Séries temporelles – Correction du TD/TP 6
Estimer, comparer et diagnostiquer des modèles ARMA

M1 Actuariat – Université Gustave Eiffel

Exercice 1 – Comprendre un AR(1) estimé

Le modèle proposé est

$$X_t = c + \phi X_{t-1} + Z_t, \qquad \widehat{c} = 21, \quad \widehat{\phi} = 0{,}65, \quad \mathrm{SE}(\widehat{\phi}) = 0{,}09, \quad \widehat{\sigma}^2 = 25.$$

On interprète des *coefficients estimés* à partir des 150 semaines, et non des paramètres qui seraient connus avec certitude.

1) **Distinguer la constante et la moyenne stationnaire.**
Étape 1 : prendre l'espérance. La stationnarité suppose en particulier que $\mathbb{E}[X_t] = \mathbb{E}[X_{t-1}] = \mu$. Les innovations étant centrées, $\mathbb{E}[Z_t] = 0$. Ainsi,

$$\begin{aligned} \mathbb{E}[X_t] &= c + \phi \mathbb{E}[X_{t-1}] + \mathbb{E}[Z_t], \\ \mu &= c + \phi\mu, \\ (1-\phi)\mu &= c. \end{aligned}$$

Comme $|\phi| < 1$, on a $1 - \phi \neq 0$, et donc

$$\boxed{\mu = \frac{c}{1-\phi}}.$$

Étape 2 : remplacer par les estimations.

$$\widehat{\mu} = \frac{21}{1-0{,}65} = \frac{21}{0{,}35} = \boxed{60 \text{ demandes par semaine}}.$$

Le collègue a donc **tort** : 21 est le terme constant de la relation de récurrence, pas la moyenne stationnaire. Une semaine moyenne correspond ici à environ 60 demandes, selon le modèle ajusté.
Étape 3 : centrer le modèle. Puisque $\widehat{c} = (1 - \widehat{\phi})\widehat{\mu}$, on peut réécrire

$$\begin{aligned} X_t &= 21 + 0{,}65 X_{t-1} + \widehat{Z}_t, \\ X_t - 60 &= 0{,}65(X_{t-1} - 60) + \widehat{Z}_t. \end{aligned}$$

Cette forme centrée permet une interprétation concrète. Si, une semaine, $X_{t-1} = 70$, soit 10 demandes *au-dessus* de la moyenne, la composante ajustée à la semaine suivante se situe $0{,}65 \times 10 = 6{,}5$ demandes au-dessus de 60, c'est-à-dire à 66,5. Le coefficient 0,65 décrit donc la **persistance des écarts à la moyenne**, et non la moyenne elle-même.

2) **Calculer les valeurs ajustées et les résidus.**
Pour chaque semaine, il faut utiliser **l'observation de la semaine précédente** :

$$\widehat{x}_t = 21 + 0{,}65 x_{t-1}, \qquad \widehat{z}_t = x_t - \widehat{x}_t.$$

Attention : dans ces calculs sur les données déjà utilisées pour estimer le modèle, $\widehat{x}_t$ est une *valeur ajustée*. Ce n'est pas, à proprement parler, une prévision hors échantillon.

Semaine 31. On connaît $x_{30} = 65$ et $x_{31} = 63$:

$$\widehat{x}_{31} = 21 + 0{,}65 \times 65 = 63{,}25, \qquad \widehat{z}_{31} = 63 - 63{,}25 = \boxed{-0{,}25}.$$

Semaine 32. Le retard pertinent est maintenant $x_{31} = 63$, et non $\widehat{x}_{31}$:

$$\widehat{x}_{32} = 21 + 0{,}65 \times 63 = 61{,}95, \qquad \widehat{z}_{32} = 70 - 61{,}95 = \boxed{8{,}05}.$$

Semaine 33. On utilise $x_{32} = 70$:

$$\widehat{x}_{33} = 21 + 0{,}65 \times 70 = 66{,}50, \qquad \widehat{z}_{33} = 56 - 66{,}50 = \boxed{-10{,}50}.$$

Semaine	31	32	33
Observation x_t	63	70	56
Observation précédente x_{t-1}	65	63	70
Valeur ajustée $\widehat{x}_t$	63,25	61,95	66,50
Résidu $\widehat{z}_t$	−0,25	+8,05	−10,50

Interprétation du signe. Le résidu est la différence « observé moins ajusté ». Un résidu négatif signifie que la caisse a reçu *moins de demandes que la valeur ajustée par le modèle* pour cette semaine. Le résidu $-10{,}5$ ne veut évidemment pas dire que le nombre de demandes est négatif.

3) **Un résidu important suffit-il à invalider le modèle ?**

Non. Même si le modèle est bien spécifié, l'innovation Z_t est aléatoire : une observation peut donc s'écarter sensiblement de sa valeur ajustée. Il faut apprécier la taille de l'écart *relativement à la dispersion habituelle*.

Ici, $\widehat{\sigma}^2 = 25$, donc l'écart-type estimé des innovations est

$$\widehat{\sigma} = \sqrt{25} = 5.$$

Pour la semaine 33, le résidu représente environ

$$\frac{|\widehat{z}_{33}|}{\widehat{\sigma}} = \frac{10{,}5}{5} = \boxed{2{,}1}$$

écarts-types estimés. C'est un écart notable : on peut aller regarder cette date plus attentivement. Mais cela n'établit pas à lui seul une mauvaise spécification. Le modèle n'ayant pas été supposé gaussien dans l'énoncé, on ne calcule pas ici une p-valeur normale à partir de 2,1.

Pour juger le modèle, on examine plutôt **l'ensemble des résidus** : répartition autour de zéro, variation de leur dispersion, autocorrélations et éventuelles valeurs atypiques répétées.

4) **Interpréter l'erreur standard et établir un diagnostic.**

Deux quantités différentes.

— $\mathrm{SE}(\widehat{\phi}) = 0{,}09$ estime l'écart-type de *l'estimateur* $\widehat{\phi}$ si l'étude était répétée sur de nouvelles séries analogues. Elle renseigne sur la *précision avec laquelle le coefficient AR est estimé*.

— $\widehat{\sigma} = 5$ estime l'écart-type des *innovations*, c'est-à-dire la dispersion des variations que le modèle ne décrit pas. Son unité est le nombre de demandes ; le coefficient ϕ et son erreur standard n'ont pas d'unité.

À titre d'illustration, une approximation normale donnerait pour ϕ l'intervalle indicatif

$$\widehat{\phi} \pm 1{,}96\,\mathrm{SE}(\widehat{\phi}) = 0{,}65 \pm 1{,}96 \times 0{,}09 \simeq [0{,}47\,;\,0{,}83].$$

Cet intervalle est uniquement une *illustration* de la notion de précision ; sa validité exige les hypothèses asymptotiques habituelles.

Le pic à trois semaines. L'ACF résiduelle présente $\widehat{\rho}_{\text{res}}(3) = -0{,}23$. Cela suggère une corrélation négative entre résidus séparés de trois semaines : après un résidu plutôt positif, celui observé trois semaines plus tard tendrait, dans une relation linéaire, à être plutôt négatif. Pour $n = 150$, les repères graphiques approximatifs $\pm 1{,}96/\sqrt{150}$ valent $\pm 0{,}16$: le pic $-0{,}23$ mérite donc l'attention. Cela ne constitue toutefois pas une preuve isolée de mauvaise spécification.

Le test global. Ljung–Box examine ici

$$H_0 : \rho_{\text{res}}(1) = \cdots = \rho_{\text{res}}(12) = 0$$

contre l'alternative selon laquelle au moins une de ces autocorrélations n'est pas nulle. Comme $p = 0{,}018 < 0{,}05$, on **rejette H_0 au seuil de** 5 %. Il reste donc une dépendance temporelle linéaire significative dans les résidus, aux retards étudiés, selon ce test.

Que faire ensuite ? On ne saute pas directement à un modèle précis. On regarde l'ACF et la PACF des données et des résidus, puis on essaie quelques modèles candidats plausibles (par exemple AR(2), AR(3), ou ARMA). On estime ces modèles, on compare AIC/BIC et on recommence le diagnostic résiduel. Un pic au retard 3 n'implique *pas automatiquement* qu'il faille ajouter un coefficient AR au retard 3.

Remarque

Conclusion de l'exercice 1. Le modèle suggère une moyenne de 60 demandes par semaine, et non 21. Les résidus quantifient les écarts entre valeurs observées et ajustées, leur dispersion ne se confond pas avec l'incertitude sur les coefficients. Le test de Ljung–Box est défavorable à cet AR(1) : son ajustement ne semble pas avoir expliqué toute la corrélation temporelle.

Exercice 2 – Faut-il suivre AIC, BIC ou les résidus ?

Les trois modèles sont ajustés *sur les mêmes* 180 *observations mensuelles* avec la même méthode. Leurs AIC/BIC sont donc comparables dans le cadre de l'exercice. Les diagnostics portent sur les résidus des modèles déjà estimés.

1) **Comprendre le désaccord entre AIC et BIC.**

Si l'on ne compare que A et B, l'AIC est minimal pour B : $408{,}0 < 410{,}0$. Le BIC est minimal pour A : $419{,}6 < 420{,}8$.

Pourquoi ces conclusions peuvent-elles différer ? B est un ARMA$(1, 1)$, tandis que A est un AR(1) : B introduit *un coefficient MA supplémentaire*. Ce coefficient peut améliorer la vraisemblance sur les données utilisées, mais il rend aussi le modèle plus complexe. Le BIC, qui pénalise plus fortement cette complexité pour $n = 180$, juge ici que le gain ne justifie pas l'ajout du paramètre. L'AIC arrive à une appréciation différente.

Il ne faut donc pas conclure que l'un des critères « se trompe » : ils appliquent *des compromis ajustement/complexité différents*. À ce stade, on ne sait pas encore si les résidus de l'un ou l'autre modèle ont un comportement satisfaisant.

2) **Pourquoi les résidus remettent-ils en cause A et B ?**

Les ACF résiduelles montrent au retard 12 des valeurs 0,30 pour A et 0,25 pour B. Ici, un retard 12 signifie *douze mois*, donc une année. Une autocorrélation positive suggère que les écarts à la valeur ajustée tendent à se ressembler pour deux mois situés à un an d'intervalle : les deux modèles ont peut-être oublié une structure annuelle.

Cette impression graphique est cohérente avec les tests de Ljung–Box à $m = 24$:

$$A : \ p = 0{,}002 < 0{,}05 \quad \Longrightarrow \text{rejet de } H_0,$$
$$B : \ p = 0{,}013 < 0{,}05 \quad \Longrightarrow \text{rejet de } H_0.$$

Le test est *global* : il examine ensemble les autocorrélations aux retards 1 à 24, et pas uniquement celle au retard 12. Les deux résultats indiquent qu'il reste de la corrélation temporelle dans les résidus. Cela rend A et B peu convaincants *malgré* leurs bonnes valeurs d'AIC ou de BIC.

Attention au raisonnement : AIC et BIC servent à *classer des modèles candidats*, pas à démontrer que les résidus d'un modèle sont compatibles avec un bruit blanc faible.

3) **En quoi le modèle C est-il adapté à une dépendance annuelle ?**

Le modèle C est SARIMA$(1,0,0)(1,0,0)_{12}$. Ses deux parties autorégressives sont

$$\phi(B) = 1 - \phi B, \qquad \Phi(B^{12}) = 1 - \Phi B^{12}.$$

La forme centrée de son équation est

$$(1 - \phi B)(1 - \Phi B^{12})(X_t - \mu) = Z_t.$$

Le polynôme saisonnier contient B^{12} : il relie donc les observations séparées de *douze mois*. Il est raisonnable de l'envisager si les diagnostics des premiers modèles suggèrent une dépendance à cette échelle temporelle. Le modèle comporte aussi une partie AR d'ordre 1, qui représente une dépendance à l'échelle d'un mois.

Pour mieux voir les retards (facultatif) : en développant le produit, on obtient

$$X_t - \mu = \phi(X_{t-1} - \mu) + \Phi(X_{t-12} - \mu) - \phi\Phi(X_{t-13} - \mu) + Z_t.$$

Le terme au retard 13 provient de l'*interaction* entre les parties ordinaire et saisonnière ; il n'est pas associé à un troisième coefficient libre.

4) **Choisir un modèle avec des arguments statistiques.**

Pour C, l'autocorrélation résiduelle au retard 12 vaut seulement 0,04. De plus, $p = 0{,}39 > 0{,}05$ au test de Ljung–Box : on **ne rejette pas** H_0 d'absence d'autocorrélation aux retards $1, \ldots, 24$. Enfin, le chronogramme est décrit comme ayant une dispersion globalement stable.

Un choix cohérent est donc de **retenir provisoirement** C. Il ne minimise ni AIC ni BIC, mais les écarts de critères restent modestes : par rapport au meilleur AIC, $408{,}4 - 408{,}0 = 0{,}4$, et par rapport au meilleur BIC, $421{,}2 - 419{,}6 = 1{,}6$. En contrepartie, son diagnostic résiduel est nettement plus favorable.

Réponse rédigée possible : « L'AIC privilégie B et le BIC privilégie A, mais les deux modèles laissent une autocorrélation résiduelle détectée par Ljung–Box. Le modèle saisonnier C obtient des critères voisins et ne présente pas ce problème aux retards étudiés. Je retiens donc C parmi les trois candidats, sous réserve d'autres vérifications. »

Limite essentielle : une p-valeur élevée ne *prouve pas* que les résidus sont indépendants ou que le modèle est parfait. Le test ne porte que sur les autocorrélations examinées ; il peut rester une dépendance non linéaire, une rupture de moyenne, des valeurs atypiques ou une variance non constante. Il faudrait aussi, dans une étude appliquée, vérifier la pertinence du modèle pour l'objectif poursuivi.

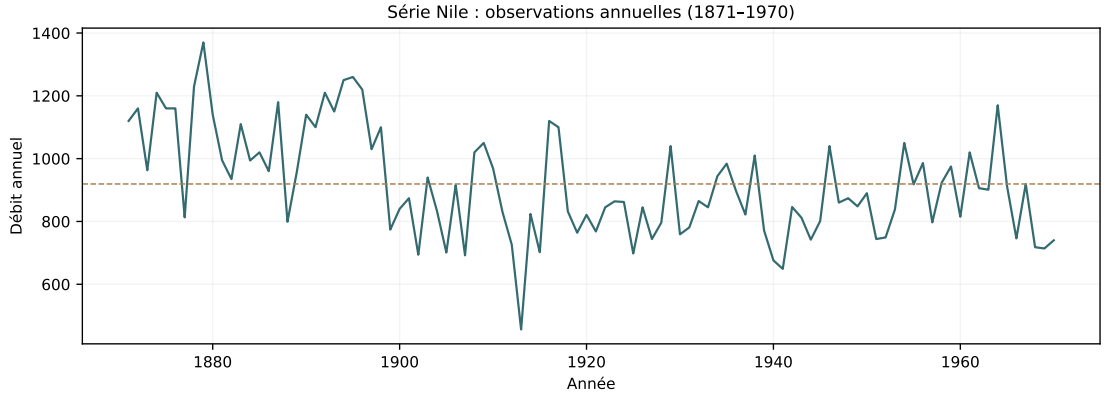
Idée à retenir

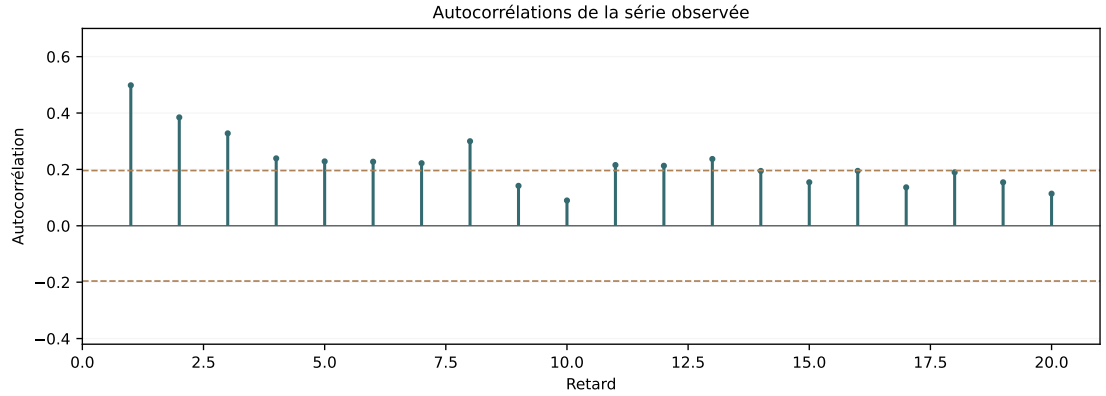
Méthode à retenir : (i) choisir quelques modèles plausibles ; (ii) confronter AIC et BIC ; (iii) vérifier les résidus ; (iv) retenir un modèle justifié *et formuler les limites de la décision*. Un modèle qui minimise un critère n'est pas nécessairement adéquat.

Exercice 3 – La série `Nile` : quel modèle conserve-t-on ?

A. Explorer la série et proposer des candidats

```
data(Nile)
x <- Nile
length(x)
start(x)
end(x)
frequency(x)
plot.ts(x, ylab = "Debit annuel", main = "Serie Nile")
acf(x, main = "ACF de Nile")
```





1) **Décrire les données avant d'estimer un modèle.**

   Lire les commandes : `length(x)` renvoie 100 ; `start(x)` et `end(x)` donnent respectivement 1871 et 1970 ; `frequency(x)` renvoie 1. On dispose donc d'**une observation par année pendant 100 ans**. Il n'existe pas ici douze observations *par an* : une saisonnalité mensuelle $s = 12$ ou trimestrielle $s = 4$ ne serait pas pertinente pour ces données annuelles.

   Lire le premier graphique : le débit varie fortement d'une année à l'autre ; on remarque aussi une possible modification du niveau vers la fin du XIX^{e} siècle ou le début du XX^{e} siècle. Le graphique ne permet pas, à lui seul, de décider si un modèle stationnaire à moyenne constante est parfaitement adapté : une rupture de niveau peut imiter une dépendance persistante.

   Lire l'ACF : plusieurs autocorrélations aux premiers retards sont positives et décroissent progressivement. Les observations d'années proches ne paraissent donc pas sans relation. On ne choisirait pas spontanément un modèle de bruit blanc comme seul candidat. Une ACF décroissante est compatible avec certains modèles AR, mais **ne prouve pas** que la vraie série suit un AR(1).

2) **Écrire ce que les candidats autorisent à expliquer.**

Pour bien distinguer leurs structures, écrivons-les autour d'une moyenne constante μ :

$$\begin{aligned}
\text{Bruit blanc :} \quad & X_t = \mu + Z_t, \\
\text{AR(1) :} \quad & X_t - \mu = \phi_1(X_{t-1} - \mu) + Z_t, \\
\text{AR(2) :} \quad & X_t - \mu = \phi_1(X_{t-1} - \mu) + \phi_2(X_{t-2} - \mu) + Z_t, \\
\text{ARMA(1,1) :} \quad & X_t - \mu = \phi_1(X_{t-1} - \mu) + Z_t + \theta_1 Z_{t-1}.
\end{aligned}$$

Le bruit blanc n'introduit aucune corrélation temporelle ; l'AR(1) utilise une observation passée ; l'AR(2) en utilise deux ; l'ARMA(1,1) combine une observation passée et une innovation passée. **Avant estimation**, la forme de l'ACF ne permet pas d'affirmer lequel sera le meilleur.

B. Estimer et comparer par AIC/BIC

```
f0 <- arima(x, order = c(0, 0, 0), method = "ML")
f1 <- arima(x, order = c(1, 0, 0), method = "ML")
f2 <- arima(x, order = c(2, 0, 0), method = "ML")
f3 <- arima(x, order = c(1, 0, 1), method = "ML")

f1
sqrt(diag(f1$var.coef))
AIC(f0, f1, f2, f3)
BIC(f0, f1, f2, f3)
```

3) **Lire les coefficients, la variance et les erreurs standards.**

La commande `order = c(1, 0, 0)` demande un AR(1), sans différenciation ($d = 0$). `method = "ML"` utilise le maximum de vraisemblance. Dans ce cas stationnaire non différencié, `arima()` estime par défaut aussi la moyenne μ ; la sortie R appelle généralement cette quantité `intercept`. **Ce n'est pas le coefficient constant** c de l'équation $X_t = c + \phi X_{t-1} + Z_t$. On a au contraire $c = (1 - \phi)\mu$.

Pour un ajustement comparable, les valeurs sont approximativement

$$\widehat{\phi}_1 \simeq 0{,}51, \qquad \widehat{\mu} \simeq 919, \qquad \widehat{\sigma}^2 \simeq 2{,}1 \times 10^4, \qquad \mathrm{SE}(\widehat{\phi}_1) \simeq 0{,}09.$$

Interprétation : 0,51 traduit une persistance positive d'une année à l'autre ; 919 est la moyenne stationnaire estimée ; $\widehat{\sigma}^2$ quantifie la variance de l'innovation du modèle ajusté. Son écart-type est d'environ

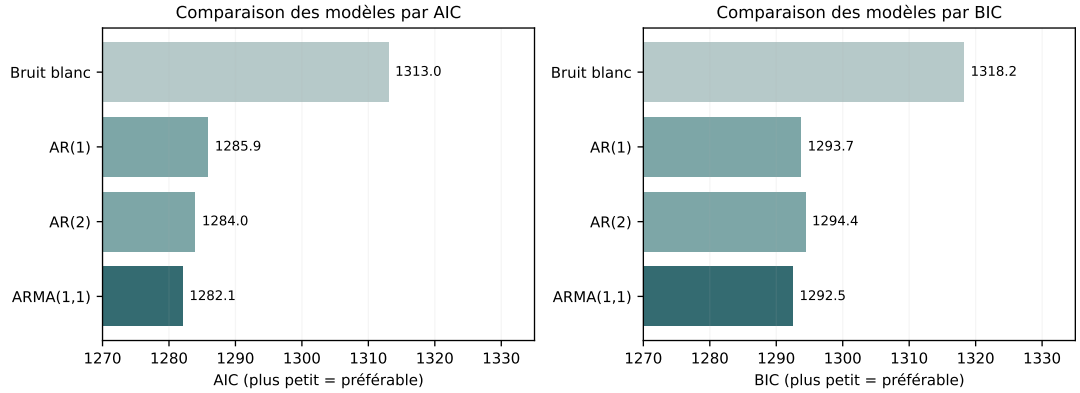
$$\sqrt{2{,}1 \times 10^4} \simeq \boxed{145 \text{ unités de débit}}.$$

L'erreur standard 0,09 porte au contraire sur *la précision du coefficient* $\widehat{\phi}_1$: ce n'est ni la variance des innovations ni une erreur de prévision. Elle est calculée ici par `sqrt(diag(f1$var.coef))` : `var.coef` est la matrice de covariance estimée des estimateurs, `diag()` en extrait la diagonale et `sqrt()` prend les racines carrées.

4) **Comparer AIC et BIC et expliquer leur classement.**

Les résultats numériques indicatifs sont les suivants :

Modèle	AIC indicatif	BIC indicatif
Bruit blanc	1313,0	1318,2
AR(1)	1285,9	1293,7
AR(2)	1284,0	1294,4
ARMA(1,1)	**1282,1**	**1292,5**



Les deux critères privilégient ici **ARMA(1,1)** parmi les candidats. Le bruit blanc est nettement moins bien classé, ce qui est cohérent avec les corrélations visibles dans l'ACF de la série brute. L'AR(1), néanmoins, n'est pas très éloigné du meilleur BIC : $1293{,}7 - 1292{,}5 = 1{,}2$ environ. Il reste donc un candidat simple à examiner sérieusement au diagnostic.

Pourquoi un paramètre supplémentaire est-il pénalisé ? Si l'on ajoute un coefficient, le meilleur ajustement sur l'échantillon peut s'améliorer, mais ce gain peut simplement refléter des fluctuations aléatoires. Les critères équilibrent ajustement et complexité. Ici,

$$\mathrm{AIC} = -2\ell(\widehat{\theta}) + 2k, \qquad \mathrm{BIC} = -2\ell(\widehat{\theta}) + k \log n.$$

Le nombre k comprend ici les coefficients du modèle *ainsi que la moyenne et la variance d'innovation estimées*. Par exemple, pour l'AR(1), $k = 3$ (AR, moyenne, variance). Avec une log-vraisemblance indicative $\ell(\widehat{\theta}) \simeq -639{,}95$:

$$\mathrm{AIC}_{AR(1)} \simeq -2(-639{,}95) + 2 \times 3 = 1285{,}9,$$
$$\mathrm{BIC}_{AR(1)} \simeq -2(-639{,}95) + 3\log(100) = 1293{,}7.$$

Ce calcul est fourni pour *expliquer la sortie de R*, et non comme un nouveau travail demandé. Les modèles ont été ajustés sur les mêmes données et selon des conventions compatibles : les comparaisons de critères ont ici un sens.

Remarque

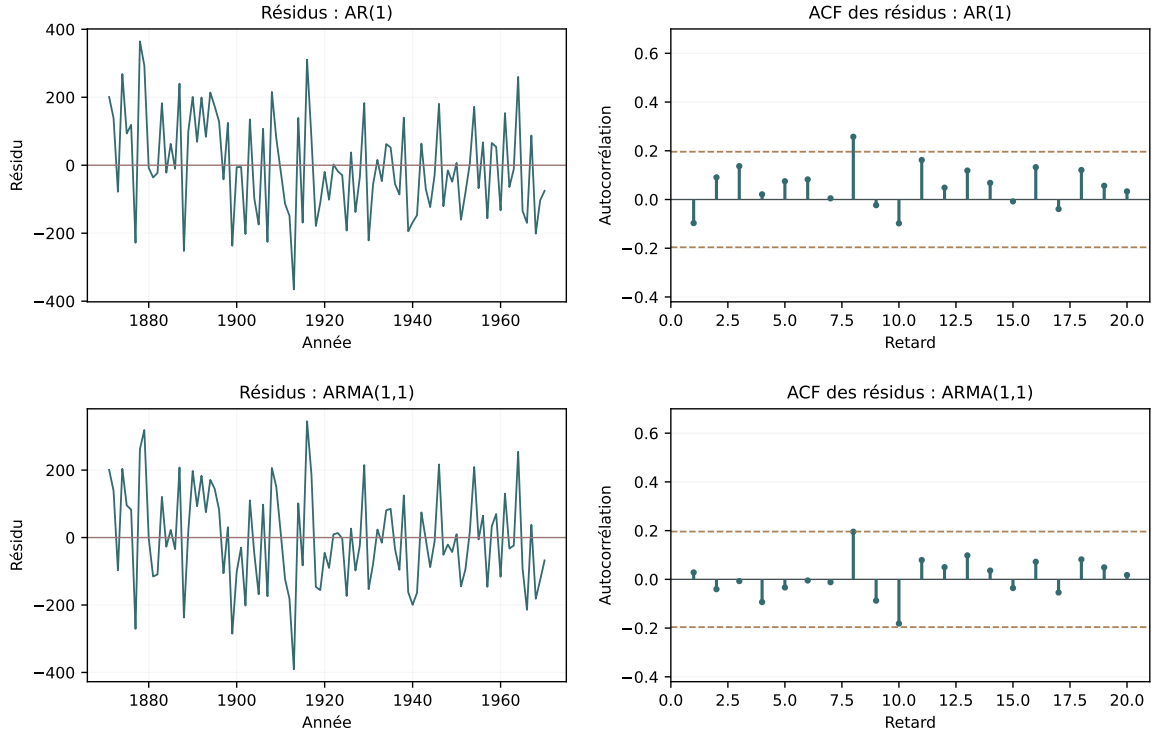
Le meilleur AIC ou BIC désigne le candidat ayant le meilleur *compromis ajustement/complexité pour le critère considéré*. Il ne dit pas à lui seul si toutes les hypothèses du modèle sont raisonnables. On doit ensuite examiner les résidus.

C. Examiner les résidus et décider

```
r1 <- residuals(f1)
r3 <- residuals(f3)

par(mfrow = c(2, 2))
plot.ts(r1, main = "Residus AR(1)")
acf(r1, main = "ACF residus AR(1)")
plot.ts(r3, main = "Residus ARMA(1,1)")
acf(r3, main = "ACF residus ARMA(1,1)")
par(mfrow = c(1, 1))

Box.test(r1, lag = 12, type = "Ljung-Box", fitdf = 1)
Box.test(r3, lag = 12, type = "Ljung-Box", fitdf = 2)
```



5) **Analyser les graphiques méthodiquement.**

Chronogrammes. Les résidus de l'AR(1) et de l'ARMA(1,1) oscillent globalement autour de zéro, sans tendance linéaire évidente de leur amplitude. On ne voit pas de croissance régulière de la dispersion au cours du temps. Il faut cependant rester prudent : la stabilité de la variance ne se *démontre* pas par un simple coup d'œil au graphique, et quelques valeurs atypiques subsistent.

ACF de la série puis des résidus. L'ACF de la série brute montrait une corrélation positive assez persistante. Après ajustement, les ACF résiduelles sont beaucoup plus proches de zéro : les deux modèles ont donc capté une part importante de la dépendance temporelle. Pour l'AR(1), on aperçoit notamment un pic isolé vers le retard 8 dans cette réalisation ; pour l'ARMA(1,1), les pics sont globalement faibles.

Comment interpréter les bandes ? Avec environ 100 résidus, les repères graphiques usuels valent approximativement $\pm 1{,}96/\sqrt{100} = \pm 0{,}196$. Ils donnent une première indication, mais un pic isolé peut apparaître par hasard lorsqu'on examine beaucoup de retards. C'est pourquoi on complète le graphique par Ljung–Box, plutôt que d'affirmer immédiatement qu'un modèle est mauvais.

6) **Lire le test de Ljung–Box.**

Pour chacun des modèles, on teste l'hypothèse nulle

$$H_0 : \rho_{\text{res}}(1) = \rho_{\text{res}}(2) = \cdots = \rho_{\text{res}}(12) = 0.$$

La statistique combine les autocorrélations résiduelles aux retards 1 à 12. La correction `fitdf` retranche les paramètres AR/MA estimés : $12 - 1 = 11$ degrés de liberté pour l'AR(1) et $12 - 2 = 10$ pour l'ARMA(1,1). Elle ne retranche pas ici la moyenne estimée.

Des ajustements gaussiens comparables donnent, *à titre indicatif* :

Modèle	Statistique $Q(12)$	Degrés de liberté	p-valeur
AR(1)	17,0	11	0,11
ARMA(1,1)	11,2	10	0,35

Pour les deux modèles, la p-valeur est supérieure à 0,05 : **on ne rejette pas H_0 au seuil de**

5 %. Les données ne mettent pas en évidence une autocorrélation résiduelle globale significative aux retards examinés.

Il faut distinguer trois affirmations :

— « On ne rejette pas l'absence d'autocorrélation aux douze premiers retards » : **correct**.

— « Tous les résidus sont rigoureusement non corrélés » : **non démontré**.

— « Les résidus sont indépendants et le modèle est vrai » : **encore moins démontré**. Le test ne vérifie pas l'indépendance.

7) **Justifier une conclusion de modélisation.**

Trois arguments possibles pour l'ARMA(1,1) :

i. Ses AIC et BIC sont les plus faibles parmi les quatre candidats étudiés.

ii. Son ACF résiduelle est globalement proche de zéro, nettement moins persistante que celle de la série brute.

iii. Son test de Ljung–Box à 12 retards ne rejette pas l'absence d'autocorrélation.

On peut donc rédiger : « Parmi les modèles ajustés, je retiens provisoirement ARMA(1,1), car il possède les plus petits AIC et BIC, les autocorrélations de ses résidus sont globalement faibles et Ljung–Box ne détecte pas de dépendance résiduelle globale significative aux douze premiers retards. »

Un autre choix peut être défendu. L'AR(1) est plus simple, son BIC est proche de celui de l'ARMA(1,1) et son test de Ljung–Box ne rejette pas non plus H_0. Un étudiant qui préfère l'AR(1) en expliquant soigneusement son choix fait donc preuve de bon raisonnement statistique. La consigne n'impose pas de traiter le minimum d'AIC/BIC comme une vérité absolue.

Une limite à citer : on soupçonne une éventuelle rupture du niveau de la série à certaines dates. Ni les seuls critères AIC/BIC ni Ljung–Box ne suffisent à éliminer cette possibilité. On n'a pas non plus vérifié toutes les formes de dépendance non linéaire, ni la qualité des prévisions sur un échantillon distinct.

D. Tableau de synthèse de l'énoncé

Un remplissage cohérent, sans inventer les tests non réalisés, serait :

Modèle	AIC	BIC	ACF résiduelle	Ljung–Box, 12 retards
Bruit blanc	1313,0	1318,2	non examinée dans la partie C	—
AR(1)	1285,9	1293,7	pics limités, dont un vers 8	non-rejet
AR(2)	1284,0	1294,4	non examinée dans la partie C	—
ARMA(1,1)	1282,1	1292,5	globalement proche de zéro	non-rejet

Les valeurs numériques de ce tableau restent indicatives. Les tirets signifient « test non réalisé », pas « absence d'autocorrélation prouvée ».

E. Vérifier et interpréter les sorties R

Le code ci-dessous donne un moyen de retrouver les nombres à la source dans sa propre session R, plutôt que de recopier des valeurs approchées du corrigé.

```
coef(f1)                  # moyenne et coefficient AR
f1$sigma2                 # variance des innovations
```

```
sqrt(diag(f1$var.coef))   # erreurs standards des coefficients

logLik(f3)                # log-vraisemblance maximisee
attr(logLik(f3), "df")    # nombre de parametres pour AIC/BIC
AIC(f3)                   # -2*logLik + 2*k
BIC(f3)                   # -2*logLik + k*log(n)

Box.test(residuals(f3), lag = 20,
         type = "Ljung-Box", fitdf = 2)
```

Pourquoi tester éventuellement 20 retards ? Cela examine une autre hypothèse globale : l'absence d'autocorrélation jusqu'au retard 20, et non jusqu'au retard 12. On peut donc obtenir une autre p-valeur, sans contradiction. Le nombre de retards doit être choisi pour une raison statistique et indiqué dans l'analyse ; on évite de chercher après coup le nombre donnant le résultat souhaité.

Idée à retenir

Fil conducteur du TP : explorer la série $\longrightarrow$ proposer plusieurs modèles $\longrightarrow$ les estimer $\longrightarrow$ comparer AIC/BIC $\longrightarrow$ diagnostiquer les résidus $\longrightarrow$ rédiger une conclusion nuancée. La sélection d'un modèle n'est pas un calcul automatique : c'est une décision argumentée à partir de plusieurs indices.