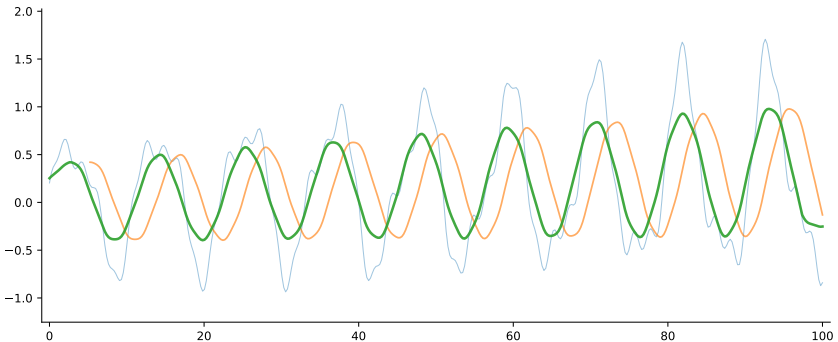
Séries temporelles

Cours 6 – Estimation, sélection et diagnostic



M1 Actuariat – Université Gustave Eiffel

Solesne Bourguin – 2026–2027

Objectifs de la séance

Au cours 5, nous avons étudié la forme des modèles ARIMA et SARIMA. Nous allons maintenant voir comment les ajuster à des observations et vérifier la pertinence de cet ajustement.

1. **Estimer** les coefficients : comment obtenir des valeurs numériques à partir des données ?
2. **Comparer** plusieurs modèles : comment arbitrer entre qualité d'ajustement et complexité ?
3. **Examiner les erreurs résiduelles** : reste-t-il une structure temporelle que le modèle n'explique pas ?

Idée à retenir. Une bonne adéquation numérique aux données observées ne suffit pas : il faut également vérifier que le modèle a capturé la dépendance temporelle.

Rappel : où en est-on après le cours 5 ?

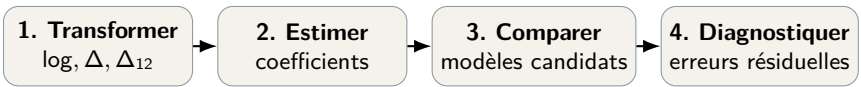
Le cours 5 a séparé deux étapes : transformer la série, puis modéliser la dépendance restante.



Question du jour. Une fois les transformations et les ordres candidats choisis, comment déterminer les coefficients du modèle ? Et comment vérifier qu'il rend correctement compte des données ?

Idée à retenir. Il nous faudra donc distinguer l'**ajustement des paramètres** et le **contrôle de la qualité de cet ajustement**.

La démarche en quatre questions



Exemple. Après différenciation, la série étudiée peut être décrite par plusieurs modèles ARMA. Comment choisir les coefficients, puis départager les modèles ?

Idée à retenir. La dernière étape peut faire **revenir en arrière** : si des dépendances subsistent dans les erreurs, il faut réexaminer le modèle.

Notations : distinguer trois objets

Observation

x_t

Valeur de la série *effectivement observée* à la date t.

Paramètre

$\vartheta = (\phi, \theta, \sigma^2, \ldots)$

Quantité inconnue à estimer sur l'échantillon.

Résidu

$\widehat{z}_t$

Erreur reconstruite *après* estimation du modèle.

Point de vigilance. Ne pas confondre le **bruit théorique** Z_t, inconnu, avec le **résidu calculé** $\widehat{z}_t$. Les deux ne coïncident pas exactement en échantillon fini.

Partie 1

Estimer les paramètres

Partir d'un modèle théorique et apprendre ses paramètres à partir des données.

Modèle, estimation, modèle ajusté

Considérons par exemple un modèle ARMA(1,1) stationnaire :

$$X_t = \phi X_{t-1} + Z_t + \theta Z_{t-1}.$$

La forme du modèle ne suffit pas : il faut encore déterminer ϕ, θ et la variance du bruit.

Modèle
Famille de lois indexée par des paramètres inconnus.

Estimation
Choix des paramètres à partir des observations.

Modèle ajusté
Modèle avec les paramètres estimés.

Idée à retenir. Les diagnostics ne portent pas sur les vrais Z_t, inconnus, mais sur des résidus calculés avec le modèle ajusté.

Du bruit théorique aux résidus calculés

Considérons l'AR(1) centré du cours précédent :

$$X_t = \phi X_{t-1} + Z_t, \qquad \mathbb{E}[Z_t \mid X_0, \dots, X_{t-1}] = 0.$$

Si le coefficient ϕ était connu, le modèle donnerait, à partir de X_{t-1},

$$\mathbb{E}[X_t \mid X_0, \dots, X_{t-1}] = \phi X_{t-1}, \qquad Z_t = X_t - \phi X_{t-1}.$$

Mais ϕ est inconnu : on le remplace par un estimateur $\widehat{\phi}$ calculé sur les données.

Définition. Le **résidu du modèle ajusté** à la date t est

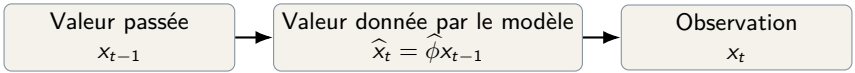
$$\boxed{\widehat{z}_t = x_t - \widehat{\phi}\, x_{t-1}.}$$

C'est la différence entre l'observation x_t et la valeur $\widehat{\phi}\, x_{t-1}$ fournie par le modèle ajusté.

Idée à retenir. Z_t est une variable aléatoire du **modèle théorique** ; $\widehat{z}_t$ est une quantité **calculée à partir de l'échantillon**.

Comment calcule-t-on un résidu ?

Pour calculer un résidu, on utilise *bien* l'observation x_t : c'est elle que l'on compare à la valeur ajustée.



Exemple. Si $\widehat{\phi} = 0{,}6$, $x_{t-1} = 1$ et $x_t = 2$, la valeur ajustée est $\widehat{x}_t = 0{,}6$ et le résidu vaut $\widehat{z}_t = 2 - 0{,}6 = 1{,}4$.

Point de vigilance. Précision importante. Le coefficient $\widehat{\phi}$ est estimé à partir de *tout* l'échantillon, qui peut inclure x_t. Le résidu ainsi calculé est donc une **erreur d'ajustement dans l'échantillon**, pas une prévision réellement réalisée avant d'observer x_t.

Un cas explicite : AR(1) par moindres carrés

Supposons

$$X_t = \phi X_{t-1} + Z_t,$$

et observons $x_0, x_1, \ldots, x_n$. Pour une **valeur candidate** a du coefficient, les erreurs sont

$$r_t(a) = x_t - a x_{t-1}.$$

La méthode des **moindres carrés** consiste à choisir a pour minimiser

$$S(a) = \sum_{t=1}^{n} (x_t - a x_{t-1})^2.$$

Estimateur des moindres carrés

$$\widehat{\phi} = \frac{\sum_{t=1}^{n} x_t x_{t-1}}{\sum_{t=1}^{n} x_{t-1}^2}, \qquad \text{si } \sum_{t=1}^{n} x_{t-1}^2 > 0.$$

Idée à retenir. Une fois $\widehat{\phi}$ obtenu, les **résidus du modèle ajusté** sont $\widehat{z}_t = r_t(\widehat{\phi})$.

Démonstration : d'où vient la formule des moindres carrés ?

On minimise la somme des carrés $S(\phi) = \sum_{t=1}^{n} (x_t - \phi x_{t-1})^2$.

$$\begin{aligned} S'(\phi) &= -2 \sum_{t=1}^{n} x_{t-1} \big(x_t - \phi x_{t-1} \big) \\ &= -2 \sum_{t=1}^{n} x_t x_{t-1} + 2\phi \sum_{t=1}^{n} x_{t-1}^2 . \end{aligned}$$

À l'optimum $S'(\widehat{\phi}) = 0$, donc

$$\boxed{\widehat{\phi} = \frac{\sum_{t=1}^{n} x_{t-1} x_t}{\sum_{t=1}^{n} x_{t-1}^2}.}$$

Idée à retenir. Puisque $S''(\phi) = 2 \sum x_{t-1}^2 > 0$, ce point critique est bien l'unique minimum (si le dénominateur est non nul).

Exemple numérique : estimer un AR(1)

On observe cinq valeurs **illustratives** d'une série centrée :

$$(x_0, x_1, x_2, x_3, x_4) = (2, 1, 2, 1, 0).$$

Calcul de l'estimateur

$$\sum_{t=1}^{4} x_t x_{t-1} = 2 + 2 + 2 + 0 = 6,$$

$$\sum_{t=1}^{4} x_{t-1}^2 = 4 + 1 + 4 + 1 = 10.$$

D'où $\boxed{\widehat{\phi} = 0{,}6}$.

Résidus à un pas

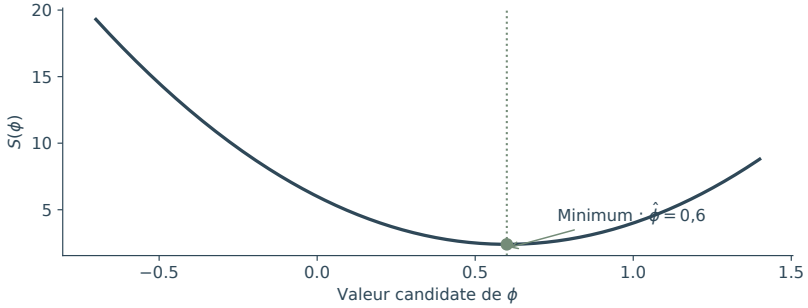
$$\widehat{z}_t = x_t - 0{,}6 x_{t-1}.$$

$$(\widehat{z}_1, \ldots, \widehat{z}_4) = (-0{,}2; 1{,}4; -0{,}2; -0{,}6).$$

Point de vigilance. Avec quatre transitions, on comprend le calcul, **mais on ne peut pas juger sérieusement le modèle**. Pour diagnostiquer, il faut davantage de données.

Le minimum se voit aussi graphiquement

Pour les cinq observations précédentes, la somme des carrés est une parabole :



Idée à retenir. Le meilleur ajustement *pour cette famille AR(1)* correspond à son minimum. Cela ne prouve pas que la famille AR(1) est adaptée à la série.

Et si la série n'est pas centrée ?

Le modèle $X_t = \phi X_{t-1} + Z_t$ suppose une moyenne nulle (dans le cas stationnaire avec $\mathbb{E}Z_t = 0$).

AR(1) de moyenne μ

$$X_t - \mu = \phi(X_{t-1} - \mu) + Z_t, \qquad \text{soit} \qquad X_t = c + \phi X_{t-1} + Z_t,$$

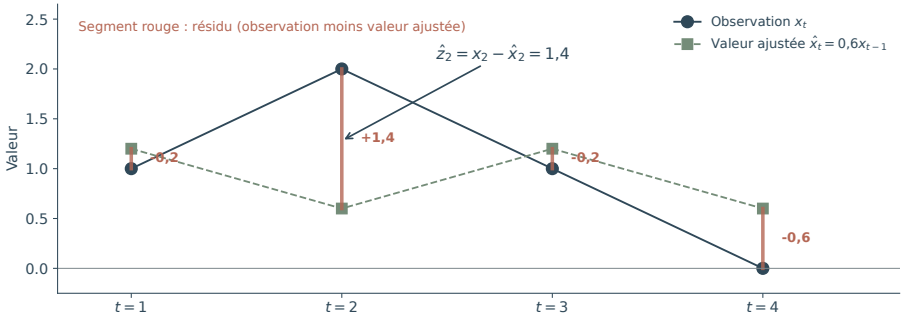
avec $c = (1 - \phi)\mu$.

- Si on **centre** les données, on peut étudier un modèle sans constante.
- Sinon, on peut estimer une **constante** en plus du coefficient AR.
- Une fois le modèle différencié ($d > 0$ ou $D > 0$), la gestion des termes constants ou de dérive demande de la prudence.

Idée à retenir. La formule explicite précédente concerne bien l'AR(1) **sans constante**. Ne pas l'appliquer automatiquement à une série avec moyenne non nulle.

Visualiser les résidus de l'exemple AR(1)

Reprenons les cinq observations de l'exemple précédent et le coefficient estimé $\widehat{\phi} = 0{,}6$.



Idée à retenir. Lecture au temps $t = 2$: $x_1 = 1$, donc la valeur ajustée est $\widehat{x}_2 = 0{,}6x_1 = 0{,}6$. Or $x_2 = 2$: le résidu est $\widehat{z}_2 = 2 - 0{,}6 = 1{,}4$. Chaque résidu se calcule de la même manière.

Maximum de vraisemblance : principe général

Définition. Pour les données $x_1, \dots, x_n$, soit f_ϑ leur densité jointe (ou leur probabilité dans le cas discret). La **vraisemblance** est

$$L(\vartheta) = f_\vartheta(x_1, \dots, x_n), \qquad \ell(\vartheta) = \log L(\vartheta).$$

On appelle **estimateur du maximum de vraisemblance** une valeur

$$\widehat{\vartheta} \in \arg\max_{\vartheta} \ell(\vartheta).$$

Idée à retenir. À données fixées, on choisit les paramètres pour lesquels les observations sont les plus plausibles selon le modèle.

Exemple : AR(1) gaussien. Pour $X_t = \phi X_{t-1} + Z_t$ avec $Z_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2)$, conditionnellement à $X_0 = x_0$:

$$\ell(\phi, \sigma^2) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{t=1}^{n}(x_t - \phi x_{t-1})^2.$$

À σ^2 fixé, maximiser en ϕ revient à minimiser la somme des carrés des résidus.

Pourquoi les carrés apparaissent-ils avec une loi gaussienne ?

Supposons $X_t = \phi X_{t-1} + Z_t$ avec $Z_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2)$, indépendants du passé.

$$\mathbb{P}(X_t \in dx \mid X_{t-1} = x_{t-1}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x - \phi x_{t-1})^2}{2\sigma^2}\right] dx.$$

En multipliant les densités conditionnelles, puis en prenant le logarithme :

$$\ell(\phi, \sigma^2 \mid x_0) = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{t=1}^{n} (x_t - \phi x_{t-1})^2.$$

Idée à retenir. **À variance fixée**, maximiser cette log-vraisemblance revient exactement à **minimiser les carrés** des résidus : c'est le lien entre maximum de vraisemblance et moindres carrés.

Et la variance du bruit, comment l'estimer ?

Dans l'AR(1) gaussien précédent, une fois $\widehat{\phi}$ estimé, il reste σ^2.

$$\ell(\widehat{\phi}, v \mid x_0) = -\frac{n}{2}\log(2\pi v) - \frac{S(\widehat{\phi})}{2v}, \quad v > 0.$$

La dérivation par rapport à v donne

$$-\frac{n}{2v} + \frac{S(\widehat{\phi})}{2v^2} = 0 \quad \Longrightarrow \quad \boxed{\widehat{\sigma}^2_{\mathrm{MV,cond}} = \frac{S(\widehat{\phi})}{n}}.$$

Point de vigilance. L'estimateur de maximum de vraisemblance utilise ici $1/n$, et **pas** une correction « sans biais ». On a travaillé *conditionnellement à* X_0.

Idée à retenir. Avec des modèles ARMA/SARIMA plus riches, ces idées restent valables mais l'optimisation ne possède généralement pas de formule aussi simple.

ARMA : le bruit passé est inconnu

Dans un ARMA(1,1) centré :

$$X_t = \phi X_{t-1} + Z_t + \theta Z_{t-1}.$$

Même si ϕ et θ étaient connus, Z_{t-1} n'apparaît pas dans les données.

Calcul récursif illustratif (pour un choix initial $\widehat{z}_0$)

$$\widehat{z}_t = x_t - \phi x_{t-1} - \theta \widehat{z}_{t-1}.$$

La valeur calculée à la date t dépend du résidu calculé à la date $t-1$.

Point de vigilance. Les logiciels utilisent une gestion précise des conditions initiales et, en maximum de vraisemblance exact, des techniques de calcul des innovations. La récurrence ci-dessus **illustre** le mécanisme, elle n'en remplace pas tous les détails.

Pour SARIMA, le principe reste le même

Le cours 5 a introduit les retards ordinaires B et saisonniers B^{12}. Pour une série mensuelle :

$$\Phi(B^{12})\phi(B)(1-B)^d(1-B^{12})^D X_t = \Theta(B^{12})\theta(B)Z_t.$$

Deux questions distinctes

1. **Choix des ordres** (p, d, q) et $(P, D, Q)_{12}$: combien de termes autorégressifs et de moyenne mobile retient-on ?
2. **Estimation des coefficients** : une fois les ordres fixés, quelles valeurs attribuer à ces coefficients et à la variance du bruit ?

Idée à retenir. Nous avons vu comment estimer les coefficients pour un modèle donné. Il reste ensuite à **comparer plusieurs modèles candidats** de complexités différentes.

Commandes R : ajuster un ARIMA/SARIMA

```
y <- log(AirPassengers)

fit <- arima(y,
             order = c(0, 1, 1),
             seasonal = list(order = c(0, 1, 1), period = 12),
             method = "ML")

fit
res <- residuals(fit)
```

Décodage

- `order = c(p,d,q)` : ordres non saisonniers.
- `seasonal = list(order = c(P,D,Q), period = 12)` : ordres saisonniers pour des données mensuelles.
- `method = "ML"` : estimation par maximum de vraisemblance.
- `residuals(fit)` : résidus calculés après l'ajustement.

Idée à retenir. `arima()` calcule numériquement les coefficients qui n'admettent généralement pas de formule explicite.

Lire une sortie de `arima()` : quoi regarder ?

L'objet retourné par `arima()` contient plusieurs informations sur le modèle ajusté.

Coefficients estimés
$\widehat{\phi}_i$, $\widehat{\theta}_j$, coefficients saisonniers : ils décrivent la dynamique retenue.

Erreurs standards
Elles quantifient approximativement l'incertitude liée à l'estimation des coefficients.

Variance estimée
`sigma2` : estimation de la variance du bruit.

Log-vraisemblance
`loglik` : valeur de la log-vraisemblance au point estimé. Nous verrons comment l'utiliser pour comparer des modèles.

Point de vigilance. Ces quantités décrivent l'ajustement. Elles ne suffisent pas, à elles seules, pour conclure que le modèle est adéquat.

Mise en pratique R : récupérer les informations utiles

```
y <- log(AirPassengers)
fit <- arima(y, order = c(0,1,1),
             seasonal = list(order = c(0,1,1), period = 12),
             method = "ML")
coef(fit)                  # coefficients estimes
sqrt(diag(fit$var.coef))  # erreurs standards approx.
fit$sigma2                # variance du bruit
fit$loglik                # log-vraisemblance maximisee
residuals(fit)            # residus calcules
```

Idée à retenir. Nous venons d'ajuster un modèle SARIMA $(0,1,1) \times (0,1,1)_{12}$. Pour ces ordres fixés, R fournit les coefficients et les résidus. Reste à savoir comment **comparer ce modèle à d'autres candidats**.

Partie 2

Choisir parmi plusieurs modèles

Un meilleur ajustement n'autorise pas à multiplier les paramètres indéfiniment.

Pourquoi ne pas toujours prendre le modèle le plus riche ?

Dans une famille de modèles emboîtés, ajouter des paramètres ne détériore pas le meilleur ajustement possible sur les données déjà observées.

Modèle trop simple

- dépendance restante ;
- résidus autocorrélés ;
- prévisions biaisées.

Modèle trop riche

- paramètres inutiles ;
- fluctuations accidentelles apprises ;
- prévision fragile.

Idée à retenir. Comparer des modèles, c'est chercher un compromis entre qualité d'ajustement et simplicité.

AIC et BIC : ajustement contre complexité

AIC : *Akaike Information Criterion* (critère d'information d'Akaike).
BIC : *Bayesian Information Criterion* (critère d'information bayésien, dit aussi de Schwarz).

Ces critères servent à **comparer des modèles candidats** : une meilleure vraisemblance favorise un modèle, mais des paramètres supplémentaires ont un coût.

Définition. Notons $\widehat{\ell}$ la log-vraisemblance maximisée, k le nombre de paramètres estimés, et n le nombre d'observations :

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

Qualité d'ajustement
$-2\widehat{\ell}$ diminue quand le modèle explique mieux les données.

Pénalité de complexité
$2k$ pour AIC ; $k \log n$ pour BIC (plus forte si $n > e^2$).

Idée à retenir. À données et vraisemblances comparables, **le plus petit score** désigne le meilleur compromis *parmi les modèles comparés*.

Combien de paramètres dans la pénalité ?

Dans AIC $= -2\widehat{\ell} + 2k$, le nombre k désigne les **paramètres estimés dans la vraisemblance**.

Exemple : ARMA(1,1) gaussien centré

$$X_t = \phi X_{t-1} + Z_t + \theta Z_{t-1}, \qquad Z_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2).$$

On estime ϕ, θ et σ^2 : cela fait $k = 3$ paramètres. Si l'on ajoute une moyenne inconnue, elle compte pour **un paramètre supplémentaire**.

Point de vigilance. Le nombre de paramètres n'est pas simplement $p + q + P + Q$ dans tous les modèles : il peut également inclure une moyenne, une dérive et la variance du bruit.

Idée à retenir. En pratique, les fonctions `AIC(fit)` et `BIC(fit)` calculent ces pénalités à partir de l'objet ajusté.

Quand ajouter un paramètre au modèle ?

Comparons deux modèles emboîtés, ajustés sur les mêmes données : le second possède **un paramètre supplémentaire** et sa log-vraisemblance augmente de $\Delta\ell > 0$.

$$\mathrm{AIC}_2 - \mathrm{AIC}_1 = -2\Delta\ell + 2, \qquad \mathrm{BIC}_2 - \mathrm{BIC}_1 = -2\Delta\ell + \log n.$$

Avec AIC
On préfère le modèle plus riche si $2\Delta\ell > 2$.

Avec BIC
On le préfère si $2\Delta\ell > \log n$.

Idée à retenir. Avec $n = 100$, on a $\log n \approx 4{,}61$: BIC exige une amélioration plus importante de la vraisemblance que AIC.

Exemple numérique : trois modèles candidats

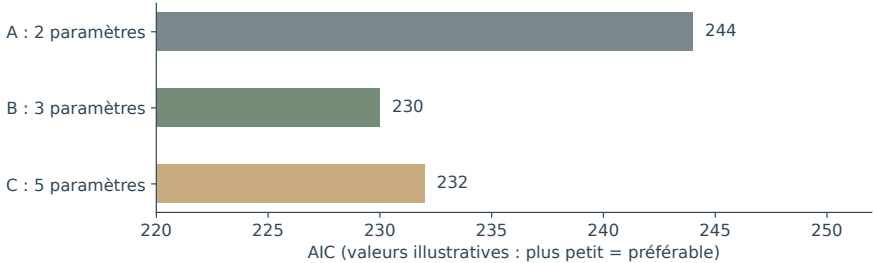
Données illustratives : même série de $n = 100$ observations, mêmes conventions de vraisemblance.

Modèle	k	$\widehat{\ell}$	AIC $= -2\widehat{\ell} + 2k$	BIC $= -2\widehat{\ell} + k \log 100$
A : simple	2	-120	244	249,21
B : intermédiaire	3	-112	**230**	**237,82**
C : complexe	5	-111	232	245,03

Question. Pourquoi C, pourtant mieux ajusté que B ($-111 > -112$), n'est-il pas préféré ?

Idée à retenir. Le gain de vraisemblance de C ne compense pas ses deux paramètres supplémentaires. **B gagne parmi ces trois modèles**, avant tout diagnostic des résidus.

Le même exemple en un coup d'œil



Idée à retenir. Les nombres servent à **comprendre** le compromis ajustement/complexité ; ce ne sont pas des AIC tirés d'une sortie de R.

Comment lire AIC/BIC en pratique ?

Les critères ne donnent pas une vérité absolue : ils classent les modèles que l'on a effectivement comparés.

Bonne lecture

- même série, même période, même transformation ;
- comparer quelques modèles plausibles ;
- retenir les modèles simples si les critères sont proches ;
- toujours diagnostiquer les résidus ensuite.

Point de vigilance. Un modèle peut avoir le meilleur AIC parmi les modèles testés et rester insuffisant si ses résidus ne ressemblent pas à un bruit blanc.

Pourquoi comparer seulement des ajustements compatibles ?

AIC et BIC n'ont de sens que si les modèles décrivent **les mêmes données observées**.

Comparaison cohérente
Trois SARIMA sur $\log X_t$, avec les mêmes d, D et la même période, estimés sur le même échantillon par la même méthode.

Comparaison à éviter telle quelle
AIC calculé sur X_t *contre* AIC sur $\log X_t$; ou AIC basés sur des échantillons transformés différents sans traitement approprié.

Point de vigilance. Une transformation change l'échelle de densité (un **jacobien** intervient). On choisit d'abord les transformations pertinentes, puis on compare les structures ARMA/SARIMA compatibles.

Commandes R : comparer par AIC

```
y <- log(AirPassengers)

fit1 <- arima(y, order = c(0, 1, 1),
              seasonal = list(order = c(0, 1, 1), period = 12),
              method = "ML")
fit2 <- arima(y, order = c(1, 1, 0),
              seasonal = list(order = c(0, 1, 1), period = 12),
              method = "ML")
fit3 <- arima(y, order = c(1, 1, 1),
              seasonal = list(order = c(0, 1, 1), period = 12),
              method = "ML")

AIC(fit1, fit2, fit3)
BIC(fit1, fit2, fit3)
```

Décodage `AIC()` et `BIC()` comparent des modèles déjà ajustés. Le résultat n'est interprétable que si les modèles sont ajustés sur les mêmes données.

Exercice minute : quel candidat conserver ?

Trois modèles sont ajustés sur **la même série** (mêmes transformations et échantillon). On obtient les résultats suivants :

Candidat	AIC	BIC
M_1	312	320
M_2	305	319
M_3	303	331

Questions. Quel modèle est préféré par AIC ? Et par BIC ? Les deux critères donnent-ils nécessairement le même classement ?

Idée à retenir. Réponse. AIC préfère M_3 et BIC préfère M_2. Les critères ne suffisent pas à conclure : il reste à **examiner les résidus** des modèles candidats.

Partie 3

Diagnostiquer le modèle ajusté

La question décisive : reste-t-il une dépendance exploitable dans les résidus ?

Ce que doivent vérifier les résidus

Dans un ARMA, ARIMA ou SARIMA bien ajusté, les résidus calculés devraient se comporter approximativement comme le bruit du modèle.

Centrage
Pas de moyenne clairement différente de zéro.

Variance
Pas de changement évident d'amplitude.

ACF
Pas de pics importants aux retards non nuls.

Point de vigilance. On ne demande pas aux résidus d'être exactement nuls ou parfaitement indépendants : on travaille avec un échantillon fini.

Bruit blanc, indépendance et normalité : ne pas confondre

Bruit blanc (au sens faible) : une série Z_t telle que

$$\mathbb{E}Z_t = 0, \qquad \mathrm{Var}(Z_t) = \sigma^2 < \infty, \qquad \mathrm{Cov}(Z_t, Z_{t+h}) = 0 \text{ pour } h \neq 0.$$

- Une ACF résiduelle proche de zéro suggère une **absence de dépendance linéaire**.
- Cela ne prouve ni **indépendance**, ni **normalité**, ni absence de dépendance non linéaire.
- Les hypothèses gaussiennes sont utiles pour certaines vraisemblances et pour les intervalles, mais **normalité** et **bruit blanc** sont deux propriétés distinctes.

Idée à retenir. Le premier diagnostic d'un ARIMA/SARIMA consiste à regarder la dépendance *encore présente* dans les erreurs.

ACF empirique des résidus

Soient $\widehat{z}_1, \ldots, \widehat{z}_n$ les résidus et $\bar{z}$ leur moyenne. L'autocorrélation empirique au retard h est

$$\widehat{\rho}_z(h) = \frac{\sum_{t=1}^{n-h} (\widehat{z}_t - \bar{z})(\widehat{z}_{t+h} - \bar{z})}{\sum_{t=1}^{n} (\widehat{z}_t - \bar{z})^2}.$$

Lecture pratique

- Si le modèle est raisonnable, les barres de l'ACF résiduelle restent petites.
- Des pics isolés peuvent arriver par hasard.
- Plusieurs pics structurés indiquent souvent une dépendance non modélisée.

La bande de confiance de l'ACF : ordre de grandeur

Pour un bruit blanc et un échantillon assez long, les autocorrélations empiriques de retards non nuls fluctuent typiquement autour de zéro, avec écart-type d'ordre $1/\sqrt{n}$.

Bandes indicatives à 95 % : $\boxed{\pm \dfrac{1{,}96}{\sqrt{n}}}$.

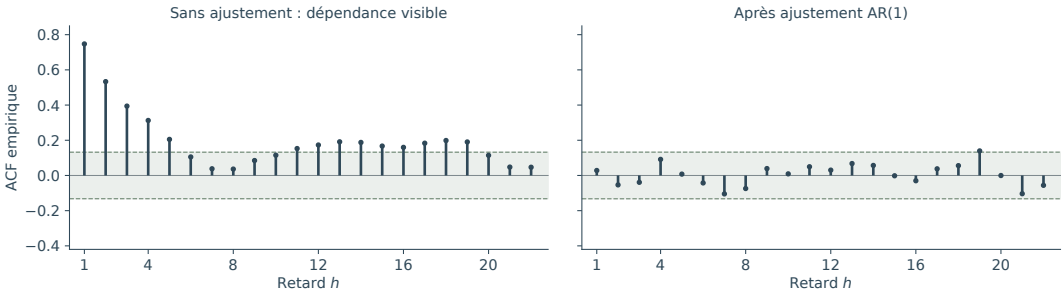
Exemple : $n = 120$ **résidus.**

$$\frac{1{,}96}{\sqrt{120}} \approx 0{,}179.$$

Un pic à 0,05 est banal ; un pic à 0,35 appelle une investigation.

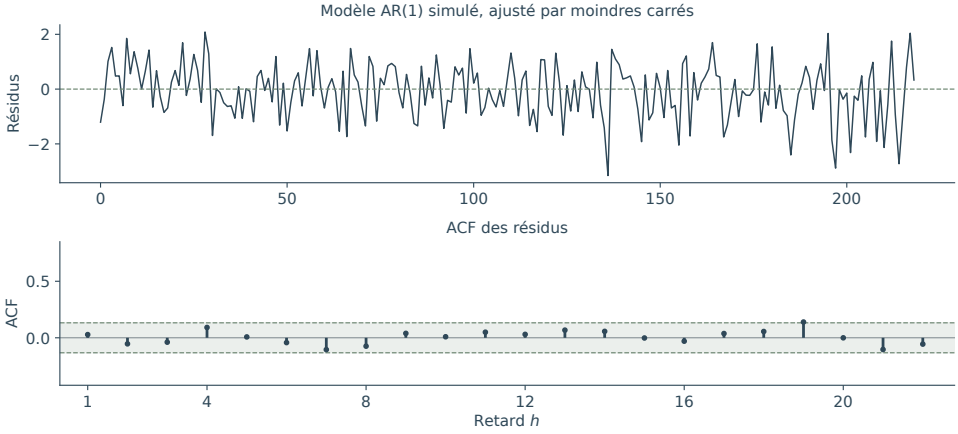
Point de vigilance. Ces bandes sont *approximatives*, surtout pour des résidus **estimés**. Examiner 20 retards peut produire un pic isolé hors bande par hasard ; d'où l'intérêt d'un test global.

Deux ACF : l'ajustement a-t-il retiré la dépendance ?



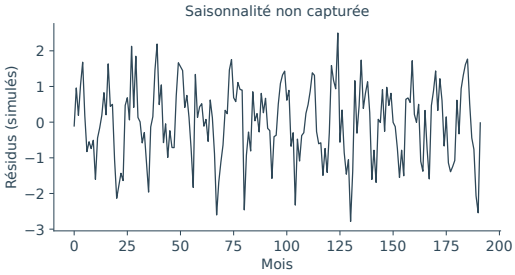
Idée à retenir. À gauche, l'ACF de la série AR(1) simulée. À droite, l'ACF *des résidus après estimation*. Ici l'AR(1) a effectivement capturé l'essentiel de la dépendance.

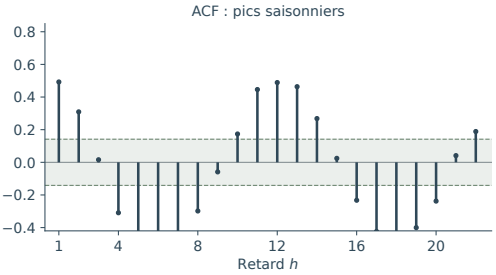
Lire un diagnostic complet sur une simulation



Idée à retenir. Deux graphiques complémentaires : **évolution des résidus** dans le temps, puis **corrélations résiduelles**. Aucun ne suffit seul.

Et si une saisonnalité subsiste dans les résidus ?





Idée à retenir. La répétition des pics près de 12, 24, . . . suggère un mécanisme saisonnier encore mal pris en compte : transformations, ordres saisonniers ou structure du modèle à réexaminer. *Exemple simulé.*

Pourquoi tester plusieurs retards simultanément ?

Sur un graphique d'autocorrélations, plusieurs valeurs **modérées** peuvent, prises ensemble, signaler une dépendance qui n'est plus due au hasard.

Regarder l'ACF
Elle indique *à quels retards* une corrélation pourrait subsister : 1, 2, 12, 24, etc.

Effectuer un test global
Il s'agit de vérifier *simultanément* si plusieurs autocorrélations peuvent être considérées comme nulles.

Idée à retenir. Nous allons construire un tel test, appelé **test de Ljung–Box**. Il complète l'analyse visuelle, sans la remplacer.

Le test de Ljung–Box : hypothèses et décision

On teste **globalement** l'absence d'autocorrélation dans les résidus jusqu'au retard m.

Hypothèse nulle $H_0 : \rho_z(1) = \cdots = \rho_z(m) = 0$.
Hypothèse alternative H_1 : au moins une de ces autocorrélations est non nulle.

$$Q_{\mathrm{LB}} = n(n+2) \sum_{h=1}^{m} \frac{\widehat{\rho}_z(h)^2}{n-h}.$$

Sous H_0 et des hypothèses de régularité, pour un modèle ARMA ajusté, on utilise souvent approximativement une loi χ^2_{m-k} (k : nombre de coefficients AR et MA estimés ; $m > k$).

p-valeur $< 0{,}05$
Au seuil de 5 %, on **rejette** H_0 : dépendance résiduelle détectée.

p-valeur $\geq 0{,}05$
On **ne rejette pas** H_0 : le test ne détecte pas d'autocorrélation résiduelle significative.

Exemple : calculer une décision Ljung–Box

Supposons $n = 120$ résidus, $m = 18$ retards et deux coefficients AR/MA estimés. On utilise l'approximation χ^2_{16}.

Situation	Statistique Q_{LB}	p-valeur (approximative)	Décision à 5 %
A	12	0,744	Ne pas rejeter H_0
B	30	0,018	Rejeter H_0

Pourquoi ? Sous H_0, de grandes valeurs de Q_{LB} sont rares : elles signifient que la somme des autocorrélations résiduelles au carré est trop importante.

Point de vigilance. Dans la situation A, **on n'a pas prouvé** que les résidus sont un bruit blanc. Le test n'a simplement pas détecté d'écart suffisamment marqué.

Choisir le nombre de retards m

Le test de Ljung–Box dépend du nombre m d'autocorrélations examinées.

- Si m est **trop petit**, on risque de manquer une dépendance à retard plus long.
- Si m est **très grand**, le test agrège beaucoup de retards potentiellement peu informatifs.
- Il faut avoir $m > k$ pour disposer de degrés de liberté positifs avec la correction.
- Pour une série mensuelle, **12 ou 24** retards peuvent être des points de départ pertinents selon la taille de l'échantillon et les questions étudiées.

Idée à retenir. Examiner **l'ACF et le test ensemble**, et ne pas choisir m après coup simplement pour obtenir une p-valeur qui nous plaît.

Point de vigilance. La correction $m - k$ est **asymptotique**. Dans un SARIMA, on prend usuellement $k = p + q + P + Q$ pour les coefficients AR/MA, sans compter d, D ni σ^2 ; c'est une approximation de diagnostic, pas une règle universelle exacte.

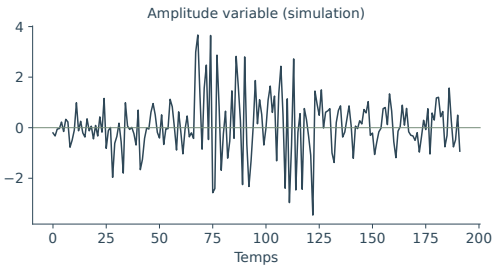
Commandes R : tout le diagnostic en quelques lignes

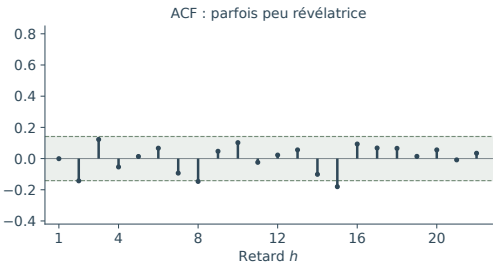
```
fit <- arima(log(AirPassengers), order = c(0,1,1),
             seasonal = list(order = c(0,1,1), period = 12),
             method = "ML")
res <- residuals(fit)
plot.ts(res, main = "Residus du modele ajuste")
abline(h = 0, lty = 2)
acf(res, lag.max = 36, main = "ACF des residus")
Box.test(res, lag = 24, type = "Ljung-Box", fitdf = 2)
```

Décodage de `fitdf = 2` : un coefficient MA ordinaire ($q = 1$) et un coefficient MA saisonnier ($Q = 1$) sont estimés. Il reste ici $24 - 2 = 22$ degrés de liberté dans l'approximation.

Attention : le test ne voit pas tout

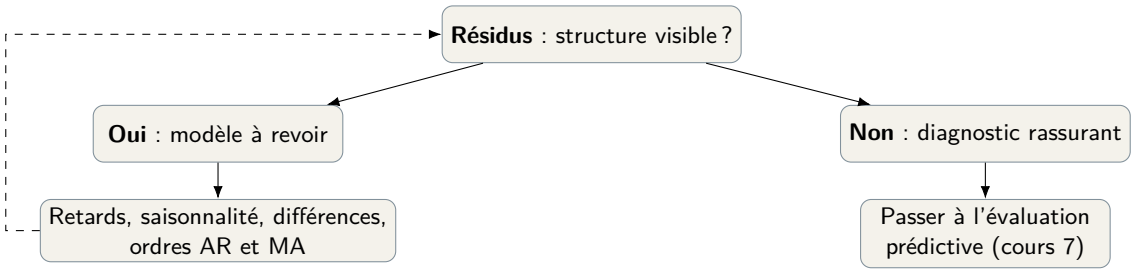
Une ACF résiduelle proche de zéro est compatible avec l'hypothèse de bruit blanc faible, mais ne suffit pas à la vérifier complètement : il faut également examiner la moyenne et la stabilité de la variance des résidus.





Point de vigilance. Cette simulation présente une variance qui change avec le temps, alors que l'ACF peut rester modérée. Ce phénomène conduit à étudier, dans d'autres modèles, la **variance conditionnelle**.

Si le diagnostic échoue, que faire ?



Idée à retenir. Itérer raisonnablement, sans ajuster des dizaines de modèles jusqu'à obtenir par hasard un test non significatif. La simplicité et la cohérence du modèle restent importantes.

Application récapitulative : une procédure reproductible

```
y <- log(AirPassengers)  # série mensuelle
# La transformation et les différences sont les mêmes partout.
f1 <- arima(y, c(0,1,1), seasonal=list(order=c(0,1,1),period=12),
            method="ML")
f2 <- arima(y, c(1,1,0), seasonal=list(order=c(0,1,1),period=12),
            method="ML")
f3 <- arima(y, c(1,1,1), seasonal=list(order=c(0,1,1),period=12),
            method="ML")
AIC(f1, f2, f3)
BIC(f1, f2, f3)
# Après sélection d'un candidat (ici f1 pour illustrer) :
r <- residuals(f1)
plot.ts(r)
acf(r, lag.max=36)
Box.test(r, lag=24, type="Ljung-Box", fitdf=2)
```

Idée à retenir. Le résultat est une démarche, et non une réponse imposée à l'avance. Les AIC, BIC et diagnostics produits par R décident de la suite.

Synthèse : ce qu'il faut retenir

1. Le modèle spécifie une famille de dynamiques ; **estimer** fixe ses coefficients à partir des observations.
2. Pour un AR(1) centré, les **moindres carrés** ont une formule explicite.
3. Sous des innovations gaussiennes, la **vraisemblance** relie naturellement estimation et somme des carrés.
4. **AIC/BIC** compensent le gain d'ajustement par une pénalité de complexité. On les compare sur des données compatibles.
5. Les **résidus d'innovation** doivent être sans structure temporelle manifeste.
6. L'**ACF** localise les corrélations ; **Ljung–Box** teste globalement plusieurs retards.
7. Une grande p-valeur **ne prouve rien** : il faut regarder les graphiques et réfléchir au modèle.

Idée à retenir. **Construire** un modèle = choisir, estimer, comparer, diagnostiquer, puis réviser si nécessaire.

Transition vers le cours 7

Après avoir estimé et diagnostiqué un modèle, on peut étudier comment **prévoir des observations futures**.

Au cours 7 : prévoir avec ARIMA/SARIMA

- Comment calcule-t-on une prévision à un pas et à plusieurs pas ?
- Pourquoi l'incertitude augmente-t-elle avec l'horizon ?
- Comment revenir à l'échelle de départ après logarithme ou différenciation ?
- Comment évaluer une prévision sur des données jamais utilisées pour l'estimation ?

Idée à retenir. La prévision est le **test final de l'utilité** d'un modèle. Un diagnostic rassurant ne garantit pas à lui seul de bonnes prévisions hors échantillon.