TP4, séries temporelles

Mamadou Lamine Diamban, Lucas Chabeau

12/11/2019

2 Modèle 2 : ventes de maisons neuves et mises en chantier.

En cours, nous avons étudié le modèle de fonction de transfert. Nous avons vu deux exemples détaillés. Le second voulait établir le lien entre le nombre de ventes de maisons neuves et le nombre de mises en chantier. Nous avons suggéré un premier modèle :

$$\nabla\nabla_{12}Y_t = \frac{0.8015B}{1-0.4005B}\nabla\nabla_{12}X_t + (10,6304B)(1-0.7849B^{12})\varepsilon_t$$

Cependant, nous avions des réserves car le corrélogramme croisé entre la série en entrée et les résidus du modèle final présentait une corrélation significativement différente de 0 au délai 0. Nous avons donc suggéré de reprendre le travail afin d'estimer le modèle :

$$\nabla\nabla_{12}Y_t = \frac{\omega_0+\omega_1 B}{1-\delta B}\nabla\nabla_{12}X_t + N_t$$

où N_t est un processus ARMA. . .

Faites une copie du fichier Scripts/Housing1.R et renommer cette copie Scripts/Housing2.R. Modifier ce fichier à partir de la ligne 69 afin d'estimer et valider le modèle (2). Cela exigera de vous d'estimer à partir de l'estimation des v_k les paramètres ω_0, ω_1 et δ. Par la suite, procéder aux filtrages appropriés pour compléter le travail.

Nous rappelons cette formule générique pour estimer v_k

$$v_k = \begin{cases} 0 & k \in \{0;...;b-1\} \\ \sum_{j=1}^{r}\delta_j v_{k-j} + \omega_0 & k = b \\ \sum_{j=1}^{r}\delta_j v_{k-j} - \omega_{k-b} & k \in \{b+1;...;b+s\} \\ \sum_{j=1}^{r}\delta_j v_{k-j} & k \geq b+s+1 \end{cases}$$

Dans notre modèle, nous avons $s = 0$, $r = 1$ et $b = 0$. Nous pouvons ainsi en déduire la relation suivante :

$$v_k = \begin{cases} \omega_0 & k = 0 \\ \delta_1 v_0 - \omega_1 & k = 1 \\ \delta_1 v_{k-1} & k \geq 2 \end{cases}$$

A partir de là nous obtenons les estimateurs suivants :

$$\begin{cases} \omega_0 = \hat{v}_0 \\ \delta_1 = \frac{\hat{v}_2}{\hat{v}_1} \\ \omega_1 = \delta_1\hat{v}_0 - \hat{v}_1 \end{cases}$$

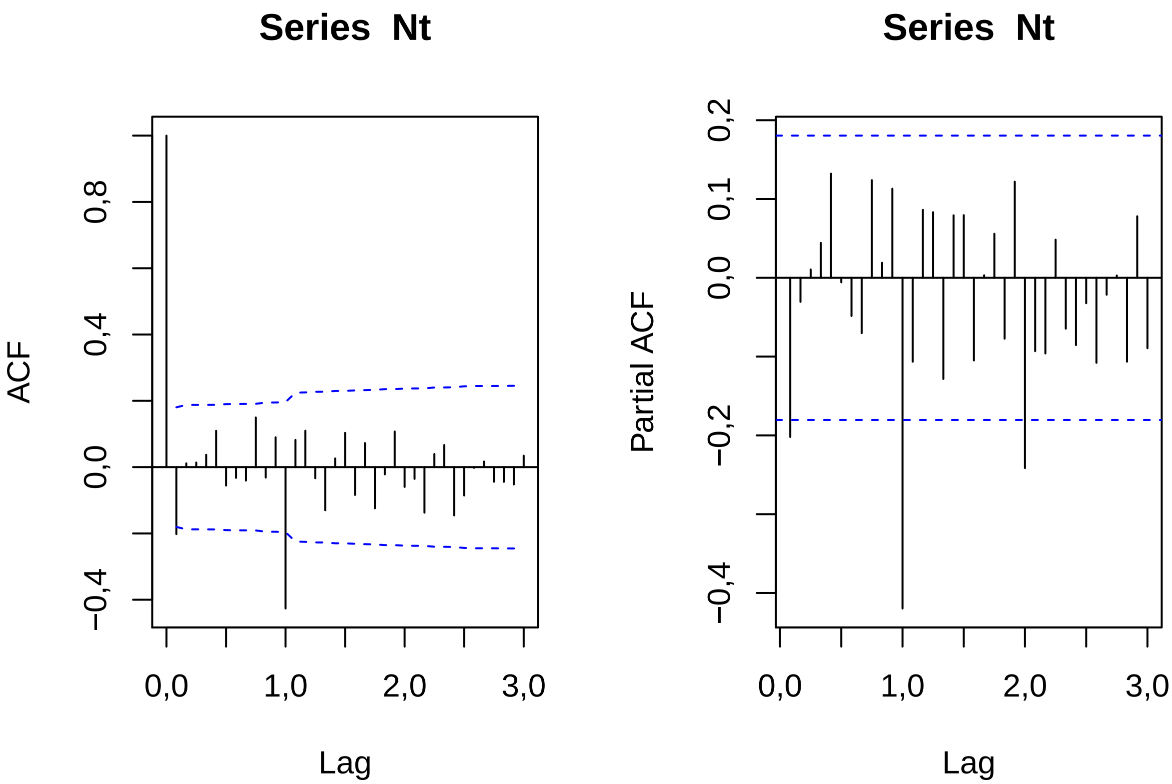
A partir de l'estimation des $\hat{v}_k$, nous obtenons donc les estimations suivantes :

$\hat{\omega}_0$ = 0,2731, $\hat{\omega}_1$ = -0,5266 et $\hat{\delta}_1$ = 0,3836

Nous allons maintenant pouvoir récupérer la série N_t et identifier son processus :

$$N_t = \nabla\nabla_{12}Y_t - \frac{\omega_0 + \omega_1 B}{1 - \delta B}\nabla\nabla_{12}X_t$$

Nous traçons ci-dessous les autocorrélogrammes complet et partiel de la série des bruits N_t.



A la vue de ces autocorrélogrammes, nous identifions un processus saisonnier MA (Moving Average) pour les raisons suivantes :

- L'autocorrélogramme simple nous montre un nombre fini de corrélations significativement différentes de zéro, et aucune ne l'est après le délai q = 12. La première corrélation est toute proche de ne pas être significativement différente de 0, c'est pour ça que nous l"'ignorons" et identifions un processus MA.
- L'autocorrélogramme partiel nous montre plusieurs corrélations significativement de 0 qui s'ammortissent.

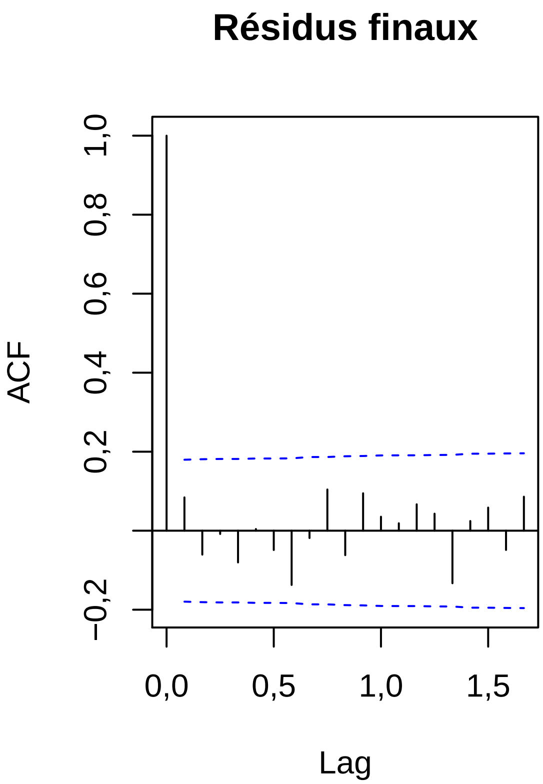
Nous allons maintenant estimer nos paramètres θ et Θ

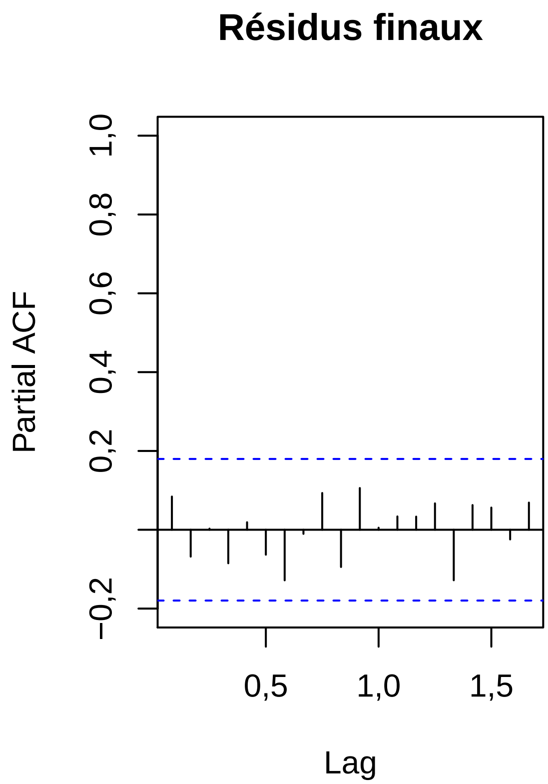
Nous estimons donc θ = -0,1725 et Θ = -0,9496

Nous cherchons maintenant à minimiser nos résidus, nous allons définir une fonction qui calcule les résidus du modèle. Puis, nous optimisons les paramètres de notre modèle pour minimiser les résidus.

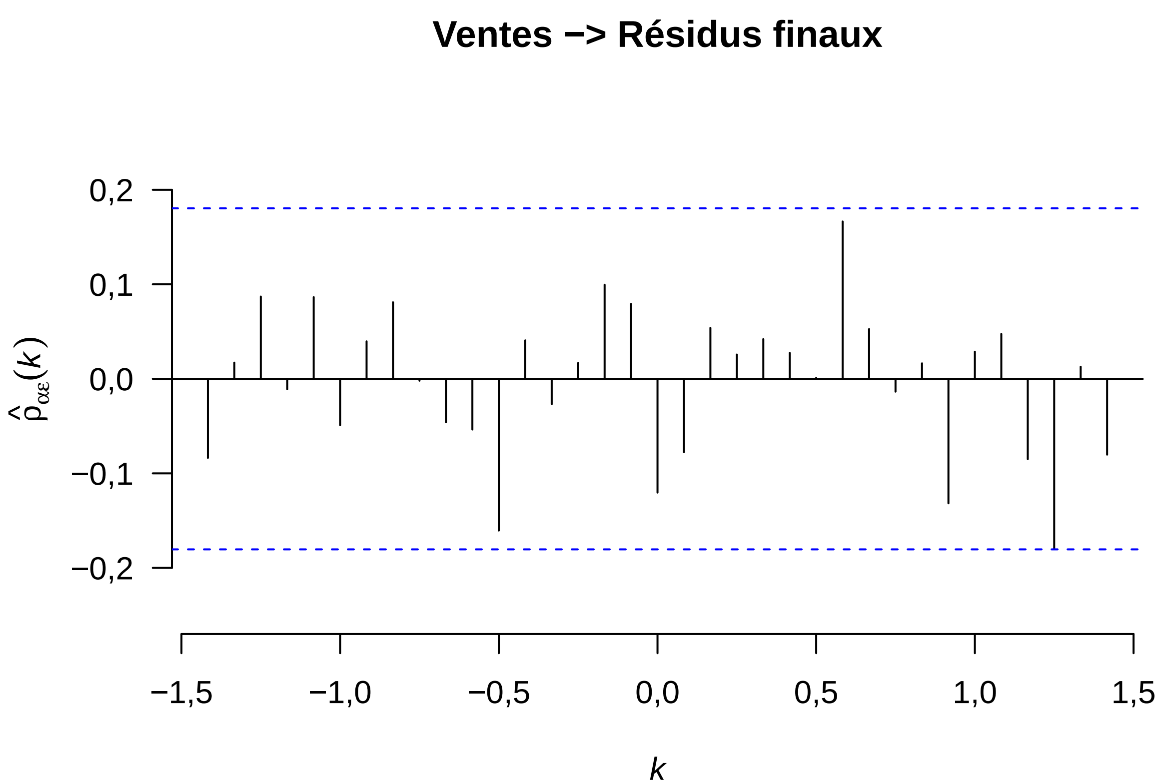
Le modèle optimal a les paramètres suivants : $\{\omega_0$ = 0,4699, ω_1 = 0,0523, δ_1 = 0,724 θ = -0,7446 et Θ = -0,7289$\}$

Nous pouvons maintenant modéliser notre série avec les paramètres optimaux obtenus. Nous avons avec ce modèle, une erreur standard moyenne de 35,3571



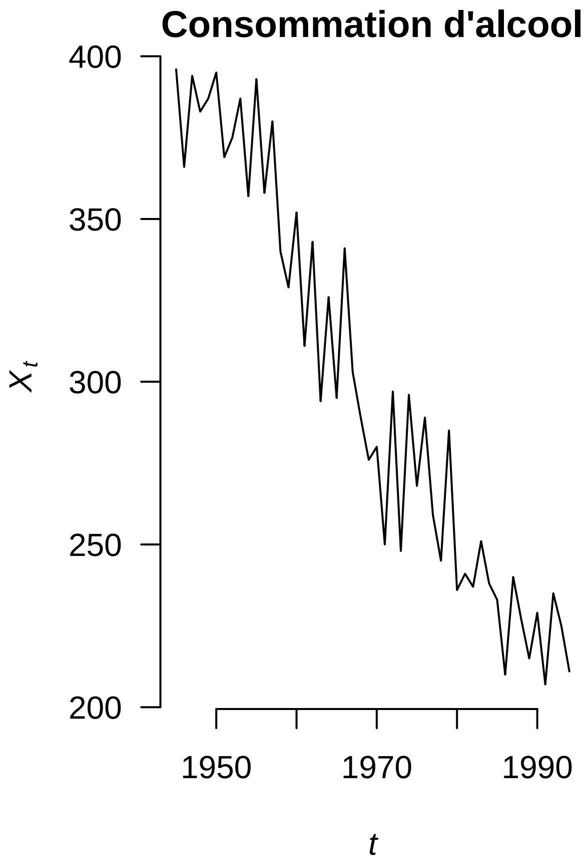


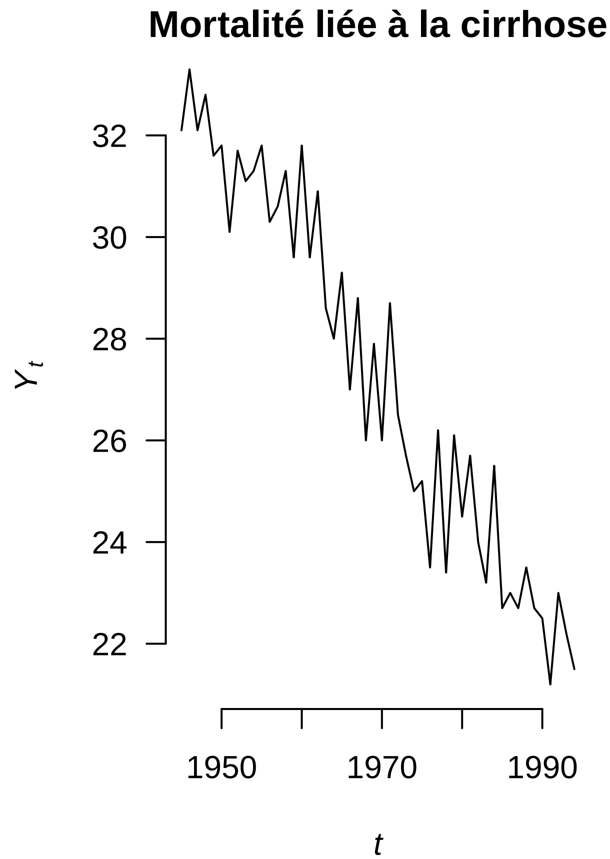
Nous voyons grâce aux autocorrélogrammes simple et partiel qu'avec ce modèle, les résidus n'ont plus d'autocorrélations significativement différentes de 0. Les résidus sont du bruit blanc.



Nous n'observons pas de délai où les résidus sont significativement corrélés à la série d'entrée (les ventes). Notre modèle est donc valide.

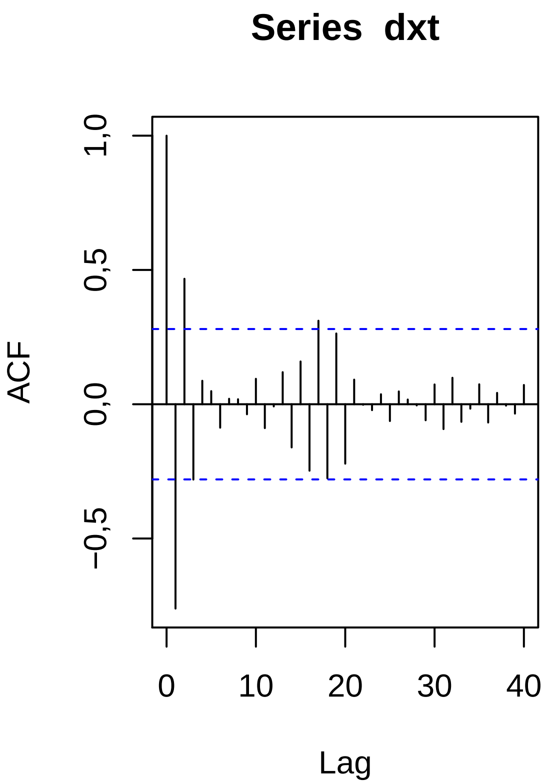
3. Consommation d'alcool et taux de mortalité par cirrhose.

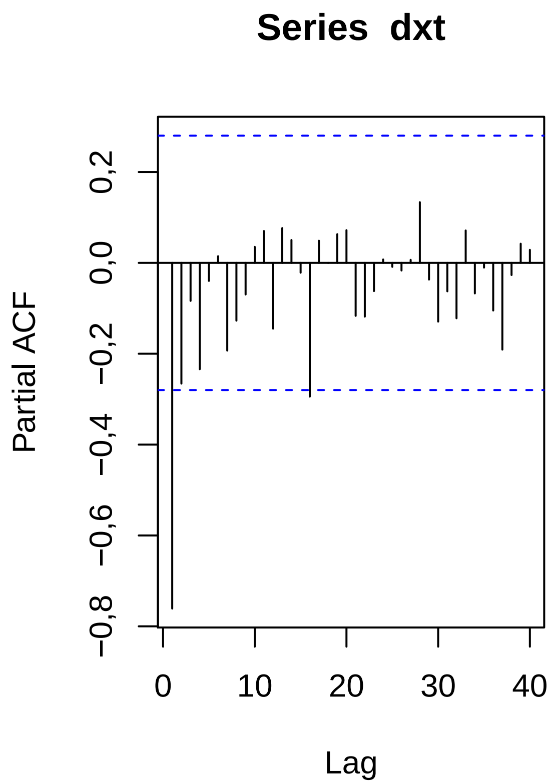




On remarque une forte similarité entre le graphique de la conssommation d'alcool et la mortalité liée à la cirrhose où l'on voit sur les deux graphiques une décroissance régulière depuis la fin des années 1940. Cette décroissance s'accélère à la fin des années 1960 avec une baisse importante de la quantité de consommation d'alcool et dans le même intervalle de temps, on note aussi une baisse de mortalité liée à la cirrhose.
Ces liens sont confirmés par une corrélation positive très signification(0,9239455) entre les deux séries.

Etant données que les séries ne sont pas stationnaires, par la suite, nous travaillons avec $x_t = \nabla X_t$ et $y_t = \nabla Y_t$





L'ACF a une une décroissance exponentielle, de plus les autocorrélations empiriques ont une forme sinisoïdale dont seul le premier retard semble très significatif. On peut donc envisager un modèle **ARMA(1,1,1)** dont les coefficients sont donnés par le tableau suivant.

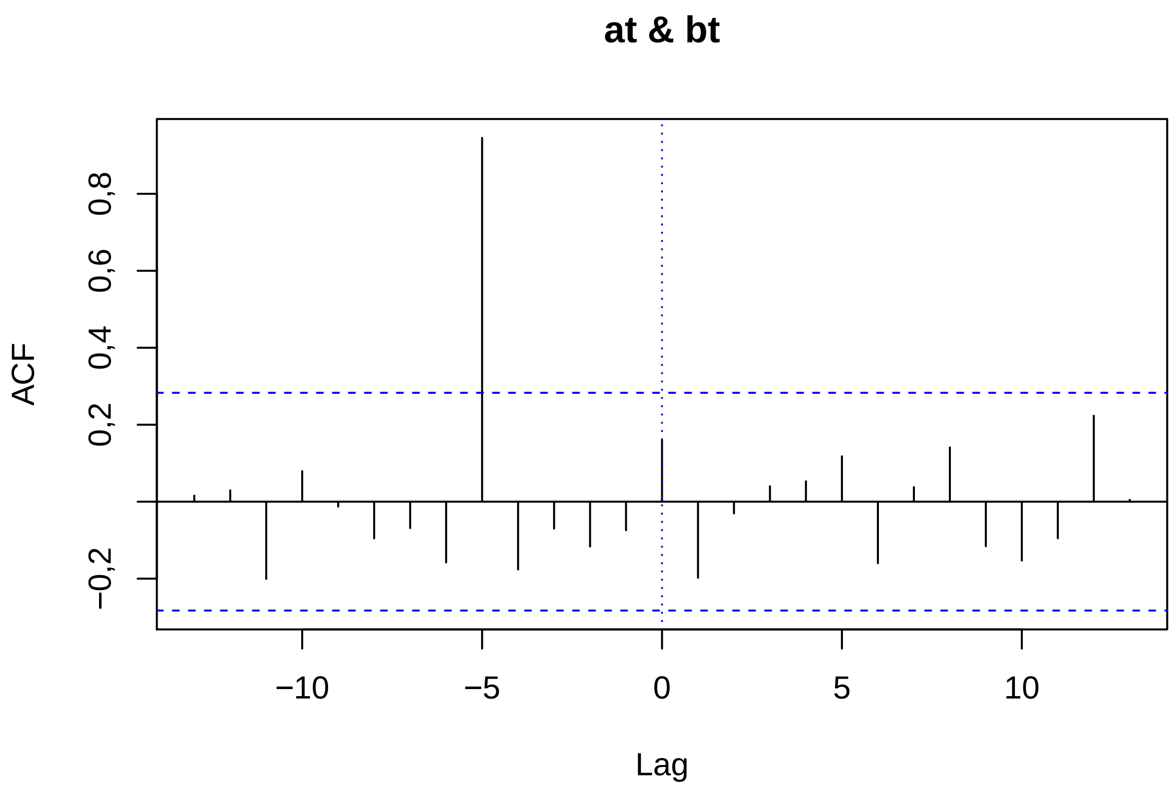
Call: arima(x = xt, order = c(1, 1, 1), method = "CSS")

Table 1: Coefficients

	ar1	ma1
	-0.6812	-0.1356
s.e.	0.126	0.1589

sigma^2 estimated as 341: part log likelihood = -212,41

A partir de là, on peut envisager un filtrage de nos 2 séries avec les paramètres trouvés.



On a qu'un seul retard qui est significatif au retard 5. Nou avons donc l'équation suivante:

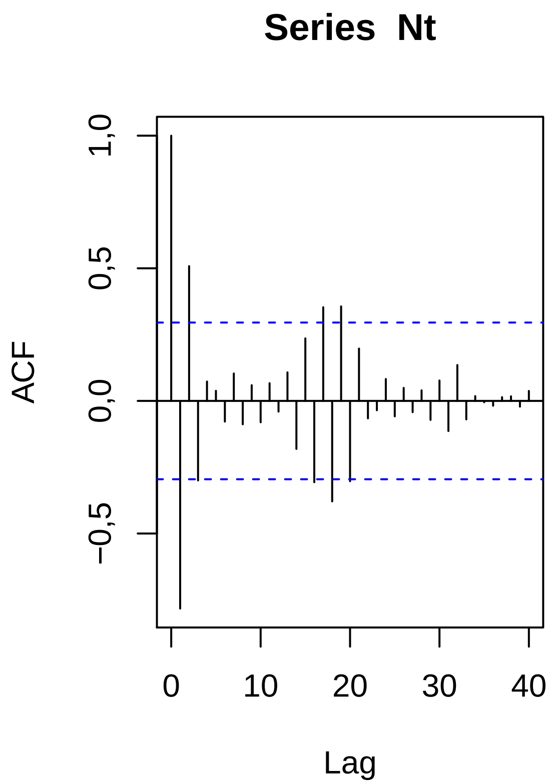
$$v(B) = \omega_0 B^5$$

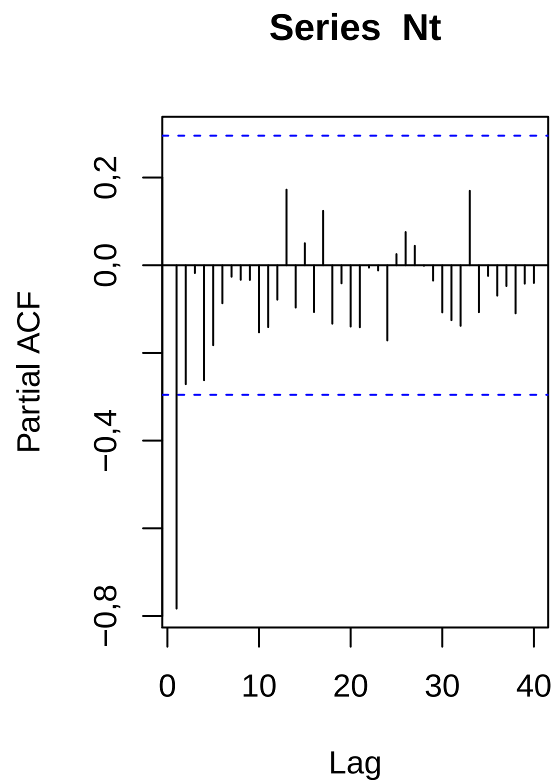
où les v_k peuvent être estimés par:

$$v_k = \sum_{j=1}^{r} \delta_j v_{k-j} + \omega_0; \quad k = b$$

pour b=5 nous avons donc :

$$\omega_0 = v_5 \; pour \; k = b = 5$$





Sur l'Acf, on remarque un changement de signe sur les retards multiples de 12. De plus, sur le Pacf, seul le retard 1 est significatif. Nous pouvons donc modéliser avec un processus **ARMA(1,0,1)**.
Les paramètres estimés sont données par le tableau suivant:

Call: arima(x = Nt, order = c(1, 0, 1))

Table 2: Coefficients

	ar1	ma1	intercept
	-0.6115	-0.4489	-0.01328
s.e.	0.1608	0.2328	0.00422

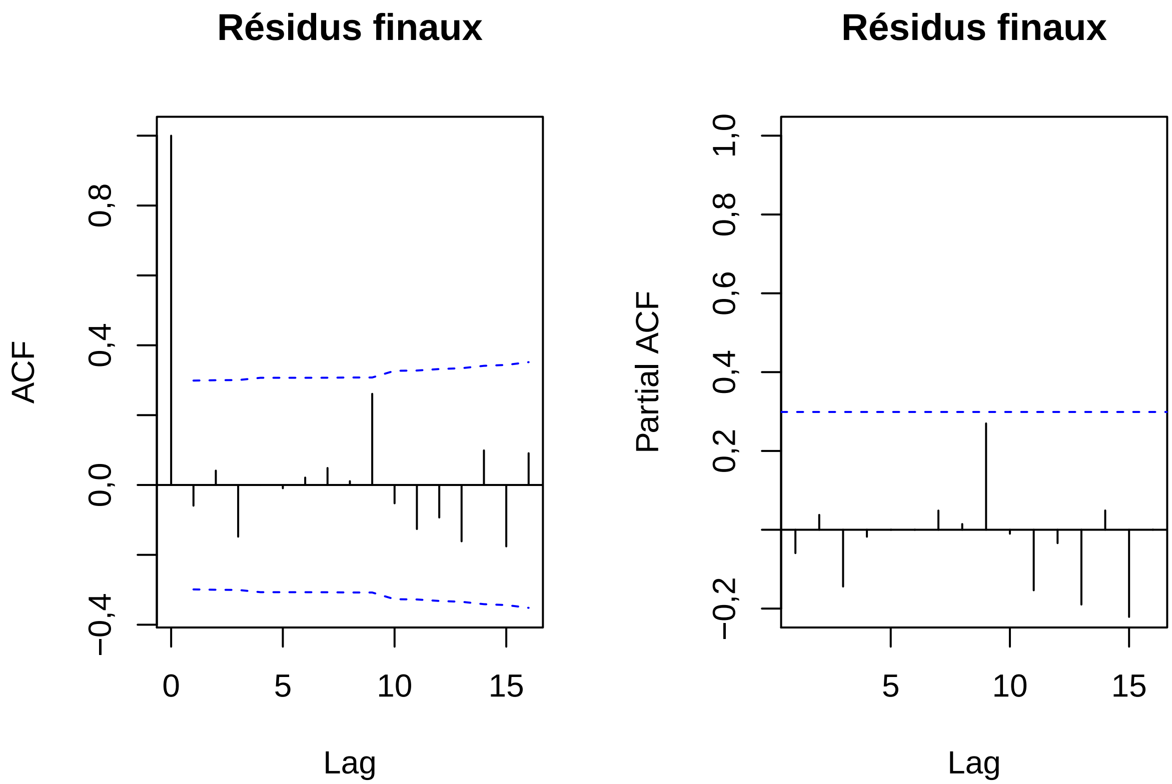
sigma^2 estimated as 0,006225: log likelihood = 48,72, aic = -89,44

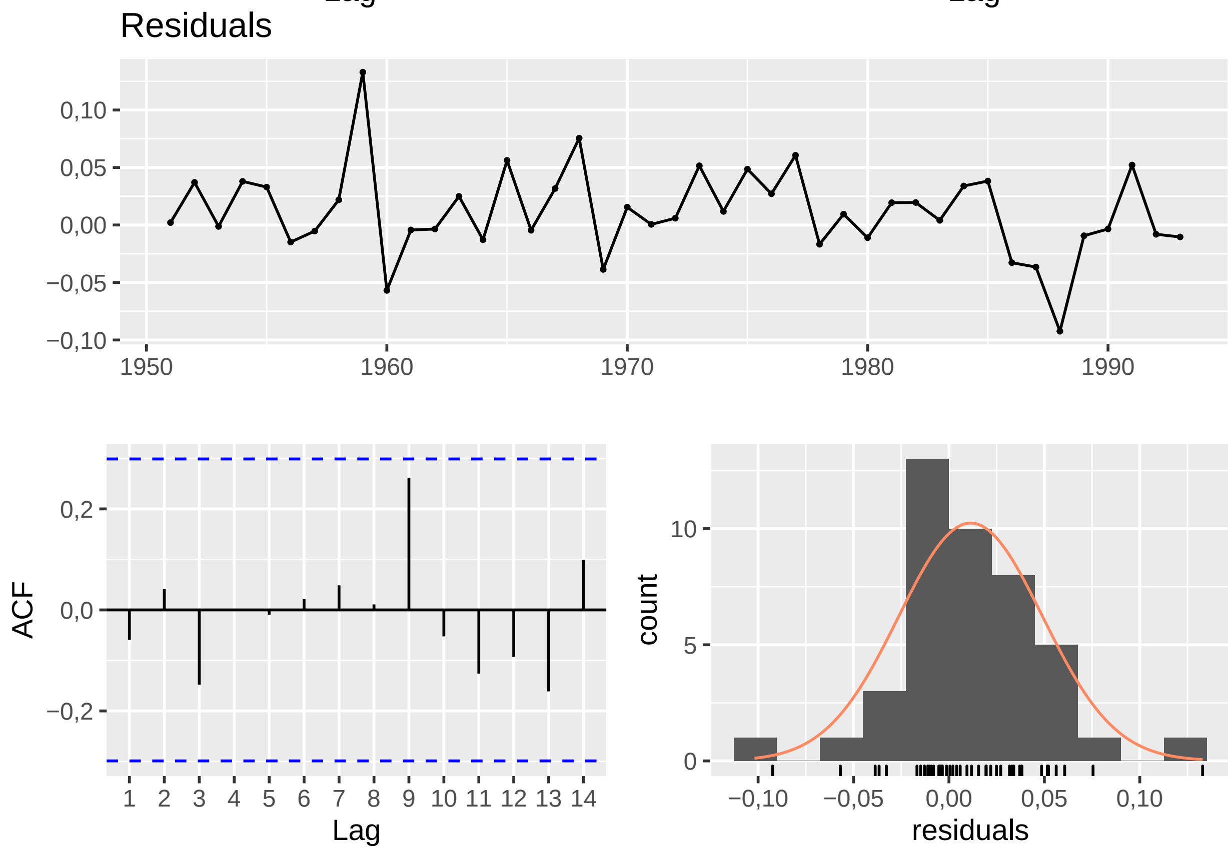
On utilise la fonction utilisée dans l'exercice Housing pour calculer les résidus.

On a ensuite les paramètres optimaux:

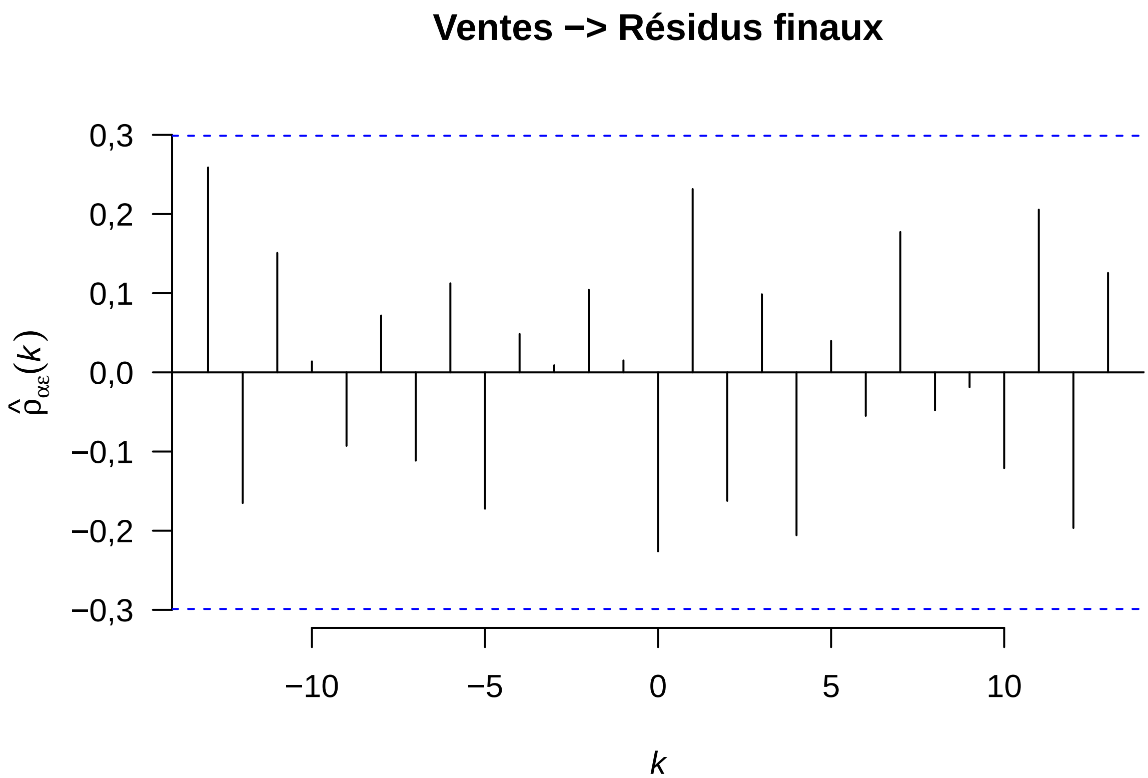
omega_0	theta_Nt	phi_Nt	betha0
0.05779	-1.119	0.1984	0.005636

Par la suite, nous avons calculé la somme au carré des erreurs qui est: 0,0015211





Les résidus semblent stationnaires et suivent une loi normale. Aucune autocorrélation des erreurs n'est significative.



Aucune autocorrélation n'est significative entre les variables prédictives et les résidus.

Notre modèle final est donc le suivant :

$$\nabla Y_t = 0.058 \times B^5 + 0.006 + 0.198 \times B \times N_t + (1 - 1.116 \times B)\epsilon_t$$