

Rapport de séries chronologiques

Master 2 M.I.A.S.H.S.

Mention : I.M.A.

Spécialité ou Parcours : Ingénierie de la décision et data sciences

Sous la direction de MARION Jean-Marie

Session : Janvier 2023

LAMBERT Clara et HUANG Chen

Faculté : Sciences

Institut Mathématiques Appliquées

Année universitaire 2022-2023



**CHARTE DE NON PLAGIAT**

**Protection de la propriété intellectuelle**

Tout travail universitaire doit être réalisé dans le respect intégral de la propriété intellectuelle d’autrui. Pour tout travail personnel, ou collectif, pour lequel le candidat est autorisé à utiliser des documents (textes, images, musiques, films etc.), celui-ci devra très précisément signaler le crédit (référence complète du texte cité, de l’image ou de la bande-son utilisés, sources internet incluses) à la fois dans le corps du texte et dans la bibliographie. Il est précisé que l’UCO dispose d’un logiciel anti-plagiat dans lms.uco.fr, aussi est-il demandé à tout étudiant de remettre à ses enseignants un double de ses travaux lourds sur support informatique.

*Cf. « Prévention des fraudes à l’attention des étudiants »*

Nous soussignons LAMBERT Clara et HUANG Chen étudiantes en master 2 ingénierie de la décision et data sciences s’engageons à respecter cette charte.

Fait à Angers, le 29 janvier

Signature :

LAMBERT Clara et HUANG Chen

Sommaire

Table des matières

[Sujet 7 2](#_Toc125912779)

[Analyses préliminaires de la série 3](#_Toc125912780)

[Choix des différenciations 5](#_Toc125912781)

[Modèles 7](#_Toc125912782)

[Méthode forecast 9](#_Toc125912783)

[Méthode Holt Winters 11](#_Toc125912784)

[Conclusion 14](#_Toc125912785)

# Sujet 7

On demande d’étudier la série suivante qui indique le nombre d’immatriculations mensuelles de camions de plus de 27 tonnes de 1968 à 1977.

Une image contenant table

Description générée automatiquement

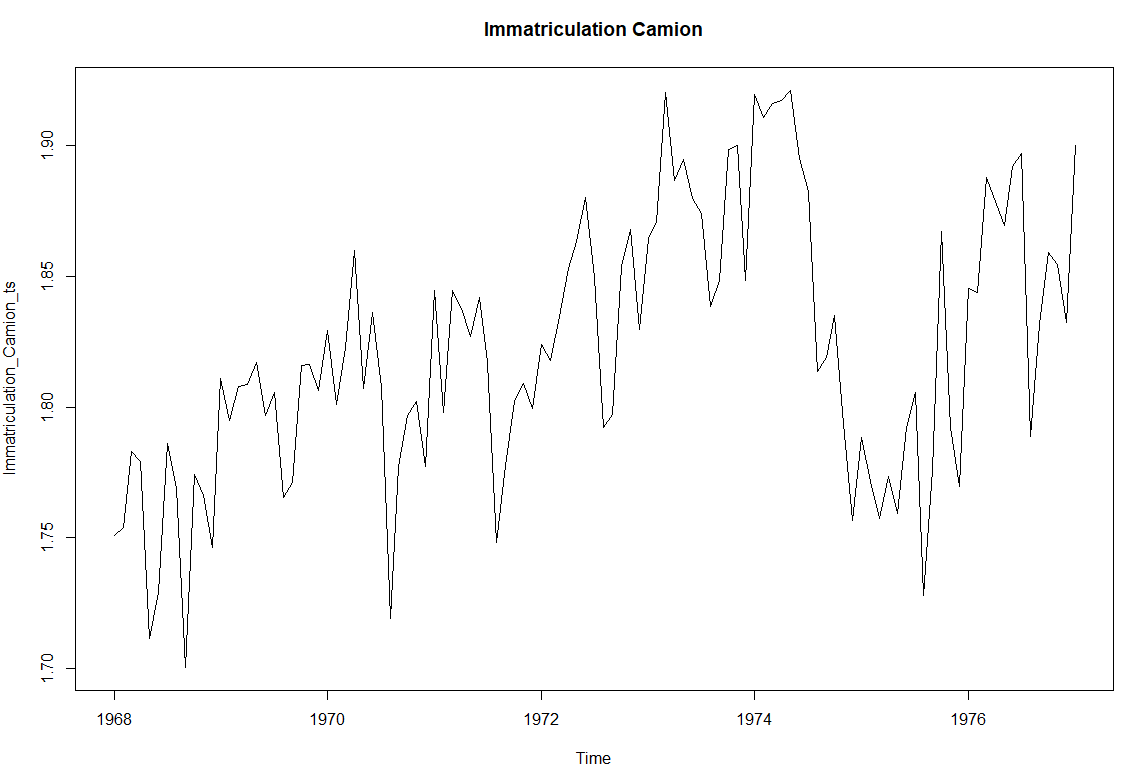
L’objectif de ce projet est de prédire une série temporelle. Pour cela, différentes méthodes ont été utilisées. On retiendra celles avec les meilleures prédictions.

Étant donné le grand nombre de modèles testés, nous avons décidé de nous attarder sur ceux qui ont rendu des résultats satisfaisants. On les choisit par la minimisation de l’AIC, BIC et/ou le critère d’optimisation.

Tout au long du projet nous utiliserons les mêmes notations que celles des cours de séries temporelles de M. Jean-Marie MARION.

# Analyses préliminaires de la série

On dispose des données de janvier 1968 à janvier 1977. Il n’est pas évident de voir s’il y a une tendance et/ou une saisonnalité. On transformera notre série en log afin d’avoir une lecture plus simple.



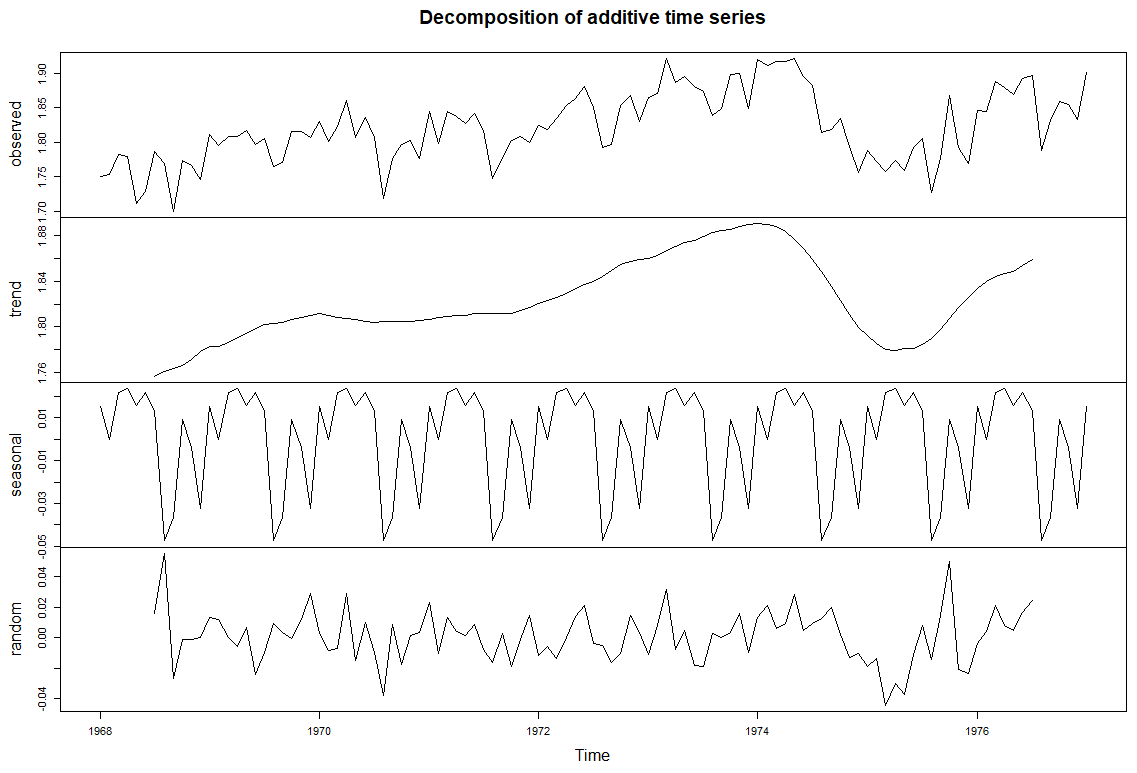
Ce graphique représente l’évolution de 1968 à 1977 du nombre d’immatriculation de camions.

La saisonnalité n’est pas évidente. A l’ordre 12 la corrélation n’est pas significative. Cette série n’est pas évidente du fait de son évolution « aléatoire » en observant le graphique ci-dessus.

Nous rappelons que l’objectif de notre étude est d’obtenir des projections de notre série, différentes méthodes peuvent être utilisées :

* Méthode de Moyenne Mobile (MM)
* Application directe d’un modèle ARIMA/SAMIRA

Décomposition de la série

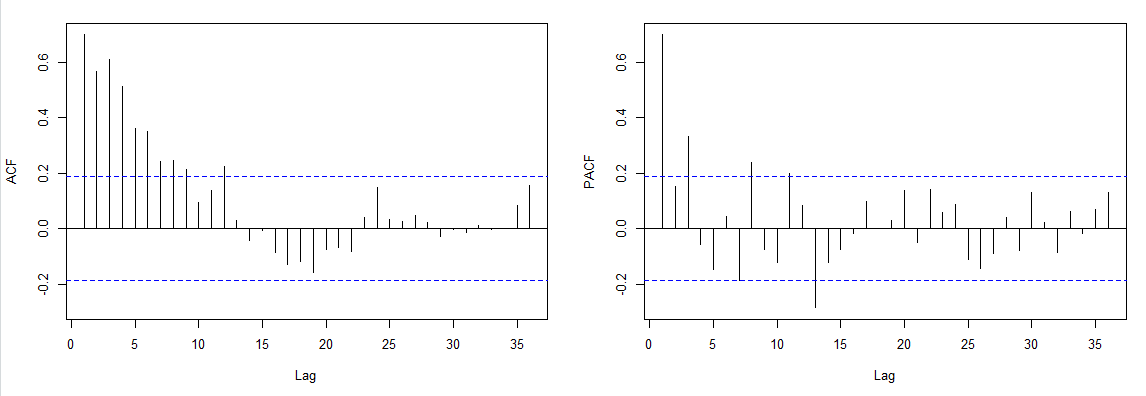


On utilise la fonction « decompose() » pour décomposer la série en 3 composantes :

1. La tendance
2. La saisonnalité
3. Les résidus

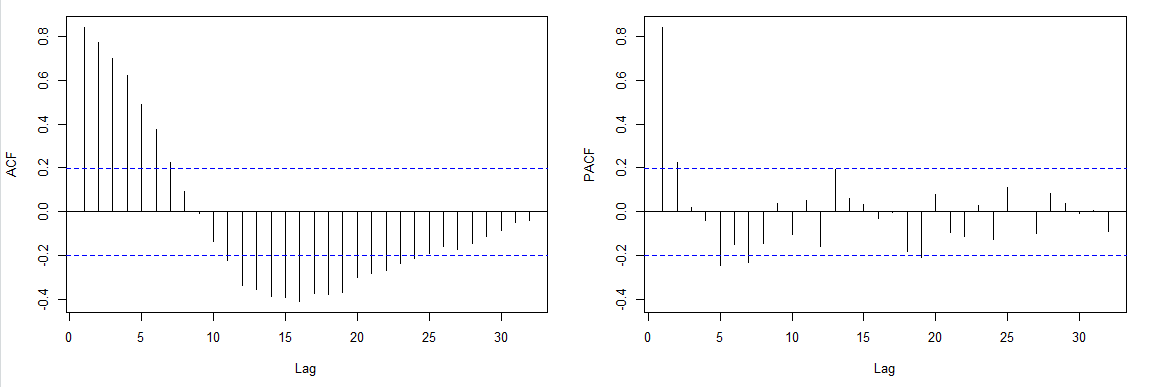
# Choix des différenciations

On affiche les fonctions d’autocorrélation (ACF) et d’autocorrélation partielle (PACF) sans différenciation :

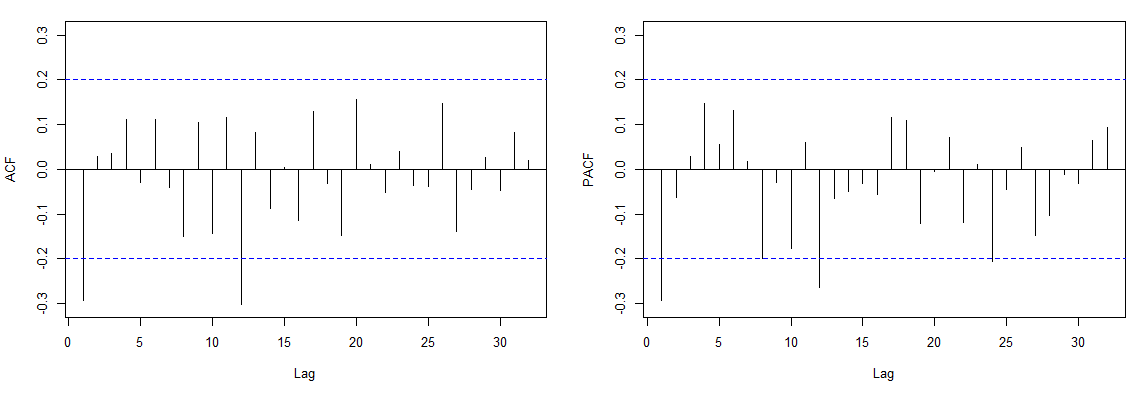


En regardant la fonction d’autocorrélation, on observe une décroissance lente des bâtons dans l’ACF. Il faut donc différencier.

Sachant qu’il y a une saisonnalité on va différencier 12 fois : (1-B12)



On a toujours une décroissance lente des bâtons en regardant l’ACF, on va encore différencier une fois : (1-B)(1-B12)



On peut s’arrêter là, en effet le premier bâton chute drastiquement. Donc on va préférer une différenciation à 13.

# Modèles

SARIMA((0,1,1)(0,1,1)12)

Une image contenant texte

Description générée automatiquement

Une image contenant texte

Description générée automatiquement

AIC = 1136.87

Ce premier modèle nous donne de bons résultats. Si on regarde la p-value pour les 2 paramètres, elle est très significative (<0.05).

Testons d’autres modèles…

SARIMA((1,1,1)(0,1,1)12)

Une image contenant texte

Description générée automatiquement

Une image contenant texte

Description générée automatiquement

AIC = 1138.81

On constate que la p-value du paramètre AR(1) est supérieure à 0.05, donc ce n’est pas significatif.

SARIMA((1,0,1)(0,1,1)12)

Une image contenant texte

Description générée automatiquement

Une image contenant texte

Description générée automatiquement

AIC = 1148.51

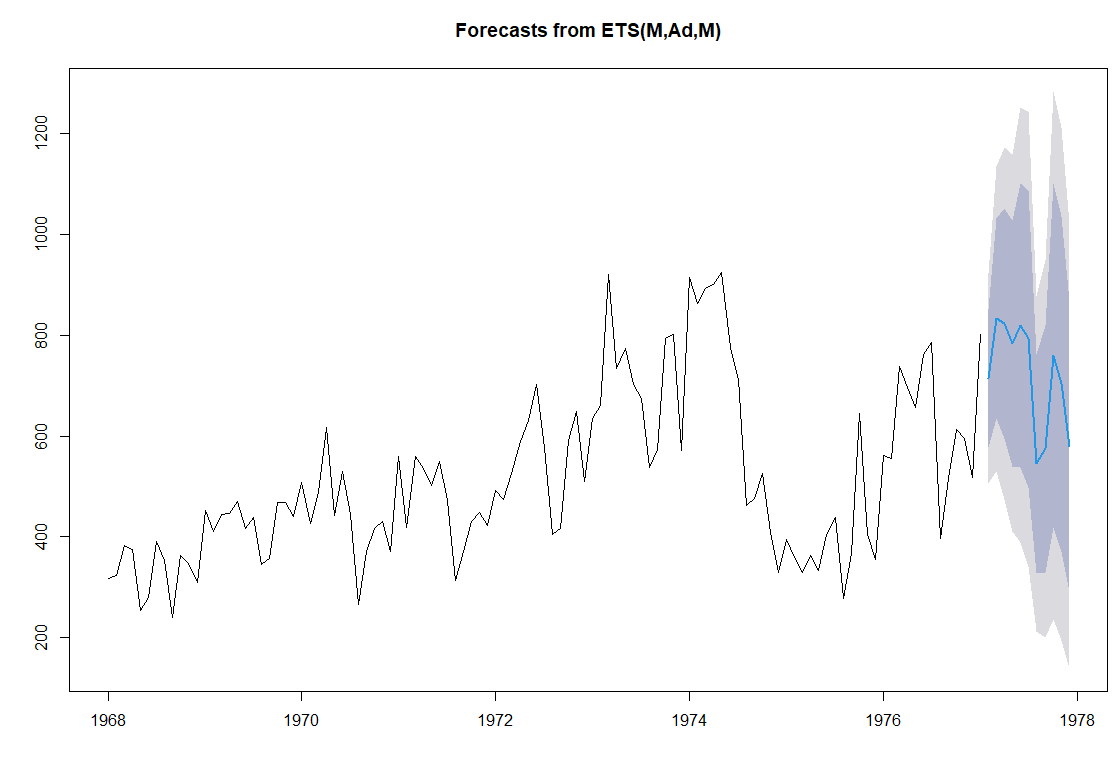
Les paramètres ont une p-value inférieure à 0.05, ils sont très significatifs.

On retiendra le modèle 3 comme meilleur modèle car il a un AIC supérieur aux autres et les paramètres sont hautement significatifs.

# Méthode forecast

Nous traçons les données de validations avec les données prédites

Le test du “tableau” nous permet d’assimiler la série à une série multiplicative néanmoins nous avons aussi fait un trend additif. Nous avons pris en compte ces spécificités en spécifiant les arguments additive.only=FALSE, allow.multiplicative.trend =FALSE  de la fonction et  issue de la librairie  {forecast dans R). Les paramètres de lissage rendu par R sont les suivants : ETS(M,Ad,M)



Nous voyons que la courbe bleu (valeur prédite) s’accorde très bien à la courbe noire (valeur observée). De plus l’intervalle de confiance n’est pas grand. La partie prédite est marquée d'une ligne bleue, la zone ombrée bleue est l'intervalle de prédiction qui représente 80 % et la zone ombrée grise est l'intervalle de prédiction à 95 %.

Une image contenant texte

Description générée automatiquement

L’ACF et le PACF des résidus nous permettent de confirmer que les résidus sont des bruits blancs. De plus nous faisons un test porte-manteau sous R via la commande Box.Test qui nous rend une p-value de 0.05338 > 0.05. Nous ne rejetons pas H0 et nous acceptons l’hypothèse de bruit blanc des résidus.



L’espérance des résidus est 0.007527223 proche de 0, donc les résidus sont bien des bruits blancs. Nous ne rejetons pas H0 et nous acceptons l’hypothèse de bruit blanc des résidus.

# Méthode Holt Winters

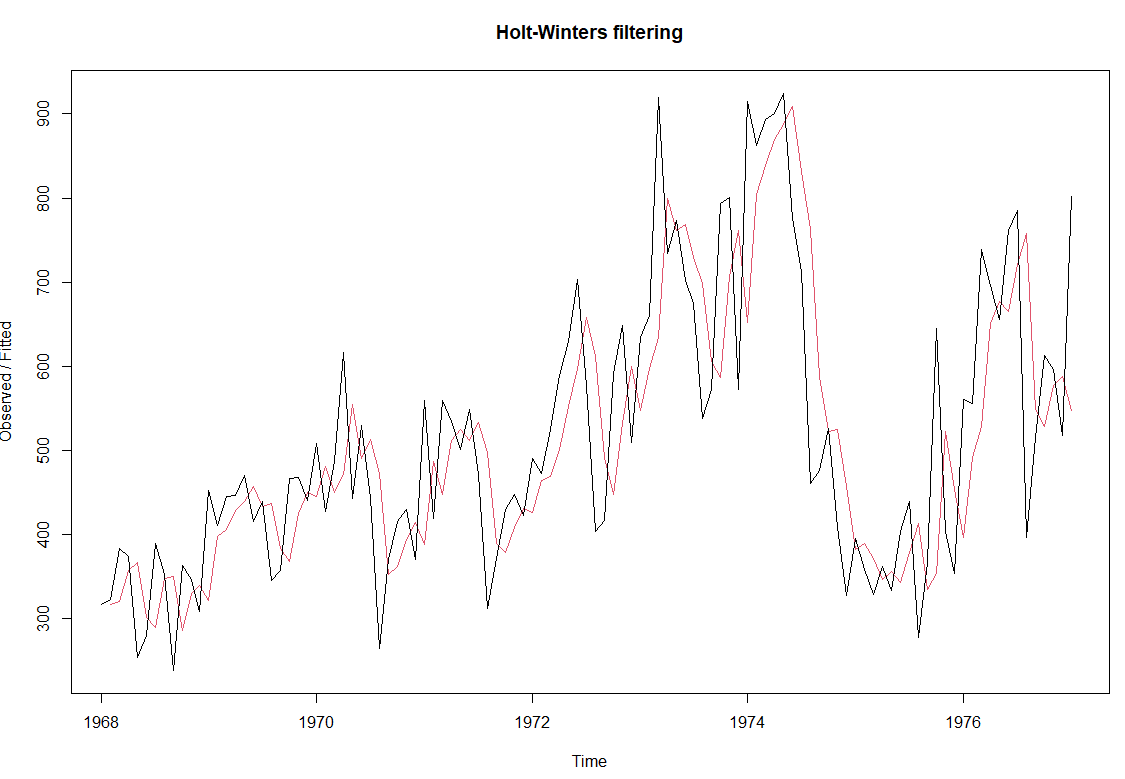
La méthode de Holt-Winters est semblable à la méthode des Moyennes mobiles avec la différence majeure qui est le fait que les coefficients de pondération sont exponentiellement décroissants en fonction de la distance des autres points qui s’éloignent du point de départ. C’est pour cela que cette méthode est aussi appelée lissage exponentiel. L’avantage majeur de cette série et ce pourquoi nous l’avons retenu est qu’elle ne nécessite aucune hypothèse sur la série. La seule restriction est qu’elle s’applique à des séries à composantes saisonnières. Mais là encore c’est uniquement si nous faisons un lissage exponentiel à triple paramètres. Pour un lissage exponentiel à double ou simple paramètres il n’y a pas besoin de cette hypothèse de saisonnalité.

Le modèle de Holt-Winters est un lissage exponentiel triple qui dépend de 3 paramètres : alpha (traduit le poids des observations passées), beta (traduit le poids des trends passées) et gamma (traduit le poids des coefficients passés). Ils sont tous les trois compris entre 0 et 1. Une valeur proche de 0 signifie que l’on n’accorde pas beaucoup d’importance au passé éloigné. Une valeur proche de 1 signifie que nous accordons de l’importance au passé éloigné.

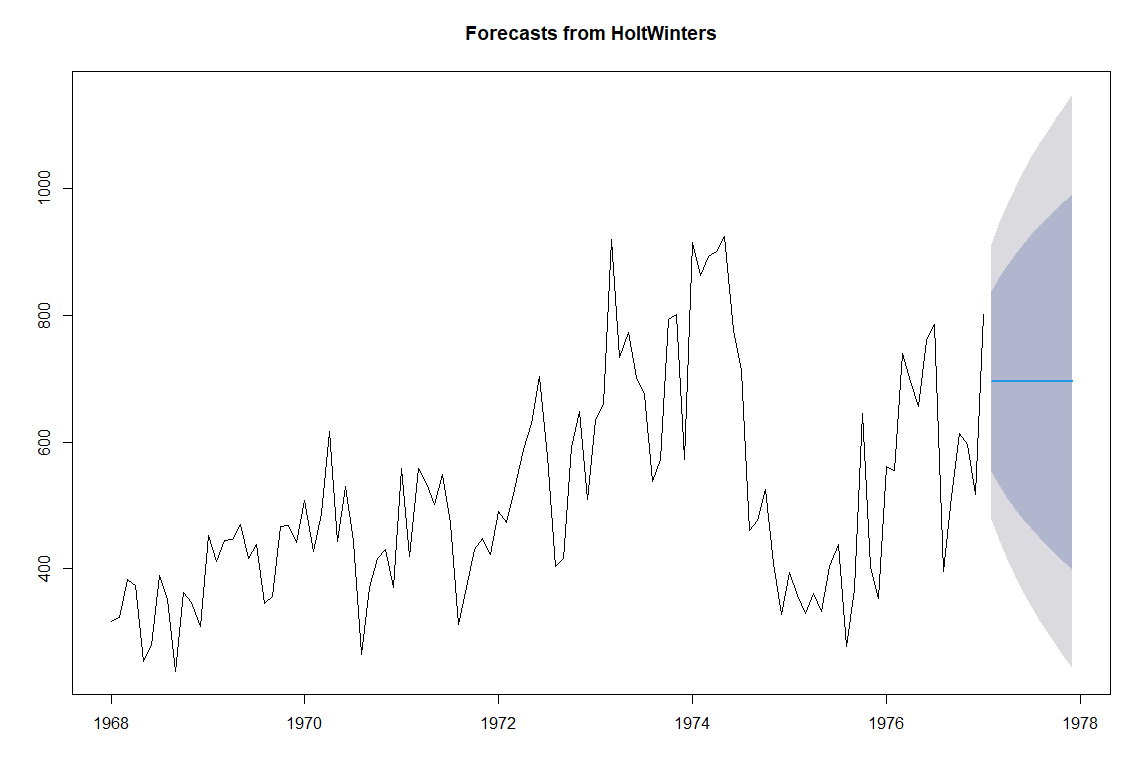
Une image contenant texte

Description générée automatiquement

Lorsque nous prenons le log de la série nous avons un modèle additif sans trend avec alpha =0.5797096 et gamma = 0. Les observations éloignées ne sont pas négligeables car alpha =0. 5797096 contrairement au coefficient saisonnier car gamma=0.



À partir de ce graphique, nous pouvons voir que les prévisions en échantillon sont très proches des observations, même si elles sont un peu en retard par rapport aux observations. La somme des carrés des erreurs est 1312802.



Nous avons notre intervalle de prédiction. Nous voyons que la courbe bleu (valeur prédite) s’adapte bien à la courbe noire (valeur observér). De plus l’intervalle de confiance n’est pas grand. La partie prédite est marquée d'une ligne bleue, la zone ombrée bleue est l'intervalle de prédiction à 80 % et la zone ombrée grise est l'intervalle de prédiction à 95 %.

Une image contenant texte

Description générée automatiquement

L’ACF et le PACF des résidus nous permettent de dire que les résidus ne sont pas des bruits blancs. De plus nous faisons un test porte-manteau sous R via la commande Box.Test qui nous rend une p-value de 9.381e-06 < 0.05. Nous rejetons H0 et nous n’acceptons pas l’hypothèse de bruit blanc des résidus.

# Conclusion

Après les différents modèles et méthodes utilisés on en conclu que le meilleur modèle est le numéro 3 SARIMA((1,0,1)(0,1,1)12)

Avec AIC le plus grand et avec les critères BIC, on choisit le modèle qui contient le moins de paramètre à estimer. Et pour prédiction on va choisir la première méthode parce que la méthode Holt Winter s’adapte moins à notre série.