Séries temporelles – TD/TP 5

Diagnostiquer une non-stationnarité et choisir les transformations

M1 Actuariat – Université Gustave Eiffel

Durée indicative : 1h30

Objectifs

Cette feuille ne reprend volontairement pas les exemples du cours. Le but est de décider soi-même *quelle* transformation est adaptée, puis de justifier ce choix. On rencontrera trois pièges : confondre tendance déterministe et tendance stochastique, confondre forte persistance et racine unitaire, et différencier davantage que nécessaire.

Partie TD

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

Soit (Z_t) un bruit blanc centré de variance σ^2. On considère d'une part un processus stationnaire (U_t) défini par

$$U_t = 0.7U_{t-1} + Z_t,$$

et d'autre part les deux séries

$$X_t = 100 + 0.5t + U_t, \qquad Y_0 = 0, \qquad Y_t = Y_{t-1} + 0.5 + Z_t.$$

Sur un échantillon fini, les deux chronogrammes peuvent montrer une croissance assez régulière.

1) Calculer $\mathbb{E}(X_t)$ et $\mathrm{Var}(X_t)$. Calculer ensuite $\mathbb{E}(Y_t)$ et $\mathrm{Var}(Y_t)$. Quelle différence essentielle apparaît ?

2) Un premier analyste propose de retirer une droite. Montrer que

$$X_t - (100 + 0.5t)$$

est stationnaire. Que devient au contraire $Y_t - 0.5t$?

3) Un second analyste propose de différencier les deux séries. Montrer que ΔY_t est stationnaire. Que vaut ΔX_t ?

4) En utilisant $(1 - 0.7B)U_t = Z_t$, montrer que la série $\Delta X_t - 0.5$ vérifie

$$(1 - 0.7B)(\Delta X_t - 0.5) = (1 - B)Z_t.$$

Pourquoi la différence rend-elle ici le modèle plus compliqué que le simple retrait de la tendance ?

5) Vous ne connaissez pas le mécanisme qui a produit une série observée. Quels graphiques ou transformations compareriez-vous avant de décider entre retrait d'une tendance et différenciation ?

Idée à retenir

Deux séries peuvent avoir un chronogramme très semblable sans avoir la même non-stationnarité. Le choix de la transformation doit être un diagnostic, pas un réflexe.

Exercice 2 – Une dynamique trimestrielle cachée

Le nombre trimestriel de sinistres d'un portefeuille est modélisé par

$$X_t - X_{t-4} = 0.6(X_{t-1} - X_{t-5}) + Z_t + 0.3Z_{t-1},$$

où (Z_t) est un bruit blanc.

1) Poser $U_t = X_t - X_{t-4}$. Réécrire le modèle uniquement en fonction de (U_t) et identifier le type de modèle suivi par cette série transformée.
2) Vérifier les conditions de causalité et d'inversibilité de ce modèle.
3) Réécrire l'équation initiale à l'aide de l'opérateur retard B.
4) Donner une notation SARIMA naturelle pour (X_t), avec période saisonnière 4.
5) Interpréter les paramètres d et D obtenus : quelle différence est nécessaire ici ? laquelle ne l'est pas imposée par le modèle ?
6) Un étudiant applique seulement $\Delta X_t = X_t - X_{t-1}$. Quel type de structure chercheriez-vous dans l'ACF de la série obtenue pour savoir qu'une saisonnalité trimestrielle reste présente ?

Exercice 3 – Le piège d'une série très persistante

On compare deux modèles :

$$V_t = 0.97V_{t-1} + Z_t, \qquad W_t = W_{t-1} + Z_t,$$

où (V_t) est supposé démarré dans son régime stationnaire et $W_0 = 0$.

1) Donner l'ACF théorique de (V_t). Calculer approximativement $\rho_V(h)$ pour $h = 1, 5, 10, 20, 50$.
2) Comparer $\mathrm{Var}(V_t)$ et $\mathrm{Var}(W_t)$. Pourquoi un chronogramme court de (V_t) peut-il pourtant ressembler à celui d'une série non stationnaire ?
3) Un analyste voit une ACF empirique qui décroît très lentement et décide immédiatement de différencier. Pourquoi cette règle est-elle dangereuse ?
4) Poser $D_t = \Delta V_t$. Montrer que

$$(1 - 0.97B)D_t = (1 - B)Z_t.$$

Que nous apprend cette équation sur l'effet d'une différence appliquée à une série qui était déjà stationnaire ?

Partie TP – Étude autonome d'une nouvelle série

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

La série `UKgas`, fournie avec `R`, contient une consommation trimestrielle de gaz au Royaume-Uni. Aucune transformation n'est imposée : c'est à vous de construire et défendre votre diagnostic.

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

Décodage du code R

- `frequency(x)` donne le nombre d'observations par unité de temps. Il permet donc d'identifier la période saisonnière naturelle.
- `start(x)` et `end(x)` donnent les dates de début et de fin de la série.
- Le chronogramme doit être analysé avant toute transformation : niveau, tendance, saisonnalité et évolution de l'amplitude.

Idée à retenir

Dans cette partie, on ne vous demande pas de retrouver de mémoire la syntaxe de `R`. Le code ci-dessous met plusieurs transformations possibles à votre disposition. **Le fait qu'une transformation soit proposée dans le code ne signifie pas qu'elle soit adaptée à la série : c'est précisément ce que vous devez décider à partir des graphiques.**

Boîte à outils `R` pour les transformations

Commencez par stocker la période saisonnière :

```
s <- frequency(x)
```

Voici les commandes dont vous pouvez avoir besoin. Vous n'êtes pas obligé de retenir toutes les transformations : elles sont données pour que le temps du TP soit consacré au *diagnostic*, et non à la recherche de syntaxe.

```
# Logarithme seul
x.log <- log(x)

# Difference ordinaire seule
x.d1 <- diff(x)

# Difference saisonniere seule
x.ds <- diff(x, lag = s)

# Logarithme puis difference ordinaire
x.log.d1 <- diff(log(x))

# Logarithme puis difference saisonniere
x.log.ds <- diff(log(x), lag = s)

# Logarithme + difference saisonniere + difference ordinaire
x.log.ds.d1 <- diff(diff(log(x), lag = s))
```

Décodage du code R

- `log(x)` applique le logarithme observation par observation.
- `diff(x)` calcule $X_t - X_{t-1}$, donc applique le filtre $1 - B$.
- `diff(x, lag = s)` calcule $X_t - X_{t-s}$, donc applique le filtre $1 - B^s$.
- On peut imbriquer les commandes : par exemple `diff(log(x), lag = s)` signifie « logarithme, puis différence saisonnière ».
- Les six objets ci-dessus sont seulement des *candidats*. Votre travail est de déterminer lesquels ont du

Décodage du code R (suite)

sens pour `UKgas`.

Pour visualiser un candidat, utilisez simplement `plot.ts(...)`. Par exemple :

```
plot.ts(x.log, main = "log(UKgas)")
plot.ts(x.log.d1, main = "log + difference ordinaire")
plot.ts(x.log.ds, main = "log + difference saisonniere")
plot.ts(x.log.ds.d1, main = "log + deux differences")
```

Vous pouvez remplacer les objets affichés par les candidats que vous souhaitez réellement comparer.

Pour comparer les ACF, la syntaxe est la même :

```
acf(x.log, main = "ACF : logarithme")
acf(x.log.d1, main = "ACF : log + diff. ordinaire")
acf(x.log.ds, main = "ACF : log + diff. saisonniere")
acf(x.log.ds.d1, main = "ACF : log + deux differences")
```

1) Décrire précisément le chronogramme. Quelle est la période saisonnière ? L'amplitude des oscillations semble-t-elle constante ? Une transformation logarithmique vous paraît-elle justifiée ?
2) Parmi les transformations proposées dans la boîte à outils, en sélectionner au moins trois qui vous paraissent plausibles. Pour chacune, écrire le filtre en B correspondant. *Le code est fourni : ce que vous devez choisir et justifier, c'est la transformation.*
3) Comparer les chronogrammes obtenus. Quelle est la transformation *la plus simple* qui vous semble produire une série à peu près stationnaire ? Justifier à partir de ce que vous voyez, et non du nom de la transformation.
4) Comparer les ACF des principales séries candidates à l'aide du code fourni. Votre choix de la question précédente est-il confirmé ? Quels retards sont particulièrement informatifs ?
5) Prenez votre transformation retenue et stockez-la sous un nom commun en remplaçant les points de suspension :

   ```
   x.retenue <- ...
   x.plus <- diff(x.retenue)
   plot.ts(x.plus, main = "une difference ordinaire en plus")
   acf(x.plus, main = "ACF apres une difference en plus")
   ```

   Que se passe-t-il dans le chronogramme et surtout dans l'ACF ? Voyez-vous un signe de sur-différenciation ?
6) À ce stade, sans estimer encore de modèle, la série transformée ressemble-t-elle à un bruit blanc ? Si non, quel serait le rôle d'un modèle ARMA/SARIMA dans la suite de l'analyse ?

Idée à retenir

Ici, le but n'est pas de retrouver une recette donnée à l'avance. Une bonne réponse doit comparer plusieurs transformations et retenir la moins agressive qui enlève les structures non stationnaires visibles.