

TP1 Séries chronologiques

Mamadou Lamine DIAMBAN

08 Octobre 2019

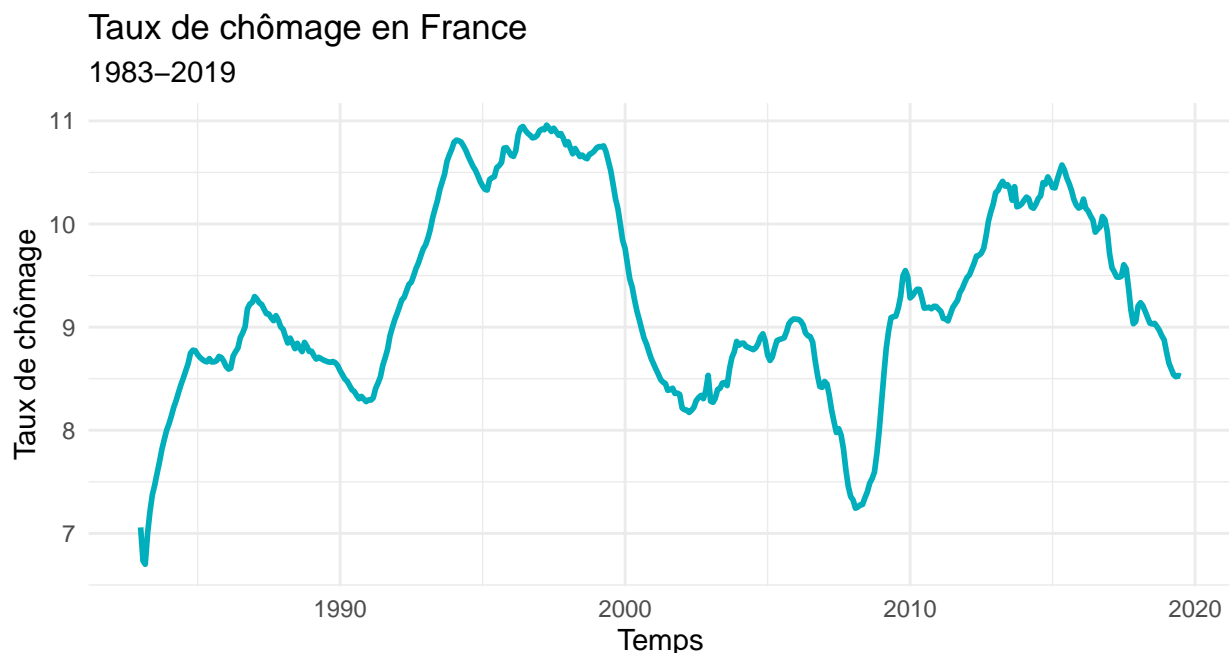
Ce rapport rentre dans le cadre de l'UE Séries temporelles du Master 2 Statistique et Science des Données à l'UGA qui est ici soumis sous forme de travaux pratiques.

Elle comprend 7 parties:

- Les deux premières parties **Etude du taux de chômage en France à partir de 1983** et **Etude du taux d'inflation en France à partir de 1997** visent à se familiariser aux techniques de lissage et à la prévision sur des données de séries temporelles.
- La troisième partie: **Etude d'un marché alimentaire** permet essentiellement la mise en pratique de *dessaisonalisation* d'une série temporelle.
- Enfin les quatre dernières parties s'appuient sur les jeux de données de la bibliothèque TSA. Elles nous permettent de voir les tendances de différentes séries chronologiques saisonnières et non stationnaires. Mais aussi de réaliser des régressions linéaires par la méthode des moindres carrés afin de pouvoir adapter le modèle adéquat à une série temporelle donnée.

1. Taux de chômage à partir de 1983

a) Tracer le graphique de cette série



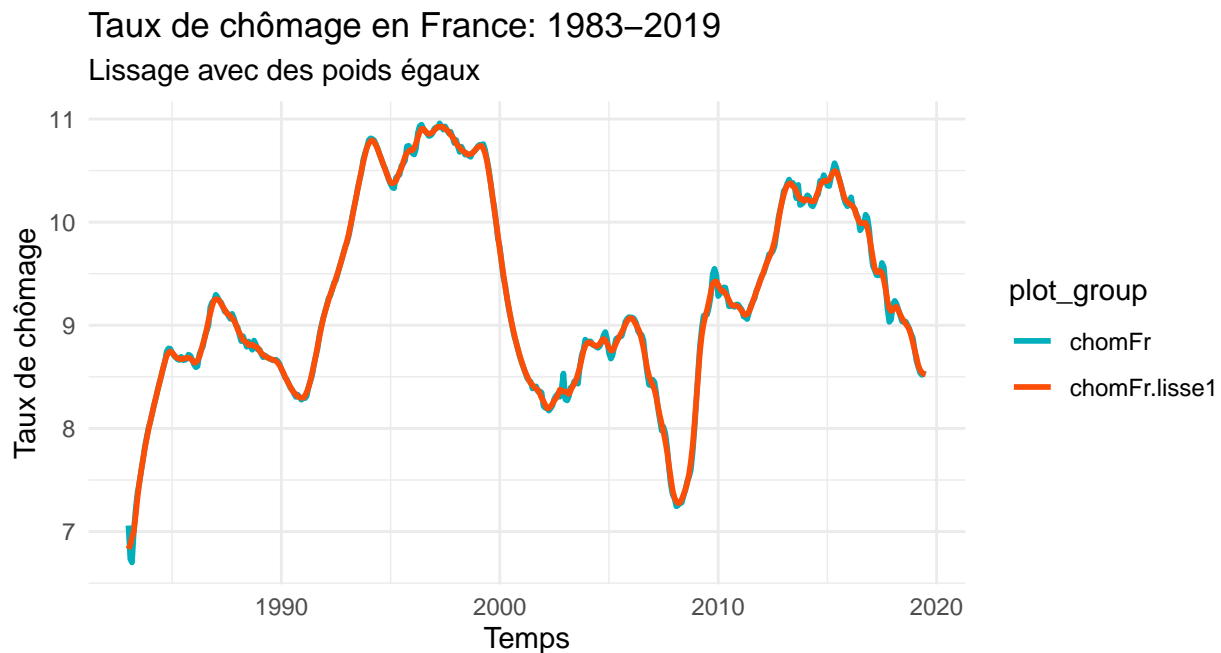
Ce chronogramme montre l'existence d'une tendance de long terme qui se déploie sur des périodes plus longues et d'un cycle lié à la politique économique. Ainsi, il apparaît que la hausse importante du taux de chômage dans les années 1980 résulte du choc pétrolier de cette époque qui s'en est suivi d'une politique monétaire désinflationniste. De même la hausse de la première moitié des années 1990 et de la fin des années 2000 prend place dans un contexte conjoncturel défavorable.

Lissez cette série en y ajustant une moyenne mobile de rayon $r = 2$, avec des poids égaux.

Etant donné que la moyenne de plusieurs observations est plus stable qu'une observation individuelle. Nous avons créé une nouvelle série temporelle basée sur la moyenne mobile. Celle-ci qui permet d'éliminer d'éventuelles fluctuations contenues dans la série d'origine.

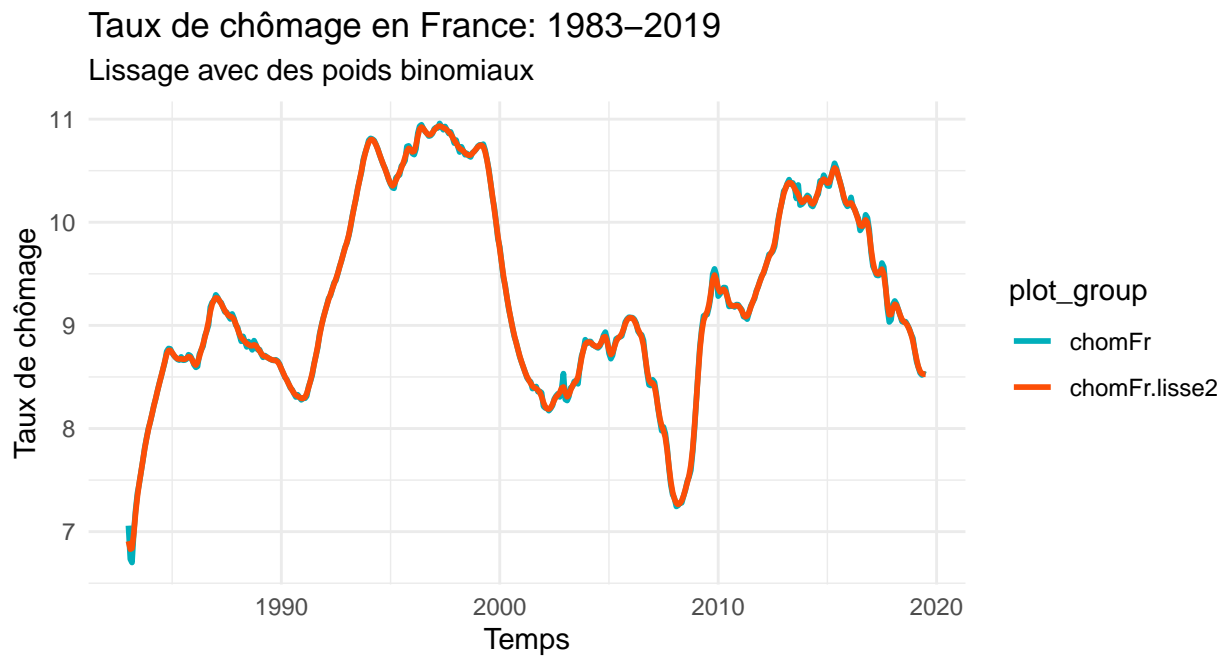
Cette nouvelle série est alors:

$$\bar{x}_t = \frac{\sum_{i=-r}^r c_i x_{t+i}}{\sum_{i=-r}^r c_i}$$



Il y'a peu de différences entre le lissage à poids équiprobable et la série d'origine.

c) Lissez la même série en y ajustant une moyenne mobile de rayon $r = 2$, avec poids binomiaux.



De même, la différence entre le lissage à poids binomiaux et la série d'origine n'est guère notable. Les valeurs entre les séries sont presque égales pour un temps donné.

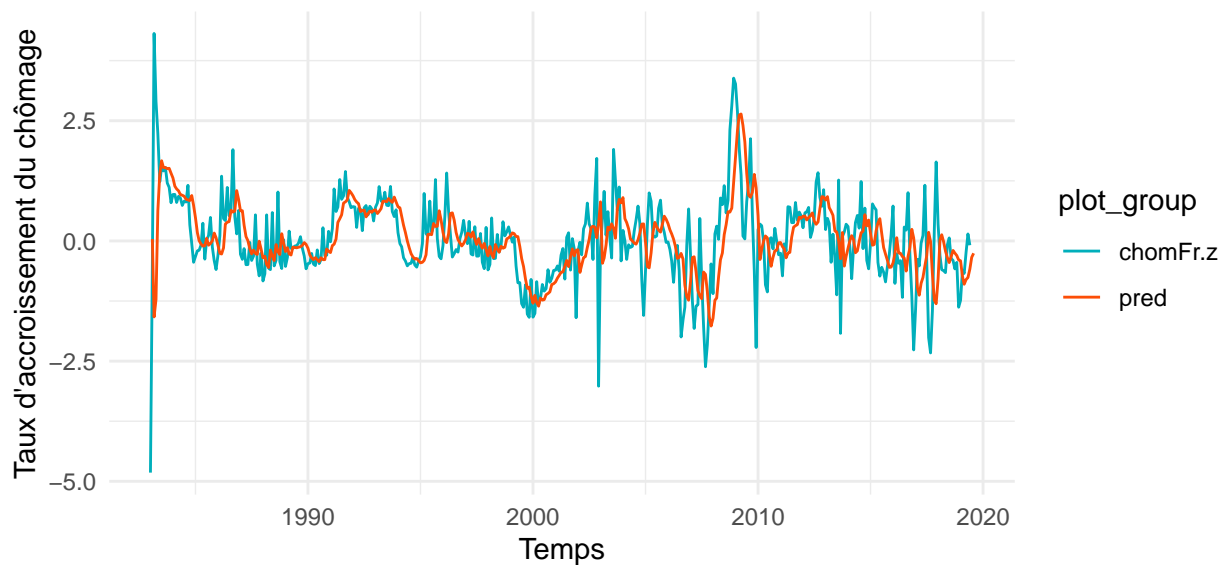
d) Appliquez à cette série un lissage exponentiel avec $\alpha = 2/3$

Afin de réaliser un lissage exponentiel, nous nous sommes d'abord intéressé à l'accroissement relatif du taux de chômage en France qui est donnée par:

$$x_t = \frac{I_{t+1} - I_t}{I_t}$$

Taux d'accroissement du chômage en France: 1983–2019

Lissage exponentiel avec $\alpha = 2/3$



Parmi les différents lissages utilisés, le lissage dit “exponentiel” paraît plus adapté. En effet grâce à la différenciation utilisée sur le lissage exponentiel, le problème de stabilité de la tendance noté sur la figure 1 s’annule et laisse place à une série plus “stationnaire” où l’on voit l’écart entre les deux courbes qui met en évidence l’existence d’erreurs de mesure de l’ordre d’environ 3 points en 2003 et de 1 point en 2010.

e) Trouvez, au dixième près, la valeur de α minimisant la SEQ.

seq	alpha
227.1	0.6

En prenant un $\alpha \in \{0.6, 0.95\}$ avec un pas de 0.1, nous constatons que la somme des erreurs quadratique $SEQ(\alpha) = \sum_{t=1}^n (z_t - \hat{z}_t)^2 = 227.1$ est minimale lorsque $\alpha = 0.6$

f) Donnez la prévision pour août 2019.

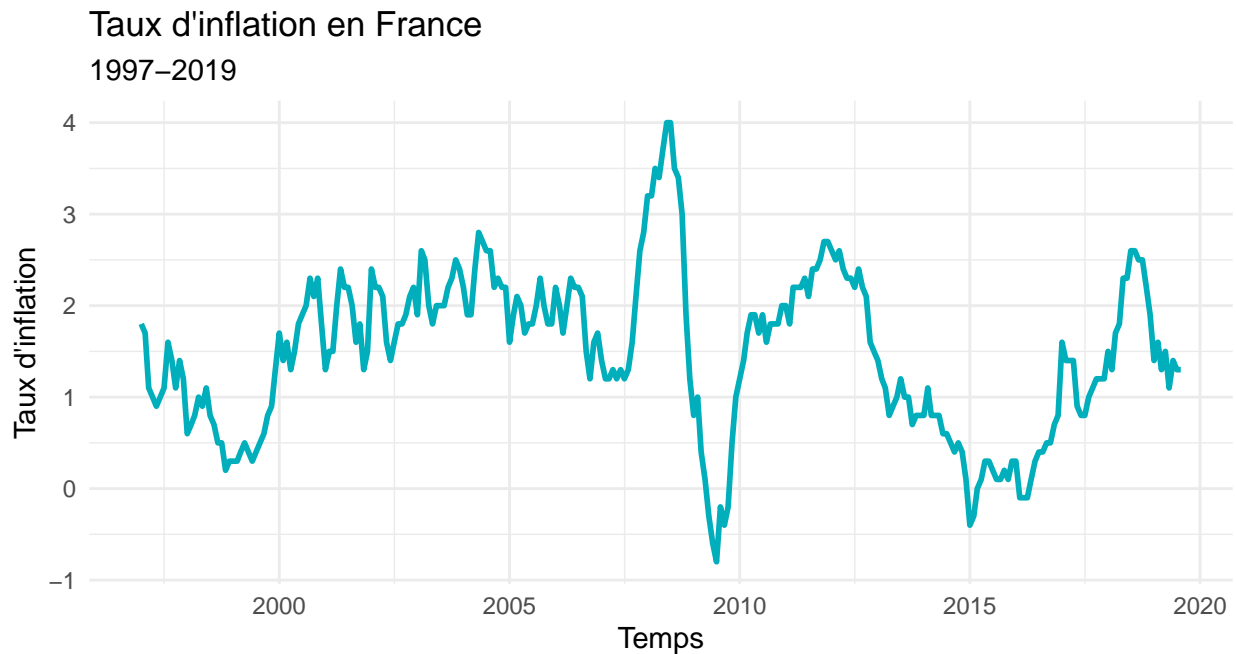
TABLE 2 – Prédiction Août

x
8.566853

La prévision du mois d’Août est de 8.56. Ce qui correspond à une baisse de 0.48% par rapport à la précédente année.

2. Taux d'inflation en France à partir de 1997

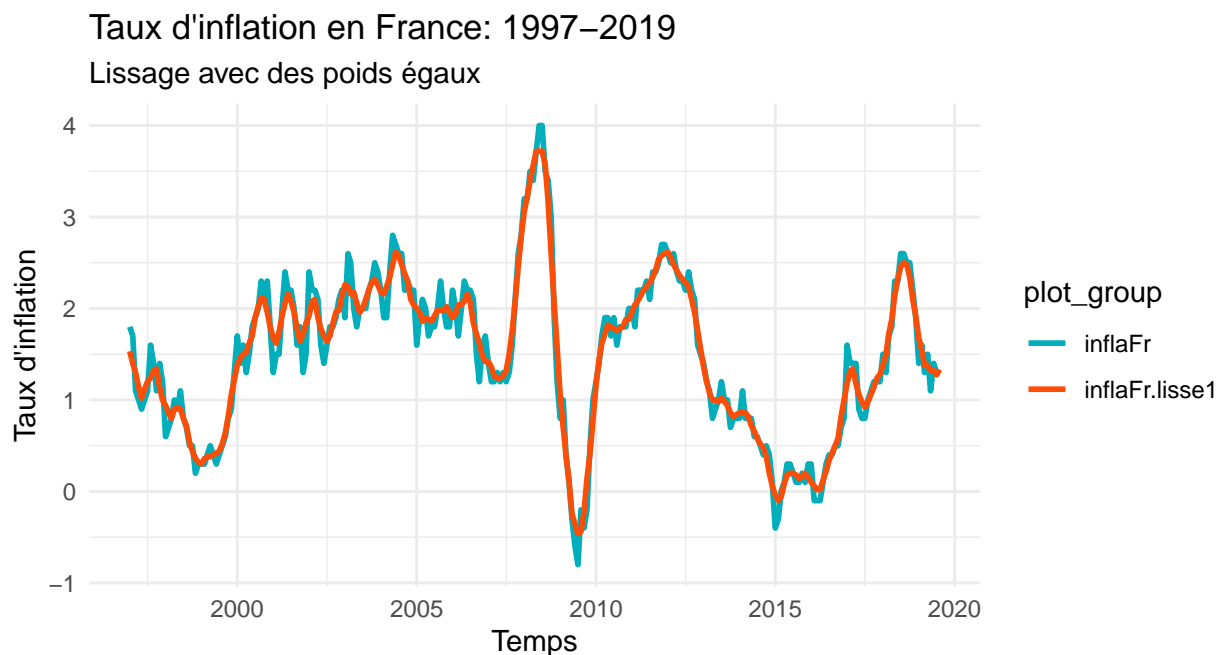
a) Tracer le graphique de cette série



b) Lissez cette série en y ajustant une moyenne mobile de rayon $r = 2$, avec des poids égaux.

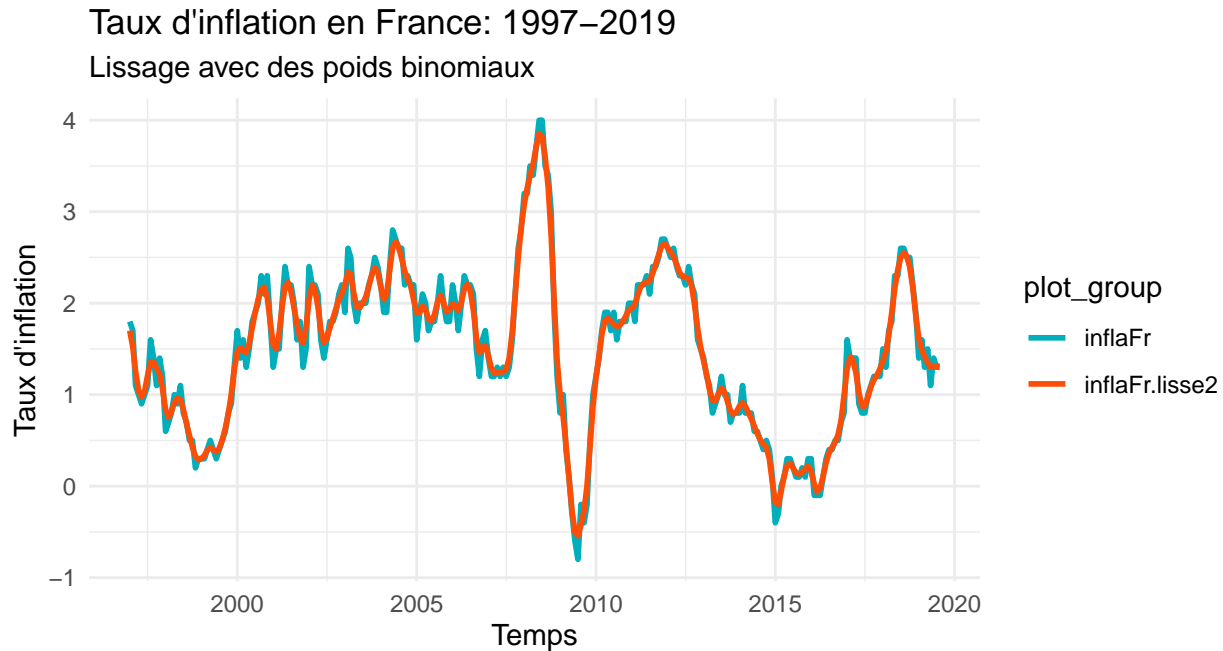
Regardons maintenant si le choix des poids affecte les résultats obtenus (on sait qu'en théorie cela doit être le cas).

Examinons d'abord le lissage d'une moyenne mobile avec des poids égaux.



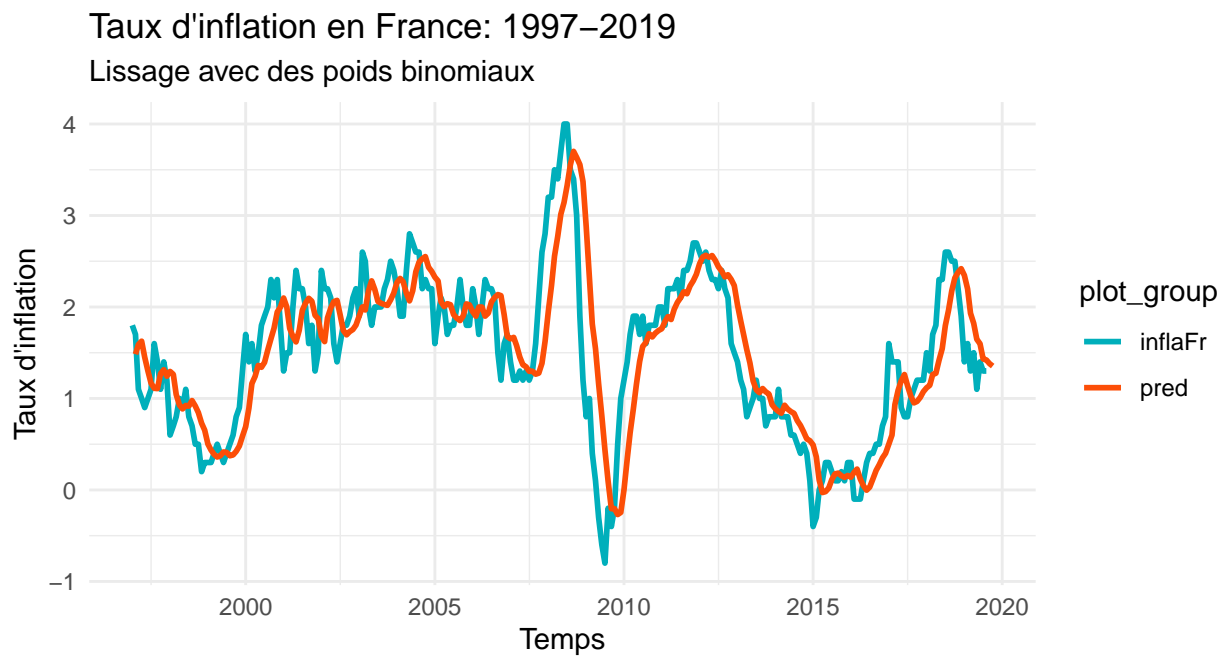
Au niveau du graphique, le résultat obtenu sur le lissage avec équipondération est plus lisse que celui avec les poids standards notamment entre les années 2000 et 2008.

c) Lissez la même série en y ajustant une moyenne mobile de rayon $r = 2$, avec poids binomiaux.



On constate la même remarque que sur le graphique précédent.

d) Appliquez à cette série un lissage exponentiel avec $\alpha = 2/3$



Comme constaté sur les données du taux de chômage en France, le lissage exponentiel semble être plus adapté que les deux lissages ci-dessus. On voit une certaine différence entre les observations individuelles et la moyenne mobile.

e) Trouvez, au dixième près, la valeur de alpha minimisant la SEQ.

seq	alpha
38.93	0.6

En prenant un $\alpha \in \{0.6, 0.95\}$ avec un pas de 0.1, nous constatons que la somme des erreurs quadratique $SEQ(\alpha) = \sum_{t=1}^n (z_t - \hat{z}_t)^2 = 38.93$ est minimale lorsque $\alpha = 0.6$

f) Donnez la prévision pour août 2019.

TABLE 4 – Prédiction Août

x
1.348175

La prévision du mois d'Août est de 1.35.

3. Marché d'alimentation

(a) En ne considérant que les jours où le commerce est ouvert, déterminez les moyennes x_j correspondant à chacun de ces 6 jours.

TABLE 5 – Moyennes de chacun des 6 jours

	x
Lundi	117.8
Mardi	80.6
Merc	95.6
Jeudi	177.0
Vend	326.5
Same	325.0

On remarque que les ventes sont deux fois plus importantes en fin de semaine qu'en milieu de semaine.

(b) Quelle est la moyenne générale s'appliquant aux jours ouvrables

TABLE 6 – Moyenne générale

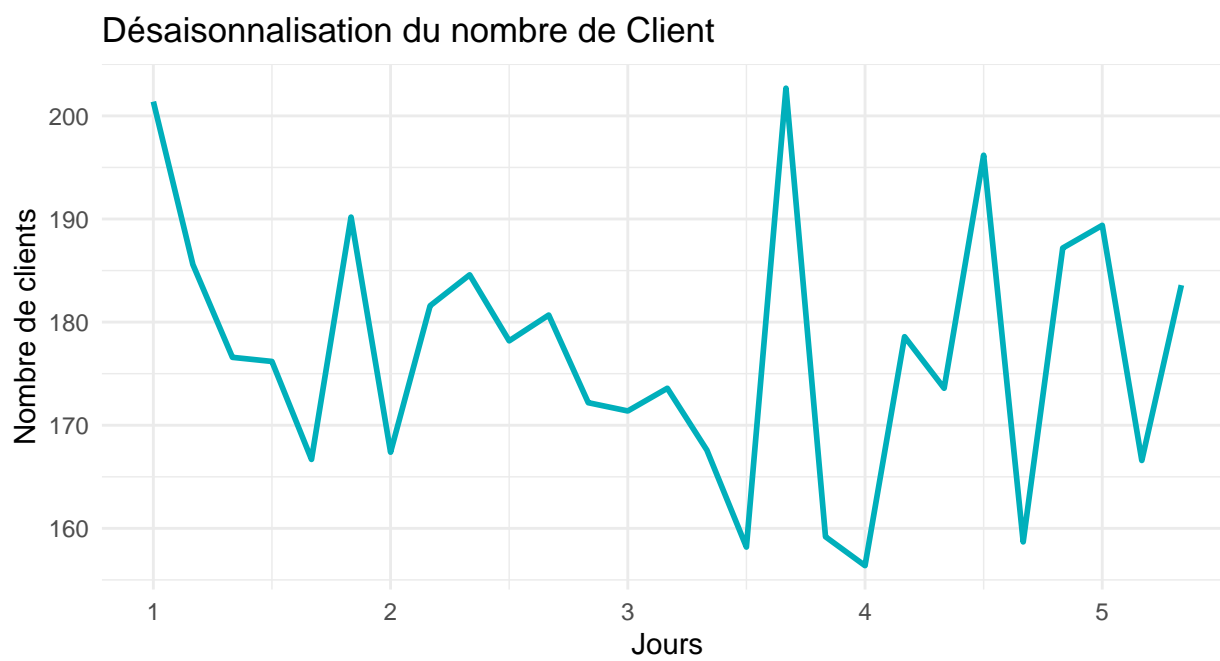
x
177.1852

(c) Quelle est l'effet de phase correspondant à chacun de ces 6 jours?

TABLE 7 – Effet de phase

	x
Lundi	-59.3851852
Mardi	-96.5851852
Merc	-81.5851852
Jeudi	-0.1851852
Vend	149.3148148
Same	147.8148148

(d) Désaisonnalisiez (selon le jour de la semaine) les 27 données du tableau.



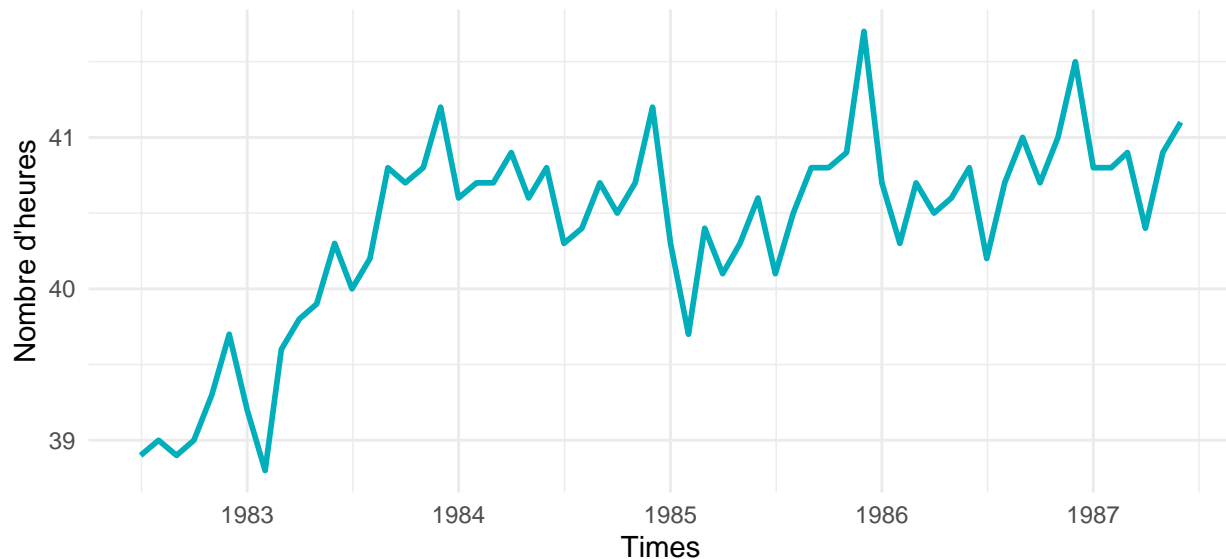
4. Séries heures

La série **hours** donne la moyenne mensuelle des heures travaillées par semaine dans le secteur manufacturier des États-Unis pour juillet 1982 à juin 1987.

a) Tracer et interpréter le graphique de cette série temporelle.

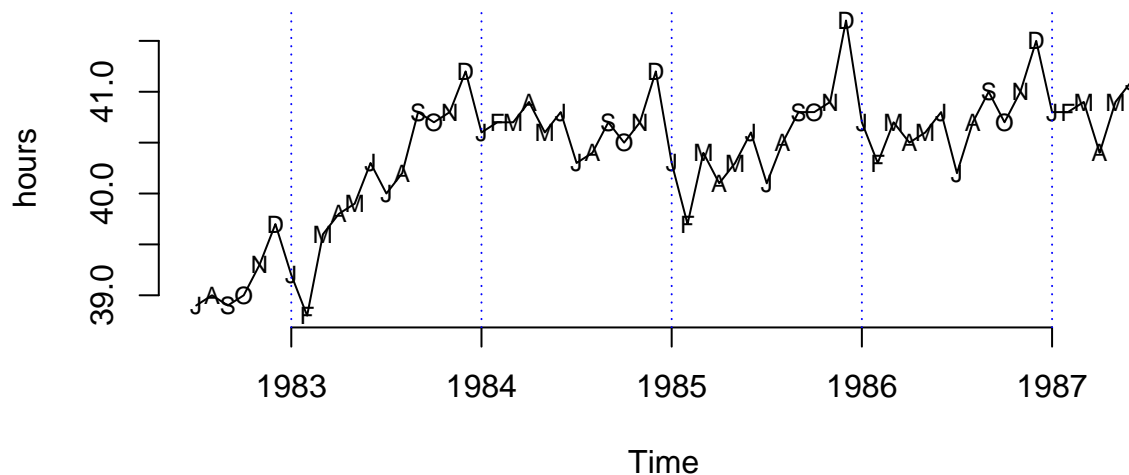
Heures travaillées par semaine

Secteur Manufacturier aux USA



Ce graphique montre la présence d'une série temporelle avec une tendance croissante non stationnaire. On note une augmentation de plus de 2 points entre le début et la fin des années 1983. Cette augmentation reste plutôt stable jusqu'à la fin des années 1984 où l'on observe une décroissance de 1.5 points. S'en suit à la fin des années 1985 un pic maximal de 41.7 heure par semaine puis une stabilité du nombre d'heures par semaine qui tourne autour de 40.5 dans les années suivantes.

b) Tracer et interpréter le graphique de cette série temporelle en identifiant chaque mois

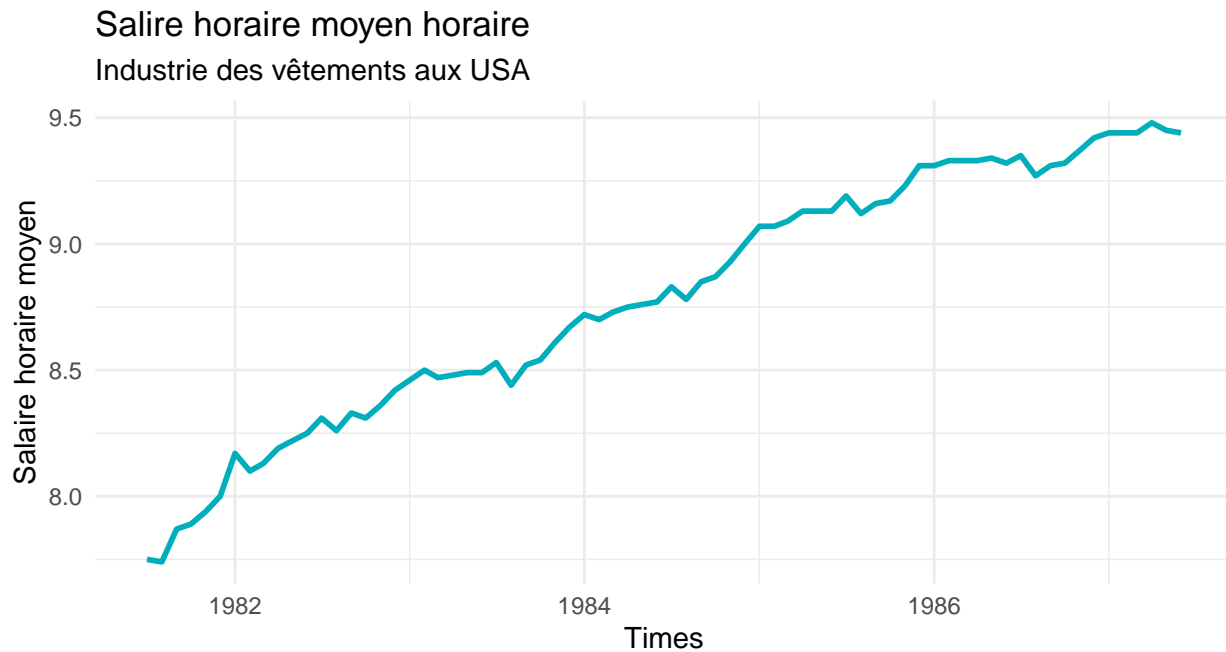


L'identification des mois nous permet de mieux voir la saisonnalité du chronogramme. En surplus, on peut dire que le nombre d'heures travaillées par semaine dans le secteur manufacturier des Etats-Unis connaît une hausse entre Octobre et Décembre puis une baisse plus ou moins significative entre Janvier et Février.

5. Série wages

La série **wages** donne les valeurs mensuelles du salaire horaire moyen (en dollars) pour les travailleurs de l'industrie des vêtements et des produits textiles aux États-Unis de juillet 1981 à juin 1987.

a) Tracer et interpréter cette série temporelle.



On note une croissance du salaire horaire moyen, qui s'accélère à la fin de chaque mois d'Août. Ceci nous permet de confirmer la saisonnalité de la série.

b) Utiliser les moindres carrés pour ajuster à une tendance linéaire

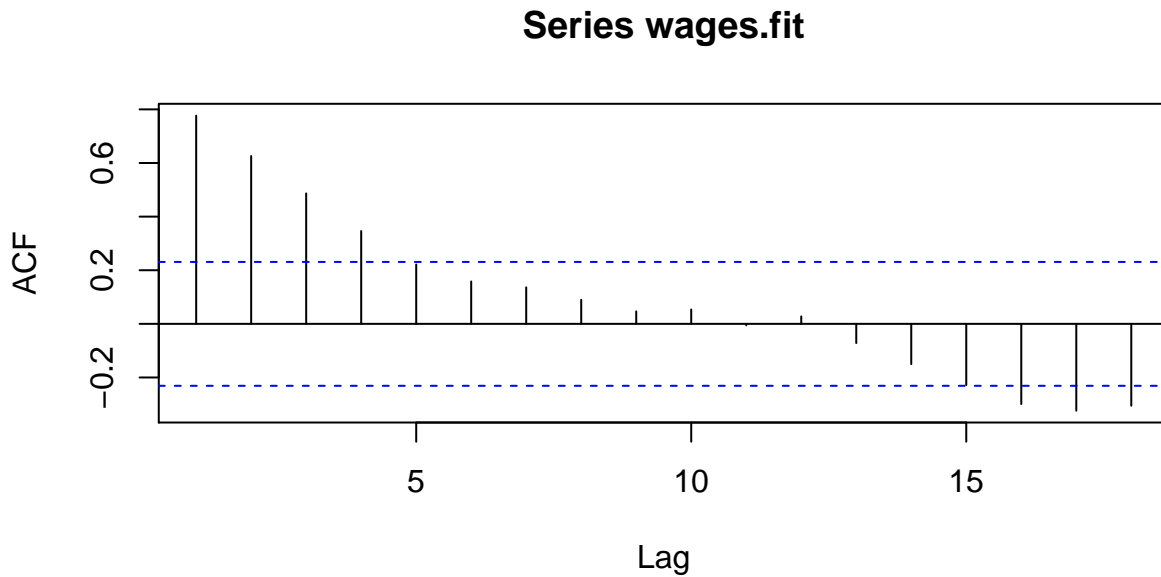
	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-549	11.15	-49.24	4.652e-56
time(wages)	0.2811	0.005618	50.03	1.578e-56

TABLE 9 – Fitting linear model: wages ~ time(wages)

Observations	Residual Std. Error	R^2	Adjusted R^2
72	0.08257	0.9728	0.9724

La p-valeur $< 2.2e - 16$, le modèle est globalement significatif. Et 97% du salaire moyen est expliqué par le temps.

c) Tracer et interpréter la série des résidus normalisés obtenus précédemment



L'ACF des résidus montre une décroissance des résidus en fonction du temps. Ce qui confirme la saisonnalité de la série.

Mais l'hypothèse de normalité appliquée sur les résidus n'est pas vérifiée: p-value = 0.00474. On en conclue que ce modèle n'est pas adéquat par rapport à la série étudiée.

d) Utiliser les moindres carrés pour estimer une tendance quadratique

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-84950	10191	-8.336	4.865e-12
time(wages)	85.34	10.27	8.309	5.439e-12
I(time(wages)^2)	-0.02143	0.002588	-8.282	6.104e-12

TABLE 11 – Fitting linear model: $\text{wages} \sim \text{time(wages)} + \text{I}(\text{time(wages)}^2)$

Observations	Residual Std. Error	R^2	Adjusted R^2
72	0.05889	0.9864	0.986

Le modèle est très significatif: p-value: < 2.2e-16 avec un taux d'ajustement plus important que le précédent(98.6%).

e) Tracer et interpréter la série des résidus normalisés de la partie précédente.

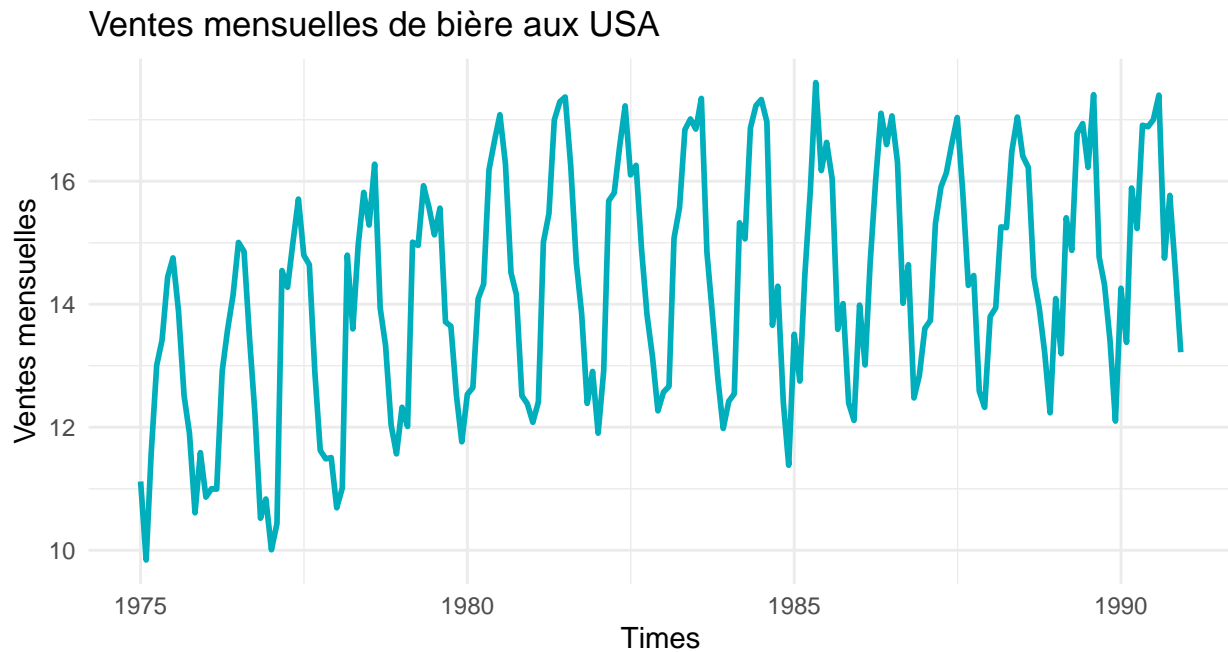


L'ACF nous indique une décroissance rapide du coefficient d'autocorrélation. Les résidus de lag 1 et ceux du lag 2 sont auto-corrélés, puisque le coefficient de corrélation dépasse la borne supérieure de l'intervalle de confiance du coefficient de corrélation de valeur nulle. De plus l'hypothèse de normalité est validée par le teste de shapiro avec un p-value = 0.7622

6. Séries beersales

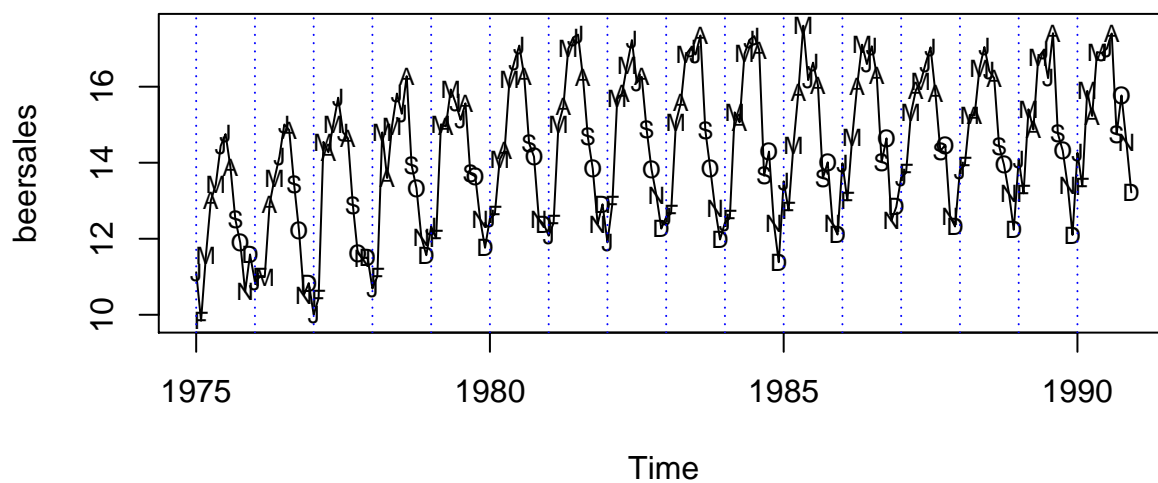
La série beersales donne les ventes mensuelles de bière aux États-Unis (en millions de barils) pour la période allant de janvier 1975 à décembre 1990.

a) Tracer et interpréter le graphique de la série



On observe une croissance de la série jusqu'aux années 1981 qui se stabilise par la suite.

b) Identifier chacun des mois sur le graphique

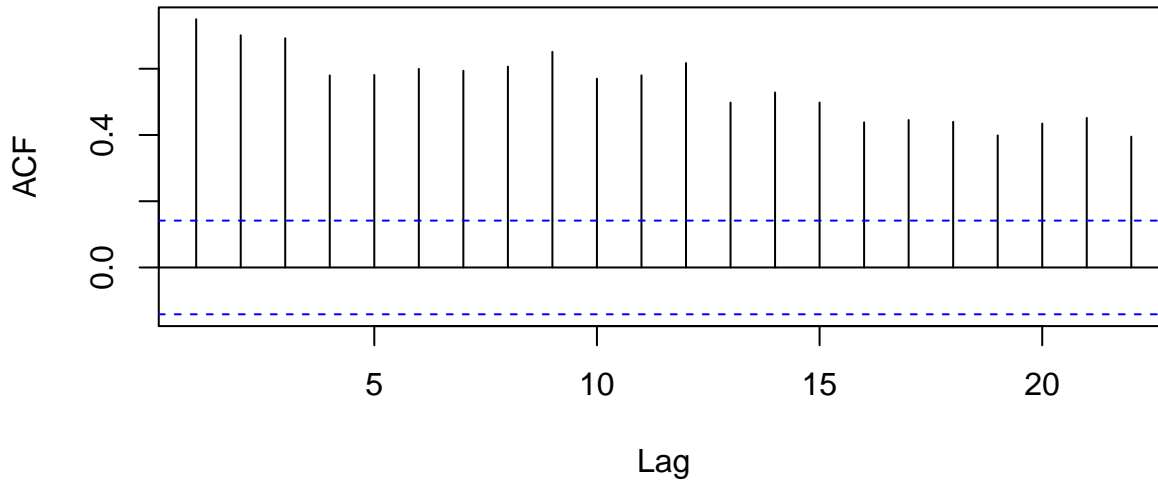


Il apparaît ainsi une tendance saisonnière assez marquée par une augmentation des ventes de bières pendant l'été et une baisse drastique entre Septembre et Décembre.

c) Utiliser les moindres carrés pour estimer la tendance saisonnière de cette série chronologique.

Nous obtenons un $R^2 = 71\%$, valeur très élevée et la regression semble très significative. Pour aller plus loin, dessinons l'ACF des résidus.

Series beersales.fit



Le chronogramme montre de longues séries de valeurs de même signe; typiquement, ce résidu n'est pas stationnaire. Le présupposé variance constante ou au moins de variance obéissant à un mécanisme déterministe ne tient pas. On ne peut pas appliquer les MCO car la relation entre les deux indices n'est pas stable. Et les résidus ne sont pas normalisés, la p-valeur du test de shapiro donne 4.892e-07.

e) Utiliser les moindres carrés afin d'introduire une tendance saisonnière plus une tendance temporelle quadratique pour cette série

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-154.2	10.71	-14.4	9.891e-32
beersales.monthFebruary	-0.1566	0.2439	-0.642	0.5217
beersales.monthMarch	2.054	0.2439	8.421	1.175e-14
beersales.monthApril	2.356	0.2439	9.656	5.06e-18
beersales.monthMay	3.543	0.2439	14.52	4.405e-32
beersales.monthJune	3.78	0.244	15.49	6.816e-35
beersales.monthJuly	3.685	0.244	15.1	9.206e-34
beersales.monthAugust	3.511	0.244	14.39	1.091e-31
beersales.monthSeptember	1.461	0.244	5.986	1.149e-08
beersales.monthOctober	1.128	0.2441	4.623	7.216e-06
beersales.monthNovember	-0.1881	0.2441	-0.7707	0.4419
beersales.monthDecember	-0.5773	0.2441	-2.365	0.01912
I(time(beersales)^2)	4.241e-05	2.723e-06	15.57	4.062e-35

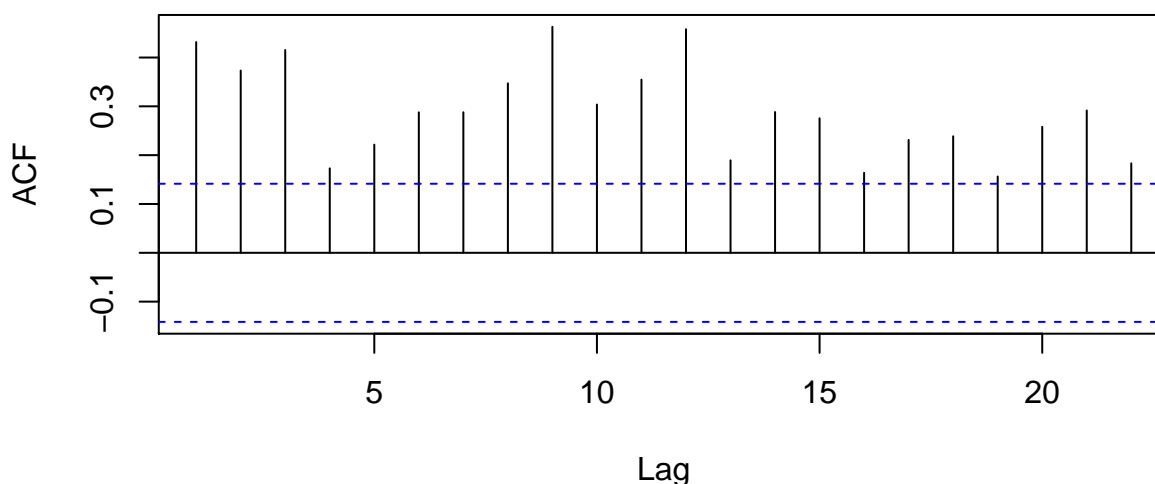
TABLE 13 – Fitting linear model: $\text{beersales} \sim \text{beersales.month} + \text{I}(\text{time}(\text{beersales})^2)$

Observations	Residual Std. Error	R^2	Adjusted R^2
192	0.6899	0.877	0.8687

Nous avons un $R^2 = 87\%$ supérieur au R^2 du précédent modèle. Ce modèle est donc plus explicatif et est très significatif. Hors mis les mois de Février et de Novembre, tous les autres mois semblent expliquer la vente de bières.

f) Tracer et interpréter la série chronologique des résidus normalisés

Series beersales.lm2.fit

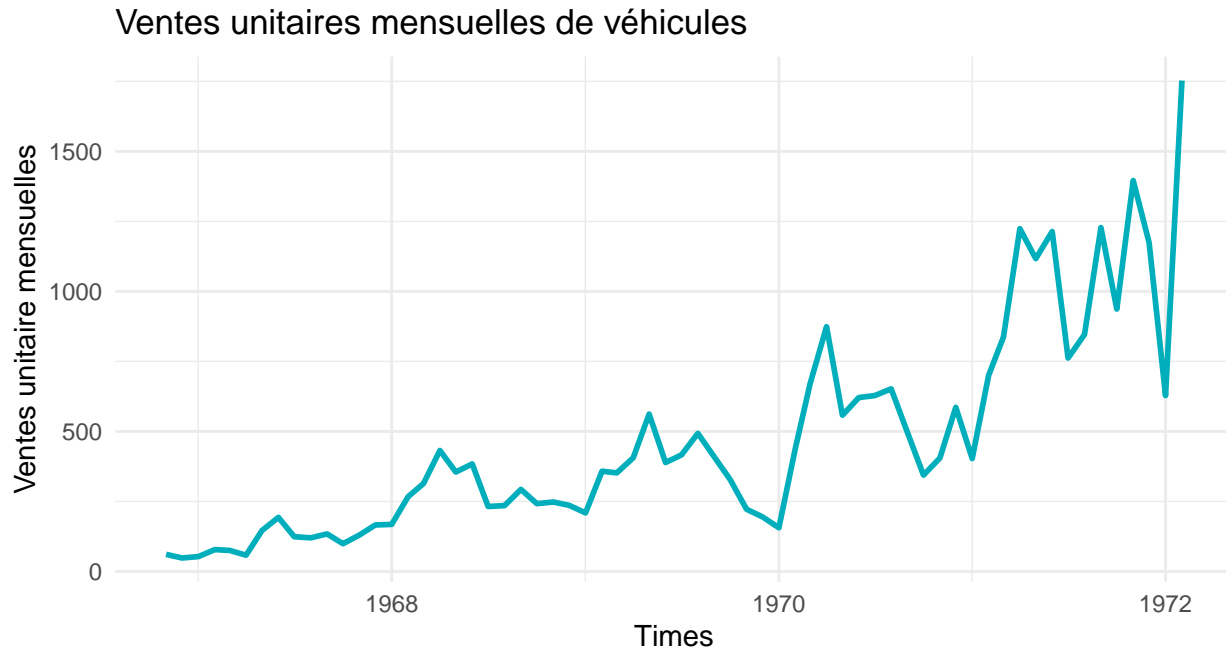


Même si les coefficients de corrélations semblent baisser sur certains lags, on remarque toujours que les résidus n'ont pas une variance constante. Cet exemple nous a conduit à effectuer une régression ayant une significativité illusoire: un R^2 élevé et une régression apparemment très significative, mais qui en réalité débouche sur un résidu non stationnaire.

7. Série winnebago

La série winnebago contient des ventes unitaires mensuelles de véhicules récréatifs Winnebago de Novembre 1966 à Février 1972

a) Tracer et interpréter cette série temporelle.



Le chronogramme montre une augmentation des ventes de véhicules selon le temps. La variabilité de la série augmente avec son niveau. A partir des années 1970, on voit une nette hausse des ventes entre Janvier et Février.

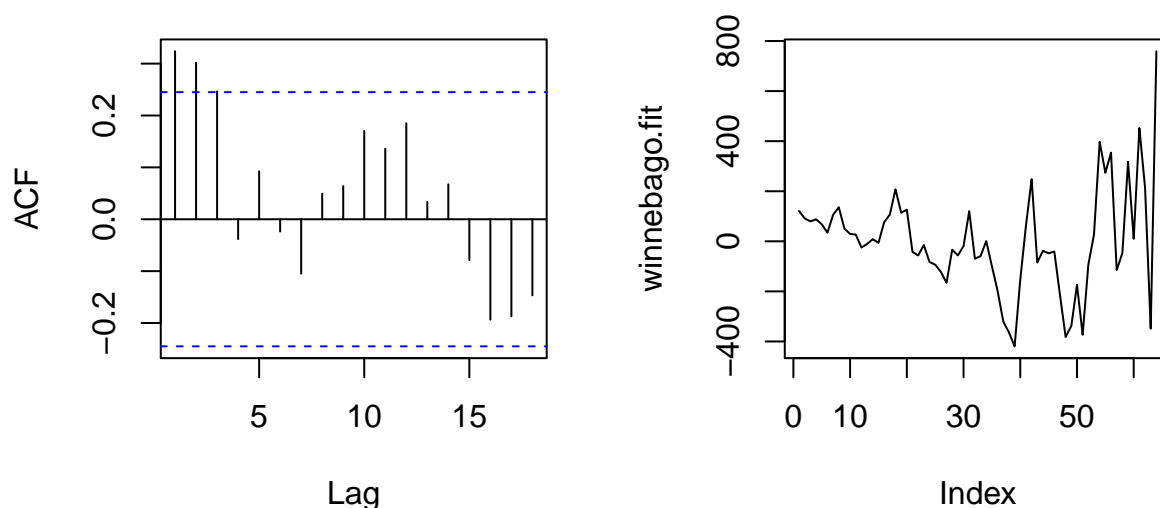
b) Utiliser les moindres carrés pour estimer une tendance linéaire dans ces données.

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-394886	33540	-11.77	1.87e-17
time(winnebago)	200.7	17.03	11.79	1.777e-17

TABLE 15 – Fitting linear model: winnebago ~ time(winnebago)

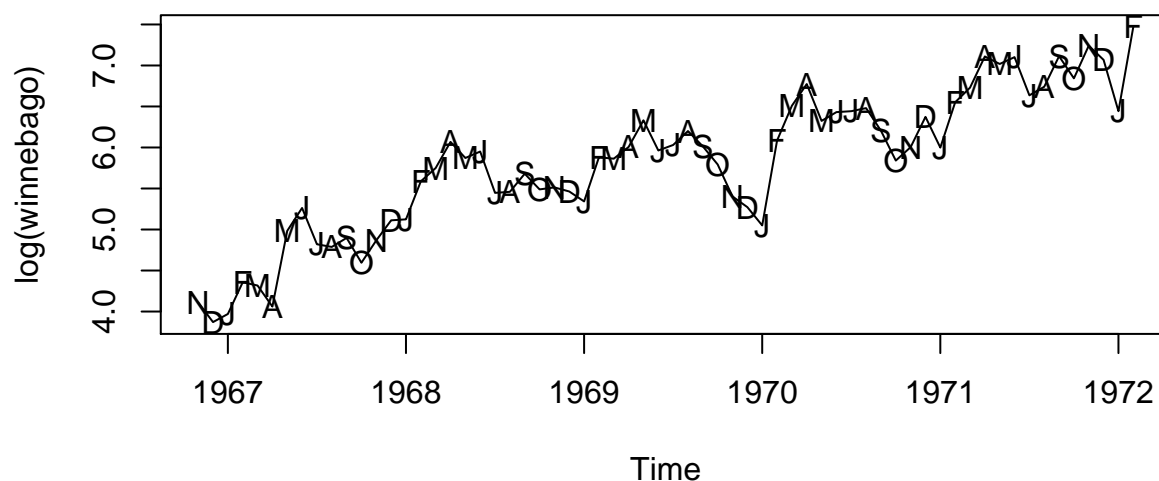
Observations	Residual Std. Error	R^2	Adjusted R^2
64	209.7	0.6915	0.6865

Series winnebago.fit



On remarque des autocorrélations très faibles et significatives sur le lag 1 et à moindre mesure sur le lag 2. On peut aussi noter une persistance de l'aspect saisonnier sur l'ACF.

c) Prendre maintenant le logarithme naturel des ventes mensuelles, tracer et interpréter cette série temporelle.



L'interprétation est presque la même que la précédente série temporelle, sauf qu'avec la transformation logarithmique, on remarque une augmentation plus linéaire des ventes.

d) Utiliser les moindres carrés pour estimer une tendance linéaire dans ces données. Tracer et interpréter la série des résidus standardisés.

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-984.9	62.99	-15.64	3.45e-23
time(log(winnebago))	0.5031	0.03199	15.73	2.575e-23

TABLE 17 – Fitting linear model: $\log(\text{winnebago}) \sim \text{time}(\log(\text{winnebago}))$

Observations	Residual Std. Error	R^2	Adjusted R^2
64	0.3939	0.7996	0.7964

Avec un $R^2 = 80\%$, ce modèle contient plus d'informations explicatives que le précédent. Et de plus, il est très significatif.

e) Utiliser maintenant les moindres carrés pour ajuster une tendance saisonnière plus une tendance linéaire aux ventes enregistrées.

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-997.3	50.64	-19.69	1.718e-25
monthFebruary	0.6244	0.1818	3.434	0.001188
monthMarch	0.6822	0.1909	3.574	0.0007793
monthApril	0.8096	0.1908	4.243	9.301e-05
monthMay	0.8695	0.1907	4.559	3.246e-05
monthJune	0.8631	0.1907	4.526	3.627e-05
monthJuly	0.5539	0.1907	2.905	0.00542
monthAugust	0.5699	0.1907	2.988	0.004305
monthSeptember	0.5757	0.1907	3.018	0.00396
monthOctober	0.2635	0.1908	1.381	0.1733
monthNovember	0.2868	0.1819	1.577	0.1209
monthDecember	0.248	0.1818	1.364	0.1785
time(log(winnebago))	0.5091	0.02571	19.8	1.351e-25

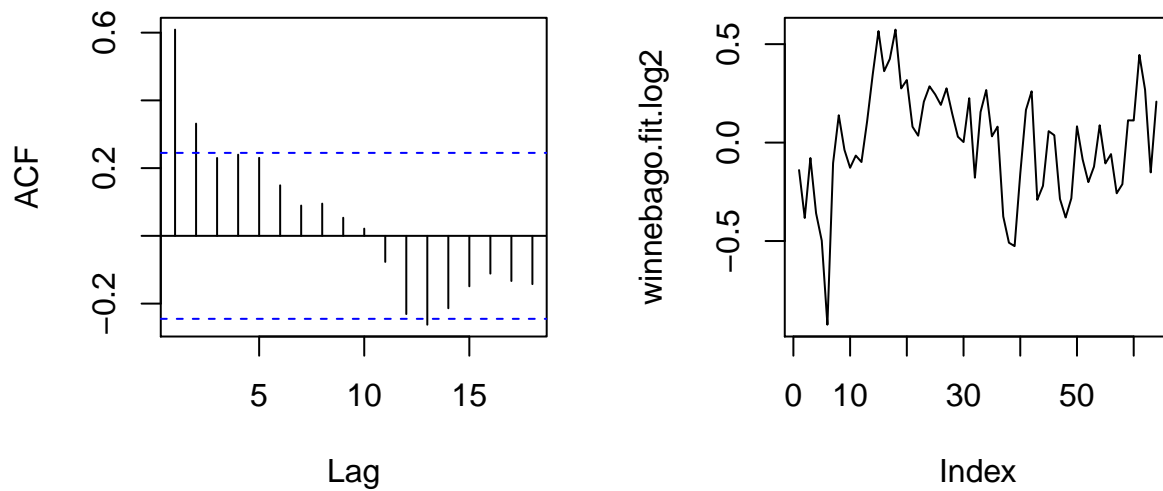
TABLE 19 – Fitting linear model: $\log(\text{winnebago}) \sim \text{month} + \text{time}(\log(\text{winnebago}))$

Observations	Residual Std. Error	R^2	Adjusted R^2
64	0.3149	0.8946	0.8699

Le modèle très significatif avec p-valeur $< 2.2\text{e-}16$ et un $R^2 = 89\%$. Tous les mois exceptés Octobre, Novembre et Décembre sont très significatifs et expliquent les ventes mensuelles de véhicules.

f) Tracer et interpréter la série chronologique des résidus normalisés.

Series winnebago.fit.log2



Avec la transformation logarithmique, on voit que les coefficients de corrélations ont légèrement augmentés avec une baisse des lags moins importante.

	(1)
(Intercept)	12.486 *** (0.264)
beersales.monthFebruary	-0.143 (0.373)
beersales.monthMarch	2.082 *** (0.373)
beersales.monthApril	2.398 *** (0.373)
beersales.monthMay	3.599 *** (0.373)
beersales.monthJune	3.850 *** (0.373)
beersales.monthJuly	3.769 *** (0.373)
beersales.monthAugust	3.609 *** (0.373)
beersales.monthSeptember	1.573 *** (0.373)
beersales.monthOctober	1.254 *** (0.373)
beersales.monthNovember	-0.048 (0.373)
beersales.monthDecember	-0.423 (0.373)
N	192
R2	0.710

*** p < 0.001; ** p < 0.01; * p < 0.05.