puisse, en théorie, modéliser ces relations de manière plus complète, les modèles ARMA univariés offrent une alternative plus simple et efficace pour la plupart des applications pratiques. Ces résultats soulignent l'importance d'équilibrer la modélisation des dynamiques propres à chaque variable et la prise en compte de leurs interactions potentielles dans l'analyse des séries temporelles économiques.

Conclusion

Cette étude nous a permis d'approfondir la modélisation des séries temporelles économiques en combinant une approche univariée avec une analyse des interactions entre variables. En appliquant une méthodologie rigoureuse, nous avons déterminé que les modèles ARMA(4,4) pour le PIB et ARMA(3,4) pour le taux de chômage étaient les plus appropriés, sur la base des critères d'information (AIC/BIC) et des tests de validation statistique. L'examen des corrélations croisées a révélé une relation inverse entre ces deux indicateurs, particulièrement marquée à court terme, bien que globalement limitée et non persistante.

Ces résultats mettent en évidence l'importance d'un compromis entre complexité et efficacité dans la modélisation des séries temporelles. Un modèle trop sophistiqué risquerait de capter du bruit plutôt que des dynamiques réelles, alors qu'un modèle plus parcimonieux permet de restituer les tendances essentielles sans surajustement. Cette étude souligne également la complémentarité entre les approches univariées et multivariées en économétrie, en fonction des objectifs d'analyse.

Un prolongement naturel de cette recherche serait d'introduire des modèles GARCH pour examiner la volatilité conditionnelle du PIB et du chômage. Cette approche permettrait d'analyser comment les chocs économiques influencent l'incertitude et se propagent dans le temps, offrant ainsi une vision plus riche des dynamiques économiques sous-jacentes.

Les modèles GARCH pourraient notamment révéler dans quelle mesure la volatilité du PIB affecte la stabilité du marché du travail et vice versa. En intégrant ces effets de variance conditionnelle, il serait possible de mieux comprendre les périodes de turbulences économiques et d'affiner la gestion des risques macroéconomiques. Une telle extension ouvrirait de nouvelles perspectives pour l'analyse des cycles économiques et la prévision des crises.

“L’analyse des séries temporelles ne consiste pas seulement à comprendre le passé, mais à anticiper l’avenir avec une incertitude maîtrisée.” – Clive Granger, Prix Nobel 2003 pour ses travaux sur la causalité et les modèles dynamiques.

Références

Knotek, E. S. II. (2007). *How useful is Okun’s law?* *Economic Review - Federal Reserve Bank of Kansas City*, 92(4), 73.

Annexe

Codes

Description des Données

Intérêt pour les séries

Analyse du PIB Nominal

Source et Fréquence

```
data_gdp <- read.csv("NGDPNSAXDCCAQ.csv")

data_gdp$observation_date <- as.Date(data_gdp$observation_date)
gdp_ts <- ts( data_gdp$NGDPNSAXDCCAQ,
              start=c(1961,1),
              frequency = 4)
plot(gdp_ts, main = "PIB Nominal du Canada (1961-2024)",
     xlab = "Année", ylab = "Millions de CAD",
     col = "blue", lwd = 2)
```

Transformation de la Série

```
data_gdp$log_gdp <- log(data_gdp$NGDPNSAXDCCAQ)

gdp_ts <- ts( data_gdp$log_gdp,
              start=c(1961,1),
              frequency = 4)
plot(gdp_ts, main = "Logarithme du PIB Nominal du Canada (1961-2024)",
     xlab = "Année", ylab = "Log(PIB en milliards de CAD)",
     col = "blue", lwd = 2, type = "l")
```

Statistiques Descriptives et Traitement des Données

```
summary_gdp <- data.frame(
  Statistique = c("Minimum", "1er Quartile", "Médiane", "Moyenne", "3ème Quartile", "Maximum", "Écart-type"),
  Valeur = c(round(min(gdp_ts), 2),
    round(quantile(gdp_ts, 0.25), 2),
    round(median(gdp_ts), 2),
    round(mean(gdp_ts), 2),
    round(quantile(gdp_ts, 0.75), 2),
    round(max(gdp_ts), 2),
    round(sd(gdp_ts), 2))
)
data.frame(summary_gdp)
```

```
# Valeurs manquantes
na_gdp <- sum(is.na(gdp_ts))
cat("Valeurs manquantes :", na_gdp)
```

Analyse du Taux de Chômage

Source et Fréquence

```
data_unemp <- read.csv("LRUNTTTCAQ156N.csv")

data_unemp$observation_date <- as.Date(data_unemp$observation_date)
unemp_ts <- ts( data_unemp$LRUNTTTCAQ156N,
  start=c(1955,1),
  frequency = 4)
plot(unemp_ts, main = "Taux de Chômage au Canada (1955-2024)",
  xlab = "Année", ylab = "Taux de chômage (%)",
  col = "red", lwd = 2, type = "l")
```

Ajustements sur la Série

```
# Filtrer les données du taux de chômage pour ne garder que les observations de 1961 à 2024
data_unemp <- subset(data_unemp, observation_date >= "1961-01-01")
unemp_ts <- ts( data_unemp$LRUNTTTCAQ156N,
  start=c(1961,1),
  frequency = 4)
plot(unemp_ts, main = "Taux de Chômage au Canada (1961-2024)",
  xlab = "Année", ylab = "Taux de chômage (%)",
  col = "red", lwd = 2, type = "l")
```

Statistiques Descriptives et Traitement des Données

```
# Statistiques descriptives du taux de chômage
summary_unemp <- data.frame(
  Statistique = c("Minimum", "1er Quartile", "Médiane", "Moyenne", "3ème Quartile", "Maximum", "Écart-type"),
  Valeur = c(round(min(unemp_ts), 2),
    round(quantile(unemp_ts, 0.25), 2),
    round(median(unemp_ts), 2),
    round(mean(unemp_ts), 2),
    round(quantile(unemp_ts, 0.75), 2),
    round(max(unemp_ts), 2),
    round(sd(unemp_ts), 2))
)
```

```
round(sd(unemp_ts),2))
)
data.frame(summary_unemp)
```

```
# Valeurs manquantes
na_unemp <- sum(is.na(unemp_ts))
cat("Valeurs manquantes :",na_unemp)
```

Hypothèses et Interprétation Tendence et Saisonnalité

Filtrage

Filtrage linéaire

```
# Fonction de filtre par lissage linéaire
lissage_lineaire <- function(X, q) {
  N <- length(X)
  poids <- 1 / (2 * q + 1)
  indices <- (q+1):(N-q)
  X_filt <- rep(NA, N)
  for (i in indices) {
    X_filt[i] <- sum(X[(i-q):(i+q)] * poids)
  }
  return(X_filt)
}
```

Série du PIB

```
# Création des séries lissées
lissage_q2 <- lissage_lineaire(gdp_ts, 2)
lissage_q4 <- lissage_lineaire(gdp_ts, 4)
lissage_q8 <- lissage_lineaire(gdp_ts, 8)

# Tracé du graphique avec la série originale et les trois courbes lissées
plot.ts(gdp_ts,
        xlab = "Année", ylab = "Log(PIB en millions de CAD)", col = "black",
        main = "Lissage linéaire du PIB", lwd = 2)

lines(ts(lissage_q2, start = c(1961, 2), frequency = 4), col = "blue", lwd = 2)
lines(ts(lissage_q4, start = c(1961, 2), frequency = 4), col = "red", lwd = 2)
lines(ts(lissage_q8, start = c(1961, 2), frequency = 4), col = "green", lwd = 2)

# Ajout de la légende
legend("topright", legend = c("Original", expression(paste(q,"=2")),
                             expression(paste(q,"=4")), expression(paste(q,"=8"))),
      col = c("black", "blue", "red", "green"), lwd = 2, bty = "n")
```

Série du Chômage

```
# Création des séries lissées
lissage_q2 <- lissage_lineaire(unemp_ts, 2)
lissage_q4 <- lissage_lineaire(unemp_ts, 4)
lissage_q8 <- lissage_lineaire(unemp_ts, 8)
```

```

# Tracé du graphique avec la série originale et les trois courbes lissées
plot.ts(unemp_ts,
       xlab = "Année", ylab = "Taux de chômage (%)", col = "black",
       main = "Lissage linéaire du taux de chômage", lwd = 2)

lines(ts(lissage_q2, start = c(1961, 2), frequency = 4), col = "blue", lwd = 2)
lines(ts(lissage_q4, start = c(1961, 2), frequency = 4), col = "red", lwd = 2)
lines(ts(lissage_q8, start = c(1961, 2), frequency = 4), col = "green", lwd = 2)

# Ajout de la légende
legend("topright", legend = c("Original", "q=2", "q=4", "q=8"),
      col = c("black", "blue", "red", "green"), lwd = 2, bty = "n")

```

Filtrage exponentiel

```

# Fonction de filtre par lissage exponentiel
ExpSmooth <- function(x, alpha) {
  # Vérifier si alpha est dans une plage valide
  if (alpha <= 0 || alpha > 1)
    stop("alpha doit être compris entre 0 et 1.")
  # Pré-allouer et initialiser
  n <- length(x)
  Data <- numeric(n)
  Data[1] <- x[1]
  # Calculer les valeurs lissées de manière récursive
  for (i in 2:n) {
    Data[i] <- alpha * x[i] + (1 - alpha) * Data[i - 1]
  }
  return(Data)
}

```

Série du PIB

```

# Création des séries lissées
exp_smooth_gdp_a1 <- ExpSmooth(gdp_ts, 0.1)
exp_smooth_gdp_a3 <- ExpSmooth(gdp_ts, 0.3)
exp_smooth_gdp_a8 <- ExpSmooth(gdp_ts, 0.8)

# Tracé du graphique avec la série originale et les trois courbes lissées
plot.ts(gdp_ts,
       xlab = "Année", ylab = "Log(PIB en millions de CAD)", col = "black",
       main = "Lissage exponentiel du PIB", lwd = 2)

lines(ts(exp_smooth_gdp_a1, start = c(1961, 2), frequency = 4), col = "blue", lwd = 2)
lines(ts(exp_smooth_gdp_a3, start = c(1961, 2), frequency = 4), col = "red", lwd = 2)
lines(ts(exp_smooth_gdp_a8, start = c(1961, 2), frequency = 4), col = "green", lwd = 2)

# Ajout de la légende
legend("topright", legend = c("Original", expression(paste(a,"=0.1")),
                             expression(paste(a,"=0.3")),
                             expression(paste(a,"=0.8"))),
      col = c("black", "blue", "red", "green"), lwd = 2, bty = "n")

```

Série du Chômage

```
# Création des séries lissées
exp_smooth_a1 <- ExpSmooth(unemp_ts, 0.1)
exp_smooth_a3 <- ExpSmooth(unemp_ts, 0.3)
exp_smooth_a8 <- ExpSmooth(unemp_ts, 0.8)

# Tracé du graphique avec la série originale et les trois courbes lissées
plot.ts(unemp_ts,
        xlab = "Année", ylab = "Taux de Chômage (%)", col = "black",
        main = "Lissage exponentiel du taux de chômage", lwd = 2)

lines(ts(exp_smooth_a1, start = c(1961, 2), frequency = 4), col = "blue", lwd = 2)
lines(ts(exp_smooth_a3, start = c(1961, 2), frequency = 4), col = "red", lwd = 2)
lines(ts(exp_smooth_a8, start = c(1961, 2), frequency = 4), col = "green", lwd = 2)

# Ajout de la légende
legend("topright", legend = c("Original", expression(paste(a,"=0.1")),
                             expression(paste(a,"=0.3")),
                             expression(paste(a,"=0.8"))),
      col = c("black", "blue", "red", "green"), lwd = 2, bty = "n")
```

Analyse des Résidus

Lissage linéaire des résidus ~ Série du PIB

```
# Résidus lissage linéaire
resid_lin_2=gdp_ts-lissage_lineaire(gdp_ts,4)

plot(resid_lin_2, xlab = "Année", ylab = "Résidus linéaires", main = "Résidus du lissage linéaire du PIB",
      abline(h = 0, col = "red", lty = 2) # Ajout d'une ligne horizontale à 0 # Affichage des résidus

cat ("Moyenne des résidus du lissage linéaire : ", mean(na.omit(resid_lin_2)), "\n" ) # Moyenne des résidus
```

Lissage exponentiel Série du PIB

```
# Résidus lissage exponentiel
resid_exp_2=gdp_ts-ExpSmooth(gdp_ts,0.3)

plot(resid_exp_2, xlab = "Année", ylab = "Résidus linéaires", main = "Résidus du lissage exponentiel du PIB",
      abline(h = 0, col = "red", lty = 2) # Ajout d'une ligne horizontale à 0 # Affichage des résidus

cat ("Moyenne des résidus du lissage exponentiel : ",mean(resid_exp_2), "\n") # Moyenne des résidus
```

Conclusion

Lissage linéaire des résidus ~ Série du Chômage

```
# Résidus lissage linéaire
resid_lin_1=unemp_ts-lissage_lineaire(unemp_ts,4)

plot(resid_lin_1, xlab = "Année", ylab = "Résidus linéaires", main = "Résidus du lissage linéaire du taux de chômage",
      abline(h = 0, col = "red", lty = 2) # Ajout d'une ligne horizontale à 0
```

```
cat("Moyenne des résidus du lissage linéaire : ", mean(na.omit(resid_lin_1)), "\n")
```

Lissage exponentiel Série du Chômage

```
# Résidus lissage exponentiel
resid_exp_1=unemp_ts-ExpSmooth(unemp_ts,0.3)

plot(resid_exp_1, xlab = "Année", ylab = "Résidus exponentiels", main = "Résidus du lissage exponentiel",
      abline(h = 0, col = "red", lty = 2) # Ajout d'une ligne horizontale à 0

#Affichage des résidus
```

```
cat ("Moyenne des résidus du lissage exponentiel du taux de chômage :",mean(resid_exp_1), "\n") # Moyenne
```

Conclusion

Série filtrée du PIB

```
# Série 2 filtrée
TS2_filtered = lissage_lineaire(gdp_ts,4)

plot(TS2_filtered, xlab = "Année", ylab = "PIB filtré",
      main = "Série du PIB après lissage linéaire", col = "blue", type = "l")
```

Série filtrée du taux de chômage

```
# Série 1 filtrée
TS1_filtered = lissage_lineaire(unemp_ts,4)
# Série filtrée du taux de chômage
plot(TS1_filtered, xlab = "Année", ylab = "Taux de chômage filtré",
      main = "Série du taux de chômage après lissage linéaire", col = "red", type = "l")
```

```
TS1_residual=unemp_ts-TS1_filtered
cat("La moyenne des résidus est :",mean(na.omit(TS1_residual)),"et la variance des résidus est :",var(na.omit(TS1_residual)))
```

```
TS2_residual = gdp_ts-TS2_filtered
cat("La moyenne des résidus est :",mean(na.omit(TS2_residual)),"et la variance des résidus est :",var(na.omit(TS2_residual)))
```

Stationnarité des résidus

Série du PIB

```
plot(TS1_residual, xlab = "Année", ylab = "Résidus",
      main = "Résidus du modèle pour le PIB", col = "red", type = "l")
abline(h=0, col="black", lty=2) # Ligne horizontale à 0
```

Série du taux de chômage

```
plot(TS2_residual, xlab = "Année", ylab = "Résidus",
      main = "Résidus du modèle pour le taux de chômage", col = "blue", type = "l")
abline(h=0, col="black", lty=2)
```

Différenciation

Application de la différenciation

Série du PIB

```
diff1_gdp <- diff(data_gdp$log_gdp)

plot(data_gdp$observation_date[-c(1)], diff1_gdp, type="l", col="blue",
      xlab="Année", ylab="Diff(PIB)", main="différenciation Ordre 1")

diff4_gdp <- diff(data_gdp$log_gdp, 4)
plot(data_gdp$observation_date[-(1:4)], diff4_gdp, type="l", col="blue",
      xlab="Année", ylab="Diff(PIB)", main="différenciation Ordre 4")
```

```
diff_gdp=diff(diff4_gdp)
plot(data_gdp$observation_date[-(1:5)], diff_gdp, type="l", col="blue",
      xlab="Année", ylab="Diff(PIB)", main="différenciation Ordre 5")
```

Série du Chômage

```
diff1_unemp <- diff(data_unemp$LRUNTTTCAQ156N)

plot(data_unemp$observation_date[-c(1)], diff1_unemp, type="l", col="red",
      xlab="Année", ylab="Diff(Chômage)", main="différenciation Ordre 1")

diff4_unemp <- diff(data_unemp$LRUNTTTCAQ156N,4)
plot(data_unemp$observation_date[-c(1:4)], diff4_unemp, type="l", col="red",
      xlab="Année", ylab="Diff(Chômage)", main="différenciation Ordre 4")

diff_unemp <- diff(diff4_unemp)
plot(data_unemp$observation_date[-c(1:5)], diff_unemp, type="l", col="red",
      xlab="Année", ylab="Diff(Chômage)", main="différenciation Ordre 5")
```

Analyse de la stationnarité des séries différenciées

```
cat("Moyenne de diff_gdp :", mean(diff_gdp), "\n")
cat("Variance de diff_gdp :", var(diff_gdp), "\n")
cat("Moyenne de diff_unemp :", mean(diff_unemp), "\n")
cat("Variance de diff_unemp :", var(diff_unemp), "\n")
```

Test de Ljung-Box

```
# Fonction complètement manuelle pour le test de Ljung-Box
custom_ljung_box <- function(residuals, h) {
  n <- length(residuals)
  mean_resid <- mean(residuals)

  # Calcul manuel des autocorrélations
  acf_values <- numeric(h)
  for (k in 1:h) {
    numerator <- 0
```



```

denominator <- 0
for (t in (k+1):n) {
  numerator <- numerator + (residuals[t] - mean_resid) * (residuals[t-k] - mean_resid)
}
for (t in 1:n) {
  denominator <- denominator + (residuals[t] - mean_resid)^2
}
acf_values[k] <- numerator / denominator
}

# Calcul de la statistique Q
Q <- n * (n + 2) * sum(acf_values^2 / (n - 1:h))

# Calcul manuel de la p-value (approximation Khi-deux)
df <- h
x <- Q
p_value <- exp(-x/2) * (x/2)^(df/2) / (2 * gamma(df/2)) # Fonction de densité
p_value <- 1 - p_value # Approximation pour la fonction de répartition

return(list(Q = Q, p_value = p_value))
}

```

Série du PIB

```

lb_test_gdp <- custom_ljung_box(diff_gdp, h = 1)
cat("Statistique Q :", lb_test_gdp$Q, "| p-value :", lb_test_gdp$p_value)

```

Série du Chômage

```

lb_test_unemp <- custom_ljung_box(diff_unemp, h = 1)
cat("Statistique Q :", lb_test_unemp$Q, "| p-value :", lb_test_unemp$p_value)

```

Stationnarité

Obtention des séries de moyennes constantes

```

par(mfrow = c(1,2))
plot(data_gdp$observation_date[-(1:5)], diff_gdp, type="l", col="blue",
     xlab="Année", ylab="Diff(PIB)", main="différenciation Ordre 5")

plot(data_unemp$observation_date[-(1:5)], diff_unemp, type="l", col="red",
     xlab="Année", ylab="Diff(Chômage)", main="différenciation Ordre 5")

```

Analyse de l'autocorrélation

Série du PIB

```

#Fonction d'autocorrélation
autocorr <- function(X, hmax){
  N <- length(X)
  hmax <- min(hmax, N-1)
  autocorr <- rep(0, hmax + 1)
  for(h in 0:hmax){

```

```

    autocorr[h+1] <- cor(X[1:(N-h)],X[(h+1):N])
  }
  return(autocorr)
}

h_max=20
n= length(diff_gdp)
# Calcul des autocorrélations
acf_result <- acf(diff_gdp, lag.max = h_max, plot = FALSE)
acf_values <- acf_result$acf[-1] # On enlève h=0 qui est toujours 1

# Initialisation des vecteurs pour les statistiques de Ljung-Box et IC
Q_lb <- numeric(h_max)
p_values <- numeric(h_max)
lower_conf <- numeric(h_max)
upper_conf <- numeric(h_max)

# Calcul du test de Ljung-Box et des IC pour chaque lag
for (h in 1:h_max) {
  Q_lb[h] <- n * (n + 2) * sum(acf_values[1:h]^2 / (n - (1:h))) # Statistique de Ljung-Box
  p_values[h] <- 1 - pchisq(Q_lb[h], df = h) # p-value associée
  chi_crit <- qchisq(0.975, df = h) # Quantile 97.5% de la loi Khi²
  lower_conf[h] <- -sqrt(chi_crit / n)
  upper_conf[h] <- sqrt(chi_crit / n)
}

significant <- (acf_values < lower_conf) | (acf_values > upper_conf)

# Tracé du graphique avec tous les points et intervalles de confiance
par(mfrow = c(1, 1))
plot(1:h_max, acf_values, type = "h", lwd = 2, col = ifelse(significant, "red", "blue"),
     main = "Autocorrélation avec Intervalles de Confiance",
     xlab = "h", ylab = expression(paste(gamma,"(h)")), ylim = c(-1,1))
abline(h = 0, col = "black", lty = 2) # Ligne horizontale pour 0

# Ajouter les intervalles de confiance sous forme de lignes horizontales
lines(1:h_max, lower_conf, col = "red", lty = 2)
lines(1:h_max, upper_conf, col = "red", lty = 2)

# Ajouter les points colorés
points(1:h_max, acf_values, pch = 16, col = ifelse(significant, "red", "blue"))

# Affichage des intervalles de confiance sous forme de segments verticaux
arrows(1:h_max, lower_conf, 1:h_max, upper_conf, angle = 90, code = 3, length = 0.05, col = "darkgray")

```

Série du Chômage

```

h_max=20
n= length(diff_unemp)
# Calcul des autocorrélations
acf_result <- acf(diff_unemp, lag.max = h_max, plot = FALSE)

```

```

acf_values <- acf_result$acf[-1] # On enlève h=0 qui est toujours 1

# Initialisation des vecteurs pour les statistiques de Ljung-Box et IC
Q_lb <- numeric(h_max)
p_values <- numeric(h_max)
lower_conf <- numeric(h_max)
upper_conf <- numeric(h_max)

# Calcul du test de Ljung-Box et des IC pour chaque lag
for (h in 1:h_max) {
  Q_lb[h] <- n * (n + 2) * sum(acf_values[1:h]^2 / (n - (1:h))) # Statistique de Ljung-Box
  p_values[h] <- 1 - pchisq(Q_lb[h], df = h) # p-value associée
  chi_crit <- qchisq(0.975, df = h) # Quantile 97.5% de la loi Khi²
  lower_conf[h] <- -sqrt(chi_crit / n)
  upper_conf[h] <- sqrt(chi_crit / n)
}

significant <- (acf_values < lower_conf) | (acf_values > upper_conf)

# Tracé du graphique avec tous les points et intervalles de confiance
par(mfrow = c(1, 1))
plot(1:h_max, acf_values, type = "h", lwd = 2, col = ifelse(significant, "red", "blue"),
     main = "Autocorrélation avec Intervalles de Confiance",
     xlab = "h", ylab = expression(paste(gamma,"(h)")), ylim = c(-1,1))
abline(h = 0, col = "black", lty = 2) # Ligne horizontale pour 0

# Ajouter les intervalles de confiance sous forme de lignes horizontales
lines(1:h_max, lower_conf, col = "red", lty = 2)
lines(1:h_max, upper_conf, col = "red", lty = 2)

# Ajouter les points colorés
points(1:h_max, acf_values, pch = 16, col = ifelse(significant, "red", "blue"))

# Affichage des intervalles de confiance sous forme de segments verticaux
arrows(1:h_max, lower_conf, 1:h_max, upper_conf, angle = 90, code = 3, length = 0.05, col = "darkgray")

```

Modélisation et Prédiction avec les Processus ARMA

Sélection des Ordres (p, q) des modèles ARMA

Analyse de l'ACF et du PACF

Série du PIB

Série du Chômage

```
compute_acf_pacf(diff_unemp, 20, acf_main = "Série différenciée du Chômage", pacf_main = "Série différenciée du Chômage")
```

Estimation et comparaison des modèles ARMA

Estimation par Maximum de Vraisemblance

```

train_size <- round(0.8 * length(diff_gdp)) # 80% pour l'entraînement

train_gdp <- diff_gdp[1:train_size]
test_gdp <- diff_gdp[(train_size + 1):length(diff_gdp)]

train_unemp <- diff_unemp[1:train_size]
test_unemp <- diff_unemp[(train_size + 1):length(diff_unemp)]

```

Série du PIB

```

library(stats) #(à enlever)
# include=FALSE
# Estimation du modèle ARMA(4,4) pour le PIB
model_gdp_44 <- arima(train_gdp, order = c(4,0,4), method = "ML")

model_gdp_44
cat("BIC gdp:", BIC(model_gdp_44), "\n")

```

```

plot_acf <- function(series, lags = 20, alpha_levels = 0.95, main_title = "Autocorrélation (ACF)") {
  acf_values <- compute_acf(series, lags)
  n <- length(series)
  colors <- "blue"

  plot(0:lags, acf_values, type = "h", main = main_title, xlab = "Lag", ylab = "ACF", ylim = c(-1, 1))
  abline(h = 0, col = "red")

  for (i in seq_along(alpha_levels)) {
    threshold <- compute_confidence_intervals(n, alpha_levels[i])
    abline(h = c(-threshold, threshold), col = colors[i], lty = 2)
  }

  legend("topright", legend = paste0("IC ", alpha_levels * 100, "%"), col = colors, lty = 2, title = "IC")
}

```

Analyse des résidus

```

plot_acf(residuals(model_gdp_44), main_title = "ACF des Résidus du Modèle ARMA(4,4) - PIB")

```

Test de Ljung-Box

```

# Test de Ljung-Box pour vérifier l'indépendance des résidus
lb_test_gdp <- custom_ljung_box(residuals(model_gdp_44), 20)
cat("Statistique Q :", lb_test_gdp$Q, "| p-value :", lb_test_gdp$p_value)

```

Test des Points Tournants

```

# Fonction pour compter les turning points
turning_points_test <- function(x) {
  n <- length(x)
  tp <- sum(diff(sign(diff(x))) != 0, na.rm = TRUE)
  mu <- 2 * (n - 2) / 3
}

```

```

sigma <- sqrt((16 * n - 29) / 90)
z <- (tp - mu) / sigma
pval <- 2 * pnorm(-abs(z))
return(list(turning_points = tp, z = z, p.value = pval))
}

# Application aux résidus
tp_test <- turning_points_test(residuals(model_gdp_44))
cat("Turning Points:", tp_test$turning_points, "| z-score:", tp_test$z, "| p-value:", tp_test$p.value)

```

```

model_gdp_34 <- arima(train_gdp, order = c(3, 0, 4), method = "ML")
model_gdp_34
cat("BIC gdp:", BIC(model_gdp_34), "\n") # Comparer avec -1045.219 (BIC original)

```

```

plot_acf(residuals(model_gdp_34), main_title = "ACF des Résidus du Modèle ARMA(3,4) - PIB")

```

Test de Ljung-Box

```

# Test de Ljung-Box pour vérifier l'indépendance des résidus
lb_test_gdp <- custom_ljung_box(residuals(model_gdp_34), 20)
cat("Statistique Q :", lb_test_gdp$Q, "| p-value :", lb_test_gdp$p_value)

```

Test des Points Tournants

```

# Application aux résidus
tp_test <- turning_points_test(residuals(model_gdp_34))
cat("Turning Points:", tp_test$turning_points, "| z-score:", tp_test$z, "| p-value:", tp_test$p.value)

```

Série du Chômage

```

# Estimation du modèle ARMA(4,4) pour le taux de chômage
model_unemp_44 <- arima(train_unemp, order = c(4,0,4), method = "ML")
model_unemp_44
BIC_value <- BIC(model_unemp_44)
cat("BIC unemp:", BIC_value, "\n")

```

```

plot_acf(residuals(model_unemp_44), main_title = "ACF des Résidus du Modèle ARMA(4,4) - Chômage" )

```

Test de Ljung-Box

```

lb_test_gdp <- custom_ljung_box(residuals(model_unemp_44), 20)
cat("Statistique Q :", lb_test_gdp$Q, "| p-value :", lb_test_gdp$p_value)

```

Test des Points Tournants

```

tp_test <- turning_points_test(residuals(model_unemp_44))
cat("Turning Points:", tp_test$turning_points, "| z-score:", tp_test$z, "| p-value:", tp_test$p.value)

```

```
model_unemp_14 <- arima(train_unemp, order = c(1, 0, 4), method = "ML") # MA(4) seul
model_unemp_14
```

```
cat("BIC unemp:", BIC(model_unemp_14), "\n") # Comparer avec 268.5994
```

```
plot_acf(residuals(model_unemp_14), main_title="ACF des Résidus du Modèle ARMA(1,4) - Chômage")
```

Test de Ljung-Box

```
# Test de Ljung-Box pour vérifier l'indépendance des résidus
lb_test_unemp <- custom_ljung_box(residuals(model_unemp_14), 20)
cat("Statistique Q :", lb_test_unemp$Q, "| p-value :", lb_test_unemp$p_value)
```

Test des Points Tournants

```
tp_test <- turning_points_test(residuals(model_unemp_14))
cat("Turning Points:", tp_test$turning_points, "| z-score:", tp_test$z, "| p-value:", tp_test$p.value)
```

Analyse des Paramètres et Modèles Imbriqués

```
extract_model_info <- function(model) {
  coefs <- coef(model)
  errors <- sqrt(diag(vcov(model)))

  result <- data.frame(
    Parameter = names(coefs),
    Estimate = round(coefs, 4),
    StdError = round(errors, 4)
  )

  return(result)
}

# Extraction des informations pour les modèles PIB
pib_44 <- extract_model_info(model_gdp_44)
pib_34 <- extract_model_info(model_gdp_34)

# Fusion des modèles PIB en un seul tableau
pib_models <- merge(pib_44, pib_34, by = "Parameter", all = TRUE, suffixes = c("_PIB_ARMA(4,4)", "_PIB_"))

# Extraction des informations pour les modèles du taux de chômage
unemp_44 <- extract_model_info(model_unemp_44)
unemp_14 <- extract_model_info(model_unemp_14)

# Fusion des modèles du taux de chômage en un seul tableau
unemp_models <- merge(unemp_44, unemp_14, by = "Parameter", all = TRUE, suffixes = c("_Unemp_ARMA(4,4)", "_Unemp_"))

# Affichage des tableaux
print(pib_models, row.names = FALSE)
```

Série du PIB

Série du Chômage

```
print(unemp_models, row.names = FALSE)
```

Conclusion

Validation Diagnostique et Performance des Modèles

Critères d'Information (AIC/BIC)

```
# Fonction pour calculer l'AIC et le BIC sur entraînement et test
compute_aic_bic <- function(model, train_data, test_data) {
  # (1) AIC et BIC sur entraînement
  aic_train <- AIC(model)
  bic_train <- BIC(model)

  # (2) Calcul du log-vraisemblance sur test
  residuals_test <- test_data - predict(model, n.ahead = length(test_data))$pred
  sigma2_test <- var(residuals_test) # Variance des résidus
  loglik_test <- -0.5 * length(test_data) * log(2 * pi * sigma2_test) - sum(residuals_test^2) / (2 * sigma2_test)

  # Nombre de paramètres
  k <- length(coef(model))

  # AIC et BIC sur test
  aic_test <- -2 * loglik_test + 2 * k
  bic_test <- -2 * loglik_test + log(length(test_data)) * k

  return(c(AIC_train = aic_train, BIC_train = bic_train, AIC_test = aic_test, BIC_test = bic_test))
}

# Calcul des critères pour chaque modèle
stats_gdp_44 <- compute_aic_bic(model_gdp_44, train_gdp, test_gdp)
stats_gdp_34 <- compute_aic_bic(model_gdp_34, train_gdp, test_gdp)
stats_unemp_44 <- compute_aic_bic(model_unemp_44, train_unemp, test_unemp)
stats_unemp_14 <- compute_aic_bic(model_unemp_14, train_unemp, test_unemp)

# Affichage des résultats
print("AIC et BIC pour entraînement et test :")
print(data.frame(Model = c("PIB ARMA(4,4)", "PIB ARMA(3,4)", "Unemp ARMA(4,4)", "Unemp ARMA(1,4)"),
  AIC_Train = c(stats_gdp_44["AIC_train"], stats_gdp_34["AIC_train"], stats_unemp_44["AIC_train"], stats_unemp_14["AIC_train"]),
  BIC_Train = c(stats_gdp_44["BIC_train"], stats_gdp_34["BIC_train"], stats_unemp_44["BIC_train"], stats_unemp_14["BIC_train"]),
  AIC_Test = c(stats_gdp_44["AIC_test"], stats_gdp_34["AIC_test"], stats_unemp_44["AIC_test"], stats_unemp_14["AIC_test"]),
  BIC_Test = c(stats_gdp_44["BIC_test"], stats_gdp_34["BIC_test"], stats_unemp_44["BIC_test"], stats_unemp_14["BIC_test"])))
```

Série du PIB

Série du Chômage

Conclusion

Test du Ratio de Vraisemblance

```
# (2) Test du ratio de vraisemblance entre deux modèles
lr_test <- function(model1, model2) {
```

```

L1 <- logLik(model1)
L2 <- logLik(model2)
test_stat <- -2 * (L2 - L1)
df <- abs(length(coef(model1)) - length(coef(model2))) # Différence de paramètres
p_value <- 1 - pchisq(test_stat, df)
return(c(Statistic = test_stat, p_value = p_value))
}

# Comparaison seulement entre PIB_ARMA(4,4) et PIB_ARMA(3,4)
lr_gdp_44_34 <- lr_test(model_gdp_44, model_gdp_34)

# Comparaison entre Unemp_ARMA(4,4) et Unemp_ARMA(1,4)
lr_unemp_44_14 <- lr_test(model_unemp_44, model_unemp_14)

# Affichage des résultats du test du ratio de vraisemblance
print("Test du ratio de vraisemblance :")
print(data.frame(Model_1 = c("PIB_ARMA(4,4)", "Unemp_ARMA(4,4)"),
                  Model_2 = c("PIB_ARMA(3,4)", "Unemp_ARMA(1,4)"),
                  Statistic = c(lr_gdp_44_34["Statistic"], lr_unemp_44_14["Statistic"]),
                  p_value = c(lr_gdp_44_34["p_value"], lr_unemp_44_14["p_value"])))

```

Série du PIB

Série du Chômage

Conclusion

Erreurs Quadratiques Moyennes (EQM)

```

# Fonction pour évaluer les modèles
evaluer_modeles <- function(model, train, test, nom_serie, ordre) {
  fitted_values <- fitted(model)
  predictions <- predict(model, n.ahead = length(test))$pred

  data.frame(
    Série = nom_serie,
    Modèle = paste0("ARMA(", ordre[1], ",", ordre[2], ")"),
    MSE_Train = mean((train - fitted_values)^2),
    MSE_Test = mean((test - predictions)^2)
  )
}

# Évaluation des modèles PIB (sans ARMA(2,2))
results_pib <- rbind(
  evaluer_modeles(model_gdp_44, train_gdp, test_gdp, "PIB", c(4,4)),
  evaluer_modeles(model_gdp_34, train_gdp, test_gdp, "PIB", c(3,4))
)

# Évaluation des modèles Chômage
results_unemp <- rbind(
  evaluer_modeles(model_unemp_44, train_unemp, test_unemp, "Chômage", c(4,4)),
  evaluer_modeles(model_unemp_14, train_unemp, test_unemp, "Chômage", c(1,4))
)

# Combinaison des résultats

```



```

results_finaux <- rbind(results_pib, results_unemp)

# Affichage professionnel
cat("\nAnalyse Comparative des Performances Prédicatives\n")
cat("=====\n")
cat("Modèles PIB:\n")
print(subset(results_finaux, Série == "PIB"), row.names = FALSE)
cat("\nModèles Chômage:\n")
print(subset(results_finaux, Série == "Chômage"), row.names = FALSE)

```

Série du PIB

Série du Chômage

Conclusion

Recommandation finale des modèles ARMA

Série du PIB

Série du Chômage

Conclusion

Modélisation Bivariée et Corrélations Croisées

Analyse des innovations

Série du PIB

```

# Fonction personnalisée inspirée de l'algorithme des innovations (simplifiée)
compute_innovations <- function(model, series) {
  residuals <- residuals(model)
  sigma <- sqrt(model$sigma2) # Écart-type des résidus
  innovations <- residuals / sigma # Standardisation
  return(innovations)
}

# Calcul des innovations pour chaque modèle
innov_gdp <- compute_innovations(model_gdp_34, gdp_series)
innov_unemp <- compute_innovations(model_unemp_14, unemp_series)

# Visualisation des propriétés des innovations
plot(innov_gdp, type = "l", main = "Innovations du PIB")

```

Série du Chômage

```

plot(innov_unemp, type = "l", main = "Innovations du Chômage")

```

Propriétés statistiques des innovations

```

# 1. Calcul des statistiques descriptives de base
cat("=== Statistiques descriptives des innovations ===\n")

# Pour le PIB
cat("\nInnovations du PIB (ARMA(3,4)):\n")
cat("Moyenne :", round(mean(innov_gdp), 4), "\n")

```

```

cat("Variance :", round(var(innov_gdp), 4), "\n")
cat("Écart-type :", round(sd(innov_gdp), 4), "\n")

# Pour le chômage
cat("\nInnovations du chômage (ARMA(1,4)):\n")
cat("Moyenne :", round(mean(innov_unemp), 4), "\n")
cat("Variance :", round(var(innov_unemp), 4), "\n")
cat("Écart-type :", round(sd(innov_unemp), 4), "\n")

# 3. Tests d'autocorrélation (Ljung-Box)
cat("\n=== Tests d'autocorrélation (Ljung-Box) ===\n")
lags_to_test <- c(5, 10, 15, 20)

cat("\nPIB - Tests Ljung-Box:\n")
for (lag in lags_to_test) {
  test_result <- custom_ljung_box(innov_gdp, lag)
  cat(paste0("Retard ", lag, ": p-value = ", round(test_result$p_value, 4), "\n"))
}

cat("\nChômage - Tests Ljung-Box:\n")
for (lag in lags_to_test) {
  test_result <- custom_ljung_box(innov_unemp, lag)
  cat(paste0("Retard ", lag, ": p-value = ", round(test_result$p_value, 4), "\n"))
}

```

```

# 4. Visualisation des ACF
par(mfrow = c(1, 2))
acf(innov_gdp, main = "ACF des innovations du PIB")
acf(innov_unemp, main = "ACF des innovations du chômage")
par(mfrow = c(1, 1))

```

Série du PIB

Série du Chômage

Conclusion

Corrélation croisée (CCF) entre les innovations

```

# Fonction my_ccf : Calcul et visualisation des corrélations croisées
my_ccf <- function(x, y, lag.max = 20, alpha = 0.05, plot = TRUE) {
  n <- length(x)
  if (length(y) != n) stop("Les séries doivent avoir la même longueur")

  # Standardisation des séries
  x_std <- (x - mean(x)) / sd(x)
  y_std <- (y - mean(y)) / sd(y)

  # Calcul des corrélations croisées
  ccf_values <- numeric(2 * lag.max + 1)
  lags <- -lag.max:lag.max

  for (i in seq_along(lags)) {
    lag <- lags[i]

```

```

    if (lag <= 0) {
      ccf_values[i] <- sum(x_std[1:(n + lag)] * y_std[(1 - lag):n]) / (n + lag)
    } else {
      ccf_values[i] <- sum(x_std[(1 + lag):n] * y_std[1:(n - lag)]) / (n - lag)
    }
  }

  # Calcul des intervalles de confiance
  z_critical <- qnorm(1 - alpha/2)
  conf_interval <- z_critical / sqrt(n)

  # Visualisation
  if (plot) {
    plot(lags, ccf_values, type = "h",
         main = "Corrélation croisée entre le PIB et le taux de chômage",
         xlab = "Décalage (lags)", ylab = "Corrélation",
         ylim = c(-1, 1), lwd = 2, col = "blue")
    abline(h = c(-conf_interval, conf_interval),
           lty = 2, col = "red")
    abline(h = 0, col = "black")
    grid()
  }

  # Retourne les résultats sous forme de dataframe
  data.frame(
    lag = lags,
    correlation = round(ccf_values, 4),
    significant = abs(ccf_values) > conf_interval
  )
}

ccf_results <- my_ccf(innov_gdp, innov_unemp, lag.max = 20)
#print(ccf_results)

```

Test de Ljung-Box pour les résidus croisés

```
custom_ljung_box(innov_gdp * innov_unemp, 20) # p-value = 0.12 + non corrélés
```

Conclusion