# Séries temporelles – Correction TD/TP 5

Diagnostiquer une non-stationnarité et choisir les transformations

M1 Actuariat – Université Gustave Eiffel

Cette correction est rédigée comme un rendu possible. Les formulations peuvent varier, mais une bonne solution doit surtout **justifier le choix des transformations**. Le point essentiel de cette feuille est qu'une série non stationnaire ne doit pas être différenciée mécaniquement : plusieurs mécanismes peuvent produire des chronogrammes proches.

## Exercice 1 – Deux séries qui se ressemblent, deux traitements différents

On a

$$U_t = 0.7U_{t-1} + Z_t, \qquad X_t = 100 + 0.5t + U_t, \qquad Y_t = Y_{t-1} + 0.5 + Z_t.$$

Le processus $(U_t)$ est un AR(1) stationnaire, car $|0.7| < 1$.

1) Puisque $(U_t)$ est centré,

$$\mathbb{E}(X_t) = 100 + 0.5t.$$

Sa variance est constante :

$$\mathrm{Var}(X_t) = \mathrm{Var}(U_t) = \frac{\sigma^2}{1 - 0.7^2} = \frac{\sigma^2}{0.51}.$$

Ainsi, la moyenne de $X_t$ dépend du temps, mais la dispersion autour de la droite reste stable.

Pour $Y_t$, l'itération de la relation donne

$$Y_t = 0.5t + \sum_{j=1}^{t} Z_j.$$

Donc

$$\mathbb{E}(Y_t) = 0.5t, \qquad \mathrm{Var}(Y_t) = t\sigma^2.$$

Ici, ce n'est pas seulement la moyenne qui évolue : l'incertitude du niveau s'accumule au cours du temps.

**Différence essentielle :** $X_t$ possède une tendance déterministe autour de laquelle fluctue une série stationnaire ; $Y_t$ possède une tendance stochastique, car les chocs passés s'accumulent dans le niveau.

2) Pour $X_t$,

$$X_t - (100 + 0.5t) = U_t,$$

qui est stationnaire. Le retrait de la bonne tendance déterministe résout donc directement le problème.

Pour $Y_t$, posons

$$R_t = Y_t - 0.5t.$$

Alors

$$\begin{aligned} R_t &= Y_t - 0.5t \\ &= Y_{t-1} + 0.5 + Z_t - 0.5t \\ &= (Y_{t-1} - 0.5(t-1)) + Z_t \\ &= R_{t-1} + Z_t. \end{aligned}$$

Ainsi, enlever la droite moyenne ne suffit pas : le résidu reste une marche aléatoire, donc n'est pas stationnaire.

3) Pour $Y_t$,
$$\Delta Y_t = Y_t - Y_{t-1} = 0.5 + Z_t.$$
La série différenciée est donc stationnaire autour de la moyenne 0.5.
Pour $X_t$,
$$\begin{aligned}\Delta X_t &= X_t - X_{t-1} \\ &= 0.5 + U_t - U_{t-1}.\end{aligned}$$
Cette série est elle aussi stationnaire, puisque différence de deux variables appartenant au même processus stationnaire. Mais cela ne signifie pas que la différence était le meilleur traitement : elle était inutile pour rendre la série stationnaire.

4) On part de
$$(1 - 0.7B)U_t = Z_t.$$
Comme
$$\Delta X_t - 0.5 = (1 - B)U_t,$$
on obtient, les polynômes en $B$ commutant,
$$\begin{aligned}(1 - 0.7B)(\Delta X_t - 0.5) &= (1 - 0.7B)(1 - B)U_t \\ &= (1 - B)(1 - 0.7B)U_t \\ &= (1 - B)Z_t \\ &= Z_t - Z_{t-1}.\end{aligned}$$
Ainsi, après différence, le résidu vérifie un ARMA(1,1) dont la partie MA est $1 - B$. On a donc transformé le simple AR(1) $U_t$ en un modèle plus compliqué. C'est une manifestation classique d'une différence inutile.

5) Une démarche raisonnable consiste à comparer plusieurs transformations :
   — ajuster une tendance déterministe et regarder le chronogramme et l'ACF des résidus ;
   — regarder la série différenciée et son ACF ;
   — vérifier si la variance semble s'accroître avec le temps ;
   — comparer la persistance de l'ACF avant et après retrait de tendance.

   Si le retrait d'une tendance donne déjà une série stable, il n'y a pas de raison de préférer une différence plus agressive.

**Idée à retenir**

Le même mouvement ascendant peut venir d'une droite déterministe ou de l'accumulation de chocs. Enlever une droite traite le premier mécanisme, pas le second.

## Exercice 2 – Une dynamique trimestrielle cachée

Le modèle est
$$X_t - X_{t-4} = 0.6(X_{t-1} - X_{t-5}) + Z_t + 0.3Z_{t-1}.$$

1) Posons
$$U_t = X_t - X_{t-4} = (1 - B^4)X_t.$$

Alors

$$U_{t-1} = X_{t-1} - X_{t-5},$$

et l'équation devient

$$U_t = 0.6U_{t-1} + Z_t + 0.3Z_{t-1}.$$

Donc $(U_t)$ suit un ARMA(1,1).

2) Le polynôme AR est

$$\Phi(z) = 1 - 0.6z,$$

dont la racine est $1/0.6 \simeq 1.67 > 1$ : le modèle est causal.

Le polynôme MA est

$$\Theta(z) = 1 + 0.3z,$$

dont la racine vaut $-1/0.3 \simeq -3.33$, donc de module supérieur à 1 : le modèle est inversible.

3) En opérateur retard,

$$(1 - 0.6B)(1 - B^4)X_t = (1 + 0.3B)Z_t.$$

Cette écriture sépare clairement la différence saisonnière $(1 - B^4)$ de la dynamique ARMA ordinaire.

4) Une notation naturelle est

$$\boxed{\text{SARIMA}(1, 0, 1)(0, 1, 0)_4.}$$

La partie ordinaire est ARMA(1,1), et la seule opération saisonnière imposée est une différence de période 4.

5) On lit

$$d = 0, \qquad D = 1.$$

Le modèle demande donc une différence saisonnière, mais aucune différence ordinaire. C'est précisément le type de situation où appliquer systématiquement $\Delta$ en plus de $1 - B^4$ risquerait de sur-différencier.

6) Si on ne fait qu'une différence ordinaire, la structure annuelle peut rester visible. Pour une série trimestrielle, on chercherait notamment dans l'ACF des pics ou une organisation particulière aux retards

$$4, 8, 12, \ldots$$

Une forte dépendance persistante à ces multiples de 4 indique que la saisonnalité n'a pas été correctement retirée.

## Exercice 3 – Le piège d'une série très persistante

On compare

$$V_t = 0.97V_{t-1} + Z_t$$

avec une marche aléatoire

$$W_t = W_{t-1} + Z_t.$$

1) Pour un AR(1) stationnaire,

$$\rho_V(h) = 0.97^{|h|}.$$

On obtient approximativement :

| $h$ | 1 | 5 | 10 | 20 | 50 |
|---|---|---|---|---|---|
| $\rho_V(h)$ | 0.970 | 0.859 | 0.737 | 0.544 | 0.218 |

Même au retard 20, la corrélation théorique dépasse encore 0.5. L'ACF décroît donc très lentement, bien que le processus soit stationnaire.

2) La variance de $V_t$ est constante :

$$\mathrm{Var}(V_t) = \frac{\sigma^2}{1-0.97^2} = \frac{\sigma^2}{0.0591} \simeq 16.92\sigma^2.$$

Pour la marche aléatoire,

$$\mathrm{Var}(W_t) = t\sigma^2.$$

La distinction théorique est nette : variance constante d'un côté, variance croissante de l'autre. Mais sur un échantillon court, un AR(1) avec coefficient 0.97 produit de longues excursions et revient très lentement vers sa moyenne ; son chronogramme peut donc ressembler à celui d'une série intégrée.

3) La règle « ACF lente ⇒ différence » est dangereuse, car une ACF lente peut simplement correspondre à un AR stationnaire avec un coefficient proche de 1. Une décision raisonnable doit combiner le chronogramme, les transformations candidates et le comportement de l'ACF après transformation.

4) Posons $D_t = (1-B)V_t$. Comme

$$(1-0.97B)V_t = Z_t,$$

on obtient

$$\begin{aligned}(1-0.97B)D_t &= (1-0.97B)(1-B)V_t \\ &= (1-B)(1-0.97B)V_t \\ &= (1-B)Z_t \\ &= Z_t - Z_{t-1}.\end{aligned}$$

Donc la différence d'une série déjà stationnaire crée ici une composante MA avec coefficient $-1$. La série reste stationnaire, mais sa structure de dépendance devient moins naturelle et présente typiquement une forte autocorrélation négative à court retard.

**Remarque**

Une série très persistante et une série à racine unitaire peuvent être difficiles à distinguer sur un petit échantillon. C'est justement pourquoi on ne doit pas transformer uniquement à partir de l'impression visuelle d'une ACF « lente ».

## Exercice 4 – `UKgas` : faut-il une ou deux différences ?

On commence par charger et examiner la série.

```
data(UKgas)
x <- UKgas
frequency(x)
start(x)
end(x)
plot.ts(x, main = "UKgas", ylab = "consommation")
```

Les commandes donnent une fréquence 4, un début en 1960 et une fin en 1986. La série est donc trimestrielle.

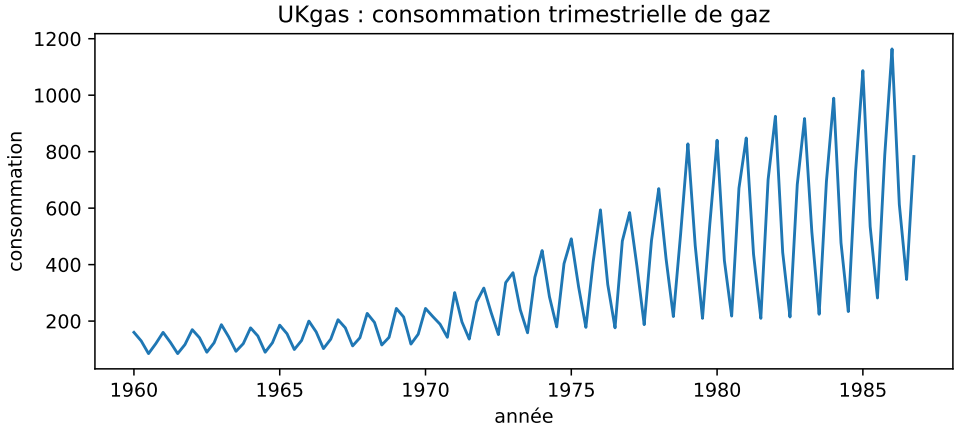



Figure 1 – Série brute `UKgas`.

1) Le chronogramme montre trois caractéristiques fortes :
   - une croissance du niveau sur la période ;
   - une saisonnalité trimestrielle très nette ;
   - une amplitude saisonnière qui augmente avec le niveau.

   Le dernier point rend le logarithme naturel : il transforme une saisonnalité approximativement multiplicative en une fluctuation plus proche d'une amplitude constante.

   On peut commencer par

```
lx <- log(x)
plot.ts(lx, main = "log(UKgas)", ylab = "log(consommation)")
```

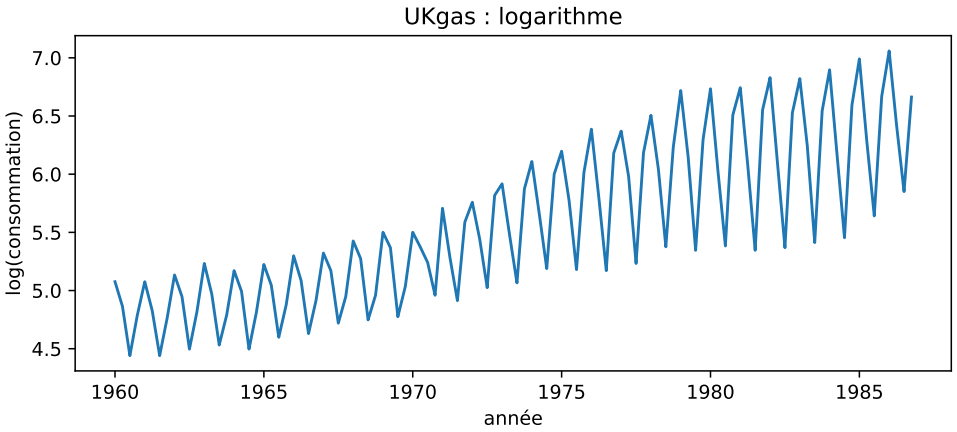



Figure 2 – Logarithme de `UKgas`.

Le logarithme stabilise mieux l'amplitude, mais la tendance et la saisonnalité restent visibles.

2) Plusieurs candidats raisonnables sont :

| | | |
|---|---|---|
| différence ordinaire | $(1-B)\log(X_t)$, | `diff(log(x))`, |
| différence saisonnière | $(1-B^4)\log(X_t)$, | `diff(log(x), lag = 4)`, |
| les deux | $(1-B)(1-B^4)\log(X_t)$, | `diff(diff(log(x), lag = 4))`. |

Le point important est de ne pas décider à l'avance que les deux différences sont nécessaires.

Le code complet peut être :

```
lx <- log(x)
d1 <- diff(lx)
d4 <- diff(lx, lag = 4)
d14 <- diff(d4)
```

3) La différence ordinaire seule ne convient pas bien. Elle enlève une partie de la croissance, mais le motif trimestriel reste extrêmement visible.

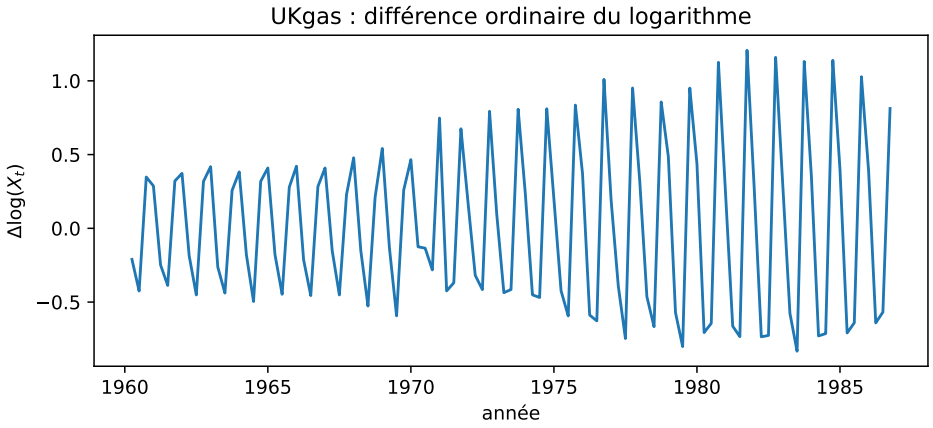



Figure 3 – Différence ordinaire du logarithme. La saisonnalité reste très forte.

La différence saisonnière du logarithme est beaucoup plus convaincante :

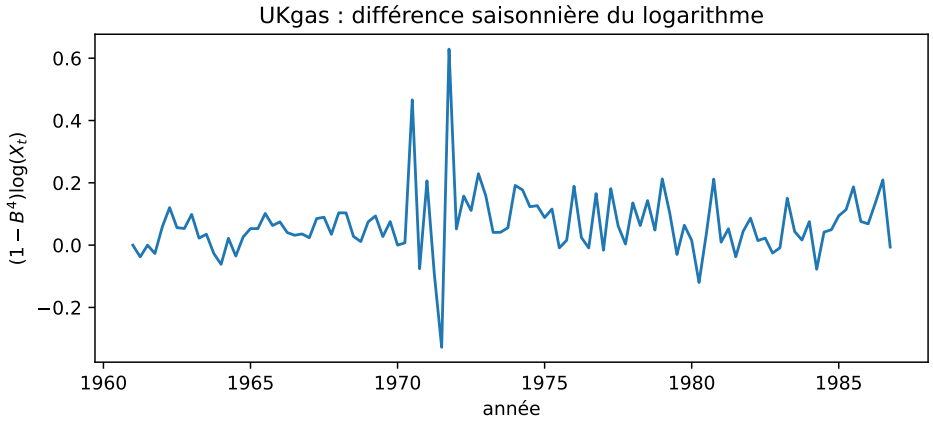



Figure 4 – Différence saisonnière du logarithme.

La série oscille maintenant autour d'un niveau approximativement stable, sans tendance durable évidente. Le choix minimal naturel est donc

$$\boxed{(1 - B^4)\log(X_t).}$$

Autrement dit, cette série suggère $D = 1$ mais pas nécessairement $d = 1$.

4) Les ACF permettent de préciser le diagnostic. On peut utiliser :

```
acf(lx, main = "ACF de log(UKgas)")
acf(d1, main = "ACF de diff(log(UKgas))")
acf(d4, main = "ACF de diff saisonniere")
```

L'ACF du logarithme présente une dépendance lente avec de forts retours aux retards saisonniers :

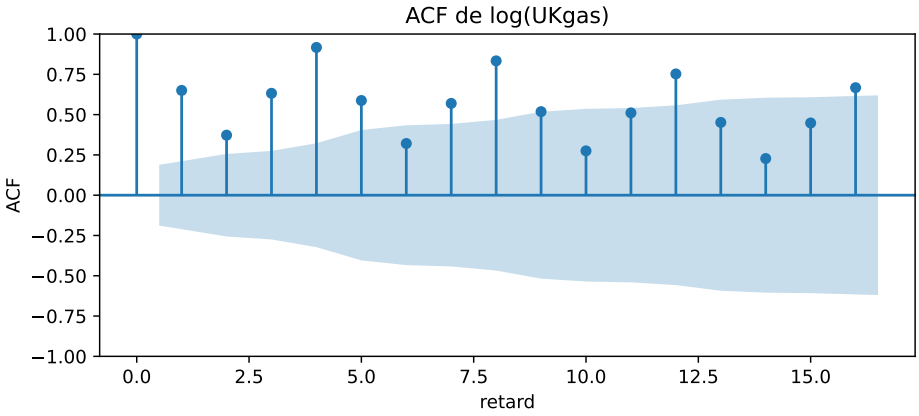



Figure 5 – ACF de log(`UKgas`).

Après différence ordinaire seulement, les retards multiples de 4 restent très fortement corrélés. Numériquement, l'autocorrélation au retard 4 est proche de 0.93 : la saisonnalité n'a clairement pas disparu.

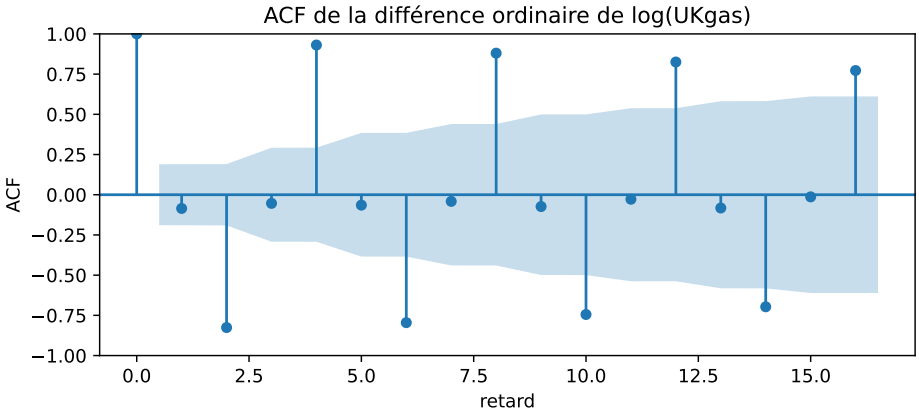



Figure 6 – ACF après différence ordinaire seulement.

Après différence saisonnière, la situation change nettement : les autocorrélations ne présentent plus la longue structure saisonnière du niveau. Il reste de la dépendance, mais elle est compatible avec l'idée d'une série stationnaire qu'un modèle ARMA/SARIMA peut ensuite décrire.

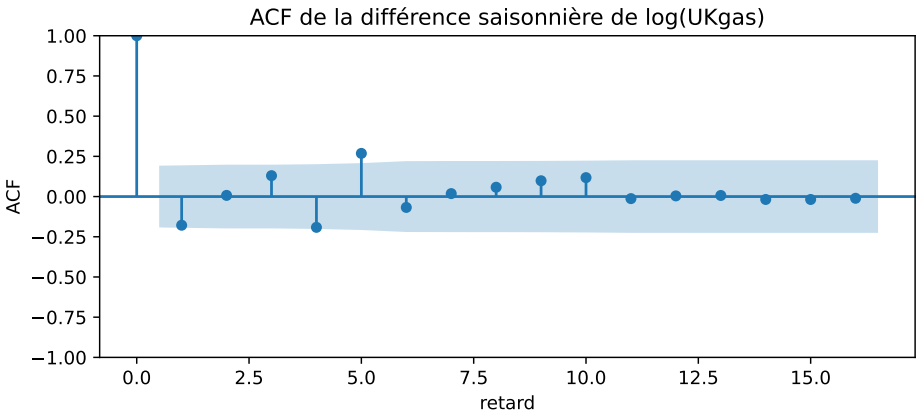



Figure 7 – ACF après différence saisonnière du logarithme.

5) Ajoutons maintenant volontairement une différence ordinaire :

```
d14 <- diff(d4)
plot.ts(d14, main = "difference saisonniere puis ordinaire")
acf(d14, main = "ACF apres deux differences")
```

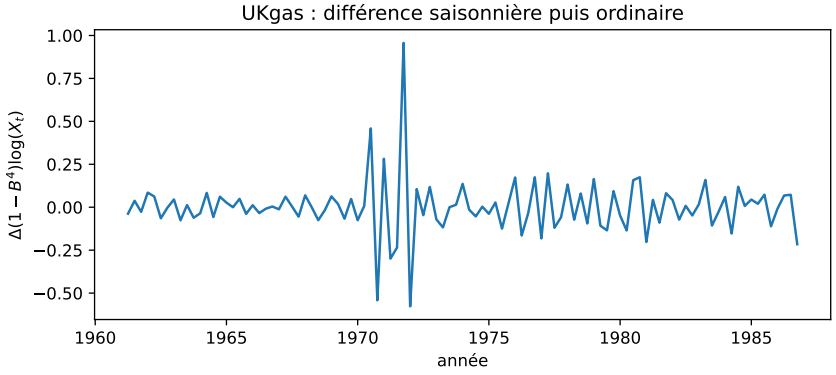



Figure 8 – Série après différence saisonnière puis ordinaire.

L'ACF fait apparaître une forte autocorrélation négative au retard 1, approximativement $-0.58$ dans cet échantillon :

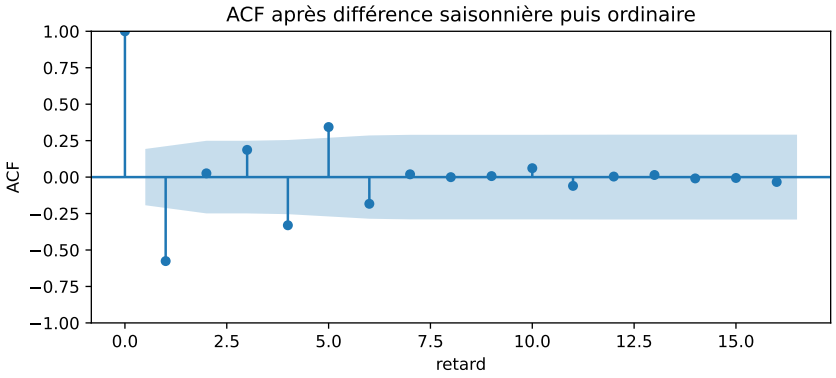



Figure 9 – ACF après l'ajout d'une différence ordinaire.

Cette forte alternance à très court terme est un signal typique d'une différence supplémentaire qui n'était pas nécessaire. On ne conclut pas à partir d'une seule barre, mais ici le contraste avec la série $d4$ est net : la différence ordinaire rend la dépendance plus compliquée au lieu de retirer une non-stationnarité visible.

6) La série $d4$ n'est pas un bruit blanc : certaines autocorrélations restent visibles. C'est précisément là que les modèles ARMA/SARIMA interviennent. La transformation sert à rendre la série à peu près stationnaire ; le modèle ARMA décrit ensuite la dépendance qui subsiste. On ne demande donc pas à la différenciation de supprimer toute autocorrélation.

**Idée à retenir**

`UKgas` fournit un bon contre-exemple à la recette « saisonnalité + tendance $\Rightarrow d = 1$ et $D = 1$ ». Après logarithme, une différence saisonnière seule est déjà très convaincante ; ajouter une différence ordinaire produit au contraire un signal de sur-différenciation.