Compte Rendu Série Temporelle

Mamadou Lamine DIAMBAN

12 Novembre 2019

La prévision

1. Écrire les modèle suivants sous forme d'équation aux différences :

Sachant que $\widetilde{z_t} = z_t - \mu$

$$(a)\widetilde{z_t} = (1 + \theta_1 B + \theta_2 B^2)a_t$$
$$(a)z_t = a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2} + \mu$$
$$(a)z_{t+1} = a_{t+1} + \theta_1 a_t + \theta_2 a_{t-1} + \mu$$

$$(b)(1 - \phi_1 B)(1 - B)\widetilde{z_t} = a_t$$
$$(b)(1 - B - \phi_1 B + \phi_1 B^2)(z_t - \mu) = a_t$$
$$(b)(1 - B - \phi_1 B + \phi_1 B^2)z_t = a_t$$

Car $B\mu = \mu$

$$(b)z_t - (1 + \phi_1)z_{t-1} + \phi_1 z_{t-2} = a_t$$
$$(b)z_t = a_t + (1 + \phi_1)z_{t-1} - \phi_1 z_{t-2}$$
$$(b)z_{t+1} = a_{t+1} + (1 + \phi_1)z_t - \phi_1 z_{t-1}$$

2. Utiliser les informations suivantes pour prévoir les valeurs aux horizons l = 1, 2 et 3 pour chacun des modèles de la question 1.

$$(a)z_t = a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2} + \mu$$

avec $(a)n = 100; \hat{a_{99}} = 1,3; \hat{a_{100}} = -2,6; \hat{\theta_1} = -0,7; \hat{\theta_2} = 0,5; \hat{\mu} = 100$

$$(a)\hat{z}_t = a_t - 0,7a_{t-1} + 0,5a_{t-2} + 100$$
$$(a)\hat{z}_t(1) = -0,7a_t + 0,5a_{t-1} + 100$$
$$(a)\hat{z}_t(1) = -0,7 * (-2,6) + 0,5 * 1,3 + 100 = 1,82 + 0,65 + 100 = 102,47$$
$$(a)\hat{z}_t(2) = 0,5a_t + 100 = 0,5 * (-2,6) + 100 = 98,7$$

$$(a)\hat{z}_t(3) = 100$$

$$(b)z_t = a_t + (1+\phi_1)z_{t-1} - \phi_1 z_{t-2}$$

avec $(b)n = 100; z_{99} = 217; z_{100} = 232; \hat{\phi}_1 = 0,3$

$$(b)z_t = a_t + (1+0,3)z_{t-1} - 0,3z_{t-2}$$

$$(b)\hat{z}_t(1) = \hat{a}_{t+1} + (1+\phi_1)z_t - \phi_1 z_{t-1}$$

$$(b)\hat{z}_t(1) = (1,3) * 232 - 0,3 * 217 = 301,6 - 65,1 = 236,5$$

$$(b)\hat{z}_t(2) = (1,3)\hat{z}_{t+1} - 0,3z_t = 307,45 - 69,6 = 237,85$$

$$(b)\hat{z}_t(3) = (1,3)\hat{z}_{t+2} - 0,3\hat{z}_{t+1} = 238,255$$

3 Trouver les valeurs des trois premiers poids pour chacun des modèles de l'exercice 2. Présenter à la fois la forme algébrique et les valeurs numériques.

$$(a)z_t = a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2} + \mu$$

C'est un ARIMA(0,2,2), donc pour $k \geq 1$:

$$(a)\psi_k = 1 + \theta_2(1 - \theta_1 - \theta_2)k$$

$$(b)\hat{\psi}_0 = 1$$

$$(a)\hat{\psi}_1 = 1 + \theta_2(1 - \theta_1 - \theta_2) = 1 + 0,5(1 + 0,7 - 0,5) = 1,6$$

$$(a)\hat{\psi}_1 = 1 + \theta_2(1 - \theta_1 - \theta_2)2 = 1 + 1(1 + 0,7 - 0,5) = 3,4$$

Pour le deuxième modèle :

$$(b)z_t = a_t + (1+\phi_1)z_{t-1} - \phi_1 z_{t-2}$$

C'est un ARIMA(1,1,0), donc pour $k \geq 1$:

$$(b)\psi_k = (1 + \phi^{k+1})/(1-\phi)$$

$$(b)\hat{\psi}_0 = 1$$

$$(b)\hat{\psi}_1 = (1 + \phi^2)/(1-\phi) = 1,557143$$

$$(b)\hat{\psi}_2 = (1 + \phi^3)/(1-\phi) = 1,467143$$

4 Déterminer l'écart type estimé de l'erreur de prévision pour chacune des prévisions produites dans l'exercice 2 en utilisant les informations suivantes

La variance de l'écart type de l'erreur de prévision à pour formule :

$$sd[e_t(l)] = \sigma \sqrt{\sum_{k=0}^{l-1} \psi_k^2}$$

$$(a) sd[e_t(1)] = 2,5\sqrt{1} = 2,5$$

$$(a) sd[e_t(2)] = 2,5\sqrt{1 + 1,6^2} = 4,716991$$

$$(a) sd[e_t(3)] = 2,5\sqrt{1 + 1,6^2 + 3,4^2} = 9,721111$$

$$(b) sd[e_t(1)] = 8\sqrt{1} = 8$$

$$(b) sd[e_t(2)] = 8\sqrt{1 + 1,557143^2} = 14.80474$$

$$(b) sd[e_t(3)] = 8\sqrt{1 + 1,557143^2 + 1,467143^2} = 18.89288$$

5 Construire des intervalles de confiance de 80% et de 95% pour chacune des prévisions produites dans l'exerice 2.

Formule de l'intervalle de confiance :

$$E[z_{t+l}|F_t] \pm c_{\alpha/2} * sd[e_t(l)]$$

Intervalle à 80 % :

```
## [1] "modele a"

##      prevision erreur-type intervalle borne min 80% intervalle borne max 80%
## [1,]    102.47    2.500000                 99.26612                 105.6739
## [2,]     98.70    4.716991                 92.65493                 104.7451
## [3,]    100.00    9.721111                 87.54189                 112.4581
##      intervalle borne min 95% intervalle borne max 95%
## [1,]                 97.57009                 107.3699
## [2,]                 89.45487                 107.9451
## [3,]                 80.94697                 119.0530
```

1. La prévision

1. Écrire les modèle suivants sous forme d'équation aux différences :

Sachant que $\widetilde{z}_t = z_t - \mu$

$$(a) \widetilde{z}_t = (1 + \theta_1 B + \theta_2 B^2) a_t$$

$$(a) z_t = a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2} + \mu$$

$$(a) z_{t+1} = a_{t+1} + \theta_1 a_t + \theta_2 a_{t-1} + \mu$$

$$(b)(1-\phi_1 B)(1-B)\widetilde{z}_t = a_t$$
$$(b)(1-B-\phi_1 B+\phi_1 B^2)(z_t-\mu) = a_t$$
$$(b)(1-B-\phi_1 B+\phi_1 B^2)z_t = a_t$$

Car $B\mu = \mu$

$$(b) z_t - (1+\phi_1)z_{t-1} + \phi_1 z_{t-2} = a_t$$
$$(b) z_t = a_t + (1+\phi_1)z_{t-1} - \phi_1 z_{t-2}$$
$$(b) z_{t+1} = a_{t+1} + (1+\phi_1)z_t - \phi_1 z_{t-1}$$

2. Utiliser les informations suivantes pour prévoir les valeurs aux horizons l = 1, 2 et 3 pour chacun des modèles de la question 1.

$$(a) z_t = a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2} + \mu$$

avec $(a) n = 100; \hat{a_{99}} = 1,3; \hat{a_{100}} = -2,6; \hat{\theta_1} = -0,7; \hat{\theta_2} = 0,5; \hat{\mu} = 100$

$$(a)\hat{z}_t = a_t - 0,7a_{t-1} + 0,5a_{t-2} + 100$$
$$(a)\hat{z}_t(1) = -0,7a_t + 0,5a_{t-1} + 100$$
$$(a)\hat{z}_t(1) = -0,7*(-2,6) + 0,5*1,3 + 100 = 1,82 + 0,65 + 100 = 102,47$$
$$(a)\hat{z}_t(2) = 0,5a_t + 100 = 0,5*(-2,6) + 100 = 98,7$$
$$(a)\hat{z}_t(3) = 100$$
$$(b) z_t = a_t + (1+\phi_1)z_{t-1} - \phi_1 z_{t-2}$$

avec $(b) n = 100; z_{99} = 217; z_{100} = 232; \hat{\phi_1} = 0,3$

$$(b) z_t = a_t + (1+0,3)z_{t-1} - 0,3z_{t-2}$$
$$(b)\hat{z}_t(1) = \hat{a}_{t+1} + (1+\phi_1)z_t - \phi_1 z_{t-1}$$
$$(b)\hat{z}_t(1) = (1,3)*232 - 0,3*217 = 301,6 - 65,1 = 236,5$$
$$(b)\hat{z}_t(2) = (1,3)\hat{z}_{t+1} - 0,3z_t = 307,45 - 69,6 = 237,85$$

$$(b)\hat{z}_t(3) = (1,3)\hat{z}_{t+2} - 0,3\hat{z}_{t+1} = 238,255$$

3 Trouver les valeurs des trois premiers poids pour chacun des modèles de l'exercice 2. Présenter à la fois la forme algébrique et les valeurs numériques.

$$(a) z_t = a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2} + \mu$$

C'est un ARIMA(0,2,2), donc pour $k \geq 1$:

$$(a) \psi_k = 1 + \theta_2(1 - \theta_1 - \theta_2)k$$
$$(b) \hat{\psi}_0 = 1$$
$$(a) \hat{\psi}_1 = 1 + \theta_2(1 - \theta_1 - \theta_2) = 1 + 0,5(1 + 0,7 - 0,5) = 1,6$$
$$(a) \hat{\psi}_1 = 1 + \theta_2(1 - \theta_1 - \theta_2)2 = 1 + 1(1 + 0,7 - 0,5) = 3,4$$

Pour le deuxième modèle :

$$(b) z_t = a_t + (1 + \phi_1) z_{t-1} - \phi_1 z_{t-2}$$

C'est un ARIMA(1,1,0), donc pour $k \geq 1$:

$$(b) \psi_k = (1 + \phi^{k+1})/(1 - \phi)$$
$$(b) \hat{\psi}_0 = 1$$
$$(b) \hat{\psi}_1 = (1 + \phi^2)/(1 - \phi) = 1,557143$$
$$(b) \hat{\psi}_2 = (1 + \phi^3)/(1 - \phi) = 1,467143$$

4 Déterminer l’écart type estimé de l’erreur de prévision pour chacune des prévisions produites dans l’exercice 2 en utilisant les informations suivantes

La variance de l’écart type de l’erreur de prévision a pour formule :

$$sd[e_t(l)] = \sigma \sqrt{\sum_{k=0}^{l-1} \psi_k^2}$$

$$(a) sd[e_t(1)] = 2,5\sqrt{1} = 2,5$$
$$(a) sd[e_t(2)] = 2,5\sqrt{1+1,6^2} = 4,716991$$
$$(a) sd[e_t(3)] = 2,5\sqrt{1+1,6^2+3,4^2} = 9,721111$$
$$(b) sd[e_t(1)] = 8\sqrt{1} = 8$$
$$(b) sd[e_t(2)] = 8\sqrt{1+1,557143^2} = 14.80474$$
$$(b) sd[e_t(3)] = 8\sqrt{1+1,557143^2+1,467143^2} = 18.89288$$

5 Construire des intervalles de confiance de 80% et de 95% pour chacune des prévisions produites dans l’exerice 2.

Formule de l’intervalle de confiance :

$$E[z_{t+l}|F_t] \pm c_{\alpha/2} * sd[e_t(l)]$$

Intervalle à 80 % :

TABLE 1 – Modèle a

prevision	erreur-type	intervalle borne min 80%	intervalle borne max 80%	intervalle borne min 95%	intervalle borne max 95%
102.47	2.500000	99.26612	105.6739	97.57009	107.3699
98.70	4.716991	92.65493	104.7451	89.45487	107.9451
100.00	9.721111	87.54189	112.4581	80.94697	119.0530

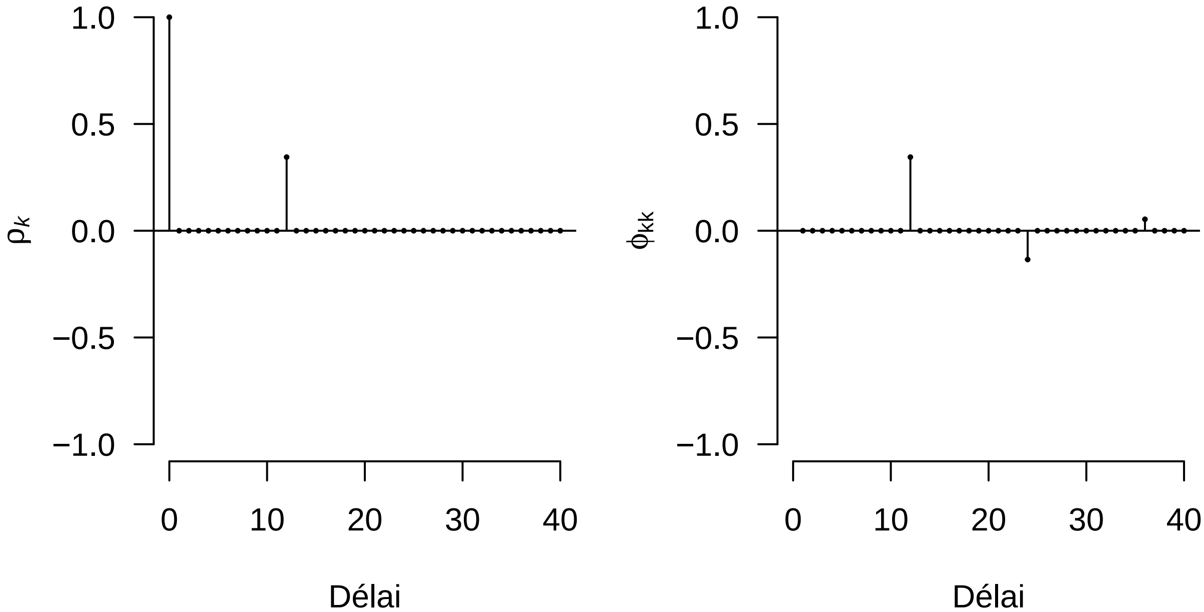
TABLE 2 – Modèle b

prevision	erreur-type	intervalle borne min 80%	intervalle borne max 80%	intervalle borne min 95%	intervalle borne max 95%
236.500	8.00000	226.2476	246.7524	220.8203	252.1797
237.850	14.80474	218.8770	256.8230	208.8332	266.8668
238.255	18.89288	214.0428	262.4672	201.2256	275.2844

3.1 Etude des corrélogrammes saisonniers

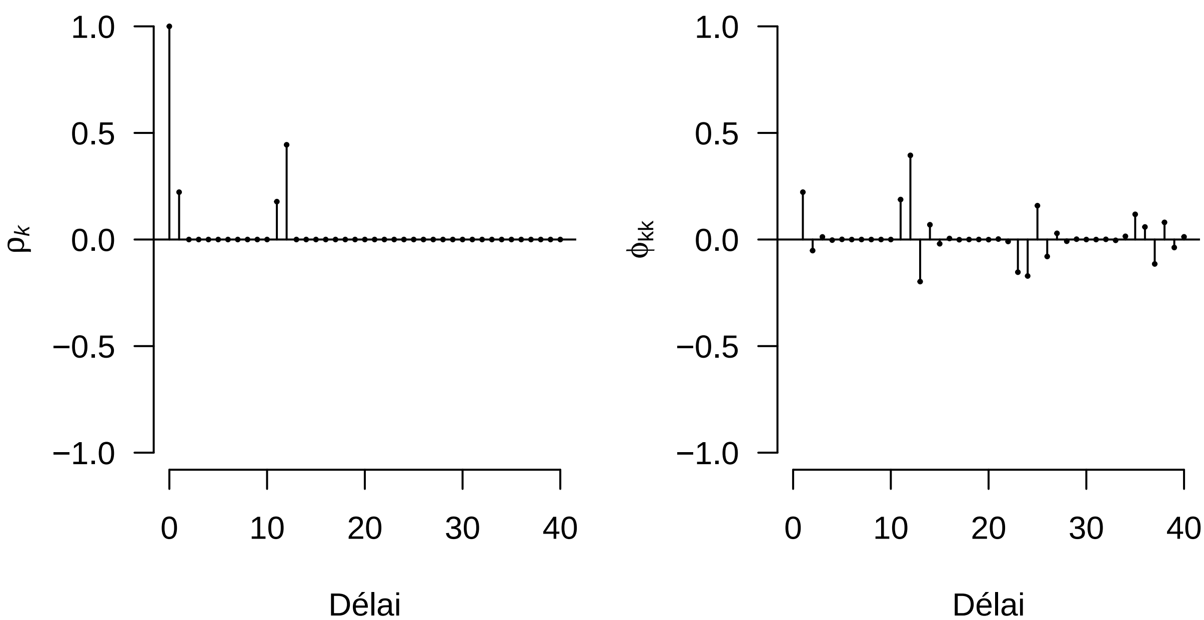
1 S'inspirant du script 1, tracer et commenter les fonctions d'autocorrélations des modèles suivants:

(a) $Y_t = (1 + \Theta B^{12})\epsilon_t$



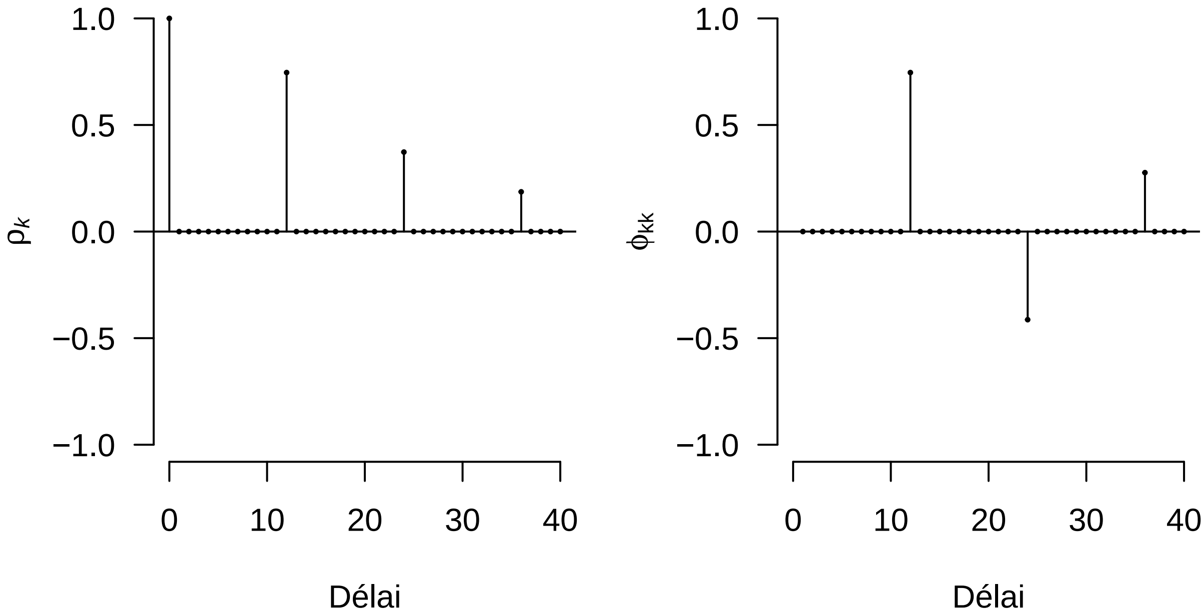
La décroissance de l'ACF de 12 en 12 est assez lente, symptôme de saisonnalité. De plus, sur le PACF, les autocorrélations s'annullent après chaque retard multiple de 12. On pourrait envisager un modèle MA(1).

(b) $Y_t = (1 + \theta B)(1 + \Theta B^{12})\epsilon_t$



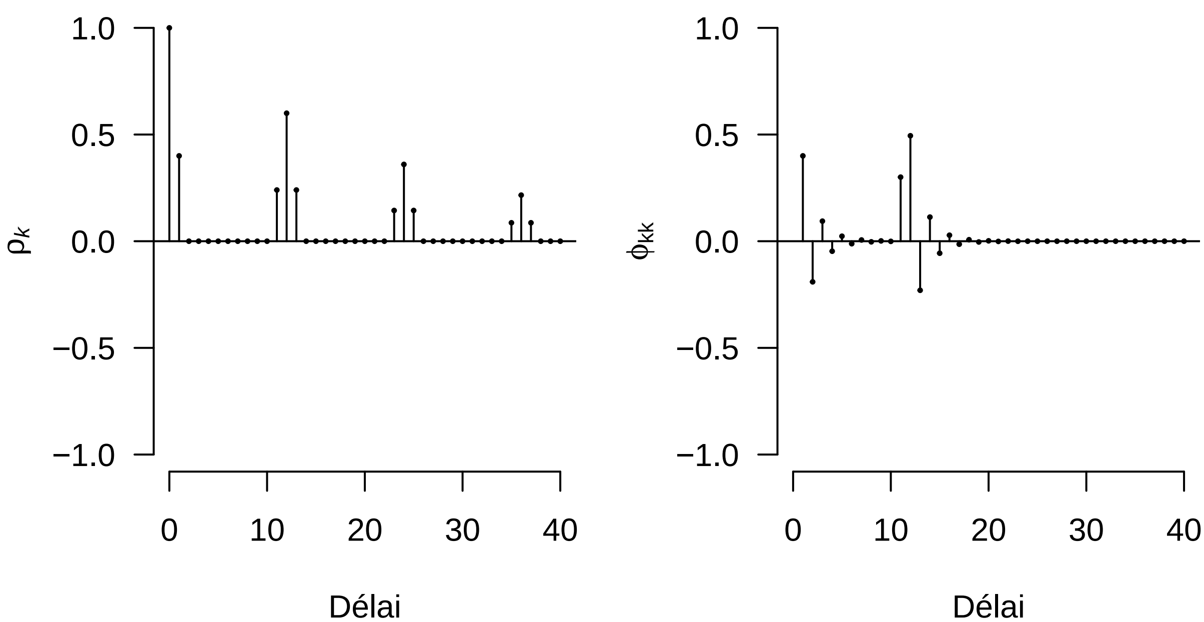
On remarque sur l'ACF une décroissance exponentielle sur le retard 0 et le retard 12.

(c) $(1 - \Phi B^{12})Y_t = (1 + \Theta B^{12})\epsilon_t$



Cela ressemble fortement à notre premier modèle, avec des pics décroissants sur les retards de 12 notés sur l'ACF et d'autre part, des pics de signes contraires sur le PACF. On pourrait donc envisager un modèle MA(2).

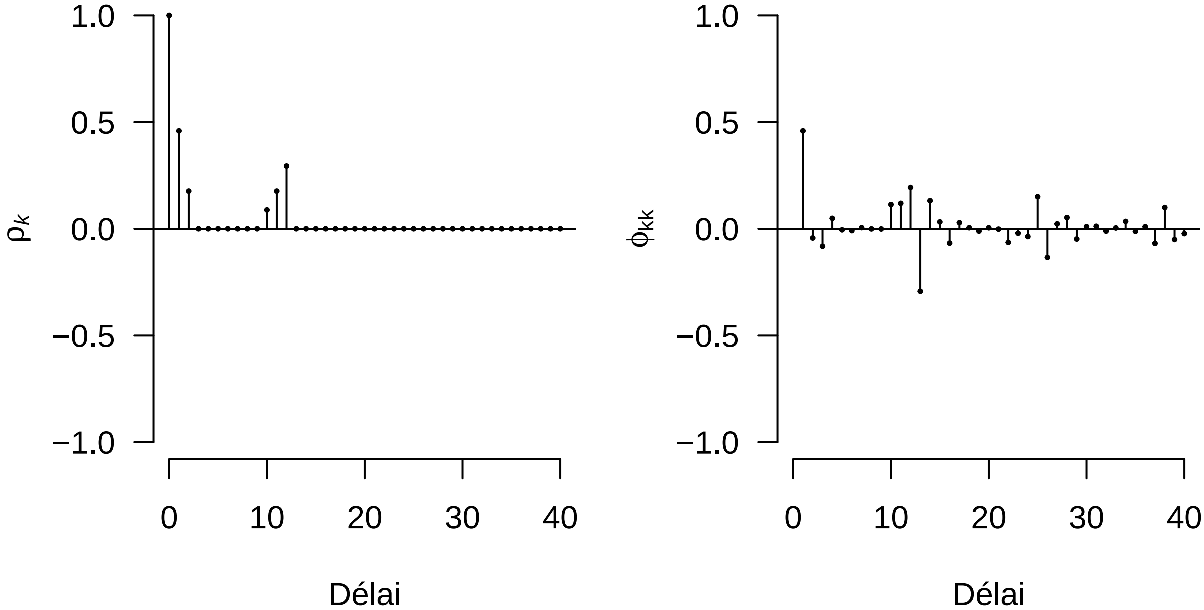
(d) $(1 - \Phi B^{12})Y_t = (1 + \theta B)\epsilon_t$



Les autocorrélations théoriques ont une forme sinisoïdale amortie. De plus la PACF présente des autocorrélations qui sembles significatives sur les retards 1, 11 et 12. On pourrait donc envisager un modèle

ARMA(1,1)

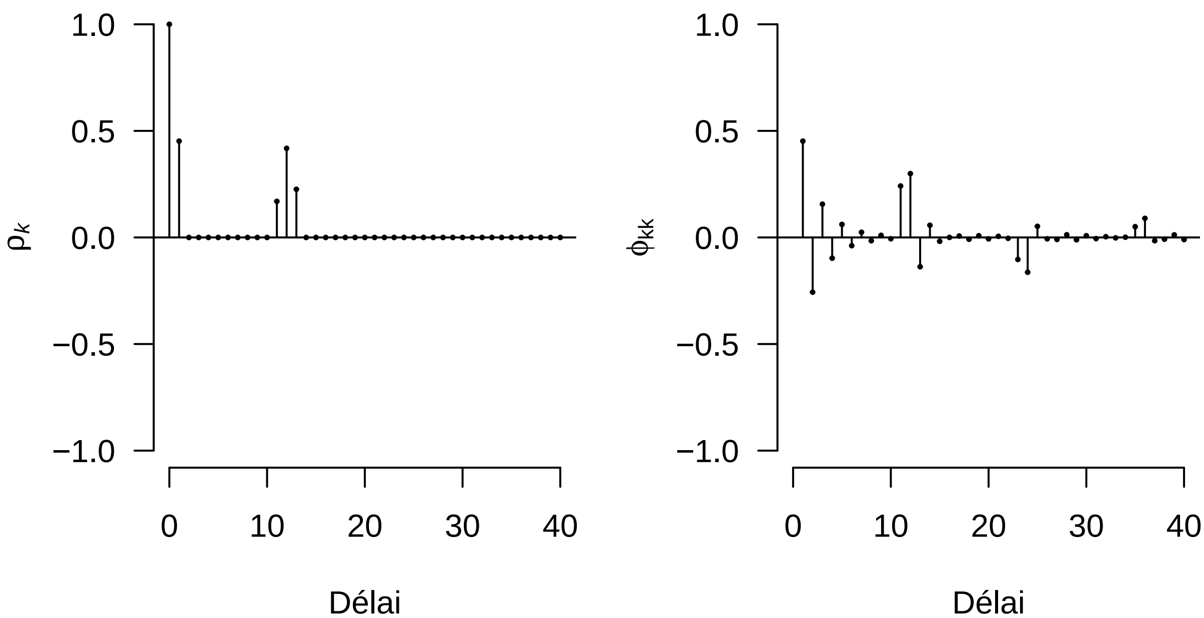
(e) $Y_t = (1 + \theta_1 B + \theta_2 B^2)(1 + \Theta_1 B^{12})\epsilon_t$



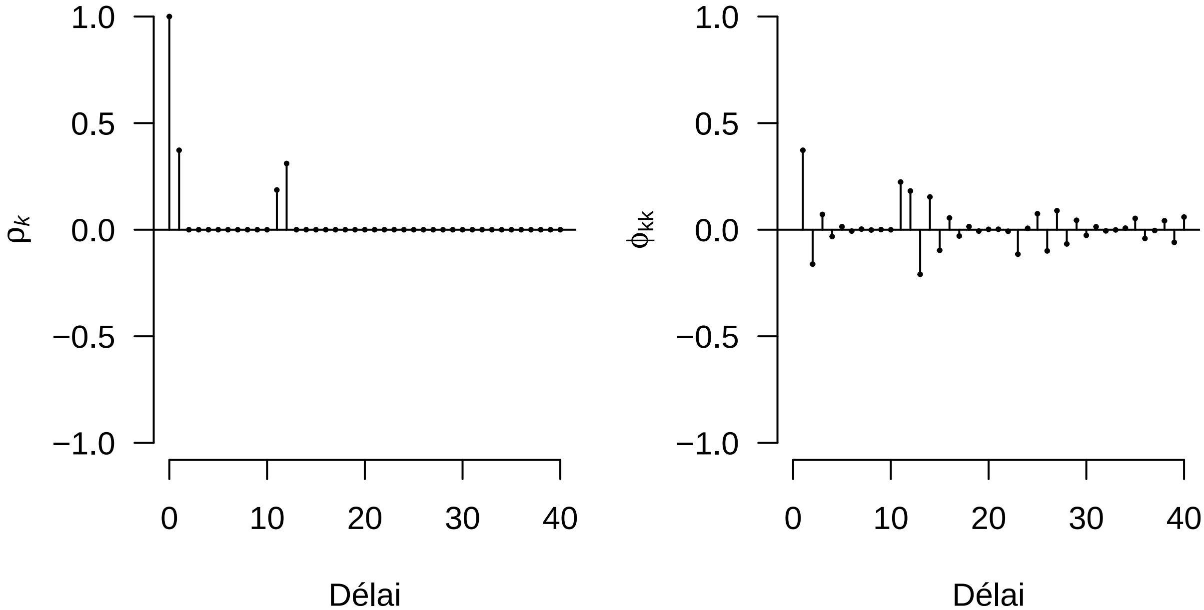
On est en présence d'un modèle multiplicatif et saisonnier. Sur la PACF les retards multuples de 12 changent de signe et seuls les retards 1 et 12 semblent significatifs. Le modèle pourrait donc être MA(2).

2. En choisissant des valeurs intéressantes pour chacun des paramètre et en s'inspirant du script 1, tracer et commenter les fonctions d'autocorrélation des modèles suivants

(a) $Y_t = (1 + \theta_1 B + \theta_{12} B^{12} + \theta_{13} B^{13})\epsilon_t$



(b) $Y_t = (1 + \theta_1 B)(1 + \Theta B^{12})\epsilon_t$

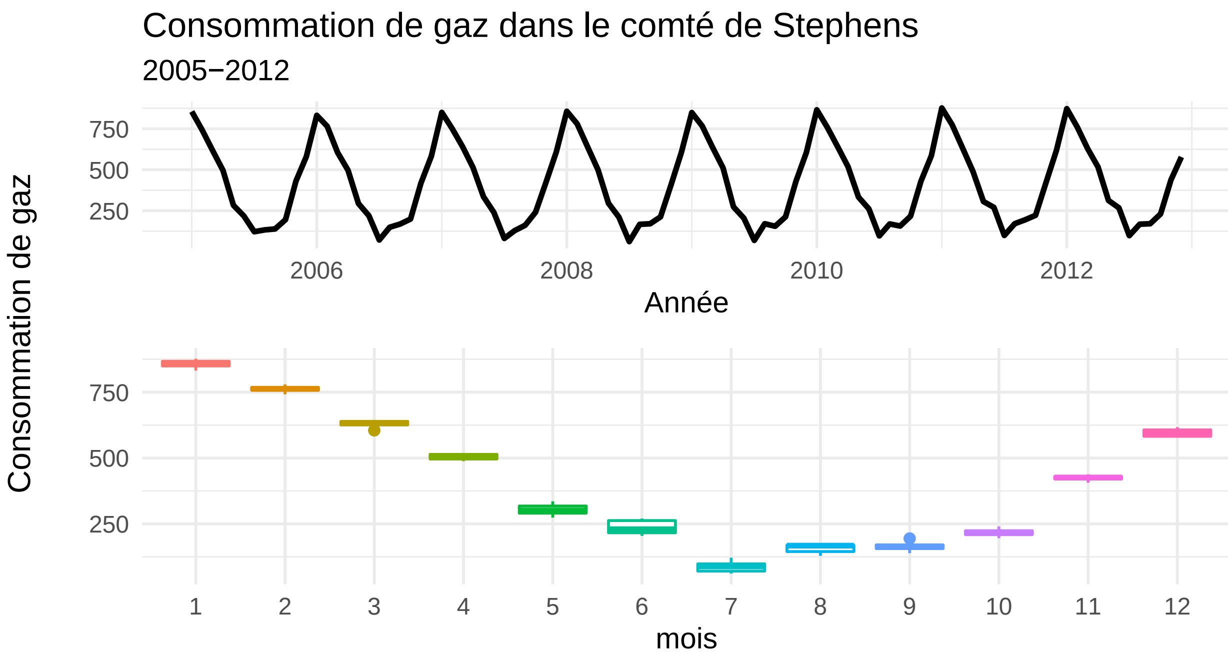


On remarque que la saisonnalité est plus marquée sur (a) et de plus les autocorrélations empiriquent s'annulent plus rapidement par rapport à (b)

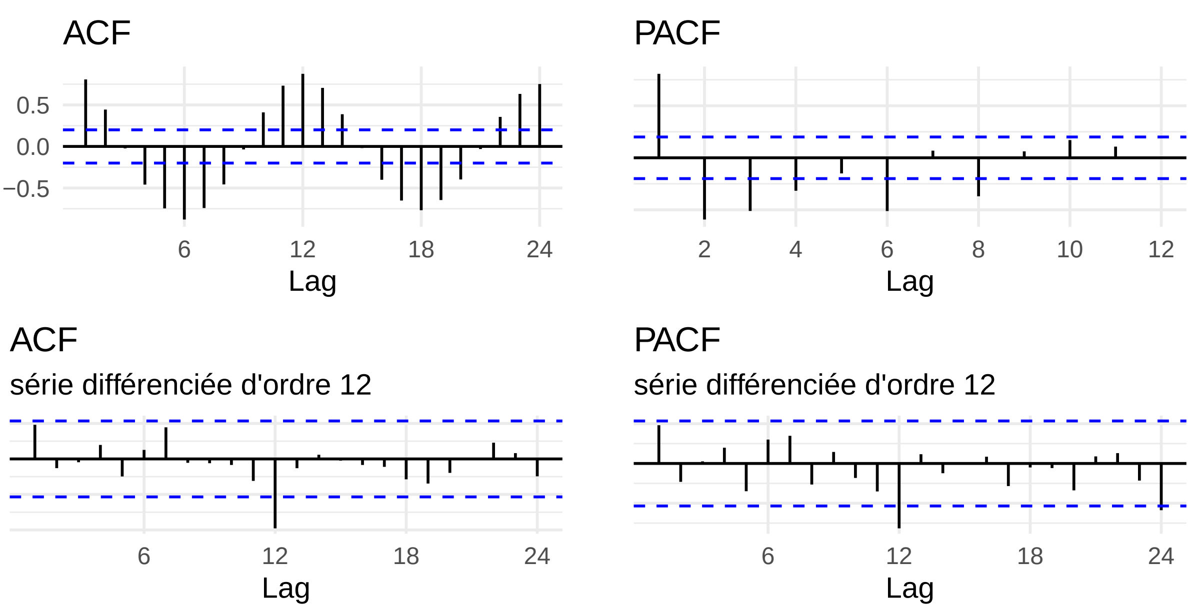
4. Etudes de cas

4.1. Consommation de gaz dans le comté de Stephens

1. Retirez les 12 dernières observations de la série et identifier un modèle pour cette série tronquée



La consommation de gaz dans le comté de Stephens a une tendance très saisonnière avec très peu de variance. On note des baisses de consommations importantes pendant l'été et à moindre mesure pendant l'automne. Ceci est peut être dû a un surplus de consommation des chaudières pendant l'hiver qui peut, en moyenne, atteindre 850 en janvier.



On observe que l'ACF présente des retards significatifs en forme sinisoïdale et des pics aux retards de 12 qui ne diminuent que très lentement. L'ACF de la série différenciée montre un pic au retard 12 suivi d'une forte atténuation, caractéristique d'une série stationnaire avec saisonalité.
On pourrait donc envisager un modèle **ARIMA(1,0,0)(2,1,0)[12]** avec une différenciation saisonnière.

2. Estimer les paramètres de ce modèle, pour cette série tronquée

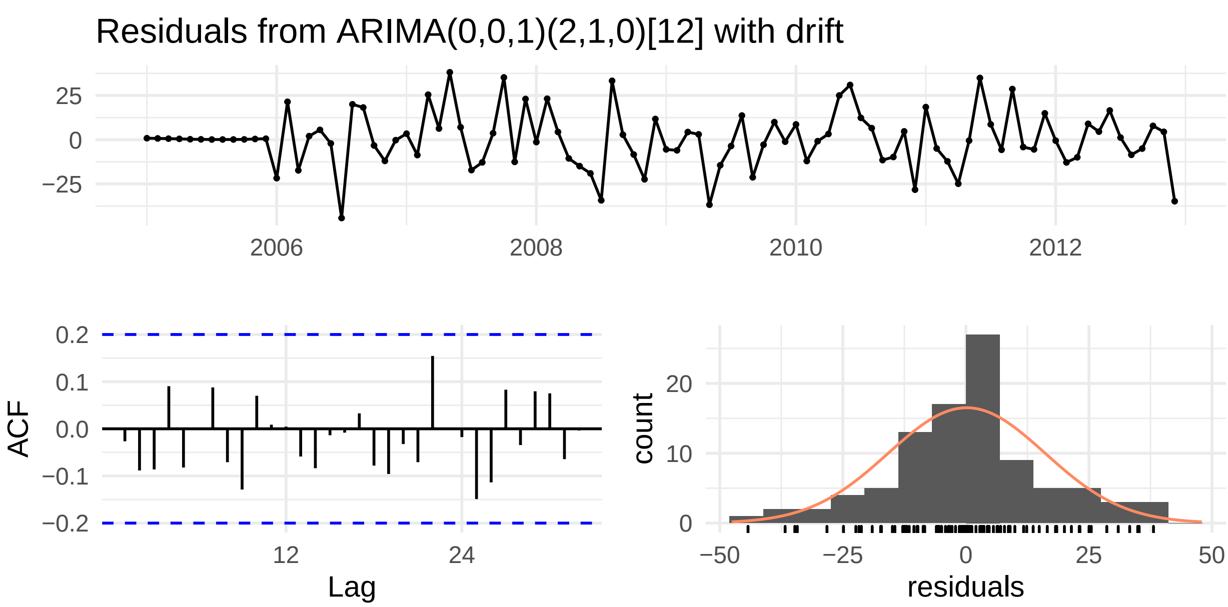
Les parmètres estimés du modèle sont donnés par le tableau suivant:

Call: Arima(y = gaz.tr, order = c(0, 0, 1), seasonal = c(2, 1, 0), include.constant = TRUE)

TABLE 3: Coefficients

	ma1	sar1	sar2	drift
	0.2104	-0.5958	-0.3681	0.2505
s.e.	0.1186	0.1094	0.1147	0.1057

sigma^2 estimated as 301.1: log likelihood = -359.88, aic = 729.76




```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(0,0,1)(2,1,0)[12] with drift
## Q* = 10.03, df = 15, p-value = 0.8178
##
## Model df: 4.   Total lags used: 19
```

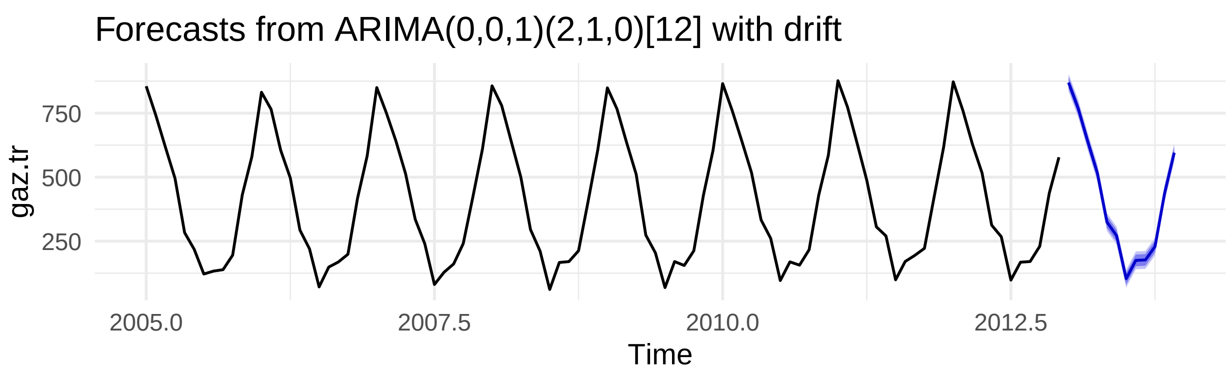
Les résidus semblent stationnaires et suivent une loi normale. Aucune autocorrélation des erreurs n'est significative et le test de Ljung-Box donne une p-valeur = 0.8178. Ce qui confirme que les erreurs ne sont pas autocorrélées.

3. Donner de manière précise l'équation du modèle

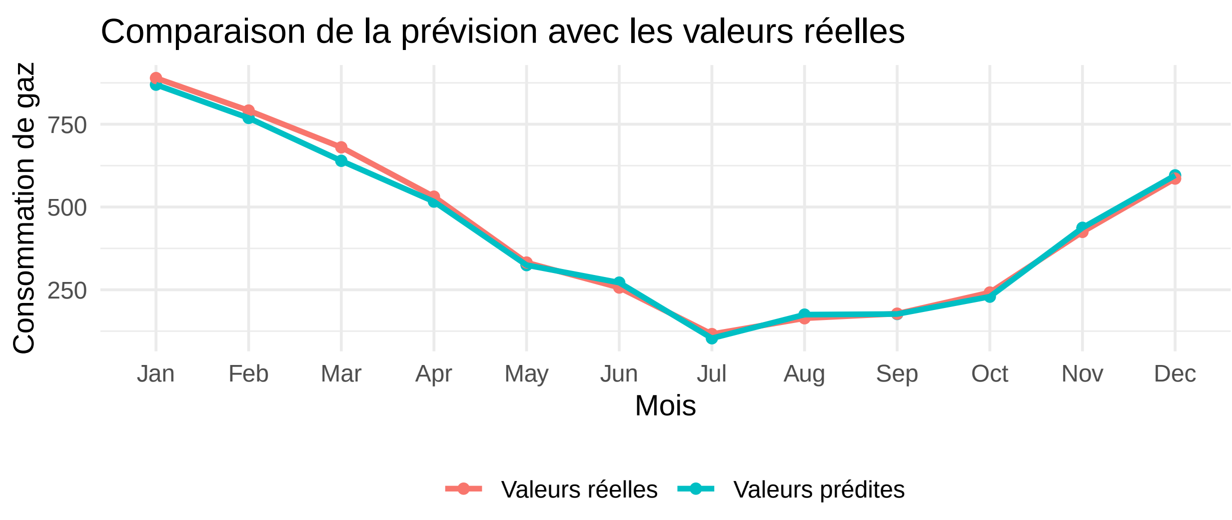
L'équation explicite du modèle est:

$$(1 - 0.5958B - 0.3681B^2)\nabla\nabla_{12}Y_t = (1 + 0.2104B^{12})\epsilon_t$$

4. Faites les prévisions des 12 valeurs à partir de janvier 2013



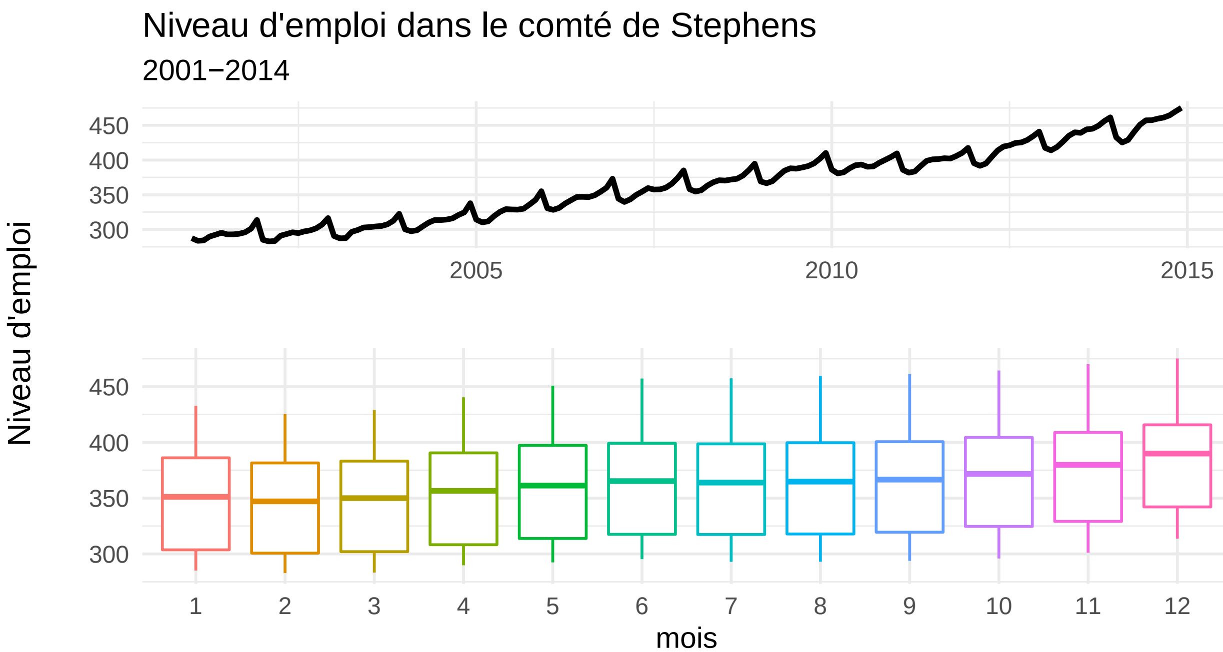
5. Comparer les prévisions aux valeurs réelles et commenter



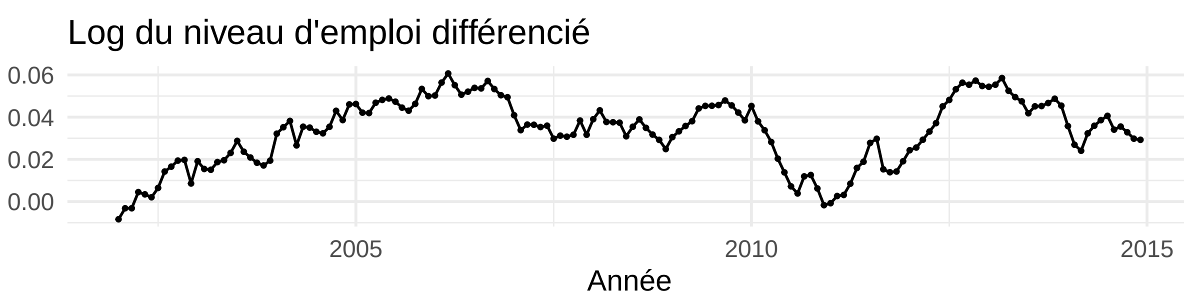
Dans l'ensemble, les valeurs prédites sont très proches des valeurs réelles nottamment entre Avril et Décembre où la variance de l'erreur est moindre.

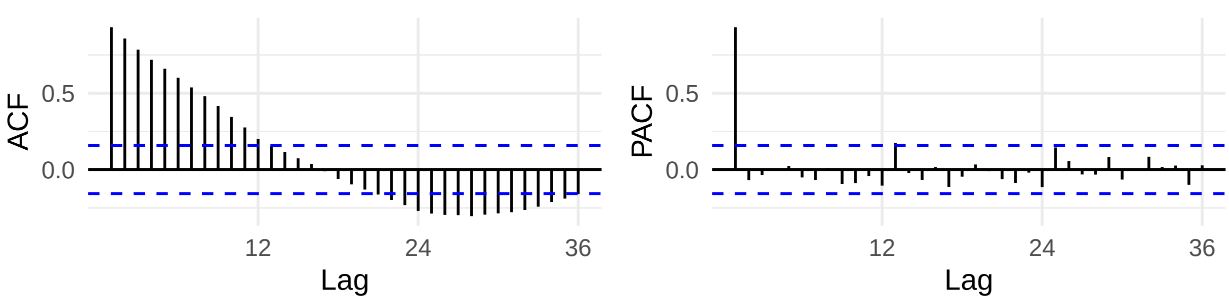
4.2. Niveau d’emploi dans le compté de Stephens

1. Retirez les observations de 2015 de cette série et identifier un modèle pour cette série tronquée



Depuis 2004, le niveau d’emploi dans le comté de Stephens manifeste une tendance à la hausse au fur des années avec des variations saisonnières assez marquées. Le nombre moyen du niveau d’emploi a tendance à diminuer entre Janvier et Mars puis s’en suit une hausse régulière qui atteint son pic maximal au mois de Décembre.
Nous remarquons aussi qu’il y’a une légère augmentation de la variance entre 2012 et 2015. Pour palier à ce phénomène, nous allons utiliser par la suite le logarithme.





L’ACF montre une dimunition linéaire des autocorrélations.

Sur le PACF, on observe une diminution brusque des autocorrélations à partir du retard 1 qui est d'ailleurs le seul à être significatif.
On peut donc prendre comme modèle un **ARIMA(0,1,1)(1,1,1)[12]**

2. Estimer les paramètres de ce modèle, pour cette série tronquée

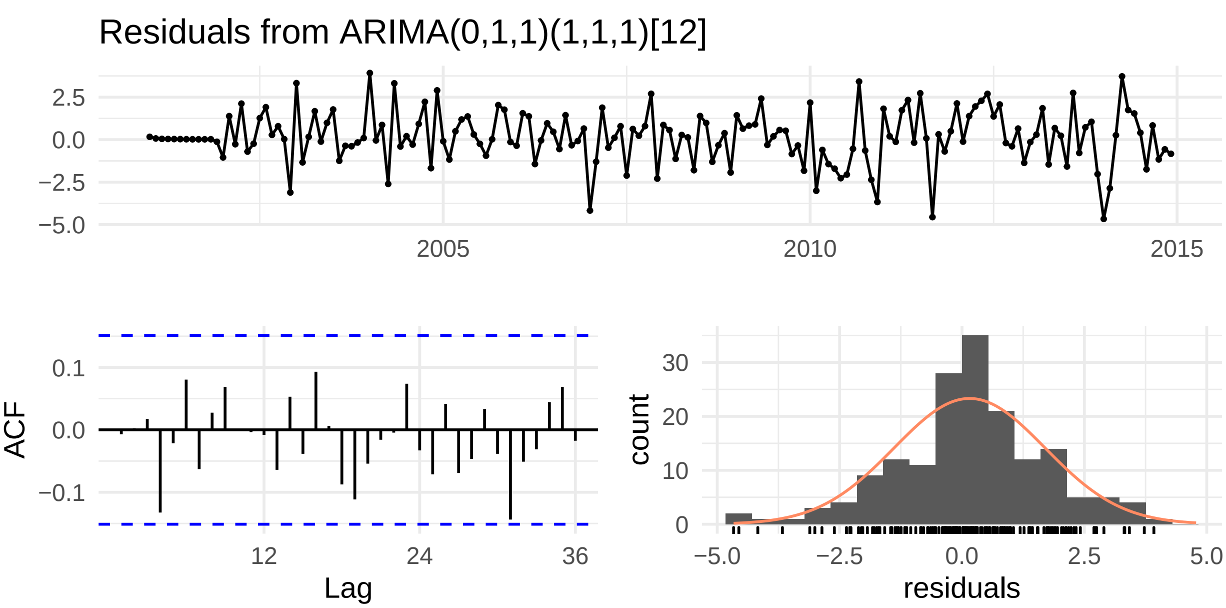
Les parmètres estimés du modèle sont donnés par le tableau suivant:

Call: Arima(y = emp.tr, order = c(0, 1, 1), seasonal = c(1, 1, 1))

TABLE 4: Coefficients

	ma1	sar1	sma1
	0.1978	0.3777	-0.7137
s.e.	0.0793	0.157	0.1225

sigma^2 estimated as 2.641: log likelihood = -294.9, aic = 597.81




```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(0,1,1)(1,1,1)[12]
## Q* = 14.984, df = 21, p-value = 0.8238
##
## Model df: 3.   Total lags used: 24
```

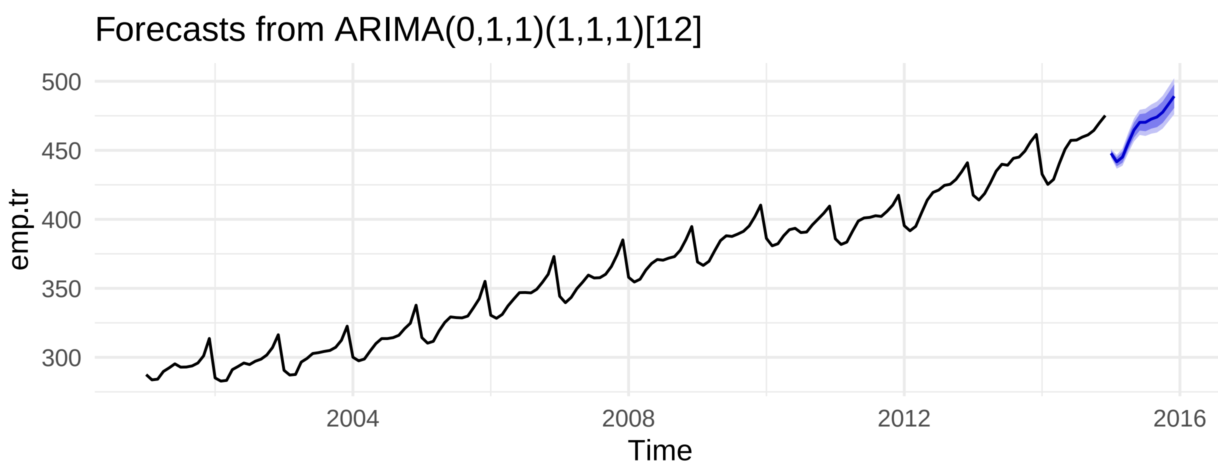
Les résidus semblent stationnaires et suivent une loi normale. Aucune autocorrélation des erreurs n'est significative et le test de Ljung-Box donne une p-valeur = 0.8238. Ce qui confirme que les erreurs ne sont pas autocorrélées.

3. Donner de manière précise l'équation du modèle estimé

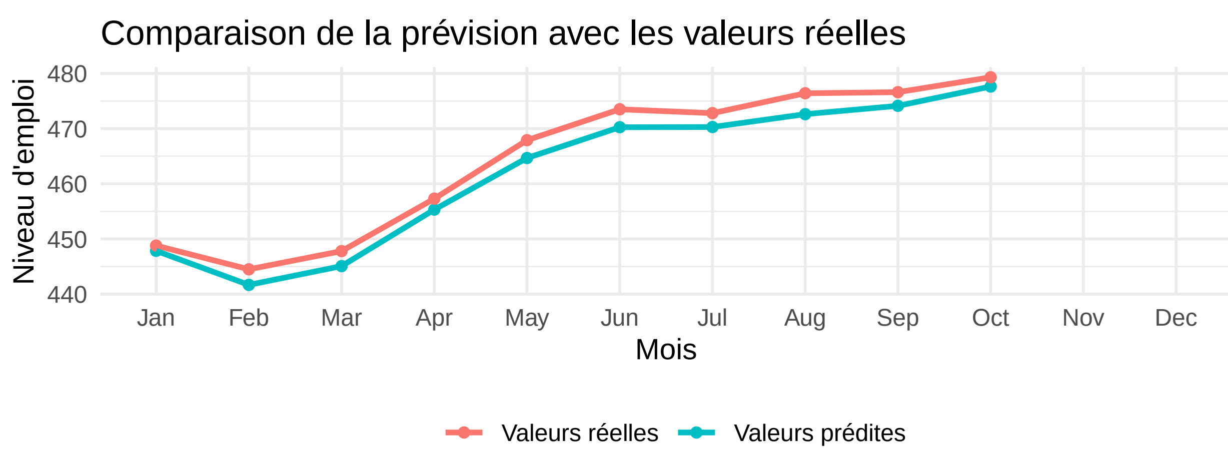
L'équation explicite du modèle est:

$$(1 + 0.3777B)\nabla_{12}Y_t = (1 + 0.1978B - 0.7137B^{12})\epsilon_t$$

4. Faites les prévisions des valeurs de 2015



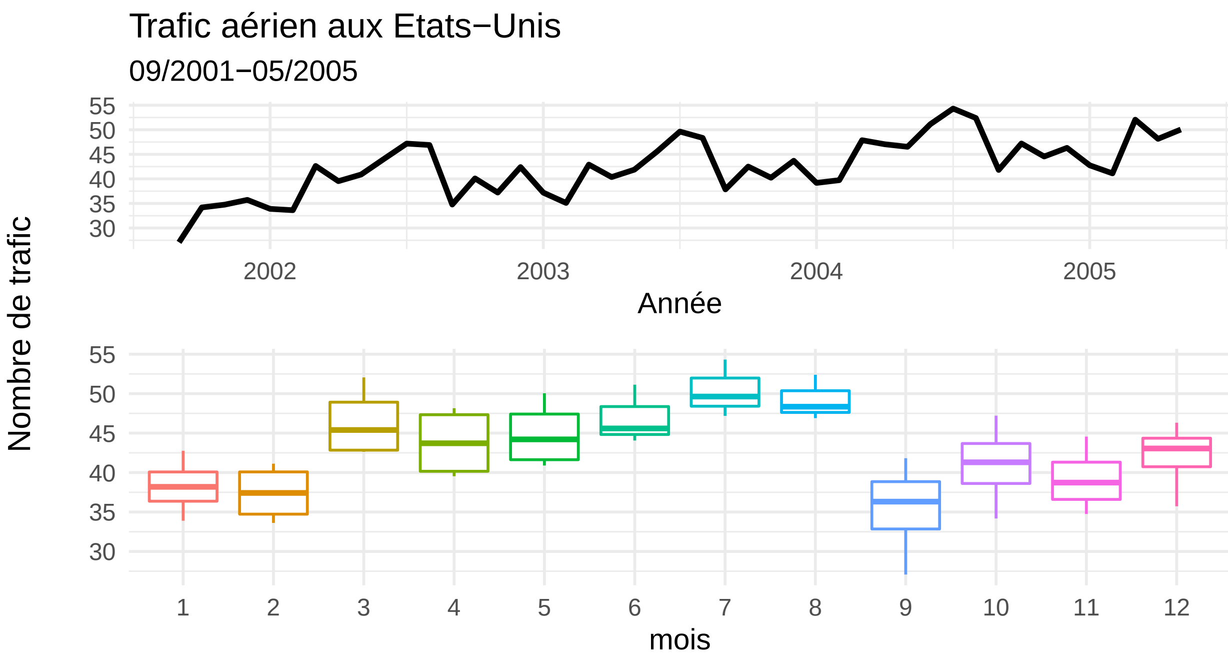
5. Comparer les prévisions aux valeurs réelles et commenter



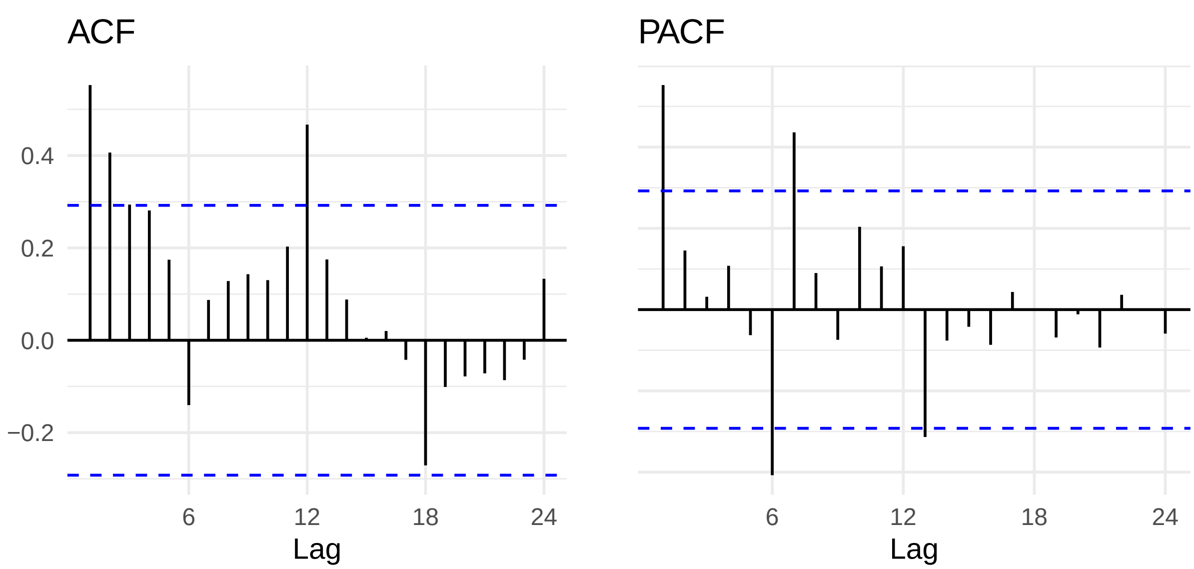
Quelque soit le mois, les valeurs prédites sont inférieures aux valeurs réelles surtout entre le mois de Mai et le mois d'Août où l'on observe une variance des erreurs plus importante.

4.3. Trafic aérien aux Etats-Unis

1. Retirez les observations à partir du mois de Septembre 2001 de cette série et identifier un modèle



Le trafic aérien aux Etats-Unis montre une tendance croissante au fur des années avec une saisonnalité prononcée. Le trafic est plus dense entre le printemps et l'été et plus particulièrement pendant le mois de Juillet où l'on observe des pics pouvant aller jusqu'à 54.7. Le mois de Septembre enregistre la plus faible activité qui peut être expliquée par la fin de saison des vacances d'été en Août.



Les autocorrélations empiriques sont significatives principalement aux retards 2 et 12 avec une décroissance exponentielle sur les 5 premiers retards. De plus la PACF semble s'atténuer plus vite que l'ACF après le retard 12.

On privilégie donc un autorégressif saisonnier d'ordre 1: **ARIMA(1,0,0)(1,1,0)[12]**

2. Estimer les paramètres de ce modèle, pour cette série tronquée

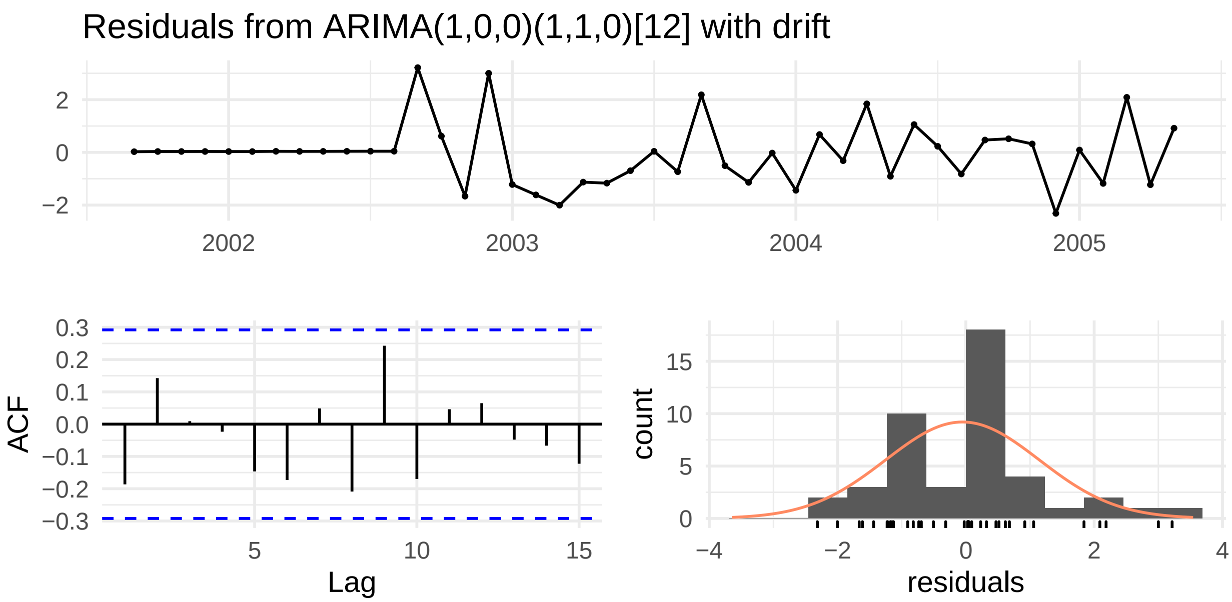
Les parmètres estimés du modèle sont donnés par le tableau suivant:

Call: Arima(y = trafic.tr, order = c(1, 0, 0), seasonal = c(1, 1, 0), include.constant = TRUE)

TABLE 5: Coefficients

	ar1	sar1	drift
	0.3995	-0.5523	0.291
s.e.	0.1923	0.1774	0.02489

sigma^2 estimated as 2.111: log likelihood = -59.84, aic = 127.69




```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(1,0,0)(1,1,0)[12] with drift
## Q* = 11.55, df = 6, p-value = 0.07278
##
## Model df: 3.   Total lags used: 9
```

Malgré que les résidus semblent dissymétrique entre la fin de l'anné 2002 et le début de l'anné 2003, le test de Ljung-Box est significatif au seuil de 5%(p-valeur = 0.07278). Ils ne sont donc pas autocorrélés et leur densité montre qu'ils suivent une loi normale.

3. Donner de manière précise l'équation de ce modèle

L'équation explicite du modèle est:

$$(1 + 0.3995B)\nabla_{12}Y_t = (1 - 0.5523B^{12})\epsilon_t$$